

Mining Tourists' Movement Patterns in a City

Luís B. Elvas^{1,2}, Miguel Nunes¹, José A. Afonso^{3,4}, Berit Irene Helgheim and Bruno Francisco¹

¹ Instituto Universitário de Lisboa (ISCTE-IUL), ISTAR, Lisboa, Portugal

² Department of Logistics, Molde University College, Molde, 6410, Norway

³ CMEMS – UMinho, University of Minho, 4800-058, Guimarães, Portugal

⁴ LABBELS –Associate Laboratory, Braga/Guimarães, Portugal

luis.elvas@iscte.pt

Abstract. Although tourists generate a large amount of data (known as "big data") when they visit cities, little is known about their spatial behavior. One of the most significant issues that has recently gained attention is mobile phone usage and user behavior tracking. A spatial and temporal data visualization approach was established with the purpose of finding tourists' footprints. This work provides a platform for combining multiple data sources into one and transforming information into knowledge. Using Python, we created a method to build visualization dashboards aiming to provide insights about tourists' movements and concentrations in a city using information from mobile operators. This approach can be replicated to other smart cities with data available. Weather and major events, for instance, have an impact on the movements of tourists. The outputs from this work provide useful information for tourism professionals to understand tourists' preferences and improve the visitors' experience. Management authorities may also use these outputs to increase security based on tourists' concentration and movements. A case study in Lisbon with 4 months data is presented, but the proposed approach can also be used in other cities based on data availability. Results from this case study demonstrate how tourists tend to gather around a set of parishes during a specific time of the day during the months under study, as well as how unusual circumstances, namely international events, impact their overall spatial behavior.

Keywords: tourist behaviour; location data; data analytics; mobile phone sensing; Internet of Things; smart cities.

1 Introduction

A network of actual, physical things that are linked to the Internet is known as the Internet of Things (IoT) [1]. These things are able to gather and exchange information about themselves and their environment. These data can be used to monitor and manage the objects, as well as to understand and improve the systems they are part of. The IoT has the potential to transform the way we live and work and make our lives more convenient and comfortable. The IoT can be used to collect big data, which can be used to understand trends and patterns. It can also be used to improve the efficiency of systems

and to make better decisions. The physical objects of the IoT can range from wireless sensor network [2] devices specifically designed for an application to more general objects such as vehicles [3] and mobile devices. The use of mobile devices to collect useful data is also called mobile phone sensing [4]. The data that may be collected using mobile devices includes location data, data from other internal sensors (e.g., accelerometers, microphone, etc.), and data from wearable devices (e.g., a Bluetooth Low Energy heart rate monitor).

In the context of smart cities, things have been done as exploring the application of data-driven approaches and visualization techniques in order to support decision makers in urban settings by providing valuable insights and tools for them. Research have been conducted in order to study areas like disaster management [5], [6], incident management [7], road accidents [8] and city-wide data analytics. The findings highlight the potential benefits of leveraging data and advanced analytics to enhance urban planning, enhance road safety, and develop effective strategies for disaster and incident management. Ultimately there's been a contribution to emphasizing the importance of data-driven approaches in empowering decision-makers to create more efficient, sustainable, and resilient urban environments.

This paper proposes an approach for understanding tourist mobility and its patterns, using mobile phone location data, which allows to analyse people's mobility quickly and accurately in a city. This information can be useful to decision makers of tourism infrastructure and marketing, as well as to negotiate with local businesses and tourism-related organizations to ensure that the economic benefits of tourism are shared fairly among all stakeholders. In addition, tracking tourist movement patterns can help cities to better manage their resources and plan for the future. By understanding how tourists are using the city, city planners can anticipate the demand for certain services and amenities, such as public transportation, accommodations, and visitor centres, and allocate resources accordingly. They can also identify areas of the city that may be experiencing overcrowding or congestion due to high levels of tourist traffic and develop strategies to mitigate the impact. Tracking tourist mobility can also help cities to better understand and address the environmental impact of tourism. By analysing location data, city planners can identify areas of the city where tourists are likely to generate the most waste or pollution and develop strategies to minimize these impacts.

The main contributions of this paper are: (1) provide platform that can merge multiple data sources into one, transforming information into knowledge; (2) provide a data visualization tool that enables all the stakeholders to interpret data in a more efficient way; (3) build a decision support system that allows users to make data-driven decisions based on a visualization component.

After a year of 2020 marked by a severe pandemic scenario in which Portugal registered a sharp drop in international tourism, there was a recovery in international tourism starting in the second half of 2021 [9]. This trend continued in the first two months of 2022, with a 769.2% increase in guests from abroad in February 2022 compared to February 2021.

To get a better understanding of how international tourists move in a city, we collaborated with the Lisbon City Council, who provided us with location data (roaming and non-roaming) from a mobile carrier company – Vodafone, in the city of Lisbon,

from September 2021 to January 2022. We also collected weather data from the Portuguese Sea and Atmosphere Institute (IPMA), a national meteorological organization, in the same timeline and in the same city. In this study, we considered people using mobile phones in roaming as possible tourists. Based on the raw data, this work analyses the spatial-temporal behaviour of Lisbon tourists with the following goals:

- Being aware of the key tourist attractions, restaurants, and accommodations;
- Finding the main concentration points and how they changed during the study period;
- Recognizing how variations in weather affect tourists' mobility patterns;
- Identifying peculiar mobility patterns in international events taking place in Portugal during the study's months.

The remainder of this paper is structured as follows: Literature Review is presented in Section 2. Section 3 introduce the methodology that we have used, as well as all the methods used in this work. Then, in Section 4 we present the results of our work and finally, Section 5 presents the conclusions.

2 Literature Review

In order to assess the relevancy of our study, we have conducted a literature review of the last 5 years, where we have used a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) [10] as our methodology for answering the following research question (RQ): What is the state of the art on tourists' behaviour analysis and tourism mobility analysis in smart cities?

The database used for the search was Scopus, and the study took place between May 8 and 12, 2022; all the findings had to be publications published within the past five years and written in English. Based on the data synthesis and analysis results mentioned above, a qualitative evaluation was conducted.

The Scopus database was thoroughly searched for published work on topic related with the concept "data analysis" or "behaviour analysis", the target population "smart cities" or "cellular network" or "tourist" or "roaming" and within the "mobility" context of the study. All keywords were searched using wildcards (such as "#" and "*"), which are used to match variations of the keywords, allowing us to perform a better search. The search query was built by intercepting all the columns, that is Concept AND Population AND Context, as depicted in Table 1. All the limitations have been applied. This query resulted in 44 documents, from which 16 papers were retrieved after a manual procedure was completed to determine major subjects on their research questions and define the outcomes. Year, region, RQ topic, and a brief description were all factors in our study systematization.

The trend line in Figure 1 shows that the issue we are examining is growing in popularity, demonstrating its importance.

Table 1. Search keywords and numbers of documents.

