City Mobility and Night Life Monitor

Luís B. Elvas[1,2](✉), Miguel Nunes[1], Bruno Francisco[1], and Nuno Domingues[3]

[1] Instituto Universitário de Lisboa (ISCTE-IUL), ISTAR, Lisbon, Portugal
luis.elvas@iscte.pt
[2] Department of Logistics, Molde University College, 6410 Molde, Norway
[3] ISEL-ADEM, Rua Conselheiro Emídio Navarro, 1, 1959-007 Lisbon, Portugal

Abstract. This paper presents an Internet of Things (IoT) system designed to collect and analyse information regarding the travel patterns and movements of individuals in densely populated locations, in the context of smart cities. People's movements are retrieved from coarse-grained aggregated cellular network data without collecting sensitive information from mobile devices and users. These data were provided by a Portuguese cellular operator to the Lisbon City Council to characterize people movements in the city. In this sense, the mobile phones act as useful sensor devices for collecting rich spatiotemporal information about human movement patterns. The purpose of this research work is to create a machine learning-based data-driven approach that is able to receive anonymised data from telecommunication operators to provide a big picture about citizen mobility in the city and to identify patterns based on the collected data, in order to provide relevant information for city planning and events coordination. Some of the main applications of the proposed system are the coordination of big events and the management and control of commuting traffic.

Keywords: machine learning · big data · mobility patterns · IoT · smart cities

1 Introduction

In contemporary cultures, the fast growth of information and communication technologies (ICTs) has produced abundant resources for spatio-temporal data mining and knowledge discovery. Understanding aggregated urban mobility patterns based on mobile phone statistics, such as identifying activity hotspots and clusters, is one important topic for municipality city management.

With the advent of powerful smartphones and creative mobile apps, the data volume generated by these tools has exploded. Due to the potential value contained within these enormous datasets, it is vital to be able to correctly and efficiently monitor and analyze these data. In this sense, several methods have been developed for comprehending associated patterns.

A. L. Martins et al. (Eds.): INTSYS 2023, LNICST 540, pp. 126–150, 2024.
https://doi.org/10.1007/978-3-031-49379-9_7

In the context of smart cities, work has been done to investigate the application of data-driven methodologies and visualization techniques in order to support decision-makers in urban settings by giving them useful information and tools. Studies on topics like incident management [1], traffic accidents [2], disaster management [3, 4], and city-wide data analytics have all been done. The results show the potential advantages of using data and sophisticated analytics to improve urban planning, increase traffic safety, and create efficient plans for disaster and event management. In the end, emphasis has been placed on the value of data-driven strategies in enabling decision-makers to build urban environments that are more effective, sustainable, and resilient.Different uses for the collected smartphone data include the recognition of mobility path patterns, traffic planning, route prediction and city-wide sensing applications. Previous research on the application of these data emphasizes their significant potential for analyzing minute differences in human mobility. However, high-level mobility information and low-level location data are often disconnected. Consequently, it is necessary to provide suitable methods for dealing with low-level location data to allow gaining valuable insights regarding the mobility patterns of users.

Mobile phones present unique qualities that entice academics and enterprises to leverage their data. The research conducted in the past has led to the development of several mobile sensing techniques that rely on position tracing or mobile positioning, which involves tracking the location coordinates of mobile phones. Numerous location-based services (LBSs) incorporate geographical information systems (GISs), based on global navigation satellite systems (GNSSs), such as the Global Positioning System (GPS), and the Internet. These LBSs record the movement, flows, and location of individuals in order to recommend social activities or provide tailored advertising.

Mobile phone location tracking may be classified as either active or passive. In the former the device position is identified, for example, using the proximity to radio-waves (cell ID tracking) and the triangulation of signals from multiple sources, such as with GPS. The latter resorts to billing data that are saved as a result of routine procedures. This solution requires the capacity for distance-based charging. Mobile phone calls and SMSs sent or received produce records including cell IDs, enabling the approximate position of the phone to be established. By receiving and analyzing this kind of location data supplied by mobile networks, mobile operators may subsequently build more effective monitoring tactics.

The remainder of this paper is structured as follows: State of the art is presented in Sect. 2. Section 3 introduce the methodology that we have used, as well as all the methods used in this work. Finally, Sect. 4 presents the conclusions.

2 State of the Art

2.1 Search Strategy and Inclusion Criteria

A systematic literature review was made by following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) Methodology [5], and with the research question (RQ) "What is the state of the art on the behaviour and mobility analysis on smart cities?".

The database searched was Scopus and the research was conducted between 8 and 12 of May, 2022; all the results had to be articles published on the last five years and written in English. Grey literature, reviews, conference papers, workshops, books, and editorials were excluded, as well as works not related to the domain.

The search strategy was based on queries made with different research focuses of. This method allowed for the observation of the number of existing articles, considering the concept and context, and the population under study. The initial selection of papers was done using the title and abstract, and, in some cases in which that information was insufficient, the full document was analyzed.

2.2 Data Extraction and Synthesis

The review data were managed and stored using Zotero and Microsoft Excel. These data were title, author, year, journal, subject area, keywords and abstract. For data synthesis and analysis, a qualitative assessment was conducted based on the results presented above. The Scopus database was searched systematically regarding the published literature work related with the purpose of "Data Analysis" or "Behavior Analysis", with the target scenario of "Smart cities" or "Cellular network", and within the "Mobility" context for the study.

2.3 Results

The number of documents obtained using the defined keywords is presented in Table 1. The query was made in the Scopus database with the same restrictions and filters.

Table 1. Documents obtained using the selection criteria.

<table>
<tr><th>Concept</th><th>Population</th><th>Context</th><th>Limitations</th></tr>
<tr><td>Data Analysis
Behaviour Analysis</td><td>smart cities cellular network</td><td>Mobility</td><td>2018–2012</td></tr>
<tr><td>453.106 Documents</td><td>35.965 Documents</td><td rowspan="2">642.769 Documents</td><td rowspan="2">Only journal papers, articles, and reviews</td></tr>
<tr><td colspan="2">923</td></tr>
<tr><td colspan="4">17 Documents</td></tr>
</table>

From this we can see that when the query was made using the keywords from all columns, it returned 17 documents. After performing a manual process towards the identification of significant subjects from research questions and identifying the outcomes, 10 documents were obtained. Our research systematization considered year, area, RQ topic and a small description.

2.4 Study Characteristics

From Fig. 1 we can notice from the trend line that there is a growth on the topic that we are studying, revealing his relevance.

Emphasizing also the relevance of the theme in our area, it's possible to confirm on Fig. 2 that the subject area where more studies are concentrated is precisely Computer Science.

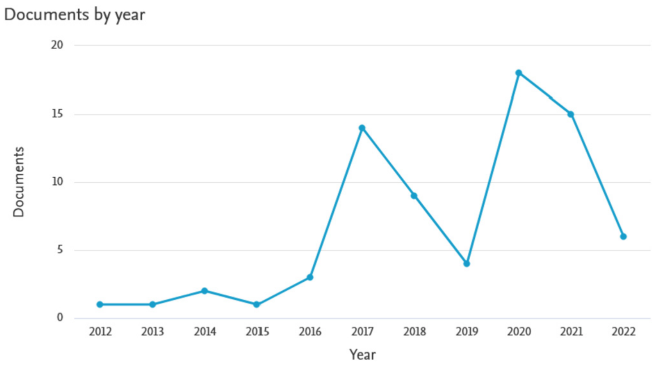

Fig. 1. Evolution of the number of documents per year.

