

## Bike-sharing mobility patterns: a data-driven analysis for the city of Lisbon

Vitória Albuquerque<sup>1</sup>, Francisco Andrade<sup>2</sup>, João Carlos Ferreira<sup>2,3,\*</sup>, Miguel Sales Dias<sup>1,2</sup> and Fernando Bacao<sup>1</sup>

<sup>1</sup>NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Campus de Campolide, 1070-312 Lisboa, Portugal

<sup>2</sup>Instituto Universitário de Lisboa (ISCTE-IUL), ISTAR, 1649-026 Lisboa, Portugal

<sup>3</sup>Inov Inesc Inovação—Instituto de Novas Tecnologias, 1000-029 Lisbon, Portugal

### Abstract

New technologies applied to transportation services in the city, enable the shift to sustainable transportation modes making bike-sharing systems (BSS) more popular in the urban mobility scenario. This study focuses on understanding the spatiotemporal station and trip activity patterns in the Lisbon BSS, based in 2018 data taken as the baseline, and understand trip rate changes in such system, that happened in the following years of 2019 and 2020. Furthermore, our paper aims to understand the COVID-19 pandemic impact in BSS mobility patterns. In this paper, we analyzed large datasets adopting a CRISP-DM data mining method. By studying and identifying spatiotemporal distribution of trips through stations, combined with weather factors, we looked at BSS improvements more suitable to accommodate users' demand. Our major contribution was a new insight on how people move in the city using bikes, via a data science approach using BSS network usage data. Major findings show that most bike trips occur on weekdays, with no precipitation, and we observed a substantial growth of trip count, during the observed time frame, although cut short by the pandemic. We believe that our approach can be applied to any city with available urban mobility data.

**Keywords:** bike-sharing system, urban mobility patterns, statistical analysis, cluster analysis

Received on 22 April 2021, accepted on 03 May 2021, published on 04 May 2021

Copyright © 2021 Vitória Albuquerque *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [Creative Commons Attribution license](https://creativecommons.org/licenses/by/4.0/), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi: 10.4108/eai.4-5-2021.169580

\*Corresponding author. Email: [joao.carlos.ferreira@iscte-iul.pt](mailto:joao.carlos.ferreira@iscte-iul.pt)

### 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 active mobility modality, showing large growth rates worldwide. Such demand, increased the number of bike-sharing companies operating in the world, becoming more effective and available in most developed cities. Moreover, citizens are shifting towards more sustainable urban transports, such as bike-sharing,

increasingly adopted and becoming more popular. Hence, understanding how and when people use bike-sharing systems and their mobility patterns over time is thus mandatory towards improving the system's efficiency.

Aligned with OECD Sustainable Development Goal [1], [2] (SGD) 11 - Sustainable Cities and Communities, Portugal national [3], regional [4] and Lisbon [5] strategies for mobility, aim to integrate bike-sharing systems in the long-term public transport plans and daily commute.

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 and our preliminary study [6],

we have adopted a data mining approach to understand station and trip patterns in its GIRA BSS and understand this service from a perspective of evolution throughout the years. To this aim, we have analyzed GIRA BSS data and environmental data to derive the spatiotemporal distribution of travel distances, speed and durations and their relationship with environmental conditions, such as weather. Moreover, we analyzed the evolution of the GIRA BSS usage rate from 2018 to 2020 and the impact of COVID 19 pandemic.

## 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 81 bike stations across the city. There are future to expand Lisbon BSS, with more stations and bikes, since bike-sharing is an important strategy in the context of urban mobility policies approved by the City Hall to achieve intelligent and sustainable urban mobility in Lisbon. The deployed system includes a data collection feature that monitors spatiotemporal users' behavior and trip patterns. By analyzing such collected data, we can learn about urban mobility, specifically on real-world bike-sharing system usage behaviors. Additionally, monitoring and analyzing user behavior changes provides a broader Lisbon public transportation network scenario, giving new opportunities and patterns for prediction and usability improvement.

## 1.2. Our Research Approach

In this study, we aim to collect spatiotemporal bike trip data, with trip id, origin and destination stations, trajectory, and time to identify spatial and temporal patterns. Hence, to correlate bike mobility patterns with weather data and external events, such as COVID 19 pandemic that affected urban mobility in 2020.

Our approach addresses the following research questions:

RQ1: What are the spatiotemporal station and trip activity patterns in Lisbon BSS in 2018?

This question statement leads to the following sub-questions:

- What are the average figures for monthly and daily Lisbon BSS use?
- What is the bike trip relation to weather conditions, specifically, to precipitation and temperature?
- How can we group the Lisbon BSS origin and destination stations?
- How can we group Lisbon BSS into clusters across the city?

RQ2: Have Lisbon BSS trip patterns changed in 2019 and 2020 from 2018, given the COVID 19 pandemics?

This leads to the following sub-question:

- How has COVID 19 pandemic affected Lisbon BSS?

Three levels of analysis were performed to address these research questions: the first, bike trip and station usage in 2018, looking at historical data of bike trips (approximately 700,000 records), and Portuguese Institute of Sea and Weather – Instituto Português do Mar e Atmosfera (IPMA) data with focus on finding usage patterns, towards service optimization.

The second level regards 2019 and 2020 bike trips, monthly and weekday usage, in comparison with 2018 to investigate bike trip usage patterns over the three years. The third one regards the analysis of bike trip counts collected from 2019 to 2020 by a sensor in Avenida Duque de Ávila.

This paper is structured as follows: section 2 presents our State of the Art survey. Section 3 introduces our data mining methodology, which adopted Cross-Industry Standard Process for Data Mining (CRISP-DM). In section 4, Results, data understanding, data pre-processing and modeling are explained and presented with analysis and visualizations. In section 5, Discussion, we discuss our results with a comparative analysis and identify research gaps and limitations. Finally, in section 6, we raise some conclusions and draw lines for further research.

## 2. State of the Art of Bike-Sharing Systems

The community agrees that BSS improves urban accessibility and sustainability, and thus more cities in the world are implementing BSS to tackle urban mobility issues and pollution problems. Since 2016, over 1000 BSS were implemented in 60 countries [7].

From BSS third-generation smart card technology [8] was used, producing station-based and trip-level data and facilitating studies that enable the adoption of these systems in urban transportation networks [9]. Evolving fourth-generation BSS, provided additional key data on users' behavior and trip patterns.

Monitoring makes possible the identification of system performance and data analysis provides insights into users' behavior [10], enabling to balance bike demand and improve bike network resilience and response.

The latest bike-sharing systems technology [11] 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, is supplied by operators and specified by external software developers. In Europe, such access is governed by the GDPR – General Data Protection Regulation [12], enforced since 2018, including provisions for personal data privacy and protection, including data anonymization. This scheme produces usage datasets, of critical importance in transportation research [11].

O'Brien [11] first analyzed 38 bike-sharing systems in Europe, the Middle East, Asia, Australasia and the Americas, identifying behavior patterns. Metrics were applied to classify BSS based on non-spatial and spatial

location attributes and temporal usage statistics, and a qualitative classification. The study also proposed applications such as demographic analysis and the role of operator redistribution activity.

Other authors have studied BSSs mobility patterns resulting in insights about station and bike trip patterns analysis. One of the most sophisticated BSS in the world is in Copenhagen, reaching a ratio of 557,920 inhabitants for 650,000 bikes, with 48,000 bike stations and 429 km of cycle lanes [13]. Overall, it is estimated that 1,27 million km are travelled daily with 5 times more bicycles entering the city than cars, resulting in 4/5 access to bikes.

Jensen [14] studied VéloV BSS, in Lyon (France) and analyzed 11,6 million journeys resulting in a visualization map. Characteristics, such as peak hour, peak usage and speed in commute were analyzed, showing that the highest speed occurs 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 [15] and Jensen [14], 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 [16] and New York, although the latter differs between weekday and weekend usage [17].

Caulfield [18] findings showed that most BSS trips in medium-sized cities were short and frequent trips. Weather conditions also had an important impact, meaning that good weather conditions corresponded to increased trips.

El-Assi [19] 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 [20] 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 [19] found similar results in Toronto BSS regarding day peaks.

On the other hand, in Montreal BIXI BSS [21], 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 are the most common, namely, hierarchical clustering [16], [22], [23], community detection clustering [16], [24], and K-means clustering, [24]–[27].

According to Caggiani [24], who analyzed the three clustering algorithms' performance, K-means has been proven to be the best algorithm to detect and rebalance bike-sharing usage patterns.

Our previous study [6] focused in 2018 Lisbon BSS data, showing that 64% of trips are done in June, July, August and September, and 82% of the trips were done in weekdays,

mainly in the peak hours (8-9 am and 6-7 pm). Regarding trip correlation with weather, it showed 97% usage in non-precipitation days and with temperature between 20° to 30°. Moreover, in the clusters analysis [6] we observed four clusters in Alvalade-Saldanha, Telheiras-Campo Grande, Marquês de Pombal-Baixa and Parque das Nações.

This study aims to take a step further and understand the evolution of Lisbon BSS mobility patterns from 2018 to 2020. Although, having only access to the total daily trip count, and not having data on stations or origin and destination trips we aim to understand how mobility patterns changed in 2019 and 2020, from 2018 results, aiming also to correlate with 2019 and 2020 weather data, and understand how COVID 19 pandemic affected Lisbon BSS usage.

### 3. Methodology

In our approach, we applied the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology (see Fig. 1). This method is structured in six phases, as follows:

Phase 1 - Business Understanding: understand and decide what to accomplish with data mining and setting criteria for the data mining aims.

Phase 2 - Data Understanding: data is collected and evaluated regarding data quality and suitability.

Phase 3 - Data Preparation: data pre-processing transforms the data into useful information used for the next phase. It involves cleaning, reduction, transformation, and integration of data.

Phase 4 - Modelling: modelling technique is selected and built the model.

