

DM

Effectiveness of Sentiment and Emotion Detection to Unveil Customer Preferences in Portuguese Restaurants Insights for Restaurant owners

MASTER DISSERTATION

Marcos Diogo Santos Silva

MASTER IN INFORMATICS ENGINEERING



UNIVERSIDADE da MADEIRA

A Nossa Universidade

www.uma.pt

January | 2025

Effectiveness of Sentiment and Emotion Detection to Unveil Customer Preferences in Portuguese Restaurants

Insights for Restaurant owners

MASTER DISSERTATION

Marcos Diogo Santos Silva

MASTER IN INFORMATICS ENGINEERING

SUPERVISOR

Fernando Manuel Rosmaninho Morgado Ferrão Dias

CO-SUPERVISOR

Sheikh Shanawaz Mostafa



FACULTY OF EXACT SCIENCES AND ENGINEERING

MASTER'S DEGREE IN INFORMATICS ENGINEERING

Marcos Diogo Santos Silva

Supervised by:

Fernando Manuel Rosmaninho Morgado Ferrão Dias

Sheikh Shanawaz Mostafa

Composition of the public examination jury:

Karolina Baras (Assistant Professor), President

Filipe Magno de Gouveia Quintal (Assistant Professor), Member

Sheikh Shanawaz Mostafa (Researcher), Member

Wednesday 19th February, 2025

Resumo

O trabalho apresentado nesta tese investiga o impacto das avaliações online de uma empresa na indústria de serviços alimentares em Portugal, enfatizando a importância das plataformas de avaliação online na formação do comportamento do consumidor e na melhoria dos serviços de restauração. A pesquisa utiliza dados do Zomato, uma plataforma de descoberta de restaurantes e entrega de comida, para analisar as avaliações dos clientes e extrair informações valiosas.

O estudo começa com uma visão geral das metodologias usadas na recolha, preparação e processamento de dados, destacando a eficácia do uso de uma base de dados NoSQL, especificamente MongoDB. É realizada Análise Exploratória de Dados para descobrir padrões e tendências, focando-se no tamanho das avaliações, deteção de idioma, tradução e deteção de género.

A deteção de emoções é explorada com uma técnica personalizada que utiliza o léxico EMO-TAIX.PT, revelando o impacto significativo do tom emocional nas classificações das avaliações. A seleção de características é realizada usando os métodos de correlação de Spearman e mRMR, identificando as características relacionadas com o sentimento como os preditores mais fortes das classificações das avaliações.

Uma contribuição essencial desta pesquisa é o desenvolvimento do ReviewSight, uma plataforma web projetada para fornecer aos proprietários de restaurantes informações acionáveis. O ReviewSight integra técnicas de processamento de dados e processamento de linguagem natural numa interface fácil de usar. A plataforma gera relatórios detalhados incluindo visualizações como gráficos de linhas, gráficos de pizza e nuvens de palavras, bem como resumos de texto e identificação de tópicos. Estes relatórios ajudam os proprietários de restaurantes a entender o sentimento dos clientes e assim tomar decisões baseadas em dados adaptados às suas necessidades.

Finalmente, é discutida a implementação do ReviewSight numa Máquina Virtual Linux na Google Cloud Platform, mostrando a robustez e escalabilidade dos serviços de cloud para soluções empresariais.

Keywords: Avaliações Online · Indústria de Serviços Alimentares · Comportamento do Consumidor · Processamento de Linguagem Natural · Análise de Sentimentos · Aprendizagem de Máquina

Abstract

The work presented in this thesis investigates the impact of online reviews of a company on the food service industry in Portugal, emphasizing the importance of online review platforms in shaping consumer behavior and improving restaurant services. The research utilizes data from Zomato, a prominent restaurant discovery and food delivery platform, to analyze customer reviews and extract valuable insights.

The study begins with an overview of the methodologies used for data collection, preparation, and processing, highlighting the effectiveness of using a NoSQL database, specifically MongoDB. Exploratory Data Analysis is conducted to uncover patterns and trends, focusing on review length, language detection, translation, and gender detection.

Emotion detection is explored using a customized technique that employs the EMOTAIX.PT lexicon, revealing the significant impact of emotional tone on review ratings. Feature selection is performed using Spearman correlation and mRMR methods, identifying sentiment-related features as the strongest predictors of review ratings.

An essential contribution of this research is the development of ReviewSight, a web platform designed to provide restaurant owners with actionable insights. ReviewSight integrates data processing techniques and natural language processing tools into a user-friendly interface. The platform generates detailed reports that include visualizations such as line charts, pie charts, and word clouds, as well as text summaries and topic detection. These reports help restaurant owners understand customer sentiment and make data-driven decisions tailored to their needs.

Finally, the deployment of ReviewSight on a Linux Virtual Machine on the Google Cloud Platform is discussed, showcasing the robustness and scalability of cloud services for enterprise solutions.

Keywords: Online Reviews · Food Service Industry · Consumer Behavior · Natural Language Processing · Sentiment Analysis · Machine Learning

Acknowledgments

The completion of this thesis would not have been possible without the support and guidance of several individuals who contributed to my academic and personal development. First, I extend my special thanks to my advisor, Professor Doctor Morgado Dias, and my supervisor, Doctor Sheikh Mostafa, for their invaluable support, experience, and guidance throughout this process and towards this project's completion. I am also grateful to my team members, Doctor Fabio Mendonça, Alexandre Branco, and Daniel Olival, as well as the members of the Zomato team for their help and support. I want to especially thank my girlfriend, Catarina Jeppesen Cruz, for her love and support through times of struggle, and Cristina Gouveia, who guided me to success even when I could not find it on my own. Without their persistence, I would not be where I am today. Finally, I want to thank my family and friends for their love and support.

Table of Contents

List of Figures	ix
List of Tables	xii
1 Introduction	1
1.1 Motivation	3
1.2 Research Questions	5
1.3 Research Objectives	5
1.4 Thesis Structure	7
2 Data Collection and Database Preparation	10
2.1 Introduction	10
2.2 State of the Art (SQL)	10
2.3 State of the Art (NoSQL)	11
2.4 Data Source	12
2.5 Data Cleaning and Pre-processing Methods	13
2.6 SQL and NoSQL Implementation	14
2.6.1 SQL	14
2.6.2 NoSQL	15
2.7 Selection Of NoSQL Database and Design	15
2.8 Conclusion	17
3 Exploratory Data Analysis	18
3.1 Introduction	18
3.2 Comparative Analysis of Sentiment Analysis Techniques	18
3.3 State of the Art	19
3.4 Rating	21
3.5 Stars rating's compatibility with a classification cluster	22
3.6 Review Length	23
3.6.1 Method: number of words	24
3.6.2 Results	24
3.7 Language Detection	24

3.7.1	Method.....	24
3.7.2	Results	25
3.8	Translation	25
3.8.1	Method.....	25
3.8.2	Results	26
3.9	Stop Word Detection/Removal	26
3.9.1	Method.....	27
3.9.2	Results	27
3.10	Gender Detection	28
3.10.1	Method.....	28
3.10.2	Results	29
3.11	Timetable Separation	30
3.11.1	Method.....	30
3.11.2	Results	30
3.12	Conclusion	31
4	Emotion Detection	32
4.1	Introduction	32
4.2	State of the Art.....	32
4.3	Emotion Detection Algorithms	33
4.3.1	Method.....	33
4.3.2	Levenshtein Distance.....	35
4.3.3	Minimum Equal Initial Characters	37
4.3.4	Stemming	37
4.3.5	Synonyms	38
4.4	SentiStrength Method.....	38
4.5	Relationship of emotion and rating	39
4.6	Results	40
4.7	Conclusion	41
5	Feature Selection	44
5.1	Introduction	44
5.2	State of the art	44
5.3	Features	46
5.3.1	Correlation	46

5.3.2	mRMR Method	47
5.3.3	Borda Count Method	48
5.4	Results	48
5.5	Conclusion	50
6	Unsupervised Topic Detection Enhanced by Human Labeling	52
6.1	Introduction	52
6.2	State of the Art	52
6.3	Unsupervised ML Topic Detection Enhanced by Human Annotation ...	53
6.3.1	Unsupervised Topic Detection Method	55
6.3.2	Topic Labeling - Manual Labeling with Survey	59
6.4	Conclusion	62
7	Visual and Textual presentation of Data according to User Choice	63
7.1	Introduction	63
7.2	Demographics and Habits	63
7.3	Data Visualization	64
7.3.1	A, B Comparison	64
7.3.2	Concept Perception Rating:	65
7.3.3	Combined vs Separated Charts	66
7.4	Preferred Text (ASTA vs LLM)	67
7.5	Comparative Analysis of Linguistic Features in ASTA and GPT	69
7.5.1	Vocabulary Features	69
7.5.2	Part-of-Speech	69
7.5.3	Entity	70
7.5.4	Perplexity	71
7.5.5	Dependency Parsing	71
7.6	Conclusion	72
8	System design for Reporting	74
8.1	Introduction	74
8.2	System Design for report generation (ReviewSight)	76
8.3	Authentication Service	76
8.3.1	User	76
8.3.2	Administrator	76

8.3.3 Password Hasher	77
8.4 Database	77
8.4.1 Data Denormalization	78
8.4.2 Data Cleaning	78
8.4.3 Populating the MongoDB Database	79
8.5 Linguistic Service	79
8.5.1 ASTA	79
8.5.2 LLM	79
8.6 Visualization Service	80
8.6.1 Map	80
8.6.2 Line chart using moving average	81
8.6.3 Tables	82
8.6.4 Pie Charts	83
8.6.5 Radar Plots	83
8.6.6 Bar Plots	84
8.6.7 Word Clouds	85
8.7 ReviewSight Structure	87
8.7.1 Topic Based Tips	90
8.7.2 Branding	90
8.8 Exportability Service	91
8.8.1 PDF generation	91
8.8.2 XLSX Generation	92
8.8.3 Downloadable Individual Images	92
8.9 Conclusion	92
9 Cloud Implementation	94
9.1 Introduction	94
9.2 Implementation	94
9.3 Conclusion	95
10 Discussion and Conclusion	97
10.1 Discussion	97
10.2 Conclusion	98
Appendix	100

A.1	Provided Files	100
A.2	Translation	102
A.3	All Features combined.....	106
A.4	Topics	107
A.5	Opinion of UMa's General Data Protection Regulation	116
A.6	Graphs Comparison Survey (English)	118
A.7	Graphs Comparison Survey (Portuguese)	124
A.8	ASTA Versus Chat GPT Survey (English)	130
A.9	ASTA Versus Chat GPT Survey (Portuguese)	134
A.10	Topic Labeling Survey.....	138
A.11	ReviewSight PDF Structure	151
A.12	PDF Report	155
A.13	EMOTAIX.PT lexicon	159
A.14	PositiveWords.csv	179
A.15	NeutralWords.csv	183
A.16	NegativeWords.csv	188
	References	193

List of Figures

1	Overview of the thesis structure and data processing flow.	9
2	Distribution of reviews according to their creation date.	12
3	a) An example of an SQL query is displayed in the phpMyAdmin interface via XAMPP, retrieving comments and ratings from all reviews associated with the Restaurant identified by ID 18347750. b) Example of a MongoDB query, visualized in MongoDB Compass, retrieves comments and ratings from all reviews associated with the Restaurant identified by ID 18347750.	16
4	a) Distribution of Zomato’s Review Ratings. b) Distribution of Zomato’s Restaurant Ratings	22
5	a): Distribution of restaurant ratings across all reviews using initial naive clustering assumption. Negative: (1.0, 1.5, and 2.0), Neutral: (2.5, 3.0, and 3.5), Positive: (4.0, 4.5, and 5.0). b): Distribution of restaurant ratings across all reviews using a revised clustering approach. Negative: (1.0, 1.5, 2.0, 2.5, and 3.0), Neutral: (4.0 and 4.5), Positive: (5.0).	23
6	Illustration of all the categories in the EMOTAIX lexicon and how they are organized. . .	35
7	Distribution of SentiStrength scores across all reviews. a) Positive (in blue) and negative (in red) scores for English reviews. b) Positive (in blue) and negative (in red) scores for Portuguese reviews.	40
8	3D scatter plot illustrating the correlation between review ratings and word sentiment. The color scheme represents different ratings, and the axes labeled as 'Positive', 'Neutral', and 'Negative' denote the relative number of words in each respective category. The plot is divided into four quadrants. The top-left quadrant provides a 3D perspective, while the remaining three quadrants each showcase a unique 2D combination of 'Negative', 'Neutral', and 'Positive' sentiments.	41
9	Displayed from left to right are three individual 3D scatter plots, each representing the centroids of clusters when using 2, 3, and 5 clusters, respectively. The color scheme corresponds to different ratings, while the axes denote the relative number of 'Negative', 'Neutral', and 'Positive' words.	42

10	Elbow plot using SSE for positive, neutral, negative, and unknown words across all reviews, derived from the EMOTAIX.PT lexicon with a maximum Levenshtein distance of one and a minimum number of equal initial characters of six across different numbers of clusters. (Red circle: Highest change in slope. Green circle: The second highest change in slope is the Purple circle: Third highest change in slope).	42
11	Distribution of the number of positive, neutral, negative, and unknown words across all reviews using EMOTAIX.PT lexicon with maximum Levenshtein distance set to one and minimum amount of equal initial characters set to six. a) Positive, neutral, and negative words box plot. b) Unknown words box plot.	43
12	Spearman correlation matrix for all features excluding those determined to be irrelevant in earlier processes and those with a high amount of Nulls.	47
13	Diagram of Unsupervised Topic Detection Enhanced by Human Labeling.	54
14	a) Overall perplexity across the number of topics. b) Perplexity per topic across the number of topics.	58
15	Example of how the Topic labeling questions were presented in the topic labeling survey	61
16	Example of a comparison between radar charts question on the Data visualization part of the survey.	65
17	Example of a question of concept perception of the data visualization survey.	66
18	Example of a question of the Text Generation section of the survey where the LLM and the ASTA are compared.	67
19	a) Part-of-Speech comparison between GPT.3.5 and ASTA paragraphs. b) Entity comparison between GPT.3.5 and ASTA paragraphs.	70
20	a) Perplexity comparison. b) Dependency Parsing comparison.	72
21	ReviewSight platform in action.	74
22	High-level diagram of the ReviewSight Platform's workflow.	75
23	a) Example Map on the ReviewSight report using OpenStreetMap. b) Example Map on the PDF report using OpenStreetMap.	81
24	a) Example of rating over time using spline. b) Example of rating over time using spline and moving average.	82
25	Table including Review Distribution Across Genders for All Reviews and Last 36 Months.	82
26	Example of Pie chart depicting Distribution of Emotion in Reviews.	83

27	a) : Radar plot illustrating a comparative analysis of emotions associated with the owner's restaurant, in blue, and those prevalent in nearby restaurants, in orange, segregated by gender and classified into three categories: Positive (M), Neutral (M), Negative (M) for males and Positive (W), Neutral (W), Negative (W) for females. b) : Radar chart illustrating the comparison of emotional responses, segregated by gender, associated with the owner's restaurant (depicted in blue). This is contrasted against the emotional responses for a hypothetical ideal best restaurant (depicted in green) and a hypothetical ideal worst restaurant (depicted in red).....	84
28	Comparative bar plots illustrating the gender distribution in two different Figures, a and b. The blue bars represent the number of men and women in each scenario. a) represents the number of men and women in the restaurant's entire existence. b) represents the same distribution in a user-specified period.	85
29	a) Word cloud representing all-time reviews. b) Word cloud representing a user-selected timeslot.	86
30	Excerpt of the Report's initial page with watermark and footer logo.	91
31	Illustration of GCP's VM Management Interface in Action: Running the Project-Specific Instance.	95
32	Ubuntu VM instance, streamed via NoMachine, showcasing the ReviewSight in operation on both console and browser, complemented by MongoDB Compass and Task Manager.	96
33	a) : Full distribution of review length. b) : Cropped (0 - 150 words) review length distribution.	102
34	Distribution of user gender across all reviews.	104
35	Visual representation of the Levenshtein distance method.	104

List of Tables

1	Overview of SA Studies: Supervised Learning (SL), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), Bag of Words (BoW)	19
2	Example of First Names of Users and respective attributed genders.....	29
3	Examples of original vs. stemmed words.....	37
4	Ranking of features based on mRMR and Spearman correlation methods, along with their Borda points. The final Borda score column combines these rankings to provide an overall feature importance ranking from best to worst.	50
5	Partial list of topics along with the nine initial words for each computed by sklearn's LDA	59
6	Partial list of topics along with the four proposed Labels presented in the survey.	60
7	List of all topics and the respective winner labels given by the topic labeling survey	61
8	Comparison of sentence counts and similarity scores for different sections	69
9	First ten words from the PositiveWords.csv, NeutralWords.csv, and NegativeWords.csv files used in Word Cloud generation.....	87
10	Table presenting a partial list of topics along with three associated tips for each topic. ..	90
11	Attributes of Zomato's Review Data.	100
12	Attributes of Zomato's Restaurant Data.....	101
13	Example of review text translated from original Portuguese to English using Google translate API.	102
14	Example of review text translated from original English to Portuguese using Google translate API.	103
15	All Features.....	106
16	Full list of all topics and respective associated words presented in the topic labeling survey.	107
17	List of all topics and the respective winner labels given by the topic labeling survey.	108
18	Full list of topics and respective suggested labels presented in the Topic labeling survey .	109

List of Codes

1	SQL code snippet to retrieve comments and ratings from all reviews associated with the restaurant identified by ID 18347750.	15
2	MongoDB code snippet retrieves comments and ratings from all reviews related to the restaurant identified by ID 18347750.	15
3	Code snippet illustrating the construction of a denormalized collection in MongoDB, derived from the 'review' and 'restaurant' collections. This collection is intended to serve as the primary database for executing all future operations. ...	17
4	Tips.csv file. This includes the list tips associated with each topic	115
5	Code snippet including 'ExamplePrompt', which is a sample prompt sent to ChatGPT 3.5, and 'ExampleAnswer', which holds the corresponding response from the model.	150
6	Snippet of code showing variables containing the information prompted to ChatGPT 3.5 to generate the nearby restaurants introduction portion of the report. NearbyRestaurantsIntroGPTDict contains, in the form of a Python dictionary, the names of the variables and an explanation of what they contain. NearbyRestaurantsIntroPrompt contains the task the LLM is supposed to perform, very carefully explained along with the maximum number of words, what to do in the case that some items in the dictionary are not available, and more.	151
7	CSV of the EMOTAIX.PT lexicon split into 3, 9, 18, and 50 classes.	179
8	PositiveWords.csv file context. This includes all the the positive words that were used to filter the Negative Word Clouds	183
9	NeutralWords.csv file context. This includes all the the Neutral words that were used to filter the both Positive and Negative Word Clouds	188
10	NegativeWords.csv file context. This includes all the the Neutral words that were used to filter the both Positive Word Clouds	192

Acronyms

ACID Atomicity, Consistency, Isolation, Durability

API Application Programming Interface

ASTA Adaptive Stochastic Textual Assembler

BASE Basically Available, Soft state, Eventual consistency

BC Borda Count

BNS Bi-Normal Separation

BoW Bag of Words

CART Classification and Regression Trees

CNN Convolutional Neural Network

COVID-19 Coronavirus Disease 2019

CSV Comma Separated Values

DB Database

DL Deep Learning

DT Decision Tree

EDA Exploratory Data Analysis

FCA Formal Concept Analysis

GB Gigabyte

GCP Google Cloud Platform

GPT Generative Pre-trained Transformer

GPU Graphical Processing Unit

GSS Generalized Sequential Sample

GUI Graphic User Interface

IP Internet Protocol

IaaS Infrastructure as a Service

LDA Latent Dirichlet Allocation

LLM Large Language Model

LR Logistic Regression

LSI Latent Semantic Indexing

LSTM Long Short-Term Memory

MB Megabyte

ML Machine Learning

MLP Multi Layer Perceptron

MRF Markov Random Field

MaxEnt Maximum Entropy

N/A Not Applicable

NB Naive Bayes

NER Named Entity Recognition

NLP Natural Language Processing

NLTK Natural Language Toolkit

NN Neural Network

NoSQL Not Only SQL

OS Operating System

PDF Portable Document Format

PNG Portable Network Graphic

PaaS Platform as a Service

RBF Radial Basis Function

RF Random Forest

RNN Recurrent Neural Network

SA Sentiment Analysis

SL Supervised Learning

SQL Structured Query Language

SSD Solid State Drive

SSE Sum of Squared Errors

SVM Support Vector Machine

SWOT Strengths, Weaknesses, Opportunities, and Threats

TF-IDF Term Frequency-Inverse Document Frequency

TPU Tensor Processing Unit

VM Virtual Machine

XAMPP Cross-Platform, Apache, MySQL, PHP, and Perl

XLSX Microsoft Excel Open XML Format Spreadsheet file

YAML Yet Another Markup Language

k-NN k-Nearest Neighbors

mRMR minimum Redundancy Maximum Relevance

1 Introduction

In 2018, Portugal’s accommodation, food service, and similar segments comprised 113,191 companies with a total revenue of \$18 billion. This impressive figure highlights the importance of the food service industry in Portugal’s economy. According to Eurostat, in 2018, Portuguese households allocated 9.2% of their total expenses towards food services. This corresponds to a 7% excess over the European Union’s mean [1], emphasizing Portuguese households’ high regard for food services and their significant role in daily life.

Furthermore, the growth and increased accessibility of web-based review platforms have resulted in online restaurant reviews significantly impacting consumer purchasing decisions [2]. Online reviews can benefit consumers and restaurants by reducing uncertainty in consumption decisions and increasing revenue for restaurants [3]. This highlights the importance of online reviews in the food service industry and their potential to shape consumer behavior.

Reviewing services for experiences and products is becoming increasingly important. These services provide transparency to customers and make products, services, or companies more competitive [4]. Evaluation systems are available in various industries, such as entertainment, food services, transportation, housing, etc. They provide customers with the best quality of service while allowing service or product providers to improve their shortcomings and improve over time. The food service industry is one of the industries that utilizes reviews to provide value to both customers and companies. Food services can benefit from ratings and comments to create a more competitive and fair service for all involved parties [5].

A prevalent approach to evaluating products and services involves the use of a rating system. Such systems typically incorporate a scale that encompasses diverse values, enabling customers to express their experience with the product or service. For example, the scale might extend from 1 to 5 in increments of 0.5, culminating in a total of 9 potential scores. Rating systems most frequently employ a scale of 5 to 9 points [6].

In addition to assigning a numerical rating, the option is often available for reviewers to supplement their rating with a comment. This comment, which provides a more comprehensive account of their experience and elucidates their rating, has proven to be highly beneficial in augmenting the value of the review for both service providers and prospective customers [7]. The significance of delivering detailed feedback in conjunction with numerical ratings is therefore emphasized. Through

the provision of a numerical rating and an elaborate comment, reviewers are able to assist service providers in enhancing their offerings and enable potential customers to make well-informed decisions regarding the utilization of the product or service. Since ratings are number-based, analyzing how they vary over time and performing advanced mathematical operations to extract relevant information is possible. This can provide valuable insights into how customers perceive the product or service and how their opinions change over time [8].

On the other hand, analyzing comments is a complex task due to the intricacies and subjectivity of human language. The subjectivity of language further complicates comment analysis. Each individual may use language differently, influenced by their personal experiences, emotions, and perspectives. There are numerous obstacles to accurately interpreting comments. For instance, certain industries may have specific slang words and expressions that are not well-documented or categorized. Additionally, the use of metaphor, irony, hyperbole, and other types of figurative language combined with the presence of grammatical and syntax errors can all make it challenging to understand the meaning of a comment accurately. Furthermore, cultural differences and regional variations in language can also pose challenges to accurately analyzing comments [9].

Despite these challenges, comment analysis can provide valuable insights into customer opinions and experiences. By using advanced Natural Language Processing (NLP) techniques, it is possible to extract meaningful information from comments and use it to improve products and services [9].

Insights can be derived from a vast volume of unstructured text data by utilizing advanced data analysis tools. These techniques also enable the execution of Sentiment Analysis (SA), topic modeling, and entity recognition, among others [10]. By employing these techniques, it becomes feasible to discern patterns and trends within a substantial quantity of unstructured text data. This facilitates the extraction of valuable insights pertaining to customer opinions, market trends, and other crucial information. Utilizing these insights can lead to the enhancement of products and services and enable the making of decisions driven by data.

Although NLP has predominantly been applied to English, a wealth of data exists in other languages such as Portuguese, Indian, Chinese, and more. The application of NLP techniques to analyze reviews in these languages can significantly benefit the industries of the countries where they are spoken [11]. Consider Portuguese as an example. It is the ninth most spoken language worldwide, with 234 million speakers [12]. As such, this research will concentrate on both English

and Portuguese and employ NLP techniques that analyze reviews in both Portuguese and English simultaneously by using multiple methods in both languages.

This research has the potential to make a substantial contribution to the field of NLP by demonstrating the value of analyzing reviews in both English and Portuguese languages. By extending NLP techniques to other languages, it is possible to extract information from a broader range of data and improve products and services in more countries globally.

Portuguese restaurants use a range of mobile and online services available for search and reservation platforms to attract new customers and increase customer satisfaction [13]. These platforms provide customers with an easy and convenient way to find and book restaurants and can help restaurants reach a wider audience. As such, it is essential to understand the needs and preferences of Portuguese restaurant customers in order to improve the industry. By analyzing customer reviews and feedback, it is possible to gain valuable information about what customers value most when choosing a restaurant.

Zomato is a restaurant discovery and food delivery platform in multiple countries, including Portugal. It provides a wealth of restaurant information, including menus, photos, ratings, reviews, and more. Through its food delivery service, users can order food from local restaurants for delivery to their homes. Founded in 2008 in India, Zomato has since expanded its operations to other countries worldwide. In addition to its website and mobile app, Zomato also maintains a presence on social media platforms such as Facebook, Twitter, and Instagram [14]. As a research collaborator, Zomato supplied all the data used in this study. This encompassed a Database (DB) containing reviews, restaurants, and all information associated with these domains, available on their platform.

1.1 Motivation

The motivation for this research is rooted in the significant role that the food service industry plays in Portugal's economy and the daily lives of its citizens. The industry's importance is highlighted by the substantial revenue it generates and the high percentage of household expenses allocated towards food services. However, despite the industry's significance, there is a gap in understanding customer sentiment and preferences, which is crucial for improving service quality and customer satisfaction. To address this gap, online reviews have emerged as a powerful tool in shaping consumer behavior and decision-making processes.

However, the analysis of these reviews, particularly the comments section, presents a complex challenge due to the intricacies and subjectivity of human language. This complexity is further compounded when considering the cultural nuances and regional variations in languages such as Portuguese.

Despite these challenges, the potential benefits of analyzing customer reviews are immense. Advanced NLP techniques can extract meaningful information from comments, providing valuable insights into customer opinions and experiences that can enhance customer satisfaction and increase revenue.

Moreover, the application of NLP techniques is not limited to English. There is a wealth of data available in other languages, such as Portuguese, that could significantly benefit from NLP analysis and improve products and services in more countries globally.

This research aims to bridge the gap in understanding customer sentiment and preferences in the food service industry by employing NLP techniques to analyze reviews in both English and Portuguese. By collaborating with Zomato, access to a wealth of data, including reviews and restaurant information, will be provided. Comprehensive insights will be gained, service offerings will be enhanced, customer satisfaction will be improved, and the growth and success of the industry will be contributed to.

A central motivation for this research is to develop a tool for restaurant owners that enables improvements in service quality. By analyzing customer reviews and feedback, the tool will offer valuable insights into the factors that customers prioritize when choosing a restaurant. This information empowers restaurant owners to make data-driven decisions that enhance service offerings, ultimately leading to higher customer satisfaction and increased competitiveness in the food service industry.

In conclusion, the motivation for this research is rooted in the desire to enhance the food service industry by gaining a deeper understanding of customer sentiment and preferences. Through the application of advanced NLP techniques, valuable insights can be extracted from customer reviews to inform service improvements and increase customer satisfaction. Extending these techniques to multiple languages has the potential to make a meaningful contribution to both the field of NLP and the global food service industry. Ultimately, this research aims to provide restaurant owners with a powerful tool to deliver superior service to their customers.

1.2 Research Questions

The present study addresses the challenge of analyzing restaurant reviews in the food service industry. The following research questions are proposed:

1. **Emotion Classification:** What is the ideal number of sentiment emotion classes for analyzing restaurant reviews, considering the complexity and diversity of human emotions expressed in these reviews?
2. **Topic Modeling:** How can the optimal number of topics for topic modeling in restaurant reviews be determined by combining statistical methods and domain-specific knowledge? What are the key topics that customers frequently discuss, and what insights do they provide about the aspects of the service that customers care about most?
3. **Data Visualization:** How can clear, concise, and interactive visualizations be used effectively to convey complex data insights derived from the analysis of restaurant reviews? How can these insights be ensured to be accessible and meaningful to a wide range of users?
4. **Alternatives to LLMs:** Are there alternative methods to Large Language Model (LLM)s for generating text for SA and topic modeling that produce similar results but have fewer disadvantages?
5. **Implementation:** How can a SA platform be optimized to categorize customer emotions from restaurant reviews, and what factors influence the successful implementation of a cloud solution?

With these research questions, the aim is to address critical challenges in analyzing restaurant reviews and present a road map for developing more effective tools and techniques in this area. The approach underscores the importance of a nuanced and tailored approach, the combination of statistical methods with domain knowledge, and adherence to practical data visualization principles.

1.3 Research Objectives

This research proposes to analyze the data provided by Zomato, along with all the available information, such as ratings, restaurants, comments, and more, to generate valuable insights for restaurant owners, managers, and other stakeholders. By analyzing this data, it is possible to understand customer preferences and experiences better and use this information to make informed

decisions that can improve the restaurant business. The proposed method for delivering this information is through a report that presents the generated insights in an easily interpretable manner. This report will provide restaurant owners with insightful information regarding the emotions expressed in their reviews, the topics that were discussed in their reviews, how their distribution of gender is formed, and how their restaurant compares with nearby restaurants that can guide them toward making improvements to their business. This approach was developed through multiple meetings with the company during the project, along with informal discussions, a literature review, and project requirements. To accomplish this objective, the following sub-objectives were defined:

1. Identify state-of-the-art methodologies for text analysis relevant to the proposed research.
2. Create a consolidated DB to better access and process data
3. Examine attributes such as review and restaurant information and analyze their interrelationships.
4. Generate new features from the existing ones and assess their correlations with each other and with established attributes.
5. Develop a method to determine user gender and investigate gender-based disparities within customer reviews.
6. Create a technique for sentiment identification to explore dataset variations.
7. Perform topic detection to unveil the subjects discussed in the reviews.
8. Create recommendations based on the identified topics.
9. Create multiple approaches for data presentation, encompassing Adaptive Stochastic Textual Assembler (ASTA), LLM, word clouds, and multiple data plotting techniques.
10. Establish a web platform to showcase all the insights above comprehensively.
11. Generate a Portable Document Format (PDF) report that encapsulates all the insights curated by the web platform.
12. Deploy the web platform on a cloud server, like Google Cloud.
13. Create and deploy a form to gather user feedback on the best forms of data visualization.
14. Create and deploy a form to gather user feedback on how to label the detected topics.

15. Create and deploy a form to gather user feedback on the best forms of written data delivery.
16. Utilize user feedback across all channels to enhance and fine-tune the web platform and PDF report generation.

1.4 Thesis Structure

This section provides an overview of the thesis structure, outlining the sequence of chapters and their content. Each chapter contributes uniquely to the research narrative, offering insights into methodologies, findings, and conclusions. The structure is designed to provide a comprehensive understanding of the methodologies, analyses, and findings of the study. The following chapter descriptions serve as a concise guide for navigating through the thesis and ensuring a cohesive understanding of the presented research. The visual representation in Figure 1 provides an overview of the thesis structure, illustrating the interconnections between the chapters and the overall data processing flow. Starting from the user review data provided, the diagram outlines the sequential steps from data collection to cloud deployment, emphasizing the methodologies, analyses, and system design stages.

- **1 Introduction:** This chapter sets the stage for the research by highlighting the significance of the food service industry in Portugal and the impact of online reviews on consumer behavior. It discusses the importance of review systems in various industries and the challenges associated with analyzing numerical ratings and textual comments. The chapter concludes by emphasizing the potential benefits of applying NLP techniques to analyze reviews in both English and Portuguese.
- **2 Data Collection and Database Preparation:** Here, an overview of the methodologies used for data collection, preparation, and processing is provided. It discusses the selection between Structured Query Language (SQL) and Not Only SQL (NoSQL) databases, the challenges faced during the denormalization process, and the effectiveness of using MongoDB for managing and analyzing Zomato's data.
- **3 Exploratory Data Analysis:** This chapter delves into the initial analysis of the dataset. It covers various aspects such as the rating system, review length, language detection, gender detection, and timetable separation. The aim is to offer a comprehensive overview of the dataset and highlight critical insights that will inform subsequent analyses.

- **4 Emotion Detection:** The exploration of the identification and categorization of emotions in restaurant reviews is the focus here. It discusses the use of the EMOTAIX.PT lexicon and the SentiStrength method for SA. The chapter highlights the importance of language choice and the challenges associated with emotion detection in multilingual reviews.
- **5 Feature Selection:** This chapter focuses on identifying the most relevant features that influence review ratings. It discusses the methodologies employed, including correlation analysis, the minimum Redundancy Maximum Relevance (mRMR) method, and the Borda Count (BC) method. The chapter concludes by addressing how these discoveries can enhance the understanding of review ratings and improve predictive models.
- **6 Unsupervised Topic Detection Enhanced by Human Labeling:** The method for detecting topics in restaurant reviews using Latent Dirichlet Allocation (LDA) is presented here. It discusses the data preparation steps, the challenges of selecting an optimal number of topics, and the human-in-the-loop approach for labeling topics. The aim is to provide a detailed analysis of customer feedback by identifying specific subjects discussed in reviews.
- **7 Visual and Textual Presentation of Data According to User Choice:** This chapter discusses the creation and deployment of surveys to enhance the system based on user opinions. It covers the demographic information collected, feedback and evaluation habits, and specific survey questions. The importance of user-friendly visualizations and the integration of both ASTA and Generative Pre-trained Transformer (GPT)-3.5 for text generation are also highlighted.
- **8 System Design for Reporting:** The ReviewSight platform, developed using Streamlit, is introduced here. The chapter discusses the platform's features, including tailored reports, a secure login system, and a connection to a MongoDB DB. The emphasis is on the platform's ability to provide valuable insights for restaurant owners and facilitate data-driven decision-making.
- **9 Cloud Implementation:** The focus of this chapter is the deployment of a Linux Virtual Machine (VM) on the Google Cloud Platform (GCP). It discusses the customization options provided by GCP, the configuration of the VM for remote operation, and the successful deployment and operation of the platform. The potential of cloud services in delivering scalable and adaptable solutions for enterprise environments is highlighted.

- **10 Conclusion:** This chapter summarizes the key findings of the research, discusses the implications of the results, and suggests potential areas for future research.
- **Appendix:** The appendix section includes additional documents, tables, images, and code snippets. It provides supplementary information that supports the main chapters of the thesis.

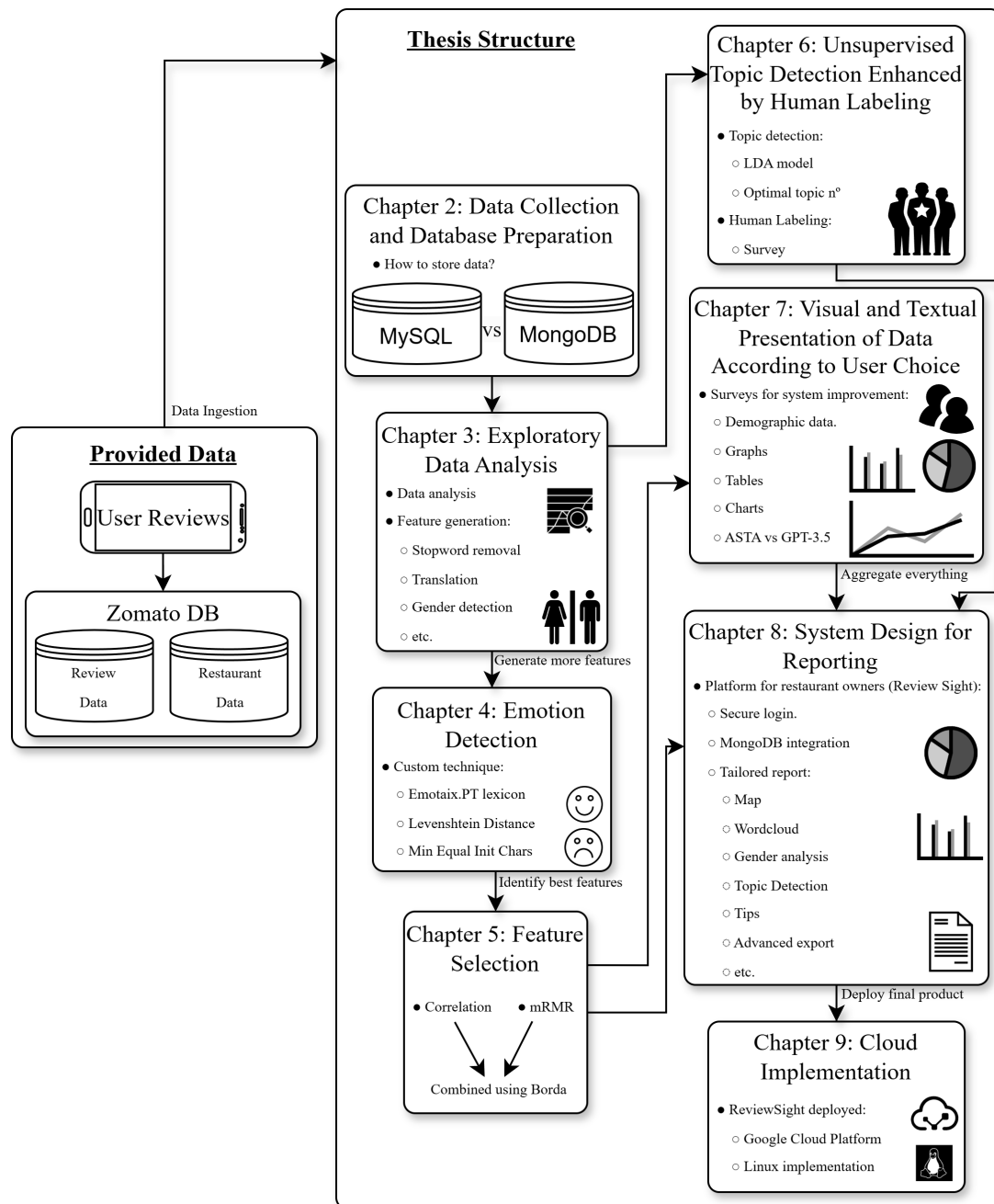


Fig. 1: Overview of the thesis structure and data processing flow.

2 Data Collection and Database Preparation

2.1 Introduction

This chapter provides an overview of the methodologies utilized for data collection, preparation, and processing in this research. The data obtained from Zomato presents a rich dataset of restaurant reviews and establishments in Portugal. The procedures for data cleaning and pre-processing are also discussed. A selection between SQL and NoSQL DBs for data management and analysis is then pondered. The specific challenges faced during the denormalization process and how these were surmounted using MongoDB are further explored. The chapter concludes with a reflection on the effectiveness of the chosen DB system in managing and analyzing Zomato’s data. A meticulous account of the measures undertaken to ensure the accuracy and reliability of the data is also provided, laying a robust foundation for subsequent analysis.

2.2 State of the Art (SQL)

Pradha et al. [15] proposed a compelling text data pre-processing technique, along with an algorithm to train Support Vector Machine (SVM), Deep Learning (DL), and Naive Bayes (NB) classifiers for SA on Twitter data. They compared various pre-processing methods, such as multiple methods of stemming, lemmatization, and spelling correction. They also developed an algorithm that assigned weights to the sentiment score based on the hashtag and cleaned text. They applied their technique and algorithm to compare the popularity of Google Now and Amazon Alexa using data from 3 million tweets. Their findings revealed that stemming was the most efficient in terms of computational speed, and SVM achieved the highest accuracy, with a rate of 90.3%.

Akrivi et al. [16] applied several pre-processing techniques to three Twitter datasets. The pre-processing methods evaluated included different data representations: unigram, bigrams, and 1-to-3 grams, and two feature extraction filters: one based on information gain and the other based on the Random Forest (RF) algorithm. Their results indicated that with appropriate feature selection and representation, SA accuracy can be improved. They suggested further investigation into the available pre-processing options to find the optimal settings. Future research plans involve embedded methods that carry out feature selection and model tuning at the same time [16].

Duong and Nguyen-Thi [17] conducted a study on SA in Vietnamese using semi-supervised learning. They implemented pre-processing techniques such as data cleaning, normalization, nega-

tion handling, and intensification handling. Data augmentation techniques were also used to enrich the original dataset. The study yielded promising results, demonstrating the effectiveness of these methods. The experiments performed presented the effects of these techniques when applied to Vietnamese datasets. For most of the experimented datasets, the accuracies of classifiers were improved. Moreover, ensembles of data augmentation techniques were also experimented with, and competitive results were obtained. These results show that the approaches are reasonable and suitable for Vietnamese, saving cost and time to build a pre-labeled dataset and gradually reaching domain independence.

Wibawa et al. [18] conducted a comprehensive study on restaurant data in Bandung, Indonesia, using web scraping techniques to extract data from the Zomato online food marketplace. They analyzed various attributes such as restaurant locations, types of cuisine, costs, reviews, and facilities. They found a positive correlation between restaurant costs and ratings, indicating that higher costs often led to better ratings. The research emphasized the importance of using big data and web scraping to gain valuable business insights into the restaurant industry.

Vajjhala et al. [19] explored the impact of online reviews on consumer purchase intentions within the food service industry, specifically focusing on the Zomato platform. Their mixed-method study, which includes both quantitative surveys and qualitative focus group discussions, reveals that online reviews and star ratings significantly influence consumers' willingness to pay. The research identifies critical moderating variables, such as information seeking, risk reduction, approval seeking, and reassurance, which collectively shape consumer perceptions and decisions. This study underscores the importance of positive online reviews in enhancing business visibility.

2.3 State of the Art (NoSQL)

Li et al. [20] highlight several advantages of NoSQL DBs in their paper. They emphasize that NoSQL DBs are designed for speed and scalability, making them well-suited for handling Big Data. These DBs often support flexible schemas, allowing for the storage of diverse data types without requiring a fixed structure. Additionally, NoSQL DBs like MongoDB and others offer features such as auto-sharding and data replication, which enhance their ability to manage large volumes of data efficiently.

Lourenço et al. [21] conducted a thorough assessment of NoSQL DBs, examining their appropriateness for various software quality attributes. The study underscores the shortcomings of traditional relational DBs in managing web-scale and Big Data systems, which has led to the rise of NoSQL technologies. The authors classify NoSQL DBs into four categories: Key-Value Stores, Document Stores, Column Stores, and Graph DBs, each providing distinct benefits for specific applications. They stress that while performance is a crucial aspect, other quality attributes such as availability, consistency, durability, and scalability are equally vital in the design of enterprise systems. This research offers valuable guidance for software engineers and architects in choosing the most suitable NoSQL DB based on specific quality attribute needs.

2.4 Data Source

With the assistance of Zomato’s technical team, multiple files were extracted from their DB. These files were categorized into two types: reviews and restaurants. On the review side, Zomato provided 101 files in Comma-Separated Values (CSV) format, each representing approximately a month of reviews, spanning from April 1st 2014 to September 2nd 2022. The merged and non-duplicate entries contained approximately 1.3 million reviews distributed, as shown in Fig. 2.

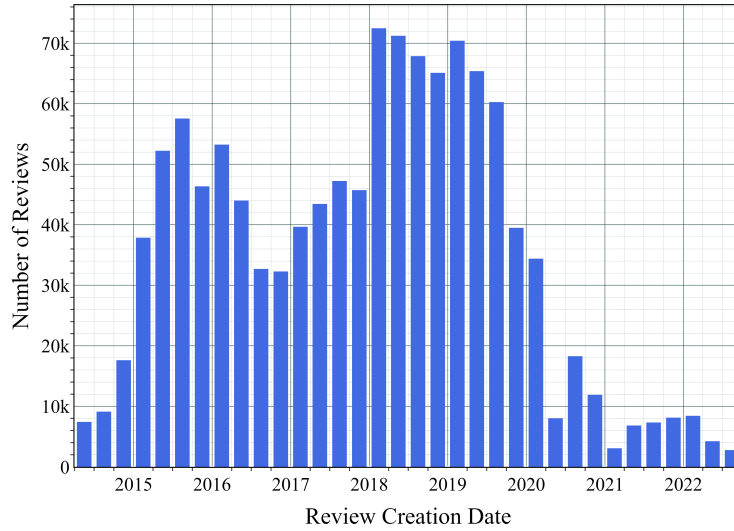


Fig. 2: Distribution of reviews according to their creation date.

The review files encompassed information about each available review on Zomato’s platform, as detailed in Appendix Table 11. This table comprises ten distinct fields, including User and restaurant names, the assigned review rating, and other pertinent information.

On the restaurant side, only one file was provided. This file encompassed all the relevant information regarding the restaurants, as presented in Appendix Table 12. The table comprises 30 distinct fields, encompassing details such as the Restaurant’s name, location, average rating, description, and other pertinent attributes.

The data from Zomato proved invaluable for this research, as it contained comprehensive information about reviews and establishments in Portugal. This facilitated all the work done in this thesis.

2.5 Data Cleaning and Pre-processing Methods

Data cleaning and pre-processing are essential in the preparation of data for analysis. They enhance the accuracy and reliability of the final dataset by discarding any unrepresentative or misleading information. This refinement process enables the research to concentrate on the most pertinent aspects of the data. For the vast amount of data and due to different tests for research, data cleaning is done in several steps.

In the realm of text mining, several specific pre-processing techniques come into play. Tokenization, for instance, disassembles text into individual units, while conversion to lowercase promotes uniformity across the text. Frequently occurring words with minimal semantic value, referred to as stopwords, are typically excluded. Stemming is another technique that simplifies words to their root form, thereby grouping semantically related terms.

Additionally, special characters that lack semantic contribution are removed. These techniques are selected based on the research question and characteristics of the corpus. They aid in mitigating data sparsity and augmenting predictive power, albeit with a potential impact on the results [22].

In the first stage, given the substantial volume of information within Zomato’s DB, the following data-cleaning measures were taken:

1. Removal of features with redundant information
2. Elimination of identification number-related features (e.g., user_id, restaurant_id, menu_id, chain_id)
3. Exclusion of reviews with out-of-range ratings (1 to 5)

4. Removal of features containing personal or irrelevant data (e.g., manager_phone, manager_email, and address_full)
5. Elimination of features with a high percentage of empty fields (e.g., seats_interior and seats_exterior)
6. Removal of test users corresponding to user_id 1, 2, and 3

2.6 SQL and NoSQL Implementation

When selecting a DB management system for storing and retrieving data in scientific research, SQL and NoSQL DBs represent two primary paradigms, each with distinct features and advantages. The choice between SQL and NoSQL DBs hinges on thoroughly evaluating the specific data requirements, scalability needs, consistency models, and the complexity of the queries to be executed. This work implemented both of them to investigate the best choice for the job.

2.6.1 SQL

SQL DBs are known for their relational structure and compliance with Atomicity, Consistency, Isolation, and Durability (ACID) properties, ensuring data integrity and reliability. On the other hand, NoSQL DBs are non-relational and focus on scalability and performance rather than strict consistency, often using Basically Available, Soft state, and Eventual consistency (BASE) properties. This makes NoSQL DBs ideal for managing large volumes of unstructured data and for applications that require flexible schema design, such as social media platforms and big data analytics [23] [24].

Utilizing an SQL DB simplifies the process of executing simple queries on data. This capability enables accessing specific DB segments without necessitating access to the entire dataset. For instance, all reviews from restaurants in Lisbon, all reviews with comments, or all reviews with ratings higher than 3 becomes feasible. To achieve this functionality, a web server application named Cross-Platform, Apache, MySQL, PHP, and Perl (XAMPP) was used. XAMPP establishes a local My Structured Query Language (MySQL) server accessible through a web client [25]. The choice of XAMPP stemmed from its user-friendly interface and extensive feature set. Within the MySQL server environment, two separate tables were established and populated using the above restaurant and review files. The following Example is in the Code excerpt. 1 query demonstrates how to extract all reviews, their respective ratings, and comments from a specific restaurant, in this case, restaurant 18347750:

```

1 SELECT review_text, rating
2 FROM review
3 WHERE res_id = 18347750
4 AND NOT (review_text = "")

```

Code 1: SQL code snippet to retrieve comments and ratings from all reviews associated with the restaurant identified by ID 18347750.

The output of this query can be seen in Fig. 3 a).

2.6.2 NoSQL

Another possible approach to managing data for this project is to use a NoSQL DB like MongoDB. NoSQL DBs have the advantage of using files to store information instead of tables, providing flexibility in handling semi-structured or unstructured data. They are designed to excel in scalability and performance in distributed environments, offering advantages in handling large volumes of data and supporting agile development practices. Since files don't need strict formatting, each can store only the necessary fields. This simplifies the DB structure and eliminates empty fields that would otherwise exist in a traditional SQL DB [26]. This approach would also be efficient for handling Zomato's data, which has a lot of variation in the number of fields that contain information.

A MongoDB DB was created using MongoDB Community Server and MongoDB Compass to explore this method. All files were imported as collections into the DB with the most efficient datatype possible to minimize size. The equivalent query to the one mentioned above using MySQL, now using MongoDB, could be represented by the Code excerpt. 2.

```

1 filter = {res_id: 18347750, review_text: {"$exists": true}},
2 project = {review_text: 1, rating: 1}

```

Code 2: MongoDB code snippet retrieves comments and ratings from all reviews related to the restaurant identified by ID 18347750.

This query generates the same information as the aforementioned MySQL example, and the results are shown in Fig. 3 b).

2.7 Selection Of NoSQL Database and Design

The selection process was straightforward due to the needs of the data. Creating a single file that aggregated the review and restaurant information simultaneously was necessary to perform the

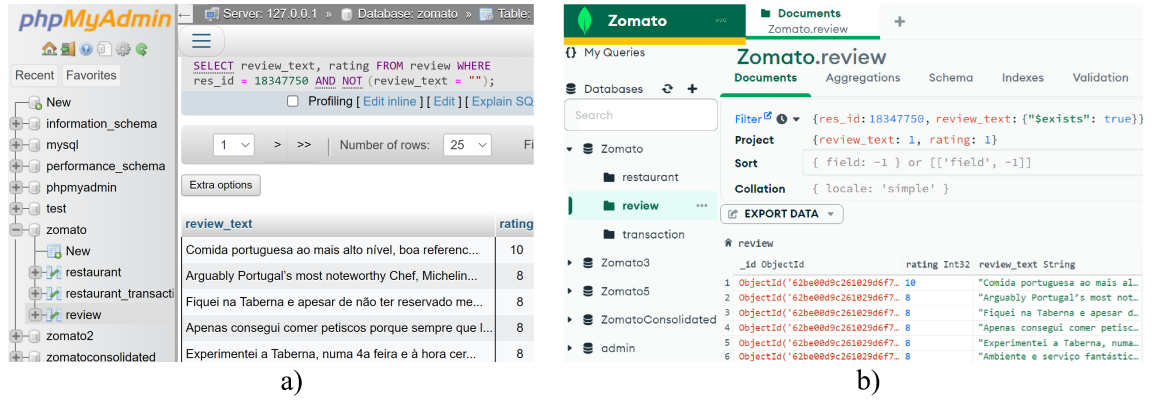


Fig. 3: **a)** An example of an SQL query is displayed in the phpMyAdmin interface via XAMPP, retrieving comments and ratings from all reviews associated with the Restaurant identified by ID 18347750. **b)** Example of a MongoDB query, visualized in MongoDB Compass, retrieves comments and ratings from all reviews associated with the Restaurant identified by ID 18347750.

intended analysis. This file would contain repeated information and would not be an efficient way to store data. However, when there is a need to perform feature creation, having a single file where all the feature-generating analyses and processes can be completed and added is much more convenient. Having a single file containing all the information is also very useful when performing NLP processes that always require access to the entire data pool. The following process was used to create this consolidated file using SQL. This process encountered difficulties due to the extensive number of reviews and restaurants present in the data. The issue was a timeout during the execution of the denormalization process. Denormalization is a DB optimization technique where redundant data is added to one or more tables, which can help avoid costly joins in a relational DB [27]. It is an optimization strategy applied after normalization. Despite implementing several measures to enhance the efficiency of the process and extend the maximum timeout duration, the error remained unresolved. To combat this, it was decided to try to perform the denormalization using MongoDB. An aggregation is needed to create a query that combines multiple collections in MongoDB. An aggregation groups values from various documents into a single document using a field as an index to compare the values in both collections in a similar way that selecting from multiple tables works in SQL DBs. To create the aggregated file that combines all the reviews with their corresponding restaurant information, the following aggregation needs to be available in the Code excerpt. 3.

```

1 $lookup:{
2   from: 'restaurant',
3   localField: res_id,
4   foreignField: ResID,
5   as: 'Restaurant'

```

```
6 },  
7 $unwind: {  
8   path: '$Restaurant'  
9 }
```

Code 3: Code snippet illustrating the construction of a denormalized collection in MongoDB, derived from the 'review' and 'restaurant' collections. This collection is intended to serve as the primary database for executing all future operations.

This aggregation process took around 4 hours, involving approximately 2.6 billion comparisons. The outcome of this process was a file of 1.84 Gigabytes containing 1.3 million lines. The file encapsulates the entire dataset from Zomato's DB in CSV format. This foundational file facilitated subsequent steps, encompassing further data cleaning and feature generation. After obtaining the resulting file and uploading it back into MongoDB, all kinds of queries could be applied to it without creating additional aggregations. The generated file represents an aggregated DB that can be used as the final denormalized DB for the project to perform all necessary analyses since all data is condensed into a single file. Smaller versions of this DB can also be generated to feed models for training with specific conditions instead of using the entire dataset. These conditions can include reviews from the last month or year, reviews from restaurants that accept payment by card or are located in Lisbon or Funchal, etc. These conditions can also be combined to create a small subset of the DB (e.g., reviews from the last year with more than 50 words written by males in restaurants from Funchal). Access to particular subsets of the dataset can help extract statistical information or train a model with specific goals.

2.8 Conclusion

The present study considered SQL and NoSQL DBs for managing and analyzing data from Zomato's DB. Each approach has its advantages and disadvantages. Ultimately, the specific needs and requirements of this study determined that using a NoSQL DB, specifically MongoDB, was the best option for this project. This was due to its ability to create an aggregated file and its convenience and efficiency in storing all necessary information in a single file. This approach allows for easy access to specific subsets of the data and provides a flexible and efficient way to manage and analyze the data from Zomato's DB.

3 Exploratory Data Analysis

3.1 Introduction

Exploratory Data Analysis (EDA) is a method of analysis that uncovers and identifies patterns in the data. EDA is a fundamental practice in data analysis that allows restaurant owners to understand their users better by using statistical graphics, plots, and information tables. This chapter is dedicated to the EDA of the dataset under consideration. The large volume of the data necessitates a thorough understanding of its intricacies. The state-of-the-art in this field is examined, focusing on an analysis of the rating system. The investigation is extended to the length of reviews, focusing on their potential influence on other features. Attention is also given to language detection and translation, emphasizing review comments in Portuguese and English. Gender detection is explored to discern any differences in the way males and females rate and review meals, as well as their experiences at restaurants. The chapter concludes with a discussion on timetable separation, a process that facilitates a more detailed analysis of the opening and closing hours of restaurants on Zomato’s platform. This chapter aims to offer a comprehensive overview of the dataset and underscore critical insights that will inform subsequent analyses of the restaurant owners.

3.2 Comparative Analysis of Sentiment Analysis Techniques

This study employs a threshold-based method to directly link reviews with their ratings, ensuring 100% accuracy by eliminating the need to infer sentiment. This approach contrasts with other studies that use ML and NLP techniques to infer sentiment from textual data, as shown in Table 1. Typically, ratings are used to establish ground truth for training ML models. However, in systems relying on cascading thresholding and ML algorithms, the ML model cannot outperform the thresholding method due to structural limitations. Introducing a ML model adds complexity, increases computational demands, and introduces errors that can propagate through the system. In food rating systems, the presence of ratings often negates the need for complex ML models. Given that ratings are consistently available and paired with text when present, sentiment classification can be effectively achieved using simple thresholding methods. Therefore, the additional complexity of an ML model is unnecessary.

For instance, Sajal et al. analyzed Coronavirus Disease 2019 (COVID-19) tweets, highlighting the importance of understanding public opinion during the pandemic. Md. Rakibul Hasan et al.

(2019) achieved 85.25% accuracy using NLP on Twitter data with a Term Frequency-Inverse Document Frequency (TF-IDF) vectorizer. Mohd Ridzwan Yaakub et al. reviewed various SA techniques, concluding that ML methods typically achieve higher accuracy. The simplicity and efficiency of this method make it useful for business, social media, and customer feedback analysis. Accurate SA provides valuable insights into customer opinions, enabling businesses to improve their products and services. Additionally, this method is easy to implement and scale, making it practical for various industries.

Table 1: Overview of SA Studies: Supervised Learning (SL), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), Bag of Words (BoW)

Reference / Author	Description	Method	Max Accuracy
[28] Mohd Ridzwan Yaakub et al.	Review of SA techniques	ML, NLP	96%
[29] Sajal et al.	SA of COVID-19 tweets	SL	94.9%
[30] Dang et al.	SA using DL	CNN, RNN, LSTM	90.45%
[31] Md. Rakibul Hasan et al.	SA on Twitter data	BoW, TF-IDF	85.25%
[32] Gupta et al.	SA of Zomato reviews	NLP, ML	Not Specified
[33] Monisha Kanakaraj et al.	SA on Twitter data	WSD, WordNet	Not Specified

This research splits reviews by a threshold based on rating, where the threshold is a predefined value that creates the split between different sentiment categories, a method also used in other studies. For example, Bern Jonathan et al. analyzed customer reviews for Bangalore restaurants on Zomato using a RF Classifier. They preprocessed 150,000 reviews and vectorized them using TF-IDF, achieving 92% accuracy. Similarly, Branco et al. conducted SA on Portuguese restaurant reviews using transformer-based models and edge computing, addressing imbalanced datasets by re-sampling the original nine label values into three classes: negative, neutral, and positive. This approach ensured a more balanced dataset, crucial for improving sentiment classification accuracy.

3.3 State of the Art

Sahayak et al. [34] conducted EDA on Twitter data. They extracted tweets and used machine learning classifiers such as NB, Maximum Entropy (MaxEnt), and SVM to classify these tweets into positive, negative, or neutral sentiments. Feature extractors like unigrams, bigrams, and unigrams with part of speech tags were used to process the data. They also suggested implementing features like emoticons, neutralization, negation handling, capitalization, and internationalization to address the complexity and nuance of opinions expressed on social media.

In their study, Kanwal et al. [35] conducted EDA on restaurant reviews using a dataset from "SWOT's Guide to Karachi's restaurants" [36]. This dataset, consisting of approximately 4000 customer posts, was processed and transformed into a structured format for analysis. The researchers employed a comprehensive methodology for their analysis. They used NLP techniques such as lexical analysis, semantic analysis, and syntactic analysis for data pre-processing. This involved tokenization of the data, removal of stop words, and lemmatization to combine inflected forms of words into meaningful sentences. To refine the data, they converted words into root words by eliminating prefixes and suffixes using the Natural Language Toolkit (NLTK) Stem component, which works on the Porter Stemmer algorithm. They applied the TF-IDF, a term-weighting scheme used for text mining, to identify the relevancy of the terms in the dataset. After pre-processing, they prepared a bag of words that represented sentences with root words, independent of any grammatical and punctuation rules. This EDA allowed them to uncover underlying patterns and trends in the data, demonstrating the potential of such techniques in extracting insights from unstructured text data.

Vajjhala et al. [19] explored how online reviews impact consumer purchase intentions, with a specific focus on the Zomato platform. Combining quantitative surveys and qualitative focus group discussions, the researchers identified vital variables influencing consumer decisions. Notably, online reviews and star ratings significantly affect consumers' willingness to pay, with factors like information-seeking, risk reduction, and reassurance playing pivotal roles. This study underscores the significance of user-generated content in shaping consumer behavior and highlights the critical role of electronic word-of-mouth in the food service industry.

Alonso et al. [37] explored the key factors influencing restaurant choice among 277 consumers, primarily from the Southeastern United States. They discovered that food quality (including taste) played a central role in respondents' decisions. Additionally, positive past experiences, a clean production/service environment, and hospitable service significantly impacted restaurant preferences. Surprisingly, despite the region's diverse food offerings and agricultural heritage, the use of local ingredients in menu development was given less consideration. The study underscores the importance of maintaining consistent quality and service to attract and retain customers.

The studies discussed above collectively demonstrate the power of EDA in uncovering valuable insights from diverse datasets. By utilizing machine learning classifiers and NLP techniques,

researchers have successfully classified sentiments and structured unstructured data and identified significant patterns in consumer behavior. Key findings highlight the importance of nuanced features in SA, the effectiveness of lexical, semantic, and syntactic analyses, and the profound impact of online reviews on consumer decisions. Additionally, factors such as food quality, past experiences, and the service environment play crucial roles in shaping consumer preferences. These insights underscore the importance of analyzing user-generated content to understand its impact on restaurant ratings and reviews.

3.4 Rating

Integer rating systems are often favored over decimal counterparts for several reasons. The simplicity and straightforwardness of integer ratings reduce complexity and facilitate quick decision-making for users. They also can prevent over-analysis and decision paralysis that might arise where numerous options or excessive precision may overwhelm the user.

Additionally, the visual rounding effect also significantly contributes to this preference. This effect is the inclination of individuals to perceive rounded numbers, such as integers, as more trustworthy and aesthetically appealing. This visual rounding effect can sway user perception and the decision-making process, rendering it more attractive to a broad audience of users. Consequently, despite the additional nuance provided by decimal ratings, integer ratings are often chosen for their simplicity, user-friendliness, and the psychological influence of the visual rounding effect [38]. In the data, this phenomenon is further observed in the 0.5 steps, which were less frequently used, indicating that 1.5, 2.5, 3.5, and 4.5 ratings were less prevalent compared to their integer equivalents, 1, 2, 3, 4, and 5. The distribution of this rating across all reviews in the DB is depicted in Fig. 4 a). However, when an average of ratings is calculated from the integer ratings, it naturally concludes with a decimal number. Since the restaurant rating is derived from an arithmetic average of all its reviews, it is possible to observe with a granularity level beyond 0.5 intervals to 0.1 Fig. 4 b).

