

# Advanced Visual Interaction with Public Bicycle Sharing Systems

Alexandra Cortez  
ViRVIG, Universitat  
Politécnica de Catalunya  
Carrer Llorens i Artigas,  
4-6 08028, Barcelona  
alexandra.cortez@upc.edu

Pere-Pau Vázquez  
ViRVIG, Universitat  
Politécnica de Catalunya  
Carrer Llorens i Artigas,  
4-6 08028, Barcelona  
pere.pau.vazquez@upc.edu

## ABSTRACT

Nowadays, public bicycle sharing systems have become popular and widespread across the world. Their usefulness largely depends on their ability to synchronize with citizens' usage patterns and optimize re-balancing operations that must be carried out to reduce outages. Two crucial factors to tackle this problem are stations' characteristics (geography, location, etc) and the availability of bikes and drop-off slots. Based on the requirements and input from regular users and experts in policy-making, system operation, and urban planning, we have created a web-based visualization system that facilitates the analysis of docking stations' behavior. This system provides the first group with the availability prediction of both bike and free slots in docking stations to assist their planning. Besides, the system helps the second group understand patterns of usage and get deeper insights (e.g. need for resizing or complementary transportation systems) to facilitate decision-making and better fulfill the citizens' needs. In a final evaluation, both groups found it highly useful, effective, and better suited than other existent applications.

## Keywords

Human-centered computing, Visualization systems and tools, Visual analytics.

## 1 INTRODUCTION

Bicycle usage is undergoing a rapidly growing popularity in urban areas thanks to its multiple environmental, economic, social, and health benefits. Bicycles are commonly used as a replacement to other transportation systems for short-to-medium distance trips and for the so-called last-mile (or first-mile) connections to other transportation modes. Most bicycle sharing systems require citizens (from now on called regular users) to borrow a bike by going to one docking station, and later return the bike to another station. This use is commonly complemented by web services or mobile applications where regular users can check availability or current occupation of the stations, as well as booking bicycles in advance. However, information available through official applications is scarce or not reliable enough and oftentimes the interfaces are poorly designed.

Moreover, applications for system monitors and domain experts in urban planning and policy-making

(from now on called advanced users) are not suited for proper monitoring of the system as a whole, since they lack information that may be relevant to these users such as the predicted occupation of a docking station or usage patterns.

The objective of this study is to pilot a web-based visualization system that takes into consideration the requirements of regular and advanced users and provides additional and novel features not existing in other systems. For instance, the prediction of slots or bicycles of each type (mechanical or electrical), analysis of docking stations suffering outages, or a prioritized list of stations close to a point of interest.

## 2 PREVIOUS WORK

Public bicycle sharing systems (from now on BSS), have received a lot of attention lately, since they are a huge opportunity for medium and large cities to reduce pollution and increase the health of their citizens, among other benefits [Ric15].

The analysis of how the BSS are used has been addressed in various ways. One common approach is to study the effect of urban configuration (e.g. elevation of stations) on bicycle flows and the destination preferences [KPL, FR14, FIE15]. Other studies mainly focus on building mathematical models to analyze the characteristics of bike trips or to make trip predictions

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

[BAF\*11, HMO18]. The effect of weather and other variables such as the calendar events has been also investigated [YZWB20]. These studies mostly illustrate their findings with charts and maps, but no visual tool has been developed to help users to drill down the data. Another area of interest is optimizing re-balancing operations to generate efficient vehicle routes, but we focus on a wider audience that includes citizens and advanced users.

With a different focus, several approaches have been developed to face the problem of BSS visualization with diverse goals in mind. In some cases, the visualizations are simply static depictions, while in others, authors provide fully interactive tools or propose the development of multiple channels to increase engagement. The individual customers' journeys have been analyzed by different authors [WBD14, BWB14] to gain insights into people using those services. In contrast, our approach focuses on the analysis of stations' behavior instead of people's. Other investigations develop multiple bicycle sharing systems [OMS18, OCB14, MDO\*20], but they are restricted to the analysis of aggregated data. Hence, insights on a single system to the detail required by advanced users cannot be extracted from their work. Several studies propose interactive visualization systems. [OST\*16] explores a BSS behavior with multiple views emphasizing the balance of both weekdays and weekends. [DTL20] facilitates the exploration of what they call patterns of usage which can be extracted using a combination of spatial, temporal and destination dimensions, and clustering. Therefore, even when we share with these studies their focus on the availability of bicycles and clustering, we cover other purposes, such as the prediction of station behavior, as well as understanding long outages or resizing needs.

### 3 ANALYSIS OF REQUIREMENTS

After the analysis of several BSS, we identified some limitations on how the information is provided (e.g. clarity on the number of available bikes), usability problems, or lack of details (prediction of future availability). Given that the reported problems and digital maturity largely vary among public BSS, we have concentrated on Barcelona's BSS named Bicing. We performed a number of semi-structured interviews with several regular and advanced users to get more details of their needs. We were especially concerned with issues that can be addressed using the publicly available data. As a result, we identified the following requirements that are typically provided by official applications:

- R1:** Clear availability of bicycles and free slots.
- R2:** Distinction between mechanical/electrical bikes.
- R3:** Closest stations to a certain position.