Phase 5 - Evaluation: the chosen modelling technique is evaluated according to its objectives according to the results produced in the process

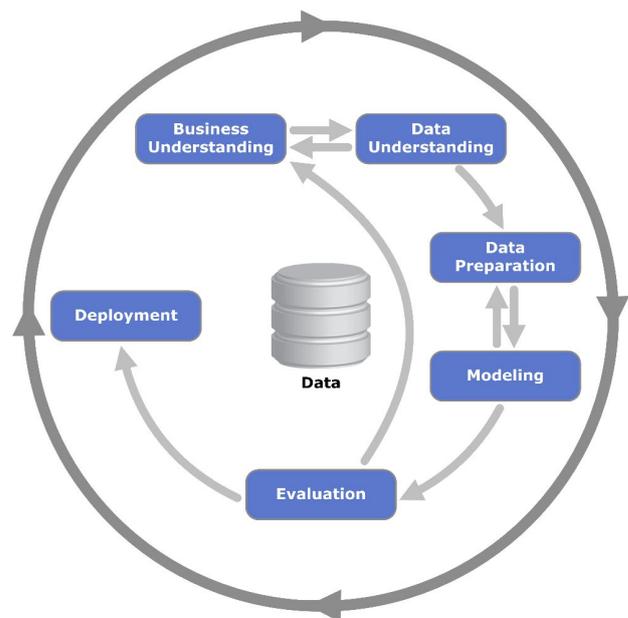


Figure 1. CRISP-DM Methodology

CRISP-DM ensures the quality of knowledge discovery in the project results [28], requires reduced skills for such knowledge discovery and reduced costs and time.

In phases 1 business understanding, we identified the objectives and framed the business issue (research questions), gathering information. In this phase we perceived the collected data's characteristics to meet the users and business needs.

In phase 2, data understanding, we investigated the collected data, understanding where the data comes from and what type of analysis could be done with it.

In data preprocessing phase, we performed data cleaning, removing noise in the data so that further analysis wouldn't be affected by the data itself.

We performed an adaptation of the ETL methodology proposed in CRISP-DM. Our ETL (see Fig. 2) was used in the data cleaning phase, performing cleaning, conformance and normalization processes in the data sets, to obtain correct, complete, consistent, accurate and unambiguous data [29].

The model phase, allowed the application of statistical and 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 visually show us the found patterns and behaviors.

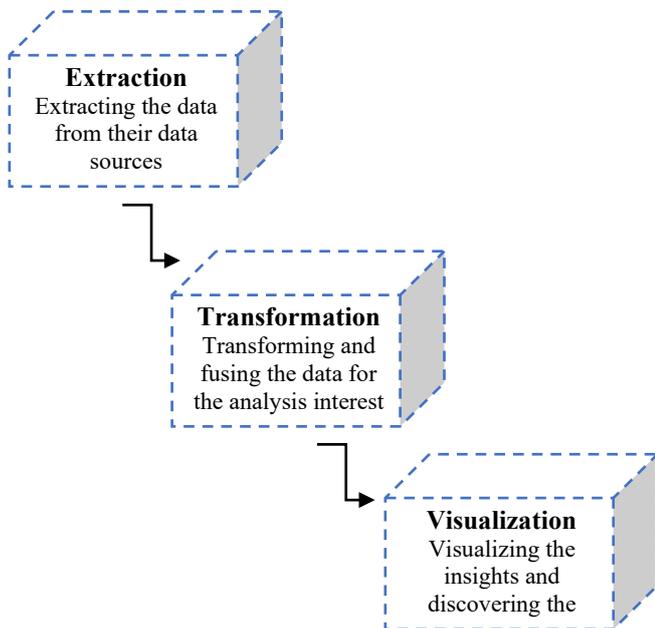


Figure 2. ETL Methodology

## 4. Mining Bike Sharing Data

In this section we apply CRISP-DM phases to our study, introducing first business and data understanding, looking at the aim and how to address the research questions and how to understand the different BSS and weather datasets. This is supported by data preprocessing, cleaning and

normalization, that provides new datasets to the model building phase, targeting the analysis and visualization of insights. Our datasets include bike trip data of 2018, 2019 and 2020, which were analysed according to the data characteristics in different levels, with the aim to understand the evolution of bike ridership and the impact of built environment and pandemics.

Our data analysis and visualization were performed in Python [30] using the Jupyter Notebook platform [31]. Data cleaning, preprocessing, analysis and visualization were performed using different Python libraries, according to the application's purpose. "Numpy" [32], "Pandas" [33], "Matplotlib" [34], "Seaborn" [35], were used for statistical analysis and visualization. "GDAL" [36], "Shapely" [37], "Folium" [38], "Fiona" [39], were used to visualize spatial analysis. Our data science algorithms used "Scikit-learn" [40] to perform K-Means, Train-test split and Accuracy Score.

### 4.1. Data Understanding

Data was provided in the scope of Lisboa Inteligente [41] challenges, namely challenges #4 "Are there mobility patterns in Lisbon BSS", and #49 "Determine COVID 19 pandemic impact in mobility and environment (<https://lisboainteligente.cm-lisboa.pt/lxdatalab/desafios/determinacao-do-impacte-da-pandemia-por-covid-19-na-mobilidade-e-ambiente/>).

Three levels of analysis were performed with the provided data: the first concerns the bike trip and station data from 2018, with a descriptive analysis regarding month, weekday, period of day and hourly usage rate of the service, following the geographical analysis of trips and stations and finally, a weather analysis. The second regards the 2019 and 2020 bike trips and the monthly and weekday usage rate comparison with 2018, to find and/or confirm behaviour patterns over the 3 analysed years. The third one regards the analysis of bike trip counts collected by a sensor in Avenida Duque de Ávila, a central avenue of Lisbon, from 2019 and 2020.

#### Lisbon BSS data of 2018

Different sources of data were provided by the Lisbon City Hall (CML), namely data on bike trips (from 25th January 2018 to 15th October 2018) and stations from the Mobility and Parking Company of Lisbon - Empresa de Mobilidade e Estacionamento de Lisboa (EMEL) and, weather data from the Portuguese Institute of Sea and Weather – Instituto Português do Mar e Atmosfera (IPMA).

Bike station data schema (Table 1) includes information about stations: commercial designation ID (desigcomercial), entity ID (entity\_id), planning ID (id\_planeamento), latitude, longitude and the station capacity (capacidade\_docas). This data was collected from 76 bike stations in Lisbon.

Table 1. Lisbon BSS 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 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

Bike trip data of 2018 (Table 2), is featured 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 (Table 3) provides the total precipitation of 3 weather stations (ID) located in Lisbon: “1200535” Lisboa Geofísica (Lisbon centre), “1200579” Lisboa Avenida Gago Coutinho and “1210762” Lisboa Tapada da Ajuda. It is structured with the following features: ANO (Year), MS (Month), DI (Day), HR (Hour).

Table 2. Lisbon BSS trip data schema from 2018

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)

### Lisbon BSS data of 2019 and 2020

Bike trip data of 2019 and 2020 (Table 4) is characterized by date (dd/mm/yyyy) and trips per day ranging from 1st January 2019 to 4th June 2020.

Table 4. Lisbon BSS trip data schema of 2019 and 2020

Characteristics	Description
data	Date
viagens	Bike trip count

The IPMA weather data (Table 5) is structured with 18 variables in 2019, and 11 variables (Table 5) in 2020. The variables marked with \* are only provided for 2019 data, and the others both to 2019 and 2020. The data ranges from 1st January of 2019 to 30th October 2019, and from 17th January 2020 to 30th June 2020. It is important to highlight that data is missing 9th, 10th and 24th March 2019; 18th and 19th April 2019; 22nd September to 30th September; November and December 2019 and the first two weeks of 2020 (1st January to 16th January). In our analysis we used features, such as, date, weather station code (the codes are the same as 2018 with a new code, “1210783” corresponding to the Alvalade Weather Station) and the temperature levels.

Table 5. IPMA data schema of 2019 and 2020

Characteristics	Description
data_hora	Date
entity_id*	Entity ID
entity_location*	Entity Location (Coordinates)
entity_ts*	
entity_type*	Type of entity
estaciones	Weather Station code
fecha	
fiware_service*	
fiware_servicepath*	
humidade	Humidity
iddireccvento	Wind direction ID
intensidadeventokm	Wind intensity
position	Station position
preacumulacada	
pressao	Atmospheric pressure
radiacao	Radiation
temperatura	Temperature
validity_ts*	

#### Avenida Duque de Ávila data of 2019 and 2020

Bike count data of 2019 and 2020 (Table 6) was collected from a sensor located in Avenida Duque de Ávila. Data provided features all trips count, from BSSs bikes and bikes owned by users. It was collected from 1st January 2019 to 1st October 2020 and is structured as follows: “Time”: entry of day, month and year (dd/mm/yyyy); “Piloto Lx” total of bike count (east and west) per day; “Piloto Lx Ciclistas Entradas”: bike count per day from the east; and “Piloto Lx Ciclistas Entradas”: bike count per day from the west.

Table 6. Avenida Duque de Ávila bike count of 2019 and 2020

Characteristics	Description
Time	Date
Piloto Lx	Avenida Duque de Ávila total bike count
Piloto Lx Ciclistas Entradas	Avenida Duque de Ávila bike count from east
Piloto Lx Ciclistas Saídas	Avenida Duque de Ávila bike count from west

## 4.2. Data Preprocessing

### Lisbon BSS Data Preprocessing of 2018

Lisbon BSS data, a fourth-generation system, provides broad and extensive information. Data extraction methods have not yet been extensively explored [42], therefore, there are limitations in the collected data, which need to be evaluated on its limitations and cleaned. Data cleaning involves handling missing data and noise removal, thus generating datasets with accurate and validated data.

The following data cleaning methods were applied to bike trip data:

- Removal of the not assigned (NA) values of the bike type (1% of the dataset).
- Removal of the geometry and number of nodes with NA values, corresponding to 50% of the data.
- Removal of the the variable speed in trips that were shorter than 1 minute.
- The missing values in the distance were filled by computing the average speed times the duration.

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), 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 (ETL) process, loading data from external databases, transforming the data by creating common columns and joining the datasets, and finally loading the data into our project. As a result, from the 3 data schemas presented in Tables 1, 2 and 3, we derived, via such ETL process, 2 datasets, namely, the “bike trips-stations temporal analysis” dataset (Table 7), the “bike trips-stations clustering” dataset (Table 8).

Table 7. Bike trip-stations temporal analysis dataset

Characteristics	Description
id	Trip ID
date_start	Start date
date_end	End date
station_start	Start station
station_end	End station
bike_rfid	Bike RFID
Tipo_Bicicleta	Bike type
duration	Trip duration
speed	Trip speed
hour	Hour
date_key	Date key
DATA_ID	Date key
rain	Precipitation (Y/N)
temp_media	Average temperature
DIA	Day
MES	Month
ANO	Year
SEMANA	Week
SEMESTRE	Semester
TRIMESTRE	Trimester
FERIADO	Holiday
DIA_DE_SEMANA	Weekday
MÊS_DSC	Month description
DATA	Date
Periodo_dia	Day period
níveis_temp	Temperature levels

Table 8. Bike trip-stations clustering dataset

station	Station ID
n_trips	Number of trips
designation	Station designation
lat	Station latitude
lon	Station longitude
c_docas	Statio capacity

After data cleaning, we retained 684,471 trips in 2018. The average number of trips per month, ranging from January to October, was 68,447 and by station, the average number of trips was 9,126. Per day, there was an average number of trips of 2,602.

