Understanding Spatiotemporal Station and Trip Activity Patterns in the Lisbon Bike-Sharing System

Vitória Albuquerque[1], Francisco Andrade[2], João Carlos Ferreira[2(✉)], and Miguel Sales Dias[1,2]

[1] NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Campus de Campolide, 1070-312 Lisboa, Portugal
{d20190115, jmdias}@novaims.unl.pt
[2] Instituto Universitário de Lisboa (ISCTE-IUL), ISTAR, 1649-026 Lisbon, Portugal
{fafaell, jcafa}@iscte-iul.pt

Abstract. The development of the Internet of Things and mobile technology is connecting people and cities and generating large volumes of geolocated and space-time data. This paper identifies patterns in the Lisbon GIRA bike-sharing system (BSS), by analyzing the spatiotemporal distribution of travel distance, speed and duration, and correlating with environmental factors, such as weather conditions. Through cluster analysis the paper finds novel insights in origin-destination BSS stations, regarding spatial patterns and usage frequency. Such findings can inform decision makers and BSS operators towards service optimization, aiming at improving the Lisbon GIRA network planning in the framework of multimodal urban mobility.

Keywords: Bike-sharing system · Mobility patterns · Statistical analysis · Cluster analysis · K-means · Urban mobility

1 Introduction

Cities are becoming more predominant in modern societies, and citizens mobility is a raising problem concerning pollution and traffic. To overcome such challenges, shared mobility approaches have been developed. In this domain, bike-sharing is a rising soft and active transportation modality, showing large growth rates around the world. In face of such demand, the number of bike-share companies operating in the world has increased also, becoming more effective and available in most developed cities. Therefore, citizens are shifting towards more sustainable urban transportation where bike-sharing is increasingly being adopted. Understanding how people use the bike and when, is thus mandatory, towards improving system efficiency.

In 2017, Lisbon implemented a fourth-generation Bike-Sharing System (BSS), which is currently expanding, under currently enforced development plans by the City Hall. Taking Lisbon as a use case, we have adopted a data mining approach to understand station and trip patterns in its GIRA bike-sharing system. To this aim, we

A. L. Martins et al. (Eds.): INTSYS 2020, LNICST 364, pp. 16–34, 2021.
https://doi.org/10.1007/978-3-030-71454-3_2

have analyzed GIRA bike data and environmental data, to derive the spatiotemporal distribution of travel distances, speed and durations and their relationship with environmental conditions, such as weather.

1.1 Historical Background

In 2017 Lisbon implemented its first bike-sharing system, GIRA. Over a year, it expanded and in 2018 there were already 140 bike stations across the city, among which 92 in the central area of the city, 27 downtown and riverfront, 15 at Parque das Nações and 6 at Avenida Fontes Pereira de Melo and Avenida da Liberdade. At that time the total available bikes were 1,410, with 940 electric. Currently, the GIRA bike-sharing system has future expansion plans, since bike-sharing is one an important strategy in the context of urban mobility policies approved by the city hall, towards achieving intelligent and sustainable mobility in Lisbon.

The deployed system includes a data collection feature, allowing to monitor spatiotemporal users' behavior and trip patterns. By analyzing such collected data, we can gain new insights about mobility in the urban fabric, specifically on real-world bicycle-sharing system usage behaviors. Additionally, monitoring and analyzing user behavior changes, provides a broader scenario of the Lisbon public transportation network, giving new opportunities and patterns for prediction and usability improvement.

1.2 Our Research Approach

Our approach started by posing the following research question: "What are the spatiotemporal station and trip activity patterns of GIRA, the Lisbon BSS, in 2018?" This question statement leads us to derive the following sub-questions: "What are the average figures of monthly and daily BSS use?"; "What is the bike trip relation to weather conditions, specifically to precipitation and temperature?"; "How can we group the BSS origin and destination stations, into clusters across the city?

To address these questions, we have applied statistical and machine learning methods, based in our literature review of the state of the art. We have looked at historical data of bike trips (approximately 700,000 records), Portuguese Institute of Sea and Weather – Instituto Português do Mar e Atmosfera (IPMA) data, and cycling network data of 2018, with a focus on finding usage patterns, towards service optimization.

The paper is structured as follows: Sect. 2 presents our survey of State of the Art. In Sect. 3, we introduce our methodology, which adopted state of the art methods. In Sect. 4, Major Findings, we discuss our results, with a comparative analysis with other cities and identify a few research gaps and limitations of our research. Finally, Sect. 5, we raise some conclusions and draw lines for further research.

2 State of the Art of Bike-Sharing Systems

The community agrees that BSS improve urban accessibility and sustainability, and thus more cities in the world are implementing BSS to tackle urban mobility and pollution problems. Since 2016 over 1000 BSS are running in 60 countries [1].

From its third-generation, BSS start using smart card technology [2], producing station-based and trip-level data, and facilitating studies that enable the adoption of these systems into urban transportation networks [3]. Evolving fourth-generation BSS provides key data on users' behavior and trip patterns.

Monitoring makes possible the identification of system performance and data analysis provides insights into users' behavior [4] enabling to balance bike demand and improve bike network resilience and response.

The latest bike-sharing systems technology [5] uses two configurations: a fixed number of bike stations to hire and return bikes and a free placement scheme. Bike stations can be monitored in real-time on online maps. Application Programming Interfaces (API), accessing the network usage data are supplied by operators and specified to be used by external software developers. In Europe, such access is governed by the GDPR – General Data Protection Regulation [6], enforced since 2018, which includes provisions for personal data privacy and protection, including data anonymization. This scheme produces usage datasets, of critical importance in transport research [5].

O'Brien [5] first analyzed 38 bike-sharing systems in Europe, the Middle East, Asia, Australasia and the Americas, identifying behavior patterns. Metrics were applied to classify bicycle sharing systems, based on non-spatial and spatial location attributes and temporal usage statistics, plus a qualitative classification. The study proposed applications such as demographic analysis and the role of operator redistribution activity.

BSSs have been studied over time by other authors, with important insights about station and bike trip patterns analysis.

