Analysis of the Tourist's Behavior in Lisbon Using Data from a Mobile Operator

Bruno Francisco[1], Ricardo Ribeiro[1,2], Fernando Batista[1,2], and João Ferreira[3,4](✉)

[1] Instituto Universitário de Lisboa (ISCTE-IUL), 1649-026 Lisbon, Portugal
{bamfo1,ricardo.ribeiro,fernando.batista}@iscte-iul.pt
[2] INESC-ID Lisboa, 1000-029 Lisbon, Portugal
[3] Instituto Universitário de Lisboa (ISCTE-IUL), ISTAR, 1649-026 Lisbon, Portugal
joao.carlos.ferreira@iscte-iul.pt
[4] Inov Inesc Inovação – Instituto de Novas Tecnologias, 1000-029 Lisbon, Portugal

Abstract. This paper aims to provide to all entities involved in Lisbon tourism activities a geospatial, statistical, and longitudinal analysis tool based on data provided by a mobile operator in cooperation with Lisbon City council, which allows obtaining knowledge about the behaviors and habits of tourists and visitors of the city. The main intention is to provide information that allows decision-makers to base their choices on real data and facts instead of empirical knowledge and non-sustained information The work was mainly developed in three distinct phases. On the first phase, it was necessary to create knowledge about the tourism business and understand the available data to understand whether they would be able to answer our questions. In the next phase, the dataset was prepared and adapted to our needs - the data given to us had information regarding both mobile phones belonging to Portuguese and foreign users. Considering that our focus was on second group, part of the information was discarded.

Through the work developed, it was possible to identify which countries and geographical areas come from Lisbon's tourists and visitors. Additionally, we were able to identify, through the available data, the most visited places, and parishes in the city, as well as the place where they eat and sleep when they are in the city. It was also possible to characterize how events such as the Web Summit or a football game influence the behavior and movements of visitors in Lisbon.

The analyses and information provided were duly validated by specialists from the Lisbon Municipal Council, through presentations and questionnaires to decision-makers and users of the developed solution.

Keywords: Tourism · Lisbon · Travel Behavior · Smart Mobility · Transportation Networks · Big Data · Data Analytics · Mobile Networks · Data driven

1 Introduction

According to [1], the Portuguese capital receives about 4.5 Million tourists every year. Considering that Lisbon has approximately 504 thousand residents, the city receives

A. L. Martins et al. (Eds.): INTSYS 2022, LNICST 486, pp. 3–22, 2023.
https://doi.org/10.1007/978-3-031-30855-0_1

around 9 tourists per resident, the same is to say that it receives nine times the fixed population. As a term of comparison, cities like Prague, Barcelona and London receive between 4 and 5 tourists per resident. Looking into this ratio and considering that when compared with any of the referred cities, Lisbon is much smaller, it is easy to understand that all the stakeholders need to have a deep knowledge of the behaviors and movements along the city, to provide the tourists, the best possible experience.

Considering this number of visitors, the empirical knowledge is not enough to manage and define the strategy for hotels, local accommodations, stores, restaurants, transports, museums, security and all the areas we can remember when we think about tourism. By providing to stakeholders like City Council and the entities responsible for tourism with the necessary information, we can have a very positive impact on the decision-making process, mitigating the risk of incorrect actions that can lead to an unpleasant stay in Lisbon and to the decrease of the financial income. The best ambassadors that Lisbon can have abroad are the previous visitors.

Cities are complex environments and there is a huge number of challenges that need to be addressed to provide everyone a better experience in a city like Lisbon.

Addressing these challenges assumes a decisive role, especially now that tourist activity begins to recover after the pandemic period. As of the second half of 2021, the recovery started, and in May (last data available), there was a recovery of around 162% [2], compared to the same month of 2021. Even so, we continue with a negative variation, compared to May 2019.

Considering the Big Data generated and available these days, it is perfectly unthinkable that this digital asset is not used and that is exactly what we intend with our work - use Big Data to respond to our challenges.

This work was developed in partnership with the Lisbon City Council [3], more specifically with the Lx Data Lab [4], which provided us with the data obtained through an established contract with a Mobile Operator.

Considering that it was not possible through the literature review to find the information available in previous works, with this work, we intend to create knowledge in relation to: 1) Where do the Top visitors of Lisbon come from? 2) What are the most visited areas of Lisbon? 3) Where are the Tourists during mealtimes and where they sleep?; and 4) Event Analysis. To achieve this, we apply a data science approach with CRISP-DM [17] using past data of mobile operators, where we use cellular grid areas with information about tourists' nationality (number only due to GRDP rules), and time stamps in periods of five minutes.

According to the General Secretariat for the Economy [5], the weight of Tourism on the Portuguese Gross Domestic Product is around 19%, being the 5th country in the world where the contribution of Tourism has the greatest weight. This fact is particularly relevant given the number of people that this sector employs in its various aspects, so any negative variations in this indicator have an extremely harmful influence, not only in economic terms, but also in social terms. Unfortunately, it was not possible to confirm from any source what percentage of financial income derived from Tourism is generated in Lisbon. Of course, we cannot just and only focus on the financial part, which is not always in the best interests of the "Customers", that, in our case, are the Tourists. It is important that whenever you visit Lisbon, you can be sure that you will find a safe city, properly sanitized, with a good transport network, enough accommodation and framed with the most visited places and events of interest that can properly complete all points of interest, such as monuments, gastronomy, climate and so on. We believe that from the moment we provide decision-makers with the data, they will be able to create a transport network that meets the demand of tourists, they will be able to better train police authorities and all professionals working in Tourism.

Given this scenario, it turns out to be simple to understand that the motivation to carry out this research work lies in the possibility of carrying out an academic work that can have practical applicability and with a positive impact on the economy and, consequently, on the lives of all those who depend on tourism. The main intention is to avoid mistakes in decision-making by stakeholders.

The reminder of this document demonstrates, mostly through visualizations, that it is fundamental to to better understand how do tourists "behave" in the city of Lisbon. The focus of the analysis will be on the evolution over the five months of data we have available, considering the number of people and origin, with particular attention to the most represented countries and continents. Attention will also be paid to the places where tourists spend the most time and where the largest numbers of visitors are found, where they sleep and where they are at mealtimes (for the time being, our datasets do not make it possible to specify the commercial establishments). Finally, the influence of events in Lisbon will be highlighted, in relation to volumes and origins by comparison with a baseline that will always be the same period of the previous or subsequent week, to understand how events such as the Web Summit (https://websummit.com/) or the games Football League of Champions League (https://www.uefa.com/uefachampionsleague/) change the usual panorama of Tourism in Lisbon.