### Data Preprocessing of Lisbon BSS in 2018, 2019 and 2020

Bike trip data of 2019 and 2020 did not require data cleaning and was ready to use.

IPMA data from 2019 and 2020 required data transformation since there were variables not relevant for our analysis. Our final dataset included the date, weather stations and temperature levels variables. The date format included the hour, and in order to merge with our bike trip data, we had to compute the daily mean of the hourly values. This was processed with the Grouper function from the Pandas [33] library.

This resulted in two datasets: one with all 2019 data (bike trip and IPMA data) and the other with all 2020 data. A temporal variable was added to each of these datasets.

### Data Preprocessing of Avenida Duque de Ávila bike count in 2019 and 2020

Avenida Duque de Ávila bike count data from 2019 to 2020 did not require data cleaning and it was ready to use.

## 4.3. Modelling

### Modelling Lisbon BSS of 2018

Studies in this field aim to understand user's profile and travel behavior [43]–[45], activity patterns in stations [15] and the impact of the built environment in the BSS [46].

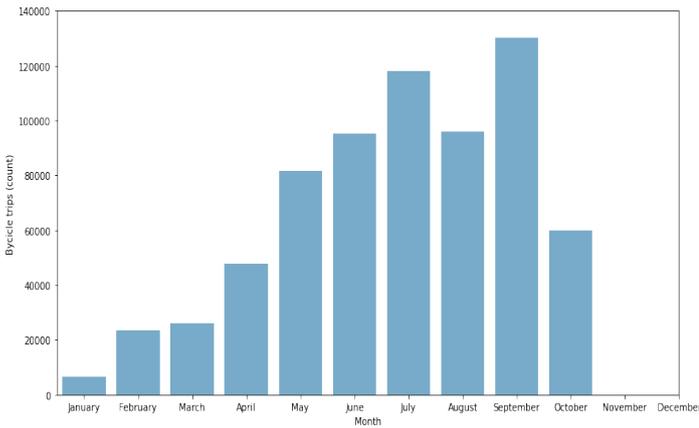
The methods applied focus in statistical methods to analyze and visualize data. To understand bike trip patterns in the urban mobility network and trip models, studies have shown the importance to correlate transport mode and trip choices and built environment characteristics [47], [48].

Many methods are applied to perform data mining, namely, to examine the relations between bike stations, bike trips and the built environment. The evaluation of BSS success depends on these relationships, leading to users' access to bike stations [49].

Clustering algorithms combining temporal and spatial attributes variables are also data mining methods used for this analysis purpose. More specifically, K-means clustering [24]–[27], used by McKenzie [50] and Zhong [51] to measure regularity at different scales and to measure spatiotemporal variation and cluster interaction.

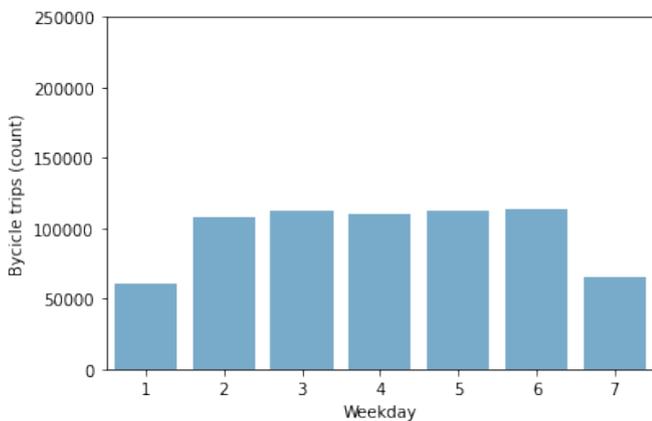
### Bike Usage Analysis

To investigate the monthly bicycle usage frequency, we merged the “bike trip dataset” with the “bike temporal basis dataset” and obtained 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 2018. In the Summer months (June, July, August and September), the more concentrated period (64% of all trips), there were a total of 439,176 trips, as depicted in Fig. 3.



**Figure 3.** Bike usage frequency per month

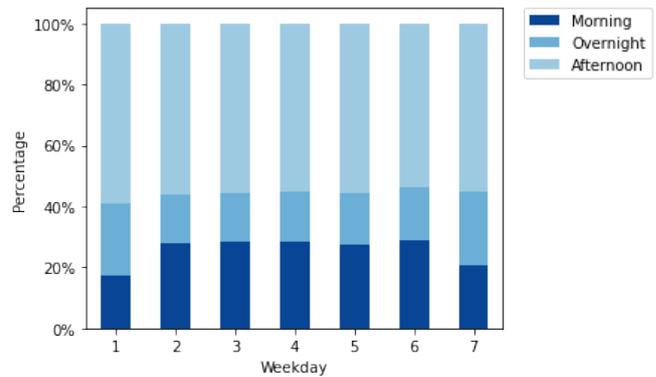
The weekday and weekend usage were also analyzed to understand the preferences of using the bike-sharing service during the week. Results are presented in Fig. 4, where weekdays are ordered from 1 to 7. The weekend is represented by 1 (Sunday) and 7 (Saturday). Our results show that most users (82%) prefer to use the service during the week, rather than during the weekend.



**Figure 4.** Bike usage per weekday

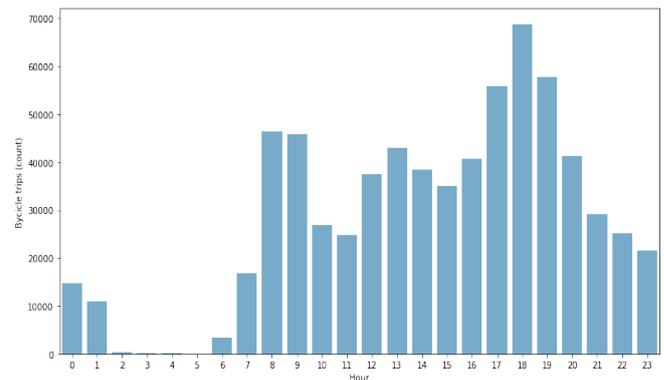
The distribution of trips throughout the different periods of the day was analyzed too. The column `date_start` was transformed into a time format and the hour was extracted in order to create the column `Periodo_dia` (Day period). The day was broken down into three-hour groups: Morning: 7am to 12am; Afternoon: 12am– 8 pm and Overnight: 8 pm– 7am. Our analysis shows that most of the trips (56%) occur during the afternoon compared to the morning and overnight periods (see Fig. 5). 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 bike trips, we addressed the differences between the weekdays versus bike type. Regarding bike 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.



**Figure 5.** Bike usage (%) per weekday within the day period

The hour rate was also analyzed. There was an extraction from the variable `“date_start”` of the hour and the creation of the variable `“hour”`. The higher usage rate corresponds to 6 pm (10%) following 5 pm and 7 pm (see Fig. 6). There is also a high usage rate at 8 am, and 9 am (13% combined). Also, it is possible to see that the citizens start to use this service from 7 am to 1 am, having no significant usage between 2 am and 6 am (see Fig. 6).

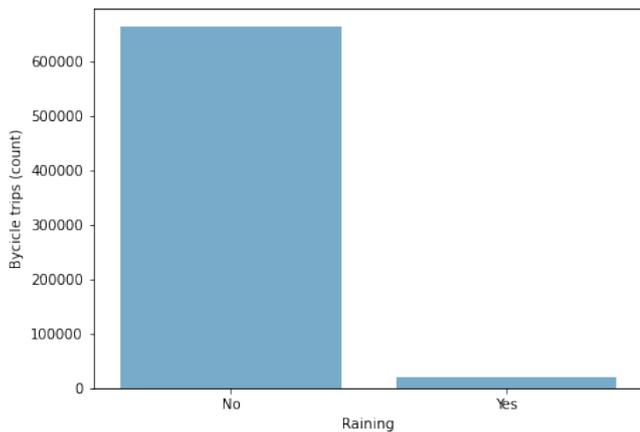


**Figure 6.** Bike usage (trip count) per hour

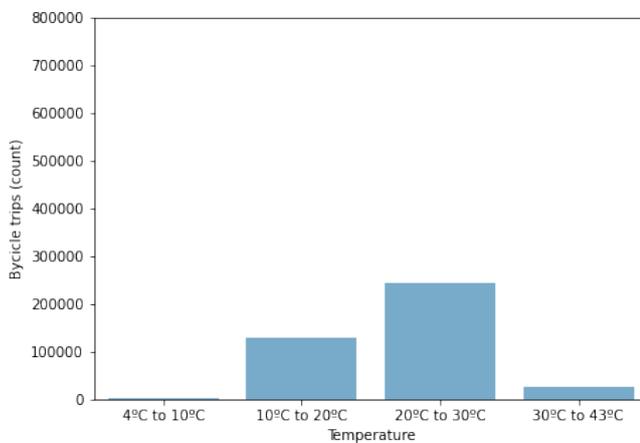
### Bike Trip Weather Analysis

We conducted an additional analysis to find behavior patterns of BSS users, influenced by the built environment variables, particularly weather variables such as atmospheric precipitation and temperature. In our 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. A new date\_key field was generated from the date\_start field of bicycle trips, to join the two datasets. From our analysis, we can conclude that the trips are mostly made when there is no precipitation (97%) (see Fig. 7). 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 categories: 0° to 10°, 10° to 20°, 20° to 30 and 30° to 42°, being 42° the maximum observed temperature value (see Fig. 8).



**Figure 7.** Bike usage frequency relation to atmospheric precipitation



**Figure 8.** Bike usage frequency relation to temperature

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.

### Bike Station Usage Analysis

Our analysis approach on bike station usage was to identify the top 5 most popular stations, the top 5 stations with the highest outflow and highest inflow, and the frequent station pairs on weekdays and weekends looking at in 2018, for each month from January to October and for the whole

period. This analysis only considered trips with a duration of over 60 seconds and less than 2 hours and 15 minutes.

In 2018, there was an evolution of bike usage. Table 9 shows trip increase throughout the year, where the months of July (115,857) and September (127,616) were the ones with more bike usage. Station usage also evolved and almost doubled between January (43 stations) and October 2018 (74 stations). This might be related to the opening of new stations in scope of BSS network expansion.