Concept	Population	Context	Limitations
data analysis	smart cities	mobility	2018-2022 Only journal papers, articles and reviews
behavior analysis	cellular network		
	touris*		
	roaming		
44 documents			

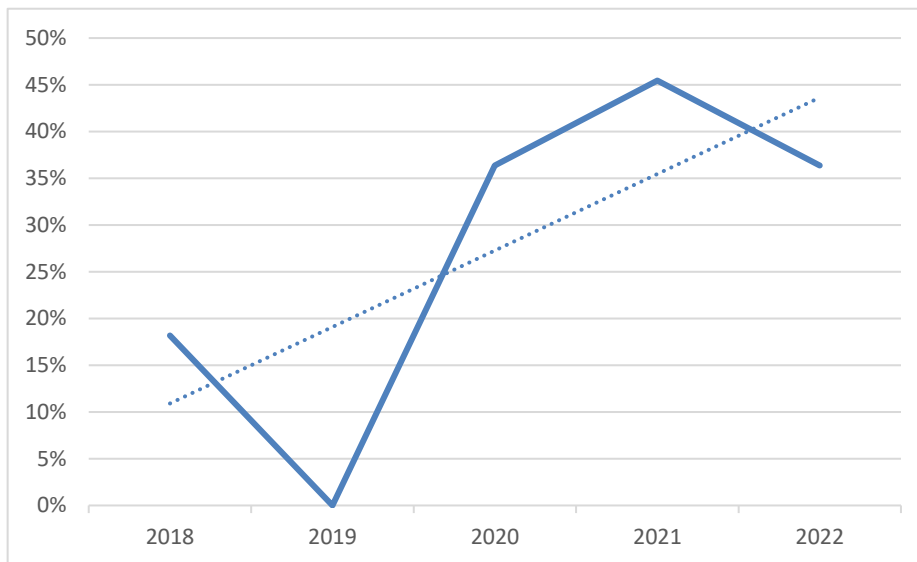
**Fig. 1.** Evolution of eligible studies by year.

Figure 2 presents the percentage of documents on each topic. This figure shows that most of the documents focus on behaviour analysis and/or used mobile phones / information and communications technology (ICT) infrastructures.

The research work presented in this paper is also based on both concepts, since we not only investigate people's behaviour using the communication infrastructure of a mobile phone operator in the city of Lisbon, but we also analyse it, in order to comprehend it and build a strategy to meet their demands.

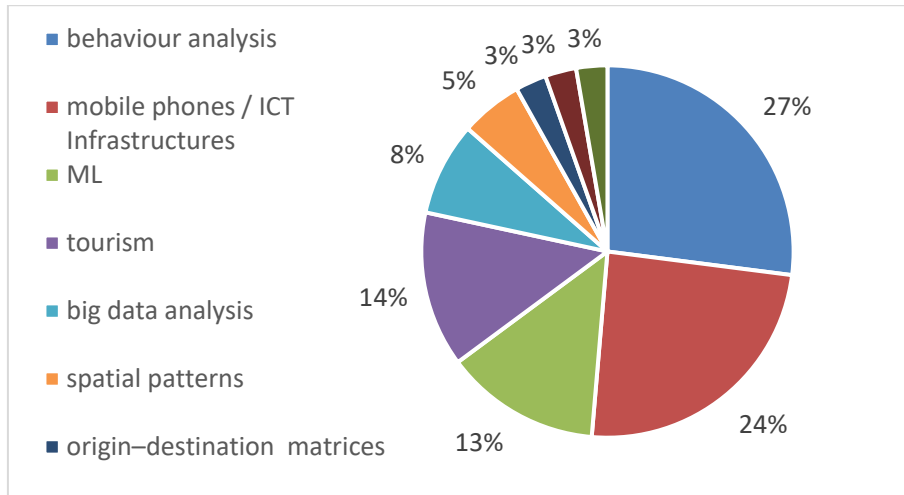


Fig. 2. Research topics on the state of the art

After reading all of the publications, it was clear that the amount of behavioural research on tourist mobility has expanded dramatically in recent years all over the world.

Authors from [11] present a method for estimating origin-destination (O-D) matrices using passively obtained cellular network signalling data from millions of anonymous mobile phone users in the Rhône-Alpes region of France, enhancing and revolutionizing the field of travel demand and traffic flow modelling.

Still on the subject, the authors of study [12] can identify pedestrian hotspots and provide future traffic signal and street layout information to make the city more pedestrian-friendly, as well as apply the knowledge gained to other data sets, such as bicycle traffic, to guide city infrastructure initiatives.

In a similar vein, but focusing on behaviour analysis, study [13] identifies a number of metrics for determining whether a person on the move is stationary, walking, or riding in a motorized private or public vehicle, with the goal of providing city users with personalized assistance messages for, among other things, sustainable mobility, health, and/or a better and more enjoyable life. This work was applied to the cities of Tuscany and Florence. The authors in [14] study and compare the density of users in Shanghai city using Weibo geolocation data and univariate and bivariate density estimation approaches, such as point density and kernel density estimation (KDE), where the main findings are based on characteristics of users' spatial behaviour, such as the centre of activity based on check-ins, and the feasibility of using check-in data to explain the relationship between users and their social media accounts. Continuing in this vein, the research in [15], based on long-term mobile phone data (from 2007 to 2012) of Beijing participants, gives a means to visualise individual mobility patterns.

The authors in [16] aim to provide a taxonomy of 5G Core Network (CN) mobility prediction frameworks, from data gathering to model provisioning, while taking into account the 3GPP architecture and interfaces. The authors provide two critical use cases in 5G CNs, in which the benefits of mobility predictions are assessed using information

from real networks, whereas study [17] focuses on building a mobile sequential recommendation system to assist auto services (e.g., taxi drivers).

On the subject of behavioural analysis, the authors in [18] present an urban travel behaviour model and assesses its feasibility for creating a greener, cleaner environment for future generations, whereas the study in [19], based on a trip survey from the São Paulo metropolitan area, one of the world's busiest traffic locations, supplements a current bundling approach to enable multi-attribute trail datasets for the visual study of urban mobility, aiding in the identification and analysis of urban mobility.

In terms of quality of life, the authors in [20] aim to investigate the structural equation model of smart city elements that influence global management of world heritage sites, as well as the quality of life for Thai tourists and inhabitants in Ayutthaya province.

Focusing on tourism and behaviour analysis, the authors in [21] use machine learning to determine the most relevant parameters influencing COVID-19 transmissions across different Chinese cities and clusters. The authors used a data-driven hierarchical modelling technique, with the “number of tourists” as one of the variables. Following the same line, the work in [22] has the goal of assessing the structure of tourist flows and examine the variables that impact their regional distribution. Similarly, the authors in [23] examine distinct intercity transportation patterns across different holidays and evaluate driving factors utilizing geographical and statistical analytic tools, in order to optimize city hierarchical structure and allocate transportation resources. The authors in [24] used machine learning and ICT to offer a position prediction system that takes into account both the spatial and temporal regularity of object movement. The object's historical trajectory data is utilized to derive personal trajectory patterns in order to determine possible future placements.

Using Airbnb data, the authors in [25] examine how visitors use local public transit in connection to the locations they visit. They employ a big data analysis of the factors influencing tourists' mobility behaviour and use of public transportation in various tourist sites to assess the attractiveness profile of 25 prominent tourist attractions throughout the world.

