NOVA IMS Information Management School

MGI

Master's degree Program in
Information Management

Urban Dynamic Indicator: A Tool for Temporal and Spatial Assessment of Mobility in Lisbon

Daniel Filipe Barreiras Santos

Project Work

presented as partial requirement for obtaining a Master's Degree in Information Management

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

Urban Dynamic Indicator: A Tool for Temporal and Spatial Assessment of Mobility in Lisbon

by

Daniel Filipe Barreiras Santos

Project Work presented as partial requirement for obtaining the Master's degree in Information Management with a specialization in Knowledge Management and Business Intelligence

Supervised by

André Figueiredo Barriguinha, PhD, NOVA Information Management School

February, 2025

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.

Daniel Santos

Lisboa, 2025

DEDICATION

This work is dedicated to my family, whose unwavering love and support have served as the bedrock of my academic endeavors. To my companions, who have consistently provided me with support and happiness. To my devoted girlfriend, whose constant encouragement, patience, and comprehension have been indispensable during this challenging time.

ACKNOWLEDGMENTS

I would like to express my sincere appreciation to all those who contributed to the success of this research. First and foremost, I want to thank my advisor, André Barriguinha, for his support, feedback, and expert guidance throughout this journey. His knowledge and expertise greatly influenced the development of this work.

I am deeply grateful to my colleagues and professors at Nova Information Management School for their valuable input and constructive discussions, which have significantly enhanced this study. Additionally, I want to extend my gratitude to Nova Cidade Urban Analytics Lab for providing the resources that facilitated my research.

I express my profound gratitude for the constant support, patience, and encouragement extended to me by my family and friends, particularly during the most challenging periods.

Finally, I am thankful to all those who have positively contributed to the successful completion of this project.

ABSTRACT

An Urban Dynamic Indicator (UDI) is an innovative composite statistic that highlights the complexities of urban activity and its interaction with environmental elements. This study analyzes urban mobility patterns in Lisbon, the capital of Portugal, driven by the urgent need for data-driven solutions to enhance urban planning and sustainability. The study aims to create and validate the UDI using high-frequency datasets, including mobility statistics, environmental sensor data, traffic reports, and meteorological information, covering the period from September 2021 to August 2022.

A methodological approach was employed, integrating factor analysis to derive latent variables and combine multiple datasets into a unified indicator. The investigation identified significant temporal trends, including sustained activity in specific parishes and notable increases throughout morning and afternoon periods. The validation of the UDI using mobility data confirmed its trustworthiness, revealing considerable positive correlations within parishes and highlighting its capacity to represent actual urban dynamics.

The study's findings underscore the UDI's potential as a strategic tool for sustainable urban planning, providing policymakers with practical information to optimize transportation, better environmental conditions, and bolster urban resilience. This research offers a fresh viewpoint to urban studies by resolving the shortcomings of current urban metrics and delivering a comprehensive temporal-spatial analysis. Future expansions may encompass more extensive datasets and predictive modeling to forecast urban trends, therefore reinforcing the UDI's contribution to the advancement of smart city efforts.

KEYWORDS

Urban Mobility; Composite Indicator; Sustainable Urban Planning; High-Frequency Data; Urban Dynamic Indicator (UDI); Geospatial Analysis

Sustainable Development Goals (SDG):

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

APD	Associação Porto Digital
CIs	Critical Infrastructures
CPI	City Prosperity Index
DSR	Design Science Research
GDP	Gross Domestic Product
ICT	Information and Communication Technologies
INE	Instituto Nacional de Estatística
IoT	Internet of Things
IPMA	Instituto Português do Mar e Atmosfera
IQR	Inter-Quartile Range
KMO	Kaiser-Meyer-Olkin
PCA	Principal Component Analysis
SDG	Sustainable Development Goals
UDI	Urban Dynamic Indicator

1. INTRODUCTION

This chapter provides an in-depth overview of the study, covering its justification, primary objectives, data sources, methodologies employed, and the general structure of the project. This work aims to present and use an Urban Dynamic Indicator (UDI) as a tool to analyze the alterations and trends in urban mobility in Lisbon.

1.1. RESEARCH CONTEXT

Urbanization has become an important topic of the 21st century, with over 50% of the global population residing in urban areas by 2009 (Alirol et al., 2011). The United Nations forecasts that this tendency will escalate, with the world urban population anticipated to nearly double, increasing from 3.3 billion in 2007 to 6.3 billion by 2050, indicating an annual growth rate of around 2.5% (Alirol et al., 2011).

In Europe, urban areas currently house almost two-thirds of the population, considerably above the worldwide average of 55%. By 2050, this percentage is anticipated to attain 70% (Suzuki et al., 2014), highlighting the continent's dependence on metropolitan centers as focal points of economic, social, and cultural engagement. These forecasts are corroborated by reports from the European Commission (2011), the United Nations (2019), and the European Union (2011).

Lisbon exemplifies these dynamics as Portugal's most heavily populated area (Pordata & INE, 2022). Despite comprising just 3.3% of the nation's land area, it accommodates 27% of the population, underscoring the strain on urban infrastructure and resources. The city is crucial to Portugal's economic development, accommodating 30% of its enterprises and employing 37% of the labor force (EURES, 2023). Moreover, Lisbon features a well-educated populace, with around 40% possessing university degrees, above the national average of 31.8% (EURES, 2023).

This concentrated growth, however, brings considerable challenges, especially in urban mobility and sustainability. The city's high population density leads to recurrent congestion in key regions, inequities in public transit access, and escalating environmental issues, including air and noise pollution (Sarroeira et al., 2023). The increasing number of commuters and the ongoing growth of suburban regions further intensify these difficulties, imposing extra demands on transportation infrastructure and urban planning (Louro et al., 2021; Lestegás, 2019).

These developments highlight the need for innovative mechanisms to monitor and govern urban dynamics effectively. Lisbon's distinctive attributes provide a good case study for the development of the Urban Dynamic Indicator (UDI), a composite metric intended to encapsulate the intricacies of urban dynamics. The UDI seeks to overcome the shortcomings of conventional urban metrics by including mobility patterns, environmental variables, and traffic data, therefore providing policymakers with actionable insights to enhance transportation planning, reduce environmental impacts, and promote sustainable urban development.

1.2. Motivation

The Urban Dynamic Indicator has been effectively used in Porto, Portugal, showcasing its capacity to assess urban dynamics through the integration of mobility, environmental, and traffic data. This program offered significant insights into everyday urban operations and facilitated data-driven decision-making in urban administration, especially during emergencies like the COVID-19 pandemic (Jardim et al., 2022).

Lisbon, the capital of Portugal and a heavily populated metropolitan center offers a chance to broaden the implementation of the UDI. The city faces challenges identical to the ones in Porto, including complex mobility patterns, high population density, and environmental concerns. This study aims to evaluate the effectiveness of the UDI in Lisbon within a diverse urban context, utilizing high-frequency data to capture temporal and spatial variations in urban dynamics.

This research aims to evaluate if the UDI can yield practical insights for urban management in Lisbon, such as its impact in Porto. The objective is to assess its adaptation to various urban environments and to investigate its potential for guiding sustainable urban management techniques. The results of this study will complement current research on urban analytics by providing a verified tool to address the complexities of contemporary urban environments (Jardim et al., 2023).

1.3. Research Focus

This thesis centers around the development and implementation of an Urban Dynamic Indicator (UDI). The purpose of this composite indicator is to offer a comprehensive examination of urban dynamics in Lisbon. It aims to capture the fluctuations in mobility, environmental and meteorological conditions, and traffic jams. The UDI plays a vital role in detecting and analyzing trends and patterns that can provide valuable insights for sustainable urban planning and policy-making. This enables the city to successfully address the needs and requirements of its expanding population.

Despite the effective implementation of this strategy by Jardim et al. (2022, 2023) in Porto, its adaptability to other urban situations has yet to be investigated. This thesis seeks to address this gap by implementing a UDI developed for one city in another, evaluating its validity in Lisbon. This offers a basis for future applications in other European capitals, enhancing the comprehension of urban dynamics.

1.4. RESEARCH GOALS

This research aims to create and apply a framework to compute an Urban Dynamic Indicator, developed by Jardim et al. (2023) to evaluate changes in Lisbon's parishes between September 1st, 2021, and August 31st, 2022. This period guarantees that the indicator values contain all variables when time series intersect, although *Chapter 3.4.1 – Temporal Reduction* has additional details.

Specifically, the study has the key goal to Develop an Urban Dynamic Indicator (UDI). This involves applying factor analysis to consolidate diverse urban variables into a cohesive composite indicator that effectively represents the day-to-day dynamics of urban regions, implementing a concept originally conceived for Porto and now applied in Lisbon. The data will undergo standardization to guarantee comparability across different factors.

1.5. DATA SOURCES

This study employs various data sources to construct the UDI. The data sources include mobility data provided by Vodafone, environmental data collected by urban sensors, meteorological data supplied by Instituto Português do Mar e Atmosfera (IPMA), and congestion data extracted from the Waze app. Collectively, these datasets offer a comprehensive perspective on the urban dynamics in Lisbon.

1.6. THESIS ORGANIZATION

This thesis is structured into five chapters, with each chapter building upon the preceding one to offer a thorough investigation of urban mobility dynamics in Lisbon. The research is centered around the creation and utilization of the Urban Dynamic Indicator.

Chapter 1 - Introduction: The initial chapter delineates the fundamental components of this study, specifying the research background, motivation, focus, objectives, and data sources. This chapter presents the justification for creating an Urban Dynamic Indicator (UDI), which functions as the primary analytical instrument for the research.

Chapter 2 - Literature Review: This chapter explores the evolution of urban indicators, highlighting their significance in comprehending and controlling urban dynamics. The analysis begins with an evaluation of traditional indicators that encompass demographic, economic, environmental, social, and infrastructural aspects, yet are criticized for their simplicity, insufficient contextualization, and restricted adaptability. The chapter emphasizes composite indicators as a comprehensive method, combining several variables into unified metrics such as the City Prosperity Index, while addressing issues like subjective weighting and data quality. Technological advancements have increased the importance of these indicators, especially in smart cities, where high-frequency and real-time data provide transformative insights for urban planning. Attention is directed towards indicators used in traffic management, public transportation, air pollution, and noise monitoring, highlighting their capacity for evidence-based decision-making and sustainable urban development. The discussion concludes with the Urban Dynamic Indicator (UDI), a composite metric designed to represent urban activity with significant spatial and temporal precision. The UDI is presented as an instrument for monitoring urban dynamics, enabling policy interventions, and promoting sustainable city management, addressing the gaps created by traditional indicators, as evidenced by two pivotal studies.

Chapter 3 – Methodology and Data: The third chapter comprehensively explains the specific methodological technique employed in this research. The study incorporates many data sources, such as Vodafone's mobility data, urban sensor's environmental data, meteorological data, and traffic congestion data from the Waze app. The chapter subsequently outlines the data preprocessing processes utilized to guarantee the integrity and compatibility of the data, including cleaning, normalization, and temporal aggregation. Subsequently, the chapter presents an overview of the methodological framework, specifically emphasizing the Design Science Research (DSR) approach. The chapter further explains the statistical techniques employed, such as factor analysis, to construct the Urban Dynamic Indicator (UDI). This chapter is essential since it offers a clear and reproducible description of the study process, guaranteeing that the outcomes may be verified and potentially utilized in various urban settings.

Chapter 4 - Results and Discussion: The fourth chapter of the study summarizes the findings, starting with the examination of the Urban Dynamic Indicator (UDI) at various spatial and temporal scales in Lisbon. The findings are thoroughly examined, with an emphasis on detecting patterns and trends in urban mobility and environmental dynamics. This chapter further assesses the UDI results by comparing them with other pertinent data sources, such as mobility data, to verify the efficacy of the indicator.

Chapter 5 - Conclusions, Contributions, Limitations, and Future Works: The main findings of the study are summarized in this final chapter, which also discusses how the Urban Dynamic Indicator contributes to the comprehension of urban dynamics in Lisbon. The potential of the UDI to inform adaptive, data-driven policy decisions that address urban mobility, and sustainability is emphasized in the discussion of its contributions to urban planning. The chapter also delineates the study's constraints, offering a critical assessment of the methodology and the scope of the findings in broader contexts. Finally, it suggests potential areas for future research, such as the integration of additional data sources, such as public transportation and environmental metrics, into the UDI and the development of predictive models to predict mobility trends and enhance urban management strategies.

2. LITERATURE REVIEW

2.1. TRADITIONAL URBAN INDICATORS

Traditional urban indicators are crucial instruments for comprehending and controlling the intricacies of urban environments (Kitchin, Lauriault, & McArdle, 2015). These indicators contain a wide range of elements, such as demographic, economic, environmental, social, and infrastructure characteristics. Integrating them into urban planning and policy-making processes promotes the building of cities that are both sustainable and livable (Kitchin, Lauriault, & McArdle, 2015).

Demographic indicators offer valuable insights into the population dynamics of urban regions. Key indicators encompass population size and growth, which quantify the total number of residents and the pace at which the population is expanding. Comprehending these patterns is essential for strategizing services and infrastructure (Rosales, 2010; Sarmento et al., 2000). Population density is a measure of the number of individuals living in a certain region, usually expressed as the number of people per hectare or square kilometer. It is used to evaluate the level of land utilization and the possible pressure on resources (Sarmento et al., 2000). Age distribution is the process of classifying the population into distinct age groups, including children (0-14 years), working-age adults (15-64 years), and the elderly (65+ years). This classification helps in the strategic planning of education, healthcare, and social services (Rosales, 2010; Klopp, 2017).

Economic indicators evaluate the economic strength and job market of urban areas. Key metrics include Gross Domestic Product (GDP) and City Product, which measure the economic production of a city, often per capita, and serve as measures of the economic well-being and productivity of urban regions (Rosales, 2010). Employment rates, which encompass both levels of employment and unemployment, are frequently analyzed by economic sectors to determine economic strengths and weaknesses (Sarmento et al., 2000; Button, K., 2002). Income levels are a measure of the average income and how it is distributed throughout the population. This is important for comprehending economic inequality and living standards (Button, K., 2002).

Environmental indicators are used to assess and track the condition and long-term viability of the urban environment. Important metrics include air quality, which assesses the amounts of pollutants such as SO_2 (sulfur dioxide), NOx (Nitrogen oxides), CO (Carbon monoxide), O_3 (ozone), and suspended particles, offering valuable information on air pollution levels and associated health effects (Sarmento et al., 2000). Water quality evaluation examines the level of purity and security of water sources, guaranteeing the provision of safe drinking water and the maintenance of a thriving ecosystem (Rosales, 2010). Garbage management assesses the effectiveness and extent of garbage collection and disposal services, which are essential for preserving cleanliness in metropolitan areas and minimizing environmental risks (Button, K., 2002).

Urban residents' well-being and quality of life are reflected by social indicators. Key metrics encompassing public health status include life expectancy, availability of healthcare facilities, and rates of child mortality, as highlighted by Rosales (2010). Education indicators, such as literacy rates, rates of school enrollment, and levels of educational achievement, are essential for socio-economic growth (Sarmento et al., 2000; Klopp, 2017). Housing indicators evaluate the standard and accessibility of housing, housing costs, and rates of homelessness, aiding in the preparation for sufficient and reasonably priced housing (Button, K., 2002).

Infrastructure indicators assess the accessibility and standard of urban services and amenities. Key indicators encompass transportation, which encompasses the assessment of the accessibility and utilization of public transportation, levels of road congestion, and average durations of commutes. These indicators are crucial for evaluating mobility and accessibility (Sarmento et al., 2000). Access to utilities such as electricity, water supply, and sewage systems is essential for evaluating the fundamental living conditions in urban settings (Rosales, 2010). Communications indicators encompass the availability of telephone and internet services, which are crucial for establishing connectedness and facilitating economic operations (Button, K., 2002).

However, traditional urban indicators, which often prioritize economic metrics like GDP, population size, and urban growth, have been subject to criticism because of notable constraints, as highlighted by Bettencourt et al. (2010). Traditional urban indicators that depend on official data or per capita measurements have been criticized for their failure to reflect the non-linear impacts of urban agglomeration and the intricacy of local dynamics (Bettencourt et al, 2010).

Furthermore, certain studies have contended that measures such as GDP are insufficient in capturing social well-being or environmental sustainability. This highlights the necessity for alternative metrics, such as the Genuine Progress Indicator, as pointed out by Bagstad and Shammin (2012). Furthermore, it is increasingly acknowledged that traditional measurements can be influenced by bias, often showing preference towards either small or large cities depending on the method of data analysis. This underscores the significance of adopting scale-adjusted or alternative indicators to gain a more comprehensive understanding of urban dynamics, as emphasized by Alves et al. (2015).

Based on these criticisms, it is evident that traditional urban indicators frequently employ an overly simplistic approach. They oversimplify and reduce sophisticated and linked urban processes into one-dimensional measurements, which fail to appropriately portray the complex and diverse features of urban life. The reductionist approach is inadequate in considering the social, cultural, and environmental variables that are essential for a thorough comprehension of cities (Astleithner & Hamedinger, 2003; Craglia et al., 2004).

Furthermore, conventional metrics frequently fail to consider the specific historical, political, and economic contexts of cities, thus disconnecting them from their distinctive backgrounds.

This approach fails to consider the unique and interconnected factors that influence urban growth, resulting in an inadequate and occasionally deceptive understanding of urban realities (Mori & Christodoulou, 2012).

Also, there is a notable problem with the extent and scope of traditional indicators in terms of geography and themes (Kitchin, R., 2015). Several of these indicators are formulated with a wide-ranging, frequently country-wide emphasis, disregarding the localized and territorial particulars that are essential for efficient urban administration and policy formulation. The absence of a "geo-localized" strategy restricts the effectiveness of these indicators in tackling urban concerns (Kitchin, R., 2015).

In addition, conventional indicators are usually intended for longitudinal research but often overlook the varying timeframes at which distinct urban activities occur. This could result in ineffective policy responses that neglect the persistent character of urban transformations and advancements (Greene, Tracey, & Cowling, 2007).

Another significant constraint is the notion that these metrics can be universally applied across various cities. This method neglects to recognize the varied paths and systems of governance that cities worldwide follow (Gruppa & Mogee, 2004). As a result, it frequently results in unsuitable comparisons and policy suggestions that do not align with the unique requirements and circumstances of each city (Gruppa & Mogee, 2004).

The conventional method of using urban indicators often faces the problem of "information overload," where a large amount of data is gathered but not efficiently used due to its complexity and the inability to interpret and apply it properly. This leads to the failure to take advantage of possibilities for making well-informed decisions and developing policies (Münnich, R., et al, 2014).

2.2. COMPOSITE URBAN INDICATORS

Composite indicators have been developed to overcome the shortcomings of traditional urban indicators by offering a more detailed and inclusive comprehension of urban dynamics (Kitchin, R. et al, 2015; SDG-Goal-11-Monitoring-Framework, 2020). Conventional indicators tend to be too reductionist, concentrating on certain characteristics of urban life such as economic productivity or population magnitude. This method does not adequately comprehend the complex and interrelated characteristics of urban environments, which encompass economic, social, and environmental aspects (Veckalne & Tambovceva, 2023).

An important progress in addressing the constraints of conventional urban indicators is the incorporation of many aspects via composite indicators. Composite indicators merge many individual indicators into a unified index, enabling the simultaneous evaluation of multiple dimensions. The City Prosperity Index (CPI) combines many factors such as productivity, infrastructure development, quality of life, equity, and environmental sustainability to offer a comprehensive assessment of urban prosperity (Kitchin, Lauriault, & McArdle, 2015). This integration facilitates a comprehensive understanding of urban life, by effectively addressing the intricate character of urban dynamics, surpassing the effectiveness of conventional indicators (Gómez-Álvarez, et al, 2018).

Weighting and normalizing procedures are important advancements in composite indicators. These strategies guarantee that many facets of urban life are adequately portrayed in the comprehensive index. Composite indicators offer a balanced and equitable evaluation of urban circumstances by giving varying weights to indicators based on their significance and normalizing data to ensure comparability. This approach aims to resolve the problem of traditional indicators being excessively simplistic and unable to fully comprehend the intricacies of urban dynamics (OECD, 2008; Gómez-Álvarez et al., 2018).

Composite indicators are also intended to be flexible to suit various urban environments, thereby overcoming the drawback of traditional indicators being detached from their specific contexts. This flexibility guarantees that the indicators remain pertinent and responsive to the specific circumstances of the local area. The MIDE Guadalajara Metropolitana initiative customizes its indicators to match the particular socio-economic and environmental circumstances of the Guadalajara metropolitan area. Context-specific composite indicators offer a more precise and significant understanding of local urban dynamics (Gómez-Álvarez et al., 2018).

Composite indicators necessitate the acquisition of dependable and top-notch data from multiple sources. Enhancing data collection, management, and integration is crucial for optimizing the efficacy of composite indicators. Implementing standardized data collection methods and improving data infrastructures can greatly minimize errors and enhance the accuracy of the indicators. Ensuring improved data quality and administration is crucial for obtaining a more thorough and dependable comprehension of urban situations (Münnich, R., et al, 2014).

Furthermore, the creation of composite indicators typically entails a participatory methodology, which requires actively involving a diverse group of stakeholders, such as local governments, communities, and specialists. The incorporation of all urban stakeholders' interests and viewpoints in the indicators guarantees inclusion, hence increasing the indicators' relevance and acceptance (Münnich, R., et al, 2014).

The progress in technology has facilitated the incorporation of real-time data into composite indicators. This connection facilitates a more flexible and prompt urban administration, empowering politicians to observe alterations and tackle problems as they emerge. Real-time dashboards and smart city efforts demonstrate how composite indicators can integrate up-to-date data to improve decision-making (Kitchin et al., 2015).

Although composite indicators offer benefits, they also encounter difficulties. An important problem arises from the absence of a globally recognized approach to assigning different indicators with appropriate weights, resulting in subjective biases in the evaluation process

(SDG-Goal-11-Monitoring-Framework, 2020). Moreover, due to the intricate nature of urban systems, it is impossible for a single composite indicator to comprehensively encompass all pertinent facets of urban sustainability. Consequently, it is necessary to consistently improve and verify these indicators to guarantee their pertinence and precision (Kitchin, R. et al, 2015).

Modern cities face growing challenges, including rising population density, climate change, and the necessity to ensure efficiency, sustainability, and social inclusion in their management practices. The difficulties stem from rapid urbanization, which forecasts that two-thirds of the global population will reside in urban areas by 2050 (United Nations, 2018). Cities depend significantly on indicators to assess, analyze, and strategize the numerous aspects of urban life. These indicators convert intricate phenomena into measurable measurements, enabling informed decision-making and fostering evidence-based urban government (Kitchin et al., 2015; Klopp & Petretta, 2017). They provide instruments for urban administrators to address systemic issues and execute remedies grounded in credible data.

Besides their technical role, urban indicators serve as a crucial link between data collecting and the development of efficient public policies. Lo Iacono-Ferreira et al. (2022) present an analytical framework synthesizing several dimensions of urban life, facilitating focused and strategic interventions. Indicators pertaining to mobility and air quality facilitate the prioritization of investments in public transportation or sustainable infrastructure. These indicators have progressively transformed from basic, singular measurements to intricate, composite models, mirroring the increasing complexity of contemporary cities and the interrelations among their diverse social, economic, and environmental facets. This development has improved their analytical abilities and highlighted the need for a unified strategy to address present urban issues.

To build these indicators, high-frequency data has become a crucial tool for understanding urban movement, enabling detailed analysis of the timing and location of human activities. Multiple studies have emphasized the utilization of high-frequency data from sources like mobile phones and Wi-Fi networks to analyze urban mobility, spatial patterns, and their consequences for urban planning and policy-making (Jardim et al., 2022; Chan, J., 2023; Ratti & Townsend, 2011). Research has demonstrated that analysis of cellular network data can uncover recurring patterns and temporal consistency in urban dynamics, hence improving our comprehension of people's movement within urban areas (Sun et al., 2011). In addition, researchers have utilized mobile phone positioning data to examine the temporal dynamics of spatial divides within urban regions. This analysis demonstrates how urban areas evolve by expanding and fragmenting over time (Zhou et al., 2016).

2.2.1. Traffic Indicators

The way cities monitor and control urban traffic has changed as a result of developments in high-frequency data and Internet of Things (IoT) technologies. According to Correia et al. (2022) and Du et al. (2021), these technologies offer a crucial foundation for real-time monitoring, traffic pattern prediction, and dynamic modifications that boost road systems' efficiency. Furthermore, this data facilitates evidence-based choices that enhance urban metrics pertaining to safety, sustainability, and mobility.

The temporal granularity required to make real-time system adjustments and short-term traffic flow predictions is provided by high-frequency data. According to Du et al. (2021) and Gao et al. (2013), neural networks combined with sophisticated algorithms like wavelets can precisely forecast congestion, enabling prompt actions. These models work especially well in situations with dynamic traffic, when things change quickly. According to Correia et al. (2022), real-time data use is crucial for ensuring higher sustainability in transportation and coordinating mobility systems with smart city objectives.

Real-time data gathering and analysis are made possible by the Internet of Things' crucial role in connecting sensors and gadgets in automobiles and urban infrastructure. According to Rathore et al. (2021), fog devices that are attached to car cameras keep an eye on traffic and identify infractions. They process data locally to increase road safety and cut down on delays. According to Mohanty et al. (2016), intelligent IoT networks allow traffic lights to be automatically adjusted, which enhances traffic flow and cuts down on wait times. Furthermore, IoT sensors track environmental indicators like CO_2 emissions, enabling cities to modify transportation policies based on verifiable data, as shown by Rathore et al. (2018).

IoT and high-frequency sensors gather temporal data, and geolocation adds vital spatial context. According to Yuan et al. (2011), GPS data from taxis functions as mobile sensors, producing geographical data in real time that can be used to spot traffic jams and provide other routes. According to Ghanim et al. (2022), combining geolocation with high-frequency data during the COVID-19 pandemic enabled cities to adapt their transportation networks to mobility limits while preserving operational effectiveness and lessening the effects of flow fluctuations. According to Camboim et al. (2019), smart city initiatives frequently use geolocation and the Internet of Things to plan public transportation routes and lessen their negative effects on the environment.

