NOVA IMS Information Management School

MDSAA

Master Degree Program in
Data Science and Advanced Analytics

EXPLORING MULTILINGUAL SENTIMENT ON URBAN MOBILITY IN LISBON

Leveraging Twitter Data and Large Language Models

Janaina Carvalho dos Santos

Dissertation

presented as partial requirement for obtaining a Master's Degree in Data Science and Advanced Analytics

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

EXPLORING MULTILINGUAL SENTIMENT ON URBAN MOBILITY IN LISBON

Leveraging Twitter data and Large Language Models

by

Janaina Carvalho dos Santos

Master Thesis presented as partial requirement for obtaining the Master's degree in Data Science and Advanced Analytics, with a specialization in Data Science

Supervised by

Bruno Jardim, PhD, NOVA Information Management School

Co-Supervised by

Miguel de Castro Neto, PhD, NOVA Information Management School

July, 2024

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledged the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Lisbon, July 2024

DEDICATION

To my family and friends,

This work is dedicated to you all. Thank you to my mom for your endless love and support, no matter the distance. Your strength has always guided me. To my dad, who although no longer with us, continues to inspire me with his values and teachings about life and learning.

To all my family and friends back home, your encouragement has meant so much. And to the new friends that Lisbon has brought me, your warmth and assistance have made my time here truly special.

I am deeply grateful to each of you for being part of this journey.

ABSTRACT

This thesis explores the various factors and interactions involved in multilingual sentiment analysis in relation to Lisbon's urban mobility, using advanced sentiment analysis techniques. By analyzing over 500,000 tweets, this study identifies public sentiments and key issues related to urban mobility in Lisbon. The methodology involves filtering tweets based on relevant keywords, employing a Chain-of-Thought (CoT) prompting method for sentiment classification, and addressing LLM hallucination through post-processing. Key findings reveal significant concentrations of negative sentiment in specific locations, recurring themes in positive and negative feedback, and the importance of considering linguistic diversity in sentiment analysis. The results highlight critical areas for targeted interventions and improvements in urban mobility infrastructure and services. Future work suggests leveraging more advanced language models, diverse data sources to further enhance sentiment analysis and support data-driven urban planning.

KEYWORDS

Sentiment Analysis; Data-Driven Urban Planning; Social Media Analytics; Chain-of-Thought Prompting ; Large Language Models; GPT; Spatiotemporal Analysis

Sustainable Development Goals (SDG):

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

NLP	Natural Language Processing
LLM	Large Language Model
CoT	Chain of Thoughts
GPT	Generative Pre-trained Transformer
BERT	Bidirectional Encoder Representations from Transformers
RoBERTa	Robustly optimized BERT approach
GPT	Generative Pre-trained Transformer

1. INTRODUCTION

Lately, urban mobility has become a focal point, particularly in busy cities like Lisbon. The increasing number of tourists, combined with the city's varied transportation choices, leads to complex mobility patterns and challenges. Understanding the sentiments of tourists, particularly as expressed on social media platforms like Twitter or Instagram, is essential to clarify the multifaceted landscape of urban mobility in Lisbon (Louro et al., 2021).

1.1. BACKGROUND CONTEXT

Lisbon, a renowned global tourist destination, attracts millions of visitors. The city's historic streets, panoramic views, and vibrant culture contribute to its appeal. Managing urban mobility effectively, however, remains a significant challenge. The city has a unique mobility situation, due to its compact streets and complex transportation network, including trams, buses, and metro lines (Louro et al., 2021). Additionally, factors like traffic congestion, accessibility to key destinations, and public transportation usage contribute to the city's mobility challenges.

1.2. THE RELEVANCE OF UNDERSTANDING TOURIST SENTIMENT

In the digital age, tourists frequently share their experiences and opinions on social media. These online expressions provide valuable insights into their perceptions of urban mobility in Lisbon. Analyzing sentiments in their tweets, posts, and images reveals their positive experiences, frustrations, and preferences (Serna et al., 2017). Such sentiment analysis can guide urban planners and policymakers in refining transportation systems and improving the overall tourist experience in Lisbon.

1.3. THE RESEARCH QUESTION AND OBJECTIVES

This research aims to answer a critical question: What is the perception of social media content related to the experience with urban transportation in Lisbon? Despite the available research on tourists' social media posts sentiment analysis (Flores-Ruiz et al., 2021; Tao et al., 2019), and specifically in urban mobility (Balla et al., 2023; Nokkaew et al., 2024; Fontes et al., 2023; Serna et al., 2017), a noticeable gap exists in understanding the multilingual and multimodal sentiments' impact in this domain. This research, therefore, seeks to analyze and recognize the patterns and peculiarities in sentiments related to Lisbon's mobility, considering multiple languages, chronological and geographical factors, to explore their impact on the city transportation.

Finally, the main objectives of this research are to collect and preprocess social media data from Twitter, and to develop a model to understand the sentiment analysis classification specifically to the context of Lisbon's urban mobility (Serna et al., 2017). The goal is to provide practical insights to enhance urban mobility infrastructure and improve tourists' experiences in Lisbon.

1.4. OVERVIEW OF THESIS STRUCTURE

The thesis is organized to fully address the research objectives, starting with an Introduction that provides the context. The section focus is on raising the main questions that will be addressed by the study.

The Literature Review follows, providing a detailed investigation of existing research in sentiment analysis, its application in urban planning, and the challenges of multilingual sentiment analysis. This section also explores the distinct analytical tools and models used in this field, placing a solid foundation for the study.

In sequence, The Methodology section provides the procedures for data collection and analysis, ensuring future reproducibility.

The Results and Discussion sections then articulates the outcomes of the sentiment analysis, interpreting and discussing their implications for urban mobility in Lisbon.

Finally, the Conclusion and Future Works section offers the summarized findings, limitations, and suggestions for the future. This comprehensive structure ensures a clear and logical progression from the study's introduction to its conclusion, aiming to contribute meaningful insights to the field of urban mobility and sentiment analysis.

2. LITERATURE REVIEW

This section provides an extensive analysis of existing literature on sentiment analysis in different domains, covering previous published work from 2017 to 2023. The review includes 33 research papers, mainly focused on exploring the capabilities of natural language processing or machine learning models for sentiment analysis.

2.1. SENTIMENT ANALYSIS IN DIFFERENT CONTEXTS

Sentiment analysis has turned into a key tool for understanding public sentiments in multiple domains (Gandhi et al., 2023). Its applications range from evaluating product reviews to interpreting political discussions, providing valuable insights into public opinions and perceptions. Table 2.1 brings 22 studies mainly based on Social Media data that used sentiment analysis in different contexts.

Table 2.1 – Reference on Sentiment Analysis

Topic	Authors/Year	Dataset
A social media analysis of travel preferences and attitudes, before and during Covid-19	(Hardt & Glückstad, 2024)	- Over one million Reddit posts on travel-related subreddits
Analyzing online public opinion on Thailand-China high-speed train and Laos-China railway mega-projects using advanced machine learning for sentiment analysis	(Nokkaew et al., 2024)	- Public Thai reviews from YouTube
Deep Learning Model for COVID-19 Sentiment Analysis on Twitter	(Contreras Hernández et al., 2023)	- Tweets collected during one of the waves of COVID-19 in Mexico.
Examining shifts in public discourse on electric mobility adoption through Twitter data	(Balla et al., 2023)	- Public tweet database from 2012-2022.
Leveraging Social Media as a Source of Mobility Intelligence: An NLP-Based Approach	(Fontes et al., 2023)	- Twitter messages collected from London, Melbourne, and New York
More than a Feeling: Accuracy and Application of Sentiment Analysis	(Hartmann et al., 2023)	- Meta-analysis of 272 datasets - 12 million sentiment-labeled text documents.

Multimodal sentiment analysis: A systematic review of history, datasets, multimodal fusion methods, applications, challenges and future directions	(Gandhi et al., 2023)	- Multimodal datasets created by collecting videos from internet video sharing platforms.
Public perception of electric vehicles on Reddit and Twitter: A cross-platform analysis	(Ruan & Lv, 2023)	- 3,437,917 Reddit posts and 7,383,327 Tweets.
Twitter Sentiment Analysis	(Islam et al., 2023)	- Real-time Twitter data by the Twitter API
Machine Learning Techniques for Sentiment Analysis of COVID-19-Related Twitter Data	(Braig et al., 2023)	- Twitter datasets with at least 5,000 labeled Tweets.
SA-MSVM: Hybrid Heuristic Algorithm-based Feature Selection for Sentiment Analysis in Twitter	(Thamil Selvi & PushpaLaksmi, 2023)	- Amazon.com dataset of Clothing E-Commerce
Discovering a tourism destination with social media data: BERT-based sentiment analysis	(Viñán-Ludeña & de Campos, 2022)	- 90,725 Instagram posts, 235,755 Twitter tweets and 30,805 reviews from TripAdvisor.
RoBERTa-LSTM: A Hybrid Model for Sentiment Analysis With Transformer and Recurrent Neural Network	(Tan et al., 2022)	- IMDb dataset and Twitter US Airline Sentiment dataset.
Text Sentiment Analysis Based on Transformer and Augmentation	(Gong et al., 2022)	- IMDb dataset and Twitter US Airline dataset.
Detecting sentiment dynamics and clusters of Twitter users for trending topics in COVID-19 pandemic	(Ahmed et al., 2021)	- COVID-19 related tweets dataset with 100 million tweets.
Sentiment analysis to measure quality and build sustainability in tourism destinations	(Borrajo-Millán et al., 2021)	- Large amounts of unstructured data or data predicted from surveys.
Using Social Media in Tourist Sentiment Analysis: A Case Study of Andalusia during the Covid-19 Pandemic	(Flores-Ruiz et al., 2021)	- Andalusian Tourism Situation Survey (ECTA) and Twitter data.
Social media data-based sentiment analysis of tourists' air quality perceptions	(Tao et al., 2019)	- 27,500 Sina Weibo comments regarding air quality of 195 tourist destinations in China .

Predicting Geo-Located Food Based Sentiment Analytics using Twitter for Healthy Food Consumption across India	(Saxena et al., 2017)	- Geolocated Twitter data.
Sustainability analysis on Urban Mobility based on Social Media Content	(Serna et al., 2017)	- Urban mobility information from Social Media data.
A review of natural language processing techniques for opinion mining systems	(Sun et al., 2017)	- Movie review dataset.
Visual Sentiment Analysis on Twitter Data Streams	(Hao et al., 2011)	- Streams, Movie and Amusement park Twitter data.

Sentiment analysis has been applied in many different areas, showing how useful it is for understanding public opinions and behaviors. For example, researchers have used sentiment analysis to study travel preferences before and during the COVID-19 pandemic by analyzing over a million Reddit posts (Hardt & Gluckstadt, 2024). Similarly, public opinion on the Thailand-China high-speed train projects was examined through YouTube reviews (Nokkaew et al., 2024).

In the fields of multimedia and tourism, sentiment analysis has also shown its worth. Researchers have explored feelings of ancestry and cultural heritage through YouTube videos (Hartmann et al., 2023), while multimodal sentiment analysis combined historical and emotional factors from various streaming platforms to provide a comprehensive view of public sentiment (Gandhi et al., 2023). In tourism, sentiment analysis helped understand visitor perspectives in Andalusia during the COVID-19 pandemic by combining survey data with Twitter data (Flores-Ruiz et al., 2021). It was also used to assess tourists' air quality perceptions in China through Sina Weibo comments (Tao et al., 2019). These examples highlight the broad impact and adaptability of sentiment analysis across different fields

2.2. APPLICATION OF SENTIMENT ANALYSIS IN URBAN PLANNING

Integrating sentiment analysis into urban planning represents a significant opportunity for transformation. Urban planners are increasingly using sentiment analysis to guide their decisions and policies, giving them the possibility to optimize transportation systems, improve infrastructure, and create environments that are more welcoming to visitors.

Research in this area has demonstrated the utility of sentiment analysis in urban planning. For instance, sentiment analysis has been employed to evaluate the effectiveness of public transportation systems, identifying areas needing improvement. In the context of smart cities, the analysis of sentiments from social media data has provided insights into citizens' experiences and opinions about urban services (Serna et al.,2017). Public perception of electric vehicles was examined on platforms like Reddit and Twitter, showcasing cross-platform analysis capabilities (Ruan & Lv, 2023). During the pandemic, Twitter sentiment analysis provided insights into public reactions in Mexico during different waves of COVID-19 (Contreras Hernández et al., 2023). Analyzing shifts in public discourse on electric mobility through Twitter data from 2012 to 2022 highlighted changing attitudes towards this topic (Balla et al., 2023). Additionally, sentiment analysis has been used to gather mobility intelligence in cities like London, Melbourne, and New York, proving its value in urban planning (Fontes et al., 2023). These studies highlight the practical applications of sentiment analysis in urban planning, highlighting its potential benefits in managing urban mobility and enhancing city life.

This study extends the application of sentiment analysis to the unique context of Lisbon's urban mobility. When analysing sentiments expressed by tourists on social media, it is expected to extract insights from their experiences and perceptions. This approach follows the growing interest in understanding the interactions of tourists with urban environments (Lansley & Longley, 2016). Sentiment analysis brings a deeper understanding of tourists' emotional responses and preferences regarding different modes of transportation and overall mobility experiences in Lisbon.

2.3. ANALYTICAL MODELS

The field of sentiment analysis has seen significant advancements with the introduction of transformer-based models and their subsequent iterations, handling the complexities of language and the nuances of emotional expressions within textual content. The following subsections will explain the incorporation of Large Language Models, the Chain of Thoughts prompting technique, and pseudo-labelling into the domain of sentiment analysis, based on recent reference works that underscore their effectiveness and utility (Zhou et al., 2023).