In [26], the authors present novel techniques for studying pedestrian mobility aspects over the whole road network using ICT and study the influence of visitor flows on the quality of life of locals and maintaining cultural resources. This study was performed in the city of Venice.

After reading all these publications, it was clear that the amount of behavioural research on tourist mobility has expanded dramatically in recent years all over the world. However, no study explores the use of mobile phone location data to create spatial or temporal graphs considering the correlation with rainy weather conditions and big events.

3 Methodology

The development of this study was conducted by the Cross Industry Standard Process for Data Mining (CRISP-DM) methodological approach, which is a common approach

for creating data mining projects to lower costs, boost dependability, and manageability, making the data mining process more effective [27]. However, for this study, given the data in question and our main goals, we opted for a different version of this methodology consisting of 4 phases, as shown in Figure 3: 1) Data Understanding; 2) Data Preparation; 3) Data Integration; and 4) Data Visualisation.

The data was provided by a mobile operator in Lisbon, worked in Python and visualised in Microsoft Power BI.



Fig. 3. Development methodology used in this paper.

3.1 Data Understanding

In order to fully understand the potential of our data and how we could optimize the added value of this study, we meticulously investigated each variable once we had collected the data for our study.

As stated previously, our main goal was understanding mobility patterns in tourists. To achieve that, the Lisbon City Council provided some data related to the mobility of people (roaming and non-roaming) in the city of Lisbon based on anonymized mobile phone data supplied by Vodafone, a British multinational telecommunications company. The data, which were collected from people who have Vodafone as a mobile carrier, was anonymized through the removal of all data related to personal information. To provide a more accurately reflect of the mobility of all the people who circulated in Lisbon between September 2021 and January 2022, Vodafone extrapolated the data, resulting in the dataset we have used (Vodafone dataset). This extrapolation was based on market share of the three operators available in Lisbon.

To get more accurate insights on the mobility of international tourists and how weather conditions change their mobility patterns, we reached out to the Portuguese Sea and Atmosphere Institute (IPMA) and requested data on the weather conditions in Lisbon between the months of September 2021 and February 2022. Therefore, there are two different datasets (Vodafone dataset and IPMA dataset), so they are addressed separately.

The dataset provided by Vodafone was divided into several files in CSV format separated by months. From this data, we selected 14 variables (see Table 2) and 126 443 863 records in total. As for the number of observations per month, Table 3.

Table 2. Metadata from used variables from the Vodafone dataset.

ID	Variable Name	Variable Description	Variable Type
-----------	----------------------	-----------------------------	----------------------

0	Grid_ID	Number of grid cells – There are 3743 squares of 200 by 200 meters in order to cover the metropolitan area of Lisbon	Nominal
1	Datetime	Time and date of occurrence	Datetime
2	C1	Number of distinct mobile phone terminals counted on each grid cell – Measured every 5 minutes	Numeric
3	C2	Number of distinct terminals in roaming counted on each grid cell during the 5 minute period – Measured every 5 minutes	Numeric
4	C3	Number of distinct terminals that remained in the grid cell counted at the end of each 5 minute period	Numeric
5	C4	Number of distinct terminals in roaming that remained in the grid cell counted at the end of each 5 minute period	Numeric
6	C5	Number of distinct terminals entering the grid cell	Numeric
7	C6	Terminals leaving the grid cell– These are the distinct terminals that left the grid cell; the calculation is made using the previous 5-minute interval as reference, also considering the crossings of the grid in the same interval	Numeric
8	C7	Number of entries of distinct terminals, in roaming, in the grid cell	Numeric
9	C8	Number of exits of distinct terminals, in roaming, in the grid cell	Numeric
10	C9	Total number of distinct terminals with active data connection in the grid cell – Measurement every 5 minutes	Numeric
11	C10	Total number of distinct terminals, in roaming, with active data connection in the grid cell – Measurement every 5 minutes	Numeric
12	C11	Number of voices calls originating from the grid cell	Numeric
13	C12	Number of devices that for 5 minutes enter the 11 street sections considered for analysis; for this purpose, a section of track is considered to be a route with a few km which begins outside and ends inside the Municipality of Lisbon.	Numeric

Table 3. Number of observations per month (Vodafone dataset).

Month	Observation Numbers
September 2021	17 233 318
October 2021	32 627 337
November 2021	21 619 292
December 2021	33 121 658
January 2021	33 344 624

3.2 Data Preparation

After acquiring a better understanding of the data, we moved to the second phase, which consists of four subphases: data selection, data cleaning, resource selection and data integration. Originally, the Vodafone dataset was spread over several files in CSV format, each month consisting of four to nine files with the same format. To handle the data more efficiently on personal computers and to proceed with preparation of the data, we decided to merge the CSV files by months, rather than compiling all the files provided into one file. Later, we did the same to the IPMA dataset, which was divided by months.

Vodafone Dataset

Location data was obtained based on the radio waves that are sent and received by the telecommunications base station. The mobile network operators immediately save this data in their log files or memory, in this case Vodafone. When a mobile phone is registered in a country but used in another one, its user can be seen as a potential tourist, and the corresponding information, such as the country of origin and location coordinates, are registered as mobile roaming.

In the data selection subphase, the information provided by Vodafone was aggregated over 3743 squares of 200 by 200 meters, with no values of less than 10 devices reported, and collected in 5-minute periods. The data becomes available after a processing period of approximately 45 minutes. This information is very important to study the mobility of tourists.

In the data cleaning subphase, we discarded missing values and removed duplicate rows.

In the resource selection subphase, one of the first things we did was selecting the variables which were not of interest to our goal and removing them from the dataset. Thus, the original dataset identified in Table 2, was reduced to only a few variables of interest for our objectives. Subsequently, we created a new dataset with only the mobility data of people in roaming (foreign tourists) from the Vodafone dataset. This was achieved by keeping only the variables related to people in roaming.

As the goal of our work focuses on the mobility of people, it was also important to have information about holidays and weekdays/weekends in our dataset. To do this we have built a function to identify which days there was a holiday. Only holidays on weekdays have been marked, since at the weekend they do not have much impact on mobility in general.

To check some events during the different times of the day, a column with distinct time intervals was also created.

We coupled a grid dataset to the Vodafone dataset to have information about the parish and latitudes and longitudes of each Grid_ID. From the merge of these datasets, it was possible to build new columns to facilitate posterior data analysis and visualisation in Microsoft Power BI. A column with zones of Lisbon was then created, in which the 24 parishes of Lisbon were grouped into 5 distinct zones according to the administrative reorganization of the parishes in 2012, namely [28]:

- North Zone (Green Zone) - Santa Clara, Lumiar, Carnide, São Domingos de Benfica, Benfica;
- Western Zone (Yellow Zone) - Alcântara, Ajuda, Belém;
- Center Zone (Orange Zone) - Campolide, Alvalade, Avenidas Novas, Santo António, Arroios, Areeiro;
- Historic Downtown Area (Purple Zone) - Campo de Ourique, Estrela, Misericórdia, Santa Maria Maior, São Vicente, Penha de França;
- Eastern Zone (Blue Zone) - Beato, Marvila, Olivais, Parque das Nações.