The incorporation of these technology directly affects urban metrics including safety, sustainability, and mobility. Rydzewski and Czarnul (2021) demonstrate how IoT and geolocation-based simulations can be used to forecast typical journey times and modify intersection capacity, improving traffic flow and easing congestion. According to Matheus et al. (2020), integrated dashboards enable real-time indicator visualization, fostering increased citizen participation and governance transparency. Smékalová and Kučera (2020) note that while big cities have greater access to these technology, cooperative efforts can assist smaller communities in overcoming financial obstacles to adopt intelligent solutions. According to Aleksander and Czarnul (2020), standardizing technologies and exchanging best practices might hasten the adoption of these solutions, fostering robust and effective transportation networks.

Urban traffic management is thus transformed by the combination of high-frequency data, IoT, and geolocation, as shown by Correia et al. (2022), Mohanty et al. (2016), and Rathore et al. (2018). More precise forecasts, quicker reactions, and evidence-based decisions that maximize mobility, reduce environmental effects, and enhance road safety are made possible by these technologies. The sustainable growth of smart cities depends on these developments.

2.2.2. Public Transportation

The literature on urban indicators and their influence on public transport illustrates the complexity and range of methods employed to comprehend and enhance urban mobility across different contexts. Rajput et al. (2020) emphasize the significance of sensors and data gathered by cellphones as effective instruments for assessing crowd density on buses, demonstrating that crowdsourcing-based solutions can provide accessible and scalable methods for controlling public transportation networks. This method underscores that accessible technologies can alleviate urban mobility challenges, particularly in areas with inadequate infrastructure, facilitating more accurate interventions in route and schedule design.

Wimbadi et al. (2021) examine the significance of urban experiments in low-carbon mobility transition programs, based on the 'Avoid-Shift-Improve' (ASI) paradigm. The authors demonstrate how urban indices associated with greenhouse gas emissions and transportation modes might inform the implementation of technology like low-emission vehicles and integrated public transport networks. The analysis indicates that urban experiments not only implement technical advances but also foster behavioral and institutional transformations essential for long-term sustainability.

Accessibility is a pivotal element in public transit, especially in island regions, as examined by Garau et al. (2022). The research on Sardinia demonstrates that the divergence between coastal and inland regions directly affects mobility and exacerbates socio-economic disparities. The authors advocate for the implementation of digital technology and multimodal planning as methods to enhance access to public transportation services, highlighting that accessibility indices must account for not only geographical proximity but also local socioeconomic and demographic factors.

A pertinent feature in the literature is the effect of crises, such as the COVID-19 pandemic, on public transportation systems. Jenelius and Cebecauer (2020) demonstrate that the pandemic has caused a 40% to 60% decline in public transport utilization in several locations of Sweden. The authors indicate that the loss was more pronounced among infrequent users, underscoring the susceptibility of public transport networks to sudden shifts in behavioral patterns. Conversely, Wilbur et al. (2023) determined that low-income regions in the United States experienced a diminished effect from the drop in demand, indicating these populations' reliance on public transportation for fundamental activities. This emphasizes the significance of urban indicators that reflect the correlation between mobility and social disparities.

Collectively, these studies illustrate that urban indicators are essential for comprehending and administering public transport networks, encompassing operational efficiency as well as equality and ecological concerns. Integrating technological, social, and environmental data is crucial for formulating solutions that address the present and future requirements of urban areas. Garau et al. (2022) and Wimbadi et al. (2021) demonstrate that the strategic application of indicators can enhance public transport networks and stimulate broader transformations towards more sustainable and inclusive urban environments.

2.2.3. Air Pollution Indicators

Air pollution significantly affects urban sustainability, impacting public health, quality of life, and climate (Asian Development Bank, 2001). Principal pollutants such as PM2.5, PM10, NO_2, and O_3, associated with vehicular emissions, industrial activities, and energy usage, are well-established indicators of environmental quality (Coelho et al., 2022; Yeo & Kim, 2022). Certain studies highlight local sources like road mobility, whilst others accentuate regional and transboundary influences, especially in densely populated regions (Yeo & Kim, 2022). Mitigation strategies encompass technology innovations like as IoT-based monitoring systems and simulation models (WRF-Chem), allowing targeted interventions like green transportation and advances in surface reflectivity (Jandaghian & Akbari, 2018). The scalability and efficacy of these initiatives differ, with detractors challenging their applicability across various urban morphologies (Lu et al., 2020).

Urban planning strategies that incorporate climate and air quality objectives, such as enhancing surface albedo and implementing green infrastructure, have potential in alleviating heat island impacts and enhancing air quality (Jandaghian & Akbari, 2018; Balogun et al., 2021). However, implementation issues emerge in low-resource environments, underscoring the necessity for customized strategies (Salman & Hasar, 2023). Prolonged exposure to pollutants reveals substantial health hazards, including chronic respiratory conditions associated with NO_2 and PM10, while promoting comprehensive policies that tackle both health and climate effects (Chang et al., 2022; Yeo & Kim, 2022). Synergistic interactions between pollutants, such as PM10 and CO, exacerbate urban health hazards, requiring comprehensive policy interventions.

Source apportionment, shown by the ClairCity project, elucidates pollution sources, allowing targeted policy interventions (Coelho et al., 2022). Nonetheless, difficulties in precisely measuring emissions, especially in urban-industrial environments, undermine its credibility (Ashrafi et al., 2013). Composite indicators, such as the Urban Dynamic Indicator (UDI), amalgamate pollution parameters with economic and mobility data, providing practical insights for urban management. However, data constraints and deficiencies in monitoring infrastructure remain, highlighting the necessity of using modern techniques such as machine learning and spatial analytics to enhance urban sustainability frameworks (Salman & Hasar, 2023; Balogun et al., 2021).

2.2.4. Noise Pollution Indicators

Recent work on urban noise pollution during COVID-19 lockdowns indicates an agreement on the decrease in exterior noise levels and the significance of this phenomena for urban design. Rumpler et al. (2020) indicate a drop of up to 4 dB in external noise levels in Stockholm on weekdays and weekends, ascribed to less urban traffic and human activities. This decline was analogous to periods of bank holidays, underscoring the influence of constraints on human-generated sound sources. Likewise, Lee and Jeong (2021) noted a decrease in external noise levels in London. They observed a notable rise in interior noise, especially from neighbors, including noises from televisions, music, and conversations, attributable to increased time spent at home. These alterations indicate the necessity for mitigation techniques that account for both physical elements and human perceptions.

Nonetheless, certain disparities in the methodologies of the research underscore significant intricacies about the issue of noise pollution. Rumpler et al. (2020) primarily emphasize quantitative data, prioritizing demonstrable alterations in external noise levels and explicitly correlating them with a decline in urban mobility. Conversely, Lee and Jeong (2021) investigate subjective impressions, demonstrating that the augmentation of time spent at home heightened residents' sensitivity to interior noise. This disparity indicates that alterations in the auditory environment do not consistently align with inhabitants' views, underscoring the necessity for integrated methods. A notable distinction pertains to the function of natural sounds: whilst Rumpler et al. (2020) neglect this dimension, Lee and Jeong (2021) ascertain that certain participants found birdsong irritating during the lockdown, perhaps attributable to the lack of other ambient noises in London.

Moreover, comprehensive investigations, exemplified by the ClairCity project, have shown the significance of both local and transboundary causes of noise pollution. Road travel was recognized as the primary source of NO_2 emissions, although fine particulate matter (PM10 and PM2.5) had a more significant impact from cross-border movement (Coelho et al., 2022). These findings underscore the significance of urban policies that advocate for sustainable transportation and enhancements in acoustic insulation, particularly in apartments, where exposure to internal noise was more evident. The utilization of cost-effective acoustic monitoring devices has demonstrated efficacy in evaluating urban noise conditions and quantifying the effects of public policy. The discrepancies across studies on human perceptions and natural sounds suggest that next research should amalgamate objective metrics and subjective assessments to provide a more holistic understanding of noise pollution. Insights gained from lockdowns provide essential frameworks for formulating urban plans aimed at enhancing acoustic comfort in both routine and extraordinary circumstances (Rumpler et al., 2020; Lee & Jeong, 2021; Coelho et al., 2022).

2.3. URBAN DYNAMIC INDICATOR

An Urban Dynamic Indicator (UDI) is a specialized composite indicator that aims to offer a comprehensive and frequent assessment of urban dynamics. The UDI was created to address the requirement for comprehensive and immediate monitoring of urban areas (Jardim et al., 2022). It collects data on the constant movement of individuals, resources, services, events, and information inside a city setting (Jardim et al., 2022). Due to its comprehensive nature, the UDI can represent the city's everyday activities and fluctuations in urban dynamics. This is particularly beneficial during times of crisis such as the COVID-19 pandemic (Jardim et al., 2022; García-Peña, Molina, Cabrera, & Sinoga, 2023).

The UDI serves multiple purposes. Initially, it observes urban dynamics by offering a daily assessment of urban activities, aiding in the monitoring of how various aspects interrelate and evolve over time. This provides essential information about the city's operating dynamics (Jardim et al., 2022; García-Peña, Molina, Cabrera, & Sinoga, 2023). Furthermore, it facilitates policy decisions by assisting policymakers in making well-informed choices, emphasizing areas that require intervention, and evaluating the effects of enacted policies. The UDI facilitates sustainable urban planning by proactively recognizing emerging trends and potential challenges, thereby preventing them from reaching a critical stage (Barbosa et al., 2024). Finally, it improves public communication by condensing intricate urban data into a unified and simply comprehensible index, making it accessible for public comprehension and involvement. Transparency has the potential to boost trust and collaboration between municipal authorities and inhabitants (Barbosa et al., 2024).

The use of the UDI showcases the capacity of composite indicators in urban administration, offering a framework that may be adjusted and executed in other cities to augment their ability to withstand and maintain their sustainability (Jardim et al., 2022; Jardim et al., 2023).

The necessity to establish the UDI comes from the difficulty of precisely and often monitoring the urban dynamics that affect the quality of living in cities. Throughout the COVID-19 epidemic, it became evident that conventional indicators, centered on economic factors and often computed at a national or weekly scale, were unable to reflect the swift alterations in urban activity at a local level. These phenomena encompass transportation, pollution, public service utilization, and the intensity of outdoor activities, all of which directly impact the everyday lives of residents. Consequently, it was essential to create a composite indicator capable of consolidating several urban characteristics into a singular representative value, while adequately reflecting local particularities and daily temporal dynamics (Jardim et al., 2022). The publication titled '*The Daily Urban Dynamic Indicator: Gauging the Urban Dynamic in Porto during the COVID-19 Pandemic*' introduces the UDI as a novel approach for assessing the urban dynamics of Porto throughout the pandemic (Jardim et al., 2022). This study employed a methodology that integrated indicators indicative of urban activity into a singular metric. The utilized data comprised five primary time series: traffic flow, public transportation usage, logins to bus Wi-Fi networks, NO_2 emissions, and noise levels. The data was sourced from local city sensors and monitoring systems and was modified for seasonal and calendar adjustments (Jardim et al., 2022).

The employed technique consisted of many processes, commencing with exploratory data analysis to discern patterns and deficiencies. A seven-day moving average approach was employed to address missing variables. Furthermore, seasonality and public holiday influences were modified utilizing the structural time series model introduced by Harvey and Peters, executed through the Prophet algorithm (Taylor & Letham, 2018). Subsequent to these modifications, factor analysis was employed to construct the UDI. This strategy discerned latent features that elucidate the covariance among the observed variables, yielding a singular indicator that accounts for 49% of the overall variance in the data (Jardim et al., 2022).

The research confirmed the appropriateness of factor analysis by the Kaiser-Meyer-Olkin (KMO) test, which yielded a score of 0.667, signifying that the data was adequate for this modeling approach. Bartlett's test of sphericity also validated the existence of substantial correlations among the variables, hence supporting the model's validity. The factor loadings of the variables in the UDI varied from 0.52 (NO2) to 0.76 (logins on bus Wi-Fi), indicating the relative importance of each component (Jardim et al., 2022). The study's results indicated that the UDI offered an extensive perspective on urban dynamics in Porto, facilitating the identification of the consequences of pandemic limitations on urban activities. While economic measures like the Daily Economic Indicator (DEI) anticipated a rebound by the end of 2020, the UDI revealed that urban dynamics remained below pre-pandemic levels. This mismatch underscored the significance of the UDI in capturing local variations absent in national measures (Jardim et al., 2022).

The UDI also demonstrated its use in advising public managers on certain domains requiring intervention, like urban transportation and pollution mitigation. The study determined that the UDI can enhance other high-frequency indicators, providing a robust instrument for swift diagnosis and aiding in the development of adaptive, evidence-based public policies. Consequently, the study substantially advanced the theory and practice of sustainable urban planning, illustrating the applicability of UDI in both crises and stable eras (Jardim et al., 2022).

The publication titled '*Urban Dynamic in High Spatiotemporal Resolution: The Case Study of Porto*' signifies a substantial advancement of the original UDI idea. The recent work seeks to enhance the indicator to better reflect fluctuations in urban dynamics with improved geographical and temporal accuracy. This study extends the analysis from a citywide daily aggregate viewpoint to census block levels and hourly intervals, providing a more thorough examination of urban dynamics in Porto (Jardim et al., 2023).

The methodological approach entailed the collection of high-resolution data across four primary dimensions: traffic, public transit use, soft mobility, and air quality. The data was sourced from several origins, including local sensor systems and digital platforms like the Waze application for traffic information and air quality sensors managed by the city (Jardim et al., 2023). Temporal granularity was established at hourly intervals, whilst spatial granularity used census block units (BGRI), enabling a comprehensive and contextual study.

The methodology started with data pre-processing, using spatial and temporal interpolation techniques to address missing values and standardize the varying temporal and geographical resolutions of the datasets. The Prophet model was utilized to account for seasonality, public holidays, and climatic influences (Jardim et al., 2023; Taylor & Letham, 2018). After preliminary processing, a factor analysis was conducted to formulate the improved UDI, which accounted for 34% of the total variance in the data. The decrease relative to the last study is attributable to the incorporation of additional factors and levels of granularity, which illustrate the intricacy of urban dynamics (Jardim et al., 2023).

The findings demonstrated distinct tendencies in Porto's urban dynamics. Aldoar, Cedofeita, and Bonfim had the highest levels of activity, particularly during weekday afternoons. Conversely, regions like Campanhã and Paranhos had the lowest ratings, indicative of mostly residential attributes and diminished integration with the city's economic and cultural dynamics. Moreover, traffic was recognized as the most significant variable in the UDI, accounting for 35 percent, followed by air quality at 34 percent, whereas soft mobility had a lesser influence at 9 percent (Jardim et al., 2023). This research further corroborated the enhanced UDI by juxtaposing it with mobile device location data supplied by Vodafone. The association between the indicator and mobility patterns validated its capacity to represent urban flows across various eras and locales. This illustrated the practical use of the UDI in facilitating data-driven decisions, ranging from localized interventions to long-term planning initiatives (Jardim et al., 2023).

When comparing both studies, it is evident that the first presents a broad and aggregated perspective on urban dynamics, whereas the second delivers a more nuanced study, adept at identifying individual intricacies across various regions and timeframes. The preliminary research developed the conceptual and methodological foundation for the UDI, demonstrating its use throughout the pandemic by underscoring the disparities between economic recovery and urban dynamics. The latest study enhanced this methodology, enabling the identification of more accurate local patterns and the validation of the indicator using location data from mobile devices. This methodological advancement illustrates the capability of the UDI as an instrument to facilitate urban planning, applicable in both crisis situations and long-term goals.

In conclusion, the two studies, despite their divergent methodologies and granularity, are complementary and demonstrate the emergence of the UDI as a crucial resource for monitoring and planning urban dynamics. The initial research was essential in confirming the indicator's feasibility and significance, but the subsequent study enhanced its application, allowing in-depth insights and localized actions. Collectively, they establish the UDI as an essential instrument for advancing intelligent and sustainable urban environments, tailored to the evolving and varied requirements of their residents.

3. METHODOLOGY AND DATA

This chapter provides a comprehensive overview of the research methodology, encompassing all requisite steps. It begins with the methodology utilized in this study, an overview of the data sets employed, and a description of the data preprocessing processes implemented to ensure data integrity. These methodologies encompass data cleansing and data integration techniques.

The study's methodological approach, which is primarily based on the Design Science Research framework, is introduced in this chapter. This method facilitates the iterative refinement and creation of the UDI, which is designed to capture and analyze the diverse factors that contribute to the urban dynamics of Lisbon. The UDI model is described in the chapter, which follows the methodological framework. The datasets that are incorporated into the model include high-frequency mobility data from Vodafone, environmental data from urban sensors, meteorological data, and congestion reports from Waze.

3.1. METHODOLOGY

The research employs a Design Science Research (DSR) methodology, as stated by Baskerville et al. (2018). This methodology promotes the integration of high relevance and robust rigor, while at the same time highlighting utility and its contribution to scientific knowledge through the improvement of existing material. The topic is appropriate for DSR as it focuses on creating innovative and efficient solutions for complex problems (Bokolo, 2023; Myeong et al., 2022). The methodology phases are depicted in Figure 1.

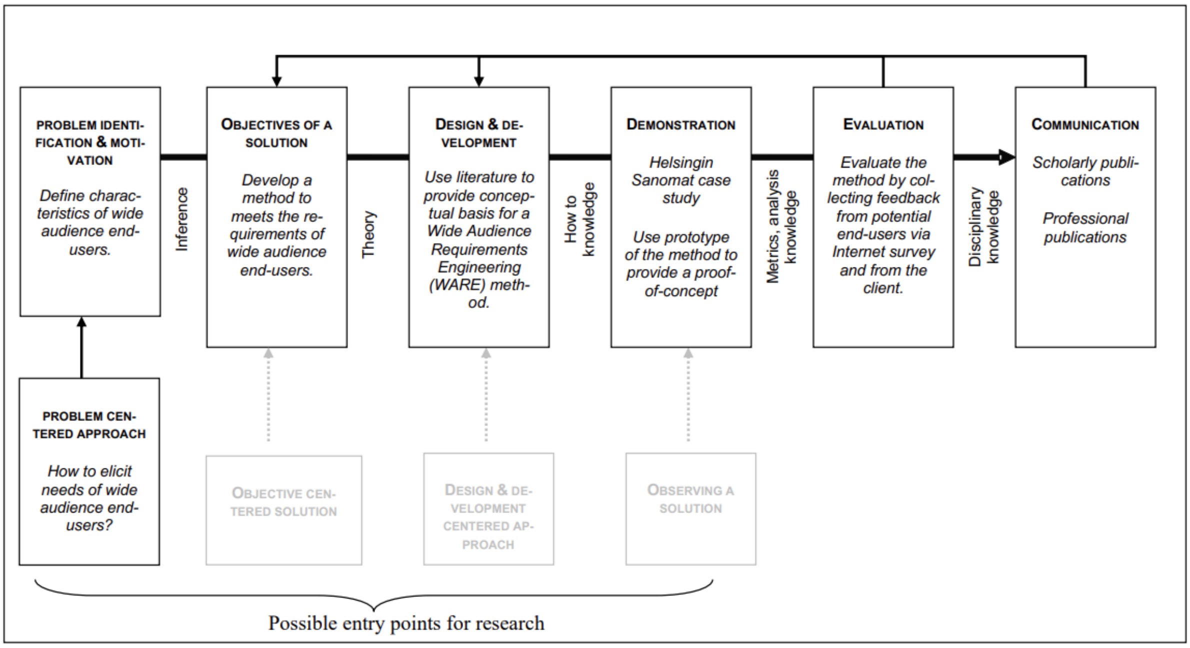

Figure 1 - DSR process case study (Peffers et al, 2007)

1. Problem Identification and Motivation
   i. The stated challenge is the need to quantify and examine the urban dynamics of Lisbon comprehensively, with a focus on both spatial and temporal aspects, at a high level of detail. The objective is to provide data that may facilitate well-informed judgments for sustainable urban development.

2. Objectives and Solution
   i. The goal is to develop a composite metric, the Urban Dynamic Indicator, that integrates multiple urban attributes into a unified measure. This indicator must be able to identify and analyze patterns and trends in urban dynamics, consequently facilitating analysis and decision-making processes.

3. Design and Development
   i. Implement factor analysis to combine many urban variables, including transportation, meteorological, and environmental data, into a singular composite indicator.
   ii. Standardize the data to ensure comparability across different variables.

4. Demonstration
   i. Employ the Urban Dynamics Indicator (UDI) to examine the period from September 1st, 2021, to August 31st, 2022, to identify and analyze patterns and trends in Lisbon's urban dynamics.

5. Evaluation
   i. Validate and interpret the UDI's findings by contrasting them with other relevant data, such as Vodafone mobility statistics.
   ii. Evaluate the effectiveness of the UDI in capturing the complexities of urban development and providing useful data for urban planning objectives.

6. Communication
   i. Document the methodology used to develop the Urban Dynamic Indicator and the results of the analysis.

3.2. STUDY AREA

Figure 2 illustrates the geographical location of the study region. Lisbon is the capital of Portugal and the municipality with the largest population, according to *INE*, with a population reaching 504 964 habitants with a population density of 5047,1 habitants per Km² in 2018 (INE, 2018).

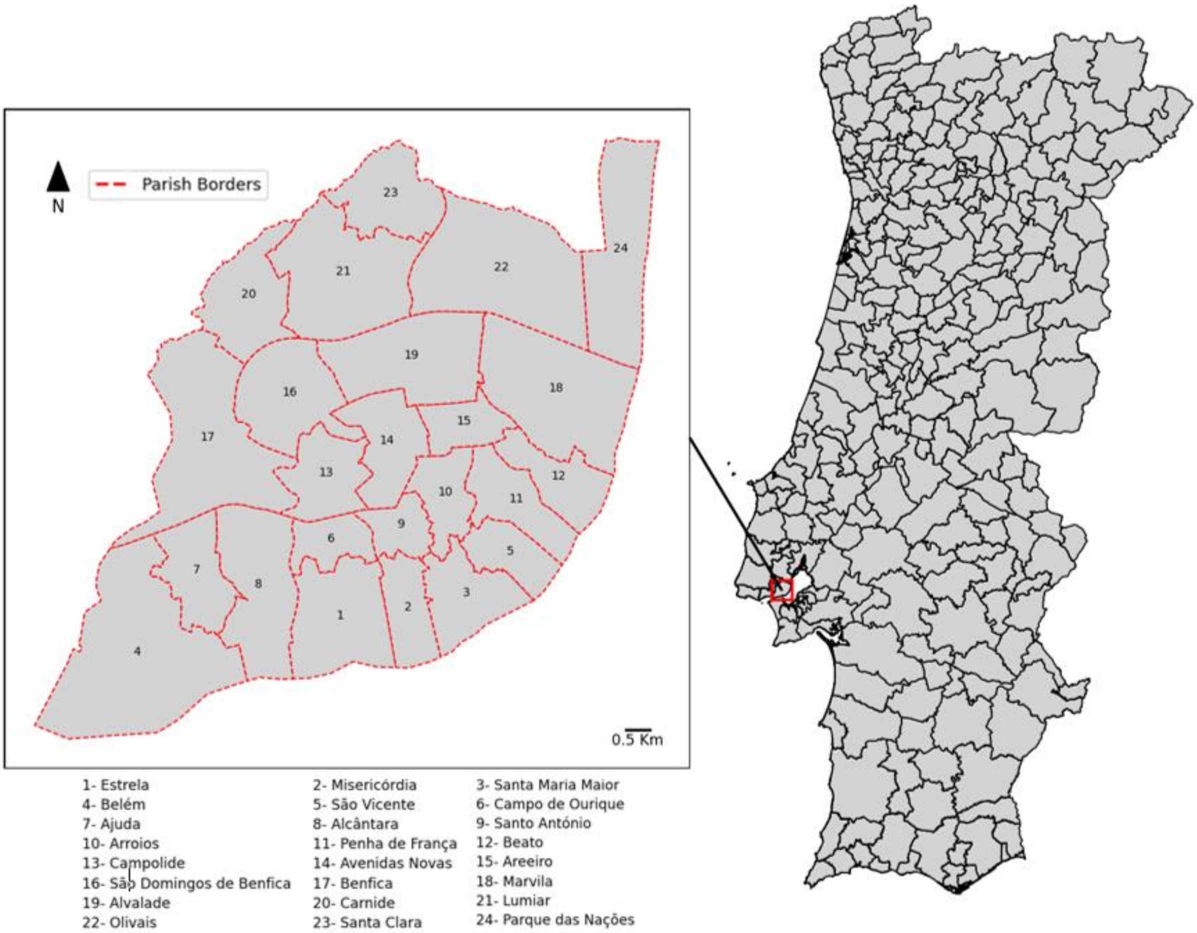

Figure 2- Portugal municipalities and Lisbon parishes (CML, 2021)

Lisbon is situated in the western region on the northern side of the Tagus River of mainland Portugal and has hot, arid summers and pleasant winters (Köppen-Geiger climatic classification Csa – (Kottek et al., 2006). The climate is Mediterranean, marked by scorching, arid summers and chilly, rainy winters. Most of the precipitation takes place between October to April. The climatological norm (1971–2000) indicates an average annual temperature of 18 °C, an average minimum temperature of 9 °C, an average maximum temperature of 27 °C, and an average annual precipitation of 726 mm (IPMA, 2010).

The European Commission website states that Lisbon has adopted a smart city strategy focused on its citizens and their requirements. The city aims to be intelligent, sustainable, competitive, participatory, creative, innovative, and citizen-centric. Lisbon has devised an urban development strategy for the upcoming decades and has committed to investing 307 million euros in related projects (Programa Operacional Regional de Lisboa 2020). The strategy's primary objectives are to improve housing quality and offer smart living services and opportunities for intelligent aging to attract additional residents, boost the economy, and increase employment by investing in research and development, attracting entrepreneurs, and broadening access to higher education, and to elevate the quality of life in the city through initiatives centered on energy efficiency, mobility, and social cohesion. Local revitalization and citizen involvement are priorities as well (European Commission, 2024).

On the European Commission website, it can also be understood that the designated demonstration area for the SHARING CITIES initiative encompasses 10 km^2 and is home to 100,000 residents. This key area extends from the riverbank to the city center, encompassing the primary historic and tourist districts. The region presents multiple obstacles, including its unique orography, the historical significance of its architecture, and its aging demographic (European Commission, 2024).