2.3.1. Transformer-Based Models on Sentiment Analysis

Transformer-based models, exemplified by models such as BERT (Bidirectional Encoder Representations from Transformers), GPT (Generative Pre-trained Transformer) and other LLM (Large Language Models), have significantly impacted sentiment analysis by addressing and overcoming limitations of previous models.

Before the launch of transformer-based models, sentiment analysis primarily relied on recurrent (RNN) and convolutional neural networks (CNN), which processed text sequentially. This linear approach often struggled with long-range dependencies, where the sentiment of a word could be influenced by elements much earlier or later in the text.

Transformers revolutionized this by enabling parallel processing of entire sequences, allowing the model to attend to all parts of a sentence simultaneously. This so-called self-attention mechanism brought about a profound change, enabling the understanding of nuanced sentiments in longer text sequences (Zhou et al., 2023).

2.3.2. Advancements in Sentiment Analysis: From BERT to RoBERTa and Beyond

The field of sentiment analysis experienced a significant transformation with the introduction of BERT. This innovative model differentiated itself by examining the context surrounding words both prior and subsequent, a departure from earlier models like word2vec or GloVe, which treated words separately. By considering the broader context, BERT was able to achieve a deeper understanding of words within sentences, effectively capturing the nuances of emotional expression that previous methods might overlook. This advancement represented a pivotal moment in the field, enhancing the precision with which sentiments could be discerned from textual data.

Subsequent to BERT, models such as RoBERTa have further advanced the methodology by refining the data learning processes, thus achieving greater accuracy in sentiment analysis. This evolution from BERT through RoBERTa to the development of more sophisticated models like GPT marks a considerable advance in the subject. These latter models possess the capability not just to interpret but also to generate text that represents human-like expression closely. The research presented by Zhou et al. (2023) and Zhong et al. (2023) elucidates this progression, demonstrating the enhanced abilities of contemporary models to navigate the complexities of human emotion as manifested in written communication.

2.3.3. GPT: Generative Sentiment Analysis

GPT's contribution to sentiment analysis lies in its generative capabilities. By predicting the next word in a sentence and being able to generate human-like text, GPT models have a deep understanding of language, which includes a capacity to detect sentiments. This understanding goes beyond the text's surface, enabling the model to infer underlying sentiments from the way sentences are constructed and words are chosen (Zhong et al., 2023).

LLMs like GPT-3 and upcoming models like GPT-3.5 and GPT 4 represent a significant advancement in sentiment analysis. With their extensive knowledge base and advanced language comprehension, these models can perform sentiment analysis in a more comprehensive way, offering a more complete picture of sentiment. They can understand idiomatic expressions, sarcasm, and complex sentence structures that earlier models struggled with.

This evolution towards more sophisticated sentiment analysis using LLMs suggests a trajectory where the depth and breadth of sentiment understanding will continue to expand, aligning more closely with human-level language comprehension. These advancements create opportunities for research that combines these advanced technologies, offering insights with unique detail and range.

2.3.4. Hallucination in Large Language Models (LLMs)

Hallucination in Large Language Models (LLMs) occurs the model generates outputs that are plausible but factually incorrect or irrelevant to the prompt. This phenomenon has been observed in various applications of LLMs, including natural language processing tasks like sentiment analysis, machine translation, and text generation. Hallucinations occur because these models are trained on vast datasets containing diverse and sometimes contradictory information. During training, the model learns patterns and associations but may also inadvertently internalize errors and biases present in the data. Consequently, when prompted, the model may produce responses that, while grammatically correct and contextually appropriate, do not align with factual accuracy or the specific requirements of the task at hand (Bender et al., 2021).

The impact of hallucination on the reliability and trustworthiness of LLMs has been a subject of significant research interest. Studies have shown that hallucinations can undermine user trust and lead to the dissemination of misinformation, particularly in sensitive domains such as healthcare, law, and finance (Maynez et al., 2020). Various strategies have been proposed to mitigate hallucinations, including refining training datasets to exclude erroneous information, incorporating external knowledge bases to validate outputs, and developing more sophisticated model architectures that better understand context and factual consistency. Moreover, evaluating the hallucination rate, or the frequency at which these incorrect outputs occur, is crucial for assessing the practical utility of LLMs in real-world applications. Understanding and addressing hallucination is essential to advancing the development of robust and reliable LLMs that can be confidently deployed across various industries (Ji et al., 2022).

2.3.5. Chain of Thoughts Prompting (CoT)

The Chain of Thoughts (CoT) prompting technique enables a sequential analytical process within language models, similar to human problem-solving strategies for complicated puzzles. This method avoids directly arriving at conclusions; instead, it breaks downs the prompt into manageable segments, addressing each sequentially. The example in Figure 1 compares the results obtained when performing Zero-Shot Prompt (Figure 2.1), or regular prompting techniques, with the Cot prompting (Figure 2.2).

As explained by Yang et al. (2024), this approach is exceptionally advantageous for complex tasks, such as interpreting sentiments or viewpoints within social media texts, where the context and linguistic nuances are important. Through following to this logical progression, language models are capable of generating responses that are not only coherent but also more comprehensible.

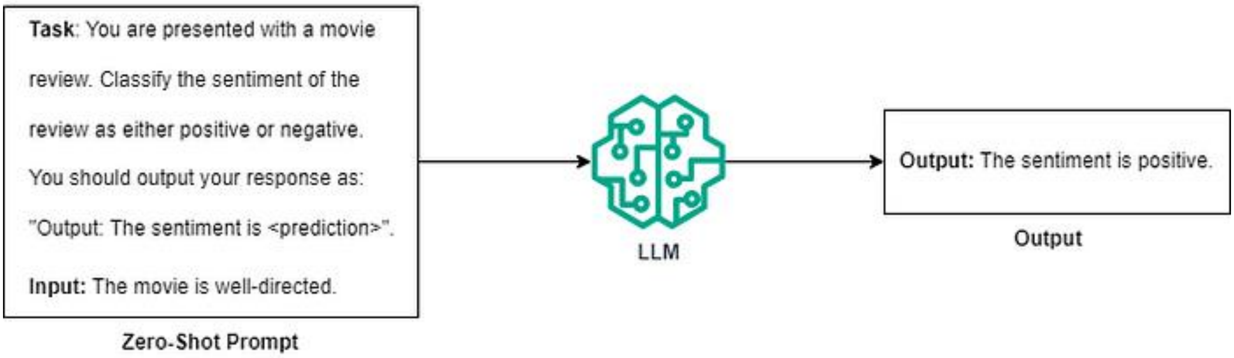

Figure 2.1 – Zero-Shot Prompt Application. Retrieved from https://tinyurl.com/4j22tfms in February 2024.

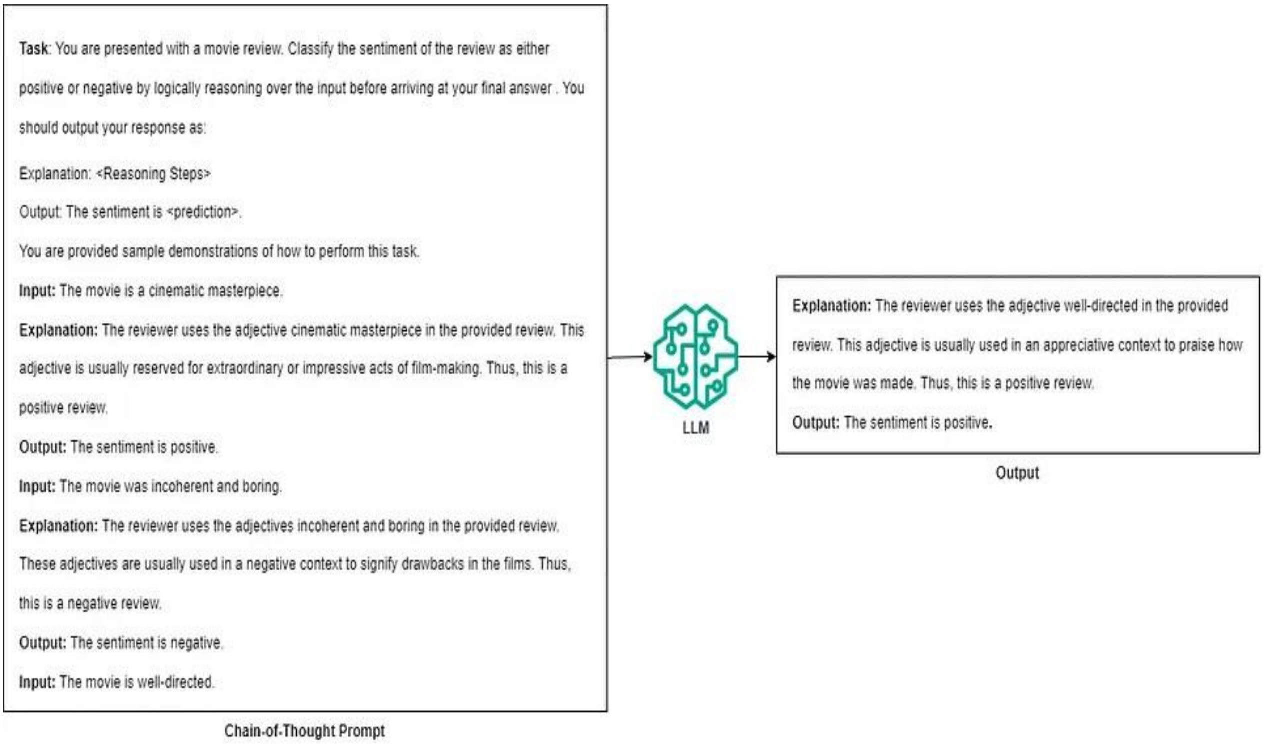

Figure 2.2 – Chain of Thoughts Application. Retrieved from https://tinyurl.com/4j22tfms in February 2024.

The Chain of Thoughts (CoT) method is notable for its ability to explain the logic underlying its responses. When employed for tasks such as identifying the emotional tone of a tweet, the CoT method starts by examining the lexical elements, subsequently introducing the contextual scenario, and lastly synthesizing these observations to predict the sentiment.

This detailed process of CoT was validated by empirical evidence, as presented by (Yang et al., 2024), for instance, improving results accuracy by 4% when compared with Zero-Shot Prompt on ChatGPT 3.5. The CoT method has shown better efficacy in addressing complex analytical challenges faced by distinct LLMs by essentially enabling the model to articulate its cognitive process, consequently enhancing the confidence in its capabilities when related to mathematics, commonsense, as sentiment analysis, and symbolic reasoning tasks as experimented by Fei et al. (2023), Wang et al. (2023) and Wei et al. (2022), as detailed in Figure 2.3 .

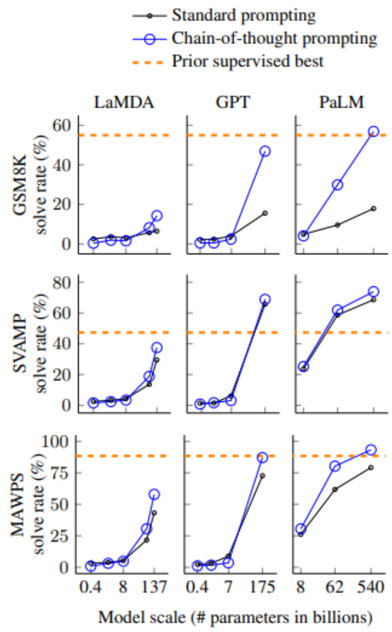

Figure 4: Chain-of-thought prompting enables large language models to solve challenging math problems. Notably, chain-of-thought reasoning is an emergent ability of increasing model scale. Prior best numbers are from Cobbe et al. (2021) for GSM8K, Jie et al. (2022) for SVAMP, and Lan et al. (2021) for MAWPS.

Figure 2.3 – Chain of Thoughts Application on math problems (from Wei et al., 2022).

2.3.6. Pseudo-Labeling Techniques

Pseudo-labeling techniques represent a sophisticated bridge in semi-supervised learning, where they employ advanced models like GPT-3.5 to assign initial labels to unlabeled data. This approach allows for the model to benefit from an expanded dataset, effectively combining labeled and unlabeled data for enhanced learning. The implementation of pseudo-labeling in sentiment analysis has been notably demonstrated by Malik et al. (2024), particularly in their work on multi-label emotion classification of French tweets. Their findings illustrate how pseudo-labeling, when paired with the computational strength of Large Language Models, markedly improves sentiment analysis, especially under the constraint of limited labeled datasets.

Expanding on this, the research conducted by Srikanth et al. (2022) explores the domain of COVID-19 discussion on Twitter, utilizing pseudo-labeling to enrich the data annotation process. This method not only addresses the challenges posed by the expansive volume of social media data but also enhances sentiment classification accuracy by using the ability to understand and adapt to the context of models like GPT-3.5. The integration of pseudo-labeling techniques into sentiment analysis frameworks signifies a significant step forward in accurately decoding the multifaceted expressions of human emotion in large-scale datasets.

Evaluating the effectiveness of models post-pseudo-labeling is crucial. While traditional performance metrics are essential, the pseudo-labeling approach itself ensures a richer and more contextually informed dataset. Studies have shown that this enriched dataset enhances the model's ability to understand and classify sentiments accurately

2.4. IMPACT OF ADDITIONAL VARIABLES ON SENTIMENT ANALYSIS

Sentiment expressed in social media posts is influenced by a range of external factors. Geographic location, major events, weather conditions, and cultural festivities can significantly affect how tourists perceive urban mobility and their overall experiences in a city (Gandhi et al., 2023.

2.4.1. Spatiotemporal Dynamics

Temporal elements, such as time of day, day of the week, and seasonality, can play a crucial role in shaping tourists' experiences related to urban mobility. Previous studies (Balla et al., 2023, Fontes et al., 2023) have highlighted how sentiment patterns fluctuate during the day and time. For instance, rush hours and weekend travel may bring different sentiments compared to off-peak times.

