

Article

A Cluster-Based Approach Using Smartphone Data for Bike-Sharing Docking Stations Identification: Lisbon Case Study[†]

Tiago Fontes^{1,2,*}, Miguel Arantes², Paulo V. Figueiredo² and Paulo Novais¹

¹ ALGORITMI Center, Informatics Department, University of Minho, 4710-057 Braga, Portugal; pjon@di.uminho.pt

² CEiiA, 4450-017 Matosinhos, Portugal; antonio.asilva@ceiia.com (M.A.); paulo.figueiredo@ceiia.com (P.V.F.)

* Correspondence: tiagorafaelfontes39@gmail.com or tiago.fontes@ceiia.com; Tel.: +351-913784736

† This paper is an extended version of our paper published in “Bike-Sharing Docking Stations Identification Using Clustering Methods in Lisbon City”.

Abstract: Urban mobility is a massive issue in the current century, being widely promoted the need of adopting sustainable solutions regarding transportation within large urban centres. The evolution of technologies has democratised smart cities to better plan and manage their mobility solutions, without compromising the social, economic, and environmental impacts. Pursuing the carbon neutrality and the climate agreement goals, soft mobility is one of the most popular emerging methods to provide greener alternatives regarding mobility. Among these transportation modes are the bicycle, which has been widely used in several public systems across the world, one of them being in Lisbon. This article provides a decision support system for bike-sharing docking stations for three council parishes of the city, namely, *Parque das Nações*, *Marvila*, and *Beato*. Taking advantage of clustering methods and GSM data from a telecommunication operator, this study pretends to highlight a novel approach to identify soft mobility hotspots, in specific bike-sharing docking stations, for suited mobility management systems in Lisbon’s city centre.

Keywords: urban mobility AI development; bike-sharing clustering models; docking stations; Lisbon; smart cities; soft mobility; practical use case



Citation: Fontes, T.; Arantes, M.; Figueiredo, P.V.; Novais, P. A Cluster-Based Approach Using Smartphone Data for Bike-Sharing Docking Stations Identification: Lisbon Case Study. *Smart Cities* **2022**, *5*, 251–275. <https://doi.org/10.3390/smartcities5010016>

Academic Editor: Giuseppe Guido

Received: 31 January 2022

Accepted: 25 February 2022

Published: 3 March 2022

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the last decades, increasing population growth has created several concerns, especially in urban centres, as it brings new environmental, economic, and social challenges, as shown by environmental issues surrounding human overpopulation [1]. Climate change and environmental degradation represent not only a European but also a global concern.

The Agenda for Sustainable Development [2], adopted by all Member States of the United Nations [3], provides a standard blueprint for peace and prosperity for people and the planet, now and for the future. The seventeen Sustainable Development Goals (SDGs) are at its heart, an urgent call to action from all countries (both developed and in development). They recognise, among several topics, stimulating economic growth while tackling climate change. The thirteenth goal is “Making cities and human settlements inclusive, safe, resilient and sustainable,” clarifying the global need to promote sustainability in one of the most fundamental ways, mobility.

In this sense, aiming to urge communities and cities to adopt sustainable habits regarding climate changes effects, the European Commission [4] (EC) has released the European Environment Pact (EEP), which is the plan outlined to define a strategy that enables a modern economy, efficient in the use of resources and competitive. This plan aims to boost the efficient use of resources through the transition to a cleaner economy and reduce green gas emissions. Mobility plays a crucial role in sustainability, as it is one of

the main focuses of the EEP. Land and Head [5] summarise sustainable urban mobility as “the ease, convenience, accessibility and low cost in terms of travel in these centres, seeking minimum impact on the environment and other inhabitants.” Hence, the latter contemplates obvious objectives in this sense, such as investing in environmentally friendly technologies and implementing cleaner and healthier public and private transport forms.

Portugal, in particular, as documented by Climate ADAPT [6], has driven several initiatives aimed to promote ecologic sustainable friendly behaviour concerning mobility, encouraging the purchase of electric vehicles by lowering taxes, limiting vehicles circulation in specific zones in major cities, increasing bicycle lanes, and increasing shared vehicles in urban centres, among others.

Portugal also has some examples of its implementation in terms of sustainable technologies, which are documented and studied in “The potential of Bike-sharing Systems” [7], as is the case of *Aveiro* and *Lisbon*, as stated by the Portuguese American Journal [8], in terms of the soft mobility modes implemented, taking advantage of Bicycle Sharing Systems (SBP). However, implementing these soft mobility solutions carries high costs, as seen in the Journal of Public Transportation [9].

As demonstrated by [10–12], the examples of cities like Warsaw, Paris, Barcelona, etc., show a need to plan the location of ecofriendly solutions to maximise their efficiency and, ideally, make them the mode of transport chosen by citizens when it comes to internal city mobility. Considering the high demographic rate and urban traffic flow in urban centres, it is essential to face these displacements as a serious sustainability challenge.

Typically, cities are more developed in the digitisation component, promoting Information and Communication Technologies (ICT) [13] mechanisms in the enhancement of efficiency and quality of operations and services. Thus, in recent years a concept has been created to define these urban centres. A *smart city* [14] is an urban area that leverages digital innovations, tools, and organisational principles in order to help the community evolve to become more sustainable, inclusive, prosperous and creative, ultimately benefiting the individual citizen. Lisbon is thus already considered a *smart city*, as reported by Smart50Awards [15]; so, by taking advantage of these intelligent ecosystems, it is possible to collect accurate information that will allow extracting and generating knowledge about them, serving as a basis for the management of goods, resources, and services efficiently.

Apart from the massive side effects from air pollution and climate change, which affects the planet every year, the European Environment [16] sent alerts to the world communities, especially for European countries, for air pollution being “the biggest environmental health risk in Europe.” As stated by the mentioned source, the “levels of air pollutants still exceed EU standards and the most stringent World Health Organisation guidelines.”

Moreover, *Lisbon* in particular, has been a special point of interest in Europe concerning air pollution and private car usage, as highlighted by Portuguese News portal [17], being elected as “the most congested city in the Iberian Peninsula” in recent years, being corroborated with TomTom [18] traffic index for the past years.

Without compromising our capacity to move, we intend to minimise the environmental impacts associated with it. To this end, it is pertinent to strengthen the participation of public transport and soft modes in the modal split, the adoption of technologies that increase the energy and environmental efficiency of transports, or the development of soft mobility, i.e., forms of travel under their own steam (cycling, walking, scooters, roller skaters, etc.), always aiming for more sustainable mobility.

Since the beginning of the century, bicycle-sharing schemes have been rapidly growing in several earth regions, being considered as an indicator of the attractiveness of such systems and their adaptability to changes according to different situations, according to Midgley regarding bike-sharing schemes [19].

Although there are benefits to society, some major obstacles are the number and location definitions for new docking stations. Defining these outputs is not an easy task since several parameters must be considered. Moreover, these parameters can change

through time, including the demographic rate, tourism interest points, transportation access hubs, bicycle paths, time of the year, year's season, etc., for each parish council.

1.1. Lisbon Bicycles System

Lisbon has one of the most common and well-known Bike-sharing Schemes (BSS) in the world. Launched in 2017, the public bike-sharing service went live in the city centre (named here as *Lisbon Bicycles*). Municipal Mobility and Parking Company have operated the system after a pilot phase in *Parque das Nações*.

At the end of 2018, the mark of one million trips was reached [20]. Nowadays, more than three and half million trips are made. Hence, the noticeable and continuous increase in popularity of the service has allowed rapid growth of docking stations, bicycles, etc. For this year, the number of docking stations and bicycles is expected to rise, following the intended investments predicted for bicycle paths for the upcoming years [21].

Furthermore, as stated before, *Lisbon* has been elected as one of the most congested cities in the last years, highlighting the importance of looking for new mobility solutions that may reduce the side effects of carbon emissions. To provide full representation of the current status of the public bike-sharing system operating in *Lisbon*, Figure 1 presents the current geographical distribution of such docking stations.

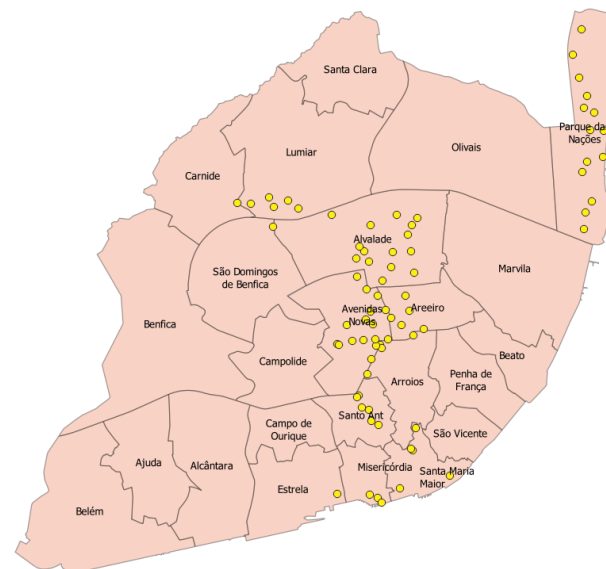


Figure 1. Lisbon bicycles docking stations (yellow dots) distribution over Lisbon city center.

Since 2017, when the system went live in *Parque das Nações*, Figure 1 highlights how the system has grown in the most recent years, since it first went live in *Parque das Nações* in 2017. By now, *Lisbon bicycles* has fourteen docking stations in *Parque das Nações*, but also it has been implemented in other council parishes in the *Lisbon* city centre, especially in the most congested areas and avenues.

For the current study, the focus will be in *Parque das Nações* (where the system is already implemented), to find potential weaknesses and suggest changes to the current configuration, and also to expand it for *Marvila* and *Beato*, providing a sustainable offer to the Lisbon citizens regarding soft mobility solutions.

1.2. Area of Interest Insights

Since it is pretended to develop a proof-of-concept for part of *Lisbon* city centre, the Area of Interest (AoI) will consider three out of the council parishes in the city, namely, where the system is and is not implemented yet.

In 2017, the investment was performed for *Parque das Nações*. Nowadays, the service is expected to be available in two more council parishes (*Marvila* and *Beato*).

These new locations do have several tourism access points, having also public transportation access hubs (subway, train, etc.), and new bike lanes are under development. From Figure 2, it is possible to identify the three council parishes named for the current study, namely, *Parque das Nações*, *Marvila*, and *Beato*.

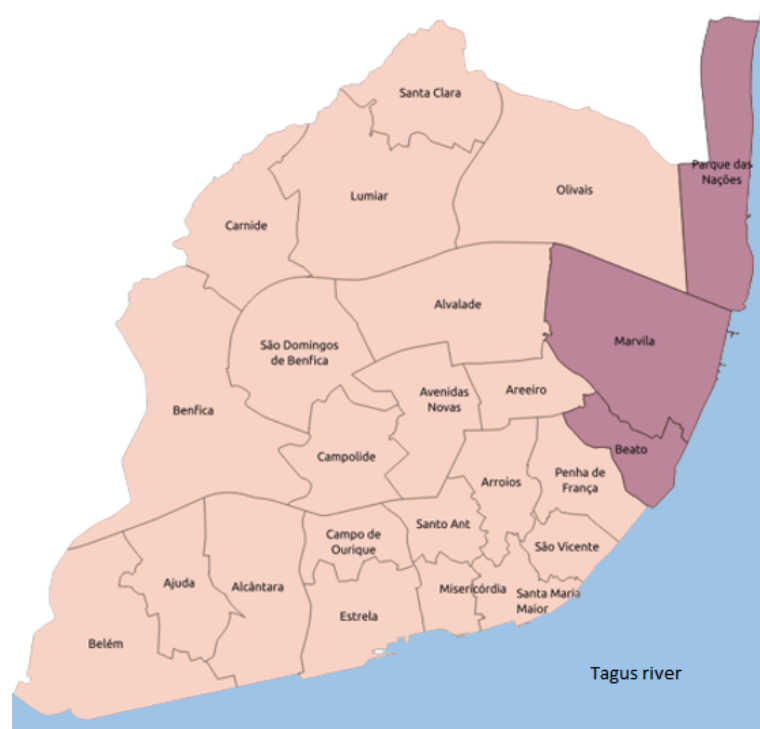


Figure 2. Map of Area of Interest (AoI)—*Beato*, *Marvila*, and *Parque das Nações*.