Lisbon also has an open data platform named Lisboa Aberta, which provides public information in an accessible manner to enhance openness and foster citizen engagement. This platform provides diverse data encompassing mobility, environmental issues, urbanism, and cultural activities, enabling people, researchers, and businesses to leverage this information to create new, evidence-based solutions. The initiative demonstrates Lisbon's dedication to digital transformation and sustainability, aligning with smart city objectives by enabling informed decision-making and fostering collaborative projects (Câmara Municipal de Lisboa, 2018).

3.3. DATA UNDERSTANDING

This section examines the data sources included in this study, emphasizing their significance and the distinct contributions that each source provides to the results. Therefore, the data provided in Table 1 was acquired for this research.

Table 1 - Distribution and Boxplot of Mobility Data

Datasets	Description	Provider	Data Formats
Mobility Data	Count of mobile terminal movement during a 5-minute interval within Lisbon	Vodafone	CSV
Environmental Data	Location of sensors and environmental data within Lisbon	LoRa	CSV
Meteorological Data	Location of stations and data of meteorological parameters within Lisbon	IPMA	CSV
Traffic Data	Jams reported by the Waze App users at certain speed and delay.	Waze	CSV
Parish Limits of Lisbon	Delineates the boundaries of Lisbon's parishes	DGT	shp

3.3.1. Mobility Data

Vodafone's data includes a series of CSV files, including several anonymised and aggregated records. These records correspond to the movement of mobile devices, like mobile phones, within the grids of Lisbon, segmented into 200x200 meter squares. This dataset collects data at five-minute intervals for each of the 3,743 grids distributed throughout Lisbon from September 2021 to August 2022. It contains significant signals pertaining to the mobility of mobile terminals, both local and roaming, as illustrated in Table 14, located in the Annex. Upon analysis, the variable "C1" was revealed as the most logical variable for examination. Consequently, the remaining variables were omitted, and a full list of variables is available in Table 2.

Table 2 – Mobility Dataset

Variable	Description
Grid_ID	Identification of the Grid Number
Geometry	Geometric multi-polygon representing the location of each grid
Parish	Parish of the centroid of the square
DateTime	Date and Time
Total	Count of unique terminals in the grid (C1)

This dataset will act as the evaluation dataset for this project, since it offers essential information for obtaining numerous relevant inputs on human movement inside Lisbon grids from September 2021 to September 2022, validating the correlation with the UDI.

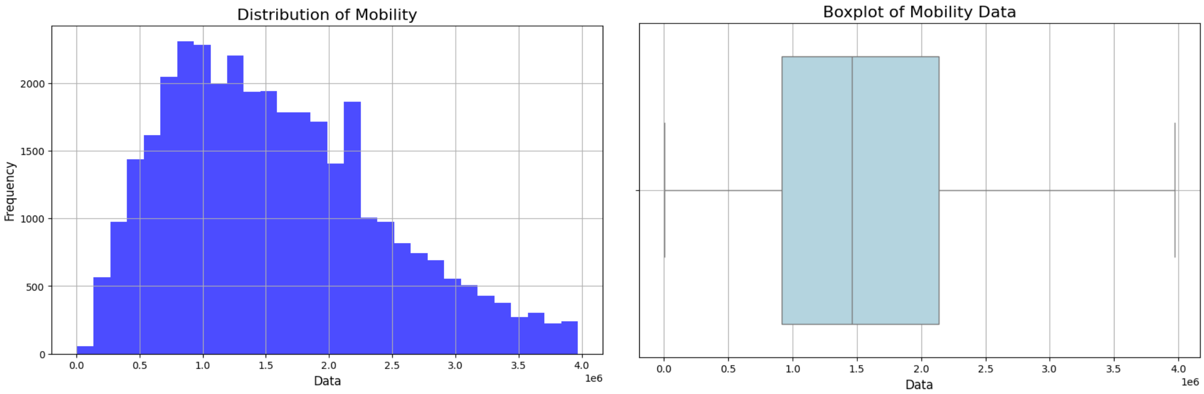

Figure 3 - Distribution and Boxplot of Mobility Data

The graphs in Figure 3 offer detailed insights into data distribution and its statistical properties. The left picture depicts a histogram of mobility data, indicating the number of documented device movements within a specified timeframe. The distribution displays right skewness, with the majority of values between 0.5 and 2 million, accompanied by a progressive decrease in the frequency of higher values. The long tail signifies a limited number of instances where movement counts attain markedly elevated values, around 4 million. The histogram distinctly depicts the distribution of the data, emphasizing the frequency of different movement value ranges and the dispersion of the data.

In contrast, the right graph, the boxplot, encapsulates the data's variability. The interquartile range (IQR) denotes the central 50% of the dataset, and the whiskers indicate the range, omitting outliers. The boxplot visualizes extreme values as potential outliers, providing a structured representation of the data's variability and emphasizing deviations from the central values.

Collectively, these statistics suggest that the mobility data is characterized by intermediate values intersected with sporadic high peaks, as illustrated in the histogram. The boxplot validates this dispersion, distinctly illustrating the data's variability and encompassing its primary statistical ranges. The amalgamation of these numbers offers an extensive perspective on the data distribution and its variability.

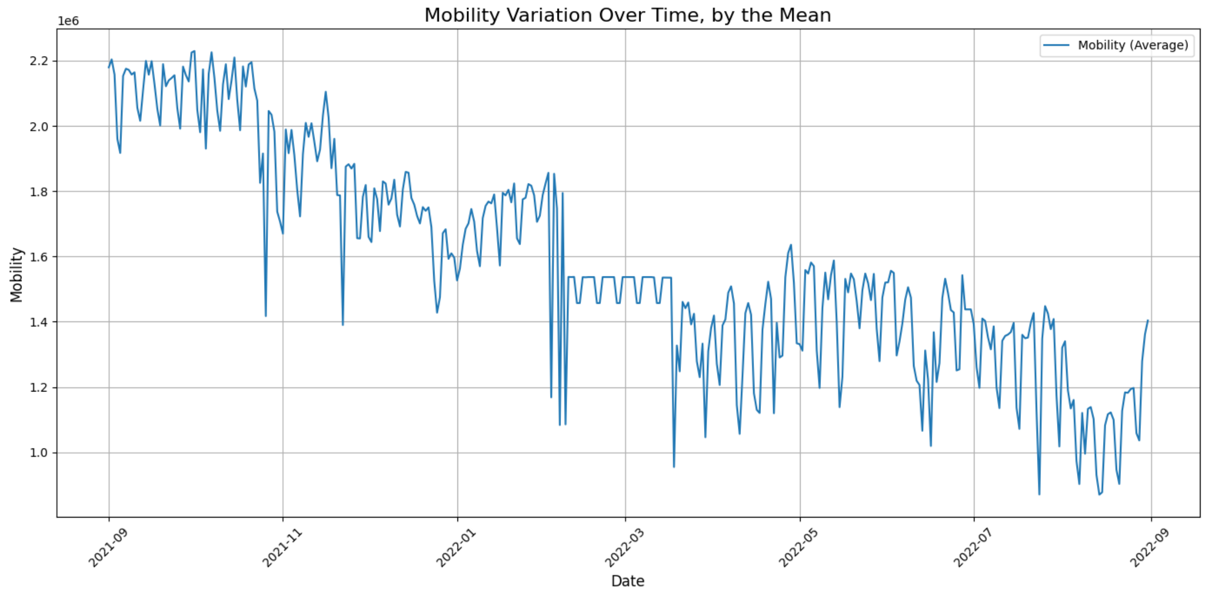

Figure 4 - Mobility Variation Over Time (Average)

Figure 4 illustrates the fluctuation in total mobility throughout a one-year duration. The graph indicates a distinct decline in mobility, characterized by several abrupt oscillations during the time. Mobility initially remains elevated through September and October 2021 before experiencing a significant fall starting in November 2021. This declining trend persists until the conclusion of 2021 and the commencement of 2022. Throughout this timeframe, mobility stays very low, with significant declines, especially in early 2022.

Between February and May 2022, a phase of relative stability occurs, characterized by diminished swings in mobility, indicating either more uniform travel patterns or a decline in activity during this period. From February to the end of March, a consistent temporal frame exists due to the imputation of missing values in the data. From June 2022 onward, fluctuations increase, indicating a resurgence in unpredictability in mobility patterns. The graph illustrates a decline in total mobility over time, featuring different phases of volatility and stability in movement patterns.

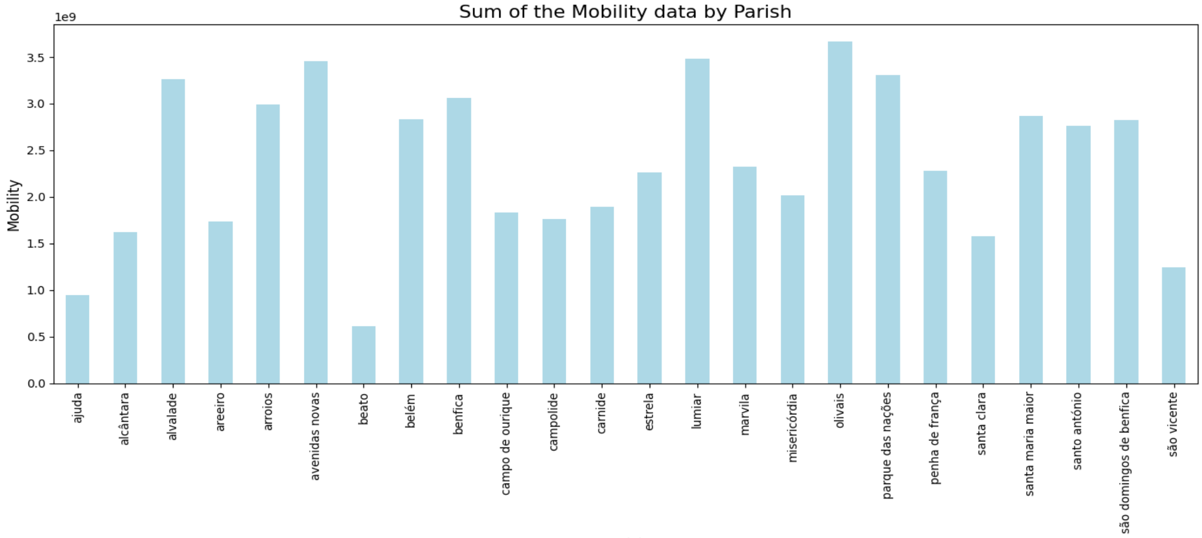

Figure 5 - Mobility Data by Parish (Sum)

Figure 5 illustrates the aggregate mobility statistics across different parishes, revealing distinct variations in traffic volume. Parishes, including Alvalade, Avenidas Novas, Lumiar, Olivais, and Parque das Nações, exhibit the highest mobility statistics, surpassing 3 billion units each. Moreover, parishes such as Areeiro, Benfica, Santo António, and Santa Maria Maior demonstrate mobility levels ranging from 2 to 3 billion units.

Parishes such as Ajuda, Beato, Santa Clara, Misericórdia, and São Vicente exhibit the lowest mobility totals, up to approximately 1.5 billion units. This chart offers a graphic representation of mobility dispersion among the different parishes.

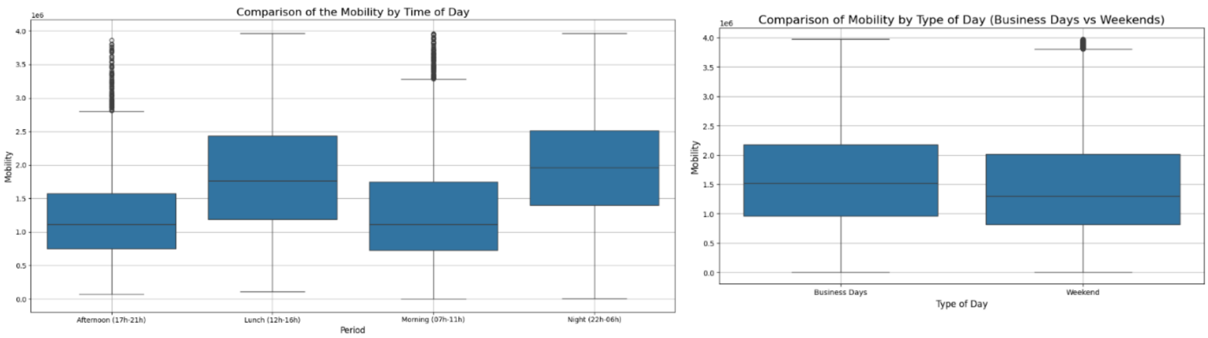

Figure 6 - Comparison of Mobility by Period and Type of Day

This analysis, presented in Figure 6, offers a succinct summary of mobility trends throughout various times of day and types of days, emphasizing significant patterns in variability and volume.

The graph on the left, illustrating mobility by time of day, reveals considerable fluctuations in mobility patterns across various time intervals. The afternoon interval (17:00-21:00) exhibits a median mobility value of approximately 1.02 million, with numerous outliers above 3.5 million, signifying significant traffic surges. The lunch period (12:00-16:00) demonstrates the greatest median, almost 2 million, and presents fewer outliers, indicating a more stable flow of movement throughout this interval. The morning time (07:00-11:00) exhibits the lowest mobility levels, with a median slightly below 1.05 million and reduced fluctuation, signifying a more stable trend. Conversely, the nocturnal period (22:00-06:00) exhibits a larger median, approaching 2 million, with comparatively fewer outliers but sporadic traffic surges. The lunchtime period experiences the biggest and most consistent traffic, whilst the afternoon exhibits increased fluctuation and sporadic peaks. The morning time is the most constant, whereas the night period has increased traffic but fewer notable peaks.

The graph on the right, illustrating mobility on business days versus weekends, reveals a distinct disparity in traffic patterns. During workdays, the median mobility approximates 2 million, exhibiting considerable variability shown by the data spread and the existence of outliers above 4 million. During weekends, the median is marginally reduced, approximating 1.5 million, with diminished variability. Nonetheless, there remain outliers above 4 million, signifying that traffic spikes may transpire on specific weekends. Weekday traffic is generally denser and more erratic, whereas weekend travel is more stable, albeit with occasional surges.

3.3.2. Environmental Data

The Lisbon Council installed several sensors around the city, as can be seen in Figure 7, to gather environmental urban data, such as ozone and noise levels.

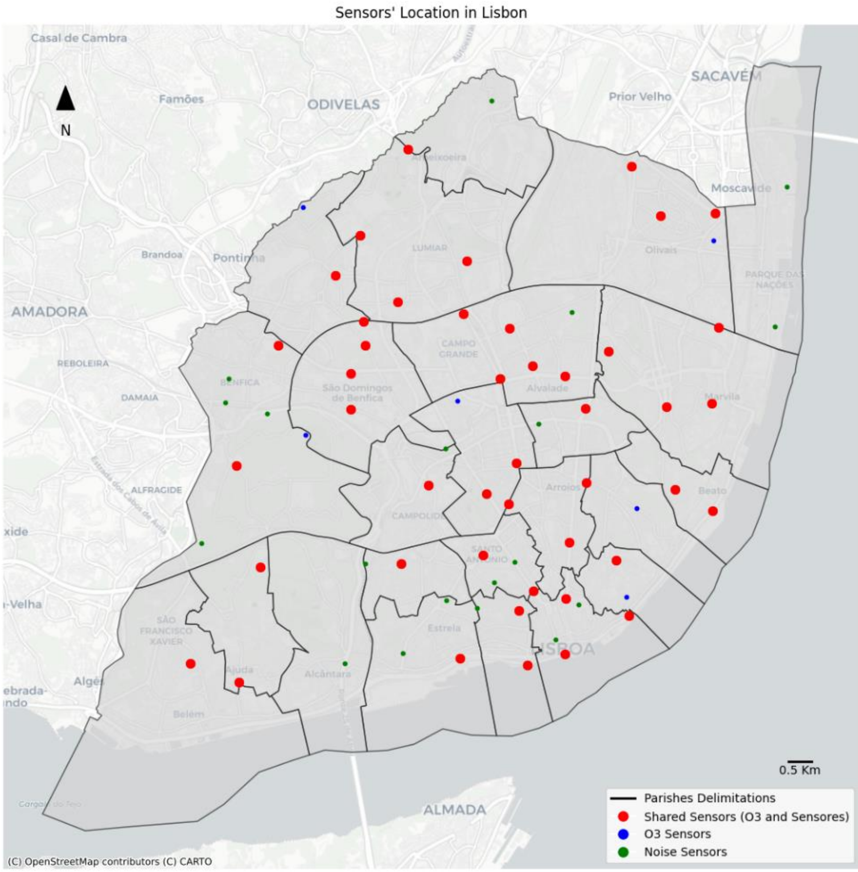

Figure 7 - Environmental Sensor's Location

The data collected is provided in two different datasets, presented in the following Table, Table 3.

Table 3 - Environmental Dataset

Datasets	**Description**
Lora_sensor	Contains the location of each sensor, and which variable each sensor collects.
Lora_values	Contains the date and the values collected from the sensors. Includes data of ozone (µg/m³) and noise (dB) levels.

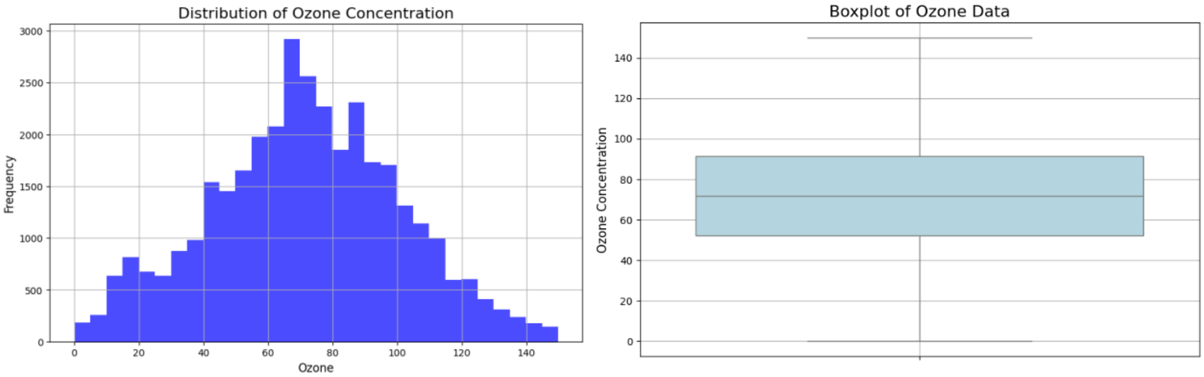

Figure 8 - Distribution of Ozone Concentration

Figure 8 presents a detailed depiction of the distribution and fluctuation of ozone concentration data. The left picture, a histogram, illustrates the distribution of ozone (O_3) values, which is nearly symmetrical, with the peak concentration of values occurring between 60 and 80 µg/m³. The maximum frequency is observed at approximately 60 µg/m³, with over 3,000 instances, clearly demonstrating that most ozone readings are centered within this range. The histogram displays an extended tail on the right, reaching around 140 µg/m³, signifying occasional occurrences of elevated ozone concentrations, albeit infrequently.

The boxplot on the right enhances the histogram by summarizing the data's variability. The interquartile range (IQR) encompasses the middle 50% of the data, extending from around 55 to 90 µg/m³, indicating that the majority of ozone values reside within this consistent range. The whiskers span from 0 to 140 µg/m³, demonstrating the complete spectrum of values without notable outliers. The boxplot elucidates data dispersion by distinctly illustrating the range and spread of ozone values, with a median concentration of approximately 75 µg/m³.

Collectively, these data indicate that ozone concentrations predominantly cluster around moderate levels, with occasional elevated peaks evident in the histogram. The boxplot validates this variability range, offering a systematic representation of the distribution and statistical attributes of the ozone data.

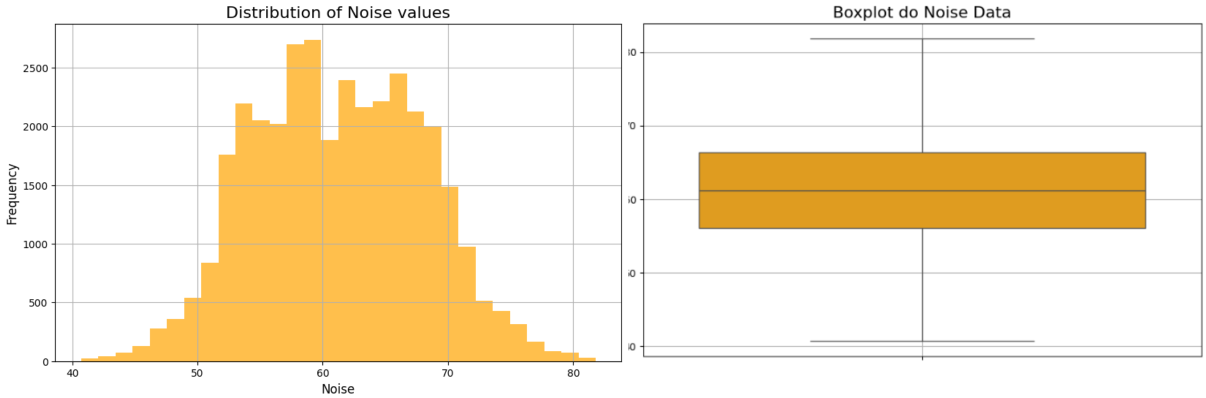

Figure 9 - Distribution of Noise Values

Figure 9 presents a comprehensive analysis of the distribution and variability of noise data. The left picture, a histogram, illustrates the distribution of noise values, which is almost symmetrical. The greatest concentration of values is found between 55 and 75 dB, with the peak frequency approximately at 60 dB. This signifies that most noise measurements are clustered within this band. The histogram indicates a significant decrease beyond 70 dB, implying that elevated noise levels are less prevalent, although they do persist.

The boxplot on the right complements the histogram by encapsulating the data's variability. The interquartile range (IQR) represents the middle 50% of the data, spanning approximately 57 to 67 dB, signifying that the majority of noise levels reside within this range. The whiskers range from roughly 40 to over 80 dB, including the broad distribution of the noise data without notable outliers. The boxplot distinctly indicates that the median noise level exceeds 60 dB, consistent with the central tendency found in the histogram.

Collectively, these statistics indicate that the noise data is predominantly centered around moderate levels, exhibiting considerable variability at the extremes. The histogram illustrates the frequency distribution, whilst the boxplot succinctly summarizes the data's range and variability, presenting a comprehensive overview of the statistical properties of the noise data.

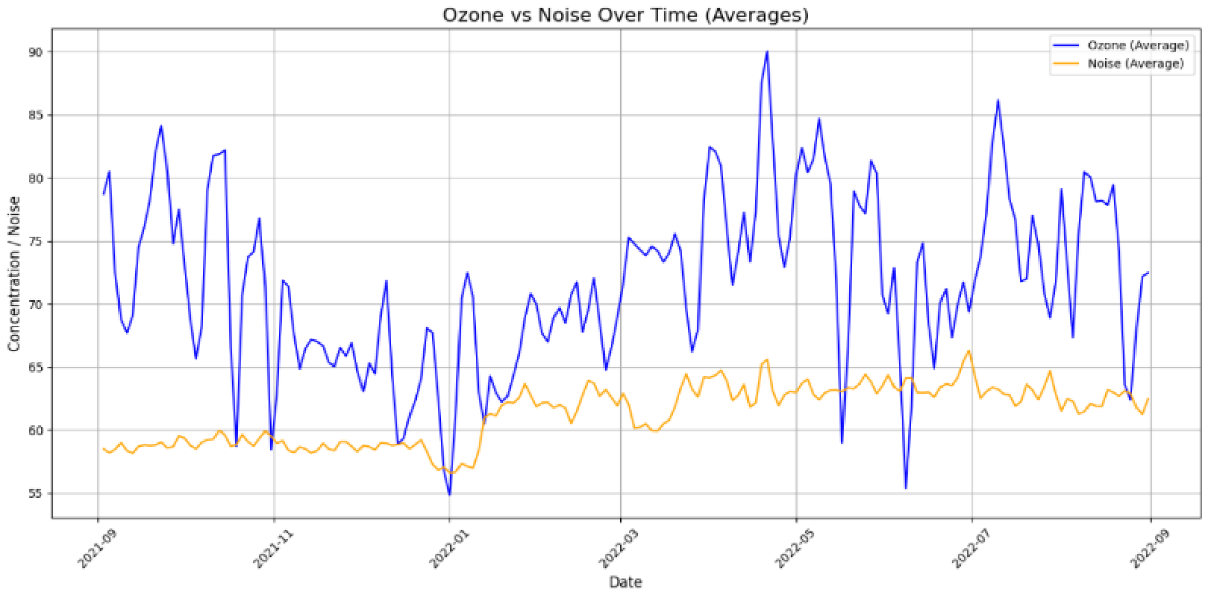

Figure 10 - Ozone and Noise Variation Over Time (Average)

Figure 10 depicts the fluctuations in ozone concentration and noise levels over one year. The blue line indicates the ozone concentration, exhibiting significant variability over time, marked by notable peaks and abrupt falls. The peak values were recorded between September and December 2021, above 80 µg/m³, followed by a swift decline at the onset of 2022. Subsequent to this interval, ozone concentrations persist in fluctuating, exhibiting a pattern of ongoing oscillations for the remainder of the year.

The orange line denotes the noise levels, which have significantly less variability than ozone concentration. Noise levels are consistently steady, fluctuating between 55 and 65 dB annually. In contrast to ozone, noise does not exhibit significant fluctuations, suggesting it is less susceptible to abrupt environmental alterations and is probably driven by more stable variables, such as urban activities and traffic congestion.