This granularity is suitable for distribution analysis, showing that the review rating distribution did not adhere to a Gaussian distribution as anticipated but instead exhibited a right-skewed Gaussian distribution. This implies that the most frequent rating was not 3, as would be expected in a 1 to 5-star rating system with 0.5 steps, but instead 4. This suggests that individuals tend to rate higher than anticipated and prefer using the integer part of the scale over the in-between values. The Distribution ratings, derived from the review ratings, follow a right-skewed Gaussian

distribution, as shown in Fig. 4. The conclusions drawn from this distribution are similar to those drawn from the review rating mentioned in the previous paragraph. However, since the restaurant rating is derived from an arithmetic average of all its reviews, it is possible to observe these conclusions with a granularity level beyond 0.5 intervals to 0.1. With this level of granularity, it is possible to observe the right-skewed Gaussian distribution.

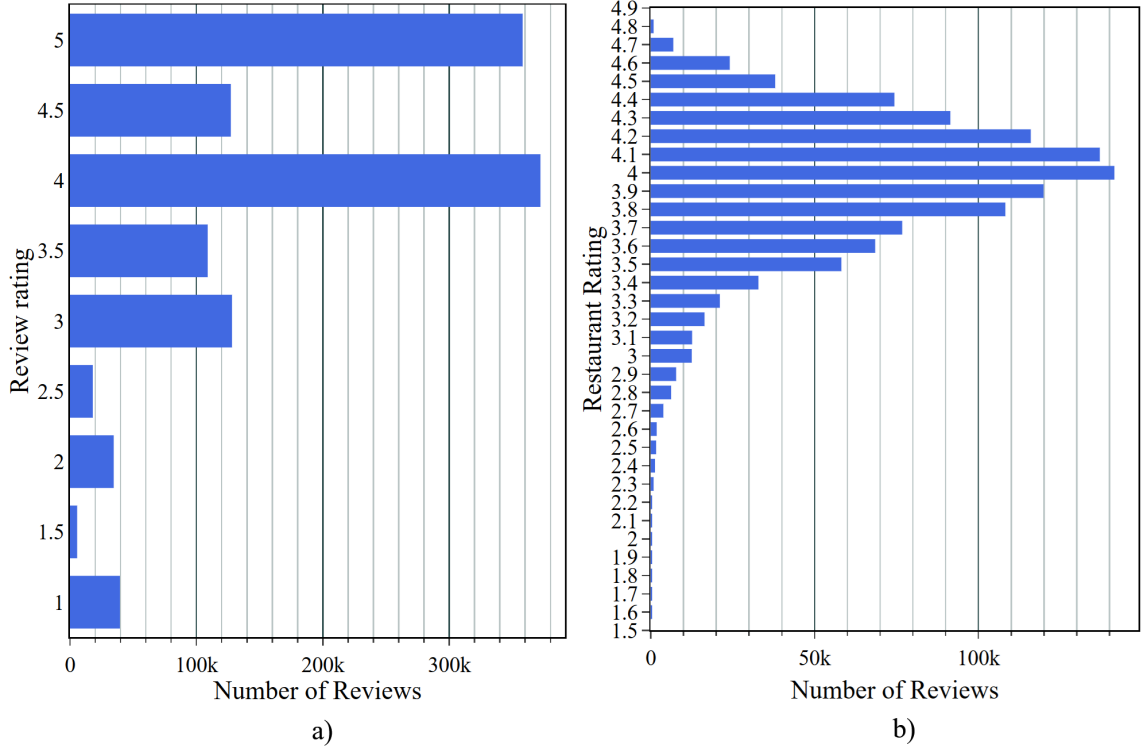


Fig. 4: a) Distribution of Zomato's Review Ratings. b) Distribution of Zomato's Restaurant Ratings

3.5 Stars rating's compatibility with a classification cluster

In SA, classifying review ratings into three categories—positive, negative, and neutral—is a common practice [39]. This classification simplifies the analysis and interpretation of the overall sentiment expressed in reviews. Positive ratings typically reflect satisfaction, negative ratings reflect dissatisfaction and neutral ratings indicate moderate satisfaction or ambivalence. By categorizing ratings into these three clusters, it becomes easier to analyze trends and patterns in customer feedback, which can provide valuable insights for improving service and understanding customer preferences. When attempting to cluster reviews into groups, one might expect that, since there are nine possible ratings, the ratings should be divided into three equally sized clusters as follows:

Cluster 1: (1.0, 1.5, 2.0), Cluster 2: (2.5, 3.0, 3.5), and Cluster 3: (4.0, 4.5, 5.0) [39] [40]. However, adhering to this initial clustering assumption results in a significantly skewed data distribution, predominantly favoring positive ratings as seen in Fig. 5 a).

There are multiple approaches to minimize the fact that the dataset is imbalanced. One commonly used technique is applying the inverse class frequency method. This method assigns higher weights to underrepresented classes and lower weights to more represented classes during the training process. This approach increases the significance of underrepresented classes, leading to a more accurate capture of patterns in the data [40]. Another approach is to cluster the reviews as closely to a standard Gaussian distribution as possible. This involves grouping them using the following groupings: Cluster 1: (1.0, 1.5, 2.0, 2.5, 3.0 and 3.5), Cluster 2: (4.0 and 4.5), and Cluster 3: (5.0). As seen in Fig. 5 b), this distribution follows a Gaussian distribution and indicates that what intuition suggests as the middle is not valid.

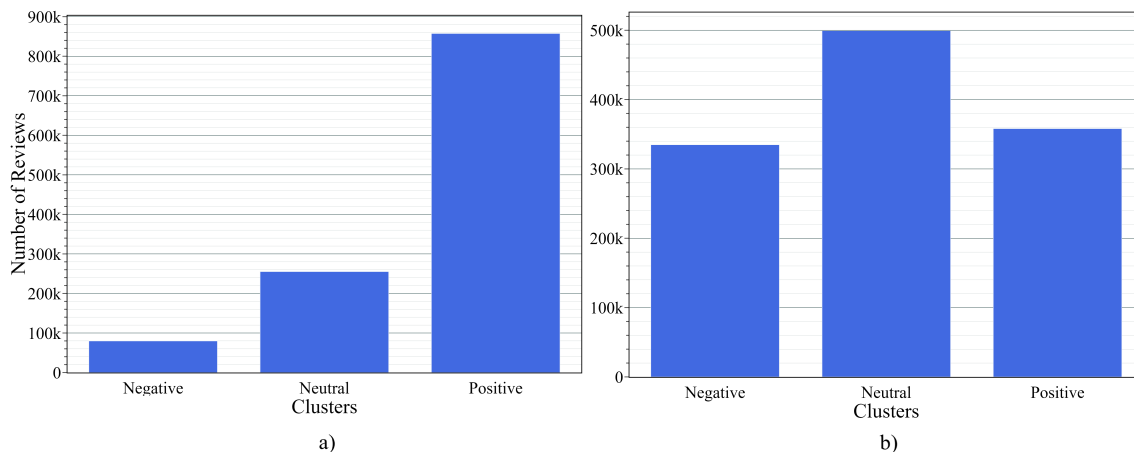


Fig. 5: **a)**: Distribution of restaurant ratings across all reviews using initial naive clustering assumption. Negative: (1.0, 1.5, and 2.0), Neutral: (2.5, 3.0, and 3.5), Positive: (4.0, 4.5, and 5.0). **b)**: Distribution of restaurant ratings across all reviews using a revised clustering approach. Negative: (1.0, 1.5, 2.0, 2.5, and 3.0), Neutral: (4.0 and 4.5), Positive: (5.0).

3.6 Review Length

This section offers a detailed examination of the lengths of reviews within the dataset. The length of a review can provide significant insights into the comprehensiveness of customer feedback and potentially correlate with the ratings provided. A word count methodology was employed to determine the length of each review, followed by an analytical assessment of the findings. In addition,

integrating word count data into the DB empowers researchers to filter reviews selectively according to specific length criteria and investigate potential associations with other characteristics.

3.6.1 Method: number of words

To calculate the number of words, a Python script was designed to process each review meticulously. This script was tasked with eliminating punctuation marks and extraneous white space, ensuring an accurate count of words in each review. Subsequently, the number of words counted by the Python script was recorded in a newly established column within the DB, uniquely associated with each review. This approach enriches the dataset and possibilities for more nuanced analyses of the textual content of reviews.

3.6.2 Results

A significant trend emerges upon analyzing review lengths. Approximately half of the 1.3 million reviews have a length of 0, leaving around 670 thousand reviews for NLP analysis. Despite this reduction, ample data remains for insightful NLP exploration. Most reviews fall within the 30 to 75-word range, with a median review length of 46 words and an average of 64 words per review, enriching the insights. An intriguing customer behavior pattern emerges on Zomato’s platform: nearly half of the patrons opt not to provide written reviews. Among the contributors, there is a preference for succinct expression, adding depth to the observations. This trend is visualized in Appendix Figure 33.

3.7 Language Detection

Though the dataset was scrapped from the Zomato Portugal DB, the reviewers wrote their reviews in different languages. For any kind of NLP, the detection of the language of a given text is recognized as a fundamental step, especially when the data is multilingual. This becomes particularly important when subsequent NLP tasks are to be performed on the detected languages. This section provides an outline of the method that was utilized for language detection and presents a discussion of the results that were derived from this process.

3.7.1 Method

The focus of this work is on NLP. As such, labels indicating the language of the review comments are crucial. To this end, the langdetect library [41] was used to create language labels in the form

of abbreviations (e.g., pt for Portuguese, en for English, es for Spanish, etc.). Not Applicable, N/A was used as a label when no language was detected. A Python process was created that utilized the langdetect library to check the language of each review and added a column to the DB with the respective language abbreviation.

3.7.2 Results

After running the process for all reviews and inspecting the distribution of languages in the review comments, it became clear that Portuguese was the most used language, with approximately 500 thousand samples. English was a distant second, with about 20 thousand samples, while all other 35 languages combined had less than 10 thousand samples. Because the distribution of languages was heavily focused on Portuguese and English, and translation between other languages and English may not always be optimal, it was decided to discard all languages other than Portuguese and English in this research.

3.8 Translation

Most of the NPL process is dependent on the language itself. Therefore, in the field of NLP, the role of translation, mainly when dealing with multilingual data, is crucial. This section is focused on the translation process of review comments between Portuguese and English. The process of translation, the methods employed, the results obtained, and their implications.

3.8.1 Method

High-quality translations of all reviews were produced to compare NLP in both Portuguese and English. This was crucial because Portuguese has received less NLP research than English [42]. The task was accomplished using GoogleTranslator from the deep_translator library [43] by translating all original Portuguese reviews into English and all actual English reviews into Portuguese. Using GoogleTranslator through the deep_translator library meant that the service was free, but there were daily limits for words per minute and Internet Protocol (IP). To circumvent this, the dataset was divided into ten different CSV files, and three computers ran the process over six days at different physical locations to expedite the process. To facilitate a comparative analysis of NLP capabilities across Portuguese and English languages—acknowledging the lesser extent of NLP research in Portuguese than English [42] a comprehensive translation of all review texts was undertaken. The objective was to generate high-quality translations of reviews between Portuguese

and English, leveraging the GoogleTranslator tool accessible via the `deep_translator` library [43]. This approach enabled the conversion of original Portuguese reviews into English and vice versa, thus ensuring a bilingual dataset for analysis.

The utilization of GoogleTranslator through the `deep_translator` library offered a cost-effective solution due to the free nature of the service, albeit with daily limitations on the number of words translated per minute and restrictions based on IP addresses. To overcome these limitations and expedite the translation process, the dataset was segmented into ten separate CSV files. These files were then processed on three different computers located at distinct physical locations, allowing the translation task to be completed over six days. This systematic approach not only circumvented the imposed restrictions but also facilitated the timely preparation of the dataset for subsequent NLP analysis. After translating all Portuguese reviews into English and all English reviews into Portuguese, they were added to the DB in two separate columns. One column contained all original Portuguese reviews and all English reviews translated into Portuguese, while the other column contained all original English reviews and all Portuguese reviews translated into English. This resulted in two complete DBs, one in Portuguese and another in English, which could be used most conveniently for each specific task.

3.8.2 Results

This method produced high-quality translations of all reviews to facilitate comparison and contrast between NLP analysis in Portuguese and English, as seen in Appendix Tables 13 and 14. Translating all reviews into both languages allows this research to analyze how different NLP techniques perform on data in other languages.

3.9 Stop Word Detection/Removal

Stop-words are words that, despite not carrying significant meanings, are often found in text. Their removal usually does not affect the overall understanding of the text. They are commonly used words such as 'is', 'at', 'which', 'on', etc. In the context of NLP and text mining, these words can be pretty distracting and could potentially drown out essential words in the text. Therefore, the removal of these stop words becomes indispensable to text analysis tasks, such as SA, where the critical focus is on identifying words that signify opinions, sentiments, and emotions [44]. Considering that Stop-word detection is an essential step in NLP, mainly when dealing with substantial

volumes of text data. This section is dedicated to this process, especially in the context of extensive text data. The primary focus here is the extraction of stop words from review comments. This process is vital for facilitating a more streamlined and targeted analysis of the principal content within these reviews.

3.9.1 Method

Stopwords are commonly used to connect words or phrases and are usually removed from text during NLP. Examples of stopwords include "a", "an", "the", "of", "in", etc. [45]. To avoid losing data, all the stopwords from the Portuguese review text were removed, and the newly generated text was added to the DB, keeping both versions of the reviews. A Python process was created to remove stopwords from the review text. This process used the NLTK library to remove Portuguese stopwords from all reviews except for the word 'not' [46]. Additionally, some other changes were done, namely removing periods, commas, ellipses, colons, semicolons, parenthesis, square brackets, any instance of two or more spaces between words, and special characters like #, &, « and others. This was done using a Python process that performs a verification that identifies and removes any instances of special characters or double spaces between words, one by one. This was achieved so that all the words are stored in string form in the DB without special characters and with a single space between words, making the following processes more accessible to perform.

3.9.2 Results

The implementation of this method facilitated NLP analysis by removing stop words from the review text. By discarding these commonly used words, which contribute minimally to the overall content of the review, the focus of the analysis could be directed toward the most pertinent words and phrases within the review text. From an efficiency perspective, this process also offered significant advantages. The median and average number of words per review were reduced from 46 and 64 to 28 and 39, respectively, equating to a 64% reduction in both metrics. This substantial reduction not only decreases the computational power required for complex tasks such as topic detection but also minimizes the amount of noise or irrelevant words in the dataset. Consequently, the prominence of the most significant words is enhanced.

3.10 Gender Detection

Chapman et al. [47] found that there are no significant gender differences in the ratings of positive and neutral words. Women consistently rated negative words more strongly than men across all three dimensions. This highlights the importance of considering gender effects in emotional processing studies.

Hess et al. [48] conducted three studies to explore gender stereotypes and self-perceptions related to emotional expressiveness. They investigated how people believe men and women react to different emotional situations and how individuals view their emotional responses. The findings showed that women are typically expected to display more sadness and retreat in adverse situations, while men are anticipated to exhibit more happiness and calmness. These stereotypes were consistent across imagined scenarios, self-reported expectations, and personal stories, indicating that cultural beliefs play a significant role in shaping both perceived and actual emotional reactions.

Gender detection is crucial for understanding differences in how males and females rate review meals, as well as their experiences at restaurants. Research indicates that while there are no significant gender differences in ratings of positive and neutral words, women tend to rate negative words more negatively compared to men. Additionally, cultural stereotypes suggest that women are generally expected to display more sadness and retreat in adverse situations, while men are anticipated to show more happiness and calmness. These insights underline the importance of considering gender effects in emotional processing and user behavior studies. This section will delve into the methods used for gender detection and analyze the results obtained.

3.10.1 Method

User gender is one piece of information that the Zomato platform does not collect. Still, it could prove helpful in identifying differences in the way males and females rate and review meals and their experiences in restaurants [49]. Men and women have distinct approaches to judging food and service quality [48]. Since there was no access to users' gender, The extraction of gender was done via the user's name using the `gender_guesser` library [50] in a Python process. Given that there are two possible names available in the reviews file, `user_pt_name` and `user_in_name`, which refer to the user names on the Portuguese and International DBs, respectively, both were used to check for the gender of the name. The Portuguese DB name was primary, and the International DB name

was used as a fail-safe. If the script could determine the gender using the Portuguese name, it would try to guess the gender of the International name. This was done for all reviews in the DB.

The `gender_guesser` library can label in six ways: `male`, `mostly_male`, `female`, `mostly_female`, and `(androgynous: a name that can refer to both genders)`, and `unknown`. To reduce the possible labels, females and mostly females were grouped into one group and males and mostly_males into another group, reducing the total number of possibilities to four. A fifth possibility was also added: `no_name` for users who didn't provide a name to the platform. Ultimately, there were five possible labels: `male`, `female`, `unknown`, and `no_name`.

3.10.2 Results

A comprehensive evaluation of the `gender_guesser` library's accuracy would necessitate the gender determination of all names in all reviews. However, this was not feasible due to the extensive size of the DB. Consequently, a sample of the reviews was analyzed for errors. The system appears to accurately identify the gender when users have a valid first name. Names containing unusual characters or symbols are labeled as `unknown` by the system. The results of this analysis are depicted in Table. 2.

Table 2: Example of First Names of Users and respective attributed genders.

First Name	Gender	First Name	Gender
João	Male	Sara	Female
Inês	Female	Ana	Female
Filipa	Female	Gisela	Female
Cátia	Female	Margarida	Female
Filipe	Male	?	Unknown
Tiago	Male	Telmo	Male
?	Unknown	Filipa	Female

Approximately one-third of the reviews were classified as `no_name`, `unknown`, or `androgynous`. Despite this, the remaining two-thirds had user names that could be identified as `male` or `female`. From the portion that could be classified, the distribution suggests that approximately 60% of reviews were made by women and the remaining 40% by men, as seen in Appendix Fig. 34. These results suggest that while a significant portion of reviews could not be classified by gender due to missing or ambiguous information, most reviews were made by women. This implies that women constitute the majority on the platform and contribute more reviews than men.

The literature suggests that men and women have different ways of quantifying emotion and choosing food [49] [48]. This could potentially influence the nature and content of their reviews, providing an additional layer of complexity to the analysis.

3.11 Timetable Separation

The procedure of separating the timetable facilitates a more in-depth examination of the operational hours of restaurants on the Zomato platform. By providing a detailed view of the opening and closing times, a comprehensive analysis is enabled. The methodologies employed for this separation are elucidated in this section, along with a discussion of the results derived from it.

3.11.1 Method

The Zomato restaurant file contained three columns named `time_slot_1`, `time_slot_2`, and `time_slot_3`. These columns held time slots for every day of the week in a similar form to a Python dictionary, e.g., 1: 12:30:00|15:00:00 ... 7: 12:30:00|15:00:00. Each time slot contained a time for every day of the week, and since there were three possible time slots, this resulted in 3 possible time slots for every day of the week.

To isolate the timetable for each day of the week, these time slots were split into 21 different columns using a Python process. Each column contained one of three time slots for each of the seven weekdays, totaling 21 different opening and closing hours. This was done to facilitate the analysis of all three timetables separately for every day of the week.

This separation enabled the organization of timetable information for each day of the week into individual columns, facilitating more detailed analysis. By splitting the time slot information into separate columns for each day, it becomes possible to analyze how various timetables relate to other features on Zomato's platform. Having each day's schedule in a distinct column may help identify specific correlations, such as whether being open or closed on certain days or during particular hours influences ratings positively or negatively.

3.11.2 Results

The analysis revealed that separating the timetable into individual columns for each day of the week did not yield significant insights. Despite the detailed breakdown of opening and closing times, no substantial correlations were found between these operational hours and the ratings given by users

on Zomato. This suggests that the specific operational hours of restaurants do not have a meaningful impact on their overall ratings. Several factors might explain these findings. The dataset could lack sufficient detail or variability, and restaurant ratings are likely influenced by various factors such as food quality, service, and ambiance, which may overshadow the effect of operational hours. Additionally, if operational hours are generally consistent across most restaurants or if external factors and complex interactions with other variables play a more significant role, these could also diminish the perceived impact of operational hours on ratings.

3.12 Conclusion

To summarize, the EDA in this chapter has provided an in-depth understanding of the dataset. Various methodologies, including review length analysis, language detection, translation, stop-word detection, gender detection, and timetable separation, have been applied to thoroughly explore the dataset. These methods could serve as a valuable guide for future analyses. The gender detection process emphasized the importance of understanding user behavior, potentially informing strategies for more personalized services. The translation process facilitated a comparison of NLP in both Portuguese and English, highlighting the significance of handling multilingual data in NLP. The stop-word detection process simplified the dataset by reducing noise and emphasizing important words. This comprehensive EDA has revealed vital insights that will guide future analyses, making a significant contribution to the broader field of data analysis and NLP.

4 Emotion Detection

4.1 Introduction

Emotion detection involves identifying and categorizing human emotions from various forms of data, including text, speech, images, and physiological signals. This is achieved through algorithms and machine learning techniques to analyze and interpret the emotional content in the input data. By recognizing emotions such as happiness, sadness, anger, and fear, emotion detection helps understand the underlying sentiment in human communication [51]. In this research, emotion detection is crucial for addressing key questions and objectives. It helps determine the ideal number of sentiment classes for analyzing restaurant reviews, supporting the development of a nuanced understanding of customer sentiments. It also aids in creating techniques for sentiment identification, optimizing a sentiment analysis platform, and contributing to effective data visualizations. The following section will explore current advancements in emotion detection and introduce the bespoke algorithm developed for this project, highlighting its functionality and role in achieving the research objectives.

4.2 State of the Art

Rahman et al. performed emotion detection on Bangla text by implementing a variety of classical machine learning algorithms. They used a dataset of 6,314 comments from Facebook groups, which were primarily reactions to ongoing socio-political issues. These comments were annotated for six basic emotions: sadness, happiness, disgust, surprise, fear, and anger. The researchers experimented with five different machine learning algorithms: NB, Decision Tree (DT), k-Nearest Neighbors (k-NN), SVM, and k-Means Clustering. They used various combinations of features, including unigrams, bigrams, trigrams, and part-of-speech tags. They also experimented with different preprocessing techniques, such as tokenization and stop-word removal. The best-performing model was the SVM with a non-linear Radial Basis Function (RBF) kernel. This model used the most important words using TF-IDF as features and achieved an accuracy of 52.98% and an F1-score of 0.3324 (macro) and 0.476 (micro) [52].

Alswaidan et al. examined multiple approaches to emotion recognition in text. They highlighted the limitations of keyword-based approaches for explicit emotion recognition and discussed the challenges associated with implicit emotion recognition. The authors noted that the effectiveness of

rule-based approaches is influenced by the quality of the text and the presence of rules representing the emotion in the rule set. They also pointed out that classical learning-based approaches require efficient features, while DL approaches necessitate large quantities of training data. The authors observed that although English is the dominant language for emotion detection, there is a growing body of work in other languages. They also noted that the most commonly used learning-based method is SVM and that the success of these approaches depends on the correct selection of features. Looking forward, the authors predict that the use of pre-trained word embeddings will become standard practice in emotion recognition and that transformer models will dominate the field of DL models [53].

Dey et al. explored various SA techniques used on social media and online platforms. They highlighted the crucial role of SA in forecasting political and business outcomes through the examination of online opinions. The authors addressed issues such as accuracy, data sparsity, and polarity shift and evaluated popular methods like NB, SVM, and Neural Networks (NN). Their study highlights the necessity of effective strategies for reliably achieving accurate SA results, serving as a significant resource for ongoing research in this field [54].

4.3 Emotion Detection Algorithms

Emotion detection is a crucial task in this research. It is a rapidly growing research area and comes in many forms. This research used a more traditional approach to perform the emotion detection process rapidly and efficiently. This method does not depend on external LLM or any other language processing tools, as it is fully executed with simple techniques that are performed locally. This allows for faster processing and analysis of the data, enabling quick identification of the polarity of the emotions present in the text.

4.3.1 Method

EMOTAIX.PT is a lexicon of Portuguese words [55] based on the original EMOTAIX, in French [56]. These lexicons are divided into three categories: Supra, Super, and Basic. The Supra category is divided into three positive valence classes: Benevolence, Well-being, and Composure, and three negative valence classes: Malevolence, Ill-being, and Anxiety. The Super category comprises 18 emotions that fit within the Supra classes, while the Basic category is divided into 50 different emotions that also fit within the Super and Supra Classes.

To utilize EMOTAIX.PT, as a lexicon for identifying emotions, a series of procedures was implemented. Initially, four different files were created, each containing lexicons with varying levels of granularity based on EMOTAIX.PT: Polarity, Supra, Super, and Basic Categories. These four CSV files were created separately to facilitate the implementation of distinct processes for each category. Multiple Python processes were then employed to analyze all the review texts and count the number of words in each phrase corresponding to each emotion. The CSV file with all the categories can be found in Appendix. 10.

For the Polarity category, an additional Neutral category was created, comprising the Surprise, Non-Specific Emotion, and Indifference basic categories. These were separated because they do not fit clearly into positive or negative valence, unlike the other categories. The categorization hierarchy starts with the Basic Categories, which contain specific emotion-related words. These are grouped into broader emotional themes called Super Categories, which are further grouped into even broader themes known as Supra Categories. At the top level, the Polarity category represents the overall valence of emotions, classified as positive, negative, or neutral.

In cases where words did not fit into any of the predefined categories, they were placed in an additional category called Unknown. All the types of the EMOTAIX lexicon are illustrated in Fig 6. This hierarchical structure ensures that the lexicons are organized from specific emotions to broader emotional themes, facilitating a comprehensive analysis of the text.

To visualize the results of the analysis, the following Portuguese review text is considered: "Comida portuguesa ao mais alto nível, boa referencia para os admiradores desta gastronomia, sem esquecer a parte das sobremesas! %% Recomendo vivamente." A Polarity sweep performed on this text returned [1,1,1,18], representing the number of positive, neutral, negative, and Unknown words, respectively. Similarly, when analyzed for Supra categories, it returned [0,0,0,0,0,0,0,0,21], representing the number of Benevolence, Malevolence, Well-being, Ill-being, Composure, Anxiety, Surprise, Non-Specific Emotion, Indifference and Unknown words respectively. The same approach was used for Super emotions and Basic emotions. Although the review is quite positive, the classifications returned by the polarity sweep do not fully capture the sentiment due to the limitations of the lexicon. This discrepancy highlights the challenges of accurately interpreting human emotions through word counting alone. This approach checks if each word of the review matches exactly with each word in every category presented above. However, this method severely limits the number

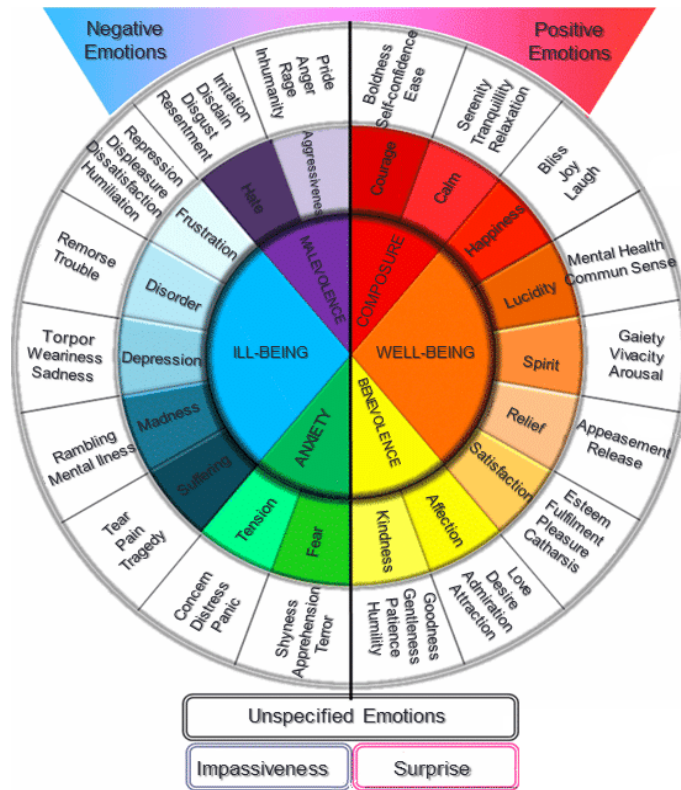


Fig. 6: Illustration of all the categories in the EMOTAIX lexicon and how they are organized.

of matches because the original lexicon only contains 3983 words. For instance, some examples of positive words present in the lexicon are "brilhante", "caloroso" and "completo". These words are present in the EMOTAIX.PT lexicon but their adverbial counterparts "brilantemente", "calorosamente" and "completamente" are not. Another example of missing words is that both masculine and feminine versions of confused exist in the DB "baralhado" and "baralhada" but only the male version of the word understanding "compreensivo" exists. These examples are only a tiny portion of the discrepancies present in the lexicon. Additional techniques were used to address these discrepancies and expand the possible matches to include more inflected forms of all words in the reviews. These techniques are shown in the following sections.

4.3.2 Levenshtein Distance

The Levenshtein distance is a popular technique that outputs the distance or difference between two given words using complex mathematical formulas. It consists of finding the minimum number of single-character modifications needed to convert one word into another. These edits can include insertions, deletions, or substitutions of individual characters [57] present in equation Eq. 1.

$$lev(a, b) = \begin{cases} |a| & \text{if } |b| = 0, \\ |b| & \text{if } |a| = 0, \\ lev(tail(a), tail(b)) & \text{if } a[0] = b[0], \\ 1 + \min \begin{cases} lev(a, tail(b)) \\ lev(tail(a), (b)) \\ lev(tail(a), tail(b)) \end{cases} & \text{otherwise,} \end{cases} \quad (1)$$

An alternative approach to identifying correspondences between the words present in the reviews and those within the EMOTAIX.PT lexicon was employed in this research. By calculating the distance or difference between words using the Levenshtein distance, potential partial matches between words in the review text and words in the EMOTAIX.PT lexicon were identified. Gender-related variations between words, characterized by a Levenshtein distance of one, are exemplified by instances such as "compreensivo" versus "compreensiva", "curado" versus "curada", and "curioso" versus "curiosa". Additionally, modifications in grammatical number are also captured, such as "brilhante" versus "brilhantes", "deliciado" versus "deliciados", and "ídolo" versus "ídolos". This method can also identify some errors, as demonstrated by the Portuguese review example mentioned above, where the word "referência" was misspelled due to an absent accent on the letter "e", resulting in a Levenshtein distance of one. These examples illustrate potential matches between words in the review text and words in the EMOTAIX.PT lexicon were identified using the Levenshtein distance.

By calculating the distance or difference between words, cases where words may have different gender or number variations but still refer to the same underlying concept were identified. Appendix Fig. 35 shows an example of a calculated Levenshtein distance of two between the words "party" and "park". The Levenshtein distance of these two words is two because two operations are needed to transform the word party into the word park: The first is a substitution that replaces the letter "T" with the letter "K". The second operation is the removal of the final "Y". After these two operations, it can be said that with only two Levenshtein modifications, the words "party" and "park" can be converted into one another.

Levenshtein distance is an exciting solution to the problem mentioned above, but a question arises: What is the ideal distance to consider for words to be equal? To answer this question, multiple Levenshtein distances ranging from 0 to 5 were tested to determine which provided the highest correlation with rating when used in conjunction with the Emotion detection method.

4.3.3 Minimum Equal Initial Characters

Using Levenshtein distance examples can help generalize the lexicon, but in some cases, errors are associated with this way of generalizing. For example, "moral" vs. "imoral", and "legal" vs. "ilegal" have a Levenshtein distance of 1, but the meanings associated with these words are entirely different. To take advantage of generalization and avoid its inherent problems, a minimum amount of equal initial characters was imposed to guarantee that the Levenshtein comparison was only performed if a minimum number of initial equal characters was guaranteed. When used with the Emotion detection method, this value was then varied to determine the highest correlation with rating.

4.3.4 Stemming

Stemming is predominantly used in NLP to transform a word into its base or stem form by removing any inflections or derivations [58]. This can be useful for increasing the generalization of words by applying stemming to both the review text and the EMOTAIX.PT lexicon, the odds of matching become much higher because every word is in its simplest form, and there is no need to account for all of its variations. To achieve this, a process was created to transform all words of every review text and all words of all categories of the EMOTAIX.PT lexicon into their stem equivalents to increase the number of matches. This was done using the nltk.stem library and the RSLPStemmer() function [46]. Stemming allows words like "maravilhoso" (wonderful), "maravilha" (wonder), and "maravilhosamente" (wonderfully) in Portuguese to be simplified to the base form "maravilh". Even though this form is not an actual word, it still holds broad meaning. However, sometimes, words derived from the same meaning have different stems. For example, the words "serviço" (service) and "servido" (served) have "servic" and "serv" as their respective stems, which is not ideal. In Table 3, multiple examples of words in their normal and stemmed forms are shown.

Table 3: Examples of original vs. stemmed words.

Nº	Original Word	Stemmed Word	Nº	Original Word	Stemmed Word
1	Informação	Inform	6	Sentados	Sent
2	Informativo	Inform	7	Cheguei	Chegu
3	Restaurante	Restaur	8	Chegamos	Cheg
4	Restauração	Restaur	9	Pedido	Ped
5	Sentar	Sent	10	Pedimos	Ped

This method was used to improve the accuracy of matching words in the review text with words in the EMOTAIX.PT lexicon by transforming all words into their base or stem form. By applying stemming to both the review text and the EMOTAIX.PT lexicon, this research can increase the odds of matching words.

4.3.5 Synonyms

Assuming all words are assigned to a correct category, one can infer that having a lexicon with more words within each category is always preferred. An attempt to try to expand the EMOTAIX.PT lexicon was performed by adding synonymous words to the lexicon. This endeavor was done by creating a Python process that searched for all synonyms of each word of every category using `pysinonimos.sinonimos` library using the `Search()` function [59]. All repeating words were removed after all synonyms were added to their respective category. This procedure was made for the Polarity Supra Super and Basic categories. This resulted in four files, one for each category granularity level, containing original and synonymous words. This process increased the matches between words in the review text and words in the EMOTAIX.PT lexicon by 6.9 times. This increased the total number of words in the EMOTAIX.PT lexicon from 3975 to 27557.

4.4 SentiStrength Method

The alternative solution is SentiStrength, which is a program that performs SA to estimate the positive and negative emotions present in the text. A Microsoft Windows version of SentiStrength was used to determine the positive and negative scores of every review [60]. This evaluation was carried out using the SentiStrength Portuguese dictionary for both Portuguese reviews and English reviews translated to Portuguese and the SentiStrength English dictionary for English reviews and Portuguese reviews translated to English. This was done to obtain scores in both languages, as the self-reported performance of SentiStrength is better using the English dictionary [60].

This method estimated the positive and negative emotions in each review by evaluating each review using Portuguese and English dictionaries. This research will compare SentiStrength’s performance in different languages and analyze how different emotions may be expressed in reviews on Zomato’s platform. The SentiStrength platform utilizes a Graphical User Interface (GUI) that allows users to tune multiple parameters. After multiple tests, the optimal parameters for maxi-

mum correlation between rating and SentiStrength score were determined. These were the default SentiStrength parameters with the following changes:

1. Enable Exclamation mark as +2
2. Export as TripAdvisorScale.
3. Include incorrect spelling
4. Disable report for classifications and translations.

In terms of the SentiStrength score output, it appears that there is a higher correlation with ratings when using the English dictionary for both English reviews and Portuguese reviews translated into English. This is in comparison to using the Portuguese dictionary for Portuguese reviews and English reviews translated into Portuguese. This observation supports SentiStrength’s claim of superior performance when utilizing the English language.

With the optimal SentiStrength settings in place, two additional columns, positive and negative, were generated and added to the DB. These columns represent the positiveness and negativeness of each review on a 5-level scale, with values ranging from 1 to 5, as illustrated in Fig. 7. These results suggest that the SentiStrength program performs better when using an English dictionary than a Portuguese dictionary.

4.5 Relationship of emotion and rating

A clustering analysis was performed in which all reviews were depicted in a 3D space. This depiction was predicated on the ratio of EMOTAIX positive, negative, and neutral words in relation to the rating. The representation indicates that the data becomes more distinguishable when negative and neutral words are utilized rather than positive ones, as illustrated in Fig 8.

The scatter plots exhibit a more apparent separation of rating using negative and neutral words and not as much when comparing the positive words. This visualization provides a comprehensive understanding of how different facets of reviews contribute to the overall rating. Furthermore, a 3-dimensional visualization of the cluster centroids was generated, which can be seen in Fig 9.

Building on the insights from the 3D visualizations and cluster centroids, elbow plots were employed to further refine the understanding of the optimal number of clusters. These plots provide a graphical representation of the Sum of Squared Errors (SSE) for varying numbers of clusters.

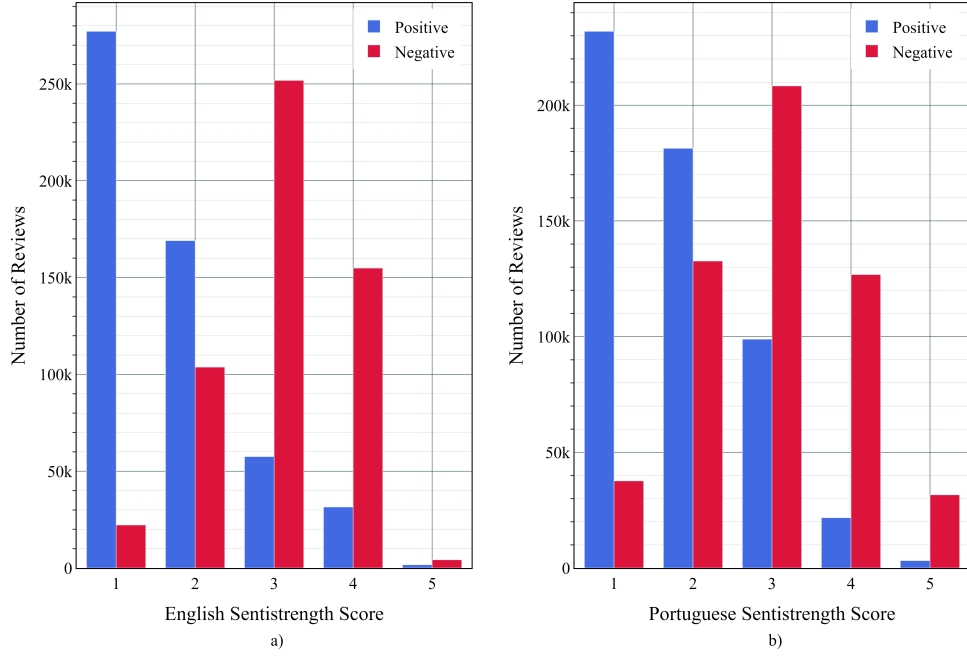


Fig. 7: Distribution of SentiStrength scores across all reviews.

a) Positive (in blue) and negative (in red) scores for English reviews.

b) Positive (in blue) and negative (in red) scores for Portuguese reviews.

They serve as a valuable tool for pinpointing the juncture at which the addition of more clusters results in only a minimal decrease in SSE. This juncture signifies an optimal balance between the compactness of the clusters and their quantity.

Figure 10 shows the elbow plot for using the polarity categories across all reviews. The red circle indicates the point with the highest change in slope, which might superficially suggest that one cluster could be considered. However, using one cluster would be meaningless. Instead, the green and purple circles show that two and three clusters are reasonable options due to the noticeable changes in slope at these points, providing a more detailed and informative segmentation of the data.

This result indicates that using two or three clusters may be the ideal number when using the EMOTAIX.PT lexicon to categorize positive, neutral, and negative emotions.

4.6 Results

After multiple tests, the ideal amount of minimum equal initial characters and Levenshtein distance that maximized correlation with rating was 6 and 1, respectively. It was also noted that stemming and using synonyms made the correlation worse. Using the aforementioned ideal values, it was noted that most reviews had few identifiable positive, neutral, and negative words and many unknown

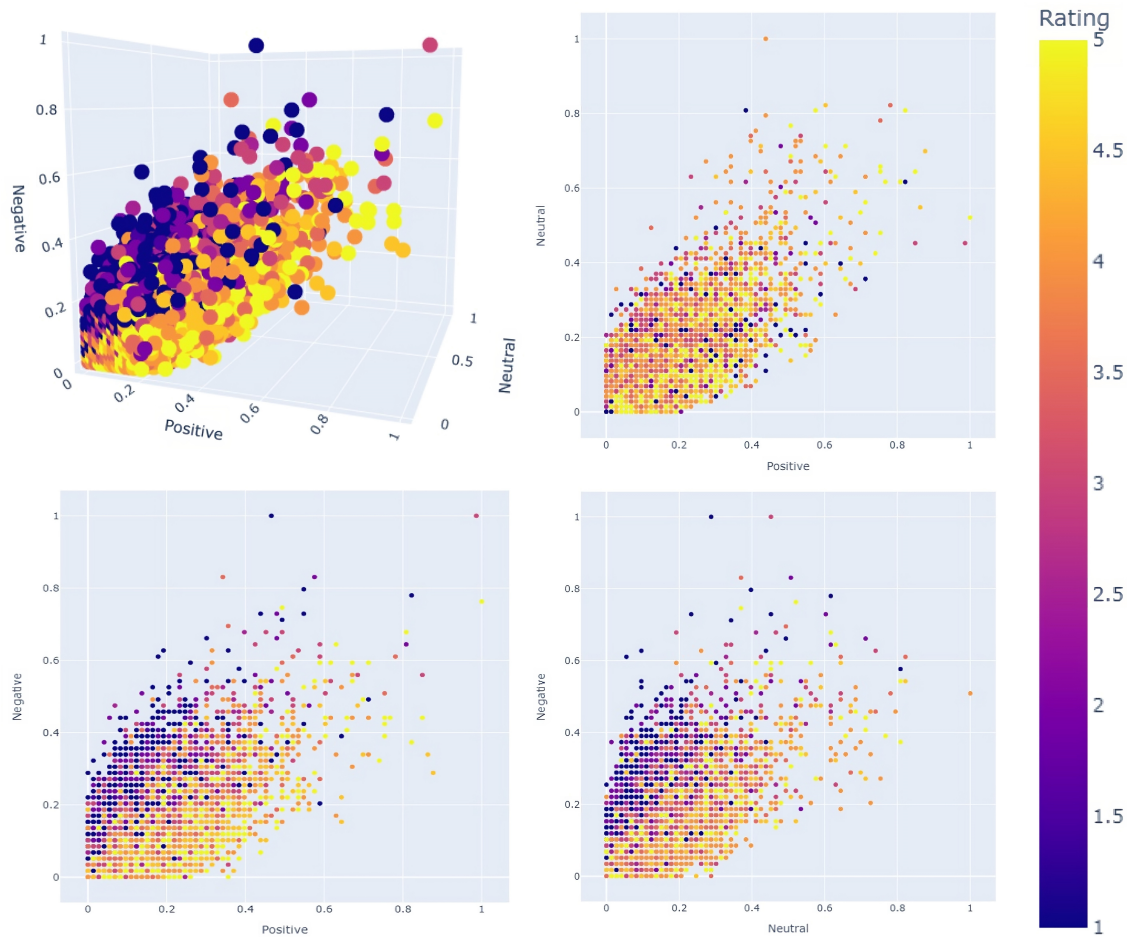


Fig. 8: 3D scatter plot illustrating the correlation between review ratings and word sentiment. The color scheme represents different ratings, and the axes labeled as 'Positive', 'Neutral', and 'Negative' denote the relative number of words in each respective category. The plot is divided into four quadrants. The top-left quadrant provides a 3D perspective, while the remaining three quadrants each showcase a unique 2D combination of 'Negative', 'Neutral', and 'Positive' sentiments.

words. It is also possible to observe that negative words are used less than neutral and positive words. This confirms what the literature says about negative words/emotions being used more sparingly than positive, which confirms the skewed nature of the data as shown by Fig. 11.

4.7 Conclusion

This chapter delved into the intricacies of emotion detection within restaurant reviews, employing a variety of techniques to enhance the accuracy and relevance of the analysis. The research began by exploring the use of the EMOTAIX.PT lexicon, a comprehensive tool for identifying and categorizing emotions in Portuguese text. To address the limitations of exact word matching, several methods were tested, including Levenshtein distance, minimum equal initial characters, stemming, and the addition of synonyms.

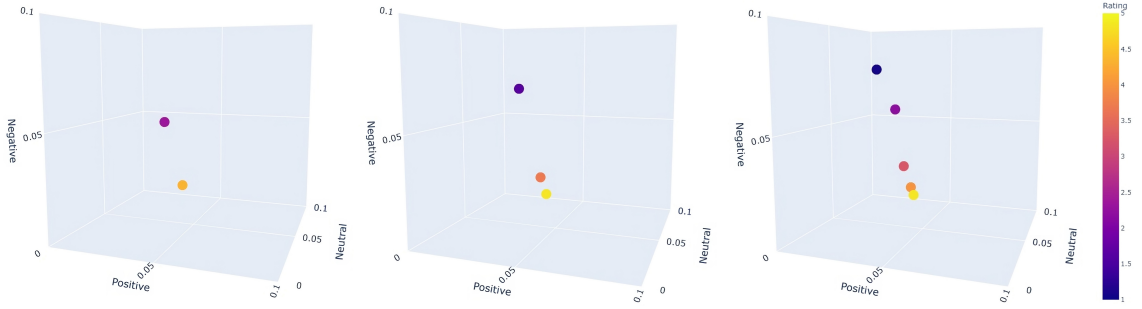


Fig. 9: Displayed from left to right are three individual 3D scatter plots, each representing the centroids of clusters when using 2, 3, and 5 clusters, respectively. The color scheme corresponds to different ratings, while the axes denote the relative number of 'Negative', 'Neutral', and 'Positive' words.

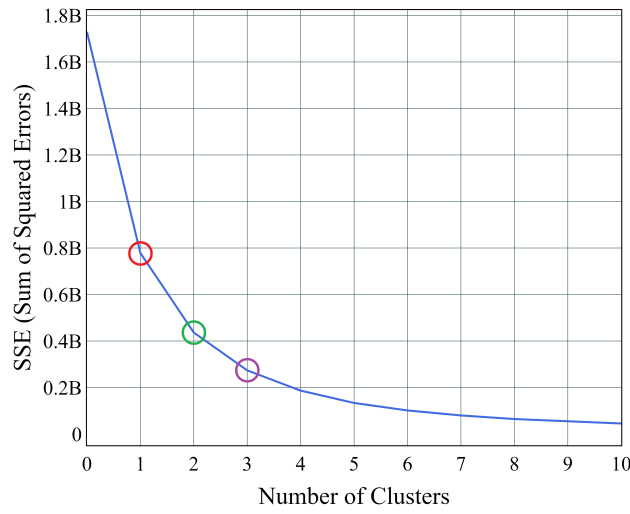


Fig. 10: Elbow plot using SSE for positive, neutral, negative, and unknown words across all reviews, derived from the EMOTAIX.PT lexicon with a maximum Levenshtein distance of one and a minimum number of equal initial characters of six across different numbers of clusters. (Red circle: Highest change in slope. Green circle: The second highest change in slope is the Purple circle: Third highest change in slope).

The optimal Levenshtein distance was found to be 1, and the minimum equal initial characters were set to 6, which maximized the correlation with review ratings. However, the application of stemming and synonym expansion did not yield better results, indicating that these techniques might lead to overgeneralization in this context.

An alternative approach using the SentiStrength program was also evaluated. SentiStrength provided positive and negative emotion scores for each review, and it was observed that the English dictionary performed better than the Portuguese dictionary, even for reviews originally written in Portuguese. This finding underscores the importance of language choice in SA tools.

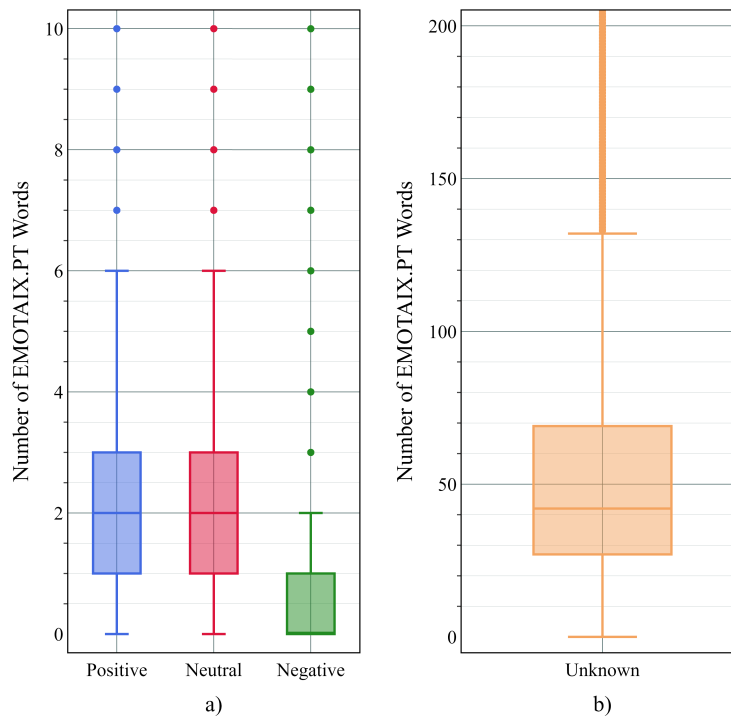


Fig. 11: Distribution of the number of positive, neutral, negative, and unknown words across all reviews using EMOTAIX.PT lexicon with maximum Levenshtein distance set to one and minimum amount of equal initial characters set to six.

- a) Positive, neutral, and negative words box plot.
b) Unknown words box plot.

The relationship between emotion and rating was further examined through clustering analysis and 3D visualizations. The data showed a clearer separation of ratings when negative and neutral words were considered, rather than positive ones. Elbow plots were used to determine the optimal number of clusters, suggesting that two or three clusters provided a more meaningful segmentation of the data.

In summary, this chapter tackled the challenges in emotion detection for restaurant reviews. The EMOTAIX.PT lexicon and specific text processing techniques were effective, but tool choice significantly impacted results, with SentiStrength performing better using the English dictionary. These insights enhance understanding of customer sentiments and improve SA platforms.

5 Feature Selection

5.1 Introduction

The primary objective of this chapter is to predict review ratings by identifying the most relevant features that significantly influence these predictions. To achieve this, we employed various feature selection methodologies, including correlation analysis, the mRMR method, and the BC method. This chapter presents the outcomes derived from the dataset analysis and the feature selection process. The analysis includes an in-depth exploration of the correlation between various variables and the review ratings. The feature selection process aims to pinpoint the most pertinent features that impact the prediction of review ratings. The results section discusses findings from numerous tests, such as the influence of stemming and synonyms on correlation, the distribution of words with positive, neutral, and negative connotations in reviews, and the application of elbow plots for cluster analysis. The feature selection section details how correlation and mRMR methods were used to identify crucial features that affect review ratings. The chapter concludes by addressing how these findings can enhance the understanding of review ratings and potentially improve future predictive models. It also explains how the correlation, mRMR, and BC methods were integrated to identify the most critical features influencing review ratings.

5.2 State of the art

In 2020, Erenel et al. [61] proposed a novel feature selection scheme for emotion recognition from text. Their approach focused on the utilization of moderately frequent terms during term selection, arguing that these terms may carry valuable discriminative information. The authors demonstrated that their proposed scheme outperformed conventional filter-based feature selection measures like Chi-Square and Gini-Text in many cases. This work represents a significant advancement in the field of feature selection for emotion recognition, highlighting the potential of moderately frequent terms in improving classification performance.

In a 2015 study by Claypo et al. [62], the researchers explored the use of K-Means clustering combined with Markov Random Field (MRF) feature selection for analyzing Thai restaurant reviews. The process began with text preprocessing, where the reviews were split into individual words, and stop words were removed. Next, the text was transformed to extract keywords and create input vectors for analysis. MRF feature selection was applied to identify the most relevant

features, which helped significantly reduce the time required for computation. The reviews were then classified into positive and negative categories using K-Means clustering. The study found that this approach outperformed other clustering methods, such as Self-Organizing Maps, Fuzzy C-Means, and Hierarchical Clustering, demonstrating superior performance in classifying the reviews.

Madasu et al. [63] conducted a comprehensive analysis of various feature selection methods used in SA. They start by extracting features using TF-IDF. They then explore the effectiveness of different feature selection techniques, including Chi-Square, Generalized Sequential Sample (GSS), and Bi-Normal Separation (BNS). These features are subsequently used to train several machine learning models, such as Logistic Regression (LR), SVM, DT, and NB. To boost performance, the study also incorporates ensemble methods like Bagging and Random Subspace. The findings reveal that combining feature selection with ensemble techniques leads to significantly better performance compared to NLs, with the added benefits of reduced training time and simpler models that do not require extensive hyperparameter tuning.

Sharm et al. [64] explored how different feature selection techniques and machine learning models impact SA and consumer behavior, specifically focusing on university students' fast food preferences in Bangladesh. Using data collected from 170 participants via a structured questionnaire, they tested various supervised machine learning models, such as Gaussian NB, Classification and Regression Trees (CART), RF classifier, and LR. Their findings revealed that the NB model had the highest accuracy, correctly predicting fast food consumption 79.4% of the time, showing its effectiveness in classifying preferences based on chosen features. Additionally, their comparison of feature selection methods for SA, especially in the context of online movie reviews, found that the Gain Ratio method was superior, with SVM achieving the best overall classification results. These results highlight the importance of specific machine learning techniques in SA and offer practical insights for the fast food industry, emphasizing the need for targeted marketing strategies to meet the preferences of young consumers.

Claypo et al. [65] introduced an opinion-mining model for classifying Thai restaurant reviews. They combined a Multi-Layer Perceptron (MLP) NN with the mRMR feature selection method. Using mRMR, they reduced the dataset's size by selecting maximally dissimilar features, improving the NL performance. The experimental results showed that the optimized MLP NN model achieved

high accuracy in categorizing reviews as positive or negative, surpassing other methods like RBF and SVM. Notably, this approach enhanced classification accuracy and reduced computational time, making it a valuable tool for analyzing customer opinions.

Emerson et al. [66] delve into the complexities of the BC voting system, particularly examining its behavior when partial voting is permitted. The traditional BC assigns points from the last choice in descending order, but modern versions often use a scale like $(n-1, n-2, \dots, 0)$. Emerson points out that partial voting can significantly impact the results, sometimes making the BC function similarly to a plurality vote if voters only rank a single candidate. The study argues that Jean-Charles de Borda’s original approach, which starts point allocation from the last preference, might be more valuable than previously thought, as it encourages complete voter participation and limits potential manipulation.

5.3 Features

All previous methods for extracting emotion produced new data, which was subsequently incorporated into the central database. Appendix Table 15 presents all the provided and generated features combined. Two methods, mRMR and Spearman’s correlation, were used to identify the most critical features, and they were subsequently combined using the Borda method. The following sections will explain how these methods work and how they were integrated.

5.3.1 Correlation

Correlation is a statistical technique that determines the relationship between two variables, assessing their strength and direction. A positive correlation between two variables indicates that as one variable increases, the other also increases. Conversely, a negative correlation means that as one variable increases, the other decreases [67]. Correlation values range from -1 to 1, with the former indicating a perfect negative correlation, the latter showing a perfect positive correlation, and the in-between value, 0, indicating no correlation at all. The closer the correlation is to -1 or 1, the stronger the relationship between the two variables.

Pearson’s, Kendall’s, and Spearman’s correlations were evaluated to account for linear and non-linear relationships, normal and non-normal distributions, and ordinal and non-ordinal scales. Since most variables in the dataset did not exhibit a normal distribution, Spearman’s Correlation was chosen. This non-parametric method measures the strength and direction of a monotonic

relationship between two variables without requiring normality or linearity [68]. As such, it is an appropriate choice for datasets where the normality assumption is not satisfied. Fig. 12 shows a Spearman's correlation matrix between all remaining features after removing those with a high amount of Nulls. The correlation matrix was implemented using Python with the pandas library and the `corr()` method [69].

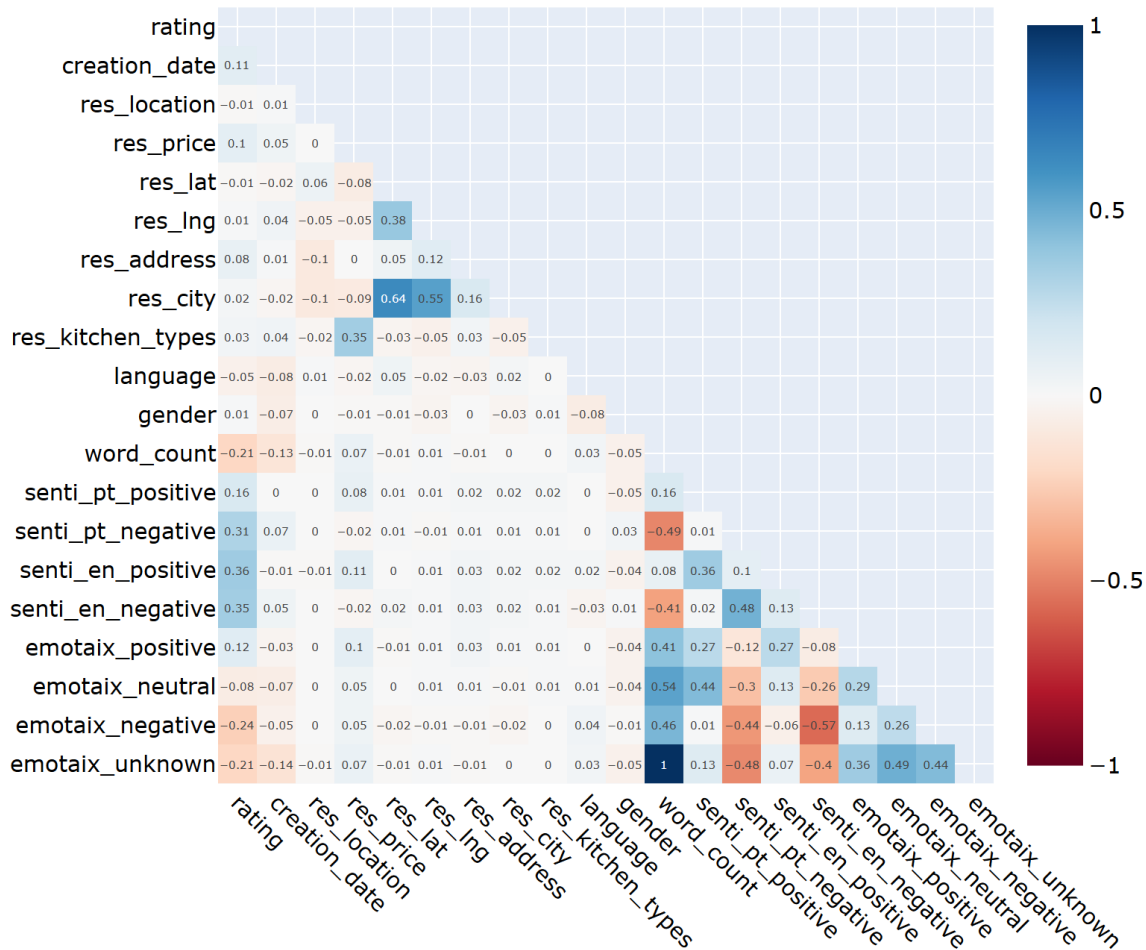


Fig. 12: Spearman correlation matrix for all features excluding those determined to be irrelevant in earlier processes and those with a high amount of Nulls.

5.3.2 mRMR Method

mRMR [70] is a feature selection method that aims to find a subset of features that are both relevant to the prediction task and maximally diverse from one another. mRMR aims to select features that reduce redundancy and increase relevance relative to a target [71]. This means that redundant values, such as age and date of birth, which can be derived from one another, could contribute to predicting a certain target feature. Still, since they are redundant, only one would

be considered by mRMR to minimize redundancy and maximize relevance. The mRMR feature selection method was implemented using a Python process, using the library [72].

5.3.3 Borda Count Method

BC is a voting method where voters rank candidates in order of preference. Each candidate earns points based on their position in each voter’s ranking [73]. For example, in an election with ten candidates, the top-ranked candidate on a ballot gets 9 points, the next gets 8 points, and so on, with the lowest-ranked candidate receiving 0 points. The candidate with the most total points wins.

This method aims to find a candidate who is broadly acceptable to the majority rather than just the most popular choice. It is a way to ensure that the winner has widespread support across all voters.

5.4 Results

The features with the highest absolute correlation suggest that the words used in the reviews are correlated with the review rating. Sentistrength dominates the top positions in the correlation matrix, and it can be observed that the English Sentistrength scores have a higher correlation when compared to their Portuguese counterparts. Despite this, Sentistrength was only used to compare with the custom-developed emotion detection algorithm presented in Chapter 4 and as a benchmark to evaluate its performance. This was done because Sentistrength requires a paid commercial license to be used for commercial purposes. To avoid any paid services, it was decided not to use Sentistrength. Therefore, the top 10 essential features to consider according to Spearman’s correlation score are senti_en_positive, senti_en_negative, senti_pt_negative, emotaix_negative, word_count, emotaix_unknown, senti_pt_positive, emotaix_positive, creation_date and res_price.

In the mRMR method, Sentistrength dominates the top positions again, but the most important feature is the creation_date. This suggests that the time of review creation significantly impacts the review rating, possibly due to factors such as changes in restaurant service quality over time or temporal trends in user rating behavior. Interestingly, the English Sentistrength scores appear to be more influential than their Portuguese counterparts. This is likely because Sentistrength’s inference is better in English, and English is less complex, especially in terms of the number of

verbal tenses. The `emotaix_negative` and `emotaix_unknown` features, derived from the custom-developed emotion detection algorithm, are also among the top 10 features. This highlights the value of custom SA tools in capturing nuanced emotional expressions that can influence review ratings.

Practical details such as the `res_address` and `res_price` also appear in the top 10, suggesting that logistical and financial aspects of the restaurant experience can significantly impact user ratings. The language feature’s presence among the top 10 could indicate that the review’s language might influence the review rating, possibly due to cultural or linguistic differences in rating behavior.

Overall, the mRMR method provides a balanced selection of features that captures various aspects of the review and restaurant experience, highlighting the complexity of factors that can influence user ratings.

The Borda method was employed to combine the two feature selection methods. In Table 4, the left and center columns list all the features considered from both the mRMR and Spearman correlation methods, ranked from best to worst, along with their respective Borda points. The right column shows the features sorted from best to worst according to the combined Borda score.

The top 10 essential features to consider according to the Borda score are `senti_en_positive`, `senti_en_negative`, `emotaix_negative`, `senti_pt_negative`, `creation_date`, `senti_pt_positive`, `emotaix_unknown`, `word_count`, `res_address`, and `res_price`. This order indicates that sentiment-related features are indeed the strongest predictors of review rating. Unsurprisingly, the English Sentistrength scores outperform the Portuguese ones. The creation date is an unexpectedly strong predictor, as is the word count. The presence of `emotaix_negative` and `emotaix_unknown` among the top features highlights that the custom-developed method is a good predictor, though not as effective as Sentistrength.

Additionally, practical details such as the restaurant address and price contribute significantly to the prediction model, suggesting that logistical and financial aspects play a role in shaping user ratings. Review language and reviewer gender are positioned lower and thus contribute less compared to the other features. This is expected, as the sentiment and emotional tone of the review are more directly related to the review rating than the language or gender of the reviewer. Furthermore, these are predicted features and likely to have an associated error.