From the variable that represents the TOP10 of country of origin of equipment (mobile phone), by order of representativeness, a split was applied in order to separate the countries that belonged to that column by distinct columns: TOP1 to TOP10, and then decreasing from TOP10 to only TOP3 of representativeness of the countries in that location, so that instead of creating 10 additional columns, only 3 columns were created, making data manipulation more efficient.

To facilitate the visualisation of the mobility of tourists through the city of Lisbon, the countries were grouped by territorial zones, except the main nationalities of tourists in Portugal: Spain, France, Germany and United Kingdom [29]. Three new columns were then created (*zonasnacionalidade1*, *zonasnacionalidade2* and *zonasnacionalidade3*), which represent the TOP1, TOP2 and TOP3, respectively, of different nationality groups.

IPMA Dataset

IPMA is the national meteorological, seismic, sea and atmospheric organization of Portugal. It generates data about the weather conditions in real time. These data were then grouped into datasets to be studied and analysed by professionals qualified to know the weather and its variations, to aid in prediction models, e.g., time series. In this case, the IPMA data were used to complement the main Vodafone dataset, due to their importance in understanding the mobility pattern of tourists.

To clean this dataset, we discarded missing values, removed duplicate rows, and deleted all rows which had an observation with the number -99 (Error Code). After that, we moved to the resource selection subphase, where we eliminated the variables that were not of interest to our goal and the variables that had no variance. Thus, the original dataset was reduced to only a few variables of interest (see Table 4), the most relevant ones being humidity, wind intensity, accumulated precipitation, and temperature.

Table 4. Remaining variables (IPMA Dataset).

ID	Variable Name	Variable Description	Variable Type
0	Day	Day of observation	Datetime
1	Hour	Hour of observation	Numeric
2	Month	Month of observation	Numeric
3	Month_txt	Month of observation in text format	Nominal
4	Station	Station where the data was collected from	Nominal

5	Position	Coordinates of the Station	Coordinates
6	Humidity	Relative Humidity in %	Numeric
7	Humidity_txt	Categorized version of the variable Humidity	Categorical
8	WindIntensity	Wind intensity Index used by IPMA	Ordinal
9	WindIntensity_txt	Categorized version of the variable WindIntensity	Categorical
10	WindIntensityKm	Wind Intensity in km/h	Numeric
11	PrecAccumulated	Accumulated Precipitation in mm	Numeric
12	PrecAccumulated_txt	Categorized version of the variable PrecAccumulated	Categorical
13	Temperature	Temperature in °C	Numeric
14	Temperature_txt	Categorized version of the variable Temperature	Categorical

3.3 Data Integration

We have created new columns that were the categorical version of our numeric variables of interest (see Table 5), basing our categories in the metadata from IPMA and the measuring system of each variable, i.e., name and measurement unit, and created a text version of the month variable to aid in the visualisation.

Table 5. New categorical variables (IPMA dataset).

Variable Name	Categories	Scale
Humidity_txt	No humidity	0-5 (%)
	Low humidity	6-20 (%)
	Moderate humidity	21-60 (%)
	High humidity	61-100 (%)
WindIntensity_txt (Index used by IPMA from 0 to 12)	No wind	0-1
	Low wind	1-4
	Moderate Wind	4-7
	Intense wind	7-12
AccPrecipitation_txt	No rain	0-5 (%)
	Low rain	6-20 (%)
	Moderate rain	21-60 (%)
	Abundant rain	61-100 (%)
Temperature_txt	Very Cold	-5-0 (°C)
	Cold	1-20 (°C)
	Pleasant	21-28 (°C)
	Warm	29-40 (°C)

4 Results

In this stage, we represented the data in graphical form, as graphs and charts let us explore and learn more about it, providing a quick way to see trends and patterns in the mobility of the tourists and to focus on the most important points [30]. Good data visualisation also makes it easier to communicate our ideas and insights to other people [30]. With Power BI we represented each parish and used the data available to visualise the movement of tourists and to identify major patterns.

4.1 Overview

As shown in Figure 4, of the four months under review, October had the most tourists. Therefore, this month was selected for an analysis of the spatial distribution of tourists in the various periods of the day, as shown in the next figures.

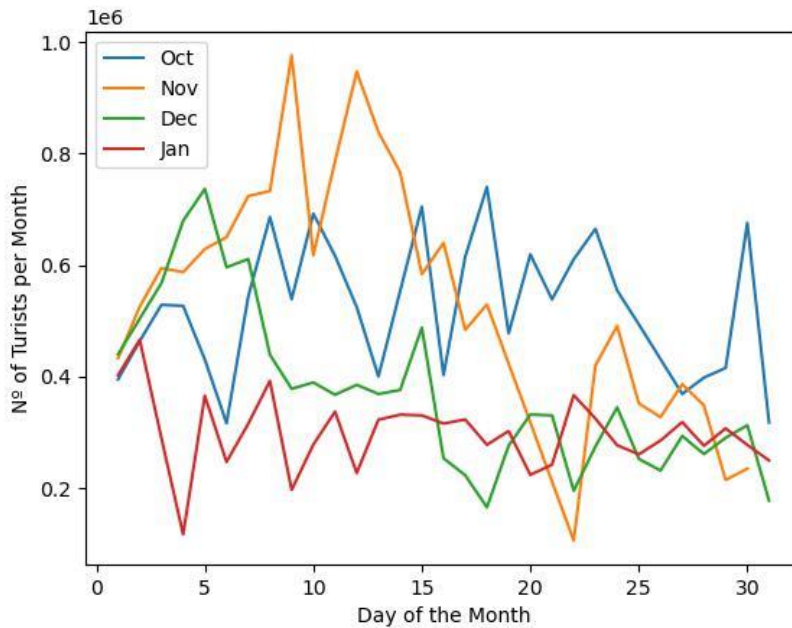


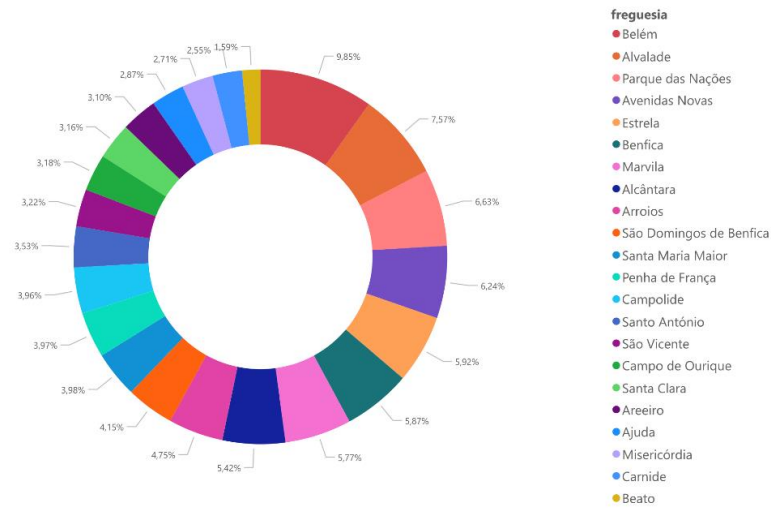
Fig. 4. Tourists in Lisbon by day, in millions, during the months of October 2021 to January 2022.

