UNIVERSITY
INSTITUTE
OF LISBON

Data Towards City Bike Mobility Patterns

Francisco António Faria Andrade

Master's in **Integrated Business Intelligence Systems**

Supervisor
PhD João Carlos Amaro Ferreira, Assistant Professor with Habilitation,
ISCTE-IUL

Co-supervisor
PhD José Miguel de Oliveira Monteiro Sales Dias, Associated Professor with Habilitation,
ISCTE-IUL

Outubro 2021

Department of Information Science and Technology

Data Towards City Bike Mobility Patterns

Francisco António Faria Andrade

Master's in **Integrated Business Intelligence Systems**

Supervisor
PhD João Carlos Amaro Ferreira, Assistant Professor with Habilitation,
ISCTE-IUL

Co-supervisor
PhD José Miguel de Oliveira Monteiro Sales Dias, Associated Professor with Habilitation,
ISCTE-IUL

Outubro 2021

Resumo

As novas tecnologias aplicadas aos serviços de transporte e a transição para meios de transporte sustentáveis tornaram os sistemas de bicicletas partilhadas mais relevantes no cenário da mobilidade urbana. O objetivo deste estudo é compreender os padrões de mobilidade de espaço e tempo das estações e viagens neste sistema de Lisboa em 2018, e também compreender as mudanças na taxa de viagens nos sistemas de Lisboa em 2019 e 2020 em comparação com 2018. Analisando a distribuição de espaço e tempo das viagens através das estações e, os fatores climáticos juntamente com a taxa de utilização ao longo dos anos, é possível melhorar e tornar o sistema mais adequado à procura dos utilizadores. Usamos um grande conjunto de dados com implementação do CRISP-DM. A principal contribuição do trabalho foi o desenvolvimento de um processo de análise e visualização de dados urbanos, especificamente dados de sistemas de bicicletas partilhadas, que permite assim, a melhor compreensão de como as pessoas se movem na cidade usando bicicletas. Além disso, é importante identificar os padrões de mobilidade que mudam com o tempo e o impacto dos eventos pandémicos. Os resultados mostram que a maior parte do uso de bicicletas partilhadas é efetuado durante a semana, sem precipitação e com temperatura amena. Houve um aumento exponencial no número de viagens, por sua vez interrompido pela pandemia do COVID-19. Esta abordagem pode ser aplicada a qualquer cidade com dados digitais disponíveis.

Keywords: sistemas de bicicletas partilhadas, padrões de mobilidade urbana, análise estatística, análise de agrupamentos.

Abstract

New technologies applied to transportation services and the shifting to sustainable modes of transportation turned bike-sharing systems more relevant in the urban mobility scenario. This thesis aims to understand the spatiotemporal station and trip activity patterns in Lisbon bike-sharing system in 2018 and understand trip rate changes in Lisbon bike-sharing system in 2019 and 2020 compared to 2018. By analyzing the spatiotemporal distribution of trips through stations and the weather factors combined with the usage rate throughout the years, it is possible to improve and make the system more suitable to the users' demand. In this research work, we used large open datasets made available by the Lisbon City Hall, that are deployed by using the CRISP-DM. Our major work contribution was the development of a data analytics process for urban data, specifically bike-sharing data, that helps to understand how people move in the city using bikes. Moreover, we aimed to understand how mobility patterns change over time and the impact of pandemic events. Major findings show that most bike-sharing happens on weekdays, with no precipitation and mild temperature. Additionally, there was an exponential increase in the number of trips, cut short by COVID-19 pandemics. The current approach can be applied to any city with digital data available.

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

Acknowledgments

This is the only part of the document, where the first person can be used. The rest of the document must be written in the third person.

To my advisor, Professor João Carlos Ferreira, I thank you for all your guidance and support throughout my master's degree helping me overcome any obstacle.

To my co-supervisor, Professor Miguel Sales Dias, I thank you for always supporting my work and always correcting my flaws, and guiding me through the right path in making this thesis.

To the family, and especially to my parents who always supported me and gave me the chance to participate in this master's degree, it wouldn't be possible without them.

To colleagues from my Master's in Business Intelligence who accompanied me with the thesis and my route in ISCTE.

To friends, Vitória Albuquerque for always pushing me forward and always giving me the drive to proceed. To Luís de Brito Elvas for always being on my side and helping me for these 2 years to achieve my objectives and to be better and ambitious and to Sara Gouveia for always supporting my work and always believing me even with I was filled with doubts.

Finally, I want to thank the school for creating this master's and accepting my application and making the creation of this research work possible. Also, I want to thank CML for making this theme available as a challenge and for being supportive in terms of questions and approaches to this work.

To all that I enumerate my sincere "Thank you".

Index

Tables Index

Figures Index

List of abbreviations

BSS – Bike-sharing system

CML – Câmara Municipal de Lisboa

ETL – Extract, Load and Transform

IPMA - Instituto Português do Mar e da Atmosfera

O-D – Origin-Destination

OECD – Organisation for Economic Co-operation and Development

1 Introduction

1.1.Topic Context

Cities are becoming more predominant in modern societies, and the mobility of citizens is raising problems concerning pollution and traffic. To overcome such challenges, shared mobility approaches have been developed. In this domain, bike-sharing is a rising active mobility mode, showing large growth rates worldwide. Such demand, increased the number of bike-sharing companies, becoming more effective and available in most developed cities. Moreover, citizens are shifting towards more sustainable urban transports, such as bike-sharing. It is 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 (BSS) in the long-term public transport plans and daily commute.

1.2.Motivation and Topic Relevance

In late 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 studies [6]–[8] we have adopted a data mining approach to understand station and trip patterns in Lisbon BSS and understand this service evolution throughout the years. To this aim, we have analyzed Lisbon BSS 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 Lisbon BSS usage rate from 2018 to 2020 and the impact of the COVID 19 pandemic.

1.3. Research Questions and Goals

This study aims to collect, analyze, and visualize spatiotemporal bike trip data, with trip id, origin and destination stations, trajectory, and time, to identify spatial and temporal patterns. The data was provided through the Lisbon City Hall (Lisboa Inteligente) challenges. Data was already formatted with the available variables 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's 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 the 2018 baseline?

This question statement leads to the following sub-questions:

- What is the bike trip's relation to weather conditions, specifically, to precipitation and temperature?
- How did the bike trip patterns change given the COVID 19 pandemics?

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 a 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 by a sensor in Avenida Duque de Ávila from 2019 to 2020.

1.4. Structure and Organization of Dissertation

The thesis is organized into four chapters that intend to reflect the different phases until its conclusion.

This paper is structured as follows: section 1, includes this Introduction, while section 2 presents our State-of-the-Art survey. Section 3 introduces our data mining methodology, Cross-Industry Standard Process for Data Mining (CRISP-DM), with data understanding, data pre-processing, modeling, and evaluation explained and presented with analysis and visualizations. In section 4, Conclusions, we discuss our results with a comparative analysis, identifying research gaps and limitations, and finally defining future research work.

2 Literature review

2.1. Methodology

The literature review methodology is structured in two steps: first, a bibliometric research tool Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), was adopted to select the most relevant papers to address in our systematic literature review (SLR). This methodology helps us define a strategy to report our SLR starting with a larger number of identified records and consequentially excluding records according to eligibility criteria [9]. Every stage of PRISMA must be reported in the PRISMA flow diagram to summarize the selection process. The second step consists in using a bibliometric research tool for network analysis, VOSviewer[1], to analyze the papers resulting from the PRISMA method implementation. This tool creates a bibliometric network based on the authors, keywords, and title and abstract text in clusters, allowing us to understand the SLR through three types of analysis: keyword, author co-authorship, and title and abstract text occurrence. These networks can be displayed by three types of visualization: network visualization, overlay visualization, and density visualization.

---

[1] www.vosviewer.com

2.2. Methodology Results

2.2.1. PRISMA

The SLR was performed to understand the most relevant papers on the bike-sharing system's mobility patterns analysis with a special emphasis on weather conditions and the COVID-19 pandemic.

We searched in the Scopus database about published work. We followed a systematic approach to select papers related to "bike sharing system" combined with "mobility patterns" to find papers on mobility behavior. These queries were the initial search criteria applied to the Scopus database, with dates ranging from 2014 to 2020. The output of these queries was manually checked to identify additional relevant studies that were missed in the database search. We excluded workshops, books, editorials, 2021 publications, and works not related to the domain. We only selected papers published in journals and conference papers.

The initial selection of papers was made by reading the title and abstract of the papery, and in the cases of insufficient information, the full document was checked. Paper data was collected, stored, and managed using Mendeley[2]. This data consisted of citation information, namely authors and co-authors, title, year, EID, source title, volume, issue, pages, citation count, source and document type, publication stage, DOI, open access, the bibliographical information such as affiliations, serial identifiers, PubMed ID, publisher, editor, the language of the original document, correspondence address, abbreviated source title, and abstract and keywords.

For data synthesis and analysis, we made a qualitative assessment based on the methodological quality, sample size, intervention characteristics, outcome, statistical significance, and direction of effects observed.

Query 1 - "bike sharing system" OR "BSS" OR "bike-sharing" AND "mobility patterns" in Scopus we had 84 documents retrieval in the said timeline. If we limited the search to journals and relevant conference proceedings, the number of documents retrieved was 77. We filtered manually the subjects related to computer science and engineering. This resulted in 40 documents to perform full-text screening, towards identifying significant content to the research questions (RQ).

[2] www.mendeley.com

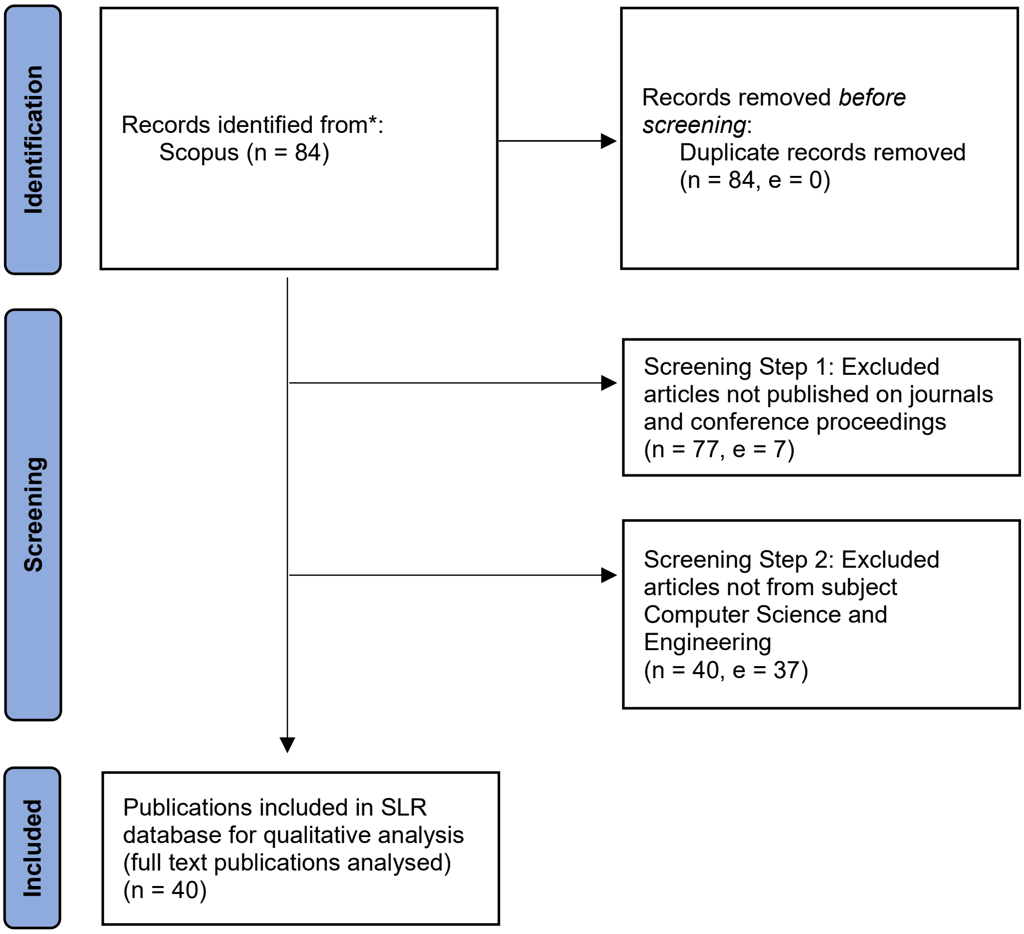

Figure 2.1: PRISMA flow diagram (n=retained; e=excluded)

2.2.2. VOSviewer

The VOSviewer bibliometric research tool for network analysis, allowed us to make a keyword occurrence analysis. The analysis was performed using a full counting method using the threshold of 3, as the minimum number of occurrences per keyword. Only 20 keywords were selected for the analysis. Most of the analyzed keywords were related to mobility patterns analysis applied to bike-sharing systems such as occupation rate and spatiotemporal variation. The five most relevant terms were bicycles (22 occurrences, 72 total link strength), sharing systems (13 occurrences, 49 total link strength), urban transportation (8 occurrences, 35 total link strength), cycle transport (6 occurrences, 33 total link strength) and urban mobility (8 occurrences, 33 link strength).

Table 2.1: Keyword occurrences ranked by total link strength

Keyword	**Occurrence**	**Total Link Strength**
bicycles	22	72
sharing systems	13	49
urban transportation	8	35
cycle transport	6	33
urban mobility	8	33
bike-sharing	5	30
mobility pattern	10	29
data mining	7	25
china	4	24
spatiotemporal analysis	4	23
public transport	5	20
smart city	4	20
travel demand	3	20
big data	3	19
forecasting	3	19
urban planning	5	19
travel behavior	3	17
new york city	3	14
travel time	3	14
cluster analysis	3	13

We found 4 clusters (Figures 2.2 and 2.3) with 20 items, 136 links, and total link strength of 264. For cluster 1, the biggest node corresponds to bicycles (red), cluster 2 corresponds to bike-sharing (green), cluster 3 corresponds to cluster analysis (blue), and cluster 4 corresponds to big data (yellow). In Figure 2.3, we can see that most of the keywords and links between keywords belong mostly to articles from 2019. This analysis showed that the topics bike-sharing and mobility patterns, referring to spatiotemporal analysis, as well as cluster analysis, show a strong connection with the main research fields of this thesis.