In addition, seasonal variations can significantly impact urban mobility sentiments. Factors like weather conditions, holidays, and tourist flood can influence perceptions and emotions related to transportation. Research such as (Hardt & Glückstad, 2024) suggests that sentiment fluctuations are observable during different seasons or adverse events, presenting opportunities for targeted adjustments.

Moreover, the geographical location of urban areas can introduce unique dynamics to urban mobility sentiments. Different cities may exhibit varying sentiment patterns due to their distinctive characteristics, infrastructure, and cultural factors. Studies (Saxena et al., 2017, Fontes et al., 2023) have explored how sentiments can vary across cities and regions, providing insights into location-specific sentiment trends.

2.4.2. Key Words and Hashtags

The extraction of key words and hashtags from text data allows for a deeper understanding of specific aspects of urban mobility, such as traffic congestion, public transportation usage, popular destinations, and trending topics. Natural language processing (NLP) techniques enable the identification of keywords and hashtags that correlate with sentiments (Sun et al., 2017; Kim et al., 2013; Hao et al., 2011). For example, mentions of "traffic jams" or the hashtag #PublicTransportIssues may indicate frustration, while references to "scenic routes" or #EfficientMetroSystems could be related to positive experiences (Kim et al., 2013).

3. METHODOLOGY

This research adopts the CRISP-DM (Cross-Industry Standard Process for Data Mining) framework (Schröer et al., 2021) to systematically explore multilingual sentiment dynamics in Lisbon’s urban mobility, leveraging the advanced capabilities of GPT-3.5 model and the innovative Chain of Thoughts (CoT) prompting technique.

The CRISP-DM framework, Figure 3, offers a structured approach to data mining projects and consists of six phases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. For this study, special emphasis is placed on the Data Understanding and Modeling phases.

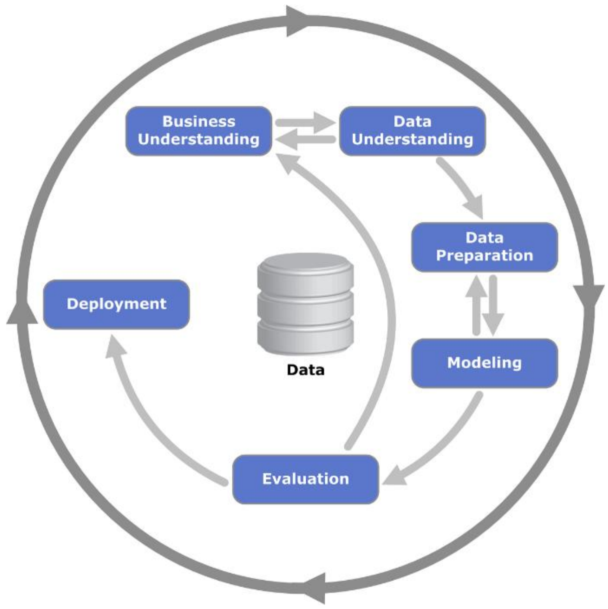

Figure 3.1 – The CRISP-DM Framework. Retrieved from https://tinyurl.com/4zawh7t6 in March 2024.

This study will utilize the Python programming language throughout the entire data science process. From data understanding and preparation to modeling, evaluation, and deployment, we will implement data analysis and graphic libraries. Additionally, GPT-3.5 and GPT 4 models using Open AI API on Python (OpenAI, 2024) will be leveraged for natural language processing tasks and Matplotlib, Seaborn and Folium Maps tools on Python for Data Visualization.

Inspired on the advanced framework from (Fontes et al., 2023), as displayed by Figure 2, this research adapts its methodology to incorporate GPT 4 on Data Preparation phase, specifically for filtering, and GPT-3.5 Turbo Model across all modelling tasks, enriching the process with the Chain of Thoughts (CoT) prompting technique as illustrated by Yang et al. (2024). This modification aims to apply the latest advanced techniques to understand the multilingual sentiment dynamics associated with Lisbon's urban mobility.

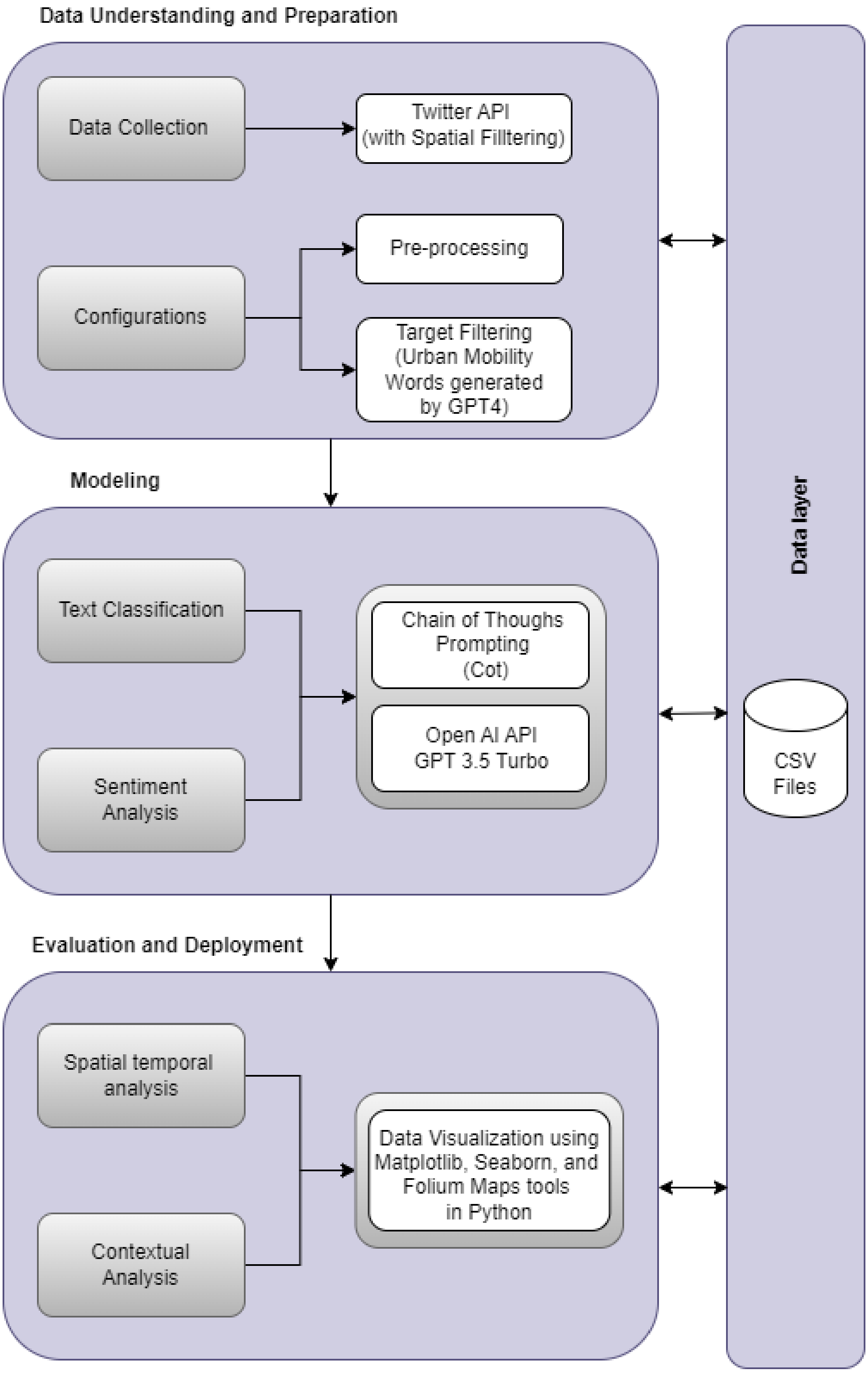

Figure 3.2 – Twitter Sentiment Analysis Framework. Based on Fontes et al., (2023).

3.1. BUSINESS UNDERSTANDING

This phase focuses on defining the project's objectives and requirements from a business perspective. This research aims to analyze sentiments expressed in multiple languages on social media, particularly Twitter, regarding urban mobility in Lisbon, and to correlate these sentiments with actual mobility patterns. Key stakeholders include urban planners, policymakers, public transport authorities, and the residents and commuters of Lisbon.

Social media platforms, especially Twitter, serve as rich sources of public opinion and sentiment. Twitter provides a platform for users to share their thoughts, opinions, and experiences regarding urban mobility issues in Lisbon. According to (Statista, 2022), Twitter is one of the most popular social media platforms in Portugal, with approximately 1.3 million users as of 2022. By analyzing data from 2022 it is possible to capture the sentiments and concerns prevalent during that period, offering valuable historical insights. These insights are crucial for urban planners and policymakers who need to understand past public reactions to inform future strategies and decisions.

Success will be determined by the ability to provide actionable insights that enable stakeholders to make informed decisions in urban planning. These insights, derived from analyzing Twitter data, will help urban planners and policymakers understand public perceptions and experiences, guiding the development of effective strategies and interventions. Ultimately, the goal is to facilitate improvements in transportation infrastructure and services in Lisbon, making the city's mobility systems more responsive to the needs and sentiments of its residents.

3.2. DATA UNDERSTANDING

Utilizing Twitter's API, tweets are collected based on Lisbon's geographical coordinates, comprehending the area between northwest corner with Latitude 39.0, Longitude -9.5 and the southeast corner with Latitude 38.5 and Longitude -8.6.

The dataset includes 33 columns related to 511,571 tweets from the year of 2022 in the Lisbon Metropolitan Area, Table 3.1, with attributes such as tweet content, reply count, retweet count, like count, language, coordinates, and user location.

A significant portion of the data contained missing values, particularly in critical columns such as 'rawContent' and coordinates. These missing values were addressed during the data preparation phase.

Table 3.1 – Lisbon's Twitter Database Columns Description

Column	Non-Null Count	Dtype
url	511571 non-null	object
date	511571 non-null	object
rawContent	495185 non-null	object
renderedContent	491738 non-null	object
id	491738 non-null	float64
user	491738 non-null	object
replyCount	491738 non-null	float64
retweetCount	491738 non-null	float64
likeCount	491738 non-null	object
quoteCount	491738 non-null	object
conversationId	491738 non-null	object
lang	491738 non-null	object
source	491737 non-null	object
sourceUrl	491736 non-null	object
sourceLabel	491735 non-null	object
links	42252 non-null	object
media	85349 non-null	object
retweetedTweet	0 non-null	float64
quotedTweet	60753 non-null	object
inReplyToTweetId	228034 non-null	object
inReplyToUser	228034 non-null	object
mentionedUsers	234030 non-null	object
coordinates	486813 non-null	object
place	486814 non-null	object
hashtags	41992 non-null	object
cashtags	247 non-null	object
card	26521 non-null	object
viewCount	19654 non-null	object
vibe	14 non-null	object
user_location	416061 non-null	object
new_date	491735 non-null	object
lat	486813 non-null	float64
lon	486813 non-null	float64

Table 3.2 presents summary statistics for the variables in the dataset. It is possible to observe, the 'rawContent' field, which contains the tweet text, has 495,185 non-null tweets from 491,738 users, indicating some users wrote more than one tweet. The hashtags field has a high number of null values, suggesting that many tweets do not include hashtags. The 'lat' and 'lon' fields, representing geographic coordinates, have fewer distinct counts and a significant number of null values, highlighting potential issues with location data accuracy. The date comprehends the period from January to December of 2022.

Table 3.2 – Lisbon's Twitter Database Summary Statistics

Variable	id	rawContent	hashtags	lat	lon	date
CountDistinct	491.736	483.375	24.131			
Count	491.738	495.185	41.992			
NullValues	19.833	16.386	469.579	24.758	24.758	16.389
Max				38,787634	-9,083400	2022-12-31 23:59:56
Min				38,663492	-9,241971	2022-01-01 00:01:15

From Figure 3.3 it is possible to observe that tweets were primarily in Portuguese, reflecting the local language, but also included languages such as English, Spanish, French, Italian and others.

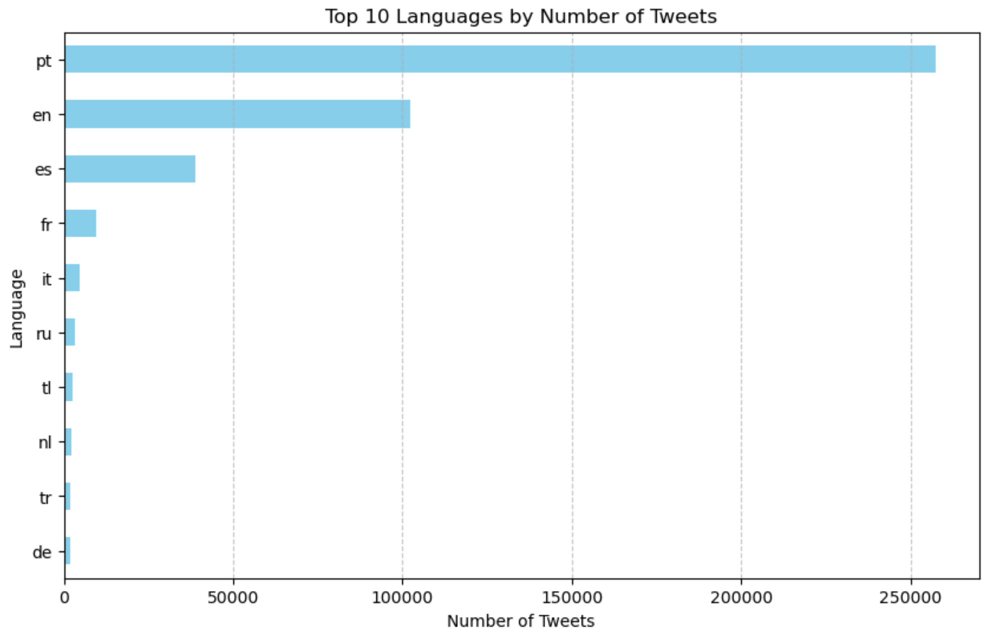

Figure 3.3 – Top 10 Languages by Number of Tweets

Engagement metrics varied widely, with most tweets having low like counts, as shown in the distribution plot on Figure 3.4.