A surprising factor found is the importance of clearly distinguish between the availability of mechanical and electrical bikes. For instance, some regular users reported that if an electrical bike is not available they will choose an alternative form of transportation (electrical motorbike, underground, ...). In addition, regular users also have some needs not fulfilled by the current systems:

**RU1:** Prediction of bicycles or slots availability in the next one or two hours.

**RU2:** Prioritized list of stations closest to a point.

**RU3:** Distance and slope to a certain station.

**RU4:** Usage ratio: number of bicycles that are picked up or dropped off in a certain time interval.

In fact, real-time information of docking stations' state is useful, but predictions are also necessary for planning trips in advance, especially short term. Furthermore, information regarding the three or four closest stations to the one of interest together with the distance and slope needed to reach them is also a must for all users. Additionally, for stations with high usage, one user also suggested having some flow ratio that helps to understand whether bicycles might be available even if a certain station is currently empty.

On the other hand, advanced users highlighted the need to address problems such as re-balancing the docking stations, the overall flows of stations along the day and the week, as well as the proximity of other means of transportation systems; adding to our list of requirements the following ones:

**RP1:** Daily docking station usage

**RP2:** Usage patterns for multiple stations.

**RP3:** Visual exploration of stations' behavior around a certain position.

**RP4:** Stations with a higher degree of outages (might require increase/decrease their capacity).

Fulfilling these needs may facilitate the understanding of how the system works as a whole and helps advanced users to better plan BSS optimization.

To address all the collected requirements, we have designed a multiple-view visualization system and several features intended to help both regular and advanced users with the tasks that are more likely going to be useful for them.

### 4 BSS EXPLORATORY ANALYSIS

Our application offers two different modes: *User* designed for regular users and *Planning* with features useful for advanced users; both are based on the requirements presented in section 3. Although currently, we do not restrict the views to access, the idea is to present as initial view, the one corresponding to the type of use expected. In both cases, the main screen consists of multiple coordinated views, as described next.

## 4.1 User mode

The user-mode has been tailored to solve regular users common tasks which include: *i*) Current status of docking stations, *ii*) prediction of future availability of bikes and free slots, *iii*) closest stations to a point of interest or position, and *iv*) information regarding the frequency of stations' usage. As in many other applications, we build the view around the map of the city, where we show the availability of bicycles, as shown in Figure 1.

Users can access the availability of free bicycles or slots in two different ways: *i*) Clicking on a docking station in the map, and *ii*) Selecting a point of interest in the left menu. The outcome of any of those interactions is a set of views that provides the following information to the user: current availability and prediction, prioritized list of closest stations, and usage ratio chart.

The availability information solves requirements **R1** and **R2**. Moreover, since this can be triggered using a nearby docking station position in the map through the points of interest, it also solves **R3**. The prediction widget is a stacked bar chart that encodes the number of available bikes and free slots for the following two hours, solving **RU1**. Given that the available bicycles may be scarce (and bike reservations are not visible through the official applications), we also provide a list of the closest stations. The list is prioritized based on two parameters: distance and altitude. This facilitates the user to make informed decisions (such as not going uphill if sweating could be an issue) and solves **RU2** and **RU3**. Finally, the usage ratio is presented as a line plot to help users decide whether bicycles might be available even if the docking station is currently empty (fulfilling requirement **RU4**). This is reported as a quite common situation typically around large transportation hubs. In order to solve the most common tasks, we have designed a set of interactions in the different views described next, based on the view that triggers them.

**Control Panel.** The user view can be configured by using different tools on the left panel. By default, we render the bicycles' availability on the map (color-coded), but the user can swap it to show free slots. Second, the user can change the day of the month and the hour and make bus, metro, and train stations visible by toggling them on. A widget that gives the altitude of the docking stations can also be used to filter out stations. For users that may not be completely familiar with the city, we also enable an option to show points of interest such as the main transportation hubs and some special highly touristic zones such as Sagrada Família or Park Güell. By selecting one of those points the closest stations are highlighted and their predicted availability is shown.

**Map navigation.** The map provides the usual navigation features such as drag and zoom. Hovering over one station on the map will highlight it, together with the closest ones by increasing their size. Moreover, the

highlighted station displays a popup that indicates its ID and current availability data (see Figure 2). Similarly, the elevation chart, located in the control panel, emphasizes the altitude of the pointed station. Selection is carried out by clicking. Upon selection, the rightmost view is populated with detailed information of bicycle availability for the following two hours (30 minutes intervals) for the selected station and the closest ones and the usage ratio chart. Figure 3 illustrates this.

**Prediction panel.** As discussed in section 3, decisions can be made taking into account three factors: *i*) availability of mechanical bicycles, *ii*) availability of electrical bikes and *iii*) free slots. Therefore, this view contains color-coded information to depict the number of elements in each category for the selected point and the four closest stations that have been prioritized according to distance and slope. Hovering over each of those categories in any of the bars highlights it in the other bar charts to facilitate quicker inspection of the data, as shown in Figure 4. A line chart indicating the activity is also displayed. This helps the user to distinguish between stations that may be empty but without activity and others that might be currently empty but have a high amount of activity, thus resulting in a potential bike available in the following minutes.

**Getting insights.** This set of features secures that users get relevant information from the BSS to make more informed decisions. For example, one can get an idea of how difficult is to get a bicycle near a certain metro or bus stop by activating the transportation layers and changing the day and hour. By filtering stations using the elevation chart and checking the activity ratio and the predicted availability, we can also see how the stations with higher elevation get empty soon and then keep low activity throughout the day. More complex cases are analyzed in Section 6.