One of the most sophisticated BSS in the world is deployed in Copenhagen, reaching a ratio of 557,920 inhabitants for 650,000 bikes, with 48,000 bike stations and 429 km of cycle lanes [7]. It is estimated that overall, 1,27 million km are travelled daily with 5 times more bicycles entering the city than cars, resulting in 4/5 access to bicycles.

Vélov, Lyon (France) bike-sharing system, studied by Jensen [8], analyzed 11,6 million journeys and visualized bike flows on map. Characteristics, such as peak usage in a strike as well as different work speeds, highest peak hours, where analyzed. The authors observed that the highest speed occurred in the morning peak.

The London BSS network (Santander Cycles) is also expanding. In 2016 it reached 11,000 bicycles for 8,416,535 inhabitants, with 750 bike stations, 402,199 km travelled daily and 131,000 bicycle trips. London BSS station data, analyzed by Lathia [9] and Jensen [8], observed usage peaks and significant weekday and weekend differences. Spatial clusters with distinctive structures were found grouping intra-day usage patterns.

Studies show that longer BSS trips are observed in larger cities such as Chicago [10] and New York, although the latter differs between weekday and weekend usage [11].

Caulfield [12] findings showed that the majority of trips of BSS in medium size cities were short and frequent trips. Weather conditions also had an important impact, meaning that good weather conditions corresponded to an increase of trips.

El-Assi [13] analyzed the variation of trip activity along the season, month, week and hour, establishing correlations between these variables. The authors found a positive correlation with temperature calculated for each season.

Other studies showed that morning and afternoon peak patterns are different in BSS.

Han study [14] on San Francisco spatial-temporal bike trip patterns showed that in the hourly metrics analysis, most of the trips were between 8:00–9:00 am and 5:00–6:00 pm, meaning that most users use bikes to commute to work. El-Assi [13] found similar results in Toronto BSS regarding day peaks.

On the other hand, in Montreal BIXI BSS [15], peaks occurs in the evening and weekends.

Clustering algorithms studies on BSS data are applied by combining temporal and spatial attributes variables. More specifically, three clustering algorithms, namely, hierarchical clustering [10, 16, 17], community detection clustering [10, 18], and K-means clustering, [18–21], are the most common.

According to Caggiani [18], who analyzed the performance of the three clustering algorithms, K-means has been proven to be the best clustering algorithm to detect and rebalance bike-sharing usage patterns.

3 Data Mining Methodology

Our data analysis and visualization were performed in Python [22] using Jupyter Notebook platform [23]. Data cleaning, preprocessing, analysis and visualization were performed using different libraries according to the purpose of the application. “Numpy” [24], “Pandas” [25], “Matplotlib” [26], “Seaborn” [27] were used for statistical analysis and visualization. “GDAL” [28], “Shapely” [29], “Folium” [30], “Fiona” [31] were used to visualize spatial analysis. Our data science algorithms used “Scikit-learn” [32] to perform K-Means, Naïve Bayes, Train-test split and Accuracy Score.

In our approach, we have adopted the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology (see Fig. 1). This method is structured in three phases that are organized in a sequence: firstly, data collection, secondly data cleaning and thirdly data mining. Chapman [33] is convinced that CRISP-DM ensures the quality of knowledge discovery in the project results, requires reduced skills for such knowledge discovery, and with reduced costs and time. Data collection consists in collecting and perceiving the characteristics of the collected data, to meet the users and business needs, understanding where the data comes from and what type of analysis can be done with it. With Data Cleaning, we remove noise in the data, so that further analysis cannot be affected by the data itself. Data Mining allows the application of statistical and/or machine learning techniques enabling discovery of behaviors that could not be possible to observe before. It also includes data visualization, with diagrams, plots and other graphical depictions, that show us visually, the found patterns and behaviors.

Our datasets included bike trip data with trip characteristics, and bike stations data holding information about the network of bike stations throughout the city. To investigate the built environment correlation with trips, we’ve used precipitation and temperature datasets for this analysis.

Two datasets were generated for our analysis: one combining precipitation and temperature data and bike trips data (see schemas in Table 2 and Table 3), and another combining bike trips data and bike station data (see schemas in Table 1 and Table 2), with the goal to generate bike paths in the city and to visualize the stations chosen by the users. The first dataset was joined through a temporal basis and the second one was joined via the stations field. To generate these datasets, we've developed an Extract, Transform and Load process (ETL), to load the external databases, transform them by creating common columns and joining the datasets, and finally by loading them into our project. I've performed an adaptation of the ETL methodology proposed in CRISP-DM. Our ETL was used in the Data Cleaning phase, as it performs some cleaning and conforming processes in the incoming data, to obtain data which is correct, complete, consistent, accurate and unambiguous [34].

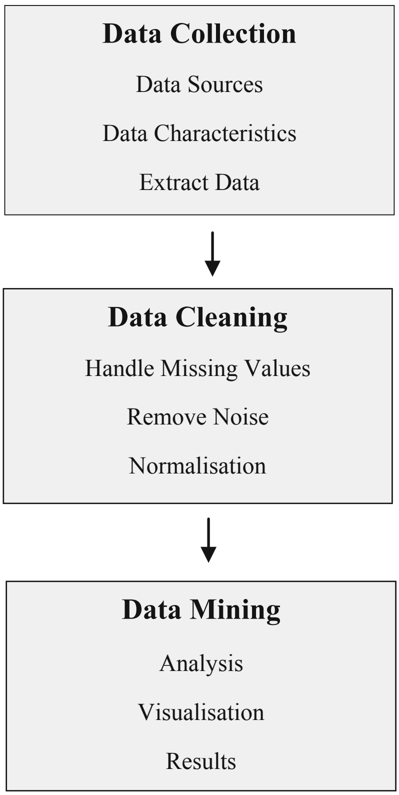

Fig. 1. CRISP-DM Methodology.