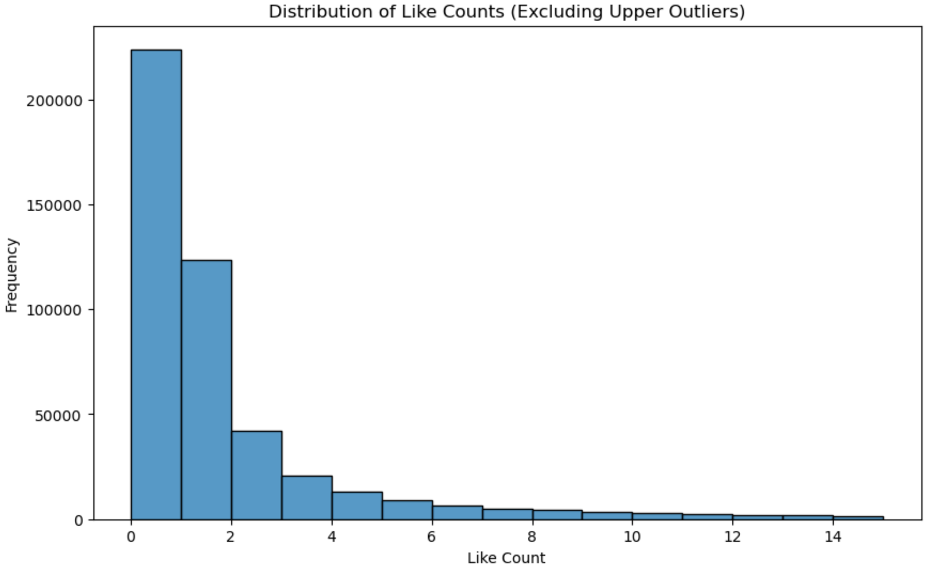

Figure 3.4 – Distribution of Like Counts

By analyzing the data geographically, Figure 3.5 shows geolocated tweets filtering to focus exclusively on Lisbon where it is possible to observe the concentration of tweets in the central area.

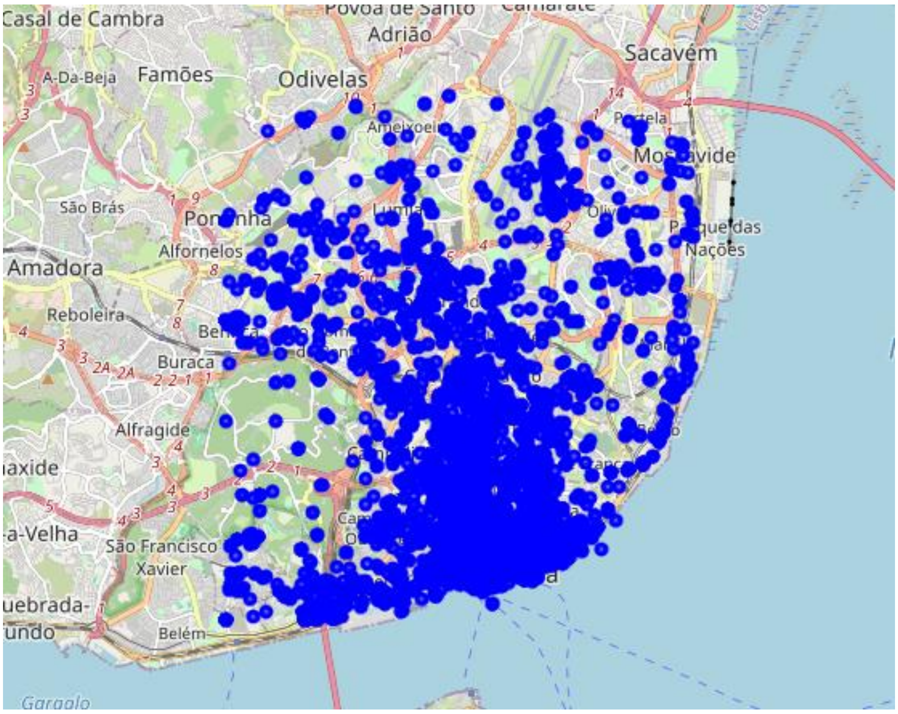

Figure 3.5 – Lisbon's Twitter Database Map

Assessing tweet volume over time and identifying peak periods of discussion about urban mobility. By performing the graphical analysis (Figure 3.6), it is possible to detect peaks of tweets, especially in April, July and November, and increase or decrease trends in specific periods.

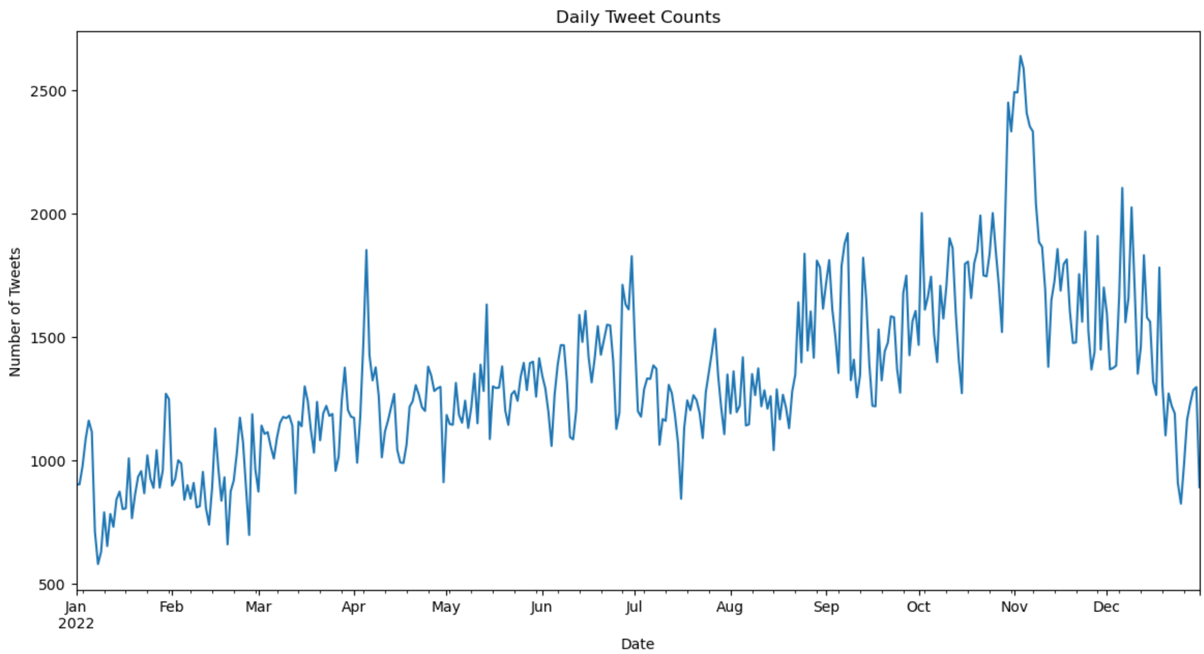

Figure 3.6 – Time Series from Lisbon's Twitter Database

Finally, based on the analysis of word clouds for hashtags and tweet content, extracting Stop Words from main languages, in Figures 3.7 and 3.8, it can be observed that main words are in Portuguese, the predominant language.

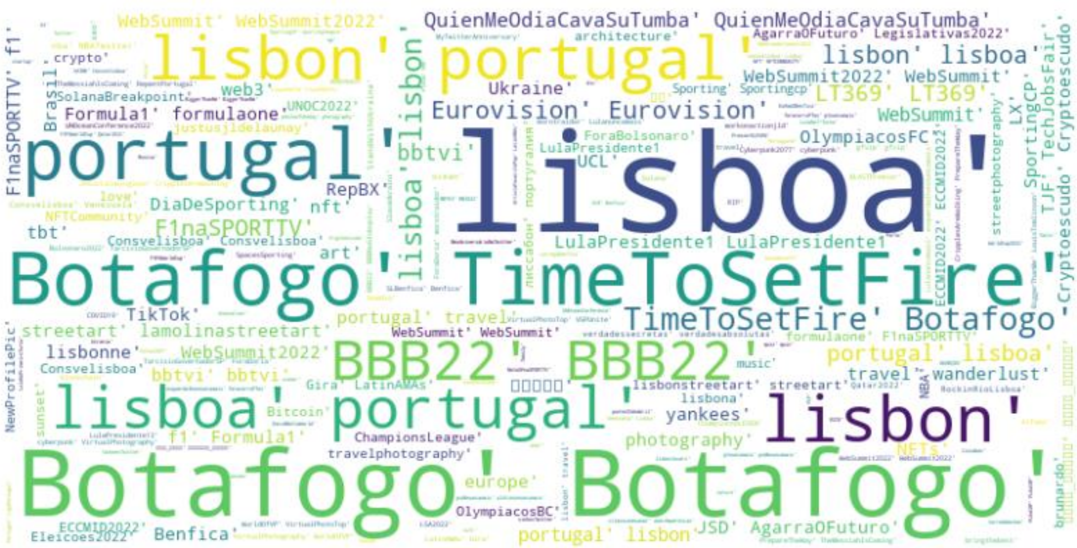

Figure 3.7 –WordCloud of Hashtags

Figure 3.8 –WordCloud of Tweets

3.3. Data Preparation

During the Data Preparation phase, data preprocessing involved several key steps to ensure the dataset was suitable for analysis. Missing values in engagement metrics were replaced with zeroes, and rows with missing critical values in rawContent and coordinates were dropped.

Variable types were converted, with likeCount and quoteCount converted to numeric types, and date converted to datetime objects for time series analysis.

Following data collection, the methodology advances to urban mobility tweets detection in order to reduce the dataset size and ensure that the sentiment analysis focuses on tweets directly related to urban mobility.

A list of keywords related to urban mobility was created using GPT-4. Tweets were then filtered to include only those relevant to urban mobility. This was necessary to focus the analysis on tweets specifically discussing urban mobility issues in Lisbon.

From the processed dataset, 486,813 tweets were identified with valid coordinates, and approximately 14,000 tweets were directly related to urban mobility after filtering.

3.4. Modeling

This phase involves developing and applying the models to analyze the data.

3.4.1. Urban Mobility Classification

After initial filtering using an extensive list of keywords, Chain-of-Thought (CoT) prompts are customized for GPT-3.5, instructing it to categorize tweets as directly related to urban mobility or not. Tweets not relevant to urban mobility are filtered out and excluded from further analysis. This step ensures that tweets containing urban mobility-related words but not directly addressing urban mobility issues are filtered, maintaining the relevance and accuracy of the sentiment analysis.

3.4.2. Sentiment Analysis

This step aims to conduct in-depth sentiment analysis to capture a spectrum of emotions related to urban mobility. The Approach will be employing GPT-3.5 for sentiment analysis. By utilizing CoT (Chain of Thoughts) techniques, the analysis defines the sentiments in positive and negative by utilizing the power of the model on understanding complex emotions and opinions.

3.4.3. Chain of Thoughts (CoT) Prompting for Labeling

The central objective of this step is to enhance the accuracy of classification and sentiment analysis models.

The following prompt was performed ensuring the identification of words related to Urban Mobility and the Sentiment related to the Tweet, leveraging CoT (Chain of Thoughts) prompting techniques, with using diverse examples and solutions, to guide GPT-3.5 in its decision-making process.

"

You are an assistant trained to identify texts related to urban mobility. Follow the provided guidelines to ensure accurate and consistent outputs. Do not skip any output. All your outputs have to be in lowercase.

Provide the following outputs based on the text:

Output 1: Detect the main language of the text. The answer should be only the language name (e.g., 'english', 'portuguese', 'spanish', 'german', 'french') or 'none' if not detectable.

Output 2: List words related to urban mobility. If there are multiple words, separate them with ';'. If there are none, put 'none'. If any word is in another language, output only its translation to English.

Output 3: Based on the text and the previous answer, determine if the tweet is directly related to Urban Mobility. The answer should be 'yes', 'no', or 'not possible to define'.

Output 4: If Output 3 is 'yes', describe the implicit opinion towards the mentioned aspect of Urban Mobility and explain why. If Output 3 is not 'yes', return 'not applicable'.

Output 5: Based on Output 4, state the sentiment polarity towards Urban Mobility as 'positive', 'negative', 'neutral', or 'not applicable'. These are the only options. If there is a mixed sentiment towards Urban Mobility but the text includes a specific sentiment about Lisbon, consider only the sentiment related to Lisbon.

Here are some examples to guide your analysis:

Text: Estou a espera que o metro do chiado abra para poder ir fazer a máquina dos Go Chill.

Output 1: portuguese

Output 2: metro
Output 3: yes
Output 4: the speaker is waiting for the metro to open, indicating reliance on public transport.
Output 5: positive

Text: Omg this might be the cutest thing in the world! A doggy school bus!!! 🥺 🥺 🥺
Output 1: english
Output 2: doggy school bus
Output 3: no
Output 4: not applicable
Output 5: not applicable

Text: eu vou sentir muita falta de andar de metro em Lisboa, o transporte público de porto alegre é um lixo
Output 1: portuguese
Output 2: metro; public transportation
Output 3: yes
Output 4: the speaker expresses fondness for lisbon's metro.
Output 5: positive

Text: I love Ice Cream.
Output 1: english
Output 2: none
Output 3: no
Output 4: not applicable
Output 5: not applicable

Text: I can't stand taking the bus in this city. It's always late, overcrowded, and dirty. Public transport needs a serious upgrade.
Output 1: english
Output 2: bus; public transport
Output 3: yes
Output 4: the speaker expresses frustration with the bus service, indicating dissatisfaction with public transport
Output 5: negative

Text: Desejos para 2023! Ruas mais seguras
Output 1: portuguese
Output 2: streets; safety
Output 3: yes
Output 4: the text implies current dissatisfaction with street safety and expresses a desire for improvement.
Output 5: negative

Now, analyze the new text and provide the following:
Output 1:

Output 2:
Output 3:
Output 4:
Output 5:
"

3.5. MODEL EVALUATION

In this study, the evaluation of the sentiment analysis model will not be conducted through traditional metrics or using labeled datasets. Instead, the robust capabilities of GPT-3.5 Turbo are relied upon for sentiment prediction. GPT-3.5 Turbo, a state-of-the-art large language model (LLM), has been trained on a diverse and extensive dataset, enabling it to perform accurate sentiment analysis across various contexts and languages.