## 4.2 Planning mode

The planning view provides more information related to groups of stations, so the advanced users can analyze their behavior collectively, instead of exploring the stations one by one. To this end, we implement useful features such as clustering based on the pickups and drop-offs (**RP2**), heatmaps that show the status of all the stations along the day (**RP4**), sliders to explore the behavior of stations per time frame (**RP1**), among others. Figure 5 shows this interface. As discussed in section 3, advanced users have many more needs than a regular user, so it is necessary the development of a larger set of higher-level features. For instance, the detection of stations with long outages can help system operators to better plan re-balancing operations, determine which regions of the city may require a new docking station, or which stations need to be resized. Besides, urban planners can complement zones with a low

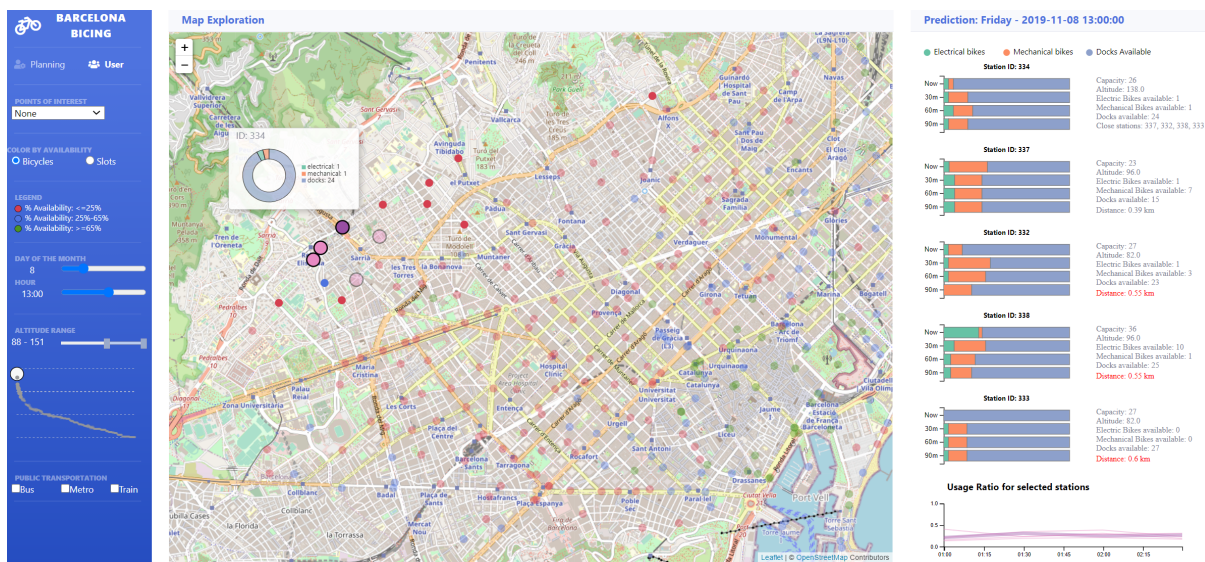


Figure 1: Interface for regular users. The user has selected the stations that are above a certain elevation on a Friday at 13:00. The map shows the availability of bicycles in stations (red colors indicate low availability). The right view displays the prediction for the next two hours. After selecting the one she is interested in, the rightmost view shows the closest stations prioritized by difficulty to access (altitude + distance) and the prediction shows that only one e-bike might arrive. Moreover, their activity is very low (bottom line chart).



Figure 2: By hovering over the map, stations are highlighted and a donut chart glyph with the ID of the station and its availability is shown.

number of stations or large outage periods with other transportation means, such as bus stops. To solve their most common requirements, this view is divided into five components: Control panel, interactive map (RP3), status, clusters, and details view.

**Clusters.** After analyzing different strategies (e.g. pickups vs availability), both the pickups and drop-offs seem to produce an understandable classification of docking stations. We also separate the clustering for weekdays and weekends, since weekends tend to have a much smaller amount of activity and completely different patterns, as shown earlier [OMS18]. More details are given in section 5. The result for weekdays is a set of 5 clusters that can be described as:

**Cluster 1:** Peak of pickups in the 8-10 a.m. range and medium usage along the day. In general, low availability during the day.

**Cluster 2:** Peak of pickups in the 8-10 a.m. range and strong intensity of use during the afternoons, especially arrivals. Relatively low availability the central hours of the day.

**Cluster 3:** Low use and availability during the day.

**Cluster 4:** A high number of arrivals from 7 a.m. and intense use along the day, no clear availability pattern.

**Cluster 5:** Peak of drop-offs in the 7-9 a.m. range and medium activity with more departures than arrivals for the rest of the day, especially in the afternoon. They offer a healthy availability balance in the central hours of the day.

The identified behavior, relatively sharp peaks in clusters early in the morning but more distributed activity along the evening, and some usage peaks around 7-10 p.m., seems to align with the fact that working times in Barcelona start early in many jobs, but the end of the working journey is less defined, as found in other reports [ftIoLC\*10].