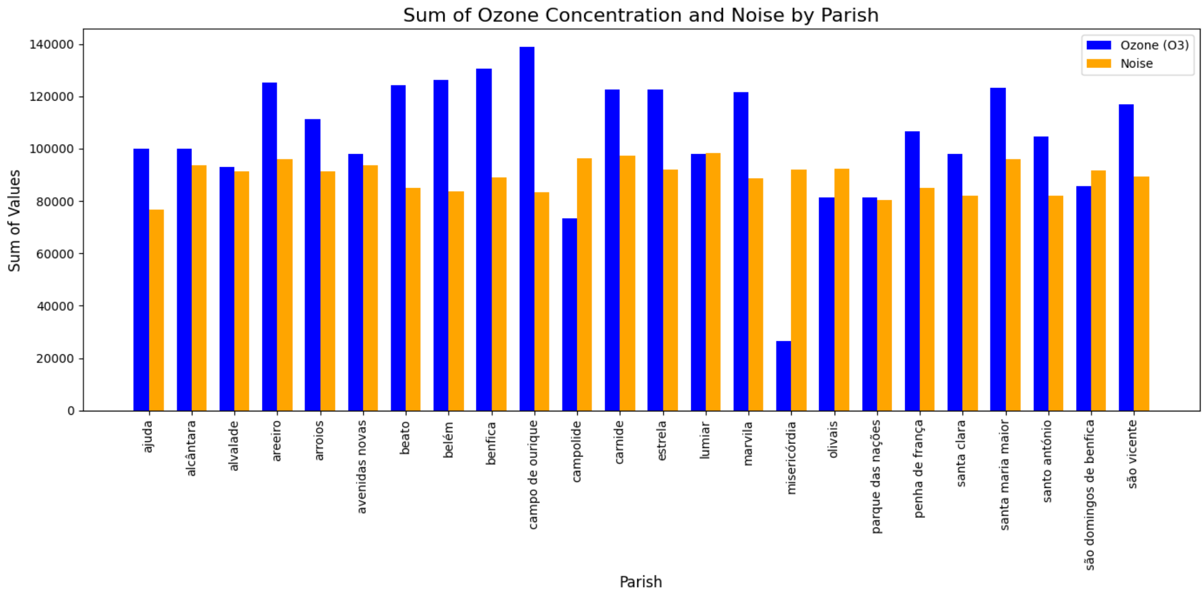

Figure 11 – Ozone and Noise values by Parish (Sum)

Figure 11 presents the aggregate sums of ozone concentration and noise levels across various parishes, showing clear differences in the values between each parish. For ozone concentration (represented by the blue bars), parishes such as Benfica, Lumiar, Marvila, and Olivais display the highest sums, each exceeding 120,000 µg/m³. In comparison, parishes like Ajuda, Beato, and São Vicente exhibit lower ozone concentration values, below 100,000 µg/m³.

This comparison elucidates the divergent behaviors of the two variables. Ozone concentrations exhibit considerable variability, with substantial temporal changes, while noise levels remain relatively steady, with negligible variation during the investigated timeframe. Collectively, these two lines offer insights into the temporal behavior of several environmental elements under identical conditions.

In terms of noise levels (represented by the orange bars), parishes show a relatively uniform pattern, with values generally lower than those for ozone. However, certain parishes like Lumiar and Areeiro have slightly higher noise sums compared to others. Overall, noise levels are lower and less variable across the parishes compared to ozone concentration.

This chart provides a comparative view of ozone and noise levels across parishes, highlighting that ozone concentrations tend to vary more widely than noise levels. The graphical representation enables a quick assessment of the parishes with the highest and lowest environmental measurements for each variable.

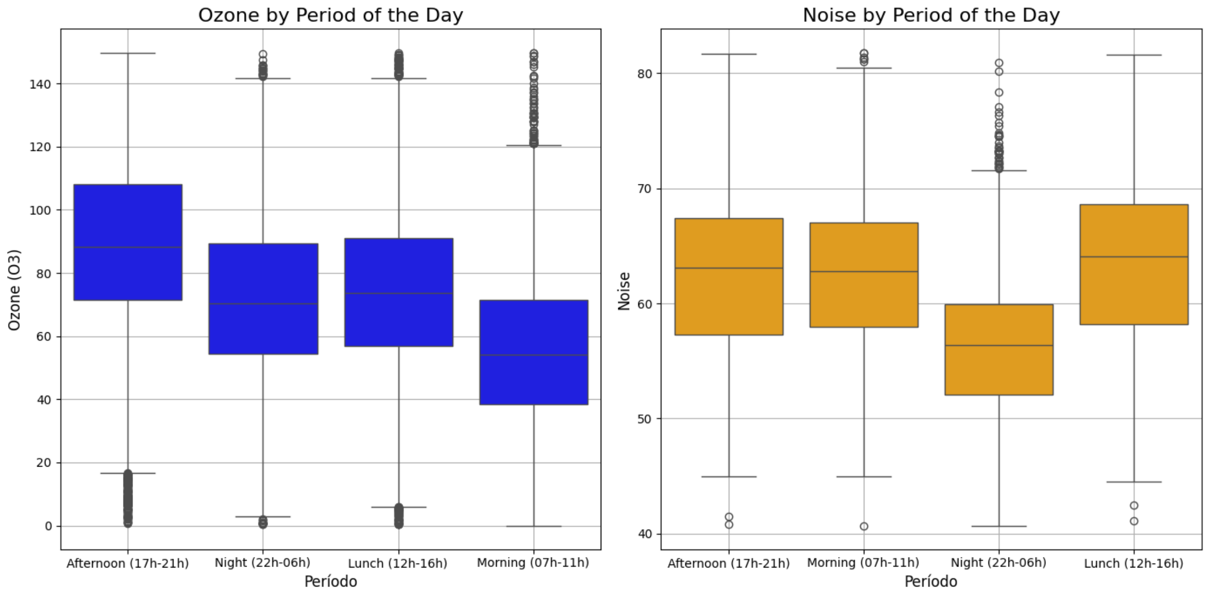

Figure 12 – Ozone and Noise values by Period of the Day

Figure 12 examines ozone concentration and noise levels throughout the day, revealing distinct patterns for each variable.

The left graph depicts the distribution of ozone (O_3) levels over various time intervals. The afternoon period (17:00-21:00) exhibits the highest median ozone concentration, approaching 90 µg/m³, accompanied by numerous outliers indicating sporadic peaks in ozone levels. The lunchtime (12:00-16:00) and night periods (22:00-06:00) exhibit moderate ozone concentrations, both with a median of 70 µg/m³, while both periods also present several outliers. The morning hour (07:00-11:00) demonstrates the lowest median ozone concentration, marginally exceeding 60 µg/m³, signifying a more consistent level during this interval. The afternoon and evening periods have more variations and elevated peaks in ozone levels, whereas the morning demonstrates increased constancy.

The right graph illustrates noise data over time, exhibiting a generally more stable distribution than that of ozone levels. The median noise level consistently ranges from 55 to 65 dB throughout all time durations. The afternoon period (17:00-21:00) exhibits marginally greater variability, with multiple outliers surpassing 80 dB, signifying intermittent surges in noise intensity. The lunch period (12:00-16:00) exhibits a slightly increased range of noise levels, albeit with fewer outliers than the afternoon. The nocturnal (22:00-06:00) and matutinal (07:00-11:00) intervals exhibit analogous distributions, characterized by a scarcity of outliers and more uniform noise levels, presumably attributable to diminished urban activity during these times.

An analysis of the two graphs indicates divergent trends for each variable over the course of the day. Ozone levels demonstrate considerable fluctuations over time, particularly in the afternoon and evening, although noise levels maintain greater stability, with sporadic surges.

The outliers in both figures indicate that, at times, ozone and noise levels can attain values beyond the conventional range, albeit for distinct reasons and at varying intervals.

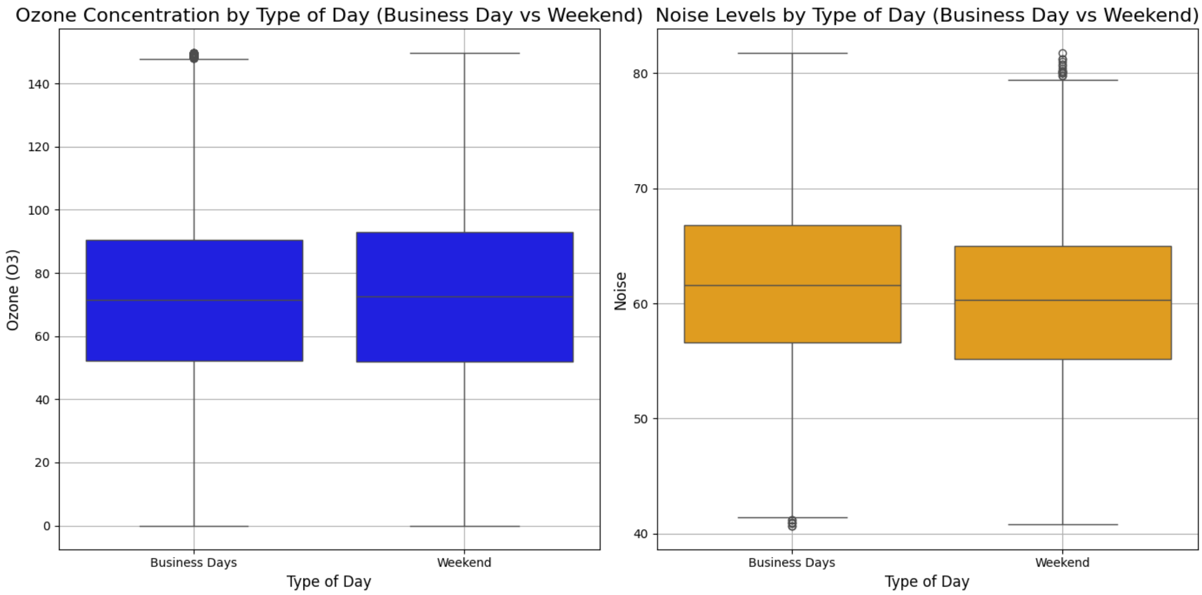

Figure 13 - Ozone and Noise values by Type of Day

Figure 13 compares ozone concentration and noise levels on weekdays versus weekends.

The left graph depicts the distribution of ozone (O_3) concentration. The median ozone values for weekdays and weekends are comparable, ranging from 53 to 90 µg/m³, suggesting that ozone concentration remains relatively stable. Nevertheless, there are some outliers on work days, with measurements over 140 µg/m³, indicating sporadic surges in ozone concentrations throughout the week.

The graph on the right illustrates noise levels categorized by day type. A little discrepancy in noise levels exists between weekdays and weekends. During weekdays, the median noise level is marginally elevated, fluctuating between 57 and 68 dB, whereas on weekends, the median decreases to approximately 60 dB. Business days exhibit more outliers, with certain values over 80 dB, signifying that commercial and daily traffic activities elevate noise levels. Conversely, weekends demonstrate less unpredictability and diminished extreme noise peaks.

Ozone concentrations are stable throughout working days and weekends, whereas noise levels often diminish on weekends, with fewer spikes and less unpredictability.

3.3.3. Meteorological Data

The Lisbon City Council in collaboration with Instituto Português do Mar e Atmosfera (IPMA) provided a CSV file that includes meteorological variables captured by 3 stations in Lisbon, presented in Figure 14. The meteorological stations measure several variables present in Table 17 in the Annex.

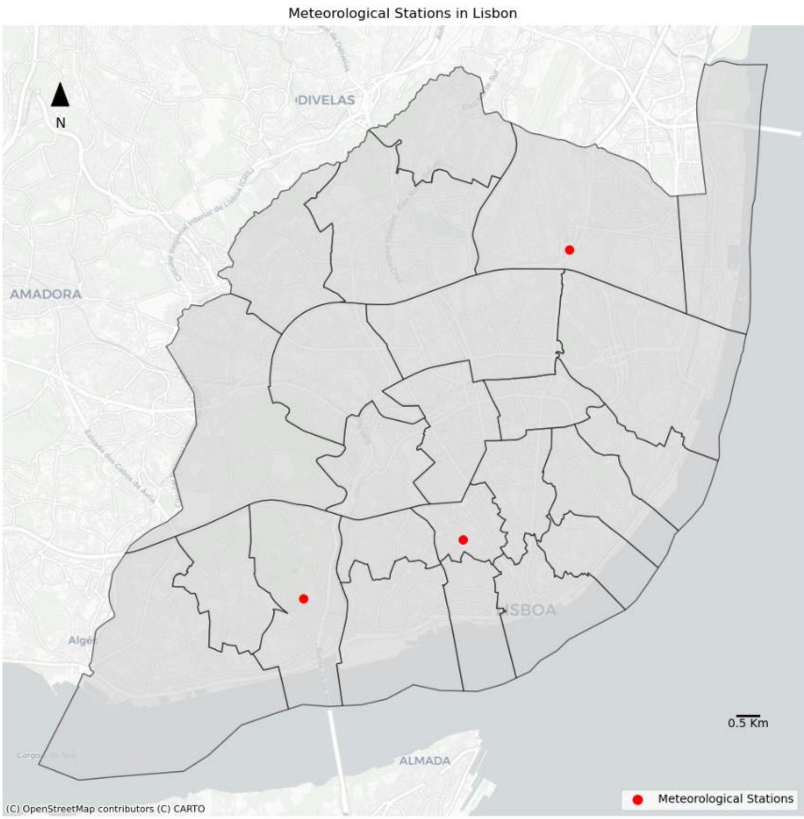

Figure 14 - Meteorological Stations

The variables presented in the dataset are provided in the following table, Table 4.

Table 4 - Meteorological Dataset

Variables	**Description**
ID	Unique Identification of Record
DateTime	Date and Time
Station	Terminals in the grid
Cumulative Precipitation	Value of Cumulative Precipitation (mm)
Temperature	Value of Temperature (°C)

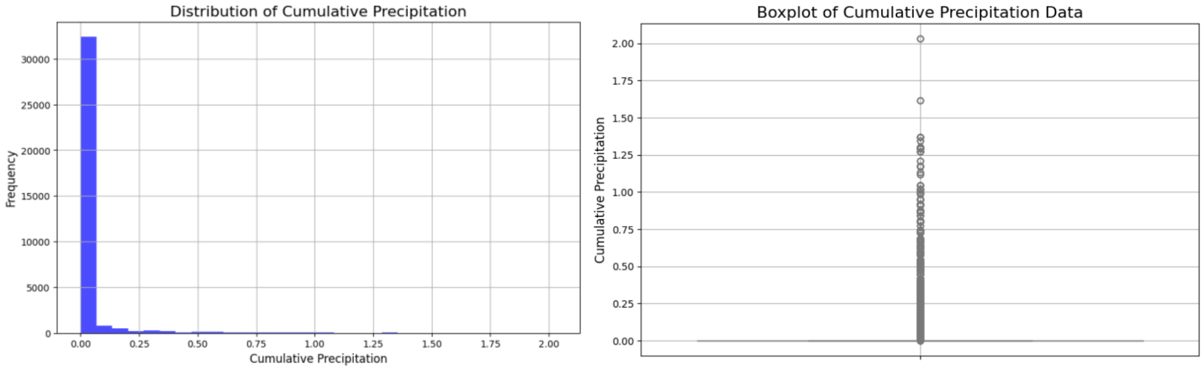

Figure 15 - Distribution of Cumulative Precipitation Values

Figure 15 illustrates the distribution and variability of cumulative precipitation data.

The histogram on the left illustrates a significantly skewed distribution, with the majority of data points clustered around 0 mm of precipitation. The majority of precipitation events documented negligible or no rainfall, with only a few occurrences over 0.25 mm. The extended tail to the right signifies infrequent occurrences of substantial precipitation.

The boxplot on the right corroborates this result, indicating that the majority of the data is densely concentrated around 0 mm. The interquartile range (IQR) is small, signifying restricted variability in the majority of the data. Nonetheless, there are multiple outliers, with measurements reaching 2 mm, indicating that although substantial rainfall events are rare, they transpire.

These graphs demonstrate that the data is primarily arid, with minimal occurrences of significant precipitation. The data is predominantly concentrated on low precipitation levels, with sporadic outliers indicating greater rainfall.

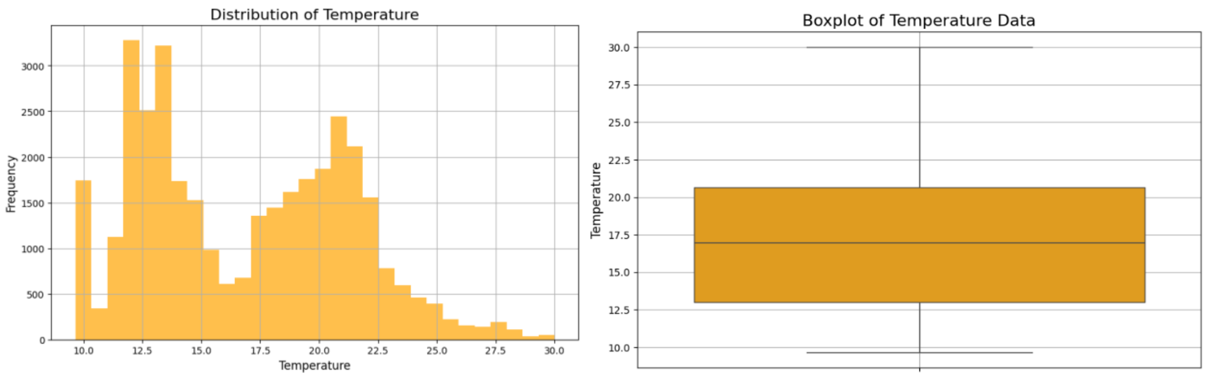

Figure 16 - Distribution of Temperature Values

Figure 16 illustrates the distribution and fluctuation of temperature data.

The histogram on the left indicates that the temperature distribution is predominantly concentrated in two ranges: 10°C to 15°C and 17,5°C to 22,5°C. These intervals indicate the predominant temperature values, implying that the dataset is characterized by slightly to moderately chilly temperatures, with limited occurrences of excessive heat. The histogram indicates that temperatures over 25°C are infrequent, whilst the predominant values reside within a moderate range.

The boxplot on the right enhances the histogram by depicting the dispersion of the temperature data. The interquartile range (IQR) extends from around 15°C to 20°C, signifying that most temperature data reside within this consistent range. The median temperature is approximately 17°C, exhibiting no substantial outliers, indicating uniformity in temperature trends. The range spans from 10°C to around 30°C, capturing the full variability of the data without significant departures.

The data reflects a consistent temperature profile, marked by moderate values and negligible extremes. The temperature data predominantly centers around moderate levels, exhibiting minimal variability, indicative of a stable environment within the studied timeframe.

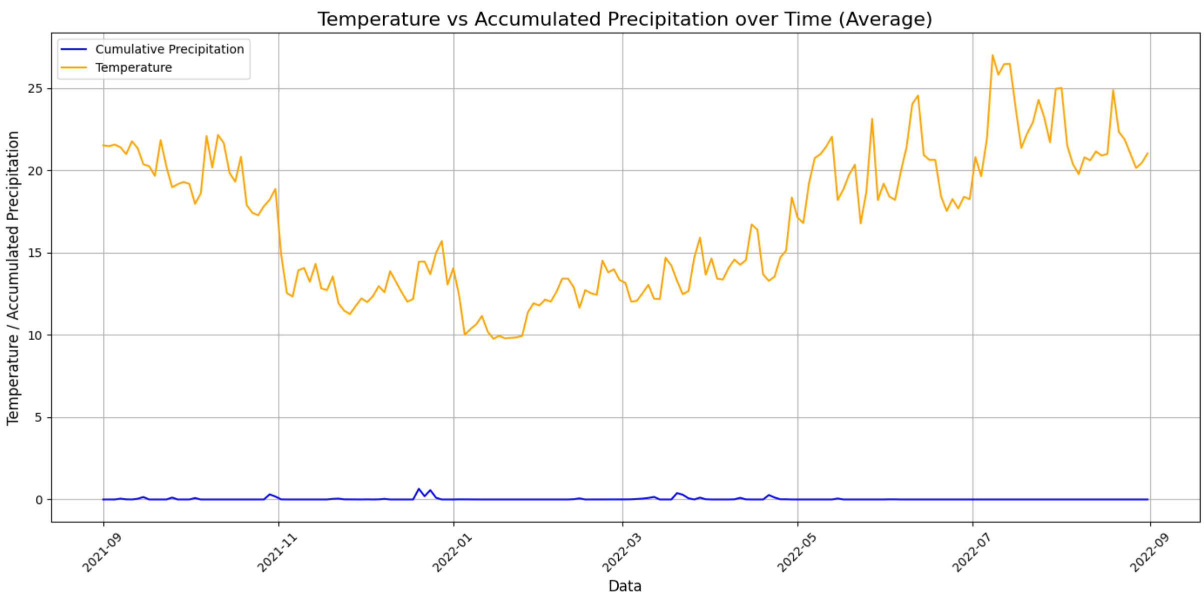

Figure 17 – Cumulative Precipitation and Temperature Variation Over Time (Average)

Figure 17 illustrates the fluctuations in temperature and accumulated precipitation throughout time. The orange line denotes temperature, which progressively declines from September to November 2021, culminating in a precipitous plunge at the year's end. Subsequent to this drop, temperature values oscillate throughout 2022, exhibiting multiple maxima approaching 25°C. The oscillations continue throughout the studied period, signifying consistent temperature variance.

The blue line illustrates cumulative precipitation, which remains relatively stable over much of the period, with a minor rise observed towards the conclusion of 2021. Nevertheless, the cumulative precipitation stabilizes rapidly, remaining near 0 mm for the duration of the month. This signifies negligible precipitation accumulation throughout the year, with rare occurrences of substantial rainfall.

The graph indicates that temperature displays consistent changes, although cumulative precipitation remains rather stable. The divergent patterns of these two variables suggest that they do not exhibit analogous tendencies during the analyzed period, with temperature displaying variability and precipitation demonstrating stability.

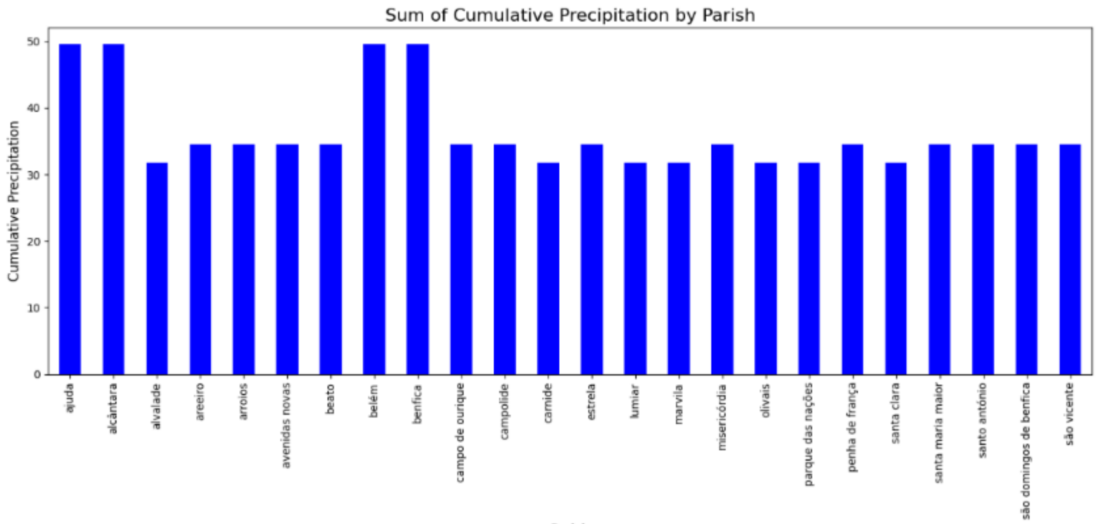

Figure 18 - Cumulative Precipitation values by Parish (Sum)

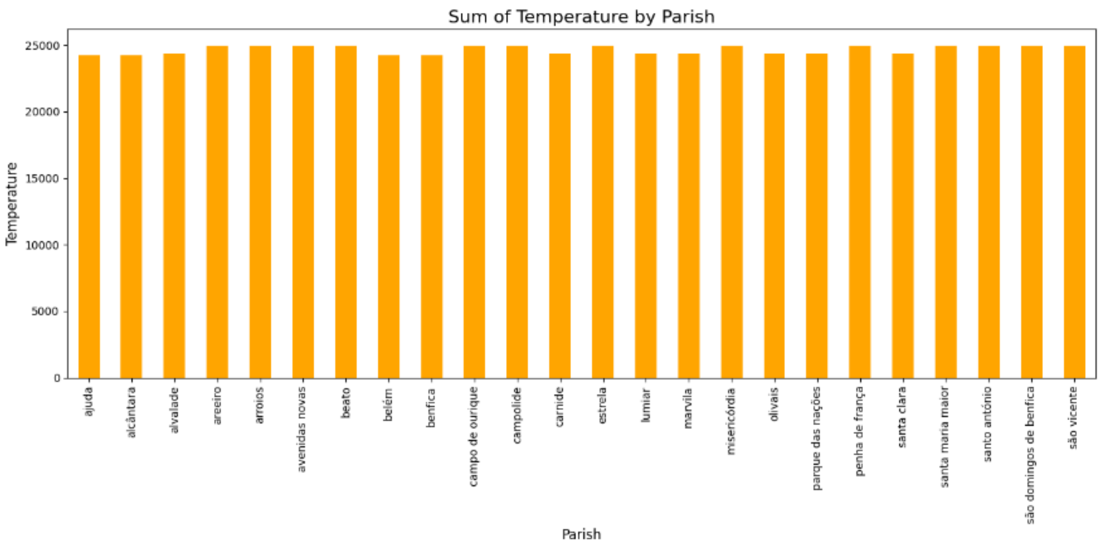

Figure 19 - Temperature values by Parish (Sum)

Figures 18 and 19 present bar graphs illustrating cumulative precipitation and temperature across various parishes, respectively.

Figure 18 illustrates cumulative precipitation, revealing significant heterogeneity among parishes. Most parishes have amassed precipitation totals ranging from 30 to 33 mm. Nonetheless, several parishes, including Ajuda, Alcântara, Benfica, and Belém, demonstrate elevated amounts, approaching 50 mm. This fluctuation indicates that certain regions receive more precipitation than others.

Figure 19, depicting temperature, illustrates a more consistent distribution among parishes. The total temperature measurements are uniform, fluctuating between 20,000°C and 25,000°C throughout all parishes. This homogeneity suggests negligible temperature variance throughout regions, signifying a steady environment devoid of substantial thermal disparities across parishes.