Figure 5 presents an analysis of the parishes frequented by tourists at lunchtime and afternoon. As shown in the figure, in the lunch hour period, the parishes that tourists frequent the most are Belém, Alvalade and Parque das Nações, whereas in the afternoon and evening periods, the most visited parishes are Belém, Alvalade and Avenidas Novas; finally, during the early morning hours the most frequented parishes are Avenidas

Novas, Belém and Arroios (see Figure 5 b)). This is aligned with the suggested places to eat available from tourist literature, such as TripAdvisor [31].

Figure 6 presents an analysis of the nationalities of the tourists that frequented the Lisbon parishes at two different periods: night and dawn.

a) % of Tourists by parishes of Lisbon at Lunchtime - October 2021



b) % of Tourists by parishes of Lisbon in the Afternoon - October 2021

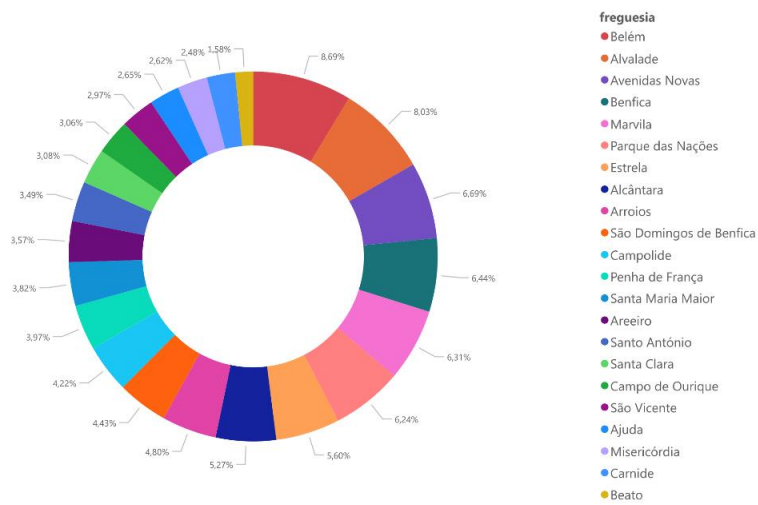


Fig. 5. Distribution of Lisbon tourists by parishes at lunchtime and afternoon in October 2021.

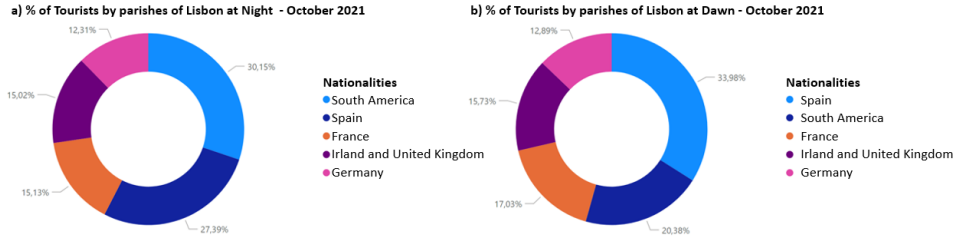


Fig. 6. Distribution of tourists' nationalities in Lisbon at night and dawn.

4.2 Case 1 – Mobility Patterns in Different Weather

One interesting case is the influence of tourists' movements based on weather conditions and the destinations they choose to visit when it is raining or sunny. When we looked at the tourist's nationalities, we discovered that during the months of October, November, and December the most common nationality in Lisbon was Spanish, followed by South American, French and Irish and English (see Figure 7). In January 2022 the nationalities remained with similar representations except for the South American nationality, which had the most representation in Lisbon with an increase of more than 6% over the previous month (see Figure 7, d)).

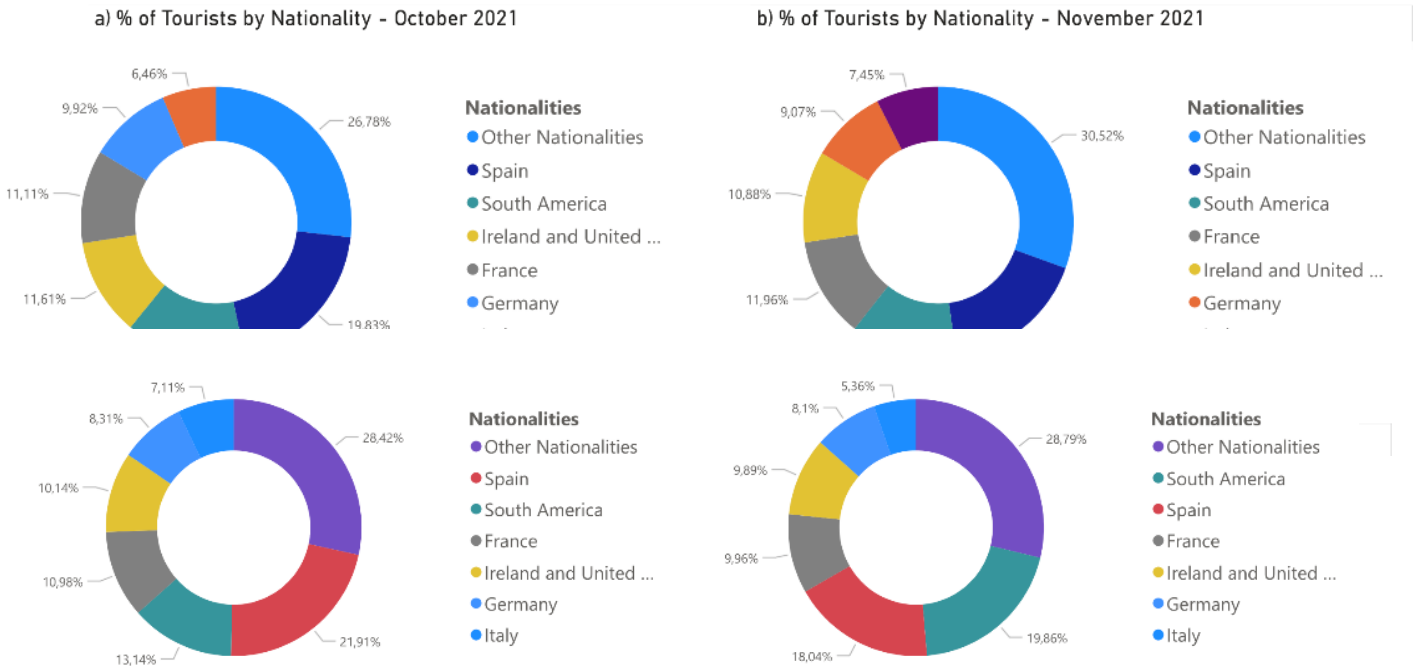


Fig. 7. Tourists by nationality in Lisbon in October – January 2022.

In Figure 8 we represented tourists' movements on sunny and rainy weeks. When it is raining, tourists tend to concentrate in certain places (the historic downtown area, the airport, near Carnide, where we found a lot of hostels and apartments for rent, and in the Eastern zone). Tourists tend to stick to monuments, their places of stay and shopping centres on rainy days (see Figure 8 a)). When it is sunny, rather than concentrating in some places, tourists tend to spread out more and visit more areas than just the “hotspots” (see Figure 8 b)). The effects of the weather on the mobility of tourists are clearly visible in these results.