2 Related Work

The high rate of use of mobile phones combined with mobility makes the data generated through the signaling exchanged between the terminal and the network a tool through which analyzes can be carried out that make it possible to identify patterns and behaviors related to mobility. According to the GSM Association [19] there were 460 million mobile subscribers in Europe in 2021, covering around 86% of the population; according to data provided by ANACOM [20], in Portugal there are around 13 million active sim cards. Therefore, this information can effectively be one of the best probes that exist for the analyses. It should also be noted that since roaming within the European Union started to have costs like those in the country of origin, people began to enjoy their mobile services much more when they are visiting another European country, which now allows them to have a sufficient volume of data and potential also to carry out research work related to tourism using this information. Therefore, the use of such a kind of data has been applied to study the way how people move in the cities.

Mariem Fekih et al. [5], 2021, explored and made use of data generated through signaling exchanged between a mobile operator in order to create Origin-Destination matrices, with the main objective of assessing whether the amount of data generated through signaling can or cannot be a reliable source of analysis of the commuting movements of individuals, proposing a system capable of transforming the data generated by the mobile network into flows that allow typifying the Origin-Destination, having these validated through inquiries made by the local authority responsible for transport. This study was conducted in Rhône in the French Alps and the data used were provided by the mobile operator Orange. Through this work, the authors showed that these data can be used to estimate the pendulum movements, having been possible to prove through the questionnaires that the conclusions are valid. It was also possible to prove that this method can be automated and that after a few days, it may be possible to typify these movements successfully, being possible to avoid the constant questionnaires. Thus, the group of researchers proved that the use of this information allows the means of transport to be optimized, benefiting users through optimizations that allow a higher quality of service.

In the article "Enhancing pedestrian mobility in Smart Cities using Big Data", Ebony Carter et al. [6], 2020, proposes the use of different datasets generated through sensors installed on the Internet of Things network of the city of Melbourne to improve accessibility and the sustainability of the Municipality. The datasets used for the study include diverse information, for example: about parking, mobility, departures and arrivals at the airport and pedestrian traffic, having been estimated that in a period of 24 hours, data of around 650 thousand people. The results and analyzes were produced through heatmaps and various graphics, which allowed interpreting and contextualizing the analyses. Through the work carried out, it was possible to characterize pedestrian movements in Melbourne, and it was proposed to City officials that they continue to develop the sensor network and Internet of Things, since it is an essential source for the continued development of knowledge. Necessary information on pedestrian movements to improve sustainability and accessibility.

Continuing the study of the work carried out using the data generated by the signaling generated by cellular networks, Claudio Badii et al. [7], 2021, developed several metrics that allow us to perceive whether a given individual is or is not in mobility and if he is in motion, how you are doing it (on foot, by bicycle or in your own or public means of motorized transport), with the objective of sending each person personalized messages that can raise awareness of issues related to sustainable mobility and healthy living habits. To achieve their goals, the authors created a multi-class classifier that proved to be more accurate than resorting to a hierarchical approach and able to handle and manage data in real time. The developed solution was implemented in Antwerp and Helsinki.

Data generated through social networks are also an important source of information for the topic of mobility and behaviors in cities, Saqib Ali Haidery et al. [8], 2020 made use of data from Weibo, a Chinese social network in the sense of analyze and typify the number and density of Weibo users in the city of Shanghai using estimation techniques with one or more variables. With this work, it was possible to use the referred data to conduct different vectors of analysis: points of concentration of people and from their location develop personalized recommendations and typify the different volumes in the 10 districts of Shanghai. From this information, it is possible to develop disaster mitigation plans as well as manage security and emergency resources.

Chiara Mizzi et al. [9], 2018, in the article "Unraveling pedestrian mobility on a road network using ICTs data during major tourist events", also used the data provided by the Italian mobile operator TIM to study the characteristics of pedestrian mobility on the road network, using the City of Venice as an example to study the impact of tourists on the lives of local citizens as well as preserving the city's cultural heritage. After having worked and transformed the data, they developed an algorithm capable of reconstructing pedestrian movements through the streets of Venice and from there, they were able to distinguish mobility patterns between tourists and locals. Additionally, it allowed stakeholders to be given important and relevant information for decision-making based on data and not on empirical knowledge.

Entering in the field of machine learning, Jaeseong Jeong et al. [10], 2021, presents a model in which people's mobility predictions are made from the traffic seen from the core side and the 5G radio. With this approach, it is possible to analyze mobility in real time and with data being generated in real time. The contribution of this work focused on the concept of NWDAF (Network data analytics function), having proposed an approach a predictive model capable of adapting the 5G network to each user and their needs, allowing to give the user a better experience regarding internet connection speeds and the reduction of latencies.

Continuing the study of the state of the art, the work developed by PENGZHAN GUO et al. [11], 2021 is also relevant, who in their article "Route Optimization via Environment-Aware Deep Network and Reinforcement Learning" studied and developed an adaptive system that allows optimizing taxi services, proposing a deep learning system capable of optimizing routes on which these vehicles provide service, identifying the "optimal path", especially in abnormal cases and unexpected situations. For this case study, the researchers used data from "yellow taxis" circulating in New York City in the pre- and post-Covid period. The model created was able to detect anomalous events such as unexpected concentrations of people and, from there, adapt its recommendations on the best route to follow between two points, having been able to increase the weekly profitability of the vehicles by 98%.

Martin ŠAUER et al. [12], 2021, developed knowledge about the intra-regional flow of tourists in Central Europe and its implications, having listed that understanding and typifying these movements is essential for strategic planning and sustainable development, particularly at the level of the most visited cities. In carrying out this work, data provided by entities responsible for tourism in various cities in Germany, Austria and the Czech Republic were used and from these data it was possible to conclude that the factors that most influence the distribution of tourists are: air connections, the attractiveness of the chosen destination and the size of the tourism market in the place where visitors come from, given that, as in the case of Lisbon, the Germans are the ones who most influence tourism.

Continuing with the study of work done in relation to mobility, Xin Lao et al. [13], 2021, in the article entitled "Comparing Intercity Mobility Patterns among Different Holidays in China: a Big Data Analysis" made use of data provided by Tencent, a Chinese Internet-related service company to model mobility patterns between Chinese cities on holidays, identifying the differences between different more traditional festive seasons such as Spring Festival, Tomb-sweeping Day, Dragon Boat Festival, and Mid-Autumn Festival and the less traditional ones in China, like Christmas and New Year's Eve. Through the work carried out in this article, they were able to prove that: a) movements are different depending on the type of holidays, b) the cities of Pearl River Delta and

Xi'an are those from which more people leave for their hometowns during the traditional festive periods and c) during less traditional holidays, travel is mainly for recreational reasons, unlike traditional holidays where people travel mainly for cultural and traditional reasons.