Table 4: Ranking of features based on mRMR and Spearman correlation methods, along with their Borda points. The final Borda score column combines these rankings to provide an overall feature importance ranking from best to worst.

mRMR			Spearman			Borda	
Rank Lower is better	Feature	Borda Points	Absolute Correlation Rank Higher is better	Feature	Borda Points	Feature	Score
1	creation_date	18	0,36	senti_en_positive	18	senti_en_positive	35
2	senti_en_positive	17	0,35	senti_en_negative	17	senti_en_negative	32
3	emotaix_negative	16	0,31	senti_pt_negative	16	emotaix_negative	31
4	senti_en_negative	15	0,24	emotaix_negative	15	senti_pt_negative	30
5	senti_pt_negative	14	0,21	word_count	14	creation_date	28
6	senti_pt_positive	13	0,21	emotaix_unknown	13	senti_pt_positive	25
7	res_address	12	0,16	senti_pt_positive	12	emotaix_unknown	22
8	language	11	0,12	emotaix_positive	11	word_count	21
9	res_price	10	0,11	creation_date	10	res_address	20
10	emotaix_unknown	9	0,11	res_price	9	res_price	19
11	gender	8	0,08	res_address	8	language	17
12	word_count	7	0,08	emotaix_neutral	7	emotaix_positive	17
13	emotaix_positive	6	0,05	language	6	emotaix_neutral	10
14	res_city	5	0,03	res_kitchen_types	5	res_city	9
15	res_kitchen_types	4	0,02	res_city	4	res_kitchen_types	9
16	emotaix_neutral	3	0,01	res_location	3	gender	8
17	res_lng	2	0,01	res_lat	2	res_location	4
18	res_location	1	0,01	res_lng	1	res_lng	3
19	res_lat	0	0,01	gender	0	res_lat	2

5.5 Conclusion

In this chapter, feature selection was conducted using mRMR and Spearman correlation methods to identify the most influential features for predicting review ratings. The mRMR method aimed to find features that are both relevant to the prediction task and maximally diverse from one another. On the other hand, Spearman’s correlation assessed the strength and direction of the relationship between two variables.

The results from these two methods were combined using the BC method. This method is a consensus-based voting system that ranks candidates based on their position in each voter’s ranking, providing an overall feature importance ranking.

The analysis revealed that sentiment-related features, particularly those derived from English Sentistrength scores, were the strongest predictors, emphasizing the role of emotional tone in reviews. The timing of the review (creation date) and length (word count) also significantly influenced ratings.

The custom-developed emotion detection algorithm, described in Chapter 4, demonstrated the potential to capture nuanced emotions. Practical aspects like the restaurant's address and price were important, indicating the impact of logistical and financial factors on ratings.

Interestingly, the language of the review and the reviewer's gender contributed less to the prediction model, suggesting that the review's sentiment and emotional tone are more directly related to the review rating.

6 Unsupervised Topic Detection Enhanced by Human Labeling

6.1 Introduction

Building on the insights gained from emotion detection, this chapter delves into the extraction of meaningful topics from large volumes of unstructured text data. One such application is in the field of restaurant reviews, where understanding customer sentiment and the topics they discuss can provide valuable insights into a restaurant’s performance. This chapter presents a method for detecting topics in restaurant reviews using LDA, an unsupervised learning algorithm.

The method involves several steps, including data preparation, tokenization, lemmatization, vectorization, and finally, running the LDA model. Each step is designed and implemented to ensure the effectiveness of the topic detection process. The method also addresses several challenges, such as the selection of an optimal number of topics and others. After running the LDA model, the topics were labeled by human annotators in order to attribute meaningful interpretations to each topic. This human-in-the-loop approach ensures that the topics identified by the model align with human understanding and intuition.

The ultimate goal of this method is to provide a more detailed analysis of a restaurant’s performance by identifying specific subjects that customers discuss in their reviews. This approach goes beyond merely understanding the polarity of a review and delves into the nuances of the topics discussed, thereby offering a more comprehensive understanding of customer feedback.

6.2 State of the Art

Zeynab et al.’s [74] review on Twitter topic detection reveals a preference for unsupervised clustering methods in identifying new and unspecified topics. These methods are capable of exploring content without needing pre-labeled data. Specified methods that apply SL to filter tweets by criteria like location or keywords require intensive manual labeling. Despite the computational challenges faced by topic modeling in real-time detection, Formal Concept Analysis (FCA) methods, being fully unsupervised, show promise for topic detection tasks.

Garcia et al. [75] applied a combination of machine learning and NLP techniques to dissect a substantial dataset of Twitter posts regarding COVID-19. They revealed ten principal topics across English and Portuguese tweets from USA and Brazil, with seven topics being shared between the languages. Their findings, dominated by negative sentiments across the identified topics, underscore

the utility of integrating topic detection and SA for deep analysis of extensive datasets. This offers a methodological blueprint for subsequent research in data analysis.

Another micro-blogging website, Sina Weibo, was used to identify and extract seasonally distributed food-related topics, similar to Twitter data, by applying text clustering and DL [74]. For understanding insights for hotel management, it employs LDA to analyze sentiments in eco-hotel reviews [76]. This text-mining approach reveals that different aspects of the hotel, such as food and hospitality, elicit primarily positive sentiments among guests. Through LDA, the study identifies vital topics characterizing guests' experiences and feelings.

Menner et al. [77] compared different unsupervised topic detection approaches, including the identification of frequent nouns and verbs, keyword clustering, and Latent Semantic Indexing (LSI) for evaluating reviews for hotels. Their research found that the keyword clustering approach with an 80% accuracy achieved the highest accuracy. However, Named Entity Recognition (NER) emerged as the most effective approach, demonstrating a balanced and high-quality topic detection capability. This has the potential to identify previously unrecognized topics and facilitate a deeper understanding of customer discussions in tourism.

Cataldi et al. [78] introduces an innovative approach to analyzing user-generated reviews by identifying sentiment at the feature level using NLP and statistical analysis. It starts by extracting critical features from reviews and then models each review as a lexical dependency graph to understand the relationships between terms. Sentiment polarities are assigned to WordNet synsets for precise meaning, allowing for the computation of feature-specific sentiments. This method was validated against human assessments, demonstrating high precision in detecting sentiments towards specific aspects of products and services. This advancement provides nuanced insights into consumer opinions, which is beneficial for both consumers and service providers.

6.3 Unsupervised ML Topic Detection Enhanced by Human Annotation

From the literature review, it is clear that the employment of unsupervised learning algorithms as a preliminary step in the data analysis process presents a cost-effective strategy for managing the complexities inherent in the manual labeling of expansive datasets. This computational approach facilitates the rapid processing and categorization of large volumes of textual data into preliminary topic clusters without the need for predefined labels, significantly mitigating the initial workload

and associated financial burdens. Such a methodology is indispensable for the efficient handling of the continuously increasing quantities of digital textual content, enabling a scalable framework for initial topic identification and pattern recognition within substantial datasets.

However, unsupervised topic labeling faces significant challenges primarily due to the absence of predefined labels that guide the algorithm in identifying coherent and meaningful topics from raw text data. In addition, automatically generated topics may not correspond to coherent or easily interpretable themes without human intervention. Assigning meaningful labels to these topics often requires domain knowledge and manual review.

Therefore, a collaborative topic detection and labeling process is proposed to solve those problems by combining the computational power and scalability of unsupervised learning with the nuanced understanding and flexibility of human annotation; collaborative topic detection and labeling can create highly effective topic detection systems that are both broad in scope and precise in their categorizations. This hybrid approach is particularly well-suited to the complexities of natural language and the diverse requirements of different NLP applications. An overview of the Unsupervised Topic Detection Enhanced by Human Labeling is shown in Fig 13. There are two main parts of the labeling process for this proposed topic. The first is the unsupervised topic detection method, followed by manual labeling with a survey, which contains internal labeling and a survey.

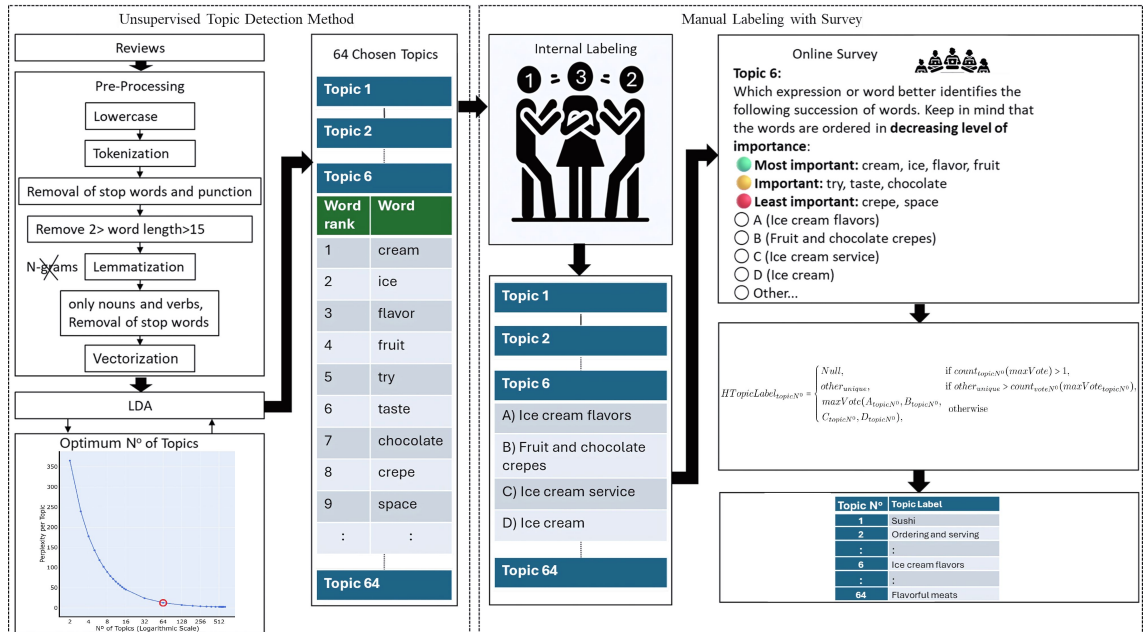


Fig. 13: Diagram of Unsupervised Topic Detection Enhanced by Human Labeling.

6.3.1 Unsupervised Topic Detection Method

LDA is an unsupervised learning algorithm that identifies topics within a text. It operates by taking a predetermined number of topics and a collection of documents as input and outputs a set of words associated with each topic. The algorithm functions by examining the co-occurrence of words within and across documents. It iteratively assigns each word in each document to one of the topics based on the probability that the word belongs to that topic given the other words in the document and their topic assignments. The algorithm updates these probabilities iteratively until it converges on a stable set of topic assignments [79].

When applied to a collection of restaurant reviews, LDA enables the identification of underlying topics being discussed. This can yield valuable insights into the specific subjects that customers are talking about, thereby facilitating a more detailed analysis of a restaurant's performance. Upon convergence, LDA provides a set of words associated with each topic, ordered by relevance. These words represent the most likely words to appear in documents discussing that topic. However, this method does not provide the name of the topic itself, only the words that are most likely associated with it. The implementation of the LDA method was done using Python. The steps involved in this process are described below.

Data Preparation: The method employed in this study utilized the English-translated versions of the reviews. The libraries used for tokenization and lemmatization were specifically chosen due to their superior performance with the English language, which is less complex and has fewer verb tenses compared to other languages. Furthermore, these techniques were primarily developed for English, making it the optimal choice for topic detection. Prior to the application of tokenization and lemmatization, the reviews underwent a pre-processing stage. This involved the removal of punctuation, special characters such as '@', double spaces, and newline characters. The elimination of stop words was not conducted at this stage as this is incorporated into the lemmatization process, which will be elaborated on subsequently.

Tokenization: Tokenization is a typical pre-processing step in NLP tasks. The goal of tokenization is to split text into individual tokens that can be used as input for other NLP tasks. These tasks include part-of-speech tagging, parsing, text classification, and topic detection [80]. In this study, the tokenizing process was performed using the Gensim library [81] [82]. After this

process, each review was converted into a list of lowercase words. These words were no smaller than two characters and no longer than 15 characters, with all punctuation removed.

N-grams: An n-gram is a contiguous sequence of n-words from a given text. For example, in the sentence "The quick brown fox jumps over the lazy dog," the bigrams (2-grams) are "The quick," "quick brown," "brown fox," and so on, while the trigrams (3-grams) are "The quick brown," "quick brown fox," "brown fox jumps," and so on. This study tested the inclusion of bigrams and trigrams by generating all possible instances and analyzing their frequency in the review text. If the frequency of a bigram or trigram fell within a predefined range, it was added to the list of words, allowing the topic detection model to identify commonly co-occurring words. However, after testing multiple maximum and minimum frequency thresholds, it was found that including bigrams and trigrams consistently resulted in worse performance than using only unigrams. As a result, no bigrams or trigrams were used in this study.

Lemmatization: Lemmatization is the process of converting words into their base or root form, effectively removing inflections. This procedure is a standard pre-processing step in NLP tasks and aims to normalize the text. Lemmatization takes into account a word's context and part of speech to determine its lemma, which is the dictionary form of the word. The Gensim library [81] [82] was utilized for this process, which was configured to allow only nouns and verbs. By excluding adjectives and adverbs, the level of noise in the data was significantly reduced, leading to an enhancement in the performance of the topic detection model. In addition to the above, the lemmatization process also took care of the removal of stop words. This step further contributed to reducing noise and improving processing efficiency by significantly reducing the number of words to be analyzed by the model.

Vectorization: Vectorization is a standard process in NLP. It involves converting ML algorithms. This is achieved by representing each word, phrase, or document as a vector of numbers, where each number represents a specific feature or characteristic of the text. Several techniques exist for vectorizing text data, including bag-of-words, TF-IDF, and word embeddings. The bag-of-words method used the CountVectorizer class from the scikit-learn library [79]. This class has multiple parameters, but the most important ones that were tested are as follows:

- **min_df = 1:** This parameter sets the minimum number of times a word must appear in the reviews. It was set to 1, meaning that words appearing in at least one review were considered.

- **max_df = 1.0**: This parameter sets the maximum number of times a word can appear in the reviews. It was set to 1.0, meaning that words appearing in up to 100% of the reviews were considered.
- **stopwords = "english"** : This parameter specifies the language for which stopwords should be removed. It was set to 'english' to remove English stopwords.
- **token_pattern = "[a-zA-Z0-9]{3,}"**: This parameter specifies the pattern for tokenizing the text. It was set to only consider words with at least three characters, consisting of letters and numbers, and avoid other characters.

LDA Model: The final stage of the topic detection process involved running the model with the pre-processed data from the previous steps. This stage presented several challenges due to the multitude of parameters requiring adjustment and the absence of a one-size-fits-all solution. To address these challenges, each parameter was meticulously analyzed to comprehend its influence on model performance, and various parameter values were experimented with to identify initial patterns and trends. The number of topics (`n_components`), which dictated the number of topics the model would extract from the data, was one of the most critical parameters.

Another significant challenge was the malfunctioning of the `perp_tol` parameter in sklearn's LDA library [79]. This function, which serves as a form of early stopping to prevent overfitting by halting training when the model's learning rate diminishes with each iteration [83], led to two potential scenarios: either the model would run until it reached the maximum number of iterations, or it would converge for a few iterations before diverging. As a result, early stopping had to be manually implemented by assessing the model's perplexity.

Perplexity is a measure of how well a probability model predicts a sample. In the context of topic modeling, it quantifies how accurately the model represents the data. Lower perplexity indicates a better fit to the data, meaning the model is more accurate in predicting the topics within the text. Essentially, perplexity measures how "surprised" the model is by the test data; lower values indicate less surprise and thus a better model. By manually assessing the perplexity for each iteration, the optimal number of iterations was determined by selecting the iteration that minimized perplexity.

Once the optimal number of iterations for each parameter combination was determined, the model was rerun with a fixed number of iterations equal to that value. This process was repeated for various combinations of `learning_method`, `n_components`, `learning_decay`, `doc_topic_prior` and `topic_word_prior`. Owing to the manual nature of this task, automated methods such as grid search [84] or genetic algorithm [84] were not employed to identify the best parameters. It should be emphasized that while the grid search method was indeed experimented with, it was ultimately not adopted for use in this process.

The graphs in Fig 14 a) and b) show the overall perplexity across a different number of topics and the perplexity per topic across the same number of topics, respectively.

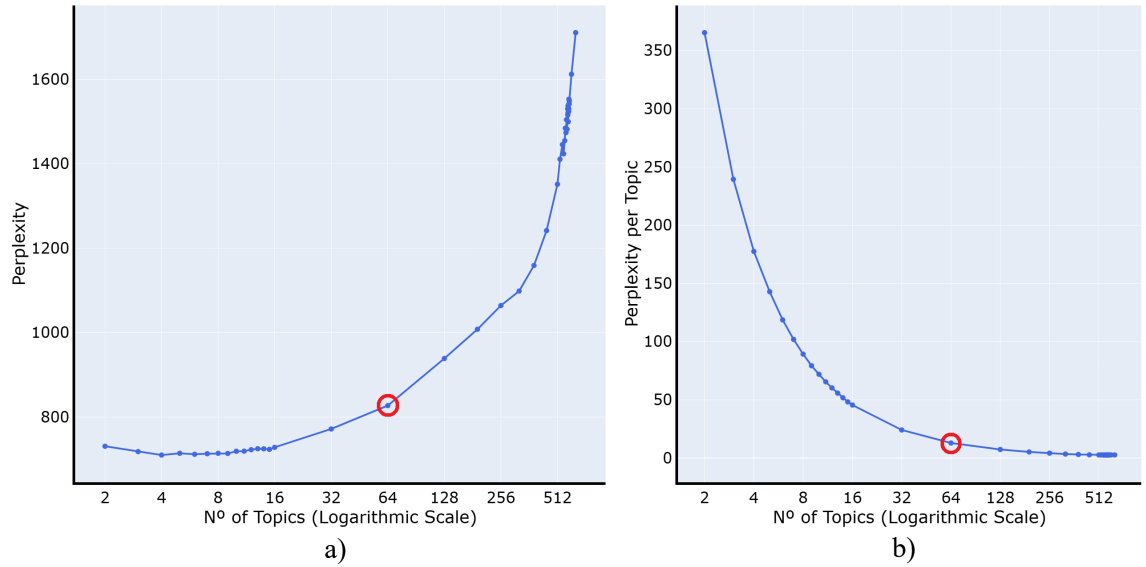


Fig. 14: **a)** Overall perplexity across the number of topics. **b)** Perplexity per topic across the number of topics.

When analyzing Fig.14, several potential paths were considered. One option was choosing the model with minimum perplexity, which had four topics and a perplexity of 710. This model would have had the advantage of having the lowest overall perplexity, indicating a good fit to the data. Another option was choosing the model that minimized perplexity per topic, which had 584 topics and a perplexity of 1500. This model would have had the advantage of having many topics, allowing for more fine-grained distinctions.

After careful consideration, a model with 64 topics was chosen instead of 4 or 584. This decision was made for several reasons. Firstly, both four and 584 topics were considered too few and

too many, respectively. Having only four topics would likely result in overly broad and general topics that would not capture nuances of data. On the other hand, having 584 topics would likely result in excessively specific or random topics that would not be meaningful or useful for human interpretation.

Secondly, when examining words associated with each topic in both the four-topic and 584-topic models, it was found that words related to the four-topic model were very general and lacked meaning. At the same time, words associated with the 584-topic model were sometimes too specific or random. This further reinforced the decision to choose a model with a moderate number of topics.

Thirdly, labeling the 584-topic model would have been very difficult for humans. It would have required a significant amount of time and effort to consistently mark such a large number of topics and a large number of people to ask for long forms to achieve statistical significance. After weighing all these factors, the 64-topic model was chosen as a sweet spot between having a relatively low perplexity of 827, a low perplexity per topic of 12.93, meaningful identified words, and a manageable number of topics for humans to label consistently. This model strikes a balance between capturing the nuances of the data while still being interpretable and manageable to label.

6.3.2 Topic Labeling - Manual Labeling with Survey

As previously mentioned, the ideal number of topics chosen for this work was 64. An example of the output generated by the model is shown in Table 5. This table displays the topics identified by the model, followed by the words that were identified as part of that topic. Note that the topic is identified only by a number, not a name, which highlights the need for topic labeling.

Table 5: Partial list of topics along with the nine initial words for each computed by sklearn's LDA

Topic	Word 0	Word 1	Word 2	Word 3	Word 4	Word 5	Word 6	Word 7	Word 8
0	piece	restaurant	buffet	eat	soup	sushi	roll	try	menu
1	ask	table	say	order	come	employee	customer	time	want
2	quality	price	ratio	service	space	food	quantity	lunch	recommend
3	time	smell	service	serve	restaurant	space	stone	customer	employee
4	service	lack	food	employee	space	customer	price	point	table
5	love	restaurant	open	door	close	try	make	time	say
6	cream	ice	flavor	fruit	try	taste	chocolate	crepe	space
7	option	menu	dish	lunch	day	space	food	service	vegan

To accomplish this, several approaches were considered. One option was to consult with a food industry professional to create accurate labels for each topic. Another option was to label the topics among ourselves manually. A third option was to create a survey in which participants could attribute their labels to each topic.

Internal labeling: To minimize the paradox of choice and individual biases, the topics were labeled internally manually in the initial phase. For the internal labeling phase, a team of three was selected. The first individual was tasked with assigning four potential labels to the topics. The second individual then reviewed these options. Finally, any conflicts were resolved by the third individual.

A subset of the labels provided created by internal labeling can be found in Table 6. The complete table can be found in Appendix Table 18.

Table 6: Partial list of topics along with the four proposed Labels presented in the survey.

Topic	Label A	Label B	Label C	Label D
0	Piece	Sushi	Soup	Buffet
1	Customer service	Ordering and serving	Employee interaction	Table and ordering
2	Price-quality ratio	Food quantity and service	Lunch recommendations	Value
3	Service and management	Customer experience	Meat and space	Waiting time
4	Service quality	Employee friendliness	Food and space	Service and food
5	Restaurant experience	Mall food	Lisbon cuisine	Good experience
6	Ice cream flavors	Fruit and chocolate crepes	Ice cream service	Ice cream
7	Vegan options	Lunch menu	Vegan restaurant	Menu options

Crowdsourcing labeling and validation: In the survey, four predefined choices (A, B, C, and D) were provided to the participants. To ensure diversity and creativity, a fifth option was also introduced. This was an open-ended option that allowed participants to create their labels if the predefined choices did not satisfy their requirements. In a study where topics were represented by words sorted by their importance, only the first nine words were displayed in three categories—most important (green), important (yellow), and least important (red) to guide participants in naming topics using the most significant words primarily, with the option to refer to the less important ones as a secondary and tertiary resources Fig. 15.

The allocation of participants to either the Topic Labeling (0-31) or Topic Labeling (32-63) section was contingent upon their selection between two visually identical, randomly ordered blank options. This methodology aimed to condense the survey’s dimensions, enhancing participant re-

Topic 6:
 Which expression or word better identifies the following succession of words.
 Keep in mind that the words are ordered in **decreasing level of importance**:
 ● **Most important:** cream, ice, flavor, fruit
 ● **Important:** try, taste, chocolate
 ● **Least important:** crepe, space
☐ A (Ice cream flavors)
☐ B (Fruit and chocolate crepes)
☐ C (Ice cream service)
☐ D (Ice cream)
☐ Other...

Fig. 15: Example of how the Topic labeling questions were presented in the topic labeling survey

sponse accuracy and diminishing the incidence of arbitrary choices, thereby mitigating the risk of obtaining irrelevant outcomes. For those participants inclined to provide additional insights, an opportunity was extended to annotate a subsequent set of 32 topics, thereby augmenting the dataset of survey responses.

Topic Labeling Results: The Topic Labeling survey was conducted with a group of 50 participants, resulting in 66 half-responses. This was due to the option for respondents to engage with either or both segments of the survey. Among these participants, 15 contributed to both sections, while 12 and 24 individuals exclusively addressed the first and second parts, respectively. This distribution resulted in a total of 33 complete responses, effectively averaging each of the 64 topics to be labeled 33 times. The labeling mechanism offered respondents four predefined choices: A, B, C, and D, alongside a fifth, open-ended option permitting custom inputs by the users. The final topic label was chosen based on the maximum voting policy. For example, for a topic with equal votes (such as topic 47, topic 6), the label "Ice cream flavors" was chosen. The actual results, including the voting distribution for each topic and the final labels, are partially shown in Table 7. The entire table can be found in Appendix Table 17.

Table 7: List of all topics and the respective winner labels given by the topic labeling survey

Topic N°	Topic Label	Topic N°	Topic Label
0	Sushi	4	Service quality
1	Ordering and serving	5	Restaurant experience
2	Price-quality ratio	6	Ice cream flavors
3	Customer experience	7	Menu options

The formula Eq. 2 is used to determine the final label for each topic. It takes into account the votes each label received from the survey participants. The formula operates under the following conditions:

- If there is a tie in the maximum votes for a topic, the topic label is set to ‘Null’.
- If a unique label suggested by a participant has more votes than the maximum votes for the predefined choices, this unique label is chosen.
- Otherwise, the label with the maximum votes among the predefined choices (A, B, C, D) is selected.

This formula ensures a fair and democratic process for topic labeling, taking into account both the majority opinion and unique insights from the participants.

$$HTopicLabel_{topicN^0} = \begin{cases} Null, & \text{if } count_{topicN^0}(maxVote) > 1, \\ other_{unique}, & \text{if } other_{unique} > count_{voteN^0}(maxVote_{topicN^0}), \\ maxVote(A_{topicN^0}, B_{topicN^0}, & \text{otherwise} \\ C_{topicN^0}, D_{topicN^0}), \end{cases} \quad (2)$$

6.4 Conclusion

In this chapter, we present a novel approach to detecting topics in reviews. This method integrates unsupervised topic detection using Latent Dirichlet Allocation (LDA) techniques with manual labeling, both internally and through crowdsourcing, to reduce load and bias. The optimal number of topics was determined to be 64 using LDA and perplexity metrics. Initially, English-translated reviews were processed using various data processing methods such as tokenization, lemmatization, and vectorization. The N-grams process was found to hinder the process and was therefore excluded from the final pipeline. For each review, the unsupervised LDA detection approach identified 64 topics with associated words. To label the topics, an online survey was conducted with 50 participants, yielding 66 partial responses. Ultimately, 63 topics were successfully labeled, while one topic resulted in a four-way tie and was subsequently discarded.

7 Visual and Textual presentation of Data according to User Choice

7.1 Introduction

In today’s data-driven world, the ability to present information in a way that aligns with user preferences is crucial. Visual and textual presentations of data cater to different cognitive styles, enhancing comprehension and engagement. By allowing users to choose their preferred format, we can improve the accessibility and effectiveness of information dissemination. To achieve this, three surveys were created and deployed to gather user opinions and minimize biases. Prior approval was obtained from the Data Protection Officer at the University of Madeira, as documented in Appendix PDF. 1. The surveys, deployed via Google Forms, ensured anonymity and consisted of three sections: Demographic Information, Feedback and Evaluation Habits, and Specific Survey Questions. The Demographic Information section collected details about participants’ country of birth, gender, age group, and English fluency to create a comprehensive profile. The Feedback and Evaluation Habits section gathered information about participants’ restaurant ownership history, usage of platforms to find restaurants, rating habits, and reasons for rating experiences. These two sections were standard across all surveys. The Data Visualization Preferences Survey and the Text Generation Comparison Survey were also available in Portuguese for inclusivity. However, the Topic Labeling Survey was conducted exclusively in English due to translation complexities.

7.2 Demographics and Habits

The survey participants were predominantly born in Portugal, with a higher proportion of men, 64%, than women, 36%. The age distribution was fairly even, with the majority falling between 21 and 40 years old. All participants had some level of English proficiency, with most being either independent or proficient users.

In terms of habits, most participants had not held a decision-making position in a food establishment. The use of platforms to find restaurants varied, with a significant number of participants using these tools sometimes or often, but also a considerable number who rarely or never use them.

When providing feedback on purchases or services, there was a clear preference for numerical evaluations over written reviews. However, a significant portion of participants rarely or never offered either type of feedback. Most participants rated products or services in alignment with

their genuine perceptions, indicating a high degree of honesty and objectivity. However, a small segment tended to assign higher ratings than they believed were earned.

The factors influencing the provision of higher or lower ratings were explored, but due to the low number of responses, no clear conclusions could be drawn. The data suggests that emotional factors, personal values, and comparative assessments may play a role in how people rate products and services.

This analysis provides a comprehensive overview of the surveyed population characteristics and behaviors, revealing a typical population with the most demographic factors and habits aligning with expected norms. The findings hint at a preference for quantifiable evaluations and a high degree of integrity in feedback. Further research could delve into the factors influencing these behaviors.

7.3 Data Visualization

The Data Visualization portion of this survey aims to gather information about preferences when it comes to visualizing data. The survey consists of standard sections followed by a specific part. In this particular part, two types of questions were asked:

7.3.1 A, B Comparison

Participants are presented with two graphs and asked to choose their preferred one. An example is illustrated in Fig. 16, where two different visualizations of the same data are shown side by side. Participants can then select the graph they find more effective or appealing.

Several forms of data visualization were created with the objective of surveying individuals about their preferred method of data visualization. The complete surveys can be found in the Appendix PDF 2 for the English version and Appendix PDF A.7 for the Portuguese version. The questions and answers asked in the survey were grouped into categories based on the type of data visualization they refer to:

Radar Charts With Various Categories: Questions in this category asked participants to choose their preferred radar chart for visualizing data. The results showed a noticeable preference for three categories over two. Based on this significant difference, three classes were chosen to represent emotion in the platform.

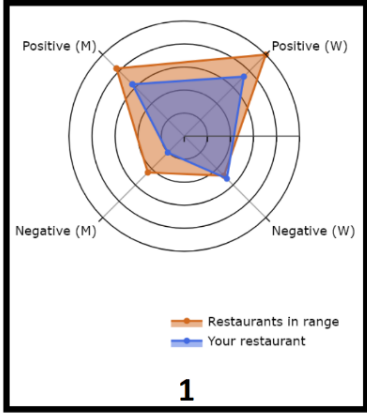
Radar chart (Your restaurant vs restaurants in range) *

Radar charts 1 and 2 compare the emotions expressed in reviews of **your restaurant**, in blue to those of **nearby restaurants**, in orange, split by gender.

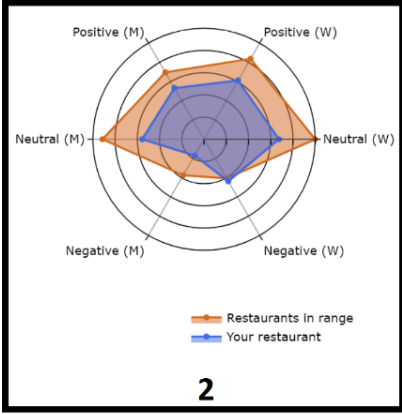
Radar chart 1 shows the **positive and negative emotions** expressed by men and women.

Radar chart 2 shows the **positive, neutral, and negative emotions** expressed by men and women.

Which radar graph do you prefer for visualizing this data?



1



2

☐ Radar chart 1
☐ Radar chart 2

Fig. 16: Example of a comparison between radar charts question on the Data visualization part of the survey.

Pie vs Bar Charts: Questions in this category asked participants to choose between pie charts and bar charts for visualizing data. The results showed a clear preference for pie charts over bar charts. Therefore, pie charts were chosen instead of bar charts on the platform.

Bar vs Table: One question asked participants to choose between a bar chart and a table to visualize gender distribution. The results showed a slight preference for the bar chart method, so the bar chart method was chosen to show gender distribution across the reviews on the platform.

7.3.2 Concept Perception Rating:

Participants are asked to quantify, on a scale of 1 to 5, how well they perceive a specific concept based on a given graph. They are shown a chart and then asked to rate their understanding of a particular conclusion drawn from that graph. An example of this type of question can be seen in Fig. 17.

Line Charts Questions in this category asked participants to rate their understanding of a specific concept based on a given line chart. The majority of users found line charts easy to understand. Therefore, line charts will be utilized in the platform.

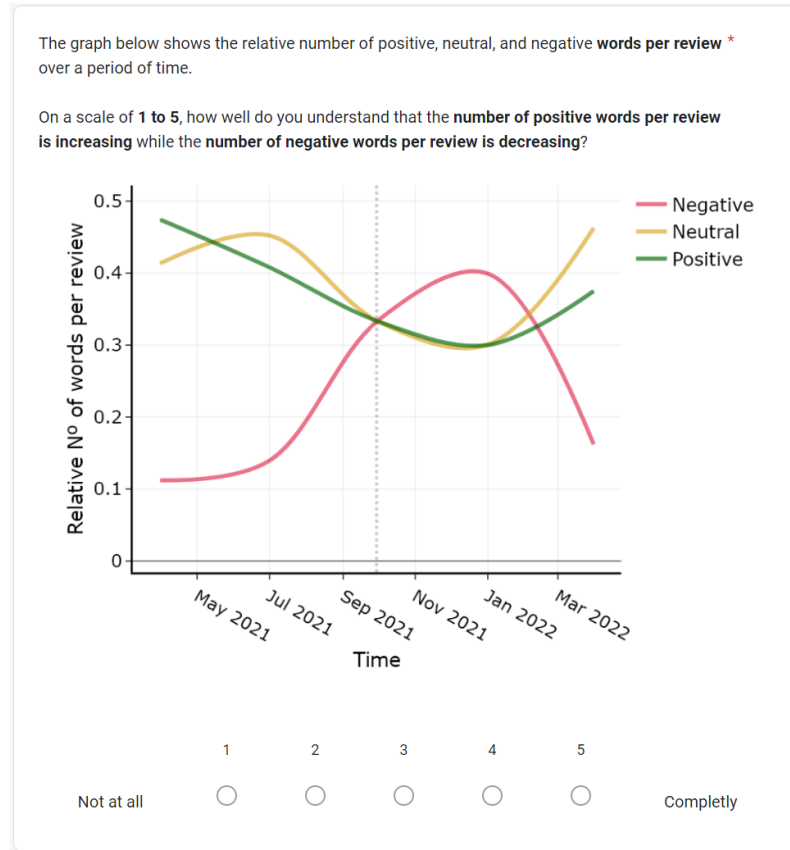


Fig. 17: Example of a question of concept perception of the data visualization survey.

7.3.3 Combined vs Separated Charts

One question asked participants to choose between a figure containing two graphs and a figure combining both graphs into a combined score. The results were evenly split, so it was decided to keep both methods of displaying emotion combined and separated on the platform.

The survey findings provide several key takeaways. Notably, the majority of users find line charts easy to understand. As for the representation of emotion over time, the custom Emotaix method and the rating threshold method are equally popular. Based on these insights, this strategy ensures that the visualizations are in line with user preferences and are easily interpretable.

7.4 Preferred Text (ASTA vs LLM)

The preferred text generation portion of the survey was conducted to collect information about preferences regarding the text generated in the paragraphs of the report. In this section, the participants were asked to choose their preferred paragraph among two options: one generated by ASTA and another by an LLM, specifically GPT-3.5, using the same information as ASTA. The purpose of this survey was to evaluate and compare the text generation capabilities of ASTA and GPT-3.5 and to understand the participant preferences in the context of the generated content. An example of one of the questions in the Text Generation Comparison survey is shown in Fig. 18.

Please select the paragraph you think is **better overall**. *

Gender section introduction A:
 This report provides an analysis of the gender of customers at ABC-XYZ. It includes bar charts showing the number of reviews by gender, pie charts splitting the reviews into positive and negative parts by gender, and radar charts comparing the emotions of customers in the restaurant with those in range, using 2, Positive and Negative and 3 classes, Positive, Negative, and Neutral. The gender that is most common among the restaurant's customers is female, and the approximate ratio of men to woman customers at the restaurant is 1 man for every 2 woman.

Gender section introduction B:
 This section provides an study of the gender distribution among patrons of ABC-XYZ and how it compares to that of the ideal best and ideal worst restaurants. It comprises bar charts that indicate the total number of reviews for each gender and radar charts split into 2 (Positive and Negative) and 3 (Positive, Neutral, and Negative) groups that display the ratio of men to women for each section. A radar chart that illustrates how ABC-XYZ compares to the ideal best and ideal worst restaurants is also available in this section. Gender analysis reveals that the restaurant is predominantly visited by individuals who identify as female, with approximately 1 men for every 2 woman.

☐ Paragraph A

☐ Paragraph B

Fig. 18: Example of a question of the Text Generation section of the survey where the LLM and the ASTA are compared.

As previously mentioned, a survey containing paragraph examples generated by both an LLM and ASTA was deployed to multiple participants. Each participant was presented with three different paragraphs, each generated by both ASTA and the LLM, specifically GPT-3.5, using the same information. The purpose of this survey was to determine the most influential textual presentation method and evaluate the quality of ASTA. The complete survey can be found in the Appendix PDF A.8 for the English version and Appendix PDF 3 for the Portuguese version.

The questions and answers asked in the survey were grouped into categories based on the type of paragraph being compared:

Summary of the Report: The Summary section of the report showed a preference for GPT-3.5 with 26 votes, while ASTA received 20 votes. This resulted in a distribution of 57% for GPT-3.5 and 43% for ASTA.

Section on Gender: When it comes to the Gender section, 25 participants favored GPT-3.5, and 21 participants chose ASTA. This translates to a 54% preference for GPT-3.5 and a 46% preference for ASTA.

Section on Nearby Restaurants: For the Nearby Restaurants section, 26 participants opted for GPT-3.5 and 20 for ASTA. This means GPT-3.5 was preferred by 57% of the participants and ASTA by 43%.

The statistics reveal a minor inclination, with an average of 55.8%, in favor of the LLM version when contrasted with 44.2% of ASTA. This implies that although the LLM is the victor, ASTA is a close contender. This suggests that the additional complexity, potential inconsistency, and extra costs linked with an LLM can be sidestepped by opting for a more straightforward method with negligible drawbacks.

The ASTA boasts several advantages, such as local operation, no associated costs, and the potential for straightforward extension or enhancement. It does not suffer from the drawback of dependence on an external service, which could potentially modify its pricing or structure in the future.

Both LLM and ASTA are to be integrated into the platform discussed in Chapter ???. However, the GPT version was excluded from the platform version that will be delivered to Zomato. Zomato made this decision as a preference not to depend on an external service was expressed. This approach ensures that the unique requirements of the project partners are met while a variety of options on the platform are offered to contribute to broader research efforts.

7.5 Comparative Analysis of Linguistic Features in ASTA and GPT

In this survey, GPT-3.5 version is noted as A and ASTA version as B. The number of sentences and the similarity scores for different sections were compared. The sections considered are the Report Summary, Gender Section, and Nearby Restaurants. The detailed comparison is in Table 8.

Table 8: Comparison of sentence counts and similarity scores for different sections

Section	GPT (N° of sentences)	ASTA (N° of sentences)	Similarity
Report Summary	7	6	0.977
Gender Section	3	4	0.973
Nearby Restaurants	3	7	0.915

7.5.1 Vocabulary Features

In this section, the vocabulary features of the collected corpus are analyzed. The focus is on how GPT and ASTA differ in the choice of words when answering the same set of questions. Since the number of answers from GPT and ASTA is unbalanced, one answer from each is randomly sampled during the statistical process. The following features were calculated: Average length (L), which is the average number of words in each question; Vocab size (V), the number of unique words used in all answers. Looking at both features of average length and vocab size, it is observed that GPT's answers are relatively longer but use a smaller vocabulary compared to ASTA. GPT's answers are usually longer than those of ASTA, but the vocab size is significantly smaller.

7.5.2 Part-of-Speech

The Part-of-Speech (POS) analysis provides insights into the grammatical categories of words used by both GPT and ASTA. The comparison of POS usage between GPT and ASTA is visually represented in Fig. 19 a).

The bar graph in Fig. 19 compares the mean percentage of various parts of speech used by GPT and ASTA. The parts of speech considered include adjectives (ADJ), adpositions (ADP), adverbs (ADV), auxiliaries (AUX), conjunctions (CCONJ), determiners (DET), nouns (NOUN), numerals (NUM), particles (PART), pronouns (PRON), proper nouns (PROPN), punctuation (PUNCT), subordinating conjunctions (SCONJ), spaces (SPACE), symbols (SYM), and verbs (VERB).

The graph shows that for most categories, the percentages are relatively close between GPT and ASTA, with some variations. For instance, GPT seems to have a higher rate in categories like AUX, DET, PRON, and VERB, while ASTA has higher percentages in ADP and CCONJ. This suggests that while both GPT and ASTA are capable of generating linguistically diverse responses, they exhibit slightly different preferences in their usage of certain parts of speech.

7.5.3 Entity

The Entity analysis provides insights into the types of entities recognized by both GPT and ASTA. The comparison of entity recognition between GPT and ASTA is visually represented in Fig. 19 b).

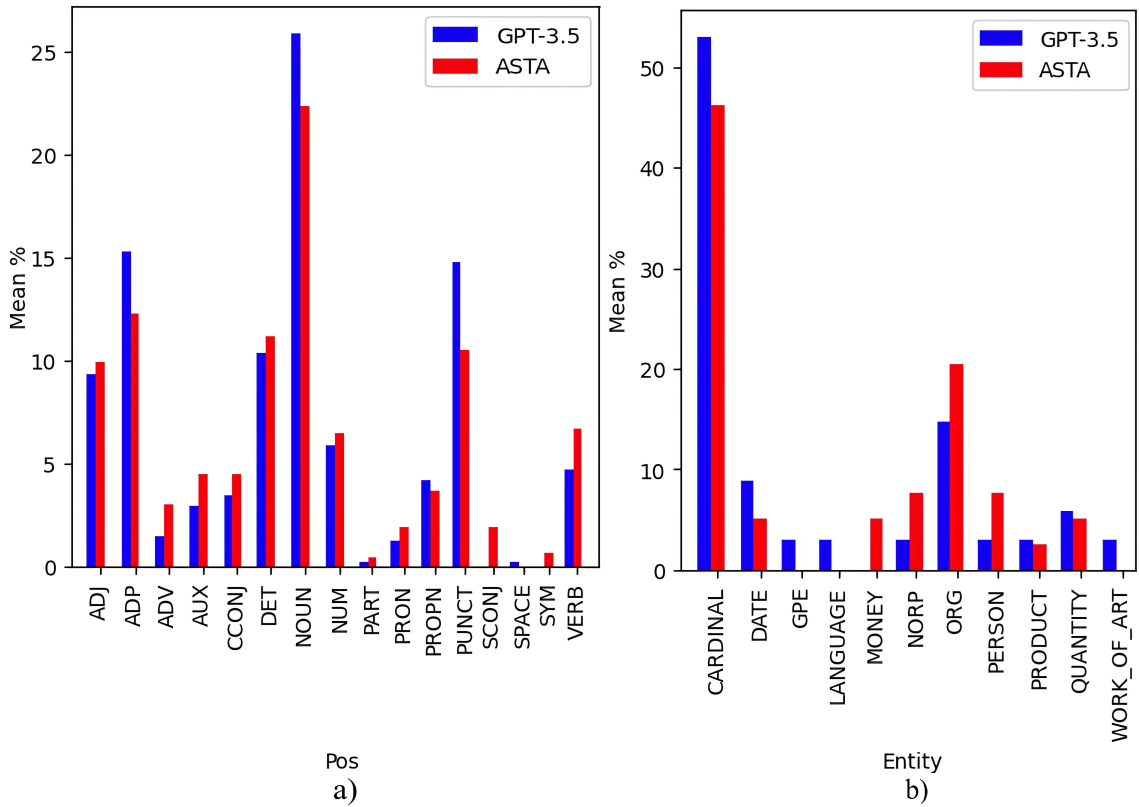


Fig. 19: **a)** Part-of-Speech comparison between GPT.3.5 and ASTA paragraphs. **b)** Entity comparison between GPT.3.5 and ASTA paragraphs.

The bar graph in Fig. 19 compares the mean percentage of various entities recognized by GPT and ASTA. The entities considered include CARDINAL, DATE, EVENT, FAC, GPE, LANGUAGE, LAW, LOC, MONEY, NORP, ORDINAL, ORG, PERCENT, PERSON, PRODUCT, QUANTITY, TIME, and WORK_OF_ART.

The graph shows that for most categories, the percentages are relatively close between GPT and ASTA, with some variations. This suggests that while both GPT and ASTA are capable of recognizing a diverse range of entities, they exhibit different preferences in their recognition of certain types of entities.

This analysis is significant as it helps understand the entity recognition capabilities of different language models in NLP tasks. It also provides a basis for further exploration into how these differences in entity recognition might impact the readability, style, and perceived quality of the generated text.

7.5.4 Perplexity

The Perplexity analysis provides insights into the predictability of both GPT and ASTA. The comparison of perplexity between GPT and ASTA is visually represented in Fig. 20 a).

The bar graph in Fig. 20 a) compares the perplexity of GPT and ASTA across three different categories: Reportssummary, GenderSection, and NearbyRestaurants. Perplexity is a measurement of how well a probability distribution or probability model predicts a sample. It is often used in NLP to evaluate language models.

In the category Reportssummary, GPT has a lower perplexity than ASTA, indicating that GPT is more predictable in this context. In GenderSection, both GPT and ASTA have high perplexity values, with ASTA slightly higher than GPT, suggesting that both models find this context more challenging to predict. For NearbyRestaurants, both GPT and ASTA have very high perplexity values, with ASTA having significantly higher perplexity than GPT; this indicates that the NearbyRestaurants' context is challenging for both models to predict, especially for ASTA.

This analysis is significant as it helps understand the predictability of different language models in various contexts. It also provides a basis for further exploration into how these differences in perplexity might impact the readability, style, and perceived quality of the generated text.

7.5.5 Dependency Parsing

The Dependency Parsing analysis provides insights into the types of syntactic relations recognized by both GPT and ASTA. The comparison of dependency parsing between GPT and ASTA is visually represented in Fig. 20 b).

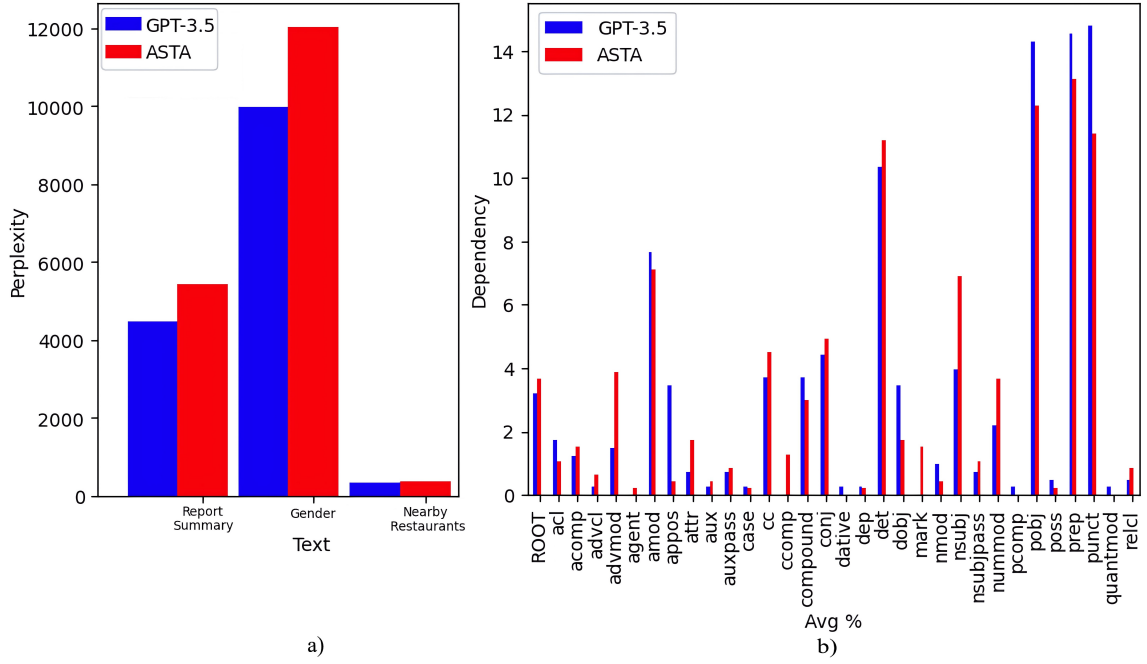


Fig. 20: **a)** Perplexity comparison. **b)** Dependency Parsing comparison.

The bar graph in Fig. 20 a) compares the average percentage of various dependency types recognized by GPT and ASTA. The dependencies considered include ‘ROOT’, ‘acl’, ‘advcl’, ‘advmod’, and so on.

The graph shows that for most categories, the percentages are relatively close between GPT and ASTA, with some variations. This suggests that while both GPT and ASTA are capable of recognizing a diverse range of dependencies, they exhibit different preferences in their recognition of certain types of dependencies.

This analysis is significant as it helps understand the dependency parsing capabilities of different language models in NLP tasks. It also provides a basis for further exploration into how these differences in dependency parsing might impact the readability, style, and perceived quality of the generated text.

7.6 Conclusion

The surveys conducted provided valuable insights into the preferences and habits of the participants, which will guide the development and enhancement of the system. The demographic and habits survey revealed a typical population with the most demographic factors and habits aligning with expected norms. The data visualization survey highlighted the importance of user-friendly

and easily interpretable visualizations, leading to the adoption of line charts, pie charts, and both combined and separated emotion representation on the platform.

The text generation comparison between ASTA and GPT-3.5 showed a slight preference for the LLM version. However, the close competition and the advantages of ASTA, such as local operation and no associated costs, led to the decision to integrate both ASTA and LLM into the platform, with the exclusion of the GPT version from the version delivered to Zomato.

The linguistic analysis, including dependency parsing, perplexity, and part-of-speech comparisons, provided a deeper understanding of the syntactic and semantic capabilities of both ASTA and GPT-3.5. These analyses revealed that while both methods exhibit strong linguistic capabilities, they have different strengths and weaknesses that can be leveraged depending on the specific application requirements.

In conclusion, the surveys and analyses have played a crucial role in understanding user preferences and shaping the system's development. The findings were used to keep the system user-centric and effective in its purpose. The integration of both ASTA and LLM ensures that the platform remains flexible and adaptable to various needs while also addressing the unique requirements of Zomato. Future research could further explore the factors influencing user preferences and the impact of different linguistic features on the perceived quality of the generated text.

8 System design for Reporting

to keep the data current and relevant. This platform acts as an invaluable tool for restaurant owners, facilitating data-driven decision-making. The following sections will explore these features in detail, illustrating their implementation and utility in enhancing the decision-making process for restaurant owners. Fig. 22 is a diagram that explains how the ReviewSight Platform works. Additionally, Fig. 21 is a new diagram that shows the platform in action.

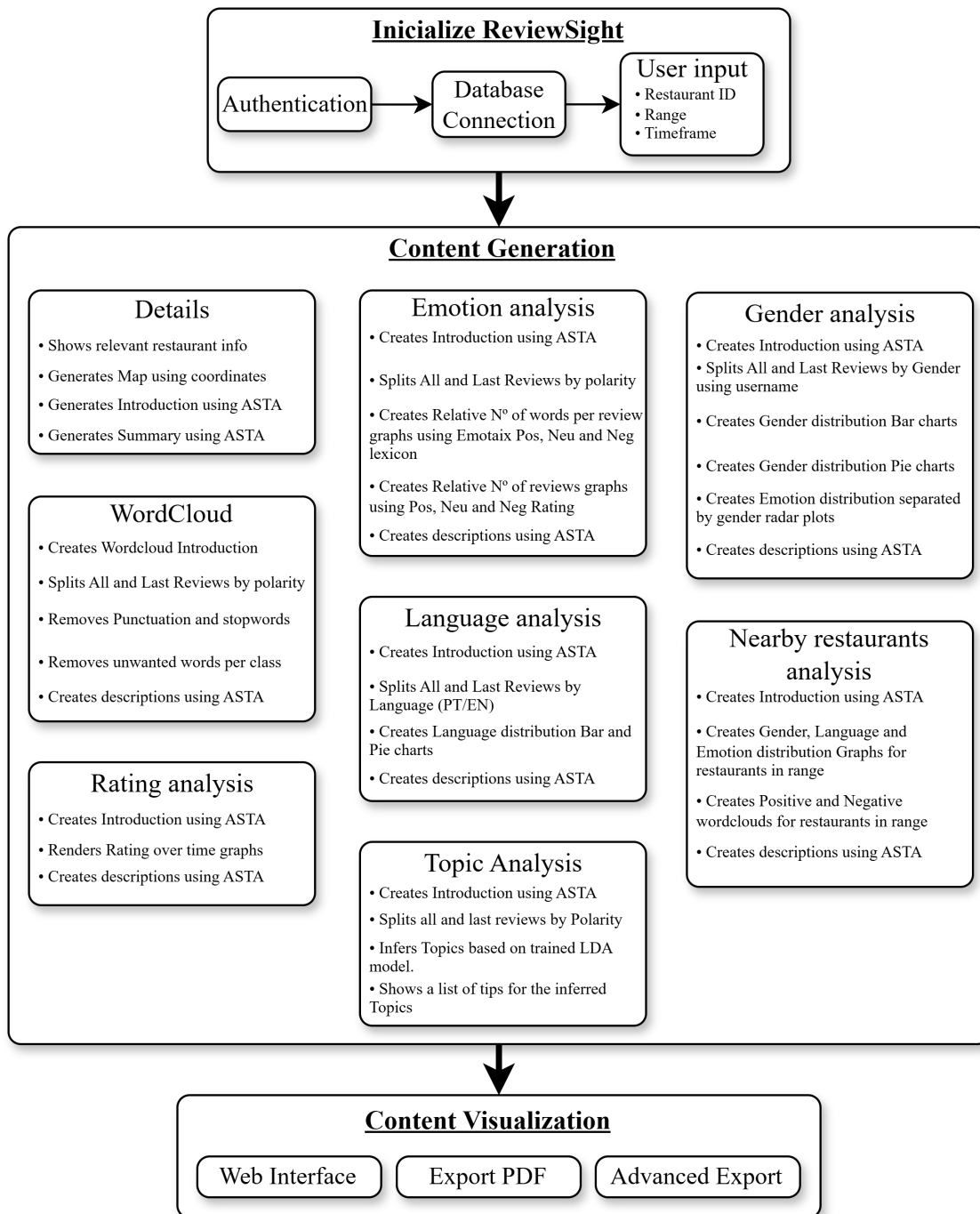


Fig. 22: High-level diagram of the ReviewSight Platform's workflow.

8.2 System Design for report generation (ReviewSight)

The ReviewSight is designed with two primary goals for restaurant owners. The first goal is to provide a comprehensive understanding of their restaurant through a range of features, including SA, topic detection, metrics visualization comparison between genders, Word Clouds, and more. The second goal is to offer insights into competitor restaurants, enabling owners to benchmark and strategize effectively. All this information is presented in a compact form, with the option to generate a PDF report or a zip file. These features are implemented using the Streamlit library in Python [85] [86]. Streamlit is an open-source library that allows developers to build visually appealing user interfaces for data applications. It is made for machine learning engineers and data scientists, enabling them to convert their data scripts into shareable web applications [87]. This makes ReviewSight a powerful tool for restaurant owners to understand and improve their business performance.

8.3 Authentication Service

8.3.1 User

To facilitate the deployment of the platform, a login system was developed using the `streamlit_authenticator` library. This system accommodates two types of users: Normal users and Administrators. Regular users can log in using their username and password and change two interface parameters. The first parameter is the "Range" in meters, which specifies the distance for comparison with nearby restaurants. This allows users to customize their search for restaurants within a specific radius. The second parameter is the "Timeframe", which determines the number of months into the past that the system should consider when analyzing data. This allows users to focus on more recent data or to include a larger historical dataset.

8.3.2 Administrator

Platform administrators can adjust the same parameters as regular users and add two additional review rating thresholds to separate reviews into negative, neutral, or positive classes. This allows platform administrators to fine-tune the classification of reviews based on their specific needs. Additionally, platform administrators can generate a PDF report and perform advanced exports, giving them greater flexibility in managing and analyzing data.

8.3.3 Password Hasher

To manage users, a Yet Another Markup Language (YAML) credentials file was used containing the following credentials: username, email, name, and password. To create a new user, a new field must be added to this file with the new credentials, where the password is stored in a hashed format for security purposes. Another file named restaurantIDS.csv exists with two fields to identify the type of user: username and restaurant ID. This file determines whether a user is an admin or a regular user. If the user is a regular user, their restaurant ID field must be a number used in the central DB and belong to a specific restaurant. This allows regular users to be associated with particular restaurants. If the user is an administrator, their restaurant ID field must be "Admin", indicating that they have administrative privileges. The system includes a cookies feature that is managed by the credentials.yaml file to enhance the user experience by automatically signing in a user if a specified period has not been exceeded.

The default period is set to 1 day, but it can be easily adjusted to a longer duration for user convenience. If a user returns to the platform within the specified period, the system will automatically log them in without requiring them to enter their credentials again. This provides a more seamless and efficient user experience while maintaining security and protecting user data. This is particularly useful for users who frequently access the platform, as it saves time and effort by eliminating the need to enter their login information manually. It also provides a more personalized experience, as users can quickly resume their work on the platform without interruptions.

8.4 Database

The ReviewSight retrieves information through two methods: by connecting to a MongoDB DB or through a pickle file. The MongoDB DB serves as the primary source of data for ReviewSight due to its flexibility and ability to add new data without the need to rewrite the entire DB. The pickle file, a binary file used for storing data, acts as a backup in case the MongoDB DB is unavailable.

Despite containing a large number of reviews (approximately 1.2 million) and being fully denormalized, the DB requires relatively little memory (around 530 megabytes). This is because the system only accesses specific columns containing text and numbers, such as review_id, rating, creation_date, res_id, res_name, res_location, res_rating, res_price, res_lat, res_lng, res_city, language, gender, word_count, translation_pt, translation_en, emotax_positive, emo-

taix_neutral, emotaix_negative, emotaix_unknown, translation_pt_no_stop_words and translation_en_no_stop_words.

To improve performance, the interface is designed to load the entire DB into system memory upon initial launch and then rely on memory for data retrieval. This approach is significantly faster than querying the DB each time new information is required. The system will be slower in its initial launch but faster in its subsequent runs.

While this is an academic endeavor, and no new data was added to the system during the thesis, the system is designed to handle new data. The processes explained in the previous chapters regarding data cleaning and feature generation would need to be performed before ingesting the new data into the system. This ensures that the ReviewSight platform can adapt to new data and continue to provide accurate and insightful reports.

8.4.1 Data Denormalization

Data was received from Zomato, as discussed in previous chapters, in two main files: restaurant and reviews. For the system to operate correctly, a denormalization process needs to occur in the same way that was described in the previous chapters. This process could vary depending on how the data is provided. If the data is given in a different format, the process would be different. The crucial part is that the final output respects the same data columns and types for the system to work correctly.

8.4.2 Data Cleaning

To make the system live, a comprehensive data cleaning process should be implemented before the data is ingested into the MongoDB DB. This process would involve several crucial steps. Firstly, the system should check for and remove any duplicate entries in the data, ensuring that each review is unique and preventing skewed results due to repeated data. Secondly, the system should identify and appropriately handle any missing or null values in the data, which could involve removing rows with missing values or imputing missing values based on the data's characteristics. The review text should then be pre-processed to remove any unnecessary characters, such as punctuation and special symbols, and converted to lowercase to ensure consistency. If the reviews are in different languages, they should be translated into Portuguese and English. Finally, the system should perform SA on

the review text to add the Emotai positive, neutral, and negative number of word columns to the DB.

8.4.3 Populating the MongoDB Database

Once the data is cleaned, it is ready to be ingested into the MongoDB DB. In this way, the MongoDB DB can be continuously updated with new data as it becomes available. This ensures that the reports generated by the ReviewSight platform are always based on the most recent and relevant data.

By maintaining a clean and up-to-date DB, the ReviewSight platform can provide restaurant owners with accurate and insightful reports, helping them make data-driven decisions to improve their business.

8.5 Linguistic Service

8.5.1 ASTA

ASTA is a system engineered to produce semi-structured written reports. The uniqueness of these reports lies in the creation of each paragraph. Rather than being a static, single paragraph, each one is sampled from a pool of multiple versions, each subtly different in tone. Moreover, critical strings within these paragraphs are replaced with values from another pool of strings. These values represent the data interpreted by the system, tailored to the specific situation and context of the restaurant. This process creates an incredibly vast number of combinations, virtually infinite, ensuring that each report is unique. This approach not only maintains the consistency of the report's content but also varies its presentation. The result is a report that feels as though it was written by a human, specifically tailored to the restaurant owner's unique situation.

8.5.2 LLM

In the context of report writing, the second method involves utilizing a LLM, specifically GPT-3.5. This approach includes providing a saved structured prompt with all the information gathered by the system for each paragraph to the GPT-3.5 servers via Python over the web. The model's response is then organized into paragraphs, ensuring content consistency while allowing for variations in presentation. The result is reports tailored precisely to the unique situation of the restaurant owner.

8.6 Visualization Service

Data visualization plays a crucial role in conveying complex information effectively. In this section, various data visualization techniques, including pie charts, radar charts, scatter plots, bar charts, and tables, were explored to display the data the platform generated. These techniques were selected for their ability to convey complex data in a visually appealing and easy-to-understand manner. Surveys were conducted to determine the most effective methods for visualizing data. These surveys were distributed to a randomized group of individuals, including restaurant owners. The goal was to identify a preferred visualization technique that would allow for the clear and concise presentation of information. All generated graphs were rendered directly on the web platform using Streamlit and other Python libraries such as Pyplot, Plotly, and more. The charts were saved as PNG files for inclusion in the PDF report and advanced zip export. The following sections will discuss the various visualization techniques used on the platform and the rationale behind their selection for each scenario. Examples of how these techniques effectively communicated the findings will also be provided.

8.6.1 Map

The report generation platform provides several options for creating a map that displays the location of the restaurant and its surroundings. This feature provides user context and enhances the platform's visual appeal. For the web version, Streamlit's `map()` function was used to create a map with a red dot indicating the restaurant's location. This simple function provides a quick and easy way to generate a map. The web version of the map can be seen in Fig. 23 b).

A more sophisticated approach was employed for the PDF version to generate a visually impactful map. The Folium library was used to create an HTML version of the map. This map version is available in the advanced export version of the Report and provides users with additional functionality, such as zooming, panning, labels of nearby locations, street names, etc. A screenshot of the map was captured and saved as a PNG file to include the map in the PDF version of the Report. This allowed for the quick incorporation of the map into the Report while maintaining its visual appeal. The folium version of the map can be seen in Fig. 23 a).

Using this heuristic, it is possible to increase the number of samples per average as the total number of samples increases. This reduces volatility in the data by grouping multiple values and displaying the results more smoothly. The following examples illustrate the Graph without the moving average and with the moving average applied in Fig. 24 a) and b), respectively.