These graphs demonstrate that, although temperature is consistently distributed, cumulative precipitation exhibits significant variation within parishes, with certain areas receiving more rainfall than others.

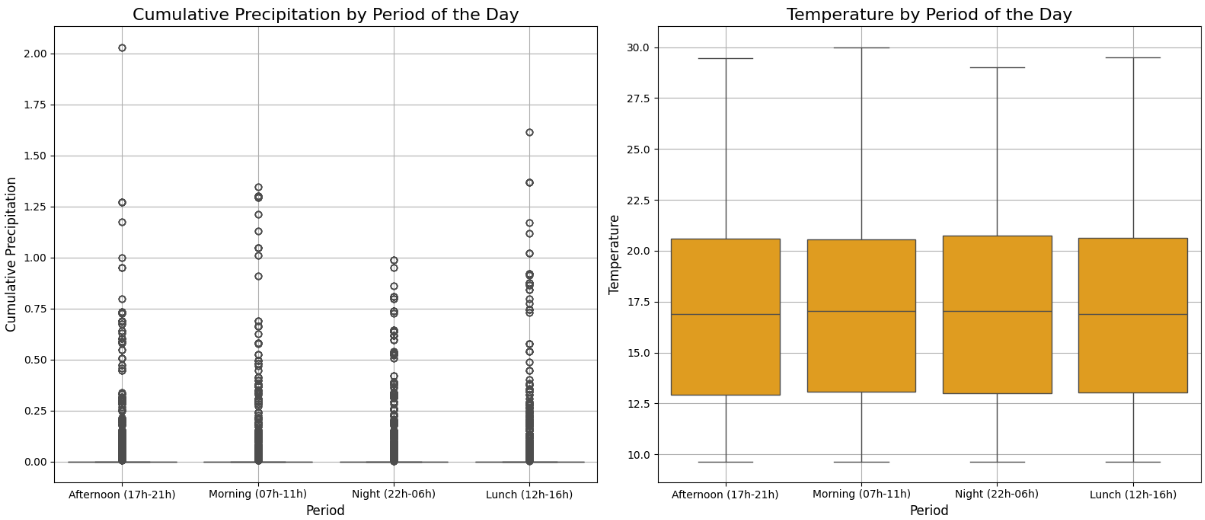

Figure 20 – Cumulative Precipitation and Temperature values by Period of the Day

Figure 20 illustrates the cumulative distribution of precipitation and temperature throughout the course of the day.

The left graph, depicting cumulative precipitation by time of day, indicates that the majority of data points cluster around minimal values, with a median approaching zero throughout all hours. Nonetheless, there are few outliers, especially in the morning (07:00-11:00) and afternoon (17:00-21:00), where cumulative precipitation levels attain up to 2 mm. The outliers suggest that although the majority of precipitation events are negligible, there are sporadic occurrences of substantial rainfall during these intervals.

The right graph, depicting temperature over time of day, exhibits a more uniform distribution throughout the various intervals. The median temperature throughout all intervals is roughly 17°C, with the interquartile range extending from about 13°C to 21°C. This indicates that temperature remains consistently steady throughout the day, exhibiting minor variations across different time intervals.

The figures indicate that cumulative precipitation is often low, with sporadic peaks, whereas temperature exhibits a more consistent and predictable pattern throughout the day. Precipitation is erratic. However, temperature demonstrates stable daily fluctuations with slight variation, indicating more reliable climatic patterns regarding temperature.

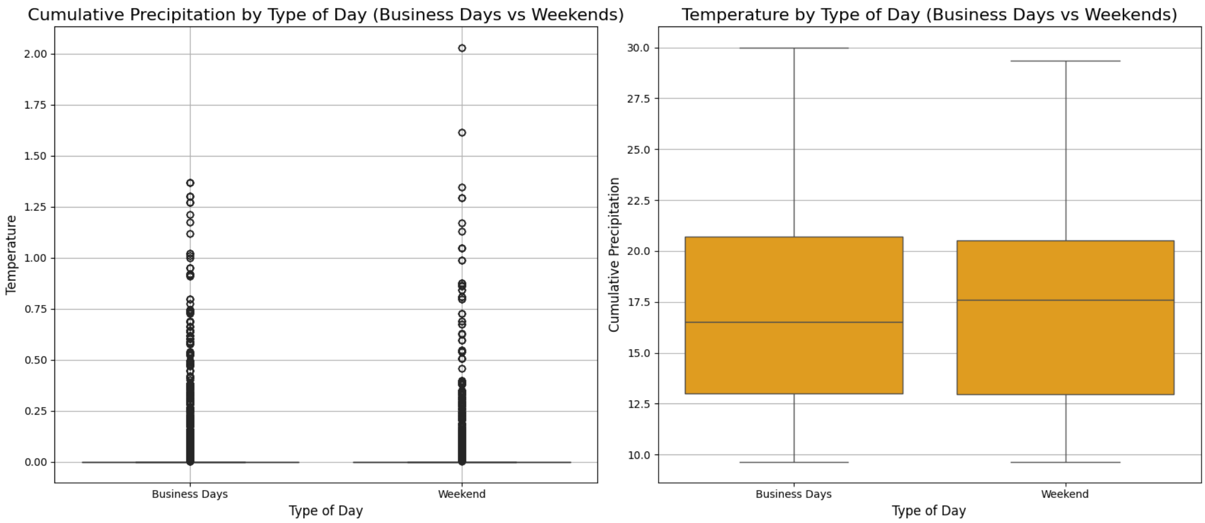

Figure 21 - Cumulative Precipitation and Temperature values by Type of Day

Figure 21 illustrates a comparison of cumulative precipitation and temperature on business days versus weekends.

The left line, depicting cumulative precipitation, indicates slight variations between weekdays and weekends. The median precipitation is marginally elevated on weekdays; nevertheless, the overall distribution stays consistent. Both categories of days demonstrate modest median values, with some outliers attaining up to 2 mm. This suggests that rainfall is rather uniform irrespective of the day type, with only sporadic occurrences of more intense precipitation.

The right graph, depicting temperature by day type, illustrates a comparably steady distribution across weekdays and weekends. The median temperature is roughly 17°C for both categories of days, with an interquartile range of approximately 13°C to 21°C. Nevertheless, weekdays exhibit a greater number of outliers, with diminished temperature readings recorded throughout this period, indicating that weekdays may undergo irregular temperature variations.

In summary, cumulative precipitation is uniform over weekdays and weekends, although temperature exhibits greater variability on weekdays, occasionally reaching lower values. Nonetheless, both variables exhibit consistent patterns with negligible fluctuations depending on the type of day.

Also, the Accumulative Precipitation variable was dropped due to its low variance of 0.013, as evidenced in the Data Understanding phase. It led us to use the temperature as the only Meteorological variable for the study, with a variance of 19.51.

3.3.4. Traffic Data

This dataset had been provided by the Lisbon City Council, which documented the reports from Waze program users concerning congestion in Lisbon. Table 5 presents only the variable employed for calculating the composite indicator. The full list of variables is available in Table 16 of the Annex.

Table 5 - Traffic Dataset

Variable Name	Description
DateTime	Date and Time
Position	Coordinates of the streets
Reports	Number of Reports
Delay	Delay of the jam, seconds
Speed	The average speed of the congestion, km/h

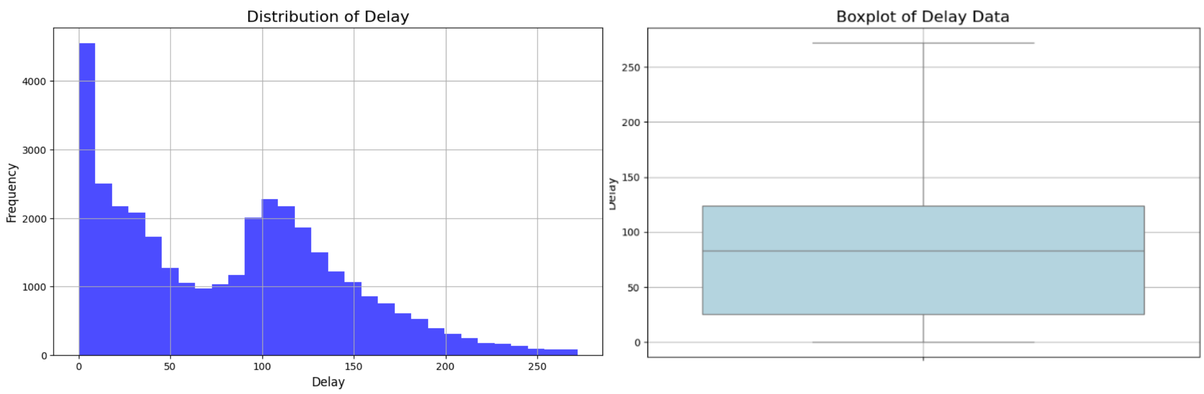

Figure 22 - Distribution of Delay Values

Figure 22 illustrates the distribution and variety of delay times.

The histogram on the left illustrates a right-skewed distribution of delay values, characterized by a significant concentration of observations close to zero. This signifies that a substantial segment of the data has negligible delays, with numerous occurrences nearing zero delays. As delay durations escalate, the frequency declines, indicating that prolonged delays are infrequent yet present within the sample. The histogram additionally reveals a secondary cluster at about 100 seconds, signifying a significant prevalence of moderate delays.

The boxplot on the right enhances the histogram by depicting the variety in the delay data. The median delay is roughly 75 seconds, with an interquartile range (IQR) of 25 to 125 seconds, indicating a significant variability in delay periods. The whiskers signify the presence of elevated delay times, even to 250 seconds. Yet, the absence of severe outliers indicates that even the maximum delays remain within a typical range for this dataset.

In summary, these numbers demonstrate that although the majority of delays are brief, a considerable percentage of delays are moderate to big. The distribution exhibits a primarily low delay pattern, interspersed with sporadic occurrences of elevated delays, signifying a variety of traffic circumstances.

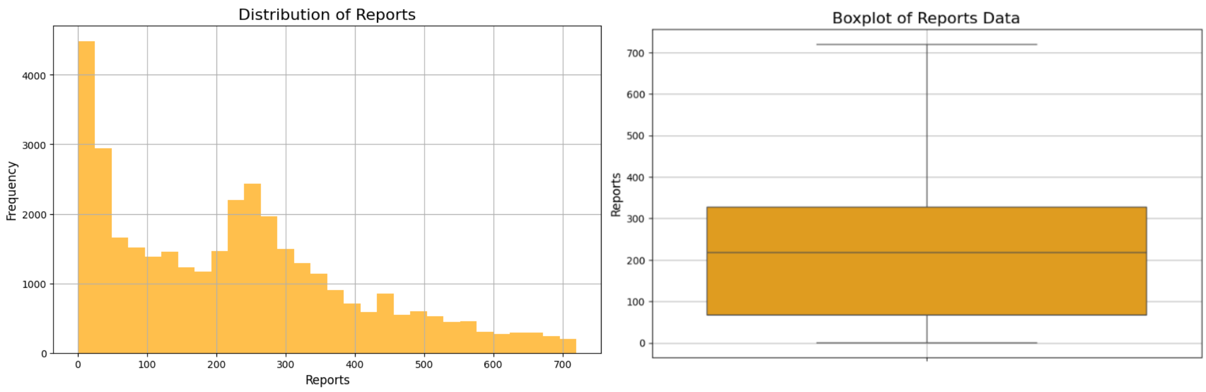

Figure 23 - Distribution of Reports Values

Figure 23 illustrates the distribution and variability of the 'Reports' data.

The histogram on the left illustrates a right-skewed distribution, characterized by a significant concentration of values at lower report counts, especially below 100. This suggests that the majority of records pertain to a limited number of reports, indicating that most reported incidents or events are relatively modest. The distribution displays a long tail toward elevated values, signifying that although high-report occurrences are infrequent, they do transpire infrequently, with values ascending to 700 reports.

The boxplot on the right summarizes the report's distribution. The median number of reports is approximately 200, indicating that half of the data fall below this figure. The interquartile range (IQR) is extensive, indicating considerable variability in the number of reports per occurrence. The presence of an elongated upper whisker indicates a greater frequency of reports; however, severe outliers are absent. This distribution indicates that although the majority of events have a moderate number of reports, there are sporadic occurrences of significantly higher report counts.

The data indicates a prevalence of low-report occurrences, with variability extending to higher values. This distribution reveals that although the majority of incidents are minor, certain instances generate a disproportionately higher volume of reports, indicating a bigger impact or scrutiny.

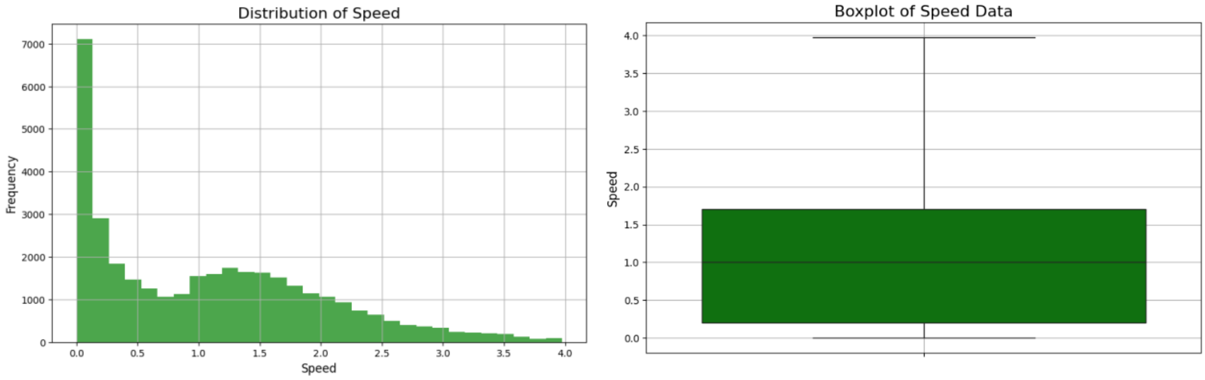

Figure 24 - Distribution of Speed Values

Figure 24 illustrates the distribution and variability of the 'Speed' variable.

The histogram on the left demonstrates a right-skewed distribution for 'Speed', characterized by a significant concentration of observations close to zero. Approximately 7,000 cases are recorded in the lowest speed category, suggesting that the majority of incidents pertain to very low speeds or stationary conditions, possibly indicative of high traffic or congestion. As speed values rise, the frequency of occurrences declines significantly, indicating that elevated speeds are rare within the dataset.

The boxplot on the right corroborates this observation, revealing a median speed of 1 km/h, signifying that over half of the data is below this speed threshold. The interquartile range (IQR) extends from 0 km/h to roughly 4 km/h, encompassing the majority of the speed data within this limited range. The range increases marginally without any noticeable outliers, demonstrating that although some variability exists, the majority of speeds remain low and do not diverge substantially from this pattern.

The distribution and boxplot of 'Speed' data indicate that most recorded incidents pertain to low speeds, exhibiting limited variance at higher speeds. This pattern indicates a scenario of recurrent slow or motionless traffic, with few occurrences of heightened speed, underscoring widespread congestion within the dataset.

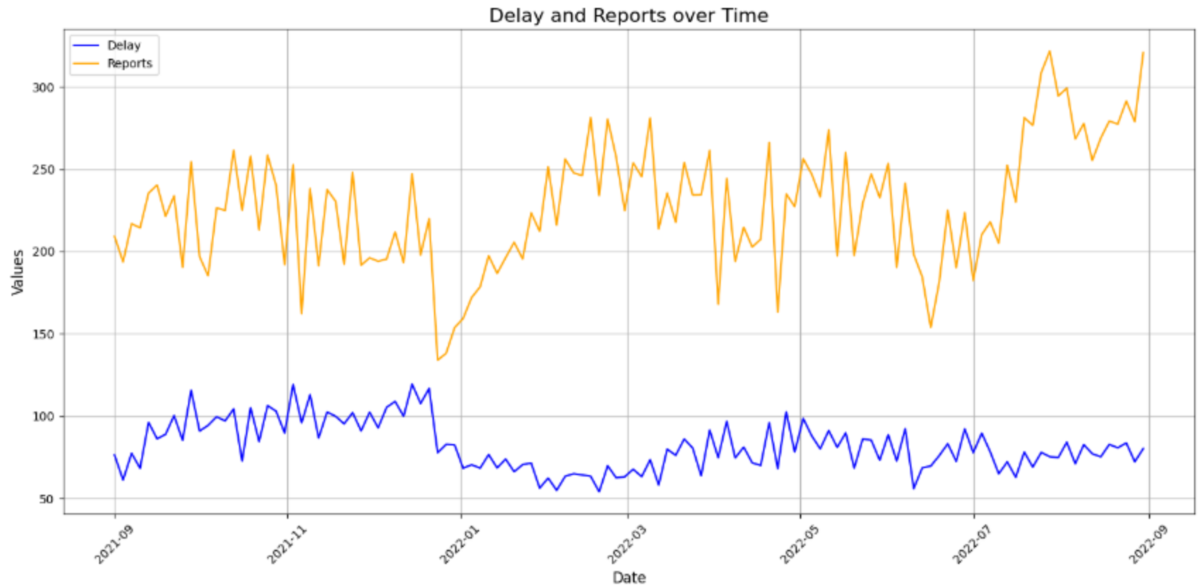

Figure 25 – Delay and Reports Variation Over Time (Average)

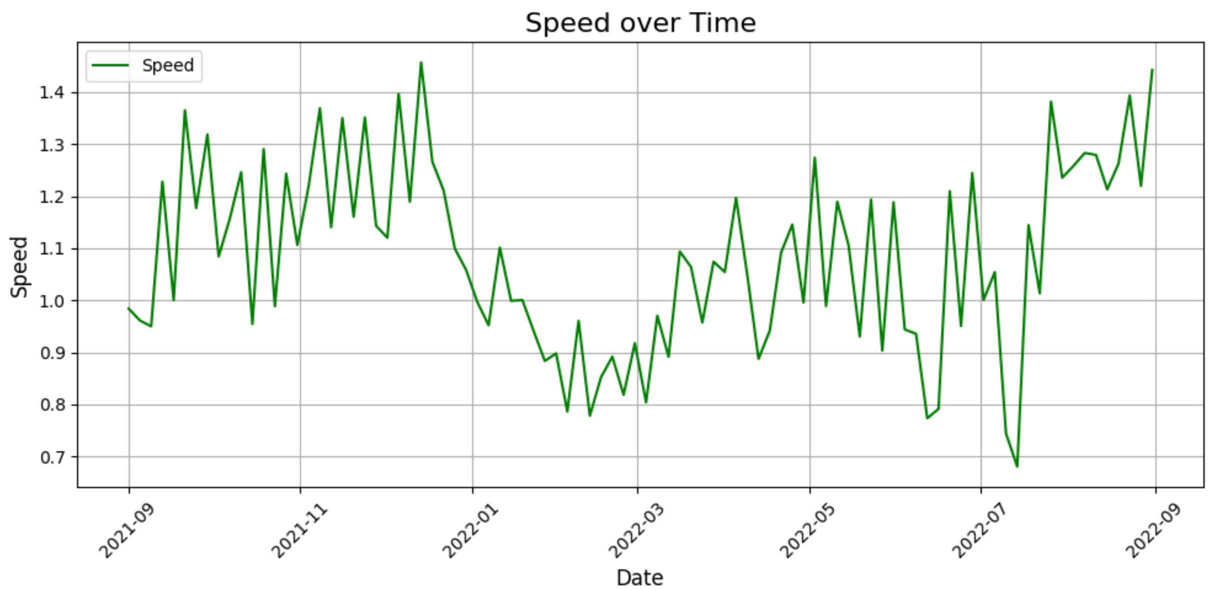

Figure 26 - Speed Variation Over Time (Average)

Figures 25 and 26 depict the temporal behavior of delay, reports, and speed, emphasizing significant tendencies within the system.

The initial graph illustrates that the orange line denotes the number of Reports, exhibiting a distinct increasing trajectory, exceeding 300 by the conclusion of the monitored timeframe.

The rise in reports indicates an escalating frequency of recorded incidents or difficulties over time. Conversely, the blue line representing delay remained comparatively consistent, exhibiting slight oscillations, which suggests that despite the increase in reported incidents, the delay values were sustained without substantial escalation. This stability indicates that the system managed the increasing frequency of reports without affecting wait times, demonstrating resilience in performance.

The second graph emphasizes speed, indicating a significant decrease commencing in December 2021 and culminating in a trough in March 2022. After this low point, velocity progressively escalates during the remainder of 2022. The initial decline may correspond with the increase in reports depicted in the first graph, while the later rise in speed could signify system adaptation or operational enhancements that resulted in a restoration of traffic flow.

Collectively, these figures demonstrate that although reports rose, signifying a higher number of recorded occurrences, the system sustained steady delay durations and ultimately enhanced speed, reflecting resilience and adaptation under heightened operational pressure.

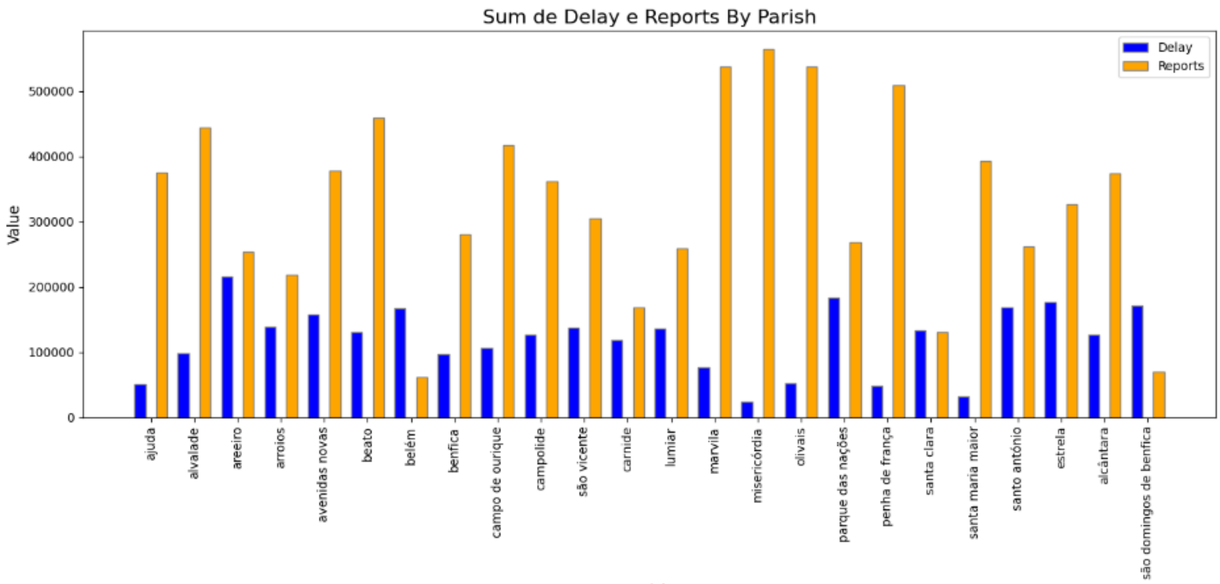

Figure 27 – Delay and Reports values by Parish (Sum)

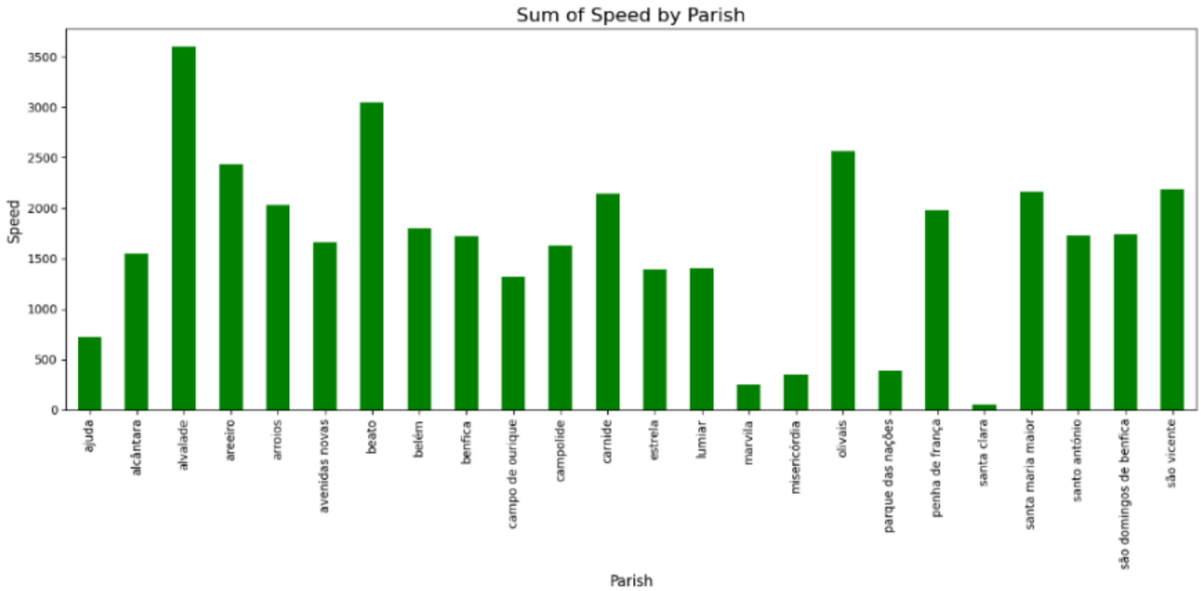

Figure 28 – Speed values by Parish (Sum)

Figures 27 and 28 present an examination of delay, reports, and speed across different parishes.

The initial graph juxtaposes the cumulative totals of delays and reports by parish. Reports (shown in orange) regularly exceed the delay values (represented in blue) throughout the majority of parishes, suggesting that although many occurrences are documented, they do not necessarily result in heightened delays. In parishes such as Santa Clara and São Domingos de Benfica, the disparity between reports and delays is minimal, indicating that increased report quantities in these regions may more directly influence traffic delays. In contrast, parishes like Olivais, Misericórdia, Penha de França, and Marvila demonstrate low values for both Reports and Delay, suggesting reduced traffic-related events and potentially superior traffic conditions.