Xin Li et al., 2018 [14], through the article entitled "Position prediction system based on spacial-temporal regularity of object mobility" proposed the creation of a system that makes predictions regarding the mobility of a given object, using historical data of the referred object that could be any type of "connected device" with GPS, from a mobile phone, to a car or any type of wearable - from the past data, the proposed model is able to predict which will be the next positions occupied by the referred object, and each possible position to be occupied in space is classified according to a score calculated on historical data, with the one with the highest classification being displayed. It should be noted that when evaluating the accuracy of the model proposed by Xin Li et al., it obtained accuracy rates 44% higher than those obtained with an algorithm based on Markov time series, thus proving the capacity of its model for predictions.

In the article entitled "The path of least resistance explaining tourist mobility patterns in destination areas using Airbnb data", Umut Turk et al. [15], 2021, resorted to data provided by the Airbnb local accommodation platform to identify which are the 25 most attractive tourist destinations the world, having as motivation to do so the fact that a lack of knowledge on the topic was identified. Initially, an assessment was made of the quality of Airbnb's offer and prices in each of the locations and subsequently the prices and quality of the transport network were evaluated in each of the locations studied. The authors confirmed that the asking price for local accommodation is directly related to its geographic location and the quality of public transport to the places of interest in each city that one of the reasons that most weigh in choosing accommodation is the proximity to good transport, especially in cities like Berlin and Frankfurt (Table 1).

Table 1. Keywords definition

<table>
<tr><th></th><th colspan="3">Number of documents</th></tr>
<tr><td>Concept</td><td rowspan="3">453106</td><td rowspan="8">3156</td><td rowspan="11">44</td></tr>
<tr><td>Data Analysis</td></tr>
<tr><td>Behavio#r Analysis</td></tr>
<tr><td>Population</td><td rowspan="5">220301</td></tr>
<tr><td>smart cities</td></tr>
<tr><td>cellular network</td></tr>
<tr><td>Touris*</td></tr>
<tr><td>Roaming</td></tr>
<tr><td>Context</td><td colspan="2" rowspan="2">642769</td></tr>
<tr><td>Mobility</td></tr>
<tr><td>Limitations</td><td colspan="2"></td></tr>
</table>

(*continued*)

Table 1. (*continued*)

	Number of documents	
Period: 2018 - 2022		
Only Journal papers, articles and reviews		

This is demonstrated by the 44 documents returned by the query (Concept AND Population AND Context AND Limitations) when we use the keywords from each column.

After completing a manual approach to identify the key subjects for their research questions and specify the outcomes, 16 publications were identified. Our study's systematization considered the year, the region, the RQ subject, and a succinct description. The 16 studies that were reviewed were selected based on the standards. The trend line in Fig. 1 reveals that the subject we're looking at is becoming more and more popular, underscoring its significance.

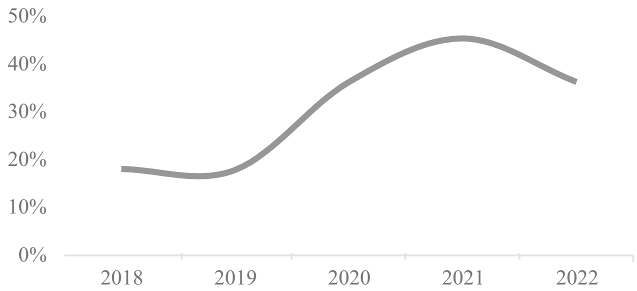

Fig. 1. Evolution of relevant studies per year

Given that the goal of this study is to identify how tourist behavior analysis and tourism mobility are used in smart cities, Table 2 and Fig. 2 provide theoretical explanations of the topics mentioned in each of the papers that were evaluated, with a focus on the use of mobile phones and tourist behavior analysis when using mobile devices. Figure 2 demonstrates how most studies examined how people used mobile phones and other ICT infrastructure and their behavior (ICT). Our research is based on both ideas since we not only analyze human behavior utilizing Lisbon's communication infrastructure as an operator, but also grasp it and develop a plan to satisfy their needs.

Table 2 provides a summary of a more thorough analysis of this review. The problems were plainly stated; therefore, it was unnecessary to ask the publications' authors for clarification. Since the studies' results were categorized based on their inclusion or exclusion in the research, they are not mutually exclusive.

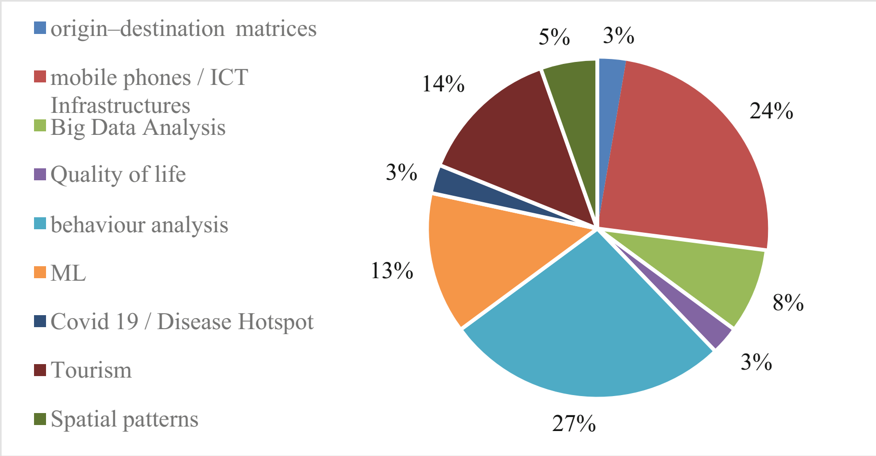

Fig. 2. Relative weight by document subject

Table 2. Summary review analysis

Topic	**Reference**
(1) origin–destination matrices	[3]
(2) mobile phones/ICT Infrastructures	[3–11]
(3) Big Data Analysis	[3, 5, 10]
(4) Quality of life	[12]
(5) behaviour analysis	[4, 8, 10, 11, 13–17]
(6) ML	[4, 6, 7, 9, 13]
(7) Covid 19/Disease Hotspot	[13]
(8) Tourism	[11, 13]– [15, 17]
(9) Spatial patterns	[14, 15]

3 Knowledge Extraction Approach

CRISP DM stands for Cross Industry Standard Process for Data Mining [19]. The goal of CRISP DM is providing us with a structured way of planning and executing a data mining project, ensuring that the best insights are retrieved from the available data. During the present work we followed this methodology as it has proven to lead to a more efficient data mining. Given the relevant data and our primary objectives, we chose a modified version of this methodology for our project that consists of 3 phases, due the need of visual dashboards for decision makers: 1) Business understanding; 2) Data understanding 3) Data preparation and 4) Visualization

3.1 Business Understanding

In this first stage, the emphasis is on comprehending the project's requirements and goals from a business standpoint. Using this knowledge, a data mining issue definition and a rough project schedule are then created to meet the goals.