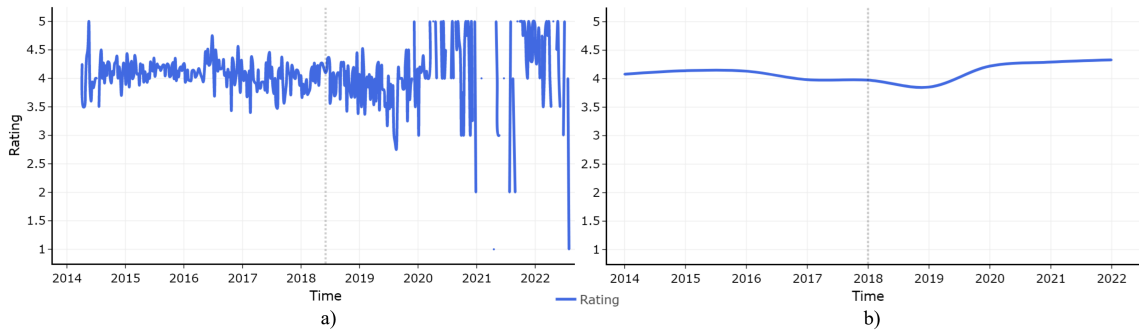


Fig. 24: **a)** Example of rating over time using spline. **b)** Example of rating over time using spline and moving average.

By comparing the graphs with and without the moving averages, it is possible to observe that including the moving average makes the Graph simpler and more interpretable. Even though the Graph does not show the actual values, it offers an approximation of the values with a smooth interpolation between them.

8.6.3 Tables

Tables offer an effective means of presenting data in an organized and structured format. They enable the easy comparison of values and can exhibit a wide range of data types in a condensed form. In this study, tables were employed to display the gender distribution in reviews overall time and in reviews from a user-selected number of months. The utilization of tables allows for the presentation of absolute values and percentages for each gender, thereby providing a more comprehensive perspective of the data, as depicted in Fig. 25.

	Nº of Man - (%)	Nº of Woman - (%)
All reviews	868 (27.0%)	2341 (73.0%)
Last 36 months	51 (32.3%)	107 (67.7%)

Fig. 25: Table including Review Distribution Across Genders for All Reviews and Last 36 Months.

8.6.4 Pie Charts

In this work, multiple pie charts were used to display data ranging from gender distribution to language and emotion distribution. Pie charts are ideal for conveying a sense of proportion between different categories. For this reason, they were implemented in this work using the Plotly library, with percentages included to aid the user in interpreting the chart, as shown in Fig. 26.

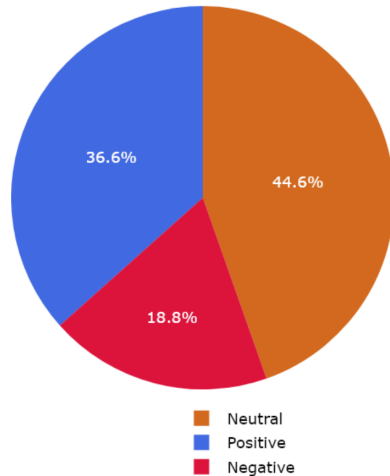


Fig. 26: Example of Pie chart depicting Distribution of Emotion in Reviews.

8.6.5 Radar Plots

Radar plots play a significant role in this work. They are utilized several times in this work to compare the distribution of emotions across different classes for each gender and to contrast the emotions of each gender with those of hypothetical ideal best and worst restaurants. This is illustrated in Fig b). 27.

The relative occurrence of positive, neutral, and negative words, as per the emotion detection process, employed by each gender when composing reviews on the platform is depicted in part b) of the referenced figure (Fig.27). These counts have been normalized, with the class possessing the highest count being established as the maximum value on the scale and the others adjusted accordingly. This approach directs attention toward the plot's shape rather than the absolute values.

The overall utilization of emotional words within the analyzed restaurant is lower compared to that in one of the adjacent establishments. While the general shape of the analyzed restaurant's profile resembles that of nearby restaurants, a discrepancy arises when scrutinizing the employment

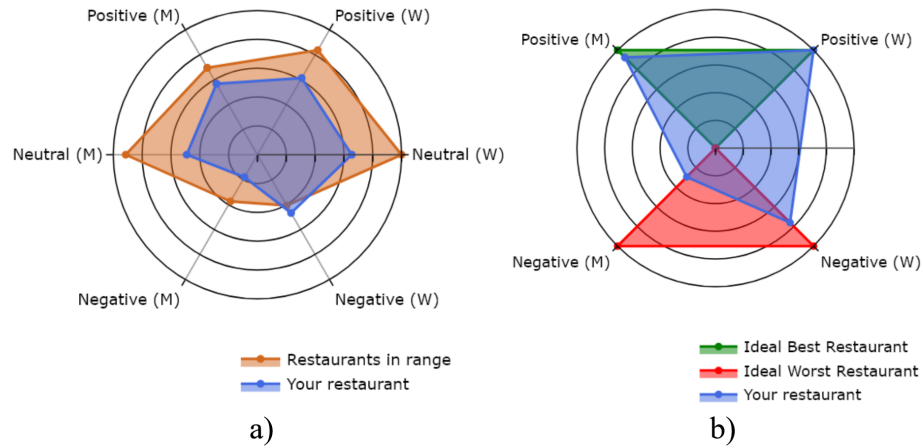


Fig. 27: **a)**: Radar plot illustrating a comparative analysis of emotions associated with the owner's restaurant, in blue, and those prevalent in nearby restaurants, in orange, segregated by gender and classified into three categories: Positive (M), Neutral (M), Negative (M) for males and Positive (W), Neutral (W), Negative (W) for females. **b)**: Radar chart illustrating the comparison of emotional responses, segregated by gender, associated with the owner's restaurant (depicted in blue). This is contrasted against the emotional responses for a hypothetical ideal best restaurant (depicted in green) and a hypothetical ideal worst restaurant (depicted in red).

of negative words by male and female reviewers. Notably, women patrons at the analyzed restaurant employ a greater frequency of negative words than their counterparts at nearby restaurants. In contrast, male patrons at the analyzed restaurant use fewer negative words than those at nearby restaurants.

This information holds significance for the restaurant owner, prompting a deeper inquiry into the factors contributing to the divergence in the utilization of negative words between their establishment and neighboring ones. Through a comprehensive analysis of this data, the restaurant owner can identify potential areas warranting improvement and initiate appropriate actions to elevate customer satisfaction and overall experience.

8.6.6 Bar Plots

Bar plots are a powerful tool for data visualization. Using vertical or horizontal bars, they can display the distribution of data points or compare metric values across different data groups. The developed platform uses 7 bar plots to illustrate the distribution of gender, language, and emotion for the entire duration of the restaurant's existence, as well as for specific months selected by the user. This information is available to both the restaurant owner and nearby restaurants, allowing for easy comparison and analysis of trends. These insights into customer behavior and preferences are valuable for decision-making. Bar plots effectively convey complex information clearly and concisely, making it easier for decision-makers to understand and act on the data. These bar plots

are created using the plotly library with custom parameters to achieve a specific visual style, as shown by Fig. 28.

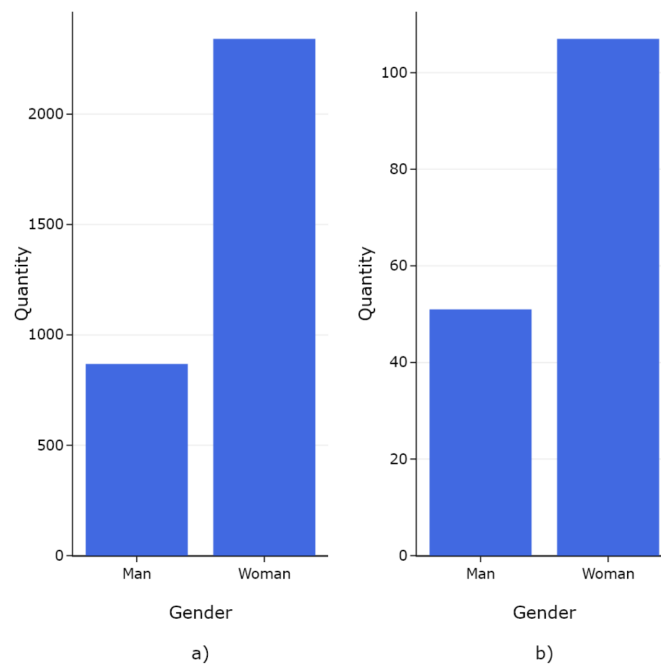


Fig. 28: Comparative bar plots illustrating the gender distribution in two different Figures, a and b. The blue bars represent the number of men and women in each scenario.

a) represents the number of men and women in the restaurant's entire existence.

b) represents the same distribution in a user-specified period.

8.6.7 Word Clouds

A Word Cloud is a visualization technique that displays words in a way that their font size is related to their frequency. This method lets users quickly identify essential words in a group of reviews and make an educated guess about the topics being discussed [88]. Word clouds help analyze large amounts of text data, such as customer reviews or social media posts. By displaying the most frequently used words, Word Clouds can help users quickly identify trends and patterns in the data.

In this work, Word Clouds provides users with a quicker way to understand the key themes and topics discussed in customer reviews. By analyzing the words used in these reviews, users can gain valuable insights into the strengths and weaknesses of the restaurant and make informed decisions about how to improve their business. Fig.29 a) and b) are examples of Word Clouds used in the platform.

word on the previously created list by using the `get_word_forms` function from the `word_forms` library [89]. For example, given the input "absolute", it would output "absolute", "absolutely", "absoluteness", "absolutenesses", and "absolutes". The three CSV files mentioned above were then generated with all variations of every word. The Report generator platform is designed to read these files every time it initializes. It removes the words from the generated word clouds, making them more representative of what is happening in the restaurants. This feature enables users to customize the word clouds by adding or removing words from the CSV files, thereby tailoring the pre-processing stage of the report generator to their specific needs.

Finally, the word-cloud files are saved using the PNG format. These files are used in the PDF version of the Report and can also be accessed directly in PNG format via advanced export. In Table 9, you can find the first examples of Negative, Neutral, and Positive words in the CSV files. The entire content of the files is in Appendix excerpts A.14 ,A.15 and A.16 for `PositiveWords.csv`, `NeutralWords.csv`, and `NegativeWords.csv`, respectively.

Table 9: First ten words from the `PositiveWords.csv`, `NeutralWords.csv`, and `NegativeWords.csv` files used in Word Cloud generation.

PositiveWords.csv	NeutralWords.csv	NegativeWords.csv
absolution	1st	abnormal
absolutions	2nd	abnormalities
absolve	3rd	abnormality
absolved	4th	abnormally
absolver	5th	absence
absolvers	6th	absences
absolves	7th	absent
absolving	8th	absented
absolviory	9th	absentee
advance	a	absentees

8.7 ReviewSight Structure

The ReviewSight system offers a web interface for real-time analysis, along with a comprehensive PDF report and a zip file.

This section will outline the structure of the PDF version. The PDF report 4 begins with a title page that showcases key statistics and specifics about the restaurant, followed by a map image, an introduction, and a summary. The summary is a concise encapsulation of all the vital information from the rest of the Report.

The Report features a section on Word Clouds, which visually highlights the most frequently mentioned words in both positive and negative reviews, providing insights into customer experiences. This is followed by a Rating Analysis section that includes graphs depicting review ratings over time, offering a visual representation of customer satisfaction and potential areas for improvement.

Next, the Report delves into Emotion Analysis, where customer feedback is analyzed to uncover emotional trends and overall satisfaction levels, which can be instrumental in enhancing the dining experience. The Gender analysis section details the gender demographics of the restaurant's customers, comparing them to ideal best and worst-case scenarios using bar charts and radar charts.

The Report also includes a Language Analysis section that explores the languages used in customer reviews. It presents bar and pie charts showing the distribution of languages over the restaurant's entire history and the last user-selected months. The Topic Analysis section examines the topics mentioned in recent reviews and provides actionable tips for the business to improve performance related to these topics.

Finally, the Report concludes with a Nearby Restaurants Analysis section. This section offers a comparison with nearby restaurants, analyzing factors such as average price, rating, total number of reviews, and customer demographics. This comprehensive analysis provides a holistic view of the restaurant's performance and areas for improvement. In Appendix A.11 is a detailed explanation of each specific section of the PDF.

As an alternative to the ASTA, a LLM was used to generate some of the paragraphs of the Report and present them in a format similar to that of the ASTA. Specifically, OpenAI's ChatGPT 3.5 was used via the Python openai library, connecting directly to ChatGPT 3.5 using an Application Programming Interface (API) and a key that was specially generated for this purpose. A prompt with all the necessary details and specific instructions on how the paragraph should be created was crafted to generate a paragraph.

The prompts were designed to direct the LLM to perform a particular task, including what information to include, how to present it, the maximum number of words, and even what to do in case some information is not available. An example of the prompt structure for the Report's nearby restaurant introduction is shown in the Appendix Code excerpt. 6. In this Excerpt, all the

specific details that were included in the prompt to achieve the best and most consistent result can be seen.

An example of the prompt used to generate the summary paragraph, which is the most extended and most complex paragraph in the entire Report, along with the corresponding response from GPT 3.5, is shown below. This is an example of an actual prompt sent through GPT 3.5 after all the information in the variables has been replaced with the actual information produced by the system, which is available in the Appendix Code excerpt 5.

As shown in the example, while the prompt carefully imposes limitations on specific parts and allows creativity in others, certain aspects of the response are poorly formulated. The first phrase of the example, "This report provides an analysis of ABC - XYZ, a restaurant located within a 250-meter radius.", contains a serious mistake that completely changes the intended meaning. Though such errors do not occur consistently, their frequency is significant. This drawback highlights the disadvantage of relying on an LLM for report generation, as opposed to the ASTA, which offers greater consistency.

The development of the functions to generate the three most extensive paragraphs using GPT 3.5 was successful, and they are fully functional and included in the main code. Although additional paragraphs were deemed unnecessary due to limited responses, the problem of hallucination remained at the time of this research. However, with the rapid advancements in LLM technology, this method might become more reliable in the near future.

The speed and cost associated with the LLM approach were critical considerations. As the LLM method involved sending prompts to external servers and awaiting responses, its performance depended heavily on the server's availability. Moreover, any changes made by the company to the LLM itself, pricing structure, or availability could adversely affect the Report's generation, potentially rendering it non-functional, depending on the extent of these changes. The reliance on external servers, inconsistency of response, and susceptibility to changes were decisive factors in disregarding the usage of this method in the system. However, the system was tailored to leverage the capabilities of GPT-3.5, and with minor adjustments, it could seamlessly transition to using GPT-4 or other LLM models like LLaMa, BERT, and others.

8.7.1 Topic Based Tips

Topic detection is a powerful tool that helps restaurant owners identify the most important themes discussed in their reviews. By analyzing these topics, owners can gain valuable insights into their establishment's strengths and weaknesses. Once these topics are identified, the next step is to address any issues and enhance the overall customer experience.

Developing a system to provide a pool of tips and suggestions based on the identified topics is the next logical step to complement our topic detection model. These tips can guide restaurant owners in improving their business, whether by addressing specific flaws or enhancing already strong areas. By taking these steps, restaurant owners can ensure their business thrives and meets the ever-changing demands of their customers. Ultimately, this can lead to increased customer satisfaction and loyalty, which are essential for the long-term success of any restaurant.

To implement this system, tips are stored in a CSV file with four columns: Topic, Tip1, Tip2, and Tip3. The Topic column contains one of 64 possible topics, while the Tip1, Tip2, and Tip3 columns contain three corresponding tips. The system is designed to perform topic detection and then retrieve one random tip from the three available for the corresponding topic from the CSV file. This adds variety and further reduces the chance of two reports being similar to one another. An example of the first three entries of the CSV file that contains the tips is shown in Table 10. The full CSV file is available in Appendix 4.

Table 10: Table presenting a partial list of topics along with three associated tips for each topic.

Topic	Tip1	Tip2	Tip3
Table	Ensure that the tables are clean and properly set, provide comfortable chairs or booths for diners, and train staff to monitor tables for any customer needs or requests.	Make sure tables are clean and free of any crumbs or debris before seating guests.	Make sure tables are sturdy and wobble-free and provide enough space for guests to comfortably dine.
Value	Offer good portion sizes for reasonable prices, consider providing daily specials, and offer a loyalty program or discounts for repeat customers.	Offer daily specials or promotions to provide extra value to customers.	Offer a loyalty program or discounts for repeat customers to show appreciation for their business.

8.7.2 Branding

In partnership with Zomato, ReviewSight offers a watermark in the generated reports. This watermark serves multiple purposes: branding, security, and flexibility. The watermarked file is included in the advanced export .zip file alongside the standard Report.

The flexibility comes from the straightforward process of changing the watermark. By providing a blank PDF file with the desired watermark, the system seamlessly overlays the Report on top of it. Additionally, a Zomato footer logo has been strategically integrated on all pages of the Report. Users can customize this image using the same method described above, ensuring compatibility with other logos, phrases, or shapes. In Fig.30, there is an example of the Report's initial page with the watermark.

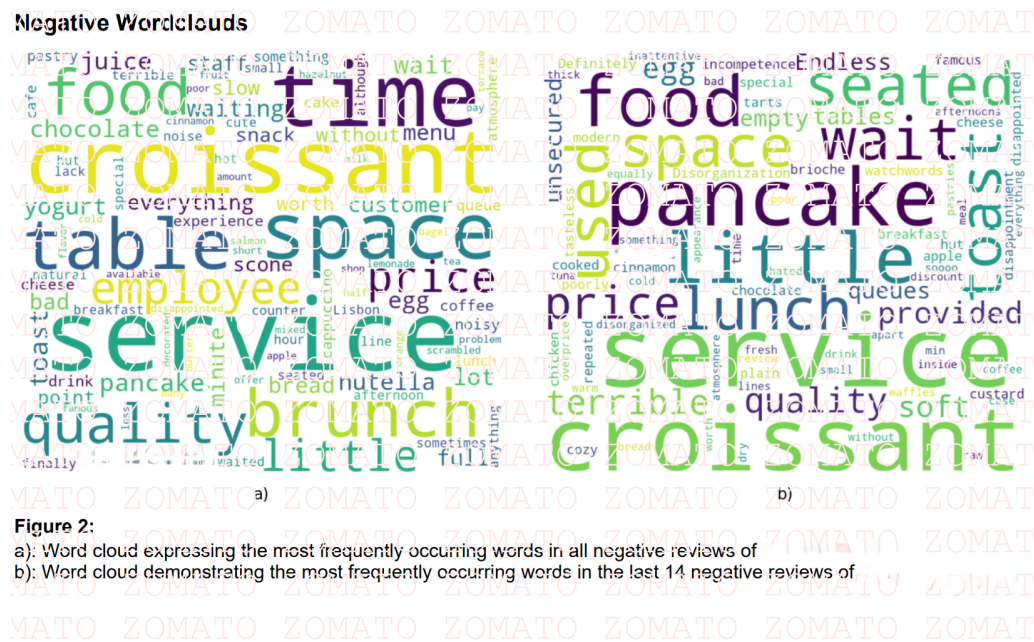


Fig. 30: Excerpt of the Report’s initial page with watermark and footer logo.

8.8 Exportability Service

8.8.1 PDF generation

The system generates a PDF report that includes all the items from the previous list, such as images, graphs, and captions, based on the selected restaurant, range, and timeframe. The introduction of each section and the main summary are created by randomly selecting one of five paraphrased versions of the same paragraph or phrase. These versions have placeholders that are replaced by actual information produced by the system.

The Report has 15 paragraphs with five paraphrased versions each. The placeholders contain ratios, ratings, predominant gender, predominant emotion, adjectives, and more. For example,

to evaluate the restaurant's ratings over time, the selected timeframe is divided into two halves, and the average of both halves is computed. Then, the ratio between the averages of the second and first halves is calculated. The word that expresses this ratio is chosen according to specific criteria. If the ratio is higher than 1.1, one of three random words is used: thriving, flourishing, or prosperous. If the ratio is lower than 0.9, one of three random words is used: declining, struggling, or worsening. If the ratio is between 0.9 and 1.1, one of three random words is used: stable, steady, or consistent. This method is applied to generate various words that convey ratios, polarity, or trichotomy throughout the entire Report. This results in a highly variable report with a solid structure that virtually never repeats.

8.8.2 XLSX Generation

In the ReviewSight system, users can download a .zip file that contains multiple folders. Within one of these folders, several XLSX files are capturing various metrics related to topics, tips, gender demographics, and more. Users can utilize this data to create customized reports, generate graphs, tables, and charts, or track restaurant performance over time. These XLSX files provide the same values as the Report but in a format that allows restaurant owners to manipulate the data according to their preferences. The intentionally open approach ensures that users are not limited to predefined methods and can create their own versions of the Report.

8.8.3 Downloadable Individual Images

Following the same principle of openness, the ReviewSight system also allows users to download individual maps, graphs, and tables. When an advanced export is generated, the resulting .zip file includes a folder containing all these items in PNG format. This convenience enables restaurant owners and other stakeholders to create their customized versions of the Report, unrestricted by predefined views.

8.9 Conclusion

This chapter introduces a comprehensive platform designed to provide restaurant owners with valuable insights extracted from customer reviews. By leveraging advanced data processing techniques and NLP tools, the platform offers a user-friendly interface that generates detailed reports. These reports include various visualizations, such as line charts, pie charts, and word clouds. Additionally,

the platform employs unique report writing methods, such as ASTA, which ensure accurate, concise, and predictable information transmission. This empowers restaurant owners to understand customer sentiment, identify areas for improvement, and make data-driven decisions tailored to their specific needs, ultimately enhancing service quality and customer satisfaction.

Given 15 paragraphs with five variations each, the number of report variations is 3.052×10^{10} , calculated as 5^{15} . Considering 64 topics with three tips each and the system's ability to display up to 5 distinct topics with 1 out of 3 suggestions, the number of possible reports increases significantly. This is calculated by multiplying 3.052×10^{10} by $\binom{64}{49}$ and 3^{10} , resulting in 3.759×10^{41} possible reports. Factoring in different terminologies for expressing ratios, 3.759×10^{41} is multiplied by 6.811×10^{11} , yielding a total of 2.560×10^{53} . Given this vast number, it is almost impossible for the system to generate two identical reports, ensuring unique human-like authorship. This large range of possibilities further highlights the platform's capacity to offer unique insights for each restaurant owner.

9 Cloud Implementation

9.1 Introduction

The focus of this chapter is the deployment of a Linux VM on the GCP, a top-tier cloud service provider. GCP delivers a variety of modular services, such as computing power, data storage, data analytics, and machine learning. The versatility of the platform empowers users to opt for an environment that best fits their needs. The project includes an analysis of how to set up a Linux VM on GCP, specifically configured to ensure compatibility with Linux, which is widely used in enterprise settings. Additionally, the diverse customization options that GCP offers are discussed. These include the choice of cores, the addition of a Graphics Processing Unit (GPU) or a Tensor Processing Unit (TPU), and the setting of RAM and storage.

9.2 Implementation

Google Cloud Platform (GCP) is a suite of cloud services built on Google’s infrastructure, ensuring high uptime and data redundancy. It offers modular services like computing power, data storage, analytics, and machine learning, with environments such as IaaS, PaaS, and serverless computing to meet diverse needs [90].

In this project, a Linux VM was set up on GCP to showcase its compatibility with enterprise Linux environments. The VM supports remote access for multiple users, making it suitable for collaborative projects and high-traffic applications. Hosting on Google’s infrastructure ensures high availability, reliability, and scalability.

GCP’s customization options include selecting cores, adding GPUs/TPUs, and adjusting RAM and storage. These settings can be modified to support multiple instances and new features. The project, funded by “Sentiment and Rating Understanding of Portuguese Review for Restaurants using Machine Learning and Natural Language Processing” (FCT/CPCA-IAC/AF/478645/2022), involved creating a Google account and configuring a VM with 16 Intel Broadwell cores, 128 GB RAM, and 40 GB SSD. The setup included installing Ubuntu, NoMachine for remote access, and necessary software like MongoDB, Anaconda, and Python dependencies.

Tests on the Ubuntu VM demonstrated GCP’s capabilities by launching multiple instances to assess workload handling. The task manager monitored performance, validating the hardware

specifications. The system managed multiple instances efficiently, ensuring smooth operation during peak usage and future growth.

Figure 31 shows the GCP interface, highlighting VM management features.

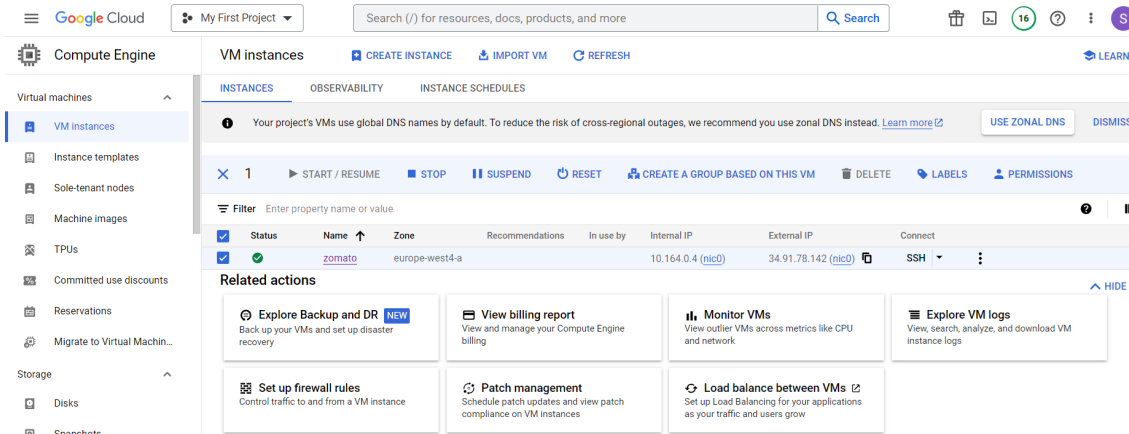


Fig. 31: Illustration of GCP’s VM Management Interface in Action: Running the Project-Specific Instance.

Upon starting the VM and using NoMachine, interaction with Ubuntu’s Operating System (OS) GUI can be facilitated by establishing a connection with the VM. Fig 32 displays several applications operating on the VM running Ubuntu using GCP, accessed via NoMachine. Google Chrome accessing the platform is visible on the top left, while the console running the Python application is on the bottom right. The task manager of Ubuntu, showing resource usage while loading the page, is on the top right, and MongoDBCompass, ensuring DB connectivity, is running on the bottom right.

9.3 Conclusion

The deployment of a Linux VM on GCP showed the power and adaptability of cloud services. The compatibility of GCP with Linux, a widely adopted enterprise OS, was successfully demonstrated.

The VM was configured for remote operation, facilitating multi-user access. This makes it ideal for collaborative projects or high-traffic applications. The extensive customization options provided by GCP, including core selection, GPU/TPU inclusion, and RAM and storage configurations, were effectively utilized.

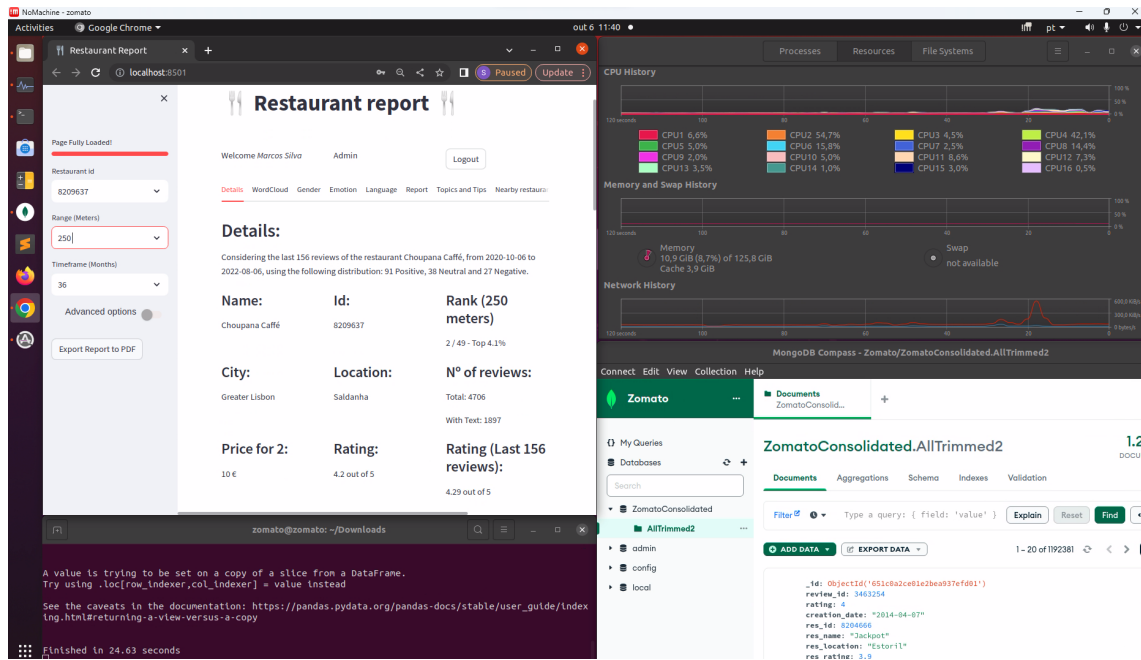


Fig. 32: Ubuntu VM instance, streamed via NoMachine, showcasing the ReviewSight in operation on both console and browser, complemented by MongoDB Compass and Task Manager.

The successful deployment and operation of the VM were ensured through a series of steps. These included the installation of Ubuntu, the configuration of NoMachine for remote access, and the installation of necessary software and dependencies.

The capabilities of GCP were demonstrated through tests conducted on the Ubuntu desktop operating on the VM. These tests validated the choice of hardware specifications for the VM, confirming the system's efficiency in managing multiple instances in memory. This is crucial for smooth operation during peak usage periods and for future growth.

In conclusion, the successful implementation of the platform on a Linux VM on GCP has highlighted the potential of cloud services in providing robust, scalable, and adaptable solutions for enterprise environments. This experience serves as a valuable reference for future projects where this system could be employed. It could be used with a public IP for real restaurant owners, for example.

10 Discussion and Conclusion

10.1 Discussion

This research has significant implications for the food service industry and the application of NLP techniques in analyzing customer reviews. By understanding the impact of emotions on ratings, restaurant owners can better address negative feedback and improve customer satisfaction. The detailed feedback provided by customers, combined with advanced NLP techniques, offers valuable insights that can enhance service offerings and tailor experiences to meet customer needs.

Chapter 1 highlighted the importance of the food service industry in Portugal's economy, emphasizing the significant revenue generated and the high percentage of household expenses allocated to food services. The growth of online review platforms has made it crucial for restaurants to understand and leverage customer feedback.

In Chapter 2, data collection and database preparation, the choice of a NoSQL database, specifically MongoDB, was justified based on its ability to manage and analyze large volumes of data efficiently.

Chapter 3 provided a thorough examination of the dataset, uncovering notable patterns and trends. It underscored the importance of understanding user behavior, language detection, and the challenges associated with handling multilingual data.

Chapter 4 introduced a customized technique for identifying and categorizing emotions in restaurant reviews. The findings underscored the significance of sentiment analysis in understanding customer opinions and experiences.

In Chapter 5, various methodologies were employed to identify the most influential features for predicting review ratings. The results highlighted the importance of sentiment-related features.

Chapter 6 presented a novel approach to detecting topics in reviews, integrating unsupervised learning techniques with manual labeling. This approach provided deeper insights into customer feedback, offering valuable information on specific subjects discussed in reviews.

Chapter 7 focused on gathering user preferences through surveys to determine the most effective forms of data visualization and text generation. The platform offers various visualization options, including line charts, pie charts, and word clouds, which convey complex data insights in an easily interpretable manner. This approach ensured that user preferences were thoroughly considered.

Chapter 8 showcased the development of the ReviewSight platform, which provides restaurant owners with valuable insights extracted from customer reviews. Its deployment on a Linux VM on GCP demonstrated the adaptability and scalability of cloud services.

Chapter 9 detailed the deployment process of the ReviewSight platform on a Linux VM on GCP, highlighting the benefits of cloud services, such as scalability, flexibility, and cost-effectiveness. The process included configuring the Linux VM and setting up the necessary software and dependencies, demonstrating the platform's ability to handle large-scale data processing and deliver reliable performance.

Online reviews have become a powerful tool in shaping consumer behavior and decision-making processes. This research illustrates the importance of analyzing customer reviews to gain valuable insights into customer preferences and experiences.

The application of NLP techniques in the food service industry has shown great potential for enhancing customer engagement and improving service delivery. This research has proven the effectiveness of these techniques in analyzing restaurant reviews and providing actionable insights for restaurant owners.

Future research could explore the application of NLP techniques to other languages and industries, thereby expanding the potential benefits. Currently, the system generates tips based on the topics detected in the text. Future work could involve combining topic detection with review polarity to produce sentiment-specific tips, thereby increasing their relevance and providing more comprehensive insights into customer experiences.

Integrating ReviewSight into a public cloud environment could offer significant benefits, including increased accessibility and scalability. Such integration would serve as a practical demonstration of the system's functionality in diverse contexts, broadening its reach and impact while improving efficiency and preparing it for large-scale deployment.

10.2 Conclusion

This thesis has explored the significant role of the food service industry in Portugal's economy and the impact of online reviews on consumer behavior and restaurant performance. By employing NLP techniques, it has provided valuable insights into customer sentiment and preferences, which can be leveraged to enhance service offerings and improve customer satisfaction.

Data was collected and processed from Zomato, a leading restaurant discovery and food delivery platform. The selection of a NoSQL database, specifically MongoDB, was based on its capability to efficiently manage and analyze large datasets.

The exploratory data analysis (EDA) revealed meaningful patterns and trends that informed subsequent analyses, highlighting the importance of understanding user behavior, language detection, and the challenges of multilingual data processing.

A customized technique for detecting and categorizing emotions in restaurant reviews was developed, emphasizing the relevance of sentiment analysis in capturing customer opinions and experiences.

Various methods were used to identify the most significant features influencing review ratings. The results underscored the importance of sentiment-related features as well as practical factors, such as the restaurant's location and pricing.

A novel topic detection approach, integrating unsupervised learning with manual labeling, offered more detailed analyses of customer feedback, shedding light on the specific topics discussed in reviews.

A user-friendly interface was developed to generate comprehensive, tailored reports for restaurant owners, enhancing the practical application of the platform.

The successful deployment of the platform on a Linux VM on GCP demonstrated the viability of cloud-based solutions, showcasing the adaptability and scalability needed for enterprise environments.

In conclusion, this thesis has made a substantial contribution to both NLP and the food service industry by demonstrating the value of analyzing reviews in both English and Portuguese. The insights derived from this research can be utilized to enhance service offerings, boost customer satisfaction, and contribute to the overall growth and success of the food service industry in Portugal and beyond.

Appendix

A.1 Provided Files

Table 11: Attributes of Zomato's Review Data.

N ^o	Name	Description
1	review_id	Review's unique identifier
2	user_pt_id	User's unique identifier in Zomato's Portuguese database
3	user_pt_name	User's name in Zomato's Portuguese Database
4	user_in_id	User's unique identifier in Zomato's Indian database
5	user_in_name	User's name in Zomato's Indian database
6	restaurant_id	Restaurant's unique identifier
7	restaurant_name	Restaurant's name
8	text	Review's comment
9	rating	Review's rating
10	created_at	Review's creation date

Table 12: Attributes of Zomato's Restaurant Data.

N°	Name	Description
1	id	Restaurant's unique identifier
2	name	Restaurant's name
3	location	Restaurant's location
4	rating	Restaurant's rating
5	price	Restaurant's average price for two people (Manually inserted by Zomato)
6	zomato_id	Restaurant's unique identifier 2
7	lat	Restaurant's latitude
8	lng	Restaurant's longitude
9	address_full	Restaurant's complete address
10	address_city	Restaurant's city
11	about	Restaurant's description
12	group_name	Restaurant's group name
13	group_type	Restaurant's type of environment (e.g., hotel, mall, complex)
14	priority	Restaurant's internal ranking number based on unknown parameters (Ranges from best, P1, to worst, P10)
15	dining_rating	Restaurant's rating 2
16	manager_email	Restaurant's manager email
17	manager_phone	Restaurant's manager phone number
18	chain_id	Restaurant's chain unique identifier
19	menu_id	Restaurant's menu unique identifier
20	et1	Restaurant's primary area/type (e.g., Casual, Pastry Shop, Bar)
21	et2	Restaurant's secondary area/type
22	time_slot_1	Restaurant's primary schedule (Monday to Sunday)
23	time_slot_2	Restaurant's secondary schedule
24	time_slot_3	Restaurant's tertiary schedule
25	company_name	Restaurant's company name
26	seats_exterior	Restaurant's n° of exterior seat
27	seats_interior	Restaurant's n° of interior seats
28	kitchenTypes_name	Restaurant's types of kitchens
29	infos_description	Restaurant's provided services/products (e.g., Reservations, WIFI, Credit Card)
30	deliveryMethods_name	Restaurant's delivery method

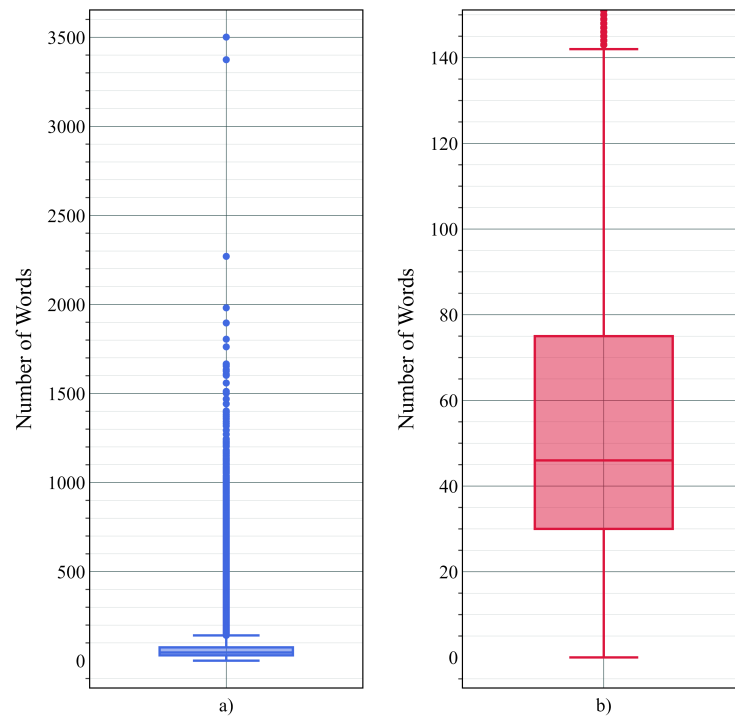


Fig. 33:

a): Full distribution of review length.

b): Cropped (0 - 150 words) review length distribution.

A.2 Translation

Table 13: Example of review text translated from original Portuguese to English using Google translate API.

Original Portuguese Review	English Translation
Nem sei muito bem o que dizer, porque adorei tudo!! Desde o ambiente á comida, adorei o espaço! Recomendando.	I don't really know what to say, because I loved everything!! From the ambiance to the food, I loved the space! I recommend.
Às vezes nos sítios mais inesperados encontramos os pratos mais fantásticos. Mesmo em frente ao célebre animatógrafo do Rossio é possível entrar num restaurante bem tradicional com o melhor caldo de camarão da cidade. E o preço? Muito Acessível. Ao almoço é um ótimo sítio para se provar boa comida caseira.	Sometimes in the most unexpected places we find the most fantastic dishes. Right in front of the famous Rossio animatograph it is possible to enter a very traditional restaurant with the best shrimp broth in town. It's the price? Very Affordable. At lunch it is a great place to taste good homemade food.
Um restaurante muito romantico que nos transporta para uma Lisboa mais antiga. Com uma vista unica na cidade é o lugar perfeito para o tal jantar romântico. A simpatia dos empregados ajuda muito a criar este momento e a cozinha não deixa a casa ficar mal. Com uma boa seleção de vinhos, se tivermos a sorte de apanha a mesa à janela da varanda, numa noite de verão é difícil querer ir embora.	A very romantic restaurant that transports us to an older Lisbon. With a unique view of the city, it is the perfect place for that romantic dinner. The friendliness of the employees helps a lot to create this moment and the kitchen doesn't let the house down. With a good selection of wines, if you're lucky enough to catch a table at the balcony window, on a summer night it's hard to want to leave.
um pouco caro mas muito bom	a little expensive but very good

Table 14: Example of review text translated from original English to Portuguese using Google translate API.

Original English Review	Portuguese Translation
This was the first time I had gone to ABC to eat... I know it's a market, but we found a nice cafe/restaurant section at one corner... Me and my friends stopped at the food section to grab a quick lunch... All my friends had gone for the traditional Portuguese cuisine but I was more interested in the pizza place. There was this nice live pizza station where the lady made the pizza which I headed over to immediately. I order the ham and pineapple as it wasn't something too risky at a live station. Somehow I was pleasantly surprised !!! It was a delicious pineapple pizza with the perfect amount of cheese and perfectly done crust.... At 3.90 euros for a pizza fit for two it was a total steal !!! A very satisfied customer of ABC !!	Esta foi a primeira vez que eu fui ao ABC para comer... Eu sei que é um mercado, mas encontramos uma boa seção de café / restaurante em uma esquina... Eu e meus amigos paramos na seção de comida para pegar um almoço rápido... Todos os meus amigos tinham ido para a cozinha tradicional portuguesa mas eu estava mais interessado na pizzaria. Havia uma bela estação de pizza ao vivo, onde a senhora fez a pizza para a qual me dirigi imediatamente. Eu peço o presunto e o abacaxi, pois não era algo muito arriscado em uma estação ao vivo. De alguma forma fiquei agradavelmente surpreso !!! Foi uma deliciosa pizza de abacaxi com a quantidade perfeita de queijo e crosta perfeitamente... A 3,90 euros para uma pizza para dois, foi um roubo total !!! Um cliente muito satisfeito do ABC!!
What a surprise! I was not fan of Magret de Pato... Now i m dreaming about that special flavour... Please, just give a try! It s worth it! ORGASMIC	Que surpresa! Eu não era fã de Magret de Pato... Agora estou sonhando com aquele sabor especial... Por favor, experimente! Vale a pena! ORGASMIC
Good place to grab a lunch in the sun. The burgers are decent. Do not go here expecting gourmet, however, if you are extremely hungry, you will have your money's worth. Come to think of it, some of us may find the place a tad bit overpriced. I had a simple chicken burger and fries. The patty was well cooked, but too small compared to the bread. It lacked sauce, hence the taste was a bit bland. The fries were nice. They do offer a big plate of fries, if only the burger would last that long :P	Bom lugar para almoçar ao sol. Os hambúrgueres são decentes. Não vá aqui esperando gourmet, no entanto, se você estiver com muita fome, terá o valor do seu dinheiro. Venha para pensar sobre isso, alguns de nós podem achar o lugar um pouco caro. Eu tinha um hambúrguer simples de frango e batatas fritas. A empada estava bem cozida, mas muito pequena em comparação com o pão. Falta molho, daí o sabor era um pouco sem graça. As batatas fritas estavam boas. Eles oferecem um grande prato de batatas fritas, se ao menos o hambúrguer durasse tanto tempo :P

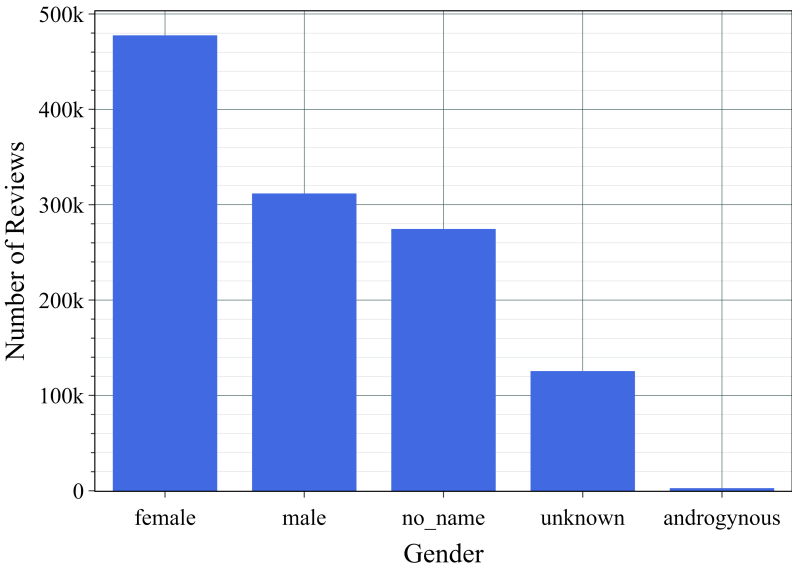


Fig. 34: Distribution of user gender across all reviews.

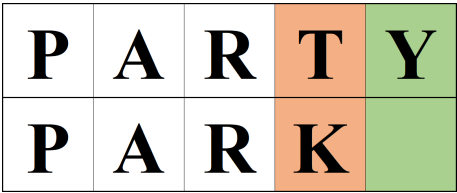


Fig. 35: Visual representation of the Levenshtein distance method.

A.3 All Features combined

Table 15: All Features.

N°	Short Name	Description	N°	Short Name	Description
1	review_id	Review's ID	35	word_count	Review's Number of words
2	user_pt_id	User's Portuguese ID	36	text_no_stopwords	Review without Stop-words
3	user_pt_name	User's Portuguese Name	37	stemmed_text	Review with Stemmed Words
4	user_in_id	User's International ID	38	translation_pt	Review Translated to Portuguese
5	user_in_name	User's International Name	39	translation_en	Review Translated to English
6	text	Review's Text	40	senti_pt_positive	Review's Portuguese Positive Sentistrength Score
7	rating	Review's Rating	41	senti_pt_negative	Review's Portuguese Negative Sentistrength Score
8	creation_date	Review's Creation Date	42	senti_en_positive	Review's English Positive Sentistrength Score
9	res_id	Restaurant's ID	43	senti_en_negative	Review's English Negative Sentistrength Score
10	res_name	Restaurant's Name	44	emotaix_positive	Review's Emotaix Positive Emotions
11	res_location	Restaurant's Location	45	emotaix_neutral	Review's Emotaix Neutral Emotions
12	res_rating	Restaurant's Rating	46	emotaix_negative	Review's Emotaix Negative Emotions
13	res_price	Restaurant's Price Range	47	emotaix_unknown	Review's Emotaix Unknown Emotions
14	res_zomato_id	Restaurant's Zomato ID	48	monday_1	Schedule 1 for Monday
15	res_lat	Restaurant's Latitude	49	tuesday_1	Schedule 1 for Tuesday
16	res_lng	Restaurant's Longitude	50	wednesday_1	Schedule 1 for Wednesday
17	res_address	Restaurant's Address	51	thursday_1	Schedule 1 for Thursday
18	res_city	Restaurant's City	52	friday_1	Schedule 1 for Friday
19	res_about	Restaurant's Description	53	saturday_1	Schedule 1 for Saturday
20	res_group_name	Restaurant's Group Name	54	sunday_1	Schedule 1 for Sunday
21	res_group_type	Restaurant's Group Type	55	monday_2	Schedule 2 for Monday
22	res_priority	Restaurant's Priority Level	56	tuesday_2	Schedule 2 for Tuesday
23	res_chain_id	Restaurant's Chain ID	57	wednesday_2	Schedule 2 for Wednesday
24	res_menu_id	Restaurant's Menu's ID	58	thursday_2	Schedule 2 for Thursday
25	res_et	Restaurant's Combined Schedule	59	friday_2	Schedule 2 for Friday
26	res_company_name	Restaurant's Company Name	60	saturday_2	Schedule 2 for Saturday
27	res_seats_exterior	Restaurant's Exterior Seating Capacity	61	sunday_2	Schedule 2 for Sunday
28	res_seats_interior	Restaurant's Interior Seating Capacity	62	monday_3	Schedule 3 for Monday
29	res_kitchen_types	Restaurant's Kitchen Types	63	tuesday_3	Schedule 3 for Tuesday
30	res_info	Restaurant's Information	64	wednesday_3	Schedule 3 for Wednesday
31	res_delivery_method	Restaurant's Delivery Method	65	thursday_3	Schedule 3 for Thursday
32	language	Review's Estimated Language	66	friday_3	Schedule 3 for Friday
33	gender	Reviewer's Estimated Gender	67	saturday_3	Schedule 3 for Saturday
34	first_name	Reviewer's Estimated First Name			

A.4 Topics

Table 16: Full list of all topics and respective associated words presented in the topic labeling survey.

Topic	Word 0	Word 1	Word 2	Word 3	Word 4	Word 5	Word 6	Word 7	Word 8
0	piece	restaurant	buffet	eat	soup	sushi	roll	try	menu
1	ask	table	say	order	come	employee	customer	time	want
2	quality	price	ratio	service	space	food	quantity	lunch	recommend
3	time	smell	service	serve	restaurant	space	stone	customer	employee
4	service	lack	food	employee	space	customer	price	point	table
5	love	restaurant	open	door	close	try	make	time	say
6	cream	ice	flavor	fruit	try	taste	chocolate	crepe	space
7	option	menu	dish	lunch	day	space	food	service	vegan
8	pizza	ingredient	service	quality	space	price	dough	crust	pasta
9	cake	bread	buy	slice	make	birthday	house	cod	place
10	juice	order	toast	croissant	egg	try	space	pancake	fruit
11	pizza	pasta	risotto	order	restaurant	dessert	try	come	starter
12	dinner	group	food	friend	restaurant	service	birthday	drink	recommend
13	book	think	try	space	sushi	price	quality	eclair	service
14	salad	dish	space	food	lunch	rice	service	meat	choose
15	music	atmosphere	space	food	hour	service	like	employee	time
16	cake	chocolate	space	service	decorate	staff	food	dessert	mousse
17	order	time	pizza	deliver	arrive	ask	service	make	mistake
18	brunch	drink	coffee	place	menu	egg	include	juice	breakfast
19	view	drink	space	place	night	terrace	afternoon	summer	service
20	coffee	cafe	snack	terrace	service	day	pig	space	lunch
21	fish	restaurant	grill	quality	price	service	meat	eat	food
22	place	food	porto	eat	service	people	staff	city	price
23	eat	meat	time	restaurant	make	point	service	cook	food
24	food	restaurant	service	make	time	space	price	quality	money
25	food	service	space	think	improve	come	restaurant	say	point
26	space	menu	lunch	food	child	hide	restaurant	staff	service
27	wine	cheese	dessert	meat	choose	flavor	accompany	space	starter
28	center	shopping	day	game	watch	eat	price	restaurant	service
29	space	variety	price	service	food	a lot of	quality	view	location
30	cod	dish	restaurant	beef	food	pork	meat	roast	octopus
31	rice	prego	fry	picanha	time	restaurant	eat	cuttlefish	service
32	tourist	space	area	price	smoke	wine	service	gin	bar
33	pastry	shop	cake	service	bread	croissant	puff	tart	bakery
34	sushi	quality	restaurant	service	price	space	staff	repeat	friendliness
35	wait	table	time	restaurant	arrive	food	people	service	reservation
36	dish	dessert	order	starter	chicken	restaurant	meat	service	potato
37	beer	snack	place	service	drink	snail	friend	space	price
38	gold	zomato	service	space	repeat	use	experience	food	cereal
39	sandwich	ham	cheese	time	think	eat	salad	order	bread
40	table	space	room	wall	chair	bar	people	floor	time
41	hamburger	burger	potato	fry	sauce	bread	meat	service	space
42	space	pancake	dog	service	staff	love	recommend	order	try
43	sauce	water	order	ask	steak	restaurant	service	come	tuna
44	wine	list	place	hustle	service	need	bustle	work	visit
45	chicken	taco	food	sauce	order	restaurant	service	guacamole	barbecue
46	francesinha	porto	sauce	space	eat	try	service	place	taste
47	order	sauce	come	like	salmon	bread	make	cheese	shrimp
48	pay	price	restaurant	time	atm	people	dish	drink	euro
49	seafood	sea	service	beach	restaurant	view	recommend	price	lunch
50	try	space	time	like	want	decide	tea	come	opportunity
51	steak	house	sauce	eat	time	know	restaurant	year	place
52	clams	prawn	service	crab	restaurant	eat	seafood	like	come
53	place	park	terrace	enjoy	raman	space	service	car	crepe
54	food	recommend	service	atmosphere	staff	visit	restaurant	welcome	space
55	space	food	service	order	establishment	staff	concept	make	experience
57	experience	space	chef	restaurant	food	make	thank	service	dish
58	restaurant	pasta	area	parking	dish	service	nacoe	parque	flavor
59	leave	desire	lot	restaurant	food	space	service	wine	make
60	dish	room	restaurant	table	end	kitchen	meal	menu	flavor
61	home	order	delivery	eat	service	recommend	restaurant	time	food
62	serve	time	come	order	table	employee	minute	service	wait
63	food	meat	flavor	service	restaurant	experience	curry	recommend	staff

Table 17: List of all topics and the respective winner labels given by the topic labeling survey.

Topic N°	Topic Label	Topic N°	Topic Label
0	Sushi	32	Tourist area
1	Ordering and serving	33	Bakery and pastry
2	Price-quality ratio	34	Quality sushi
3	Customer experience	35	Restaurant wait times
4	Service quality	36	Restaurant dishes
5	Restaurant experience	37	Beer and snacks
6	Ice cream flavors	38	Zomato gold
7	Menu options	39	Sandwich shop
8	Pizza	40	Furniture and decor
9	Cake and bread	41	Hamburgers and fries
10	Breakfast	42	Dog-friendly
11	Pizza and pasta	43	Steak dish
12	Group dinner	44	Wine place
13	Booking	45	Mexican
14	Salad and rice	46	Francesinha in porto
15	Music atmosphere	47	4 way Tie. (Topic discarded)
16	Chocolate cake	48	Payment and price
17	Pizza delivery	49	Seafood restaurant
18	Brunch menu	50	Trying out
19	View and drinks	51	Steak house
20	Coffee terrace	52	Seafood
21	Grilled fish	53	Location
22	Porto food	54	Recommended food
23	Meat restaurant	55	Experience
24	Restaurant service	56	Dish recommendation
25	Service	57	Experience
26	Lunch space	58	Pasta
27	Wine and cheese	59	Bad experience
28	Shopping center	60	Dining rooms
29	Food variety	61	Home delivery
30	Portuguese food	62	Serving time
31	Prego	63	Flavorful meats

Table 18: Full list of topics and respective suggested labels presented in the Topic labeling survey

Topic	Label A	Label B	Label C	Label D
0	Piece	Sushi	Soup	Buffet
1	Customer service	Ordering and serving	Employee interaction	Table and ordering
2	Price-quality ratio	Food quantity and service	Lunch recommendations	Value
3	Service and management	Customer experience	Meat and space	Waiting time
4	Service quality	Employee friendliness	Food and space	Service and food
5	Restaurant experience	Mall food	Lisbon cuisine	Good experience
6	Ice cream flavors	Fruit and chocolate crepes	Ice cream service	Ice cream
7	Vegan options	Lunch menu	Vegan restaurant	Menu options
8	Pizza quality	Pizza ingredients	Pasta	Pizza
9	Cake and bread	Birthday	Cod	Cake
10	Juice and toasts	Croissant and egg	Pancake breakfast	Breakfast
11	Pizza and pasta	Risotto and dessert	Starter and dessert	Pizza and risotto
12	Group dinner	Birthday celebration	Dinner with friends	Friends
13	Sushi quality	Éclair	Sushi booking	Booking
14	Salad and rice	Lunch meat dishes	Dish	Salad
15	Music atmosphere	Service waiting time	Background music	Music
16	Chocolate cake	Dessert	Cake decoration	Cake
17	Pizza delivery	Mistake	Order time	Pizza ordering
18	Brunch drinks	Coffee and eggs	Brunch menu	Brunch
19	View and drinks	Night terrace	Summer afternoon	View
20	Coffee	Coffee snacks	Coffee terrace	Coffee snack
21	Grilled fish	Fish quality	Sushi and seafood	Fish
22	Porto food	Porto people	Porto	Porto city
23	Meat quality	Meat restaurant	Cooked meat	Meat
24	Food value	Restaurant service	Food quality	Food
25	Improvement needed in food	Improvement needed	Service	Food
26	Kid-friendly menu	Lunch space	Kid-friendly restaurant	Space
27	Wine and cheese	Dessert flavors	Wine accompaniment	Wine
28	Shopping center	Sports game watching	Game	Service
29	Space variety	Food variety	Space	Price variety
30	Cod and beef	Roasted meats	Pork and octopus	Portuguese food
31	Rice and beans	Fried picanha	Cuttlefish and prego	Prego
32	Bar scene	Tourist area	Drinks and smoking	Tourists
33	Bakery and pastry	Cake shops	Bread and pastry	Pastry
34	Sushi	Quality sushi	Friendly sushi places	Good sushi
35	Restaurant wait times	Table reservations	Food waiting time	Waiting for table
36	Restaurant dishes	Desserts and starters	Meat and chicken	Dish
37	Beer and snacks	Drinks with friends	Snail bar	Beer
38	Zomato gold	Zomato service	Cereal restaurant	Zomato discount
39	Sandwich shop	Ham and cheese sandwiches	Salad sandwiches	Sandwich
40	Room table and chair	Furniture and decor	Bar space	Table
41	Hamburgers and fries	Hamburguer	Potato dishes	Fries
42	Dog-friendly	Pancake	Friendly staff	Nice staff
43	Sauce	Steak dish	Tuna dish	Water
44	Wine list	Wine place	Hustle and bustle	Wine
45	Tacos	Chicken tacos	Sauces and guacamole	Mexican
46	Francesinha in porto	Sauce tasting	Eating space	Porto city
47	Salmon and shrimp order	Cheese sauces	Bread dish	Ordering
48	Paying	Atm and prices	Price	Payment and price
49	Seafood restaurant	Beach view	Sea view	Seafood
50	Trying out	Tea	Opportunities to eat	Space
51	Steak house	Sauce	House	Steak
52	Clams and prawns	Crab	Seafood	Clam
53	Terrace	Raman crepe	Parking	Location
54	Recommended food	Good service	Friendly staff	Good food
55	Experience	Service	Order	Space
56	Family restaurant	Dish recommendation	Friendly atmosphere	Dish
57	Chef	Space	Food	Experience
58	Pasta	Parking area	Nacoe flavors	Dish
59	Insufficient	Food	Wine	Bad experience
60	Dining rooms	Kitchen menus	Flavorful meals	Dish
61	Home delivery	Eating at home	Recommendation	Ordering
62	Table service	Serving time	Ordering time	Waiting time
63	Flavorful meats	Curry	Friendly staff	Food

1 Topic, Tip1, Tip2, Tip3

2 Sushi, "Offer a variety of sushi options, including vegetarian and
vegan options", "Use high-quality, fresh ingredients and prepare
the sushi in front of the customer", "Offer a sushi-making class or
workshop to engage customers and provide a unique experience"

- 3 Ordering and serving, "Implement an efficient ordering system, such as an online or mobile app, to reduce wait times", "Train staff to provide attentive and prompt service", "Offer complimentary items, such as bread or an amuse-bouche, while customers wait for their food"
- 4 Price-quality ratio, "Ensure that the prices accurately reflect the quality of the ingredients and the skill of the chefs", "Offer daily specials or promotions to attract customers", "Provide a range of menu options at different price points to cater to different budgets"
- 5 Customer experience, "Provide excellent customer service by training staff to be friendly, attentive, and knowledgeable about the menu", "Create a welcoming and comfortable atmosphere through decor and ambiance", "Offer loyalty programs or rewards to encourage repeat business"
- 6 Service quality, "Train staff to provide prompt and attentive service", "Ensure that staff are knowledgeable about the menu and can make recommendations", "Address any customer complaints or issues promptly and professionally"
- 7 Restaurant experience, "Create a unique and memorable dining experience through decor, ambiance, and menu offerings", "Offer live music or other entertainment to enhance the dining experience", "Provide outdoor seating or a patio area for customers to enjoy their meal al fresco"
- 8 Ice cream flavors, "Offer a variety of ice cream flavors, including classic favorites and seasonal or unique options", "Use high-quality, fresh ingredients to create delicious and creamy ice cream", "Offer toppings and mix-ins for customers to customize their ice cream"
- 9 Menu options, "Provide a diverse range of menu options to cater to different tastes and dietary requirements", "Update the menu regularly to incorporate seasonal ingredients and keep customers interested", "Offer daily specials or chef's recommendations to showcase the skills of the kitchen staff"
- 10 Pizza, "Use high-quality ingredients, such as fresh mozzarella and San Marzano tomatoes, to create delicious pizzas", "Offer a variety of pizza toppings, including classic favorites and unique combinations", "Provide a range of crust options, such as thin-crust or deep-dish, to cater to different preferences"
- 11 Cake and bread, "Bake cakes and breads fresh daily using high-quality ingredients", "Offer a variety of cake flavors and styles, including classic favorites and seasonal options", "Provide options for special dietary requirements, such as gluten-free or vegan cakes"
- 12 Breakfast, "Offer a range of breakfast options, including classic dishes like pancakes or eggs benedict as well as healthy options like smoothie bowls or avocado toast", "Use high-quality ingredients, such as free-range eggs and locally-sourced produce,

- to create delicious breakfast dishes","Provide coffee and tea options from local roasters or tea shops"
- 13 Pizza and pasta,"Offer a variety of pizza and pasta dishes using high-quality ingredients like fresh mozzarella and San Marzano tomatoes","Provide options for different dietary requirements, such as gluten-free pasta or vegan pizza toppings","Incorporate seasonal ingredients into pizza and pasta dishes to keep the menu fresh and interesting"
- 14 Group dinner,"Provide large-format dishes or family-style meals for groups to share","Offer group discounts or promotions to attract larger parties","Create a welcoming atmosphere with large tables or private dining rooms for groups"
- 15 Booking,"Implement an easy-to-use booking system, such as an online reservation platform or mobile app","Provide confirmation emails or text messages to remind customers of their reservation","Offer incentives for booking in advance, such as a complimentary appetizer or dessert"
- 16 Salad and rice,"Offer a variety of salad options using fresh, locally-sourced produce","Incorporate grains like quinoa or farro into salads for added texture and nutrition","Provide dressing options on the side so customers can customize their salad"
- 17 Music atmosphere,"Create a playlist of music that fits the theme and ambiance of the restaurant","Consider hiring live musicians or DJs to provide entertainment","Adjust the volume and tempo of the music to match the mood and energy of the dining room"
- 18 Chocolate cake,"Use high-quality chocolate and other ingredients to create a rich and decadent cake","Offer a variety of chocolate cake styles, such as flourless or molten lava cake","Pair the chocolate cake with complementary flavors, such as raspberry or coffee"
- 19 Pizza delivery,"Implement an efficient delivery system, such as partnering with a third-party delivery service or hiring dedicated delivery staff","Ensure that pizzas are delivered hot and fresh by using insulated delivery bags","Offer promotions or discounts for customers who order delivery"
- 20 Brunch menu,"Offer a range of brunch dishes, including classic favorites like eggs benedict or pancakes as well as unique options like shakshuka or breakfast pizza","Provide bottomless mimosa or bloody mary options for customers to enjoy","Incorporate seasonal ingredients into brunch dishes to keep the menu fresh and interesting"
- 21 View and drinks,"Create a comfortable and inviting outdoor space with a view for customers to enjoy drinks","Offer a variety of drink options, including cocktails, beer, wine, and non-alcoholic beverages","Provide small plates or snacks for customers to enjoy with their drinks"
- 22 Coffee terrace,"Create a comfortable and inviting outdoor space for customers to enjoy coffee","Offer a variety of coffee options, including espresso drinks and pour-over coffee from local roasters"