3.1 Data Sources and Data Characteristics

Three different sources of data provided by the Lisbon City Hall and the Portuguese Institute of Sea and Weather – Instituto Português do Mar e Atmosfera (IPMA), were used in our research: bike station data of 2018, bike trip data (from 25th January 2018 to 15th October 2018) and IPMA data of 2018.

Bike station data schema includes information about the stations, such as commercial designation ID (desigcomercial), entity ID (entity_id), planning ID (id_planeamento), latitude, longitude and the station capacity (capacidade_docas). This data was collected in 76 bicycle stations around Lisbon.

Table 1. Bike station data schema

Characteristics	Description
desigcomercial	Commercial designation
entity_id	Entity ID
id_planeamento	Planning ID
latitude	Latitude
longitude	Longitude
capacidade_docas	Station capacity

Table 2. Bike trip data schema

Characteristics	Description
id	Column ID
date_start	Start date and time
date_end	End date and time
distance	distance
station_start	Start station ID
station_end	End station ID
Bike_rfid	Bike ID
geom	Travel trajectory geometry
num_vertices	Number of nodes
Tipo_bicicleta	Bike type (conventional or electric)

Bike trip data of 2018, is characterized by origin-destination (O-D) trip that includes id (column ID), date_start (start date and time), date_end (end date and time), distance (distance in metres), station_start (start station ID), station_end (end station ID), bike_rfid (bike ID), geom (geometry), num_vertices (number of nodes), and tipo_bicicleta (bike_type).

The IPMA weather data of 2018 consists in the total precipitation in 2018, and its schema includes the fields ANO (Year), MS (Month), DI (Day), HR (Hour), “1200535”, “1200579” and “1210762”. These 3 last fields represent the reference (ID) of the 3 weather stations located in Lisbon, giving information about the precipitation, where “1200535” is Lisboa Geofísica (Lisbon centre), “1200579” is Lisboa Avenida Gago Coutinho and “1210762” is Lisboa Tapada da Ajuda.

Table 3. IPMA data schema

Characteristics	Description
ANO	Year
MS	Month
DI	Day
HR	Hour
1200535	Lisboa Geofísica Weather Station #1
1200579	Lisboa Avenida Gago Coutinho Weather Station #2
1210762	Lisboa Tapada da Ajuda Weather Station #3

From the 3 data sources whose schemas where presented in Tables 1, 2 and 3, we derived, via an ETL process, 2 datasets, namely, the "bike temporal analysis dataset" and the "bike trips-stations dataset", used in our data mining approach.

3.2 Data Cleaning

EMEL GIRA data, a fourth generation BSS, provides broad and extensive information. Data extraction methods have not yet been extensively explored [35], therefore there are limitations in the collected data, which needs to be evaluated on its limitations and cleaned, if appropriate. Data cleaning involves processes of handling missing data and noise removal, to generate datasets with accurate and validated data. In the GIRA dataset, we have found incoherent data, namely sparse, discontinuities and nonuniformities of data. On the contrary, bike station data and IPMA data did not require data cleaning and were ready to use.

The following data cleaning methods, where applied to bike trip data:

- We have removed the not assigned (NA) values of the bike type (1% of the dataset).
- We have removed the geometry and number of nodes which had NA values, corresponding to 50% of the data.
- The variable speed was removed due to the trips that were shorter than 1 min.
- The missing values in the distance were filled by computing the average speed times the duration.

After data cleaning, the total number of trips using the GIRA BSS in 2018 was 684,471. In that year, the average number of trips per month, ranging from January to October, was 68,447. In terms of stations, the average number of trips was 9,126 throughout year. Per day, there was an average number of trips of 2,602, starting from January 25th until October 15th.

3.3 Data Mining

Literature studies have tried to understand user's profile and travel behavior [36–38], activity patterns of bike stations [9] and the impact of the built environment in the BSS [39].

The methods applied are statistical methods to analyze data and visualization techniques. To understand bike trip patterns in the urban mobility network and trip models, studies have shown the importance to correlate transport mode and trip choices and built environment characteristics [40, 41].

Many methods are applied to perform data mining namely, to examine the relations between bike stations, bike trips and built environment. The evaluation of BSS success depends in these relationships, most of them leading to users' access to the bike stations [42].

Clustering algorithms combining temporal and spatial attributes variables are also data mining methods used for this analysis purpose. More specifically K-means clustering [18–21], used by McKenzie [43] and Zhong [44] to measure regularity at different scales and to measure spatiotemporal variation and cluster interaction.

3.3.1 Bike Usage Analysis

To investigate the monthly bicycle usage frequency, we have merged the "bike trip dataset" with the "bike temporal basis dataset" and got a new relation with columns ANO (Year), MÊS (Month), DIA (Day), FERIADO (Holiday), SEMANA (Week), SEMESTRE (Semester), TRIMESTRE (Trimester), DIA_DE_SEMANA (Weekday) and MÊS_DSC (Month description). This was our trips schema, with data spanning from January to October from 2018. In the Summer months (June, July, August and September), there were a total of 439,176 trips being the more concentrated period (64% of all trips), as depicted in Fig. 2.

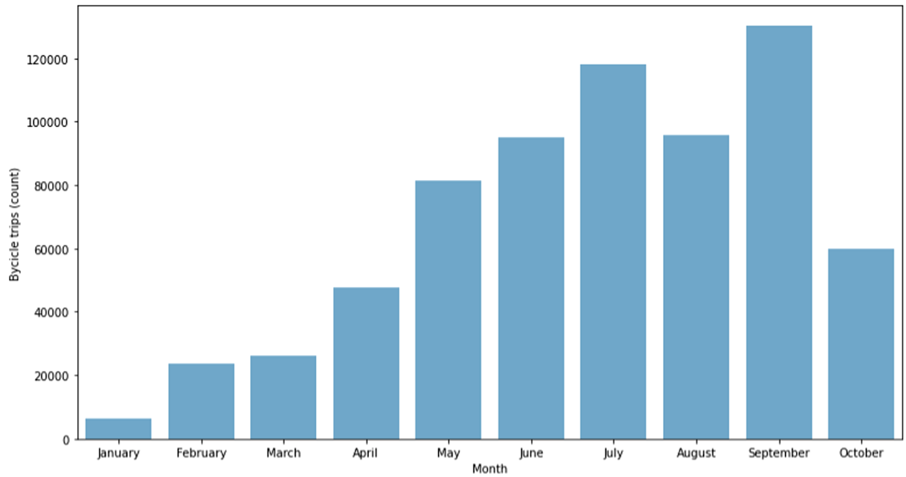

Fig. 2. Bicycle usage frequency per month.

The weekday and weekend usage were also analyzed to understand the preferences of using the bike-sharing service during the week. Results are presented in Fig. 3, where weekdays are ordered from 1 to 7. The weekend is represented by 1 (Sunday) and 7 (Saturday). Our results show that most users (82%) prefer to use the service during the week, rather than during the weekend.

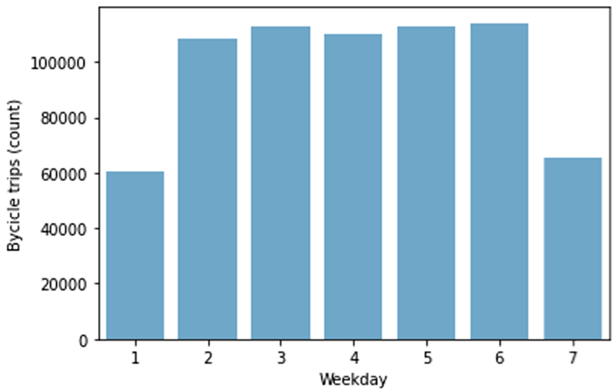

Fig. 3. Bicycle usage per weekday.

The distribution of trips throughout the different time periods of the day was analyzed too. The column date_starts was transformed into a time format and the hour was extracted in order to create the column Periodo_dia (Day period). The day was broken down into three-hour groups: Morning: 7:00am to 12:00am; Afternoon: 12:00am– 20:00 pm and Overnight: 20:00 pm–7:00am. Our analysis shows that most of the trips (56%) occur during the afternoon, when comparing with the morning and overnight periods (see Fig. 4). Additionally, during working weekdays, after the afternoon, the morning period comes second. In the weekends, users still prefer to ride during the afternoon, but overnight rides come second, rather than morning ones.

When analyzing the behavior and patterns regarding the distance and the duration of the bicycle trips, we addressed the differences between the weekdays versus bicycle type. Regarding bicycle type (Electric or Conventional), we have observed no noticeable differences in terms of trip distance and duration, during weekdays. There also no noticeable difference, in terms of speed and duration, across the different days of the week, in average.

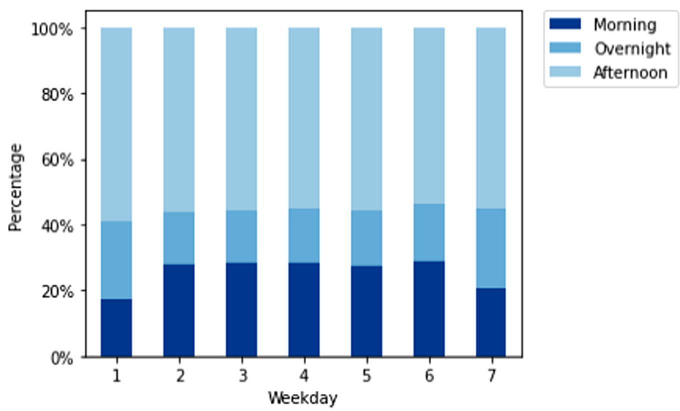

Fig. 4. Bicycle usage (%) per weekday within the day period.

3.3.2 Station Usage Analysis

Our analysis shows that the most used start stations are located in the main axis of the city. The top 6 start stations (see Fig. 5) are 105 (CC Vasco da Gama), 307 (Marquês de Pombal), 417 (Avenida Duque de Ávila), 421 (Alameda D. Afonso Henriques), 446 (Avenida da República/Interface de Entrecampos) and 481 (Campo Grande/Museu da Cidade).

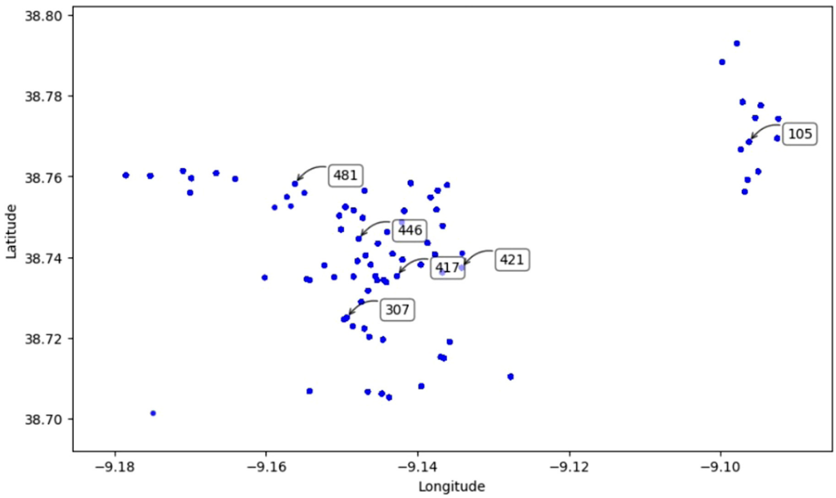

Fig. 5. Bike trip start stations (top 6)

A heatmap was created with "Folium" [30] Python package with built-in OpenStreetMap [45] tileset to visualize patterns of bike trip most used start stations (see Fig. 6). A bike dataset with the latitude and longitude start of each station was used for this purpose. As observed in Fig. 6, the major concentration of start stations demand is located in the Lisbon center.

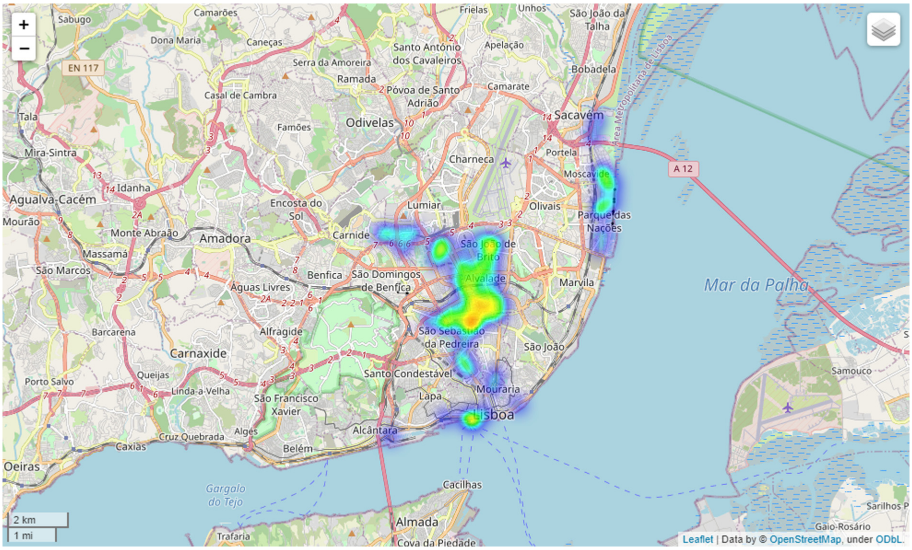

Fig. 6. Bike trip start station heatmap. The red color corresponds to a higher concentration of bike trip start stations, whereas purple, to a lower. (Color figure online)

Regarding end stations of bike trips, we found that the most used ones, are located in the main axis of the city (see Fig. 7), namely, stations 105 (CC Vasco da Gama), 403 (Avenida Fontes Pereira de Melo), 417 (Avenida Duque de Ávila), 421 (Alameda D. Afonso Henriques), 446 (Avenida da República/Interface de Entrecampos) and 481 (Campo Grande/Museu da Cidade).

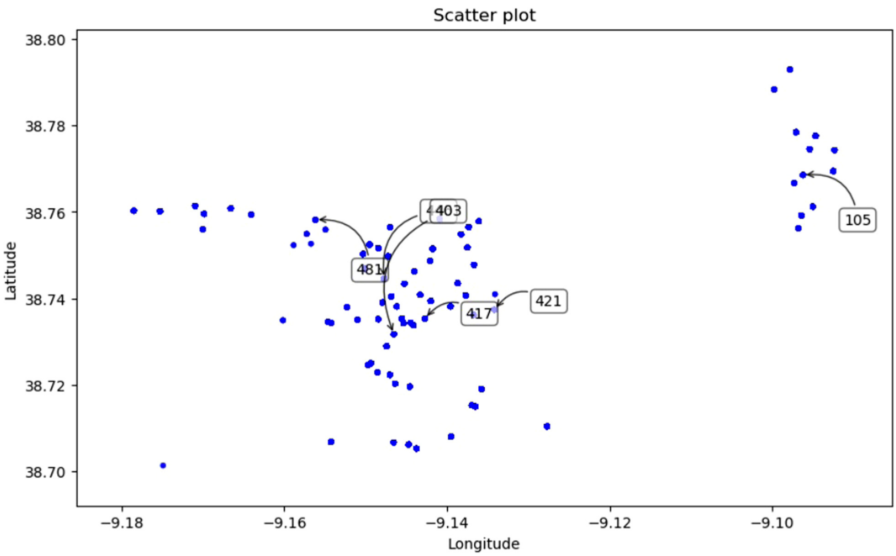

Fig. 7. Bike trip end stations (top 6)

3.3.3 Bike Trip Weather Analysis

We conducted an additional analysis aiming at finding behavior patterns between the users of BSS and the built environment variables, particularly, weather variables such as atmospheric precipitation and temperature. For the analysis in terms of atmospheric precipitation, we created a Boolean variable "rain" indicating if it was raining or not, in any of the three weather stations. To join the two datasets, we created a new date_key field from the date_start field of the bicycle trips. From our analysis, we can conclude that the trips are mostly done when there is no precipitation (97%) (see Fig. 8.). Regarding temperature analysis, the negative values were removed, and we calculated the average values of the three stations. Then, we divided the dataset into four

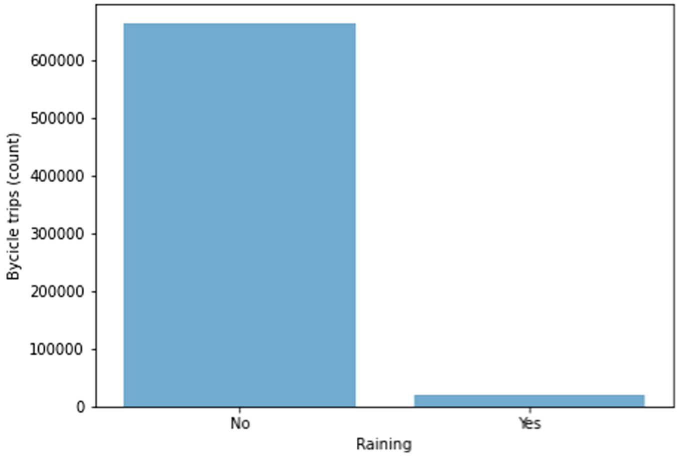

Fig. 8. Bicycle usage frequency relation to atmospheric precipitation.

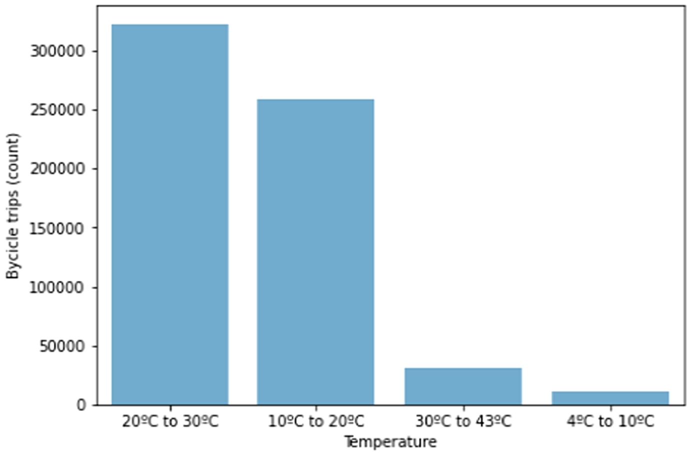

Fig. 9. Bicycle usage frequency relation to temperature.

categories: 0º to 10º, 10º to 20º, 20º to 30 and 30º to 42º, being 42º the maximum observed temperature value (see Fig. 9).

The trip speed was also analyzed in order to check if there was any observed change when raining, concluding that users are faster in their trips when it was not raining.

3.3.4 Spatial Cluster Analysis

In our research we seek to understand the behaviors of BSS users, particularly the movement of the bicycles in terms of the starting and ending of a trip, in each station, as well as and the frequency of usage of each station, across the available data. For that purpose, we performed a clustering analysis on the "bike trips dataset". Initially our study focused in identifying geographical (WGS84) patterns throughout the city of Lisbon in terms of trips starting and ending in each station. K-means was used to produce geographic clustering. An additional data pre-processing step was needed, before applying K-means. To get the clusters of trips and stations, we found the need to get all the trips from all the stations, irrespectively if a given station is a trip start or trip end station. To this aim, we split the original "bike trips dataset" in two, one having the station_start variable and the other having the station_end variable. Then, the variables name station_start and station_end are changed to "station" in the corresponding datasets. Both datasets are afterwards concatenated within the variable station. After, we computed the count of the trips in each station. The final result was a dataset with six variables: station, number of trips, station designation, latitude, longitude and dock capacity.

When applying K-means, we used the Elbow algorithm [46] to find the optimal K number through the calculation of the SSE (Sum of Squared Errors). The algorithm found four spatial clusters of the stations where bike trips start and end (see Fig. 8): one representing the center of Lisbon going from Alvalade to Saldanha (Purple), a second one representing the east side of Lisbon with just a few stations going from Telheiras to Cidade Universitária (Yellow), a third one representing the lower part of Lisbon going from Marquês de Pombal to Baixa (Green), and a forth one representing the Parque das Nações (Blue). Table 4 shows the coordinates of the center of each cluster for the geographic clustering generated by K-means (Fig. 10).

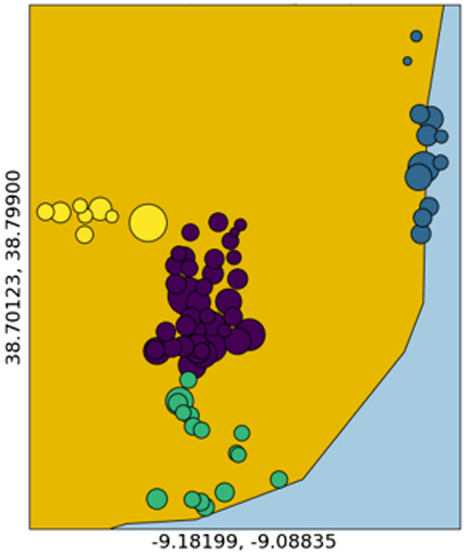

Fig. 10. Spatial clustering of stations where bike trips start and/or end (Yellow: Telheiras-Campo Grande (Museu da Cidade); Purple: Alvalade-Saldanha; Green: Marquês de Pombal-Baixa; Blue: Parque das Nações) (Color figure online)

Table 4. Cluster centroids

Latitude	Longitude
38.743263	−9.144271
38.772288	−9.095947
38.715984	−9.143659
38.759463	−9.168919

A second analysis was focused on station usage clustering. For that purpose, a pre-processing step required the creation of a new field n_trips (Number of trips per station), representing how many trips occurred on that specific station. Then we applied also K-means, with the same type of approach to find the optimal K number, as in the prior geographical cluster analysis. This time K-means was applied to find the main spatial clusters across the city, in terms of the number of trips that start and/or end in each station. We found four of such clusters (see Fig. 11). We can observe also that four of the most frequently used stations (labeled in green) are located in the center of the city, while a fifth one lies in the east side.

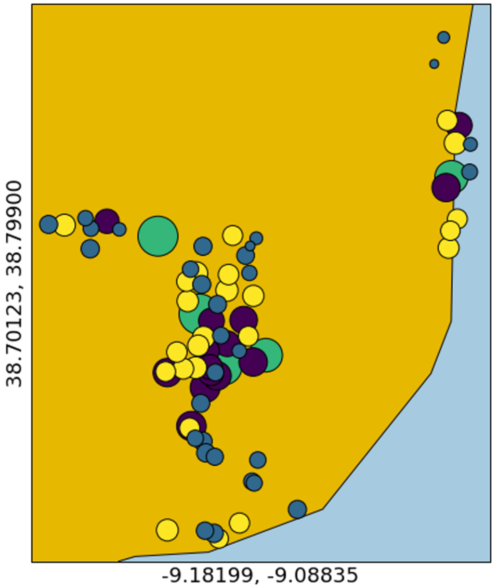

Fig. 11. Stations clustering by the number of trips that start and/or end on a given station (Green: 1st most used stations; Purple: 2nd most used stations; Yellow: 3rd most used stations; Blue: 4th most used stations) (Color figure online)

4 Major Findings

Our analysis shows that the total number of trips of EMEL GIRA BSS, from January 15th to October 25th, 2018 was 684,471. The average number of trips per month was 68,447, while the average station number of trips figure was 9,126. Moreover, we found that the daily average number of trips was 2,602.

The analysis shows also that the months of June, July, August and September had the most concentration of trips during 2018, with 439,176 trips, representing 64% of all trips. Regarding the day of the week, we have observed that users choose working weekdays to travel in the city (82%) compared to the weekend. Also, it is possible to affirm that the users prefer to use the service on working weekdays than during weekends.

Our findings also show that most of the trips happen, during 2018, in the afternoon (56%), followed by the morning period. We have also observed that during weekdays, the users prefer to ride during the afternoon but in the weekend, users mostly bike overnight.

Most used O-D stations were observed in two axis: one from Campo Grande to Marquês de Pombal and another in Parque das Nações, showing that bike demand start and end stations are located in Lisbon office areas. There are four major concentrations in the city for the number of trips. The main areas where users unlock GIRA bikes belong to Parque das Nações, the city center – Alvalade, Avenidas Novas, Santa Maria Maior - meaning that the center of Lisbon is where the most bike trips happen. There is also a close relation of the number of trips with the station capacity, and the station cluster with more trips is associated with the stations with the greater capacity.

We have also found that built environment factors such as precipitation, affect GIRA BSS usage, showing that almost 97% of trips take place when there is no rain. This observation was complemented by a correlation with speed analysis showing that higher speed is reached when there is no precipitation. Regarding temperature, it is possible to observe that most users prefer to travel when temperatures are between 20º and 30º (52%). There is also a significant number of users cycling when the temperature is between 10º and 20º (42%).

Finally, we also found that there was no significant difference, in terms of speed and duration of bike trips across the different days of the week, in average, nor of bicycle type (Electric or Conventional). Therefore, our research suggest that the type of bike is not a decisive factor when it comes to analyze bike trips.

5 Conclusions

This paper provides new insights on the recent implemented GIRA BSS in Lisbon. Overall, it was interesting to observe a strong use of BSS in a city that did not have a cycling culture, until recently.

Major findings show that most GIRA BSS trips take place on working weekdays, in the afternoon, which suggest a usage pattern that correlates well with working-home commute practices. We have also observed that weather conditions [12, 13] had an important impact on travel behavior. No rain was consistent with an increase of ridership, and temperatures between 10º to 30º were consistent with such behaviors too.

Lisbon GIRA BSS trip patterns are thus similar to other observed BSS mobility patterns of medium-size cities [12] discussed in the State of the Art section, such as patterns found in short and frequent trips and ride peak observed, observed in a case study in the city of Cork (Ireland) [12].

Parallels with larger cities can be established as well. In Canada, for instance, Montreal's BIXI BSS [15] is mostly used in weekdays evenings and weekends. In Toronto, bike trips are shorter in the weekdays mornings [13].

Large American cities BSS studies [10, 11, 14] show a frequent bike use in the morning and afternoon peaks [10] and different usage patterns between weekdays and weekends, identifying longer trips in the weekend [10].

In European large cities, weekday morning trips in the peak hour [8, 44] reach higher speed than trips over the weekdays and weekends.

As for the Lisbon GIRA BSS there is a strong possibility of overtime change, as future BSS network expansion plans are implemented in the city in the coming years.

Further work needs to be conducted regarding GIRA BSS in the scope of urban analytics [47] and literature discussion and parallel comparison with other BSS implemented nationally and internationally.

Future work needs to be conducted regarding bike station managing models, prediction of potential network demand to improve network planning, optimization of stations and locations, bikes rebalancing operation overtime and integration of BSS with multimodal transport systems, in the context of the first and last mile.

Acknowledgements. This study was performed in the scope of Lisboa Inteligente LxDataLab [48] – Lisbon City Council Data Centre – challenges to the academia. Vitória Albuquerque and Francisco Andrade work were partially supported by Fundação para a Ciência e a Tecnologia (FCT), under the project "IoT for Smart Cities and R&D Projects on Smart {Cities, Buildings, Education and Health} Summer School 2019/2020", with a Research Grant supporting the following projects: UIDB/04466/2020 and UIDP/04466/2020 [49].

References

1. Meddin, R., DeMaio, P.: The Meddin Bike-sharing World Map. Google Maps. https://bikesharingworldmap.com/#/all/2.3/-1.57/33.92/. Accessed 04 Oct 2020
2. DeMaio, P.: Bike-sharing: history, impacts, models of provision, and future. J. Public Transp. **12**(4), 41–56 (2009)
3. Zhang, J., Pan, X., Li, M., Yu, P.: Bicycle-sharing system analysis and trip prediction. In: 2016 IEEE International Conference on Mobile Data Management, pp. 174–179. IEEE (2016)
4. Padgham, M.: Human movement is both diffusive and directed. PLoS ONE **7**, e37754 (2012)
5. O'Brien, O., Cheshire, J., Batty, M.: Mining bicycle sharing data for generating insights into sustainable transport systems. J. Transp. Geogr. **34**, 262–273 (2014)
6. General Data Protection Regulation. https://gdpr-info.eu/. Accessed 04 Oct 2020
7. Kaplan, S., Manca, F., Nielsen, T., Prato, C.: Intentions to use bike-sharing for holiday cycling: an application of the Theory of Planned Behavior. Tour. Manag. **47**, 34–46 (2015)
8. Jensen, P., Rouquier, J.-B., Ovtracht, N., Robardet, C.: Characterizing the speed and paths of shared bicycle use in Lyon. Transp. Res. Part D Transp. Environ. **15**, 522–524 (2010)
9. Lathia, N., Ahmed, S., Capra, L.: Measuring the impact of opening the London shared bicycle scheme to casual users. Transp. Res. Part C Emerg. Technol. **22**, 88–102 (2012)
10. Zhou, X.: Understanding spatiotemporal patterns of biking behavior by analyzing massive bike sharing data in Chicago. PLoS ONE **10**, e0137922 (2015)
11. Nolan, R., Smart, M., Guo, Z.: Bikeshare trip generation in New York City. Transp. Res. Part A Policy Pract. **94**, 164–181 (2016)
12. Caulfield, B., O'Mahony, M., Brazil, W., Weldon, P.: Examining usage patterns of a bike-sharing scheme in a medium sized city. Transp. Res. Part A Policy Pract. **100**, 152–161 (2017)
13. El-Assi, W., Salah Mahmoud, M., Nurul Habib, K.: Effects of built environment and weather on bike sharing demand: a station level analysis of commercial bike sharing in Toronto. Transportation **44**(3), 589–613 (2017). https://doi.org/10.1007/s11116-015-9669-z
14. Han, X., Wang, P., Gao, J., Shah, M., Ambulgekar, R., Jarandikar, A., Dhar, S.: Bike sharing data analytics for silicon valley in USA. In: 2017 IEEE SmartWorld Ubiquitous Intelligence and Computing, Advanced and Trusted Computed, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People and Smart City Innovation, SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI, pp. 1–9. IEEE (2018)
15. Faghih-Imani, A., Eluru, N., El-Geneidy, A., Rabbat, M., Haq, U.: How land-use and urban form impact bicycle flows: evidence from the bycicle sharing system (BIXI) in Montreal. Transp. Geogr. **41**, 306–314 (2014)
16. Sarkar, A., Lathia, N., Mascolo, C.: Comparing cities' cycling patterns using online shared bicycle maps. Transportation **42**(4), 541–559 (2015). https://doi.org/10.1007/s11116-015-9599-9