Understanding the project's goals and requirements is the focus of the business understanding phase and it is divided in 4 sub-tasks. Except for the third task, the remaining three tasks in this phase are fundamental project management procedures that apply to most projects:

a) Define business objectives
b) Assess situation
c) Define data mining goals
d) Create a project plan

In our specific case, and considering that we have no knowledge regarding the tourism business, in addition to what is common sense, we mostly resort to the help of the Lisbon city Council and the LX Data Lab. This was the way found to ensure that we were able to obtain enough knowledge to allow us to interpret the data and the results obtained, this would not be possible without knowing the business or, at least, the analyzes would be more superficial and eventually with less added value.

3.2 Data Understanding

The data made available by Lisbon city Council (Câmara Municipal de Lisboa) is supplied under an agreement with a mobile operator and is generated using the information provided by the cellular network and the mobile devices of each user. The information contained in the dataset is duly anonymized for legal and privacy reasons. In this way, it is not possible in any way to make any specific analysis of a particular user. There is not even any key that relates a given user to an event, and it is only possible to carry out analyzes involving volumes.

All the data available is aggregated in 3743 grids of 200 × 200 meters, being collected in periods of 5 minutes. Due to privacy constraints, if a certain grid doesn't have at least 10 users in the 5 minutes frame, it won't be reported. Data is made available on the big data platform, for a period of about 45 minutes after being collected. This means that we can have a maximum of 1 hour delay between the collection and the availability of the data. However, it is important to say that for the scope of the present work, we will use a snapshot of the data and, therefore, we will not be leverage of the online data stream. Although we won't be using them all in our project, Table 3 presents the 24 indicators/dimensions available in the data provided by the Mobile Operator.

In addition to the dataset that contains the data provided by the city Council through the agreement established with Mobile Operator, a dataset that contains information related to each of the 3743 grids was also used. These are the data that allow us to geo-reference the main dataset since it contains the coordinates of the centroid of each grid, the parish, or parishes in which the grid is inserted, the name, the geometry and the WKT. With this information and using the "Grid_ID" key, it becomes possible to

Table 3. Mobile operator dataset variables

ID	Name	Description	Type
0	Grid_ID	Number of the grid There are 3743 squares of 200 by 200 m to cover the metropolitan area of Lisbon	Nominal
1	Datetime	Time and date of occurrence	Datetime
2	C1	Number of distinct terminals counted on each grid cell during the 5-min period – Measured every 5 min	Metric
3	C2	Number of distinct terminals in roaming counted on each grid cell during the 5-min period– Measured every 5 min	Metric
4	C3	No. of distinct terminals that remained in the grid cell counted at the end of each 5-min period	Metric
5	C4	No. of distinct terminals in roaming that remained in the grid cell counted at the end of each 5-min period	Metric
6	C5	No. of distinct terminals entering the grid	Metric
7	C6	Terminals leaving the grid – These are the distinct terminals that left the grid. The calculation is made using the previous 5-min interval as reference, also considering the crossings of the grid in the same interval	Metric
8	C7	Number of entries of distinct terminals, in roaming, in the grid	Metric
9	C8	Number of exits of distinct terminals, in roaming, in the grid	Metric
10	C9	Total no. of distinct terminals with active data connection in the grid cell – Measurement every 5 min	Metric
11	C10	Total no. of distinct terminals, in roaming, with active data connection in the grid cell – Measurement every 5 min	Metric
12	C11	No. of voices calls originating from the grid cell	Metric
13	C12	Entering the city: No. of devices that for 5 min enter the 11 street sections considered for analysis. For this purpose, a section of track is a route with	Metric
14	C13	Entering the city: No. of devices that for 5 min enter the 11 street sections considered for analysis. For this purpose, a section of track is a route with	Metric
15	D1	Top 10 origin Countries of the devices in Roaming	Metric
16	E1	Number of voice calls that ended in the grid within the 5-min	Metric
17	E2	Average download speed per grid within the 5-min	Metric
18	E3	Average load speed per grid within the 5-min	Metric
19	E4	Peak download speed on the grid within the 5-min	Metric
20	E5	Peak upload speed on the grid within the 5-min	Metric
21	E6	Top 10 apps used on the grid within the 5-min	Metric

(*continued*)

Table 3. (*continued*)

ID	Name	Description	Type
22	E7	Lowest permanence period on the grid within the 5-min	Metric
23	E8	Average permanence on the grid within the 5-min	Metric
24	E9	Maximum permanence period on the grid within the 5-min	Metric
25	E10	Count of devices sharing the internet connection in the grid within the 5-min	Metric

insert the events in the space and, from there, trigger our analyses, after collecting the data for our study, we carefully examined it and investigated each variable to understand its potential and how we could increase the added value of this research. As previously mentioned, our key objective is to comprehend how tourists move around. To do so, the Lisbon city Council provided us with a dataset about people's movement in the city of Lisbon (both roaming and non-roaming), based on mobile phone data produced. The mobile operator extrapolated the data to create the currently accessible dataset to provide a more accurate depiction of the mobility of all individuals who moved around Lisbon between September 2021 and January 2022.

3.3 Data Preparation

This process was oriented to the dashboard visualization of geographic and temporal data to obtain a clear image that would be able to help us understand the data and address the questions we set out to answer. In the course of our work, we realized that the result would be as rich as the more information we were able to provide to stakeholders

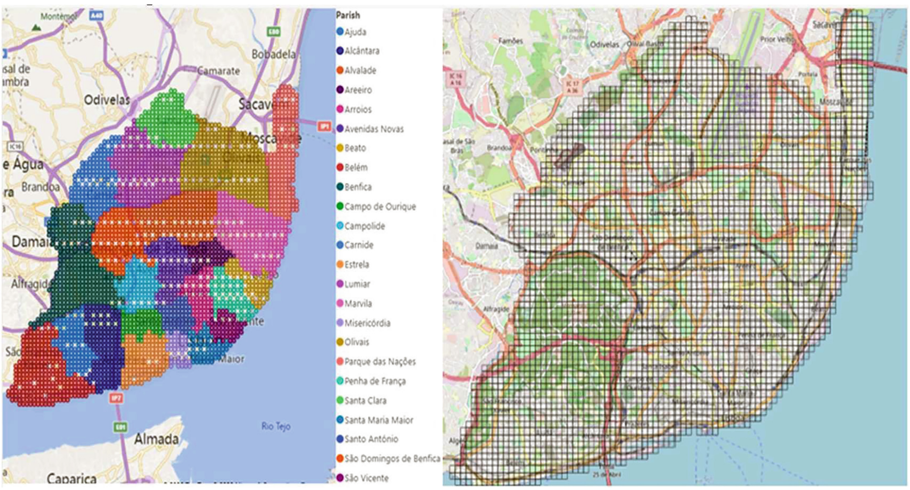

Fig. 3. Lisbon Districts, and Operator grid and respective cells

4 Insights and Visualizations

Once the data had been worked on and prepared, it was time to start creating the graphics and visualizations that effectively allow us to respond to the questions we set ourselves. Therefore, in this section, we will continue our analysis. For this phase, two "high-level" tools were used, namely, Microsoft Power BI (https://en.wikipedia.org/wiki/Microsoft_Power_BI) and Microsoft Excel (https://en.wikipedia.org/wiki/Microsoft_Excel). The latter, not being exactly a massive data analysis tool, with due work, allowed us to obtain interesting results.