- ","Provide pastries or other baked goods for customers to enjoy with their coffee"
- 23 Grilled fish,"Use high-quality, fresh fish and seafood for grilling
","Offer a variety of grilled fish options, including whole fish and fillets","Pair grilled fish with complementary flavors, such as lemon or herbs"
- 24 Porto food,"Incorporate traditional Porto dishes into the menu, such as francesinha or tripas à moda do Porto","Use locally-sourced ingredients to create authentic Porto flavors","Offer wine pairings from the Porto region to complement the food"
- 25 Meat restaurant,"Offer a variety of high-quality meat options, including beef, pork, lamb, and poultry","Prepare meat using different cooking techniques, such as grilling or roasting","Pair meat dishes with complementary sides and sauces"
- 26 Restaurant service,"Train staff to provide excellent customer service by being friendly, attentive, and knowledgeable about the menu","Ensure that staff are able to handle customer complaints or issues promptly and professionally","Provide regular training and development opportunities for staff to improve their skills"
- 27 Service,"Train staff to provide prompt and attentive service","Ensure that staff are knowledgeable about the menu and can make recommendations","Address any customer complaints or issues promptly and professionally"
- 28 Lunch space,"Create a comfortable and inviting space for customers to enjoy lunch","Offer a range of lunch options, including sandwiches, salads, and hot dishes","Provide quick and efficient service for customers on a tight schedule"
- 29 Wine and cheese,"Offer a variety of wine and cheese pairings for customers to enjoy","Provide information about the origin and flavor profile of each wine and cheese option","Create a comfortable and inviting atmosphere for customers to relax and enjoy their wine and cheese"
- 30 Shopping center,"Consider opening a location in a shopping center to attract customers who are shopping or running errands","Offer quick and convenient menu options for customers on-the-go","Provide comfortable seating for customers who want to take a break from shopping"
- 31 Food variety,"Offer a diverse range of menu options to cater to different tastes and dietary requirements","Update the menu regularly to incorporate seasonal ingredients and keep customers interested","Provide daily specials or chef's recommendations to showcase the skills of the kitchen staff"
- 32 Portuguese food,"Incorporate traditional Portuguese dishes into the menu, such as bacalhau or pastéis de nata","Use locally-sourced ingredients to create authentic Portuguese flavors","Offer wine pairings from Portuguese regions to complement the food"
- 33 Prego,"Offer a variety of prego options, including different types of meat and toppings","Use high-quality ingredients, such as fresh

- bread and tender meat, to create delicious pregos", "Pair pregos with complementary sides, such as fries or salad"
- 34 Tourist area, "Consider opening a location in a tourist area to attract visitors", "Offer menu options that showcase local cuisine and culture", "Provide information about the local area, such as maps or recommendations for attractions"
- 35 Bakery and pastry, "Bake breads and pastries fresh daily using high-quality ingredients", "Offer a variety of baked goods, including classic favorites and seasonal options", "Provide options for special dietary requirements, such as gluten-free or vegan baked goods"
- 36 Quality sushi, "Use high-quality, fresh ingredients to create delicious sushi", "Offer a variety of sushi options, including classic rolls and unique combinations", "Prepare sushi in front of the customer to showcase the skill of the chefs"
- 37 Restaurant wait times, "Implement an efficient seating system to reduce wait times for customers", "Offer complimentary items, such as bread or an amuse-bouche, while customers wait for their table", "Provide a comfortable waiting area with seating and refreshments for customers"
- 38 Restaurant dishes, "Offer a diverse range of menu options to cater to different tastes and dietary requirements", "Use high-quality ingredients and cooking techniques to create delicious dishes", "Update the menu regularly to incorporate seasonal ingredients and keep customers interested"
- 39 Beer and snacks, "Offer a variety of beer options, including local craft beers and international favorites", "Pair beer with complementary snacks, such as nuts or pretzels", "Create a comfortable and inviting atmosphere for customers to relax and enjoy their beer and snacks"
- 40 Zomato gold, "Consider participating in the Zomato Gold program to attract customers who are members", "Offer exclusive deals or promotions for Zomato Gold members", "Ensure that staff are trained to handle Zomato Gold redemptions smoothly"
- 41 Sandwich shop, "Offer a variety of sandwich options using high-quality ingredients like fresh bread and deli meats", "Provide options for different dietary requirements, such as vegetarian or gluten-free sandwiches", "Incorporate seasonal ingredients into sandwich offerings to keep the menu fresh and interesting"
- 42 Furniture and decor, "Choose furniture and decor that fits the theme and ambiance of the restaurant", "Ensure that furniture is comfortable and functional for customers", "Update decor regularly to keep the restaurant looking fresh and inviting"
- 43 Hamburgers and fries, "Use high-quality ingredients like fresh ground beef and hand-cut fries to create delicious hamburgers and fries", "Offer a variety of burger toppings, including classic favorites like cheese and bacon as well as unique options like avocado or fried egg", "Provide options for different dietary requirements, such as vegetarian or gluten-free burgers"

- 44 Dog-friendly, "Create a dog-friendly outdoor space for customers to enjoy their meal with their furry friends", "Provide amenities for dogs, such as water bowls or treats", "Ensure that staff are trained to handle dogs safely and respectfully"
- 45 Steak dish, "Offer a variety of steak cuts cooked to the customer's preference", "Use high-quality beef from reputable sources to ensure delicious flavor and tenderness", "Pair steak dishes with complementary sides and sauces"
- 46 Wine place, "Offer a diverse selection of wines from different regions and varietals", "Provide information about each wine, including tasting notes and food pairings", "Create a comfortable and inviting atmosphere for customers to relax and enjoy their wine"
- 47 Mexican, "Incorporate traditional Mexican dishes into the menu, such as tacos or enchiladas", "Use fresh ingredients like cilantro, lime, and avocado to create authentic Mexican flavors", "Offer margaritas or other Mexican-inspired cocktails to complement the food"
- 48 Francesinha in porto, "Offer francesinha on the menu using traditional Porto ingredients like linguiça sausage and tomato-beer sauce", "Pair francesinha with complementary sides like fries or salad", "Provide information about the history and origin of francesinha for customers who are unfamiliar with the dish"
- 49 Payment and price, "Offer a variety of payment options, including cash, credit/debit cards, and mobile payments", "Ensure that prices are clearly displayed and accurately reflect the quality of the food and service", "Provide a range of menu options at different price points to cater to different budgets"
- 50 Seafood restaurant, "Use high-quality, fresh seafood to create delicious dishes", "Offer a variety of seafood options, including fish, shellfish, and crustaceans", "Pair seafood dishes with complementary flavors, such as lemon or herbs"
- 51 Trying out, "Offer a tasting menu or small plates for customers to try a variety of dishes", "Provide recommendations for dishes to try based on the customer's preferences", "Encourage customers to try new or unique dishes by offering daily specials or chef's recommendations"
- 52 Steak house, "Offer a variety of steak cuts cooked to the customer's preference", "Use high-quality beef from reputable sources to ensure delicious flavor and tenderness", "Pair steak dishes with complementary sides and sauces"
- 53 Seafood, "Use high-quality, fresh seafood to create delicious dishes", "Offer a variety of seafood options, including fish, shellfish, and crustaceans", "Pair seafood dishes with complementary flavors, such as lemon or herbs"
- 54 Location, "Choose a location that is easily accessible and visible to potential customers", "Ensure that the restaurant has adequate parking or is located near public transportation", "Consider the surrounding area and target demographic when choosing a location"

- 55 Recommended food,"Provide recommendations for dishes based on the customer's preferences or dietary requirements","Train staff to be knowledgeable about the menu and able to make recommendations","Offer daily specials or chef's recommendations to showcase the skills of the kitchen staff"
- 56 Experience,"Create a unique and memorable dining experience through decor, ambiance, and menu offerings","Provide excellent customer service by training staff to be friendly, attentive, and knowledgeable about the menu","Offer live music or other entertainment to enhance the dining experience"
- 57 Dish recommendation,"Provide recommendations for dishes based on the customer's preferences or dietary requirements","Train staff to be knowledgeable about the menu and able to make recommendations","Offer daily specials or chef's recommendations to showcase the skills of the kitchen staff"
- 58 Experience,"Create a unique and memorable dining experience through decor, ambiance, and menu offerings","Provide excellent customer service by training staff to be friendly, attentive, and knowledgeable about the menu","Offer live music or other entertainment to enhance the dining experience"
- 59 Pasta,"Offer a variety of pasta dishes using high-quality ingredients like fresh pasta and San Marzano tomatoes","Provide options for different dietary requirements, such as gluten-free pasta or vegan sauce options","Incorporate seasonal ingredients into pasta dishes to keep the menu fresh and interesting"
- 60 Bad experience,"Address any customer complaints or issues promptly and professionally","Train staff to handle difficult situations calmly and respectfully","Offer compensation or incentives for customers who have had a bad experience to encourage them to return"
- 61 Dining rooms,"Create comfortable and inviting dining rooms with appropriate lighting and decor","Ensure that tables are spaced appropriately for privacy and comfort","Provide options for private dining rooms or large group seating"
- 62 Home delivery,"Implement an efficient delivery system, such as partnering with a third-party delivery service or hiring dedicated delivery staff","Ensure that food is delivered hot and fresh by using insulated delivery bags","Offer promotions or discounts for customers who order delivery"
- 63 Serving time,"Implement an efficient ordering system to reduce wait times for food","Train kitchen staff to prepare dishes quickly without sacrificing quality","Offer complimentary items, such as bread or an amuse-bouche, while customers wait for their food"
- 64 Flavorful meats,"Use high-quality meats from reputable sources to ensure delicious flavor and tenderness","Prepare meats using different cooking techniques, such as grilling or roasting, to enhance their flavor","Pair meat dishes with complementary sides and sauces"

Code 4: Tips.csv file. This includes the list tips associated with each topic

A.5 Opinion of UMa's General Data Protection Regulation



PARECER

Na qualidade de Encarregado de Proteção de Dados da Universidade da Madeira, e após análise pela equipa de proteção de dados desta instituição, do Projeto de Investigação ***“Understanding users’ choice of topic naming and reporting style”*** é nosso parecer que o mesmo cumpre com o Regulamento Geral da Proteção de Dados.

Funchal, 30 de junho de 2023

Encarregado de Proteção de Dados

Assinado por: **Filipa Isabel Sousa Mota**
Num. de Identificação: 10993715
Data: 2023.06.30 10:48:44+01'00'



Pdf 1: Opinion of UMa's General Data Protection Regulation

A.6 Graphs Comparison Survey (English)

Comparison between multiple types of data presentation for the RRSO (Restaurant Review Sentiment Output) Project

Welcome to the comparison between multiple types of data visualization for M.Sc. thesis and RRSO (Restaurant Review Sentiment Output) project. Your participation is valuable and will help us understand the best way to generate multiple forms of data visualization for an automatically generated report. The goal is to assist establishment owners, such as restaurants, in interpreting their review data in a concise and objective manner using the most intuitive form of conveying information. Completing this form should take approximately 7 minutes.

In this form, you will be presented with various forms of data visualization and will be asked to choose your preferred one. Additionally, you will also be asked about the ease with which you were able to extract information on a given topic using the provided visual representations.

The information collected will be used exclusively within the context of the project and to ensure confidentiality, your name will not be associated with the collected data. **By responding to this form, you are agreeing to the aforementioned terms.** Please remember that you can **choose to exit the form at any time.**

We appreciate your cooperation in advance.

* Indicates required question

Demographic information

This section aims to inquire the participant about relevant demographic data for the survey.

<https://docs.google.com/forms/d/19APbzRLKND1Wqa4tS8tRqr-VXegTR-m3h3uJdEBzCQ/edit>

1/22

8/28/23, 4:01 PM

Comparison between multiple types of data presentation for the RRSO (Restaurant Review Sentiment Output) Project

3. How old are you? *

Mark only one oval.

- ☐ 15 or less
- ☐ 16 - 20
- ☐ 21 to 30
- ☐ 31 to 40
- ☐ 41 to 50
- ☐ 50 or more
- ☐ I'd rather not say

4. What is your level of fluency in English? *

If you don't know about English proficiency levels, please click [here](#)

Mark only one oval.

- ☐ No English knowledge
- ☐ Basic User (A1, A2)
- ☐ Independent User (B1, B2)
- ☐ Proficient User (C1, C2)
- ☐ Native

Feedback and Evaluation Habits

This section aims to gather information about the participant's habits and behavior when providing feedback and evaluating restaurants and other purchases or contracted services.

<https://docs.google.com/forms/d/19APbzRLKND1Wqa4tS8tRqr-VXegTR-m3h3uJdEBzCQ/edit>

3/22

1. Please indicate your country of birth. *

Mark only one oval.

- ☐ England
- ☐ France
- ☐ Germany
- ☐ Italy
- ☐ Portugal
- ☐ Spain
- ☐ I'd rather not say
- ☐ Other: _____

2. What is your gender? *

Mark only one oval.

- ☐ Man
- ☐ Woman
- ☐ Other
- ☐ I'd rather not say
- ☐ Other: _____

<https://docs.google.com/forms/d/19APbzRLKND1Wqa4tS8tRqr-VXegTR-m3h3uJdEBzCQ/edit>

2/22

8/28/23, 4:01 PM

Comparison between multiple types of data presentation for the RRSO (Restaurant Review Sentiment Output) Project

5. Have you ever owned, operated or held a decision-making position in a food establishment, either in the past or currently? *

Mark only one oval.

- ☐ Yes
- ☐ No

6. How often do you use platforms to find restaurants? e.g. TripAdvisor, Zomato, Foursquare, The Fork, Yelp or equivalents *

Mark only one oval.

- ☐ Never
- ☐ Rarely
- ☐ Sometimes
- ☐ Often
- ☐ Always

7. How often do you give your opinion in the form of a numerical evaluation or equivalent (e.g. 1 - 5 or ★★★★★) about your purchases or contracted services? *

Mark only one oval.

- ☐ Never
- ☐ Rarely
- ☐ Sometimes
- ☐ Often
- ☐ Always

<https://docs.google.com/forms/d/19APbzRLKND1Wqa4tS8tRqr-VXegTR-m3h3uJdEBzCQ/edit>

4/22

8. **How often** do you give your opinion in **textual form** (e.g. comment, review) about * your purchases or contracted services?

Mark only one oval.

- ☐ Never
☐ Rarely
☐ Sometimes
☐ Often
☐ Always

9. When you **rate a product or service**, which of the following best describes **your approach?** *

Mark only one oval.

- ☐ I tend to rate it lower than what I think it deserves Skip to question 10
☐ I rate it exactly as I think it deserves Skip to question 12
☐ I tend to rate it higher than what I think it deserves Skip to question 11

I tend to rate lower than what my assessment suggests

10. Please indicate any factors that may influence you to provide a **lower rating** for * a product/service **than your actual assessment** of its quality.

Check all that apply.

- ☐ I want to publicly shame them
☐ I want to come across as strong.
☐ I want to discourage the product/service supplier
☐ I desire an explanation from the supplier regarding any negative aspects of their product/service.
☐ Other: _____

Skip to question 12

I tend to rate higher than what my assessment suggests

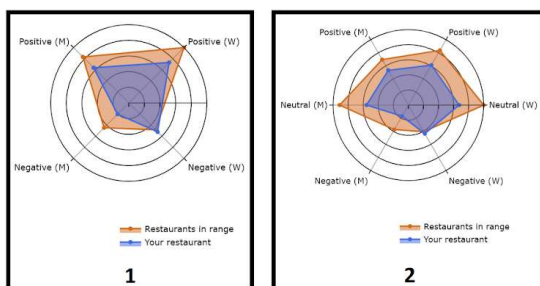
12. **Radar chart (Your restaurant vs restaurants in range)** *

Radar charts 1 and 2 compare the emotions expressed in reviews of **your restaurant**, in **blue** to those of **nearby restaurants**, in **orange**, split by gender.

Radar chart 1 shows the **positive and negative emotions** expressed by men and women.

Radar chart 2 shows the **positive, neutral, and negative emotions** expressed by men and women.

Which radar graph do you prefer for visualizing this data?



Mark only one oval.

- ☐ Radar chart 1
☐ Radar chart 2

11. Please indicate any factors that may influence you to provide a **higher rating** * for a product/service **than your actual assessment** of its quality.

Check all that apply.

- ☐ I don't want to publicly shame them
☐ I want to come across as kind.
☐ I want to encourage the product/service supplier
☐ I don't want to be contacted by the product/service supplier
☐ Other: _____

Skip to question 12

Choosing the Most Effective Graph for a Restaurant Owner

The purpose of this form is to identify the data visualization method that effectively communicates various metrics to restaurant owners, providing them with valuable insights about their establishments.

For each question, **choose the image** that you believe **best represents the described information**.

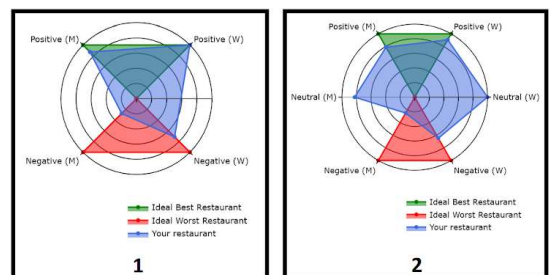
13. **Radar chart (Your restaurant vs ideal best and worst)** *

Radar Charts 1 and 2 compare the emotions expressed in reviews of **your restaurant**, in **blue** to those of **ideal best**, in **green** and **ideal worst**, in **red**, restaurants.

Radar chart 1 shows the **positive and negative emotions** expressed by men and women, **compared to** the emotions of an **ideal best** and **ideal worst** restaurant.

Radar chart 2 shows the **positive, neutral, and negative emotions** expressed by men and women, compared to the emotions of an **ideal best** and **ideal worst** restaurant.

Which radar graph do you prefer for visualizing this data?



Mark only one oval.

- ☐ Radar chart 1
☐ Radar chart 2

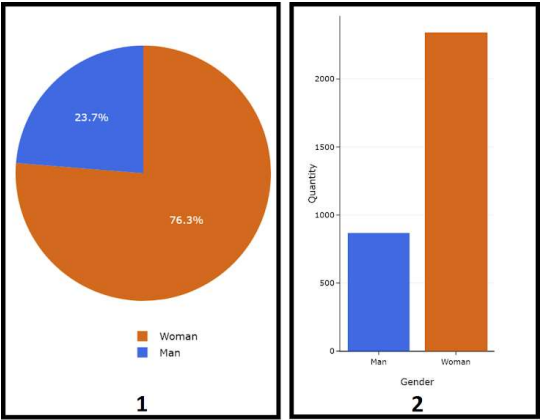
14. Gender distribution (Pie vs Bar) *

Graphs 1 and 2 show the **gender distribution** in reviews of your restaurant.

Graph 1 presents this information in the form of a **pie chart**.

Graph 2 presents this information in the form of a **bar chart**.

Which chart do you prefer for visualizing this data?



Mark only one oval.

- ☐ Chart 1
- ☐ Chart 2

15. Gender distribution Split (Bar vs Table) *

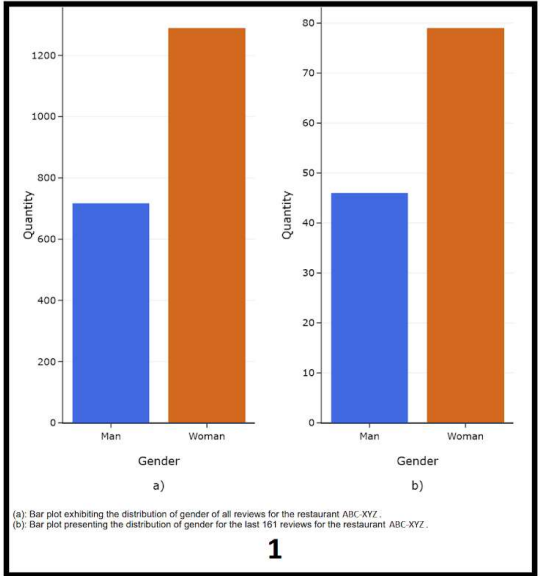
Figure 1 and Table 2 show the **gender distribution** of customers at your restaurant over its **entire existence** and over the **past 24 months**.

Figure 1 presents this information in the form of 2 bar charts.

Table 2 presents this information in the form of a table.

Which figure do you prefer for visualizing this data?

- ☐ Table 2



(a): Bar plot exhibiting the distribution of gender of all reviews for the restaurant ABC XYZ.
(b): Bar plot presenting the distribution of gender for the last 161 reviews for the restaurant ABC XYZ.

	Nº of Man - (%)	Nº of Woman - (%)
All reviews	717 (35.7%)	1289 (64.3%)
Last 24 months	46 (36.8%)	79 (63.2%)

Mark only one oval.

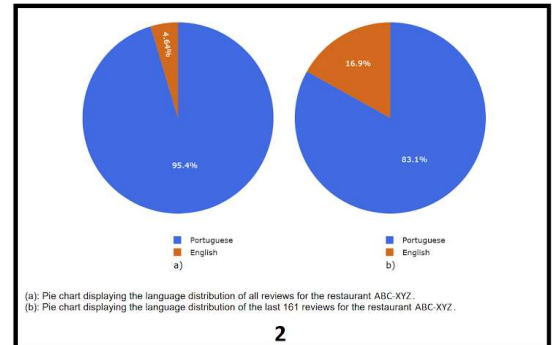
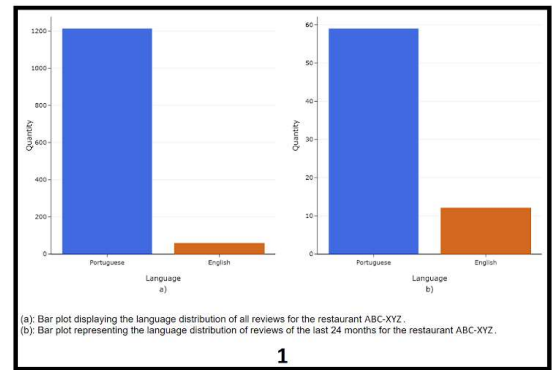
- ☐ Figure 1

16. Figures 1 and 2 offer a **distribution of the language** used in reviews of your restaurant, * considering its **entire existence** and the **last 24 months**.

Figure 1 represents this data in bar form.

Figure 2 represents this data in a pie chart.

Which figure do you find preferable?



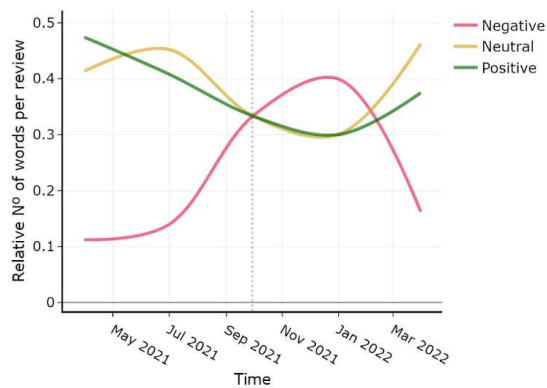
Mark only one oval.

☐ Figure 1

☐ Figure 2

17. The graph below shows the relative number of positive, neutral, and negative words **per review** over a period of time. *

On a scale of **1 to 5**, how well do you understand that the **number of positive words per review is increasing** while the **number of negative words per review is decreasing**?



Mark only one oval.

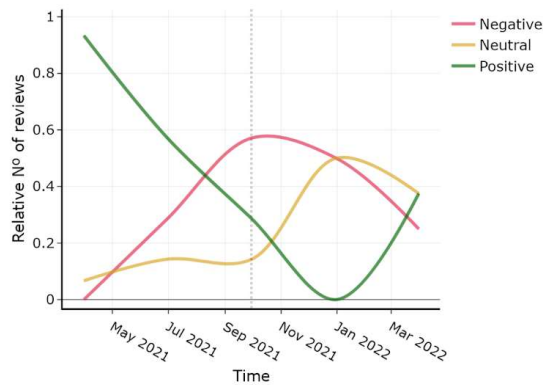
Not at all

- 1 ☐
- 2 ☐
- 3 ☐
- 4 ☐
- 5 ☐

Completly

18. The graph below shows the relative number of positive, neutral, and negative review ratings over a period of time. *

On a scale of 1 to 5, how well do you understand that the number of positive reviews is increasing while the number of negative reviews is decreasing?



Mark only one oval.

Not at all

1 ☐

2 ☐

3 ☐

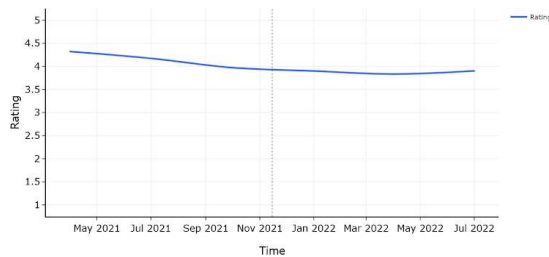
4 ☐

5 ☐

Completely

20. The graph below shows the average review rating of your restaurant over a period of time. *

On a scale of 1 to 5, how well do you understand that your review ratings have demonstrated a slight decrease since November 2021 compared to the period before?



Mark only one oval.

Not at all

1 ☐

2 ☐

3 ☐

4 ☐

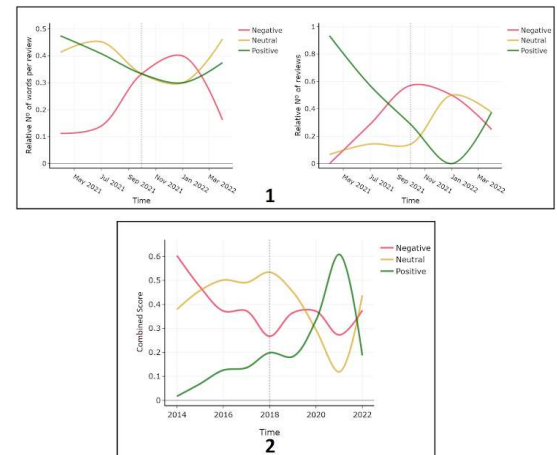
5 ☐

Completely

19. Which figure would you prefer to visualize? *

Figure 1: Contains two graphs - one showing the relative number of positive, neutral, and negative words per review, and another showing the relative number of positive, neutral, and negative review ratings.

Figure 2: Contains a single graph combining both graphs in Figure 1 into a Combined score.



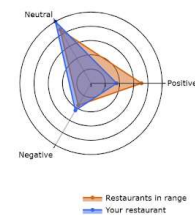
Mark only one oval.

☐ Figure 1

☐ Figure 2

21. The radar chart below compares the distribution of positive, neutral, and negative emotions in the last 17 reviews of your restaurant, in blue, in comparison to others, in orange, within a 250-meter radius. *

On a scale of 1 to 5, how well do you understand that your restaurant is lacking in positive emotions when compared to the restaurants in range?



Radar chart comparing the distribution of emotions in the last 17 reviews for the restaurant ABC-XYZ with those of others in a 250 meters range.

Mark only one oval.

Not at all

1 ☐

2 ☐

3 ☐

4 ☐

5 ☐

Completely

Pdf 2: Graphs Comparison Survey (English)

A.7 Graphs Comparison Survey (Portuguese)

Formulário de Comparação entre vários tipos de visualização de dados para o Projeto RRSO (Restaurant Review Sentiment Output)

Bem-vindo ao formulário de comparação entre vários tipos de visualização de dados para tese de Mestrado e o projeto RRSO (Restaurant Review Sentiment Output). A sua participação é valiosa e tem como objetivo nos ajudar a compreender a melhor forma de visualizar dados para um relatório gerado automaticamente. O objetivo é auxiliar os proprietários de estabelecimentos, como restaurantes, a interpretar seus dados de forma concisa e objetiva, utilizando a forma mais intuitiva de transmitir informação. Completar este formulário deve demorar aproximadamente 7 minutos.

Neste formulário, serão apresentadas várias formas de visualização de dados e será solicitado que escolha a sua preferida. Além disso, também será questionado sobre a facilidade com que conseguiu extrair informações sobre um determinado tópico usando as representações visuais fornecidas.

As informações recolhidas serão usadas exclusivamente no contexto do projeto e, para garantir a confidencialidade, o seu nome não será associado aos dados recolhidos. **Ao responder a este formulário, está concordando com os termos acima mencionados.** Lembre-se de que **pode optar por sair do formulário a qualquer momento.**

Agradecemos antecipadamente a sua cooperação.

* Indicates required question

Informação demográfica

Esta seção tem como objetivo inquirir o participante sobre dados demográficos relevantes para a pesquisa.

3. Por favor, indique a sua idade *

Mark only one oval.

- ☐ 15 ou menos
- ☐ 16 a 20
- ☐ 21 a 30
- ☐ 31 a 40
- ☐ 41 a 50
- ☐ 50 ou mais
- ☐ Prefiro não dizer

4. Por favor indique o seu nível de fluência em inglês. *
Se não conhece os níveis de fluência em inglês, clique [aqui](#)

Mark only one oval.

- ☐ Sem conhecimento de inglês
- ☐ Utilizador Básico (A1, A2)
- ☐ Utilizador Independente (B1, B2)
- ☐ Utilizador Proficiente (C1, C2)
- ☐ Utilizador Nativo

Feedback e hábitos de avaliação

Esta seção tem como objetivo reunir informações sobre os hábitos e comportamentos do participante ao fornecer feedback e avaliar restaurantes e outras compras ou serviços contratados.

5. Já foi proprietário, operou ou ocupou um cargo de tomada de decisão num estabelecimento no ramo da alimentação, seja no passado ou atualmente? *

Mark only one oval.

- ☐ Sim
- ☐ Não

1. Por favor, indique o seu país de nascimento. *

Mark only one oval.

- ☐ Inglaterra
- ☐ França
- ☐ Alemanha
- ☐ Itália
- ☐ Portugal
- ☐ Espanha
- ☐ Prefiro não dizer
- ☐ Other: _____

2. Por favor, indique o seu género *

Mark only one oval.

- ☐ Homem
- ☐ Mulher
- ☐ Outro
- ☐ Prefiro não dizer
- ☐ Other: _____

6. Com que frequência usa plataformas para encontrar restaurantes? *
Por exemplo, TripAdvisor, Zomato, Foursquare, The Fork, Yelp ou equivalentes

Mark only one oval.

- ☐ Nunca
- ☐ Raramente
- ☐ Às vezes
- ☐ Frequentemente
- ☐ Sempre

7. Com que frequência dá a sua opinião sob a forma de avaliação numérica ou equivalente (por exemplo, 1 - 5 ou ★★★★★) sobre as suas compras ou serviços contratados? *

Mark only one oval.

- ☐ Nunca
- ☐ Raramente
- ☐ Às vezes
- ☐ Frequentemente
- ☐ Sempre

8. Com que frequência dá a sua opinião sob a forma textual (por exemplo, comentário, avaliação) sobre suas compras ou serviços contratados? *

Mark only one oval.

- ☐ Nunca
- ☐ Raramente
- ☐ Às vezes
- ☐ Frequentemente
- ☐ Sempre

9. Quando **avalia um produto ou serviço**, qual das seguintes opções melhor descreve a **sua abordagem**? *

Mark only one oval.

- ☐ Tendo a avaliá-lo abaixo do que acho que merece Skip to question 10
- ☐ Avalio-o exatamente como acho que merece Skip to question 12
- ☐ Tendo a avaliá-lo acima do que acho que merece Skip to question 11

Eu tendo a avaliar abaixo do que minha avaliação sugere

10. Por favor, **mencione os fatores** que podem fazer com que avalie um produto/serviço com uma **nota inferior à que acha que o mesmo realmente merece**. *

Check all that apply.

- ☐ Quero envergonhá-los publicamente
- ☐ Quero parecer forte
- ☐ Quero desencorajar o fornecedor do produto/serviço
- ☐ Desejo uma explicação do fornecedor sobre quaisquer aspectos negativos do seu produto/serviço.
- ☐ Other: _____

Skip to question 12

Eu tendo a avaliar acima do que minha avaliação sugere

11. Por favor, **mencione os fatores** que podem fazer com que avalie um produto/serviço com uma **nota superior à que acha que o mesmo realmente merece**. *

Check all that apply.

- ☐ Não quero envergonhá-los publicamente
- ☐ Quero parecer gentil
- ☐ Quero encorajar o fornecedor do produto/serviço
- ☐ Não quero ser contactado pelo fornecedor do produto/serviço
- ☐ Other: _____

Skip to question 12

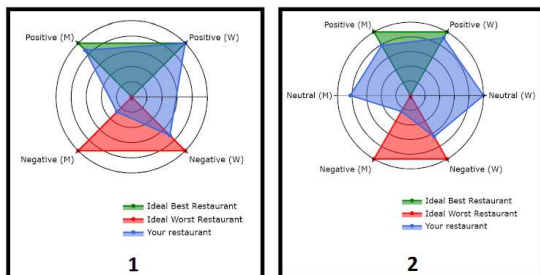
13. **Gráfico de radar (O seu restaurante vs ideal melhor e ideal pior)** *

Os gráficos de radar 1 e 2 comparam as emoções expressas nas avaliações do **seu restaurante, em azul**, com as de **restaurantes ideais melhores, em verde**, e **ideais piores, em vermelho**.

O **gráfico de radar 1** mostra as **emoções positivas e negativas** expressas por homens e mulheres, comparadas às **emoções de um restaurante ideal melhor e ideal pior**.

O **gráfico de radar 2** mostra as **emoções positivas, neutras e negativas** expressas por homens e mulheres, comparadas às **emoções de um restaurante ideal melhor e ideal pior**.

Qual o gráfico de radar que prefere para visualizar estes dados?



Mark only one oval.

- ☐ Gráfico de radar 1
- ☐ Gráfico de radar 2

Escolha o Gráfico Mais Eficaz para um Proprietário de Restaurante.

O objetivo deste formulário é identificar o método de visualização de dados que comunica efetivamente várias métricas aos proprietários de restaurantes, fornecendo-lhes informações valiosas sobre seus estabelecimentos.

Para cada pergunta, **escolha a imagem** que acredita **representar melhor as informações descritas**.

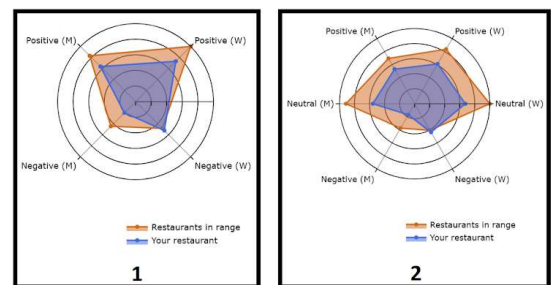
12. **Gráfico de radar (O seu restaurante vs restaurantes na área)** *

Os gráficos de radar 1 e 2 comparam as emoções expressas nas avaliações do **seu restaurante, em azul**, com as de **restaurantes próximos, em laranja**, divididas por género.

O **gráfico de radar 1** mostra as **emoções positivas e negativas** expressas por homens e mulheres.

O **gráfico de radar 2** mostra as **emoções positivas, neutras e negativas** expressas por homens e mulheres.

Qual o gráfico de radar que prefere para visualizar estes dados?



Mark only one oval.

- ☐ Gráfico de radar 1
- ☐ Gráfico de radar 2

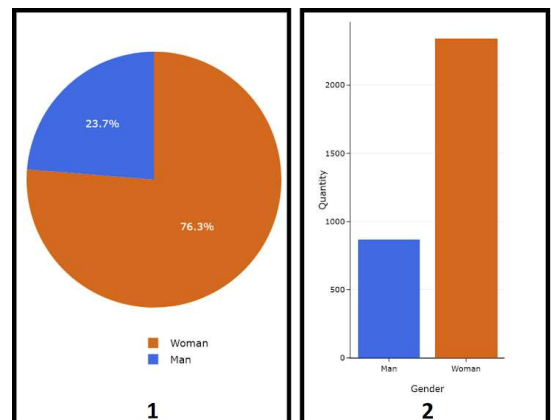
14. **Distribuição de género (Gráfico Circular vs Barra)** *

Os gráficos 1 e 2 mostram a **distribuição de género** nas avaliações do seu restaurante.

O **gráfico 1** apresenta esta informação na forma de um **gráfico circular**.

O **gráfico 2** apresenta esta informação na forma de um **gráfico de barras**.

Qual o gráfico que prefere para visualizar estes dados?



Mark only one oval.

- ☐ Gráfico 1
- ☐ Gráfico 2

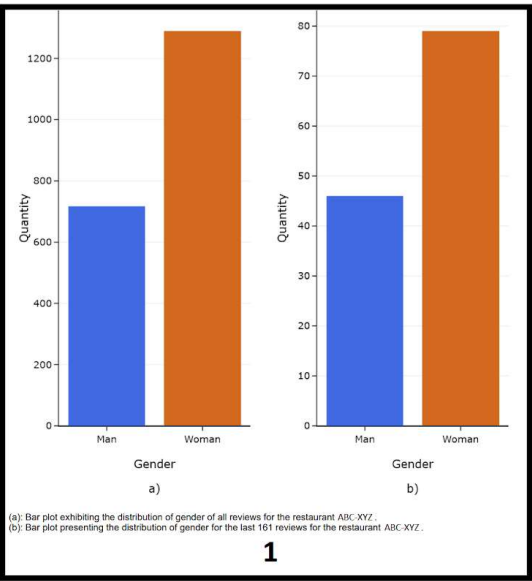
15. Distribuição de género dividida (Barra vs Tabela)

A Figura 1 e a Tabela 2 mostram a **distribuição de género** dos clientes do seu restaurante durante **toda a sua existência** e **nos últimos 24 meses**.

A **Figura 1** apresenta esta informação na forma de **2 gráficos de barras**.

A **Tabela 2** apresenta esta informação na forma de uma **tabela**.

Que figura prefere para visualizar estes dados?



2

	Nº of Man - (%)	Nº of Woman - (%)
All reviews	717 (35.7%)	1289 (64.3%)
Last 24 months	46 (36.8%)	79 (63.2%)

Mark only one oval.

☐ Figura 1

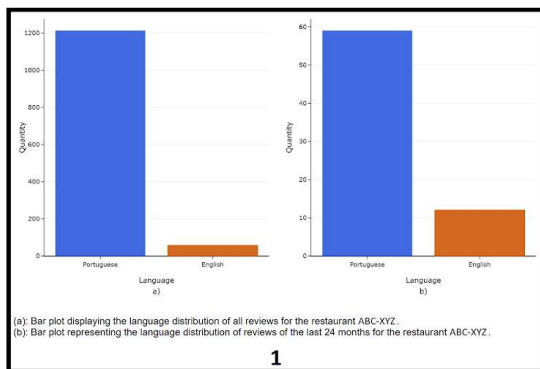
☐ Tabela 2

16. As Figuras 1 e 2 mostram a **distribuição da idioma** usada nas avaliações do seu restaurante, considerando **toda a sua existência** e **os últimos 24 meses**.

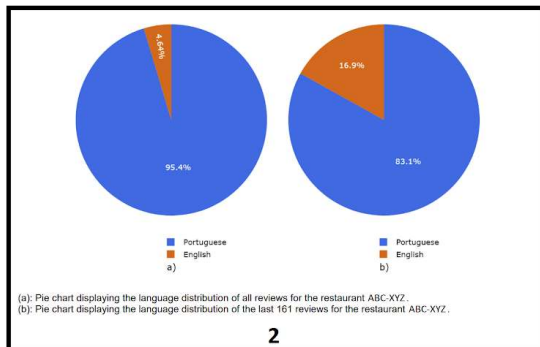
A **Figura 1** representa esses dados na forma de **gráfico de barras**.

A **Figura 2** representa esses dados na forma de um **gráfico circular**.

Que figura prefere?



1



2

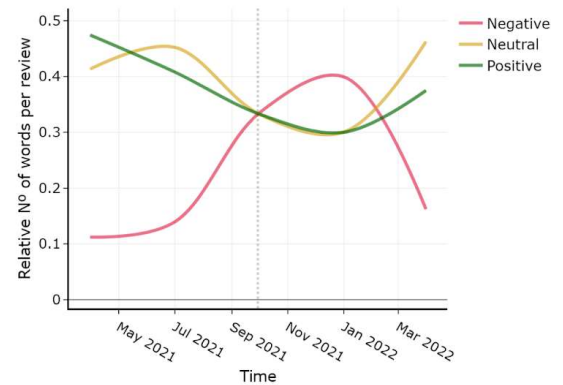
Mark only one oval.

- ☐ Figura 1
☐ Figura 2

Completamente

17. O gráfico abaixo mostra o **número relativo de palavras positivas, neutras e negativas** por avaliação ao longo de um período de tempo. *

Numa escala de 1 a 5, quão bem percebe que o **número de palavras positivas por avaliação está a aumentar** enquanto o **número de palavras negativas por avaliação está a diminuir**?



Mark only one oval.

Nada

1

2

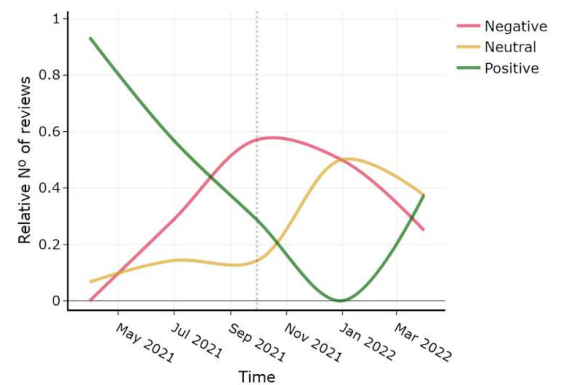
3

4

5

18. O gráfico abaixo mostra o **número relativo de avaliações positivas, neutras e negativas** ao longo de um período de tempo. *

Numa escala de 1 a 5, quão bem percebe que o **número de avaliações positivas está a aumentar** enquanto o **número de avaliações negativas está a diminuir**?



Mark only one oval.

Nada

1

2

3

4

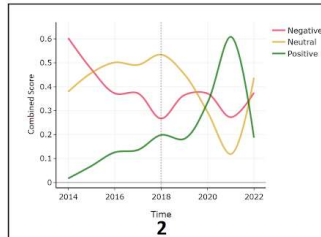
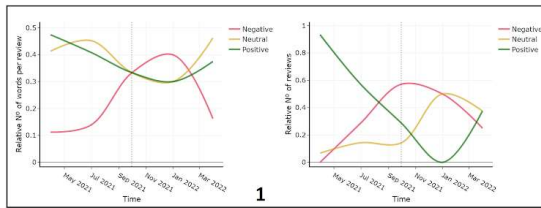
5

Completamente

19. Escolha a figura que prefere. *

Figura 1: Contém dois gráficos - um que mostra o **número relativo de palavras** positivas, neutras e negativas por avaliação e outro que mostra o **número relativo de avaliações** positivas, neutras e negativas.

Figura 2: Contém um único gráfico que **combina ambos os gráficos** da Figura 1 numa **pontuação combinada**.

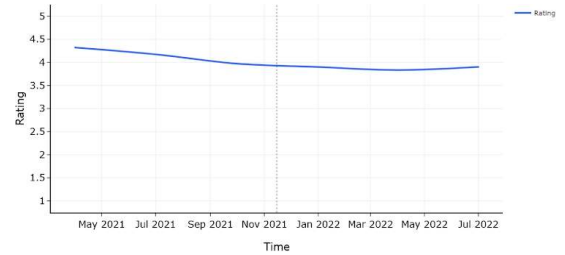


Mark only one oval.

- ☐ Figura 1
☐ Figura 2

20. O gráfico abaixo mostra a **classificação média** de avaliação do seu restaurante ao longo de um período de tempo. *

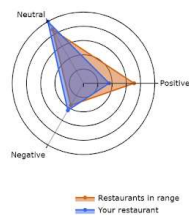
Numa escala de **1 a 5**, quão bem percebe que as pontuação das suas avaliações demonstra uma **ligeira diminuição desde novembro de 2021 em comparação com o período anterior?**



Mark only one oval.

- Nada
☐
 1 ☐
 2 ☐
 3 ☐
 4 ☐
 5 ☐
 Completamente

21. O gráfico de radar abaixo compara a **distribuição de emoções positivas, neutras e negativas** nas últimas 17 avaliações do seu restaurante, em azul, em comparação com outros, em laranja, num raio de 250 metros. *
- Numa escala de **1 a 5**, quão bem percebe que o seu restaurante **tem menos emoções positivas em comparação com os restaurantes nas imediações?**



Radar chart comparing the distribution of emotions in the last 17 reviews for the restaurant ABC-XYZ with those of others in a 250 meters range.

Mark only one oval.

- Nada
☐
 1 ☐
 2 ☐
 3 ☐
 4 ☐
 5 ☐
 Completamente

A.8 ASTA Versus Chat GPT Survey (English)

Comparison Form between Texts Generated by LLMs and Texts Created with String Combination for the RRSO (Restaurant Review Sentiment Output) Project

Welcome to the comparison form between texts generated by LLMs and texts created with string combination for M.Sc. thesis and the RRSO (Restaurant Review Sentiment Output) project. Your participation is valuable and will help us understand the best way to generate sentences for an automatically generated report. The goal is to assist establishment owners, including restaurants, in interpreting their review data in a concise and objective manner using Natural Language Processing (NLP). Completing this form should take approximately 10 minutes.

In this form, you will be presented with multiple versions of example paragraphs and will be asked to choose your preferred one.

The information collected will be used exclusively within the context of the project and to ensure confidentiality, your name will not be associated with the collected data. **By responding to this form, you are agreeing to the aforementioned terms.** Please remember that you can **choose to exit the form at any time.**

We appreciate your cooperation in advance.

* Indicates required question

Demographic information

This section aims to inquire the participant about relevant demographic data for the survey.

3. How old are you? *

Mark only one oval.

- ☐ 15 or less
- ☐ 16 to 20
- ☐ 21 to 30
- ☐ 31 to 40
- ☐ 41 to 50
- ☐ 50 or more
- ☐ I'd rather not say

4. What is your level of fluency in English? *

If you don't know about English proficiency levels, please click [here](#)

Mark only one oval.

- ☐ No English knowledge
- ☐ Basic User (A1, A2)
- ☐ Independent User (B1, B2)
- ☐ Proficient User (C1, C2)
- ☐ Native

Feedback and Evaluation Habits

This section aims to gather information about the participant's habits and behavior when providing feedback and evaluating restaurants and other purchases or contracted services.

1. Please indicate your country of birth. *

Mark only one oval.

- ☐ England
- ☐ France
- ☐ Germany
- ☐ Italy
- ☐ Portugal
- ☐ Spain
- ☐ I'd rather not say
- ☐ Other: _____

2. What is your gender? *

Mark only one oval.

- ☐ Man
- ☐ Woman
- ☐ Other
- ☐ I'd rather not say
- ☐ Other: _____

5. Have you ever **owned, operated or held a decision-making position in a food establishment**, either in the **past or currently**? *

Mark only one oval.

- ☐ Yes
- ☐ No

6. **How often do you use platforms to find restaurants?** e.g. TripAdvisor, Zomato, Foursquare, The Fork, Yelp or equivalents *

Mark only one oval.

- ☐ Never
- ☐ Rarely
- ☐ Sometimes
- ☐ Often
- ☐ Always

7. **How often** do you give your opinion in the form of a **numerical evaluation or equivalent** (e.g. 1 - 5 or ★★★★★) about your purchases or contracted services? *

Mark only one oval.

- ☐ Never
- ☐ Rarely
- ☐ Sometimes
- ☐ Often
- ☐ Always

8/28/23, 4:00 PM

Comparison Form between Texts Generated by LLMs and Texts Created with String Combination for the RRSO (Restaurant ...

8. **How often** do you give your opinion in **textual form** (e.g. comment, review) about ***** your purchases or contracted services?

Mark only one oval.

- ☐ Never
☐ Rarely
☐ Sometimes
☐ Often
☐ Always

9. **How often** do you use **LLMs** (Large Language Models) like ChatGPT, GPT-3, GPT-***** 4, BARD, LLaMA or others?

Mark only one oval.

- ☐ Never
☐ Rarely
☐ Sometimes
☐ Often
☐ Always

10. When you **rate a product or service**, which of the following best describes **your** ***** **approach**?

Mark only one oval.

- ☐ I tend to rate it lower than what I think it deserves *Skip to question 11*
☐ I rate it exactly as I think it deserves *Skip to question 13*
☐ I tend to rate it higher than what I think it deserves *Skip to question 12*

I tend to rate lower than what my assessment suggests

<https://docs.google.com/forms/d/1btHv1EyOwT0Hmk3kxTALX-0PH7mIFuOLl0ovfcBWkbtw/edit>

5/11

8/28/23, 4:00 PM

Comparison Form between Texts Generated by LLMs and Texts Created with String Combination for the RRSO (Restaurant ...

13. Please select the paragraph you think is **better overall**: *****

Report summary A:

This report provides an in-depth analysis of ABC-XYZ, a restaurant located in Portugal. Data was collected from reviews between May 25th, 2020 and August 6th, 2022, a period of 36 months. 182 reviews were selected for analysis, and customers gave the restaurant an average rating of 4.29 out of 5. The predominant emotion expressed in reviews was Neutral, and the predominant gender among customers was female, with a ratio of 1 man for every 2 woman. The predominant language used in reviews was Portuguese. Positive reviews were 1.74 times more common than negative reviews, with the most common topics being Chocolate cake and Breakfast. Comparisons were made to other restaurants within a 250 meter radius, which had a predominant gender of Female, predominant emotion of Positive, predominant language of Portuguese, average price of 25.5, average rating of 3.9, and average number of reviews of 447.

Report summary B:

An evaluation of the past 182 reviews for ABC-XYZ from 2020-05-25 to 2022-08-06, spanning the last 36 months, indicates that the restaurant has performed reasonable, with an average rating of 4.29/5. The majority of customers are female, with a ratio of 0.48 men for every woman, and the main language used in reviews is Portuguese. Customer emotions lean towards the optimistic side, with roughly 1 male customer for every 2 female customer. The most frequently mentioned topics in positive reviews are Chocolate cake and Breakfast, while in negative reviews they are Space and Waiting for table. An analysis of restaurants within a radius of 250 meters reveals that the predominant gender is Female, the most commonly used language is Portuguese, and the dominant emotion is Positive. The average price is 25.5 Euros and the average rating is 3.9/5, based on a total of 21908 reviews and an average of 447 reviews per restaurant.

Mark only one oval.

- ☐ Paragraph A
☐ Paragraph B

<https://docs.google.com/forms/d/1btHv1EyOwT0Hmk3kxTALX-0PH7mIFuOLl0ovfcBWkbtw/edit>

7/11

8/28/23, 4:00 PM

Comparison Form between Texts Generated by LLMs and Texts Created with String Combination for the RRSO (Restaurant ...

11. Please indicate any factors that may influence you to provide **a lower rating** for ***** a product/service **than your actual assessment** of its quality.

Check all that apply.

- ☐ I want to publicly shame them
☐ I want to come across as strong.
☐ I want to discourage the product/service supplier
☐ I desire an explanation from the supplier regarding any negative aspects of their product/service.
☐ Other: _____

Skip to question 13

I tend to rate higher than what my assessment suggests

12. Please indicate any factors that may influence you to provide **a higher rating** for a product/service **than your actual assessment** of its quality. *****

Check all that apply.

- ☐ I don't want to publicly shame them
☐ I want to come across as kind.
☐ I want to encourage the product/service supplier
☐ I don't want to be contacted by the product/service supplier
☐ Other: _____

Skip to question 13

Choosing the Most Effective Paragraph for a Restaurant Owner

This form aims to determine which paragraph most effectively communicates information to a restaurant owner. The paragraphs are part of a report that will inform a restaurant owner about various subjects related to their establishment.

For each question, **choose the paragraph (A or B)** that you believe **best represents the information requested** in the question.

<https://docs.google.com/forms/d/1btHv1EyOwT0Hmk3kxTALX-0PH7mIFuOLl0ovfcBWkbtw/edit>

6/11

8/28/23, 4:00 PM

Comparison Form between Texts Generated by LLMs and Texts Created with String Combination for the RRSO (Restaurant ...

14. Please select the paragraph you think is **better overall**: *****

Gender section introduction A:

This report provides an analysis of the gender of customers at ABC-XYZ. It includes bar charts showing the number of reviews by gender, pie charts splitting the reviews into positive and negative parts by gender, and radar charts comparing the emotions of customers in the restaurant with those in range, using 2, Positive and Negative and 3 classes, Positive, Negative, and Neutral. The gender that is most common among the restaurant's customers is female, and the approximate ratio of men to woman customers at the restaurant is 1 man for every 2 woman.

Gender section introduction B:

This section provides an study of the gender distribution among patrons of ABC-XYZ and how it compares to that of the ideal best and ideal worst restaurants. It comprises bar charts that indicate the total number of reviews for each gender and radar charts split into 2 (Positive and Negative) and 3 (Positive, Neutral, and Negative) groups that display the ratio of men to women for each section. A radar chart that illustrates how ABC-XYZ compares to the ideal best and ideal worst restaurants is also available in this section. Gender analysis reveals that the restaurant is predominantly visited by individuals who identify as female, with approximately 1 men for every 2 woman.

Mark only one oval.

- ☐ Paragraph A
☐ Paragraph B

<https://docs.google.com/forms/d/1btHv1EyOwT0Hmk3kxTALX-0PH7mIFuOLl0ovfcBWkbtw/edit>

8/11

15. Please select the paragraph you think is **better overall**: *

Nearby restaurants section introduction A:
This report will provide the restaurant owner of ABC-XYZ with an in-depth analysis of the nearby restaurants within a radius of 250 meters. The Nearby Restaurant section of this report includes bar charts showing the distribution of gender, language and emotion of customers, as well as wordclouds of the positive and negative reviews of these restaurants. Additionally, this section contains information such as the average price of nearby restaurants (25.5), their average rating (3.9), the total number of reviews (21908), the average number of reviews per restaurant (447), the number of positive reviews (4777) and the number of negative reviews (6526).

Nearby restaurants section introduction B:
In this section, an analysis is conducted regarding nearby restaurants. The analysis includes average price and rating, total number of reviews, and more. Graphs showing the distribution of language, gender, and emotion are also provided, along with word clouds for positive and negative reviews. The examination of the restaurants within a 250 meters range of ABC-XYZ shows that the predominant gender is Female, the most spoken language is Portuguese, and the dominant emotion is Neutral, as seen in the graphs below. The average price is 25.5 Euros and the average rating is 3.9 out of 5. With a total of 21908 reviews considered and an average of 447 reviews per restaurant, this provides a detailed view of the area's restaurant scene. The WordCloud graphs based on 4777 positive and 6526 negative reviews can also be seen here.

Mark only one oval.

- ☐ Paragraph A
- ☐ Paragraph B

A.9 ASTA Versus Chat GPT Survey (Portuguese)

Formulário de Comparação entre Textos Gerados por LLMs e Textos Criados com Combinação de Strings para o Projeto RRSO (Restaurant Review Sentiment Output)

Bem-vindo ao formulário de comparação entre textos gerados por LLMs e textos criados com combinação de strings para tese de Mestrado e projeto RRSO (Restaurant Review Sentiment Output). A sua participação é valiosa e ajudará a perceber a melhor forma de gerar frases para um relatório gerado automaticamente. O objetivo é auxiliar os proprietários de estabelecimentos, incluindo restaurantes, na interpretação dos seus dados de avaliação de forma concisa e objetiva usando Processamento de Linguagem Natural (NLP). Completar este formulário deve demorar aproximadamente 10 minutos.

Neste formulário, serão apresentadas várias versões de parágrafos exemplo e será solicitado que escolha o seu preferido.

As informações recolhidas serão usadas exclusivamente no contexto do projeto e, para garantir a confidencialidade, o seu nome não será associado aos dados recolhidos. **Ao responder a este formulário, está a concordar com os termos mencionados.** Lembre-se de que **pode optar por sair do formulário a qualquer momento.**

Agradecemos antecipadamente a sua cooperação.

* Indicates required question

Informação demográfica

Esta seção tem como objetivo inquirir o participante sobre dados demográficos relevantes para a pesquisa.

3. Por favor, indique a sua idade *

Mark only one oval.

- ☐ 15 ou menos
- ☐ 16 a 20
- ☐ 21 a 30
- ☐ 31 a 40
- ☐ 41 a 50
- ☐ 50 ou mais
- ☐ Prefiro não dizer

4. Por favor indique o seu nível de fluência em inglês. *
Se não conhece os níveis de fluência em inglês, clique [aqui](#)

Mark only one oval.

- ☐ Sem conhecimento de inglês
- ☐ Utilizador Básico (A1, A2)
- ☐ Utilizador Independente (B1, B2)
- ☐ Utilizador Proficiente (C1, C2)
- ☐ Utilizador Nativo

Feedback e hábitos de avaliação

Esta seção tem como objetivo reunir informações sobre os hábitos e comportamentos do participante ao fornecer feedback e avaliar restaurantes e outras compras ou serviços contratados.

1. Por favor, indique o seu país de nascimento. *

Mark only one oval.

- ☐ Inglaterra
- ☐ França
- ☐ Alemanha
- ☐ Itália
- ☐ Portugal
- ☐ Espanha
- ☐ Prefiro não dizer
- ☐ Other: _____

2. Por favor, indique o seu género *

Mark only one oval.

- ☐ Homem
- ☐ Mulher
- ☐ Outro
- ☐ Prefiro não dizer
- ☐ Other: _____

5. Já foi proprietário, operou ou ocupou um cargo de tomada de decisão num estabelecimento no ramo da alimentação, seja no passado ou atualmente? *

Mark only one oval.

- ☐ Sim
- ☐ Não

6. Com que frequência usa plataformas para encontrar restaurantes? *
Por exemplo, TripAdvisor, Zomato, Foursquare, The Fork, Yelp ou equivalentes

Mark only one oval.

- ☐ Nunca
- ☐ Raramente
- ☐ Às vezes
- ☐ Frequentemente
- ☐ Sempre

7. Com que frequência dá a sua opinião sob a forma de avaliação numérica ou equivalente (por exemplo, 1 - 5 ou ★★★★★) sobre as suas compras ou serviços contratados? *

Mark only one oval.

- ☐ Nunca
- ☐ Raramente
- ☐ Às vezes
- ☐ Frequentemente
- ☐ Sempre

8. Com que frequência dá a sua opinião sob a **forma textual** (por exemplo, comentário, avaliação) sobre suas compras ou serviços contratados? *

Mark only one oval.

- ☐ Nunca
- ☐ Raramente
- ☐ Às vezes
- ☐ Frequentemente
- ☐ Sempre

9. Com que frequência usa LLMs (Large Language Models) como ChatGPT, GPT-3, GPT-4, BARD, LLaMA ou outros? *

Mark only one oval.

- ☐ Nunca
- ☐ Raramente
- ☐ Às vezes
- ☐ Frequentemente
- ☐ Sempre

10. Quando avalia um produto ou serviço, qual das seguintes opções melhor descreve a sua abordagem? *

Mark only one oval.

- ☐ Tendo a avaliá-lo abaixo do que acho que merece Skip to question 11
- ☐ Avalio-o exatamente como acho que merece Skip to question 13
- ☐ Tendo a avaliá-lo acima do que acho que merece Skip to question 12

Eu tendo a avaliar abaixo do que minha avaliação sugere

13. Por favor, selecione o parágrafo que acha **melhor no geral**: *

Sumário do relatório (A):

Este relatório fornece uma análise aprofundada do ABC-XYZ, um restaurante localizado em Portugal. Os dados foram recolhidos a partir de avaliações entre 25 de maio de 2020 e 6 de agosto de 2022, um período de 36 meses. 182 avaliações foram selecionadas para análise, e os clientes deram ao restaurante uma classificação média de 4,29 em 5. A emoção predominante expressa nas avaliações foi Neutra, e o gênero predominante entre os clientes foi o feminino, com uma proporção de 1 homem para cada 2 mulheres. O idioma predominante usado nas avaliações foi o português. As avaliações positivas foram 1,74 vezes mais comuns do que as negativas, com os tópicos mais comuns sendo Bolo de chocolate e Pequeno almoço. Foram feitas comparações com outros restaurantes num raio de 250 metros, que tinham um gênero predominante Feminino, emoção predominante Positiva, idioma predominante Português, preço médio de 25,5, classificação média de 3,9 e número médio de avaliações de 447.

Sumário do relatório (B):

Uma avaliação das últimas 182 avaliações do ABC-XYZ de 2020-05-25 a 2022-08-06, abrangendo os últimos 36 meses, indica que o restaurante teve um desempenho razoável, com uma classificação média de 4,29/5. A maioria dos clientes é do sexo feminino, com uma proporção de 0,48 homens para cada mulher, e o idioma principal usado nas avaliações é o português. As emoções dos clientes tendem para o lado otimista, com cerca de 1 cliente masculino para cada 2 clientes femininas. Os tópicos mais mencionados em avaliações positivas são Bolo de chocolate e Pequeno almoço, enquanto que em avaliações negativas são Espaço e Tempo de espera por mesa. Uma análise dos restaurantes num raio de 250 metros revela que o gênero predominante é o Feminino, o idioma mais usado é o Português e a emoção dominante é Positiva. O preço médio é de 25,5 Euros e a classificação média é de 3,9/5, baseado em um total de 21908 avaliações e uma média de 447 avaliações por restaurante.

Mark only one oval.

- ☐ Parágrafo A
- ☐ Parágrafo B

11. Por favor, mencione os fatores que podem fazer com que avalie um produto/serviço com uma **nota inferior à que acha que o mesmo realmente merece**. *

Check all that apply.

- ☐ Quero envergonhá-los publicamente
- ☐ Quero parecer forte
- ☐ Quero desencorajar o fornecedor do produto/serviço
- ☐ Desejo uma explicação do fornecedor sobre quaisquer aspectos negativos do seu produto/serviço.
- ☐ Other: _____

Skip to question 13

Eu tendo a avaliar acima do que minha avaliação sugere

12. Por favor, mencione os fatores que podem fazer com que avalie um produto/serviço com uma **nota superior à que acha que o mesmo realmente merece**. *

Check all that apply.

- ☐ Não quero envergonhá-los publicamente
- ☐ Quero parecer gentil
- ☐ Quero encorajar o fornecedor do produto/serviço
- ☐ Não quero ser contactado pelo fornecedor do produto/serviço
- ☐ Other: _____

Skip to question 13

Escolha o Parágrafo Mais Eficaz para um Proprietário de Restaurante

Este formulário tem como objetivo determinar qual parágrafo comunica de forma mais eficaz informações a um proprietário de restaurante. Os parágrafos fazem parte de um relatório que informará o proprietário de restaurante sobre vários assuntos relacionados ao seu estabelecimento.

Para cada pergunta, escolha o parágrafo (A ou B) que considera melhor no geral.

