INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Analysis of the mobility of people and tourists in nightlife areas in the city of Lisbon

Bernardo Filipe Marques dos Santos

Master's degree in Integrated Business Intelligence Systems

Supervisor:
Doctor João Carlos Amaro Ferreira, Assistant Professor with Habilitation
ISCTE - University Institute of Lisbon

Co-Supervisor:
Doctor Ana Madureira, Coordinator Professor
ISEP - Porto School of Engineering

July 2023

TECNOLOGIAS
E ARQUITETURA

Department of Information Science and Technology

Analysis of the mobility of people and tourists in nightlife areas in the city of Lisbon

Bernardo Filipe Marques dos Santos

Master's degree in Integrated Business Intelligence Systems

Supervisor:
Doctor João Carlos Amaro Ferreira, Assistant Professor with Habilitation
ISCTE - University Institute of Lisbon

Co-Supervisor:
Doctor Ana Madureira, Coordinator Professor
ISEP - Porto School of Engineering

July 2023

Acknowledgements

This dissertation is the result of a lot of effort, hard work and dedication. My first dream was to enter university, where a master's degree was far away and out of the equation. Tired of studying, I really thought I would stick to a degree in management. But in January 2021, I started thinking: what will I do with a degree in management? Is this what I really like? And that's when I started thinking about what I liked. Passionate about technology and computers, I decided to look for a course that linked management to technology, but that was not too technical.

And that's when I found ISCTE master's degree in integrated business intelligent systems. It was definitely the right choice and a happy journey. First of all, I would like to thank my supervisor Doctor João Carlos Ferreira, for all the help and availability during this work. I also would like to thank Professor Luís Brito Elvas; my co-supervisor Doctor Ana Madureira; and Bruno Francisco for all the tips and help throughout this process.

Secondly, a special thanks to my girlfriend, Margarida, for all the patience and fellowship to overcome the hardest difficulties, for all the sleepless nights and for making me grow and learn to never give up. A special thank you to my parents in law and family, their vital support was essential for the success of this dissertation. Finally, a special thanks to all my friends and colleagues, especially to André Mesquita, André Santos and Rafaela Besugo, for all teamwork across all subjects.

Bernardo Filipe Marques dos Santos

Resumo

A mobilidade da população pode ser melhor compreendida através da análise de atividades específicas do telemóvel. Assim, esta dissertação tem como objetivo analisar a mobilidade de pessoas e turistas em zonas de diversão noturna, na cidade de Lisboa, utilizando dados de dispositivos móveis de cada utilizador, fornecidos por uma operadora. Estes dados são obtidos através de um acordo entre a operadora e a Câmara Municipal de Lisboa. O principal objetivo é fornecer à equipa Laboratório de Dados Urbanos de Lisboa (LxDataLab) informações baseadas em dados que possam ser utilizadas, para melhor gerir e afetar recursos relativamente às áreas de diversão noturna. Estes espaços têm um grande impacto na vida da cidade, e a sua gestão é muito importante para satisfazer os interesses dos vários intervenientes: comerciantes, residentes e utilizadores. O desenvolvimento desta investigação centrou-se em três etapas: 1) criar conhecimento relativamente à vida noturna de Lisboa e perceber se os dados respondem às questões de investigação; 2) compreender, extrair, limpar e transformar os dados, para desenvolver um conjunto de dados úteis, capazes de responder às nossas necessidades; 3) visualização dos dados, onde foi possível fazer uma análise completa dos dados, extraindo valor, conhecimento e as respostas necessárias para fornecer aos decisores.

Relativamente aos entregáveis, o trabalho criado inclui também ficheiros python onde foi feito o processamento dos dados, além deste documento. Para as visualizações e dashboards foi utilizado o Microsoft PowerBI.

As análises e conclusões retiradas foram validadas pela equipa do LxDataLab através de uma apresentação de resultados online.

Palavras-chave: Mobilidade Inteligente; Análise Comportamental; Vida Noturna Lisboa; Análise de Dados; Rede Móvel; Cidades Inteligentes; Análise de Clusters.

Abstract

The population's mobility can be better understood by looking into specific mobile phone activities. Thus, this dissertation aims to analyze the mobility of people and tourists in nightlife areas in the city of Lisbon using data from each user's mobile device provided by a mobile operator. This data is obtained through an agreement between the mobile operator and the Lisbon city council. The main purpose is to provide the Lisbon Urban Data Lab team (LxDataLab) with data-based information that they can use to better manage and allocate resources regarding nightlife areas with strong public space occupation. These spaces have a great impact on the life of the city, and their management is very important in order to fulfill the interests of the several intervenients such as merchants, residents and users. The development of this research has focused on three stages: 1) create knowledge regarding Lisbon's nightlife and understand if the data can answer the research questions; 2) Understand, extract, clean and transform data, to develop a set of clean and useful data capable of responding to our needs; 3) Data visualization where it was possible to do a complete analysis of the data, extracting the value, knowledge and answers needed to provide to decision makers.

Regarding deliverables, the work created also includes python files where data processing was done in addition to this document. For the visualizations and dashboards Microsoft PowerBI was used.

The analysis and conclusions drawn were validated by the LxDataLab team through an online presentation of results.

Keywords: Smart Mobility; Behavior Analysis; Lisbon Nightlife; Data Analytics; Cellular Network; Smart Cities; Clustering Analysis.

Index

Abbreviations

BTS	**B**ase **T**ransceiver **S**tation
CRISP-DM	**C**ross **I**ndustry **S**tandard **P**rocess for **D**ata **M**ining
dB	**D**ecibel
GDP	**G**ross **D**omestic **P**roduct
Gb	**G**igabytes
GPS	**G**lobal **P**ositioning **S**ystem
ICTs	**I**nformation and **C**ommunication **T**echnologies
IoT	**I**nternet **O**f **T**hings
LAeq	Equivalent Continuous Sound Level
LXDataLab	**L**isbon Urban **Data Lab**oratory
PRISMA	**P**referred **R**eporting **I**tems for **S**ystematic Reviews and **M**eta-**A**nalyzes
SIM	**S**ubscriber **I**dentity **M**odule
TEG	**T**ourism **E**conomic **G**rowth
USD	**U**nited **S**tates **D**olar
Wkt	**W**ell-**K**nown **T**ext

List of Figures

List of Tables

CHAPTER 1

Introduction

Big data can reveal the geographical pattern and intercity network of mobility more effectively than conventional data, such as travel diary surveys and population census [1]. GPS and mobile devices have made substantial use of location-based big data. The Internet of Things (IoT) is currently a common method for creating applications for tourism and mobility [2]. The population's mobility and economic activity can be better understood by looking into specific mobile phone activities [3].

Mobile devices are quite widespread among the general public. 96% of people in the population had active mobile subscriptions in 2014 [4]. Mobile penetration is presently at 121% in developed nations, while it has reached 85% in underdeveloped nations [5]. Mobile phones allow for the collection of additional data on the travels of the phone user over time, as opposed to fixed phones, which might be used to list receivers and analyze the frequency and length of calls [3].

People moving across time and space is what defines tourism [6]. Tourists travel away from home to experience new things, discover new locations, do business, or look for themselves [6]. Recent research demonstrates that visitors want to distinguish themselves from conventional tourists and their motivations [7]. Over 12.3 billion journeys were made by tourists worldwide in 2019, an increase of 4.6% from 2018. The entire revenue from tourism worldwide was $5.8 trillion USD, or 6.7% of the global GDP (World Tourism Economy Trends Report [8]). The primary driver of tourism economic growth (TEG) is mobility, and the transportation infrastructure is crucial for ensuring a steady stream of tourists [9].

Understanding the behavior of visitors and locals, as well as the functionality of urban regions, requires analysis of human movement patterns. Urban strategists can compare two cities and apply their expertise from one to the other by transferring information from one to the other. For example, they can build new tourist attractions or adopt certain effective urban strategies in brand-new cities (Barcelona) [10]. Depending on how long they stay, tourists' mobility in destination cities may fluctuate significantly [11]. Accessibility, safety and security and accommodation availability are also important variables to consider.

1.1. Overview

This work was developed based on a challenge proposal of the Lisbon Urban Data Laboratory (LXDataLab) team, a department of Lisbon City Council [12]. LxDataLab is an initiative of the Lisbon Management and Urban Intelligence Center, of the Lisbon City Council, which aims to use the data generated in the city of Lisbon to generate analytical solutions capable of solving real problems and improving the services provided to those who live, work, and visit it.

It consists of a partnership between the municipality of Lisbon and several entities, namely higher education, and scientific research institutions. The academy benefits from access to real data about the city and the Municipality of Lisbon has the opportunity to test analytics and data visualization solutions capable of promoting innovation, efficiency and proactivity in the services provided to citizens. With this in mind, and with several challenges available on the "Lisboa Aberta" website, challenge 74 was chosen: to study the mobility of people in nighttime areas, in the city of Lisbon.

Nightlife areas with strong occupation of public space have a great impact on the life of the city of Lisbon and their management is important in order to respond to the interests of the various stakeholders, namely merchants, residents, and users.

In this way, and using data from mobile devices, we intend to know the movement and permanence of people between nightlife areas. Given the economic interest of tourism and the impact of tourists on nightlife areas, this analysis will also include data on roaming mobile devices. It is also interesting to know the effects of these movements of people on the outdoor noise environment, namely, to know the evolution of the values of the equivalent continuous sound level (LAeq) parameter. Data from mobile devices is provided to Lisbon City Council by a mobile operator through an agreement between the parties. This study will be conducted through the creation and analysis of dashboards.

1.2. Motivation

With Lisbon making a big buzz around the world for its famous food, beaches and culture, today Lisbon is one of the most visited European cities. This beautiful city is located along the stunning coastline of Portugal and the city has a great and rich history. The ability to watch and comprehend mobility behavior at a previously unseen level of detail is made possible by the advent of new Big Data sources including call records from mobile phones, smart card data, and geo-coded social media records [13]. In the age of Big Data, everything is measurable, and it is now feasible to automatically follow the movement of people in Lisbon because to the enormous quantity of behavioral data that is now being produced by various sources, such as mobile phones [1].

Understanding people's mobility may be extremely useful for urban planners and politicians when developing infrastructure. Authorities can make educated judgements regarding the transport infrastructure, public transit systems, and road networks by studying how people move inside and around a metropolitan capital. This information contributes to the improvement of overall mobility efficiency, traffic congestion relief, and the optimization of transportation services. Moreover, the study of mobility patterns aids in achieving sustainable development objectives. Urban planners can find chances to encourage sustainable mobility alternatives like public transit, walking, and cycling by examining the modes of transportation people use and the distances they travel. Promoting environmentally friendly forms of transportation lowers greenhouse gas emissions, reduces air pollution, and enhances city life in general [14].

As a city resident, I am very interested in understanding people’s mobility patterns based on real data. With this in mind, I was given the opportunity to work on a real challenge in partnership with LxDataLab to develop knowledge on the mobility of residents and tourists in the city of Lisbon based on mobile phone data. In order to assist decision makers, this work will conduct a geospatial, longitudinal, and statistical analysis through the development of dashboards. By providing the stakeholders with the greatest information possible through the use of this information, it is possible to improve the nightlife areas experience, for merchants, residents and users. To better manage nightlife areas with high occupancy of public space it is necessary to determine not only the time period, but also the zones, in which the public space is not usual and becomes related to night-time entertainment activities.

1.3. Objectives

The goal of this dissertation is to analyze the mobility of people and tourists in nightlife areas, with possible transfers between areas and relating it to the noise levels from environmental sensors. In order to achieve this purpose we will use CRISP-DM approach [15] that will be further explained in the following section 1.4 of this dissertation.

In addition to other knowledge that may result throughout the analysis, we aim to determine the time period, and which squares in which the use of space ceases to be normal and becomes related to night-time entertainment activities in public spaces. Following this reasoning, for the squares and time periods identified, we want to characterize, among other things, the following variables throughout the night-time entertainment period:

- Permanence of residents and tourists;
- Mobility in night-time entertainment areas;
- Assessment of possible transfers of users between nightlife areas;
- Relate the movements of people with noise levels recorded in the environmental sensors located in the area of study considered.

1.4. Methodology

We used the CRISP-DM [15] (Cross Industry Standard Process for Data Mining) approach to conduct this study on analyzing the mobility of national users and tourists in nightlife areas. The goal of CRISP-DM is to give data workers a formal framework for organizing and carrying out data mining projects. We decided to use this technique since it has shown to be an effective way to examine and glean value from data of all kinds. To allow for adaptability by the working groups, CRISP-DM is not intended to be and does not claim to be a set of rigid rules. Thus, in order to better serve our goals and the fact that this is an academic work, we made certain modifications to the technique. As a result, and briefly, we decided to create our data exploration in four stages: 1) Business understanding; 2) Data understanding; 3) Data preparation; 4) Visualization and decision-supporting dashboards; which we feel are essential for us to be able to respond to our requests and supply information to stakeholders.

1.5. Dissertation challenges

This dissertation ran into five major challenges: data understanding; data cleaning; data volume/processing; data enrichment; and noise sensors location.

Considering the quantity of data on hand, it was necessary to carefully analyze each dimension to understand how it may be utilized to address the research questions and objectives established in sub-chapter 1.3.

Regarding data cleaning, decisions had to be made on how to clean and delete data that was not going to be used. The solution found was to define concrete objectives within the scope of the dissertation to facilitate the data selection process.

The volume of data was a big issue as the raw data was around 42GB and computing capacity was limited. The strategy found to solve this problem was to split the data by month, summing up to 8 chunks of data.

As far as data enrichment is concerned, the data provided did not have all the variables necessary to achieve the proposed objectives. Thus, variables were created such as: whether it is a holiday or not; separating the week from the weekend; areas of the city; categorizing the period of the day based on the hours; nationalities by region of the world. These variables were important in that it allowed not only to better understand the data, but also to draw more conclusions and results.

As for noise sensors, there are only 14 sensors throughout the city, and these are not distributed in the most relevant areas for the proposed objectives. The solution found was to study the readings of environmental noise sensors within a specific famous event, the night of Santo António.

1.6. Dissertation outline

With the objectives and methodology defined, the structure of this dissertation consists of five chapters, including the Introduction (Chapter 1). The remaining structure is composed by:

Chapter 2 is the chapter of the related work, which is composed of 5 subchapters, regarding to the literature review: method; data extraction; searching results; goals and outcomes analysis; and state of art.

Chapter 3 corresponds to the extraction of knowledge from the data, where the business understanding was done first. Then, the description and understanding of the data was done with its subsequent preparation. The data preparation was subdivided by its selection, cleaning and feature engineering.

Chapter 4 concerns to data visualization. This chapter has the in-depth and detailed analysis and answers the research questions as well as the defined objectives. Here the knowledge was created through graphs and dashboards and their compliance validation.

Chapter 5 is the last and consists of the final conclusions drawn as well as proposals for further developments and future work.

CHAPTER 2

Related work

2.1. Method

This study was developed on the basis of the PRISMA methodology [16], which stands for Preferred Reporting Items for Systematic Reviews and Meta-Analyzes. PRISMA intends to assist authors in enhancing and better evaluate the systematic review and meta-analysis reporting process by allowing readers to comprehend the authors' research and conclusions. Reporting standards are improved, and peer reviews are made easier.

2.2. Data extraction

This research was developed using SCOPUS database, in January 2023. Searching keywords were divided into 3 categories: Concept, Population and Context. Table 2.1 provides an illustration of how the search query was created by intersecting the three columns, i.e, Concept AND Population AND Context. A time limitation from 2011 to 2022 was also implemented and only journal papers, reviews and articles were considered. Figure 2.1 illustrates the evolution of the volume of documents over the study period and the study's focus. As demonstrated, interest in the subject of the study has been rising.