At this stage, we portrayed the data graphically to make it easier to focus on the most crucial information and quickly identify trends and patterns in the tourist population's mobility. We can study and discover more about data using graphs and charts.

4.1 Where Do the Top Visitors of Lisbon Come from?

Bearing in mind that our work focuses on tourists in the city of Lisbon and their habits, it makes perfect sense that we start exactly by typifying their origins, whether from the country or from the Continent/Geographical Area where they come from. In Fig. 2, representing the average number of Visitors in each 200 × 200 meters cell from the grid, it is visible that the top six of the origin of the Tourists in the 5 months of analyzed data, are the same, although the rankings change between the months. According to available data, the largest number of visitors to Lisbon come from Germany, Spain, France, Italy, the United Kingdom, and Brazil. The analyzed month with more visitors was November, eventually due to Web Summit.

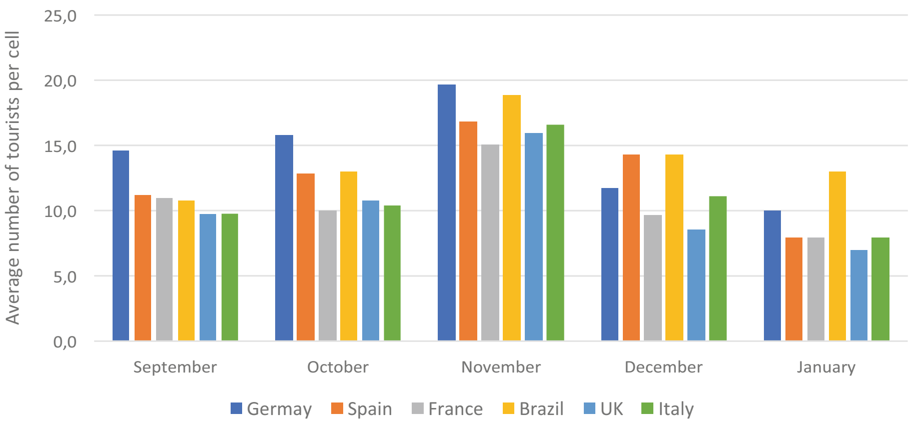

Fig. 4. Top six Lisbon tourists by citizenship

We also consider it important to go up a level in terms of geographic aggregation and for that, we grouped some of the countries in large geographic areas whose result is shown in Fig. 3, representing the average number of Visitors in each grid cell (200 × 200 meters). Through this visualization, we can see that the most represented areas of

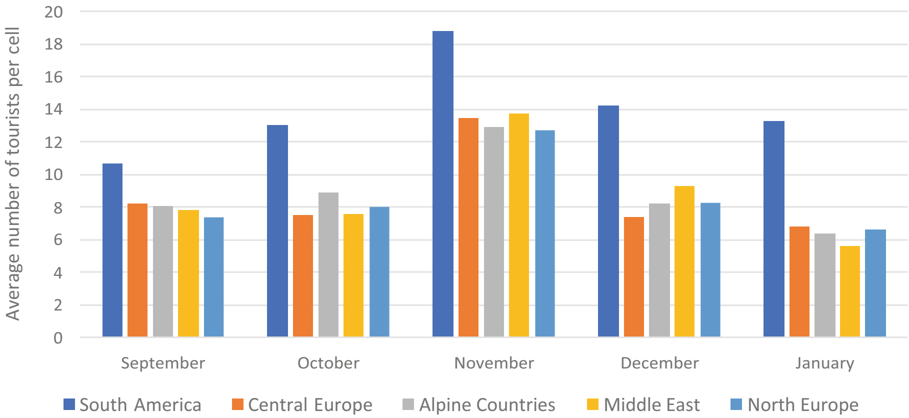

Fig. 5. Evolution of Lisbon tourists by geographic area

touristic origin are South America, Europe (South/North/Center), the Alpine Countries and the Middle East. This information is supported by the per Country analysis performed before. Once again, we can see that November was the month with more Tourists, most probably because of the *Web Summit*.

4.2 What Are the Most Visited Areas of Lisbon?

As previously mentioned, we grouped the parishes of Lisbon and our first analysis of the most visited areas of the city falls precisely on this grouping. Not surprisingly, the areas most visited by tourists are the Historic Center and the City Center of Lisbon (Fig. 4), representing the average number of Visitors in each of the 200 $\times$ 200 meters grids. Even so, it is interesting to check the average number of visitors in each of the grids and their evolution over the months.