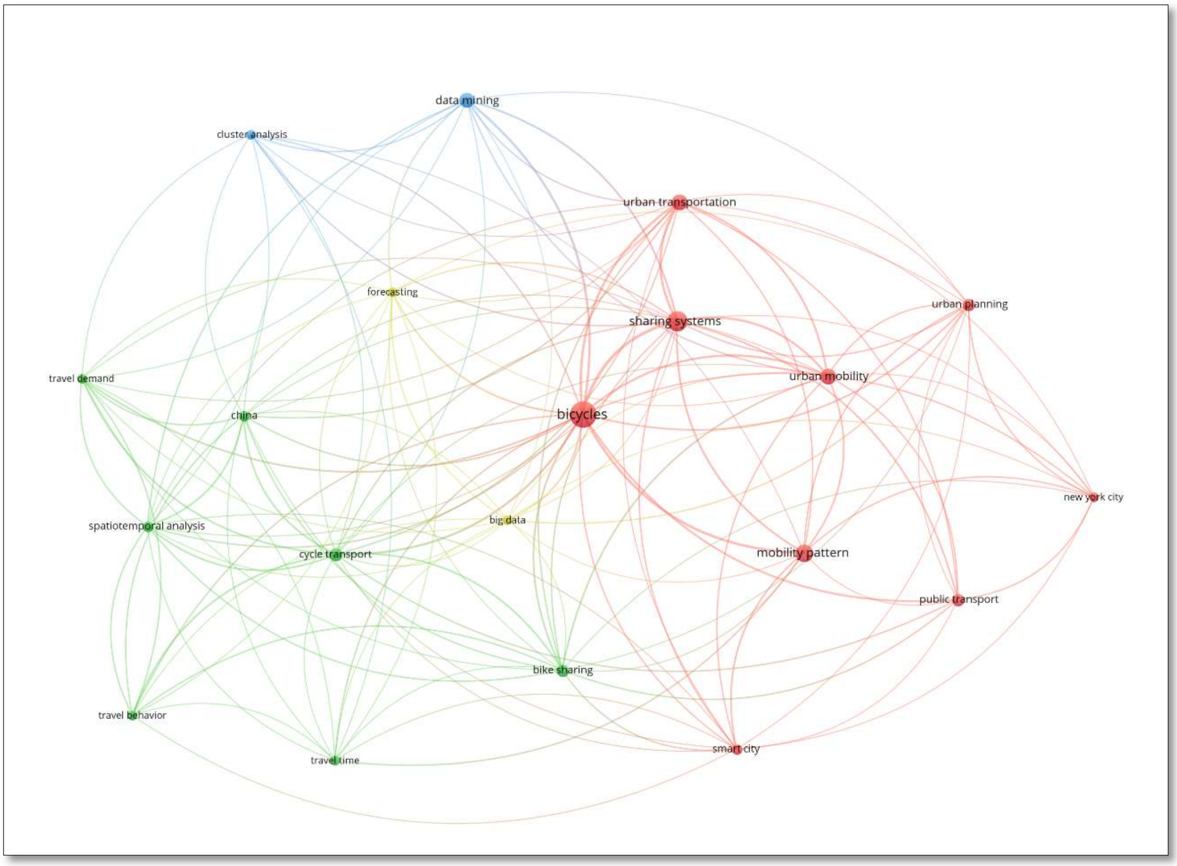

Figure 2.2: Keyword occurrence network visualization

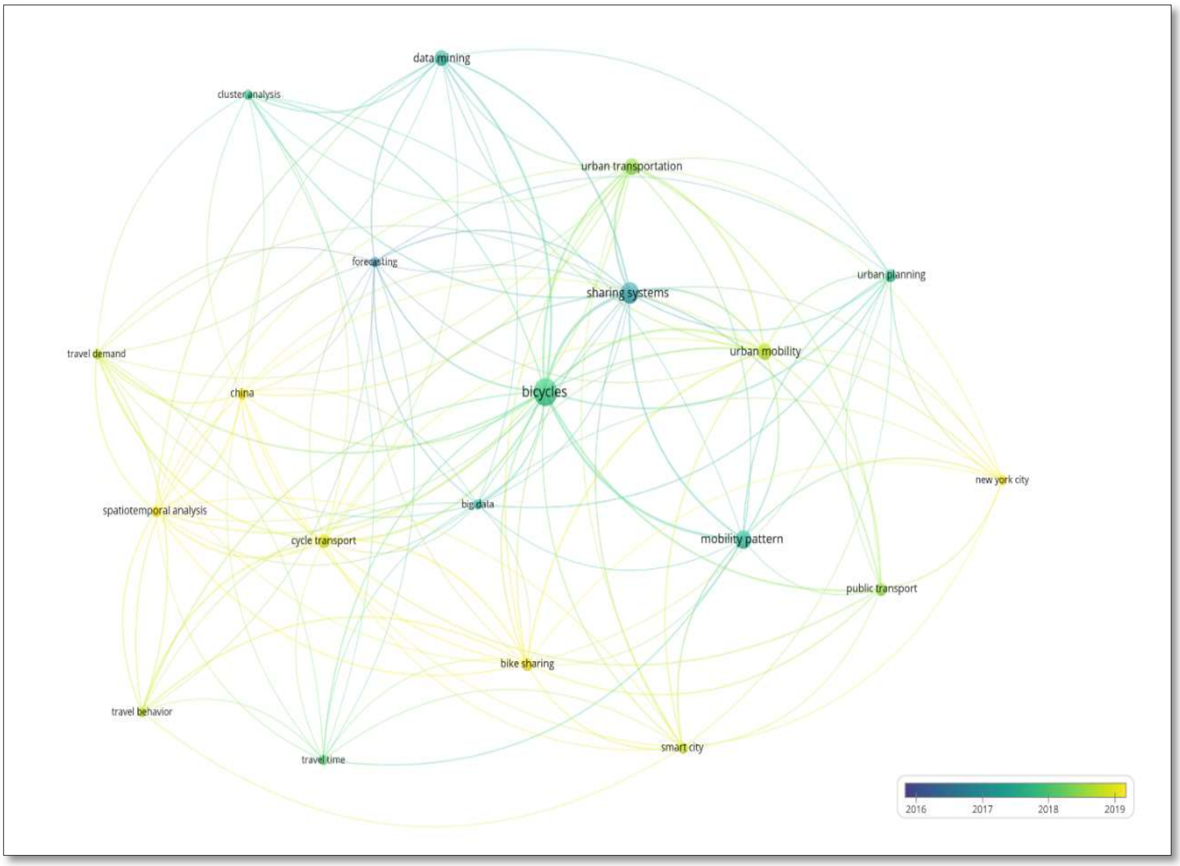

Figure 2.3: Keyword occurrence network overlay visualization

Secondly, we made an author co-authorship type of analysis using the counting method of full counting. We selected the maximum number of authors per document at 10 with the minimum number of documents of an author of 1 meeting the 131 authors. Finally, we selected the number of authors of 90 (Figure 2.4).

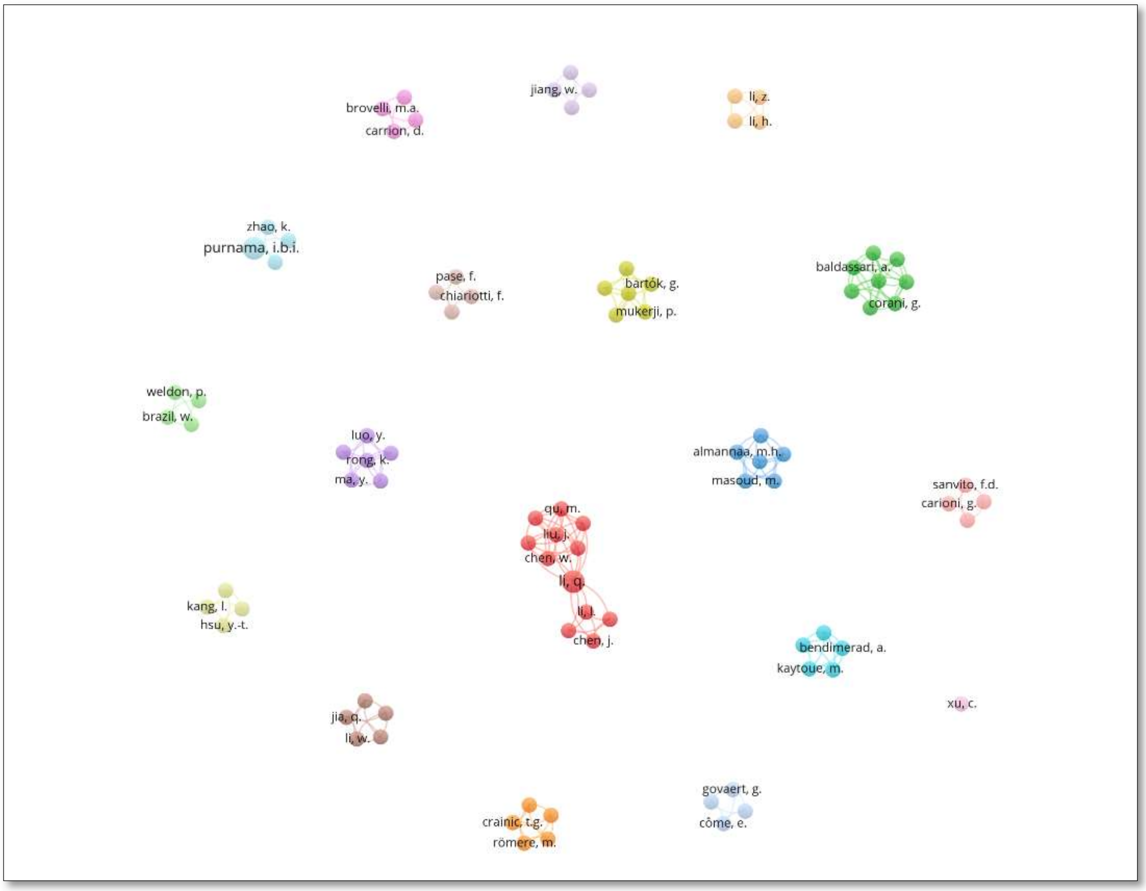

Figure 2.4: Author and co-author network visualization

The top identified author was Li, Q. [10], [11] with a total of 2 documents and with a link strength of 11. It also identified the authors Baldassari, A. [12], Cellina, F. [12], Chen, W. [10], Corani, G. [12], Fu, Y.[10], Förster, A. [12], Guidi, R. [12], Liu, J. [10], Pampuri, L. [12], Qu, M. [10], Rizzoli, A. E. [12], Rudel, R. [12], Xiong, H. [10], Yang, J. [10] and Zhong, H. [10] all with a total of 1 document and a link strength of 7. In this analysis, it was identified 18 clusters with 90 items and 195 links. In the overlay visualization, we can see that the first cluster corresponds to authors who published articles from 2016 to 2020. Three clusters correspond to authors who published articles in 2014/2015. In Figure 2.5 it is possible to see that the majority of the co-authorship publications were made in the range of 2018 to 2020.

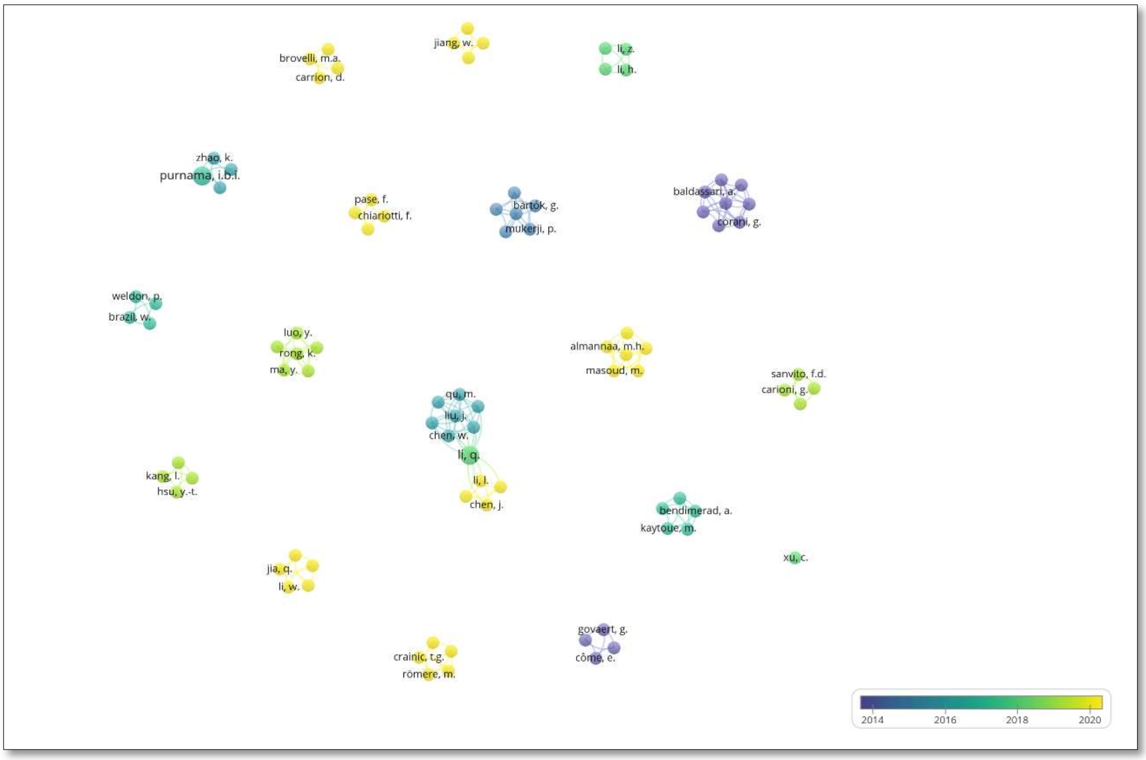

Figure 2.5: Author and co-author network overlay visualization

Table 2.2: Author and co-author ranked by total link strength

Author	Documents	Link Strength
Li, Q.	2	11
Baldassari, A.	1	7
Celina, F.	1	7
Chen, W.	1	7
Corani, G.	1	7
Fu, Y.	1	7
Förster, A.	1	7
Guidi, R.	1	7
Liu, J.	1	7
Pampuri, L.	1	7
Qu, M.	1	7
Rizzoli, A.E.	1	7
Rudel, R.	1	7
Xiong, H.	1	7
Yang, J.	1	7
Zhong, H.	1	7
Almannaa, M.H.	1	5
Ashqar, H.I	1	5
Bartók, G.	1	5
Elhenawy, M	1	5
Krause, A.	1	5
Luo, Y.	1	5
Ma, Y.	1	5
Mangalagiu, D.	1	5
Masoud, M.	1	5
Meenen, M.	1	5
Mukerji, P.	1	5
Rakha, H.	1	5
Rakotonirainy, A.	1	5
Rong, K.	1	5
Santoni, M.	1	5
Singla, A	1	5
Thornton, T.F	1	5
Wang, Y.	1	5
Bendimerad, A.	1	4

Third, we also analyzed the terms extracted from the title and abstract fields using a full counting method with a minimum number of occurrences threshold of 10. The number of terms selected was 19. In total, we have 14 items grouped in 5 clusters. These 5 clusters have 89 links within them and have a total of 3711 total link strength. Cluster 1 consists in 9 items such as "bike", "city", "data", "demand", "pattern", "station", "system", "trip" and "user". Cluster 2 consists of “bike-sharing”, cluster 3 in “day”, cluster 4 in “approach” and cluster 5 in “mobility pattern”. Our overlay analysis shows that these terms are found in articles between the range of 2017 and 2018, being 2017 more related with the "system", "station" and "user" and 2018 more related with "mobility patterns", "day" and "city".