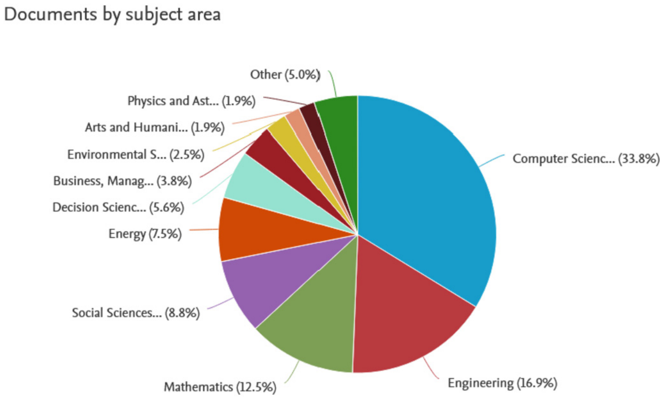

Fig. 2. Classification of the document in subject areas.

Considering that the main goal of this paper is to analyse the user behaviour and mobility in smart cities, a list of the main topics discussed on each of the 10 reviewed articles is illustrated on Fig. 3, where it is noticeable the focus on use of mobile phones and behaviour analysis. Notice that some articles cover multiple topics.

From Fig. 3, we can see that most of the studies focus on mobile phones/information and communications technology (ICT) infrastructures and/or on behaviour analysis. The pertinence of our study falls in between these two themes, because we not only use data from the cellular operators' ICT infrastructure for the city of Lisbon to study people's behaviour, but we also perform a behaviour analysis to understand and help on planning the measures to meet the needs of the citizens.

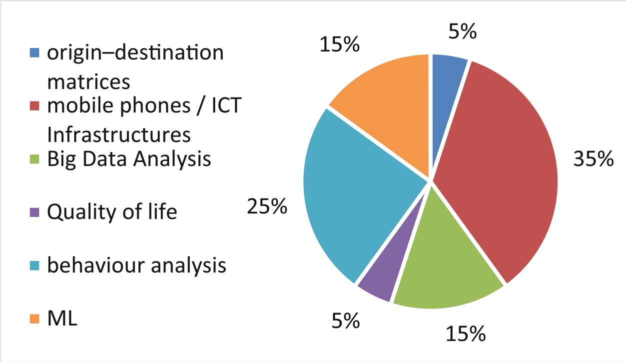

Fig. 3. Topics of the reviewed articles.

A more detailed analysis of this review is summarized in Table 2. As mentioned before, the classification of the studies regarding the outcome is not mutually exclusive, given that these were attributed due to presence/absence in the study.

Table 2. Main topics of reviewed articles.

Topic	Reference
(1) Origin–destination matrices	[6]
(2) Mobile phones / ICT infrastructures	[6–12]
(3) Big data analysis	[6, 8, 12]
(4) Quality of life	[13]
(5) Behaviour analysis	[7, 10, 12, 14, 15]
(6) Machine Learning	[7, 9, 11]

From the topic of mobile phones, authors from [6] offer a technique 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 this topic, authors on study [8] identify pedestrian hotspots and provide future traffic signal and street layout information to make the city more pedestrian friendly, allowing also the use of this obtained knowledge to other datasets, such as bicycle traffic, to guide city infrastructure initiatives.

From the same topic of mobile phones, but focusing on behaviour analysis, the study in [7], which was applied to Tuscany and Florence, 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 sustainable mobility, health, and/or a better and more enjoyable life, among other things.

On this chapter Still on the combination of topics 2 and 5, the goal of [10] is to 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 concern the characteristics of users' spatial behaviour, such as the center of activity, based on check-ins, the feasibility of using check-in data to explain the relationship between users and social media, and the presentation of clear results for regulatory or managing authorities in urban planning. Continuing in these two topics, study [12], based on long-term mobile phone data (from 2007 to 2012) of volunteers from Beijing, presents a way to illustrate individual movement patterns.

Blending topic 2 with machine learning, study [9] intends to give a taxonomy of 5G Cellular Network mobility prediction frameworks, from data gathering to model providing, while taking into account the 3GPP architecture and interfaces.; and we The authors provide two critical use cases in 5G Cellular Networks (CNs), in which the benefits of mobility predictions are assessed using information from real networks. On the other hand, study [11] focuses on building a mobile sequential recommendation system to help auto service companies to increase their profits (e.g., taxi drivers).

On the subject of behavioural analysis, study [14] provides an urban travel behaviour model and evaluates its feasibility for creating a greener environment for future generations. Study [15], based on a trip survey from the São Paulo metropolitan area, which is one of the world's busiest traffic locations, supplement a current bundling approach to enable multi-attribute trail datasets for the visual study of urban mobility, helping to identify and analyse distinct mobility patterns for various data variables, such as peak hours, socioeconomic strata, and transportation modes, according to the findings.

Regarding quality of life, The aim from study [13] is to look at the structural equation model of smart city factors that impact worldwide management of world heritage sites, as well as the quality of life, for Thai visitors and residents in the Ayutthaya province.

3 Methodology

This project follows the Cross-Industry Standard Process for Data Mining (CRISP-DM) Methodology. CRISP-DM is a well-known and commonly utilized methodology for successful data mining operations. CRISP-DM provides a complete framework that leads data mining practitioners through the full data mining process, from identifying business objectives through delivering the findings. The approach is divided into six key phases: Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation and Deployment. Each phase includes a collection of goals and activities that guarantee that data mining initiatives are approached in a methodical and iterative manner. CRISP-DM highlights the significance of understanding the business context and aligning data mining goals with organizational objectives, which increases the likelihood of success in extracting meaningful insights and actionable information from data. Organizations may improve their decision-making processes, optimize resource allocation, and promote innovation by using the CRISP-DM technique [16].

3.1 Business Understanding

The main goal behind the exploration of this dataset is to identify and explore patterns, behaviors and reactions in the population related to the measures taken by the government to lift the restrictions that were imposed to deal with the COVID-19 pandemic. Some of our focus points were studying and exploring the adherence of the population to some activities that were not so usual in the last couple of years, such as the comeback of the nightlife, going to work or going to school.

To create value from the collected data, the following objectives were defined:

- To study and correlate the changes that have occurred in nightlife in periods with different pandemic restrictions. To do this, the movements that occur during night time hours in different establishments and areas with regular nightlife were analysed. Heat maps were constructed, in which the differences in the movement of the population in certain periods can be observed.
- To study the evolution of mobility in Parque das Nações, a redeveloped area of Lisbon which was the site of the 1998 Lisbon World Exposition (Expo '98), during the different months of the pandemic. This area was selected for being composed of business and leisure zones and for containing the only active vaccination centre in the city of Lisbon during the months of study. With this approach it will be possible to study eventual patterns and isolate their causes. After mapping the zone, it was necessary to collect data on factors such as events that occurred in the area, government measures on teleworking and hybrid regimes and vaccination to study population movements in more detail.

3.2 Maintaining the Integrity of the Specifications

The dataset under analysis consists of data provided by Vodafone that contain information about the cellular connections (segmented in a total of 3743 different squares or grid cells of 200 by 200 m) established between mobile devices and base station towers in Lisbon, between September 2021 and January 2022 for every 5 min. It is formed by 3 different types of files. The main files, which contain more information, are composed by 27 different variables. Those variables provide different information about the devices that were detected in every 200 square meter area around the city during those 4 months. The other file types consist of information about the coordinates of every 200-square meter area and the coordinates of 11 selected roads that connect the city to other areas and that help to explore the entrances and the exits of the city during these months.