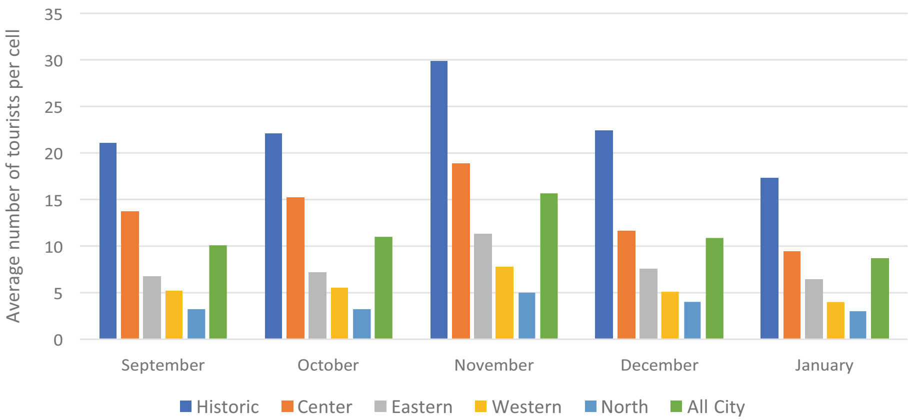

Fig. 6. Evolution of Lisbon Tourists by Month and City Area

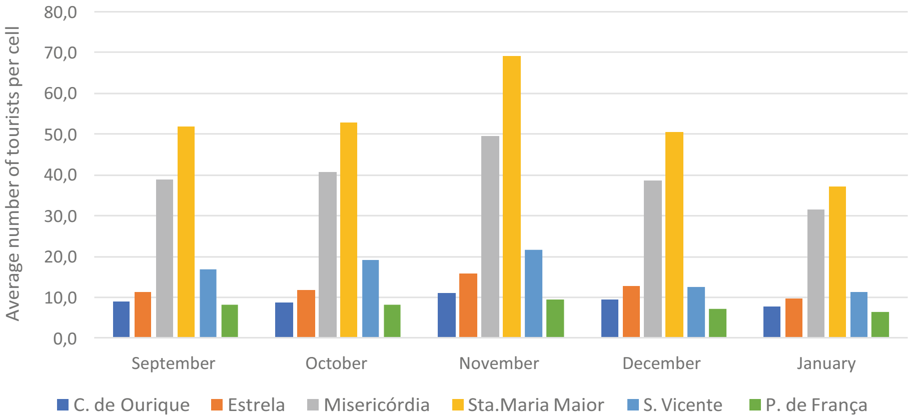

Fig. 7. Evolution of Lisbon Tourists by Month in Historical Parishes

According to the available data and through Fig. 5 and focusing on parishes in the Historic Center of Lisbon that receive the most foreign visitors we can see that *Santa Maria Maior* and *Misericórdia* are by far the most visited ones. This insight is consistent with the fact that it is in these parishes that very emblematic monuments of the city are located.

Through a georeferenced analysis, using the centroids included in the dataset of the mobile operator, we were able to understand the exact locations in the parishes of the historic center of Lisbon where Tourists travel the most. In Fig. 6, it is possible to see highlight in Castelo, *Alfama*, *Baixa Pombalina* and *São Vicente de Fora*. These visualizations provide useful insights to decision-makers detailed information at street and hourly level.

Continuing with the analysis of the Parishes most visited by Tourists in Lisbon, it is still very important to characterize those belonging to the City Centre. Among the 6, and as we can see in Fig. 7, representing the average number of Visitors in each grid cell (200 × 200 meters), there are 3 that stand out for clearly having several tourists well above the others and they are *Santo António*, *Arroios* and *Avenidas Novas*. Through the geographic analysis (Fig. 8), it is possible to clearly perceive that Avenida da Liberdade and the Saldanha area are where more tourists move in the parishes of the center of the city.

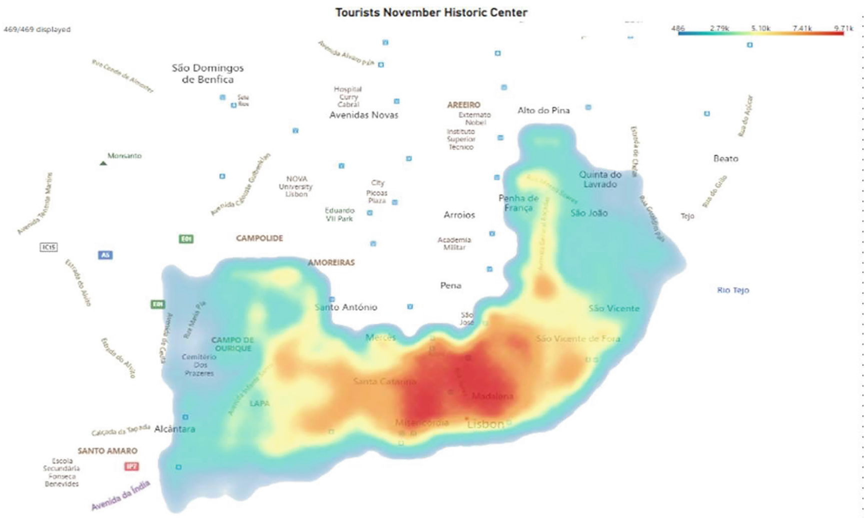

Fig. 8. Heatmap of Lisbon Tourists in Historical Parishes (November)

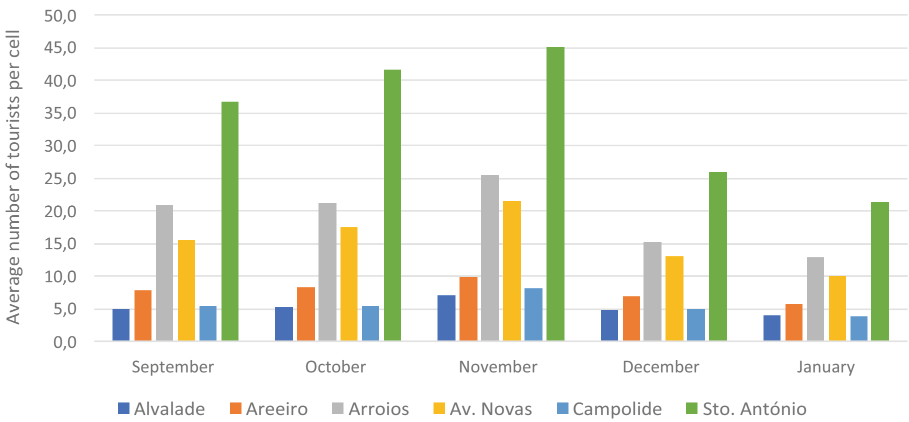

Fig. 9. Evolution of Lisbon Tourists by Month in City Center Parishes

4.3 Where Are the Tourists During the Mealtimes and Where Do They Sleep?

Using the centroids present in the dataset of the mobile operator again and applying a time filter to the data, considering that the lunch period is between 12am and 2.30pm and dinner is between 7.30pm and 10pm, we created the visualization in Power BI below (Fig. 9) which shows the concentration of tourists in the afore mentioned periods. Carrying out the exercise for the month of September, even though the data are prepared to do so for any of the months, we can see that, with the exception of the West zone of Lisbon (Belém/Alcântara), visitors have lunch and dinner in the same places in the city, which ends up making sense since it does not make much sense for them to travel to have their meals, unless the restaurant it is also a point of interest. Again, the data and