**Control panel.** By default, the planning view displays weekdays data clustered and the cluster panel is showing the arrivals. Likewise, stations on the first cluster are selected and colored in the map panel using cluster information, and the stations with outages are emphasized. However, users can swap between different views with the tools on the left panel to show the information that suits their interest at most. As before, the

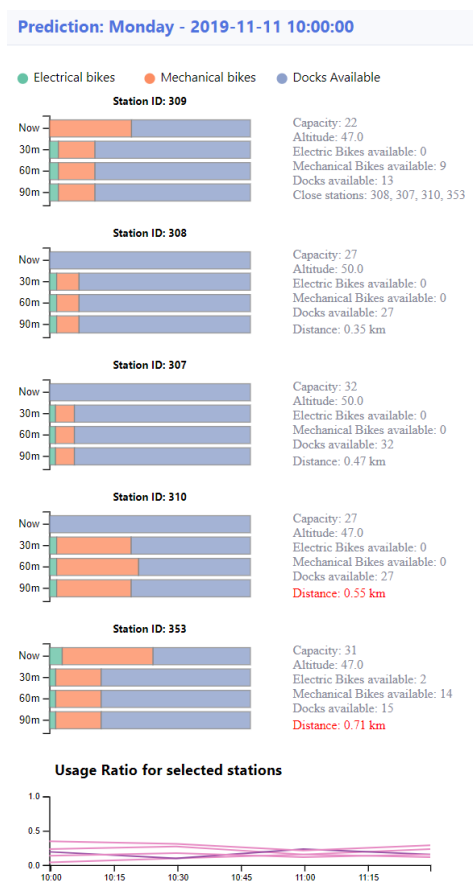


Figure 3: The prediction view provides the current status of the selected station and the closest ones, and the availability prediction for the following 2 hours in 30 minutes intervals. Finally, the bottom activity chart indicates whether the station is highly active or not.

day, hour, and transportation hubs can be changed, and stations can be filtered by their altitude, but there is no option to display the points of interest.

**Map Navigation.** As in user view, it is possible to drag and zoom. Using the control panel, it is easy to view the information for one cluster or all the stations and change the station's color by their cluster or by their availability. Hovering over one station will highlight it together with the closest stations, and will show a popup with the current availability. Station selection is carried out by clicking. This operation also emphasizes the pointed station in the elevation chart. The selected station and the closest ones are also highlighted in the status and cluster panel and the information panel is updated with the predicted availability for the next two hours at 30 min intervals and the usage ratio for the selected station.

**Status panel.** It is composed of three heatmaps that show arrivals, departures, and availability of all stations in the selected cluster along the day to facilitate the exploration per time frame and easily identify patterns.

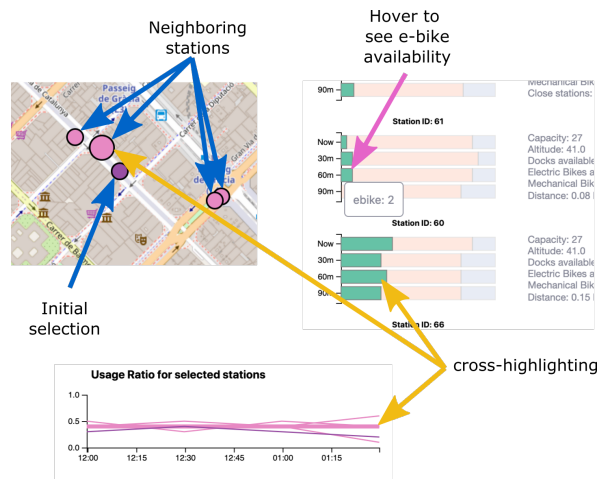


Figure 4: Exploring the status and prediction of a docking station and its neighborhood. After selecting a station on the map, we can hover over the prediction charts to see which ones will have electrical bikes, and the activity of the station is also highlighted, together with its location on the map.

Stations in a cluster are sorted according to the similarity between them, taking as an initial reference the medoid. Station and hour selection is done by hovering, this action also shows a popup with the percentage value of arrivals, departures, or availability and highlights the closest stations. The pointed station and the closest are emphasized in the elevation chart and the cluster and map panels. The control panel can be used to highlight the stations with extremely large outages.

**Clusters behavior.** It shows the information for all stations divided by clusters along the day and is useful to check different patterns between them. By default, the information of arrivals is displayed but it can be changed for departures or availability using the control panel. When a cluster is selected, the color of its corresponding plot changes to easily identify it. If users hover over a station, this and the closest ones are emphasized, even if they belong to a different cluster. This action also highlights the same stations in the map, status, and elevation chart.

**Information panel.** This panel is divided into three sections. The first one contains the predicted availability, for the selected station in the following two hours divided by electrical, mechanical bikes, and slots availability. The second part corresponds to a popup with information regarding its capacity, altitude, and current status. The last section shows the usage ratio chart.

**Getting insights.** The above-mentioned features may reveal interesting patterns that would be difficult to achieve with regular apps. For example, we can infer that rebalancing operations happen when we see sudden availability changes at night (or followed by very low activity ratios). Line charts can also reveal which dock-