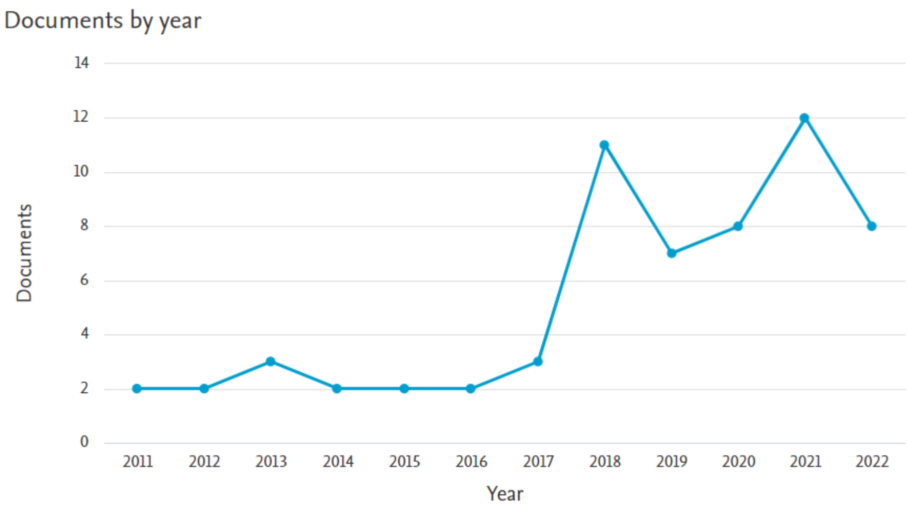


Figure 2.1 - Evolution of the subject under study over the years

Table 2.1 - Keywords selection

<table>
<tr><th>Concept</th><th>Population</th><th>Context</th><th>Limitations</th></tr>
<tr><td>data analysis
behavio#r analysis</td><td>touris*</td><td>mobility
smart#cit*</td><td rowspan="3">Years: 2011 - 2022

Only journal papers, articles, and reviews</td></tr>
<tr><td>480 631 documents</td><td>30 208 documents</td><td rowspan="2">667 025 documents</td></tr>
<tr><td colspan="2">2 322 documents</td></tr>
<tr><td colspan="4">61 documents</td></tr>
</table>

2.3. Searching results

In our search we first obtained a total of 61 documents from SCOPUS. In order to build searches with unidentified characters, multiple spellings, or multiple endings, we used wildcards [17] and truncation symbols, as shown in Table 2.1. Secondly, the documents were subjected to a filtering process, where after a preliminary analysis we noticed that 6 articles were not significant for the study. All processes are represented in Figure 2.2.

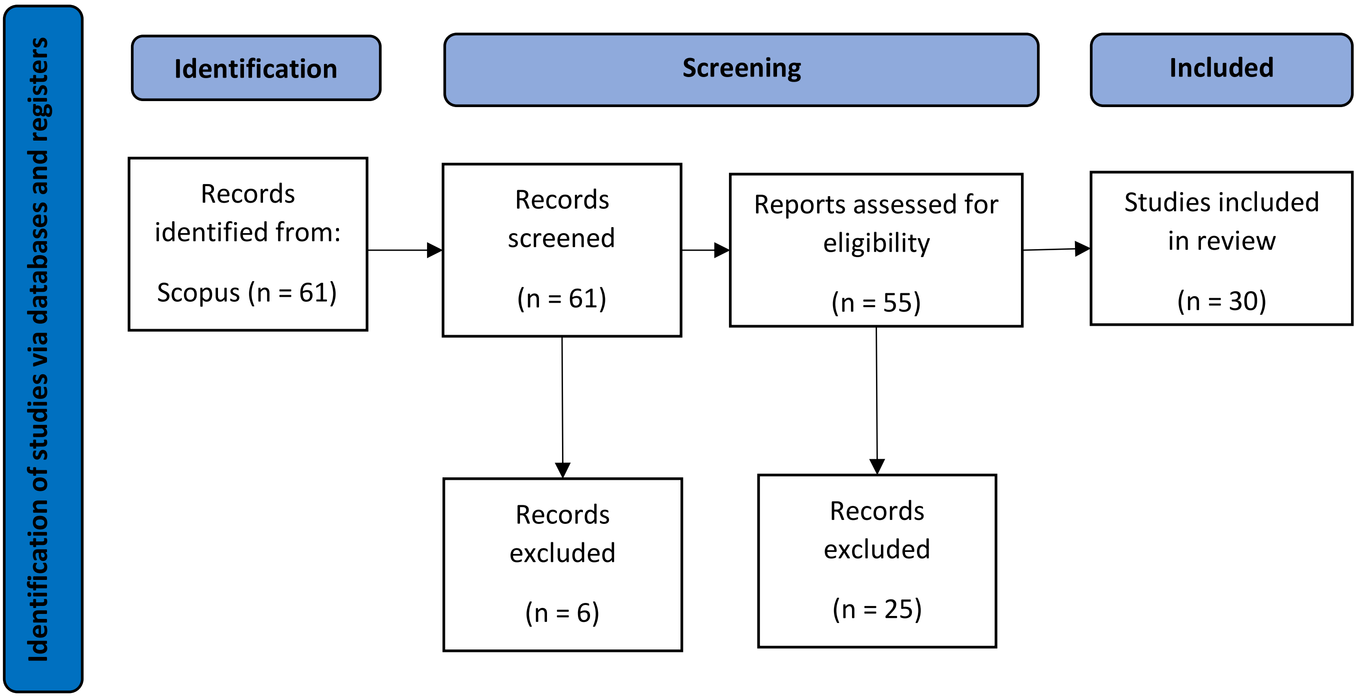


Figure 2.2 - PRISMA workflow diagram

After this brief analysis, of the 61 documents, 55 remained to be analyzed. To perform the analysis of the remaining 55 documents, title and abstract were analyzed and each paper was categorized according to the defined keywords. The set of documents were then analyzed according to their purpose. The main purpose expressed in the 36 out of 55 documents is "Mobility". Other important and used purposes are "Smart Cities", "Cellular Network/Mobile Phones" and "Touris*". Differently intended documents were discarded. Finally, after this selection based on purpose, the final set of documents included 30 references.

When analyzing the topics present in the 30 final documents, it can be observed in Figure 2.3 the most present topics which are, respectively, "Mobility", "Tourism/Tourist" and "Behavior Analysis". The topic "Mobility" appears in 26 documents, which represents 87% of the document set. "Tourism/Tourist" and "Behavior Analysis" topics are present in 23 and 22 documents, respectively, 77% and 73% of the total.

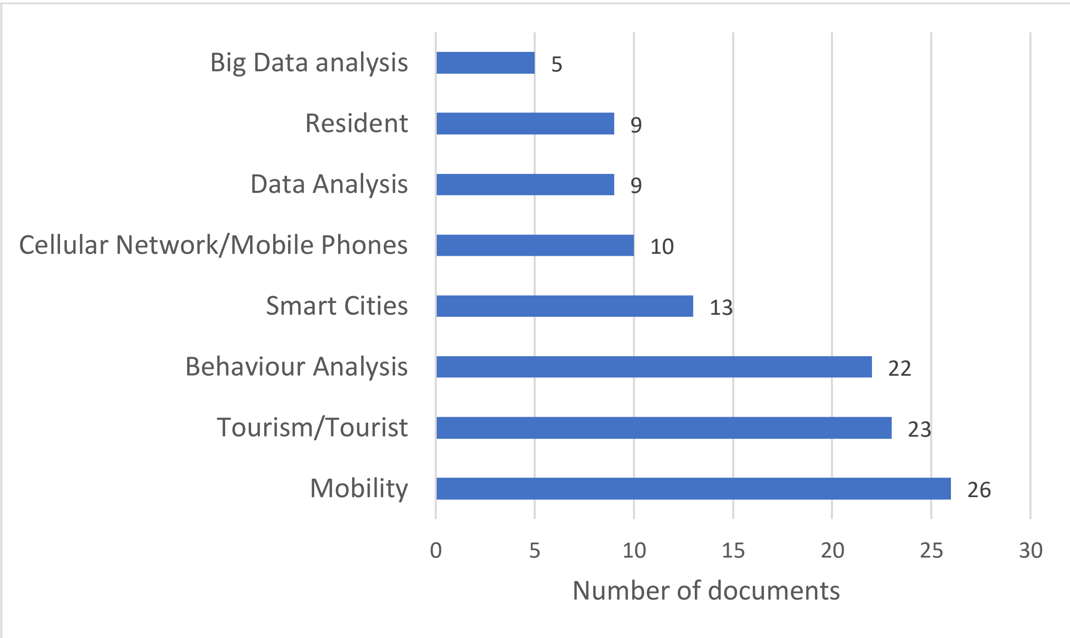


Figure 2.3 - Number of documents by topics

Figure 2.4 was produced using VOSviewer to show the correlations more clearly between the papers' keywords and their importance. Mobility and Tourist were the two main keywords we could find.

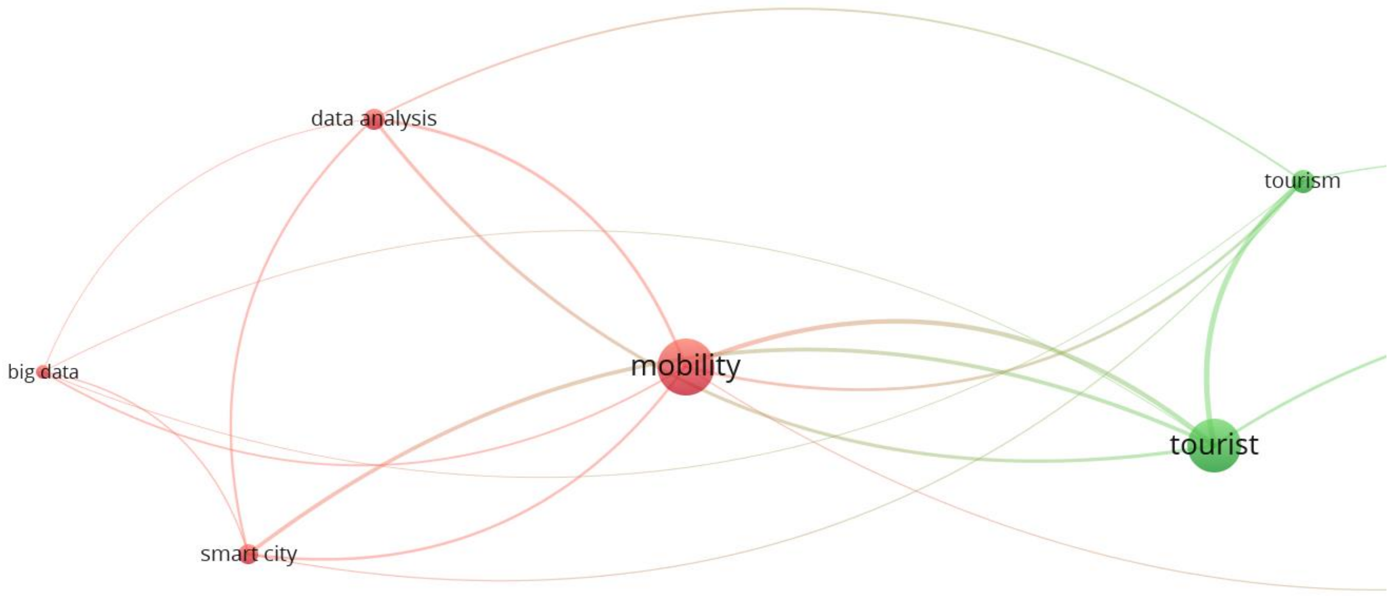


Figure 2.4 - Keywords network VOSviewer

2.4. Goals and outcomes analysis

With a focus on Mobility, Tourism/Tourist, and Behavior Analysis, Figure 2.5 offers theoretical explanations for each of the themes discussed in each of the publications that were reviewed. These three subjects form the basis of our research since the aim of this study is to analyze the mobility of national users and tourists in Lisbon's nightlife areas.

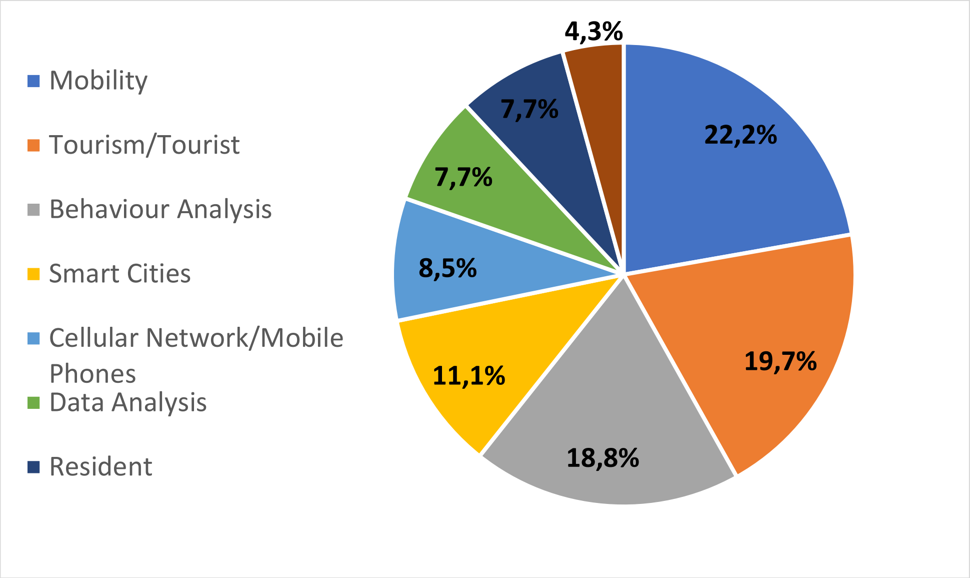


Figure 2.5 - Relative weight by document subject

The analysis of the chosen articles, where subjects were developed based on the identified keywords, is summarized in Table 2.2 It seeks to clarify how absolute and relative frequency relate to the subjects being studied.

Table 2.2 - Summary of references by topics aborded

Topic	References	# Doc	% Doc
Mobility	[1], [4], [10], [6], [9], [10], [18]–[24], [24]–[36]	26	87%
Tourism/Tourists	[2], [5], [6], [9], [10], [13], [18], [20], [22], [23], [27]–[39]	23	77%
Behavior Analysis	[3], [6], [9], [10], [18]–[22], [24]–[35], [37]	22	73%
Smart Cities	[2], [5], [10], [18], [24], [27], [30], [31], [33], [36], [39]–[41]	13	43%
Cellular Network/Mobile Phones	[3], [5], [18], [19], [25], [26], [31], [34], [38], [42]	10	33%
Data Analysis	[2], [5], [23], [26], [36], [38]–[40], [42]	9	30%
Residents	[10], [18], [22]–[24], [28], [29], [33], [38]	9	30%
Big Data analysis	[5], [20], [24], [39], [42]	5	17%

2.5. State of art

One of the most significant ecological and social issues of the twenty-first century is human mobility [25]. Mobility has emerged as a key geographic concept, and the switch from traditional data (census and survey data) to location-based big data opens up new possibilities for movement research since the latter can more accurately reflect intercity mobility patterns in real-time than the former [43].

The ability to watch and comprehend mobility behavior at a previously unseen level of detail is made possible by the advent of new Big Data sources including call records from mobile phones, smart card data, and geo-coded social media records [1]. Big data can track daily population movements throughout the year, thus while doing research, academics usually choose a typical time period or utilize the annual aggregate and average number [33]. Big data can reveal the geographical pattern and intercity network of population movements more effectively than conventional data due to its real-time dynamic monitoring of people mobility across multiple spatial scales and different periods with vast data volume and high spatial-temporal resolution [44]. With the quick advancement of big data technologies, population mobility studies have made substantial use of location-based big data obtained from GPS and mobile devices [45].

Cities are networks of human interactions that space and systems weave together. To create better cities, we must first understand how people, space and systems interact—the fundamental nodes, networks, and linkages that create the fabric of cities [40]. New kinds of interactions using wireless sensor networks are made possible by the Internet of Things (IoT). It is currently a common method for creating applications for tourism and mobility [2]. The idea of a "smart city" is to use the Internet of Things (IoT) and linked devices to gather information that can be used to better manage the city's assets, services, and resources. As the importance of enabling Smart Cities has grown, numerous municipal councils have started looking at the use of IoT sensors and other technologies to better understand how traffic moves around the city [24]. The domains of smart cities, environmental monitoring, and information security all greatly benefit from the application value and practical relevance of data mining based on trajectories [19].

The volume of trajectory data produced by GPS-equipped devices is continually growing as a result of the quick advancement of location-based technologies. This enormous amount of trajectory data is frequently employed in various fields to follow moving objects [46]. A useful method for population monitoring may be the spatiotemporal distribution of mobile devices geolocated to base transmitting towers (BTS). The prospect of being able to map transferring human population distributions and movement over relatively short timescales with data being collected by mobile network providers (while maintaining the anonymity of individual mobile users) opens the door for new applications and a near real-time understanding of patterns and processes in human geography [47].

As a result of the ever-increasing volume of data produced by smartphones, telecom corporations have access to vast amounts of data, which makes them sitting on a gold mine. Data that have been unlocked may be used to create competitive advantages and new income streams. Customers of telecom companies use their devices and networks to supply everyday information. Because operators utilize users' intelligent phones, they have a deeper understanding of their clients than any other sector of business [42].

Mobile devices are quite widespread among the general public. 96% of people in the population had active mobile subscriptions in 2014 [4]. Mobile penetration is presently at 121% in developed nations, while it has reached 85% in underdeveloped nations and has surpassed the total population in developed nations [5]. Mobile phone networks regularly and frequently determine the location of mobile phone devices, even while they are just in standby, in order to offer service to the users. The closest cell antenna to the user's handset is used to determine the user's location. This provides a precision corresponding to the region covered by the cell antenna, which in metropolitan areas can be as large as a few hundred meters [1]. Mobile phones allow for the collection of additional data on the travels of the phone user over time, as opposed to fixed phones, which might be used to list receivers and analyze the frequency and length of calls. The population's mobility and economic activity can be better understood by looking into specific mobile phone activities [3].