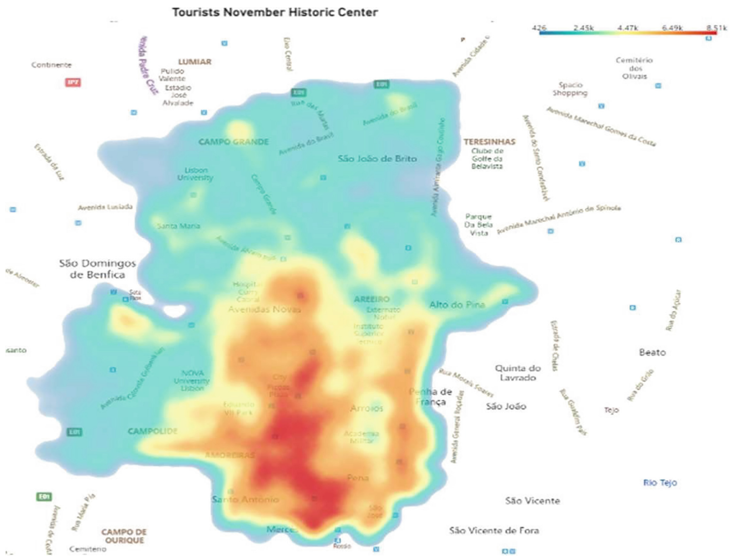

Fig. 10. Heatmap of Lisbon Tourists in City Center Parishes (November)

visualizations are prepared to a level of detail that allows you to go down to street and time level.

4.4 Event Analysis

As is well known, the number of major international events has been growing, bringing an even greater number of visitors to the city, in addition to the already large number of tourists. For the present work, we believe that it would make perfect sense to compare the volumes and origins on the days of the event, with the same days of the previous or subsequent week, this being the most correct way we found to make the comparison.

Therefore, we developed the analysis of the Web Summit (https://websummit.com/) that took place between the 1st and 4th of November 2021, and for the counterpart days of the week before and the week after the event. Through Fig. 10, it is possible to see that on the days of the event, tourists were mostly concentrated in the Lisbon International Fair (https://www.fil.pt/) while on the same days of the counterpart weeks were more spread out above by the points of interest of the Parish, namely in the Oceanário. It is also possible to see that the number of Tourists on the days of the event almost tripled compared to the same period last year. It is also possible to verify that, in percentage terms, the geographic areas of origin of the visitors are also quite different. During the Web Summit week about 21% of the Tourists were from South America while in the previous and following were from the Eastern Europe.

Additionally, we also analyze the influence of a sporting event, in this case a Champions League football match held on October 20th - Benfica - Bayern. For this case, we only use visitors from Germany, as Bayern Munich is a German Club. For this analysis, we applied a time filter between 07pm and 10pm, in the parishes of Benfica, Carnide and São Domingos de Benfica. On Fig. 11, it is noticeable that, compared to the same day and time of the previous and the after week, the number of Germans in the analyzed parishes grew almost 20 times and that their concentration was almost exclusively in the centroids of Estádio da Luz. The fact that there is a much higher than usual volume of German tourists in the area surrounding the stadium on the day of a Champions League game against a German team, does not represent anything new, being exactly what would be expected from find in this analysis. The exercise intends to demonstrate the capability of a tool like the one that was developed to help quickly and in real time take on events that may require, for example, the intervention of security forces (Figs. 12 and 13).

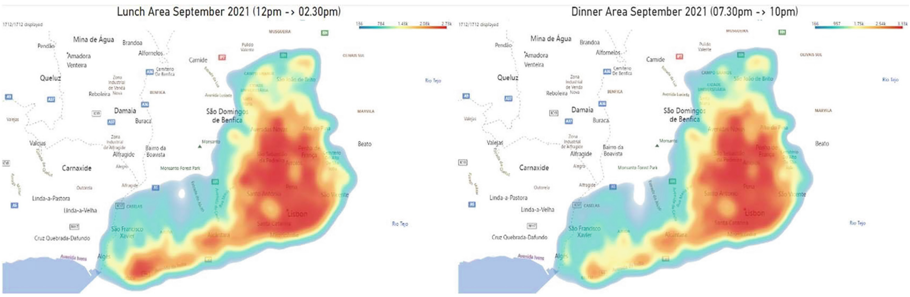

Fig. 11. Heatmap of Lisbon Tourists at Lunch and Dinner Time (September)

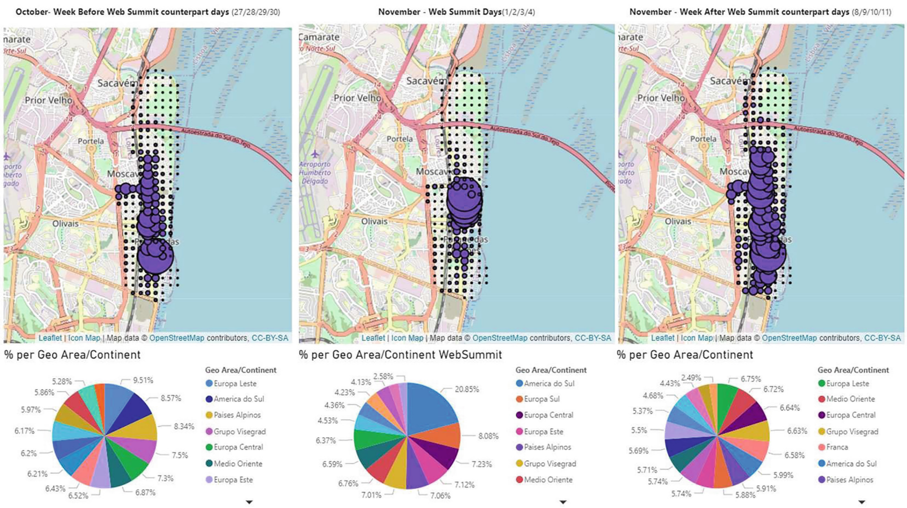

Fig. 12. Web Summit Analysis

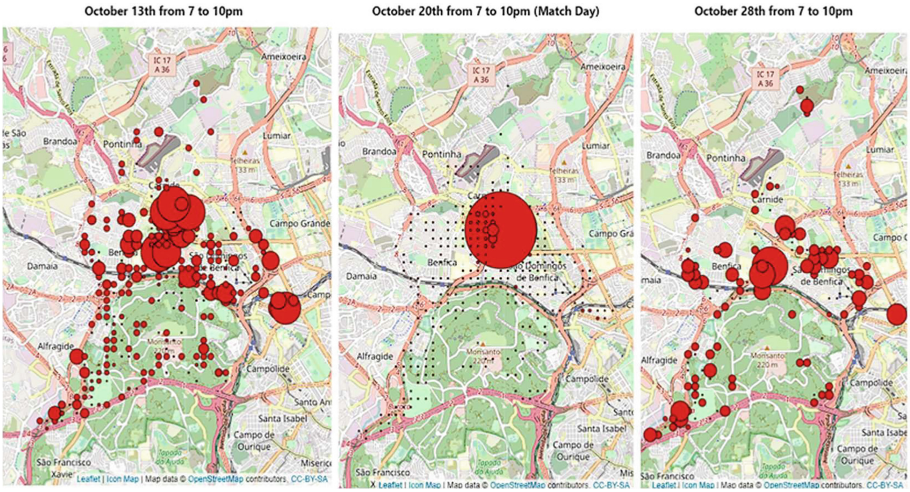

Fig. 13. Benfica vs Bayern

5 Conclusions

It is essential for public policy to comprehend the spatial spread of urban tourism. As a result, in areas where there is a high concentration of tourists, local authorities might think about initiatives to improve the tourist experience, like creating pedestrian-only lanes or enlarging sidewalks, increasing the number of public spaces with free Wi-Fi, and positioning new tourist information centres, among other things.