Several studies have highlighted the effective performance of large language models similar to GPT-3.5 Turbo in sentiment analysis tasks. For instance, extensive research has shown that models like GPT-3.5 Turbo perform exceptionally well in both zero-shot prompting and few-shot prompting scenarios, maintaining high performance across different datasets and tasks (Brown et al., 2020; López Espejel et al., 2023; Ye et al., 2023). Similarly, other studies demonstrated the effectiveness of GPT-3 and GPT-3.5 and GPT4 in various NLP tasks, using CoT prompting, showing its ability to generate accurate predictions (Fei et al., 2023; López Espejel et al., 2023; Wang et al., 2023; Wei et al., 2022). Moreover, the GPT-3.5 series, particularly GPT-3.5 Turbo, has demonstrated strong capabilities in understanding and generating human-like text, further enhancing its utility in sentiment analysis.

However, it is important to note that validating the sentiment analysis results in this study is challenging due to the absence of a ground truth. Without labeled datasets or predefined benchmarks, it is not possible to quantitatively validate the model's predictions. Additionally, manual annotation of data, which is essential for creating labeled datasets, was not feasible due to time constraints and the need for multiple annotators to ensure consistency and reliability. While pseudo-labeling could have been an alternative approach, it still requires an initial set of manually labeled data to bootstrap the process, which was not possible within the timeframe of this study.

Given these constraints, the advanced natural language understanding and sentiment analysis features of GPT-3.5 Turbo are relied upon to provide reliable and sufficient predictions for this research. Nonetheless, the absence of traditional validation metrics and manual annotations remains a significant limitation, necessitating caution in interpreting the results. Future work could focus on creating labeled datasets specific to this context to enable more rigorous evaluation of model performance.

3.6. DEPLOYMENT

Based on the approach outlined by Fontes et al. (2023), this phase focuses on deploying visualization techniques to effectively present the insights obtained from sentiment analysis.

By leveraging maps and graphs, temporal, geographical, and linguistic trends are highlighted. This deployment strategy provides a robust framework for understanding the main factors driving public sentiment related to urban mobility in Lisbon.

Effective visualization techniques are essential for analyzing and presenting the findings on sentiment dynamics. To examine temporal patterns, heatmaps and time series plots are employed. These visual tools display how sentiments change throughout the day and across different seasons, helping to identify trends and fluctuations influenced by daily activities or seasonal variations.

For spatial analysis, choropleth maps are used to illustrate the distribution of sentiments across various districts in Lisbon. These maps highlight geographic differences in sentiment, providing a clear view of how public opinions are spread throughout the city.

In analyzing language-specific sentiments, bar charts are used to show the proportions of sentiments expressed in different languages. This allows for a comparative analysis of sentiment across linguistic groups, offering insights into the multilingual nature of public opinion.

Lastly, key terms analysis is conducted using word clouds and sentiment bar charts. Word clouds visually represent the most frequently used terms, emphasizing their prominence, while sentiment bar charts quantify the positive, negative, or neutral sentiments associated with these terms.

4. RESULTS

The results section presents the outcomes of the sentiment analysis conducted on the collected Twitter data, focusing on Lisbon's urban mobility. The analysis covers various dimensions, including time (both daily and seasonal), geographical location, language, and key terms related to urban mobility.

4.1. DATA FILTERING AND LLM HALLUCINATION

During the data processing phase, an initial dataset of over 500,000 tweets was filtered based on keywords related to urban mobility, generated by GPT-4 (Appendix A). After this initial filtering, only 14,110 tweets remained. These tweets were then analyzed using a Chain-of-Thought (CoT) prompting method to determine their relevance to urban mobility. The CoT prompting was used to classify whether the tweets were directly related to urban mobility. Out of these, only 4,607 tweets were classified as relevant to be performed the sentiment analysis.

Table 4.1 – Tweets Filtering Results Summary Table

	Total	% from Total
Tweets collected	511571	100%
Tweets after Preprocessing (Dealing with null values and Filtering Containing Urban Mobility related Words)	14110	3%
Tweets after Filtering Urban Mobility Directly Retated by CoT Prompting	4607	1%
Tweets with Detected Sentiment	4217	1%

Figure 4.1 highlights the hallucination rate of the sentiment classification process. However, some entries contained errors, where the prompt was asking for responses strictly in "positive," "negative," or "neutral" terms. These errors point to instances of model hallucination, where the LLM generated responses that did not align with the expected format, as shown in Table 4.2. Hallucination in LLMs occurs when the model produces outputs that are plausible but incorrect or irrelevant, which can significantly affect the quality of the analysis (Maynez et al., 2020). In this study, the observed hallucination rate was 8.5%, that indicates the proportion of instances where the model diverges from the expected response format.

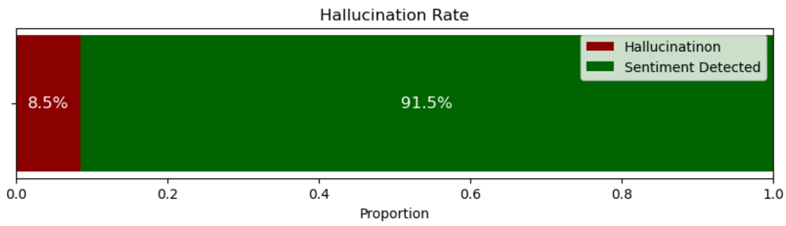

Figure 4.1 – Hallucination Rate

To address part of results that could still be detected, as shown in Table 4.2, a post-processing step was implemented, where the sentiment classification was adjusted based on the first word of the LLM output. If the first word indicated a clear sentiment (positive or negative), the tweet was classified accordingly. If the output was mixed or unclear, it was classified as neutral. This step helped mitigate the impact of hallucination.

Table 4.2 – Sentiment Detection Distribution Table

SentimentPrep	Sentiment	Number of Rows
negative	negative	1.958
negative	negative (although this is not related to urban mobility)	1
negative	negative (as the speaker almost missed their station, it implies frustration with the situation)	1
negative	negative (if related to urban mobility) or not applicable (if just an aesthetic preference)	1
negative	negative.	5
neutral	mixed (negative towards pedestrian infrastructure, positive towards	1
neutral	mixed (negative towards traffic safety but positive	1
neutral	mixed (positive towards	1
neutral	mixed (positive towards Gir	1
neutral	mixed (positive towards bike	1
neutral	mixed (positive towards exercise gained, potentially negative towards reliance	1
neutral	mixed (positive towards metro, negative towards public transport in Porto Alegre)	1
neutral	mixed (positive towards public transportation, negative towards car ownership)	1
neutral	mixed (positive towards the bikes, negative towards the app)	1
neutral	neutral	972
neutral	neutral/negative (depending on interpretation)	1
positive	positive	1.264
positive	positive.	5

The sentiment distribution (Figure 4.2) shows that a significant proportion of tweets are negative (46.6%), followed by positive (30.1%) and neutral (23.3%). This high percentage of negative sentiment indicates prevalent dissatisfaction among users regarding urban mobility in Lisbon.

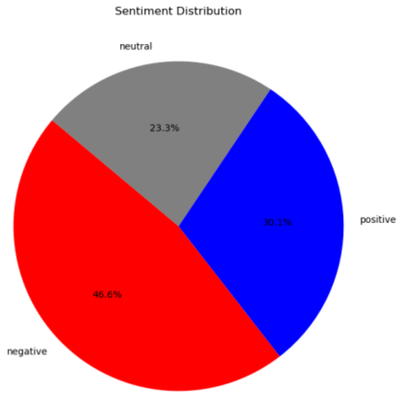

Figure 4.2 – Sentiment Distribution

4.2. Sentiment Over Time

4.2.1. Hourly Sentiment Trends

Analyzing tweet counts by the hour of the day and sentiment (Figure 4.3) reveals distinct patterns. Negative sentiments peak during the early morning (7-9 AM) and evening rush hours (5-7 PM). These peaks suggest that commuters experience significant stress during these periods, likely due to traffic congestion or inefficiencies in public transportation. Positive sentiments, although less frequent, also rise during these times, possibly reflecting positive experiences by some users amidst the general dissatisfaction. Neutral sentiments remain relatively stable throughout the day, indicating a consistent volume of indifferent opinions.

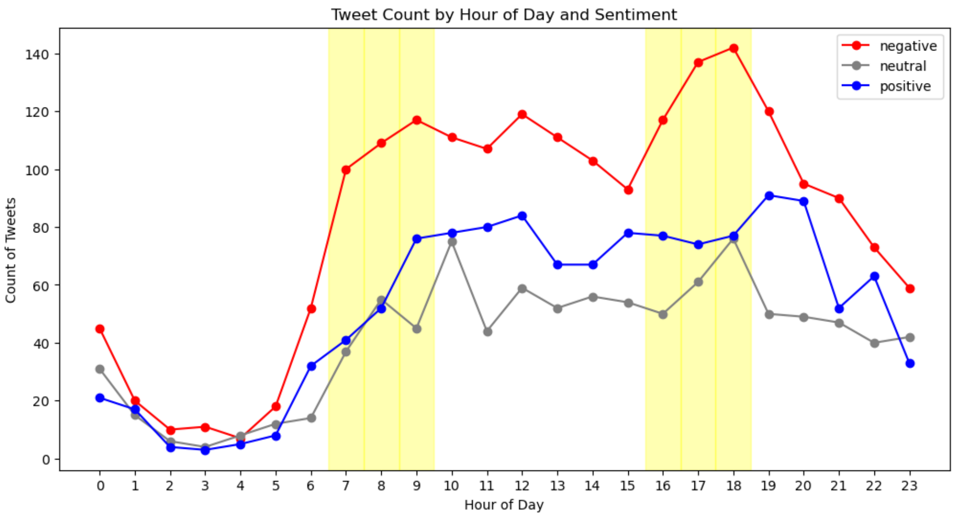

Figure 4.3 – Tweet Count by Hour of Day and Sentiment

4.2.2. Weekly Sentiment Trends

The weekly sentiment trends (Figure 4.4) indicate spikes in tweet counts around specific events and holidays. For instance, negative sentiments are notably high during major events such as the Lisbon Marathon and Rock in Rio festival, possibly due to increased traffic congestion and transportation disruptions. Positive sentiments tend to rise slightly around weekends and major holidays, suggesting a better travel experience during these times.

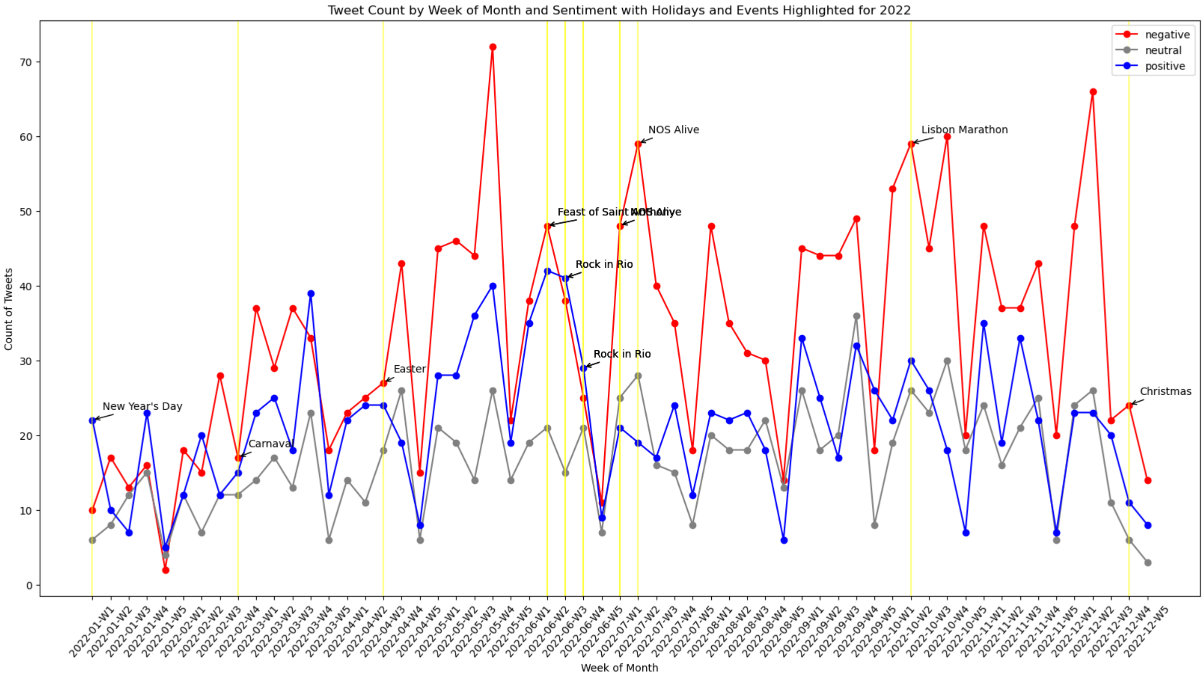

Figure 4.4 – Tweet Count by Week of Month and Sentiment

4.2.3. Seasonal Sentiment Analysis

The bar chart on Figure 4.5 displays the seasonal distribution of sentiments in tweets about Lisbon's urban mobility, showing that fall has the highest total count of sentiments, with significantly more negative sentiments. Spring and summer also have a large number of negative sentiments. Winter has the lowest total sentiment count but a higher proportion of negative sentiments, pointing to particular difficulties during this season.