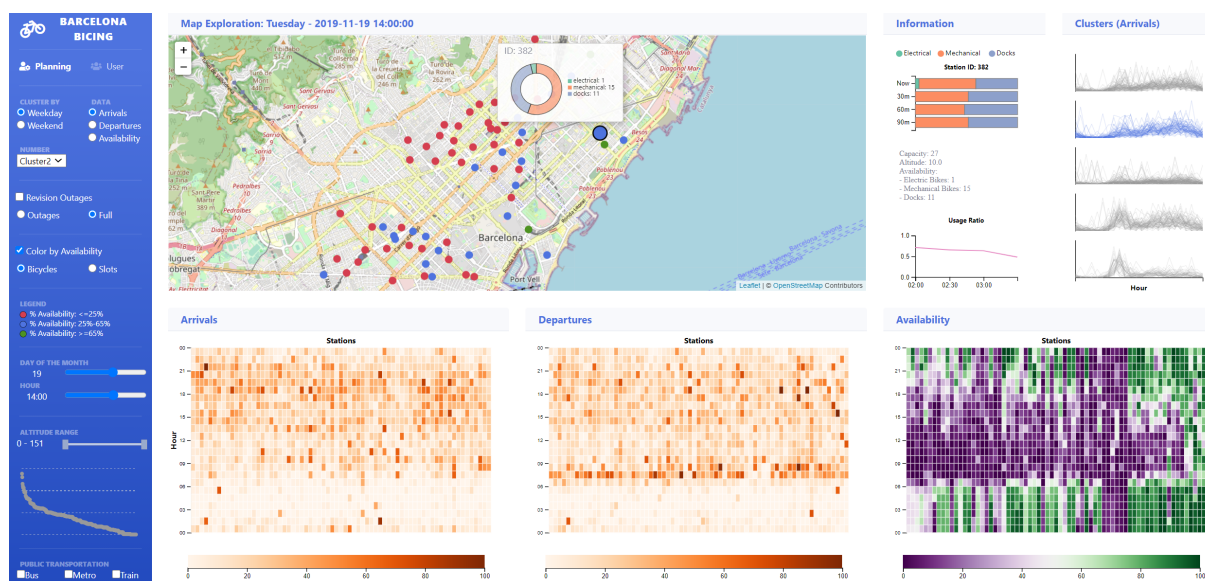


Figure 5: The planning interface. The map may encode the different clusters or the availability, together with information of potential outages. The heatmaps show different aspects of the usage: arrivals, departures, and availability. On the right, the view displays the behavior of each cluster, using the same usage state defined in the left panel. In this case, she has selected cluster 2, which groups several stations in the center of Barcelona. By hovering over them, we can see that they keep a high degree of activity along the central hours of the day, as may be inferred from the bottom heatmaps.

ing stations belong to mostly residential zones (e.g. high number of pickups in the morning and evening and very low activity in the middle of the day).

## 5 IMPLEMENTATION

**Interface.** The interface is built using HTML and it's composed of two different tabs: *PlanningHome* and *UserHome*. The code is built using Flask for managing HTTP calls and rendering templates. Views are implemented using D3, and Leaflet library for the maps. Data management is carried out in Python, using *pandas* and *numpy*. The forecasting and initial data processing, which includes cleansing, clustering, and data derivation algorithms are performed using R.

**Data gathering and processing.** The original samples are updated approximately every 5 minutes having more than 3.5 million records every month. Considering that 2020 data had an abnormal behavior due to the lockdown period and several mobility restrictions, we use data from 2019 to train and evaluate the models and deploy the use cases. Moreover, the Bicing provider company has changed during 2019 which means there is no continuity in the information during the first months since the change occurred gradually. Therefore we focus on the last months of 2019 from September to November.

The Barcelona BSS has 424 docking stations. We first clean the data to avoid including information of docking stations that could appear in the set but are still not used, a special field in the set denotes it, as well as

others that were not working properly. The final result is a set of 409 stations. Also, garbled information is cleaned such as dates outside the analyzed range or negative data availability. From this data, we calculate additional information necessary to deploy the visual interface: day of the week (and month), the number of total docks available in each station, and we summarize data about the number of pickups and drop-offs (absolute and relative). Additionally, the Bicing usage data is combined with other sources that include information such as the GPS position of the docking stations as well as their altitude. We also add other public data such as the location of all metro, bus, and train stations and some information regarding the large transportation hubs in Barcelona city, such as Barcelona-Sants or Plaça Catalunya.

After the data is cleaned, we apply different algorithms to derive the data that is later used in our visualization. These includes: *i*) Clustering docking stations, *ii*) closest stations calculation for all stations, *iii*) calculation of usage ratios, *iv*) calculation of outages, and *v*) prediction of usage per station.

**Clustering.** It can be addressed in different ways considering several aspects. First, data partitioning can be done for different time intervals (months, weeks, days, hours...). In our case, after some experimental evidence, we find that the behavior changes depending on whether the day is a weekday or a weekend [OMS18]. Second, we can use data of pickups, drop-offs, availability, or some ratio of usage. After testing different

combinations, we discover that data from pickups and drop-offs give us better results clustering the stations to discover their behavior. There are many clustering algorithms [RCC\*19] that may be suitable for different tasks. Key to the clustering algorithm is the selection of the distance measure, and we have many options such as the Euclidean, correlation, temporal correlation, or dynamic time warping. Depending on the measure selected, our similarity measure will be more or less resistant to noise, and to time shifts [SGP16]. After testing some of the most widely used algorithms (PAM, K-means, DBSCAN, AGNES), we find that both Partitional Around Medoids (PAM or k-medoid) [KAU87] and K-means [M\*67] outperform both internal and stability results and are suitable for our problem. PAM method has the objective of minimizing the average dissimilarity around the nearest medoid (always an element of the original dataset) which is recomputed in each iteration. K-means works in the same way, but groups the elements of a cluster around the nearest mean. Using the PAM method (which gives better results for the silhouette coefficient), we build two clusterization sets: one based on the activity score during weekdays and the other for the weekends. A similarity matrix is defined based on Euclidean distance, and the silhouette analysis is executed to test the degree of separation between the clusters.