Through the method developed and the tools used, we were able to answer the questions we wanted to. It was possible to identify the origin, and the number of Tourists in Lisbon as well as the evolution over the period under study (September 2021 to January 2022), making a separation not only by Country but also by Continent/Large Geographical Area. We were able to understand and demonstrate which areas and parishes are most visited in the City of Lisbon and the places where they are during the typical meal and sleep times. Finally, we were able to understand in a way that a major international event such as the Web Summit or a football match in the Champions League changes the panorama of tourists/visitors in Lisbon.

In that sense, properly adapting our method to handle a constant flow of data, public policy makers, like the Lisbon Municipal Council or the National Tourists Office, can take advantage of an aggregated view that in real time manages the resources of various departments ranging from transport, hygiene, and safety. By processing this data in real-time, everyone involved in the management of the city in its various vectors will be able, on the one hand, to provide a much more pleasant experience for visitors, but also to avoid security breaches and any type of unwanted events Additionally, our work can encompass an analysis tool based on real data that allows, in the medium and long term, to plan accommodation and commerce.

It is also important to mention that with the available data, we can have much more insights than the ones we refer to in this work. However, due to scope limitations, we have chosen the one that seems to have the best fit. The developed tool was designed with the possibility of using a series of filters that allows the Lisbon City Council to make its own analysis on topics that were not explored in this dissertation, making more final analysis both from a temporal and geographical point of view. If users deem it necessary, they can still use the solution developed to load new data, if they follow the same structure, making it possible to use it in analyses with new information coming from the same source.

To process an amount of data of this order to obtain real value for the benefit not only of Tourism, but of all those who travel through Lisbon for work or leisure, it is essential that there is a large computational and analytical capacity. For the first case, there should be recourse to cloud technology, which, as is well known, although with high costs, allows processing large amounts of information, without delay and without the need to install local capacity. From an analytical point of view, it makes sense to develop machine learning mechanisms, capable of highlighting patterns and helping data analysis, providing decision-makers with automated reports and dashboards to support decision making.

References

1. LISBONEIDEE Homepage. http://www.lisbonne-idee.pt/p5358-lisboa-recebe-mais-turistas-poresidente-que-londres.html. Accessed 23 July 2022
2. Câmara Municipal de Lisboa Homepage. https://www.lisboa.pt/. Accessed 23 July 2022
3. LxDataLab Homepage. https://lisboainteligente.cm-lisboa.pt/. Accessed 23 July 2022
4. Secretaria Geral da Economia Homepage. https://www.sgeconomia.gov.pt/noticias/portugal-e-o-5-pais-com-mais-forte-contributo-do-turismo-para-o-pib.aspx. Accessed 24 July 2022
5. Fekih, M., Bellemans, T., Smoreda, Z., Bonnel, P., Furno, A., Galland, S.: A data-driven approach for origin–destination matrix construction from cellular network signalling data: a case study of Lyon region (France). Transportation **48**(4), 1671–1702 (2020). https://doi.org/10.1007/s11116-020-10108-w
6. Carter, E., Adam, P., Tsakis, D., Shaw, S., Watson, R., Ryan, P.: Enhancing pedestrian mobility in smart cities using big data. J. Manag. Anal. **7**(2), 173–188 (2020). https://doi.org/10.1080/23270012.2020.1741039
7. Badii, C., Difino, A., Nesi, P., Paoli, I., Paolucci, M.: Classification of users' transportation modalities from mobiles in real operating conditions. Multimed. Tools Appl. **81**(1), 115–140 (2021). https://doi.org/10.1007/s11042-021-10993-y
8. Haidery, S.A., Ullah, H., Ullah Khan, N., Fatima, K., Shahla Rizvi, S., Kwon, S.J.: Role of big data in the development of smart city by analyzing the density of residents in Shanghai. Electron. Switz. **9**(5) (2020). https://doi.org/10.3390/electronics9050837
9. Mizzi, C., et al.: Unraveling pedestrian mobility on a road network using ICTs data during great tourist events. EPJ Data Science **7**(1), 1–21 (2018). https://doi.org/10.1140/epjds/s13688-018-0168-2
10. Jeong, J., et al.: Mobility prediction for 5G core networks. IEEE Commun. Stand. Mag. **5**(1), 56–61 (2021). https://doi.org/10.1109/MCOMSTD.001.2000046
11. Guo, P., Xiao, K., Ye, Z., Zhu, W.: Route optimization via environment-aware deep network and reinforcement learning. ACM Trans. Intell. Syst. Technol. **12**(6) (2021). https://doi.org/10.1145/3461645
12. Šauer, M., Vystoupil, J., Novotná, M., Widawski, K.: Central European tourist flows: intraregional patterns and their implications. Morav. Geogr. Rep. **29**(4), 278–291 (2021). https://doi.org/10.2478/mgr-2021-0020

13. Lao, X., Deng, X., Gu, H., Yang, J., Yu, H., Xu, Z.: Comparing intercity mobility patterns among different holidays in china: a big data analysis. Appl. Spat. Anal. Policy (2022). https://doi.org/10.1007/s12061-021-09433-z
14. Li, X., et al.: Position prediction system based on spatio-temporal regularity of object mobility. Inf. Syst. **75**, 43–55 (2018). https://doi.org/10.1016/j.is.2018.02.004
15. Türk, U., Östh, J., Kourtit, K., Nijkamp, P.: The path of least resistance explaining tourist mobility patterns in destination areas using Airbnb data. J. Transp. Geogr. **94** (2021). https://doi.org/10.1016/j.jtrangeo.2021.103130
16. Data Science Process Alliance. https://www.datascience-pm.com/crisp-dm-2. Accessed 24 July 2022
17. GSM Association Homepage. https://www.gsma.com/mobileeconomy/europe/. Accessed 20 Dec 2022
18. ANACOM Homepage. https://www.anacom.pt/streaming/ServicosMoveis1T22.pdf?contentId=1722575&field=ATTACHED_FILE. Accessed 20 Dec 2022