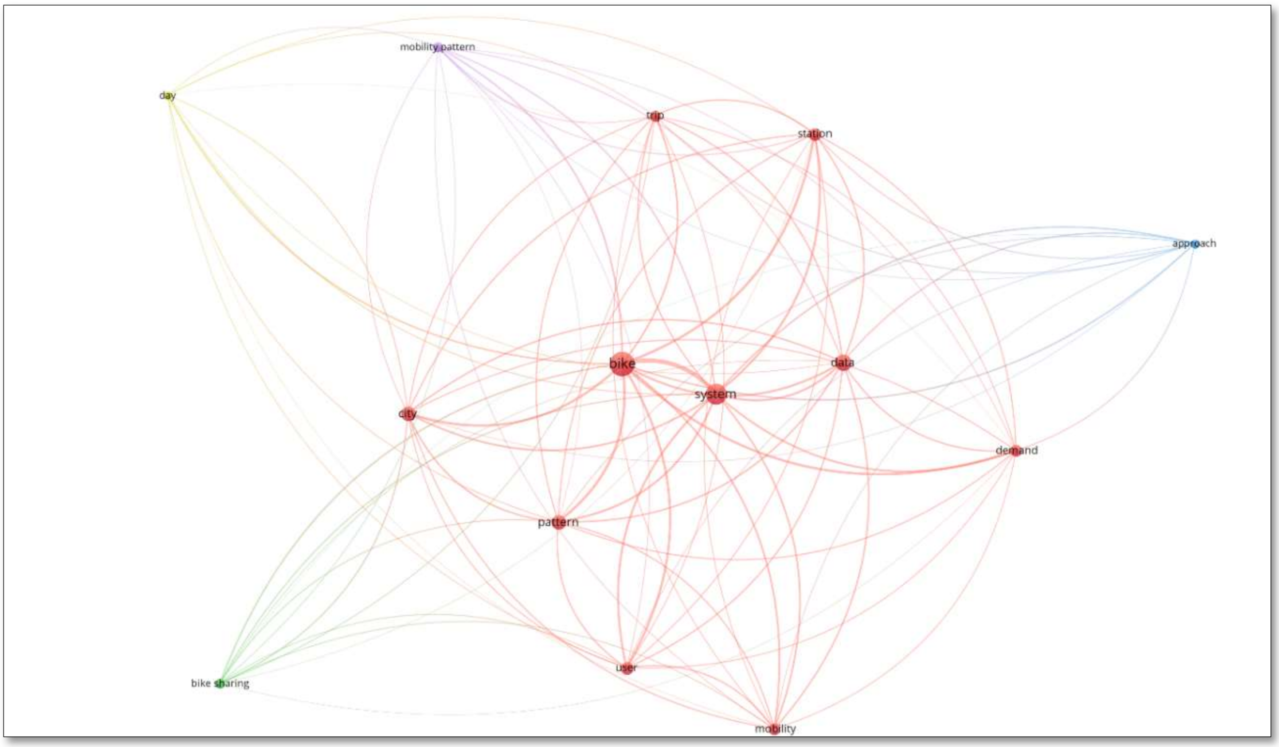

Figure 2.6: Title and abstract text occurrence network visualization

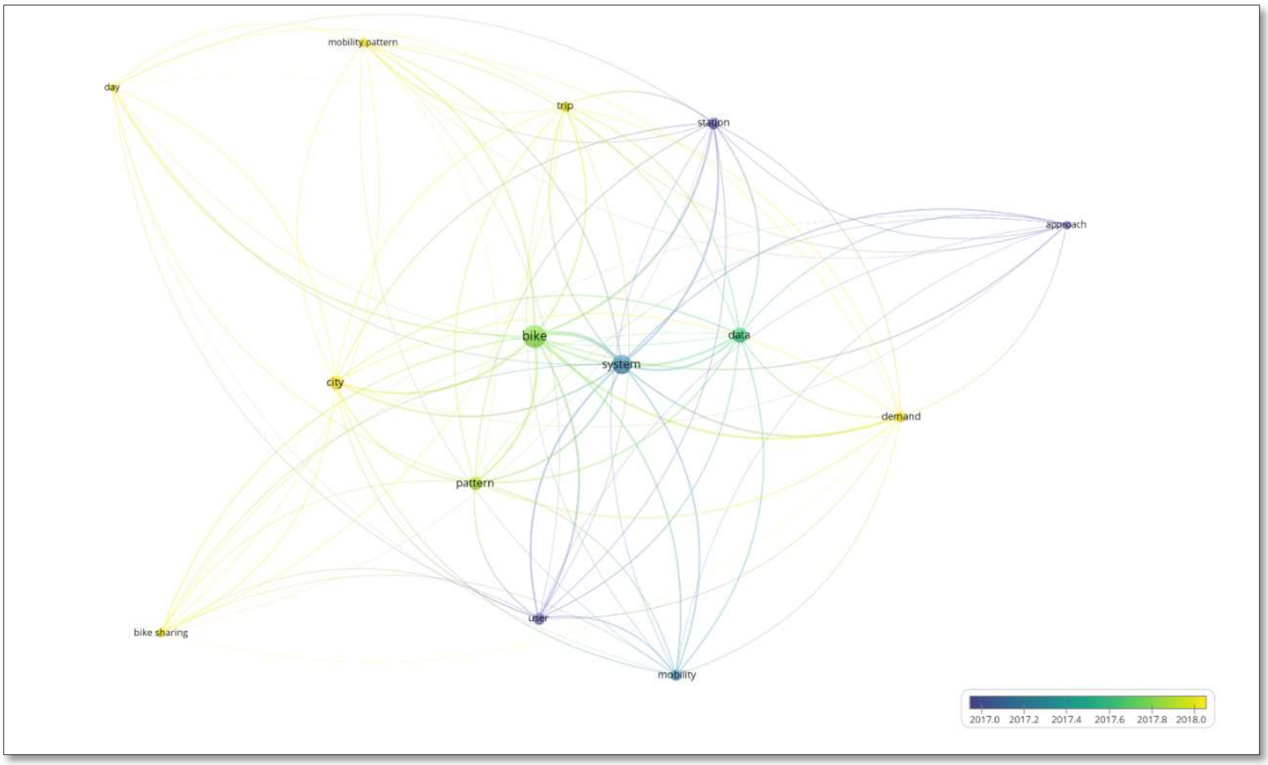

Figure 2.7: Title and abstract text occurrence network overlay visualization

2.2.3 Results Synthesis

To help to identify the most cited publications, and relevant methods and applications in the literature, an analysis was made shown in Table 2.3. For this analysis, we only used journal papers since it was our main publication type.

Table 2.3: Methods and applications ranked by number of citations

Authors	Title	Year	Cited by	Publisher	Document Type	Method	Application
Etienne C., Latifa O.	Model-based count series clustering for bike sharing system usage mining: A case study with the vélib' system of Paris	2014	104	Association for Computing Machinery	Article	Statistical analysis; Cluster analysis	Station usage through departure and arrival counts.
Caulfield B., O'Mahony M., Brazil W., Weldon P.	Examining usage patterns of a bike-sharing scheme in a medium sized city	2017	94	Elsevier Ltd	Article	Statistical analysis.	Trip patterns analysis.
Xu C., Ji J., Liu P.	The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets	2018	81	Elsevier Ltd	Article	Statistical analysis and Deep learning (LSTM NN)	Analyze and forecasting
Teixeira J.F., Lopes M.	The link between bike sharing and subway use during the COVID-19 pandemic: The case-study of New York's Citi Bike	2020	56	Elsevier Ltd	Article	Statistical analysis.	Trip patterns and intermodal analysis.
Nikitas A.	Understanding bike-sharing acceptability and expected usage patterns in the context of a small city novel to the concept: A story of 'Greek Drama'	2018	56	Elsevier Ltd	Article	Statistical analysis.	Survey analysis.

For table 2.3, we considered the top 5 journal articles by number of citations. We found that Etienne C. and Latifa O. [13] paper is the most cited one with 104 citations, followed by the paper of authors, Caulfield, B., O'Mahony, M., Brazil, W., Weldon, P [14] with 94 citations, then Xu C., Ji J., Liu P. [15]. third with 81 citations Teixeira J.F., Lopes M. [16] and Nikitas A. [17] fourth and fifth both with 56 citations.

Methods used in these journal papers are related to statistical and cluster analysis applied to trip patterns and station analysis. These methods and applications are aligned with the objectives proposed in this thesis in section 1.

Our thesis provides new insights into Lisbon BSS, first implemented in 2017 and still evolving. It was interesting to analyze the evolution and strong BSS demand in a city that did not have a cycling culture until recently. BSS as well as electric mobility transport modes can have a positive effect on transportations and environments [18] and there are similarities in the patterns and, in turn, in the characteristics of the people who use this type of transport. A similar study like this one was made in Italy [19], [20] with data ranging from 2015 to 2018, starting from the analysis of the mobile patterns with weather conditions combined, and then ending in the station clustering using K-means. Also, it was made a comparative study regarding mobility patterns between New York and London [21] regarding the usage patterns.

We also found that weather conditions [14], [22], [23] had an important impact on travel behavior. In Mexico City, there are three clusters and the largest number of inbound and outbound trips corresponds to the same station [24]–[27]. The rebalancing of the stations [28], [29] was not possible due to the lack of data on bikes per station. For instance, in Italy [30] and Canada [31] there was a study that analyzed the stations in an attempt to rebalance the stations in a closed queue system as well as in Munich [32] through GPS data.

The impact of the COVID 19 pandemics can affect the urban mobility patterns as seen in New York [33], [34]. There could be a strong impact in terms of patterns due to COVID 19 and the weakening of the economy for example different routes because of clearer traffic or even because there is more traffic. The challenge is to provide more accurate data in real-time so that the best path can be chosen [35].

Lisbon BSS trip patterns are thus similar to other observed BSS of medium-size cities [14] discussed in the State of the Art section, such as patterns found in short and frequent trips and ride peaks observed in the morning and afternoon, as in the case study of the city of Cork (Ireland) [14].

Parallels with larger cities can be established as well. In Canada, for instance, Montreal's BIXI BSS [36] is mainly used on weekdays, evenings, and weekends. In Toronto, bike trips are shorter on weekdays mornings [22]. In China, the morning and evening peaks correspond to weekdays and on the weekend, we see often the evening peaks [37], [38], [39].

Large USA cities BSS studies [40]–[42] show frequent bike use in the morning and afternoon peaks [40] and different usage patterns between weekdays and weekends, identifying longer trips on the weekend [40].

In large European cities, weekday morning trips in the peak hour [43], [44] reach a higher speed than trips over the weekdays and weekends.

In Austin [45] there is a difference between e-bikes and e-scooters and in e-bikes, there is a difference in average speed in commuting and for recreational purposes.

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, not only in analyzing the patterns but also in creating crowdsourcing techniques to apply to smart cities [46]. Further work needs to be conducted regarding Lisbon BSS in the scope of urban analytics [47] and parallel comparison with other BSS implemented nationally and internationally. The rebalancing of the bikes through stations is also a challenge for this BSS knowing the clusters in which the trips occur and the stations specifically [48].

Lisbon BSS's 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. Prediction of mobility patterns [49] with machine learning algorithms [50] is also a future work possibility for further years.

3 Data Mining Process

3.1. Methodology

In our approach, we applied the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology [51] (Figure 3.1) being appropriate to manage big data. 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 preprocessing transforms the data into useful information used for the next phase. It involves cleaning, reduction, transformation, and integration of data.

Phase 4 – Modelling: modeling technique is selected and built the model.

Phase 5 – Evaluation: the chosen modeling technique is evaluated according to its objectives according to the results produced in the process.

CRISP-DM ensures the quality of knowledge discovery in the project results, requires reduced skills for such knowledge discovery, and reduced costs and time [52].

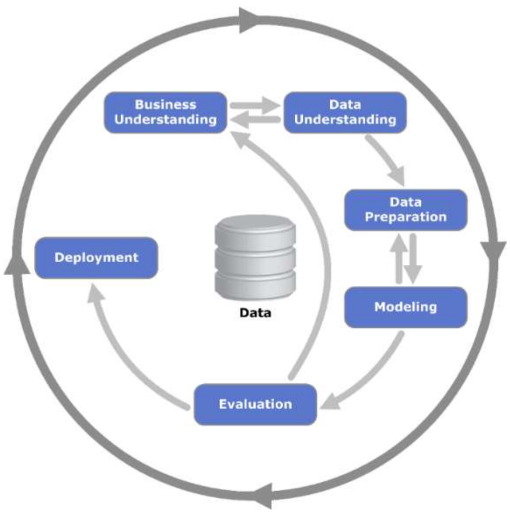

Figure 3.1: CRISP-DM Methodology Diagram

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 user's 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 the data pre-processing phase, we performed data cleaning, removing noise in the data so that further analysis would not be affected by the data itself.

The model phase, allows the application of statistical and machine learning techniques, enabling the 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.

For the evaluation phase, the CML evaluators responded to an inquiry with the following criteria: Utility, Understandability, Accessibility, Level of detail, Consistency.

For our data cleaning (Figure 3.2) there are 3 phases: extraction, transformation, and then visualization. To obtain the insights, it is relevant to perform cleaning, conformance, and normalization processes in the data sets, to obtain correct, complete, consistent, accurate, and unambiguous data [53].

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 pre-processing, cleaning, and normalization, which 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 analyzed according to the data characteristics in different levels, intending to understand the evolution of bike ridership and the impact of built environment and pandemics.

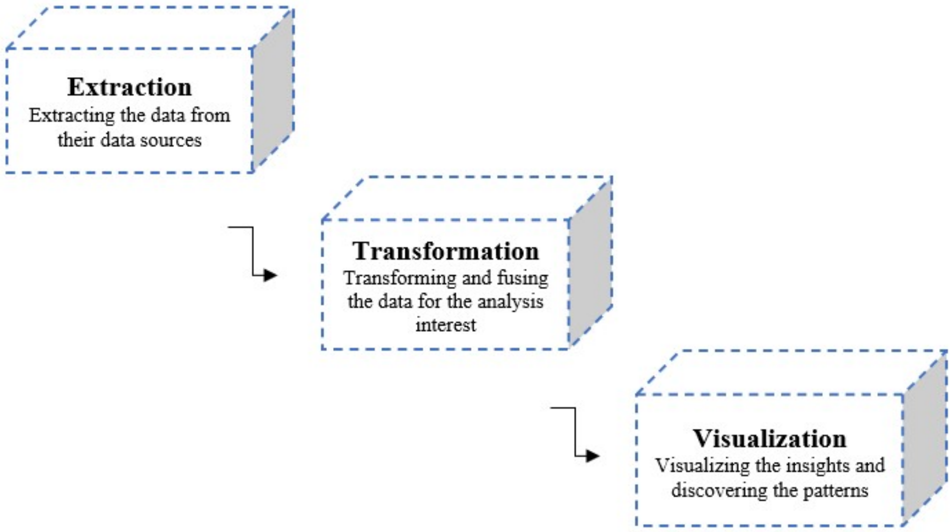

Figure 3.2: Data preparation phases (ETL adaptation)

Our data analysis and visualization were performed in Python [54] using the Jupyter Notebook platform [55]. Data cleaning, pre-processing, analysis, and visualization were performed using different Python libraries, according to the application's purpose. “Numpy” [56], “Pandas” [57], “Matplotlib” [58], “Seaborn” [59], were used for statistical analysis and visualization. “GDAL” [60], “Shapely” [61], “Folium” [62], “Fiona” [63], were used to visualize spatial analysis. Our data science algorithms used “Scikit-learn” [64] to perform K-Means, Train-test split, and Accuracy Score.