14. Por favor, selecione o parágrafo que acha **melhor no geral**: *

Introdução à seção de gênero A:

Este relatório fornece uma análise de gênero dos clientes do ABC-XYZ. Inclui gráficos de barras mostrando o número de avaliações por gênero, gráficos circulares dividindo as avaliações em partes positivas e negativas por gênero e gráficos de radar comparando as emoções dos clientes no restaurante com as dos restaurantes em volta, usando 2 classes, Positivo e Negativo e 3 classes, Positivo, Negativo e Neutro. O gênero mais comum entre os clientes do restaurante é feminino, e a proporção aproximada de homens para mulheres no restaurante é de 1 homem para cada 2 mulheres.

Introdução à seção de gênero B:

Esta seção fornece um estudo da distribuição de gênero entre os clientes do ABC-XYZ e como esta se compara à dos restaurantes ideais melhores e piores. Esta secção contém gráficos de barras que indicam o número total de avaliações para cada gênero e gráficos de radar divididos em 2 (Positivo e Negativo) e 3 (Positivo, Neutro e Negativo) grupos que exibem a proporção de homens para mulheres em cada seção. Um gráfico de radar que ilustra como o ABC-XYZ se compara aos restaurantes ideais melhores e piores também está disponível nesta seção. A análise de gênero revela que o restaurante é predominantemente visitado por clientes do gênero feminino, com aproximadamente 1 homem para cada 2 mulheres.

Mark only one oval.

- ☐ Parágrafo A
- ☐ Parágrafo B

15. Por favor, selecione o parágrafo que **acha melhor no geral**. *

Introdução à seção de restaurantes próximos A:

Este relatório fornecerá ao proprietário do restaurante ABC-XYZ uma análise aprofundada dos restaurantes próximos num raio de 250 metros. A seção Restaurantes Próximos deste relatório inclui gráficos de barras mostrando a distribuição de gênero, idioma e emoção dos clientes, bem como nuvens de palavras das avaliações positivas e negativas desses restaurantes. Além disso, esta seção contém informações como o preço médio dos restaurantes próximos (25,5), a sua classificação média (3,9), o número total de avaliações (21908), o número médio de avaliações por restaurante (447), o número de avaliações positivas (4777) e o número de avaliações negativas (6526).

Introdução à seção de restaurantes próximos B:

Nesta seção, é realizada uma análise sobre os restaurantes próximos. A análise inclui preço e classificação médios, número total de avaliações e muito mais. Gráficos que mostram a distribuição de idioma, gênero e emoção também são fornecidos, juntamente com nuvens de palavras para avaliações positivas e negativas. A análise dos restaurantes num raio de 250 metros do ABC-XYZ mostra que o gênero predominante é o Feminino, o idioma mais falado é o Português e a emoção dominante é Neutra, como visto nos gráficos abaixo. O preço médio é de 25,5 Euros e a classificação média é de 3,9 em 5. Com um total de 21908 avaliações consideradas e uma média de 447 avaliações por restaurante, isso fornece uma visão detalhada dos restaurante da área. As nuvens de palavras são baseadas em 4777 avaliações positivas e 6526 avaliações negativas e também podem ser vistos aqui.

Mark only one oval.

- ☐ Parágrafo A
- ☐ Parágrafo B

This content is neither created nor endorsed by Google.

Google Forms

Pdf 3: ASTA Versus Chat GPT Survey (Portuguese)

A.10 Topic Labeling Survey

Topic labeling form for the RRSO (Restaurant Review Sentiment Output) project

Welcome to the topic labeling form for M.Sc. thesis and for the RRSO (Restaurant Review Sentiment Output) project. Your participation is valuable and will help improve the topic identification method to assign the correct name to each topic. This will provide restaurant owners with a better process for identifying topics in their reviews using Natural Language Processing (NLP) models. Note that this form may contain misspelled words due to the nature of its data collection method based on responses written by users. Completing this form should take approximately 15 to 25 minutes.

The information collected will be used exclusively in the context of the project and to ensure confidentiality, your name will not be associated with the data collected. **By responding to this form, you are agreeing to the aforementioned terms.** Please remember that you can **choose to exit the form at any time.**

We thank you in advance for your collaboration.

* Indicates required question

Demographic information

This section aims to inquire the participant about relevant demographic data for the survey.

1. Please indicate your country of birth. *

Mark only one oval.

- ☐ England
- ☐ France
- ☐ Germany
- ☐ Italy
- ☐ Portugal
- ☐ Spain
- ☐ I'd rather not say
- ☐ Other: _____

Feedback and Evaluation Habits

This section aims to gather information about the participant's habits and behavior when providing feedback and evaluating restaurants and other purchases or contracted services.

5. Have you ever **owned, operated or held a decision-making position in a food establishment**, either in the **past or currently**? *

Mark only one oval.

- ☐ Yes
- ☐ No

6. **How often do you use platforms to find restaurants?** e.g. TripAdvisor, Zomato, Foursquare, The Fork, Yelp or equivalents. *

Mark only one oval.

- ☐ Never
- ☐ Rarely
- ☐ Sometimes
- ☐ Often
- ☐ Always

2. What is your gender? *

Mark only one oval.

- ☐ Man
- ☐ Woman
- ☐ Other
- ☐ I'd rather not say
- ☐ Other: _____

3. How old are you? *

Mark only one oval.

- ☐ 15 or less
- ☐ 16 to 20
- ☐ 21 to 30
- ☐ 31 to 40
- ☐ 41 to 50
- ☐ 50 or more
- ☐ I'd rather not say

4. What is your level of fluency in English? *

If you don't know about English proficiency levels, please click [here](#)

Mark only one oval.

- ☐ No English knowledge
- ☐ Basic User (A1, A2)
- ☐ Independent User (B1, B2)
- ☐ Proficient User (C1, C2)
- ☐ Native

Skip to question 5

7. **How often** do you give your opinion in the form of a **numerical evaluation or equivalent** (e.g. 1 - 5 or ★★★★★) about your purchases or contracted services? *

Mark only one oval.

- ☐ Never
- ☐ Rarely
- ☐ Sometimes
- ☐ Often
- ☐ Always

8. **How often** do you give your opinion in **textual form** (e.g. comment, review) about your purchases or contracted services? *

Mark only one oval.

- ☐ Never
- ☐ Rarely
- ☐ Sometimes
- ☐ Often
- ☐ Always

9. When you **rate a product or service**, which of the following best describes **your approach**? *

Mark only one oval.

- ☐ I tend to rate it lower than what I think it deserves *Skip to question 10*
- ☐ I rate it exactly as I think it deserves *Skip to question 12*
- ☐ I tend to rate it higher than what I think it deserves *Skip to question 11*

I tend to rate lower than what my assessment suggests

10. Please indicate any factors that may influence you to provide a **lower rating** for a product/service **than your actual assessment** of its quality. *

Check all that apply.

- ☐ I want to publicly shame them
- ☐ I want to come across as strong.
- ☐ I want to discourage the product/service supplier
- ☐ I desire an explanation from the supplier regarding any negative aspects of their product/service.
- ☐ Other: _____

Skip to question 12

I tend to rate higher than what my assessment suggests

11. Please indicate any factors that may influence you to provide a **higher rating** for a product/service **than your actual assessment** of its quality. *

Check all that apply.

- ☐ I don't want to publicly shame them
- ☐ I want to come across as kind.
- ☐ I want to encourage the product/service supplier
- ☐ I don't want to be contacted by the product/service supplier
- ☐ Other: _____

Skip to question 12

Random selection

12. Please select one of the following options at **random** to continue with the survey. *

Mark only one oval.

- ☐ Skip to question 13
- ☐ Skip to question 46

Topic Labeling (0 - 31)

This form is designed to help identify the expression or word that best represents a given topic. Each topic is defined by a set of **9 words**, arranged in **decreasing order of importance from top to bottom and left to right**. The **most important** words are indicated by a **green circle** (●), while **important** words are indicated by a **yellow circle** (●) and the **least important** words are indicated by a **red circle** (●).

For each question, please **select the topic name (A, B, C or D)** that you feel **best represents the given set of words**. If none of the predefined answers are satisfactory, you may provide your own suggestion using the "Other" option.

13. **Topic 0:** *
- Which expression or word better identifies the following succession of words. Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** piece, restaurant, buffet
- **Important:** eat, soup, sushi
- **Least important:** roll, try, menu

Mark only one oval.

- ☐ A (Piece)
- ☐ B (Sushi)
- ☐ C (Soup)
- ☐ D (Buffet)
- ☐ Other: _____

14. **Topic 1:** *
- Which expression or word better identifies the following succession of words. Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** ask, table, say
- **Important:** order, come, employee
- **Least important:** customer, time, want

Mark only one oval.

- ☐ A (Customer service)
- ☐ B (Ordering and serving)
- ☐ C (Employee interaction)
- ☐ D (Table and ordering)
- ☐ Other: _____

15. **Topic 2:** *
- Which expression or word better identifies the following succession of words. Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** quality, price, ratio
- **Important:** service, space, food
- **Least important:** quantity, lunch, recommend

Mark only one oval.

- ☐ A (Price-quality ratio)
- ☐ B (Food quantity and service)
- ☐ C (Lunch recommendations)
- ☐ D (Value)
- ☐ Other: _____

16. **Topic 3:** *
- Which expression or word better identifies the following succession of words. Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** time, smell, service
- **Important:** serve, restaurant, space
- **Least important:** stone, customer, employee

Mark only one oval.

- ☐ A (Service and management)
- ☐ B (Customer experience)
- ☐ C (Meat and space)
- ☐ D (Waiting time)
- ☐ Other: _____

17. **Topic 4:** *
- Which expression or word better identifies the following succession of words. Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** service, lack, food
- **Important:** employee, space, customer
- **Least important:** price, point, table

Mark only one oval.

- ☐ A (Service quality)
- ☐ B (Employee friendliness)
- ☐ C (Food and space)
- ☐ D (Service and food)
- ☐ Other: _____

8/28/23, 3:58 PM

Topic labeling form for the RRSO (Restaurant Review Sentiment Output) project

18. **Topic 5:**

Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** love, restaurant, open
- **Important:** door, close, try
- **Least important:** make, time, say

Mark only one oval.

- ☐ A (Restaurant experience)
- ☐ B (Mall food)
- ☐ C (Lisbon cuisine)
- ☐ D (Good experience)
- ☐ Other: _____

19. **Topic 6:**

Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** cream, ice, flavor, fruit
- **Important:** try, taste, chocolate
- **Least important:** crepe, space

Mark only one oval.

- ☐ A (Ice cream flavors)
- ☐ B (Fruit and chocolate crepes)
- ☐ C (Ice cream service)
- ☐ D (Ice cream)
- ☐ Other: _____

<https://docs.google.com/forms/d/1zgtZRIKFpZSdUghltztyMfRaW9FsTPbkyJY6MLsYZyA/edit>

9/39

8/28/23, 3:58 PM

Topic labeling form for the RRSO (Restaurant Review Sentiment Output) project

20. **Topic 7:**

Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** option, menu, dish
- **Important:** lunch, day, space
- **Least important:** food, service, vegan

Mark only one oval.

- ☐ A (Vegan options)
- ☐ B (Lunch menu)
- ☐ C (Vegan restaurant)
- ☐ D (Menu options)
- ☐ Other: _____

21. **Topic 8:**

Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** pizza, ingredient, service
- **Important:** quality, space, price
- **Least important:** dough, crust, pasta

Mark only one oval.

- ☐ A (Pizza quality)
- ☐ B (Pizza ingredients)
- ☐ C (Pasta)
- ☐ D (Pizza)
- ☐ Other: _____

<https://docs.google.com/forms/d/1zgtZRIKFpZSdUghltztyMfRaW9FsTPbkyJY6MLsYZyA/edit>

10/39

8/28/23, 3:58 PM

Topic labeling form for the RRSO (Restaurant Review Sentiment Output) project

22. **Topic 9:**

Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** cake, bread, buy
- **Important:** slice, make, birthday
- **Least important:** house, cod, place

Mark only one oval.

- ☐ A (Cake and bread)
- ☐ B (Birthday)
- ☐ C (Cod)
- ☐ D (Cake)
- ☐ Other: _____

23. **Topic 10:**

Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** juice, order, toast
- **Important:** croissant, egg, try
- **Least important:** space, pancake, fruit

Mark only one oval.

- ☐ A (Juice and toasts)
- ☐ B (Croissant and egg)
- ☐ C (Pancake breakfast)
- ☐ D (Breakfast)
- ☐ Other: _____

<https://docs.google.com/forms/d/1zgtZRIKFpZSdUghltztyMfRaW9FsTPbkyJY6MLsYZyA/edit>

11/39

8/28/23, 3:58 PM

Topic labeling form for the RRSO (Restaurant Review Sentiment Output) project

24. **Topic 11:**

Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** pizza, pasta, risotto
- **Important:** order, restaurant, dessert
- **Least important:** try, come, starter

Mark only one oval.

- ☐ A (Pizza and pasta)
- ☐ B (Risotto and dessert)
- ☐ C (Starter and dessert)
- ☐ D (Pizza and rissoto)
- ☐ Other: _____

25. **Topic 12:**

Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** dinner, group, food
- **Important:** friend, restaurant, service
- **Least important:** birthday, drink, recommend

Mark only one oval.

- ☐ A (Group dinner)
- ☐ B (Birthday celebration)
- ☐ C (Dinner with friends)
- ☐ D (Friends)
- ☐ Other: _____

<https://docs.google.com/forms/d/1zgtZRIKFpZSdUghltztyMfRaW9FsTPbkyJY6MLsYZyA/edit>

12/39

8/28/23, 3:58 PM

Topic labeling form for the RRSO (Restaurant Review Sentiment Output) project

26. **Topic 13:** *
- Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:
- **Most important:** book, think, try
 - **Important:** space, sushi, price
 - **Least important:** quality, eclair, service

Mark only one oval.

- ☐ A (Sushi quality)
- ☐ B (Éclair)
- ☐ C (Sushi booking)
- ☐ D (Booking)
- ☐ Other: _____

27. **Topic 14:** *
- Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:
- **Most important:** salad, dish, space
 - **Important:** food, lunch, rice
 - **Least important:** service, meat, choose

Mark only one oval.

- ☐ A (Salad and rice)
- ☐ B (Lunch meat dishes)
- ☐ C (Dish)
- ☐ D (Salad)
- ☐ Other: _____

<https://docs.google.com/forms/d/1zgtZRIKFpZSdUghltztymlMRaW9FsTPbkyJY6MLsYZyA/edit>

13/39

8/28/23, 3:58 PM

Topic labeling form for the RRSO (Restaurant Review Sentiment Output) project

28. **Topic 15:** *
- Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:
- **Most important:** music, atmosphere, space
 - **Important:** food, hour, service
 - **Least important:** like, employee, time

Mark only one oval.

- ☐ A (Music atmosphere)
- ☐ B (Service waiting time)
- ☐ C (Background music)
- ☐ D (Music)
- ☐ Other: _____

29. **Topic 16:** *
- Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:
- **Most important:** cake, chocolate, space
 - **Important:** service, decorate, staff
 - **Least important:** food, dessert, mousse

Mark only one oval.

- ☐ A (Chocolate cake)
- ☐ B (Dessert)
- ☐ C (Cake decoration)
- ☐ D (Cake)
- ☐ Other: _____

<https://docs.google.com/forms/d/1zgtZRIKFpZSdUghltztymlMRaW9FsTPbkyJY6MLsYZyA/edit>

14/39

8/28/23, 3:58 PM

Topic labeling form for the RRSO (Restaurant Review Sentiment Output) project

30. **Topic 17:** *
- Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:
- **Most important:** order, time, pizza
 - **Important:** deliver, arrive, ask
 - **Least important:** service, make, mistake

Mark only one oval.

- ☐ A (Pizza delivery)
- ☐ B (Mistake)
- ☐ C (Order time)
- ☐ D (Buffet)
- ☐ Other: _____

31. **Topic 18:** *
- Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:
- **Most important:** brunch, drink, coffee
 - **Important:** place, menu, egg
 - **Least important:** include, juice, breakfast

Mark only one oval.

- ☐ A (Brunch drinks)
- ☐ B (Coffee and eggs)
- ☐ C (Brunch menu)
- ☐ D (Brunch)
- ☐ Other: _____

<https://docs.google.com/forms/d/1zgtZRIKFpZSdUghltztymlMRaW9FsTPbkyJY6MLsYZyA/edit>

15/39

8/28/23, 3:58 PM

Topic labeling form for the RRSO (Restaurant Review Sentiment Output) project

32. **Topic 19:** *
- Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:
- **Most important:** view, drink, space
 - **Important:** place, night, terrace
 - **Least important:** afternoon, summer, service

Mark only one oval.

- ☐ A (View and drinks)
- ☐ B (Night terrace)
- ☐ C (Summer afternoon)
- ☐ D (View)
- ☐ Other: _____

33. **Topic 20:** *
- Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:
- **Most important:** coffee, cafe, snack
 - **Important:** terrace, service, day
 - **Least important:** pig, space, lunch

Mark only one oval.

- ☐ A (Coffee)
- ☐ B (Coffee snacks)
- ☐ C (Coffee terrace)
- ☐ D (Coffee snack)
- ☐ Other: _____

<https://docs.google.com/forms/d/1zgtZRIKFpZSdUghltztymlMRaW9FsTPbkyJY6MLsYZyA/edit>

16/39

8/28/23, 3:58 PM

Topic labeling form for the RRSO (Restaurant Review Sentiment Output) project

34. **Topic 21:**

Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** fish, restaurant, grill
- **Important:** quality, price, service
- **Least important:** meat, eat, food

Mark only one oval.

- ☐ A (Grilled fish)
- ☐ B (Fish quality)
- ☐ C (Sushi and seafood)
- ☐ D (Fish)
- ☐ Other: _____

35. **Topic 22:**

Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** place, food, porto
- **Important:** eat, service, people
- **Least important:** staff, city, price

Mark only one oval.

- ☐ A (Porto food)
- ☐ B (Porto people)
- ☐ C (Porto)
- ☐ D (Porto city)
- ☐ Other: _____

<https://docs.google.com/forms/d/1zgtZRIKfPzSdUghltzYmMRaW9FsTPbkyJY6MLsYZyA/edit>

17/39

8/28/23, 3:58 PM

Topic labeling form for the RRSO (Restaurant Review Sentiment Output) project

36. **Topic 23:**

Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** eat, meat, time
- **Important:** restaurant, make, point
- **Least important:** service, cook, food

Mark only one oval.

- ☐ A (Meat quality)
- ☐ B (Meat restaurant)
- ☐ C (Cooked meat)
- ☐ D (Meat)
- ☐ Other: _____

37. **Topic 24:**

Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** food, restaurant, service
- **Important:** make, time, space
- **Least important:** price, quality, money

Mark only one oval.

- ☐ A (Food value)
- ☐ B (Restaurant service)
- ☐ C (Food quality)
- ☐ D (Food)
- ☐ Other: _____

<https://docs.google.com/forms/d/1zgtZRIKfPzSdUghltzYmMRaW9FsTPbkyJY6MLsYZyA/edit>

18/39

8/28/23, 3:58 PM

Topic labeling form for the RRSO (Restaurant Review Sentiment Output) project

38. **Topic 25:**

Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** food, service, space
- **Important:** think, improve, come
- **Least important:** restaurant, say, point

Mark only one oval.

- ☐ A (Improvement needed in food)
- ☐ B (Improvement needed)
- ☐ C (Service)
- ☐ D (Food)
- ☐ Other: _____

39. **Topic 26:**

Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** space, menu, lunch
- **Important:** food, child, hide
- **Least important:** restaurant, staff, service

Mark only one oval.

- ☐ A (Kid-friendly menu)
- ☐ B (Lunch space)
- ☐ C (Kid-friendly restaurant)
- ☐ D (Space)
- ☐ Other: _____

<https://docs.google.com/forms/d/1zgtZRIKfPzSdUghltzYmMRaW9FsTPbkyJY6MLsYZyA/edit>

19/39

8/28/23, 3:58 PM

Topic labeling form for the RRSO (Restaurant Review Sentiment Output) project

40. **Topic 27:**

Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** wine, cheese, dessert
- **Important:** meat, choose, flavor
- **Least important:** accompany, space, starter

Mark only one oval.

- ☐ A (Wine and cheese)
- ☐ B (Dessert flavors)
- ☐ C (Wine accompaniment)
- ☐ D (Wine)
- ☐ Other: _____

41. **Topic 28:**

Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** center, shopping, day
- **Important:** game, watch, eat
- **Least important:** price, restaurant, service

Mark only one oval.

- ☐ A (Shopping center)
- ☐ B (Sports game watching)
- ☐ C (Game)
- ☐ D (Service)
- ☐ Other: _____

<https://docs.google.com/forms/d/1zgtZRIKfPzSdUghltzYmMRaW9FsTPbkyJY6MLsYZyA/edit>

20/39

8/28/23, 3:58 PM

Topic labeling form for the RRSO (Restaurant Review Sentiment Output) project

42. Topic 29:

Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** space, variety, price
- **Important:** service, food, lot
- **Least important:** quality, view, location

Mark only one oval.

- ☐ A (Space variety)
- ☐ B (Food variety)
- ☐ C (Space)
- ☐ D (Price variety)
- ☐ Other: _____

43. Topic 30:

Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** cod, dish, restaurant
- **Important:** beef, food, pork
- **Least important:** meat, roast, octopus

Mark only one oval.

- ☐ A (Cod and beef)
- ☐ B (Roasted meats)
- ☐ C (Pork and octopus)
- ☐ D (Portuguese food)
- ☐ Other: _____

<https://docs.google.com/forms/d/1zgtZRIKFpZSdUghltztlmMRaW9FsTPbkYJY6MLsYZyA/edit>

21/39

8/28/23, 3:58 PM

Topic labeling form for the RRSO (Restaurant Review Sentiment Output) project

46. Topic 32:

Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** tourist, space, area
- **Important:** price, smoke, wine
- **Least important:** service, gin, bar

Mark only one oval.

- ☐ A (Bar scene)
- ☐ B (Tourist area)
- ☐ C (Drinks and smoking)
- ☐ D (Tourists)
- ☐ Other: _____

47. Topic 33:

Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** pastry, shop, cake
- **Important:** service, bread, croissant
- **Least important:** puff, tart, bakery

Mark only one oval.

- ☐ A (Bakery and pastry)
- ☐ B (Cake shops)
- ☐ C (Bread and pastry)
- ☐ D (Pastry)
- ☐ Other: _____

<https://docs.google.com/forms/d/1zgtZRIKFpZSdUghltztlmMRaW9FsTPbkYJY6MLsYZyA/edit>

23/39

8/28/23, 3:58 PM

Topic labeling form for the RRSO (Restaurant Review Sentiment Output) project

44. Topic 31:

Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** rice, prego, fry
- **Important:** picanha, time, restaurant
- **Least important:** eat, cuttlefish, service

Mark only one oval.

- ☐ A (Rice and beans)
- ☐ B (Fried picanha)
- ☐ C (Cuttlefish and prego)
- ☐ D (Prego)
- ☐ Other: _____

45. In this form you can only do a **maximum of 2 topic labeling sections** *

Do you wish to **continue labeling topics**?

Mark only one oval.

- ☐ Yes, I want to continue labeling *Skip to question 46*
- ☐ No, I want to stop

Topic Labeling (32 - 63)

This form is designed to help identify the expression or word that best represents a given topic. Each topic is defined by a set of **9 words**, arranged in **decreasing order of importance from top to bottom and left to right**. The **most important** words are indicated by a **green circle** (●), while **important** words are indicated by a **yellow circle** (●) and the **least important** words are indicated by a **red circle** (●).

For each question, please **select the topic name (A, B, C or D)** that you feel **best represents the given set of words**. If none of the predefined answers are satisfactory, you may provide your own suggestion using the **"Other"** option.

<https://docs.google.com/forms/d/1zgtZRIKFpZSdUghltztlmMRaW9FsTPbkYJY6MLsYZyA/edit>

22/39

8/28/23, 3:58 PM

Topic labeling form for the RRSO (Restaurant Review Sentiment Output) project

48. Topic 34:

Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** sushi, quality, restaurant
- **Important:** service, price, space
- **Least important:** staff, repeat, friendliness

Mark only one oval.

- ☐ A (Sushi)
- ☐ B (Quality sushi)
- ☐ C (Friendly sushi places)
- ☐ D (Good sushi)
- ☐ Other: _____

49. Topic 35:

Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** wait, table, time
- **Important:** restaurant, arrive, food
- **Least important:** people, service, reservation

Mark only one oval.

- ☐ A (Restaurant wait times)
- ☐ B (Table reservations)
- ☐ C (Food waiting time)
- ☐ D (Waiting for table)
- ☐ Other: _____

<https://docs.google.com/forms/d/1zgtZRIKFpZSdUghltztlmMRaW9FsTPbkYJY6MLsYZyA/edit>

24/39

8/28/23, 3:58 PM

Topic labeling form for the RRSO (Restaurant Review Sentiment Output) project

50. **Topic 36:** ★
Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:
- **Most important:** dish, dessert, order
 - **Important:** starter, chicken, restaurant
 - **Least important:** meat, service, potato

Mark only one oval.

- ☐ A (Restaurant dishes)
- ☐ B (Desserts and starters)
- ☐ C (Meat and chicken)
- ☐ D (Dish)
- ☐ Other: _____

51. **Topic 37:** ★
Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:
- **Most important:** beer, snack, place
 - **Important:** service, drink, snail
 - **Least important:** friend, space, price

Mark only one oval.

- ☐ A (Beer and snacks)
- ☐ B (Drinks with friends)
- ☐ C (Snail bar)
- ☐ D (Beer)
- ☐ Other: _____

<https://docs.google.com/forms/d/1zgtZRIKfPzSdUghltzYmMRaW9FsTPbkyJY6MLsYZyA/edit>

25/39

8/28/23, 3:58 PM

Topic labeling form for the RRSO (Restaurant Review Sentiment Output) project

52. **Topic 38:** ★
Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:
- **Most important:** gold, zomato, service
 - **Important:** space, repeat, use
 - **Least important:** experience, food, cereal

Mark only one oval.

- ☐ A (Zomato gold)
- ☐ B (Zomato service)
- ☐ C (Cereal restaurant)
- ☐ D (Zomato discount)
- ☐ Other: _____

53. **Topic 39:** ★
Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:
- **Most important:** sandwich, ham, cheese
 - **Important:** time, think, eat
 - **Least important:** salad, order, bread

Mark only one oval.

- ☐ A (Sandwich shop)
- ☐ B (Ham and cheese sandwiches)
- ☐ C (Salad sandwiches)
- ☐ D (Sandwich)
- ☐ Other: _____

<https://docs.google.com/forms/d/1zgtZRIKfPzSdUghltzYmMRaW9FsTPbkyJY6MLsYZyA/edit>

25/39

8/28/23, 3:58 PM

Topic labeling form for the RRSO (Restaurant Review Sentiment Output) project

54. **Topic 40:** ★
Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:
- **Most important:** table, space, room
 - **Important:** wall, chair, bar
 - **Least important:** people, floor, time

Mark only one oval.

- ☐ A (Room table and chair)
- ☐ B (Furniture and decor)
- ☐ C (Bar space)
- ☐ D (Table)
- ☐ Other: _____

55. **Topic 41:** ★
Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:
- **Most important:** hamburger, burger, potato
 - **Important:** fry, sauce, bread
 - **Least important:** meat, service, space

Mark only one oval.

- ☐ A (Hamburgers and fries)
- ☐ B (Hamburguer)
- ☐ C (Potato dishes)
- ☐ D (Fries)
- ☐ Other: _____

<https://docs.google.com/forms/d/1zgtZRIKfPzSdUghltzYmMRaW9FsTPbkyJY6MLsYZyA/edit>

27/39

8/28/23, 3:58 PM

Topic labeling form for the RRSO (Restaurant Review Sentiment Output) project

56. **Topic 42:** ★
Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:
- **Most important:** space, pancake, dog
 - **Important:** service, staff, love
 - **Least important:** recommend, order, try

Mark only one oval.

- ☐ A (Dog-friendly)
- ☐ B (Pancake)
- ☐ C (Friendly staff)
- ☐ D (Nice staff)
- ☐ Other: _____

57. **Topic 43:** ★
Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:
- **Most important:** sauce, water, order
 - **Important:** ask, steak, restaurant
 - **Least important:** service, come, tuna

Mark only one oval.

- ☐ A (Sauce)
- ☐ B (Steak dish)
- ☐ C (Tuna dish)
- ☐ D (Water)
- ☐ Other: _____

<https://docs.google.com/forms/d/1zgtZRIKfPzSdUghltzYmMRaW9FsTPbkyJY6MLsYZyA/edit>

28/39

8/28/23, 3:58 PM

Topic labeling form for the RRSO (Restaurant Review Sentiment Output) project

58. **Topic 44:** *
- Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:
- **Most important:** wine, list, place
 - **Important:** hustle, service, need
 - **Least important:** bustle, work, visit

Mark only one oval.

- ☐ A (Wine list)
- ☐ B (Wine place)
- ☐ C (Hustle and bustle)
- ☐ D (Wine)
- ☐ Other: _____

59. **Topic 45:** *
- Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:
- **Most important:** chicken, taco, food
 - **Important:** sauce, order, restaurant
 - **Least important:** service, guacamole, barbecue

Mark only one oval.

- ☐ A (Tacos)
- ☐ B (Chicken tacos)
- ☐ C (Sauces and guacamole)
- ☐ D (Mexican)
- ☐ Other: _____

<https://docs.google.com/forms/d/1zgtZRIKfPzSdUghltzYmMRaW9FsTPbkyJY6MLsYZyA/edit>

29/39

8/28/23, 3:58 PM

Topic labeling form for the RRSO (Restaurant Review Sentiment Output) project

60. **Topic 46:** *
- Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:
- **Most important:** francesinha, porto, sauce
 - **Important:** space, eat, try
 - **Least important:** service, place, taste

Mark only one oval.

- ☐ A (Francesinha in porto)
- ☐ B (Sauce tasting)
- ☐ C (Eating space)
- ☐ D (Porto city)
- ☐ Other: _____

61. **Topic 47:** *
- Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:
- **Most important:** order, sauce, come
 - **Important:** like, salmon, bread
 - **Least important:** make, cheese, shrimp

Mark only one oval.

- ☐ A (Salmon and shrimp order)
- ☐ B (Cheese sauces)
- ☐ C (Bread dish)
- ☐ D (Ordering)
- ☐ Other: _____

<https://docs.google.com/forms/d/1zgtZRIKfPzSdUghltzYmMRaW9FsTPbkyJY6MLsYZyA/edit>

30/39

8/28/23, 3:58 PM

Topic labeling form for the RRSO (Restaurant Review Sentiment Output) project

62. **Topic 48:** *
- Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:
- **Most important:** pay, price, restaurant
 - **Important:** time, atm, people
 - **Least important:** dish, drink, euro

Mark only one oval.

- ☐ A (Paying)
- ☐ B (Atm and prices)
- ☐ C (Price)
- ☐ D (Payment and price)
- ☐ Other: _____

63. **Topic 49:** *
- Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:
- **Most important:** seafood, sea, service
 - **Important:** beach, restaurant, view
 - **Least important:** recommend, price, lunch

Mark only one oval.

- ☐ A (Seafood restaurant)
- ☐ B (Beach view)
- ☐ C (Sea view)
- ☐ D (Seafood)
- ☐ Other: _____

<https://docs.google.com/forms/d/1zgtZRIKfPzSdUghltzYmMRaW9FsTPbkyJY6MLsYZyA/edit>

31/39

8/28/23, 3:58 PM

Topic labeling form for the RRSO (Restaurant Review Sentiment Output) project

64. **Topic 50:** *
- Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:
- **Most important:** try, space, time
 - **Important:** like, want, decide
 - **Least important:** tea, come, opportunity

Mark only one oval.

- ☐ A (Trying out)
- ☐ B (Tea)
- ☐ C (Opportunities to eat)
- ☐ D (Space)
- ☐ Other: _____

65. **Topic 51:** *
- Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:
- **Most important:** steak, house, sauce
 - **Important:** eat, time, know
 - **Least important:** restaurant, year, place

Mark only one oval.

- ☐ A (Steak house)
- ☐ B (Sauce)
- ☐ C (House)
- ☐ D (Steak)
- ☐ Other: _____

<https://docs.google.com/forms/d/1zgtZRIKfPzSdUghltzYmMRaW9FsTPbkyJY6MLsYZyA/edit>

32/39

8/28/23, 3:58 PM

Topic labeling form for the RRSO (Restaurant Review Sentiment Output) project

66. **Topic 52:**

Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** clams, prawn, service
- **Important:** crab, restaurant, eat
- **Least important:** seafood, like, come

Mark only one oval.

- ☐ A (Clams and prawns)
- ☐ B (Crab)
- ☐ C (Seafood)
- ☐ D (Clam)
- ☐ Other: _____

67. **Topic 53:**

Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** place, park, terrace
- **Important:** enjoy, raman, space
- **Least important:** service, car, crepe

Mark only one oval.

- ☐ A (Terrace)
- ☐ B (Raman crepe)
- ☐ C (Parking)
- ☐ D (Location)
- ☐ Other: _____

<https://docs.google.com/forms/d/1zgtZRIKfPzSdUghltzYmMRaW9FsTPbkyJY6MLsYZyA/edit>

33/39

8/28/23, 3:58 PM

Topic labeling form for the RRSO (Restaurant Review Sentiment Output) project

68. **Topic 54:**

Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** food, recommend, service
- **Important:** atmosphere, staff, visit
- **Least important:** restaurant, welcome, space

Mark only one oval.

- ☐ A (Recommended food)
- ☐ B (Good service)
- ☐ C (Friendly staff)
- ☐ D (Good food)
- ☐ Other: _____

69. **Topic 55:**

Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** space, food, service
- **Important:** order, establishment, staff
- **Least important:** concept, make, experience

Mark only one oval.

- ☐ A (Experience)
- ☐ B (Service)
- ☐ C (Order)
- ☐ D (Space)
- ☐ Other: _____

<https://docs.google.com/forms/d/1zgtZRIKfPzSdUghltzYmMRaW9FsTPbkyJY6MLsYZyA/edit>

34/39

8/28/23, 3:58 PM

Topic labeling form for the RRSO (Restaurant Review Sentiment Output) project

70. **Topic 56:**

Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** food, dish, restaurant
- **Important:** service, recommend, atmosphere
- **Least important:** family, staff, serve

Mark only one oval.

- ☐ A (Family restaurant)
- ☐ B (Dish recommendation)
- ☐ C (Friendly atmosphere)
- ☐ D (Dish)
- ☐ Other: _____

71. **Topic 57:**

Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** experience, space, chef
- **Important:** restaurant, food, make
- **Least important:** thank, service, dish

Mark only one oval.

- ☐ A (Chef)
- ☐ B (Space)
- ☐ C (Food)
- ☐ D (Experience)
- ☐ Other: _____

<https://docs.google.com/forms/d/1zgtZRIKfPzSdUghltzYmMRaW9FsTPbkyJY6MLsYZyA/edit>

35/39

8/28/23, 3:58 PM

Topic labeling form for the RRSO (Restaurant Review Sentiment Output) project

72. **Topic 58:**

Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** restaurant, pasta, area
- **Important:** parking, dish, service
- **Least important:** nacoee, parque, flavor

Mark only one oval.

- ☐ A (Pasta)
- ☐ B (Parking area)
- ☐ C (Nacoe flavors)
- ☐ D (Dish)
- ☐ Other: _____

73. **Topic 59:**

Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** leave, desire, lot
- **Important:** restaurant, food, space
- **Least important:** service, wine, make

Mark only one oval.

- ☐ A (Insufficient)
- ☐ B (Food)
- ☐ C (Wine)
- ☐ D (Bad experience)
- ☐ Other: _____

<https://docs.google.com/forms/d/1zgtZRIKfPzSdUghltzYmMRaW9FsTPbkyJY6MLsYZyA/edit>

36/39

8/28/23, 3:58 PM

Topic labeling form for the RRSO (Restaurant Review Sentiment Output) project

74. **Topic 60:**

Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** dish, room, restaurant
- **Important:** table, end, kitchen
- **Least important:** meal, menu, flavor

Mark only one oval.

- ☐ A (Dining rooms)
- ☐ B (Kitchen menus)
- ☐ C (Flavorful meals)
- ☐ D (Dish)
- ☐ Other: _____

75. **Topic 61:**

Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** home, order, delivery, eat
- **Important:** service, recommend
- **Least important:** restaurant, time, food

Mark only one oval.

- ☐ A (Home delivery)
- ☐ B (Eating at home)
- ☐ C (Recommendation)
- ☐ D (Ordering)
- ☐ Other: _____

<https://docs.google.com/forms/d/1zgtZRIKFpZSdUghltztyMlRaW9FsTPbkyJY6MLsYZyA/edit>

37/39

8/28/23, 3:58 PM

Topic labeling form for the RRSO (Restaurant Review Sentiment Output) project

76. **Topic 62:**

Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** serve, time, come
- **Important:** order, table, employee
- **Least important:** minute, service, wait

Mark only one oval.

- ☐ A (Table service)
- ☐ B (Serving time)
- ☐ C (Ordering time)
- ☐ D (Waiting time)
- ☐ Other: _____

77. **Topic 63:**

Which expression or word better identifies the following succession of words.
Keep in mind that the words are ordered in **decreasing level of importance**:

- **Most important:** food, meat, flavor
- **Important:** service, restaurant, experience
- **Least important:** curry, recommend, staff

Mark only one oval.

- ☐ A (Flavorful meats)
- ☐ B (Curry)
- ☐ C (Friendly staff)
- ☐ D (Food)
- ☐ Other: _____

<https://docs.google.com/forms/d/1zgtZRIKFpZSdUghltztyMlRaW9FsTPbkyJY6MLsYZyA/edit>

38/39

8/28/23, 3:58 PM

Topic labeling form for the RRSO (Restaurant Review Sentiment Output) project

78. In this form you can only do a **maximum of 2 topic labeling sections** *

Do you wish to **continue labeling topics**?

Mark only one oval.

- ☐ Yes, I want to continue labeling *Skip to question 13*
- ☐ No, I want to stop

This content is neither created nor endorsed by Google.

Google Forms

<https://docs.google.com/forms/d/1zgtZRIKFpZSdUghltztyMlRaW9FsTPbkyJY6MLsYZyA/edit>

39/39

```

1 #Example prompt after all variables are populated:
2 ExamplePrompt = "Write a summary with no more than 150 words about a
    restaurant report intended for a restaurant owner to better
    understand their business. This summary will precede the actual
    report and provide a concise overview of the key information
    contained in the report. Keep in mind that the report contains the
    information in the following Python dictionary: {'Number of past
    reviews that were selected for analysis': 174, 'Name of the
    restaurant being analyzed': 'ABC-XYZ', 'Earliest date for which
    data was collected': '2020-07-27', 'Most recent date for which
    data was collected': '2022-08-06', 'Number of months prior to the
    final date for which data was collected': 36, 'Average rating (out
    of 5) that customers gave to the restaurant': 4.28, 'Emotion that
    is most common among the restaurant's customers': 'Neutral', '
    Gender that is most common among the restaurant's customers': '
    female', 'Approximate ratio of men to woman customers at the
    restaurant': '1 man for every 2 woman', 'Language that is most
    commonly used in reviews of the restaurant': 'Portuguese', 'Ratio
    between positive and negative emotions expressed in reviews of the
    restaurant': 1.62, 'First most common topic found in positive
    reviews of the restaurant': 'Breakfast', 'Second most common topic
    found in positive reviews of the restaurant': 'Chocolate cake', '
    First most common topic found in negative reviews of the
    restaurant': 'Space', 'Second most common topic found in negative
    reviews of the restaurant': 'Waiting for table', 'Radius (in
    meters) around the restaurant that will be considered when
    comparing it to other restaurants': 250, 'Predominant gender among
    restaurants within a specified radius (in meters) around the
    restaurant being analyzed': 'Female', 'Predominant emotion among
    restaurants within a specified radius (in meters) around the
    restaurant being analyzed': 'Neutral', 'Predominant language among
    restaurants within a specified radius (in meters) around the
    restaurant being analyzed': 'Portuguese', 'Average price among
    restaurants within a specified radius (in meters) around the
    restaurant being analyzed': 25.5, 'Average rating among
    restaurants within a specified radius (in meters) around the
    restaurant being analyzed': 3.9, 'Average number of reviews among
    restaurants within a specified radius (in meters) around the
    restaurant being analyzed': 447}. Mention all the content of the
    dictionary unless it is an empty string. In that case, don't
    mention that specific field. This summary will be the first
    paragraph that the user sees after some brief restaurant
    information. Please remove any quoting forms when copying the
    information."
3
4 #Example answer:
5 ExampleAnswer = "This report provides an analysis of ABC-XYZ, a
    restaurant located in a 250-meter radius. Data was collected from
    174 reviews between July 27, 2020 and August 6, 2022, spanning 36

```

months. The average rating given by customers was 4.28 out of 5, with the most common emotion expressed being 'Neutral'. The predominant gender among customers was female, with a ratio of 1 man for every 2 woman. Reviews were predominantly in Portuguese, with a positive-to-negative emotion ratio of 1.62. Positive reviews most commonly mentioned 'Breakfast' and 'Chocolate cake', while negative reviews most commonly mentioned 'Space' and 'Waiting for table'. Comparatively, the predominant gender among restaurants in the same radius was female, with a predominant emotion of 'Neutral' and a predominant language of Portuguese. The average price, rating, and number of reviews among restaurants in the same radius were 25.5, 3.9, and 447 respectively."

Code 5: Code snippet including 'ExamplePrompt', which is a sample prompt sent to ChatGPT 3.5, and 'ExampleAnswer', which holds the corresponding response from the model.

```

1 #Nearby Restaurants Introduction GPT Dictionary
2 NearbyRestaurantsIntroGPTDict = {
3     "Radius (in meters) around the restaurant for comparison with other
      restaurants": res_range,
4     "Name of the restaurant being analyzed": restaurantName,
5     "Predominant gender among customers of nearby restaurants":
      RestaurantsInRangePredomGender,
6     "Predominant language among customers of nearby restaurants":
      RestaurantsInRangePredomLanguage,
7     "Predominant emotion among customers of nearby restaurants":
      RestaurantsInRangePredomEmotion,
8     "Average price of nearby restaurants": RestaurantsInRangePrice,
9     "Average rating of nearby restaurants": RestaurantsInRangeAvgRating,
10    "Total number of reviews for nearby restaurants": RestaurantsInRange.
      shape[0],
11    "Average number of reviews per restaurant for nearby restaurants":
      RestaurantsInRangeAvgNReviews,
12    "Number of positive reviews for nearby restaurants":
      RestaurantsInRangePos.shape[0],
13    "Number of negative reviews for nearby restaurants":
      RestaurantsInRangeNeg.shape[0]
14 }
15
16
17 #Nearby Restaurants Introduction Prompt String
18 NearbyRestaurantsIntroPrompt = "Write an introduction with no more
      than 150 words for the Nearby Restaurant section of a restaurant
      report intended for a restaurant owner to better understand their
      business. This introduction should inform the owner that the
      section includes bar charts showing the distribution of gender,
      language and emotion of customers and also wordclouds of the
      positive and negative reviews of restaurants in the specified

```

```
radius. Keep in mind that this section also contains the
information in the following Python dictionary: " + str(
NearbyRestaurantsIntroGPTDict) + ". Mention all the content of the
dictionary unless it is an empty string. In that case, don't
mention that specific field. Please remove any quoting forms or
parenthesis when copying the information!"
```

Code 6: Snippet of code showing variables containing the information prompted to ChatGPT 3.5 to generate the nearby restaurants introduction portion of the report.

NearbyRestaurantsIntroGPTDict contains, in the form of a Python dictionary, the names of the variables and an explanation of what they contain. NearbyRestaurantsIntroPrompt contains the task the LLM is supposed to perform, very carefully explained along with the maximum number of words, what to do in the case that some items in the dictionary are not available, and more.

A.11 ReviewSight PDF Structure

1. Title:

- (a) **Title:** The title of the report is displayed here.
- (b) **Restaurant details:** Essential information about the restaurant, such as its name, location, average price for two people, rating, and other relevant details, are included.
- (c) **Map:** An OpenStreetMap image shows the restaurant's geographical location.
- (d) **Introduction:** This section provides an overview of the report, outlining the subjects covered and the benefits it offers to the restaurant owner.
- (e) **Summary:** This section summarizes the report's main findings, incorporating information from all other sections.

2. Word Clouds

- (a) **Introduction:** This section introduces Word Clouds and explains their relevance to the Report.
- (b) **Positive Word Cloud [Image and caption]:** Two Word Cloud images (a and b) from a specific restaurant are included, along with their captions. Image a) represents a Word Cloud generated from all positive reviews, while image b) represents a Word Cloud generated from the most recent X negative reviews.

- (c) **Negative Word Cloud [Image and caption]:** Two Word Cloud images (a and b) are provided, each with its caption. Image a) displays a Word Cloud generated from all negative reviews, while image b) shows a Word Cloud based on the most recent X negative reviews.

3. Rating analysis

- (a) **Introduction:** This section introduces the rating analysis and evaluates the stability of the restaurant's rating over the past X months.
- (b) **Rating overtime (All reviews) [Graph and caption]:** Considering all reviews, a line plot illustrates the restaurant's rating over time. The accompanying caption explains the Graph.
- (c) **Rating overtime (Last reviews) [Graph and caption]:** A line plot depicting the restaurant's rating over time, with a focus on the most recent X reviews. The caption provides insights into the Graph's trends.

4. Emotion analysis

- (a) **Introduction:** This section presents an emotional analysis of the restaurant's reviews and gives an overview of the proportion of positive, neutral, and negative emotions expressed by the customers.
- (b) **Relative N° of words per review (Positive, Neutral and Negative) [Graphs and caption]:** Two line plots (a and b) are shown, depicting the relative number of words with positive, neutral, and negative emotions per review, as well as the relative number of reviews with positive, neutral, and negative emotions for all reviews.
- (c) **Relative N° of reviews (Positive, Neutral and Negative) [Graph and caption]:** Two line plots (a and b) are shown, depicting the relative number of words with positive, neutral, and negative emotions per review, as well as the relative number of reviews with positive, neutral, and negative emotions for the most recent X months.
- (d) **Gender Distribution [Table and caption]:** A table is displayed showing the absolute and relative number of male and female reviewers for all reviews and the most recent X months.

- (e) **Emotion Distribution Pie Charts [Graphs and caption]:** Two pie charts (a and b) are shown, displaying the distribution of emotions in three categories (positive, neutral, and negative) for all reviews (chart a) and for the most recent X months (chart b).
- (f) **Emotion Distribution Radar Chart [Graph and caption]:** A radar chart is shown that compares the distribution of emotions in three categories (positive, neutral, and negative) between the selected restaurant and other restaurants within a specified range.

5. Gender analysis

- (a) **Introduction:** This section presents a gender analysis of the restaurant's reviews and gives an overview of the proportion of male and female customers who have reviewed the restaurant.
- (b) **Gender distribution Bar Charts [Charts and caption]:** Two bar plots (a and b) are shown, depicting the absolute number of reviews by each gender (male and female) for the selected restaurant. Plot A shows the values for all reviews, while plot b shows the values for the most recent X months.
- (c) **Gender distribution Positive Pie Charts [Charts and caption]:** Two pie charts (a and b) are shown, displaying the relative number of positive reviews by male and female customers for all time (chart a) and for the most recent X months (chart b).
- (d) **Gender distribution Negative Pie Charts [Charts and caption]:** Two pie charts (a and b) are shown, displaying the relative number of negative reviews by male and female customers for all time (chart a) and for the most recent X months (chart b).
- (e) **Emotion of genders Radar Charts (6 classes) [Charts and caption]:** Two radar charts (a and b) are shown. Diagram A depicts the difference in emotions between genders across positive, neutral, and negative words per review and compares it to other restaurants within a specified range. The results are analyzed in the caption. Chart b depicts the difference in emotions between genders across three categories (positive, neutral, and negative words per review) and compares it to fictitious ideal best and worst restaurants. The results are analyzed in the caption.

- (f) **Emotion of genders Radar Charts (4 classes) [Charts and caption]:** Two radar charts (a and b) are shown. Chart a depicts the difference in emotions between genders across positive and negative words per review and compares it to other restaurants within a specified range. The results are analyzed in the caption. Chart b depicts the difference in emotions between genders across two categories (positive and negative words per review) and compares it to fictitious ideal best and worst restaurants. The results are analyzed in the caption.

6. Language analysis

- (a) **Introduction:** This section presents a language analysis of the restaurant's reviews.
- (b) **Language distribution Bar Charts [Charts and caption]:** Two bar charts (a and b) are shown, depicting the absolute number of reviews in Portuguese and English. Chart A shows the values for all reviews, while Chart B shows the values for the most recently selected number of months.
- (c) **Language distribution Pie Charts [Charts and caption]:** Two pie charts (a and b) are shown, displaying the relative number of reviews in Portuguese and English. Chart A shows the values for all reviews, while Chart B shows the values for the most recently selected number of months.

7. Topic analysis

- (a) **Introduction:** This section presents an overview of the topic analysis and the main topics identified in the restaurant reviews.
- (b) **Randomly selected Positive review:** A snippet from a positive review randomly selected from the restaurant's reviews is included.
- (c) **Randomly selected Negative review:** A snippet from a negative review randomly selected from the restaurant's reviews is included.
- (d) **Topics in the last X months of Positive reviews:** Up to 5 of the most relevant topics found in the positive reviews from the last X months are listed.
- (e) **Tips from topics in Positive reviews:** A tip related to each of the topics found in the positive reviews is provided.

- (f) **Topics in the last X months of Negative reviews:** Up to 5 of the most relevant topics found in the negative reviews from the last X months are listed.
- (g) **Tips from topics in Negative reviews:** A tip related to each of the topics found in the negative reviews is provided.
- (h) **Topics in the last X months of Inconsistent reviews:** Up to 5 of the most relevant topics that appeared in both positive and negative reviews are listed. These topics were excluded from the positive and negative topics and placed in this section.
- (i) **Tips from topics in Inconsistent reviews:** A tip related to each of the topics found in the inconsistent reviews is provided.

8. Nearby restaurants analysis

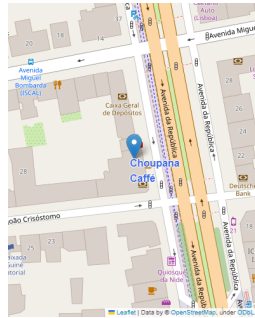
- (a) **Introduction:** This section presents an overview of the analysis of nearby restaurants within a specified range, including crucial information such as predominant gender, average rating, and total number of reviews.
- (b) **Nearby restaurants Gender, Language and Emotion Bar charts [Charts and caption]:** Three bar charts (a, b, and c) are displayed, showing the absolute numbers for the selected month period. Chart A depicts the average gender distribution among the nearby restaurants, Chart B represents the average language distribution among the nearby restaurants, and Chart C shows the average emotion distribution among the nearby restaurants, including average positive, neutral, and negative words per review.
- (c) **Nearby restaurants Positive and Negative Word Clouds [Charts and caption]:** Two Word Cloud images (a and b) from a specific restaurant are included, along with their captions. Image a is generated from all positive reviews from restaurants within the specified range, while image b is generated from all negative reviews from restaurants within the selected range.

A.12 PDF Report

Zomato Automated Report

Restaurant details:

Name:	Choupana Caffé
Date Range:	2020-09-02 to 2022-08-06
City:	Greater Lisbon
Location:	Saldanha
N° of reviews:	Total: 4706 - With text: 1897
Price for 2:	10 Euros
Price for 2 (Rank 250m range):	43° of 49 - Top 87.8%
Rating of all time:	4.2 out of 5
Rating (Last 36 months):	4.28 out of 5
Rating (Rank 250m range):	2° of 49 - Top 4.1%



Introduction:

This report provides an in-depth analysis of Choupana Caffé's performance. It is organized into multiple sections that examine different aspects of performance including ratings, customer emotions, topic detection, and gender demographics. Through this report, you will gain a better understanding of Choupana Caffé's strengths and areas for improvement.

Summary

An evaluation of the past 169 reviews for Choupana Caffé from 2020-09-02 to 2022-08-06, spanning the last 36 months, indicates that the restaurant revealed a unsatisfactory, with an average rating of 4.28 out of 5. The majority of customers are female, with a ratio of 0.47 men for every woman, and the main language used in reviews is Portuguese. Customer emotions lean towards the favorable side, with roughly 1 male customer for every 2 female customer. The most frequently mentioned topics in positive reviews are Breakfast and Chocolate cake, while in negative reviews they are Space and Waiting for table. An analysis of restaurants within a radius of 250 meters reveals that the predominant gender is Female, the most commonly used language is Portuguese, and the dominant emotion is Neutral. The average price is 25.5 Euros and the average rating is 3.9 out of 5, based on a total of 21908 reviews and an average of 447 reviews per restaurant.

2. WordClouds

In this section, you will find WordClouds that offer insights into both positive and negative customer experiences at Choupana Café, encompassing the entire history of your restaurant on this platform and the last 36 months. Words that appear more frequently in the reviews are displayed in a larger font size.

Positive Wordclouds



Figure 1:
(a): Word cloud depicting the most frequently occurring words in all positive reviews of Choupana Caffé.
(b): Word cloud exhibiting the most frequently occurring words in the last 97 positive reviews of Choupana Caffé.

Negative Wordclouds

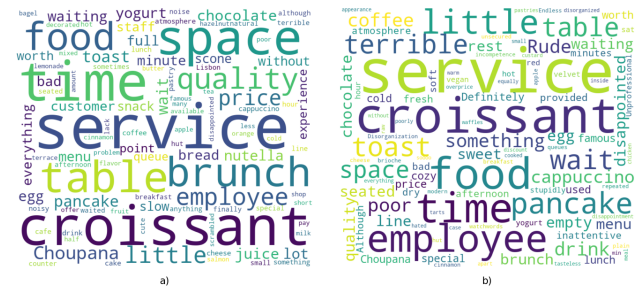


Figure 2:
a): Word cloud expressing the most frequently occurring words in all negative reviews of Choupana Caffé.
b): Word cloud demonstrating the most frequently occurring words in the last 30 negative reviews of Choupana Caffé.

3. Rating analysis

This section presents graphs that show the review ratings over time using data from all reviews and from the chosen timeslot. These graphs provide a visual representation of changes in review ratings and can help identify trends and areas for improvement.

By separating the specified timeline into two halves and analyzing the business's performance, it is evident that the business was consistent in the past 36 months.

Rating overtime (All reviews)

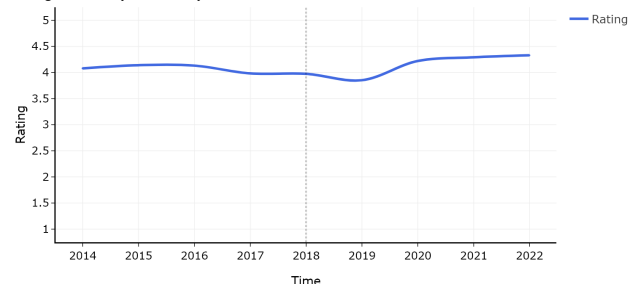


Figure 3:
Scatter plot displaying the rating of all reviews for the restaurant Choupana Caffé over time.

Rating overtime (Last reviews)

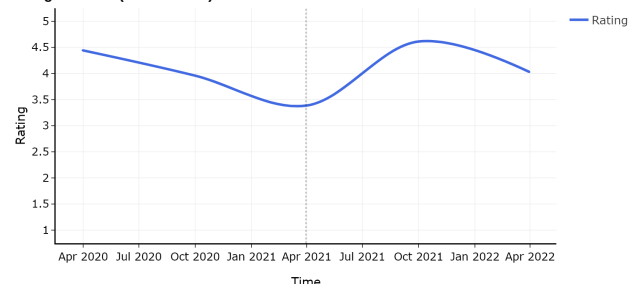


Figure 4:
Scatter plot displaying the rating of the last 36 months for the restaurant Choupana Caffé.

4. Emotion analysis

Through careful analysis of customer feedback and reviews, this section provides valuable insights into their emotional trends and overall satisfaction. This information can be used to identify areas for improvement and enhance the dining experience, leading to increased customer loyalty and repeat business. Overall, the reviews are somewhat favorable, with 1.58 positive emotions for every negative emotion.

Relative N° of words per review (Positive, Neutral and Negative)

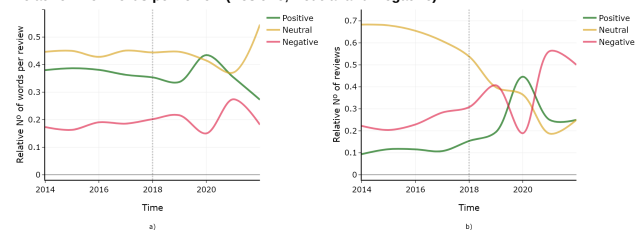


Figure 5:
a): Scatter chart illustrating the temporal distribution of emotions across all reviews of Choupana Caffé restaurant based on relative n° of positive, neutral, and negative words per review, respectively indicated by green, yellow, and red lines.
b): Scatter chart illustrating the temporal distribution of emotions across all reviews of Choupana Caffé restaurant based on the relative n° of positive, neutral, and negative review ratings, respectively indicated by green, yellow, and red lines.

Relative N° of reviews (Positive, Neutral and Negative)

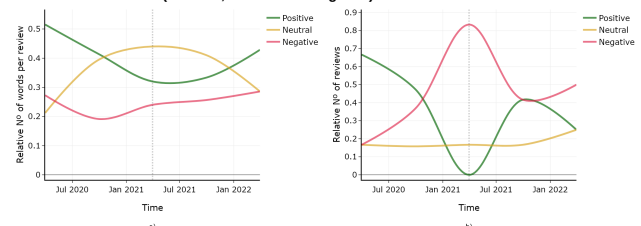


Figure 6:

a) Scatter chart illustrating the temporal distribution of emotions across the last 169 reviews of Choupana Caffé restaurant based on relative n° of positive, neutral, and negative words per review, respectively indicated by green, yellow, and red lines.

b) Scatter chart illustrating the temporal distribution of emotions across the last 169 reviews of Choupana Caffé restaurant based on relative n° of positive, neutral, and negative review ratings, respectively indicated by green, yellow, and red lines.

Gender Distribution Table

	Nº of Man - (%)	Nº of Woman - (%)
All reviews	868 (27.0%)	2341 (73.0%)
Last 36 months	49 (31.8%)	105 (68.2%)

Table 1:

Table expressing the distribution across gender of all and last 36 months of reviews for the restaurant Choupana Caffé.

Emotion Distribution Pie Charts

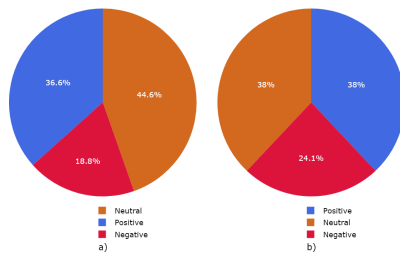


Figure 7:

(a): Pie chart expressing the emotion distribution of all reviews for the restaurant Choupana Caffé.
(b): Pie chart displaying the emotion distribution of the last 169 reviews for the restaurant Choupana Caffé.

Emotion Distribution Radar Chart

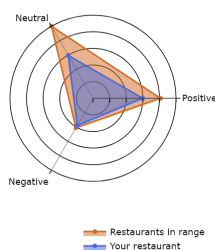


Figure 8:

Radar chart comparing the distribution of emotions in the last 169 reviews for the restaurant Choupana Caffé with those of others in a 250 meters range.

Page 5/12

zomato

5. Gender analysis

This section offers a detailed look at the gender demographics of Choupana Caffé's customers and how they compare to those of the ideal best and ideal worst restaurants. It contains bar charts that show the total count of reviews for each gender and radar charts divided into 2 (Positive and Negative) and 3 (Positive, Neutral, and Negative) classes that represent the percentage of men and women in each segment. A comparison of Choupana Caffé to the ideal best and ideal worst restaurants is also provided in this section through a radar chart.

Taking gender analysis into account, it seems that the female group tends to visit the restaurant more frequently, with a ratio of roughly 1 male customer for every 2 female customer.

Gender distribution Bar Charts

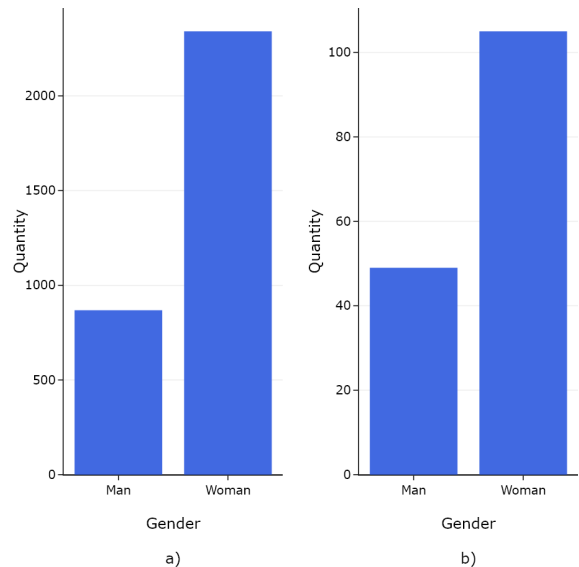


Figure 9:

(a): Bar plot exhibiting the distribution of gender of all reviews for the restaurant Choupana Caffé.
(b): Bar plot presenting the distribution of gender for the last 169 reviews for the restaurant Choupana Caffé.

Page 6/12

zomato

Gender distribution Positive Pie Charts

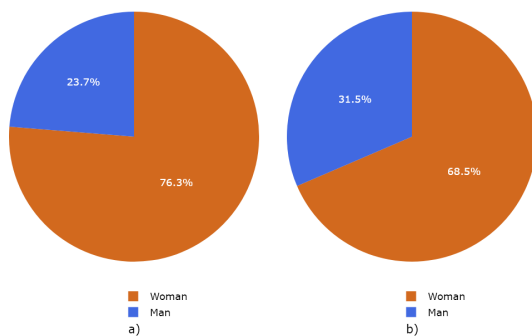


Figure 10:

(a): Pie chart illustrating the distribution of gender of all positive reviews for the restaurant Choupana Caffé.
(b): Pie chart showcasing the distribution of gender for the last 97 positive reviews for the restaurant Choupana Caffé.

Gender distribution Negative Pie Charts

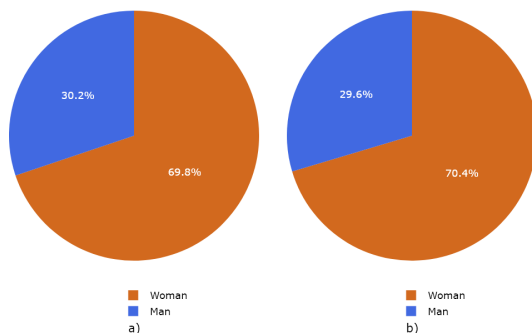


Figure 11:

(a): Pie chart representing the gender distribution among all negative reviews of Choupana Caffé.
(b): Pie chart exhibiting the distribution of genders in the last 30 negative reviews of Choupana Caffé.