Understanding the behavior of visitors and locals, as well as the functionality of urban regions, requires analysis of human movement patterns. Urban strategists can compare two cities and apply their expertise from one to the other by transferring information from one to the other. For example, they can build new tourist attractions or adopt certain effective urban strategies in brand-new cities [10]. The spatial behavior of visitors at tourism locations is not a random occurrence. Tourists typically want to see a lot of sights in a short amount of time [20]. People moving across time and space is what defines tourism. Tourists travel away from home to experience new things, discover new locations, do business, or look for themselves [6].

Recent research demonstrates that visitors want to distinguish themselves from conventional tourists and their motivations [7]. Depending on how long they stay, tourists' mobility in destination cities may fluctuate significantly. When travelling for one day or less, visitors tend to move farther between attractions; however, when travelling for three days or longer, tourists tend to move closer to the attractions and over small distances [11].

Due to a large level of heterogeneity in visitors' motivations, historical-cultural amenities, financial spending capacity, age and gender characteristics, travel preferences, socio-cultural interests, etc., visitors' erratic movement patterns in a given region are frequently quite complex [48]. The people that visit tourist destinations are diverse in terms of their country of origin, age, gender, education, income, duration of stay, and reasons for visiting. For instance, [49] demonstrate that factors influencing destination choice in Hangzhou, China include the length of the trip, the type of companion (alone, with family, friends, or group), and car ownership.

The tourist business has continued to grow quickly in recent years. Over 12.3 billion journeys were made by tourists worldwide in 2019, an increase of 4.6% from 2018. The entire revenue from tourism worldwide was $5.8 trillion USD, or 6.7% of the global GDP (World Tourism Economy Trends Report [8]). By boosting jobs, enhancing infrastructure, and generating foreign cash for destinations, tourism has significantly impacted economic growth [9]. With the decrease and elimination of barriers to international human mobility as well as significant advancements and cost savings in the transportation sector, the tourism sector has emerged as a significant source of domestic income for a large number of countries worldwide. As a result, several nations have created and put into effect a variety of strategies to entice travelers and so increase their tourism revenues [22].

The primary driver of tourism economic growth (TEG) is mobility, and the transportation infrastructure is crucial for ensuring a steady stream of tourists [9]. The public transportation networks of several cities and metropolitan regions have produced rising amounts of data over the past three decades that might be very useful in the examination of a variety of mobility challenges [23]. For locals, city employees, and visitors, smart cities must optimize their transportation alternatives. The ability to walk is frequently overlooked, but as most short excursions inside a city are made on foot, improving the efficiency of this traffic can improve a city's liveability [50]. In fact, for both environmental and health reasons, car-free zones in compact center cities are increasingly encouraging walking as the primary mode of transportation [24].

On the other hand, locals frequently take vacations within their own nation. The leisure-oriented lifestyle between urban and rural homes is crucial. As an example, in Finland one of the top countries in the world for second house ownership [21].

Continuing the study of the work done, using the data generated by mobile phones, it was possible to find similar studies that meet the purpose of ours. Four studies serve as examples: “Unraveling pedestrian mobility on a road network using ICTs data during great tourist events” [18]; “Big data analytics – Geolocation from the perspective of mobile network operator” [5]; “Understanding tourist behavior using large-scale mobile sensing approach: A case study of mobile phone users in Japan [25]”; “Exploratory analysis of urban mobility: from mobile phone usage data to tourist behavior” [31].

Using the City of Venice as an example, the author [18] looked at the characteristics of pedestrian mobility on the road network in order to study the effects of tourists on the lives of local citizens as well as the preservation of the city's cultural heritage. Also used the data provided by the Italian Mobile Operator TIM. After analyzing and processing the data, they created an algorithm that could recreate pedestrian movements across Venice's streets. From there, they were able to discriminate between the mobility patterns of tourists and locals. Additionally, it made it possible to provide stakeholders with crucial and pertinent information for making decisions based on statistics rather than practical knowledge [51].

Now using the country of Czech Republic as an instance, the author [5] studied 4 different case studies: 1) Pilot Case Study of Sumava National Park; 2) Czech Mountain Ski Resorts; 3) Use of Mobility Data for the Preparation of City Territorial and Development Plan; 4) Václav Havel Airport. The largest mobile network provider in the Czech Republic is T-Mobile, which regularly communicates with six million terminals and generates hundreds of millions of signal recordings daily. Each case study's data came from the operator. The case studies were a success and allowed the author to deliver relevant information based on data rather than empirical knowledge. For example, the study allowed them to conclude where the tourists came from, how long they stayed, where they stayed and what places they visited.

In Japan, 130,861 mobile phone users' data was consulted by the author [25]. The information was given by a Japanese operator, and it had already been made anonymous. Through the use of an algorithm to find GPS traces on travelers, they looked at how tourists moved around and stayed in the city. They came at their conclusions after looking at data on the frequency of travels, the length of stays at each location, the types of transportation employed, and the connection between travel behavior and personal mobility.

The author [31] used information from the operator Orange to research the movement of visitors in the city of Paris. They overcame the issue of individual data by using aggregation and network visualization techniques to restore the fluxes between metropolitan locations. They concluded which are the most touristic and visited places and the differences in mobility in different time periods.

CHAPTER 3

Knowledge extraction approach

3.1. Business understanding

In the first phase of CRISP-DM [15] the most important thing is to deeply understand the business needs and make the requirements gathering [52]. There are four steps that make up this phase, although the third stage is not mandatory: 1) Define business goals; 2) Evaluate the situation; 3) Define data mining objectives; 4) Create a data mining project plan.

For the first step, it was necessary to understand what the stakeholders' objectives were and they can be summarized as follows:

- Analyzing the mobility of national users and tourists in nightlife areas.
- Period of stay of national users and tourists.
- Assessment of possible areas of user transfers between nightlife areas.
- To relate the movements of people with the noise levels registered in the environmental sensors located in the study area.

Regarding the evaluation of the situation, it is necessary to understand what resources are available. We were given 3 datasets to work on: 1) Number of mobile phones entering, staying, and leaving per 200m/200m grid cell in a 5-minute period in Lisbon; 2) Mapping of Vodafone's grids; 3) Location of establishments with noise limiters.

For the last step, it was necessary to define which technologies and tools to use. The technologies used were Python, Microsoft Excel, and Microsoft Power BI.

3.2. Data description and understanding

Mobile Operators continuously gather information about how customers and users utilize the service, whether for technical, monetary, or even legal reasons [3]. In this regard, the network must constantly gather different data and metrics regardless of whether the client is using 2G, 3G, 4G, or 5G in order to, for example, allow the continuation of a phone conversation or a mobile data session. There is always a requirement for a base station that is at a given position and that serves a certain geographical region to be servicing the customer, even though an operator may have many networks depending on the technology employed [1]. This base station communicates with the mobile device while also communicating with the network's central hub. It gathers information on mobility-related events as well as voice calls, internet sessions, and written messages through this interface with the hub. It should be highlighted that when we talk about "data from a Mobile Operator," we don't just mean the mobile data service; rather, we're talking about particular details about all the communication that is sent between the user's mobile device and the network that is supporting it. The analyzes and work produced for this study were based on the signaling exchanged in the city of Lisbon between a Mobile Operator's network and each roaming and non-roaming users during the course of eight months, from January 2022 to August 2022.

A metadata file was utilized in addition to the datasets the Mobile Operator provided in order to enhance them and make it feasible to correlate each event with a specific location. The information given by the Lisbon City Council is created using data from each user's mobile device and cellular network as part of an arrangement with a Mobile Operator. For legal and privacy considerations, the data in the dataset has been appropriately anonymized. This makes it impossible to perform any type of precise analysis on a certain user. Analysis can only be done using volumes and there isn't even a key that links a specific person to an event.

There are three datasets available. The one provided by the mobile phone operator is compiled in 3743 grids of 200x200 meters, being gathered in 5-minute intervals. Due to privacy restrictions, if a particular grid has less than 10 users over a 5-minute period, it won't be published. After being collected, data is made available on the big data platform for around 45 minutes. This means that the time between data collection and availability can only be delayed by a maximum of one hour. However, it is crucial to note that in order to conduct the current study, we will just use a snapshot of the data and won't be utilizing the online data stream.

In Table 3.1 we can consult the 24 dimensions of the dataset, with the respective description, provided by the Mobile Operator. However, for the scope of our study, we will only use 3 of the columns, marked in light blue. The term used by the Mobile Operator terminals corresponds to a SIM card. Therefore, in this research variable C3 will be described as national users and variable C4 will be described as tourists because they are terminals in roaming.

Table 3.1 - Mobile operator variables list dataset

ID	Variable Name	Variable Description	Variable Type
1	Grid_ID	Grid number. Lisbon's metropolitan region is divided into 3743 squares of 200 by 200 meters.	Nominal
2	Datetime	Time and date of occurrence.	Datetime
3	C1	No. of distinct terminals counted on each grid cell during the 5-minute period - Measured every 5 min.	Metric
4	C2	No. of distinct terminals in roaming counted on each grid cell during the 5-minute period– Measured every 5 min.	Metric
5	C3	No. of distinct terminals that remained in the grid cell counted at the end of each 5-minute period	Metric
6	C4	No. of distinct terminals in roaming that remained in the grid cell counted at the end of each 5-minute period	Metric
7	C5	No. of distinct terminals entering the grid	Metric
8	C6	No. of distinct terminals leaving the grid. The calculation is made using the previous 5-minute interval as reference, also considering the crossings of the grid in the same interval.	Metric
9	C7	No. of entries of distinct terminals, in roaming, in the grid.	Metric
10	C8	No. of exits of distinct terminals, in roaming, in the grid.	Metric
11	C9	No. of distinct terminals with active data connection in the grid cell – Measured every 5 min.	Metric
12	C10	No. of distinct terminals, in roaming, with active data connection in the grid cell – Measured every 5 min.	Metric
13	C11	No. of voices calls originating from the grid cell.	Metric
14	C12	No. of entries into Lisbon along the 11 main roads.	Metric
15	C13	No. of exits from Lisbon along the 11 main roads.	Metric
16	D1	Top 10 origin Countries of the devices in roaming.	Metric
17	E1	No. of voice calls that ended in the grid within the 5-minutes.	Metric
18	E2	Average download speed per grid within the 5-minutes.	Metric
19	E3	Average load speed per grid within the 5-minutes.	Metric
20	E4	Peak download speed on the grid within the 5-minutes.	Metric
21	E5	Peak upload speed on the grid within the 5-minutes.	Metric
22	E6	Top 10 apps used on the grid within the 5-minutes.	Metric
23	E7	Lowest permanence period on the grid within the 5-minutes.	Metric
24	E8	Average permanence on the grid within the 5-minutes.	Metric
25	E9	Maximum permanence period on the grid within the 5-minutes.	Metric
26	E10	Count of devices sharing the internet connection in the grid within the 5-minutes	Metric

A dataset that provides information about each of the 3743 grids was also utilized in addition to the dataset that comprises the data given by the Lisbon City Council through the agreement reached with Mobile Operator. Since they included the coordinates of each grid's centroid, the parish, or parishes in which the grid is put, the name, the geometry, and the WKT, these data enable us to geo-reference the main dataset. With this knowledge, the events may be added to the space using the "Grid_ID" key, which then launches our analysis. Table 3.2 presents the 11 geographical variables, with their respective description.

Table 3.2 - Geographical variables dataset

ID	Variable Name	Variable Description	Variable Type
1	grelha_id	Grid number. Lisbon's metropolitan region is divided into 3743 squares of 200 by 200 meters.	Nominal
2	dicofre	Identification of the parish. Assigned by administrative entities.	Nominal
3	entity_id	Identification of the data source where the information was generated.	Nominal
4	entity_type	Identification of the data source where the information was generated.	Nominal
5	freguesia	Parish to which the largest area where the grid belongs.	Nominal
6	freguesias	Parishes in which the grid is inserted.	Nominal
9	latitude	Centroid Latitude.	Metric
10	longitude	Centroid Longitude.	Metric
12	objectid	Id of the object in the database.	Nominal
13	position	Grid in geometry format.	Metric
14	wkt	Grid in WKT format.	Metric

The last dataset provided by Lisbon City Council contains information on the readings of 11 environmental noise sensors and their location, in the city of Lisbon, measured on an hourly basis. There are readings from May 2022 to September 2022 and the readings are taken hourly. However, for the scope of our study, the period defined was from January 2022 to August 2022. Thus, for this dataset, only the readings from May 2022 to August 2022 will be considered. The dataset's 7 variables, together with their descriptions and variable types, are listed in Table 3.3.

Table 3.3 - Environmental noise sensors dataset

ID	Variable Name	Variable Description	Variable Type
1	SMO_DTM	Date and time of the sensor reading.	Datetime
2	SMO_VAL	Sensor reading value.	Continuous
3	SMO_UN	Unit of measurement (always in deciles as unit of measurement).	Nominal
4	SMO_NV_LABEL	User-friendly indicator of the intensity of the level measured (Moderate, Normal, High, and Very High).	Ordinal
5	SLO_LOCAL	Textual location of the sensor.	Nominal
6	SLO_LGTX	Sensor Longitude	Metric
7	SLO_LATY	Sensor Latitude	Metric

After collecting all the data necessary for our analysis, it was necessary to examine and investigate the datasets to understand which data would bring value for our purpose. In Table 3.4 we can see the total number of records per month for the Mobile Phone Operator dataset. For the scope of our study, 3 columns will be used since these are the ones that will allow us to achieve our purpose, as previously mentioned, to analyze the mobility of national users and tourists in nightlife areas, in the city of Lisbon. These columns were selected with the help of the LxDataLab team, who are the responsible stakeholders, to align the purpose of the study with their objectives.

Table 3.4 - No. of records per month mobile operator dataset

Month	Number of Records
January 2022	33 353 532
February 2022	6 643 825
March 2022	14 346 919
April 2022	31 972 706
May 2022	32 365 721
June 2022	29 105 489
July 2022	33 416 473
August 2022	33 417 504

3.3. Data preparation

3.3.1. Data selection

The first step was to analyze the Mobile Operator dataset and understand which columns we would use to meet our objective. Thus, 3 columns were selected: C3: No. of distinct terminals that remained in the grid cell counted at the end of each 5-minute period; C4: Same but with roaming; and D1: Top 10 origin countries of the devices in roaming, all in the 3743 grids. Since we intend to analyze not only the mobility of national users but also tourists, it was necessary to select the data with roaming. Roaming refers to when a mobile phone is registered in a network other than the home network. So, we may presume that "roamers" are tourists or other visitors.

For the dataset provided by the LxDataLab team, we also did not use all the columns of the dataset. The parish id (dicofre), entity id of the data source (entity_id), entity type of the data source (entity_type) and object id (objectid) were left aside since they wouldn't be useful for our study. There are also 2 identical columns, parish (freguesia), and parishes (freguesias) where the second one is a string that joins more than one parish. The parish variable is sufficient to proceed with the study. In this dataset, we were provided with two columns where the 3743 grids are in geographic format (position) and another one in wkt format (wkt). Since we are going to use latitude and longitude, these coordinates already correspond to the centroid of each grid.

Regarding the ambient noise sensors dataset, out of the seven columns, we will only exclude the SMO_UN column because it only contains one value, which is the decibel, the unit of measurement.

3.3.2. Data cleaning

Regarding data cleaning, the data was analyzed for null values and duplicate records. Of the 3 datasets provided and the data selected, only variable D1, "Top 10 origin countries of the devices in roaming" from the Mobile Operator dataset had null values, with around 19 million of 33 million records. After questioning the decision-makers, we concluded that these values would correspond to national users, i.e., Portuguese people. There were no duplicate records in any of the databases, therefore cleaning wasn't necessary.

3.3.3. Feature engineering

The process of choosing, modifying, and creating features from raw data for usage is known as feature engineering. A solid understanding of the business problem and the data sources at hand is the foundation for effective feature engineering. Thus, some variables were created in order to better understand and facilitate the visualization and analysis of the data. Since the data was already given with the right data types, it was not required to change any.

We created a variable based on the top 10 origin of the tourists, where only their main origin was selected. To do this, it was necessary to separate these records, as many were strings with several countries. Next, we also created a continent variable, which grouped the tourists' origins by continents: Europe, Africa, Asia, Oceania, North America, Central America, and South America. The purpose of creating this variable is to drill up on the origins, for a more global view.

Because the primary emphasis of our research is on people's mobility, it was essential to separate holidays from other days of the week in our dataset. We simply marked the holidays that fell on the weekdays using a holidays library.

Similarly, to what was done with the origins of the tourists, it was also necessary to group the parishes of Lisbon by zones: north, south, east, west, and historical zone, to have a more global vision of the city of Lisbon.