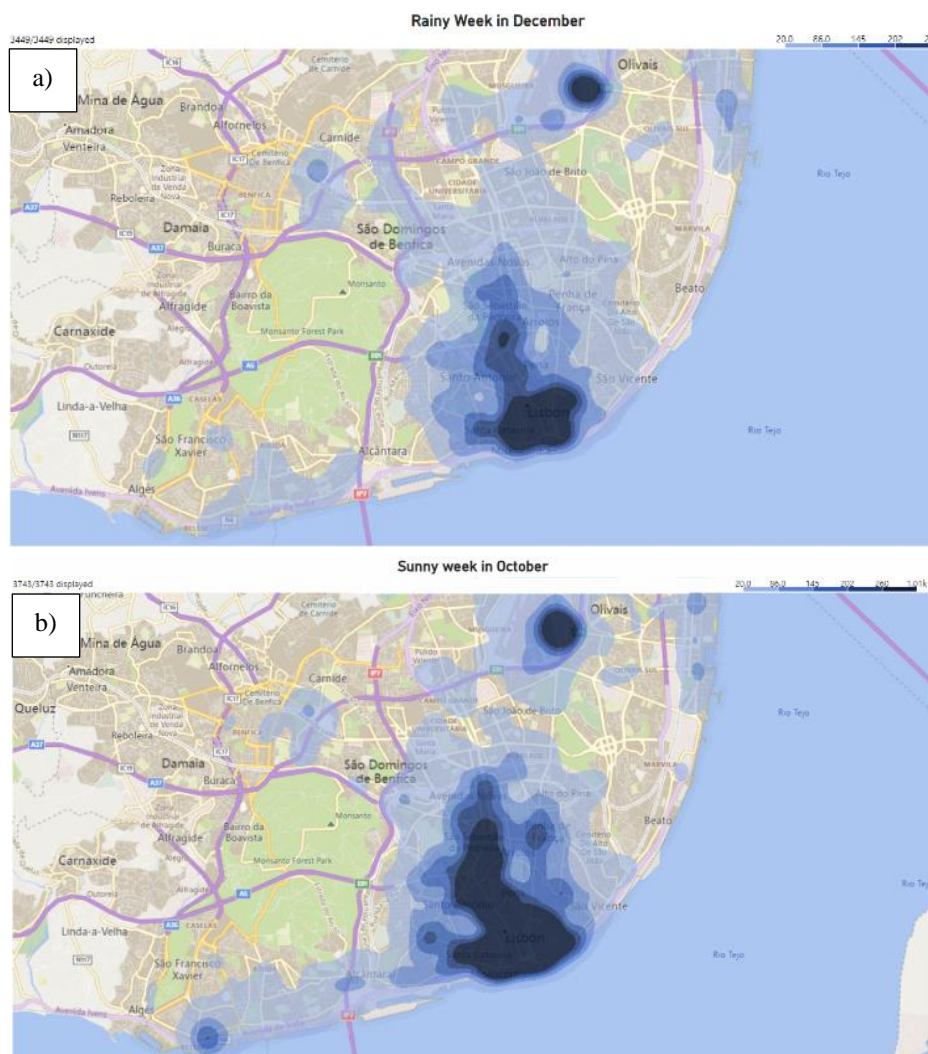


Fig. 8. Tourists' mobility on a rainy week in December (a)) vs a sunny week in October (b)). These heat maps about tourists' location go from deep blue (high concentration of tourists) to light blue (low concentration).

4.3 Case 2 – The Effect of Big Events: Web Summit

The Web Summit is an annual technology conference held in Lisbon, Portugal, and the biggest in Europe [32]. In 2021, while still under the effects of the COVID-19 pandemic, the event gathered over 40,000 visitors. This influx of people allowed us to visualise the impact that the event had in the area where it was held (FIL – Lisbon International Fair).

To analyse this, we mapped the movement of roaming users and considered those who were staying in the same grid for more than 5 minutes when the event was taking place, putting those results side by side with the remaining days of the month. On the Web Summit period (1st – 4th November 2021), we were able to register an average of 2852 tourists at any given time in the event area, compared to the 262 for the rest of the month of November (5th – 30th November 2021). It's a 987% increase in tourist activity, Figure 9.

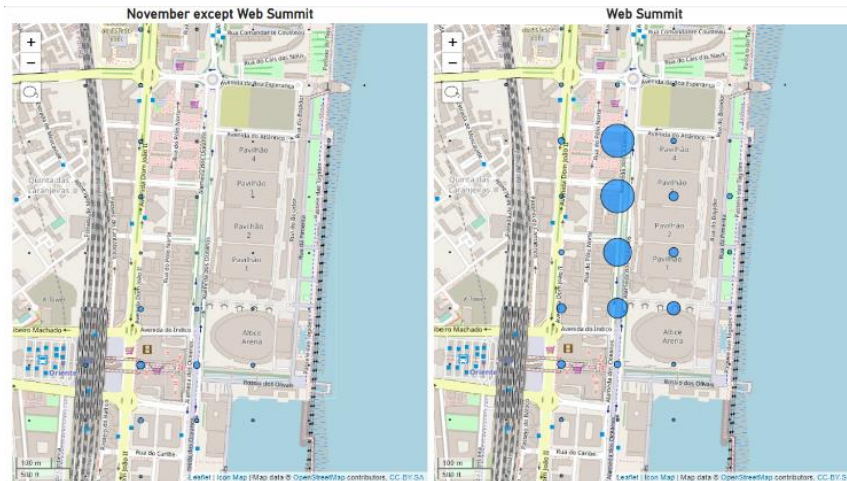


Fig. 9. Comparison of tourist distribution outside of Web Summit (left) and during the Web Summit (right).

4.4 Case 3 – Big Events Monitor Process – Case of a Football Game Day

Football tourists are those who travel to attend a football event, usually a game. The proposed method enables for real-time surveillance of tourist movements around the city. Since the UEFA Champions League games create big movements of people from around the world, we tracked 4 major games in Lisbon.

From the four games (UEFA Champion League in Lisbon, Benfica-Barcelona, Benfica-Bayern Munich, Sporting-Besiktas and Sporting-Dortmund) we monitored, we

chose the one that brought more tourists from a less common nationality (Turkey) to Lisbon during the course of the study. This aspect allowed us to monitor the football fans' movements a lot more effectively and clearly than if we had a considerable number of tourists just visiting the city in general.

As can be seen in Figure 10 a), we were able to identify the time of arrival of the football fans at Lisbon airport at 9 am on the day of the match (marked in the figure).

During the day, these tourists went mainly to the historic downtown area, arriving there as 2 pm, as shown in Figure 10 b). Then at the game time, we identify around 5000 tourists from Turkey at the Sporting stadium, as shown in Figure 10 c). After the game, since there are no flights, they returned to the historic downtown area (Figure 10 d)), and in the next day, around 9 am, they went back to the airport to depart from Lisbon (Figure 10 e)). This case highlights the importance of this analysis for parishes to understand tourists' movements and to better manage big events.



Fig. 10. Turkish tourist's movements during the football match in Lisbon, a) to d) represents the match day and e) the day after when they left the country.