The second graph emphasizes Speed, demonstrating significant heterogeneity across parishes. Alvalade, Arroios, and Avenidas Novas exhibit the greatest cumulative speed metrics, indicating a more fluid traffic flow in these regions. Conversely, parishes such as Lumiar, Penha de França, and Marvila exhibit reduced speeds, indicating more congested or less efficient traffic conditions. Speed serves as a direct indicator of traffic fluidity, rendering it a significant metric for assessing congestion.

An analysis of the two figures uncovers intriguing patterns. Parishes such as Avenidas Novas and Benfica have elevated report numbers alongside comparatively high speeds, suggesting that, despite recurrent occurrences, traffic flow continues predominantly unimpeded. In contrast, parishes like Lumiar and Penha de França have elevated report counts alongside reduced speed values, indicating that these incidents may play a more substantial role in traffic congestion in these regions.

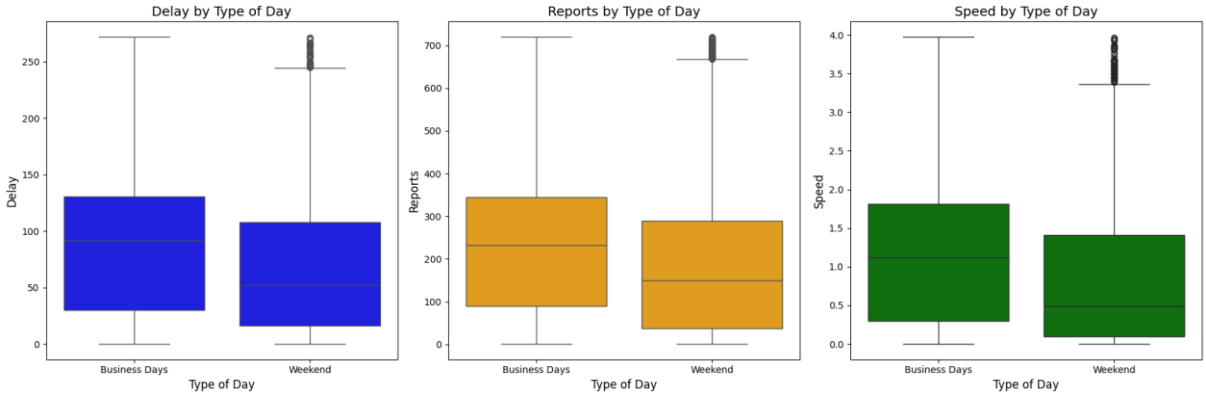

Figure 29 – Delay, Reports, and Speed values by Type of Day

Figure 29 presents three boxplots that compare Delay, Reports, and Speed between weekdays and weekends, facilitating an analysis of traffic patterns and reported incidents over several days.

The initial boxplot, illustrating Delay, indicates that median delay values are somewhat elevated on weekdays compared to weekends. Both categories of days, however, demonstrate a broad spectrum of delay values. Significant outliers on weekends indicate occasional large delays, even when traffic is normally lighter. This indicates that some events can result in significant delays irrespective of the day.

The second boxplot, illustrating Reports, indicates comparable median values for weekdays and weekends, with a little increase on weekdays. Weekends exhibit a greater frequency of outliers, signifying irregular occurrences that result in an increased volume of reports on particular weekends. The constancy in median reports, along with rare peaks, indicates that reported concerns are predominantly constant across days, with certain deviations on weekends.

The final boxplot, depicting Speed, reveals a significant disparity between weekdays and weekends. The median speed is significantly elevated on weekends, exhibiting a more concentrated distribution that reflects less variability in speeds relative to weekdays. Weekends exhibit a greater number of outliers, with certain occurrences of exceptionally high speeds presumably indicative of reduced traffic conditions on those days.

In summary, these boxplots demonstrate that although delays and reports remain generally consistent during weekdays and weekends, speed notably increases on weekends, indicating improved traffic conditions. Outliers in delays and weekend reports indicate sporadic incidents affecting traffic, while elevated speeds on weekends reflect the usual decrease in congestion characteristic of these days.

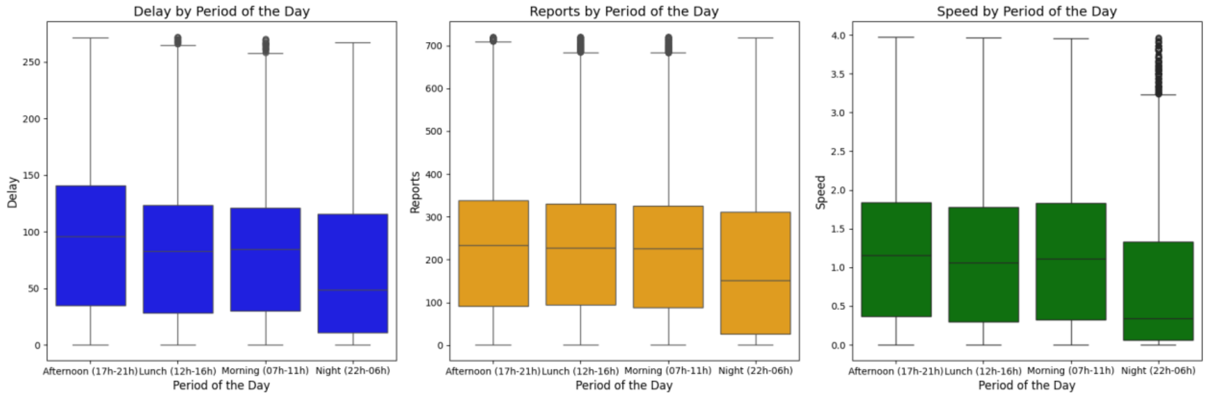

Figure 30 - Delay, Reports, and Speed values by Period of the Day

Figure 30 contrasts delay, reports, and speed during several times of the day: morning, lunch, afternoon, and night, highlighting unique patterns for each variable.

The delay boxplot indicates consistent delay times across the morning, lunch, and afternoon intervals, with similar medians and a wide range of values. Nonetheless, delay durations are markedly reduced during nighttime, exhibiting a significant decrease in variability, which signifies fewer traffic interruptions and a more uniform traffic flow during this interval.

The boxplot of reports demonstrates uniformity during the morning, lunch, and afternoon intervals, with a minor decline during the night. The existence of outliers during daytime hours indicates sporadic surges in reported incidents or traffic problems at particular intervals. The nighttime period exhibits a lower volume of reports, suggesting a reduced incidence of recorded events or difficulties.

The speed boxplot demonstrates a generally stable range of speeds throughout daylight hours, with low change. Nevertheless, velocities significantly decrease at night, exhibiting a reduced range and an elevated prevalence of outliers, indicating occasional occurrences of heightened speed. This pattern indicates reduced traffic volumes during nighttime, resulting in both slower and faster-moving automobiles.

In summary, the morning, lunch, and afternoon intervals display analogous features, characterized by increased delays, a greater number of reports, and moderate speeds. Conversely, nighttime exhibited diminished delays, fewer incidents, and marginally reduced speeds, signifying lighter traffic conditions and fewer occurrences during these hours.

3.4. DATA PREPROCESSING

This section offers a detailed explanation of the diverse methodologies and technologies employed for cleaning, integrating, and transforming data.

According to García et al. (2015), failing to prepare data can lead to significant problems, such as data inconsistency, information loss, and increased noise. These challenges may consequently result in inaccurate or biased conclusions during data analysis.

3.4.1. Temporal Reduction

Table 6 lists the time series, temporal dimension, frequency, and aggregation provided by multiple sources.

Table 6 - Temporal Reduction

Dataset	Time Frame	Frequency	Aggregation
Mobility Data	01/09/2021 to 31/08/2022	5 minute-interval	sum
Waze Data	01/01/2018 to 31/12/2023	millisecond	Sum; mean
Environmental Data	01/06/2021 to 15/01/2024	hourly	mean
Meteorological Data	01/12/2018 to 31/12/2023	hourly	mean

The data points included for constructing the composite indicator were restricted to the period from September 1, 2021, to August 31, 2022. This guarantees that the indicator values uniformly encompass all variables by addressing the timeframe during which the diverse time series converge.

Furthermore, due to the extensive data from Vodafone and Waze, comprising over 2 billion data points gathered at 5-minute and millisecond intervals, respectively, it was essential to consolidate this data to minimize the computational resources needed for analyzing the composite indicator. Consequently, the data has been aggregated by day, time frames, and parishes, with the designated times being Morning (7:00-11:00), Lunch (12:00-16:00), Afternoon (17:00-21:00), and Night (22:00-06:00).

3.4.2. Data Imputation

The Nearest Neighbors Approach offers several advantages for the imputation of missing data in urban studies, making it the most suitable choice for this research. Firstly, it upholds spatial autocorrelation, thereby ensuring the preservation of natural geographic connections among neighboring regions, which is crucial for comprehending urban dynamics. This methodology enhances context-sensitive imputation by considering geographic closeness and the resemblance in essential variables, including mobility and environmental factors, which has been demonstrated to elevate imputation precision in spatial datasets (Li & Parker, 2014).

In contrast to techniques like mean imputation or regression, which may overlook spatial relationships, the Nearest Neighbors Approach ensures that imputed values are relevant to their context. This method effectively diminishes the risk of bias or oversimplification of complex urban patterns, as evidenced by its capacity to maintain data structure and reduce distortions in analogous situations (Beretta & Santaniello, 2016). Moreover, its capacity to uphold both geographical and data-driven continuity has been substantiated through spatial data similarity measures that guarantee precise classification and decision-making in spatial issues (Liao, Hou, & Jiang, 2019).

The Nearest Neighbors Approach posits that geographically adjacent regions exhibit analogous urban traits, rendering it an efficient technique for imputing absent data in spatial datasets (Li, Y. et al., 2014).

Implementation details:

- A spatial join was performed using GeoPandas to link the environmental sensor data with the geographic coordinates of the parishes. This stage guaranteed the precise alignment of each data point with its respective spatial entity, facilitating the integration of mobility, environmental, and traffic data across Lisbon.
- Proximity Assessment: The closest neighbors for each parish were determined using geometric distance and data similarity. This methodology ensured that imputation was informed by geographic and contextual similarities, a practice corroborated by research in environmental monitoring and urban analysis (Liao et al., 2019).
- Data Imputation: Missing data points from environmental and meteorological datasets, where there weren't sensors and stations in those parishes, were filled by averaging values from the closest nearby parishes, a technique demonstrated to effectively maintain local spatial patterns and reduce bias (Beretta & Santaniello, 2016). This methodology was crucial for preserving the precision of high-frequency datasets, including movement and environmental data, where sensor failures or partial records frequently occurred. By imputing values according to spatial and contextual similarities, the integrity of the dataset's spatial correlations was maintained.

3.4.3. Data Treatment and Cleaning

3.4.3.1. Outliers Treatment

The Interquartile Range (IQR) method is a prevalent statistical strategy for detecting and addressing outliers in datasets (Han, Kamber e Pei, 2012). The measurement is derived from the dispersion of data between the first quartile (Q1), indicating the threshold below which 25 percent of the data resides, and the third quartile (Q3), denoting the threshold below which 75 percent of the data resides. The disparity between these two values is referred to as the interquartile range (IQR), which indicates the middle dispersion of the data while omitting the most extreme values (Han, Kamber e Pei, 2012).

Outliers are identified by calculating the lower and upper boundaries using the interquartile range (IQR). The lower boundary is defined as $Q1 - 1.5\ IQR$, and the upper boundary is $Q3 + 1.5\, IQR$. Any figure that falls below the lower threshold or exceeds the upper threshold is classified as an outlier. This method effectively identifies values that deviate from the data set's dispersion pattern without depending on distribution assumptions (Han, Kamber e Pei, 2012).

3.4.3.1.1. Mobility Dataset

Outliers in the mobility dataset were determined using the interquartile range (IQR). Values below the lower boundary, $Q1 - 1.5\ IQR$, and the upper boundary, $Q3 + 1.5\, IQR$, were deemed outliers. For any value over this range, the algorithm computed a moving average utilizing the values from the 7 days preceding and 7 days succeeding the day of the outlier, omitting the present date and any additional values that also exceed the restrictions. The technique was iteratively executed until all outliers were eradicated. If any values exceeded the top limit at the conclusion of this operation, they were substituted with the maximum permissible value within the range established by the boxplot, so guaranteeing that no values remained too elevated.

3.4.3.1.2. Environmental Dataset

In the dataset environmental dataset, which includes the variable O_3 and Noise, the treatment adhered to the same rationale. Outliers were determined using the interquartile range, and the extreme values were substituted with the average of the adjacent values within a 14-day period (7 days prior and 7 days subsequent), eliminating both the outliers and the value for the current day. The technique was iteratively repeated until no further outliers were detected. Upon conclusion of the treatment, all values beyond the upper limit were substituted with the maximum value within the boxplot range, so preserving the integrity of the data distribution.

3.4.3.1.3. Traffic Dataset

Ultimately, in the traffic dataset, which encompasses the delay, reports, and speed variables, outlier treatment was implemented for each of these columns. The interquartile range limits for each variable were computed to identify outliers, which were subsequently substituted with the average of the values within the 14-day interval around the original value. The procedure was reiterated until no outliers persisted, and ultimately, values over the upper threshold were substituted with the maximum permissible value shown by the boxplot. This method guaranteed that outliers did not adversely affect subsequent studies, maintaining the integrity of the delay, reports, and speed variables while assuring a regulated distribution of the data.

The outlier treatment method was uniformly applied to all datasets with outliers, enabling extreme values to be substituted with estimates derived from moving averages, thus mitigating the influence of outliers and enhancing the data quality for more rigorous studies.

3.4.3.1.4. Meteorological Data

Despite there weren't any outliers, there were missing values in this dataset. The temperature missing values have been replaced by the average monthly temperature for the month, taken from the IPMA monthly Bulletin (IPMA, 2021; IPMA, 2022).

3.5. Methods

This chapter provides a detailed explanation of the methods employed to compute the urban dynamic indicator. It elucidates the justification for selecting critical components and the methodology for modeling seasonal impacts, factor analysis, and Principal Component Analysis using varimax rotation. The analysis preserves clarity and relevance by concentrating exclusively on the chosen variables in urban transportation dynamics.

For transparency, variables that were initially evaluated but eventually excluded are recorded in Appendix A, each with a rationale for its absence from the calculation of UDI. This chapter is designed to provide a targeted and methodical approach that corresponds with the project's goals.

3.5.1. Seasonality Effects

Following the work by Jardim, *et al* (2023), this work utilized the Prophet algorithm to predict seasonality in urban dynamics, selected for its adaptability in managing missing data and its superior capability in modeling non-linear trends compared to conventional time series methods such as ARIMA or STL (Taylor & Letham, 2018). The power of Prophet resides in its capacity to partition time series into discrete components—long-term trends, daily or weekly patterns, and annual seasonality—rendering it especially adept for high-frequency urban data characterized by irregular and complicated temporal cycles.

It employed a decomposable time series model (Harvey and Peters, 1990) that consists, in this case, of two primary components: trend and seasonality:

$$y(t)=g(t)+s(t)+\epsilon t$$

In this case, periodic variations (such as daily and weekly seasonality) are represented by s(t), and the trend function g(t) depicts nonperiodic changes in the time series value.

Any peculiar changes that the model cannot account for are represented by the error term εt; we will later assume that εt is normally distributed.

Numerous expected seasonal trends were derived from previous research on urban mobility. Weekdays frequently experience elevated activity levels due to commuting. However, weekends typically display reduced mobility levels. Morning and evening rush-hour peaks, aligned with typical work and school schedules, were anticipated to induce considerable variations in activity. Moreover, seasonal fluctuations, including increased tourist influx during the summer, were anticipated to affect urban mobility. The Prophet's capacity to simulate daily, weekly, and annual seasonality facilitated the effective capturing of these intricate patterns.

The application of Prophet in this study was essential for breaking the time series into components and successfully isolating trend, daily, weekly, and annual seasonal impacts. The program managed outliers without compromising the long-term patterns. Moreover, Prophet's adaptable management of absent data and irregular intervals maintained the temporal integrity of the data, yielding more precise insights into Lisbon's urban dynamics than conventional models such as ARIMA, which necessitate stationarity and are less appropriate for irregular data (Harvey & Peters, 1990).

The outcomes from the Prophet model were illustrated and represented in Figure 31, which shows the variables used in the UDI. This graphic illustrates the modifications implemented to address seasonal impacts, demonstrating the decomposition of the raw time series into its constituent elements. By analyzing these temporal trends, the study developed a more robust and precise UDI, considering the cyclical characteristics of urban movement over daily, weekly, and yearly intervals.

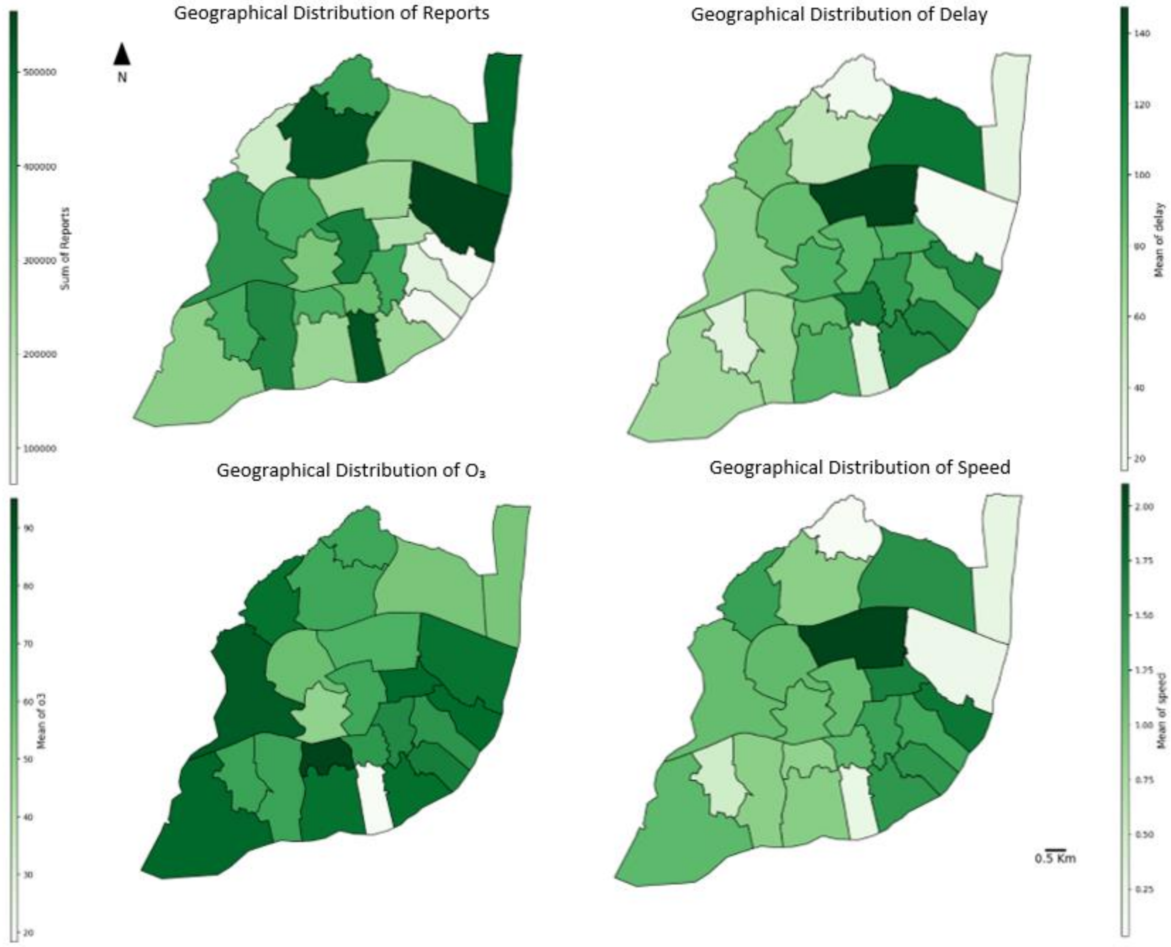

Figure 31 - Spatial distribution of the post-processed values for each series that was employed to estimate the Urban Dynamic Indicator

Addressing seasonal influences is crucial when assessing urban dynamics, as it guarantees a more precise comprehension of the fundamental patterns influencing mobility and environmental alterations. Urban activities fluctuate considerably according to daily schedules, weekly work patterns, and annual occurrences such as vacations and tourism seasons. Disregarding these temporal fluctuations may result in erroneous interpretations, obscuring the genuine determinants of urban behavior. This method guarantees that the identified patterns are not distorted by transient abnormalities but accurately represent the authentic temporal trends influencing Lisbon's urban environment.

3.5.2. Factor Analysis

Factor analysis is a multivariate statistical technique commonly employed to diminish the dimensionality of extensive datasets and to reveal hidden variables that elucidate the underlying patterns within the data (Hashemi *et al,* 2021). This study employed factor analysis to analyze urban dynamics in Lisbon, utilizing many observed factors including traffic delays, noise levels, temperature, and mobility data. Utilizing component analysis, was used to simplify the intricacies of these interrelated factors and create a composite measure—the Urban Dynamic Indicator (UDI)—that encapsulates the overarching trends of urban mobility and environmental impacts (Šoštarić, et al, 2021).

The factor analysis model is mathematically expressed as (Jardim, de Castro Neto, & Calçada, 2023):

$$X=\Lambda F+\xi$$

- X = [X1, …, Xn] represents a collection of n observed time series (or variables), such as the number of traffic reports, temperature, noise levels, and delays. These observed variables can be decomposed into two orthogonal components: the common component and the unique component.
- The common component consists of:
  - F = (f1, …, fi), a vector of i latent factors (or common factors). These factors represent the underlying dimensions that drive the observed variability in the data, such as "urban mobility" or "environmental conditions."
  - Λ = (Λ11, …, Λni), a matrix of factor loadings that quantifies how strongly each latent factor contributes to each observed variable. The factor loadings measure the extent to which the common factors (F) are present in the observed variables (X). Higher factor loadings indicate a stronger relationship between a specific observed variable and a latent factor.
- The unique component is captured by ξ, an error matrix representing residuals. The matrix ξ comprises the discrepancies between the observed variables (X) and their estimated values based on the common factors. This component reflects the unique variance of each variable that cannot be explained by the common factors, thus accounting for the randomness or noise in the data

In simple terms, the objective of factor analysis is to reduce the impact of the unique component ξ while enhancing the explanatory capacity of the common component (F and Λ). This enables us to streamline the data structure, pinpointing the essential latent components that influence urban dynamics while minimizing noise and redundancy in the observed variables (Jolliffe, 2002; Harman, 1976).

3.5.3. Principal Component Analysis with Varimax Rotation

This study utilized Principal Component Analysis (PCA) with varimax rotation to discern the latent elements influencing urban dynamics. The dataset comprised transportation and environmental characteristics, including traffic reports, delays, ozone, and noise levels.

PCA is a dimensional reduction method that was selected for its capacity to diminish dimensionality by converting the original variables into a reduced collection of uncorrelated components that encapsulate the majority of the data's variance (Tran, Burdejová, Ospienko, & Härdle, 2014).

In urban mobility research, elements including traffic congestion, environmental circumstances, and mobility patterns frequently intersect, rendering PCA an effective instrument for uncovering the fundamental structures within the data (Šoštarić, et al, 2021). Nonetheless, a significant drawback of PCA is that the unrotated components may be challenging to read because of intricate loadings (F. R. On et al., 2016).

To resolve this issue, variablemax rotation was employed to the factor loadings. Varimax is an orthogonal rotation technique that aims to enhance the interpretability of factors by maximizing the variance of squared loadings among the factors (Forina et al., 2005). This methodology guarantees that each variable is predominantly linked to a single component, resulting in more lucid interpretations of the latent elements influencing urban transportation and environmental circumstances.

The application of varimax rotation facilitated the interpretation of latent components and reduced ambiguity (Forina et al., 2005). This orthogonal rotation guaranteed that the retrieved components were uncorrelated, hence simplifying the model for further analyses, including the calculation of the Urban Dynamic Indicator (UDI).

3.5.4. Suitability of Data for Factor Analysis

Before executing PCA, two assessments were undertaken to evaluate the dataset's appropriateness for factor analysis: the Kaiser-Meyer-Olkin (KMO) test and Bartlett's test of sphericity.

- The KMO test assesses sampling adequacy by evaluating the correlation and suitability of variables for factor analysis (Kaiser, 1981). The KMO value of 0.56 in this study suggested that the dataset was sufficient for factor analysis, however just below the ideal threshold of 0.7 for more robust outcomes. This score indicates that the variables possess adequate common variance for the identification of significant components.
- Bartlett's test of sphericity was employed to confirm that the correlation matrix was not an identity matrix, indicating the presence of links among the variables. A significant

Bartlett's test ($p < 0.05$) validated that the correlations among the observed variables were sufficiently robust to warrant the application of factor analysis (Bartlett, 1950). This outcome demonstrated that the dataset possessed adequate structure to discern significant latent components.

Both tests validated the suitability of employing PCA on the dataset, confirming that the variables were well connected and that factor analysis would yield dependable insights.

3.5.5. Eigenvalues and Factor Selection

The determination of the number of variables to maintain in PCA is contingent upon the eigenvalues of the principal components. An eigenvalue quantifies the proportion of total variation in the dataset attributed to each element. Factors possessing eigenvalues exceeding 1 are often preserved, following Kaiser's criterion (Kaiser, 1960). In this investigation, the initial two factors exhibited eigenvalues over 1, so validating their selection for subsequent analysis. The application of eigenvalues for factor selection guaranteed the retention of just the most significant components, rendering the UDI an effective instrument for summarizing urban activity. The two components collectively accounted for a substantial fraction of the total variance in the dataset and were viewed as indicative of the mobility and environmental dimensions of urban dynamics.

A scree plot, present in Figure 32, was created to depict the eigenvalues visually. The plot exhibited a significant decline following the second component, corroborating the appropriateness of retaining two factors. The scree plot facilitated the retention of only the most significant components—those that contributed substantially to the variation—while eliminating factors that accounted for negligible variance.