3.2 Business Understanding

Data were provided in the scope of Lisboa Inteligente [65] challenges of the Lisbon City Hall, namely challenges #4 “Are there mobility patterns in Lisbon BSS”, and #49 “Determine COVID 19 pandemic impact in mobility and environment [66].

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 the 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 behavior patterns over the 3 analyzed 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.

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).

3.3 Data Understanding

3.3.1 GIRA 2018

Bike station data schema (Table 3.1) 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 3.1: 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

Bike trip data of 2018 (Table 3.2), 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).

Table 3.2: Lisbon BSS bike trip data of 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)

3.3.2 GIRA 2019 and 2020

Bike trip data of 2019 and 2020 (Table 3.3) is characterized by date (dd/mm/yyyy) and trips per day ranging from 1st January 2019 to 4th June 2020.

Table 3.3: Lisbon BSS bike trip data schema of 2019 and 2020

Characteristics	Description
data	Date
viagens	Bike trip count

3.3.3 Avenida Duque de Ávila 2019 and 2020

Bike count data of 2019 and 2020 (Table 3.4) 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 the 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 3.4: 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

3.3.4 IPMA 2018

The IPMA weather data of 2018 (Table 3.5) 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 3.5: IPMA data schema of 2018

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

3.3.5 IPMA 2019 and 2020

The IPMA weather data is structured with 18 variables in 2019 and 11 variables (Table 3.6) in 2020. The variables marked with * are only provided for 2019 data, and the others are both for 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 3.6: IPMA data schema of 2019/2020

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
intensidadedevento km	Wind intensity
position	Station position
preacumulacada	
pressao	Atmospheric pressure
radiacao	Radiation
temperatura	Temperature
validity_ts*	

3.4 Data Preparation

Lisbon BSS data, from a fourth-generation system, provides extensive information. However, data extraction methods have not yet been extensively explored [67], 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 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 3.2 and Table 3.6), and another combining bike trips data and bike station data (see schemas in Table 3.1 and Table 3.2), 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 station's field. We´ve developed an Extract, Transform and Load (ETL) process to generate these datasets, load data from external databases, transform the data by creating common columns and joining the datasets, and finally load the data into our research work. As a result, from the 3 data schemas presented in Tables 3.1, 3.2, and 3.6, we derived, via such ETL process, 2 datasets, namely, the “bike trips-stations temporal analysis” dataset (Table 3.7), the “bike trips-stations clustering” dataset (Table 3.8).

Table 3.7: 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 3.8: 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.

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 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 [57] 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.

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

3.5 Data Modeling

3.5.1 GIRA 2018

Looking into the literature, studies in this field aim to understand user´s profile and travel behavior [68]–[70], activity patterns in stations [71], and the impact of the built environment in the BSS [72].

The methods applied focus on 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 [73], [74].

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 [75].

Clustering algorithms combining temporal and spatial attributes variables are also data mining methods used for this analysis purpose. More specifically, K-means clustering [76]–[79], used by McKenzie [80] and Zhong [43] to measure regularity at different scales and to measure spatiotemporal variation and cluster interaction.

3.5.1.1 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 Figure 3.3.

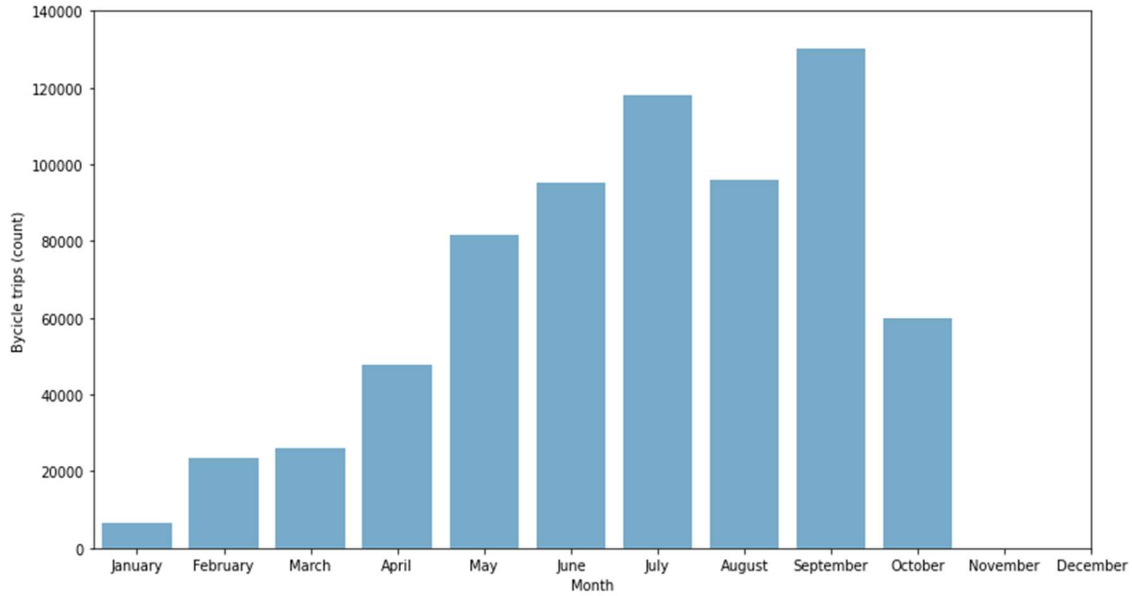

Figure 3.3: Bike usage frequency per month

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 Figure 3.4, 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.

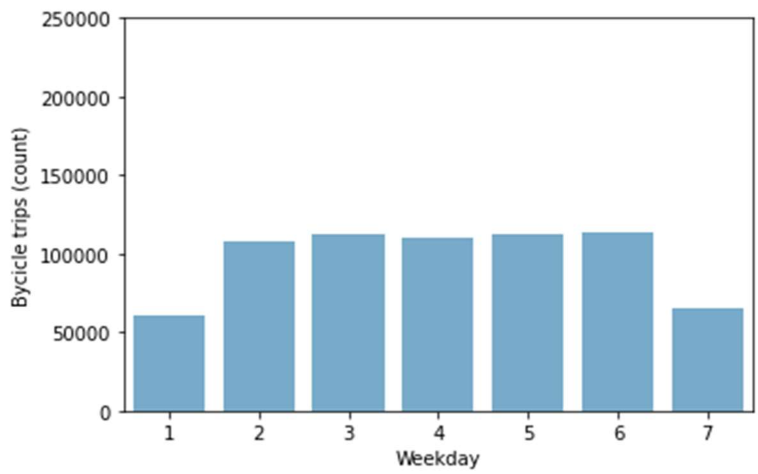

Figure 3.4: Bike usage per weekday

The distribution of trips throughout the different periods of the day was analyzed too. The column date_starts was transformed into a time format, and the hour was extracted to create the column Periodo_dia (Day period). The day was broken down into three-hour groups: Morning: 7 am to 12 am; Afternoon: 12 am– 8 pm and Overnight: 8 pm–7 am. Our analysis shows that most of the trips (56%) occur during the afternoon compared to the morning and overnight periods (Figure 3.5). Additionally, during working weekdays, after the afternoon, the morning period comes second. On 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 types. Regarding bike type (Electric or Conventional), we have observed no noticeable differences in terms of trip distance and duration during weekdays. There is also no noticeable difference in average in terms of speed and duration across the different days of the week.

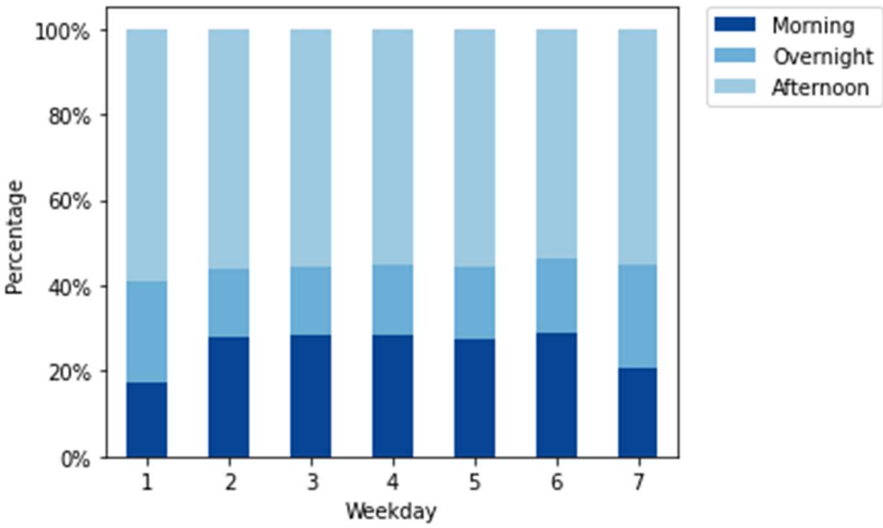

Figure 3.5: Bike usage (%) per weekday within the day

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 (Figure 3.6). 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 Figure 3.6).

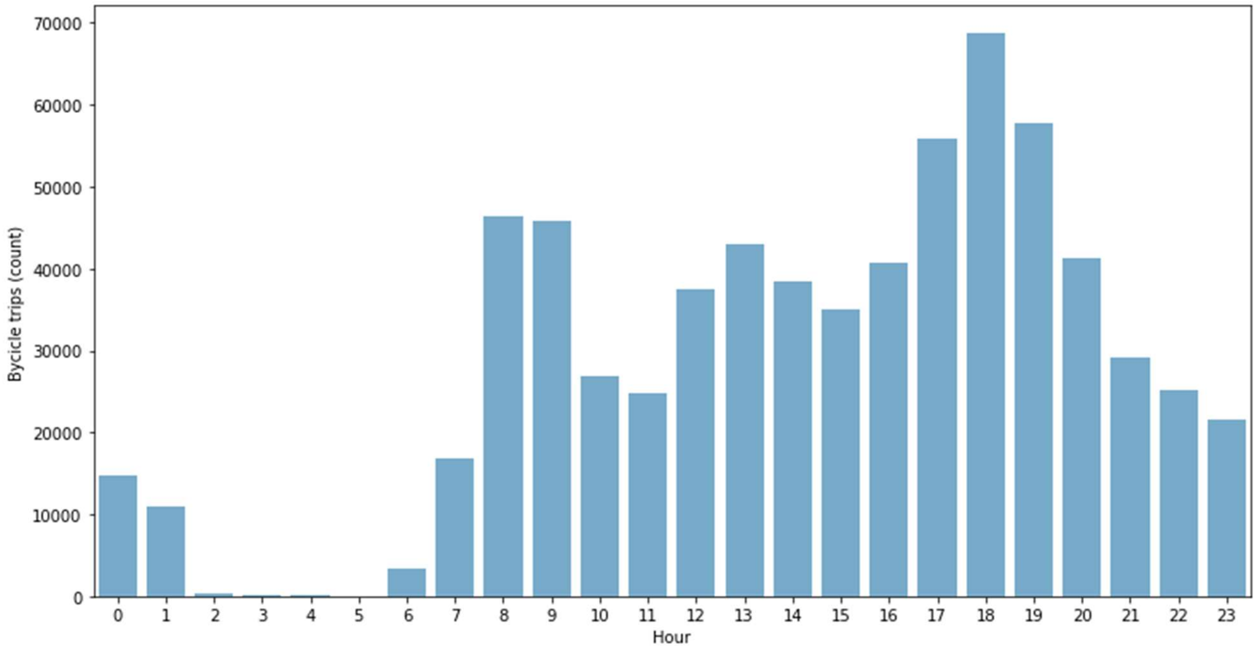

Figure 3.6: Bike usage (trip count) per hour

3.5.1.2 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%) (Figure 3.7). 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: 0º to 10º, 10º to 20º, 20º to 30 and 30º to 42º, being 42º the maximum observed temperature value (Figure 3.8).

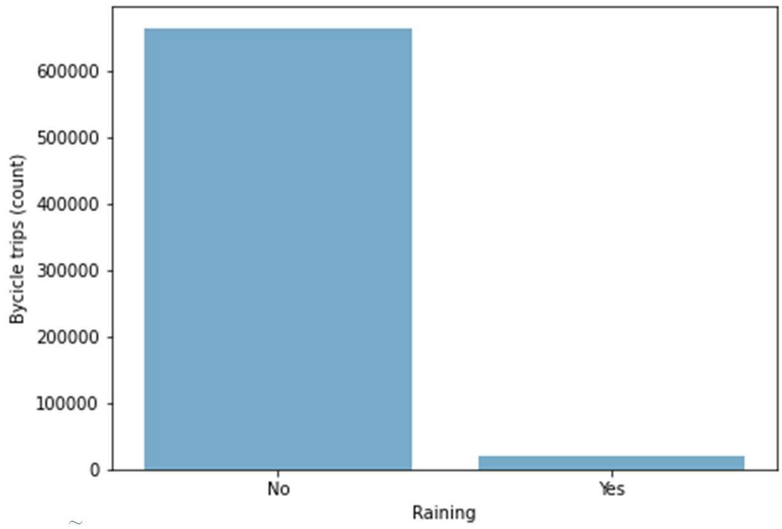

Figure 3.7: Bike usage frequency relation to atmospheric precipitation

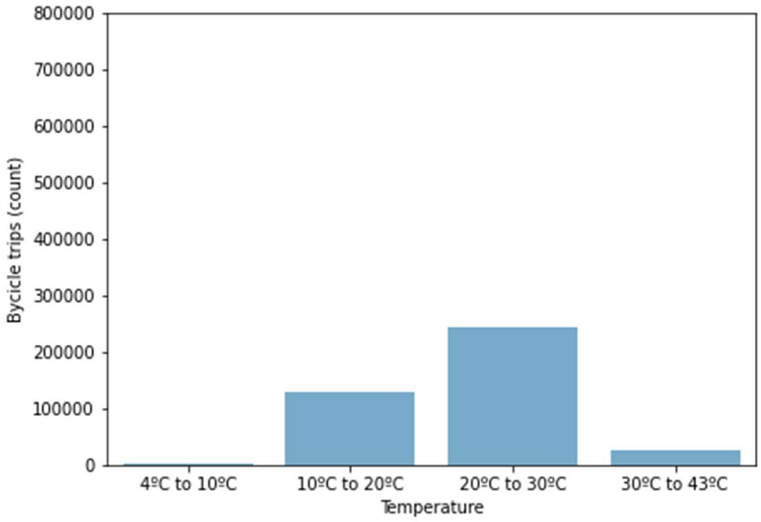

Figure 3.8: Bike usage frequency relation to temperature

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

3.5.1.3 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 in bike usage. Table 3.9 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 the scope of BSS network expansion.

Table 3.9: 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 3.10, 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 rankings in the analyzed months in 2018. The top 5 most popular stations are shown and numbered in Figure 3.9, mapped with all the network bike stations.