**Table 9.** Trips and stations

Month	Trips	Stations
January	6,326	43
February	23,324	43
March	25,872	56
April	47,122	58
May	80,417	72
June	93,296	74
July	115,857	74
August	94,007	81
September	127,636	74
October	58,459	74
Jan - Oct	672,316	81

The expansion of the bike station network in 2018 did not change the top 5 most popular stations pattern. As shown in Table 10, the top 5 most popular stations correspond to stations 446 – Avenida da República/Interface de Entrecampos, 481 – Campo Grande/Museu da Cidade, 417 – Avenida Duque de Ávila, 421 – Alameda D. Afonso Henriques, and 105 – Centro Comercial Vasco da Gama. These top 5 most popular stations are observed with different ranking in the analyzed months in 2018.

The top 5 most popular stations are shown and numbered in Figure 8, mapped with all the network bike stations.

Table 10. Top 5 most popular stations

Month	#1	#2	#3	#4	#5
January	446	105	481	417	403
February	446	417	481	105	403
March	446	417	481	105	403
April	446	481	417	105	420
May	446	481	417	420	421
June	446	481	421	417	105
July	446	481	421	417	105
August	446	481	421	417	105
September	481	446	417	421	105
October	481	421	446	417	443
Jan - Oct	446	481	417	421	105

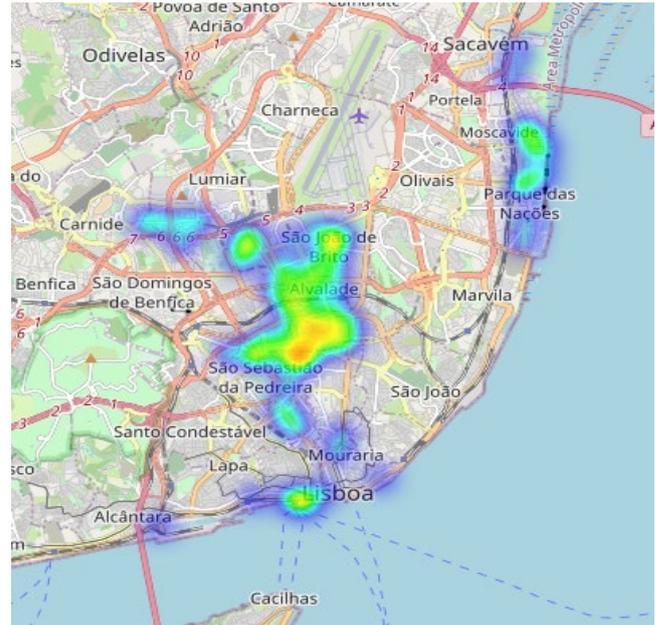


Figure 10. Station trip heatmap. The orange color corresponds to a higher number of station trips, whereas purple, to a lower.



Figure 9. Lisbon BSS stations map with the top 5 most popular stations ID number identified

This is also shown in the station trip heatmap (see Fig. 10), where the orange color corresponds to a higher number of station trips in a gradient to yellow, green and purple lower station trips.

If we look at the flow level (highest inflow and outflow), we observe similarities to the top 5 most popular stations. The top 5 stations with the highest inflow in 2018 (from January to October), are listed in Table 11, as follows: 481 – Campo Grande/Museu da Cidade, 446 - Interface de Entrecampos, 417 – Avenida Duque de Ávila, 421 – Alameda D. Afonso Henriques, and 105 – Centro Comercial Vasco da Gama.

Table 11. Top 5 stations with highest inflow

Month	#1	#2	#3	#4	#5
January	446	481	105	417	403
February	446	481	403	417	105
March	446	417	481	105	403
April	481	446	417	105	403
May	481	446	417	420	421
June	481	421	446	417	105
July	446	421	481	417	105
August	446	481	421	417	105
September	481	417	421	446	105
October	481	421	417	446	443
Jan - Oct	481	446	417	421	105

The top 5 stations with the highest outflow in 2018 (see Table 12) are: 446 - Interface de Entrecampos, 481 – Campo

Grande/Museu da Cidade, 417 – Avenida Duque de Ávila, 421 – Alameda D. Afonso Henriques, and 105 – Centro Comercial Vasco da Gama. We conclude that although the highest inflow and outflow top 5 stations are the same, the first two are ranked differently.

Table 12. Top 5 stations with highest outflow

Month	#1	#2	#3	#4	#5
January	446	105	481	417	403
February	446	417	481	105	403
March	446	417	481	105	403
April	446	481	417	105	420
May	446	481	417	420	421
June	446	481	421	417	105
July	446	481	421	417	105
August	446	481	421	417	105
September	481	446	417	421	105
October	481	421	446	417	443
Jan - Oct	446	481	417	421	105

Looking at the top 5 most frequent station pairs on weekdays and weekend, we observe that in the weekdays (Table 13), most trips take place in Parque das Nações and in the axis Campo Grande-Saldanha. In Parque das Nações, the most used station pair from station 109 – Alameda dos Oceanos/Rua do Zambeze to station 105 – Centro Comercial Vasco da Gama and in the opposite direction. Most frequent station pairs in weekdays are also observed in the Campo Grande to Saldanha axis. This also correspond to top 5 popular stations as well as inflow and outflow stations, namely, from station 446 – Avenida da República/Interface de Entrecampos to station 403 – Avenida Fontes Pereira de Melo, and from station 446 – Avenida da República/Interface de Entrecampos to station 481 – Campo Grande/Museu da Cidade.

Table 13. Top 5 frequent stations pairs in weekdays

Month	#1	#2	#3	#4	#5
January	105-109	403-446	109-105	105-110	446-403
February	105-109	109-105	446-403	403-446	105-107
March	109-105	446-403	105-109	403-446	107-105
April	109-105	105-109	446-403	110-105	107-105
May	105-109	109-105	446-403	446-481	403-446
June	105-109	109-105	107-105	105-107	446-403
July	109-105	105-109	105-107	446-481	107-105
August	109-105	105-109	107-105	105-107	446-481
September	105-109	109-105	446-481	107-105	105-107
October	109-105	105-109	446-481	481-446	421-421
Jan - Oct	109-105	105-109	446-403	107-105	446-481

The top 5 frequent station pairs in the weekends (Table 14), can be found in Parque das Nações, like wise as in the weekdays from station 109 – Alameda dos Oceanos/Rua do Zambeze to station 105 – Centro Comercial Vasco da Gama and in the opposite direction. Also, from station 105 – Centro Comercial Vasco da Gama to station 107 – Rotunda dos Vice-Reis and in the opposite direction. Another frequent station pair on weekends is from station 110 - Rua de Moscovide to station 105 – Centro Comercial Vasco da Gama.

Table 14. Top 5 frequent stations pairs in weekend

Month	#1	#2	#3	#4	#5
January	105-109	109-105	110-105	105-110	464-464
February	109-105	105-109	105-107	446-403	403-446
March	109-105	105-109	107-105	110-105	105-110
April	109-105	105-109	110-105	105-110	107-105
May	105-109	109-105	481-481	484-488	110-105
June	109-105	105-109	105-107	481-481	107-105
July	109-105	105-109	107-105	105-107	421-421
August	109-105	105-109	105-107	107-105	208-208
September	109-105	105-109	105-107	107-105	481-481
October	109-105	105-109	107-105	104-102	443-443
Jan - Oct	109-105	105-109	107-105	105-107	110-105

Overall, in 2018 there was a total of 672,316 trips in 81 stations, where the most popular pair of origin-destination stations had over 1,000 trips, reaching a total of 5,000 trips (see Fig. 11). Patterns previously identified are highlighted in the origin-destination matrix (see Fig. 11).

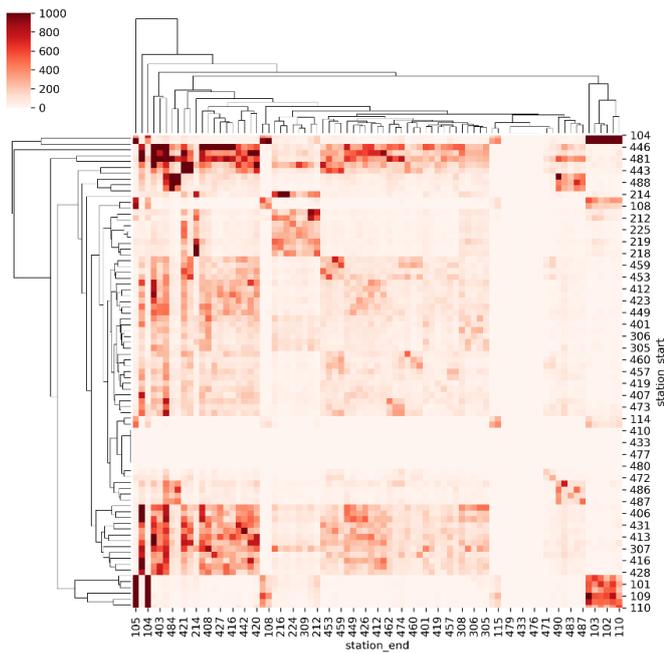


Figure 11. Origin Destination Matrix in 2018

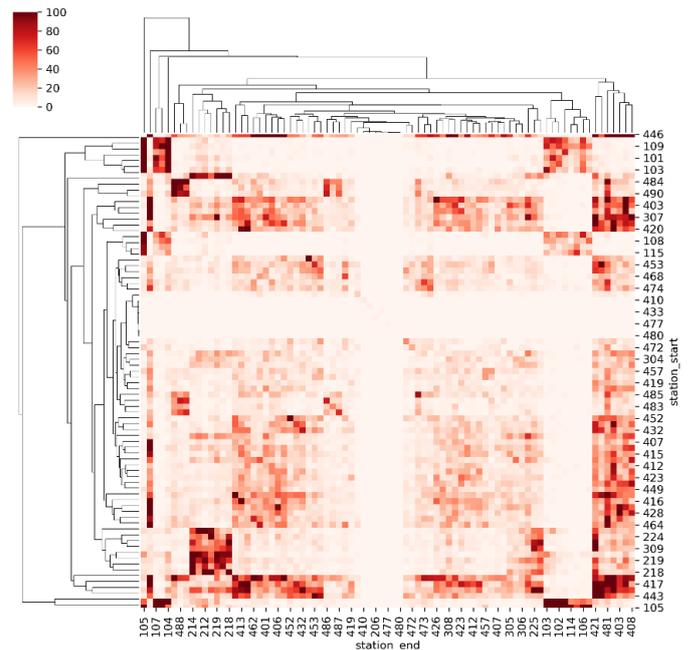


Figure 12. Origin Destination Matrix for August 2018

Looking at the months where most trips occurred, July, August and September, we observe that August shows different station patterns from July and September. This is due to August a holiday month and July and September are working months.