**Closest stations.** When a docking station is empty and the user needs to take a bike, or when it is full and the user needs to drop off one, BSS do not provide information on the closest ones. To build this feature, we need to have some information on what are the closest stations. In our case, this is precalculated and attached to each element. Computation is pretty simple and has a quadratic cost. For each docking station, the four closest are found using the earth distance metric from *fossil* package in R and ranked using their distance to the original point and their altitude.

**Usage ratios.** Usage ratios are derived from the frequency of pickups and drop-offs. This ratio helps us to understand if a docking station status corresponds to a high dynamic behavior or a static behavior. It is calculated as follows:

$$use_{ijk} = \sum arrivals_{ijk} + \sum departures_{ijk}$$

Where  $i = StationID$ ,  $j = day$ ,  $k = hour$

To better interpret this ratio, we re-scale the value to make all the elements lie between 0 and 1, having a common scale:

$$UsageRatio_{ijk} = \frac{(use_{ijk} - MinUse)}{MaxUse - MinUse}$$

Where  $i = StationID$ ,  $j = day$ ,  $k = hour$

**Outages ratios.** Two different ratios are calculated: one used to identify stations that have a low number of bicycles available (10% or less) most of the day and the other to label those stations that have a high number of bikes available (75% or more) during working hours. These ratios are estimated with a daily frequency dividing the numbers of hours when a station  $i$  had an outage of type  $k$  by the number of total hours (24 hours for  $k=empty$ , and 11 hours for  $k=full$ ). The generic formula used is presented as follows:

$$outageRatio_{ijk} = \frac{\sum outage_{ijk}}{Totalhours_k}$$

Where  $i = StationID$ ,  $j = day$ ,  $k = type\ of\ outage$

**Usage prediction.** Usage forecasting is extremely important for both types of users. For advanced users, it is key for re-balancing stations and studying mobility patterns while for regular users is essential to plan future trips. Several studies have been conducted to predict bike-sharing trips using different methods and data [LHP18, XJL18, WK18]. These contribute valuable insights to understand the factors that affect bike-sharing demand. We use different features to train the models, which can be classified as station-related and weather features. Station-related features are the day, time slot, number of available bikes, and station's capacity. The weather features are composed of temperature, wind, humidity, atmospheric pressure, and weather classified in four categories: sunny, cloudy, light rain, and heavy rain. As mentioned before, the provider company was changed during 2019 causing data inconsistency during this transition. To tackle this, the models are trained using data from September and October 2019 and test with November data.

We test two forecast models: Random Forest [FW17] and Prophet [TL17]. As there are mechanical and electrical bikes available in each station, we run both models for each station and predict both mechanical and electrical bikes. Besides the predicted variable (availability) is transformed using the natural logarithm but results are not improved. Different measures are used to compare both models, but we keep RMSE since others (e.g. MAE, MAPE) have a very strong correlation with it and do not add valuable insights. Random Forest is the model which gives better results with lower RMSE levels for both scenarios especially for electrical bikes forecasting using non-transformed data. It is optimized in terms of the number of trees (500 trees). Additionally, we create a different error measure considering the scenarios that affect the users: no bicycles available or no slots to return bikes. The stations are divided into three categories: *i*) Available bikes less than 10% *ii*) Available slots less than 10% *iii*) Healthy status (bikes between 11-89%). Finally, we evaluate how many times our model predicts a different status. The

results are promising, for mechanical bikes only 33% of times the status is different while for electrical bikes the error is less than 6%. These results suggest the model applied has reasonably good prediction accuracy. However, it can be improved as systems mature and data is collected across a longer period.

## 6 USE CASES

To illustrate the use of the application, we design two different use cases that correspond to the two views designed: one for a regular user-oriented profile, and another for a planning-oriented profile.

### 6.1 Availability and free slots prediction

The first use case consists of a multi-query situation: the user is leaving from a meeting and she wants to know where is the closest docking station and whether it has electrical bikes because she is going to an elevated part of the city. Then, she wants to check the availability of free slots near her destination in one hour or later.

To solve the problem, she has to look for a docking station using the map. By hovering with the mouse, the visual marks indicating the closest stations scale up and are emphasized with an outline. Given that they are colored according to availability, it is easy to infer whether there are free bikes. If she is close to one of the most visited areas in the city, she can also use the left panel to select the closest stations to that point. By clicking on the station of interest, a list of the closest stations will appear on the right panel. Each station shows a horizontal bar chart where the first bar (top) is the current availability and the others are the prediction for 30 minutes slots. Electrical bikes are encoded in green, thus, it is easy to identify their availability in any of the neighboring stations. By hovering over the green bar, the concrete number can be read. Additionally, the usage ratio chart at the bottom is also highlighted for the currently selected station. Therefore, if the availability is low, she can have extra information on whether it is due to the lack of movement or the high throughput by checking it. Similarly, to determine the best place to leave the bicycle she can use the map. By hovering over a docking station near her destination, she will be able to see whether there are free slots. If there are no slots available, by clicking on it, she can see a prediction of future available slots in the selected station as well as in neighboring docking stations. The list of docking stations specifies the slope and the distance to cover. This way, she can make more educated decisions on which stations are adequate for her needs.

### 6.2 Analyzing stations with severe outages

In this case, an urban planner wants to analyze docking stations that suffer outages. If that is the case, she will be interested in knowing where, and how other closest

stations behave in the same time intervals. She will use a deductive method thanks to the characteristics provided by the application.