Each month was splitted into multiple datasets, where each of these datasets contained information about a set of grid cells, divided into 29 columns (Table 3).

After a brief exploratory analysis, it was found that there were several missing data, such as weeks without any records or even missing grid cells. In other words, there was no information about some areas of Lisbon.

Due to the high volume of data, it was necessary to download the data in stages, to ensure that the data were not corrupted during the process and to check one by one if everything was complete. After all this analysis we proceeded with the data cleaning and data procedures.

Table 3. Description of each variable present in the main files.

Variable	Description
Grid_ID	Identification of the grid cell number
Datetime	Date and Time
extract_year_2	Year
extract_month_3	Month
extract_day_4	Day
C1	Number of distinct terminals in the grid, during the 5 min
C2	Number of distinct terminals, roaming, in the grid, during the 5 min
C3	Number of distinct terminals remaining in the grid at the end of each 5 min period
C4	Number of distinct terminals remaining in the grid, roaming, at the end of each 5 min
C5	Number of distinct terminals entries in the grid
C6	The number of distinct terminals exits in the grid
C7	Number of distinct terminal entries in the grid, roaming
C8	Number of distinct terminals exits in the grid, roaming
C9	Number of distinct terminals with an active data connection, in the grid cell, during the 5 min
C10	Number of distinct terminals with an active data connection, roaming, in the grid cell, during the 5 min
C11	Number of voice calls originating from the grid
C12	Number of entries into Lisbon along the 11 main roads
C13	Number of exits into Lisbon along the 11 main roads
D1	Top 10 home countries of terminal equipment roaming
E1	Number of voice calls terminated in the grid
E2	The average downstream rhythm of the grid
E3	The average upstream rhythm of the grid
E4	Peak downstream rhythm of the grid
E5	Peak upstream rhythm of the grid
E6	Top 10 apps (semicolon separated)
E7	Duration of the minimum stay within the grid
E8	Duration of the average stay within the grid
E9	Duration of the maximum stay within the grid
E10	Number of devices performing grid connection sharing during the 5 min period

3.3 Data Preparation

To give a uniform treatment to all the datasets of the analysis, a general Python script was created so that it was possible to apply to all datasets the same process of data cleaning and treatment.

In the first stage of this process, the existence of missing values and duplicate values was verified. It was found that variables "D1" and "E6" always contained more than 50% of missing data. However, the elimination of these rows implied a big loss, so it was decided to eliminate the columns with these variables. There were also many duplicated values, so we eliminated them.

Since each dataset contained millions of observations and our machines are not powerful enough to handle such a large amount of data, we decided to eliminate the irrelevant variables for our analysis, together with the two variables mentioned in the previous paragraph.

It was also verified that the variable "Datetime" was not in the correct format. Consequently, the value recorded in each entry of that variable was broken down into three new variables:

- Date - represents the date on which the device had its registration effected in a day/month/year format;
- Time - represents the time of day on which the record was made in an Hour: Minutes format;
- Hour - selects the time recorded in the "Time" variable to facilitate its use and optimise potential hourly analysis in Microsoft Power BI.

Subsequently, a new variable, called "Weekday", was created using the Pandas library, which was able to represent the day of the week where the record occurred, from the "Date" variable. After this treatment and cleaning phase, all datasets were joined.

In a second phase, to meet the defined objectives and because it would not be computationally feasible to work with all records in the database, the different periods, times and areas of the city of Lisbon that were to be analysed were separated. In this case, the analysis of the zones of Bairro Alto, Santos-o-Velho e Docas were restricted, from 2 am to 4 am on Thursdays, Fridays and Saturdays to analyse the impact of COVID-19 restrictions on nightlife. For the analysis of the mobility in the urban centre of Parque das Nações, we restricted the grid squares to this area. Both analyses comprise all the months under study.

With these two areas, we managed to have a study that covers both the movements of people during the day and the early hours of nightlife. However, to ensure that the right grid cells are chosen, we mapped both areas by going to the locations in person and using tools such as Google Earth or Google Maps. The reason for this decision was that there was a time discrepancy between the records present in these tools and what currently exists at the sites, which required a few weeks of exploration to delimit the areas for study.

After all the mapping was completed, we were faced with another challenge: each of the grid cells where there are records covers an area of 40 000 square meters and there are several types of buildings and/or activities that are covered. For this reason, to avoid an inaccurate and fallacious analysis, we classified each of the cells used in a category that represented the predominantly registered activity.

Since the longitude and latitude are important for the visualisation of the results, it was necessary to add these two variables to the datasets, since they were in different datasets. Finally, as the elimination of irrelevant variables was not enough to reduce the size of the data, it was decided to create samples of the datasets with 20% of the total size so that it would be possible to visualise them.

3.4 Visualization

The Impact of COVID-19 on Nightlife

One of the main objectives of this research was to understand how the measures imposed by the Portuguese government to control the COVID-19 pandemic affected Lisbon's nightlife. For this purpose, this first study started in September 2021 and ended in December 2021. During this period, there were some relevant changes in the measures imposed to control the pandemic. At the start of the study, during September, nightclubs and bars were closed to avoid the propagation of the COVID-19 virus. On the 1st of October, these places were allowed to reopen upon presentation, by the customers, of a digital certificate proving that they had been vaccinated against COVID-19 or otherwise that a test proving the customer was not infected had been performed. This measure lasted until December 24th, when new measures have been imposed making it obligatory for these places to stay closed.

Based on these measures, two distinct periods were selected for analysis: 1) the month of September, in which the measures imposed still prevented nightlife venues from opening; 2) the following three months (October, November and December), in which the nightclubs and bars were allowed to reopen. Within these periods, the areas that were selected for analysis were very popular Lisbon's nightlife areas: the Urban Beach nightclub and the Bairro Alto area, which has several bars that stay open until very late.

Firstly, the research was directed to the study of how movements occurred in the space near the Urban Beach nightclub in September (Fig. 4). It was possible to realize that during the night, the movement reached the minimum daily values, which corroborated the existing limitations on the opening and operation of nightlife spaces.

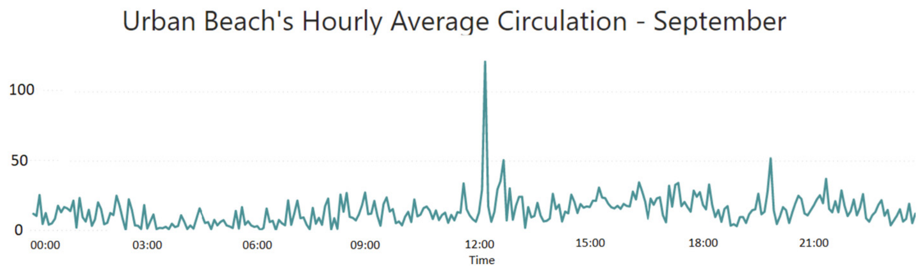

Fig. 4. Evolution of mobility in Urban Beach by hours, in September 2021.

It was also analysed how the movements in this nightclub occurred over the weeks (Fig. 5). During the periods between Friday and Sunday - which are the days when there are typically more events - these movements were low, which suggests that is a direct consequence of the measures implemented by the government to control the pandemic of COVID-19 that were in force during this month.