A closer analysis of August and September station pattern shift, we observe that in August there was 94,007 trips in 81 stations (Table 9), where the 5 most popular stations (Table 10) are ranked: 446 - Avenida da República/Interface de Entrecampos, 481 – Campo Grande/Museu da Cidade, 421 – Alameda D. Afonso Henriques, 417 – Avenida Duque de Ávila/Jardim Arco do Cego, and 105 – Centro Comercial Vasco da Gama. Moreover, we observed that the top 5 stations with the highest outflow and inflow (see Fig. 12) are the same as top 5 most popular stations.

Regarding the top 5 frequent station pairs in weekdays and weekends we found that most pair stations are in Parque das Nações, as we also observed in 2018 analysis. Top 5 frequent stations pairs in weekdays (Table 13) are 109-105, 105-109, 107-105, 105-107, and 446-481. The top 5 frequent stations pairs in weekends (Table 14) are 109-105, 105-109, 105-107, 107105, and 208-208 (Cais das Pombas/Cais do Sodré). This shows that in August weekend cycling occur along the river in Parque das Nações and Cais do Sodré.

In September, there was an increase of trips compared to August, with 127,636 trips in 74 stations (Table 9) that might be related to work activity return. We identified that the 5 most popular stations (see Fig. 13) are the same as in August but ranked as follows (Table 10): 481 – Campo Grande/Museu da Cidade, 446 – Avenida da República/Interface de Entrecampos, 417 – Avenida Duque de Ávila/Jardim Arco do Cego, 421 – Alameda D. Afonso Henriques, and 105 – Centro Comercial Vasco da Gama. The top highest outflow and inflow stations are the same but highest inflow is ranked as follows, 481 - Campo Grande/Museu da Cidade, 417 – Avenida Duque de Ávila/Jardim Arco do Cego, 421 – Alameda D. Afonso Henriques, 446 - Avenida da República/Interface de Entrecampos, and 105 – Centro Comercial Vasco da Gama; and highest outflow stations are 481 - Campo Grande/Museu da Cidade, 446 - Avenida da República/Interface de Entrecampos, 417 – Avenida Duque de Ávila/Jardim Arco do Cego, 421 – Alameda D. Afonso Henriques, and 105 – Centro Comercial Vasco da Gama.

Most frequent station pairs in weekdays and weekends show similarities with previous analyzed months. In weekdays most station pairs are located in Parque das Nações, intercalated with Campo Grande and Entrecampos as follows: 105-109, 109-105, 446-481, 107-105, and 105-107. In weekends, station pairs are mostly located in Parque das Nações, 109-105, 105-109, 105-107, 107-105, and Campo Grande/Museu da Cidade 481-481.

trips, station designation, latitude, longitude and dock capacity. Latitude and longitude variables were used for the geographical analysis.

Table 15. Clustering dataset first row entry

station	n_trips	designation	lat	lon	c_docas
446	62600	446 – Av. República/Interface Entrecampos	38.744560	9.147730	40

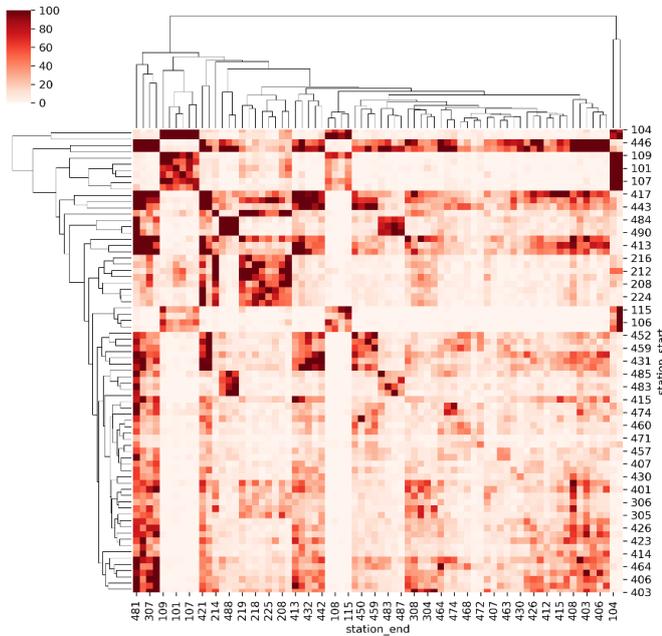


Figure 13. Origin Destination Matrix for September 2018

### Spatial Cluster Analysis

In our research, we seek to understand BSS users' behaviors, particularly the inflow and outflow of trips in each station and the frequency of stations' usage. Hence, we aim to perform clustering analysis identifying geographical patterns in Lisbon BSS. Datum system of latitude and longitude coordinates was normalized and processed in World Geodesic System 1984 (WGS84) regarding station trips data. Geographic clustering was performed with K-means was performed and an additional data pre-processing step was required before performing it. To generate station trips cluster, we first counted all trips of every stations, irrespectively if a given station is the origin or destination of a trip. To this aim, we splited the original "bike trips dataset" in two, one with the 'station\_start' variable and the other with the 'station\_end' variable. Then, the 'station\_start' and 'station\_end' variables were changed to "station" in the corresponding datasets. Afterwards, both datasets were concatenated within the station variable and trip count was computed for each station. It resulted in a dataset (Table 15) with six variables: station, number of

To perform K-means, we used the Elbow algorithm [52], to find the optimal K number through the SSE (Sum of Squared Errors) calculation. As shown in Figure 14, the K value of four corresponds to the minimum SSE of the K optimal value. The four spatial clusters of bike station trips (see Fig. 15) are: first, in the center of Lisbon, in the axis from Alvalade to Saldanha (in blue), second in the northwest side of Lisbon from Telheiras to Campo Grande/Museu da Cidade (in yellow), third in Lisbon downtown area, from Marquês de Pombal to Baixa (in purple), and a fourth, in northwest of Lisbon, in Parque das Nações (in green). Table 16 shows cluster centroids of the geographic clustering generated by K-means.

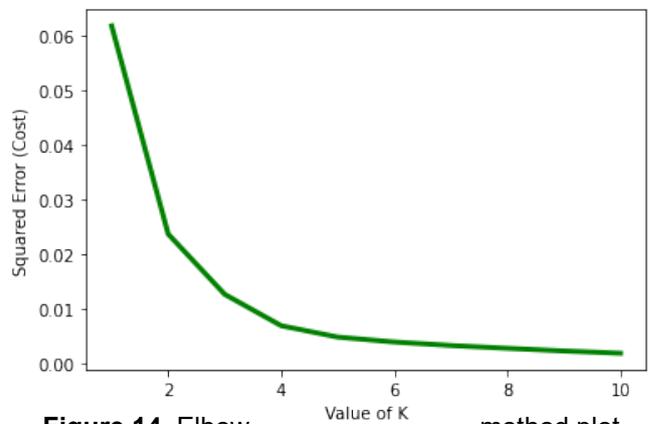
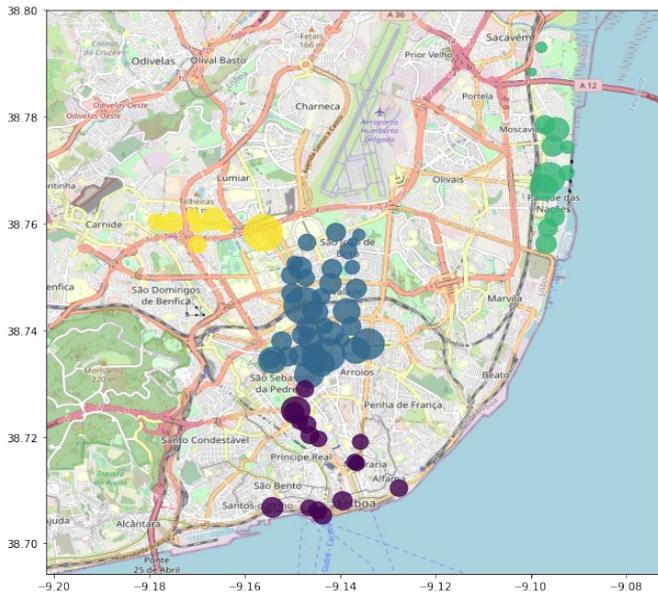
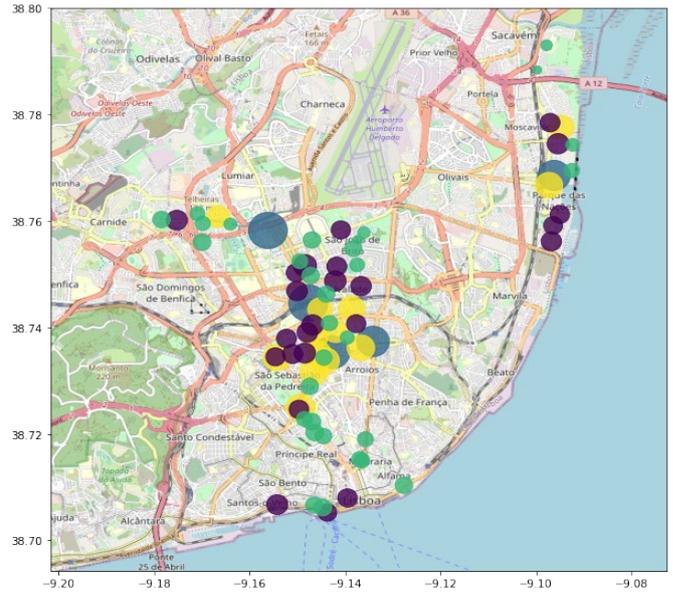


Figure 14. Elbow method plot



**Figure 15.** Spatial clustering of stations where bike trips start and/or end (yellow: Telheiras-Campo Grande/Museu da Cidade; blue: Alvalade-Saldanha; purple: Marquês de Pombal-Baixa; green: Parque das Nações).



**Figure 16.** Spatial clustering by the number of trips that start and/or end on a given station (blue: first most used stations; yellow: second most used stations; Purple: third most used stations; Green: fourth most used stations).