By selecting the planning view and the outages option, the stations that have outages or are full an important part of the day are highlighted on the availability heatmap. To increase clarity, she may set the color of the stations on the map to encode the availability instead of the cluster color, using the left control panel. By hovering over the availability heatmap the station under the mouse and the closest stations are emphasized to further investigate. For example, she sees a group of 3 stations that have outages, but one of them (number 3) behaves slightly differently along the day, as shown in Figure 6-left. She may hypothesize that it might be related to the presence of public transportation. Effectively, if she toggles metro stations on, there is a metro station, La Llacuna, which is closer to the docking stations that have severe outages (see Figure 6-right), while station number 3 is slightly farther. If the bus stops layer is activated, she sees that station 3 lies next to two bus stations. It turns out that those are precisely two stations belonging to two different (and orthogonal) lines that have highly frequent buses (the types H and V, which travel horizontally and vertically). Given that the same transportation ticket can be used (within one hour and a half) to take a bus, those passengers might not need to change to a bicycle because they can take a bus instead.

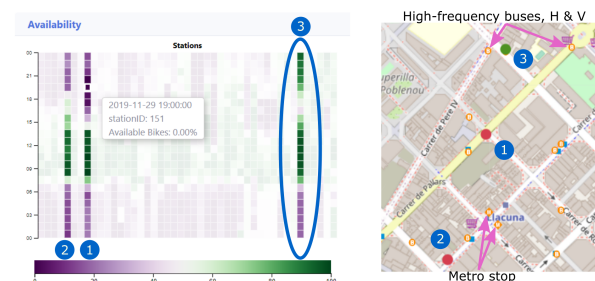


Figure 6: Inspecting the behavior of four neighboring docking stations. Station 3 does not suffer outages for so long in the afternoon. By activating the transportation layer, we find that 1 and 2 are next to a metro station. Moreover, we can also see that 3, is next to two bus stations that have highly frequent buses (H and V type) which travel in orthogonal directions, and can be used with the same transportation ticket.

### 6.3 Evaluation

We evaluated our application in two different ways: *i*) an informal user study with a few subscribers of the Barcelona Bicing system, and *ii*) interviews with domain experts in urban planning, system operation, and policy-making.

For the user study, we had an interactive remote session (due to the limitations in mobility and social contacts



during the pandemic) where we demonstrated the features of the application. This was followed by an interview where we discussed the possibilities and limitations of the system. Then, we asked them to complete a standard questionnaire offline. The questionnaire consisted of a series of questions that should be evaluated using a 1-7 Likert scale where 1 means totally disagree and 7 means totally agree. Questions were related to their experience with the Bicing system application and how the features of our application could benefit them. The average of obtained results are quite positive:

Question	Grade
It is useful to better plan usage	6.7
I would use it to check station status	6.7
I would use it to check future availability	5.7
Usage prediction would help me decide	7.0
Importance: stations altitude & closeness	6.0
Utility: check availability in relevant areas	6.0
Utility: prioritized list of closest stations	6.7
Information is relevant for a Bicing user	7.0
Ease of use	6.0
More useful than current website or app	6.3

Table 1: Results of users' interviews

We interviewed three domain experts, an architect with experience in urban planning, a politician with responsibilities in policy-making in the area of Barcelona, and a technician that supervises the bike-sharing system of a small city. We asked them to complete a questionnaire regarding potential features of the system. These advanced users agreed that being able to see the usage patterns is strongly useful to better understand citizens' behavior. Some features were highly rated such as the clustering based on stations' behavior and their representation using heatmaps (especially useful for the creation of mobility policies). They also considered the availability heatmap more relevant to monitor the system as a whole. Bikes' availability, altitude, and distance to the neighboring stations were quite interesting for re-balancing purposes. Detailed information on times of high usage of docking stations and availability prediction was found very useful for transportation planning and policy-making. Stations that are most of the time full or empty were considered extremely important for mobility planning. After the system was almost finished, we showed an interactive demo to a technician that supervises the BSS in his city, and he found that the usage patterns shown by the application, and no present in others, are highly useful to make decisions such as changing the placement of a docking station, improving the public transportation in certain locations, getting regular reports on the status and usage of the system, and monitoring the real usage (beyond to what the system operator company actually reports). All of them are aspects that they cannot analyze with their current means. As seen, both regular and ad-

vanced users were enthusiastic with the application, and have provided very important feedback for future improvements.

## 7 CONCLUSIONS

We implement a full web-based visualization system to explore and analyze public BSS. Barcelona Bicing system is used as our first scenario. Compared to current alternatives, we provide much-desired features for regular users such as availability prediction and nearest docking stations prioritized by distance and altitude. For advanced users, we also show station patterns along the day, emphasized stations with long outages, and cluster the stations based on their behavior.

However, the information we used is constrained to the one that is available through the Barcelona Open Data website, and unfortunately, there is only historical data available through this platform. The real potential of this application can be leveraged with real-time data that would require access to the original data sources and permanent computer resources. Due to pandemic and temporal limitations, the sample taken to evaluate the system has been reduced and a larger quantitative study would be needed to characterize it with greater precision. Additionally, the interaction with the system operator was challenging due to the change of operator and the pandemic.

There are several development lines we are considering. Regarding the prediction, we have already tested Random Forest and Prophet models but we plan to evaluate other algorithms' performance. The used data is limited in time due to the change in the operator and the effect of the 2020 lockdown. So, we plan to perform a deeper study on the most suitable prediction algorithm whenever the system is not interrupted again and recovers normal usage patterns. In addition, the current databases do not contain detailed information regarding the trips, that is, the concrete pickup and drop-off docking stations of each bike. This would open other possibilities to perform more advanced analysis. We have had contact with the Barcelona Open Data director to gain access to this information. Finally, another possibility would be to establish relationships between BSS and other transportation systems, for instance, information of travellers' arriving and leaving to Barcelona via train stations.

## 8 ACKNOWLEDGMENTS

Partially supported by project TIN2017-88515-C2-1-R(GEN3DLIVE), from the Spanish Ministerio de Economía y Competitividad, by 839 FEDER (EU) funds. The authors would like to thank Francisco Ridruejo and Ignacio Pascual for their invaluable contributions and feedback, as well as the users who participated in our user study.

## 9 REFERENCES

- [BAF\*11] Borgnat P., Abry P., Flandrin P., Robardet C., Rouquier J.-B., Fleury E., Shared bicycles in a city: A signal processing and data analysis perspective. *Advances in Complex Systems* 14, 03, 2011, 415–438.
- [BWB14] Beecham R., Wood J., Bowerman A., Studying commuting behaviours using collaborative visual analytics. *Computers, Environment and Urban Systems* 47, 2014, 5–15.
- [DTL20] Dai H., Tao Y., Lin H., Visual analytics of urban transportation from a bike-sharing and taxi perspective. *Journal of Visualization* 23, 6, 2020, 1053–1070.
- [FIE15] Faghih-Imani A., Eluru N., Analysing bicycle-sharing system user destination choice preferences: Chicago’s divvy system. *Journal of transport geography* 44, 2015, 53–64.
- [FR14] Frade I., Ribeiro A., Bicycle sharing systems demand. *Procedia-Social and Behavioral Sciences* 111, 2014, 518–527.
- [ftloLC\*10] European Foundation for the Improvement of Living Conditions, Comparative analysis of working time in the European union, 2010.
- [FW17] Feng Y., Wang S., A forecast for bicycle rental demand based on random forests and multiple linear regression. In 16th IEEE/ACIS International Conference on Computer and Information Science, May 24–26, 2017, Zhu G., Yao S., Cui X., Xu S., (Eds.), IEEE Computer Society, pp. 101–105.
- [HMO18] Holmgren J., Moltubakk G., O’Neill J., Regression-based evaluation of bicycle flow trend estimates. *Procedia computer science* 130, 2018, 518–525.
- [KAU87] KAUFMANN L., Clustering by means of medoids. *Proc. Statistical Data Analysis Based on the L1 Norm Conference*, Neuchatel, 1987, 405–416.
- [KPL] Kim I., Pelechris K., Lee A. J., The anatomy of the daily usage of bike sharing systems: Elevation, distance and seasonality, *ACM SIGKDD workshop on Urban Computing*, 2020.
- [LHP18] LIN L., HE Z., PEETA S., Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach. *Transportation Research Part C: Emerging Technologies* 97 (Dec 2018), 258–276.
- [M\*67] MacQueen J., et al., Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, 1967, vol. 1, Oakland, CA, USA, pp. 281–297.
- [MDO\*20] Meddin R., DeMaio P., O’Brien O., Rabello R., Yu C., Gupta R., Seamon J., The meddin bike-sharing world map, 2020,
- [OCB14] O’Brien O., Cheshire J., Batty M., Mining bicycle sharing data for generating insights into sustainable transport systems. *Journal of Transport Geography* 34, 2014, 262 – 273.
- [OMS18] Oppermann M., Möller T., Sedlmair M., Bike sharing atlas: visual analysis of bike-sharing networks. *Int. J. Transp* 6, 2018, 1–14.
- [OST\*16] Oliveira G. N., Sotomayor J. L., Torchelsen R. P., Silva C. T., Comba J. L., Visual analysis of bike-sharing systems. *Computers & Graphics* 60, 2016, 119–129.
- [RCC\*19] Rodriguez M. Z., Comin C. H., Casanova D., Bruno O. M., Amancio D. R., Costa L. d. F., Rodrigues F. A., Clustering algorithms: A comparative approach. *PloS one* 14, 1, 2019, e0210236.
- [Ric15] Ricci M., Bike sharing: A review of evidence on impacts and processes of implementation and operation. *Research in Transportation Business & Management* 15, 2015, 28–38.
- [SGP16] Saas A., Guitart A., Perriñez A., Discovering playing patterns: Time series clustering of free-to-play game data. In 2016 IEEE Conference on Computational Intelligence and Games (CIG), 2016, IEEE, pp. 1–8.
- [TL17] Taylor S., Letham B., Forecasting at scale. *The American Statistician* 72, Sep, 2017.
- [WK18] WANG B., KIM I., Short-term prediction for bike-sharing service using machine learning. *Transportation Research Procedia* 34(Jan. 2018), 171–178. *International Symposium of Transport Simulation and International Workshop on Traffic Data Collection and its Standardization 2018, ISTS IWTDCS 2018*;
- [WBD14] Wood J., Beecham R., Dykes J., Moving beyond sequential design: Reflections on a rich multi-channel approach to data visualization. *IEEE transactions on visualization and computer graphics* 20, 12, 2014, 2171–2180.
- [XJL18] XU C., JI J., LIU P., The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets. *Transportation Research Part C: Emerging Technologies* 95 (2018), 47–60.
- [YZWB20] Younes H., Zou Z., Wu J., Baiocchi G., Comparing the temporal determinants of dockless scooter-share and station-based bike-share in washington, dc. *Transportation Research Part A: Policy and Practice* 134, 2020, 308–320.