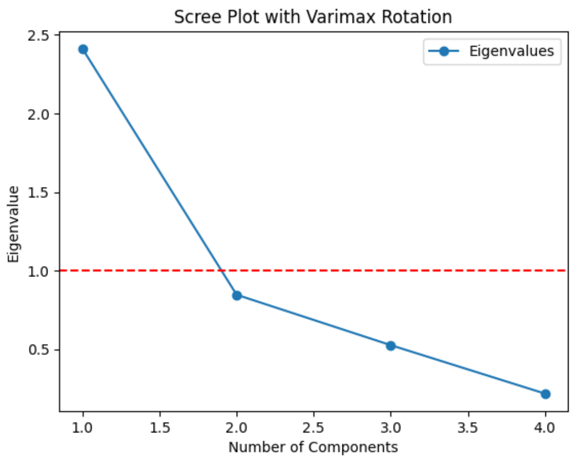

Figure 32 - Scree Plot with Varimax Rotation

PCA was executed in the code with the `FactorAnalyzer` package, which initially normalized the variables to ensure comparability in scale. This stage is essential in PCA, as variables evaluated on disparate scales (such as traffic reports and temperature) can unduly influence the outcomes. Standardization guarantees that all variables contribute uniformly to the identification of factors.

Following normalization, Principal Component Analysis (PCA) was conducted to identify latent factors, and eigenvalues were computed for each factor. Only factors with eigenvalues exceeding 1 were preserved, by Kaiser's criterion. Subsequently, Varimax rotation was employed on these factors to facilitate the interpretation of the loadings, ensuring that each variable was mostly linked to a single component.

The factor loadings (Λ) in Table 7 are derived from the varimax rotation that was used to calculate the Urban Dynamic Indicator.

Table 7 - Factor Loadings

Variable	Loading
Reports	0.55
Traffic Flow	0.49
Air Quality	0.52
Noise Level	0.43

The formula for the UDI derived from the factor loadings was:

UDI = (0.55 × Reports) + (0.49 × Traffic Flow) + (0.52 × Air Quality) + (0.43 × Noise Level) + ξ

In this formula:

- Reports and traffic flow signify elements of urban mobility. The positive loading of 0.55 for Reports signifies that a rise in recorded incidents or events within the city is associated with elevated urban activity. The positive loading of 0.49 for Traffic Flow indicates that elevated traffic levels correlate with enhanced urban mobility, underscoring the intensity of transportation activity throughout the city.
- Air Quality and Noise Levels include environmental impacts on urban dynamics. The positive loading of 0.52 for O_3 levels indicates that elevated ozone concentrations are associated with certain urban activities. Simultaneously, the positive loading of 0.43 for Noise Level suggests that heightened noise correlates with a more dynamic urban setting, yet may also imply disturbances in urban livability.

4. RESULTS AND DISCUSSION

4.1. URBAN DYNAMIC INDICATOR

Table 8 provides the UDI's value by parish, per day, and time frame in Lisbon. Also, Figure 33 provides the average UDI value represented in the map.

Table 8 - UDI per parish, day type, and period of the day (Average)

	Business Days				Weekend			
	Morning	Lunch (12h-	Afternoon	Nigh (22h-	Morning (7h-11h)	Lunch (12h-	Afternoon	Nigh (22h-
Ajuda	0.58	0.72	0.85	0.45	0.42	0.56	0.69	0.29
Alcântara	0.58	0.71	0.85	0.44	0.42	0.55	0.69	0.28
Alvalade	0.58	0.72	0.86	0.45	0.42	0.56	0.70	0.29
Areeiro	0.58	0.71	0.84	0.44	0.41	0.55	0.68	0.28
Arroios	0.57	0.71	0.84	0.45	0.40	0.55	0.67	0.28
Avenidas Novas	0.57	0.71	0.84	0.44	0.41	0.54	0.69	0.28
Beato	0.57	0.70	0.80	0.44	0.41	0.54	0.68	0.28
Belém	0.57	0.71	0.83	0.43	0.41	0.54	0.67	0.27
Benfica	0.57	0.68	0.84	0.44	0.39	0.53	0.66	0.26
Campo de	0.56	0.70	0.83	0.42	0.40	0.54	0.67	0.27
Campolide	0.57	0.69	0.83	0.43	0.41	0.53	0.68	0.29
Carnide	0.56	0.69	0.82	0.43	0.40	0.51	0.66	0.26
Estrela	0.55	0.69	0.82	0.42	0.39	0.51	0.66	0.26
Lumiar	0.56	0.68	0.81	0.43	0.38	0.53	0.67	0.26
Marvila	0.55	0.69	0.82	0.42	0.39	0.52	0.66	0.26
Mesericórdia	0.55	0.68	0.82	0.42	0.38	0.52	0.66	0.26
Olivais	0.56	0.68	0.82	0.41	0.39	0.53	0.65	0.25
Parque das	0.57	0.71	0.84	0.41	0.40	0.52	0.67	0.27
Penha de	0.54	0.69	0.81	0.41	0.38	0.52	0.65	0.25
Santa Clara	0.54	0.66	0.81	0.41	0.37	0.51	0.65	0.25
Santa Maria	0.54	0.67	0.80	0.41	0.38	0.52	0.63	0.24
Santo António	0.54	0.67	0.80	0.40	0.39	0.51	0.64	0.24
São Domingos de Benfica	0.53	0.67	0.80	0.41	0.37	0.51	0.64	0.24
São Vicente	0.53	0.67	0.80	0.40	0.37	0.50	0.64	0.24
Total Mean	0.558	0.691	0.823	0.424	0.397	0.529	0.663	0.263
	0.624				0.463			

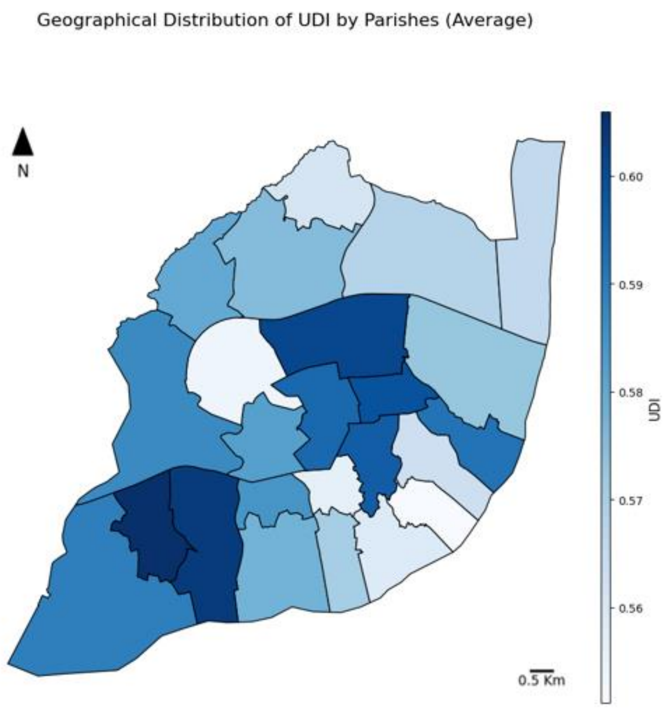

Figure 33 – Urban Dynamic Indicator in Lisbon (average by parish)

An analysis of the data in Table 8 and the map depicting the geographical distribution of the UDI by Lisbon parishes indicates visible variations in urban activity intensity, influenced by the time of day, differences between weekdays and weekends, and disparities among parishes. In Table 8, it is possible to see the analysis of the average UDI values throughout various times of the day reveals that, on weekdays, the afternoon period regularly exhibits the greatest average values (0.823), whereas the evening period records the lowest values (0.424). This trend is similarly evident on weekends, with the afternoon representing the peak activity period (0.663) and the evening denoting the lowest activity period (0.263). The overall average UDI value substantiates this pattern, with weekdays exhibiting a higher average (0.624) compared to weekends (0.463).

Upon individual analysis of the parishes, it is evident that several areas, such Ajuda, Alcântara, Belém, and Areeiro, have elevated UDI values during the afternoon on weekdays, with averages equal to or exceeding 0.85. Conversely, parishes like São Vicente and Santa Clara typically have lower numbers, particularly during the evening hours. These patterns reveal substantial disparities in urban activity dynamics among parishes, likely reflecting unique land use and occupancy characteristics in each region.

The UDI geographical distribution map depicts the average per parish using a color scale, with darker hues signifying higher averages and lighter hues indicating lower averages. Parishes such as Benfica and Belém exhibit elevated average UDI values, indicating comparatively vigorous urban activity relative to other regions. Conversely, parishes like São Vicente and Santa Clara exhibit diminished values, signifying a reduced intensity of urban activity.

The map indicates that urban activity is concentrated in the south-west of Lisbon, where greater UDI values are observed, whereas the northern and eastern regions exhibit lower values. This fluctuation, along with the temporal analysis of the data, provides an intricate perspective on urban dynamics in Lisbon, enabling the identification of distinct activity patterns that fluctuate based on the time of day, the nature of the day, and the location of the parishes. This data enhances the understanding of the unique characteristics of each parish regarding urban activities throughout time and can provide a robust foundation for future analysis and urban planning in the city.

4.2. VARIABLE WEIGHT

In order to comprehend the elements influencing the values of the Urban Dynamics Indicator, we examine the impact of variables such as the Number of Reports, Traffic Flow, Noise Level, and Air Quality on the urban dynamic of each parish. These factors influence the UDI readings under their loading values, where the number of reports is the variable with the highest average weight, followed by delay, temperature, and noise, as shown in Table 9.

Table 9 – Weight (%) of each variable in the UDI per day type and period

	Business Days				Weekend			
	Morning (7h-11h)	Lunch (12h-16h)	Afternoon (17h-21h)	Nigh (22h-6h)	Morning (7h-11h)	Lunch (12h-16h)	Afternoon (17h-21h)	Nigh (22h-6h)
Reports	52,35	51,97	51,62	52,80	49,13	48,93	48,75	49,37
Traffic Flow	18,81	19,25	19,63	18,31	16,77	17,43	18,00	15,98
Noise Level	13,59	13,12	12,72	14,12	15,78	15,07	14,46	16,72
Air Quality	15,25	15,66	16,03	14,77	18,32	18,57	18,79	18,02

Table 9 presents the percentage distribution of the variables constituting the UDI (Urban Dynamic Indicator) at various periods during the day, contrasting weekdays with weekends. During weekdays, the morning interval is marked by a significant contribution from 'Reports' (52.35%), indicating a boost in urban activity associated with the day's commencement. This score remains elevated across all working days, exhibiting little fluctuations, which suggests that the frequency of activity reports is both consistent and substantial. Conversely, the impacts of 'Traffic Flow,' 'Noise Level,' and 'Air Quality' are comparatively negligible, with 'Traffic Flow' being more pronounced between lunchtime and in the afternoon indicating a rise in the movement of individuals and vehicles during these intervals. During the weekend, the distribution of 'Reports' stays elevated, albeit marginally reduced compared to weekdays, whilst the 'Air Quality' variable assumes higher significance, especially in the morning and afternoon, potentially indicating diminished environmental stress and an increased focus on recreational areas.

Table 10 - Weight (%) of each variable in the UDI per parish

	Reports	Traffic Flow	Noise Level	Air Quality
Ajuda	51,79	18,08	13,44	16,69
Alcântara	51,76	18,12	13,47	16,65
Alvalade	51,71	18,16	13,51	16,61
Areeiro	51.68	18,20	13,54	16,56
Arroios	51,66	18,24	13,58	16,52
Avenidas Novas	51,63	18,28	13,62	16,48
Beato	51,59	18,32	13,65	16,44
Belém	51,55	18,36	13,69	16,40
Benfica	51,52	18,40	13,73	16,36
Campo de Ourique	51,52	18,44	13,77	16,32
Campolide	51,45	18,48	13,80	16,27
Carnide	51,43	18,55	13,87	16,25
Estrela	51,38	18,56	13,88	16,18
Lumiar	51,35	18,60	13,92	16,14
Marvila	51,31	18,64	13,96	16,10
Mesericórdia	51,27	18,68	13,99	16,05
Olivais	51,23	18,72	14,03	16,01
Parque das Nações	51,20	18,76	14,07	15,96
Penha de França	51,16	18,81	14,11	15,92
Santa Clara	51,12	18,85	14,15	15,87
Santa Maria Maior	51,09	18,89	14,19	15,82
Santo António	51,05	18,94	14,23	15,78
São Domingos de Benfica	51,01	18,98	14,27	15,73
São Vicente	50,97	19,02	14,32	15,69

Table 10 illustrates the percentage contributions of each variable to the UDI within each parish. The weight of 'Reports' is consistently around 51 percent throughout all parishes, showing its major role in evaluating urban dynamics, irrespective of parish. Parishes like Alvalade, Areeiro, and Arroios, characterized by heightened urban activity, exhibit moderate levels of 'Noise Level' and 'Traffic Flow', indicating efficient management of both traffic and noise, elements that likely enhance their elevated UDI values. Conversely, parishes like São Vicente and Misericórdia, characterized by elevated 'Noise Level' and 'Traffic Flow' metrics, may have difficulties in mobility management and noise regulation, indicative of diminished UDI performance.

In summary, the combined analysis of Tables 9 and 10 reveals that, during weekdays, 'Reports' is the predominant variable affecting the UDI throughout the day, with minor fluctuations in the lunch and afternoon hours attributed to 'Traffic Flow'. The variable distribution among parishes indicates that regions characterized by significant urban activity and effective traffic and noise management, such as Alvalade and Arroios, attain superior UDI outcomes. Conversely, parishes significantly affected by 'Noise Level' and 'Traffic Flow' generally have a diminished UDI, underscoring the necessity for localized measures to enhance mobility and urban life quality in these areas.

4.3. EVALUATE THE MODEL WITH MOBILITY DATA

The combined examination of Table 11 and Figure 34 indicates a substantial association between the UDI and urban activity, as quantified by mobility data, across several parishes during different times of the day, on both weekdays and weekends. The correlation values denote the strength and direction of the link, with asterisks indicating significance levels: '***' for a very significant correlation (p < 0.001), '**' for a moderately significant correlation (p < 0.01), and '*' for a lesser significance (p < 0.05).

Table 11 – Correlation between mobility and the UDI by parish, day type, and period

	Business Days				Weekend			
	Morning (7h-11h)	Lunch (12h-16h)	Afternoon (17h-21h)	Nigh (22h-6h)	Morning (7h-11h)	Lunch (12h-16h)	Afternoon (17h-21h)	Nigh (22h-6h)
Ajuda	0.38***	0.50***	0.34***	0.48***	0.40***	0.48***	0.35***	0.47***
Alcântara	0.45***	0.18***	0.46***	0.13**	0.47***	0.18*	0.42***	0.13
Alvalade	0.44***	0.01	0.44***	0.30***	0.39***	0.04	0.46***	0.27***
Areeiro	0.48***	0.36***	0.46***	0.48***	0.46***	0.39***	0.51***	0.47***
Arroios	0.09	0.50***	0.27***	0.44***	0.11	0.47***	0.33***	0.46***
Avenidas Novas	0.18***	0.43***	0.01	0.49***	0.25**	0.46***	-0.04	0.51***
Beato	0.46***	0.51***	0.34***	0.33***	0.49***	0.46***	0.29***	0.38***
Belém	0.48***	0.05	0.53***	0.03	0.48***	0.02	0.54***	0.06
Benfica	0.51***	0.15**	0.46***	0.30***	0.51***	0.19*	0.42***	0.30***
Campo de Ourique	0.45***	0.39***	0.52***	0.50***	0.39***	0.42***	0.53***	0.47***
Campolide	-0.05	0.45***	-0.03	0.43***	-0.06	0.48***	0.00	0.45***
Carnide	0.32***	0.47***	0.12*	0.45***	0.34***	0.42***	-0.00	0.50***
Estrela	0.47***	0.45***	0.43***	0.20***	0.50***	0.42***	0.41***	0.24**
Lumiar	0.42***	-0.04	0.44***	0.04	0.44***	0.02	0.40***	-0.02
Marvila	0.47***	0.26***	0.45***	0.43***	0.45***	0.33***	0.47***	0.38***
Mesericórdia	0.38***	0.47***	0.44***	0.46***	0.36***	0.49***	0.45***	0.48***
Olivais	0.09	0.45***	0.16***	0.44***	0.07	0.44***	0.18*	0.41***
Parque das Nações	0.37***	0.43***	0.24***	0.45***	0.41***	0.45***	0.20**	0.46***
Penha de França	0.49***	0.43***	0.46***	0.13**	0.46***	0.52	0.65	0.18*
Santa Clara	0.43***	0.10	0.46***	0.18***	0.46***	0.51	0.65	0.07
Santa Maria Maior	0.47***	0.33***	0.45***	0.45***	0.47***	0.51	0.64	0.42***
Santo António	0.18***	0.49***	0.41***	0.47***	0.02	0.51	0.64	0.43***
São Domingos de Benfica	-0.04	0.42***	-0.07	0.43***	-0.05	0.51	0.64	0.46***
São Vicente	0.42***	0.48***	0.29***	0.43***	0.42***	0.50	0.64	0.45***

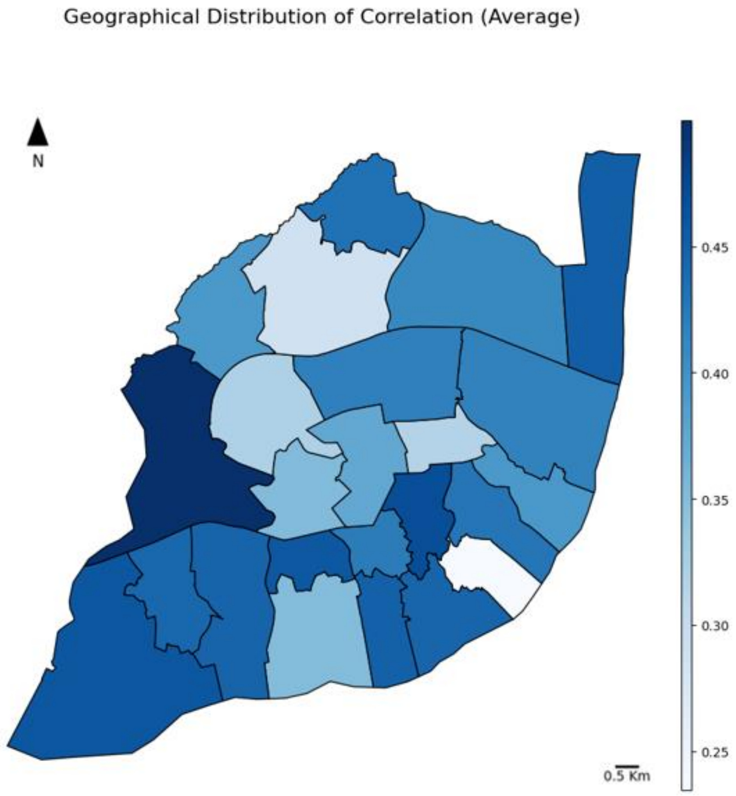

Figure 34 – Correlation between mobility and the UDI by parish (average by parish)

Table 11 indicates that on weekdays, the morning and afternoon hours exhibit the most robust correlations between the UDI and urban activity in various parishes, with values frequently surpassing 0.4 and demonstrating high significance. Parishes like Benfica and Campo de Ourique demonstrate strong and statistically significant connections across nearly all times of the day, indicating a persistent pattern of urban activity in these areas that is effectively represented by the UDI. This trend is validated by the geographical distribution in Figure 28, where these parishes have elevated average correlation values.

The nocturnal hours on weekdays exhibit marginally diminished correlations in certain parishes, including Alcântara (0.13**) and Penha de França (0.13**), potentially indicating a reduced alignment between the UDI and urban activity during these intervals. Nevertheless, certain parishes, including Areeiro and Campo de Ourique, exhibit elevated correlations during nighttime, likely attributable to nocturnal activities in these locales.

During weekends, the correlation between the UDI and urban activity remains significant across various periods, but with some fluctuations, especially during lunchtime (12h-16h), when certain parishes, such as Olivais and São Domingos de Benfica, exhibit diminished correlations. This indicates a potential decline in residential or recreational activity movements, which may affect the UDI's capacity to accurately reflect urban dynamics in these regions and during these periods.

The findings demonstrate that the UDI is a valid metric for depicting urban activity across various temporal and geographical situations. Figure 34 presents a summary of the average correlations throughout all times of the day for each parish, emphasizing regions such as Benfica, Belém, and Campo de Ourique, which have a more robust average association with urban activity. The consistency evident in the relationships presented in Table 11 and Figure 34 substantiates the UDI's reliability as a robust predictor of urban dynamics.

5. CONCLUSIONS, CONTRIBUTIONS, LIMITATIONS, AND FUTURE WORKS

The primary aim of this project was to develop a composite metric, termed the Urban Dynamics Indicator (UDI), designed to evaluate urban mobility in Lisbon from September 2021 to August 2022. This indicator was created to provide a holistic perspective on the interplay and influence of diverse urban elements on the city's dynamics, incorporating both spatial and temporal aspects.

The study discerned distinct patterns of urban activity among parishes and over daily intervals. Parishes like Alvalade, Areeiro, and Arroios sustained elevated levels of activity, distinguishing themselves as hubs of metropolitan engagement. Conversely, regions like São Vicente exhibited lower activity levels, signifying mostly residential attributes. A distinct temporal trend was noted, characterized by heightened activity during the morning peak hours and a subsequent peak in the afternoon. Notwithstanding the overall decline in nocturnal activity, several parishes sustained moderate levels due to vibrant nightlife and recreational pursuits.

The UDI model was validated by a Pearson correlation analysis that compared urban activity, as indicated by Vodafone mobility data, with UDI values. The analysis demonstrated a consistent and positive correlation between these factors throughout different periods and diverse parishes. The majority of connections demonstrated a high level of significance ($p < 0.001$), showing a high degree of confidence in the observed association between the variables of the UDI model in precisely capturing urban dynamics.

This research significantly enhances urban planning by providing a comprehensive instrument for the efficient assessment and monitoring of urban dynamics. The UDI model facilitates a comprehensive understanding of urban activity patterns and provides a reproducible methodology that can be tailored for other cities seeking to enhance their urban planning strategies. Future research may expand data sources by integrating public transport validation data, conducting studies for more recent periods to account for post-pandemic changes, developing predictive models to forecast mobility patterns, and applying the UDI model to other facets of urban life, including public transport and sustainable mobility initiatives.

Policymakers at Lisbon City Council (CML) could use findings to guide targeted strategies. For parishes with persistently low UDI values, like São Vicente during specific periods, it is imperative to examine the underlying factors, such as inadequate public transportation, insufficient commercial activity, or deficient infrastructure. Particular policies may enhance transportation links or promote commercial development through incentives. For parishes exhibiting elevated UDI values, such as Alvalade during peak hours, it is vital to devote additional resources, either augmenting public transport frequency or enhancing pedestrian and cyclist infrastructure. Temporal fluctuations in UDI may inform strategies for nocturnal periods, such as zoning for recreational areas, improving security measures, or upgrading public illumination. Furthermore, the UDI could facilitate the execution of sustainable initiatives, including bike-sharing programs, low-emission zones, or pedestrian-friendly thoroughfares in regions with elevated activity and environmental strains.

Notwithstanding the progress, the research encountered certain limits. The influence of government restrictions enacted in response to the COVID-19 epidemic precluded comparisons with prior and 'normal' movement trends. Data from the post-pandemic period, from September 2022 to August 2023, may illuminate these variations. The mobility data exclusively encompasses devices linked to the operator Vodafone, so omitting users from other operators in Portugal constrains the data's coverage.

Additional issues emerged from the environmental data, which exhibited several outliers and absent values. In certain parishes, sensors failed to record all variables, necessitating the application of values from adjacent parishes to address these deficiencies. Future endeavors should focus on creating a predictive model that can forecast movement patterns between several parishes, so aiding in the anticipation of population densities and potential congestion. Furthermore, including the UDI into additional elements, such as public transportation and sustainable mobility initiatives (e.g., bike-sharing programs), will enhance the analysis with new dimensions.

Principal component analysis (PCA) using varimax rotation is useful for dimensional reduction, although it has limitations, including subjectivity in factor interpretation and variable selection. The identified factor structure, determined using eigenvalue criteria and varimax rotation, may not be universally applicable to other cities or timeframes, necessitating recalibration of the UDI for varying settings due to the heterogeneity of urban traits and difficulties. Inconsistencies in the correlations among variables suggest that certain modifications may be required for some places and timeframes to enhance the model's precision in these circumstances.

In summary, the UDI model has demonstrated its efficacy as a helpful and dependable instrument for analyzing urban dynamics in Lisbon, providing actionable insights and guiding urban policies grounded in empirical facts. This study establishes a robust basis for subsequent research and practical implementations in sustainable urban design.

BIBLIOGRAPHICAL REFERENCES

Albino, V., Berardi, U., & Dangelico, R. (2015). Smart Cities: Definitions, Dimensions, Performance, and Initiatives. Journal of Urban Technology, 22, 21 - 3. https://doi.org/10.1080/10630732.2014.942092.

Aleksander, R., & Paweł, C. (2020). Recent advances in traffic optimisation: Systematic literature review of modern models, methods, and algorithms. IET Intelligent Transport Systems, 14(11), 1740–1758. https://doi.org/10.1049/iet-its.2020.0328.

Alirol, E., Gétaz, L., Stoll, B., Chappuis, F., & Loutan, L. (2011). Urbanisation and infectious diseases in a globalised world. The Lancet. Infectious Diseases, 11, 131 - 141. https://doi.org/10.1016/S1473-3099(10)70223-1.

Alves, L., Mendes, R., Lenzi, E., & Ribeiro, H. (2015). Scale-Adjusted Metrics for Predicting the Evolution of Urban Indicators and Quantifying the Performance of Cities. PLoS ONE, 10. https://doi.org/10.1371/journal.pone.0134862.

Ashrafi, O., Yerushalmi, L., & Haghighat, F. (2013). Greenhouse gas emissions in wastewater treatment plants of the pulp and paper industry: Modeling and simulation. International Journal of Greenhouse Gas Control, 17, 462–472. https://doi.org/10.1016/j.ijggc.2013.06.006.

Astleithner, F., & Hamedinger, A. (2003). Urban Sustainability as a New Form of Governance: Obstacles and Potentials in the Case of Vienna 1. Innovation: The European Journal of Social Science Research, 16, 51 - 75. https://doi.org/10.1080/13511610304510.