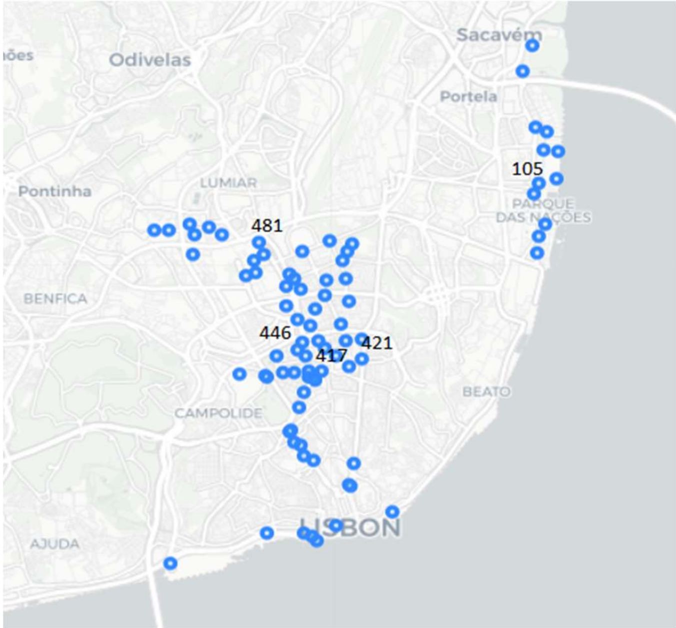

Figure 3.9: Lisbon BSS stations map with the top 5 most popular stations id number identified

Table 3.10: 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

This is also shown in the station trip heatmap (see Figure 3.10), where the orange color corresponds to a higher number of station trips in a gradient to yellow, green, and purple lower station trips.

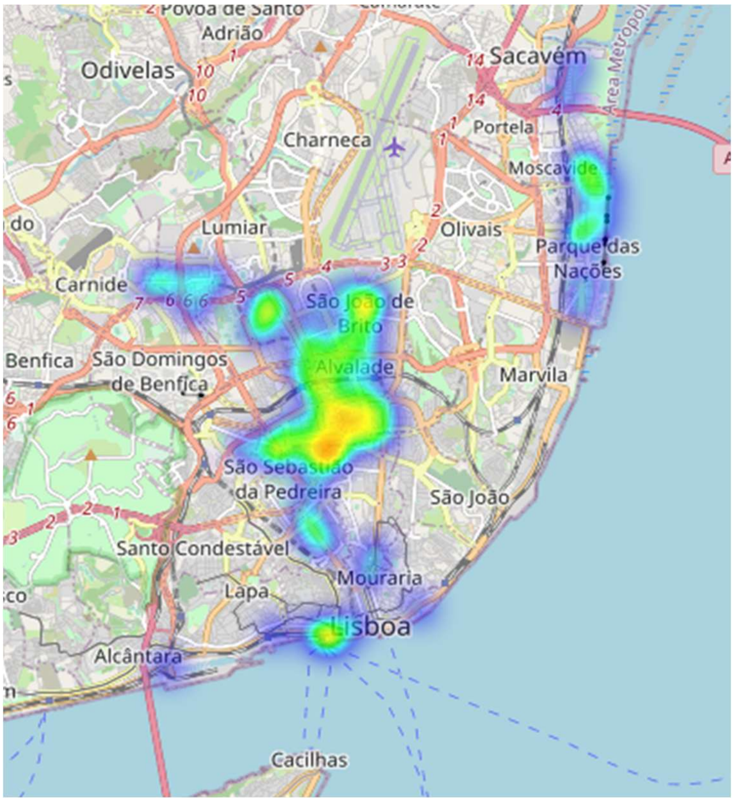

Figure 3.10: Station trip heatmap (The orange color corresponds to a higher number of station trips, whereas blue to a lower).

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 3.11, 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 3.11: 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 3.12) 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 3.12: 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 weekends, we observe that in the weekdays (Table 3.13), 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 on weekdays are also observed in the Campo Grande to Saldanha axis. This also corresponds to the 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 3.13: 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 on the weekends (Table 3.14), can be found in Parque das Nações, likewise 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 Moscavide to station 105 – Centro Comercial Vasco da Gama.

Table 3.14: 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 (Figure 3.11). Patterns previously identified are highlighted in the origin-destination matrix (Figure 3.11).

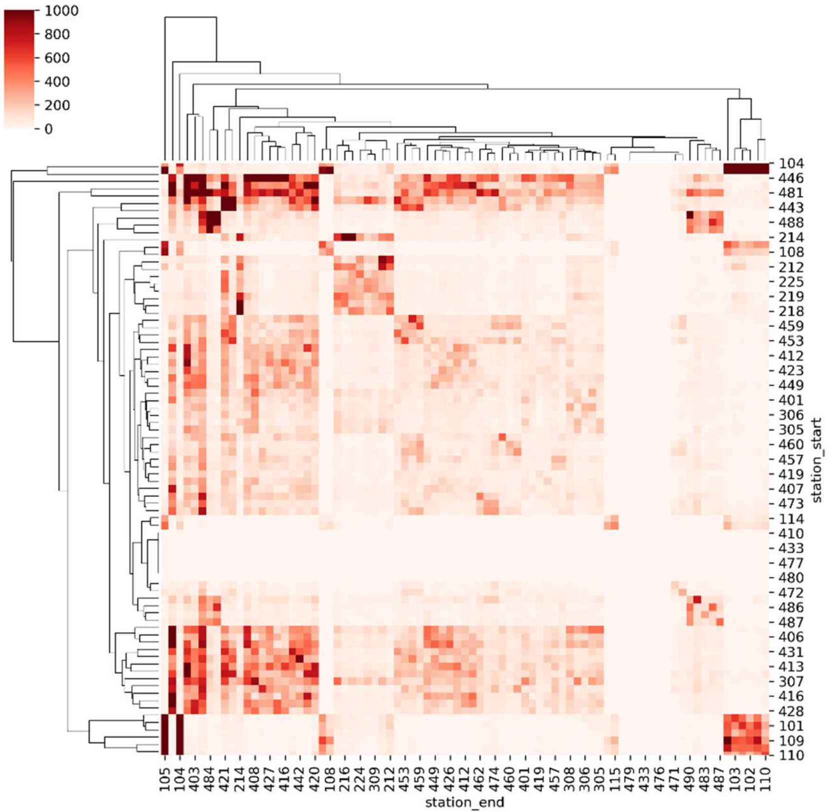

Figure 3.11: Origin-Destination Matrix in 2018

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 being 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 3.8), where the 5 most popular stations (Table 3.9) 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. 3.11) are the same as the top 5 most popular stations.

Regarding the top 5 frequent station pairs on weekdays and weekends, we found that most pair stations are in Parque das Nações, as we also observed in the 2018 analysis. Top 5 frequent stations pairs in weekdays (Table 3.12) are 109-105, 105-109, 107-105, 105-107, and 446-481. The top 5 frequent stations pairs on weekends (Table 3.13) are 109-105, 105-109, 105-107, 107105, and 208-208 (Cais das Pombas/Cais do Sodré). This shows that in August, weekend cycling occurs along the river in Parque das Nações and Cais do Sodré.

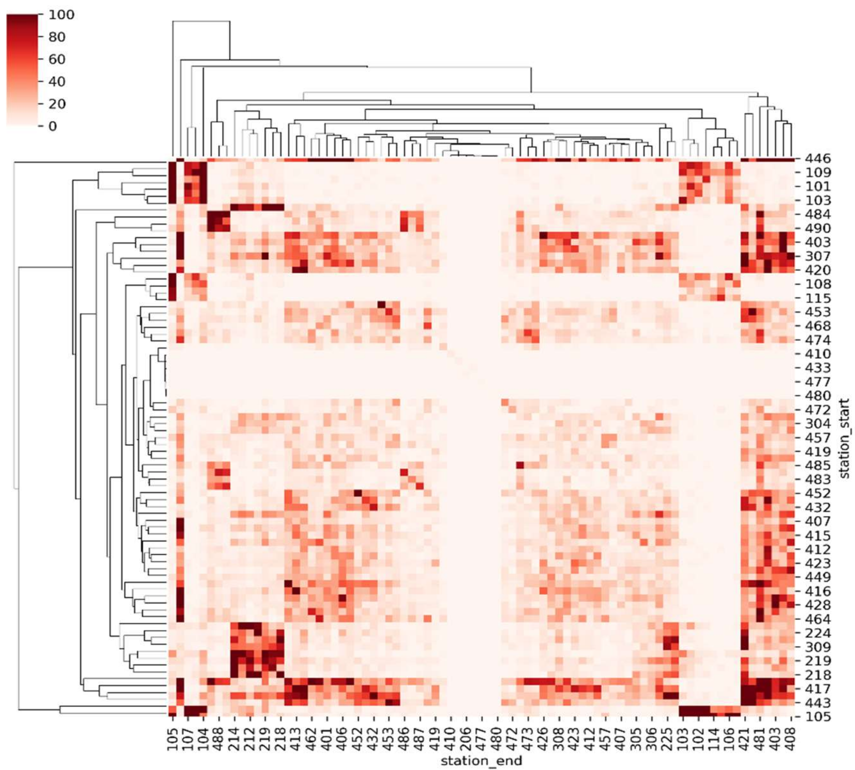

Figure 3.12: Origin-Destination Matrix for August 2018

In September, there was an increase of trips compared to August, with 127,636 trips in 74 stations (Table 3.8) that might be related to work activity return. We identified that the 5 most popular stations (Figure 3.13) are the same as in August but ranked as follows (Table 3.9): 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 on weekdays and weekends show similarities with previously analyzed months. On 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.

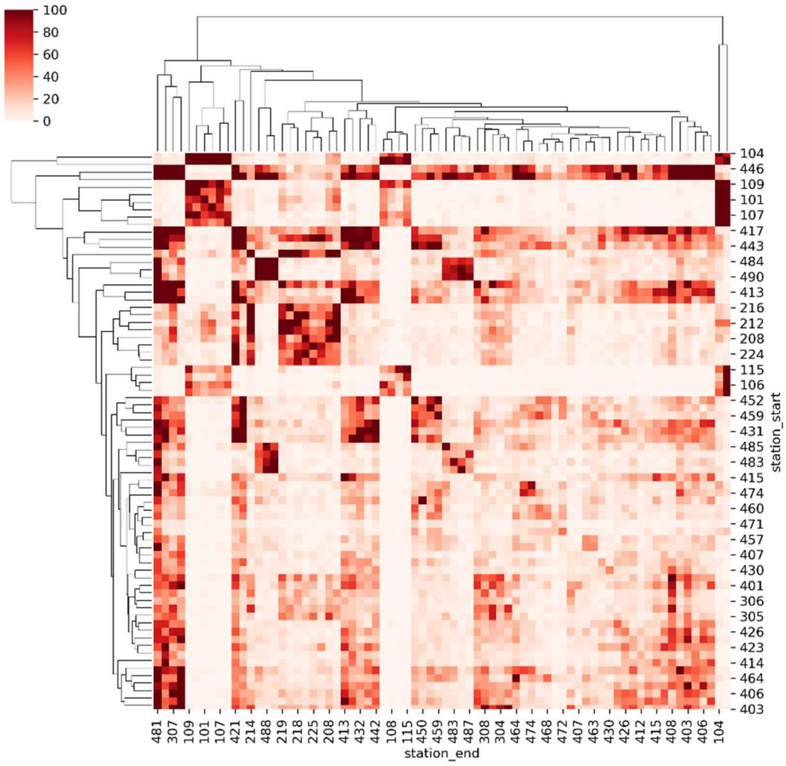

Figure 3.13: Origin-Destination Matrix for September 2018

3.5.1.4 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 a station trips cluster, we first counted all trips of every station, irrespectively if a given station is the origin or destination of a trip. To this aim, we split 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. Afterward, both datasets were concatenated within the station variable and trip count was computed for each station. It resulted in a dataset (Table 4.15) with six variables: station, number of trips, station designation, latitude, longitude, and dock capacity. Latitude and longitude variables were used for the geographical analysis.

Table 3.15: 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

To perform K-means, we used the Elbow algorithm [81], to find the optimal K number through the SSE (Sum of Squared Errors) calculation. As shown in Figure 3.14, the K value of four corresponds to the minimum SSE of the K optimal value. Thus, the four spatial clusters of bike station trips (Figure 3.15) 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 the northwest of Lisbon, in Parque das Nações (in green). Table 3.16 shows cluster centroids of the geographic clustering generated by K-means.

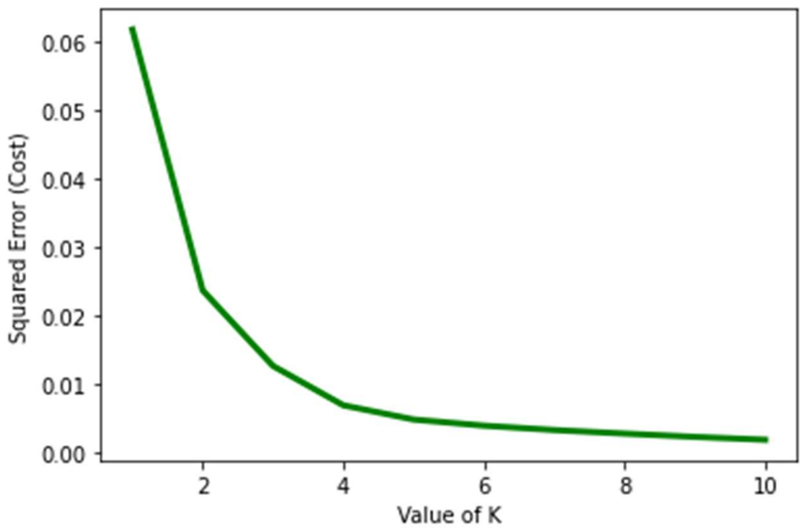

Figure 3.14: Elbow method plot

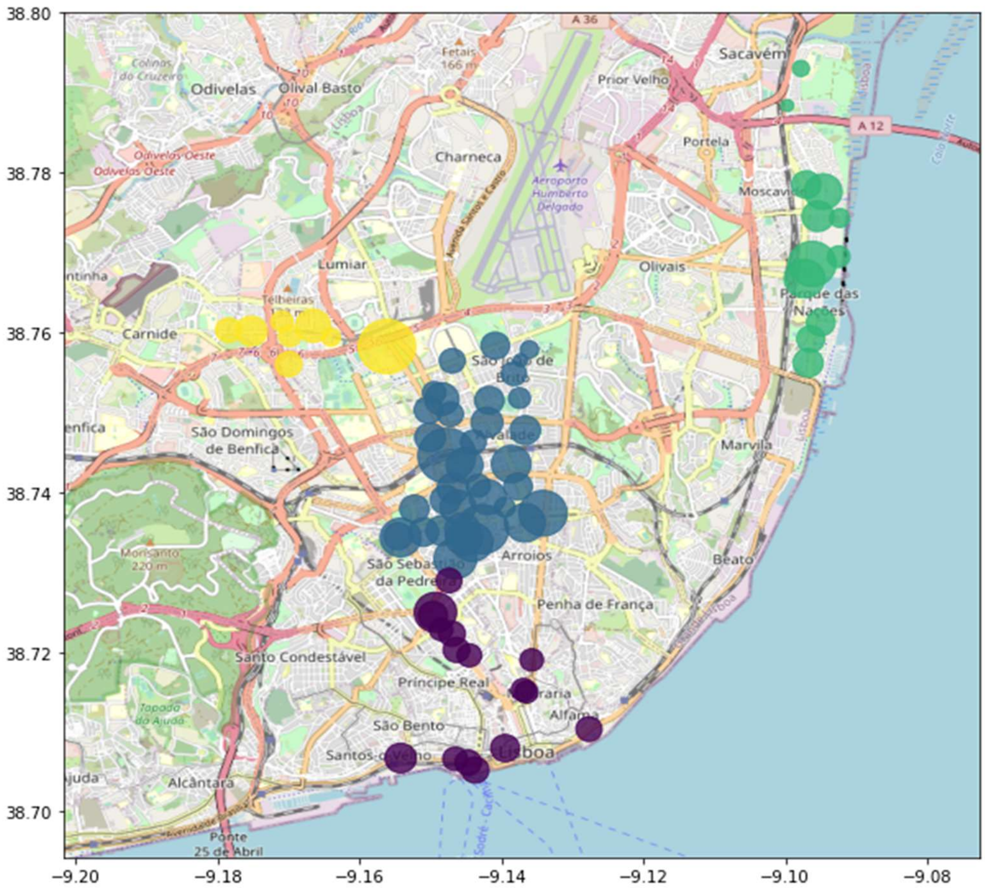

Figure 3.15: Spatial clustering of stations by geography throughout Lisbon (yellow: Telheiras-Campo Grande/Museu da Cidade; blue: Alvalade-Saldanha; purple: Marquês de Pombal Baixa; green: Parque das Nações)

Table 3.16: 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 Figure 3.16) and the four most frequently used stations (labeled in blue) are located in the city center, while the fifth one is in the northeast. These stations correspond to the top five most popular stations, identified in the previous sub-section.

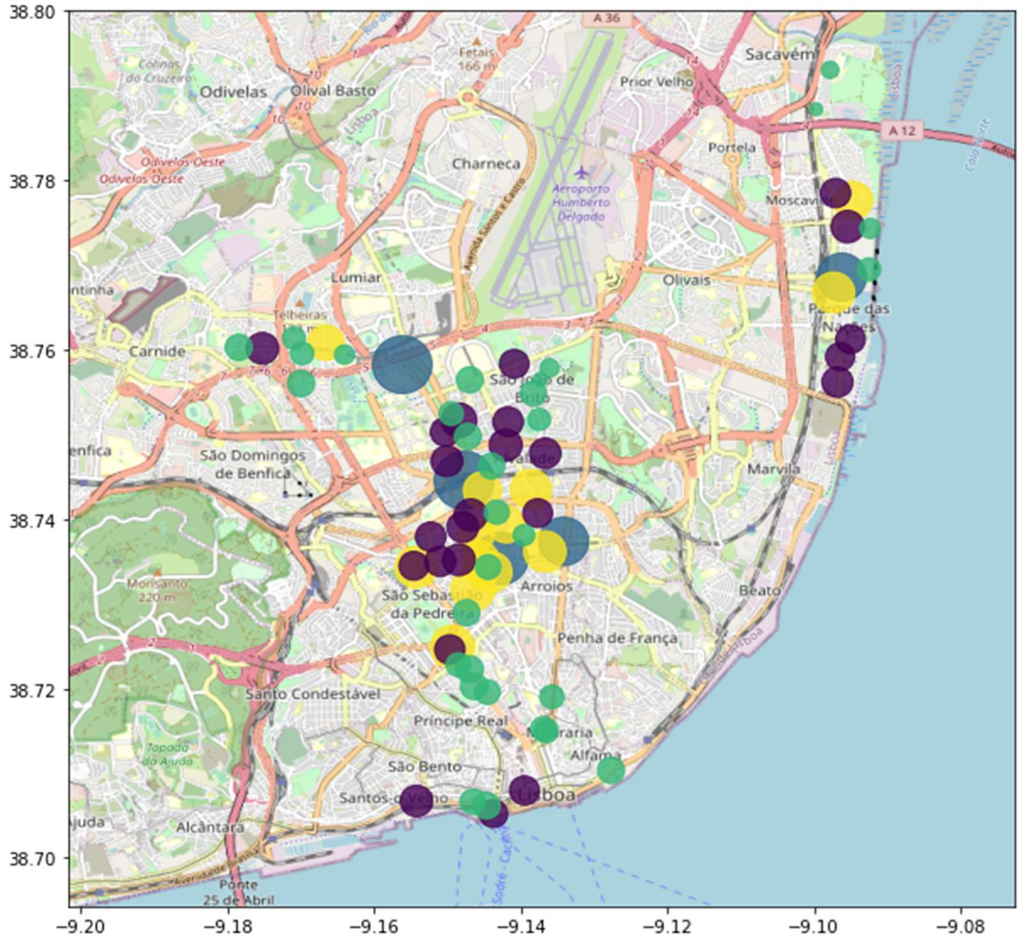

Figure 3.16: Stations clustering of stations by the number of trips throughout Lisbon (blue: first most used stations; yellow: second most used stations; purple: third most used stations; green: fourth most used stations)

3.5.1.5 Dashboard

To summarize the analysis and to answer business questions interactively, we created a dashboard for the 2018 GIRA trips and stations data which can be seen in Figure 3.17. This dashboard was created within Power BI and it shows the 2018 total number of trips and stations on the top left, the monthly, daily, and hourly trips in a bar chart which can be changed by drilling up or down on the chart, a circular chart showing the bike type percentage of use in terms of "Electric" and "Conventional", a funnel chart with the top 5 stations in terms of trips starting and/or ending in that station, and a map with the longitude and latitude of the stations. This dashboard can be used interactively meaning that all the charts are connected. For example, by clicking in a circle of the map (corresponding to a station), it is possible to see the information of the other charts that correspond to the station that is clicked.

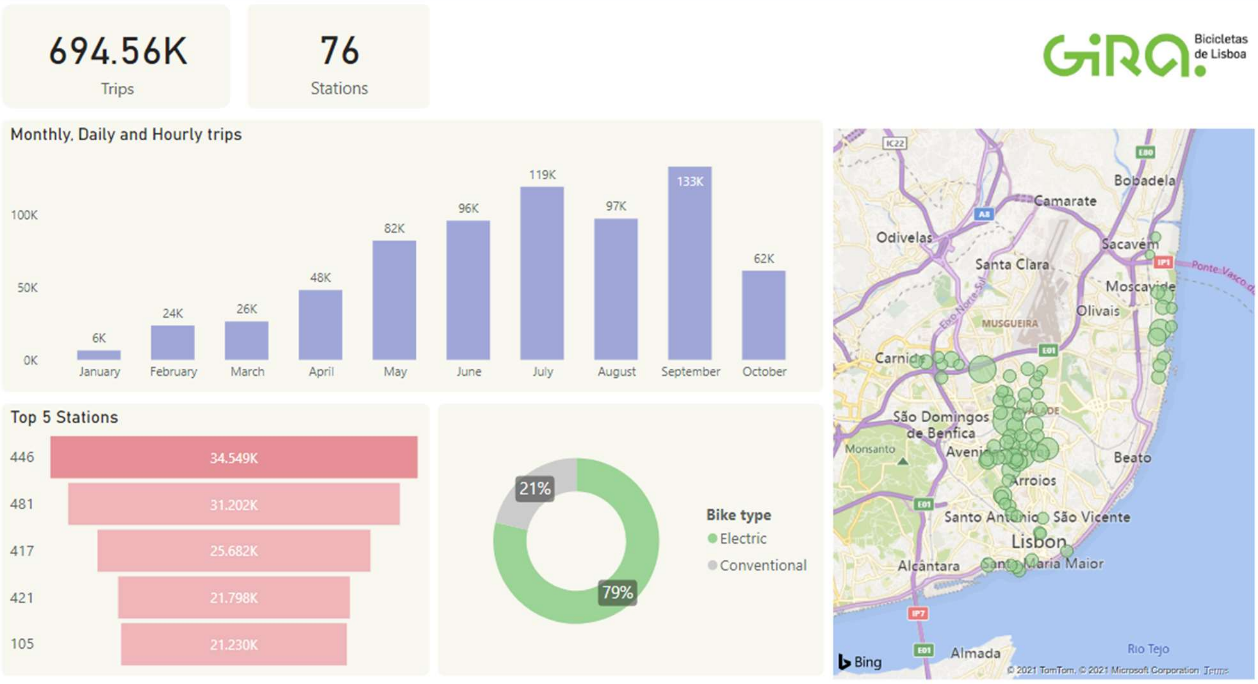

Figure 3.17: GIRA 2018 Dashboard

3.5.2 GIRA 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, into 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, there were 555,429 trips (40%) as seen in Figure 3.18. On the other hand, the months with the lowest usage rate are May, June, July, and August, corresponding to the late Spring and Summer months, with a total of 363,343 trips (26%).

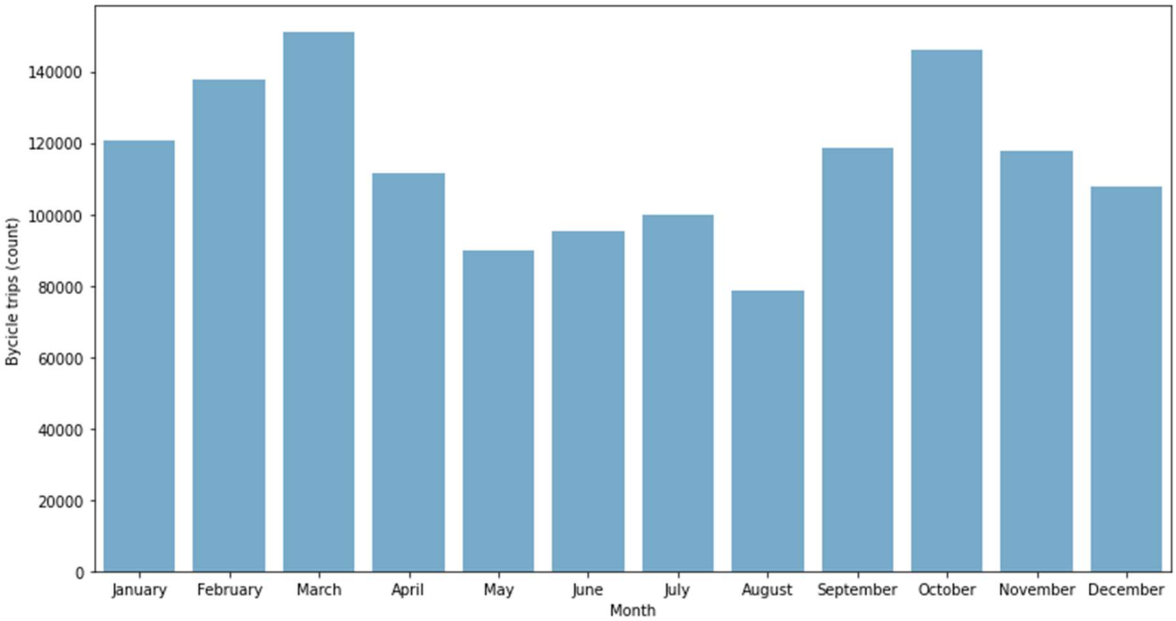

Figure 3.18: Bike count by month in 2019

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 (Figure 3.19) of about 50% (meaning from 80,803 to 40,082 trips). Afterward, 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 the Covid 19 pandemic.

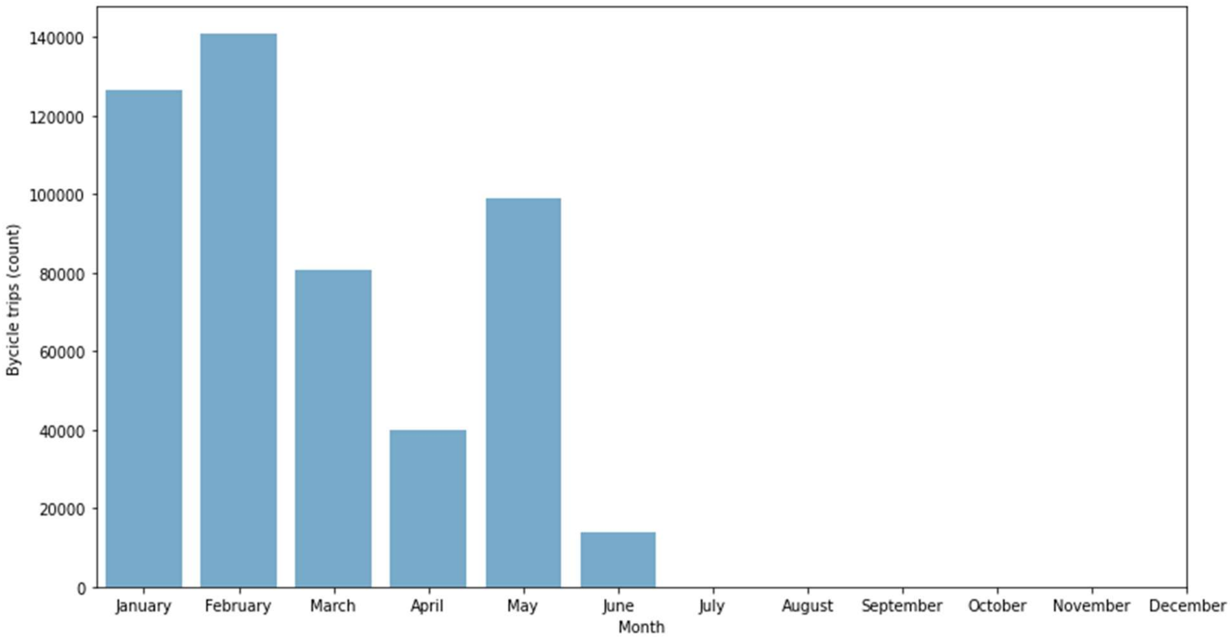

Figure 3.19: Bike count by month in 2020

Moreover, we performed a weekday analysis, applying the same method as in 2018. In 2019, the weekday analysis results showed (see Figure 3.20) that users tend to use BSS mainly on weekdays, representing a total of 1,134,365 trips (83%).

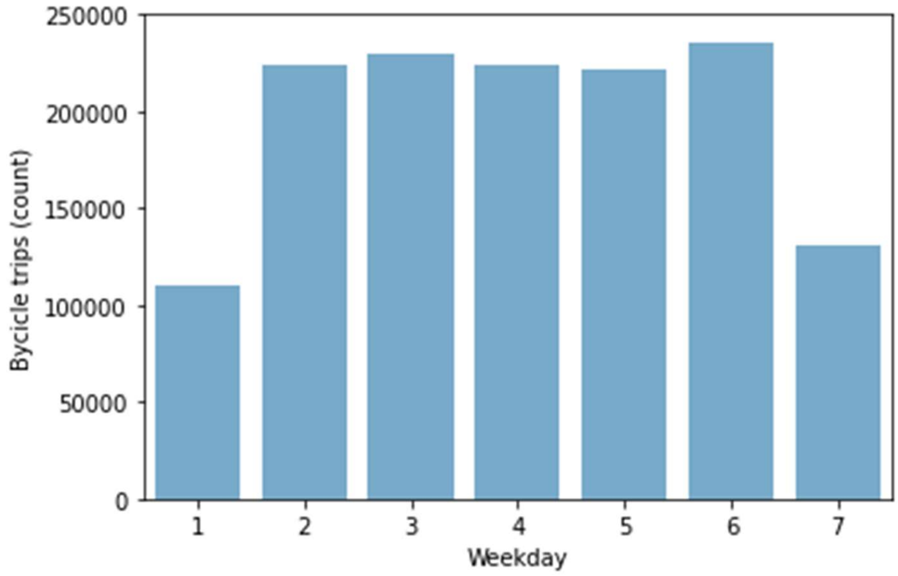

Figure 3.20: Bike trip count by weekday in 2019

In 2020, we can see the same pattern (see Figure 3.21), as in 2018 and 2019, meaning users ride BSS on weekdays. The total number of trips on the weekdays of 2020 was 395,103 (78%).

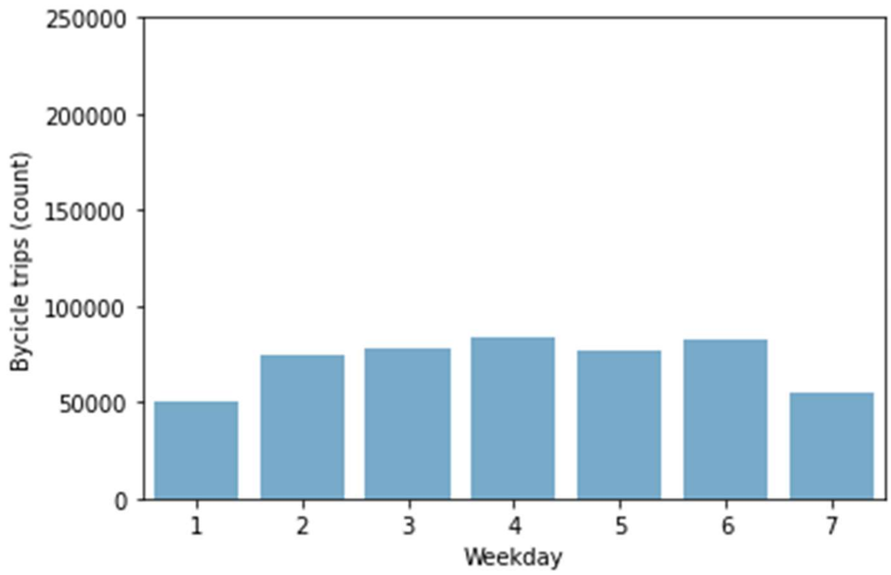

Figure 3.21: Bike trip count by weekday in 2020

Bike trips' count and temperature analysis for 2019 and 2020, are depicted in Figures 3.22 and 3.23. Data were 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: 0ºC to 10ºC, 10ºC to 20ºC, 20ºC to 30ºC, and 30ºC to 43ºC. Results show that the maximum temperature observed in 2019 and 2020 were, respectively, 27ºC and 24,5ºC. Overall, most users prefer to cycle with mild temperatures. A BSS users pattern was observed in 2018, 2019, and 2020.

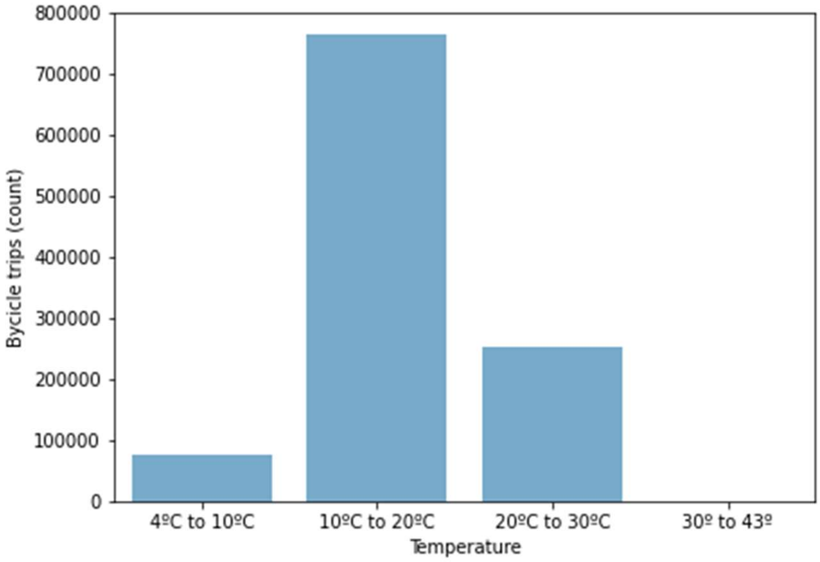

Figure 3.22: Bike usage frequency relation with temperature (2019)

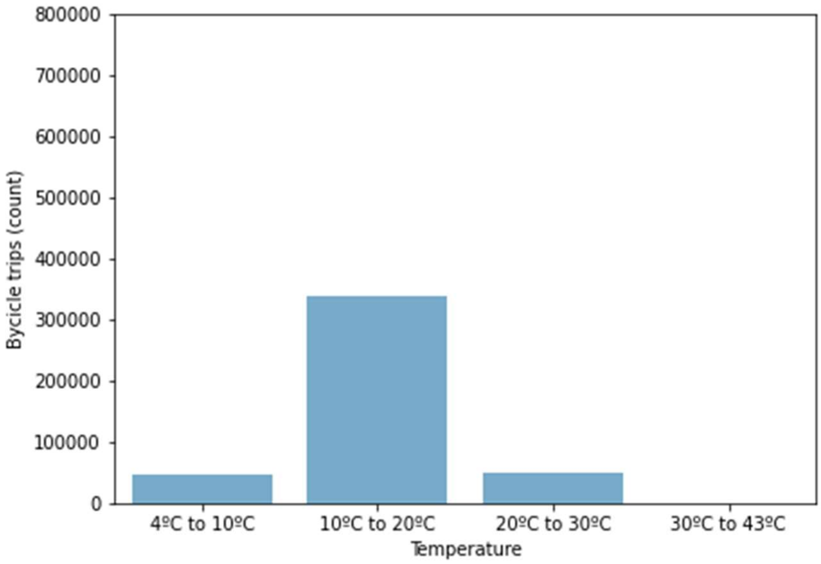

Figure 3.23: Bike usage frequency relation with temperature (2020)

3.5.3 Avenida Duque de Ávila 2019 and 2020

Our analysis shows that the number of weekly trips from East and West is remarkably similar. The average total weekly trips are 815, where East and West range between 412 and 403. Overall, this analysis (see Figure 3.24) 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 misfunction. The second, from the middle of March to May 2020, when the first lockdown restrictions were implemented, due to the COVID 19 pandemic, showing that such an event had a strong impact on Lisbon BSS mobility patterns.

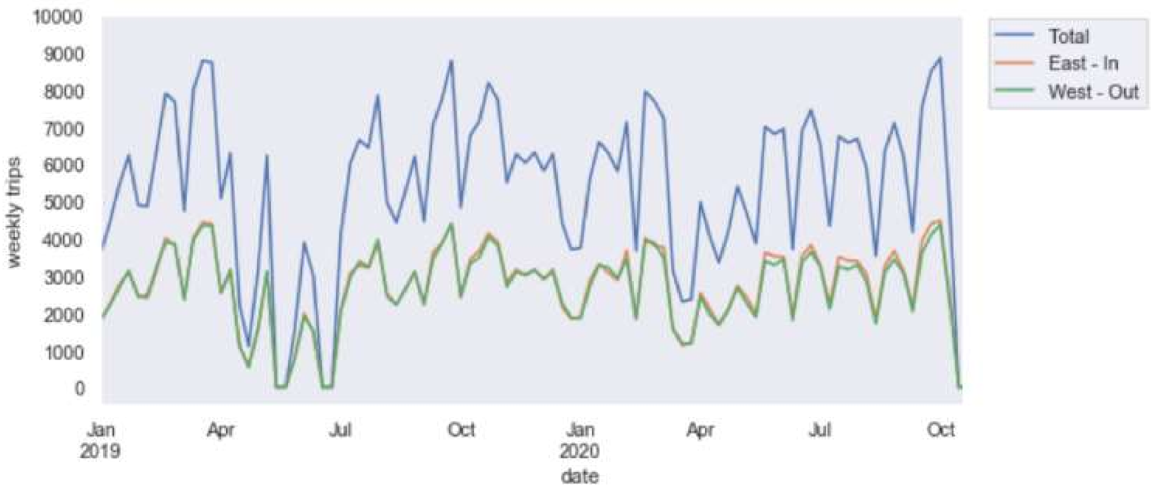

Figure 3.24: Avenida Duque de Ávila weekly bike trip count east, west, and total in 2019 and 2020 captured by sensors

The monthly analysis (Figure 3.25) shows an average of approximately 2,400 total trips. The two-trip count decrease phenomenon was 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 previously mentioned first lockdown.

Figure 3.25: Avenida Duque de Ávila monthly bike trip count east, west, and total in 2019 and 2020

3.6 Evaluation

The evaluation ensures the results match the proposed objectives of this research as well as the veracity of the business needs. Two presentations were made to CML while working on this thesis. This ensured that CML experts followed and supported this thesis development with their knowledge and expertise in the field.

At the end of this research, a questionnaire was sent to the CML experts who attended the presentation, with questions regarding the criteria presented in section 3.1.

Table 3.17: Method Assessment Questionnaire

Criteria	Objective statement	Evaluator 1#	Evaluator 2#
Utility	It can help business decisions regarding BSS.	LA	LA
Understandability	Provides understandable results.	FA	LA
Accessibility	Can be used without training.	FA	FA
Level of detail	Provides knowledge from the mobility patterns.	FA	FA
Consistency	Gives consistent results (compared with another country or company).	FA	LA
Robustness	Has enough detail to be used in other cases of BSS.	LA	PA

This questionnaire was created based on the ISO/IEC TS 33061[3] standards which is a reference model that was mainly used to evaluate software development processes. For evaluation, it was used the NLPF four levels:

- Not Achieved (NA) - [0-15%]
- Partially Achieved (PA) -]15-50%]
- Largely Achieved (LA) -]50-85%]
- Fully Achieved (FA) -]85-100%]

For the criteria of Accessibility and Level of detail, both evaluators defined it as Fully Achieved, meaning that this work can be used without any training and provides information on Lisbon's mobility patterns. For the criteria of Understandability and Consistency, the first evaluator classified it as Fully Achieved while the second evaluator classified it as Largely Achieved meaning that the evaluators are satisfied with the mobility patterns analysis results. In terms of Utility, both evaluators agree that this research work can help with business decisions, classifying it as Largely Achieved. For

[3] https://www.iso.org/standard/80362.html

Robustness, the first evaluator classified it as Largely Achieved and the second evaluator classified it as Partially Achieved meaning that there is a satisfaction level of this method to be used in other cases.

Overall, the research results match the goals and needs of this work.

4 Conclusions

4.1.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 also showed that June, July, August, and September had the most concentration of trips during 2018, of 439,176, representing 64% of all trips. Moreover, we observed that BSS users mostly chose weekdays to ride in the city (82%) rather than on the weekend. Another interesting fact regards the hourly usage rate that shows users ride bikes during weekday peak hours, from 8 am to 9 am, from 4.30 pm to 6 pm, and at lunchtime 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 on the weekend, users prefer to ride overnight.

Addressing our sub-question on weather conditions affecting Lisbon BSS mobility patterns, we found that precipitation strongly impacts 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 20º and 30º (52%), and a significant number of users cycle when the temperature is between 10º and 20º (42%).

Sub-question regarding Lisbon BSS origin and destination station groups, we have observed that the most used were observed in two axes: 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 being 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 relationship between 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 representing 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 rides 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 the temperature is between 10º and 20º (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 in 2018 and 2019.

Preliminary results of our study were presented at EAI INSTYS 2020 - 4th EAI International Conference on Intelligent Transport Systems and was awarded best paper of EAI INTSYS 2020. Following the conference, the paper was published in a Springer book chapter [6] 'Understanding Spatiotemporal Station and Trip Activity Patterns in the Lisbon Bike-Sharing System'

Furthermore, an invitation to an extended version of the previous work was done and published on EAI Smart Cities, 'Bike-sharing mobility patterns: a data-driven analysis for the city of Lisbon' [8]. This paper presented a more developed version of results and comparison of data analysis of 2018 to 2019 and 2020.

4.2.Research Limitations

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 the 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 for 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.

4.3.Future Work

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 over time, and integration of BSS with multimodal urban transportation systems, in the context of the first and last mile.

References

[1] OECD, *Greening Household Behaviour*. OECD, 2011.

[2] OECD, (Organisation for Economic Co-operation and Development). A territorial approach to the Sustainable Development Goals in Paraná, Brazil. 80 (2021).

[3] *Estratégia Nacional para a Mobilidade Ativa Ciclável 2020-2030*. Conselho de Ministros, 2019.

[4] A. C. LVT, “Estratégia Regional de Lisboa AML 2030,” Área Metropolitana de Lisboa; Comissão de Coordenação de Desenvolvimento Regional de Lisboa e Vale do Tejo, Lisboa, 2020.

[5] C. M. de Lisboa, *MOVE Lisboa 2030 Visão estratégica para a mobilidade 2030*. Lisboa, 2020.

[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] V. Albuquerque, A. Oliveira, J. L. Barbosa, R. S. Rodrigues, F. Andrade, M. S. Dias, and J. C. Ferreira, “Smart cities: Data-driven solutions to understand disruptive problems in transportation—the lisbon case study,” *Energies*, vol. 14, no. 11. 2021. doi: 10.3390/en14113044.

[8] V. Albuquerque, F. Andrade, J. Ferreira, M. Dias, and F. Bacao, “Bike-sharing mobility patterns: a data-driven analysis for the city of Lisbon,” *EAI Endorsed Trans. Smart Cities*, p. 169580, 2018, doi: 10.4108/eai.4-5-2021.169580.

[9] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, R. Chou, J. Glanville, J. M. Grimshaw, A. Hróbjartsson, M. M. Lalu, T. Li, E. W. Loder, E. Mayo-Wilson, S. McDonald, L. A. McGuinness, L. A. Stewart, J. Thomas, A. C. Tricco, V. A. Welch, P. Whiting, and D. Moher, “The PRISMA 2020 statement: an updated guideline for reporting systematic reviews,” *BMJ*, p. n71, May 2021, doi: 10.1136/bmj.n71.

[10] J. Liu, Q. Li, M. Qu, W. Chen, J. Yang, H. Xiong, H. Zhong, and Y. Fu, “Station site optimization in bike sharing systems,” 2016, vol. 2016-Janua, pp. 883–888. doi: 10.1109/ICDM.2015.99.

[11] X. Tan, X. Zhu, Q. Li, L. Li, and J. Chen, “Tidal phenomenon of the dockless bike-sharing system and its causes: the case of Beijing,” *Int. J. Sustain. Transp.*, 2020, doi: 10.1080/15568318.2020.1871129.

[12] A. E. Rizzoli, R. Rudel, A. Förster, G. Corani, F. Cellina, L. Pampuri, R. Guidi, and A. Baldassari, “Investigating mobility styles using smartphones: Advantages and limitations according to a field study in Southern Switzerland,” 2014, vol. 1, pp. 274–282.

[13] C. Etienne and O. Latifa, “Model-based count series clustering for bike sharing system usage mining: A case study with the vélib’ system of Paris,” *ACM Trans. Intell. Syst. Technol.*, vol. 5, no. 3, 2014, doi: 10.1145/2560188.

[14] 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 Policy Pract.*, vol. 100, pp. 152–161, 2017, doi: 10.1016/j.tra.2017.04.023.

[15] C. Xu, J. Ji, and P. Liu, “The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets,” *Transp. Res. Part C Emerg. Technol.*, vol. 95, pp. 47–60, 2018, doi: 10.1016/j.trc.2018.07.013.

[16] J. F. Teixeira and M. Lopes, "The link between bike sharing and subway use during the COVID-19 pandemic: The case-study of New York's Citi Bike," *Transp. Res. Interdiscip. Perspect.*, vol. 6, 2020, doi: 10.1016/j.trip.2020.100166.
[17] A. Nikitas, "Understanding bike-sharing acceptability and expected usage patterns in the context of a small city novel to the concept: A story of 'Greek Drama,'" *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 56, pp. 306–321, 2018, doi: 10.1016/j.trf.2018.04.022.
[18] F. Liao and G. Correia, "Electric carsharing and micromobility: A literature review on their usage pattern, demand, and potential impacts," *Int. J. Sustain. Transp.*, 2020, doi: 10.1080/15568318.2020.1861394.
[19] J. F. Toro, D. Carrion, M. A. Brovelli, and M. Percoco, "Bikemi bike-sharing service exploratory analysis on mobility patterns," 2020, vol. 43, pp. 197–204. doi: 10.5194/isprs-archives-XLIII-B4-2020-197-2020.
[20] M. Noussan, G. Carioni, F. D. Sanvito, and E. Colombo, "Urban mobility demand profiles: Time series for cars and bike-sharing use as a resource for transport and energy modeling," *Data*, vol. 4, no. 3, 2019, doi: 10.3390/data4030108.
[21] I. B. I. Purnama, "Spatiotemporal mining of BSS data for characterising seasonal urban mobility dynamics," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 8, no. 4, pp. 1270–1276, 2018, doi: 10.18517/ijaseit.8.4.6510.
[22] W. El-Assi, M. Salah Mahmoud, and K. Nurul Habib, "Effects of built environment and weather on bike sharing demand: a station level analysis of commercial bike sharing in Toronto," *Transportation*, vol. 44, no. 3, pp. 589–613, 2017, doi: 10.1007/s11116-015-9669-z.
[23] C. Guido, K. Rafal, and A. Constantinos, "A low dimensional model for bike sharing demand forecasting," 2019. doi: 10.1109/MTITS.2019.8883283.
[24] Y.-H. Wu, L. Kang, Y.-T. Hsu, and P.-C. Wang, "Exploring trip characteristics of bike-sharing system uses: Effects of land-use patterns and pricing scheme change," *Int. J. Transp. Sci. Technol.*, vol. 8, no. 3, pp. 318–331, 2019, doi: 10.1016/j.ijtst.2019.05.003.
[25] D. Mao, Z. Li, H. Li, and F. Wang, "Bike-Sharing Dynamic Scheduling Model Based on Spatio-Temporal Graph," 2018, pp. 483–486. doi: 10.1109/BigComp.2018.00077.
[26] A. N. Randriamanamihaga, E. CÔme, L. Oukhellou, and G. Govaert, "Clustering the Vélib' dynamic Origin/Destination flows using a family of Poisson mixture models," *Neurocomputing*, vol. 141, pp. 124–138, 2014, doi: 10.1016/j.neucom.2014.01.050.
[27] L. A. Moncayo-Martínez and A. Ramirez-Nafarrate, "Visualization of the mobility patterns in the bike-sharing transport systems in Mexico City," 2016, vol. 2016-Decem, pp. 1851–1855. doi: 10.1109/IEEM.2016.7798198.
[28] A. Singla, M. Santoni, G. Bartók, P. Mukerji, M. Meenen, and A. Krause, "Incentivizing users for balancing bike sharing systems," 2015, vol. 1, pp. 723–729.
[29] D. Tomaras, I. Boutsis, and V. Kalogeraki, "Lessons Learnt from the analysis of a bike sharing system," 2017, vol. Part F1285, pp. 261–264. doi: 10.1145/3056540.3076206.
[30] G. C. Calafiore, F. Portigliotti, and A. Rizzo, "A Network Model for an Urban Bike Sharing System," *IFAC-Pap.*, vol. 50, no. 1, pp. 15633–15638, 2017, doi: 10.1016/j.ifacol.2017.08.1902.
[31] B. A. Neumann-Saavedra, T. G. Crainic, B. Gendron, D. C. Mattfeld, and M. Römere, "Integrating resource management in service network design for bike-

sharing systems,” *Transp. Sci.*, vol. 54, no. 5, pp. 1251–1271, 2020, doi: 10.1287/TRSC.2019.0950.
[32] S. Reiss and K. Bogenberger, “GPS-Data Analysis of Munich’s Free-Floating Bike Sharing System and Application of an Operator-based Relocation Strategy,” 2015, vol. 2015-Octob, pp. 584–589. doi: 10.1109/ITSC.2015.102.
[33] F. Pase, F. Chiariotti, A. Zanella, and M. Zorzi, “Bike sharing and urban mobility in a post-pandemic world,” *IEEE Access*, vol. 8, pp. 187291–187306, 2020, doi: 10.1109/ACCESS.2020.3030841.
[34] K. Wang, G. Akar, and Y.-J. Chen, “Bike sharing differences among Millennials, Gen Xers, and Baby Boomers: Lessons learnt from New York City’s bike share,” *Transp. Res. Part Policy Pract.*, vol. 116, pp. 1–14, 2018, doi: 10.1016/j.tra.2018.06.001.
[35] A. M. Pnevmatikou, C. P. Milioti, and M. G. Karlaftis, “Economic recession and the impact of ICT on integrated mobility during disruptive events,” 2014, pp. 681–695.
[36] A. Faghih-Imani, N. Eluru, A. M. El-Geneidy, M. Rabbat, and U. Haq, “How land-use and urban form impact bicycle flows: Evidence from the bicycle-sharing system (BIXI) in Montreal,” *J. Transp. Geogr.*, vol. 41, no. August 2012, pp. 306–314, 2014, doi: 10.1016/j.jtrangeo.2014.01.013.
[37] W. Li, S. Wang, X. Zhang, Q. Jia, and Y. Tian, “Understanding intra-urban human mobility through an exploratory spatiotemporal analysis of bike-sharing trajectories,” *Int. J. Geogr. Inf. Sci.*, vol. 34, no. 12, pp. 2451–2474, 2020, doi: 10.1080/13658816.2020.1712401.
[38] D. Zhuang, J. G. Jin, Y. Shen, and W. Jiang, “Understanding the bike sharing travel demand and cycle lane network: The case of Shanghai,” *Int. J. Sustain. Transp.*, vol. 15, no. 2, pp. 111–123, 2020, doi: 10.1080/15568318.2019.1699209.
[39] X.-M. Fu and Z.-C. Juan, “Unraveling Mobility Pattern of Dockless Bike-sharing Use in Shanghai,” *Jiaotong Yunshu Xitong Gongcheng Yu XinxiJournal Transp. Syst. Eng. Inf. Technol.*, vol. 20, no. 3, pp. 219–226, 2020, doi: 10.16097/j.cnki.1009-6744.2020.03.033.
[40] X. Zhou, “Understanding spatiotemporal patterns of biking behavior by analyzing massive bike sharing data in Chicago,” *PLoS ONE*, vol. 10, no. 10, 2015, doi: 10.1371/journal.pone.0137922.
[41] R. B. Noland, M. J. Smart, and Z. Guo, “Bikeshare trip generation in New York City,” *Transp. Res. Part Policy Pract.*, vol. 94, pp. 164–181, 2016, doi: 10.1016/j.tra.2016.08.030.
[42] X. Han, P. Wang, J. Gao, M. Shah, R. K. M. Ambulgekar, A. Jarandikar, and S. Dhar, “Bike sharing data analytics for silicon valley in USA,” *2017 IEEE SmartWorld Ubiquitous Intell. Comput. Adv. Trust. Comput. Scalable Comput. Commun. Cloud Big Data Comput. Internet People Smart City Innov. SmartWorldSCALCOMUICATCCBDComIOPSCI 2017* -, pp. 1–9, 2018, doi: 10.1109/UIC-ATC.2017.8397557.
[43] C. Zhong, M. Batty, E. Manley, J. Wang, Z. Wang, F. Chen, and G. Schmitt, “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.
[44] P. Jensen, J. B. Rouquier, N. Ovtracht, and C. Robardet, “Characterizing the speed and paths of shared bicycle use in Lyon,” *Transp. Res. Part Transp. Environ.*, vol. 15, no. 8, pp. 522–524, 2010, doi: 10.1016/j.trd.2010.07.002.

[45] M. H. Almannaa, H. I. Ashqar, M. Elhenawy, M. Masoud, A. Rakotonirainy, and H. Rakha, “A comparative analysis of e-scooter and e-bike usage patterns: Findings from the City of Austin, TX,” *Int. J. Sustain. Transp.*, 2020, doi: 10.1080/15568318.2020.1833117.
[46] D. Tomaras, “Crowdsourcing techniques for smart urban mobility,” 2018, pp. 460–461. doi: 10.1109/PERCOMW.2018.8480244.
[47] M. Batty, “Urban analytics defined,” *Environ. Plan. B Urban Anal. City Sci.*, vol. 46, no. 3, pp. 403–405, 2019, doi: 10.1177/2399808319839494.
[48] D. Papanikolaou, “Data-driven state space reconstruction of mobility on demand systems for sizing-rebalancing analysis,” 2018, vol. 50, pp. 275–282.
[49] I. B. I. Purnama, N. Bergmann, R. Jurdak, and K. Zhao, “Characterising and predicting urban mobility dynamics by mining bike sharing system data,” 2016, pp. 159–167. doi: 10.1109/UIC-ATC-ScalCom-CBDCom-IoP.2015.46.
[50] D. Bacciu, S. Gnesi, and L. Semini, “Using a machine learning approach to implement evaluate product line features,” 2015, vol. 188, pp. 75–83. doi: 10.4204/EPTCS.188.8.
[51] *CRISP-DM methodology - Smart Vision Europe*.
[52] P. C. Chapman Janet; Kerber, Randy; Khabaza, Tom; Reinartz, Thomas; Russel H. Shearer, C.; Wirth, Robert, “CRISP-DM 1.0: Step-by-step data mining guide,” *Comput. Sci.*, 2000.
[53] S. H. A. H. El-Sappagh Abdeltawab M. Ahmed; Bastawassy, Ali Hamed El, “A proposed model for data warehouse ETL processes,” *J. King Saud Univ. - Comput. Inf. Sci.*, no. 23, pp. 91–104, 2011.
[54] “Python,” Mar. 29, 2021. https://www.python.org/
[55] “Jupyter,” Mar. 29, 2021. https://jupyter.org/
[56] “Numpy,” Mar. 29, 2021. https://numpy.org/
[57] “Pandas,” Mar. 29, 2021. https://pandas.pydata.org/
[58] “Matplotlib,” Mar. 29, 2021. https://matplotlib.org/
[59] “Seaborn,” Mar. 29, 2021. https://seaborn.pydata.org/
[60] “GDAL,” Mar. 29, 2021. https://pypi.org/project/GDAL/
[61] “Shapely,” Mar. 29, 2021. https://pypi.org/project/Shapely/
[62] “Folium,” Mar. 29, 2021. https://pypi.org/project/folium/
[63] “Fiona,” Mar. 29, 2021. https://pypi.org/project/Fiona/
[64] “Scikit-learn,” Mar. 29, 2021. https://scikit-learn.org/stable/
[65] “LxDataLab – Laboratório de Dados Urbanos de Lisboa,” Mar. 29, 2021. https://lisboainteligente.cm-lisboa.pt/lxdatalab/
[66] “Determinação do impacte da pandemia por COVID 19 na mobilidade e ambiente,” *LxDataLab*. May 23, 2021.
[67] 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, 2016, doi: 10.1371/journal.pone.0168604.
[68] D. Fuller, L. Gauvin, Y. Kestens, M. Daniel, M. Fournier, P. Morency, and L. Drouin, “Use of a new public bicycle share program in Montreal, Canada,” *Am. J. Prev. Med.*, vol. 41, no. 1, pp. 80–83, 2011, doi: 10.1016/j.amepre.2011.03.002.
[69] S. Shaheen, H. Zhang, E. Martin, and S. Guzman, “China’s Hangzhou Public Bicycle: Understanding early adoption and behavioral response to bikesharing,” *Transportation Research Record*, no. 2247. Transportation Sustainability Research Center, University of California, Berkeley, Building 190, 1301 South 46th Street, Richmond, CA 94804-4648, United States, pp. 33–41, 2011. doi: 10.3141/2247-05.

[70] S. Shaheen, A. Cohen, and E. Martin, "Public bikesharing in North America," *Transportation Research Record*, no. 2387. National Research Council, Transportation Sustainability Research Center, University of California, Berkeley, Building 190, 1301 South 46th Street, Richmond, CA 94804-4648, United States, pp. 83–92, 2013. doi: 10.3141/2387-10.
[71] 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, 2012, doi: 10.1016/j.trc.2011.12.004.
[72] D. Buck and R. Buehler, "Bike lanes and other determinants of capital bikeshare trips," *91st Transp. Res. Board Annu. Meet.*, pp. 703–706, 2012.
[73] R. Ewing and R. Cervero, "Travel and the built environment," *J. Am. Plann. Assoc.*, vol. 76, no. 3, pp. 265–294, 2010, doi: 10.1080/01944361003766766.
[74] A. Kemperman and H. Timmerman, "Influences of built environment on walking and cycling by latent segments of aging population," *Transportation Research Record*, no. 2134. Urban Planning Group, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, Netherlands, pp. 1–9, 2009. doi: 10.3141/2134-01.
[75] 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, pp. 73–78, 2012, doi: 10.1016/j.sbspro.2012.04.079.
[76] 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, 2018, doi: 10.1016/j.trc.2018.01.001.
[77] 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," *ACM Int. Conf. Proceeding Ser.*, pp. 15–19, 2019, doi: 10.1145/3325730.3325748.
[78] 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.
[79] 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.
[80] G. McKenzie, "Docked vs. Dockless bike-sharing: Contrasting spatiotemporal patterns," *Leibniz Int. Proc. Inform. LIPIcs*, vol. 114, 2018, doi: 10.4230/LIPIcs.GIScience.2018.46.
[81] C. Y. Yuan Haitao, "Research on K-Value Selection Method of K-Means Clustering Algorithm," *Multidiscip. Sci. J.*, 2019.

Annexes and appendix

Annex A – Questionnaire

Criteria	Objective statement	Evaluator 1#	Evaluator 2#
Utility	It can help business decisions regarding BSS.	**LA**	**LA**
Understandability	Provides understandable results.	**FA**	**LA**
Consistency	Can be used without training.	**FA**	**FA**
Level of detail	Provides knowledge from the mobility patterns.	**FA**	**FA**
Consistency	Gives consistent results (compared with another country or company).	**FA**	**LA**
Robustness	Has enough detail to be used in other cases of BSS.	**LA**	**PA**

Not Achieved (NA) - [0-15%]
Partially Achieved (PA) -]15-50%]
Largely Achieved (LA) -]50-85%]
Fully Achieved (TA) -]85-100%]