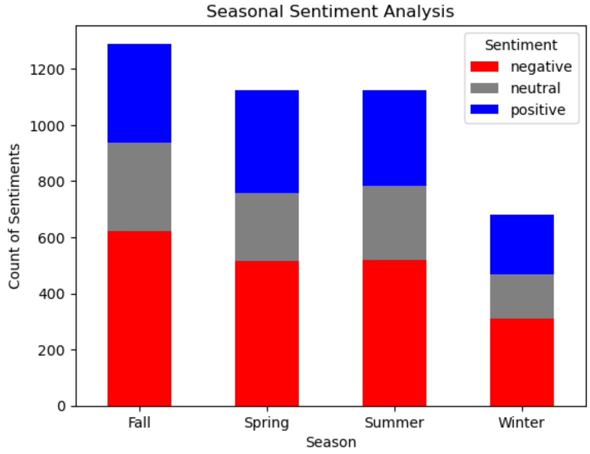

Figure 4.5 – Seasonal Sentiment Analysis

4.2.4. Heatmaps of Sentiment by Hour and Season

The heatmaps on Figure 4.6 show how positive and negative sentiments in tweets about Lisbon's urban mobility vary by hour and season. Positive sentiments are more frequent in the summer, especially in the late morning and early afternoon, with a noticeable presence in the fall during similar hours. Spring has a moderate level of positive tweets throughout the day, while winter shows fewer positive sentiments overall. This suggests that people are generally happier with their commuting experience during midday hours in the warmer months.

On the other hand, negative sentiments peak during traditional commuting hours, around 8:00 to 10:00 and 17:00 to 19:00, across all seasons. Fall and summer have the highest levels of negative sentiments, indicating more significant issues with urban mobility during these times. The high frequency of negative tweets during rush hours points to common frustrations with traffic congestion and delays in public transportation. To improve urban mobility, it's crucial to address these peak times with better traffic management and more efficient public transit services.

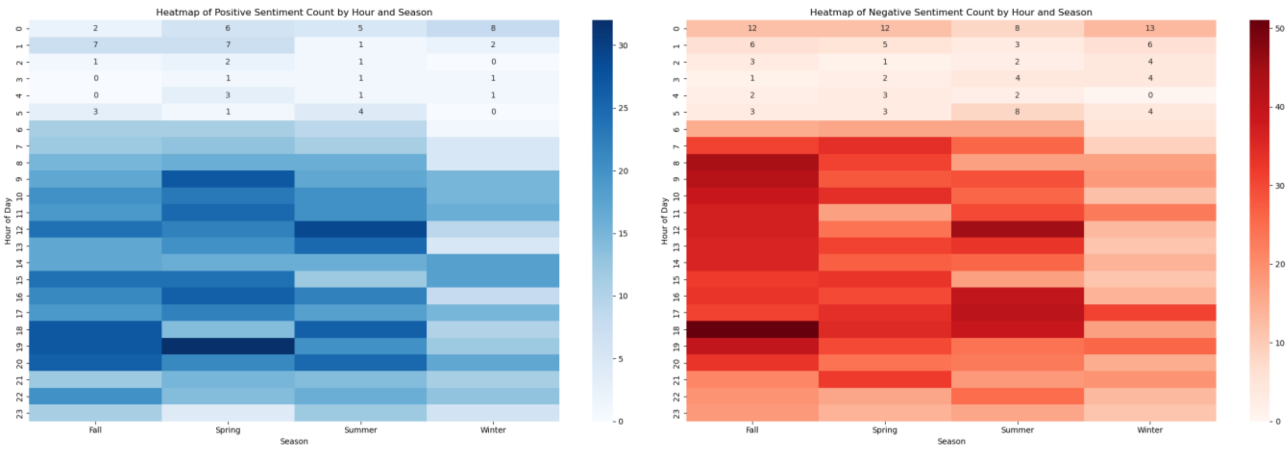

Figure 4.6 – Heatmaps of Positive and Negative Sentiments Count by Hour and Season

4.3. GEOGRAPHICAL DISTRIBUTION

The heatmap in Figure 4.7 visually represents the distribution of positive (blue) and negative (red) sentiments across different areas of Lisbon. This map highlights the spatial patterns of public sentiment towards urban mobility. Notably, the heatmap shows a higher concentration of negative sentiments (red areas) in central Lisbon and some suburban areas. Positive sentiments (blue areas) are also present but appear to be more dispersed compared to negative sentiments.

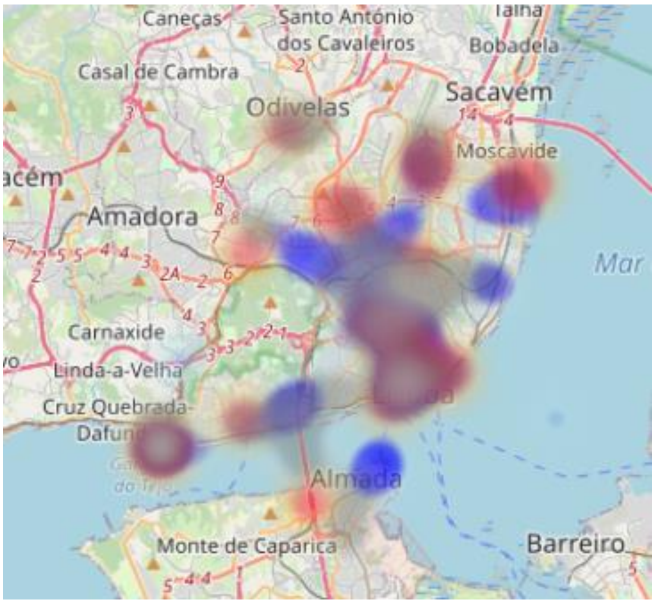

Figure 4.7 – Spatial Distribution Heatmap of Positive, Neutral and Negative Sentiments across Lisbon

Figure 4.8 provides a more detailed analysis of the sentiment distribution by plotting the exact locations of tweets with positive and negative sentiments. The size of the dots corresponds to the number of tweets from each location. A significant observation in this figure is the large cluster of negative tweets (1938) from a single location. This concentration suggests an anomaly or a specific issue at that particular spot, which may not accurately represent the overall sentiment of the area. The distribution shows that while there are many negative tweets, they are heavily influenced by this single location. Positive tweets are more evenly distributed across different locations, indicating more consistent sentiment across the city.

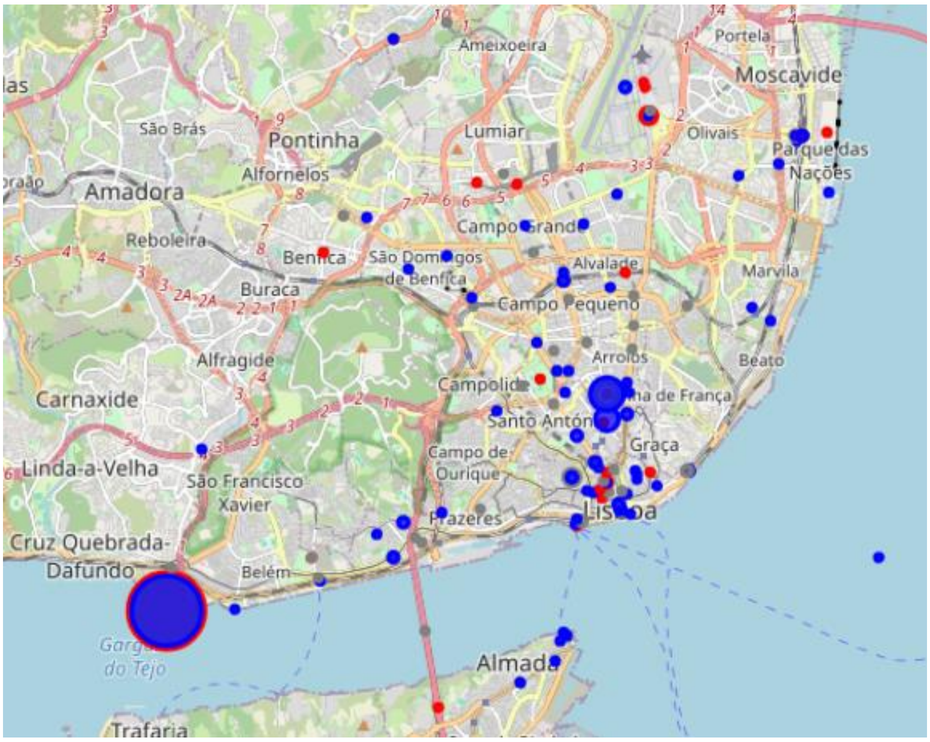

Figure 4.8 –Distribution Map of Positive, Neutral and Negative Sentiments across Lisbon

4.4. KEY TERMS SENTIMENT ANALYSIS

The analysis of key terms associated with urban mobility (Figures 4.9 and 4.10) further breaks down the top positive and negative sentiments. The top positive terms include "metro", "public transportation", "bus", "train", and "bike lane" These terms reflect aspects of the transportation system that users find beneficial. On the other hand, the top negative terms such as "metro", "bus,", "public transportation", "traffic", and "uber" highlight common pain points.

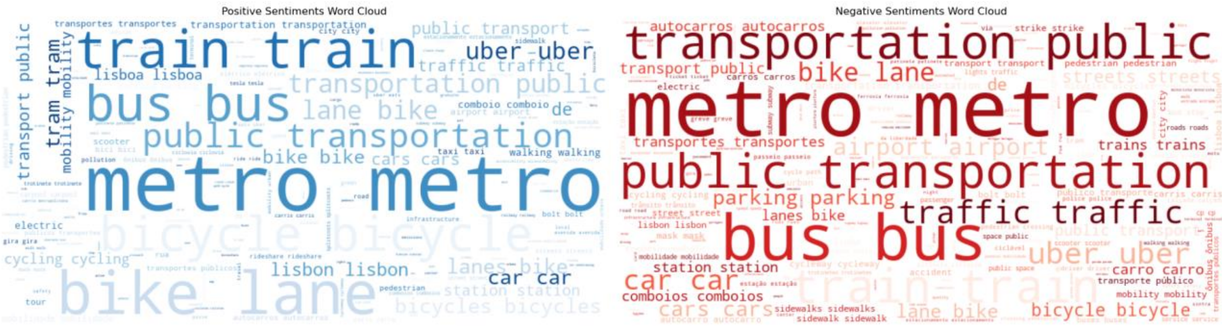

Figure 4.9 – Positive and Negative Sentiments Key Words Word Clouds

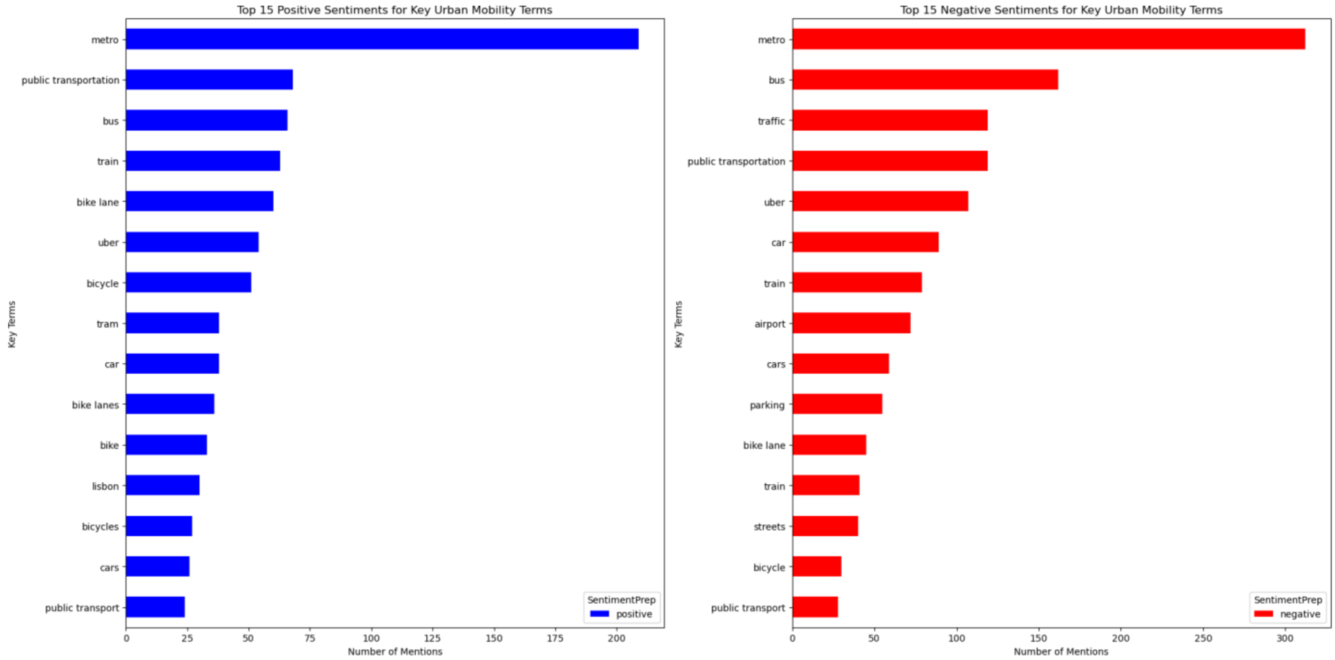

Figure 4.10 –Top 15 Key Words for Positive and Negative Sentiments

4.5. Sentiment Distribution by Language

The sentiment distribution by language (Figure 4.11) reveals that the vast majority of tweets related to Urban Mobility are in Portuguese, followed by English and Spanish, presenting a similar distribution from the initial dataset. Notably, negative sentiments are present across all languages, with Portuguese tweets showing the highest volume of negative sentiment, while as English tweets show a slightly higher positive proportion. Additionally, other languages such as French, Catalan, Dutch, Russian, and Italian exhibit smaller volumes of tweets, with varying sentiment distributions.