Also for authorities can be useful to monitor football away fans activity to prevent confrontations between them and the local fans. Unfortunately a common event.

4.5 Case 4 – The Effect of Events – Shopping Activity

Shopping activity can be monitored using the communication antennas that cover a shopping mall. The number of tourist visitors is useful information for authorities and store owners. Visit patterns and abnormal behaviour during promotion days (e.g., during Black Friday) are examples of useful information that can be extracted from the data. We can check where tourists shop more and see if the promotions influence their behaviour. The data collected shows this influence, and it is possible to witness this behaviour based on the nationality. Figure 11 and Figure 12 show this effect and the main places where tourists' shop. Promotions increase concentration in shopping areas, even for tourists.

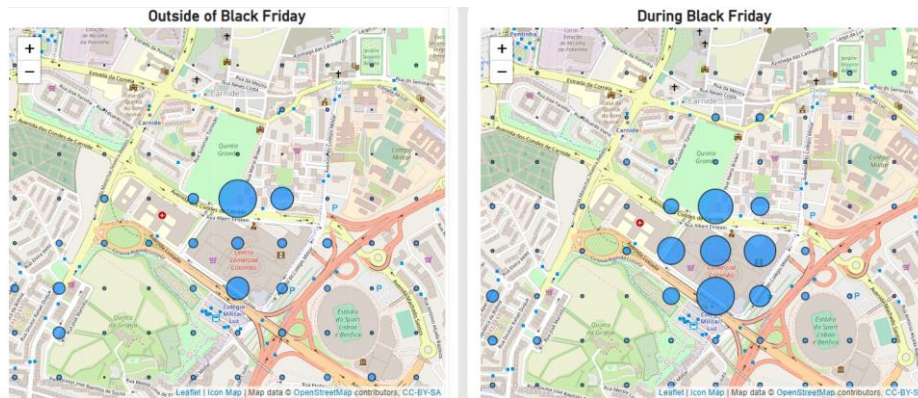


Fig. 11. Effect of Black Friday in Colombo Mall.

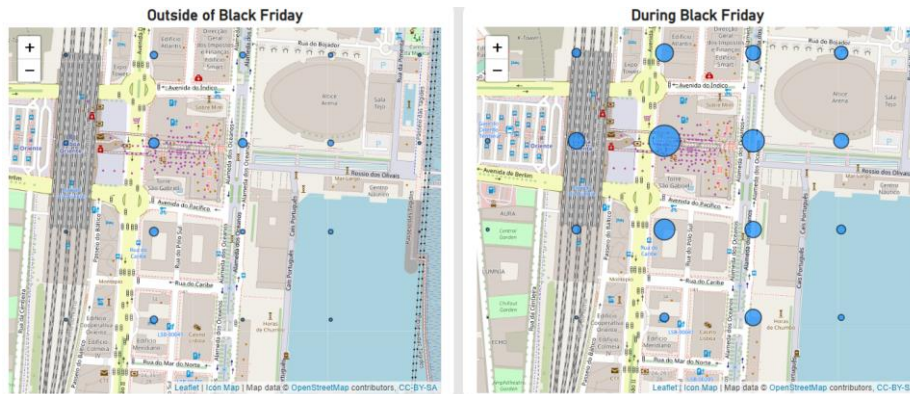


Fig. 12. Effect of Black Friday in Vasco da Gama Mall.

5 Conclusions

This work presented an analysis of Lisbon tourists' mobility patterns, allowing decision makers to perform better and more efficient decisions considering the dashboards. Mobile operators' availability on the market enables the representation of a city's population. The approach described in this paper can be replicated to other cities using similar mobile phone location data. From the provided data visualisation, it is possible to see that tourists visit the central, historic downtown district and travel to the interesting coastal regions.

Our strategy began with data cleaning and processing created in Python, followed by data visualisation created in Microsoft Power BI in the form of maps and dashboards. It is possible to replicate this in other locations, and it is a useful tool for city management authorities to comprehend tourist concentrations and movements and modify facilities and processes.

The research presented in this paper provides comprehension of tourist patterns and movements as they relate to events, weather, and nationality. Results demonstrate that tourists are deterred from exploring the city attractions by the rain. They frequently stay in "hotspots" like the city's centre and historic downtown, where they may visit attractions like museums and shopping centres or simply relax in the comfort of their accommodations. When it's sunny, tourists behave differently, spreading out more throughout the city and stopping in more places. Major international events draw a lot of tourists to the host city and particularly the neighbourhood.

Public policy must consider the spatial spread of urban tourism. In locations with a high concentration of tourists, local authorities should take measures to improve the tourist experience, such as setting up new tourist information centres, constructing pedestrian-only lanes, extending walkways, or expanding the availability of public spaces with free Wi-Fi hotspots. The real time insight supplied by current developed work may enable local authorities to understand the peoples' movements to improve both the visitors and the local inhabitants experience, as well as to generate security alerts in case of harmful concentrations in major events.

Another critical topic is the modelling of tourist behaviour, which might be done in future research with a time-series analysis. This research can also be improved by including an anonymous ID in the data to track actual human mobility patterns, allowing researchers to determine the most frequently utilized city entrances and exits, design decongestion tactics, and understand the affluence of people at bus terminals. This study could be improved by having more months of data to compare tourist mobility and patterns by season.

Author Contributions

C.C. conducted all data mining processes and the system development; L.B.E. wrote the state of the art; C.C. performed all the management interface and contributed to the writing of the article; A.L.M contributed with the writing and review; J.C.F. coordinated the research and contributed to the article’s writing; J.A.A contributed with the writing on review and editing. All authors have read and agreed to the published version of the manuscript.

Acknowledgment

We thank the Lisbon City Council for providing us with the data necessary for this study, namely Mr. António Costa (Lisbon City Council), Mrs. Helena Martins (Lisbon City Council) and Mrs. Paula Melicias (Lisbon City Council). We also thank IPMA for providing us with the data necessary to complement our study. We also thank to FCT—Fundação para a Ciência e Tecnologia, I.P., under the project FCT UIDB/04466/2020 and UIDP/04466/2020. Luís Elvas holds a Ph.D. grant, funded by FCT with Ref. UI/BD/151494/202.

Funding

This work was supported by EEA Grants Blue Growth Programme (Call #5). Project PT-INNOVATION-0069 – Fish2Fork.

References

1. Author, F.: Article title. *Journal* 2(5), 99–110 (2016).
 2. Author, F., Author, S.: Title of a proceedings paper. In: Editor, F., Editor, S. (eds.) CONFERENCE 2016, LNCS, vol. 9999, pp. 1–13. Springer, Heidelberg (2016).
 3. Author, F., Author, S., Author, T.: Book title. 2nd edn. Publisher, Location (1999).
 4. Author, F.: Contribution title. In: 9th International Proceedings on Proceedings, pp. 1–2. Publisher, Location (2010).
- LNCS Homepage, <http://www.springer.com/lncs>, last accessed 2016/11/21.
- [1] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, “Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications,” *IEEE Commun. Surv. Tutor.*, vol. 17, no. 4, pp. 2347–2376, 2015, doi: 10.1109/COMST.2015.2444095.