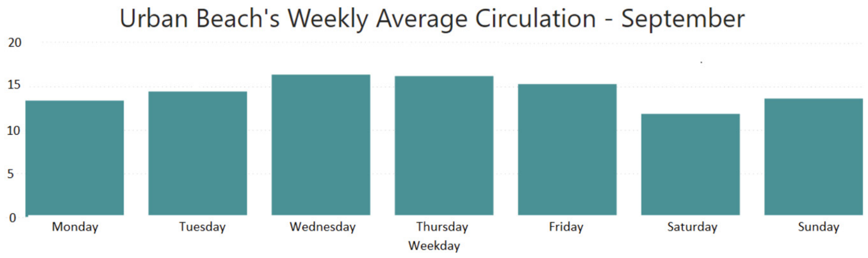

Fig. 5. Evolution of mobility in Urban Beach by weekday.

To continue this analysis and for comparison purposes, a graph was reproduced (Fig. 6), comprising the period between October and December. This graph aimed to get a better understanding of how the fluctuations of movements occurred throughout the hours. For this period, the results were practically opposite to those obtained for September. The period in which the busiest movements are registered in the Urban Beach nightclub is during the late-night period, revealing that the measures imposed by the Government to lift the restrictions had an effect that had immediate repercussions.

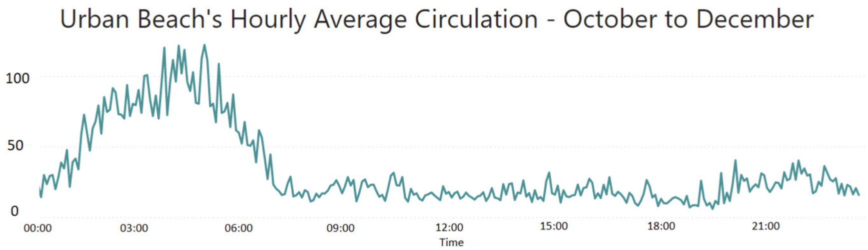

Fig. 6. Evolution of mobility in Urban Beach by hour, between October and December 2021.

For comparison purposes with the graph reproduced in Fig. 5, a visualization (Fig. 7) of the frequency with which people visited Urban Beach during the weeks in the observed months was also reproduced. This visualization made it possible to see that with the implemented deconfinement measures the pattern changed substantially, with the periods between Friday and Sunday becoming especially representative concerning the frequency of visits that the nightclub received.

As for Bairro Alto, it was possible to notice that during September it had different patterns from those that were verified in the Urban Beach nightclub. During the night periods, there was already some affluence of people in this area (Fig. 8), which resulted from several gatherings and parties that were illegally held during this period.

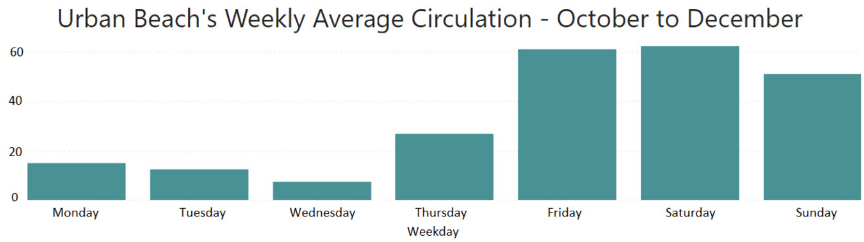

Fig. 7. Evolution of mobility in Urban Beach by weekday, between October and December 2021.

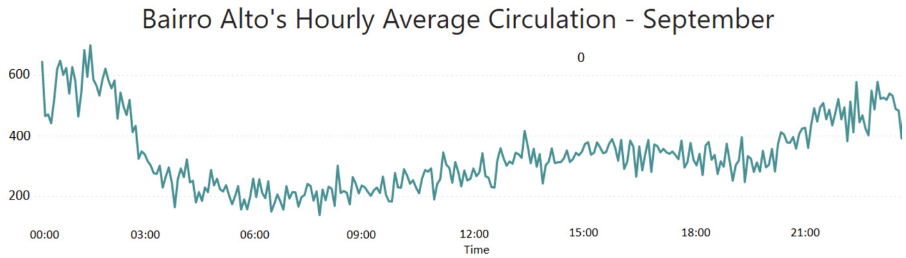

Fig. 8. Evolution of mobility in Bairro Alto by hours, in September 2021.

The weekly evolution of people's concentration at Bairro Alto (Fig. 9) was also under study. It was noticed that movement reaches its peak between Friday and Sunday, days that correspond to the periods when people are resting from their jobs and/or studies.

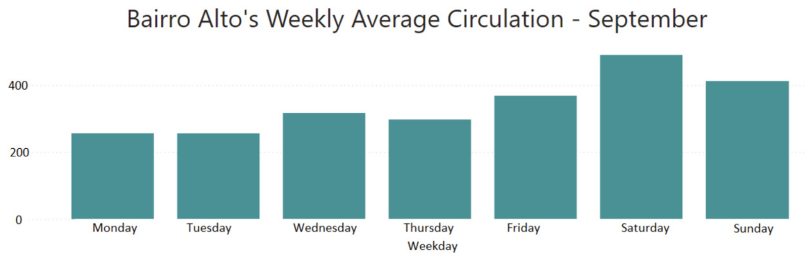

Fig. 9. Evolution of mobility in Bairro Alto by weekday, in September 2021.

For comparative purposes, it was visualized how the concentrations of people at Bairro Alto occur between October and December during 24h periods (Fig. 10) and how they vary throughout the week (Fig. 11).

Given the hourly variations, it was possible to verify that they exhibited very similar behaviour to September's, reaching the peaks of people circulation in the late-night period, with emphasis on the even higher average number of people that were registered in this zone during the period in question.

Finally, analysing the flow of people variations throughout the week it was again verified a similar trend to the one obtained in September, with the peaks of circulation

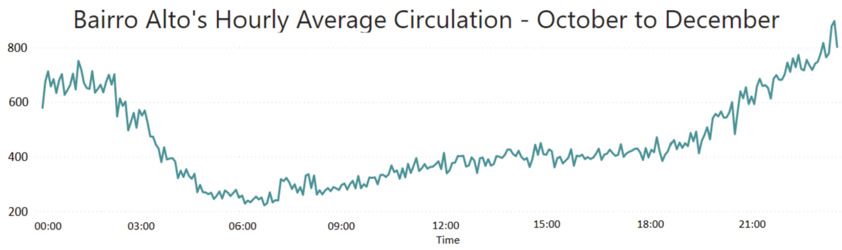

Fig. 10. Evolution of mobility in Bairro Alto by hour, between October and December 2021.

occurring again between Friday and Sunday. This similarity comes from the fact that, despite there being some measures to prevent nightclubs and bars to stay open during the night, these measures were bypassed by some people that went to illegal parties and concentrations.

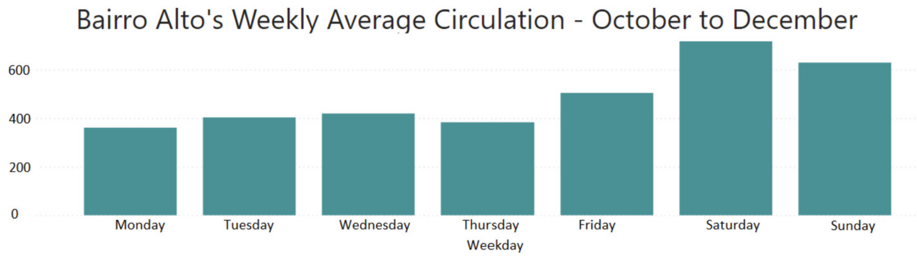

Fig. 11. Evolution of mobility in Bairro Alto by weekday, between October and December 2021.

To gain better insights and knowledge from the data studied, a visualization tool was developed in Microsoft Power BI. This tool consists of a dynamic heatmap that allows observing nightlife activity patterns at specific times, days, weeks, or months and to observe the movement flows that happen through the nights in town. With recursion to this tool, it was once again possible to demonstrate the patterns of increased nightlife circulation and activity that happened after September.

In Fig. 12 it is easily observed that between October and December the nightlife activity gained a new focus that is located near Avenida 24 de Julho, an area where there are several nightlife establishments such as bars, discos and restaurants. This recovery is due, once again, to the deconfinement measure that allowed bars and discos to reopen to the public and operate during the early hours of the morning.

In addition to this pattern, it was possible to see (in Fig. 13) that between 2 AM and 4 AM the concentration of people in areas near the river (again, where several discos and bars are located) tended to increase considerably.

Finally, it was also analysed the set of days - between December 24th and January 9th - on which a period of containment dictated the closure of discos and bars. Through an analysis of the visualization in Fig. 14, it is possible to see that the pattern of behaviour and nocturnal activity changes again in this set of days when compared with the general set of days under study. On this set of days, the riverside areas of the city show a lower

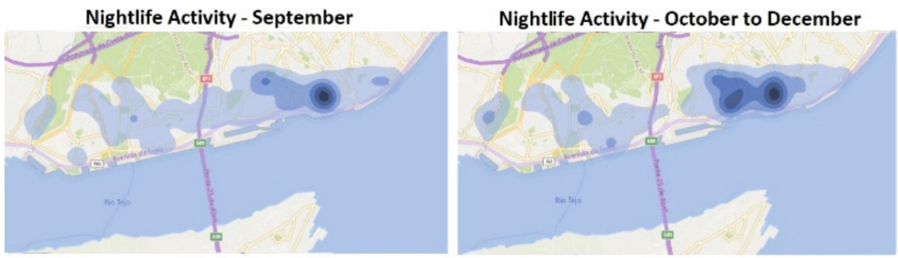

Fig. 12. Heatmap of evolution of mobility in nightlife activity in Lisbon (Sep x Oct-Dec).

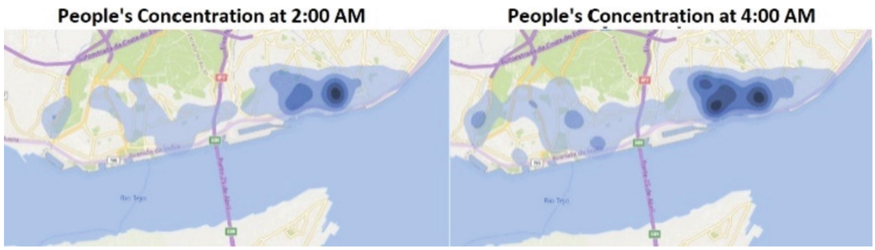

Fig. 13. Heatmap of evolution of mobility in nightlife activity in Lisbon (2 AM x 4 AM).

degree of concentration of people which increases considerably on the days following this period.

Fig. 14. Heatmap of evolution of mobility in nightlife activity in Lisbon.

Despite the conclusions drawn, it is undeniable that they are not enough to explain all the impacts of the COVID-19 pandemic on the mobility of the inhabitants of the county of Lisbon. For this reason, it was considered necessary to study with the same level of detail the movements of the population, during the day, in an area concentrating on other types of activities and infrastructures. In this way, it will be possible to complement the results obtained previously.

Evolution of Mobility in Parque Das Nações

Taking into account the size of the county of Lisbon and the volume of data associated with it, it would not be computationally feasible to study people's movements throughout the entire county, so the analysis will be much less in-depth according to the size of the database used. After some deliberation, the possibility of delimiting the analysis of mobility in an area of the city that concentrates a high volume of people daily was considered. This creative process led to the selection of a representative area of the parish of Parque das Nações which is illustrated in red on Fig. 15.

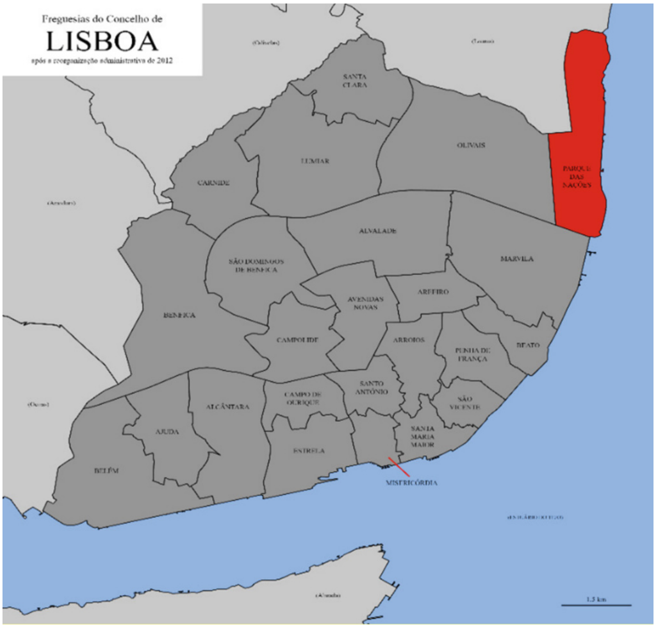

Fig. 15. Representative area – Parque das Nações (red zone). (Color figure online)

After the well-known International Exhibition - Expo in 1998, this area of the municipality of Lisbon, where various infrastructures associated with industrial activity used to be concentrated, such as refineries and rubbish dumps, underwent intense urban rehabilitation, gaining new prominence and becoming one of the busiest areas of the municipality.

The construction of the Gare do Oriente and the Vasco da Gama bridge played a very important role in the rehabilitation of this area since it started to connect this area of the municipality to the metro, train and bus networks, as well as making the connection to the other side of the river. An extremely high-quality urban space was created, with plenty of services and the integration of the latest technologies in the infrastructures. As a result, the number of companies that have their offices in this area has been growing over the last few years. On the other hand, the new residential buildings that have been built are among the most coveted by the Portuguese due to their proximity to diversified support services such as commerce and restaurants, schools and public and leisure spaces of high quality, and also cultural spaces such as the Pavilhão do Conhecimento.

Within the parish of Parque das Nações, the cells corresponding to the coastal area and the main streets from the Lisbon Oceanarium to the pavilions of the Lisbon International Fair were selected. The criterion for the delimitation of the study area was made through data analysis and explorations made presently by the group members in several weeks, thus ensuring that only areas with relevant infrastructures and significant movements were selected.

In this way, it was possible to make a very detailed mapping of the area that could facilitate the subsequent analysis of the mobility of the population. As each cell corresponds to an area of 200 by 200 m, resulting in an area large enough to cover several infrastructures. During the analysis of the site, it was possible to see that some residential buildings concentrated commercial or catering activities on the ground floors, which could alter our visualisation of the results.

Fig. 16. Parque das Nações divided by zones.

Taking these factors into consideration, a variable was created that classifies each cell studied according to the type of activity that is predominant within the delimited area. After a few days of work on site the area presented on Fig. 16 was mapped.

After completion of the data mapping and classification process, we proceeded to analyse the data regarding the effects of the measures taken by the government to combat the COVID-19 pandemic on people's mobility.

In a first stage, it was studied how the different "zones" previously created could help in the analysis of the mobility of the population. The number of devices was counted, during the 5 months under study, according to the type of zone in order to understand how each "zone" contributes to the concentration of people.

The impossibility of counting the devices in each "cell" only once meant that an alternative method had to be used to reduce as much as possible the margin of error in the visualizations that would be made. For this reason, the value recorded in each cell during a day is equal to the sum of all devices counted in the 5-min periods in which data collection occurred, divided by the 24 h that make up the day. In this way, an estimated value of the number of devices that have been counted is obtained that is closer to reality.

This approach is not perfect, however after several meetings with the Municipality of Lisbon, an agreement was reached. It was recognised that, with the limited resources available to us, this would be the best option to obtain results that would meet their expectations, assuming each device corresponds to one person.

Once all these treatment processes were completed, the data were analysed in greater depth. First, the number of devices counted per month from September 2021 to January 2022 was analysed to understand how the number of devices varied as the pandemic developed (Fig. 17).

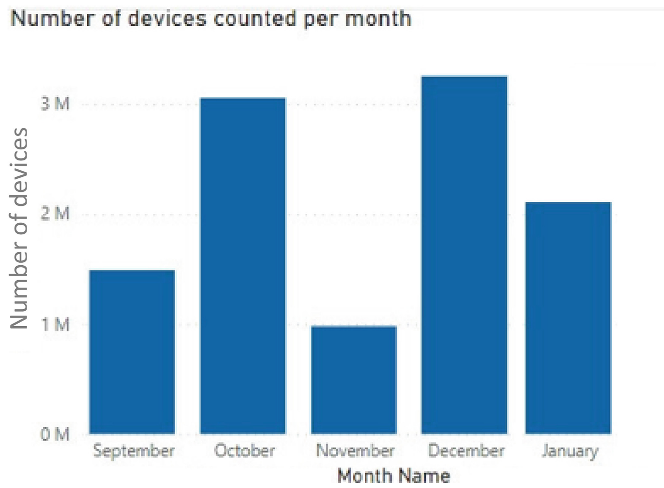

Fig. 17. Number of Devices counted per month.

By analysing Fig. 18, it was possible to see that there is an uneven distribution in the number of devices counted over the months. Although a lower value in the number of devices was expected in January 2022 due to the teleworking and confinement measures announced in that month by the government, the values for September and November were subject to further analysis.

Taking into account the study variables present in the database, the variation in the number of devices counted over each month was analysed, with special attention to September and November (Fig. 17 and Fig. 18).

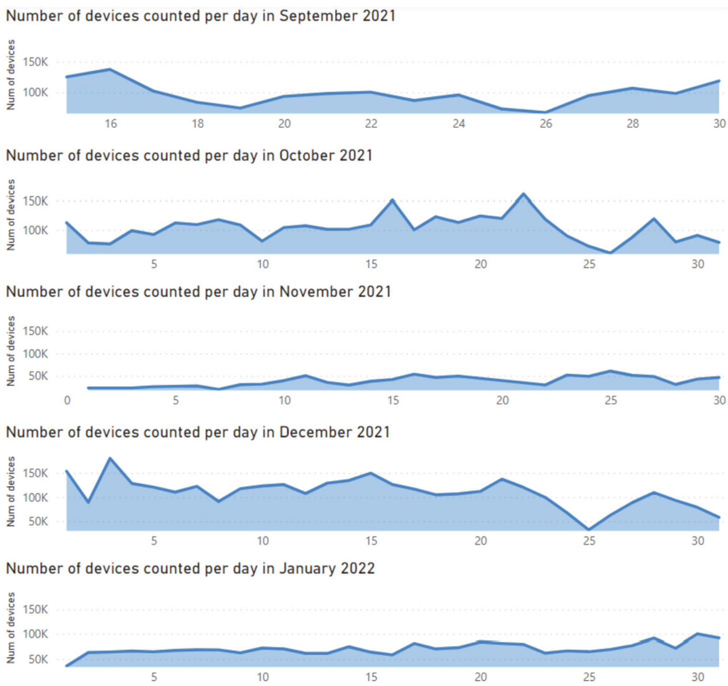

Fig. 18. Number of devices counted by day, per month.

Using the visual elements created in Microsoft Power BI it was possible to verify that, in September, devices were only accounted for from the second half of the month. This reduction in the time window can be pointed out as the main cause of the discrepancy between the values recorded in that month and the following one.

However, the reduced number of devices counted in November cannot be justified by the previous reason as devices were counted during the whole month. Even so, the values remained suspicious as during that time of the year several events were organised at Parque das Nações, some of which were of an international nature, such as the Web Summit, which attracts thousands of companies and start-ups to this part of the county every year.

In order to ascertain the true cause of the November values, the concentration of devices in each of the cells during that month was analysed to check the possibility of a problem having occurred in some parts of the study area, Fig. 19.

Considering the results obtained in the previous heatmap it was verified that most of the cells in Parque das Nações did not exist, which prevented the counting of devices. The municipality of Lisbon is composed by 3999 cells and during the month of November data was only collected in the first 3000 cells.

This situation was discussed with the Lisbon City Council who explained that the data existed, however, at the time the project was carried out, the data was corrupted, and it was not possible to present a deadline for it to be treated and delivered for study. After some conversations, and taking into consideration the problems raised previously,

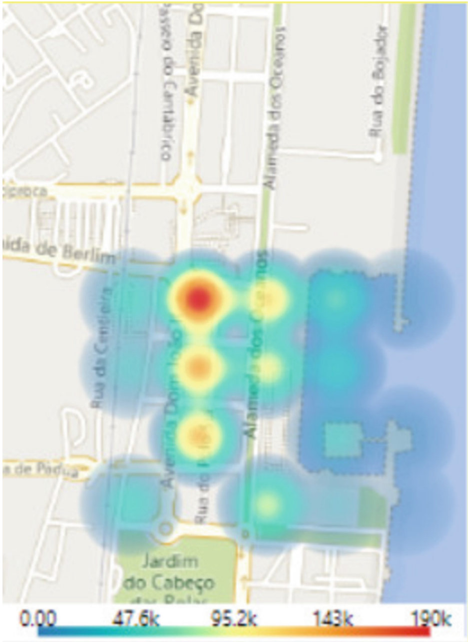

Fig. 19. Heatmap of Parque das Nações in November.

it was agreed that the study of mobility in Parque das Nações would be more focused in the months of October, December and January.

During the month of October, the only measure declared by the Directorate General of Health that would have some impact on the mobility of the Portuguese was the end of the limitations for people in closed spaces or events and the conclusion of the mandatory status of telework (Resolution of the Council of Ministers no. 135-A/2021). For this reason, more movements were recorded at Parque das Nações, given that several companies were able to choose between the permanent face-to-face regime and the hybrid regime, having more employees in their offices.

As a result of these measures, 3 055 047 people were counted during that month (Fig. 20). Taking into consideration the previous graph, it is possible to verify that most of the people that went to Parque das Nações during that month were employees from the several companies that are based in that office zone.

The second highest number of people was recorded in the residential zone, since in addition to residents, people passing through establishments located on the ground floors of residential buildings were also counted. A practical example are the cafes and small shops that during the day attract some people who are not exclusively residents in that area.

With the lifting of restrictions on the limit of people in enclosed spaces and the holding of fairs and exhibitions in spaces such as the Feira Internacional de Lisboa and Pavilhão do Conhecimento, the cultural zones counted more people than the Vasco da Gama shopping centre.

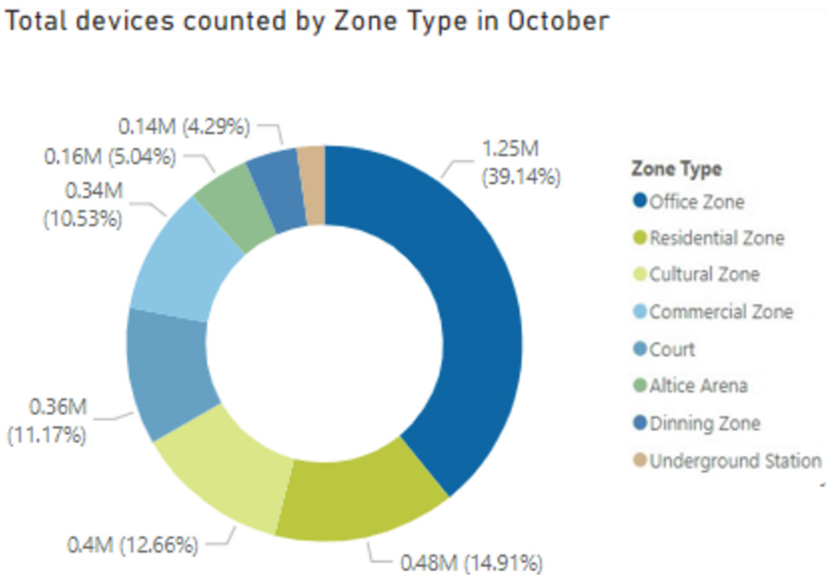

Fig. 20. Total Devices counted divided by zone in October.

During the analysis of the number of devices counted throughout the month it was verified if there was any pattern according to the type of zone a peculiar pattern was detected. When isolating the area corresponding to the Parque das Nações Restaurant Zone it is possible to verify the existence of a peak in the number of devices between days 21 and 22, followed by a significant drop until day 26 where the value rises again substantially (Fig. 21).

These fluctuations are the result of different days on which workers are paid, the civil service generally receives their salary on the 20th of the month and private companies usually pay their employees between the 24th and 26th. It is possible to conclude that on the days before workers are paid, fewer people go out to eat in restaurants, but after receiving their salary on the following days, they try to do so.

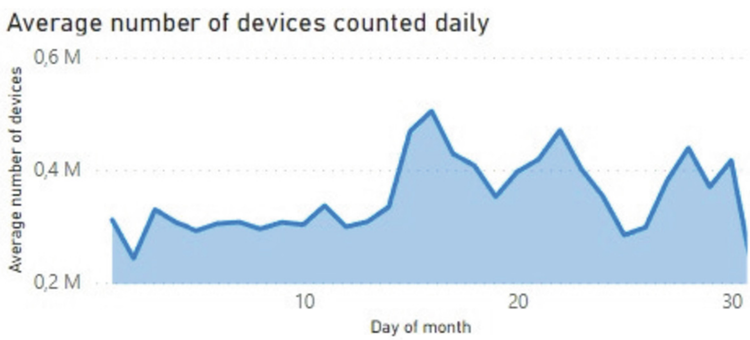

Fig. 21. Daily Count of devices.

As a complement to the previous results, a heatmap of that month was created in which it was possible to detect that the cells that registered more people were located in Avenida D. João II, corresponding to the shopping center, the offices of Vodafone, Axians, Sony Portugal and the Justice Campus. It is important to note that the concentration was quite high at the Justice Campus because only two cells included the courts and the Registration and Notary Centre, which concentrate a high volume of people daily (Fig. 22).

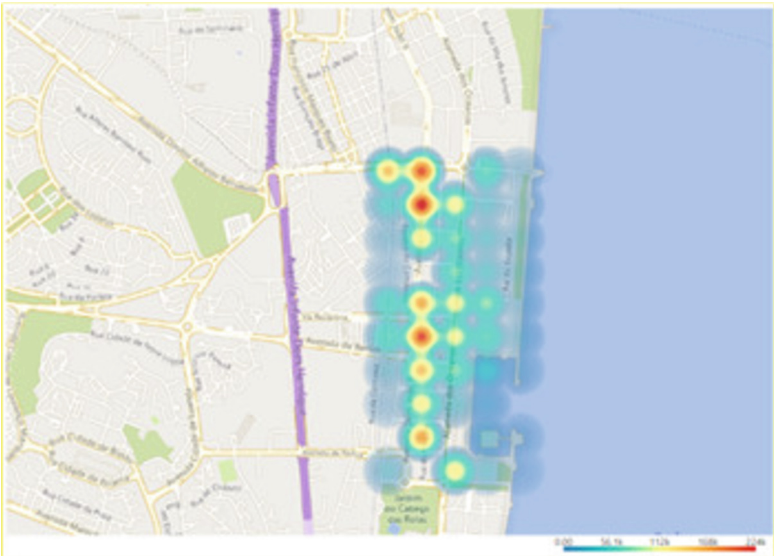

Fig. 22. Heatmap from October.

During the month of November to the beginning of December the number of cases of COVID-19 suffered a very significant increase to the point that the Directorate General of Health had to take more demanding measures. Among the various measures, it was highlighted the obligation to present a negative test to enter some spaces, the need to have the full vaccination certificate to visit restaurants and events and the obligation of teleworking between 25 December and 9 January.

For these reasons an analysis was made that directly compares the values registered between the months of December and January as more restrictive measures were implemented that may cause a more significant impact on the mobility of people in Lisbon.

Considering the graphs above, it is possible to verify a significant difference in the number of devices accounted for in December and January. In December, in addition to a greater concentration of people, there is also a substantial drop in the number of devices accounted for on the 25th (Fig. 23), which may be explained by the festive period celebrated on this day. Like the October analysis, the highest concentration of people is in the office zone, followed again by the residential zone.

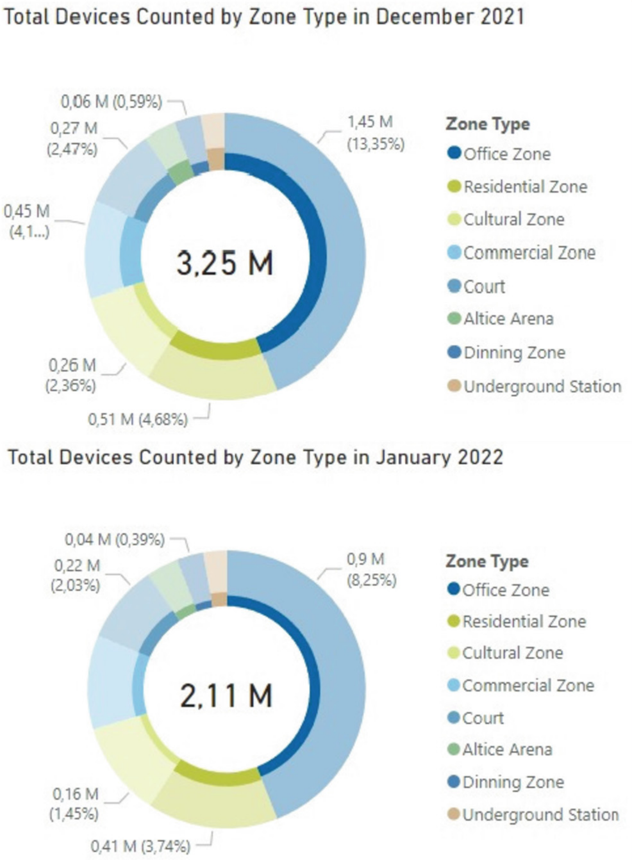

Fig. 23. Total Devices by zone Type between December and January.

After the 9th of January, there is a slight increase of devices accounted for until the end of the month. This result can be explained by the sales period that started on the 10th and by the easing of measures that took place throughout the month.

After observing the December heatmap (Fig. 25) and considering the analysis of the number of devices counted throughout the month (Fig. 24), there was a large affluence in the Vasco da Gama Shopping Centre area. This result was expected, since in the weeks before Christmas, the flow of people increases substantially in shopping areas for the festive season.

In the meetings that took place with the Lisbon City Council to monitor the project, the various results were presented through the Microsoft Power BI business analysis service, since it allows the provision of interactive visualizations with a simple interface. Throughout these meetings, this entity showed interest in this tool because it made it possible to filter data quickly. The question was raised whether it would be possible to implement this type of analysis on other types of data, using the same initial treatment. To answer this problem, a visualization tool with interactive dashboards was developed. This

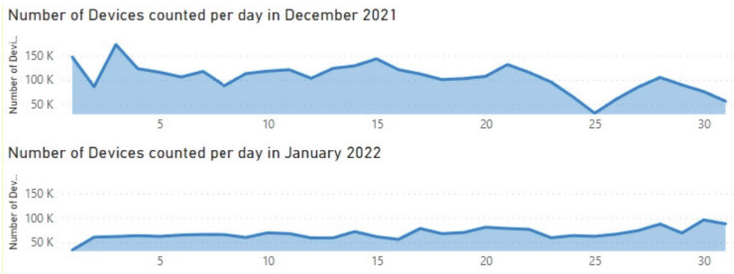

Fig. 24. Number of devices counted daily between December and January.

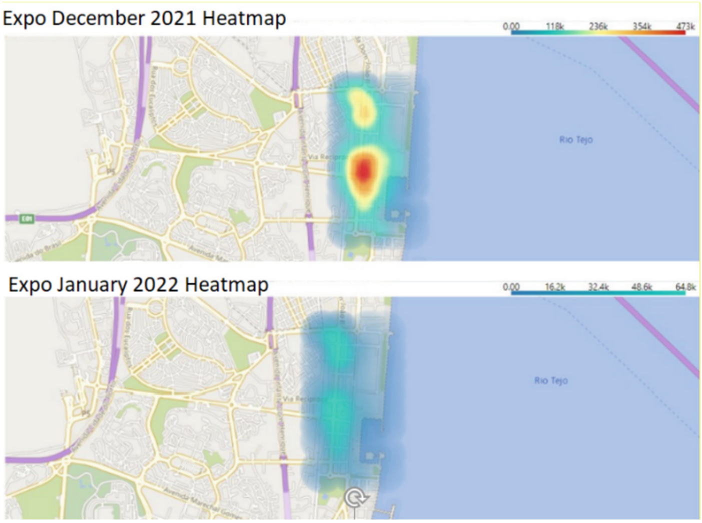

Fig. 25. Heatmaps from December and January.

allows staff, even if they have no experience with this software, to navigate intuitively and gain insights into large concentrations of people, as well as explore nightlife activity in the city.

4 Conclusions

Smart city planners and destination managers need to comprehend how people travel from one site to another. Possibilities for the development of relevant, evidence-based insights for decision-makers have been provided by the abundance of data supplied by social networking platforms. While prior research have offered observational data analysis techniques for social media data, there is still a need for method development - especially for capturing the movement patterns and behavioral aspects of individuals.

This research that outlines a novel way for analyzing people's activities, behaviors, and movements for monitoring and planning reasons. Our strategy employs information from mobile operators that establish a partnership with local municipality in Lisbon providing anomized data and we create a process to work with these data towards big pictures visualization process using Microsoft Power BI. This visualization allows city municipality identify big persons concentration, explore night life activity that is always a problem in a city and provide useful information how people move.

Funding. This work was supported by EEA Grants Blue Growth Programme (Call #5). Project PT-INNOVATION-0069–Fish2Fork. This research also received funding from ERAMUS+ project NEMM with grant 101083048.

References

1. Elvas, L.B., Marreiros, C.F., Dinis, J.M., Pereira, M.C., Martins, A.L., Ferreira, J.C.: Data-driven approach for incident management in a smart city. Appl. Sci. Switz. **10**(22), 1–18 (2020). https://doi.org/10.3390/app10228281
2. Mesquitela, J., Elvas, L.B., Ferreira, J.C., Nunes, L.: Data analytics process over road accidents data—a case study of Lisbon City. ISPRS Int. J. Geo-Inf. **11**(2), 143 (2022). https://doi.org/10.3390/ijgi11020143
3. Elvas, L.B., Mataloto, B.M., Martins, A.L., Ferreira, J.C.: Disaster management in smart cities. Smart Cities 4(2), Art. no. 2 (2021). https://doi.org/10.3390/smartcities4020042
4. Elvas, L.B., Gonçalves, S.P., Ferreira, J.C., Madureira, A.: Data fusion and visualization towards city disaster management: Lisbon case study. EAI Endorsed Trans. Smart Cities **6**(18), e3–e3 (2022). https://doi.org/10.4108/eetsc.v6i18.1374
5. Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G.: Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ **339**, b2535 (2009). https://doi.org/10.1136/bmj.b2535
6. 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 (2021). https://doi.org/10.1007/s11116-020-10108-w
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 (2022). https://doi.org/10.1007/s11042-021-10993-y
8. 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
9. 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
10. Ali Haidery, S., Ullah, H., Khan, N.U., Fatima, K., Rizvi, S.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), 837 (2020). https://doi.org/10.3390/electronics9050837
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), 1–21 (2021). https://doi.org/10.1145/3461645
12. Li, C., Hu, J., Dai, Z., Fan Z., Wu, Z.: Understanding individual mobility pattern and portrait depiction based on mobile phone data. ISPRS Int. J. Geo-Inf. 9(11), 666 (2020). https://doi.org/10.3390/ijgi9110666

13. Keawsomnuk, P.: 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. **21**(1), 35–42 (2021). https://doi.org/10.14456/hasss.2021.4
14. Leow, N.X., Krishnaswamy, J.: Smart cities need environmental consciousness and more social responsibilities as an outcome of COVID-19 – reflections from urban road commuters. Foresight **24**(2), 276–296 (2022). https://doi.org/10.1108/FS-02-2021-0035
15. Martins, T.G., Lago, N., Santana, E.F., Telea, A., Kon, F., de Souza, H.A.: Using bundling to visualize multivariate urban mobility structure patterns in the São Paulo Metropolitan Area. J. Internet Serv. Appl. 12(1), 1–32 (2021). https://doi.org/10.1186/s13174-021-00136-9
16. Wirth, R., Hipp, J.: CRISP-DM: towards a standard process model for data mining. In: Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining (2000)