Page 7/12

zomato

Emotion of genders Radar Chart (6 classes)

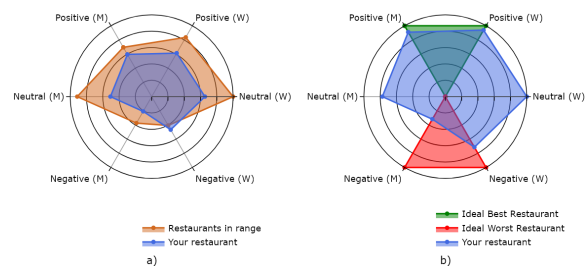


Figure 12:

(a): Radar chart expressing the proportion of positive, neutral, and negative emotions among men (M) and women (W) in the last 169 reviews of Choupana Caffé and comparing it to other restaurants in a 250 meters range. When compared to other restaurants, your restaurant elicits reduced positive emotions, reduced neutral emotions, and lower negative emotions in men. For women, the data indicates lower positive emotions, reduced neutral emotions, and increased negative emotions.
(b): Radar chart showing the relative distribution of positive, neutral, and negative emotions among male and female customers at Choupana Caffé, based on the last 169 reviews. The chart also compares Choupana Caffé to the ideal best and worst restaurants. Analysis indicates that no specific gender tend to express more positive emotions, and woman often show more negative emotions.

Emotion of genders Radar Chart (4 classes)

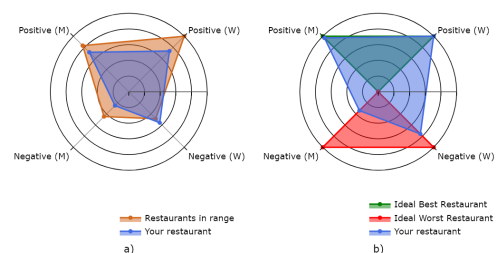


Figure 13:

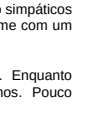
(b): Radar chart depicting the relative distribution of positive and negative emotions among male and female customers at Choupana Caffé, based on the last 169 reviews. The chart also compares Choupana Caffé to the ideal best and worst restaurants. Analysis indicates that no specific gender tend to express more positive emotions, and woman often show more negative emotions.
(b): Radar chart portraying the distribution of positive and negative emotions among male and female customers of Choupana Caffé restaurant, in comparison to the ideal best and ideal worst restaurants based on the last 169 reviews.

Page 8/12

zomato

and negative
in relation

es of both a

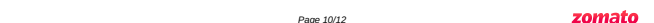


on how to

- restrictions and
fresh juices
redients, and

To improve

- ing options
expected for



average price
showing the
and negative
reveals that
the dominant
the average
reviews per
oud graphs



a 250 meters

a 250 meters

A.13 EMOTAIX.PT lexicon

- 1 POSITIVO;POSITIVO;POSITIVO;POSITIVO;POSITIVO;POSITIVO;POSITIVO;
POSITIVO;NEGATIVO;NEGATIVO;NEGATIVO;NEGATIVO;NEGATIVO;NEGATIVO;
NEGATIVO;NEGATIVO;POSITIVO;POSITIVO;POSITIVO;NEGATIVO;NEGATIVO;
NEGATIVO;POSITIVO;POSITIVO;NEGATIVO;NEGATIVO;POSITIVO;POSITIVO;
POSITIVO;NEGATIVO;NEGATIVO;NEGATIVO;POSITIVO;POSITIVO;NEGATIVO;
NEGATIVO;POSITIVO;POSITIVO;POSITIVO;NEGATIVO;NEGATIVO;NEGATIVO;
POSITIVO;POSITIVO;NEGATIVO;NEGATIVO;POSITIVO;POSITIVO;NEGATIVO;
NEGATIVO;NEUTRO;NEUTRO;NEUTRO
- 2 BENEVOLÊNCIA;BENEVOLÊNCIA;BENEVOLÊNCIA;BENEVOLÊNCIA;BENEVOLÊNCIA;
BENEVOLÊNCIA;BENEVOLÊNCIA;BENEVOLÊNCIA;MALEVOLÊNCIA;MALEVOLÊNCIA;
MALEVOLÊNCIA;MALEVOLÊNCIA;MALEVOLÊNCIA;MALEVOLÊNCIA;MALEVOLÊNCIA;
MALEVOLÊNCIA;BEM ESTAR;BEM ESTAR;BEM ESTAR;MAL ESTAR;MAL ESTAR;MAL
ESTAR;BEM ESTAR;BEM ESTAR;MAL ESTAR;MAL ESTAR;BEM ESTAR;BEM ESTAR
;BEM ESTAR;MAL ESTAR;MAL ESTAR;MAL ESTAR;BEM ESTAR;BEM ESTAR;MAL
ESTAR;MAL ESTAR;BEM ESTAR;BEM ESTAR;BEM ESTAR;MAL ESTAR;MAL ESTAR;
MAL ESTAR;SEGURANÇA;SEGURANÇA;ANSIEDADE;ANSIEDADE;SEGURANÇA;
SEGURANÇA;ANSIEDADE;ANSIEDADE;SURPRESA;EMOÇÕES NÃO ESPECÍFICAS;
INDIFERENÇA
- 3 AFEIÇÃO;AFEIÇÃO;AFEIÇÃO;AFEIÇÃO;GENTILEZA;GENTILEZA;GENTILEZA;
GENTILEZA;ÓDIO;ÓDIO;ÓDIO;ÓDIO;AGRESSIVIDADE;AGRESSIVIDADE;
AGRESSIVIDADE;AGRESSIVIDADE;FELICIDADE;FELICIDADE;FELICIDADE;
SOFRIMENTO;SOFRIMENTO;SOFRIMENTO;LUCIDEZ;LUCIDEZ;LOUCURA;LOUCURA;
ENTUSIASMO;ENTUSIASMO;ENTUSIASMO;DEPRESSÃO;DEPRESSÃO;DEPRESSÃO;ALÍ
VIO;ALÍVIO;PERTURBAÇÃO;PERTURBAÇÃO;SATISFAÇÃO;SATISFAÇÃO;SATISFAÇ
ÃO;INSATISFAÇÃO;INSATISFAÇÃO;INSATISFAÇÃO;CORAGEM;CORAGEM;MEDO;MEDO
;CALMA;CALMA;TENSÃO;TENSÃO;;;
- 4 AMOR;DESEJO;ADMIRAÇÃO;ATRACÇÃO;BONDADE;DOÇURA;PACIÊNCIA;HUMILDADE;
RESSENTIMENTO;DESGOSTO;DESPREZO;IRRITAÇÃO;CRUELDADE;RAIVA;IRA;
ORGULHO;FELIZ;PRAZER;RISO;DRAMA;DOR;CHORO;SAÚDE MENTAL;EQUILÍBRIO;
DOENÇA MENTAL;DESIQUILÍBRIO;ALEGRIA;VIVACIDADE;ALERTA;TRISTEZA;
FADIGA;APATIA;APAZIGUAR;LIBERTAÇÃO;AGITAÇÃO;REMORSO;ESTIMA;
SACIEDADE;AGRADO;HUMILHAÇÃO;FRUSTRAÇÃO;DESAGRADO;AUDÁCIA;EXTROVERS
ÃO;PAVOR;INTROVERSÃO;TRANQUILIDADE;RELAXAMENTO;ANGÚSTIA;INQUIETUDE
;;;
- 5 abraçar;ambição;aclamação;aprazível;acolhedor;acarinhar;conformação;
brio;animosidade;asco;abandalhação;amuado;abruptamente;abusar;có
lera;absoluto;afortunado;arreatador;anedota;adversativo;agonia;
choradeira;bem-estar;astuto;abstinência;aberração;alegrar;activo;
alerta;abandonado;abarrotado;amorfo;acalmar;autenticidade;abalada;
arrepender;abençoado;cheio;agradar;afronta;ausência;aborrecer;
afrontar;autonomia;abominado;acanhado;acalmia;desanuviamento;aflig
ção;alarido;abananado;afectado;alexitimia
- 6 abraço;ambicionar;aclamar;atractivo;acolhedor ;aconchegar;conformado;
briosa;antipatia;ascorosidade;abandalhado;amuar;abrupto;abusivo;

- colérico; altivez; beatitude ; desfrutação; burlesco; adversidade; agoniar; choramingar; equilibrar; bom senso; abuso; absurdo; alegre; alacridade; alertar; abandonada; abarrotar; apatia; ajuda; autêntico; abalado; arrependido; abençoar; completa; agradáveis; afrontado; ausentar-se; aborrecidamente; ânimo; autônomo; abominar; acanhamento; brando; desanuviar; afligir; alarmante; abanar; afectar; ataraxia
- 7 adoração; ambicioso; admiração; atraente; acolher; amenidade; conformar; brioso; antipático; ascoroso; abandalhamento; amuo; agreste; agredir; encolerizar; altivo; beato ; desfrutar; comédia; adverso; agônico; choramingas; razão; brilhante; adicto; anormal; alegremente; animação; atenção; abandonar; abater; apática; ajudar; cândido; abalar; arrependimento; apreço; completo; agradável; afrontamento; ausente; aborrecido; animoso; comunicativo; abominável; complexado; brandura; descompressão; aflitivo; alarmar; admiração; agudizar; desmotivação
- 8 adorada; anseio; admirado; atrair; acolhida; angelical; esperar; carácter; antipatizar; aversão ; abandallar; arrepiar; artificial; agressão; indignação; autólatria; contentamento; desfrute; cômico; azar; agonizante ; chorão; razoável; brilhantismo; adoecer; anormalidade; alegrete; animado; atentamente; abandono; abatida; apático; aliviar; desbloquear; abalo; censurável; condescendência; fartura; agradavelmente; afrontosamente; carência; aborrecimento; arrojado; desinibição; amedrontado; complexo; calma; descomprimir; aflito; alarme; admirado; aperfeiçoamento; desmotivado
- 9 adorado; ânsia; admirar; charme ; acolhimento; angélico; longânime; caracterial; charlatão; badalhocar; abominação; assediado; artificialidade; agressivamente; indignado; autolatria ; contentar; desfruto; desentediado; azarar; agonizar; chorar; sanidade; cautela; adoentado; apanhado; alegria; animar; atento; afectação; abatido; atonia; alívio; desbloqueio; abalroação; compunção; condescendente; fastio; agrado; afrontoso; carenciar; adulteração; atrever-se; desinibido; amedrontar; embaraço; calmante; descontração; amargor; alarmismo; embasbacado; aperfeiçoar; desmotivar
- 10 adorar; ansiar; admirável; decência; afabilidade; caloroso; longanimidade; despretensão; ciúme; baixeza; absolutista; assediador; artificialismo; agressividade; ira; bazófia; contente; estimulação; engraçado; azarento; agravação; choroso; saúde mental; cauteloso; afasia; aparvalhado; bem-humorado; aperaltada; concentração; amargoso; abatimento; atônico; apaziguado; descargo; abalroada; condenada; condescender; justeza; aprimorado; aniquilação; castração; adulterar; atrevido; desinibir; apavorado; embaraçoso; calmo; descontraído; amargura; alarmista; embasbacar; arrepiado; estóico
- 11 adorável; apetência; admiravelmente; deslumbrada; afável; carícia; pachorra ; despretensiosamente; ciumento; conspurcação; admoestação; assediar; áspero; agressivo; iracundo; egocêntrico; êxito; estimulante; espirituoso ; calamidade; agravamento; desengraçado; ; circunspeção; agnosia; aparvalhar; bom humor; aperaltado; concentrado; cabisbaixo; abulia; atordado; apaziguamento; desculpabilização; abalroamento; condenado; consideração; justiça; aprimoramento; aniquilamento; coibição; adultério; atrevimento; efusivamente; apavorante; inibição; descansado;

- descontrair; amarguradamente; apreensão; espantado; arrepio; impassibilidade
- 12 afectividade; apetite; adulação; deslumbrado; afavelmente; desculpa; pachorrentamente; despretenso; desunião; conspurcar; admoestar; assédio; asquerosidade; agressor; irar; egocentrismo; felicidade; estimular; galhofa; calamitoso; agravante; desengraçar;; circunspecto; agorafobia; bêbado; brincadeira; aperaltar; concentrar; cancro; abúlico; atordoamento; apaziguar; desculpabilizar; abalroar; condenável; considerado; justiceiro; aprimorar; aniquilar; coibido; adúltero; audácia; efusivo; apavorar; inibido; descansar; distender; amargurar; apreensivo; espantar; arrepios; impassível
- 13 afectivo; apetitoso; adular; deslumbramento; altruísmo; desculpar; pachorrento; discreto; desunir; desconsolação; advertência; carrancudo; asqueroso; amaldiçoar; irascível; egocentrista; felicitação; euforia; galhofada; catástrofe; agravar; lacrimação;; equilibrado; alienação; bebedeira; brincalhão; arrebatado; despertado; contristar; adinamia; atordoante; apoiado; desenlaçar; abanação; contrição; considerar; justo; bela; avermelhado; coibir; chateado; audaciosamente; expansão; assombrado; inibir; descanso; distendido; angústia; contracção; espanto; atenção; impávido
- 14 afecto; aspiração; aplaudir; deslumbrante; altruísta; desenraivar; paciência; descrição; detestação; desconsolar; advertir; desbocado; astúcia; ameaça; rebelde; relação; felicitar; eufórico; galhofar; catastrófico; agravo; lacrimante;; equilíbrio; alienado; confusão mental; brincar; arrebatamento; despertar; cremação; adinâmico; atordoar; apoiar; desenlace; abanadela; contrito; cumprimentar; plenitude; belo; avermelhar; defeito; chatear; audacioso; expansividade; assombrar; insocial; despreocupação; distensão; angustiado; contraído; espantosamente; atento; imperturbabilidade
- 15 afectuosa; aspirante; aplauso; deslumbrar; amabilidade; desenraivecer; paciente; humildade; detestar; desconsolo; afastado; desbocar; atrocidade; ameaçado; rebeldia; empertigado; feliz; êxtase; galhofeiro; derrota; atormentado; lacrimejar;; estabilidade; alienante; contra-senso; buliçoso; arrebatado; prevenção; cremar; astenia; atroador; apoio; desenlear; abanado; culpa; cumprimento; pleno; bonita; blasfemar; desistência; chatice; audaz; expansivo; assombro; introversão; fleuma; ócio; angustiador; contrair; espantoso; atitude; imperturbável
- 16 afectuosamente; aspirar; aprovação; encantada; amável; dócil; pacientemente; humilde; detestável; descontentamento; afastamento; desencantação; atroz; ameaçador; rebeldismo; empertigar; felizmente; extasiado; galhofento; derrotismo; atormentar; lacrimoso;; estável; alienar; delirante; desenfadar; brejeiro; prevenido; crematório; asténico; desconcentração; assistência; desenleio; afã; culpabilidade; deferência; realizada; bonito; blasfêmia; desistir; chato; autodeterminação; expressividade; assustada; introvertido; fleumático; ociosamente; angustiante; desinquietação; espectacular; avermelhado; inabalável
- 17 afectuosidade; avidamente; aprovar; encantado; amavelmente; docilidade; permissão; humildemente; embusteiro; descontentar; afastar; desencantamento; austeridade; ameaçar; rebelião; enfatuar-se; fortuna;

- extasiar; gargalhada; desastre; comovedor; lágrima;; fino; alienável;
delirar; desenfado; capitoso; vigiar; deprimente; atrofia;
descuidadamente; assistir; desimpedido; agitação; culpabilizar;
deferente; realizado; deleitação; blasfemo; dificuldade; contrariado;
autodeterminismo; expressivo; assustadiço; mistério; harmonia;
ociosidade; angustiar; desinquietar; espetacularmente; avermelhar;
indiferença
- 18 afectuoso; avidez; aprovativamente; encantar; atencioso; docilmente;
permissivamente; inferior; enciumar; descontente; amor próprio ;
desmesura; austero; bruscamente; retilho; envaidecer; prosperar; extá
tico; gargalhar; desastoso; comovente; lamentação;; gênio; aluado; delí
rio; desenfastiadamente; desembaraçado; vigo; deprimida; atrofiar;
descuidado; beneficiente; desintegração; agitado; culpado; dignamente;
repleto; deleitado; corar; dificultar; contrariante; aventura; extravers
ão; assustado; misteriosamente; harmonioso; ocioso; angustioso;
desinquieto; estupefacto; avivar; indiferente
- 19 afeição; ávido; aprovativo; encanto; beleza; fofo; permissividade;
inferioridade; execração; desgostar; arrogância; desprimor;
barbaramente; brusco; refilar; envaidecidamente; próspero; exultação;
graça; descalabro; comover; lamentar;; lucidamente; alucinação; desdizer
; desenfastiar; desembaraçamento; vigilância; deprimido; burnout;
descuidar; benefício; desintegrar; agitador; culposamente; dignidade;
saciado; deleitar; descrédito; estorvar; contrariar; aventurar;
extrovertido; assustador; misterioso; harmonização; relaxação; ânsias;
esbaforido; imprevisto; beatificação; indolência
- 20 afeiçoada; capricho; atônito; excitação; benéfico; indulgência; permissivo;
inferiorizar; execrar; desgosto; arrogante; desprimorar; barbaridade;
brusquidão; revolta; envaidecido; sorte; exultar; gracejar; descambar;
comovido; lamentável;; lucidez; alucinações; desembestadamente; diversã
o; desembaraçar; vigilante; deprimir; cansaço; descuido; benemerência;
desintrincar; agitar; desleal; digno; saciar; deleitável; descrença;
estorvo; contrariedade; aventureiro; incitabilidade; assustar;
pudicamente; harmonizar; relaxado; ansiedade; esbaforir;
inesperadamente; beatificar; indolente
- 21 afeiçoado; caprichoso; bestial; excitado; benevolência; indulgente;
permitir; modestamente; execrável; desgostoso; autoridade; embirra;
barbarismo; brutal; revoltado; envaidecimento; sortuda; festa; gracejo;
desdita; condolência; lamentavelmente;; lúcido; alucinado;
desenfreadamente; divertido; dessempecar; vigília; desamparado; cansado;
descuidosamente; benemérito; desirmanar; azáfama; deslealdade; encômio;
saciável; leite; descrer; falha; decepção; aventureiro; incitação;
atemorizado; púdico; impassibilidade; relaxamento; ansiosa; hipertensão
; inesperado; beato; indolentemente
- 22 afeiçoar; cobiça; bestialidade; excitante; benevolente; indulgentemente;
perseverança; modéstia; falsário; desprazer; autoritário; embirração; bá
rbaro; brutalidade; revoltante; fanfar; sortudo; folia; graciosidade;
desditoso; contorção; lamento;; meticulosamente; alucinante; dessempear
; divertimento; dessempecilhar;; desamparar; cansar; descuradamente;
benfeitor; deslaçamento; azafamar; desonestar; enobrecedor; saciedade;

- deleitoso; desdouro; falhar; decepcionado; bravo; incitado; atemorizar; pudor; impassível; relaxante; ansiosamente; hipertenso; pasmado; bênção; indolor
- 23 afinidade; curiosamente; bestialmente; excitar; benévolo; mansidão; perseverante; modesto; falsidade; ignóbil ; canceroso; embirrantemente; besta; brutalizar; revoltar; fanfarrão; ventura; jovial; gracioso; desdramatizar; contorcer; lamentoso;; metuculoso; alucinar; desequilibrado; divertir; diligência;; desamparo; cansativo; descurar; benfeitora; deslaçar; barafunda; expiação ; enobrecimento;; delicioso; desempregar; falido; decepcionante; bravura; incitador; aterrador; pusilânime; límpido; relaxar; ansioso; inquietação; pasmar; bendito; insensibilidade
- 24 agradecer; curiosidade; contemplação; excitável; bondade; meigamente; perseverantemente; obedecer; falsidão; ignobilidade; castigar; embirrar; cruel; brutamontes; revolto; fanfarrear; venturoso; jovialidade; graça; desfavorável; deploração; lamúria;; minúcia; amnésia; desequilibrar; enreter; diligenciar;; desanimada; colapso; desinteressado; catarse; desmantelamento; barafusta; expiar; estima;; elogiar; deslustrar; falir; decepcionar; capaz; incitamento; aterrar; pusilanimidade; manso;; apertar; inquietante; pasmo; boémia; insensível
- 25 agradecida; curioso; contemplar; fascinação; bondoso; meigo; perseverar; obediência; falsificação; impureza; castigo; embirrativo; crueldade; bruto; revolução; fanfarrice;; jovialmente; hílare; desfavorecer; deplorar; lamuriar;; minuciosamente; anorética; desequilíbrio; entretenimento; diligente;; desanimado; consternação; desinteressar; confortável; desmantelar; barafustar; falível; estimar;; elogio; deslustre; faltar; degradante; confiança; incitante; aterrorizado; reservadamente; melódico;; coagido; inquietar; sideração; boêmio; invalidez
- 26 agradecido; desejar; encantador; fascinante; caridade; meiguice; persistência; obediente; falsificador; impuro; censura; embirrento; descortês; conflito; revolucionar; fanfarronar;; jubilar; hilariante; desfortuna; deplorável; lamuriento;; minucioso; anorexia; desestabilização; folgar; dinâmico;; desanimar; consternar; desinteresse; conforto; despegar; baralhada; falivelmente; homenagear;; engrajar; desmerecer; fome; desagradado; confiante; incitar; aterrorizador; reservado; melodiosamente;; coagir; inquieto; siderado; borrachão; inválido
- 27 agradecimento; desejável; encantamento; fascinar; caridoso; mimar; persistente; princípios; falsificar; imundíce ; censurado; embravecer; descortesia; conflituoso; revolucionário; farsola;; júbilo; hilaridade; desgraça; deploravelmente; lástima;; moderado; asperger; desestabilizar; folgaz; dinamismo;; desesperado; debilitação; desleixado; consertar; despego; baralhado; falta; homenagem;; enriquecedor; desmerecimento; frustração; desagradar; coragem; incitativamente; aterrorizante; timidamente; melodioso;; compelido; inquietude; siderar; borracheira; neutralidade
- 28 alcovitar; desejo; encantos; fascínio; clemência ; mimo; persistentemente; pundonor; falso; imundo ; censurar; embravecimento; desnaturado; disputa; sublevação; farsolar;; jubiloso; humor; desgraçadamente; desolação;

- pranto;; parcimônia; autismo; desmentido; folgazão; dinamizar;;
 desesperança; debilitante; desleixamento; consolação; desprender;
 baralhar; improbidade; honorificamente;; enriquecedora; desonra;
 frustrada; desagradável; corajosamente; incitativo; aterrorizar;
 timidez; pacatamente;; compelir; opositor; sobressaltado; comovente;
 neutro
- 29 alcoviteira; desejoso; esplêndido; fenomenal; clemente ;mimos; persistir;
 pundonoroso; hipocrisia; javardo; chacota; encanzinar; desnaturar;
 disputador; sublevar; fatuidade;; lépido; humorismo; desgraçado;
 desolada; soluçar;; parcimoniosamente; autista; desmentir; genial;
 empolgamento;; desesperante; debilitar; desleixar; consolada;
 desprendimento; bizarria; perda; honorífico;; esbelteza; desonrado;
 frustrado; desagradavelmente; corajoso; incitável; cagaço; tímido;
 pacatez;; constrangedor; perigo; sobressaltar; comover;
- 30 alcoviteiro; desiderato; esplendor; fogsidade; comiseração; mimoso;
 resignação; séria; hipócrita; lixeira; cínico; encrespado; desonestidade
 ; disputar; tempestuoso; fátuo;; prazer; humorista; desgraçar; desolado;
 soluço;; parcimonioso; bipolar; desmiolado; genialmente; empolgante;;
 desesperar; decepar; desleixo; consolado; direito; bizarro; pesar; honra
 ;; esbelto; desonrar; frustrante; desagrado; desaforado; independência;
 cagarolas; timorato; pacato;; constranger; perigosamente; sobressalto;
 comovido;
- 31 alcovitice; erótico; esplendorosamente; galar; compadecer; pequerrucho;
 resignadamente; seriedade; hipocritamente; lixo; cinismo; encrespamento
 ; déspota; endiabrado; vandalismo; fausto;; regozijar; humorístico;
 desgracioso; desolador;; perspicácia; bulimia; desordeiro; jucundo;
 empolgar;; desespero; desânimo; desmaiado; consolador; emancipação;
 bloqueio; remorsos; honrada;; estupendamente; desonroso; frustrar;
 desapontado; desaforar; independente; caguefe;; pacificamente;;
 constrangido; perigosidade; surpreendente; corar;
- 32 amada; erotismo; esplendoroso; interessado; compaixão; perdão; resignado; sé
 rio; hostil ;moléstia; crítica; encrespar; despotismo; endiabrar; vâ
 ndalo; faustoso;; regozijo; jocosidade; desventura; desolar;;; perspicaz
 ; bulímica; despautério; ledos; emulação;; desproteger; desconforto;
 desmaiar; consolar; emancipar; confundir; reprovável; honradamente;;
 estupendo; despeito; impedimento; desapontamento; desavergonhado;
 lingüareiro; cobarde;; pacífico;; constrangimento; perigoso;
 surpreendentemente; criatividade;
- 33 amado; erotização; fabulosamente; interessante; compassivo; perdoar;
 resignar; simplicidade; hostilidade; nojento; criticar; encruado;
 destruição; energúmeno; zanga; gabar;; sonhar; jocoso; desventurado; doen
 ça;;; ponderação; catatonia; desvairado; optimismo; emular;;
 desprotegido; desencorajador; desmaio; cura; francamente; confusão;
 resipiscência; honradez;; formoso; desprestigiante; imperícia;
 desapontar; descarado; sobressair; cobardia;; pacifismo;; constringir;
 perseguição; surpreender; criativo;
- 34 amante; esperado; fabuloso; interessar; complacência; perdoável; resistê
 ncia; simplificação; hostilizar; nojo; depreciação; encruamento;
 destruidor; enfurecer; zangado; gabarola;; sonho; laracha; desventurar;

- doente;;;ponderado;catatônico;desvairamento;optimista;energico;;
 embotamento;desencorajamento;desmazelar;curado;franco;confuso;
 ruminação;honrado;;;formosura;desprestigiar;inábil;descabidamente;
 descaramento;socialização;cobardolas;;;pacifista;;;desasossegado;
 perseguir;surpreendido;descoberta;
- 35 amar;esperança;fama;interesse;complacente;suave;resistente;
 simplificar;hostilmente;obscenidade;depreciar;encruar;destruir;
 enfurecimento;zangar;gabarolice;;;suportável;palhaçada;drama;doer
 ;;;ponderar;ciclotimia;desvairar;recreio;entreter;;;embotar;
 desencorajar;desmazelo;curar;honestamente;convulsão;rumnar;honrar
 ;;;fruitivo;desprestígio;inabilidade;descabido;descoco;socializar;
 decapitação;;;paz;;;desasossegar;preocupação;surpresa;descobridor;
- 36 amiga;esperançado;fantástico;irresistível;comprazer;suavemente;
 resistir;simplista;impostor;obsceno;descasca;enervação;destrutiva;
 enraivar;;;imodestamente;;;triumfal;palhaço;dramaticamente;dolente
 ;;;precaução;ciclotimica;desvairo;viver;entusiasmada;;;endecha;
 desesperançar;dormitar;desafogado;honestidade;convulsionado;;;
 honrosamente;;;gabarito;difamação;inabilitar;desencantado;desembara
 ço;socialmente;degolar;;;placidamente;;;desasossego;preocupada;
 surpreso;descobrimento;
- 37 amiga;esperançar;glória;irresistivelmente;compreensivo;suavidade;
 suportar;;;impostura;porcaria;descascar;enervado;destrutivo;
 enraivecer;;;imodéstia;;;triumfalismo;patético;dramático;dolorosa;;;
 prudência;ciclotimico;desvario;;;entusiasmado;;;enfado;desfalecer
 ;embrutecedor;desafogar;honesto;convulsionar;;;honroso;;;gosto;
 difamar;inabilmente;desencantar;desembaralhar;sociável;demoníaco;;
 placidez;;;desconchavar;preocupadamente;;;descobrir;
- 38 amigável;esperançoso;glorificação;lascivamente;cordial;suavizar;toler
 ância;;;imposturar;refugar;descompor;enervamento;destrutor;
 enraivecido;;;imodesto;;;triumfalista;pitoresco;dramatização;
 dolorosamente;;;prudente;cleptomania;devaneio;;;entusiasmante;;
 entediado;desfalecido;embrutecer;desafogo;impoluto;convulsivamente
 ;;;proteção;;;gostosamente;difamatório;inaptidão;desencanto;desfaç
 atez;tagarela;desconfiado;;;plácido;;;desconchavo;preocupado;;
 descompassadamente;
- 39 amigavelmente;esperar ;glorificar;lascívia;cordialidade;ternamente;
 tolerante;;;imposturice;refugo ;descompostura;enervante;
 desumanidade;fúria;;;imponência;;;triumfalmente;ridículo;dramatizar;
 doloroso;;;prudentemente;compulsivo;discórdia;;;entusiasmar;;
 entediar;desfalecimento;embrutecido;desbravar;indultar;convulsivo
 ;;;proteccionismo;;;gostoso;embaraçado;inapto;desenganador;
 deslavamento;tagarelar;desconfiança;;;repousado;;;desembestar;
 preocupante;;;descompassado;
- 40 amigo;ganância;gloriosamente;lascivo;cordialmente;terno;tolerar;;
 infidelidade;repelir;desconceituar;enervar;desumano;furibundo;;
 imponente;;;triumfante;rir;fatal;dor;;;resguardar;delirium;
 disparatar;;;entusiasmo;;;entenebrece;desgastante;embrutecimento;
 descarregar;indulto;crise;;;protector;;;inaudito;embaraçar;

- incapacidade; desenganar; deslavar; tagarelice; desconfiar;; repousante;; desesperação; preocupar;; desconhecer;
- 41 amistoso; ganancioso; glorioso; sedução; cortês; ternura; tolerável;; infiel; repugnância; desconsideração; escabrosidade; disfaçado; furiosamente;; importância;; triunfar; risada; fatalidade; dor aguda;; resguardo; demência; disparate;; entusiasta;; entristecedor; desgastar; ensonado; desembravecer; inocência; desassossego;; proteger;; incrível; encalistar; incapacitar; desengano; desplante;; diabo;; repousar;; enlouquecer; pressagiar;; desconhecimento;
- 42 amizade; inveja; herói; sedutor; cortesia;;; ingratamente ; repugnante; desconsiderar; escabroso; disfarçar; furioso;; importante;; triunfo; riso; fatalismo; dorido;; responsabilidade; demencial; divagação;; entusiástico;; entristecer; desgaste; entorpecer; desencrespar; inocentar; descomposto;; respeitabilidade;; jeitoso; encalistrar; incapaz; desiludido; destemido;; diabólico;; repouso;; enlouquecido; presságio;; desencontrar;
- 43 amor; invejar; heroicamente; seduzir; delicadeza;;; ingratitude ; repugnar; descorado; exacerbação; disfarce; furor;; intratável; volúpia; risonho; fatalista; enlutado;; responsabilizar; demente; divagar;; exaltação;; entristecimento; desmoralização; entorpecimento; desopressão; inocente; desconcertada;; respeitador;; linda; endrômina; incompetência; desiludir; destemor;; esgazeado;; serenamente;; enlouquecimento; pressagioso;; desencontro;
- 44 amorosamente; invejável; heróico; sensual; delicado;;; ingrato; repulsa; descorar; exacerbar; dissimulação; luta;; jactância;; voluptuosamente; risota; fatalmente; enlutar;; responsável; dependência; doidice;; exaltado;; esfriamento; desmoralizador; farrapo; desoprimir; integridade; desconcertado;; respeitar;; lindamente; endrominar; incompetente; desilusão; desvergonha;; esgazear;; serenidade;; esmagador; pressão;; desgrudar;
- 45 amoroso; invejosamente; heroína; sensualidade; doação;;; inimigo; repulsivo; desdém; exageradamente; dissimulado; lutar;; jactancioso;; voluptuosidade; sorridente; fatídico; gemente;; sagacidade; dependente; doidinho;; exaltante;; esfriar; desmoralizante; ignavo; disciplina; integro; desconcertante;; respeitável;; lindeza; enegrecer; indigência; despeitado; desvergonhar;; espavorido;; sereno;; esmagamento; pressentimento;; desígnio;
- 46 apaixonada; invejoso; heroísmo; sensualmente; doador;;; inimizade; sujar; desdenhar; exagerado; dissimular; maldizente;; jactar-se;; voluptuoso; sorrir; fracassar; gemer;; sagaz; depressão; doidivas;; exaltar;; esmorecer; desmoralizar; inanimado; disciplinar; isenção; desconcertar;; respeitavelmente;; lindo; enegrecimento; indigente; despeitar; determinação;; espavorir;; sossega;; esmagar; pressentir;; desigualdade;
- 47 apaixonada; predisposto; idólatra; sexual; doar;;; insídia; sujeira; desdenhável; exagerar; ditador; maldizer;; majestade;; sorriso; fracasso; gemido;; sensatamente; depressiva; doido;; expedito;; esmorecido; embebedado; inânime; encorajamento; leal; desconcerto;;

- respeito;; louvação; envergonhado; indigentemente; desprazimento;
determinado;; estremecer;; sossegada;; forçado; pressionado;; desmaiar;
48 apaixonadamente; tendência; idolatradamente; sexualidade; embelezar;;;;
insidiosamente; sujidade; desdenhoso; exagero; ditadura; maltratar;;
majestosamente;;; imolação; imprecação;;; sensatez; depressivo;
ebriedade;; farra;; esmorecimento; embebedar; inconsciente; encorajar;
lealdade; desconfortante;; respeitosa; louvado; envergonhar;
injustamente; despropositado; determinar;; estremecido;;
sossegadamente;; forçar; pressionar;; desmaio;
- 49 apaixonado; vontade; idolatrar; sexualmente; empatia;;; intrusão; sujo;
deseducar; exasperação; esfacelamento; praquejar;; majestoso;;; imolar;
imprecar;;; sensato; desorganizado; ébrio;; fervente;; exasperadamente
enfraquecedor; indolência; moderação; lealmente; desconfortar;; reverê
ncia;; louvável; envilecer; injustiça; enfadamento; enfrentar;;
estremecimento;; sossegado;; opressão; stress;; desmascaramento;
- 50 apaixonante;; idolatria; subjugação; empático;;; intrujar; vil;
deseducativo; exasperante; esfacelar; provocação;; obstinação;;;;
infausto; insuportável;;; despersonalização; ensandecer;; festas;;
exaspero; enfraquecer; indolente; pacificador; liberdade; desconfortá
vel;; reverenciosamente;; louvavelmente; envilecimento; injustificável
enfadar; esforçar;; fantasma;; sossegador;; opressivamente; stressado
;; desmascarar;
- 51 apaixonar;; ídolo; subjugar; enternecedor;;; intrujice; vilania; despótico
exasperar; espancamento; provocador;; obstinadamente;;; infelicidade
insustentável;;; dispareunia; entontecer;; festeiro;; falecer;
enfraquecimento; indolentemente; pacificadora; libertação;
desconfortavelmente;; reverencioso;; louvor; enxotar; injusto; enfado;
esforço;; fantasmagoria;; sossegar;; opressivo; stressante;;
desmesuradamente;
- 52 apego;; ilustre; tentação; enternecer;;; mágoa; vilão; desprezador; frender
espancar; provocante;; obstinado;;; infeliz; intolerância;;;;
disruptiva; entontecimento;; festejar;; falecimento; enlanguescer;
indolor; pacificar; libertar; desentendimento;; salvaguarda;; luzir;
enxovalhar; insaciabilidade; enfastiado; fiável;; fantasmagórico;;
sossego;; opressor; stressar;; despedida;
- 53 beijar;; lautamente; tentador; enternecimento;;; magoado; vileza;
desprezar; frustar; espertalhão; provocantemente;; obstinar;;;;
infelizmente; intolerante;;; disruptivo; esquisitamente;; festejo;;
funestamente; enlascuecido; inércia; reconciliação; libertino; desordem
;; salvaguardar;; popular; enxovalho; insaciado; enfastiante; fúria;;
fobia;; tranquilamente;; oprimente; suspeita;; despedimento;
- 54 beijinho;; lauto; tentar; entreajuda;;; magoar;; desprezível; fulo;
esperteza; provocar;; onipotência;;; infortunadamente; intolerável
;;; dissociação; esquisitice;; festim;; funesto; esbodegado; lerdo;
reconciliar; proibição; distração;; zelador;; popularidade; escrava;
insaciável; enfastiar; insolência;; frouxidão;; tranquilidade;;
oprimido; suspeitar;; despedir;
- 55 beijo;; lisonja;; entreajudar-se;;; mal-agradecido;; desprezivelmente;
grosseiro; extinção; raiva;; onipotente;;; infortunado;

- intoleravelmente;;; dissociativa; esquisito;;; festival;;; inconsolação
; esbodegar; letargia; reconfortante; puramente; distraído;;; zelar;;;
popularizar; escravatura; insatisfação; estrago; insolente;;; frouxo;;;
tranquilização;;; oprimir; suspeito;;; deus;
- 56 beijocar;;; lisonjeador;;; escrupuloso;;; mal-amada;;; desprezo; grosseria;
exterminar; raivosamente;;; opulência;;; infortúnio; intransigência
;;; dissônia; estonteante;;; festivaleiro;;; inconsolado; esfalfamento;
macilento; reconfortar; pureza; distrair;;; popularmente; escravidão;
insatisfeito; feder; intrepidamente;;; horrendo;;; tranquilizado;;;
orprimido; suspeitoso;;; devoção;
- 57 beijoqueiro;;; lisonjear;;; favorável;;; mal-amado;;; desrespeitar;
grotescamente; extermínio; raivoso;;; opulento;;; insucesso;
intransigente;;; distimica; estontear;;; festividade;;; inconsolável;
esfalfar; madraço; reconforto; puro; dúvida;;; positivamente; escraviza
ção; lazeira; fedor; intrepidez;;; horripilante;;; tranquilizador;;; pânico
; suspense;;; difícil;
- 58 camarada;;; lisonjeiramente;;; favorecer;;; mal-encarado;;; desrespeito;
grotesco; extinção; rixa;;; orgulhar;;; maldição; lastimar;;; distímico
; estranho;;; festivo;;; isoladamente; esgotado; malandrice; remedeio;
rectidão; duvidar;;; positivo; escravizar; mal acabado; fedorento; intr
épido;;; horripilar;;; tranquilizante;;; sobressaltado; suspenso;;; diminui
ção;
- 59 camaradagem;;; lisonjeiro;;; formidável;;; mentir;;; desvalorização;
grunhido; extinguir; violação;;; orgulho;;; maldito; luto;;; distorção;
excentricidade;;; fogosamente;;; isolado; esgotamento; malandro; remediar
; redenção; duvidoso;;; regalado; escravo; malogrado; feio; intrometer;;;
horrível;;; tranquilo;;; sobressaltar; tensa;;; diminuir;
- 60 carente;;; louvar;;; formidavelmente;;; mentira;;; desvalorizar; grunhir;
extinto; violar;;; orgulhosa;;; malfadado; maltratar;;; distorcer; excên
trico;;; fogoso;;; isolamento; esgotante; maltrapilho; reparar; redimir;
enovelamento;;; regalar; estigma; malograr; frivolamente; intrometido
;;; horriavelmente;;; sobressalto; tensão;;; divindade;
- 61 carinho;;; luxo;;; galante;;; mentiroso;;; difamar ;impertinência; facínora;
violência;;; orgulhosamente;;; melodrama; mau trato;;; distúrbio;
extravagância;;; folião;;; isolar; esgotar; mandrião; serenar;
sinceramente; enovelar;;; regalo; estigmatismo; malogro; frivolidade;
intrometimento;;; horror;;; sufocação; tensão nervosa;;; divinização;
- 62 carinhoso;;; luxuosamente;;; galanteador;;; mesquinhez;;; egoísmo ;
impertinente; fera; violenta;;; orgulhoso;;; provação; padecente;;;
doentio; extravagante;;; galdério;;; lancinante; estafa; mandriar;
socorrer; sinceridade; ensarilhar;;; satisfação; estigmatizado;
mendigar; frívolo; intromissão;;; horrorizar;;; sufocamento; tenso;
divinizar;
- 63 companheiro;;; luxuoso;;; galantear;;; mesquinheza;;; egoísta ; implicação;
ferida; violentamente;;; ostensivamente;;; tragédia; padecer;;;
encoprese; fantasia;;; genica;;; letal; estafado; mandriice; socorrismo;
sincero; escandalizar;;; satisfatoriamente; estigmatizar; mendigo; fú
til; intruso;;; horroroso;;; sufocante;;; divino;

- 64 conveniência;;luxúria;;galanteio;;;mesquinho;;egolatria;implicar;
ferido;violento;;ostensivo;;;tragicamente;padecimento;;;enfermo;
fantasioso;;impetuosamente;;lipemania;estafante;marasmo;socorrista
;virtude;escândalo;;;satisfatório;fiasco;miserabilismo;futilidade
;lampeiro;;infernai;;;sufocar;;eliminação;
- 65 conviva;;luxuriante;;generosidade;;;odiar;;egotismo;importunar;
ferimento;vituperação;;ostentação;;;trágico;penar;;;enurese;
histerismo;;impetuosidade;;lôbrego;estafar;moribundo;socorro;
virtuosamente;escandaloso;;;satisfazer;gozar;miserabilista;
imperdoável;metediço;;inferno;;;sufoco;;;eliminar;
- 66 conviver;;magnanimamente;;generoso;;;odiento;;elitismo;importuno;
ferir;vituperador;;ostentar;;;trama;penoso;;;erotomania;idiota;;
impetuoso;;lúgubre;estoirado;negligência;suporte;virtuosidade;
espalhafato;;;satisfeita;gozo;miseramente;implicância;ousadamente
;;insegurança;;;transe;;emergência;
- 67 coração;;magnata;;gentil;;;ódio;;elitista;impropriamente;ferocidade;
vituperar;;peneirento;;;tramar;queixa;;;erotomaniaco;idiotia;;
maganão;;melancolia;exaustação;negligenciar;sustentar;virtuosismo;
espalhafatoso;;;satisfeito;humilhação;miséria;implicante;ousadia
;;insegurança;;;emoção;
- 68 enamorada;;magnificamente;;gentileza;;;odiosamente;;embuste;impró
prio;feroz;vituperável;;pertinácia;;;queixoso;;;erotomano;
idiotice;; másculo;;melancolicamente;exaustão;negligente;
tranquilizar;virtuoso;exaltadamente;;;suficiente;humilhado;misé
rias;implicativo;ousado;;inseguro;;;emocionada;
- 69 enamoradamente;;magnificência;;gentilmente;;;odioso;;enganador;
inconveniência;ferozmente;zangão;;pertinaz;;;queixume;;;
esquizoafectiva;idiotismo;;patusco;;melancólico;exaustar;
negligentemente;;freima;;;usufruir;humilhante;necessidade;
impopular;ousar;;intimidador;;;emocionado;
- 70 enamorado;;magnífico;;humanitário;;;preconceito;;enganar;
inconveniente;fingido;;pertinência;;;sofredor;;;esquizofrenia;
ignaro;;reanimada;;monotonamente;exaustivo;paralisado;;frenesi
;;;usufruto;humilhar;obstáculo;impopularidade;poder;;intimidante
;;;emocional;
- 71 enamorar;;maravilha;;humanitarismo;;;racismo;;engano;inculto;
fingimento;;;pertinente;;;sofrer;;;esquizofrênico;imaginação;;
reanimado;;monotonia;exausto;pasmaceira;;;frenesim;;;ignomínia;
pelintra;importunação;poderes;;intimidar;;;emocionalmente;
- 72 família;;maravilhada;;impecabilidade;;;racista;;enganoso;indecência
;fingir;;;possessão;;;sofrimento;;;esquizofreniforme;
imaginativo;;reanimar;;monótono;extenuação;passivamente;;
freneticamente;;;ignominiosamente;pelintrice;importunidade;
poderio;;intimidativo;;;emocionante;
- 73 familiar;;maravilhado;;impecável;;;rancor;;engazupar;indecente;
frechar;;;possessivo;;;sofrível;;;exibicionismo;imbecil;;
renascer;;morrer;extenuado;passividade;;frenético;;;ignominioso
;penúria;incomodado;poderosamente;;macabro;;;emocionar;

- 74 familiarmente;;maravilhar;;impecavelmente;;;rancorosamente;;engodo;
indecorosamente;frialdade;;;possesso;;;sofrivelmente;;;factícia
;imbecilidade;;renascido;;morte;extenuante;passivo;;;guerra;;;;
impropêrio;pobre;incomodar;poderoso;;medo;;;;emotividade;
- 75 fidelidade;;maravilhosa;;ingenuidade;;;rancoroso;;enjeitado;
indecoroso;friamente;;;possessor;;;súplica;;;festorvo;
imbecilizar;;renascimento;;mortífero;extenuar;percluso;;;guerrear
;;;indignar;pobremente;incomodativo;robustecer;;medonho;;;;
emotivo;
- 76 fiel;;maravilhosamente;;ingênuo;;;ressentida;;enjeitar;inoportuno;
frieza;;;prepotência;;;suplicante;;;fetichismo;imprudência;;
revigorar;;morto;fadiga;preguiça;;;hesitação;;;indignidade;
pobreza;incomodidade;robustez;;medricas;;;empalidecer;
- 77 fielmente;;maravilhoso;;lamechas;;;ressentido;;enojar;insistente;
frigidez;;;prepotente;;;suplicar;;;frotteurismo;imprudente;;
revitalização;;nostalgia;fadigar;preguiçar;;;hesitante;;;indigno
;pocilga;incômodo;robustez;;medrosamente;;;encarceração;
- 78 fraternal;;milagre;;magnanimemente;;;ressentimento;;enojo;
irritabilidade;frígido;;;presunção;;;supliciado;;;
hiperactividade;imprudentemente;;revitalizar;;nostálgico;fatigante
;preguiçosamente;;;hesitar;;;inestimável;podre;injustiçada;
robusto;;medroso;;;encarceramento;
- 79 fraternalmente;;milagrosamente;;magnanimidade;;;ressentir;;
ensoberbecer;irritação;frigir;;;presunçosamente;;;supliciar;;;
hiperactivo;impudência;;reviver;;pena;fatigar;preguiçoso;;
hiperemotividade;;;infama;podres;injustiçado;segura;;obscuridade
;;;encarcerar;
- 80 fraternidade;;milagroso;;magnânimo;;;rival;;escarnecer;irritamente;
frio;;;presunçoso;;;suplício;;;hipersónia;impudentemente;;
revivida;;pessimismo;fracalhão;quietamento;;;hiperemotivo;;;
infamante;podridão;insuficiente;segurança;;obscuro;;;;
encarecidamente;
- 81 fraternização;;motivação;;misericórdia;;;rivalidade;;escárnio;
irritante;gélido;;;pretensão;;;tenebroso;;;hipocondria;
inebriado;;revivido;;pessimista;fraco;quieto;;;hipermotiva;;;
infamar;privado;maçada;seguro;;papão;;;encarecimento;
- 82 fraterno;;motivar;;misericordiosamente;;;rivalizar;;escória;
irritantemente;hediondo;;;pretensioso;;;tormento;;;hipocondrí
aco;inebriante;;revivificação;;pungente;frágil;quietude;;;hipersen
ível;;;infame;privar;macambúzio;temerário;pavor;;;;
enclausurar;
- 83 gostar;;nobre;;misericordioso;;;traição;;esgar;irritar;imoral;;
relutância;;;torturador;;;hipomaníaco;inebriar;;revivificar;;
pungir;fragilidade;sono;;;hipersensibilidade;;;infâmia;proibição
;maçar;temeridade;;pavorosamente;;;enfestado;
- 84 gratidão;;nobrememente;;perfeccionismo;;;traíçoeiramente;;esquecer;
irritável;imoralidade;;relutante;;;trauma;;;histeria;inépcia;;
sadio;;solidão;fraquejar;sonolência;;;imponderado;;;injúria;
proibido;malcheiroso;tenacidade;;pavoroso;;;enfazar;

- 85 grato;;nobreza;;perfeccionista;;;traíçoeiro;;esquecido;labrego;
imperfecção;;;relutar;;;traumática;;;histérico;inepto;;saudável
;;solitária;fraqueza;sonolento;;;indecisão;;;injuriar;proibir;
materialismo;tenaz;;pesadelo;;;enforcado;
- 86 maternal;;predilecção;;perfeição;;;traidor;;esquecimento;ladrão;
imperfeito;;soberanamente;;;traumático;;;indiderenciado;insâ
nia;;saúde;;solitariamente;impelir;tetania;;;indeciso;;;;
injurioso;proibitivamente;materialista;topete;;pesadelos;;;;
enforcamento;
- 87 maternalmente;;predilecto;;perfeitamente;;;trair;;excomungar;
ladroagem;impiamente;;;soberania;;;traumatismo;;;insônia;insano
;;vigor;;solitário;impotência;tetanização;;;indisciplina;;;;
insultar;proibitivo;negação;valente;;poltrão;;;enforçar;
- 88 materno;;prestigiante;;perfeito;;;vigarice;;excomunhão;ladroeira;
impiedade;;;soberano;;;traumatizante;;;intoxicação;insensata;;
vigorosamente;;sombrio;impotente;tetanizar;;;indisciplinar;;;;
insulto;recalcado;negar;valentemente;;poltronaria;;;enfrascado
;
- 89 namorado;;prestigiar;;piegas;;;vigário;;expulsão;lamecha;
impiedosamente;;soberba;;;traumatizar;;;lunático;insensatez;;
vigoroso;;sozinho;inexpressividade;tórpido;;insubordinação;;;;
insultuosamente;recalcamento;negativa;valentia;;recear;;;;
engendração;
- 90 namorados;;prestígio;;pieguice;;;vigarista;;expulsar;lapuz;impiedoso
;;soberbamente;;;mania;insensato;;vitalidade;;
taciturnamente;inexpressivo;torpor;;insubordinado;;;insultuoso;
recalcar;negativamente;;receio;;;engendrar;
- 91 namorar;;prestigioso;;porreiro;;;vigarizar;;expulso;larapiar;ímpio
;;soberbo;;;maníaco;instabilidade;;viva;;taciturno;langor
;;;insubordinar;;;irrespeito;renúncia;negativismo;;
receosamente;;;engrandecer;
- 92 namoriscar;;privilegiado;;prestável;;;vingador;;fustigação;larápio;
implacabilidade;;superior;;;manicômio;instável;;vivacidade
;;tédio;languidez;;;intriga;;;irrespeitosamente;renunciar;
negativo;;receoso;;;engrandecimento;
- 93 namorisco;;privilegiar;;provocar dó;;;vingança;;fustigar;má cara;
implacável;;superioridade;;;masochismo;irreflectidamente;;
vivamente;;triste;lânguido;;;intrigar;;;irrespeitoso;represa;
porco;;sinistramente;;;engravidar;
- 94 namoro;;privilégio;;pueril;;;vingar;;gentalha;maçador;
implacavelmente;;superiorizar;;;masoquismo;irreflectido;;
vivaz;;tristemente;lassidão;;;intriguista;;;labêu;represar;
promiscuidade;;sinistro;;;engravidar;
- 95 noiva;;;puerilidade;;;vingativo;;gentinha;mal-educado;incivilidade
;;superiormente;;;masoquista;irreflexão;;viveza;;tristeza;
lasso;;;mal-estar;;;lacaio;repressão;promíscuo;;susto;;;enorme;
- 96 noivo;;;sacrificar;;;xenofobia;;ignóbil;malcriado;incivilmente;;
supremo;;;megalomania;irresponsabilidade;;vivo;;tristonho;mo

	ído;;;perdida;;;;mácula;reprimido;rançoso;;;temente;;;;
	enriquecer;
97	paixão;;;sacrifício;;;;xenófobo;;ignobilmente;mau humor;inclemência ;;;teima;;;;;;megalómano;irresponsável;;zelo;;vazio;moleza;;;; perdidamente;;;;mazela;reprimir;tendenciosamente;;;temer;;;;
	enriquecimento;
98	paixoneta;;;;sem-malícia;;;;;ignorado;melga;inclemente;;;teimar ;;;;melomania;irresponsavelmente;;zeloso;;;morosamente;;;; perdido;;;;melindrar;vadiagem;tendencioso;;;temeroso;;;;
	entravamento;
99	passional;;;sociável ;;;;;ignorância;nervos;indelicadeza;;;; teimosamente;;;;melómano;jerico;;;;morosidade;;;; periclitante;;;;melindre;vadiar;;;temível;;;;entravar;
100	piedade;;;;ignorante;nervosa;inexorável;;teimosia;;;; mericismo;levianamente;;;;moroso;;;periclitlar;;;; melindrosamente;vadio;;;temivelmente;;;;entreve;
101	piedosamente;;;;ignorar;nervosamente;lesão;;teimoso;;;; mitomania;leviandade;;;;neurastenia;;;perplexidade;;;; melindroso;vagabundagem;;;temor;;;;entrelaçamento;
102	piedoso;;;;ignoto;nervosidade;lesar;;;tirano;;;;mitomani aco;leviano;;;;neurastênico;;;perplexo;;;;mortificação; vagabundear;;;ter pesadelos;;;;entrelaçar;
103	platonicamente;;;;imperial;nervosismo;lesivo;;vaidade;;;; mitómano;louco;;;;prostração;;;perturbação;;;;mortificante; vagabundo;;;terrificante;;;;envolvimento;
104	platonismo;;;;imperialmente;nervoso;má;;vaidosamente;;;; morbidade;loucura;;;;prostrado;;;perturbado;;;;mortificar;;;; terrificar;;;;erradicação;
105	preferência;;;;imperiosamente;oponente;mal;;vaidoso;;;; morbidez;maluca;;;;psicastenia;;;perturbador;;;;ofender;;;; terrífico;;;;erradicar;
106	preferido ;;;;;imperioso;opor;maldade;;;;morbidez; maluco;;;;psicastênico;;;perturbadora;;;;ofendido;;;;terrível ;;;escrita;
107	preferir;;;;imprestável;oportunismo;maldosa;;;;mórbido; maluquice;;;;vulnerabilidade;;;perturbar;;;;ofensa;;;; terrivelmente;;;;escurecer;
108	querido ;;;;;inaceitável;oportuno;maldosamente;;;; mutismo;marciano;;;;vulnerante;;;problema;;;;ofensivo;;;; terror;;;;escuridão;
109	romântico;;;;inadmissível;oposicionista;maldoso;;;; narcisismo;marginal;;;;vulnerável;;;supra-sensível;;;;ofensor ;;;escuro;
110	romantismo;;;;inconsideração;ordinariamente;maleficência ;;;narcisista;marginalidade;;;;tontura;;;;palerma ;;;esmerar;
111	sensibilidade;;;;indiferença;ordinário;malefício;;;; narciso;mulherengo;;;;tormenta;;;;palermice;;;; espasmo;

112	sensibilização;;;;;indiferente;praguejar;maléfico;;;;;; narcolepsia;obcecação;;;;;transtornado;;;;;perder;;;;;; espiritismo;
113	sensibilizada;;;;;interesseiro;rabugento;malevolamente ;;;;;neurose;obcecadamente;;;;;transtornar;;;;;perdição ;;;;;espiritista;
114	sensibilizado;;;;;inútil;resmungão;malevolência;;;;;neuró tico;obcecar;;;;;transtorno;;;;;prostrar;;;;;espírito;
115	sensibilizar;;;;;inutilidade;resmungar;malevolente;;;;;; obsessão;parvamente;;;;;tumulto;;;;;rebaixamento;;;;;; espiritual;
116	sensível;;;;;ironia;resmunguice;malévolo;;;;;obsessivo; parvo;;;;;vaguear;;;;;rebaixar;;;;;espiritualidade;
117	sentimental;;;;;ironicamente;rezingão;malfeitor;;;;;; parafilia;parvoíce;;;;;submeter;;;;;espiritualismo ;
118	sentimentalismo;;;;;irônico;rezingar;malícia;;;;;paranóia ;paspalhão;;;;;submetido;;;;;espiritualista;
119	sentimentalmente;;;;;irrepreensível;rosnador;maliciosamente ;;;;;paranóico;paspalho;;;;;submissão;;;;;; espiritualizar;
120	sentimentos;;;;;irrepreensivelmente;rosnar;malicioso;;;;;; paranóide;patego;;;;;submisso;;;;;espiritualmente;
121	simpatia;;;;;irreverência;rosnido;malignamente;;;;;parass ônia;pateta;;;;;subordinação;;;;;espirituosamente;
122	simpática;;;;;irreverenciosamente;rude;maligno;;;;;patoló gico;pateticamente;;;;;subordinado;;;;;estado;
123	simpático;;;;;irreverente;rudemente;malquerença;;;;;; pedofilia;patetice;;;;;subordinar;;;;;estagnação;
124	simpatizante;;;;;irreverentemente;rudez;malvadamente;;;;;; perversão;patifaria;;;;;ultraçado;;;;;estagnar;
125	simpatizar;;;;;irrisão;rudeza;malvadez;;;;;perversidade; patife;;;;;ultrajante;;;;;estremecer;
126	união;;;;;irrisório;rugido;malvadeza;;;;;perverso;problem ática;;;;;ultrajar;;;;;estremecido;
127	unir;;;;;judiação;rugir;malvado;;;;;pica;problemático ;;;;;ultraje;;;;;estremecimento;
128	veneração;;;;;judiaria;ruminante;manha;;;;;piromania; problematizar;;;;;vergonha;;;;;estupidamente;
129	venerado;;;;;lastimável;ruminar;manhoso;;;;;psicadélico; suicidar-se;;;;;vergonhosamente;;;;;estupidez;
130	venerador;;;;;lastimavelmente;sensitivo;manipulação;;;;;; psicopata;suicídio;;;;;vergonhoso;;;;;estupidificar;
131	venerar;;;;;lastimoso;sensível;manipular;;;;;psicopatia; tara;;;;;vexado;;;;;estúpido;
132	veracidade;;;;;lograr;sensivelmente;maquiavélico;;;;;; psicopático;tarado;;;;;vexame;;;;;eutanásia;

verdade;,,,,,;logro;susceptibilidade;maquiavelismo;,,,,,;;
 psicose;tolice;,,,,,;vexante;,,,,,;evangelho;
 xenofilia;,,,,,;ludibriar;susceptível;maroteira;,,,,,;;
 psicossomático;tolo;,,,,,;vexar;,,,,,;exceder;
 xenófilo;,,,,,;ludíbrio;troglodita;maroto;,,,,,;psicótico;
 tontice;,,,,,;vitupério;,,,,,;excessivamente;
 ,,,,,,;mandão;zumbido;massacrar;,,,,,;psíquico;tonto
 ,,,,,,;excessivo;
 ,,,,,,;mandar;massacre;,,,,,;psiquismo
 ,,,,,,;excesso;
 ,,,,,,;menosprezar;matador;,,,,,;residual
 ,,,,,,;excitabilidade;
 ,,,,,,;menosprezo;matança;,,,,,;sádico
 ,,,,,,;excitação;
 ,,,,,,;miserável;matar;,,,,,;sadismo
 ,,,,,,;excitado;
 ,,,,,,;miseravelmente;matreirice;,,,,,;somasoquismo
 ,,,,,,;excitante;
 ,,,,,,;mísero;matreiro;,,,,,;somasoquista
 ,,,,,,;excitar;
 ,,,,,,;mostrengo;mau;,,,,,;somático
 ,,,,,,;excitável;
 ,,,,,,;ostracismo;megera;,,,,,;somatização
 ,,,,,,;expectativa;
 ,,,,,,;pejorativamente;molesto;,,,,,;somatoforme
 ,,,,,,;expressão;
 ,,,,,,;pejorativo;monstro;,,,,,;sonambulismo
 ,,,,,,;expressividade;
 ,,,,,,;petulância;monstruosamente;,,,,,;sonâmbulo
 ,,,,,,;expressivo;
 ,,,,,,;petulante;monstruosidade;,,,,,;suicída
 ,,,,,,;fácil;
 ,,,,,,;pindérico;monstruoso;,,,,,;suicidar
 ,,,,,,;facilidade;
 ,,,,,,;pulha;nazi;,,,,,;tourette;,,,,,;fado;
 ,,,,,,;pulhice;nazismo;,,,,,;tricotilomania
 ,,,,,,;fanático;
 ,,,,,,;punição;nefasto;,,,,,;vaginismo
 ,,,,,,;fanatismo;
 ,,,,,,;punir;nocivamente;,,,,,;viciante
 ,,,,,,;fanatizar;
 ,,,,,,;punitivo;nocividade;,,,,,;vinculação
 ,,,,,,;fartamente;
 ,,,,,,;punível;nocivo;,,,,,;voyeurismo
 ,,,,,,;fartar;
 ,,,,,,;recriminação;perversamente
 ,,,,,,;farto;

181 ;;;;;;;;;;snob;;severidade;;;;;;;;;;;;;;
 imortalidade;
 182 ;;;;;;;;;;snobe;;severo;;;;;;;;;;;;;;
 imortalização;
 183 ;;;;;;;;;;snobismo;;terrorismo;;;;;;;;;;;;;;
 imortalizar;
 184 ;;;;;;;;;;trapaça;;terrorista;;;;;;;;;;;;;;
 importante;
 185 ;;;;;;;;;;trapacear;;tortuosamente
 ;;;;;;;;;;;;;impressão;
 186 ;;;;;;;;;;trapaceiro;;tortuosidade
 ;;;;;;;;;;;;;impressionante;
 187 ;;;;;;;;;;troça;;tortuoso;;;;;;;;;;;;;;
 impressionar;
 188 ;;;;;;;;;;troçar;;tortura;;;;;;;;;;;;;;
 impressionável;
 189 ;;;;;;;;;;vaiar;;torturante;;;;;;;;;;;;;;
 impulsão;
 190 ;;;;;;;;;;vão;;torturar;;;;;;;;;;;;;;
 impulsivamente;
 191 ;;;;;;;;;;vilipendiar;;usurpação
 ;;;;;;;;;;;;;impulsividade;
 192 ;;;;;;;;;;vilipêndio;;usurpador
 ;;;;;;;;;;;;;impulsivo;
 193 ;;;;;;;;;;zombar;;usurpar;;;;;;;;;;;;;;
 inferir;
 194 ;;;;;;;;;;veneno;;;;;;;;;;;;;;inspiração;
 195 ;;;;;;;;;;venenoso;;;;;;;;;;;;;;inspirado;
 196 ;;;;;;;;;;víbora;;;;;;;;;;;;;;inspirar;
 197 ;;;;;;;;;;violador;;;;;;;;;;;;;;inteligê
 ncia;
 198 ;;;;;;;;;;violentar;;;;;;;;;;;;;;
 inteligente;
 199 ;;;;;;;;;;judas;
 200 ;;;;;;;;;;machismo;
 201 ;;;;;;;;;;machista;
 202 ;;;;;;;;;;máfia;
 203 ;;;;;;;;;;mafioso;
 204 ;;;;;;;;;;maior;
 205 ;;;;;;;;;;maioral;
 206 ;;;;;;;;;;maioria;
 207 ;;;;;;;;;;maioritário;
 208 ;;;;;;;;;;mais;
 209 ;;;;;;;;;;mal;
 210 ;;;;;;;;;;masculinidade;
 211 ;;;;;;;;;;masculino;
 212 ;;;;;;;;;;máxima;
 213 ;;;;;;;;;;maximamente;
 214 ;;;;;;;;;;maximizar;

215	;;;;máximo;
216	;;;;médio;
217	;;;;mediocre;
218	;;;;mediocrementemente;
219	;;;;mediocridade;
220	;;;;melhor;
221	;;;;melhoramento;
222	;;;;melhorar;
223	;;;;menor;
224	;;;;menoridade;
225	;;;;menos;
226	;;;;mercê;
227	;;;;minoria;
228	;;;;moralidade;
229	;;;;moralista;
230	;;;;moralização;
231	;;;;moralizante;
232	;;;;moralizar;
233	;;;;moralmente;
234	;;;;muita;
235	;;;;muitas;
236	;;;;muito;
237	;;;;muitos;
238	;;;;mundanidade;
239	;;;;mundanismo;
240	;;;;mundano;
241	;;;;natural;
242	;;;;naturalidade;
243	;;;;naturalismo;
244	;;;;naturalista;
245	;;;;naturalmente;
246	;;;;natureza;
247	;;;;ninguém;
248	;;;;notoriamente;
249	;;;;notoriedade;
250	;;;;notório;
251	;;;;obstrução;
252	;;;;obstrucionismo;
253	;;;;obstruir;
254	;;;;ofegante;
255	;;;;ofegar;
256	;;;;orgulho;
257	;;;;orgulhosamente;
258	;;;;orgulhoso;
259	;;;;palidamente;
260	;;;;palidez;
261	;;;;pálido;
262	;;;;palpitação;
263	;;;;palpitante;

264	palpitar;
265	parcial;
266	parcialidade;
267	parcialmente;
268	pecado;
269	pecador;
270	pecaminosamente;
271	pecaminoso;
272	pecar;
273	pensador;
274	pensamento;
275	pensante;
276	pensar;
277	pensativo;
278	pequenez;
279	pequeneza;
280	pequeno;
281	personalidade;
282	persuadir;
283	persuasão;
284	persuasivo;
285	pesado;
286	pior;
287	piorar;
288	pouco;
289	prevenir;
290	preventivamente;
291	preventivo;
292	pulsão;
293	reação;
294	reagir;
295	realista;
296	reflectir;
297	reflexão;
298	saudade;
299	sensível;
300	sentimental;
301	sentimentalidade;
302	sentimento;
303	sentir;
304	séria;
305	sério;
306	simples;
307	simplesmente;
308	suar;
309	sudação;
310	suor;
311	superexcitabilidade;

```

312 ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;superexcitação;
313 ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;superexcitado;
314 ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;superexcitante;
315 ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;superexcitar;
316 ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;superexcitável;
317 ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;temperamental;
318 ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;temperamento;
319 ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;tempestivamente;
320 ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;tempestivo;
321 ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;tiritação;
322 ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;tiritante;
323 ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;tiritar;
324 ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;transpiração;
325 ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;transpirado;
326 ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;transpirar;
327 ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;tremelicar;
328 ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;trememente;
329 ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;tremor;
330 ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;tremor;
331 ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;trémulo;
332 ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;trepidação;
333 ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;trepidante;
334 ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;trepidar;
335 ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;uivador;
336 ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;uivar;
337 ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;urro;
338 ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;veemência;
339 ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;veemente;
340 ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;veementemente;
341 ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;vermelhidão;
342 ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;vermelho;
343 ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;vibração;
344 ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;vibrante;
345 ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;vibrar;
346 ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;

```

Code 7: CSV of the EMOTAIX.PT lexicon split into 3, 9, 18, and 50 classes.