- [2] S. R. Jino Ramson and D. J. Moni, “Applications of wireless sensor networks — A survey,” in *2017 International Conference on Innovations in Electrical, Electronics, Instrumentation and Media Technology (ICEEIMT)*, Feb. 2017, pp. 325–329. doi: 10.1109/ICEEIMT.2017.8116858.
- [3] M. Gerla and L. Kleinrock, “Vehicular networks and the future of the mobile internet,” *Comput. Netw.*, vol. 55, no. 2, pp. 457–469, Feb. 2011, doi: 10.1016/j.comnet.2010.10.015.
- [4] W. Z. Khan, Y. Xiang, M. Y. Aalsalem, and Q. Arshad, “Mobile Phone Sensing Systems: A Survey,” *IEEE Commun. Surv. Tutor.*, vol. 15, no. 1, pp. 402–427, 2013, doi: 10.1109/SURV.2012.031412.00077.
- [5] L. B. Elvas, S. P. Gonçalves, J. C. Ferreira, and A. Madureira, “Data Fusion and Visualization towards City Disaster Management: Lisbon Case Study,” *EAI Endorsed Trans. Smart Cities*, vol. 6, no. 18, pp. e3–e3, Jun. 2022, doi: 10.4108/eetsc.v6i18.1374.
- [6] L. B. Elvas, B. M. Mataloto, A. L. Martins, and J. C. Ferreira, “Disaster Management in Smart Cities,” *Smart Cities*, vol. 4, no. 2, Art. no. 2, Jun. 2021, doi: 10.3390/smartcities4020042.
- [7] L. B. Elvas, C. F. Marreiros, J. M. Dinis, M. C. Pereira, A. L. Martins, and J. C. Ferreira, “Data-driven approach for incident management in a smart city,” *Appl. Sci. Switz.*, vol. 10, no. 22, pp. 1–18, 2020, doi: 10.3390/app10228281.
- [8] J. Mesquitela, L. B. Elvas, J. C. Ferreira, and L. Nunes, “Data Analytics Process over Road Accidents Data—A Case Study of Lisbon City,” *ISPRS Int. J. Geo-Inf.*, vol. 11, no. 2, 2022, doi: 10.3390/ijgi11020143.
- [9] LCG, “Toursim numbers in Portugal | March 2022.” <https://travelbi.turismodeportugal.pt/turismo-em-portugal/turismo-em-numeros-marco-2022/> (accessed May 13, 2022).
- [10] 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.
- [11] M. Fekih, T. Bellemans, Z. Smoreda, P. Bonnel, A. Furno, and S. Galland, “A data-driven approach for origin–destination matrix construction from cellular network signalling data: a case study of Lyon region (France),” *Transportation*, vol. 48, no. 4, pp. 1671–1702, 2021, doi: 10.1007/s11116-020-10108-w.
- [12] E. Carter, P. Adam, D. Tsakis, S. Shaw, R. Watson, and P. Ryan, “Enhancing pedestrian mobility in Smart Cities using Big Data,” *J. Manag. Anal.*, vol. 7, no. 2, pp. 173–188, 2020, doi: 10.1080/23270012.2020.1741039.
- [13] C. Badii, A. Difino, P. Nesi, I. Paoli, and M. Paolucci, “Classification of users’ transportation modalities from mobiles in real operating conditions,” *Multimed. Tools Appl.*, vol. 81, no. 1, pp. 115–140, 2022, doi: 10.1007/s11042-021-10993-y.
- [14] S. A. Haidery, H. Ullah, N. Ullah Khan, K. Fatima, S. Shahla Rizvi, and S. J. Kwon, “Role of big data in the development of smart city by analyzing the density of residents in shanghai,” *Electron. Switz.*, vol. 9, no. 5, 2020, doi: 10.3390/electronics9050837.
- [15] C. Li, J. Hu, Z. Dai, Z. Fan, and Z. Wu, “Understanding individual mobility pattern and portrait depiction based on mobile phone data,” *ISPRS Int. J. Geo-Inf.*, vol. 9, no. 11, 2020, doi: 10.3390/ijgi9110666.
- [16] J. Jeong *et al.*, “Mobility Prediction for 5G Core Networks,” *IEEE Commun. Stand. Mag.*, vol. 5, no. 1, pp. 56–61, 2021, doi: 10.1109/MCOMSTD.001.2000046.

- [17] P. Guo, K. Xiao, Z. Ye, and W. Zhu, "Route Optimization via Environment-Aware Deep Network and Reinforcement Learning," *ACM Trans. Intell. Syst. Technol.*, vol. 12, no. 6, 2021, doi: 10.1145/3461645.
- [18] N. X. Leow and J. Krishnaswamy, "Smart cities need environmental consciousness and more social responsibilities as an outcome of COVID-19 – reflections from urban road commuters," *Foresight*, vol. 24, no. 2, pp. 276–296, 2022, doi: 10.1108/FS-02-2021-0035.
- [19] T. G. Martins, N. Lago, E. F. Z. Santana, A. Telea, F. Kon, and H. A. de Souza, "Using bundling to visualize multivariate urban mobility structure patterns in the São Paulo Metropolitan Area," *J. Internet Serv. Appl.*, vol. 12, no. 1, 2021, doi: 10.1186/s13174-021-00136-9.
- [20] 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," *Humanit. Arts Soc. Sci. Stud.*, vol. 21, no. 1, pp. 35–42, 2021, doi: 10.14456/hass.2021.4.
- [21] A. Cheshmehzangi *et al.*, "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.
- [22] 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.
- [23] 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.
- [24] X. Li *et al.*, "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.
- [25] U. Türk, J. Östh, K. Kourtit, and P. Nijkamp, "The path of least resistance explaining tourist mobility patterns in destination areas using Airbnb data," *J. Transp. Geogr.*, vol. 94, 2021, doi: 10.1016/j.jtrangeo.2021.103130.
- [26] C. Mizzi *et al.*, "Unraveling pedestrian mobility on a road network using ICTs data during great tourist events," *EPJ Data Sci.*, vol. 7, no. 1, 2018, doi: 10.1140/epjds/s13688-018-0168-2.
- [27] R. Wirth and J. Hipp, "CRISP-DM: Towards a Standard Process Model for Data Mining," p. 11.
- [28] "Lisbon Administrative Reform - First Quarterly Monitoring Report." <https://www.amlisboa.pt/451600/1/008910,000505/index.htm> (accessed Oct. 27, 2022).
- [29] INE, "Tourism Statistics - 2020." https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_publicacoes&PUBLICACOESpub_boui=280866098&PUBLICACOESmodo=2 (accessed Oct. 21, 2022).
- [30] K. Healy, *Data Visualization: A Practical Introduction*. Princeton University Press, 2018.
- [31] "Tripadvisor," *Tripadvisor*. <https://www.tripadvisor.pt/> (accessed Aug. 09, 2022).
- [32] "Web Summit: Lisbon hosting the largest tech event in the world - Portugal - Portuguese American Journal." <https://portuguese-american-journal.com/web-summit-lisbon-hosting-the-largest-tech-event-in-the-world-portugal/> (accessed May 13, 2022).