This analysis underscores the importance of considering linguistic diversity in sentiment analysis and highlights a common dissatisfaction with urban mobility across different language groups.

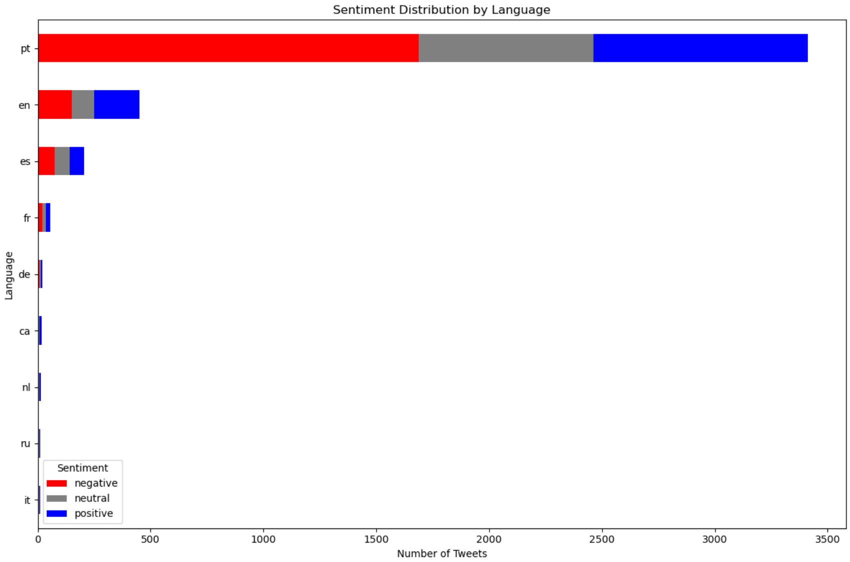

Figure 4.11– Sentiment Distribution by Language

5. DISCUSSION

The findings from the sentiment analysis of tweets regarding Lisbon's urban mobility provide a multifaceted view of public opinion, revealing several critical insights and areas for improvement, as detailed in Table 5.1. Additionally, these findings also highlight the challenges and considerations when working with social media data and large language models.

Table 5.1 – Main Findings

Finding	Description
Dataset Refinement	Only 1 % of tweets where relevant, addressing the challenge of refining large datasets.
LLM Hallucination Issue	Addressed incorrect outputs, emphasizing the need for better LLM training and prompt refinement.
Geographic Analysis	Identified hotspots of negative sentiment; potential issues with location tagging noted.
Key Terms Analysis	Positive terms: "bike lane", "metro", "train", "bus". Negative terms: "metro", "bus", "traffic".
Disparity in User Experiences	"Metro" and "bus" found in both positive and negative contexts, indicating varied experiences.
Sentiment Distribution by Language	Portuguese tweets: >50% negative. English tweets: balanced. Spanish tweets: mixed sentiments.
Common Challenges Across Languages	Negative sentiment across languages indicates shared issues in urban mobility.
Sentiment Over Time	Negative sentiments peak during rush hours, and major events cause spikes in negative sentiment.
Seasonal Variations	Fall has the highest count of negative sentiments, while summer has more positive sentiments during midday.

Firstly, the initial filtering of over 500,000 tweets to a final dataset of 4,600 relevant tweets, around 1 %, demonstrates the challenge of refining large datasets for specific analysis. The issue of LLM hallucination, where the model generates plausible but incorrect or irrelevant outputs, was addressed by adjusting the sentiment classification based on the first word of the LLM output. This step improved the accuracy of the sentiment classification but also underscores the need for continuous improvement in LLM prompting and training techniques.

Hourly, weekly, and seasonal sentiment trends show distinct patterns. Negative sentiments peak during morning and evening rush hours, indicating commuter stress. Weekly trends reveal spikes in negative sentiment around major events and slight rises in positive sentiment during weekends and holidays. Seasonally, fall has the highest count of negative sentiments, followed by spring and summer, while winter shows fewer but more negative sentiments. Heatmaps indicate more positive sentiments in summer and higher negative sentiments during commuting hours across all seasons. Thus showing the necessity of addressing peak times with better traffic management and efficient public transit.

Moreover, the geographic analysis reveals significant concentrations of negative sentiment in specific locations. While these findings highlight potential hotspots for targeted interventions, the anomaly of a high concentration of negative tweets from a single location suggests potential issues with location tagging or a disproportionate number of tweets from that area. This calls for a more nuanced approach in future analyses to ensure that the data accurately represents city-wide sentiment.

Furthermore, the word cloud and key terms analysis offer a detailed look into the aspects of urban mobility that generate positive or negative feedback. Positive terms like "bike lane", "metro", "train", and "bus" suggest that infrastructure improvements in these areas are well-received. In contrast, negative terms such as "metro," "bus," "public transportation", "traffic", and "uber" highlight persistent pain points. This dual insight helps in understanding what aspects of the transportation system are appreciated and what needs immediate attention.

Interestingly, some terms, such as "metro" and "bus," appear in both positive and negative contexts. This indicates a disparity in user experiences and perceptions, which could be more deeply explored in future work. Investigating the underlying reasons for this disparity could provide valuable insights into specific issues and areas for improvement within the urban mobility system.

Finally, the sentiment distribution by language reveals significant insights into the multilingual landscape of urban mobility sentiment. Portuguese tweets dominate the dataset, showing a diverse sentiment distribution with a substantial portion of negative sentiment which exceeds 50%, compared to approximately 25% positive sentiment. This contrast suggests significant dissatisfaction among Portuguese speakers regarding urban mobility. English tweets exhibit a more balanced distribution, with a slightly prevalence of positive sentiment compared to neutral and negative sentiments. This indicates that English speakers might have a relatively better perception of urban mobility, although negative and neutral sentiments are still present. Spanish tweets, while fewer in number, display a mix of sentiments, highlighting both positive and negative experiences.

The presence of negative sentiment across these languages indicates that there are common challenges in urban mobility that need to be addressed, irrespective of the language spoken. Additionally, other languages exhibit smaller volumes of tweets, with varying sentiment distributions. These findings highlight the necessity of considering linguistic diversity in sentiment analysis to capture a complete perspective of public opinion.

6. CONCLUSIONS AND FUTURE WORKS

This study has successfully applied advanced NLP techniques, including GPT-3.5 and Chain of Thoughts (CoT) prompting, to conduct a comprehensive sentiment analysis of urban mobility in Lisbon. The findings reveal significant seasonal and daily variations in sentiment, geographical disparities, and differences across languages. The insights gained from this analysis provide a robust foundation for policymakers and urban planners to develop targeted interventions aimed at improving public sentiment towards urban mobility.

However, several limitations need to be acknowledged. Firstly, the analysis was based only on data from 2022, which may not capture longer-term trends or changes in sentiment over different years. Secondly, the negative tweets were predominantly located in one specific area, possibly due to issues with geolocation accuracy. This raises concerns about the representativeness of the geographical sentiment distribution. Thirdly, the study did not include manually labeled tweets due to time constraints and the need for multiple annotators to ensure consistency. This limitation affects the ability to validate the sentiment analysis results against a ground truth, which is crucial for establishing the accuracy and reliability of the model's predictions.

The study also emphasizes the challenges associated with LLM hallucination and the importance of refining analytical techniques to ensure accurate and reliable outputs. Despite these limitations, the integration of advanced language models and innovative prompting techniques demonstrates the potential for sophisticated sentiment analysis frameworks to contribute meaningfully to urban mobility planning and policy development. By understanding and addressing the public's concerns, authorities can make informed decisions that enhance the urban mobility experience for residents and visitors alike.

6.1. FUTURE WORKS

Future work should aim to enhance the accuracy and depth of sentiment analysis by implementing more advanced language models and varied prompting strategies. To begin with, beyond GPT-3.5 Turbo, exploring LLM models like GPT-4 can provide deeper insights and more accurate sentiment analysis. Each model's unique strengths in natural language understanding and generation should be explored.

Secondly, incorporating additional data sources, such as real-time traffic data, public transportation usage statistics, weather conditions, and event schedules, can enrich the analysis. Combining these data sources with social media sentiments will provide a more comprehensive view of urban mobility challenges and public perceptions.

Moreover, continuous improvement of LLM prompting techniques and training can help mitigate hallucination issues. Implementing feedback loops and rigorous validation processes will ensure more accurate and reliable outputs. This can involve refining sentiment classification based on contextual understanding and improving the robustness of the models.

Additionally, future work should involve creating a labeled dataset to enable manual validation of sentiment analysis results. This approach requires multiple annotators to ensure consistency and reliability. Pseudo-labeling could also be employed, but an initial set of manually labeled data is needed to bootstrap the process. Addressing this will significantly enhance the robustness of the sentiment analysis framework.

Implementing an interactive dashboard could further improve the analysis by allowing for better filtering and visualization of different contexts. Such a tool would enable stakeholders to dynamically explore sentiment data, adjust parameters, and gain deeper insights into specific issues related to urban mobility.

Finally, future research could validate the findings by replicating the study in different cities or contexts. This will help in understanding the generalizability of the results and refining the methodologies for broader application.

By addressing these areas, future work can build upon the current research to provide a more robust framework for sentiment analysis in urban mobility and other contexts. This will not only validate the relationship between sentiment and various variables but also support data-driven decision-making and public policy.

BIBLIOGRAPHICAL REFERENCES

Ahmed, M. S., Aurpa, T. T., & Anwar, M. M. (2021). Detecting sentiment dynamics and clusters of Twitter users for trending topics in COVID-19 pandemic. *PLoS ONE, 16*(8 August). https://doi.org/10.1371/journal.pone.0253300

Balla, S. N., Pani, A., Sahu, P. K., & González-Feliu, J. (2023). Examining shifts in public discussion on electric mobility adoption through Twitter data. *Transportation Research Part D: Transport and Environment, 121*. https://doi.org/10.1016/j.trd.2023.103843

Borrajo-Millán, F., Alonso-Almeida, M. del M., Escat-Cortes, M., & Yi, L. (2021). Sentiment analysis to measure quality and build sustainability in tourism destinations. *Sustainability (Switzerland), 13*(11). https://doi.org/10.3390/su13116015

Braig, N., Benz, A., Voth, S., Breitenbach, J., & Buettner, R. (2023). Machine learning techniques for sentiment analysis of COVID-19-related Twitter data. *IEEE Access, 11*, 14778-14803. https://doi.org/10.1109/ACCESS.2023.3242234

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems, 2020-December*.

Contreras Hernández, S., Tzili Cruz, M. P., Espínola Sánchez, J. M., & Pérez Tzili, A. (2023). Deep learning model for COVID-19 sentiment analysis on Twitter. *New Generation Computing, 41*(2), 189-212. https://doi.org/10.1007/s00354-023-00209-2

Dr. Md Lutful Islam, Momin Mustufa, Sahil Shaikh, Uzma Shaikh, & Zaiba Kazi. (2023). Twitter sentiment analysis. *International Journal of Advanced Research in Science, Communication and Technology, 476*-481. https://doi.org/10.48175/ijarsct-9448

Fei, H., Li, B., Liu, Q., Bing, L., Li, F., & Chua, T. S. (2023). Reasoning implicit sentiment with chain-of-thought prompting. *Proceedings of the Annual Meeting of the Association for Computational Linguistics, 2*. https://doi.org/10.18653/v1/2023.acl-short.101

Flores-Ruiz, D., Elizondo-Salto, A., & Barroso-González, M. D. L. O. (2021). Using social media in tourist sentiment analysis: A case study of Andalusia during the Covid-19 pandemic. *Sustainability (Switzerland), 13*(7). https://doi.org/10.3390/su13073836

Fontes, T., Murcos, F., Carneiro, E., Ribeiro, J., & Rossetti, R. J. F. (2023). Leveraging social media as a source of mobility intelligence: An NLP-based approach. *IEEE Open Journal of Intelligent Transportation Systems, 4*, 663-681. https://doi.org/10.1109/OJITS.2023.3308210

Gandhi, A., Adhvaryu, K., Poria, S., Cambria, E., & Hussain, A. (2023). Multimodal sentiment analysis: A systematic review of history, datasets, multimodal fusion methods, applications, challenges and future directions. *Information Fusion, 91*, 424-444. https://doi.org/10.1016/j.inffus.2022.09.025

Gong, X., Ying, W., Zhong, S., & Gong, S. (2022). Text sentiment analysis based on Transformer and augmentation. *Frontiers in Psychology, 13*. https://doi.org/10.3389/fpsyg.2022.906061

Habib, M. A., & Anik, M. A. H. (2023). Impacts of COVID-19 on transport modes and mobility behavior: Analysis of public discussion in Twitter. *Transportation Research Record, 2677*(4), 65-78. https://doi.org/10.1177/03611981211029926

Hamzah, M., & Vu, T. T. (2018). A taxonomy of Twitter data analytics techniques. *Proceedings of the 32nd International Business Information Management Association Conference, IBIMA 2018 - Vision 2020: Sustainable Economic Development and Application of Innovation Management from Regional Expansion to Global Growth*.