CHAPTER 4

Data visualization

In this chapter, we show the results of the analysis carried out as well as the answers to give solution to the proposed objectives. Thus, subchapter 4.1.1 describes which tool was chosen to develop the visualizations. Subchapter 4.1.2 gives us an overview of the grids of the city of Lisbon. Regarding subchapter 4.1.3 we have an overview of the nationality of tourists. Next, we examined the monthly and average evolution of national users and tourists, subchapter 4.1.4. These first subchapters were an overview of the evolution of the data, where from here, we started to specific target the topics to answer the research questions.

Following this, on subchapter 4.1.5, an analysis of hourly evolution of distinct national users and tourists was done in order to understand what the busiest periods of the day are. With the same reasoning, in chapter 4.1.6 we analyze which areas of the city are the busiest and at what time of day. Having discovered that the central and historical areas are the busiest, both for residents and tourists, in subchapter 4.1.7 the monthly evolution of the mobility of people for these zones was observed.

Since the main goal of this dissertation is to analyze the mobility of people and tourists in nightlife areas, with possible transfers between areas and relating it to the noise levels from environmental sensors. From here, we have carried out a deeper analysis, as we studied the mobility of national users in night areas using heatmaps, subchapter 4.1.8. Following subchapter 4.1.9, we did the same for tourists. The next subchapter 4.1.10 is related to the study of environmental noise sensors, where an overview of this dataset was developed.

The last subchapter, 4.1.11 corresponds to the analysis of the most emblematic festive night in Lisbon, the night of Santo António. Here, not only was a cross-checking of the data with the recorded values of the noise sensors, but also a time-lapse of heatmaps was developed, to understand the patterns and transfers of the mobility of people between the most famous neighborhoods of Lisbon. This study was developed hour by hour, for the defined period from 05:00 in the afternoon until 07:00 in the morning. This study proved to be the most interesting and important for the stakeholders since conclusions were drawn that they were unaware of, and it is a night that requires great strategic planning on the side of the municipal council of Lisbon. The outcomes were validated at the results presentation meeting by LXDataLab team.

4.1. Insights and report analysis

4.1.1. Visualization tool

For the visualization, the software used was Microsoft Power BI. In order to grasp the data and respond to the questions we set out to answer, this method was focused on the dashboard representation of spatial and temporal data. As we worked, it became clear that the richer the outcome would be the more information we could give the stakeholders. It is vital to note that understanding of the tool utilized was required in order to construct the full visualization layer built for this research.

4.1.2. Grids overview

With the goal of better knowing the parishes and zones of the city of Lisbon, two maps were created through the arcGIS and Maps tools, in PowerBI, where it was possible to analyze the 3743 grids (200 x 200 meters), divided by parishes (Figure 4.1) and zones (Figure 4.2). These maps were very important in the analysis, since they allowed us to have a georeferenced view of the grids, as well as the zones and parishes.

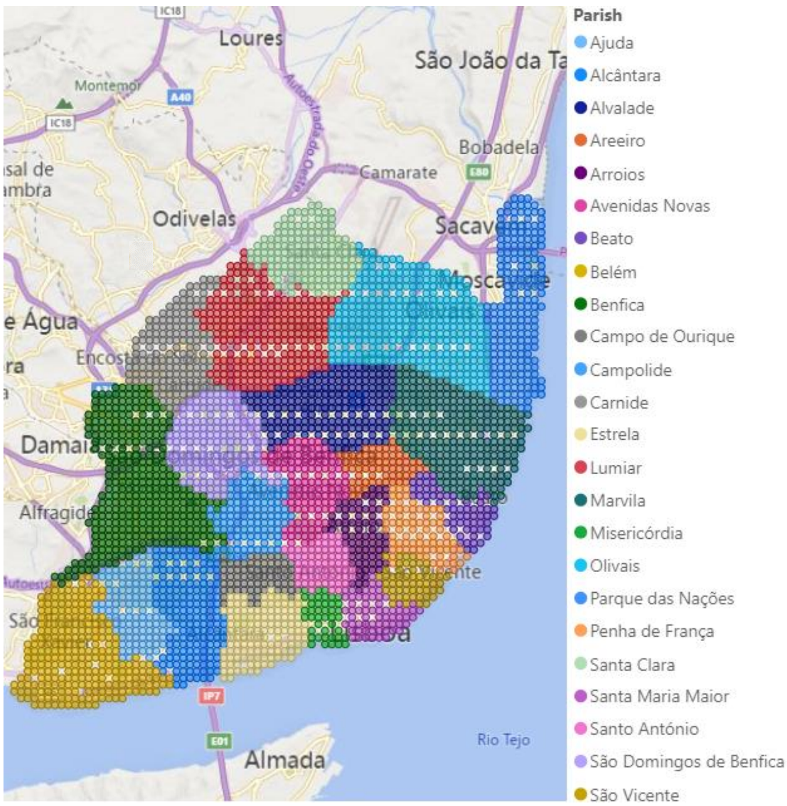


Figure 4.1 - Lisbon operator grids by parishes

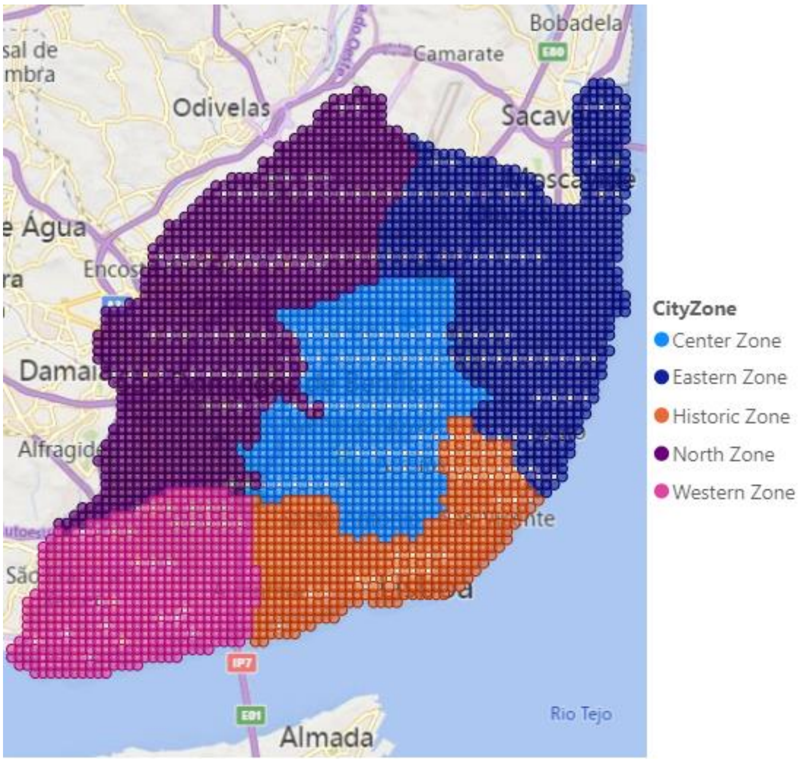


Figure 4.2 - Lisbon operator grids by city zone

4.1.3. Tourists nationalities

Since the objective of this dissertation is to analyze the mobility of national users and tourists in the nightlife areas of the city of Lisbon, we considered that it would be important, in first instance, to know the top 6 main origins of the tourists. To this end, we created the graph in Figure 4.3, which shows us the average number of tourists per cell, for each month under study, where we conclude that the largest number of visitors come from: Brazil, France, Germany, Spain, United Kingdom, and United States of America. We can see that there are fluctuations from month to month, and not always the same nationality is the most common.

Nightlife areas with strong occupation of public space have a great impact on the life of the city of Lisbon, and its management is important in order to respond to the interests of the various stakeholders: merchants, residents, users. The Lisbon City Hall is often responsible for the security and management of these spaces. As such, they can use this information to improve the services provided, from security on public streets, police allocation, cleaning and maintenance and night noise control. For merchants for instance, by knowing the origin of the tourists, they can customize their offer, doing targeted marketing and thus improve their service.

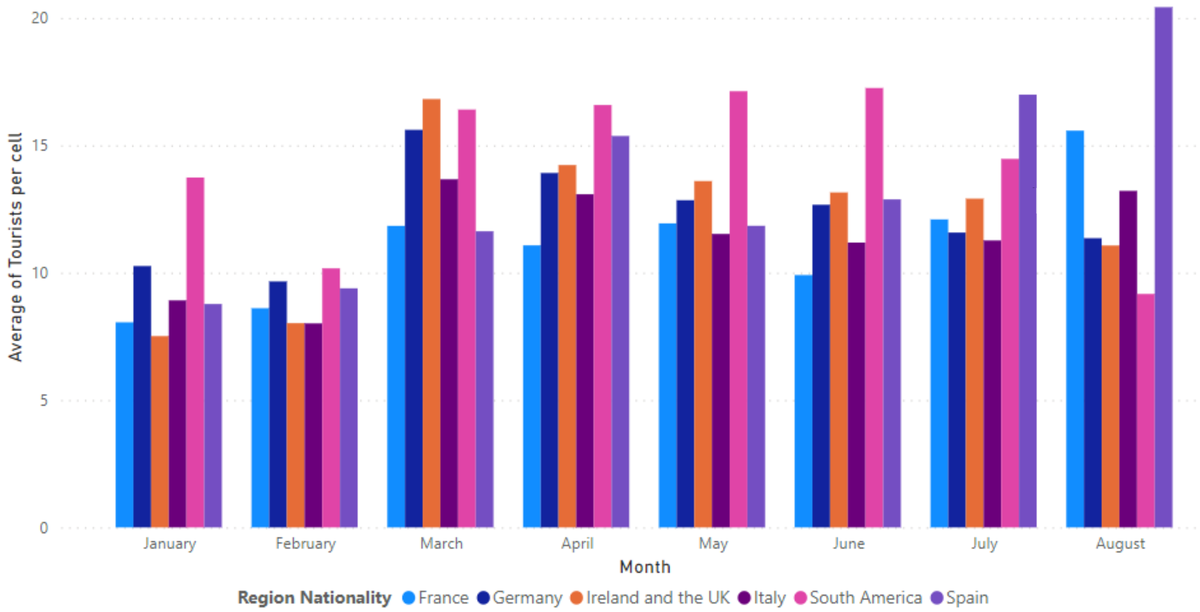


Figure 4.3 - Montly evolution of top 6 nationalities

4.1.4. Monthly and weekly average evolution of national users and tourists

Regarding the average monthly evolution of people in the city of Lisbon, it was found that the highest number of national users is in January and February and has a decreasing trend until August, as shown in Figure 4.4. However, the opposite happens with tourists. January and February are the months with less average tourists per grid. This analysis is particularly interesting, as the mobility of national users and tourists has a negative correlation. A possible justification for this is the fact that most people in Portugal take their holidays in spring and summer, as the temperatures are usually warmer. As such, tourists also have a preference for visiting the city of Lisbon in summer, as can be seen in the graph of Figure 4.5.

Concerning the analysis of the mobility of people per weekday, it was found that this variable has no impact on the mobility of people in the city of Lisbon. It would be expected that there would be some difference between the week and the weekend, but this did not occur.

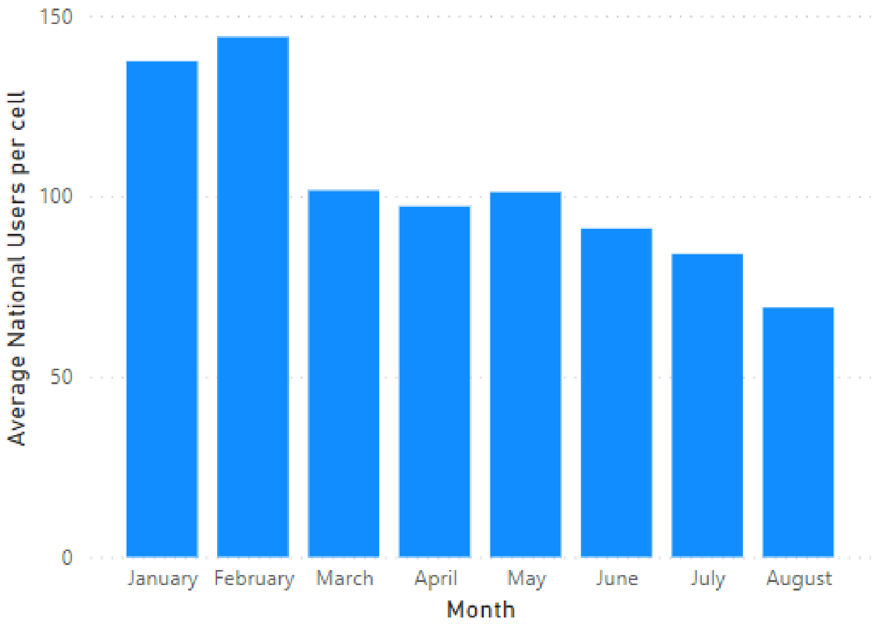


Figure 4.4 - Monthly average evolution of national users

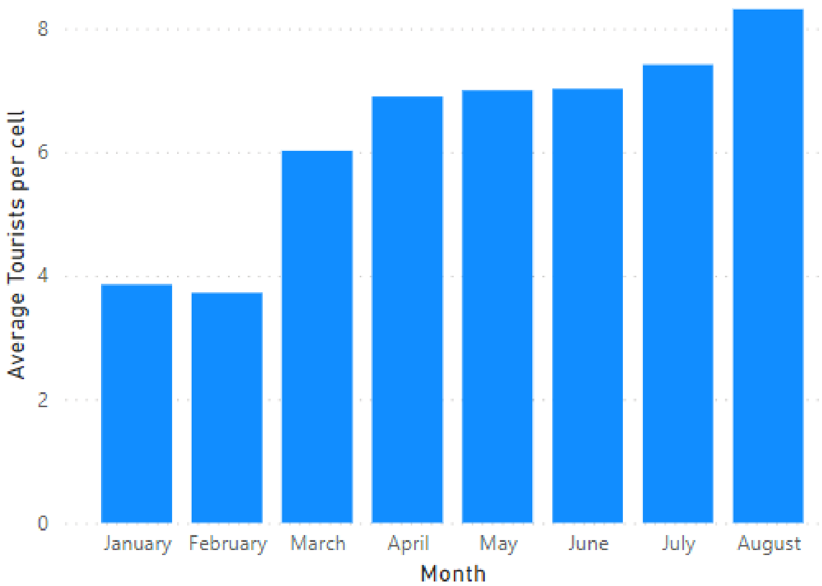


Figure 4.5 - Monthly average evolution of tourists

4.1.5. Hourly evolution of distinct national users and tourists

Understanding the busiest time of the day was important in order to study the mobility of local users and visitors at night. Thus, it was feasible to draw the conclusion, based on Figure 4.6 that between the hours of 11 a.m. and 4 p.m., during the day, there is the most affluence. The busiest nighttime is between 7 p.m. and midnight.

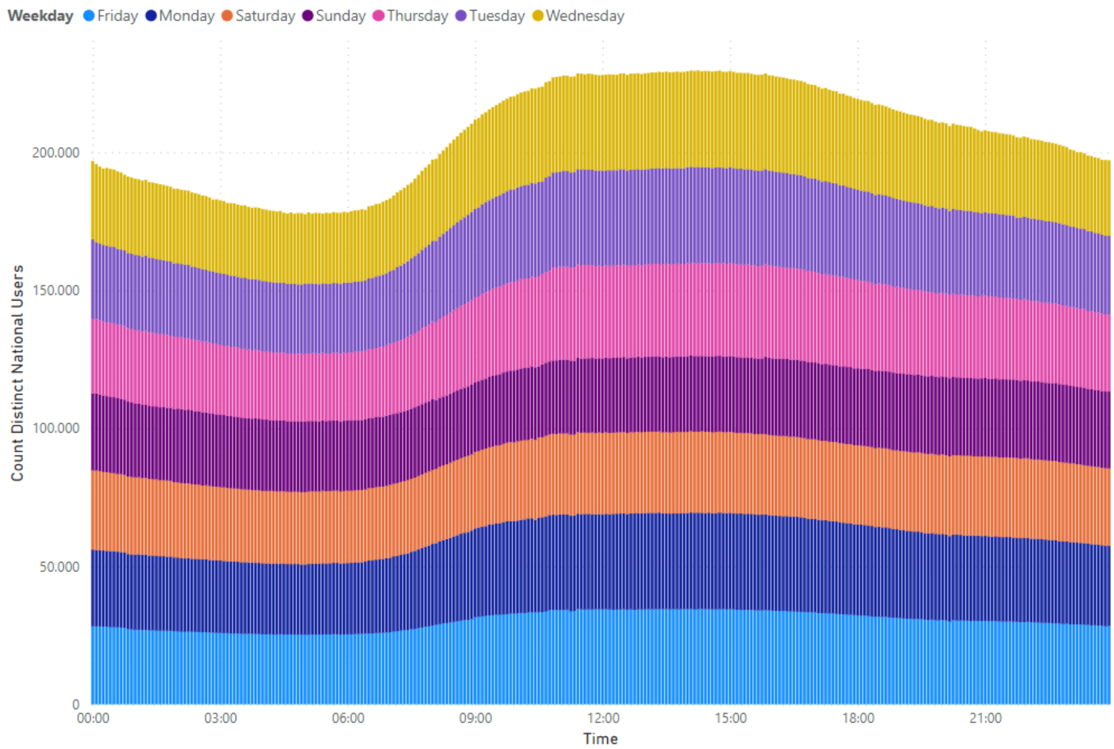


Figure 4.6 - Hourly evolution of national users by weekday

As seen previously, the weekday variable does not have a major impact on the mobility of either national users or tourists. However, we analyzed the hourly evolution by day of the week and, while for local people there are no major differences, there are for tourists. As can be seen in Figure 4.7, there is a huge decrease between midnight and 8 a.m.

The causes for this drop may not be determined with certainty from the data at hand, just as those in charge of LxDataLab did not give us any more details about this topic. However, since the goal of this dissertation is to analyze the mobility of people in Lisbon's nightlife, this descent can give us a clue about the behavior of national users and tourists during the nightlife period. This perception was the starting point for us to begin a more detailed analysis about the mobility of tourists.

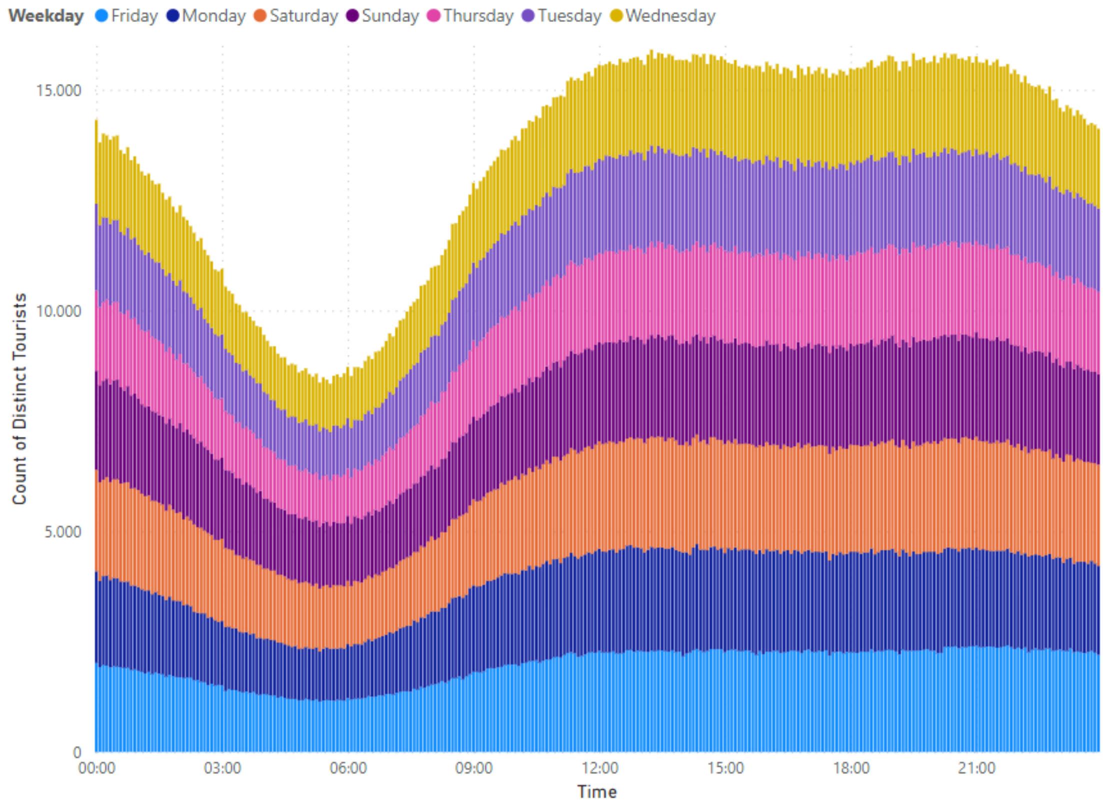


Figure 4.7 - Hourly evolution of tourists by weekday

4.1.6. Which are the busiest zones in Lisbon? And at what time at night?

To study Lisbon by zones, it was necessary to create the variable City Zone. The original data has only the city's parishes. Therefore, the city was divided into 5 zones: north, east, west, center and historic. The northern zone is composed of the following parishes: Santa Clara; Lumiar; Carnide; São Domingos de Benfica and Benfica. The west zone is constituted by: Alcântara; Ajuda and Belém. The eastern zone has the following parishes: Beato; Marvila; Olivais; Parque das Nações. The central zone is composed of: Campolide; Alvalade; Avenidas Novas; Areeiro; Arroios; Santo António. Finally, the historical zone has the following parishes: Campo de Ourique; Estrela; Misericórdia; Santa Maria Maior; São Vicente and Penha de França.

Initially, a bar graph study of the average monthly number of visitors and national users by zone of the city allowed for the conclusion that the historical and center zones were the busiest. To have a more exact location, we made the same analysis for the parishes, where we prepared the graph of Figures 4.8 and 4.9, with the top 6 busiest parishes. The parish dimension allows us to understand the mobility of people in the city of Lisbon in greater detail. As expected, the parishes with the highest average concentration of people per grid are parishes in the central and historical areas.

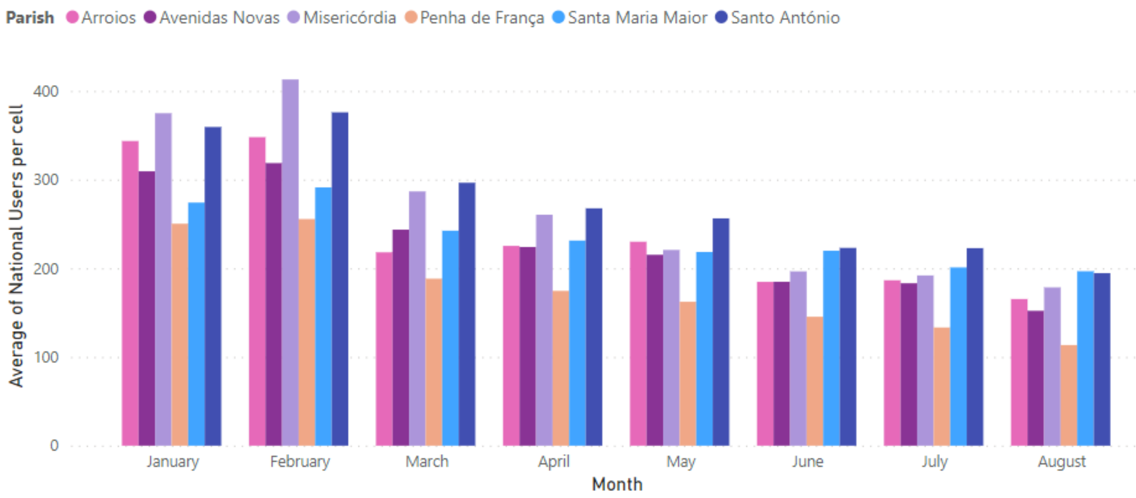


Figure 4.8 - Top 6 most visited parishes by national users

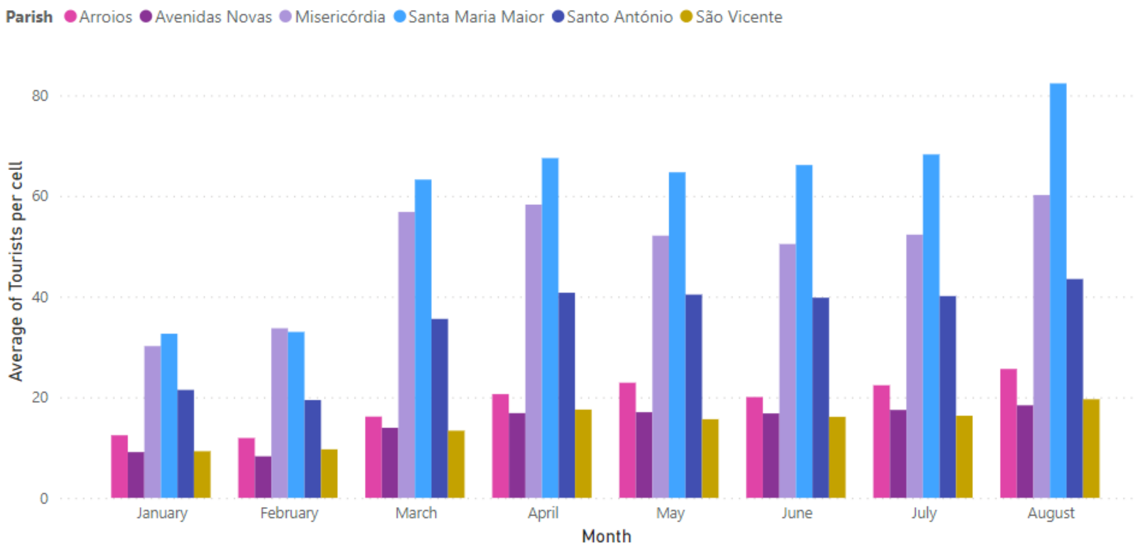


Figure 4.9 - Top 6 most visited parishes by tourists

Regarding the historical zone, the two parishes with the highest average number of national users and tourists are: Santa Maria Maior and Misericórdia. These results are consistent in that these parishes are home to some of the most emblematic monuments of the city of Lisbon. For the central zone, the two busiest parishes are: Santo António and Arroios. In these parishes are located important locations such as: Avenida da Liberdade, Marquês de Pombal and Alameda.

In order to analyze nightlife mobility, it was necessary to create a variable that categorized the time-of-day as shown in Table 4.1 The scope of the study focuses on the hourly period between 05:00 pm and 06:00 am. Through the analysis of the graphs in Figures 4.10 and 4.11, it was possible to verify that the busiest night periods, both for national users and tourists, are dinner (07:00 pm to 10:00 pm), after dinner (10:00 pm to 12:00 pm) and midnight period (12:00 pm to 02:00 am). Late afternoon, although not considered as nighttime, is the last period of the day before night and has therefore been considered for analysis. It is also concluded that on average, the number of people per cell has an increasing trend from 5 p.m. to midnight. Passing midnight, with each passing night period the number of average people per grid decreases until dawn.

Table 4.1 - Day moment period variable

ID	Day Moment	Time Range
0	Midnight Period	00:00 AM - 01:59 AM
1	Party Period	02:00 AM - 03:59 AM
2	After Party	04:00 AM - 05:59 AM
3	Dawn	6:00 AM - 07:59 AM
4	Morning	8:00 AM - 11:59 AM
5	Lunch Time	12:00 PM - 13:59 PM
6	Afternoon	14:00 PM - 16:59 PM
7	Late Afternoon	17:00 PM - 18:59 PM
8	Dinner Time	19:00 PM - 21:59 PM
9	After Dinner	22:00 PM - 23:59 PM

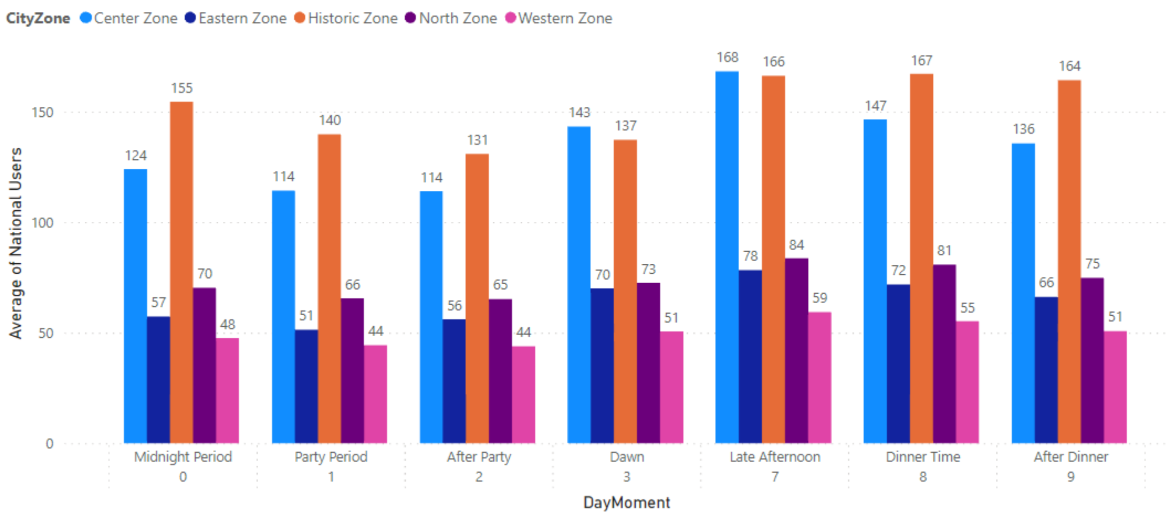


Figure 4.10 - Average evolution of national users by day moment period

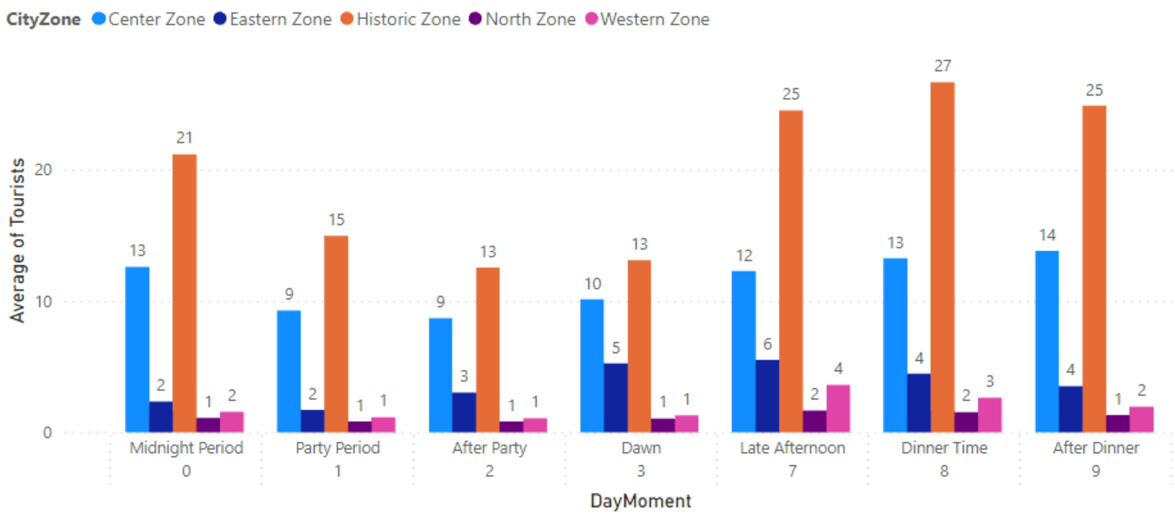


Figure 4.11 - Average evolution of tourists by day moment period

4.1.7. Monthly evolution of people's mobility in the historical and central zones

After concluding which were the two busiest zones, it was important to analyze in more detail each of these (historic and center). Table 4.2 was produced to this goal and displays the average number of national users and visitors per grid for each parish's zone. Within the historical area it was possible to conclude that the parish of Misericórdia and Santa Maria Maior were the most frequented by locals and visitors. As for center zone, the most frequented parishes are Santo António and Arroios.

Table 4.2 - Average national users and tourists per historic and center zone parishes

	Parish	Average of National Users	Average of Tourists
Historic Zone	Santa Maria Maior	226,91	62,56
	Misericórdia	246,61	50,43
	São Vicente	128,81	15,38
	Estrela	104,32	7,24
	Campo de Ourique	161,38	6,04
	Penha de França	167,97	5,88
Center Zone	Santo António	261,01	36,89
	Arroios	226,78	20,09
	Avenidas Novas	217,3	15,54
	Areeiro	152,89	6,05
	Campolide	92,6	4,25
	Alvalade	111,34	3,45

To complete the analysis the months under study were added, where we see the same trend as in Figure 4.5: there is a greater number of tourists as summer approaches and the opposite for locals, Figure 4.4. At the beginning of the year there is a higher average number of national users per grid, and it decreases as the months progress, Figures 4.12 and 4.13.

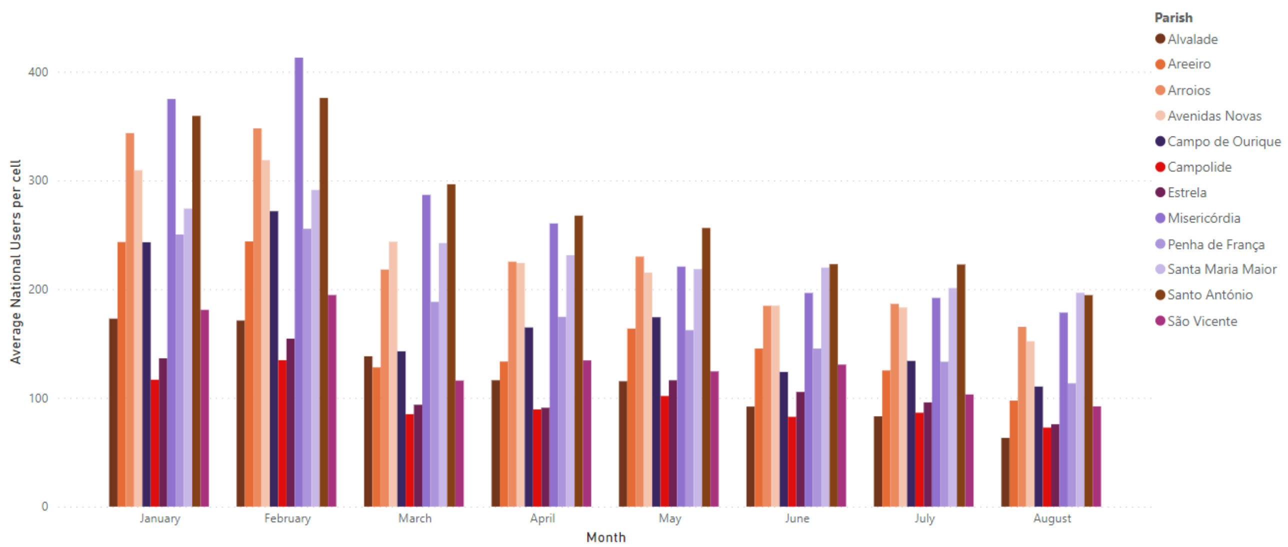


Figure 4.12 - Evolution of national users in center and historic parishes

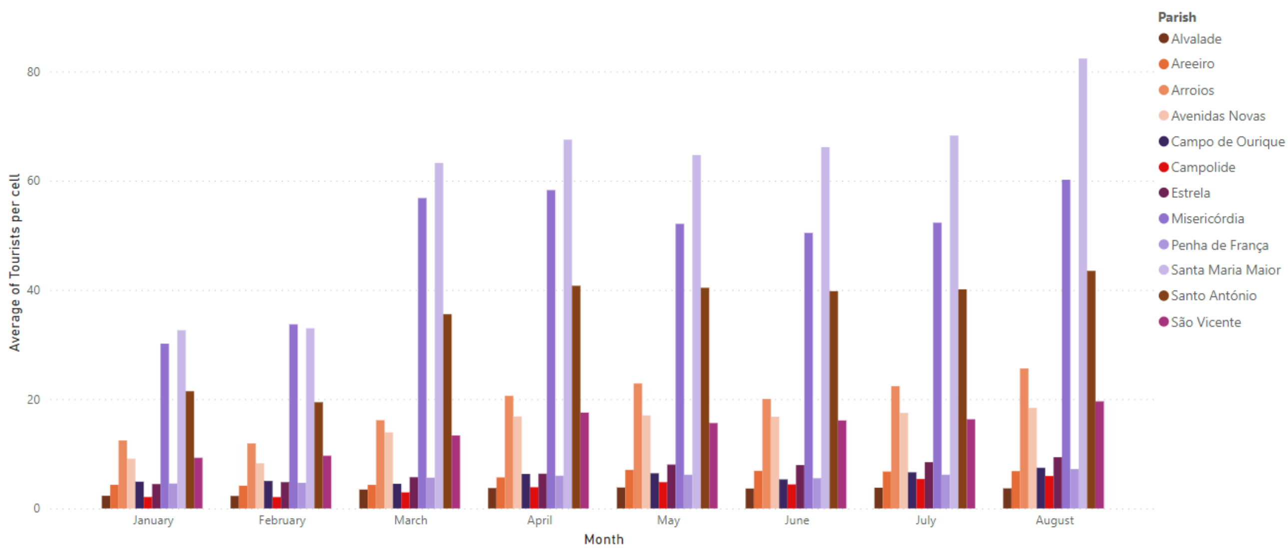


Figure 4.13 - Evolution of tourists in center and historic parishes

4.1.8. Heatmaps analysis: in-depth cell grids mobility for national users in and between nightlife areas.

As stated before, the objective of this dissertation is to analyze the mobility of people and tourists in nightlife areas, with possible transfers between areas and relating it to the noise levels from environmental sensors. For this, the nightlife period considered was between 05:00 pm and 06:00 am.

As for national users, it is possible to verify that the greatest concentration of people during the nightlife period is in the Bairro Alto, Largo Camões, and Pink Street areas. These areas, especially Bairro Alto and Pink Street are some of the most famous areas of Lisbon in terms of nightlife, such as bars, discos, music, so the results present in Figure 4.14 make perfect sense.

It is also possible to verify that there is some concentration of users around Penha de França area. After presenting these results to the LXDataLab department team from Lisbon City Council, we were told that they were unaware of the reason for the concentration of people in this location, since it is not an area with large nightlife business. The same can't be said of the Bairro Alto area, being undoubtedly the area with the highest concentration of people regarding nightlife, and the LXDataLab team informed us that it is one of the locations that involves the most strategic planning regarding public safety means and resources. Since Bairro Alto is located in the Misericórdia parish, the continuation of the analysis will be for this parish.

The next stage was to create a heatmap for each nightlife period in order to track how users moved across different nightlife areas. As can be seen in Figure 4.14, in the late afternoon period (05:00 pm to 07:59 pm), the great concentration of users is at Praça Dom Pedro IV, Rossio; in the Elevador de Santa Justa area, between Rua Áurea and Rua do Carmo, and in the emblematic Rua Augusta.

Regarding the dinner period (07:00 pm to 09:59 pm), it is possible to verify that there is a transfer of the average concentration of national users to the Bairro Alto, Largo Camões, Baixa-Chiado and Pink Street areas.

After the dinner period, we can see that in the next two periods, after dinner (10:00 pm to 11:59 pm) and midnight period (midnight to 1:59 am), the average concentration of people increases at these locations, peaking between 10:00 pm and 11:59 pm. The crowds don't start to thin out until the party period (02:00 am to 03:59 pm). However, there is still a considerable concentration. Only after 4:00 am, the after-party period, we start to see the concentration of people decreasing considerably, and they start moving towards Estrela and Av. 24 de Julho, quite famous for being the most famous nightclub area in Lisbon.

Figure 4.14 - National users heatmap analysis in historic zone at nighttime

4.1.9. Heatmaps analysis: in-depth cell grids mobility for tourists in and between nightlife areas.

In section 4.1.8 of this dissertation the analysis of the mobility of national users during the nightlife period (between 05:00 pm and 06:00 am) was carried out. In this section we will make the same analysis, but for tourists, in order to answer the research question: What is the impact of the mobility of people and tourists, in the nightlife areas, in the city of Lisbon?

Initially, the graph in Figure 4.15 was created, where it is possible to see the heatmap with the average number of tourists per grid for the central and historical zones, during the night period. From these heatmaps it was possible to verify that the zones busiest by tourists are: Bairro Alto, Largo Camões and Baixa-Chiado metro, Rua Augusta, Praça Dom Pedro IV and Praça da Figueira and Santa Justa lift.

Moving on to a more detailed analysis, i.e., by analyzing the average concentration of tourists per period of the day, it is possible to verify not only the permanence of tourists but also the transfers between zones. For the late afternoon period (05:00 pm to 06:59 pm) it is possible to verify that the highest average concentration of tourists is in Praça Dom Pedro IV, Rua Augusta and Santa Justa lift. During the dinner hour (07:00 pm to 09:59 pm), although the flow remains in these areas, tourists move to the Bairro Alto and Largo Camões areas. This trend continues until midnight, peaking also in the period after dinner (10:00 pm to 23:59 pm), as it did with national users.

Regarding the party period (02:00 am to 03:59 am) it is possible to verify a decrease in the concentration of tourists, although they remain mainly in the Bairro Alto area. One of the main differences that can be observed between national users and visitors is the fact that in the last period, after-party (04:00 am to 05:59 am), the concentration of tourists remains in Bairro Alto. This was not verified in the graph of Figure 4.15, where it can be seen that for the same hour period, national users had already moved to another place. The second difference found is related to the transfer of visitors to downtown areas (Rua Augusta, Praça da Figueira, Praça Dom Pedro IV). Most probably, this transfer back to the downtown area is related to the tourists' overnight stay in Lisbon.

Figure 4.15 - Tourists heatmap analysis in historic zone at nighttime

4.1.10. Analysis of environmental noise sensors in the city of Lisbon

To analyze the environmental noise sensors, the latest dataset provided by the Lisbon Municipality was used. Through noise sensors placed around the city of Lisbon, the noise levels are recorded hourly. Four types of noise levels are identified: Normal (29 to 54 dB); Moderate (55 to 59 dB); High (60 to 64 dB) and Very High (>65 dB). The first step was to understand the total number of noise sensors in the city of Lisbon. To do so, a count of the number of different latitudes or longitudes present in the dataset was done, concluding there are 14 in total. The first step was to understand the total number of noise sensors in the city of Lisbon. To do so, a count of the number of different latitudes or longitudes present in the dataset was done, concluding there are 14 in total. Immediately after, all noise sensors were identified in a map, shown in Figure 4.16.

Figure 4.16 - Location map of environmental noise sensors

To get a first insight into the recorded noise values, Table 4.3 was made as a starting point for a brief exploratory analysis of the data, where we analyzed the following measures: count, minimum, maximum, and average.

Table 4.3 - Noise Sensors Indicators

Label	Count of Value	Min of Value	Max of Value	Average of Value
Normal	11067	29	54	50,33
Moderate	8548	55	59	56,99
High	12033	60	64	62,3
Very High	29674	65	92	70,37
Total	61322	29	92	63,31

The very high label count, which accounts for 48% of the total records, was one of the measurements that immediately stood out. From this indicator we can see that the city of Lisbon is quite noisy, and that only 37% of the time it registers values considered normal. If we add up the value of the labels high and very high, we conclude that, in relative terms, the city of Lisbon has high or very high noise levels in 68% of the time. This means that only 32% of the time, Lisbon registers noise levels considered normal or moderate.

The first graph to be developed was the graph of Figure 4.17. It shows us the average noise of each location in Lisbon that has an environmental noise sensor. It was concluded that the 3 locations with higher average noise levels are: Praça do Comércio (75.91 dB); Avenida Infante Santo (72.30 dB) and; Downtown - Rua do Ouro (68.91 dB).

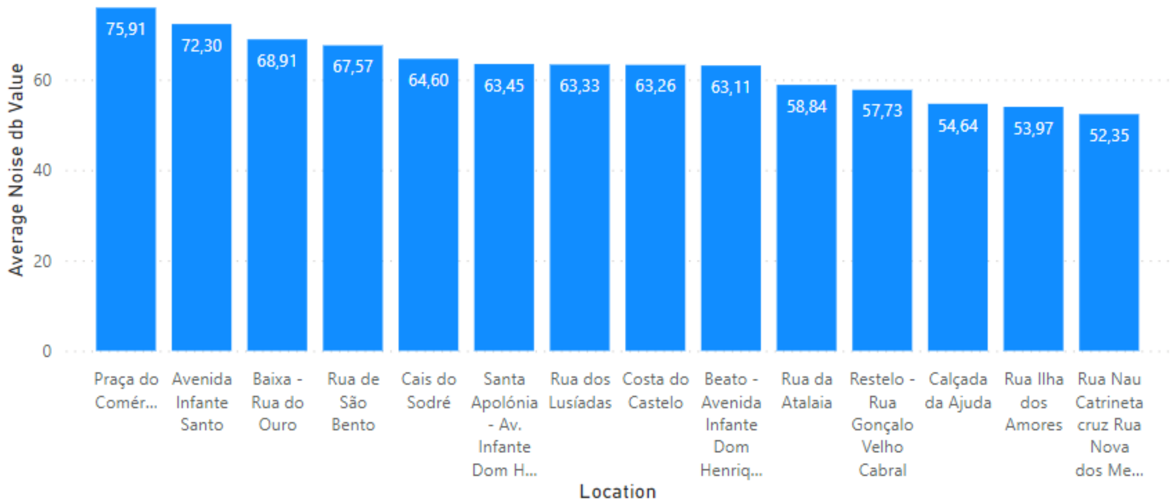


Figure 4.17 - Average noise in each Lisbon noise sensor location

The goal for this dataset is to relate the mobility of people to the noise levels recorded at the environmental sensors in the nightlife period. For this, the graph in Figure 4.18 was created, with the objective of analyzing the hourly evolution of noise levels for each location. In general, it can be concluded that the period of lower noise in the city of Lisbon is between 01:00 am and 06:00 am. As for the period of highest noise in the city is between 07:00 am and 11:00 am. The streets with the least noise are Rua Nau Catrineta Cruz; Rua Ilha dos Amores and Calçada da Ajuda. The first two are located in Parque das Nações, which is a peripheral and more residential area, which justifies the fact that the average noise is among the lowest.

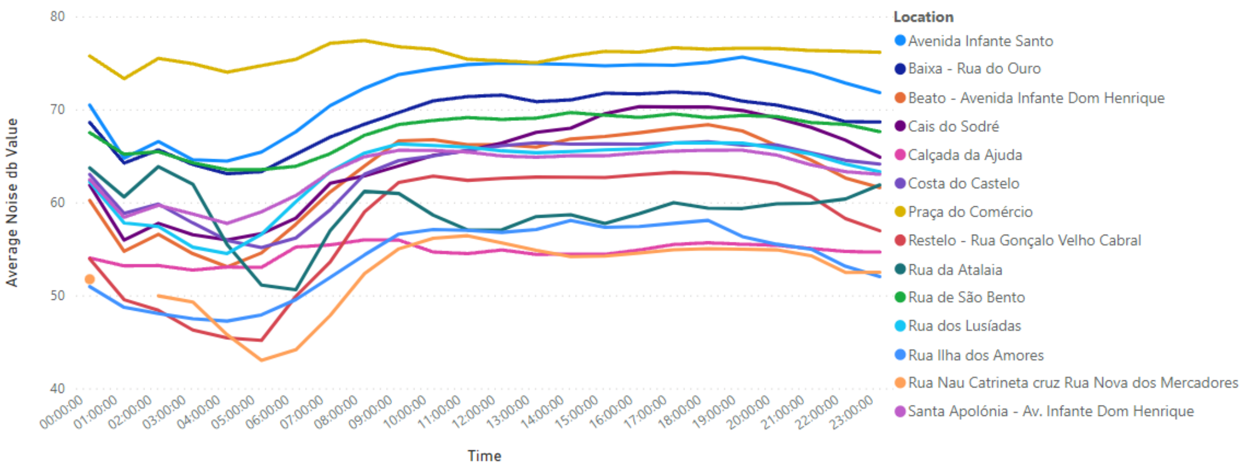


Figure 4.18 - Hourly evolution of noise in Lisbon locations

Although it has been possible to draw some conclusions regarding noise sensors, there is another one regarding the integration of data from this dataset with the main one, provided by the Mobile Operator. The two datasets cannot be combined since there isn't a common column between them.

In addition to this problem, there are still a very small number of sensors. In the busiest areas of the city, there are no environmental noise sensors at all. This information was reported to the responsible team at Lisbon City Hall, suggesting that they either implement a larger number of sensors in the streets, or readjust the existing ones. The idea given was to place the sensors in the busiest areas of the city, either by locals or tourists. These zones were identified throughout this chapter and are: Bairro Alto; Baixa-Chiado Metro; Largo Camões; Pink Street; Santa Justa Lift; Rua Augusta; Rua Áurea; Rua do Carmo; Praça Dom Pedro IV, Avenida da Liberdade; Praça da Figueira; and in the most emblematic neighborhoods of Lisbon's popular saints' festivities. The Lisbon City Council team appreciated the idea and said they will implement it in the future, as soon as possible.

4.1.11. Event Analysis: Popular Saints festivities in Lisbon (Santo António night)

To do the event analysis, the selected event was Santo António night 2022, June 12 to 13. This night is possibly the busiest night in the city of Lisbon, since it is the largest popular and cultural festival in Lisbon. The time period under scope is from 07:00 pm on the 12th to 07:00 am on the 13th. June 13rd is a Lisbon holiday.

On this night, Lisboners greet their patron saint, Santo António and fill the streets of the city's typical neighborhoods, drinking beer and eating sardines. There is a popular march on Avenida da Liberdade, where each neighborhood has its own parade.

After this background, the first indicator to be analyzed was the noise from the environmental sensors. Through Figure 4.19 we can analyze a heatmap with the maximum noise registered in each sensor in the city. Although there are no noise sensors in the most famous areas, such as Avenida da Liberdade, Graça, Alfama and Bica, we realize that the downtown area and Cais do Sodré registered very high noise levels (between 80 and 88 dB). The noise sensor closest to the most famous neighborhood, Graça, is on the castle coast, where there was also a very high maximum noise level.