17. Du, Y., Deng, F., Liao, F.: A model framework for discovering the spatio-temporal usage patterns of public free-floating bike-sharing system. Transp. Res. Part C Emerg. Technol. **103**, 39–55 (2019)
18. Caggiani, L., Camporeale, R., Ottomanelli, M., Szeto, W.: A modeling framework for the dynamic management of free-floating bike-sharing systems. Transp. Res. Part C Emerg. Technol. **87**, 159–182 (2018)
19. Zhao, Y., Dai, L., Peng, L., Song, Y., Zhou, Z.: Analysis of spatial distribution of China's station-free bike-sharing by clustering algorithms. In: ACM International Conference Proceeding Series, pp. 15–19. ACM Press (2019)
20. Li, D., Zhao, Y., Li, Y.: Time-series representation and clustering approaches for sharing bike usage mining. IEEE Access **7**, 177856–177863 (2019)
21. Guo, Y., Shen, X., Ge, Q., Wang, L.: Station function discovery: exploring trip records in urban public bike-sharing system. IEEE Access **6**, 71060–71068 (2018)
22. Python. https://www.python.org/. Accessed 04 Oct 2020
23. Jupyter. https://jupyter.org/. Accessed 04 Oct 2020
24. Numpy. https://numpy.org/. Accessed 04 Oct 2020
25. Pandas. https://pandas.pydata.org/. Accessed 04 Oct 2020
26. Matplotlib. https://matplotlib.org/. Accessed 04 Oct 2020
27. Seaborn. https://seaborn.pydata.org/. Accessed 04 Oct 2020
28. GDAL. https://pypi.org/project/GDAL/. Accessed 04 Oct 2020
29. Shapely. https://pypi.org/project/Shapely/. Accessed 04 Oct 2020
30. Folium. https://pypi.org/project/folium/. Accessed 04 Oct 2020
31. Fiona. https://pypi.org/project/Fiona/. Accessed 04 Oct 2020
32. Scikit-learn. https://scikit-learn.org/stable/. Accessed 04 Oct 2020
33. Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Wirth, R.: CRISP-DM 1.0: Step-by-step data mining guide. Computer Science (2000)
34. El-Sappagh, S., Hendawi, A., Bastawassy, A.: A proposed model for data warehouse ETL processes. In: J. King Saud Univ. Comput. Inf. Sci. **23**, 91–104 (2011)
35. Zhang, Y., Thomas, T., Brussel, M., Van Maarseveen, M.: Expanding bicycle-sharing systems: lessons learnt from an analysis of usage. PLoS ONE **11**, e0168604 (2016)
36. Fuller, D., et al.: Use of a new public bicycle share program in Montreal Canada. Am. J. Prev. Med. **41**, 80–83 (2011)
37. Shaheen, S., Zhang, H., Martin, E., Guzman, S.: China's Hangzhou public bicycle. Transp. Res. Rec. J. Transp. Res. Board **2247**, 33–41 (2011)
38. Shaheen, S., Cohen, A., Martin, E.: Public bikesharing in North America: early operator understanding and emerging trends. Transp. Res. Rec. J. Transp. Res. Board **2387**, 83–92 (2013)
39. Buck, D., Buehler, R.: Bike lanes and other determinants of capital bikeshare trips. In: 91st Transportation Research Board Annual Meeting, pp. 703–706 (2012)
40. Ewing, R., Cervero, R.: Travel and the built environment. J. Am. Plann. Assoc. **76**, 265–294 (2010)
41. Kemperman, A., Timmermans, H., Timmerman, H.: Influences of built environment on walking and cycling by latent segments of aging population. Transp. Res. Rec. J. Transp. Res. Board **2134**, 1–9 (2009)
42. Liu, Z., Jia, X., Cheng, W.: Solving the last mile problem: ensure the success of public bicycle system in Beijing. Procedia Soc. Behav. Sci. **43**, 73–78 (2012)
43. McKenzie, G.: Docked vs. Dockless bike-sharing: contrasting spatiotemporal patterns. In: Leibniz International Proceedings in Informatics, vol. 114 (2018)

44. Zhong, C., Batty, M., Manley, E., Wang, J., Chen, F., Schmitt, G.: Variability in regularity: mining temporal mobility patterns in London, Singapore and Beijing using smart-card data. PLoS ONE **11**, e0149222 (2016)
45. OpenStreetMap. https://wiki.openstreetmap.org/wiki/Main_Page/. Accessed 04 Oct 2020
46. Yuan, C., Yang, H.: Research on K-value selection method of K-means clustering algorithm. J. Multidiscip. Sci. **2**, 226–235 (2019)
47. Batty, M.: Urban analytics defined. Environ. Plan. B Urban Anal. City Sci. **46**, 403–405 (2019)
48. LxDataLab – Laboratório de Dados Urbanos de Lisboa. https://lisboainteligente.cm-lisboa.pt/lxdatalab/. Accessed 03 Aug 2020
49. Summer School 2020 – IoT for Smart Cities. https://istar.iscte-iul.pt/summerschool2020/. Accessed 03 Aug 2020