Hao, M., Rohrdantz, C., Janetzko, H., Dayal, U., Keim, D. A., Haug, L. E., & Hsu, M. C. (2011). Visual sentiment analysis on Twitter data streams. *VAST 2011 - IEEE Conference on Visual Analytics Science and Technology 2011, Proceedings*. https://doi.org/10.1109/VAST.2011.6102472

Hardt, D., & Glückstad, F. K. (2024). A social media analysis of travel preferences and attitudes, before and during Covid-19. *Tourism Management, 100*. https://doi.org/10.1016/j.tourman.2023.104821

Hartmann, J., Heitmann, M., Siebert, C., & Schamp, C. (2023). More than a Feeling: Accuracy and application of sentiment analysis. *International Journal of Research in Marketing, 40*(1), 75-87. https://doi.org/10.1016/j.ijresmar.2022.05.005

Kim, H. G., Lee, S., & Kyeong, S. (2013). Discovering hot topics using Twitter streaming data social topic detection and geographic clustering. *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2013*. https://doi.org/10.1145/2492517.2500286

Lansley, G., & Longley, P. A. (2016). The geography of Twitter topics in London. *Computers, Environment and Urban Systems, 58*. https://doi.org/10.1016/j.compenvurbsys.2016.04.002

Louro, A., da Costa, N. M., & da Costa, E. M. (2021). From livable communities to livable metropolis: Challenges for urban mobility in Lisbon metropolitan area (Portugal). *International Journal of Environmental Research and Public Health, 18*(7). https://doi.org/10.3390/ijerph18073525

López Espejel, J., Ettifouri, E. H., Yahaya Alassan, M. S., Chouham, E. M., & Dahhane, W. (2023). GPT-3.5, GPT-4, or BARD? Evaluating LLMs reasoning ability in zero-shot setting and performance boosting through prompts. *Natural Language Processing Journal*, *5*. https://doi.org/10.1016/j.nlp.2023.100032

Malik, U., Bernard, S., Pauchet, A., Chatelain, C., Picot-Clemente, R., & Cortinovis, J. (2024). Pseudo-labeling with large language models for multi-label emotion classification of French tweets. *IEEE Access*, *11*. https://doi.org/10.1109/ACCESS.2023.0322000

Maynez, J., Narayan, S., Bohnet, B., & McDonald, R. (2020). On faithfulness and factuality in abstractive summarization. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. https://doi.org/10.18653/v1/2020.acl-main.173

Nelson, J. K., Quinn, S., Swedberg, B., Chu, W., & MacEachren, A. M. (2015). Geovisual analytics approach to exploring public political discussion on twitter. *ISPRS International Journal of Geo-Information*, *4*(1). https://doi.org/10.3390/ijgi4010337

Nokkaew, M., Nongpong, K., Yeophantong, T., Ploykitikoon, P., Arjharn, W., Siritaratiwat, A., Narkglom, S., Wongsinlatam, W., Remsungnen, T., Namvong, A., & Surawanitkun, C. (2024). Analyzing online public opinion on Thailand-China high-speed train and Laos-China railway mega-projects using advanced machine learning for sentiment analysis. *Social Network Analysis and Mining*, *14*(1). https://doi.org/10.1007/s13278-023-01168-8

OpenAI. (2024). OpenAI documentation. Retrieved February 15, 2024, from https://platform.openai.com/docs/overview

Ruan, T., & Lv, Q. (2023). Public perception of electric vehicles on Reddit and Twitter: A cross-platform analysis. *Transportation Research Interdisciplinary Perspectives*, *21*. https://doi.org/10.1016/j.trip.2023.100872

Saxena, A., Tibra, M., Caytiles, R. D., & N.Ch.S.N, I. (2017). Predicting geo-located food based sentiment analytics using Twitter for healthy food consumption across India. *International Journal of Bio-Science and Bio-Technology*, *9*(3). https://doi.org/10.14257/ijbsbt.2017.9.3.07

Schröer, C., Kruse, F., & Gómez, J. M. (2021). A systematic literature review on applying CRISP-DM process model. *Procedia Computer Science*, *181*, 526-534. https://doi.org/10.1016/j.procs.2021.01.199

Serna, A., Gerrikagoitia, J. K., Bernabé, U., & Ruiz, T. (2017). Sustainability analysis on urban mobility based on social media content. *Transportation Research Procedia*, *24*. https://doi.org/10.1016/j.trpro.2017.05.059

Srikanth, J., Damodaram, A., Teekaraman, Y., Kuppusamy, R., & Thelkar, A. R. (2022). Sentiment analysis on COVID-19 Twitter data streams using deep belief neural networks. *Computational Intelligence and Neuroscience*, *2022*. https://doi.org/10.1155/2022/8898100

Statista. (2022). Number of Twitter users in Portugal from 2019 to 2028. Retrieved June 20, 2024, from https://www.statista.com/forecasts/1143785/twitter-users-in-portugal

Sun, S., Luo, C., & Chen, J. (2017). A review of natural language processing techniques for opinion mining systems. *Information Fusion*, *36*, 10-25. https://doi.org/10.1016/j.inffus.2016.10.004

Tan, K. L., Lee, C. P., Anbananthen, K. S. M., & Lim, K. M. (2022). RoBERTa-LSTM: A hybrid model for sentiment analysis with transformer and recurrent neural network. *IEEE Access*, *10*, 21517-21525. https://doi.org/10.1109/ACCESS.2022.3152828

Tao, Y., Zhang, F., Shi, C., & Chen, Y. (2019). Social media data-based sentiment analysis of tourists' air quality perceptions. *Sustainability (Switzerland)*, *11*(18). https://doi.org/10.3390/su11185070

Thamil Selvi, C. P., & PushpaLaksmi, R. (2023). SA-MSVM: Hybrid heuristic algorithm-based feature selection for sentiment analysis in Twitter. *Computer Systems Science and Engineering*, *44*(3), 2439-2456. https://doi.org/10.32604/csse.2023.029254

Viñán-Ludeña, M. S., & de Campos, L. M. (2022). Discovering a tourism destination with social media data: BERT-based sentiment analysis. *Journal of Hospitality and Tourism Technology*, *13*(5). https://doi.org/10.1108/JHTT-09-2021-0259

Wang, B., Min, S., Deng, X., Shen, J., Wu, Y., Zettlemoyer, L., & Sun, H. (2023). Towards understanding chain-of-thought prompting: An empirical study of what matters. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, *1*. https://doi.org/10.18653/v1/2023.acl-long.153

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E. H., Le, Q. v., & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, *35*.

Yang, J., Jin, H., Tang, R., Han, X., Feng, Q., Jiang, H., Zhong, S., Yin, B., & Hu, X. (2024). Harnessing the power of LLMs in practice: A survey on ChatGPT and beyond. *ACM Transactions on Knowledge Discovery from Data*. https://doi.org/10.1145/3649506

Ye, J. F., Chen, X. F., Xu, N. F., Zu, C. F., Shao, Z., Liu, S. F., Cui, Y. F., Zhou, Z. F., Gong, C. F., Shen, Y. F., Zhou, J. F., Chen, S., Gui, T., Zhang, Q. F., & Huang F, X. F. (2023). A comprehensive capability analysis of GPT-3 and GPT-3.5 series models. https://platform.openai.com/docs/model-index-for-researchers

Zhong, Q., Ding, L., Liu, J., Du, B., & Tao, D. (2023). Can ChatGPT understand too? A comparative study on ChatGPT and fine-tuned BERT. https://github.com/WHU-ZQH/ChatGPT-vs.-BERT

Zhou, C., Li, Q., Li, C., Yu, J., Liu, Y., Wang, G., Zhang, K., Ji, C., Yan, Q., He, L., Peng, H., Li, J., Wu, J., Liu, Z., Xie, P., Xiong, C., Pei, J., Yu, P. S., & Sun, L. (2023). A comprehensive survey on pretrained foundation models: A history from BERT to ChatGPT.

APPENDIX A

URBAN MOBILITY TERMS – generated by GPT 4

metro	conducao autonoma	expressways	puentes	trolleys
metropolitano	viagem compartilhada	lane	paso elevado	voiture
metropolitanos	viagens compartilhadas	lanes	pasos elevados	voitures
comboio	carro sem motorista	intersection	paso subterraneo	velos
comboios	carros sem motorista	intersections	pasos subterraneos	piste cyclable
onibus	mobilidade como servico	roundabout	estacionamiento	pistes cyclables
autocarro	maas	roundabouts	estacionamientos	pieton
autocarros	mobilidade partilhada	bridge	aparcamiento	pietons
eletrico	transporte coletivo	bridges	aparcamientos	transports publics
eletricos	transporte de massa	overpass	parqueo	transports en commun
carro eletrico	transporte ferroviario	overpasses	parqueos	bicyclette
carros eletricos	transporte rodoviario	parking lot	plaza de aparcamiento	bicyclettes
bicicleta	transporte de carga	parking lots	plazas de aparcamiento	cyclisme
bicicletas	transporte de passageiros	parking space	metro	trottinette
ciclovia	infraestrutura	parking spaces	metros	trottinettes
ciclovias	infraestrutura de transporte	parking garage	ubahn	taxi
pedestre	infraestrutura de mobilidade	parking garages	ubahnen	taxis
pedestres	terminal	commuter	zug	uber
transito	terminais	commuters	zuge	lyft
congestionamento	estacao	transit system	bus	carpool
engarrafamento	estacoes	transit systems	busse	partage de voiture
engarrafamentos	parada	transit network	tram	partager voiture
estacionamento	paradas	transit networks	trams	location de voiture
vaga	ponto de onibus	transportation	trolley	location de voitures
vagas	pontos de onibus	transport system	trolleys	vehicule
garagem	linha de trem	transport systems	strassenbahn	vehicules
garagens	linhas de trem	transport network	strassenbahnen	van
via	linha de onibus	transport networks	offentlich verkehr	fourgon

vias	linhas de onibus	urban mobility	offentlich verkehrsmittel	fourgons
travessia	linha de metro	infrastructure	fahrrad	marche
travessias	linhas de metro	transport infrastructure	fahrrader	marches
calcada	corrida compartilhada	mobility hub	radweg	pietonnier
calcadas	corridas compartilhadas	electric vehicle	radwege	pietonniere
transporte	boleia	electric vehicles	radfahren	traversee
transporte publico	boleias	ev	scooter	traversees
transporte urbano	pedestre	autonomous vehicle	scooters	chaussee
mobilidade	pedestres	autonomous vehicles	elektroroller	chaussees
mobilidade urbana	zona de pedestre	self driving car	elektroroller	stationnement
carro	zonas de pedestre	self driving cars	fahrgemeinschaft	stationnements
carros	estacionamento rotativo	shared mobility	mitfahrzentrale	garage
caminhao	estacionamentos rotativos	mobility as a service	taxi	garages
caminhoes	estacionamento subterraneo	sustainable transport	taxis	pont
taxi	estacionamentos subterraneos	sustainable mobility	cab	ponts
taxis	estacionamento publico	smart city	cabs	autoroute
carona	estacionamentos publicos	smart transportation	uber	autoroutes
caronas	gare	metro	lyft	route
caminhada	terminais de passageiros	metros	auto	routes
caminhadas	plataforma	subterraneo	autos	voie
caminhonete	plataformas	subterraneos	autoteilen	voies
caminhonetes	aeroporto	tren	carsharing	infrastructure
veiculo	aeroportos	trenes	autovermietung	mobilite
veiculos	transporte aereo	bus	autovermietungen	mobilite urbaine
avenida	transporte maritimo	autobus	zu fuss gehen	transit
avenidas	transportes	autobuses	gehen	systeme de transport
rua	subway	tranvia	fussganger	systemes de transport
ruas	subways	tranvias	fussgangerzonen	reseau de transport
passeio	train	trolebus	uberquerung	reseaux de transport
passeios	trains	trolebuses	uberquerungen	vehicule electrique
pista	streetcar	transporte publico	gehweg	vehicules electriques

pistas	streetcars	bicicleta	gehwege	voiture autonome
via expressa	public transport	bicicletas	parkplatz	voitures autonomes
vias expressas	public transportation	bici	parkplatze	voiture sans conducteur
faixa	bicycle	bicis	garage	voitures sans conducteur
faixas	bicycles	carril bici	garagen	mobilite partagee
cruzamento	bike lane	carriles bici	brucke	mobilite durable
cruzamentos	bike lanes	ciclovia	brucken	ville intelligente
rotatoria	cycle	ciclovias	autobahn	transport intelligent
rotatorias	cycles	scooter	autobahnen	embouteillage
pont	cycling	scooters	strasse	embouteillages
pontos	scooter	patinete	strassen	bouchon
passarela	scooters	patinetes	spur	covoiturage
passarelas	electric scooter	taxi	spuren	uber
passagem subterranea	electric scooters	taxis	infrastruktur	bolt
passagens subterraneas	ride sharing	cabina	mobilitat	free now
engarramento	ride hailing	cabinas	stadtmobilitat	kapten
pedagio	cab	uber	transit	via verde
pedagios	cabs	lyft	transitsystem	
viaduto	carpool	coche	transitsysteme	
viadutos	car sharing	coches	verkehrsnetz	
mobilidade sustentavel	car hire	carpooling	verkehrsnetze	
carro autonomo	car rental	compartir coche	elektrofahrzeug	
carros autonomos	zipcar	alquiler de coches	elektrofahrzeuge	
carro compartilhado	vanpool	vehiculo	autonomes fahrzeug	
carros compartilhados	shuttle bus	vehiculos	autonome fahrzeuge	
carro alugado	shuttle buses	camion	selbstfahrendes auto	
carros alugados	walking	camiones	selbstfahrende autos	
ridesharing	walk	peaton	geteilte mobilitat	
transporte inteligente	walks	peatones	nachhaltige mobilitat	
transporte sustentavel	pedestrian crossing	peatonal	intelligente stadt	
bicicleta compartilhada	pedestrian crossings	peatonales	intelligenter verkehr	
bicicletas compartilhadas	pedestrian zone	trafico	metro	
bike	sidewalk	congestion	metros	
bikes	sidewalks	carretera	souterrain	
moto	crosswalk	carreteras	souterrains	

motos	crosswalks	autopista	train
motocicleta	traffic jam	autopistas	trains
motocicletas	congestion	calle	bus
mototaxi	roads	calles	trams
mototaxis	expressway	puente	trolley

NOVA IMS
Information Management School

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa