BIKEMI BIKE-SHARING SERVICE EXPLORATORY ANALYSIS ON MOBILITY PATTERNS

J. F. Toro [1], D. Carrion [1], M. A. Brovelli [1, *], M. Percoco [2]

[1] Politecnico di Milano, Civil and Environmental Engineering Department - Section of Geodesy and Geomatics, Milan, Italy -
(juanfernando.toro, daniela.carrion, maria.brovelli)@polimi.it
[2] Università Bocconi, Department of Social and Political Science and GREEN, Milan, Italy - marco.percoco@unibocconi.it

Commission IV, WG IV/3

KEY WORDS: bike-sharing, urban mobility, spatial analysis, data visualization

ABSTRACT:

Bike Sharing Systems (BSS) are growing worldwide for the social and environmental benefits that they can provide. Thanks to the increasing popularity of the BSS and the availability of monitoring technologies, there is a continuous production of data that can help to understand bike usage and to improve its design and management. This study aims at exploring BSS users' mobility patterns habits and the demand for the service. To reach this scope, the available data have been preprocessed in order to allow data mining and data visualization with open source tools based on Python. The study case regards the BikeMi BSS of the City of Milan between June 2015 to December 2018. The suggested approach proceeded, first, with the categorization of the user typology based on the frequency of use of the service; at a second stage, the influence of the day typology on the use of the service has been explored; third, the spatial and temporal patterns of the BSS use among the stations has been analysed; fourth, the influence of meteorological conditions on the use of the service has been considered; at last, the clustering of the stations with similar bikes use activity through K-Means has been performed. As expected, it was observed that the service is extensively used for commuting to work-related activities. Regular users compose a large part of the BSS community making use of the service mostly during weekdays. In addition, it was noted that only 'strong' meteorological conditions can impact the use of the service. Both the identification of the demand for the service and of the external factors that can affect its use support the clustering activities, allowing for the elimination of not relevant information and facilitating the interpretation of the obtained clusters.

1. INTRODUCTION

Bike Sharing Systems (BSS) growth worldwide is related to the environmental and social benefits that they can provide, in terms of CO2 emissions, traffic relief and user's health. In fact, Pucher and Buehler (2017) present cycling as the "most sustainable urban transportation mode". Nowadays, there is an increasing competition inside the micro-mobility transportation-sharing market, due to the offer of new services, such as dock-less bikes, scooters, etc. Therefore, there is the need to optimize the allocation of resources for the BSS (spatial configuration and fleet distribution) in order to maintain its competitiveness.

Thanks to the increasing popularity of the BSS and to the availability of monitoring technologies large amounts of data describing the use of the service can be recorded. The analysis of these data can help understanding the bike usage and to improve its design and management. Various studies have addressed the use of the bikes analysing data coming from BSS in different cities (Andrienko et al, 2016; Beecham et al. 2014; Tao et al., 2014). The research in this field is supported by the parallel development of different tools, both in visual analytics and data mining techniques, as reviewed by Sobral et al. (2019) on the current developments in data visualization for intelligent transport systems.

This study aims at exploring BSS users' mobility patterns habits and the demand for the service. The goal of the analysis is to provide indications to optimize the redistribution of the BSS resources for improving its operations. To reach this scope, the available data have been preprocessed in order to allow data mining and data visualization with open source tools based on Python.

1.1 State of the Art

In general, for the volumes of data retrieved for a BSS it is complex to assess the spatio-temporal relationships in between the stations with respect to their use, e.g. Origin/Destination matrices. Because of this, the use of unsupervised machine learning techniques for clustering became a common practice for analysing mobility patterns to optimize the BSS. Some of the explored clustering techniques for analysing BSS are: k-means (Vogel et al., 2011; Chabchoub and Fricker, 2014; Feng et al, 2017; Ma et al., 2019), Expectation-Maximization (Vogel et al., 2011), sequential information-bottleneck (Vogel et al., 2011) and hierarchical clustering (Feng et al., 2017). These algorithms require a subsequent subjective interpretation of the clusters which usually leads to their labelling with respect to the closest point of interest (e.g. "Employment", "Residential") or on the bikes availability at the docks ("balanced", "overloaded", "underloaded"). In addition, in many cases the selection of the number of clusters is difficult to assess and it may lead to results which are not helpful for the interpretation of the mobility patterns. This is mentioned by Feng et al. (2017), where for the study case of the Vélib BSS in France different pattern analyses provided very different sets of clusters which could result in different conclusions. Vogel et al. (2011) and Feng et al. (2017) highlighted the need for appropriate testing approaches for the number of clusters selection (for example

---

* Corresponding author

indexes such as Davies-Bouldin, Dunn, Silhouette) in order to reduce the variability in the interpretation phase. Kim (2018) suggested an approach to overcome the complexity of labelling the clusters. The study comments that with an initial categorization by users' typologies and the isolation of other external factors disrupting the regular use of a BSS, it is possible to identify specific temporal patterns associated with each type of user. These subsets for each user typology, can help in interpreting the use of the bike stations and, therefore, in identifying the spatiotemporal relationships between stations through the clustering methods.

1.2 Motivation

In order to optimize a BSS service, multiple variables must be considered, such as mobility patterns, transport management, meteorological conditions, users' motivation, among others, to face the complexity of the matter. This paper suggests a procedure to process the data of the use of the BSS service, in order to improve the efficiency of the bike distribution over the stations addressing the mentioned variables, when possible, including correlation analyses. The methodology is summarized in the following:

- records are categorized according to the day typology (weekdays, weekends and holidays) and the user's typology to identify the mobility habits. Regular or occasional users are classified according to the frequency of use and separated using a threshold determined with a sensitivity analysis;
- the impact of meteorological conditions (rainfall and temperature) is assessed through a correlation analysis with respect to the use of the service: this allows to understand the behaviour of users with respect to the weather and to treat separately records which do not correspond to regular mobility patterns;
- a cluster-based analysis is applied to identify temporal-spatial patterns for specific users' typologies by using the k-means clustering algorithm.

2. CASE STUDY AND DATA PROCESSING

The study case regards BikeMi docked BSS provided by the local transport service of the municipality of Milan. The data, which were made available for this work, have been retrieved between June 2015 to December 2018, corresponding to 13'789'569 records. Each record represents an origin-destination pair for a trip with timestamp and station ID. Currently, BikeMi BSS encompasses a fleet of 3650 traditional bikes and 1150 electric bikes (150 with a child seat). The service operation hours are between 07:00 to 23:59. Since they have changed in time, the latest available locations of the stations available at the Open Data portal for the Lombardy Region (dati.lombardia.it) have been used.

The BSS has been steadily developing over the years with the integration of different stations. Figure 1 reveals the expansion of the BSS from the centre towards the periphery of the city, displaying a denser distribution of stations in the centre, which becomes sparser radially. On the other hand, in Figure 2 it is shown that the size of the docks is larger out of the city centre, having its peak in a ring right outside the area with limited traffic ("Area C"[1]).

---

[1] Area C is an area within the historic centre of Milan with restricted access to some vehicles typologies. https://www.comune.milano.it/aree-tematiche/mobilita/area-c
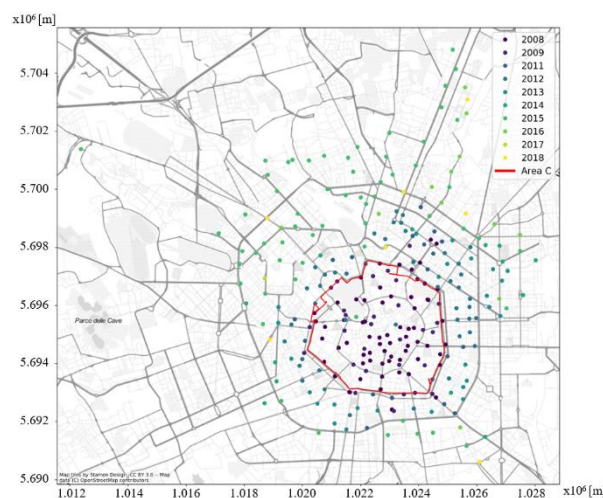

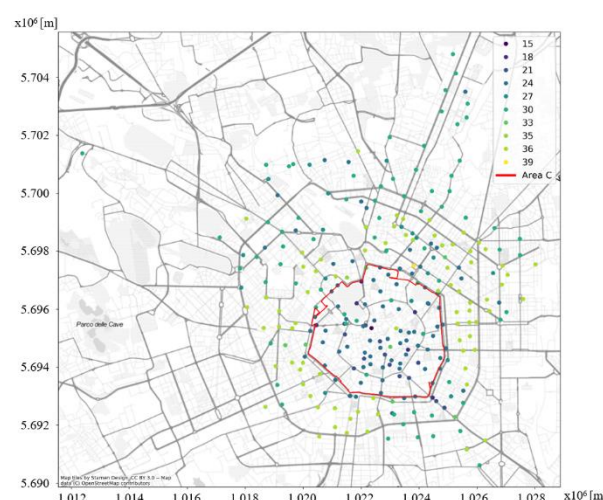Figure 1. BikeMi BSS - stations installation year. (EPSG:3857)

Figure 2. BikeMi BSS - number of docks per station. (EPSG:3857)

2.1 Data Preprocessing and Tools

It has been necessary to filter the dataset to avoid specific inconsistencies in the records, such as:

- records outside the service working hours (some extraordinary schedules have taken place during the time span, but they were not documented, for this, it was preferred to remove the records);
- trips which do not have a duration larger than 1 minute;
- records whose arrival or departure station did not match with the existing stations.

After the filtering, approximately only 3% of the observations were neglected.

A key point of the analysis is the categorization of the users under different typologies. BikeMi users can access the service through the acquisition of different types of memberships (daily, weekly, monthly, yearly), which come along with an identification number. In this case, this identification number is assumed as a proxy variable for referring to a unique user (disregarding that a single individual could be assigned different IDs in case of acquiring different memberships).

The analysis of mobility patterns involves the handling of large amounts of data and the understanding of a multivariate problem. It has been necessary to represent the data, possibly in a dynamic way, with an extensive use of data visualization tools. This study relied on the use of python by leveraging its geostatistical analysis capabilities. Python libraries such as bokeh, HoloViews and PyViz (which are based on bokeh), served as tools for developing an interactive data visualization to identify mobility patterns inside the city, followed by the use of the scientific toolbox "scikit-learn" for the clustering analysis.

2.2 Time Series Analysis

In Figure 3, one can see a decrease in the use of the service in 2018, which can be attributed to the increase in competitiveness for the transport-sharing market. McKenzy (2019), evidenced that the introduction of different transportation modes as dockless bike-sharing and scooter sharing produced a modal shift due to the overlapping of the services coverage. In this case, in August 2017 there was the introduction of a dockless BSS into the bike-sharing market (Meddin and DeMaio, 2020).

In addition, Figure 3 shows the monthly fluctuations in the use of the service, which can be explained with respect to the seasonal use, i.e. dependence on weather conditions. Looking at Figure 3b and Figure 3c, it is possible to see that the use of the service reduces significantly during colder months (November, December, January and February), as it could be expected, while it is almost equivalent during the other months, apart from August, which corresponds to Italian holidays. In addition, from Figure 3c, which shows monthly relative frequencies normalised with respect to the length of the time series per year, it is possible to see that the monthly use of the service is comparable over the years.

A particular case of the evolution in the use of the service regards April 2017. This month displayed an unusual behaviour concerning the similarity of the monthly use of the service between the years. Such irregularity possibly indicates the influence of external conditions on the service. To address this behaviour, two external variables have been reviewed: the day typology and the meteorological conditions. Their inspection noted three points: 1) April 2016 presents fewer holidays compared to the other years, which reflected in a higher number of trips; 2) the presence of heavy rainfall events led to a decrease in the use of the service ; 3) the impact on the service use under rainfall conditions vary depending on the day typology (e.g. a rainfall event during a weekday reduces the use of service at a larger scale compared to the same scenario in a holiday). These indications provided additional motivations for the analysis presented in the following sections.
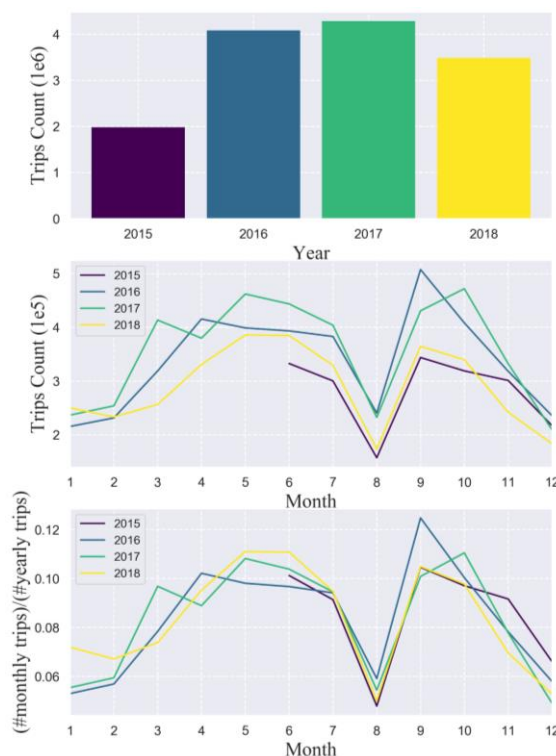

Figure 3. Evolution of BikeMi use in time. (a) Frequency of use yearly; (b) Frequency of use on a monthly basis per year; (c) Relative frequency of use on a monthly basis per year

2.2.1 Users typology: It is important to recognize that the mobility patterns could be considerably affected depending on the type of membership that the user has. This study concerned the division of the users typology based on the frequency of use of the service by each user.

Users were categorized according to their frequency of use of the service, relying on the assumption that the regular users tend to use the service more often. In Figure 4, the frequency of use of the service is summarized into a boxplot which made explicit the number of trips to categorize a certain percentile as belonging to a more (or less) frequent user. Hereafter, the users which use a service above a defined threshold of trips will be referred to as "regular" users (otherwise, they will be "occasional" users).
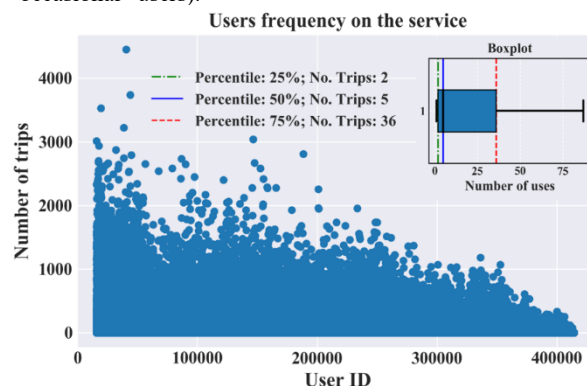

Figure 4. Frequency of use of the service per user

The selection of a threshold for the trips was defined by a sensitivity analysis on the number of users classified among the two different categories. Figure 5 displays the change in the

number of 'regular'/'occasional' users depending on the threshold of trips selected. The figure displays that the rate of change between the number of users with respect to the number of trips is quite low until the threshold for the user categorization reaches the 75th percentile (36 trips). For this reason, the distinction between the two types of users consider 36 trips as a threshold for the analysis. Furthermore, with this classification it was observed that the 'regular' users represent a larger data subset for the analysed records.
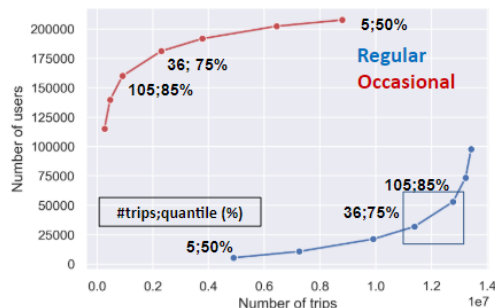
Figure 5. Sensitivity analysis on the number of "regular"/"occasional" users relative to the number of trips performed
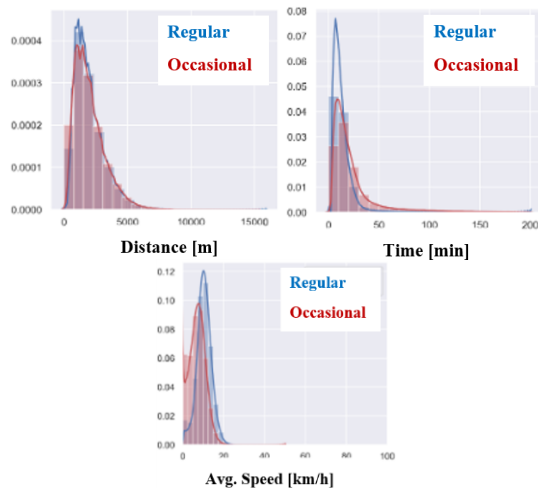

Figure 6. Characteristics for the trips according to the users typology

Correspondingly, to avoid an arbitrary selection for the thresholds for categorizing the users, Figure 6 show other characteristics relative to the trips performed by each user typology. This representation revealed that while the length of trips does not vary between users typology, there is a difference between the duration of the trips and, as a consequence, the average speed. 'Regular' users tend to have faster trips with respect to the 'occasional' users.
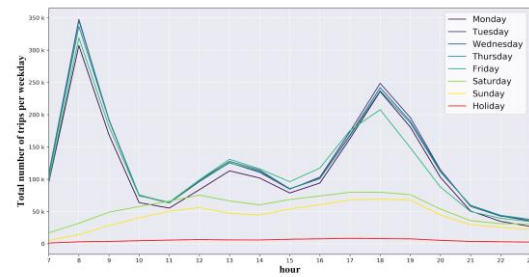

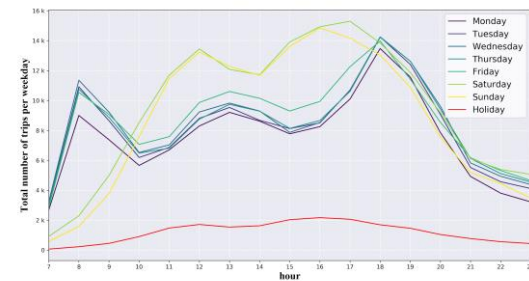Figure 7. Hourly trips count during the week based on day and user typology (regular users)

Figure 8. Hourly trips count during the week based on day and user typology (occasional users)

After the categorization according to the user typology, it was possible to perform a frequency analysis over the weekdays, weekend and holidays for each group (Figure 7 and Figure 8). This analysis confirms what found in literature, that the behavioural patterns in between the typologies of users is different. Figure 7 shows that 'regular' users, as expected, use the service mainly during weekdays, with two different peaks along the day, in the morning and in the afternoon, consistent with the commuting hours. It is interesting to see that there is also a smaller peak at lunchtime; this could be linked to part-time work, school or it could indicate people who do not have lunch close to the workplace. In contrast, 'occasional' users normally use the service during weekends and holidays, with a larger activity close to lunchtime and evening hours, similarly to the patterns described by Beecham & Wood (2014). It is also evident that the 'occasional' users during the weekend have the same behaviour as holiday time.

Given the different behaviour of the users' typologies, they will be analysed separately in the following steps of the processing. In particular, the spatial analysis will target the larger segment for the demand of the service, that corresponds to the "regular" users of the BSS.

2.3 Spatial Analysis

Here, the data visualization followed the mobility patterns of the 'regular' users. The analysis focuses on the connection to other public transportation modes. In particular, on the rail-bike connection.

Bokeh library for python allowed for interactive visualization of the mobility flows (in Figure 9 screenshots are shown). Firstly, and expected, it is possible to see that the mobility flows in the morning are oriented from the periphery to the centre of the city, while in the afternoon it is the opposite. Secondly, it is possible to show the spatial distribution of the required actions for reallocation at each station during the day. At last, one can

see the concentration of trips at the bike stalls surrounding the main train stations, resulting in a large number of bikes used and in starving stalls.
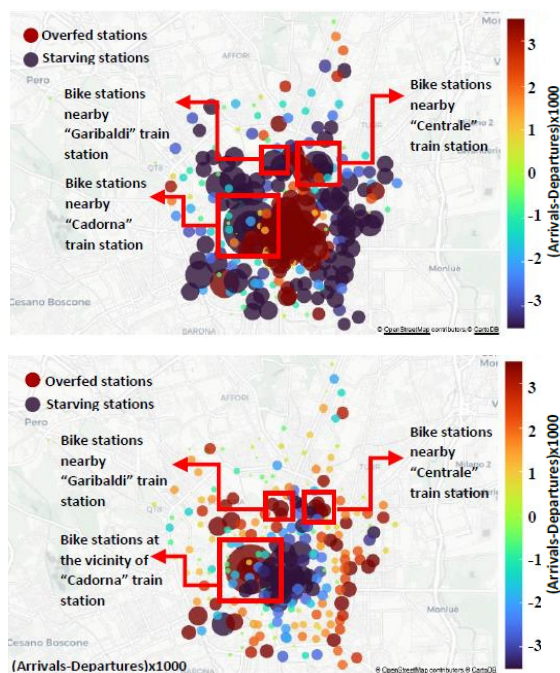

Figure 9. Mobility flows during weekdays, regular users at 08:00 (top) and at 18:00 (bottom)

The bike stations nearby the main train stations, during weekdays, for regular users, capture approximately 12% of the trips. This encourages further analysis regarding the effects on the use of the BSS. According to Comune di Milano (2011) 22% commuters arriving to Milan for work related activities choose ecological transportation modes for reaching their destinations (with a total of 457.700 individual users). These indications suggest, as expected for a city like Milan, which is an attracting pole for work, that commuters represent most of the demand of the service. Considering this, some special events, like strikes on the train service, may influence the demand for the BSS and could be further investigated.

Figure 10 displays the individual mobility patterns from a station towards others indicating the possible trips motivations depending on the proximity to points of interest. For example, for "Cadorna 3" bike station, in the morning, the flows go from Cadorna train station towards, mainly, workplaces in the city. Data representations, such as Figure 9 and Figure 10, can support the later interpretation of the clustering algorithms on the mobility habits for each station.
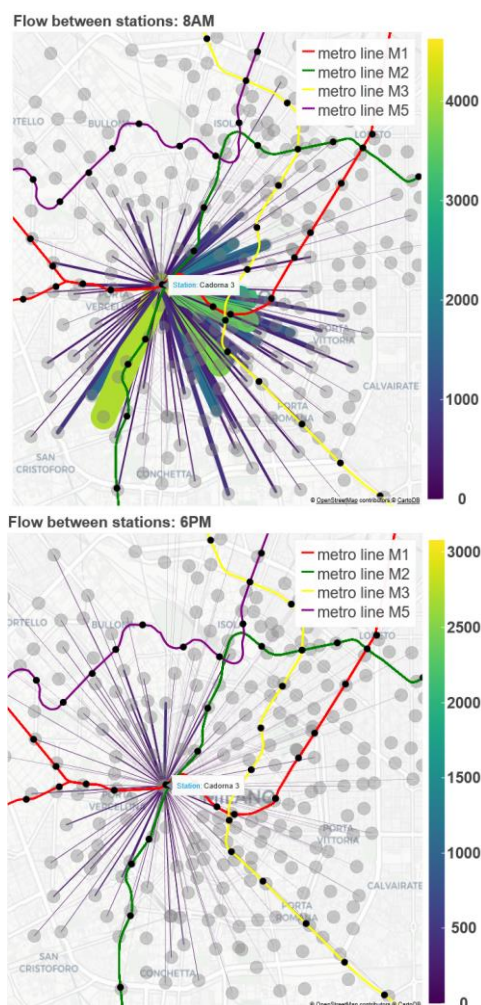

Figure 10. Mobility flows from "Cadorna 3" station, considering regular users only (trips count). 08:00 (top). 18:00 (bottom).

2.4 Weather Conditions

Given the sparse distribution in space of the meteorological sensors, with respect to the bike stations, a spatial correlation analysis between the phenomena and the number of trips at the stations may be biased. Because of this, the assessment was initially carried out among meteorological stations for each phenomenon. In both cases, the correlation analysis indicated relative low variability in space (this evidence was weaker between rainfall sensors; the furthest stations estimate 56% for the correlation statistic). Hence, the selection of the 'most extreme' event for the phenomena is considered on the impact over the use along the stations.
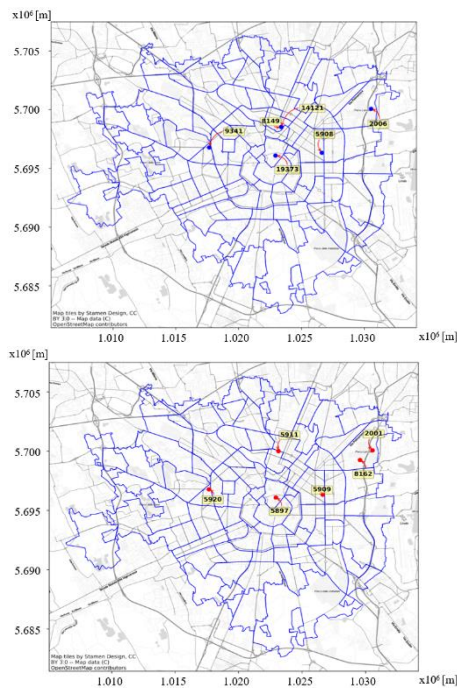
Figure 11. Meteorological sensors network from ARPA Lombardia (within Milan administrative boundaries). Rainfall sensors (top). Temperature sensor (bottom). (EPSG:3857)
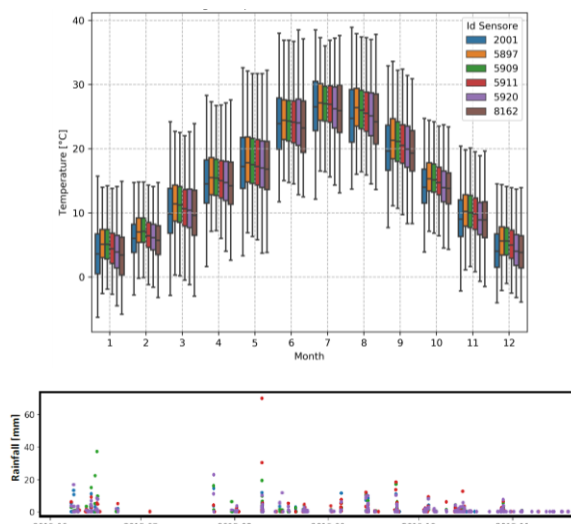

Figure 12. Temperature variability over the months (top). Rainfall events sample for the second semester of 2015 (bottom)

When taking into account different meteorological conditions, it was found that the service is not significantly affected by them (the correlation statistics estimates were 11.6% and 17.8% for the rainfall and the temperature, respectively). However, the results pointed out that most extreme events affect the service. For example, in Figure 13, it is possible to see that if rainfall tops 20 mm there is a dramatic decrease in the number of trips and the exceptional event above 60 mm did not result in any use of the service.
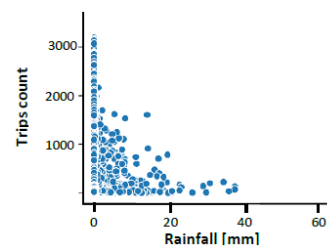

Figure 13. Rainfall vs use of the service

2.5 K-Means Clustering

This study proposes the use of a partitioning clustering method, K-Means, to assess the bikes transactions for each station. First, the aggregate count for the departing trips from each station was normalised. Afterwards, it was possible to categorise the 280 stations according to the agreement of their departure's activity in three different clusters (see Figure 16). The computed clusters follow the foreseen mobility patterns from the outer to the inner part of the city in the mornings (and vice-versa in the afternoon). On the other hand, the clustering method identified a set of stations bordering the limited traffic zone. This method allowed us to identify the bike needs at each of the stations in time and eventually to understand how they can complement each other.
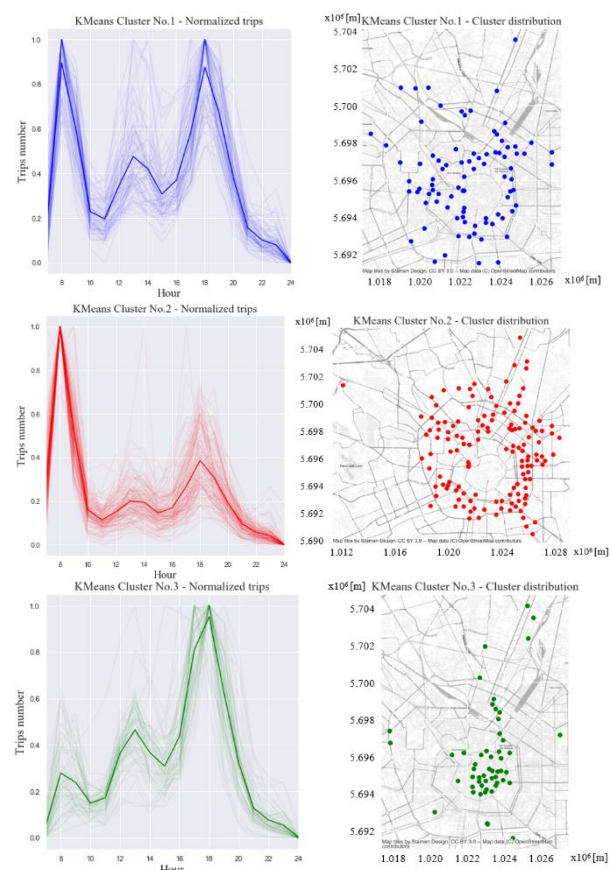

Figure 16. K-Means (k=3) over the normalised departures from the stations. (EPSG:3857)
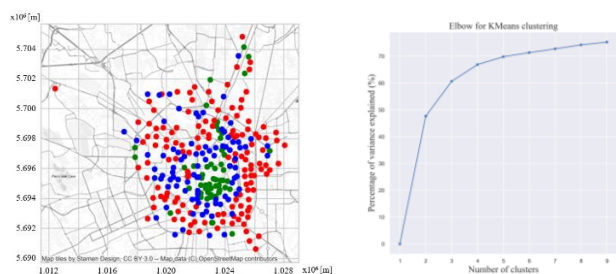
Figure 17. K-Means clustering on the normalized departures. (Left side) Stations clusters spatial distribution (k=3). (Right side) Explanatory capacity for adding a cluster. (EPSG:3857)

3. CONCLUSIONS

The use of data visualization and data mining techniques can provide extensive support for the analysis of BSS. In particular, to understand the variability of the use of the service in time. This performed analysis on a dataset of three years data for the BSS of the city of Milan allowed to follow different patterns within the datasets that translated into different typologies of trips (namely, users and days typologies). Also, it supported the identification of irregular mobility activities that were related to diverse external factors, mainly extraordinary meteorological conditions. On the other hand, with the possibility to exploit origin-destination data for each station, allowed to identify the mobility flows between stations. The spatial analysis showed the presence of clusters concerning the bike flows during the day. For these reasons, it is encouraged the exploratory analysis on the BSS records based on data visualization and data mining.

REFERENCES

Andrienko, N., Andrienko, G., Fuchs, G., & Jankowski, P., 2015. Scalable and privacy-respectful interactive discovery of place semantics from human mobility traces. *Information Visualization*, 15(2), 117–153. doi: 10.1177/1473871615581216

Beecham, R., & Wood, J., 2014. Characterising group-cycling journeys using interactive graphics. *Transportation Research Part C: Emerging Technologies*, 47, 194–206. doi: 10.1016/j.trc.2014.03.007

Chabchoub, Y., & Fricker, C., 2014. Classification of the vélib stations using Kmeans, Dynamic Time Wraping and DBA averaging method. *2014 International Workshop on Computational Intelligence for Multimedia Understanding (IWCIM)*. doi: 10.1109/iwcim.2014.7008802

Comune di Milano, 2011: 'I DATI DEL CENSIMENTO 2011 A MILANO - Analisi del pendolarismo per studio e per lavoro a Milano'.
https://www.comune.milano.it/documents/20126/2220553/Pendolarismo+a+Milano.pdf/ece83eab-3103-e28f-f894-fec26b2504de?t=1555347870095

Feng, Y., Affonso, R. C., & Zolghadri, M., 2017. Analysis of bike sharing system by clustering: the Vélib' case. *IFAC-PapersOnLine*, 50(1), 12422–12427. doi: 10.1016/j.ifacol.2017.08.2430

Kim, K., 2018. Investigation on the effects of weather and calendar events on bike-sharing according to the trip patterns of bike rentals of stations. *Journal of Transport Geography*, 66, 309–320. doi: 10.1016/j.jtrangeo.2018.01.001

Ma, X., Cao, R., & Jin, Y., 2019. Spatiotemporal Clustering Analysis of Bicycle Sharing System with Data Mining Approach. *Information*, 10(5), 163. doi: 10.3390/info10050163

Mckenzie, G., 2019. Spatiotemporal comparative analysis of scooter-share and bike-share usage patterns in Washington, D.C. *Journal of Transport Geography*, 78, 19–28. doi: 10.1016/j.jtrangeo.2019.05.007

Meddin, R., DeMaio, P., 2020: The bike-sharing world map. http://www.bikesharingmap.com/

Pucher, J., & Buehler, R., 2017. Cycling towards a more sustainable transport future. *Transport Reviews*, 37(6), 689–694. doi: 10.1080/01441647.2017.1340234

Sobral, T., Galvão, T., & Borges, J., 2019. Visualization of Urban Mobility Data from Intelligent Transportation Systems. *Sensors*, 19(2), 332. doi: 10.3390/s19020332

Tao, S., Rohde, D., & Corcoran, J., 2014. Examining the spatial–temporal dynamics of bus passenger travel behaviour using smart card data and the flow-comap. *Journal of Transport Geography*, 41, 21–36. doi: 10.1016/j.jtrangeo.2014.08.006

Vogel, P., Greiser, T., & Mattfeld, D. C., 2011. Understanding Bike-Sharing Systems using Data Mining: Exploring Activity Patterns. Procedia - Social and Behavioral Sciences, 20, 514–523. doi: 10.1016/j.sbspro.2011.08.058