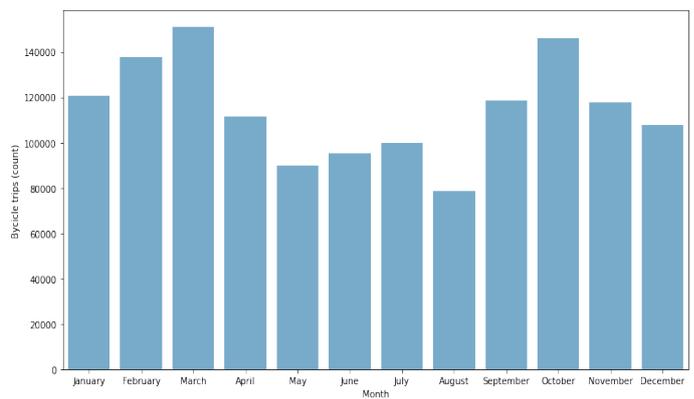
Table 16. Cluster Centroids

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, the variable *n\_trips* was used for clustering, representing the number of station trips. Then K-means was performed, with the same type of approach to find the optimal K number, as in the prior geographical cluster analysis. Four clusters were computed (see Fig. 16) and the four most frequently used stations (labelled in blue) are located in the city centre, while the fifth one is in the northeast. These stations correspond to the top five most popular stations, identified in the previous sub-section

**Bike Usage Analysis of 2019 and 2020**

The same bike usage analysis method implemented for 2018 data, was also applied in 2019 and 2020 data. We divided the 2019 and 2020 data, in two separate datasets by the year and merged each one with the temporal dataset. In 2019, data ranges from 1st January to 31st December, and we observe that in January, February, March, and October (see Fig. 17), there were a total of 555,429 trips (40%). On the other hand, the months with the lowest usage rate are May, June, July, and August, corresponding to late Spring and Summer months, with a total of 363,343 trips (26%).



**Figure 17.** Bike trip count by month in 2019

In 2020, data ranges from 1st January to 4th June, and we conclude that most users cycle in January and February with a total of 267,390 trips, representing 53% of all trips. There is a decrease in trips from March to April (see Fig. 18) of

about 50% (meaning from 80,803 to 40,082 trips). Afterwards, there is an accentuated decrease of trips in May and June, of 86%. This shows a strong impact of the lockdown on BSS mobility patterns, due to Covid 19 pandemic.

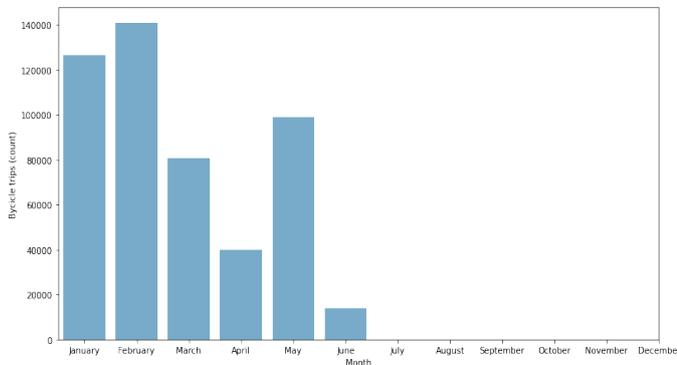


Figure 18. Bike trip count by month in 2020

Moreover, we performed a weekday analysis, applying the same method as in 2018. In 2019, the weekday analysis results show (see Fig. 19) that users tend to use BSS mainly in weekdays, representing a total of 1,134,365 trips (83%).

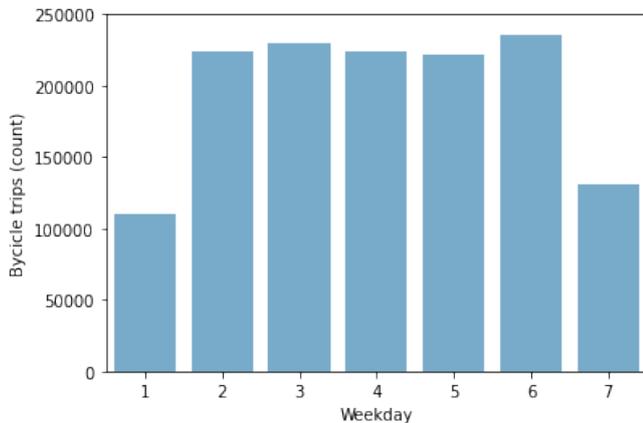


Figure 19. Bike trip count by weekday in 2019

In 2020, we can see the same pattern (see Fig. 20), as in 2018 and 2019, meaning that users mostly ride BSS in weekdays. The total trips of the weekdays of 2020 was 395,103 (78%).

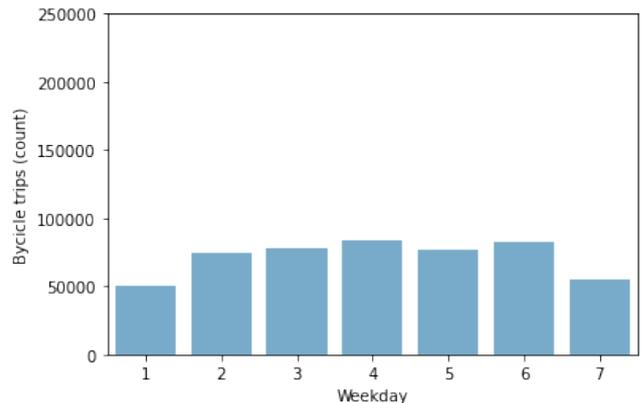


Figure 20. Bike trip count by weekday in 2020

### Relation of Bike Usage Frequency with Temperature, in 2019 and 2020

Bike trips count and temperature analysis for 2019 and 2020, are depicted in Figures 21 and 22. Data was pre-processed, where negative values were removed, and the average values per day (from the hour) were calculated from the four Lisbon weather stations. Using the same method applied to 2018 data, we divided the dataset into four temperature categories: 0°C to 10°C, 10°C to 20°C, 20°C to 30°C and 30°C to 43°C. Results show that the maximum temperature observed in 2019 and 2020 were, respectively, 27°C and 24,5°C. Overall, most users prefer to cycle with mild temperature. A BSS users pattern observed in 2018, 2019 and 2020.

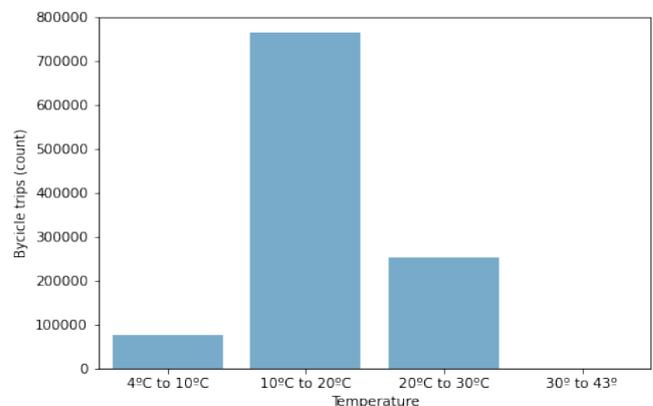
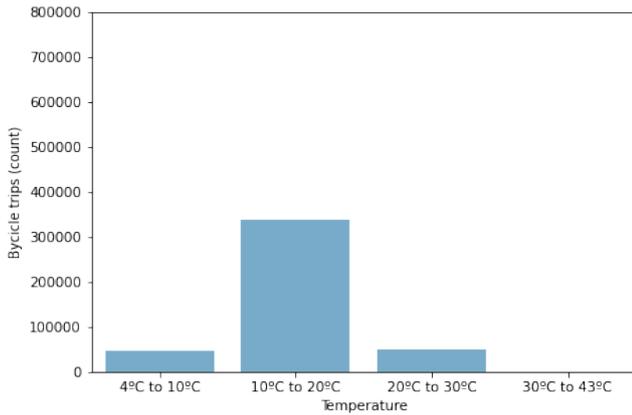
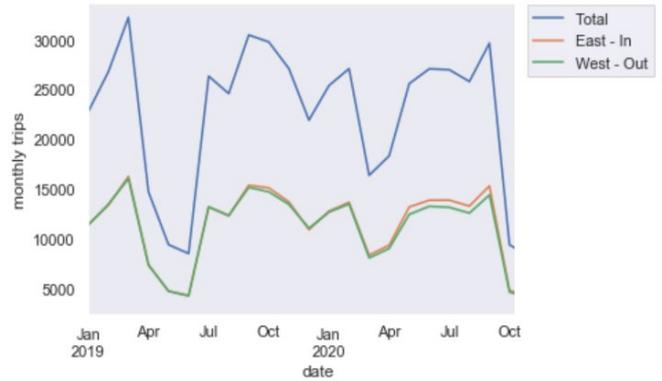


Figure 21. Bike usage frequency relation with temperature (2019)



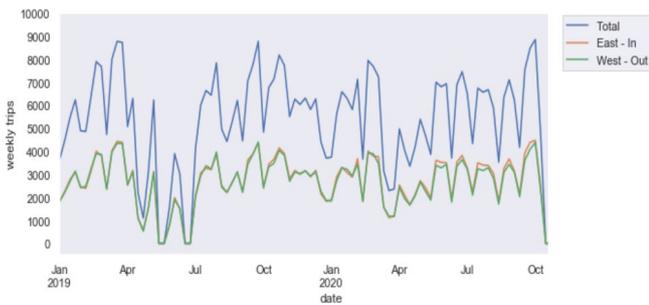
**Figure 22.** Bike usage frequency relation with temperature (2020)



**Figure 24.** Avenida Duque de Ávila monthly bike trip count east, west and total in 2019 and 2020

**Avenida Duque de Ávila bike count analysis from 2019 and 2020**

Our analysis shows that the number of weekly trips from East and West is remarkably similar. The average of total weekly trips is 815, where East and West ranges between 412 and 403. Overall, this analysis (see Fig. 22) shows a regular pattern of weekly trips in 2019 and 2020, where the most frequent trips took place during the weekdays. We observed two periods of decrease in the number of trips. The first, in 2019, between April and July, and although we do not have information, we can argue that there was a data collection malfunction. The second, from middle of March to May 2020, when the first lockdown restrictions were implemented, due to COVID 19 pandemic, showing that such event had a strong impact in Lisbon BSS mobility patterns.



**Figure 23.** Avenida Duque de Ávila weekly bike trip count east, west and total in 2019 and 2020

The monthly analysis (see Fig. 23) shows an average of approximately 2,400 total trips. The two-trip count decrease phenomenon were confirmed with previous analysis results, the first drop observed between April and June 2019, and the second between the middle of March and May 2020, corresponding to the previous mentioned first lockdown.

**5. Discussion**

Our study started with the aim to understand spatiotemporal station and trip activity patterns in Lisbon BSS in 2018, as stated in our RQ1. Our preliminary study [6] addressed our first sub-question on the average monthly and daily Lisbon BSS usage. The analysis showed that the total number of Lisbon BSS trips, from January 15th to October 25th, 2018 was 684,471 and, that the average number of trips per month was 68,447, while the average station number of trips was 9,126. Moreover, we found that the daily average number of trips was 2,602.