A.14 PositiveWords.csv

```

1  absolution,absolutions,absolve,absolved,absolver,absolvers,absolves,
   absolving,absolvitory,advance,advanced,advancement,advancements,
   advancer,advancers,advances,advancing,advantage,advantaged,
   advantageous,advantageously,advantageousness,advantageousnesses,
   advantages,advantaging,affirm,affirmable,affirmation,affirmations,
   affirmative,affirmatively,affirmativeness,affirmativenesses,
   affirmatives,affirmatory,affirmed,affirmer,affirmers,affirming,
   affirms,agile,agilely,agilities,agility,ahead,amaze,amazed,
   amazement,amazements,amazes,amazing,amazingly,approbate,approbated

```

, approbates , approbating , approbation , approbations , approbative ,
 approbatory , approval , approvals , approve , approved , approver , approvers
 , approves , approving , arrival , arrivals , arrive , arrived , arriver ,
 arrivers , arrives , arriving , attend , attendance , attendances , attendant ,
 attendants , attended , attender , attenders , attending , attends , attention
 , attentions , attentive , attentively , attentiveness , attentivenesses ,
 auspicious , auspiciously , auspiciousness , auspiciousnesses , award ,
 awarded , awarding , awards , beautiful , beautifully , best , bested , besting ,
 bests , bet , bets , betted , better , bettered , bettering , betterment ,
 betterments , betters , betting , bettor , bettors , big , bigness , bignesses ,
 calm , calmed , calming , calmly , calmness , calmnesses , calms , celebrant ,
 celebrants , celebrate , celebrated , celebrates , celebrating , celebration
 , celebrations , celebrator , celebrators , celebratory , center , centering ,
 centerings , centers , central , centralities , centrality , centrally ,
 centrals , centric , central , commend , commendation , commendations ,
 commended , commending , commends , commodities , commodity , compensate ,
 compensated , compensates , compensating , compensation , compensations ,
 complete , completed , completely , completeness , completenesses ,
 completes , completing , completion , completions , compliment ,
 complimentary , complimented , complimenting , compliments , component ,
 components , compose , composed , composer , composers , composes , composing ,
 composings , composition , compositions , compositor , compositors ,
 composure , composes , congratulate , congratulated , congratulates ,
 congratulating , congratulation , congratulations , congratulatory ,
 convict , convicted , convicting , conviction , convictions , convicts ,
 correct , corrected , correcting , correction , correctional , corrections ,
 corrective , correctives , correctness , correctnesses , corrects ,
 corrigible , dazzle , dazzled , dazzles , dazzling , dear , dearly , dearness ,
 dearnesses , dears , decencies , decency , decent , decently , delicious ,
 deliciously , divination , divinations , divinatory , divine , divined ,
 divinely , diviner , diviners , divines , divining , divinities , divinity ,
 easiness , easinesses , easy , edibilities , edibility , edible , edibleness ,
 ediblenesses , edibles , efficiencies , efficiency , efficient , efficiently
 , elevate , elevated , elevates , elevating , elevation , elevations , elevator
 , elevators , empathetic , empathetically , empathies , empathise , empathize
 , empathized , empathizes , empathizing , empathy , endorse , endorsement ,
 endorsements , endorser , endorsers , enjoin , enjoined , enjoining ,
 enjoinings , enjoinment , enjoinments , enjoins , enlighten , enlightened ,
 enlightening , enlightenment , enlightenments , enlightens , entrust ,
 esteem , esteemed , esteeming , esteems , excel , excelled , excellence ,
 excellences , excellencies , excellency , excellent , excellently ,
 excelling , excels , except , excepted , excepting , exception , exceptional ,
 exceptionally , exceptions , excepts , facilitate , facilitated ,
 facilitates , facilitating , facilitation , facilitations , facilitative ,
 facilitator , facilitators , facilitatory , familiar , familiarities ,
 familiarity , familiarly , familiars , fantasies , fantasist , fantasists ,
 fantasize , fantasized , fantasizes , fantasizing , fantastic , fantastical ,
 fantasy , fast , fasted , fasting , fastings , fastness , fastnesses , fasts ,
 favorite , favorites , fortunate , fortunately , fortune , fortunes , free ,

freed, freeing, freeings, freely, frees, friend, friendlies, friendliness
 , friendlinesses, friendly, friends, friendship, friendships, gain,
 gained, gainer, gainers, gaining, gains, gift, gifted, gifting, gifts,
 golden, good, goodness, goodnesses, goods, great, greatly, greatness,
 greatnesses, greats, grew, grow, grower, growers, growing, growings, grown
 , grows, growth, growths, had, hadn't, harmonic, harmonical, harmonically,
 harmonics, harmonies, harmonious, harmoniously, harmoniousness,
 harmoniousnesses, harmonisation, harmonisations, harmonise, harmoniser
 , harmonisers, harmonization, harmonizations, harmonize, harmonized,
 harmonizer, harmonizers, harmonizes, harmonizing, harmony, has, hasn't,
 have, haven't, haves, having, holies, holiness, holinesses, holy, homemade
 , honest, honestly, honestness, honestnesses, honor, honorable,
 honorableness, honorablenesses, honorably, honorary, honored, honoree,
 honorees, honoring, honorings, honors, idea, ideal, idealisation,
 idealisations, idealise, idealism, idealisms, idealist, idealistic,
 idealists, idealities, ideality, idealization, idealizations, idealize,
 idealized, idealizes, idealizing, ideally, ideals, ideas, ideate, ideated
 , ideates, ideating, ideation, ideational, ideations, impeccabilities,
 impeccability, impeccable, impeccably, injunction, injunctions, intact,
 intactness, intactnesses, integer, integers, integral, integralities,
 integrality, integrally, integrals, integrate, integrated, integrates,
 integrating, integratings, integration, integrations, integrative,
 integrator, integrators, intellect, intellects, intellectual,
 intellectually, intellectuals, intelligence, intelligences,
 intelligent, intelligently, intimate, intimated, intimately, intimates,
 intimating, intimation, intimations, just, justly, justness, justnesses,
 keep, keeper, keepers, keeping, keepings, keeps, kept, large, largeness,
 largenesses, larges, laugh, laughable, laughably, laughed, laughter,
 laughers, laughing, laughs, leave, leaved, leaver, leavers, leaves,
 leaving, left, lefts, leisure, leisureliness, leisurelinesses, leisurely
 , leisures, long, longed, longer, longers, longest, longing, longings,
 longness, longnesses, longs, lovable, love, loveable, loved, lover, lovers
 , loves, loving, made, make, maker, makers, makes, making, makings, manage,
 manageabilities, manageability, manageable, manageableness,
 manageablenesses, manageably, managed, management, managements, manager
 , managerial, managerially, managers, managership, managerships, manages
 , managing, mediate, mediated, mediateness, mediatenesses, mediates,
 mediating, mediation, mediations, mediator, mediators, merit, meritable,
 merited, meriting, meritorious, meritoriously, meritoriousness,
 meritoriousnesses, merits, mind, minded, minding, minds, much, muches,
 muchness, muchnesses, new, newness, newnesses, novel, novelette,
 novelettes, novelist, novelists, novels, organization, organizations,
 organize, organized, organizer, organizers, organizes, organizing,
 outstanding, outstandingly, perfect, perfected, perfecter, perfecters,
 perfectibilities, perfectibility, perfectible, perfecting, perfection,
 perfectionism, perfectionisms, perfectionist, perfectionists,
 perfections, perfectly, perfects, pinnacle, pinnacled, pinnacles,
 pinnacling, pleasance, pleasancess, pleasant, pleasantly, pleasantness,
 pleasantnesses, please, pleased, pleaser, pleasers, pleases, pleasing,

pleasingly, pleasings, pleasure, pleasures, positive, positively, positiveness, positivenesses, positives, positivism, positivisms, positivist, positivistic, positivists, positivities, positivity, potencies, potency, potent, potential, potentialities, potentiality, potentially, potentials, potently, powerful, powerfully, powerfulness, powerfulnesses, practical, practicalities, practicality, practically, practice, practiced, practices, practicing, prefer, preference, preferences, preferent, preferential, preferentially, preferment, preferments, preferred, preferring, prefers, profess, professed, professes, professing, professings, profession, professional, professionally, professionals, professions, proficiencies, proficiency, proficient, proficiently, promote, promoted, promoter, promoters, promotes, promoting, promotion, promotions, promotive, proper, properly, properness, propernesses, prosperities, prosperity, prosperous, prosperously, quick, quickly, quickness, quicknesses, quicks, quiet, quieted, quieting, quietly, quietness, quietnesses, quiets, recommend, recommendation, recommendations, recommended, recommending, recommends, refer, referee, refereed, refereeing, refereeings, referees, reference, referenced, references, referencing, referent, referential, referents, referral, referrals, referred, referring, refers, regain, regained, regaining, regainings, regains, rejuvenate, rejuvenated, rejuvenates, rejuvenating, rejuvenation, rejuvenations, relax, relaxant, relaxants, relaxation, relaxations, relaxed, relaxer, relaxers, relaxes, relaxing, remark, remarked, remarking, remarks, remember, remembered, remembering, rememberings, remembers, remembrance, remembrances, repair, repaired, repairer, repairers, repairing, repairs, reparation, reparations, retain, retained, retaining, retains, retention, retentions, retentive, retentively, retentiveness, retentivenesses, retentivities, retentivity, reward, rewarded, rewarding, rewards, rich, riches, richly, richness, richnesses, right, righted, rightful, rightfully, rightfulness, rightfulnesses, righting, rightness, rightnesses, rights, safe, safeness, safenesses, safes, satisfaction, satisfactions, satisfactorily, satisfactoriness, satisfactorinesses, satisfactory, satisfied, satisfier, satisfiers, satisfies, satisfy, satisfying, save, saved, saver, savers, saves, saving, savings, savior, saviors, skillful, skillfully, skillfulness, skillfulnesses, smooth, smoothed, smoother, smoothers, smoothing, smoothly, smoothness, smoothnesses, smooths, strong, strongly, substance, substances, substantial, substantialities, substantiality, substantially, substantialness, substantialnesses, succeed, succeeded, succeeder, succeeders, succeeding, succeeds, succession, successions, successive, successively, successiveness, successivenesses, successor, successors, suggest, suggested, suggester, suggesters, suggestibilities, suggestibility, suggestible, suggesting, suggestion, suggestions, suggestive, suggestively, suggests, super, supers, tall, tallness, tallnesses, talls, taste, tasted, tasteful, tastefully, tastefulness, tastefulnesses, taster, tasters, tastes, tastily, tastiness, tastinesses, tasting, tastings, tasty, thorough, thoroughly, thoroughness, thoroughnesses, tranquil, tranquilities, tranquility, tranquillities, tranquillity, tranquilly, transcend,

```

transcended , transcendence , transcendences , transcendencies ,
transcendency , transcendent , transcending , transcends , unbind ,
unbinding , unbinds , unblock , unbound , unfreeze , unfreezes , unfreezing ,
unfroze , unfrozen , unregretful , vantage , vantages , variabilities ,
variability , variable , variableness , variablenesses , variables ,
variably , variance , variances , variant , variants , variate , variates ,
variation , variations , varied , varies , varieties , variety , vary , varying ,
veracious , veracities , veracity , veridical , veritable , welcome , welcomed ,
welcomer , welcomers , welcomes , welcoming , well , welled , welling ,
wellness , wellnesses , wells , whole , wholeness , wholenesses , wholes ,
wholly , win , winner , winners , winning , winnings , wins , won , wonder ,
wondered , wonderer , wonderers , wondering , wonderment , wonderments ,
wonders , wondrous , wondrously , young , youngness , youngnesses , youngs

```

Code 8: PositiveWords.csv file context. This includes all the the positive words that were used to filter the Negative Word Clouds

A.15 NeutralWords.csv

```

1 1st , 2nd , 3rd , 4th , 5th , 6th , 7th , 8th , 9th , a , abilities , ability , able , ably ,
  about , absolute , absolutely , absoluteness , absolutenesses , absolutes ,
  accept , acceptabilities , acceptability , acceptable , acceptableness ,
  acceptablenesses , acceptance , acceptances , acceptant , acceptance ,
  acceptations , accepted , accepting , acceptive , accepts , accord ,
  accordance , accordances , accordant , accorded , according , accords , actual ,
  actualities , actuality , actualization , actualizations , actualize ,
  actualized , actualizes , actualizing , actually , add , addable , added , adder ,
  adders , addible , adding , addition , additions , additive , additives , adds ,
  advantage , advantaged , advantageous , advantageously , advantageousness ,
  advantageousnesses , advantages , advantaging , advice , advices ,
  advisabilities , advisability , advisable , advise , advised , advisee ,
  advisees , advisement , advisements , adviser , advisers , advises , advising ,
  advisor , advisories , advisors , advisory , after , afterward , ago , alike ,
  alikeness , alikenesses , all , almost , already , always , am , am not , annotate ,
  annotated , annotates , annotating , annotation , annotations , annotator ,
  annotators , another , answer , answerabilities , answerability , answerable ,
  answerableness , answerablenesses , answered , answerer , answerers ,
  answering , answers , anymore , anyway , are , aren't , around , arrival ,
  arrivals , arrive , arrived , arriver , arrivers , arrives , arriving ,
  articular , articular , articulate , articulated , articulately ,
  articulateness , articulatenesses , articulates , articulating ,
  articulation , articulations , articulative , articulator , articulators ,
  articulatory , as , ask , asked , asker , askers , asking , asks , ate , attempt ,
  attempted , attempter , attempters , attempting , attempts , aught , aughts ,
  away , awayness , awaynesses , back , backed , backer , backers , backing ,
  backings , backs , be , became , become , becomes , becoming , been , before , began ,
  begin , beginner , beginners , beginning , beginnings , begins , begun , behind ,
  behinds , being , beings , belief , beliefs , believabilities , believability ,
  believable , believably , believe , believed , believer , believers ,

```

believes ,believing ,bit ,bite ,bited ,biter ,biters ,bites ,biting ,bits ,
 bitten ,bitty ,blend ,blended ,blender ,blenders ,blending ,blends ,bought
 ,brand ,branded ,branding ,brands ,bring ,bringing ,brings ,brought ,buy ,
 buyer ,buyers ,buying ,buyings ,buys ,call ,callable ,called ,caller ,
 callers ,calling ,callings ,calls ,came ,certain ,certainly ,chance ,
 chanced ,chances ,chancing ,chancy ,change ,changed ,changer ,changers ,
 changes ,changing ,choose ,chooses ,choosing ,chose ,chosen ,clarities ,
 clarity ,clear ,clearance ,clearances ,cleared ,clearing ,clearings ,
 clearly ,clearness ,clearnesses ,clears ,close ,closed ,closely ,
 closeness ,closenesses ,closer ,closers ,closes ,closing ,closings ,
 closure ,closured ,closures ,closuring ,come ,comer ,comers ,comes ,coming
 ,comings ,comparative ,comparatively ,comparatives ,compare ,compared ,
 compares ,comparing ,comparings ,comparison ,comparisons ,complete ,
 completed ,completely ,completeness ,completenesses ,completes ,
 completing ,completion ,completions ,confess ,confessed ,confesses ,
 confessing ,confession ,confessions ,confessor ,confessors ,consider ,
 consideration ,considerations ,considered ,considering ,considers ,
 continuance ,continuances ,continuant ,continuants ,continuation ,
 continuations ,continue ,continued ,continues ,continuing ,countries ,
 country ,cut ,cuts ,cutter ,cutters ,cutting ,cuttingly ,cuttings ,dailies
 ,daily ,day ,days ,decide ,decided ,decides ,deciding ,decidings ,decision
 ,decisions ,decisive ,decisively ,decisiveness ,decisivenesses ,
 definite ,definiteness ,definitenesses ,deserve ,deserved ,deserves ,
 deserving ,desire ,desired ,desires ,desiring ,desirous ,despite ,
 despites ,did ,didn't ,differ ,differed ,difference ,differences ,
 different ,differential ,differentially ,differentials ,differentiate ,
 differentiated ,differentiates ,differentiating ,differentiation ,
 differentiations ,differentiator ,differentiators ,differently ,
 differing ,differs ,direct ,directed ,directing ,directive ,
 directiveness ,directivenesses ,directives ,directivities ,directivity
 ,directly ,directness ,directnesses ,director ,directors ,directorship ,
 directorships ,directs ,distinction ,distinctions ,distinguish ,
 distinguished ,distinguishes ,distinguishing ,do ,doer ,doers ,does ,
 doesn't ,doing ,don't ,done ,due ,dues ,easiness ,easinesses ,easy ,eat ,
 eaten ,eater ,eaters ,eating ,eatings ,eats ,eight ,eights ,either ,end ,
 ended ,ending ,endings ,ends ,enter ,entered ,entering ,enterings ,enters ,
 entire ,entirely ,entireness ,entirenesses ,entires ,entrance ,entranced
 ,entrancement ,entrancements ,entrances ,entrancing ,entrant ,entrants ,
 entree ,entrees ,entries ,entry ,especial ,especially ,establish ,
 established ,establishes ,establishing ,establishment ,establishments ,
 even ,evened ,evening ,evenly ,evenness ,evennesses ,evens ,ever ,every ,ex
 ,exchange ,exchanged ,exchanger ,exchangers ,exchanges ,exchanging ,exes
 ,exist ,existed ,existence ,existences ,existent ,existential ,
 existentialism ,existentialisms ,existentialist ,existentialists ,
 existing ,exists ,expect ,expectancies ,expectancy ,expectant ,
 expectantly ,expectation ,expectations ,expected ,expecting ,expects ,
 explain ,explained ,explaining ,explains ,explanation ,explanations ,
 explanatory ,extreme ,extremely ,extremes ,extremities ,extremity ,fact ,
 factor ,factored ,factoring ,factorings ,factorisation ,factorisations ,

factorise , factorization , factorizations , factorize , factorized ,
 factorizes , factorizing , factors , facts , feel , feeler , feelers , feeling ,
 feelings , feels , felt , find , finder , finders , finding , findings , finds ,
 finish , finished , finisher , finishers , finishes , finishing , finishings ,
 first , firsts , five , fives , follow , followed , follower , followers ,
 following , followings , follows , forget , forgets , forgetting , forgot ,
 forgotten , fortunate , fortunately , found , four , fours , fourth , fourths ,
 front , frontage , frontages , frontal , frontally , frontals , fronted ,
 fronting , fronts , gave , general , generalities , generality , generally ,
 generals , generalship , generalships , get , gets , getting , gettings , give ,
 given , giver , givers , gives , giving , givings , go , goer , goers , goes , going ,
 goings , gone , good , goodness , goodnesses , goods , got , gotten , great ,
 greatly , greatness , greatnesses , greats , hand , handed , handing , hands ,
 happen , happened , happening , happenings , happens , hard , hardened ,
 hardness , hardnesses , head , headed , header , headers , heading , headings ,
 heads , headship , headships , heady , hear , hearable , heard , hearer , hearers ,
 hearing , hearings , hears , help , helped , helper , helpers , helping , helpings ,
 helps , high , highly , highness , highnesses , highs , hint , hinted , hinting ,
 hints , honest , honestly , honestness , honestnesses , hope , hoped , hoper ,
 hoppers , hopes , hoping , imagination , imaginations , imaginative ,
 imaginatively , imaginativeness , imaginativenesses , imagine , imagined ,
 imagines , imagining , immediate , immediately , immediateness ,
 immediatenesses , impersonate , impersonated , impersonates ,
 impersonating , impersonation , impersonations , impersonator ,
 impersonators , impress , impressed , impresses , impressible , impressing ,
 impression , impressionism , impressionisms , impressionist ,
 impressionistic , impressionists , impressions , impressive , impressively ,
 impressiveness , impressivenesses , impressment , impressments , improve ,
 improved , improvement , improvements , improver , improvers , improves ,
 improving , in , inform , informant , informants , information , informations ,
 informative , informatively , informatory , informed , informer , informers ,
 informing , informings , informs , ins , instead , intend , intended , intending ,
 intends , intent , intention , intentional , intentionalities ,
 intentionality , intentionally , intentions , intently , intentness ,
 intentnesses , intents , is , isn't , job , jobbed , jobbing , jobs , keep , keeper ,
 keepers , keeping , keepings , keeps , kept , knew , know , knowable , knower ,
 knowers , knowing , knowingly , knowingness , knowingnesses , knowings , known ,
 knows , last , lasted , lasting , lasts , late , lateness , latenesses , later ,
 least , leasts , leave , leaved , leaver , leavers , leaves , leaving , lessor ,
 lessors , let , lets , letter , lettered , letterer , letterers , lettering ,
 letters , letting , level , leveler , levelers , levelled , leveller , levellers ,
 levelling , levels , life , lifer , lifers , likable , like , likeable , liked ,
 likeness , likenesses , likes , liking , literal , literally , literalness ,
 literalnesses , literals , livable , live , liveable , lived , liveness ,
 livenesses , liver , liveries , livers , livery , lives , living , livings ,
 locate , located , locater , locaters , locates , locating , location ,
 locations , locator , locators , long , longed , longer , longers , longest ,
 longing , longings , longness , longnesses , longs , look , looked , looker ,
 lookers , looking , looks , loose , loosed , loosely , looseness , loosenesses ,

looses, loosing, lose, loser, losers, loses, losing, losings, lost, losts, made, make, maker, makers, makes, making, makings, mark, marked, marker, markers, marking, markings, marks, mat, mater, maters, mats, matted, matting, mattings, may, maybe, mays, mention, mentioned, mentioner, mentioners, mentioning, mentions, message, messages, middle, middled, middles, middling, mind, minded, minding, minds, modal, modalities, modality, modals, mode, modes, month, monthlies, monthly, months, much, muches, muchness, muchnesses, multiple, multiples, multiplicities, multiplicity, multiply, must, mustiness, mustinesses, musts, musty, name, named, namer, namers, names, naming, namings, need, needed, needies, neediness, needinesses, needing, needn't, needs, needy, neither, never, new, newness, newnesses, next, nice, nicely, niceness, nicenesses, nine, nines, no, noes, non, none, nones, normal, normalities, normality, normals, notate, notated, notates, notating, notation, notations, note, noted, notes, nothing, nothings, notice, noticeabilities, noticeability, noticeable, noticeableless, noticeablelesses, noticeably, noticed, noticer, noticers, notices, noticing, notifiable, notification, notifications, notified, notifies, notify, notifying, noting, now, nowadays, nows, numbed, number, numbered, numbering, numberings, numbers, numbing, numbs, numerical, numerically, o.k., often, ok, okay, okayed, okaying, okays, okeh, okehs, okey, okeys, on, one, oneness, onenesses, ones, open, opened, opener, openers, opening, openings, openly, openness, opennesses, opens, opine, opined, opines, opining, opinion, opinionative, opinions, opportune, opportunely, opportuneness, opportunenesses, opportunities, opportunity, opt, optative, optative_mood, optative_moods, optatives, opted, opting, option, options, opts, order, ordered, orderer, orderers, ordering, orderings, orderlies, orderliness, orderlinesses, orderly, orders, other, otherness, othernesses, overall, overalls, part, parted, partial, partialities, partiality, partially, partialness, partialnesses, partials, parting, partings, partition, partitioned, partitioning, partitionings, partitionist, partitionists, partitions, partitive, partitives, parts, past, pastness, pastnesses, pasts, pause, paused, pauses, pausing, people, peopled, peoples, peopling, perhaps, person, persona, personae, personhood, personhoods, personification, personifications, personified, personifies, personify, personifying, place, placed, placement, placements, places, placing, possibilities, possibility, possible, possibleless, possiblelesses, possibles, possibly, practical, practicalities, practicality, practically, practice, practiced, practices, practicing, preparation, preparations, preparative, preparatory, prepare, prepared, prepares, preparing, presence, presences, present, presentation, presentations, presented, presenter, presenters, presenting, presently, presentment, presentments, presentness, presentnesses, presents, press, pressed, presses, pressing, pressingly, pressings, pressure, pressured, pressures, pressuring, pressurise, pressurize, pressurized, pressurizes, pressurizing, previous, previously, probabilism, probabilisms, probabilistic, probabilistically, probabilities, probability, probable, probables, probably, produce, produced, producer, producers, produces, producing, product, production, productions, products, put, puts, putting

, question, questioned, questioner, questioners, questioning, questioningly, questionings, questions, quite, raise, raised, raiser, raisers, raises, raising, raisings, real, realities, reality, realization, realizations, realize, realized, realizes, realizing, really, realness, realnesses, reals, reason, reasoned, reasoner, reasoners, reasoning, reasonings, reasons, recencies, recency, recent, recently, recentness, recentnesses, regard, regarded, regarding, regards, regular, regularities, regularity, regularly, regulars, relative, relatively, relatives, relativism, relativisms, relativistic, relativistically, relativities, relativity, remember, remembered, remembering, remembering, remembers, remembrance, remembrances, replied, replies, reply, replying, respective, respectively, restaurant, restaurants, return, returned, returning, returns, right, righted, righting, rightness, rightnesses, rights, rigor, rigorous, rigorously, rigorousness, rigorousnesses, rigors, ring, ringer, ringers, ringing, ringings, rings, rung, said, save, saved, saver, savers, saves, saving, savings, savior, saviors, saw, say, saying, sayings, says, second, seconded, seconder, seconders, seconding, secondment, secondments, seconds, see, seeing, seeings, seem, seemed, seeming, seems, seen, seer, seers, sees, sensation, sensations, sense, sensed, senses, sensibilities, sensibility, sensible, sensibleness, sensiblenesses, sensibly, sensify, sensing, sensings, sensitive, sensitively, sensitiveness, sensitivenesses, sensitives, sensitivities, sensitivity, sensor, sensorial, sensors, sensuous, sensuously, sensuousness, sensuousnesses, servant, servants, serve, served, server, servers, serves, serviceabilities, serviceability, serviceable, serviceableness, serviceablenesses, serving, servings, servitor, servitors, seven, sevens, several, show, showed, showily, showiness, showinesses, showing, showings, shown, shows, showy, side, sided, sides, siding, simple, simple-minded, simpleness, simplenesses, simples, simplicities, simplicity, simply, situate, situated, situates, situating, situation, situations, six, sixes, someone, someones, sometime, sorriness, sorrinesses, sorry, speak, speaker, speakers, speaking, speakings, speaks, speech, speeches, speechified, speechifier, speechifiers, speechifies, speechify, speechifying, spoke, spoken, stay, stayed, stayer, stayers, staying, stays, still, stilled, stilling, stillness, stillnesses, stills, stop, stoppage, stoppages, stopped, stopper, stoppered, stoppering, stoppers, stopping, stops, storage, storages, store, stored, stores, storing, strong, strongly, super, supers, supposal, supposals, suppose, supposed, supposes, supposing, supposition, suppositional, suppositions, suppositious, supposititious, sure, surely, sureness, surenesses, surprisal, surprisals, surprise, surprised, surpriser, surprisers, surprises, surprising, take, taken, taker, takers, takes, taking, takings, tale, tales, talk, talkative, talkatively, talkativeness, talkativenesses, talked, talker, talkers, talkily, talking, talkings, talks, talky, tell, teller, tellers, telling, tellingly, tellings, tells, ten, tens, term, termed, terming, terms, test, tested, testee, testees, tester, testers, testified, testifier, testifiers, testifies, testify, testifying, testimonial, testimonials, testimonies, testimony, testing, testings, tests, thank, thanked, thanking, thanks,

```

thing, things, think, thinker, thinkers, thinking, thinkings, thinks,
third, thirds, though, thought, thoughts, three, threes, today, todays,
told, tomorrow, tomorrows, took, top, topped, topper, toppers, topping,
tops, total, totalities, totality, totalled, totalling, totally, totals,
trial, trials, tried, trier, triers, tries, try, trying, turn, turned,
turner, turneries, turners, turnery, turning, turnings, turns, twice, two,
twoes, type, typed, types, typic, typification, typifications, typified,
typifies, typify, typifying, typing, typings, typist, typists, u,
unacceptably, understand, understandabilities, understandability,
understandable, understanding, understandingly, understandings,
understands, understood, unintentional, unintentionally, unusual,
unusually, unusualness, unusualnesses, us, usable, usableness,
usablenesses, usage, usages, usance, usances, use, useable, useableness,
useablenesses, user, users, uses, using, usings, usual, usually, usualness,
usualnesses, utile, utilisation, utilisations, utilise, utilitarian,
utilitarians, utilities, utility, utilizable, utilization, utilizations,
utilize, utilized, utilizer, utilizers, utilizes, utilizing, visit,
visitant, visitants, visitation, visitations, visited, visiting,
visitings, visitor, visitors, visits, want, wanted, wanter, wanters,
wanting, wants, warn, warned, warner, warners, warning, warnings, warns,
was, wasn't, way, ways, well, wellled, welling, wellness, wellnesses, wells,
went, were, weren't, will, willed, willing, willingly, willingness,
willingnesses, willings, wills, wish, wished, wishes, wishing, wishings,
within, won't, word, worded, wording, wordings, words, work, workable,
worked, worker, workers, working, works, year, yearlies, yearly, years, yes,
yeses, yesterday, yesterdays, yet

```

Code 9: NeutralWords.csv file context. This includes all the the Neutral words that were used to filter the both Positive and Negative Word Clouds

A.16 NegativeWords.csv

```

1 abnormal, abnormalities, abnormality, abnormally, absence, absences, absent
  , absented, absentee, absentees, absenting, absently, absents, abstain,
  abstained, abstainer, abstainers, abstaining, abstains, abstention,
  abstentions, abstentious, abstinence, abstinences, abstinent,
  abstinents, age, aged, agedness, agednesses, ageds, ages, aging, agitate,
  agitated, agitates, agitating, agitative, agitator, agitators, amateur,
  amateurs, arguable, arguably, argue, argued, arguer, arguers, argues,
  arguing, argument, argumentation, argumentations, argumentative,
  argumentatively, arguments, bad, badly, badness, badnesses, bads, bear,
  bearable, bearer, bearers, bearing, bearings, bears, begrudge, begrudged,
  begrudges, begrudging, bind, bindable, binder, binderies, binders,
  bindery, binding, bindings, binds, blamable, blame, blameable, blamed,
  blames, blameworthiness, blameworthinesses, blameworthy, blaming, bland,
  blandly, blandness, blandnesses, block, blockage, blockages, blocked,
  blocker, blockers, blocking, blockings, blocks, blocky, borne, bounce,
  bounced, bouncer, bouncers, bounces, bounciness, bouncinesses, bouncing,
  bouncy, bound, boundaries, boundary, bounded, bounder, bounders, bounding

```

, bounds , brawl , brawled , brawler , brawlers , brawling , brawls , break , breakabilities , breakability , breakable , breakableness , breakablenesses , breakables , breakage , breakages , breaker , breakers , breaking , breaks , bribable , bribe , bribed , briber , briberies , bribers , bribery , bribes , bribing , broke , broken , bullied , bullies , bully , bullying , captivate , captivated , captivates , captivating , captivation , captivations , captive , captives , captivities , captivity , center , centering , centerings , centers , central , centralities , centrality , centrally , centrals , centric , central , charge , charged , charger , chargers , charges , charging , clutter , cluttered , cluttering , clutters , collapse , collapsed , collapses , collapsible , collapsing , comfortable , comfortableness , comfortablenesses , comfortably , complain , complainant , complainants , complained , complainer , complainers , complaining , complains , complaint , complaints , complex , complexes , complexifier , complexifiers , complexify , complexities , complexity , complexly , complexness , complexnesses , confine , confined , confinement , confinements , confines , confining , conform , conformance , conformances , conformation , conformations , conformed , conforming , conformism , conformisms , conformist , conformists , conforms , confront , confrontation , confrontational , confrontations , confronted , confronting , confronts , confuse , confused , confuses , confusing , confusion , confusions , constipate , constipated , constipates , constipating , constipation , constipations , cried , crier , criers , cries , cry , crying , cryings , damage , damaged , damages , damaging , dead , deadly , deadness , deadnesses , deads , death , deathly , deaths , de cease , deceased , deceases , deceasing , decedent , decedents , defecate , defecated , defecates , defecating , defecation , defecations , defecator , defecators , deficiencies , deficiency , deficient , deform , deformation , deformational , deformations , deformed , deforming , deforms , delay , delayed , delaying , delays , demote , demoted , demotes , demoting , demotion , demotions , deplorable , deplorably , deprivation , deprivations , deprive , deprived , deprives , depriving , despair , despaired , despairing , despairs , desperate , desperately , desperates , desperation , desperations , destroy , destroyed , destroyer , destroyers , destroying , destroys , destructibilities , destructibility , destructible , destruction , destructions , destructive , destructively , destructiveness , destructivenesses , dice , die , died , dies , difficult , difficultness , difficultnesses , disadvantage , disadvantaged , disadvantageous , disadvantageously , disadvantages , disadvantaging , disallow , disallowed , disallowing , disallows , disappoint , disappointed , disappointing , disappointment , disappointments , disappoints , disapproval , disapprovals , disapprove , disapproved , disapproves , disapproving , disbelieve , disbelieved , disbeliever , disbelievers , disbelieves , disbelieving , discompose , discomposed , discomposes , discomposing , discomposure , discomposures , dishonest , dishonestly , dislike , disliked , dislikes , disliking , disorder , disordered , disordering , disorders , disorient , disorientation , disorientations , displease , displeased , displeases , displeasing , displeasure , displeasures , disprove , disproved , disprover , disprovers , disproves , disproving , disregard , disregarded , disregarding , disregards , disrespect , disrespected , disrespecting ,

disrespects , dissatisfaction , dissatisfactions , dissatisfactory ,
 dissatisfied , dissatisfies , dissatisfy , dissatisfying , distress ,
 distressed , distresses , distressing , distressingly , doom , doomed ,
 dooming , dooms , dying , dyings , enforce , enforced , enforcement ,
 enforcements , enforcer , enforcers , enforces , enforcing , evil , evilly ,
 evilness , evilnesses , evils , excusatory , excuse , excused , excuser ,
 excusers , excuses , excusing , fall , fallen , faller , fallers , falling , falls
 , false , falsely , falseness , falsenesses , falsification , falsifications ,
 falsified , falsifier , falsifiers , falsifies , falsify , falsifying ,
 falsifyings , falsities , falsity , fell , flaw , flawed , flawing , flaws , foe ,
 foes , forbade , forbid , forbiddance , forbiddances , forbidden , forbidding ,
 forbiddingly , forbiddings , forbids , foreign , foreignness , foreignnesses
 , forget , forgets , forgetting , forgot , forgotten , freeze , freezer ,
 freezers , freezes , freezing , freezings , froze , frozen , fuck , fucked ,
 fucker , fuckers , fucking , fuckings , fucks , gossip , gossip , gossipers ,
 gossiping , gossipings , gossipped , gossiping , gossips , gossip , grueling
 , gruelling , hate , hated , hater , haters , hates , hating , hesitance ,
 hesitances , hesitancies , hesitancy , hesitant , hesitantly , hesitate ,
 hesitated , hesitater , hesitaters , hesitates , hesitating , hesitation ,
 hesitations , hesitator , hesitators , hid , hidden , hide , hideous , hideously
 , hideousness , hideousnesses , hides , hiding , hidings , hind , hinder ,
 hindered , hindering , hinders , hindrance , hindrances , hinds , homeless ,
 homelesses , homelessness , homelessnesses , horrible , horribly , hostile ,
 hostilely , hostiles , hostilities , hostilities , hostility , ignorance ,
 ignorances , ignorant , ignorantly , ignorantness , ignorantnesses , ignore ,
 ignored , ignores , ignoring , ill , illness , illnesses , ills , imperfect ,
 imperfectly , imperfectness , imperfectnesses , imperfects ,
 impossibilities , impossibility , impossible , impossibility ,
 impossibility , impossibilities , impossibility , impossibility , impotence , impotences ,
 impotencies , impotency , impotent , impotently , impractical ,
 impracticalities , impracticality , improbabilities , improbability ,
 improbable , improbability , improbability , improbability , improper ,
 improperly , improperness , impropernesses , inauspicious , inauspiciously
 , inauspiciousness , inauspiciousnesses , incorrect , incorrectly ,
 incorrectness , incorrectnesses , inedible , inefficiencies , inefficiency
 , inefficient , inefficiently , inopportune , inopportune ,
 inopportune , inopportune , inopportune , inopportune , insubstantial , insubstantialities
 , insubstantiality , interrogate , interrogated , interrogates ,
 interrogating , interrogation , interrogations , interrogative ,
 interrogative_mood , interrogative_moods , interrogatively ,
 interrogatives , interrogator , interrogatories , interrogators ,
 interrogatory , interrupt , interrupted , interrupter , interrupters ,
 interrupting , interruption , interruptions , interrupts , inverse ,
 inverses , irregular , irregularities , irregularity , irregularly ,
 irregulars , issue , issued , issuer , issuers , issues , issuing , issuings ,
 lack , lacked , lacking , lacks , lament , lamentable , lamentation ,
 lamentations , lamented , lamenter , lamenters , lamenting , laments , lean ,
 leaner , leaners , leaning , leanness , leannesses , leans , leant , leave ,
 leaved , leaver , leavers , leaves , leaving , left , lefts , line , lineal ,

lineally, lineation, lineations, lined, liner, liners, lines, lining,
 linings, little, littleness, littlenesses, littles, lodge, lodged,
 lodgement, lodgements, lodger, lodgers, lodges, lodging, lodgings,
 malfunction, malfunctioned, malfunctioning, malfunctions, manipulate,
 manipulated, manipulates, manipulating, manipulation, manipulations,
 manipulative, manipulatively, manipulator, manipulators, mean, meaning,
 meanings, meanly, meanness, meannesses, means, meant, mislaid, mislay,
 mislaying, mislays, misplace, misplaced, misplacement, misplacements,
 misplaces, misplacing, miss, missed, misses, missing, negate, negated,
 negates, negating, negation, negations, negative, negatively,
 negativeness, negativenesses, negatives, negativities, negativity,
 nonexempt, nonprofessional, obstruct, obstructed, obstructer,
 obstructers, obstructing, obstruction, obstructions, obstructive,
 obstructively, obstructor, obstructors, obstructs, obviate, obviated,
 obviates, obviating, obviation, obviations, old, oldness, oldnesses, olds,
 paranormal, pee, peed, peeing, peeings, pees, penalisation,
 penalisations, penalise, penalties, penalty, perish, perished, perishes,
 perishing, perplex, perplexed, perplexes, perplexing, persist, persisted,
 persistence, persistences, persistencies, persistency, persistent,
 persistently, persisting, persists, piss, pissed, pisser, pissers, pisses,
 pissing, pissings, pit, pits, pitted, pitting, pittings, poor, poorly,
 poorness, poornesses, poors, positive, positively, positiveness,
 positivenesses, positives, positivism, positivisms, positivist,
 positivistic, positivists, positivities, positivity, privation,
 privations, problem, problematic, problematical, problematically,
 problems, promiscuities, promiscuity, promiscuous, promiscuously,
 promiscuousness, promiscuousnesses, raw, rawness, rawnesses, raws,
 refusal, refusals, refuse, refused, refuses, refusing, resent, resented,
 resenting, resentment, resentments, resents, rust, rusted, rustiness,
 rustinesses, rusting, rusts, rusty, sacrifice, sacrificed, sacrificer,
 sacrificers, sacrifices, sacrificial, sacrificing, sanction,
 sanctionative, sanctioned, sanctioning, sanctions, scream, screamed,
 screamer, screamers, screaming, screamingly, screamings, screams, severe,
 severely, severeness, severenesses, severities, severity, shoot,
 shooter, shooters, shooting, shootings, shoots, short, shorted, shorting,
 shortly, shortness, shortnesses, shorts, shot, shots, shout, shouted,
 shouter, shouters, shouting, shoutings, shouts, shut, shuts, shutter,
 shuttered, shuttering, shutters, shutting, shuttings, slash, slashed,
 slasher, slashers, slashes, slashing, slow, slowed, slowing, slowings,
 slowly, slowness, slownesses, slows, small, smallness, smallnesses,
 smalls, smash, smashed, smasher, smashers, smashes, smashing, smashings,
 spill, spillage, spillages, spiller, spillers, spilling, spills, spilt,
 storm, stormed, stormily, storminess, storminesses, storming, storms,
 stormy, strange, strangely, strangeness, strangenesses, stray, strayed,
 strayer, strayers, straying, strays, strike, striker, strikers, strikes,
 striking, strikingly, strikingness, strikingnesses, strikings, struck,
 suck, sucked, sucker, suckers, sucking, suckings, sucks, suction, suctions,
 suctorial, tasteless, tastelessly, tastelessness, tastelessnesses, tax,
 taxabilities, taxability, taxable, taxation, taxations, taxed, taxer,

```

taxers , taxes , taxing , terrible , terribleness , terriblenesses , terribly ,
tough , toughly , toughness , toughnessees , toughs , ugliness , uglinesses ,
ugly , uneasily , uneasiness , uneasinesses , uneasy , unfamiliar ,
unfamiliarities , unfamiliarity , unfortunate , unfortunately ,
unfortunates , unfree , unfriendliness , unfriendlinesses , unfriendly ,
unholiness , unholinesses , unholy , unmade , unmake , unmakes , unmaking ,
unmanageable , unmanageableness , unmanageablenesses , unmanageably ,
unpleasant , unpleasantly , unpleasantness , unpleasantnesses ,
unprofessional , unreceptive , unregretful , unretentive , unsatisfied ,
unseal , unsealed , unsealing , unseals , unsure , untrimmed , unusual ,
unusually , unusualness , unusualnesses , unwelcome , urinary ,
urinary_system , urinary_systems , urinate , urinated , urinates , urinating
, urination , urinations , urinator , urinators , urine , urines , wait , waited ,
waiter , waiters , waiting , waitings , waits , weak , weakly , weakness ,
weaknesses , wear , wearable , wearables , wearer , wearers , wearing , wearings
, wears , wore , worn , worse , worses , worst , worsted , worsting , worsts , wrong ,
wronged , wrongful , wrongfully , wrongfulness , wrongfulnesses , wronging ,
wrongly , wrongness , wrongnesses , wrongs

```

Code 10: NegativeWords.csv file context. This includes all the the Neutral words that were used to filter the both Positive Word Clouds

References

- [1] “Portugal: The Portuguese Food Service Sector,” <https://www.fas.usda.gov/data/portugal-portuguese-food-service-sector>.
- [2] Z. Zhang, Q. Ye, R. Law, and Y. Li, “The impact of e-word-of-mouth on the online popularity of restaurants: A comparison of consumer reviews and editor reviews,” *International Journal of Hospitality Management*, vol. 29, no. 4, pp. 694–700, Dec. 2010.
- [3] C. Wu, H. Che, T. Y. Chan, and X. Lu, “The Economic Value of Online Reviews,” *Marketing Science*, vol. 34, no. 5, pp. 739–754, Sep. 2015.
- [4] Z. Xiang, Q. Du, Y. Ma, and W. Fan, “A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism,” *Tourism Management*, vol. 58, pp. 51–65, Feb. 2017.
- [5] N. R. Mavani, J. M. Ali, S. Othman, M. A. Hussain, H. Hashim, and N. A. Rahman, “Application of Artificial Intelligence in Food Industry—a Guideline,” *Food Engineering Reviews*, vol. 14, no. 1, pp. 134–175, Mar. 2022.
- [6] “International Journal of Academic Research in Management,” vol. 8, no. 1, 2019.
- [7] M. Cheng and X. Jin, “What do Airbnb users care about? An analysis of online review comments,” *International Journal of Hospitality Management*, vol. 76, pp. 58–70, Jan. 2019.
- [8] F. A. Acheampong, C. Wenyu, and H. Nunoo-Mensah, “Text-based emotion detection: Advances, challenges, and opportunities,” *Engineering Reports*, vol. 2, no. 7, p. e12189, 2020.
- [9] P. Nandwani and R. Verma, “A review on sentiment analysis and emotion detection from text,” *Social Network Analysis and Mining*, vol. 11, no. 1, p. 81, Aug. 2021.
- [10] M. Khanbhai, P. Anyadi, J. Symons, K. Flott, A. Darzi, and E. Mayer, “Applying natural language processing and machine learning techniques to patient experience feedback: A systematic review,” *BMJ Health & Care Informatics*, vol. 28, no. 1, p. e100262, Mar. 2021.
- [11] “The NLP divide: English is not the only natural language,” Jun. 2021.

- [12] Lingua, “The 20 most spoken languages in the world in 2022,” <https://lingua.edu/the-20-most-spoken-languages-in-the-world-in-2022/>, Jun. 2022.
- [13] E. F. d. S. M. Lopes, “The importance of mobile and online services in the Portuguese restaurant industry : A current overview of restaurant search and online reservation platforms,” Master’s thesis, Feb. 2016.
- [14] “Zomato,” *Wikipedia*, Jan. 2023.
- [15] S. Pradha, M. N. Halgamuge, and N. Tran Quoc Vinh, “Effective Text Data Preprocessing Technique for Sentiment Analysis in Social Media Data,” in *2019 11th International Conference on Knowledge and Systems Engineering (KSE)*. Da Nang, Vietnam: IEEE, Oct. 2019, pp. 1–8.
- [16] A. Krouska, C. Troussas, and M. Virvou, “The effect of preprocessing techniques on Twitter sentiment analysis,” in *2016 7th International Conference on Information, Intelligence, Systems & Applications (IISA)*. Chalkidiki, Greece: IEEE, Jul. 2016, pp. 1–5.
- [17] H.-T. Duong and T.-A. Nguyen-Thi, “A review: Preprocessing techniques and data augmentation for sentiment analysis,” *Computational Social Networks*, vol. 8, no. 1, p. 1, Jan. 2021.
- [18] E. G. Wibawa, P. K. Dewa, and T. Siswantoro, “Restaurant Business Insights based on Zomato Online Food Marketplace Big Data Scraping,” 2021.
- [19] V. Vajjhala and M. Ghosh, “Decoding the Effect of Restaurant Reviews on Customer Choice: Insights from Zomato,” *Journal of Foodservice Business Research*, vol. 25, no. 5, pp. 533–560, Sep. 2022.
- [20] Y. Li and S. Manoharan, *A Performance Comparison of SQL and NoSQL Databases*, Aug. 2013.
- [21] J. R. Lourenço, B. Cabral, P. Carreiro, M. Vieira, and J. Bernardino, “Choosing the right NoSQL database for the job: A quality attribute evaluation,” *Journal of Big Data*, vol. 2, no. 1, p. 18, Aug. 2015.
- [22] L. Hickman, S. Thapa, L. Tay, M. Cao, and P. Srinivasan, “Text Preprocessing for Text Mining in Organizational Research: Review and Recommendations,” *Organizational Research*

Methods, vol. 25, no. 1, pp. 114–146, Jan. 2022.

- [23] “How to Choose The Right Database for Your Application?” Dec. 2020.
- [24] M. Dave, “SQL and NoSQL Databases,” *International Journal of Advanced Research in Computer Science and Software Engineering*, Aug. 2012.
- [25] “XAMPP Installers and Downloads for Apache Friends,” <https://www.apachefriends.org/>.
- [26] L. Guo, “How to Efficiently Choose the Right Database for Your Applications,” <https://www.pingcap.com/blog/how-to-efficiently-choose-the-right-database-for-your-applications/>, Feb. 2021.
- [27] T. Vajk, P. Feher, K. Fekete, and H. Charaf, “Denormalizing data into schema-free databases,” in *2013 IEEE 4th International Conference on Cognitive Infocommunications (CogInfoCom)*. Budapest, Hungary: IEEE, Dec. 2013, pp. 747–752.
- [28] M. R. Yaakub, M. I. A. Latiffi, and L. S. Zaabar, “A Review on Sentiment Analysis Techniques and Applications,” *IOP Conference Series: Materials Science and Engineering*, vol. 551, no. 1, p. 012070, Aug. 2019.
- [29] R. S. Jagdale and S. S. Deshmukh, “Sentiment Classification on Twitter and Zomato Dataset Using Supervised Learning Algorithms,” in *2020 International Conference on Smart Innovations in Design, Environment, Management, Planning and Computing (ICSIDEMPC)*, Oct. 2020, pp. 330–334.
- [30] N. C. Dang, M. N. Moreno-García, and F. De la Prieta, “Sentiment Analysis Based on Deep Learning: A Comparative Study,” *Electronics*, vol. 9, no. 3, p. 483, Mar. 2020.
- [31] M. R. Hasan, M. Maliha, and M. Arifuzzaman, “Sentiment Analysis with NLP on Twitter Data,” in *2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2)*, Jul. 2019, pp. 1–4.
- [32] R. Gupta, S. Sameer, H. Muppavarapu, M. K. Enduri, and S. Anamalamudi, “Sentiment Analysis on Zomato Reviews,” in *2021 13th International Conference on Computational Intelligence and Communication Networks (CICN)*, Sep. 2021, pp. 34–38.
- [33] M. Kanakaraaj and R. M. R. Guddeti, “NLP based sentiment analysis on Twitter data using

- ensemble classifiers,” in *2015 3rd International Conference on Signal Processing, Communication and Networking (ICSCN)*, Mar. 2015, pp. 1–5.
- [34] “[PDF] Sentiment Analysis on Twitter Data - Free Download PDF,” <https://silo.tips/download/sentiment-analysis-on-twitter-data>.
- [35] K. Zahoor, N. Z. Bawany, and S. Hamid, “Sentiment Analysis and Classification of Restaurant Reviews using Machine Learning,” in *2020 21st International Arab Conference on Information Technology (ACIT)*, Nov. 2020, pp. 1–6.
- [36] “(20+) The SWOT’s Guide™ to KARACHI & ALL SINDH EATERIES®© | Facebook,” <https://www.facebook.com/groups/swotsrestguide/>.
- [37] A. Duarte Alonso, M. O’neill, Y. Liu, and M. O’shea, “Factors Driving Consumer Restaurant Choice: An Exploratory Study From the Southeastern United States,” *Journal of Hospitality Marketing & Management*, vol. 22, no. 5, pp. 547–567, Jul. 2013.
- [38] H. M. Jia, E. W. Wan, and W. Zheng, “Stars versus Bars: How the Aesthetics of Product Ratings “Shape” Product Preference,” *Journal of Consumer Research*, vol. 50, no. 1, pp. 142–166, May 2023.
- [39] F. L. Dos Santos and M. Ladeira, “The Role of Text Pre-processing in Opinion Mining on a Social Media Language Dataset,” in *2014 Brazilian Conference on Intelligent Systems*, Oct. 2014, pp. 50–54.
- [40] A. Branco, D. Parada, M. Silva, F. Mendonça, S. S. Mostafa, and F. Morgado-Dias, “Sentiment Analysis in Portuguese Restaurant Reviews: Application of Transformer Models in Edge Computing,” *Electronics*, vol. 13, no. 3, p. 589, Jan. 2024.
- [41] M. M. Danilak, “Langdetect: Language detection library ported from Google’s language-detection.”
- [42] S. Ranathunga and N. de Silva, “Some Languages are More Equal than Others: Probing Deeper into the Linguistic Disparity in the NLP World,” <https://arxiv.org/abs/2210.08523v2>, Oct. 2022.
- [43] N. Baccouri, “Deep-translator: A flexible free and unlimited python tool to translate between

different languages in a simple way using multiple translators.”

- [44] D. J. Ladani and N. P. Desai, “Stopword Identification and Removal Techniques on TC and IR applications: A Survey,” in *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Mar. 2020, pp. 466–472.
- [45] S. Sarica and J. Luo, “Stopwords in technical language processing,” *PLOS ONE*, vol. 16, no. 8, p. e0254937, Aug. 2021.
- [46] S. Bird, E. Klein, and E. Loper, “Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit,” O’Reilly Media, Inc., 2009.
- [47] R. M. Chapman, M. N. Gardner, and M. Lyons, “Gender differences in emotional connotative meaning of words measured by Osgood’s semantic differential techniques in young adults,” *Humanities and Social Sciences Communications*, vol. 9, no. 1, pp. 1–11, Apr. 2022.
- [48] U. Hess, S. Senécal, G. Kirouac, P. Herrera, P. Philippot, and R. E. Kleck, “Emotional expressivity in men and women: Stereotypes and self-perceptions,” *Cognition & Emotion*, vol. 14, no. 5, pp. 609–642, Sep. 2000.
- [49] M. Thelwall, “Gender bias in sentiment analysis,” *Online Information Review*, vol. 42, no. 1, pp. 45–57, Jan. 2018.
- [50] I. S. Pérez, “Gender-guesser: Get the gender from first name.”
- [51] P. Nandwani and R. Verma, “A review on sentiment analysis and emotion detection from text,” *Social Network Analysis and Mining*, vol. 11, no. 1, p. 81, Aug. 2021.
- [52] M. A. Rahman and M. H. Seddiqui, “Comparison of Classical Machine Learning Approaches on Bangla Textual Emotion Analysis,” Jul. 2019.
- [53] N. Alswaidan and M. E. B. Menai, “A survey of state-of-the-art approaches for emotion recognition in text,” *Knowledge and Information Systems*, vol. 62, no. 8, pp. 2937–2987, Aug. 2020.
- [54] R. Dey, D. Sarddar, I. Sarkar, R. Bose, and S. Roy, “A Literature Survey on Sentiment Analysis Techniques involving Social Media and Online Platforms,” *International Journal of Scientific & Technology Research*, vol. 9, pp. 1–8, Jun. 2020.
- [55] “EMOTAIX.PT.”

- [56] A. Piolat and R. Bannour, “EMOTAIX : Un scénario de Tropes pour l’identification automatisée du lexique émotionnel et affectif,” *L’Année psychologique*, vol. 109, no. 04, p. 655, Dec. 2009.
- [57] L. Yujian and L. Bo, “A Normalized Levenshtein Distance Metric,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 1091–1095, Jun. 2007.
- [58] A. G. Jivani, “A Comparative Study of Stemming Algorithms,” *International journal of computer technology and applications*, vol. 2, 2011.
- [59] D. Fernando, “Pysinonimos,” Aug. 2022.
- [60] “SentiStrength - sentiment strength detection in short texts - sentiment analysis, opinion mining,” <http://sentistrength.wlv.ac.uk/>.
- [61] H. Zhao, Z. Liu, X. Yao, and Q. Yang, “A machine learning-based sentiment analysis of online product reviews with a novel term weighting and feature selection approach,” *Information Processing & Management*, vol. 58, no. 5, p. 102656, Sep. 2021.
- [62] N. Claypo and S. Jaiyen, “Opinion mining for thai restaurant reviews using K-Means clustering and MRF feature selection,” in *2015 7th International Conference on Knowledge and Smart Technology (KST)*, Jan. 2015, pp. 105–108.
- [63] A. Madasu and S. Elango, “Efficient feature selection techniques for sentiment analysis,” *Multimedia Tools and Applications*, vol. 79, no. 9, pp. 6313–6335, Mar. 2020.
- [64] A. Sharma and S. Dey, “A comparative study of feature selection and machine learning techniques for sentiment analysis,” in *Proceedings of the 2012 ACM Research in Applied Computation Symposium*, ser. RACS ’12. New York, NY, USA: Association for Computing Machinery, Oct. 2012, pp. 1–7.
- [65] N. Claypo and S. Jaiyen, “Opinion mining for Thai restaurant reviews using neural networks and mRMR feature selection,” in *2014 International Computer Science and Engineering Conference (ICSEC)*, Jul. 2014, pp. 394–397.
- [66] P. Emerson, “The original Borda count and partial voting,” *Social Choice and Welfare*, vol. 40, no. 2, pp. 353–358, Feb. 2013.

- [67] “Correlation,” https://www.jmp.com/en_ca/statistics-knowledge-portal/what-is-correlation.html.
- [68] “Correlation (Pearson, Kendall, Spearman),” <https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/correlation-pearson-kendall-spearman/>.
- [69] “Pandas.DataFrame.corr — pandas 1.5.2 documentation,” <https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.corr.html>.
- [70] C. Ding and H. Peng, “Minimum Redundancy Feature Selection From Microarray Gene Expression Data,” in *Journal of Bioinformatics and Computational Biology - JBCB*, vol. 3, Sep. 2003, pp. 523–528.
- [71] P. A. Mundra and J. C. Rajapakse, “SVM-RFE With MRMR Filter for Gene Selection,” *IEEE Transactions on NanoBioscience*, vol. 9, no. 1, pp. 31–37, Mar. 2010.
- [72] smazzanti, “Smazzanti/mrmr,” Aug. 2023.
- [73] “Borda count,” *Wikipedia*, Jul. 2024.
- [74] Z. Mottaghinia, M.-R. Feizi-Derakhshi, L. Farzinvas, and P. Salehpour, “A review of approaches for topic detection in Twitter,” *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 33, no. 5, pp. 747–773, Sep. 2021.
- [75] K. Garcia and L. Berton, “Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA,” *Applied Soft Computing*, vol. 101, p. 107057, Mar. 2021.
- [76] A. C. Calheiros, S. Moro, and P. Rita, “Sentiment Classification of Consumer-Generated Online Reviews Using Topic Modeling,” *Journal of Hospitality Marketing & Management*, vol. 26, no. 7, pp. 675–693, Oct. 2017.
- [77] T. Menner, W. Höpken, M. Fuchs, and M. Lexhagen, “Topic Detection: Identifying Relevant Topics in Tourism Reviews,” in *Information and Communication Technologies in Tourism 2016*, A. Inversini and R. Schegg, Eds. Cham: Springer International Publishing, 2016, pp. 411–423.
- [78] M. Cataldi, A. Ballatore, I. Tididi, and M.-A. Aufaure, “Good location, terrible food: Detecting

- feature sentiment in user-generated reviews,” *Social Network Analysis and Mining*, vol. 3, no. 4, pp. 1149–1163, Dec. 2013.
- [79] “Sklearn.decomposition.LatentDirichletAllocation,” <https://scikit-learn/stable/modules/generated/sklearn.decomposition.LatentDirichletAllocation.html>.
- [80] G. Grefenstette, “Tokenization,” in *Syntactic Wordclass Tagging*, ser. Text, Speech and Language Technology, H. van Halteren, Ed. Dordrecht: Springer Netherlands, 1999, pp. 117–133.
- [81] R. Řehůřek and P. Sojka, “Software Framework for Topic Modelling with Large Corpora,” University of Malta, Valetta, MT, pp. 45–50, May 2010.
- [82] “Gensim: Topic modelling for humans,” <https://radimrehurek.com/gensim/>.
- [83] Y. Bai, E. Yang, B. Han, Y. Yang, J. Li, Y. Mao, G. Niu, and T. Liu, “Understanding and Improving Early Stopping for Learning with Noisy Labels,” in *Advances in Neural Information Processing Systems*, vol. 34. Curran Associates, Inc., 2021, pp. 24 392–24 403.
- [84] P. Liashchynskyi and P. Liashchynskyi, “Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS,” Dec. 2019.
- [85] “Streamlit • A faster way to build and share data apps,” <https://streamlit.io/>, Jan. 2021.
- [86] “Welcome to Streamlit ,” Streamlit, Oct. 2023.
- [87] S. Patil and V. Loksha, “Live Twitter Sentiment Analysis Using Streamlit Framework,” Rochester, NY, May 2022.
- [88] F. Heimerl, S. Lohmann, S. Lange, and T. Ertl, “Word Cloud Explorer: Text Analytics Based on Word Clouds,” in *2014 47th Hawaii International Conference on System Sciences*. Waikoloa, HI: IEEE, Jan. 2014, pp. 1833–1842.
- [89] D. Chakravorty, “Gutfeeling/word_forms,” Nov. 2023.
- [90] A. Gupta, P. Goswami, N. Chaudhary, and R. Bansal, “Deploying an Application using Google Cloud Platform,” in *2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*. Bangalore, India: IEEE, Mar. 2020, pp. 236–239.