From Figure 2, it is possible to observe some interesting characteristics from the AoI considered. Firstly, the three council parishes under study are close to the Tagus river, and they are neighbours of each other. Considering this fact, it would be important to analyse the main avenues or roadways that cross these council parishes, since they will be very important for the commuters among these areas.

Considering the characteristics of each council parish, they present some interesting attributes. For instance, *Parque das Nações* is widely known for its commercial, business, and leisure development, providing several infrastructures such as shopping malls (*Vasco da Gama Shopping*), *Lisbon Oceanarium*, Expo 1998 Museum, etc. From this point of view, *Parque das Nações* attracts several citizens, especially during the day, since their workplaces are located in this council parish.

In contrast, *Marvila* and *Beato* come as more residential areas, with several schools, neighbourhoods, and residential buildings. This fact can be justified by the lower cost of living in these council parishes, when compared with *Parque das Nações*, for example.

1.3. Related Work

Bike-sharing systems are in extensive and continuous investigation. Hence, security plays an important part for all users [22]. This work also presents bike-sharing usage as a replacement for short trips and first/last mile usage, correlating data with other existing mobility services. The importance of primary forms of soft mobility, such as bicycles, are well defined and historically seen as a primary solution, as reported in [23]. There is a direct relationship between bicycle usage and high-quality bicycle paths, which means that bicycle research is essential when planning bicycle-sharing docking stations [24].

As [25] demonstrated, there is a need to model and understand the spatio-temporal relationships between the different stations of the BSS, which, with recourse to past data, it was possible to perceive the uses of the various stations and other indicators to apply

forecasting techniques to determine usage rates according to weekdays, weekends, and also holiday periods. They also noted the importance of the temporal granularity, that is, grouping in different time intervals led to different results. In addition to this evidence, data on weather conditions and the state of the weather were used. In this project context, data availability regarding time intervals is limited and slightly restrictive, so that it is pretended to focus on the geographical component of such a study and the geospatial clustering application.

Thus, clustering is one of the most important areas of study regarding Artificial Intelligence and Machine Learning applications. As stated by [26], data clustering is one of the most used techniques when it is set to group a given input data into groups of elements that are similar. Several researchers have used cluster techniques for urban mobility in the last years [27]. Using clustering algorithms like K-Medoids [28], the spectral clustering algorithm [29], and the agglomerative clustering [30], it was possible to find trajectories with similar location patterns [31].

In [32], a clustering process was performed in order to categorise all the stations into several groups. In addition, this experience demonstrated temporal patterns of rush-hour clusters (weekdays), gathering groups of bicycle-sharing docking stations [33].

Regarding some of the studies in mobility context and clustering techniques application, two major categories from clustering paradigm [34] are highlighted for aggregation geospatial data: partitioning clustering [35] and density-based algorithms. One of the most applied algorithms is K-Means [36]. This algorithm has been elected in several cases when compared for example with the K-Medoids [37] since it is not mandatory that founded clusters need to belong to the initial data input set.

One of the other alternative partitioning algorithms is CLARANS [38], and, following the authors, the algorithm was introduced as an extension of K-Medoids; it uses only random samples of the input data (instead of the entire dataset) and computes the best medoids in those samples. It thus works better than K-Medoids for crowded datasets. As mentioned by Comparative Study on Partitioning Techniques [39], K-Means has been revealed as a more efficient algorithm when compared with K-Medoids but also with the ease of implementation when compared to CLARANS, being also well performed on large datasets.

Consequently, the referred algorithms do need the initial parameter of the K value, i.e., the advanced specification of the number of clusters that data is set to be clustered. On other hand, density-based algorithms, such as DBSCAN [40], which does not need the initial configuration for the K value, but instead the ϵ and the MinPts, denotes the Eps-neighborhood of a point, and the MinPts denotes the minimum points in an Eps neighbourhood.

This study was intended to study and discover new bike-sharing docking stations, but considering the state of the art of the public system *Lisbon bicycles*, we pretended to directly compare the findings of our methodology with the existing docking stations already in-site. For that purpose, we believe that the initial parameter of K value to perform the cluster in this context is suitable. Among some of the studied algorithms, K-Means has revealed a proper trade-off among the efficiency, complexity, and performance with larger datasets and other key factors, which in this research project fit properly.

Nevertheless, some other distinct approaches outside the clustering sphere have been developed in the latest years, as is the example of another study regarding the mixed fleet biking system also in Lisbon [41]. In a more generalized view, a survey presented regarding the Machine Learning (ML) applications to a bike-sharing system [42] was also considered, presenting some of the most relevant applications in this field.

This paper is an extender version of conference paper regarding bike-sharing docking stations identification in Lisbon [43].

2. Materials and Methods

For the development of this work, two main subsections were developed. First, the data are verified, explaining the origin, parameters, and constraints. Then, the process is fully detailed, explaining the used methodology and algorithms.

It is known that there are already existent in-situ mobility sharing services (*Parque das Nações, Beato and Marvila*) with several docking stations. Therefore, for the development of this study, these docking services were considered and used as a baseline.

2.1. Data

Telecommunication operators all over the world generate high data volumes. Each mobile device can act as a tracking device, providing vital information regarding where each device went, which can be used to analyse patterns such as location, Points of Interest (PoI), people clusters, specific events, etc. However, these data are susceptible since they deal with personal information and shall not compromise personal human rights.

There are several telecommunication operators in Portugal, in which all of them are subjected to General Data Protection Regulation (GDPR) [44] and also ethical and legacy principles. In the context of this study, more than 70,000 entries from a given Portuguese telecommunication operator was used, regarding the month of January 2020.

Due to the ethical and privacy guidelines implemented, the representation of the several entries must be done considering aggregation, in order to avoid the tracking of devices and consequently identification of the user. For that reason, the geographical representation of those devices, in terms of geospatial characterisation, is using a vector form (as a polygon) of 10×10 m over a Geographic Information System (GIS), also named as Bin or S2 cell [45]. That means that in a given polygon were detected a given number of devices in a certain timestamp.

Although there was a large data set, this information also presents several constraints. Considering the scope of this work, the following constraints were considered:

1. A device might be detected and considered in many entries of the dataset
2. Geographical definition of the area (bin) might be affected by GPS errors and limitations
3. Velocity is the only available attribute assumed for mobility purposes (e.g., walking, running, cycling, stopping, etc.)
4. Limited time window does not allow to fully generalise several mobility patterns, such as holidays and seasonality, among other factors.

Following the GDPR guidelines, the data from each entry can be summarised in Table 1, presented below.

Table 1. Data structure and meaning of each attribute.

Variable Name	Data Type	Description
Identifier	Integer	Unique identifier for each record in dataset. It has no correlation with devices identification.
Day	Date	Date when device was detected in the given area.
Hour	Integer	Corresponds for all the available time frames during the day. Temporal granularity for all records.
Speed	Float	Means the average speed of the devices detected in a given area.
Number of Devices	Integer	Indicates the number of aggregated devices for specified entry in the dataset.
Bin	MultiPoint	S2 Cell defined by list of points given in ESPG:4326. Defines an area similar to a square of 10×10 m.

In order to provide a realistic visualisation of the initial dataset (for *Lisbon* city centre), Figure 3 illustrates the difference between the devices concentration over the *Lisbon* city centre (wire representation of devices than the three council parishes in the study). For that purpose, a dynamic opacity optimisation, considering the number of devices detected and the intensity of the colour of the data points presented in Figure 3 was performed, i.e., darker areas represent a higher concentration of devices (since the data points are represented by a superior opacity). For this illustration, two distinct time frames were used: midnight and noon. Thus, the variables used in this optimisation consist of the number

of devices detected (independent variable) and the respective opacity level (dependent variable). So, normalising the number of devices considered into a normal distribution ranging [0,1], a continuous distribution for the opacity was found, according to the number of devices registered. Thus, darker areas represent a higher amount of devices, and lighter areas represent lower devices concentration. This allowed us to graphically analyse hotspot areas (high smartphone activity registered) and identify the darker areas, namely, the ones where it can be found with more devices in place.

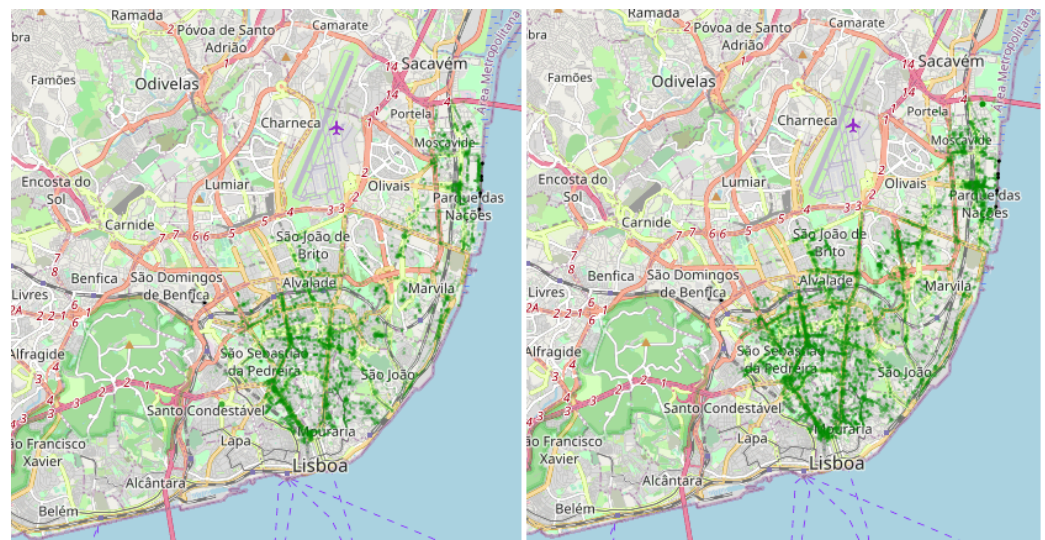


Figure 3. Geographical distribution of devices in *Lisbon* city center. 0 a.m. (on the left) and 12 p.m. (on the right) devices distribution, according to the telecommunication operator. Lighter means fewer devices, and darker means a higher concentration of devices. Background map source: OpenStreetMap [46].

From Figure 3 it is possible to extract some interesting insights, such as:

- Main avenues and roadways present significant devices concentration in both time frames but especially during the day;
- The main residential areas of *Lisbon* city centre are located in council parishes such as *São Sebastião da Pedreira*, *Mouraria*, and *Parque das Nações* (latest belongs to the AoI);
- *Parque das Nações* with interesting devices concentration, justifying the potential of such a council parish regarding mobility studies.

However, since the study was focused on *Parque das Nações*, *Marvila* and *Beato*, we followed a deeper graphical analysis to the distribution of the devices on several time frames, being the output of such analysis detailed in the following figures.

From Figure 4 it is possible to observe the four identified areas, regarding the distribution of the devices in this council parish.

Area 1, as presented in the figure above, represents one of the main focuses during the nighttime frames. It is described as an interesting residential hotspot, as there is Area 2 and Area 3. However, Areas 2 and 3 have some PoI regarding mobility (such as subway and train stations), which justifies, even more, the higher distribution of devices, either during night or day. Lastly, Area 4 represents also a very important area for *Parque das Nações*, since it represents some of the most common places for leisure activities and with interesting buildings, such as the *Lisbon Oceanarium*, for example.

Regarding *Marvila*, there are also four important areas identified during the visual analysis of this article, which led to the creation of four distinct areas of devices agglomeration, as presented in Figure 5.

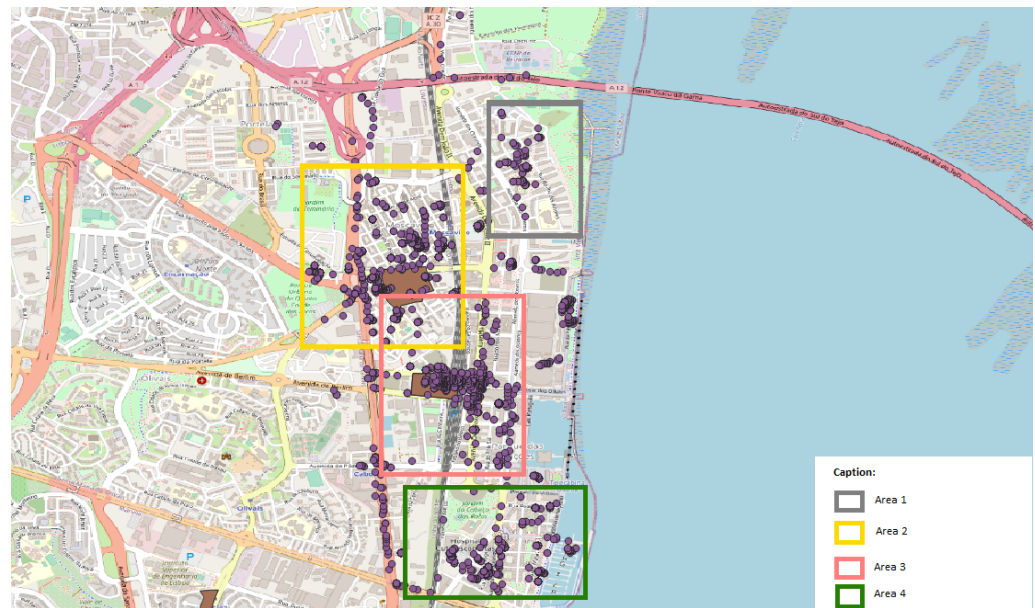


Figure 4. Interesting areas of devices concentration in *Parque das Nações*. Subway and train stations marked as brown. Background map source: OpenStreetMap [46].

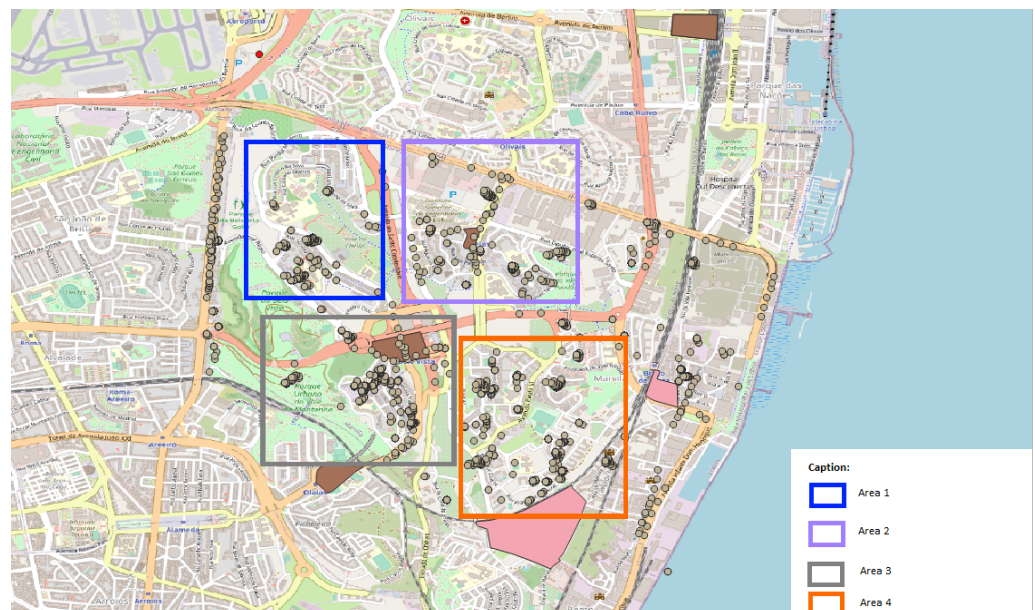


Figure 5. Interesting areas of devices concentration in *Marvila*. Subway stations—brown; train stations—pink. Background map source: OpenStreetMap [46].

For instance, Area 1 shows an interesting device distribution due to residential neighbourhoods in this council parish. Regarding the influence of mobility PoI, we have identified Areas 2, 3, and 4 as the most attended areas, due to the closeness to the existing mobility infrastructures (subway and train stations).

In analogous reasoning, the same approach was performed for *Beato*. In this council parish, we highlight two areas, as presented in Figure 6.

The first area refers to the higher concentration of devices nearby the *Olaias* subway station but also due to the residential buildings presented in the surroundings. Regarding Area 2, this is related to the greater amount of devices during the day, due to the closeness to schools and residential buildings too, which are located close to this area.

After realising the most potential areas of agglomeration in our AoI, the methodology in order to identify the new bike-sharing docking stations is described in detail in Section 2.2.

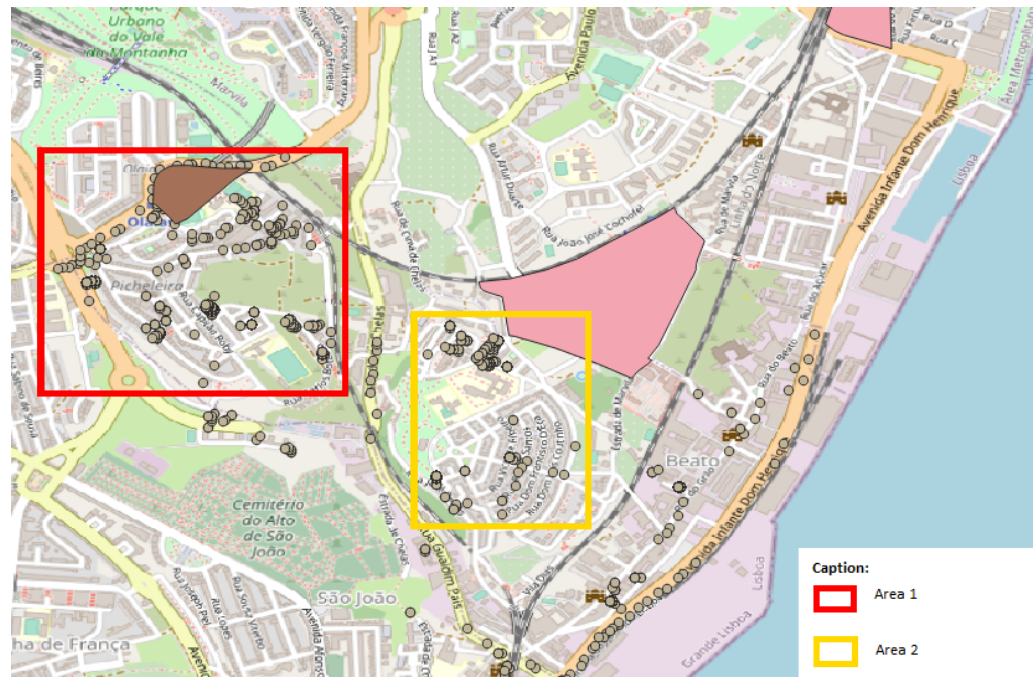


Figure 6. Interesting areas of devices concentration in Beato. Subway stations—brown; train stations—pink. Background map source: OpenStreetMap [46].

2.2. Process

As stated previously, mobile data were received by a Portuguese telecommunication operator. In this context, the first step consisted of developing a methodology pipeline, which started with acquiring a data-gathering mechanism; then, the data were cleaned and analysed.

Some parameters were kept to be used during the analysis and cleaning process, and others were removed (e.g., time frame availability of the data, external factor, public holiday, etc., since they would affect mobility patterns).

Due to privacy policies and according to GDPR, it was not possible to work with trackable and exact geographic coordinates, as mentioned in Section 2.1. Therefore, it was necessary to consider each *bin* as the geographical characterisation of each record (area). Using area parameter for each record, it was verified that it added extreme complexity to the analysis since more than one point was being considered for each record, resulting in a *MultiPoint* structure type. This point led to a necessity for complexity reduction (centroid calculation). The centroid calculation was performed aided by GIS software (finding the centre of each polygon for sets of individual points). The process of generating a single point from the several polygons can be described as follows:

$$C_x = \frac{1}{6A} \sum_{i=0}^{n-1} (x_i + x_{i+1})(x_i y_{i+1} - x_{i+1} y_i)$$

$$C_y = \frac{1}{6A} \sum_{i=0}^{n-1} (y_i + y_{i+1})(x_i y_{i+1} - x_{i+1} y_i)$$

and where A is the polygon's signed area,

$$A = \frac{1}{2} \sum_{i=0}^{n-1} (x_i y_{i+1} - x_{i+1} y_i)$$

Moreover, after this geometrical transformation, it was necessary to create an equitable distribution of our device points (since they have been aggregated); then, an unlist operation was performed. In this sense, each entry was repeated throughout the dataset N times, N

being the number of devices considered in that capture. Figure 7 presents the methodology used in this work.

Due to different geographical asymmetric characteristics in our AoI (Section 1), and considering the existence of soft mobility solutions just for one of the studied parishes (*Parque das Nações*), a comparison analysis was necessary for algorithm validation. The algorithm is expected to perform similar results in BSS stations location.

Aiming the individual study in each parish, a geoprocessing operation was performed—this procedure allowed to delimit the geographic distribution of all docking points. Thus, after applying the mentioned data transformations, the converted data points were intersected to the vectorial information of each council parishes (i.e., *shapefiles*) to obtain individual subdatasets to perform the clustering techniques.

Then, to focus the current study on soft mobility solutions and potential users of these modes of transportation, a velocity filter was applied to the given datasets. Considering the average velocity for these bike-sharing systems, according to [47,48], we selected 20 km/h, independent of the user gender, physical condition, and weather conditions, among others.

With this kind of filtering, it is possible to focus on several interesting devices distribution, such as :

- Locating the main traffic jams—only considering low-speed devices will allow us to focus on the traffic jams that occur in our AoI, enhancing the definition of the clusters in the next step.
- By identifying residential areas, considering the night time frames, it will be possible to also provide and consider the data points that are referenced to residential buildings, social neighbourhoods, etc. These places must be important, as we saw in Section 2.1.
- Identifying workplaces in the daytime, low-speed detection might mean the geographical location of workplaces. This is very important considering the number of commuters that go from home–job and job–home.
- Focusing on soft mobility solutions and bicycle-lanes—filtering devices under 20 km/h allows us to consider also pedestrians and cyclists. This will be important to also collect important data points nearby or over bicycle lanes and close to existing docking stations in our AoI.

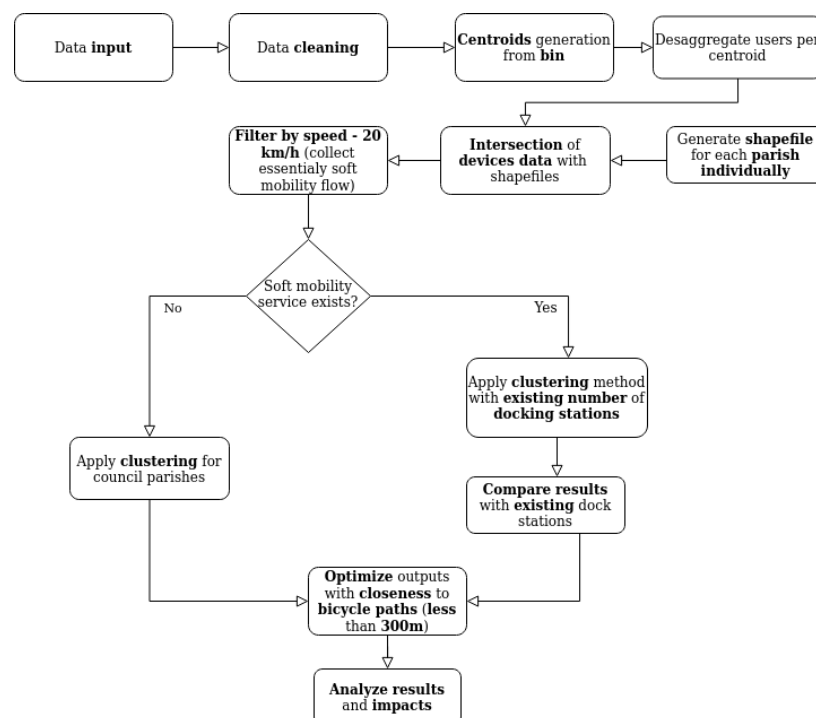


Figure 7. Methodology applied in the current study. Source: [43].

For this work, to perform the clustering over the prepared data, the K-Means [49] algorithm was used. In particular, this algorithm was chosen among several geospatial clustering algorithms, after considering several factors. Hence, considering important characteristics such as complexity, efficiency, ease of implementation, and also the ability to handle large datasets, K-Means clustering presented suitable characteristics to solve the problem presented.

Hence, the first step when running K-Means consisted of defining the number of clusters that should match the same number already installed of BSS docks. So far, there are fourteen docking stations in *Parque das Nações*, as shown in Section 1.1. The same number of docking stations was considered when running K-Means.

As is widely known, K-Means takes advantage of the notion of centroid. As described by [50], “the centroid point is the point that represents its cluster.” Hence, the centroid point is the average of all the points in the set and will change in each step and will be computed by:

$$C_i = \frac{1}{\|S_i\|} \sum_{x_j \in S_i} x_j$$

For the above equation:

- C_i : i 'th centroid
- S_i : All points belonging to set i with centroid as C_i
- x_j : j 'th point from the set
- $\|S_i\|$: number of points in set i

As mentioned, for *Parque das Nações* the output of the described process is going to be compared to the in-situ docking stations. However, for *Marvila* and *Beato*, the *Lisbon bicycles* system is not operating yet, so it is necessary to define the location of such docking stations but also suggest the number of docking stations that, according to the used data, comes as needed for our devices distribution. For that reason, the Sum of Squared Errors (SSE) emerges as a very important metric in order to analyse the variation of its value.

The main goal of the K-Means algorithm is to find K centroid points (C_1, C_2, \dots, C_k) by minimising the sum over each cluster of the sum of the square of the distance between the point and its centroid. Hence, it is pretended to get the lowest values for the metric of SSE, which is calculated as presented below:

$$C_1, C_2, \dots, C_k = \operatorname{argmin} \sum_{i=1}^k \sum_{x \in S_i} \|x - C_i\|^2$$

As stated, the mentioned metric will be used in order to analyse the number of clusters, i.e., docking stations identified either for *Parque das Nações* but especially for *Beato* and *Marvila*. To determine the number of docking stations for these two council parishes, the Elbow Method was used for the K-Means algorithm, providing the suitable trade-off towards the distribution of the SSE and the number of docking stations implemented.

Lastly, and in order to optimise the location of the centroids that are output from the clustering algorithm, and considering the importance of safety and security given by the users and potential users of BSS, Algorithm 1 describes an optimising process in order to better locate the docking stations, considering the existing bicycles lanes over the city. This will allow us to locate as near as possible the docking stations from the bicycle lanes, promoting a better environment to use these kinds of mobility solutions.

The optimised docking stations are the ones that have a distance less than 300 m from the bicycles lanes, following a common and acceptable distance from the literature review in the soft mobility context.

Algorithm 1: Optimization for bike-sharing docking stations

Result: List of docking stations nearby bike paths

Apply K-Means algorithm on specific council parish;

Define threshold for otimization (in meters);

For each centroid(c) output from k-Means **do** **For each** segment(seg) of bike-path **do** Project c in seg, output is np , following nearest point definition;

 Calculate distance(d) from np and c ;

If the minimum distance from d to np is lower than threshold;

 Adopt the point in segment as optimal point (located in bike-path);

EndDo;

EndDo;

Print optimised points;

3. Discussion and Results

In this section, we are going to present the results that have been reached, following the methodology presented in Section 2.2. First, the results regarding *Parque das Nações* are going to be presented and analysed, comparing the algorithm output with the existing *Lisbon Bicycles* docking stations already implemented. After validating this methodology, an analysis of the achieved results for *Marvila* and *Beato* is presented, highlighting the number of docking stations suggested and, most important, a suitable location for those docking stations, presenting some important insights from data used that help to corroborate the presented results.

3.1. Parque Das Nações

As stated in Section 1, the current bike-sharing system implemented in *Lisbon* city has fourteen docking stations available in *Parque das Nações*. For that reason, the K-Means Clustering Algorithm was applied to the provided GSM data, with the 14 clusters as output expected. This will allow us to directly compare the output centroids with the in-site docking stations, to evaluate the generalisation ability of our system.

As a result, Figure 8 presents the geographical distribution of the centroids (dark circles) and the respective subset of data that allowed the generation of those centroids.

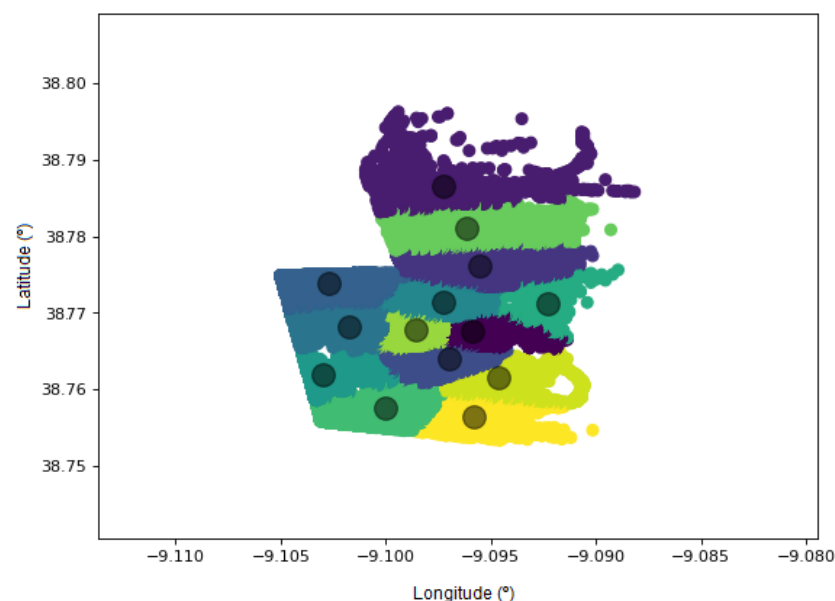


Figure 8. K-Means centroids output for *Parque das Nações*.

As Figure 8 presents, it is possible to identify the fourteen clusters and the data points belonging to each centroid. From the Figure above it is possible to identify that the most north centroid of *Parque das Nações* is simultaneous the one that presents a higher dispersion of data points. That can be explained due to a lower concentration of devices, as we are going to see further, and for that reason, the centroid generated is also affected in terms of the spatial distribution of those data points.

Regarding the other clusters, it is possible to observe a consistent balance among the subsets and the centroids generated. However, this kind of perspective is harder to compare with the existing docking stations, and for that reason, Figure 9 illustrates the comparison between our: (1) direct output from the K-Means algorithm, (2) K-Means centroids optimised (final findings), and (3) existing docking stations installed. For deeper analysis, three different areas have been identified.

In the first area, the *Lisbon Bicycles* system has implemented two different docking stations over two segments of bicycle lanes in this area. However, we have identified this area as an interesting section for analysis and discussion due to these particular attributes. In this area, there is a considerable “inaccessible area” due to existing buildings: a huge fenced terrain that takes a huge part of the mentioned area and also a Waste-Water Treatment Plant (WWTP), which, obviously, is not accessible for common mobile phone users. For that reason, this area has shown particular features that have turned into a specific sub-case inside the *Parque das Nações* analysis area.

For *Parque das Nações*, the distribution of the devices alongside the first section was very residual. For that reason, considering that we have expected the fourteen most concentrated in that region, we find it acceptable to identify only one docking station in that area.



Figure 9. Bike-sharing docking stations in Parque das Nações: Existing docks in-site are represented by yellow dots; red dots represent centroid clusters; dark blue dots represent the optimised locations for K-Means outputs (docking stations over bike paths). Subway has the pink colour polygon, while the purple polygon regards the train station. The map was divided in three segments. Background map source: OpenStreetMap [46].

Regarding the second area, we have the major number of docking stations identified. This kind of event can be justified by some interesting geographical reasons, such as:

- Closeness to the river—this part of the map is alongside the west bank of the Tagus river. Due to this, several maritime activities are then located in this area, such as *Porto de Lisboa*, for example.
- Commercial buildings—in area number two, as marked in Figure 9, we can find some attraction PoI such as restaurants and *Vasco da Gama Shopping*. Due to the existence of

such attraction services, the amount of data points there is much higher than in the first area (in the north), for example.

- Leisure places—as referenced above, it is intrinsic to the closeness of area/section two to the river. Consequently, several activities that often depend on the water can be found there, such as *Lisbon Oceanarium* and Lisbon marina. Additionally, we can find some EXPO 1998 buildings too.
- Workplaces—in this area we can find several offices and, for example, in this sense, these surrounding areas are more often affluent during the day, causing the common known rush-hours, but during the night the concentration of the devices gets much lower.

Describing several reasons for the higher number of docking stations identified in Area 2 of Figure 9, it is now time to visually analyse our outputs. In a general observation, we can verify that dark blue dots, i.e., our findings, are very close to the existing docking stations (yellow dots). That means that our workplan and followed methodology represents a reliable and accurate approach to find soft mobility hotspots, in particular, BSS docking stations. However, it is still possible to identify an isolated dark blue point, which has no close neighbours. After noticing that fact, we looked for the existing nearby services and other PoI that might justify that identification, apart from one of major factors in mobility: bicycle lanes closeness. This upper east point is the nearby a juvenile garden and very close to Tagus river, and also it is also linked to the existing bicycle lanes available in *Parque das Nações*. For that reason, this hotspot is seen as a serious and potential candidate for further additions in the current system operating in Lisbon. Lastly, we have now area/section three. This area, as described in the map, is very important because it contains several interesting conclusions that are possible to be drawn. The first of them is the fact that this region is currently uncovered by the *Lisbon bicycles* system. In addition, four out of fourteen clusters are located inside this area, meaning that 30% of our findings for *Parque das Nações* highlights the need to increase the geographical coverage of the current system in this council parish.

Taking a look at the exact location of such clusters (dark blue dots in area three), we can express the correlation between the train station in *Moscavide* (pink polygon) and also the train and simultaneously the subway station *Gare do Oriente*, which aggregates three out of the four dots in this area. This fact underlines the importance of locating docking stations nearby these kinds of mobility PoI, regarding such agglomerations of devices found. For *Gare do Oriente*, in particular, we can see two docking stations identified for each side of the station, pointing out the massive number of devices nearby this location, as we have already analysed in Section 2.1.

To study the identification of the lower-left side dot of the section (the Southeast one), it was unveiled the location of such point: *Avenida Infante Dom Henrique*. As explored before, this is one of the most important roadways in Lisbon's city centre and, moreover, in our AoI. Since this avenue crosses all three parishes of study, and as seen previously such traffic jams and congestions within this access in *Parque das Nações*, the identification of this "lonely" dot provides a key role for future improvements in this system. As an increase, the location of the latest point is in accordance with the public plan released by *Câmara Municipal de Lisboa* to extend the current bicycle lanes network, as shown by [21].

In order to better understand and comprehend the distribution of the number of docking stations suitable for each area presented in Figure 9, the concentration of the different devices for individual areas are presented in Figure 10.

In order to measure how far the centroids output distance from the existing bike-sharing docking stations, an evaluation algorithm was applied. For each centroid, the closest GIRA docking stations were then identified, and after that, the Haversine distance (explained in Section 2.2) between the two coordinates was computed. The global results are presented in Figure 11.

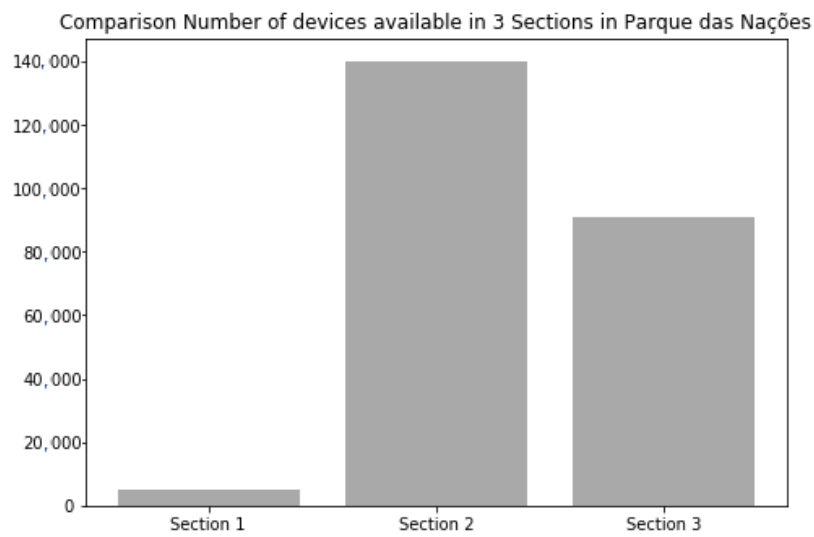


Figure 10. Comparison: Number of Devices in the 3 different areas in *Parque das Nações*.

As we can see in Figure 11, our results are, in its majority, very close to the existing docking stations. As we can see in the plot above, we identified devices’ concentrations nearby the existing docking-stations in *Parque das Nações*, corroborating what was shown in Figure 9.

From Figure 11 it also possible to extract other important conclusions regarding this evaluation metric. One important insight is the fact that there are many docking stations identified very close to the existing ones. For example, six of the predict docking stations are distancing under 200 m from the existing ones, as shown in Figure 12 being here a perfect match between the algorithm output and the existing docking stations. In this sense, we can conclude that area/section two produced the best results, according to this plot.

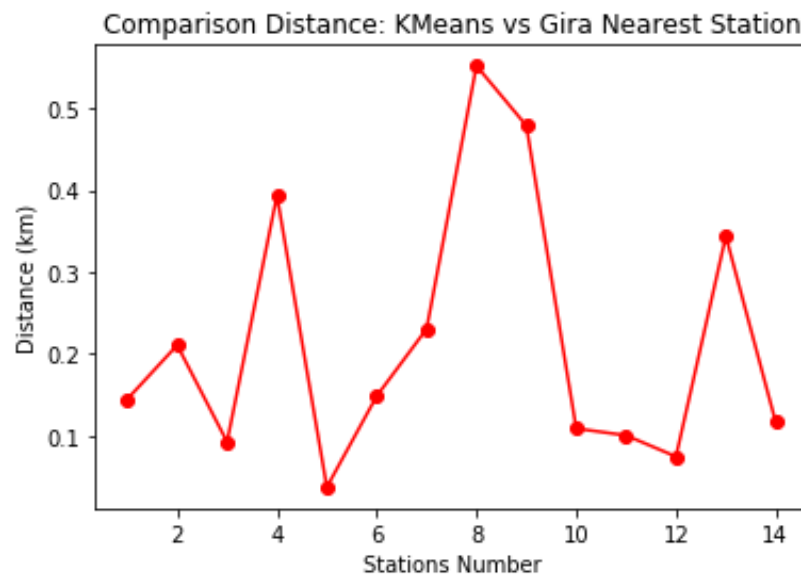


Figure 11. Distancing comparison: K-Means output vs. Gira bicycles docking stations.

On the other hand, the most away stations in *Parque das Nações* output belongs to area three, being the north-east and the south-east points of this area that are the most far away from the existing ones. However, this kind of results does not mean any kind of poor performance from the algorithm and the methodology, since the reasons have been studied and, as mentioned, they represent the identification of new docking stations in the uncovered area.

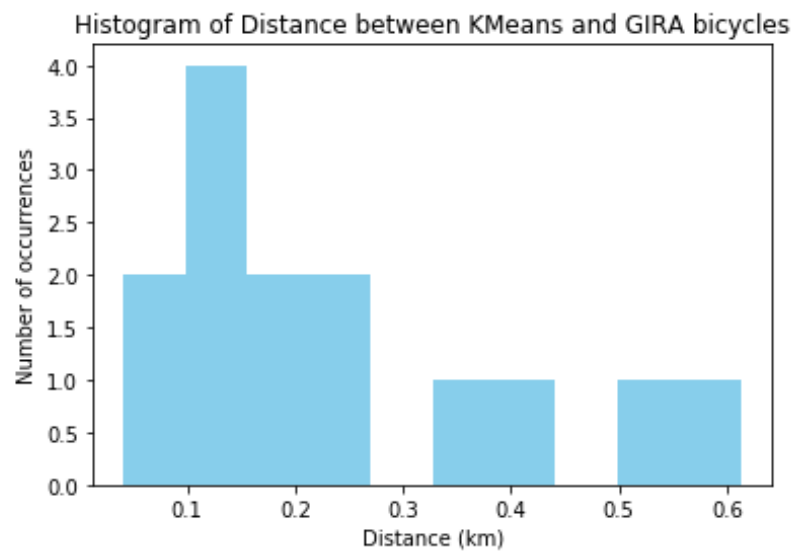


Figure 12. Histogram of Distances comparison for K-Means and GIRA stations.

An important note for this evaluation process is the fact that the Haversine distance has considered every clusters output from the algorithm and compared them with the nearest docking station from the current system implemented. Due to this fact, this analysis might consider more than once the same docking station. After verifying this, we concluded that the repetition of the nearest docking stations occurred mainly for section three, where the system is not implemented yet.

From Figure 13 it is possible to verify the variation of the Sum of Squared Errors according to the number of clusters (k) defined. As stated in Section 1, Lisbon has a public bicycle-sharing system operating in *Parque das Nações* with fourteen docking stations on-site at the moment.

As the plot suggests, the current number of docking stations available does not match with the Elbow Method that is used for *Beato* and *Marvila* cluster’s number definition. That said, it is extremely important to consider the mobility context we are studying and, for that reason, to enhance the need of having a wider geographical coverage in order to maximise the efficiency and usage of such system, as reported in Section 1.

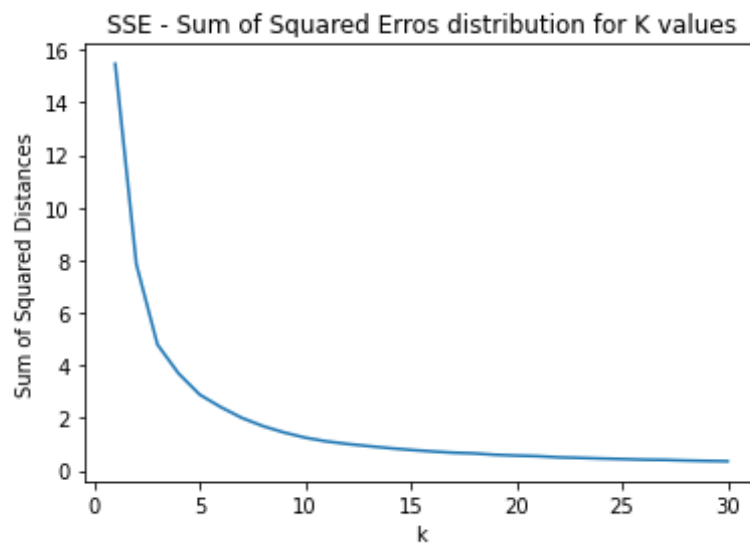


Figure 13. SSE Plot for *Parque das Nações*.

3.2. Beato and Marvila

After applying the described methodology in *Parque das Nações*, and considering the results achieved for that area, the implemented system presents the ability to identify accurate locations of docking stations for Lisbon's BSS.

Hence, considering that the system is intended to be expanded to new areas in Lisbon's city centre, *Marvila* and *Beato* represent two of the adjacent council parishes for *Parque das Nações*. Thus, *Beato* and *Marvila* represent an interesting challenge in the mobility context, in order to replicate the methodology and then discover the location for these two council parishes.

In order to identify the location of new docking stations in *Beato*, and then in *Marvila*, an additional step was taken into account, when compared to *Parque das Nações*: defining then the number of docking stations that should be installed in this council parish. For that reason, the K-Means algorithm was applied for several K values in order to determine the optimal K, i.e., the most suitable K value for the *Beato* data distribution. In Figure 14, the distribution of Sum of Squared Errors (SSE) for the respective K value is presented, in *Beato's* case.

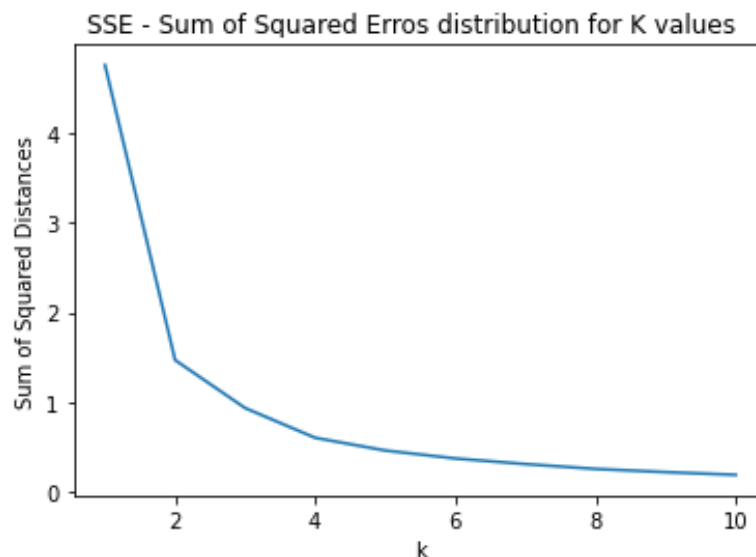


Figure 14. SSE variation according to K values in Beato.

So, in order to identify the suggesting number of docking stations for *Beato*, one of the most well-known methodologies for the identification of the best k value in clustering problems is with the Elbow Method. So, analysing the SSE plot for K values in *Beato*, we can assume that $K = 3$ represents the “Elbow point,” since at this value the error decreases in a significant way, in comparison with the error value. However, as discussed in Figure 13, the stabilisation of the error value is something that is important, considering the spatial distribution of devices and also the context of this problem.

In this sense, we can see from the plot above that the error gets lower distortion when $K = 4$. At this point, the SSE value gets less than 1, which indicates now a very significant reduction in this metric value if we consider $K = 2$ or $K = 3$. Hence, since the value is getting lower, that means the cluster identification process is getting more “accurate” considering the dataset. That means it can be used as a measure of variation within a cluster, and considering the K-Means algorithm assigns all data points to the closest centre based on their Euclidean distance. Since considering a geospatial problem, it is measuring the deviation from the centroid to the data points that were used for its generation. So that, and according to the measurable results obtained, it means if choosing a minimum number of docking stations to be installed as $K = 5$, this represents an SSE of 0.462912. From the *Parque das Nações* experience, the average distancing for the SSE value in this council

parish was described; so, based on similar reasoning, here in Beato we are obtaining, in a general way, half of the SSE when compared to Parque das Nações, which means obtaining centroids distancing, on average, 200/300 m from the real location of devices, which is perfectly aligned with some of the generic guidelines for BSSs.

Regarding *Marvila*, the reasoning applied was very similar to the one in *Beato*. Figure 15 illustrates the variation of SSE for this council parish, in analogous reasoning followed for *Beato*.

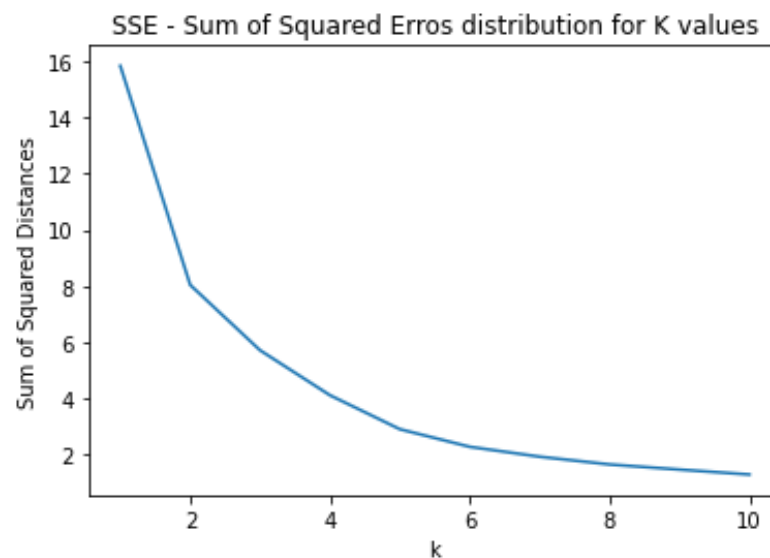


Figure 15. SSE variation according to K values in Marvila.

Following the same argument as stated for *Beato*, as should be expected, the SSE reflects a curve with a sharp decrease as the K values increase. Moreover, it is still possible to suggest that the “Elbow point” for *Marvila* is somewhere between $K = 3$ or $K = 4$, but the corresponding SSE value is very high, considering the physical representation of such values in terms of distancing from the device’s location. In this sense, and taking into account the real-world context and urban mobility planning issues, it is suggested to consider a higher K value in order to that way decrease the value of SSE to a similar scale to the previous ones, i.e., for the other two parishes. Nevertheless, it is notorious for the higher dispersion of data (better shown in 65), and for that reason, here it is output higher values from the SSE. The discussion for this behaviour is going to be discussed next. Thus, in order to keep available sparse solutions regarding the docking stations installation, and following the the same concept of SSE stabilisation, it is suggested considering $K = 8$ as a minimum number of docking stations to be installed in *Marvila*, and it still outcomes an SSE error of 1.662275, which represents a higher distribution of of devices (data points), when compared to the centroids generated. This number ($K = 8$) was chosen following the same criteria before, pointing for a flat stage of the metric in analysis, and it also minimises the effects of a higher geographical distribution of points.

In order to provide a brief summary of the achieved results, and considering the described method to identify the suitable number of clusters for each of the council parishes, Table 2 illustrates the number of clusters identified and the corresponding SSE value.

Table 2. SSE—Sum of Squared Error for best K, following *Elbow method*.

Council Parish Name	Number of Clusters	SSE—Sum of Squared Errors
Parque das Nações	14	0.853283
Marvila	8	1.662275
Beato	5	0.462912

Being a real-world problem, and in particular, one that aims to define the locations of the bike-sharing docking station, one of the most important outputs is the position of such docking stations regarding the dataset used, and consequently, the real location of such points. For that reason, a brief analysis of the clusters obtained for *Beato* and *Marvila* is presented below.

As detailed before, for *Beato* the addition of five new docking stations is suggested. If we look at the X and Y axes, they represent the longitude and the latitude, respectively. Considering the variation from such axis, we can verify that, for example, the longitude variation occurs in two hundredths. Mapping this variation into the real world, it is possible to note the geographical distribution of the points is according to the natural boundaries of *Beato*.

Regarding *Marvila*, this council parish has been revealed as a more challenging scenario. Considering a direct comparison with *Beato*, it is possible to verify a significant higher variance in the X and Y axis, since *Marvila* has more area than *Beato*. For instance, Figure 16 expresses a variance of two-hundredths on longitude scale, whereas Figure 17 presents a variance of five decimals. Transposing such facts to the real-world context, that is verified due to the extension of these two parishes, where *Beato* presents an area of 2.46 km², while *Marvila* comes with 7.12 km², according to [51].

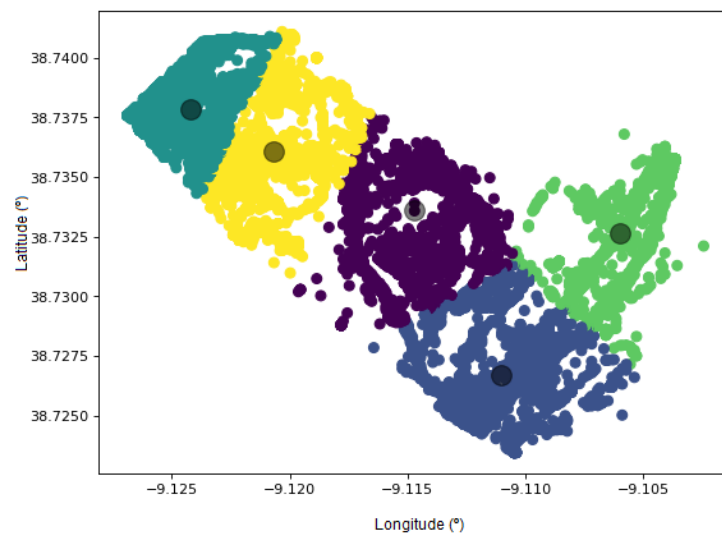


Figure 16. K-Means centroids output for Beato.

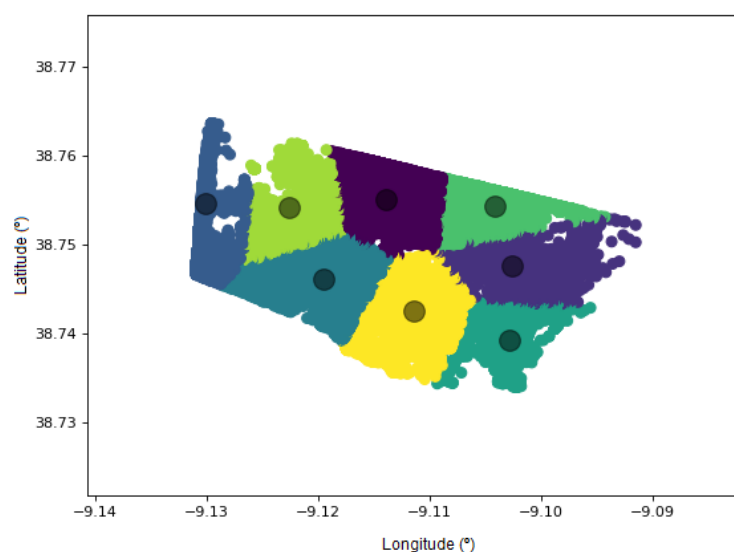


Figure 17. K-Means centroids output for Marvila.

Considering the distribution of data points and clusters identification in Figure 17, it is possible to observe data points distribution is confined to the geographical boundaries of *Marvila*, as we can see in Figure 19, validating the previous process of data preparation. Alongside this conclusion, it is possible to call attention for the balanced distribution of centroids according to the points, being possible to admit they are in a central area of their neighbours, which allows us to conclude they will provide such positive impact in the mobility context, since they are equitably distancing from all the devices nearby (and belonging to such centroid). In a pragmatic point of view, it is notorious that clustering data points from *Marvila* are harder than clustering for any other parish from our AoI, which is reflected in the SSE plot and the number of docking stations suggested, but also here, from the visual analysis of such clusters. Hence, even with a balanced distribution of the clusters, they are representing wider dispersion of geographical points, which can be justified by the much higher area from *Marvila*, when compared to *Parque das Nações* or *Beato*, and consequently, greater distance between the data points.

That said, it is important to underline that, even with a higher dispersion and error for *Marvila*, the number of devices within the dataset do not suggest the need of increasing the number of docking stations regarding this parish has fewer devices when compared to *Parque das Nações*, but they are further away, leading to the the greater value of SSE, as explained before. Considering this, and realising that the definition of the minimum number of docking stations have taken into account the geographical distribution of devices and number of devices detected, this process implies a trade-off towards the geographical coverage of provided datasets, but also the potential usage rate considering the installation of docking stations: *Parque das Nações* provides more devices when compared to *Marvila*, while *Beato* is the least agglomerated parish (according to GSM data provided).

After realising the minimum amount of docking stations set to install or add in the different council parishes, it is time to analyse the suggested locations for these soft mobility hotspots. Concerning *Beato*, as documented before, there was found out that five docking stations would be a minimum number for this parish, according to the different criteria mentioned. The output is presented in Figure 18.

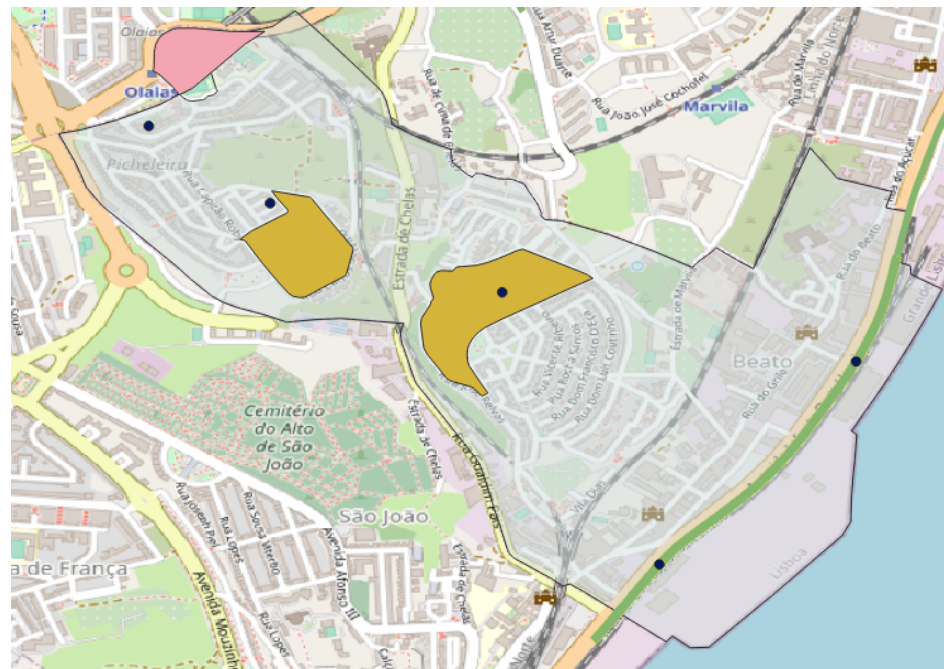


Figure 18. Bike-sharing docking stations prediction for Beato. Dark blue dots represent the optimised locations for K-Means outputs (docking stations over bike paths). Subway has the pink colour polygon while the purple polygon regards the train station. Schools are coloured yellow and bike paths green. Background map source: OpenStreetMap [46]. Source: [43].

As possible to see in Figure 18, our docking stations have already been optimised, since the two docking stations close to the Tagus river are already over the bicycle lanes available in this parish. Here in *Beato*, the identification of hotspots has been corroborated by some insights we have concluded from *Parque das Nações*. For that reason, and since *Marvila* output shares mainly these characteristics, a detailed analysis is going to be made for both parishes.

Firstly, from the observation of Figures 18 and 19, it is noticeable the importance of the mobility PoI in the agglomeration of devices. For instance, and considering *Olaias* subway station for *Beato* and *Chelas* and *Bela Vista* subway stations in *Marvila*, we can conclude the high affluence on such areas, also intensified in *Olaias* subway station due to having the *Olaias Plaza* shopping centre in the same location. In some sense, educational PoI still represents influence on the clusters identification process, as there get populated by students but also from professors and staff for example, which represents usually hotspots of devices. In *Beato* and *Marvila* this is something possible to conclude, with *Escola Antonio Verney* or *Escola Básica Duarte Pacheco* for *Beato*, and for example *Centro Artes Marvila* and *Escola Básica de Marvila* being also identified.

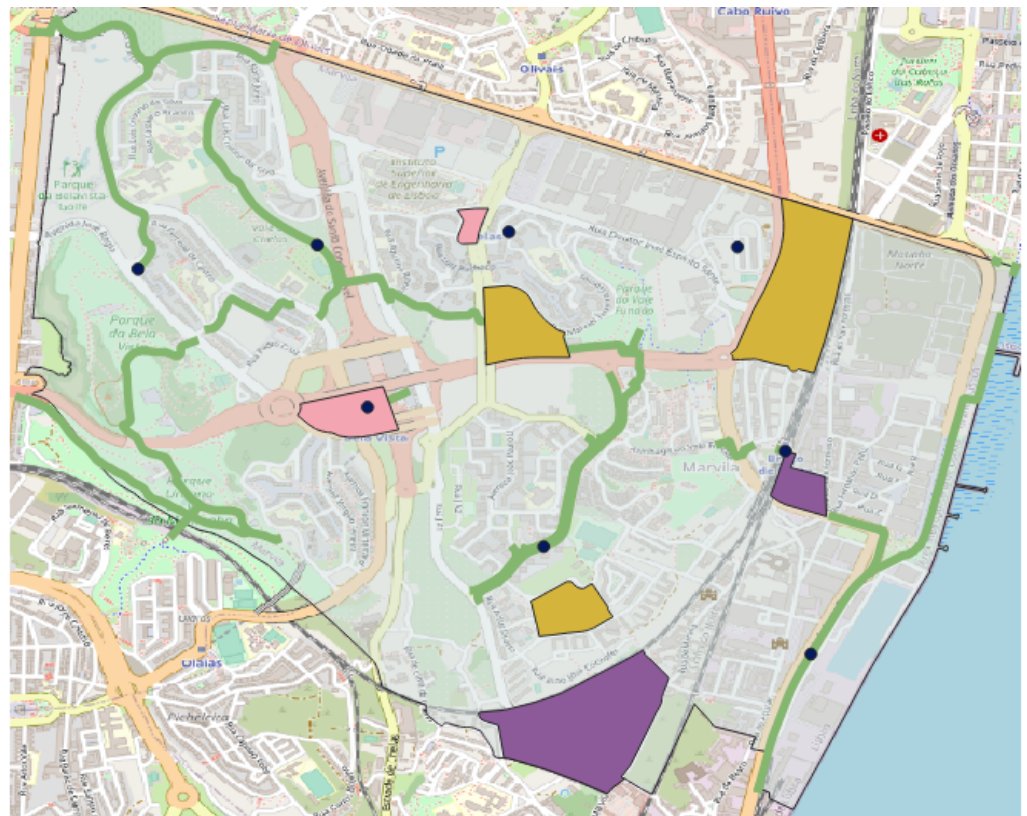


Figure 19. Bike-sharing docking stations prediction for Marvila. Dark blue dots represent the optimised locations for K-Means outputs (docking stations over bike paths). Subway has the pink colour polygon while the purple polygon regards the train station. Schools are coloured yellow and bike lanes green. Background map source: OpenStreetMap [46]. Source: [43]

In summary, this approach led to a cluster-based process in order to identify new bike-sharing docking stations in *Lisbon's* city centre, using for it a well-known clustering algorithm, the K-Means. Consequently, this article presents the suggested locations and a suitable number of docking for the AoI, according to the dataset used. The final findings are summarised in Table 3.

Table 3. Number of docking stations suggested to be added for each of the council parishes under study.

Council Parish Name	Number of Docking Stations to Be Added
Parque das Nações	4
Marvila	8
Beato	5

Moreover, the suggestion for *Parque das Nações* illustrates the addition of the four docking stations located over Area3 (the uncovered region of the council parish), which evidences the need of the system on to start operating in this area. Considering the preliminar data provided by CENSOS 2021 [52], *Parque das Nações* comes with 22,382 inhabitants. For *Marvila* and *Beato* we suggest the addition of, at least, eight and five docking stations, respectively. Thus, the operating system is intended to serve 35,482 citizens in *Marvila* and 12,185 citizens in *Beato*.

4. Conclusions

Urban mobility represents an issue of extreme importance for climate change purposes alongside the United Nations sustainable goals. Hence, it is important to use sustainable modes of transportation and also to increase the efficiency of existent systems already developed.

The main objective of this study was to identify new docking stations for bicycle-sharing systems in three different council parishes of *Lisbon* city: *Beato*, *Marvila*, and *Parque das Nações*. In this latest council parish, and as in other regions of *Lisbon's* city centre, there is already a bicycle-sharing system operating, denominated *Lisbon Bicycles*. In this sense, and considering the possible growth and optimisation of the current system, a cluster-based approach was implemented in order to discover the new docking stations to be installed in the mentioned area.

As presented, and starting from *Parque das Nações*, the council parish used to validate the followed methodology and our data processing pipeline, it was found interesting results, since it was possible to directly compare our finding with the existing ones, presenting small distancing from the in-site docking stations. Besides, it was found an important uncovered area in *Parque das Nações*, since the current system is not implemented in part of the council parish, and our study has highlight the necessity of increase the available system to those locations with several docking stations identified.

And after validating our results in the previous council parish, the same methodology was applied to *Beato* and *Marvila*, where the service is not operating yet. And for these two council parishes, it was used the Elbow Method to find a suitable number of docking stations to be installed, being identified five and eight docking stations for *Beato* and *Marvila*, respectively.

In order to optimise the location of such docking stations, an optimising process considering the existing bicycle lanes in *Lisbon* was considered, in order to locate as close as possible the docking stations from the bicycle lanes. This was a very important step in this study, since the security and safety are two out of the most important factors for bicycle-sharing system users and population, in general.

Moreover, the importance of the BSS in the context of facing carbon emissions and air pollution is underline by the concept of soft mobility. This concept includes carbon free modes of transportation, and for that reason, eco-friendly. In this sense, soft mobility is one of the ways to reduce the climate changes effects and consequently, the carbon emissions.

In specific, *Lisbon* presents an interesting case study regarding urban mobility planning and traffic optimisation, since this city has been recognised as an hotspot for traffic jams and congestions.

Considering the global approach and the major data source for this research project, it is possible to assume that due to the sensibility of the GSM data and the GDPR, the necessity of aggregating such data created some constraints an limitations. Firstly, the time window used was limited and consequently, the device distribution of the considered period does not allow to generalise for a wider time interval, considering important factors

in mobility such as holidays, weather impact, among others. Additionally, the geographical aggregation represents also a constraint, since the mean speed was also calculated considering the devices in the same S2 cell. Considering these limitations, this study is intended to go further using other data sources, for example motion sensors to be installed in the most critical and identified areas in Lisbon, so that it can be avoid the devices distribution attached to the given telecommunication operator and use a wider time interval. Looking forward, it is also set to study the possibility to apply different clustering algorithms, such as KMedoids or even density-based algorithms such as DBSCAN to compare the findings with the current presented in this paper. The possibility to include other factors such as the weather impact or the influence of bus stops and other important PoI in the city are considered.

Nevertheless, the development of the described decision support system may help the competent authorities, in this case *Lisbon Municipality*, to better planing the geographical distribution of the available BSS. However, the importance of that study and the findings that were shared, reveal a potential generalisation of the process and consider, in the future, the installation of other soft-mobility solutions such as scooters or roller-skating in the identified locations, since it can serve the population in different mobility choices but keeping the identified locations that comes from this study.

Author Contributions: Conceptualization, T.F., M.A., P.V.F. and P.N.; Methodology, T.F. and M.A.; software, T.F.; Validation, T.F., M.A., P.V.F. and P.N.; Investigation, T.F., M.A.; Resources, T.F., M.A., P.V.F. and P.N.; Data curation, T.F.; writing—original draft preparation, T.F.; writing—review and editing, T.F., M.A., P.V.F. and P.N.; Visualization, T.F.; supervision, M.A.,P.V.F. and P.N.; Project Administration, M.A.,P.V.F. and P.N.; funding acquisition, M.A. and P.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by ANI—Agência Nacional de Inovação within the R&D Units Project Scope: POCI-01-0247-FEDER-045919/2019.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Singh, R.P.; Singh, A.; Srivastava, V. (Eds.) *Environmental Issues Surrounding Human Overpopulation*; Advances in Environmental Engineering and Green Technologies, Information Science Reference; Hershey, PA, USA, 2017.
2. United Nations. 2030 Agenda for Sustainable Development. 2015. Available online: <https://sdgs.un.org/2030agenda> (accessed on 27 January 2021).
3. Nations, U. 17 Sustainable Development Goals. 2015. Available online: <https://sdgs.un.org/goals> (accessed on 29 January 2021).
4. Commission, E. European Climate Pact. 2020. Available online: https://ec.europa.eu/clima/eu-action/european-green-deal/european-climate-pact_pt (accessed on 29 January 2022).
5. Lam, D.; Head, P. Sustainable urban mobility. In *Energy, Transport, & the Environment*; Inderwildi, O., King, S.D., Eds.; Springer: London, UK, 2012; pp. 359–371. [CrossRef]
6. Climate ADAPT. National Circumstances Relevant to Adaptation Actions. 2021. Available online: <https://climate-adapt.eea.europa.eu/countries-regions/countries/portugal> (accessed on 2 February 2022).
7. Liberato, L. O Potencial dos Sistemas de Bicicletas Partilhadas: Uma Contribuição para a Construção do Paronama Português. Master's Thesis, Faculdade de Engenharia-Universidade do Porto, Porto, Portugal, 2018.
8. Portuguese American Journal. Lisbon: Bike Sharing Service Available for Locals and Visitors—Portugal. 2021. Available online: <https://portuguese-american-journal.com/lisbon-bike-sharing-service-available> (accessed on 2 February 2021).
9. DeMaio, P. Bike-sharing: history, impacts, models of provision, and future. *J. Public Transp.* **2009**, *12*, 41–56. [CrossRef]
10. Macioszek, E.; Świerk, P.; Kurek, A. The Bike-Sharing System as an Element of Enhancing Sustainable Mobility—A Case Study based on a City in Poland. *Sustainability* **2020**, *12*, 3285. [CrossRef]
11. Jan Wisniewski-RESETI. Paris's Public Bike Sharing Schemes: Win or Fail for the Environment? 2019. Available online: <https://en.reset.org/blog/pariss-public-bike-sharing-schemes-win-or-fail-environment-09302019> (accessed on 5 February 2021).

12. Winslow, J.; Mont, O. Bicycle Sharing: Sustainable Value Creation and Institutionalisation Strategies in Barcelona. *Sustainability* **2019**, *11*, 728. [CrossRef]
13. Mary K. Pratt-TechTarget. ICT (Information and Communications Technology, or Technologies), 2019. Available online: <https://searchcio.techtarget.com/definition/ICT-information-and-communications-technology-or-technologies/> (accessed on 8 February 2021).
14. Albino, V.; Berardi, U.; Dangelico, R.M. Smart cities: Definitions, dimensions, performance, and initiatives. *J. Urban Technol.* **2015**, *22*, 3–21. [CrossRef]
15. Smar50Awards. Lisboa Distinguida Como Smart City. 2019. Available online: <https://www.smartplanet.pt/news/smart-cities/lisboa-distinguida-como-smart-city> (accessed on 7 February 2021).
16. Agency, E.E. Air Pollution is the Biggest Environmental Health Risk in Europe. 2021. Available online: <https://www.eea.europa.eu/themes/air> (accessed on 30 September 2021).
17. News, T.P. Lisbon is Most Congested City in the Iberian Peninsula. 2017. Available online: <https://www.theportugalsnews.com/news/lisbon-is-most-congested-city-in-the-iberian-peninsula/41154> (accessed on 27 October 2021).
18. TomTom. TomTom Traffic Index. 2019. Available online: https://www.tomtom.com/en_gb/traffic-index/ranking/ (accessed on 17 February 2022).
19. Midgley, P. *Bicycle-Sharing Schemes: Enhancing Sustainable Mobility in Urban Areas*; United Nations, Department of Economic and Social Affairs: New York, NY, USA, 2011; Volume 8, pp. 1–12.
20. EMEL. Lisbon Bicycles 2018 Statistics. 2018. Available online: <https://www.emel.pt/pt/noticias/bicicletas-gira-ja-rolaram-1-milhao-de-viagens-2-2/> (accessed on 16 May 2021).
21. Motor24. The Planned 200 km of Bicycle Planes in Lisbon. 2019. Available online: <https://www.motor24.pt/sites/wattson/estamos-a-dois-anos-de-termos-200-km-de-ciclovias-em-lisboa/542831/> (accessed on 22 September 2021).
22. Wang, J.; Huang, J.; Dunford, M. Rethinking the Utility of Public Bicycles: The Development and Challenges of Station-Less Bike Sharing in China. *Sustainability* **2019**, *11*, 1539. [CrossRef]
23. Zayed, M. Towards an index of city readiness for cycling. *Int. J. Transp. Sci. Technol.* **2017**, *5*, 210–225. [CrossRef]
24. National Association of City Transportation Officials. *Equitable Bike Share Means Building Better Places for People to Ride*; National Association of City Transportation Officials: New York, NY, USA, 2016. Available online: https://nacto.org/wp-content/uploads/2016/07/NACTO_Equitable_Bikeshare_Means_Bike_Lanes.pdf(accessedon28February2022).
25. Yang, Z.; Hu, J.; Shu, Y.; Cheng, P.; Chen, J.; Moscibroda, T. Mobility modeling and prediction in bike-sharing systems. In Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services, Singapore, 26–30 June 2016; ACM: Singapore, 2016; pp. 165–178. [CrossRef]
26. Bano, S.; Khan, N. A Survey of Data Clustering Methods. *Int. J. Adv. Sci. Technol.* **2018**, *113*. [CrossRef]
27. Myeong Lee, G.M.; Aghi, R. *Exploratory Cluster Analysis of Urban Mobility Patterns to Identify Neighborhood Boundaries*; Platial Analysis Lab, Department of Geography, McGill University: Montreal, QC, Canada, 2017.
28. Kaufmann, L.; Rousseeuw, P. Clustering by Means of Medoids. In *Data Analysis Based on the L1-Norm and Related Methods*; Faculty of Mathematics and Informatics, Vrije Universiteit Brussel: Brussel, Belgium, 1987; pp. 405–416.
29. Ng, A.Y.; Jordan, M.I.; Weiss, Y. On Spectral Clustering: Analysis and an Algorithm. In Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic, Vancouver, BC, Canada, 3–8 December 2001; MIT Press: Cambridge, MA, USA, 2001; pp. 849–856.
30. Day, W.H.E.; Edelsbrunner, H. Efficient algorithms for agglomerative hierarchical clustering methods. *J. Classif.* **1984**, *1*, 7–24. [CrossRef]
31. Ben-Gal, I.; Weinstock, S.; Singer, G.; Bambos, N. Clustering Users by Their Mobility Behavioral Patterns. *Acm Trans. Knowl. Discov. Data* **2019**, *13*, 1–28. [CrossRef]
32. Feng, Y.; Affonso, R.C.; Zolghadri, M. Analysis of bike sharing system by clustering: the Vélib' case. *IFac-Papersonline* **2017**, *50*, 12422–12427. doi: 10.1016/j.ifacol.2017.08.2430. [CrossRef]
33. Ma, X.; Cao, R.; Jin, Y. Spatiotemporal Clustering Analysis of Bicycle Sharing System with Data Mining Approach. *Information* **2019**, *10*, 163. [CrossRef]
34. Everitt, B.; Landau, S.; Leese, M.; Stahl, D. *Cluster Analysis*; John Wiley & Sons (Wiley), Ltd.: Hoboken, NJ, USA, 2011. [CrossRef]
35. Jin, X.; Han, J. *Partitional Clustering*; Springer: Boston, MA, USA, 2011. [CrossRef]
36. Likas, A.; Vlassis, N.; J. Verbeek, J. The global k-means clustering algorithm. *Pattern Recognit.* **2003**, *36*, 451–461. [CrossRef]
37. Jin, X.; Han, J. K-Medoids Clustering. In *Encyclopedia of Machine Learning*; Sammut, C., Webb, G.I., Eds.; Springer: Boston, MA, USA, 2010; pp. 564–565. [CrossRef]
38. Ng, R.; Han, J. CLARANS: a method for clustering objects for spatial data mining. *IEEE Trans. Knowl. Data Eng.* **2002**, *14*, 1003–1016. [CrossRef]
39. Gandhi, G.; Srivastava, R. Review Paper: A Comparative Study on Partitioning Techniques of Clustering Algorithms. *Int. J. Comput. Appl.* **2014**, *87*, 10–13. [CrossRef]
40. Sander, J., Density-Based Clustering. In *Encyclopedia of Machine Learning*; Sammut, C., Webb, G.I., Eds.; Springer: Boston, MA, USA, 2010; pp. 270–273. [CrossRef]
41. Martinez, L.M.; Caetano, L.; Eiró, T.; Cruz, F. An Optimisation Algorithm to Establish the Location of Stations of a Mixed Fleet Biking System: An Application to the City of Lisbon. *Procedia-Soc. Behav. Sci.* **2012**, *54*, 513–524. [CrossRef]

42. Albuquerque, V.; Sales Dias, M.; Bacao, F. Machine Learning Approaches to Bike-Sharing Systems: A Systematic Literature Review. *Isprs Int. J.-Geo-Inf.* **2021**, *10*, 62. [[CrossRef](#)]
43. Fontes, T.; Arantes, M.; Figueiredo, P.V.; Novais, P. Bike-Sharing Docking Stations Identification Using Clustering Methods in Lisbon City. In *Distributed Computing and Artificial Intelligence, Volume 1: 18th International Conference*; Springer International Publishing: Cham, Switzerland, 2022; pp. 200–209.
44. European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016. 2016. Available online: <https://eur-lex.europa.eu/eli/reg/2016/679/oj> (accessed on 10 February 2021).
45. S2 Geometry-Google. S2 Cells. Available online: <https://s2geometry.io/> (accessed on 25 January 2022).
46. OpenStreetMap. OpenStreetMap. Available online: <https://www.openstreetmap.org/> (accessed on 16 February 2021).
47. RoadBiker. What's the Average Speed of a Beginner. Available online: <https://www.roadbikerider.com/whats-the-average-speed-of-a-beginner-cyclist/> (accessed on 27 January 2022).
48. Jensen, P.; Rouquier, J.B.; Ovtracht, N.; Robardet, C. Characterizing the speed and paths of shared bicycle use in Lyon. *Transp. Res. Part Transp. Environ.* **2010**, *15*, 522–524. [[CrossRef](#)]
49. Ahmed, M.; Seraj, R.; Islam, S.M.S. The k-means Algorithm: A Comprehensive Survey and Performance Evaluation. *Electronics* **2020**, *9*, 1295. [[CrossRef](#)]
50. Kumar, S. Understanding K-Means, K-Means++ and, K-Medoids Clustering Algorithms. 2020. Available online: <https://towardsdatascience.com/understanding-k-means-k-means-and-k-medoids-clustering-algorithms> (accessed on 21 June 2021).
51. de Lisboa, M. Mapas e Informação Turística de Lisboa. 2021. Available online: <https://www.mapalisboa.net/> (accessed on 13 May 2021).
52. Instituto Nacional de Estatística. Censos 2021. Available online: https://www.ine.pt/scripts/db_censos_2021.html (accessed on 20 February 2022).