The analysis showed also that the months of June, July, August, and September had the most concentration of trips during 2018, of 439,176, representing 64% of all trips. We also observed that BSS users mostly chose weekdays to ride in the city (82%) rather than in the weekend. Another interesting fact regards the hourly usage rate that shows users ride bikes during weekday peak hours, from 8am to 9 am, from 4.30 pm to 6 pm, and at lunch time from 12 to 2 pm. We can affirm that users ride bikes in the daily commute from home to work and work to home and during lunch hours for short travel.

Our findings also show that during 2018, most of the trips are taken in the afternoon (56%), followed by the morning period and that in the weekend, users prefer to ride overnight.

Addressing our sub-question on weather conditions affecting Lisbon BSS mobility patterns, we found that precipitation has a strong impact in bike usage, showing that almost 97% of trips take place when there is no precipitation. This observation was complemented by a correlation with speed analysis showing that higher speed is reached when there is no precipitation. Regarding temperature, most users prefer to travel when temperature ranges between 20° and 30° (52%), and a significant number of users cycle when the temperature is between 10° and 20° (42%).

Sub-question regarding Lisbon BSS origin and destination station groups, we have observed that the most used were observed in two axis: one from Campo

Grande/museu da Cidade to Saldanha and another in Parque das Nações, showing that bike demand start and end stations are located in Lisbon office areas.

Moreover, most common stations pairs are in Parque das Nações both on weekdays and weekends, due to be a busy office area on weekdays and a leisure area at weekends. The most popular station in this area is 105 – Centro Comercial Vasco da Gama.

Most popular stations are located in the axis of Campo Grande/Museu da Cidade and Saldanha - Avenida Duque de Ávila/Jardim Arco do Cego. This area corresponds to a busy office area also surrounded by universities. We have also found that one of the most frequent station pair was between Avenida da República/Interface de Entrecampos and Campo Grande/Museu da Cidade, corresponding to two transportation interfaces. We can raise the hypothesis that users are choosing to commute between interfaces by Lisbon BSS.

Still in RQ1, and regarding the Lisbon BSS clusters sub-question, we found four major concentrations in the city for the number of station trips. The main areas where users unlock BSS correspond to Parque das Nações (1), the city center: Alvalade-Saldanha (2), Telheiras-Campo Grande (3), Marquês de Pombal-Baixa (4) - meaning that the center of Lisbon is where the most trips occur. There is also a close relation of the number of trips with the station capacity. The station cluster with more trips is associated with the stations with the greater bike capacity. We also found a correlation of clusters with the origin and destination station groups.

Regarding RQ2, on addressing how Lisbon BSS trip patterns have changed in 2019 and 2020 from 2018, our study shows that the total number of trips reached 1,374,751 in 2019 (1st January to 31st December) which is an increase of 101% compared to 2018. In 2020 (from 1st January to 4th June) the total number of trips was 501,037 and represents a decrease of approximately 64% from the previous year. This is highlighted by the average number of trips per month in 2019 that was 114,562 and in 2020 was 83,506. The daily trip average observed in 2019 was 3,766 and 2020 was 3,253.

Furthermore, in 2019 and 2020, the summer months are no longer the highest trip rate of monthly usage, as observed in 2018. February, March and October, in both 2019 and 2020, were the months where most trips took place. Also, we can see that the usage is distributed over all months, and there is no discrepancy between the summer months and the other months of the year, as in 2018. Findings show that users are shifting to bike ride during Summer and Winter, preferring to use BSS to other transportation modes. Meaning, Lisbon BSS is becoming a preferred transport mode to commute in Lisbon especially for the last mile.

Regarding temperature, the usage pattern has changed between 2019 and 2020. Users prefer to cycle when temperature is between 10° and 20° (56% and 67% respectively), confirming as well that users tend to ride all year long instead of just in the summer months.

Finally, we also found no significant difference regarding speed and duration of bike trips across the weekdays by bike type (Electric or Conventional). Therefore, our research

suggests that the type of bike is not a decisive factor in the bike trip analysis.

Avenida Duque de Avila bike count showed results with similar mobility patterns of weekly and monthly bike usage as in Lisbon BSS analysis. Bike users are more active during weekdays and the counting is almost the same regarding its direction of origin and destination (East and West).

On the impact of the COVID 19 pandemics event, we observed a clear correlation with BSS usage. In 2020, the trip decrease between March and April can be explained by the State of Emergency lockdown declared in Portugal from 18th March 2020 to April, and then renewed on 3rd April 2020 until 2nd May 2020. This explains the decrease in bike trips in 2020, compared to the same period of time in 2018 and 2019.

The major limitation found in our study was the unavailability of bike usage features in provided data for 2019 and 2020 but made available in the 2018 datasets. In 2019 and 2020 data presents the aggregated total counting per day, not specifying either the origin and destination stations or trip time (hour, minute and second). This prevents the authors to understand and obtain insights, regarding the years of 2019 and 2020, such as the evolution of users' mobility patterns, the most popular stations and stations pairs, inflow and outflow in the various stations and trajectories.

## 6. Conclusions

This paper provides new insights into Lisbon BSS, first implemented in 2017 and evolving till nowadays. It was interesting to analyse the evolution and strong BSS demand in a city that did not have a cycling culture until recently.

Significant findings show that most Lisbon BSS trips in 2018 occur in weekdays during the afternoon, which correlates with the daily afternoon commute (6 pm – 7 pm). We also found that weather conditions [18], [19] had an important impact on travel behavior. No precipitation was consistent with ridership increase, as well as mild temperatures between 10° to 30°. In 2018 June, July, August and September trips represent 64% of all bike trips in 2018. The most frequent trip station pair origin and destination are along Campo Grande and Saldanha axis, and in Parque das Nações, where most offices and universities are located in Lisbon. The cluster analysis highlighted the previous results in four clusters located in Alvalade-Saldanha, Telheiras-Campo Grande, Marquês de Pombal-Baixa and Parque das Nações areas.

Although limitations of 2019 and 2020 data did not allow us to perform a spatiotemporal analysis, we performed a monthly, weekday and weather correlation analysis. In 2019, the months February, March and October represent 40% of all trips since there is a high usage during all year. In 2020, most trips were taken in January and February representing 53% of all trips. This is a striking difference compared with 2018 when trips mostly occurred in Summer months. Meaning BSS is becoming a frequent mode in Lisbon commute. In 2019 trips doubled from 2018, with

good demand rates in 2020, although the complete year data is required to its analysis. Avenida Duque de Ávila bike count data of 2019 and 2020 added a broader scenario to the analysis with a case study, and confirmed previous findings in 2018, 2019 and 2020 Lisbon BSS that trips are more frequent on weekdays.

The impact of the COVID 19 pandemics in urban mobility patterns has a clear correlation with Lisbon BSS usage but more data is needed to understand better the phenomenon along 2020.

Lisbon BSS trip patterns are thus similar to other observed BSS of medium-size cities [18] discussed in the State of the Art section, such as patterns found in short and frequent trips and ride peak observed in the morning and afternoon, as in the case study of the city of Cork (Ireland) [18].

Parallels with larger cities can be established as well. In Canada, for instance, Montreal's BIXI BSS [21] is mainly used on weekdays, evenings and weekends. In Toronto, bike trips are shorter on the weekdays mornings [53].

Large USA cities BSS studies [16], [17], [20] show frequent bike use in the morning and afternoon peaks [16] and different usage patterns between weekdays and weekends, identifying longer trips in the weekend [16].

In large European cities, weekday morning trips in the peak hour [14], [51] reach a higher speed than trips over the weekdays and weekends.

As for the Lisbon 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 Lisbon BSS in the scope of urban analytics [54] and parallel comparison with other BSS implemented nationally and internationally.

Lisbon BSS future work also requires bike data availability of 2019 and 2020 and coming years, with the same features as in 2018 data, to achieve the level of analysis regarding stations and cluster analysis.

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

### Acknowledgements.

This study was performed in the scope of Lisboa Inteligente LxDataLab [41] – Lisbon City Council Data Centre – challenges to the academia. Vitória Albuquerque and Francisco Andrade work was 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 [55].

### References

- [1] OECD, *Greening Household Behaviour*. OECD, 2011.
- [2] “A Territorial Approach to the Sustainable Development Goals,” Paris, 2020. [Online]. Available: <https://doi.org/10.1787/e86fa715-en>.
- [3] *Estratégia Nacional para a Mobilidade Ativa Ciclável 2020-2030*. Conselho de Ministros, 2019, pp. 46–81.
- [4] A. C. LVT, “Estratégia Regional de Lisboa AML 2030,” Lisboa, 2020. [Online]. Available: [https://www.aml.pt/susProjects/susWebBackOffice/uploads/wt1wwpgf\\_aml\\_sus\\_pt\\_site/componentText/SUS5F9C0ECCBA6EF/ESTRATEGIA\\_REGIONAL\\_DE\\_LISBOA\\_AML\\_LISBOA\\_2030.PDF](https://www.aml.pt/susProjects/susWebBackOffice/uploads/wt1wwpgf_aml_sus_pt_site/componentText/SUS5F9C0ECCBA6EF/ESTRATEGIA_REGIONAL_DE_LISBOA_AML_LISBOA_2030.PDF).
- [5] C. M. de Lisboa, “MOVE Lisboa 2030 Visão estratégica para a mobilidade 2030,” 2020. [https://www.lisboa.pt/fileadmin/cidade\\_temas/mobilidade/documentos/BrochuraMOVE\\_2030.pdf](https://www.lisboa.pt/fileadmin/cidade_temas/mobilidade/documentos/BrochuraMOVE_2030.pdf).
- [6] V. Albuquerque, F. Andrade, J. C. Ferreira, and M. S. Dias, “Understanding Spatiotemporal Station and Trip Activity Patterns in the Lisbon Bike-Sharing System,” 2021, pp. 16–34.
- [7] R. Meddin and P. J. DeMaio, “The Meddin Bike-sharing World Map,” *Google Maps*, 2020. <https://bikesharingworldmap.com/#/all/2.3/-1.57/33.92/> (accessed Oct. 03, 2020).
- [8] P. DeMaio, “Bike-sharing: History, Impacts, Models of Provision, and Future,” *J. Public Transp.*, vol. 12, no. 4, pp. 41–56, Dec. 2009, doi: 10.5038/2375-0901.12.4.3.
- [9] J. Zhang, X. Pan, M. Li, and P. S. Yu, “Bicycle-sharing system analysis and trip prediction,” in *Proceedings - IEEE International Conference on Mobile Data Management*, Jun. 2016, vol. 2016-July, pp. 174–179, doi: 10.1109/MDM.2016.35.
- [10] M. Padgham, “Human Movement Is Both Diffusive and Directed,” *PLoS One*, vol. 7, no. 5, p. e37754, May 2012, doi: 10.1371/journal.pone.0037754.
- [11] O. O’Brien, J. Cheshire, and M. Batty, “Mining bicycle sharing data for generating insights into sustainable transport systems,” *J. Transp. Geogr.*, vol. 34, pp. 262–273, Jan. 2014, doi: 10.1016/j.jtrangeo.2013.06.007.
- [12] “General Data Protection Regulation.” <https://gdpr-info.eu/> (accessed Mar. 29, 2021).
- [13] S. Kaplan, F. Manca, T. A. S. T. A. S. Nielsen, and C. G. C. G. Prato, “Intentions to use bike-sharing for holiday cycling: An application of the Theory of Planned Behavior,” *Tour. Manag.*, vol. 47, pp. 34–46, Apr. 2015, doi: 10.1016/j.tourman.2014.08.017.
- [14] P. Jensen, J.-B. Rouquier, N. Ovtracht, and C. Robardet, “Characterizing the speed and paths of shared bicycle use in Lyon,” *Transp. Res. Part D Transp. Environ.*, vol. 15, no. 8, pp. 522–524, Dec. 2010, doi: 10.1016/j.trd.2010.07.002.
- [15] N. Lathia, S. Ahmed, and L. Capra, “Measuring the impact of opening the London shared bicycle scheme to casual users,” *Transp. Res. Part C Emerg. Technol.*, vol. 22, pp. 88–102, Jun. 2012, doi: 10.1016/j.trc.2011.12.004.
- [16] X. Zhou, “Understanding Spatiotemporal Patterns of Biking Behavior by Analyzing Massive Bike Sharing Data in Chicago,” *PLoS One*, vol. 10, no. 10, p. e0137922, Oct. 2015, doi: 10.1371/journal.pone.0137922.
- [17] Z. Nolan, R. B.; Smart, M. J.; Guo, “Bikeshare trip generation in New York City,” *Transp. Res. Part A Policy Pract.*, no. 94, pp. 164–181, 2016.
- [18] B. Caulfield, M. O’Mahony, W. Brazil, and P. Weldon,