Balogun, A.-L., Tella, A., & Baloo, L. (2021). A review of the inter-correlation of climate change, air pollution, and urban sustainability using novel machine learning algorithms and spatial information science. Urban Climate, 40, 100989. https://doi.org/10.1016/j.uclim.2021.100989.

Bagstad, K., & Shammin, R. (2012). Can the Genuine Progress Indicator better inform sustainable regional progress?—A case study for Northeast Ohio. Ecological Indicators, 18, 330-341. https://doi.org/10.1016/J.ECOLIND.2011.11.026.

Baskerville, R., Baiyere, A., Gregor, S., Hevner, A., & Rossi, M. (2018). Design Science Research Contributions: Finding a Balance between Artifact and Theory. J. Assoc. Inf. Syst., 19, 3. https://doi.org/10.17705/1jais.00495.

Baskerville, R., Kaul, M., & Storey, V. (2018). Aesthetics in design science research. European Journal of Information Systems, 27, 140 - 153. https://doi.org/10.1080/0960085X.2017.1395545.

Beretta, L., & Santaniello, A. (2016). Nearest neighbor imputation algorithms: a critical evaluation. BMC Medical Informatics and Decision Making, 16. https://doi.org/10.1186/s12911-016-0318-z.

Bettencourt, L., Lobo, J., Strumsky, D., & West, G. (2010). Urban Scaling and Its Deviations: Revealing the Structure of Wealth, Innovation and Crime across Cities. PLoS ONE, 5. https://doi.org/10.1371/journal.pone.0013541.

Bokolo, A. (2023). Data driven approaches for smart city planning and design: a case scenario on urban data management. Digital Policy, Regulation and Governance. https://doi.org/10.1108/dprg-03-2022-0023.

Button, K. (2002). CITY MANAGEMENT AND URBAN ENVIRONMENTAL INDICATORS. Ecological Economics, 40, 217-233. https://doi.org/10.1016/S0921-8009(01)00255-5.

Barbosa, V. L. M., Faria, L. C., & Cardoso-Leite, E. (2024). Composite-Indicators for Urban Ecosystem Services at Site-Level: A Systematic Literature Review. REVSBAU, 19, e1903. https://doi.org/10.5380/revsbau.v19i0.94342.

Camboim, G., Zawislak, P., & Pufal, N. (2019). Driving elements to make cities smarter: Evidences from European projects. Technological Forecasting and Social Change. https://doi.org/10.1016/J.TECHFORE.2018.09.014.

Câmara Municipal de Lisboa (2018). Lisboa Aberta. Retrieved December 29, 2024, from https://lisboaaberta.cm-lisboa.pt/index.php/pt/.

Chan, J. (2023). Impact of Smart City Initiatives on Urban Planning Strategies in Singapore: An In-Depth Analysis of Technology-Driven Solutions and Their Influence on Sustainable Development and Quality of Life. Journal of Strategic Management. https://doi.org/10.53819/81018102t4200.

Chang, K., Lin, Y., & Su, M. (2022). Effects of long-term exposure to major outdoor air pollutants for the risk of chronic inflammatory airway diseases in adults. Atmospheric Environment, 288, 119296. https://doi.org/10.1016/j.atmosenv.2022.119296.

Coelho, S., Ferreira, J., Rodrigues, V., & Lopes, M. (2022). Source apportionment of air pollution in European urban areas: Lessons from the ClairCity project. Journal of Environmental Management, 320, 115899. https://doi.org/10.1016/j.jenvman.2022.115899.

Correia, D., Marques, J. L., & Teixeira, L. (2022). The state-of-the-art of smart cities in the European Union. Smart Cities, 5(4), 1776-1810.

Craglia, M., Leontidou, L., Nuvolati, G., & Schweikart, J. (2004). Towards the Development of Quality of Life Indicators in the 'Digital' City. Environment and Planning B: Planning and Design, 31, 51 - 64. https://doi.org/10.1068/b12918.

Du, W., Zhang, Q., Chen, Y., & Ye, Z. (2021). An urban short-term traffic flow prediction model based on wavelet neural network with improved whale optimization algorithm. Sustainable Cities and Society, 69, 102858.

EURES, European Commission. (2023). Labour market information: Portugal. https://eures.europa.eu/living-and-working/labour-market-information/labour-market-information-portugal_en.

European Commission. (2011). Report on sustainable urban mobility in the European Union. Brussels: European Commission. https://transport.ec.europa.eu/index_en.

European Union. (2011). European demographic report 2011. Luxembourg: Publications Office of the European Union. https://ec.europa.eu/eurostat/web/products-statistical-books/-/KS-RA-11-009.

European Commission. Sharing Cities - Site Lisbon. Smart Cities Marketplace. Retrieved December 29, 2024, from https://smart-cities-marketplace.ec.europa.eu/projects-and-sites/projects/sharing-cities/sharing-cities-site-lisbon.

F. R. On, R. Jailani, S. L. Hassan and N. M. Tahir, Analysis of sparse PCA using high dimensional data. 2016 IEEE 12th International Colloquium on Signal Processing & Its Applications (CSPA), 340-345. https://doi.org/10.1109/CSPA.2016.7515857.

Forina, M., Armanino, C., Lanteri, S., & Leardi, R. (2005). Methods of varimax rotation in factor analysis with applications in clinical and food chemistry. Journal of Chemometrics, 3. https://doi.org/10.1002/cem.1180030504.

Gao, J., Leng, Z., Qin, Y., , Z., & Liu, X. (2013). Short-term traffic flow forecasting model based on wavelet neural network. 2013 25th Chinese Control and Decision Conference (CCDC), 5081-5084. https://doi.org/10.1109/CCDC.2013.6561856.

Garau, C., Desogus, G., Barabino, B., & Coni, M. (2022). Accessibility and Public Transport Mobility for a Smart(er) Island: Evidence from Sardinia (Italy). Sustainable Cities and Society, 87, 104145. https://doi.org/10.1016/j.scs.2022.104145.

García, S., Luengo, J., & Herrera, F. (2015). Data Preprocessing in Data Mining (Vol. 72). Springer International Publishing. https://doi.org/10.1007/978-3-319-10247-4.

García-Peña, C., Molina, J., Cabrera, S., & Sinoga, J. (2023). Vulnerability to COVID-19: Cluster analysis of census tracts in Malaga, Spain. Journal of Urban Affairs. https://doi.org/10.1080/07352166.2023.2187301.

Ghanim, M. S., Muley, D., & Kharbeche, M. (2022). ANN-based traffic volume prediction models in response to COVID-19 imposed measures. Sustainable Cities and Society, 81, 103830. https://doi.org/10.1016/j.scs.2022.103830.

Gómez-Álvarez, D., López-Moreno, E., Bilsky, E., Blanco Ochoa, K., & Osorio Lara, E. (2015). Indicators for measuring urban sustainability and resilience. In How smart is your city?: Technological innovation, ethics and inclusiveness (pp. 163-179). Cambridge University Press. https://doi.org/10.1017/9781316647554.010.

Greene, F. J., Tracey, P., & Cowling, M. (2007). Recasting the city into city-regions: Place promotion, competitiveness benchmarking and the quest for urban supremacy. Growth and Change, 38(1), 1-22. https://doi.org/10.1111/j.1468-2257.2007.00350.x

Gruppa, H. and Mogee, M.E. (2004) 'Indicators for national science and technology policy: How robust are composite indicators?', Research Policy 33(9): 1373–1384.

Harvey, A. C., & Peters, S. (1990). "Estimation Procedures for Structural Time Series Models," Journal of Forecasting, 9, 89–108. https://doi.org/10.1002/for.3980090203.

HAN, Jiawei; KAMBER, Micheline; PEI, Jian. Data Mining: Concepts and Techniques. 3. ed. Burlington: Morgan Kaufmann, 2012. DOI: 10.1016/C2009-0-61819-5.

Hashemi, F., Hashemi, F., Naderi, M., Jamalizadeh, A., & Bekker, A. (2021). A flexible factor analysis based on the class of mean-mixture of normal distributions. Comput. Stat. Data Anal., 157, 107162. https://doi.org/10.1016/j.csda.2020.107162.

IPMA (2010). Climate Normals from 1981 to 2010. Retrieved from https://www.ipma.pt/en/oclima/normais.clima/1981-2010/.

IPMA. (2022). Boletins climatológicos mensais. https://www.ipma.pt/pt/publicacoes/boletins.jsp?cmbDep=cli&cmbTema=pcl&cmbAno=2022&idDep=cli&idTema=pcl&curAno=2022.

IPMA. (2021). Boletins climatológicos mensais. https://www.ipma.pt/pt/publicacoes/boletins.jsp?cmbDep=cli&cmbTema=pcl&cmbAno=2021&idDep=cli&idTema=pcl&curAno=2021.

Jandaghian, Z., & Akbari, H. (2018). The effect of increasing surface albedo on urban climate and air quality: A detailed study for Sacramento, Houston, and Chicago. Climate, 6(19), 19. https://doi.org/10.3390/cli6020019.

Jardim, B., Castro Neto, M., Alpalhão, N., & Calçada, P. (2023). High spatiotemporal resolution urban dynamics: A case study of Porto. Sustainable Cities and Society, 98, 104867. https://doi.org/10.1016/j.scs.2023.104867.

Jardim, B., de Castro Neto, M., & Calçada, P. (2022). The daily urban dynamic indicator: Gauging the urban dynamic in Porto during the COVID-19 pandemic. *Sustainable Cities and Society, 79*, Article 103714. https://doi.org/10.1016/j.scs.2022.103714.

Jenelius, E., & Cebecauer, M. (2020). Impacts of COVID-19 on public transport ridership in Sweden: Analysis of ticket validations, sales, and passenger counts. Transportation Research Interdisciplinary Perspectives, 8, 100242. https://doi.org/10.1016/j.trip.2020.100242.

Kitchin, R., Lauriault, T., & Mcardle, G. (2015). Knowing and governing cities through urban indicators, city benchmarking and real-time dashboards. Regional Studies, Regional Science, 2, 28 - 6. https://doi.org/10.1080/21681376.2014.983149.

Klopp, J. M., & Petretta, D. L. (2017). The urban sustainable development goal: Indicators, complexity and the politics of measuring cities. Cities, 63, 92-97. https://doi.org/10.1016/j.cities.2016.12.019.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., Rubel, F., 2006. World map of the Köppen-Geiger climate classification updated. Meteorol. Z. 15 (3). https://doi.org/10.1127/0941-2948/2006/0130.

Lee, P. J., & Jeong, J. H. (2021). Attitudes towards outdoor and neighbour noise during the COVID-19 lockdown: A case study in London. Sustainable Cities and Society, 67, 102768. https://doi.org/10.1016/j.scs.2021.102768.

Lestegás, I. (2019). Lisbon After the Crisis: From Credit-fuelled Suburbanization to Tourist-driven Gentrification. International Journal of Urban and Regional Research. https://doi.org/10.1111/1468-2427.12826.

Liao, W., Hou, D., & Jiang, W. (2019). An Approach for a Spatial Data Attribute Similarity Measure Based on Granular Computing Closeness. Applied Sciences. https://doi.org/10.3390/APP9132628.

Li, Y., & Parker, L. E. (2014). Nearest neighbor imputation using spatial-temporal correlations in wireless sensor networks. Information Fusion, 15, 64-79. https://doi.org/10.1016/j.inffus.2012.08.007.

Louro, A., Da Costa, N., & Da Costa, E. (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. https://doi.org/10.3390/ijerph18073525.

Lu, J., Li, B., & Li, H. (2020). Expansion of city scale, traffic modes, and air pollution: A review. Cities, 108, 102974. https://doi.org/10.1016/j.cities.2020.102974.

Malik, K. R., Sam, Y., Hussain, M., & Abuarqoub, A. (2018). A methodology for real-time data sustainability in smart city: Towards inferencing and analytics for big-data. *Sustainable Cities and Society, 39*, 548-556. https://doi.org/10.1016/J.SCS.2017.11.031.

Matheus, R., Janssen, M., & Maheshwari, D. (2020). Data science empowering the public: Data-driven dashboards for transparent and accountable decision-making in smart cities. Government Information Quarterly, 37(3), 101284. https://doi.org/10.1016/j.giq.2018.01.006

Myeong, S., Park, J., & Lee, M. (2022). Research Models and Methodologies on the Smart City: A Systematic Literature Review. Sustainability, 14(3), 1687. https://doi.org/10.3390/su14031687.

Mohanty, S. P., Choppali, U., & Kougianos, E. (2016). Everything you wanted to know about smart cities: The Internet of things is the backbone. IEEE Consumer Electronics Magazine, 5(3), 60–70. https://doi.org/10.1109/MCE.2016.2556879.

Mori, K., & Christodoulou, A. (2012). Review of sustainability indices and indicators: Towards a new City Sustainability Index (CSI). Environmental Impact Assessment Review, 32, 94-106. https://doi.org/10.1016/J.EIAR.2011.06.001.

Münnich, R., & Seger, J. (2014). Impact of survey quality on composite indicators. Sustainability Accounting, Management and Policy Journal, 5, 268-291. https://doi.org/10.1108/SAMPJ-10-2013-0045.

OECD. (2008). Handbook on Constructing Composite Indicators: Methodology and User Guide. OECD Publishing. https://doi.org/10.1787/9789264043466-en.

Peffers, K., Tuunanen, T., Rothenberger, M., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems, 24, 45 - 77. https://doi.org/10.2753/MIS0742-1222240302.

Pordata & INE, 2022. "População residente: Total. Onde há mais e e menos pessoas". https://www.pordata.pt/municipios/populacao+residente+total-359.

Rajput, P., Chaturvedi, M., & Patel, V. (2020). Opportunistic sensing-based detection of crowdedness in public transport buses. Pervasive and Mobile Computing, 68, 101246. https://doi.org/10.1016/j.pmcj.2020.101246.

Rathore, M. M., Paul, A., Hong, W.-H., Seo, H. C., & Awan, I. (2018). Exploiting IoT and big data analytics: Defining Smart Digital City using real-time urban data. Sustainable Cities and Society, 40, 600-610.

Rathore, M. M., Paul, A., Rho, S., Khan, M., & Shah, S. A. (2021). Smart traffic control: Identifying driving-violations using fog devices with vehicular cameras in smart cities. Sustainable Cities and Society, 71, 102986.

Rosales, N. (2010). Towards a design of sustainable cities: Incorporating sustainability indicators in urban planning. In Proceedings of the 46th ISOCARP Congress. Retrieved from https://www.isocarp.org.

Rumpler, R., Venkataraman, S., & Göransson, P. (2020). An observation of the impact of CoViD-19 recommendation measures monitored through urban noise levels in central Stockholm, Sweden. Sustainable Cities and Society, 63, 102469. https://doi.org/10.1016/j.scs.2020.102469.

Salman, M. Y., & Hasar, H. (2023). Review on environmental aspects in smart city concept: Water, waste, air pollution, and transportation smart applications using IoT techniques. Sustainable Cities and Society, 94, 104567. https://doi.org/10.1016/j.scs.2023.104567.

Sarmento, R., Zorza, F. M. B., Serafim, A. J., & Allmenroedr, L. B. (2000). Urban Environmental Quality Indicator. WIT Transactions on Ecology and the Environment, 39, 8.https://www.witpress.com/elibrary/wit-transactions-on-ecology-and-the-environment/39/4851.

Sarroeira, R., Henriques, J., Sousa, A., Da Silva, C., Nunes, N., Moro, S., & Botelho, M. (2023). Monitoring Sensors for Urban Air Quality: The Case of the Municipality of Lisbon. Sensors (Basel, Switzerland), 23. https://doi.org/10.3390/s23187702.

SDG-Goal-11-Monitoring-Framework. (2020). A guide to assist national and local governments to monitor and report on SDG goal 11 indicators.

Smékalová, L., & Kučera, F. (2020). Smart city projects in the small-sized municipalities: Contribution of the cohesion policy. Scientific Papers of the University of Pardubice, Series D: Faculty of Economics and Administration, 28(2), 1067. https://doi.org/10.46585/sp28021067.

Šoštarić, M., Vidović, K., Jakovljević, M., & Lale, O. (2021). Data-Driven Methodology for Sustainable Urban Mobility Assessment and Improvement. Sustainability. https://doi.org/10.3390/su13137162.

Sun, J., Yuan, J., Wang, Y., Si, H., & Shan, X. (2011). Exploring spacetime structure of human mobility in urban space. Fuel and Energy Abstracts. https://doi.org/10.1016/J.PHYSA.2010.10.033.

Suzuki, H., Cervero, R., & Iuchi, K. (2013). Transforming cities with transit: Transit and land-use integration for sustainable urban development. Washington, DC: The World Bank. https://doi.org/10.1596/978-0-8213-9745-9.

Taylor, S. J., & Letham, B. (2018). Forecasting at scale. The American Statistician, 72(1), 37–45. https://doi.org/10.1080/00031305.2017.1380080.

Tran, N., Burdejová, P., Ospienko, M., & Härdle, W. (2014). Principal Component Analysis in an Asymmetric Norm. Sustainability at Work eJournal. https://doi.org/10.2139/ssrn.2854822.

United Nations. (2019). World Population Prospects 2019: Highlights. New York: United Nations Department of Economic and Social Affairs, Population Division. https://population.un.org/wpp/Publications/Files/WPP2019_10KeyFindings.pdf.

United Nations, Department of Economic and Social Affairs, Population Division. (2018). World Urbanization Prospects: The 2018 Revision. https://population.un.org/wup/Publications/Files/WUP2018-Report.pdf.

Veckalne, R., & Tambovceva, T. (2023). Evaluating urban sustainability in Uzbekistan: A novel formula for empirical analysis. Sustainability, 15(7035). https://doi.org/10.3390/su15097035.

Wang, Q., & Li, S. (2021). Nonlinear impact of COVID-19 on pollutions – Evidence from Wuhan, New York, Milan, Madrid, Bandra, London, Tokyo and Mexico City. *Sustainable Cities and Society, 65*, Article 102629. https://doi.org/10.1016/J.SCS.2020.102629.

Wilbur, M., Ayman, A., Sivagnanam, A., Ouyang, A., Poon, V., Kabir, R., Vadali, A., Pugliese, P., Freudberg, D., Laszka, A., & Dubey, A. (2023). Impact of COVID-19 on public transit accessibility and ridership. Transportation Research Record, 2677(4), 531–546. https://doi.org/10.1177/03611981231160531.

Wimbadi, R. W., Djalante, R., & Mori, A. (2021). Urban experiments with public transport for low carbon mobility transitions in cities: A systematic literature review (1990–2020). Sustainable Cities and Society, 72, 103023. https://doi.org/10.1016/j.scs.2021.103023.

Yeo, M. J., & Kim, Y. P. (2022). Long-term trends and affecting factors in the concentrations of criteria air pollutants in South Korea. Journal of Environmental Management, 317, 115458. https://doi.org/10.1016/j.jenvman.2022.115458.

Yuan, J., Zheng, Y., Xie, X., & Sun, G. (2011). Driving with knowledge from the physical world. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 316–324. https://doi.org/10.1145/2020408.2020462.

Zhou, M., Yue, Y., Li, Q., & Wang, D. (2016). Portraying Temporal Dynamics of Urban Spatial Divisions with Mobile Phone Positioning Data: A Complex Network Approach. ISPRS Int. J. Geo Inf., 5, 240. https://doi.org/10.3390/ijgi5120240.

APPENDIX A – VARIABLES' TEST

To construct the UDI, multiple variables reflecting distinct facets of urban activity and environmental quality in the parishes were initially evaluated. This appendix outlines the tests and criteria utilized to ascertain the inclusion or deletion of variables in the final model.

Each variable was examined regarding its theoretical and practical significance for computing the UDI. Correlation and factor analyses were employed to determine the factors that most effectively represent urban dynamics. The study initially included the variables, also represented in the Methodology and Data chapter, of:

- Reports
- Traffic Flow (Delay)
- Noise Level (Noise)
- Air Quality (O_3)
- Atmospheric Temperature
- Car's Velocity (Speed)

The initial test concentrated on examining the connection between these variables and the UDI, in addition to assessing the appropriateness of the variables for component analysis. A strong association was noted between Delay and Speed, indicating that both metrics reflect comparable facets of urban dynamics. It was determined to eliminate one of these factors to prevent duplication and multicollinearity, with Speed being selected to drop. This selection was predicted on Speed's high correlation with Delay, with its minimal contribution to the variance elucidated in the descriptive analysis of the data, and also with the minimal impact of the variable in the UDI.

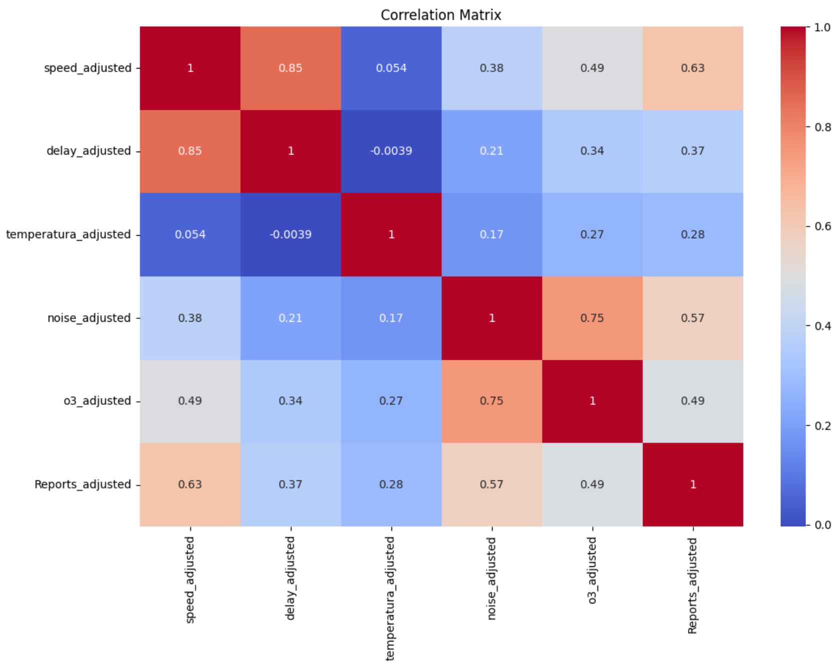

Figure 35 – Correlation between variables

Table 12 - Weight (%) of each variable in the UDI per day type and period

	Business Days				Weekend			
	Morning	Lunch	Afternoon	Nigh	Morning	Lunch	Afternoon	Nigh
Reports	50,33	50,08	49,85	50,63	46,90	46,86	46,82	46,96
Traffic Flow	18,09	18,55	18,95	17,56	16,00	16,70	17,29	15,20
Noise Level	13,06	12,65	12,28	13,54	15,07	14,43	13,89	15,81
Air Quality	14,66	15,09	15,48	14,16	17,49	17,79	18,04	17,14
Atmospheric Temperature	3,62	3,39	3,18	3,89	4,35	4,03	3,75	4,73
Car's Velocity	0,24	0,25	0,26	0,23	0,19	0,21	0,22	0,17

It was also determined to disregard the Temperature variable due to its minimal and insignificant impact on UDI, as can be seen in Table 12.

At the end, we end up with four variables to compute the UDI, the Reports, Traffic Flow (Delay), Noise Level (Noise), and Air Quality (O_3).

ANNEXES

Cartography of Lisbon Dataset

Table 13 – Cartography Dataset

Variable name	Description
ID	Parish Identifier
Name	Parish Name
Parishes	Common Name of Parishes
Area_M2	Area of the Parish
Geometry	Geometric multipolygon, with Parishes delimitation

Vodafone Dataset

Table 14 - Mobility Dataset

Variable Name	Description
Grid_ID	Identification of the Grid Number
DateTime	Date and Time
C1	Count of unique terminals in the grid
C2	Count of distinct terminals that are connected to the grid in roaming.
C3	Count of unique terminals that remained in the grid
C4	Count of unique terminals remained within the grid when roaming.
C5	Count of unique terminal entrances in the grid
C6	Count of outputs from multiple terminals in the grid
C7	Count of entries from distinct terminals when in roaming within the grid cell.
C8	Count of exit points from different terminals while in roaming within the grid cell.
C9	Count of distinct terminals currently having an operational data connection, contained within the grid cell.

C10	Count of distinct terminals in roaming, actively transmitting data.
C11	Number of voice calls originating from the grid
D1	List of the top 10 countries in roaming.
E1	Count of terminated voice calls inside the network grid
E2	Average downstream rhythm inside the grid.
E3	Average upstream rhythm in the grid
E4	Peak downstream rhythm in the grid
E5	Peak upstream rhythm in the grid
E6	Top 10 Apps (separated by ";")
E7	Duration of minimum stay within the grid
E8	Duration of the average stay in the grid
E9	Duration of maximum stay within the grid
E10	Count of devices that share the connection in the grid

Grid Location Dataset

Table 15 – Grid Location Dataset

Variable name	Description
ID	Grid identifier
Latitude	Grid Latitude
Longitude	Grid Latitude
Grid_Are	Area of the grid (40.000 m2)
Grelha_per	Perimeter of the grid (800 m)
Parish	Parish of the centroid of the square
Name	The region defined by the grid

Grid_X	An alternate coordinate system for the x-axis, where the values are integers starting from 0.
Grid_Y	An alternate coordinate system for the y-axis, where the values are integers starting from 0.
Geometry	Geometric multipolygon
WKT	Item presented in the Well-known text (WKT) format, which allows for easier manipulation in many software applications.

Waze Dataset

Table 16 - Waze Dataset

Variable Name	Description
DateTime	Date and Time
Street	Street name where the report was identified
Position	Coordinates of the streets
Delay	Delay of the jam
Speed	Average speed of the congestion

Meteorological Dataset

Table 17 - Meteorological Dataset

Variable Name	Description
Entity_ts	Time Stamp
precipitation	Value of Cumulative Precipitation
temperature	Value of Temperature

Position	Coordinates of the sensor that captured the value

Environmental Dataset

Table 18 - Environmental Dataset

Variable Name	Description
ID	Record ID
DateTime	Date and Time
O3	Value of ozone levels
Noise	Value of noise levels
Position	Coordinates of the sensor that captured the value

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

Universidade Nova de Lisboa