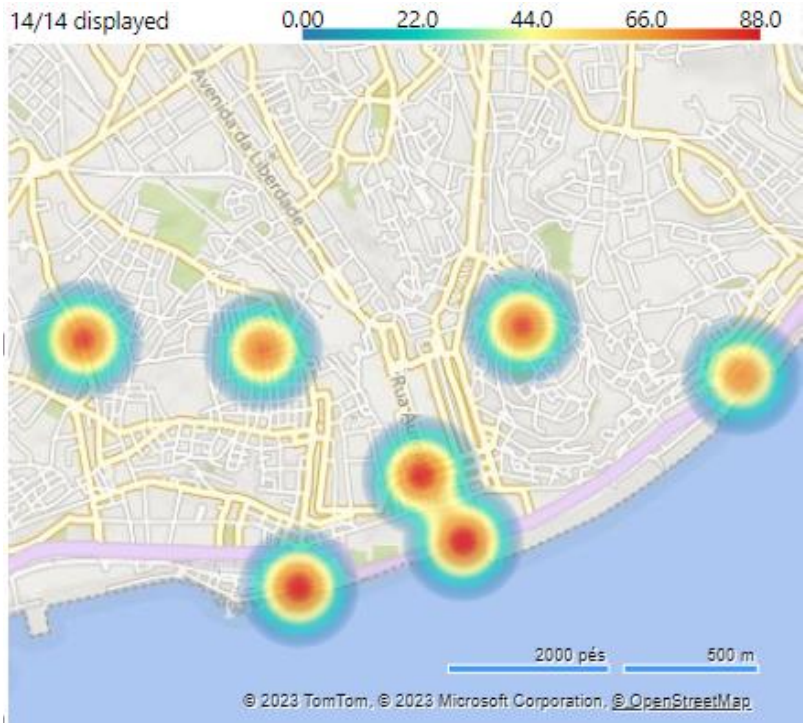


Figure 4.19 - Maximum noise in Santo António night 2022 in each noise location

Analyzing now the evolution of maximum noise, through Figure 4.20 it is possible to conclude that it has a very increasing trend from 09:00 pm to 00:00 am, with maximum values between 74 dB and 85 dB. Passing midnight, the noise has a decreasing trend until 05:00 am, reaching 72 dB, which although lower, is still considered too much noise.

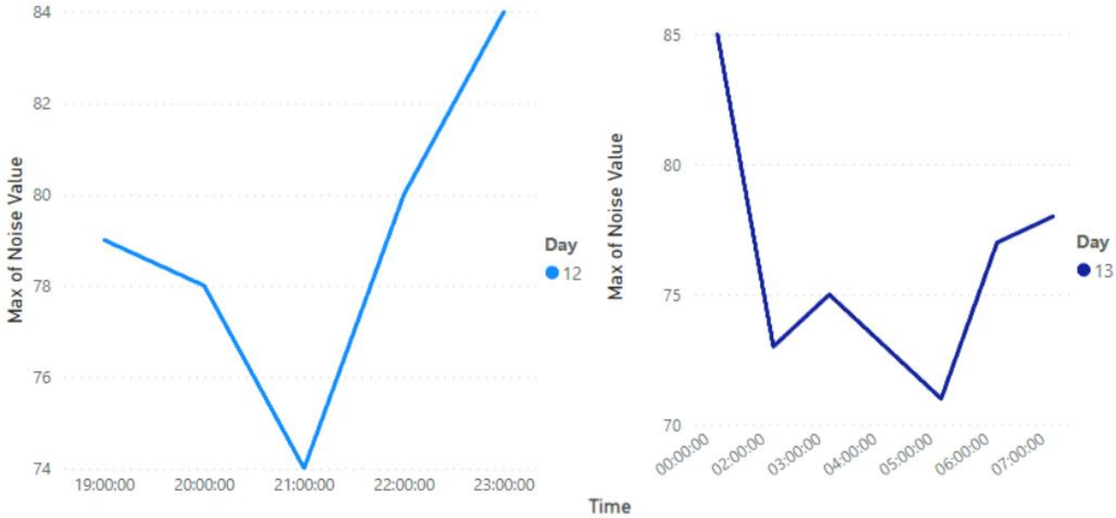


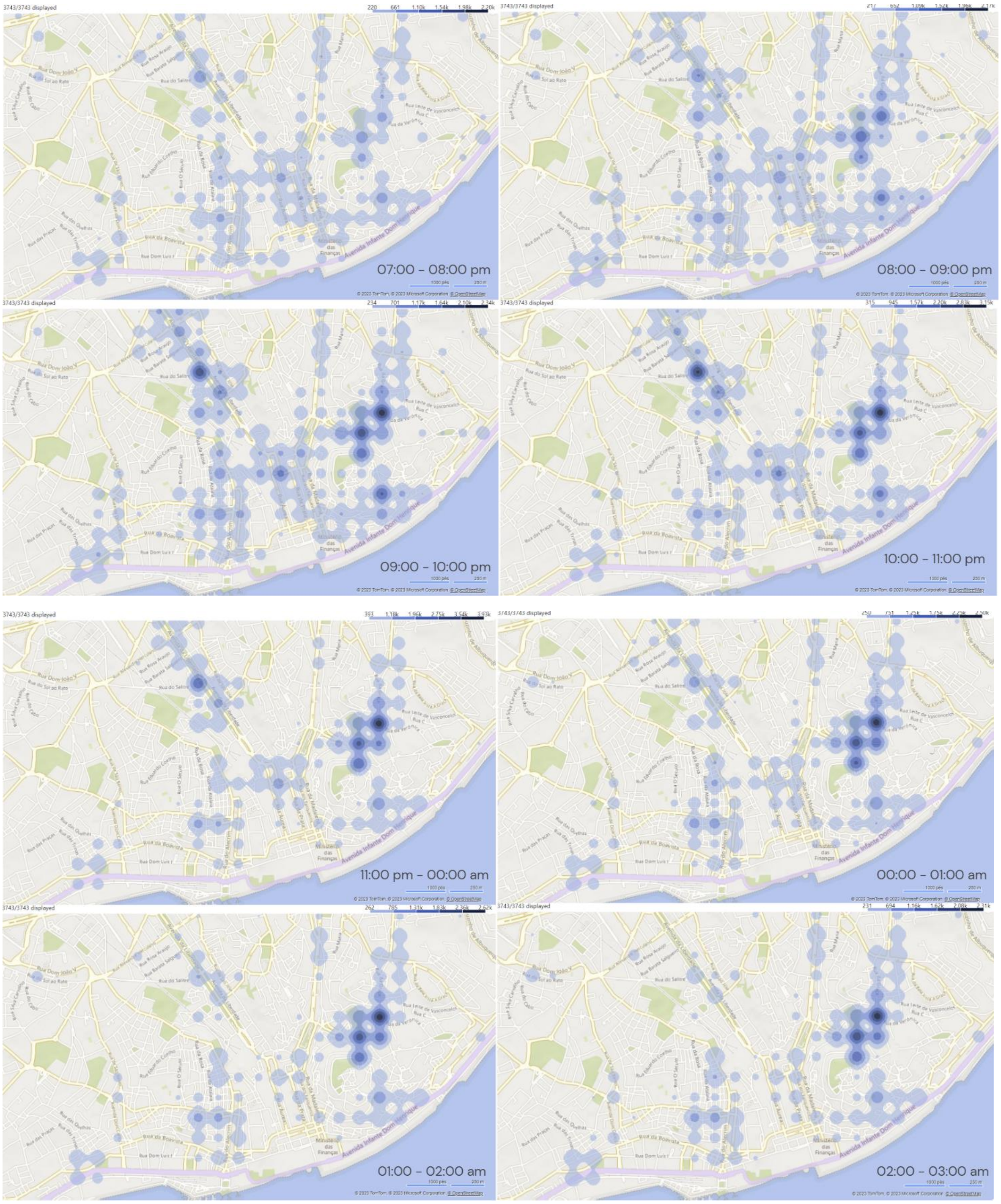
Figure 4.20 - Hourly evolution of max noise in Santo António night 2022

Regarding the mobility of people, an hourly heatmap was performed, during the period under analysis, to understand the patterns and movements between neighborhoods and party areas, as shown in Figure 4.21 Between 07:00 pm and 08:00 pm there are still no large clusters of people. However, moving to the next hour 08:00 pm to 09:00 pm, you can see that people start to concentrate in the more typical neighborhoods, they are: Graça, Vila Berta, Alfama, Bairro Alto and Bica. From 09:00 pm to 10:00 pm, people start to concentrate on Avenida da Liberdade, since it is around this time that the big parade of the popular marches begins. For the following period, between 10:00 pm and 11:00 pm, in addition to the concentration in the neighborhoods of Graça and Vila Berta, there begins to be more and more concentration on Avenida da Liberdade, to observe the marches. As for the period between 11:00 pm and 00:00 pm, the popular marches end and this is reflected in people's mobility, where they move to the Graça neighborhood. This period is also the one that registered the highest number of people.

Starting the holiday, the 13th, people stayed in Graça and Vila Berta neighborhoods, between 00:00 am and 05:00 am. However, the peak of people in Graça occurred between 03:00 am and 04:00 am. For the next hour, from 04:00 am until 05:00 am, the concentration of people decreased, although they remained in the same place. The next hour in the dawn, is when the party ends, and you can clearly see that people start to move away and go in opposite directions. In the last hour, the flow of people is much lower, where most have already left.

In the results presentation meeting they were surprised for example, that the busiest neighborhood was Graça, when they thought it was Alfama. We were told that there was a greater allocation of police forces in Alfama rather than Graça.

Thus, we can conclude that it makes more sense to reinforce the public security forces in the Graça neighborhood. Data-based knowledge for this specific night is crucial, as the city council has a lot of work to do in preparing and managing resources for this night. From police forces, teams to close the roads, security teams ambulances, firefighters, to merchants and residents. The result of this event analysis was one of the most relevant for the study, as it allowed the Lisbon city council to be able to better manage all the preparations for this night. From now on, they will be able to better manage all the preparations for this night.


07:00 - 08:00 pm
08:00 - 09:00 pm
09:00 - 10:00 pm
10:00 - 11:00 pm
11:00 pm - 00:00 am
00:00 - 01:00 am
01:00 - 02:00 am
02:00 - 03:00 am

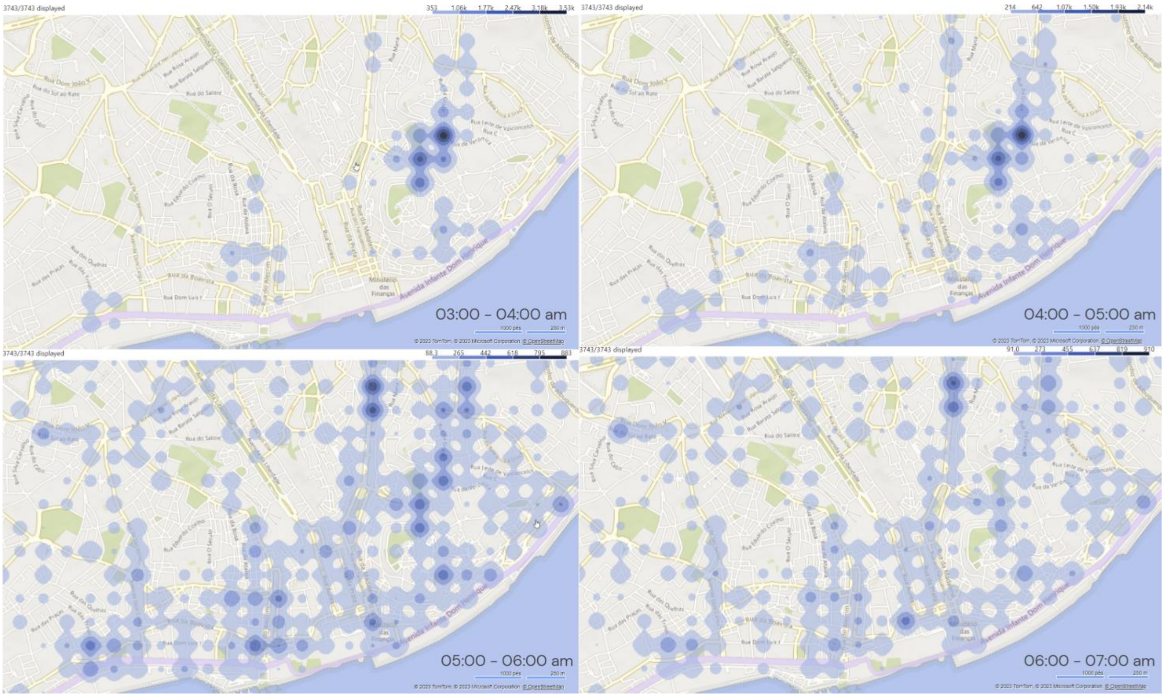


Figure 4.21 - Santo António night 2022 hourly heatmap analysis

CHAPTER 5

Conclusions and future work

5.1. Conclusions

Human mobility is one of the most important ecological and social challenges of the twenty-first century. Moving from traditional data (census and survey data) to location-based big data, opens up new opportunities for movement study. Mobility has emerged as a significant geographic concept. Cities are systems and spaces that connect networks of human connections. Understanding the underlying nodes, networks, and links that make up the fabric of cities is the first step towards building better cities. The ability to watch and comprehend mobility behavior at a previously unseen level of detail is made possible by the advent of new Big Data sources including call records from mobile phones, smart card data, and geo-coded social media records. As a result of the ever-increasing volume of data produced by smartphones, telecom corporations have access to vast amounts of data, which makes them sitting on a gold mine.

The aim of this dissertation is to analyze the mobility of people and tourists in nightlife areas, with possible transfers between areas and relating it to the noise levels from environmental sensors. To achieve the objective, an in-depth analysis was developed, and we were able to address all the questions we wanted to. Through the construction and analysis of dashboards, firstly, it was possible to identify the origin of tourists who visited the city of Lisbon between January and August 2022. Soon after, the weekly and monthly evolution of national users and tourists was studied. Here it was possible to conclude that there is a greater number of national users in January and February, and as the months go by, the number of people decreases until August. On the other hand, the lowest number of tourists in the city of Lisbon is in January and February and it keeps growing until July and August, where it reaches its peak. It was also found that the day of the week has no impact on the mobility of people.

In order to understand the night period with the highest influx of people, the hourly evolution of distinct national users and tourists was explored. It was concluded that the busiest night period is between 07:00 pm and midnight. It was also curious to realize that there is a big drop in the number of tourists between midnight and 08:00 am, although we do not have the concrete reasons for this drop, since the data does not provide this type of information.

Next, we looked at which areas of the city were busiest and at what time. The most crowded parishes in the historic and central areas are Misericórdia and Santa Maria Maior; Santo António and Arroios, respectively. Thus, it was possible to conclude that the busiest areas of the city are the historic and central zones. The busiest night periods in these areas, both for national users and tourists, are the dinner period, from 07:00 pm to 10:00 pm; the period after dinner, between 10:00 pm and midnight; and finally, the midnight period, between midnight and 02:00 am. For each additional night period, the average number of people decreases until dawn.

The most crucial step was to determine which grids in Lisbon are busiest throughout the nighttime hours and which are the transfers of users between nighttime zones; with the examination of heatmaps. Thus, it was possible to conclude that the busiest cells in the city of Lisbon, both by tourists and by national users are, in descending order: Bairro Alto; Largo Camões and Baixa-Chiado metro; Rua Augusta; Praça Dom Pedro IV; Praça da Figueira and Santa Justa elevator. Throughout the night period, i.e., between 07:00 pm and 07:00 am, Bairro Alto was undoubtedly the location with the highest concentration of people, peaking between 10:00 pm and midnight. Regarding the transfer of people between nightlife areas, it was possible to conclude that tourists leave Bairro Alto in the after-party period, between 04:00 am and 06:00 am and head towards the downtown area: Rua Augusta; Praça da Figueira and Praça Dom Pedro IV. This movement should correlate with the overnight stays of tourists in downtown hotels, although it is not possible to validate this information, based only on the data provided.

Regarding the data from the environmental noise sensors, although there are only 14 sensors in the city and they are not located in the most crowded areas at night, it was possible to conclude that, from January to August 2022, 48% of the total readings were greater than 65dB, considered as very high noise. Only 32% of the time, the city of Lisbon has noise levels considered normal or moderate, between 29 and 59 dB. It was also concluded that the streets with the highest average noise are: Praça do Comércio (75.91 dB); Avenida Infante Santo (72.30 dB) and; Downtown - Rua do Ouro (68.91 dB). This information was reported to the responsible team at Lisbon City Hall, suggesting that they either implement a larger number of sensors in the streets, or readjust the existing ones. The idea given was to place the sensors in the busiest areas of the city, either by locals or tourists. The Lisbon City Council team appreciated the idea and said they will implement it in the future, as soon as possible.