- “Examining usage patterns of a bike-sharing scheme in a medium sized city,” *Transp. Res. Part A Policy Pract.*, vol. 100, pp. 152–161, Jun. 2017, doi: 10.1016/j.tra.2017.04.023.
- [19] W. El-Assi, M. Salah Mahmoud, K. Nurul Habib, M. Mahmoud, and K. Habib, “Effects of built environment and weather on bike sharing demand: Station level analysis of commercial bike sharing in Toronto,” *Transportation (Amst.)*, vol. 44, no. 3, pp. 1–25, May 2015, doi: 10.1007/s11116-015-9669-z.
- [20] X. Han *et al.*, “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 2017 -*, Aug. 2018, pp. 1–9, doi: 10.1109/UIC-ATC.2017.8397557.
- [21] U. Faghih-Imani, Ahmadreza; Eluru, N.; El-Geneidy, A. M.; Rabbat, M.; Haq, “How land-use and urban form impact bicycle flows: evidence from the bicycle sharing system (BXI) in Montreal,” *Transp. Geogr.*, no. 41, pp. 306–314, 2014.
- [22] A. Sarkar, N. Lathia, and C. Mascolo, “Comparing cities’ cycling patterns using online shared bicycle maps,” *Transportation (Amst.)*, vol. 42, no. 4, pp. 541–559, Jul. 2015, doi: 10.1007/s11116-015-9599-9.
- [23] Y. Du, F. Deng, and F. Liao, “A model framework for discovering the spatio-temporal usage patterns of public free-floating bike-sharing system,” *Transp. Res. Part C Emerg. Technol.*, vol. 103, pp. 39–55, Jun. 2019, doi: 10.1016/j.trc.2019.04.006.
- [24] L. Caggiani, R. Camporeale, M. Ottomanelli, and W. Y. Szeto, “A modeling framework for the dynamic management of free-floating bike-sharing systems,” *Transp. Res. Part C Emerg. Technol.*, vol. 87, pp. 159–182, Feb. 2018, doi: 10.1016/j.trc.2018.01.001.
- [25] Y. Zhao, L. Dai, L. Peng, Y. Song, and Z. Zhou, “Analysis of spatial distribution of China’s station-free bike-sharing by clustering algorithms,” in *ACM International Conference Proceeding Series*, 2019, pp. 15–19, doi: 10.1145/3325730.3325748.
- [26] D. Li, Y. Zhao, and Y. Li, “Time-Series Representation and Clustering Approaches for Sharing Bike Usage Mining,” *IEEE Access*, vol. 7, pp. 177856–177863, 2019, doi: 10.1109/ACCESS.2019.2958378.
- [27] Y. Guo, X. Shen, Q. Ge, and L. Wang, “Station Function Discovery: Exploring Trip Records in Urban Public Bike-Sharing System,” *IEEE Access*, vol. 6, pp. 71060–71068, 2018, doi: 10.1109/ACCESS.2018.2878857.
- [28] R. Chapman, Peter; Clinton, Janet; Kerber, Randy; Khabaza, Tom; Reinartz, Thomas; Russel H. Shearer, C.; Wirth, “CRISP-DM 1.0: Step-by-step data mining guide,” *Comput. Sci.*, 2000.
- [29] A. H. El El-Sappagh, Shaker H. Ali; Hendawi, Abdeltawab M. Ahmed; Bastawassy, “A proposed model for data warehouse ETL processes,” *J. King Saud Univ. - Comput. Inf. Sci.*, no. 23, pp. 91–104, 2011.
- [30] “Python.” <https://www.python.org/> (accessed Mar. 29, 2021).
- [31] “Jupyter.” <https://jupyter.org/> (accessed Mar. 29, 2021).
- [32] “Numpy.” <https://numpy.org/> (accessed Mar. 29, 2021).
- [33] “Pandas.” <https://pandas.pydata.org/> (accessed Mar. 29, 2021).
- [34] “Matplotlib.” <https://matplotlib.org/> (accessed Mar. 29, 2021).
- [35] “Seaborn.” <https://seaborn.pydata.org/> (accessed Mar. 29, 2021).
- [36] “GDAL.” <https://pypi.org/project/GDAL/> (accessed Mar. 29, 2021).
- [37] “Shapely.” <https://pypi.org/project/Shapely/> (accessed Mar. 29, 2021).
- [38] “Folium.” <https://pypi.org/project/folium/> (accessed Mar. 29, 2021).
- [39] “Fiona.” <https://pypi.org/project/Fiona/> (accessed Mar. 29, 2021).
- [40] “Scikit-learn.” <https://scikit-learn.org/stable/> (accessed Mar. 29, 2021).
- [41] “LxDataLab – Laboratório de Dados Urbanos de Lisboa.” <https://lisboainteligente.cm-lisboa.pt/lxdatalab/> (accessed Mar. 29, 2021).
- [42] Y. Zhang, T. Thomas, M. J. G. Brussel, and M. F. A. M. Van Maarseveen, “Expanding bicycle-sharing systems: Lessons learnt from an analysis of usage,” *PLoS One*, vol. 11, no. 12, p. e0168604, Dec. 2016, doi: 10.1371/journal.pone.0168604.
- [43] D. Fuller *et al.*, “Use of a new public bicycle share program in Montreal, Canada,” *Am. J. Prev. Med.*, vol. 41, no. 1, pp. 80–83, Jul. 2011, doi: 10.1016/j.amepre.2011.03.002.
- [44] S. A. Shaheen, H. Zhang, E. Martin, and S. Guzman, “China’s Hangzhou Public Bicycle,” *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2247, no. 1, pp. 33–41, Jan. 2011, doi: 10.3141/2247-05.
- [45] S. A. Shaheen, A. P. Cohen, and E. W. Martin, “Public Bikesharing in North America: Early Operator Understanding and Emerging Trends,” *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2387, no. 1, pp. 83–92, Jan. 2013, doi: 10.3141/2387-10.
- [46] D. Buck and R. Buehler, “Bike lanes and other determinants of capital bikeshare trips,” *91st Transp. Res. Board Annu. Meet.*, pp. 703–706, 2012, [Online]. Available: <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Bike+Lanes+and+Other+Determinants+of+Capital+Bikeshare+Trips#0>.
- [47] R. Ewing and R. Cervero, “Travel and the built environment,” *J. Am. Plan. Assoc.*, vol. 76, no. 3, pp. 265–294, Jun. 2010, doi: 10.1080/01944361003766766.
- [48] A. Kemperman, H. Timmermans, and H. Timmerman, “Influences of Built Environment on Walking and Cycling by Latent Segments of Aging Population,” *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2134, no. 1, pp. 1–9, Jan. 2009, doi: 10.3141/2134-01.
- [49] Z. Liu, X. Jia, and W. Cheng, “Solving the Last Mile Problem: Ensure the Success of Public Bicycle System in Beijing,” *Procedia - Soc. Behav. Sci.*, vol. 43, no. 43, pp. 73–78, 2012, doi: 10.1016/j.sbspro.2012.04.079.
- [50] G. McKenzie, “Docked vs. Dockless bike-sharing: Contrasting spatiotemporal patterns,” in *Leibniz International Proceedings in Informatics, LIPIcs*, 2018, vol. 114, doi: 10.4230/LIPIcs.GIScience.2018.46.
- [51] C. Zhong *et al.*, “Variability in regularity: Mining temporal mobility patterns in London, Singapore and Beijing using smart-card data,” *PLoS One*, vol. 11, no. 2, 2016, doi: 10.1371/journal.pone.0149222.
- [52] H. Yuan, Chunhui; Yang, “Research on K-Value Selection Method of K-Means Clustering Algorithm,” *Multidiscip. Sci. J.*, 2019.
- [53] W. El-Assi, M. Mahmoud, and K. Habib, “Effects of built environment and weather on bike sharing demand: Station

- level analysis of commercial bike sharing in Toronto,”  
*Transportation (Amst)*, pp. 1–25, 2015.
- [54] M. Batty, “Urban analytics defined,” *Environ. Plan. B Urban Anal. City Sci.*, vol. 46, no. 3, pp. 403–405, Mar. 2019, doi: 10.1177/2399808319839494.
- [55] “Summer School 2020 – IoT for Smart Cities.”  
<http://istar.iscte-iul.pt/summerschool2020/> (accessed Aug. 03, 2020).