The last conclusions drawn concern Lisbon's most famous festive event: the night of Santo António. LXDataLab team were surprised for example, that the busiest neighborhood was Graça, when they thought it was Alfama. We were told that there was a greater allocation of police forces in Alfama rather than Graça.

Thus, we can conclude that it makes more sense to reinforce the public security forces in the Graça neighborhood. Data-based knowledge for this specific night is crucial, as the city council has a lot of work to do in preparing and managing resources for this night. From police forces, teams to close the roads, security teams ambulances, firefighters, to merchants and residents. The result of this event analysis was one of the most relevant for the study, as it allowed the Lisbon city council to be able to better manage all the preparations for this night. From now on, they will be able to better manage all the preparations for this night.

Finally, it is also crucial to note that, given the data at our disposal, we are capable of gaining many more insights than those we use in the present research. However, given our limited options, we have decided on the selection that seems to suit the best. The created tool has been developed with the ability to use a number of filters, allowing the Lisbon City Council to do its own study on subjects not included in this dissertation and to conduct more thorough final analysis from a chronological and geographic perspective.

5.2. Future work

We consider it a suggestion for future work, in the sense of utilizing this data in two different ways, given the real-time processing and storage capacity already available.

Development of a system that enables georeferenced data to be analyzed in real time. There are often riots and disturbances in the nightclub and bar areas of Lisbon. As such, the possibility of knowing in real time the movements and large agglomerations of people would be very useful, as it is easier to act, if necessary. From public security forces, emergencies, transportation and other urgent services. This information could prevent many serious and problematic situations, if used properly.

Plan for the medium to long future using this data. For this proposal, we recommend using the data to do urban planning, namely public services. With such a rich dataset it is possible to predict and plan, for example, which areas would need more public transport or better access. City planning based on people mobility data is the key to sustainable city development; greater social inclusion and accessibility; better public services and welfare; and ultimately, economic growth.

Both proposals aim to improve infrastructure, and enhance the overall livability of the city, for residents and tourists.

References

[1] C. Anda, A. Erath, and P. J. Fourie, 'Transport modelling in the age of big data', *International Journal of Urban Sciences*, vol. 21, no. sup1, pp. 19–42, Aug. 2017, doi: 10.1080/12265934.2017.1281150.

[2] D. Massimo, E. Not, and F. Ricci, 'User behaviour analysis in a simulated IoT augmented space*', in *Int Conf Intell User Interfaces Proc IUI*, Association for Computing Machinery, 2018. doi: 10.1145/3180308.3180316.

[3] I. Arhipova, G. Berzins, E. Brekis, J. Binde, and M. Opmanis, 'Mobile phone data statistics as proxy indicator for regional economic activity assessment', presented at the FEMIB 2019 - 1st International Conference on Finance, Economics, Management and IT Business, 2019, pp. 27–36. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85067460029&partnerID=40&md5=c7874103c9efc3f793e551733ccd16a2

[4] 'World Telecommunication Development Conference (WTDC-14): Final Report', *ITU*. https://www.itu.int:443/en/publications/ITU-D/Pages/publications.aspx (accessed Jan. 24, 2023).

[5] A. Pavlicek, P. Doucek, R. Novák, and V. Strizova, *Big data analytics – Geolocation from the perspective of mobile network operator*, vol. 310. in 11th IFIP WG 8.9 Working Conference on Research and Practical Issues of Enterprise Information Systems, CONFENIS 2017, vol. 310. Springer Verlag, 2018, p. 131. doi: 10.1007/978-3-319-94845-4_11.

[6] M. Šauer, J. Vystoupil, M. Novotná, and K. Widawski, 'Central European Tourist Flows: Intraregional Patterns and Their Implications', *Morav. Geogr. Rep.*, vol. 29, no. 4, pp. 278–291, 2021, doi: 10.2478/mgr-2021-0020.

[7] R. Doran, S. Larsen, and K. Wolff, 'Different but Similar: Social Comparison of Travel Motives Among Tourists', *International Journal of Tourism Research*, vol. 17, no. 6, pp. 555–563, 2015, doi: 10.1002/jtr.2023.

[8] 'WTCF. World Tourism Economy Trends Report; 2020', Oct. 01, 2020. http://travel.cctv.com/2020/01/10/ARTIGAy04Rv3zmZ4e7WO0SEz200110.shtml (accessed Jan. 10, 2023).

[9] J. Liu, M. Yue, F. Yu, and Y. Tong, 'The contribution of tourism mobility to tourism economic growth in China', *PLoS ONE*, vol. 17, no. 10 October, 2022, doi: 10.1371/journal.pone.0275605.

[10] C. Haikal, P. Alizadeh, C. Rodrigues, and B. Chongke, 'Place embedding across cities in location-based social networks', in *Proc ACM Symp Appl Computing*, Association for Computing Machinery, 2022, pp. 539–546. doi: 10.1145/3477314.3506992.

[11] C. Jin, J. Cheng, and J. Xu, 'Using User-Generated Content to Explore the Temporal Heterogeneity in Tourist Mobility', *Journal of Travel Research*, vol. 57, no. 6, pp. 779–791, Jul. 2018, doi: 10.1177/0047287517714906.

[12] 'LxDataLab', *LISBOA ABERTA*. https://lisboaaberta.cm-lisboa.pt/index.php/pt/lx-data-lab/apresentacao (accessed Jul. 10, 2023).

[13] S. Phithakkitnukoon, T. Horanont, A. Witayangkurn, R. Siri, Y. Sekimoto, and R. Shibasaki, 'Understanding tourist behavior using large-scale mobile sensing approach: A case study of mobile phone users in Japan', *Pervasive and Mobile Computing*, vol. 18, pp. 18–39, 2015, doi: 10.1016/j.pmcj.2014.07.003.

[14] C. Haikal, P. Alizadeh, C. Rodrigues, and B. Chongke, 'Place embedding across cities in location-based social networks', presented at the Proceedings of the ACM Symposium on Applied Computing, 2022, pp. 539–546. doi: 10.1145/3477314.3506992.

[15] C. Schröer, F. Kruse, and J. M. Gómez, ‘A Systematic Literature Review on Applying CRISP-DM Process Model’, *Procedia Computer Science*, vol. 181, pp. 526–534, Jan. 2021, doi: 10.1016/j.procs.2021.01.199.
[16] D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman, ‘Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement’, *BMJ*, vol. 339, p. b2535, Jul. 2009, doi: 10.1136/bmj.b2535.
[17] ‘Searching with Wildcards in EDS and EBSCOhost’. https://connect.ebsco.com/s/article/Searching-with-Wildcards-in-EDS-and-EBSCOhost?language=en_US (accessed Jan. 26, 2023).
[18] C. Mizzi, A. Fabbri, S. Rambaldi, F. Bertini, N. Curti, S. Sinigardi, R. Luzi, G. Venturi, M. Davide, G. Muratore, A. Vannelli, and A. Bazzani, ‘Unraveling pedestrian mobility on a road network using ICTs data during great tourist events’, *EPJ Data Science*, vol. 7, no. 1, 2018, doi: 10.1140/epjds/s13688-018-0168-2.
[19] X. Li, C. Yu, L. Ju, J. Qin, Y. Zhang, L. Dou, and S. Yuqing, ‘Position prediction system based on spatio-temporal regularity of object mobility’, *Inf. Syst.*, vol. 75, pp. 43–55, 2018, doi: 10.1016/j.is.2018.02.004.
[20] U. Türk, J. Östh, K. Kourtit, and P. Nijkamp, ‘The path of least resistance explaining tourist mobility patterns in destination areas using Airbnb data’, *Journal of Transport Geography*, vol. 94, 2021, doi: 10.1016/j.jtrangeo.2021.103130.
[21] M. J. Hiltunen and A. Rehunen, ‘Second home mobility in Finland: Patterns, practices and relations of leisure oriented mobile lifestyle’, *Fennia*, vol. 192, no. 1, pp. 1–22, 2014, doi: 10.11143/8384.
[22] Y. Bayar and B. Yener, ‘Political stability and tourism sector development in Mediterranean countries: A panel cointegration and causality analysis’, *European Journal of Tourism Research*, vol. 21, pp. 23–32, 2019.
[23] B. Zaragozí, S. Trilles, A. Gutiérrez, and D. Miravet, ‘Development of a common framework for analysing public transport smart card data’, *Energies*, vol. 14, no. 19, 2021, doi: 10.3390/en14196083.
[24] E. Carter, P. Adam, D. Tsakis, S. Shaw, R. Watson, and P. Ryan, ‘Enhancing pedestrian mobility in Smart Cities using Big Data’, *Journal of Management Analytics*, vol. 7, no. 2, pp. 173–188, 2020, doi: 10.1080/23270012.2020.1741039.
[25] S. Phithakkitnukoon, T. Horanont, A. Witayangkurn, R. Siri, Y. Sekimoto, and R. Shibasaki, ‘Understanding tourist behavior using large-scale mobile sensing approach: A case study of mobile phone users in Japan’, *Pervasive Mob. Comput.*, vol. 18, pp. 18–39, 2015, doi: 10.1016/j.pmcj.2014.07.003.
[26] F. Qi and F. Du, ‘Trajectory data analyses for pedestrian space-time activity study’, *Journal of Visualized Experiments*, no. 72, 2013, doi: 10.3791/501302013.
[27] E. Ferrara, M. Uras, L. Atzori, O. Bagdasar, and A. Liotta, ‘Mobility Analysis during the 2020 Pandemic in a Touristic city: The Case of Cagliari’, presented at the 2021 IoT Vertical and Topical Summit for Tourism, 2021. doi: 10.1109/IEEECONF49204.2021.9604867.
[28] P. Keawsomnuk, ‘A structural equation model of factors relating to smart cities that affect the management of the world heritage site as well as the quality of life of tourists and villagers in Ayutthaya, Thailand’, *Humanities, Arts and Social Sciences Studies*, vol. 21, no. 1, pp. 35–42, 2021, doi: 10.14456/hasss.2021.4.
[29] C. Ezel and H. Arasli, ‘Movement from emerging economies to small island states: Motivations of nigerian educational tourists in north cyprus’, *Island Studies Journal*, vol. 16, no. 1, pp. 135–154, 2021, doi: 10.24043/isj.87.
[30] Passerini G. and Borrego C., Eds., ‘25th International Conference on Urban Transport and the Environment’, in *WIT Trans. Built Environ.*, WITPress, 2019. Accessed: Jun. 25, 2019. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85077321730&partnerID=40&md5=6a0bed10728ba9cc40cef22f85715854

[31] F. Pfaender and J. Fen-Chong, ‘Exploratory analysis of urban mobility: from mobile phone usage data to tourist behaviour’, in *IET Conf Publ*, Shanghai, 2011, p. 15. doi: 10.1049/cp.2011.0310.

[32] K. Çakar and Ş. Aykol, ‘Understanding travellers’ reactions to robotic services: a multiple case study approach of robotic hotels’, *Journal of Hospitality and Tourism Technology*, vol. 12, no. 1, pp. 155–174, 2021, doi: 10.1108/JHTT-01-2020-0015.

[33] X. Lao, X. Deng, H. Gu, J. Yang, H. Yu, and Z. Xu, ‘Comparing Intercity Mobility Patterns among Different Holidays in China: a Big Data Analysis’, *Appl. Spat. Anal. Policy*, 2022, doi: 10.1007/s12061-021-09433-z.

[34] P. Krataithong, C. Anutariya, and M. Buranarach, ‘A Taxi Trajectory and Social Media Data Management Platform for Tourist Behavior Analysis’, *Sustainability*, vol. 14, no. 8, 2022, doi: 10.3390/su14084677.

[35] X. Luo, J. Hu, L. C. Wan, and X. S. Yi, ‘Restriction reduction: The effects of mobility restriction on consumers’ preferences for advertisements of tourism products’, *Tourism Management*, vol. 94, 2023, doi: 10.1016/j.tourman.2022.104638.

[36] O. Tizouiar and A. Boussoualim, *For a Smarter and More Sustainable City: Tipasa with Its Potentialities*, vol. 361 LNNS. in Lecture Notes in Networks and Systems, vol. 361 LNNS. 2022, p. 11. doi: 10.1007/978-3-030-92038-8_1.

[37] P. Krataithong, C. Anutariya, and M. Buranarach, ‘A data management platform for taxi trajectory-based tourist behavior analysis’, presented at the ACM International Conference Proceeding Series, 2021, pp. 15–21. doi: 10.1145/3444757.3485104.

[38] L. Sala, S. Wright, C. Cottrill, and E. Flores-Sola, ‘Generating demand responsive bus routes from social network data analysis’, *Transportation Research Part C: Emerging Technologies*, vol. 128, 2021, doi: 10.1016/j.trc.2021.103194.

[39] A. Aguirre Montero and J. A. López-Sánchez, ‘Intersection of Data Science and Smart Destinations: A Systematic Review’, *Frontiers in Psychology*, vol. 12, 2021, doi: 10.3389/fpsyg.2021.712610.

[40] A. Cheshmehzangi, Y. Li, H. Li, S. Zhang, X. Huang, X. Chen, Z. Su, M. Sedrez, and A. Dawodu, ‘A hierarchical study for urban statistical indicators on the prevalence of COVID-19 in Chinese city clusters based on multiple linear regression (MLR) and polynomial best subset regression (PBSR) analysis’, *Sci. Rep.*, vol. 12, no. 1, 2022, doi: 10.1038/s41598-022-05859-8.

[41] P. Krataithong, C. Anutariya, and M. Buranarach, ‘A Taxi Trajectory and Social Media Data Management Platform for Tourist Behavior Analysis’, *Sustainability (Switzerland)*, vol. 14, no. 8, 2022, doi: 10.3390/su14084677.

[42] V. Maria and F. Mone, ‘Big data services based on mobile data and their strategic importance’, in *Int. Conf. Comput. Commun. Control, ICCCC - Proc.*, Filip F.G., Dzitac D., Dzitac I., Manolescu M.-J., Dzitac S., and Oros H., Eds., Institute of Electrical and Electronics Engineers Inc., 2018, pp. 276–281. doi: 10.1109/ICCCC.2018.8390471.

[43] M.-P. Kwan and T. Schwanen, ‘Geographies of Mobility’, *Annals of the American Association of Geographers*, vol. 106, no. 2, pp. 243–256, Mar. 2016, doi: 10.1080/24694452.2015.1123067.

[44] J. Xu, A. Li, D. Li, Y. Liu, Y. Du, T. Pei, T. Ma, and C. Zhou, ‘Difference of urban development in China from the perspective of passenger transport around Spring Festival’, *Applied Geography*, vol. 87, pp. 85–96, Oct. 2017, doi: 10.1016/j.apgeog.2017.07.014.

[45] J. Li, Q. Ye, X. Deng, Y. Liu, and Y. Liu, ‘Spatial-Temporal Analysis on Spring Festival Travel Rush in China Based on Multisource Big Data’, *Sustainability*, vol. 8, no. 11, Art. no. 11, Nov. 2016, doi: 10.3390/su8111184.

[46] P. Krataithong, C. Anutariya, and M. Buranarach, ‘A data management platform for taxi trajectory-based tourist behavior analysis’, in *ACM Int. Conf. Proc. Ser.*, Association for Computing Machinery, 2021, pp. 15–21. doi: 10.1145/3444757.3485104.

[47] P. Deville, C. Linard, S. Martin, M. Gilbert, F. R. Stevens, A. E. Gaughan, V. D. Blondel, and A. J. Tatem, 'Dynamic population mapping using mobile phone data', *Proc Natl Acad Sci U S A*, vol. 111, no. 45, pp. 15888–15893, Nov. 2014, doi: 10.1073/pnas.1408439111.
[48] B. P. Y. Loo, 'Transport geography: Towards a more people-oriented approach in the last 25 years', *Journal of Transport Geography*, vol. 81, p. 102596, Dec. 2019, doi: 10.1016/j.jtrangeo.2019.102596.
[49] X. Tang, D. Wang, Y. Sun, M. Chen, and E. O. D. Waygood, 'Choice behavior of tourism destination and travel mode: A case study of local residents in Hangzhou, China', *Journal of Transport Geography*, vol. 89, p. 102895, Dec. 2020, doi: 10.1016/j.jtrangeo.2020.102895.
[50] M. A. E. Kamel, 'Encouraging walkability in GCC cities: smart urban solutions', *Smart and Sustainable Built Environment*, vol. 2, no. 3, pp. 288–310, 2013, doi: 10.1108/SASBE-03-2013-0015.
[51] B. A. Mateus Francisco, 'Analysis of the Tourists behavior in Lisbon using Data from a Mobile Operator', ISCTE - University Institute of Lisbon, 2022.
[52] N. Hotz, 'What is CRISP DM?', *Data Science Process Alliance*, Sep. 10, 2018. https://www.datascience-pm.com/crisp-dm-2/ (accessed Jan. 27, 2023).