Analysis and proposed data-driven enhancements for the Valenbisi bike network

Anna Kapitánová[1] and Pablo Díaz-Masa[2]

[1]Department of Artificial Intelligence, Czech Technical University in Prague.

[2]Department of Data Science, Universitat Politècnica de Valéncia.

Contributing authors: kapitann@fit.cvut.cz; pdiamas@etsinf.upv.es;

Abstract

This study addresses the growing demand for environmentally friendly urban transportation, focusing on Valencia's Valenbisi bike rental company. While the company offers a network of 2,700 bikes and 275 stations, it faces challenges such as uneven bike availability and station capacity, particularly in different parts of the city and during peak times. To address these issues, the study proposes an analysis of historical data and the creation of visualizations to identify problem areas and propose solutions. Factors such as weather conditions, seasonal events like the Fallas and holidays, and the influx of tourists during the summer are also examined to understand their impact on the network's performance and user experience. The study aims to provide actionable insights to improve the quality and usability of Valencia's bike rental network.

Keywords: Smart Mobility, Bike rentals, Bike Rental Analyses, Bike Sharing, Weather influence

1 Introduction

Nowadays there is an increasing demand for more environmentally friendly ways of transportation in the cities. People are more open to options like public transportation or sharing travel. Smart cities like Valencia in Spain are facilitating citizens with the possibility of using bikes by creating a dense bike lines network. This situation creates a need for bike rental companies.

In our work, we will focus on the Valenbisi bike rental company in Valencia. Valenbisi is a company run by the city and it is a great option if you want to move by bike. The network contains about 2,700 bikes and 275 different stations around the city.

However, there are several issues the company is dealing with, which could significantly impair the user experience. To be more specific, the number of free bikes you may borrow or empty spaces, where you can return a bike, highly depends on the daytime and seasonality. Different parts of the city are dealing with different problems. For example, it is usually a challenging task if you want to borrow a bike in the city center in the evening or at night. On the other hand, the possibility of returning a bike near a beach promenade during the summer season is nearly impossible.

We want to explore the situation in more depth, analyze the historical data, and based on that create some valuable visualizations that will present clearly what the main issues of different stations are and how particular parts of the city differentiate. We will give specific instructions on how feasible measures can be taken to enhance the quality and user experience of the network.

Additionally, we aim to examine the impact of various factors such as weather conditions, notably during the city holiday Fallas in spring, and other festivities like Easter or Christmas, on the city's dynamics. We also intend to delve deeper into the effects of the summer season, which attracts numerous tourists to Valencia.

2 Literature review

We have read about different features the authors decided to use. Some of them for example included weather conditions (humidity, wind speed, and temperature, among others), seasonality, different parts of the year, and how it influences the demand. Others were exploring how people are satisfied with the system based on surveys or by showing an increasing number of new users each year.

It was also interesting to read about different approaches for forecasting and the different models they used, e.g. Random Forest, Linear, and Ridge regression,... And different evaluation metrics they used like RMSE and RMSLE scores.

2.1 Previous Work

2.1.1 Bike rental demand forecast using ML techniques (July 2022)

The authors of this paper [1] focused on bike rental demand forecasting in Seoul's bicycle rental network. Besides the occupancy of the parking stations they also concentrated on how the weather conditions influence the demand. Specifically, they examined humidity, wind speed, and temperature in every one, two, and three-hour periods. Moreover, they tried to inspect how exactly the situation changes depending on working and non-working days and they found a correlation with the bike demand.

2.2 Related Work

2.2.1 Investigating city bike rental usage and wet-bulb globe temperature (November 2021)

Researchers of this work [2] investigated the bike rental companies in the USA. Based on data that were taken between the years 2000 and 2010 they found that bicycle activity increased by 60 %. They pondered whether there is a relationship between the increasing world global temperature and shared bicycle activity in New York City and San Francisco. However, they discovered that many bike renters did not change their behaviors according to heat advisories. Furthermore, bike rentals peaked during commuting rush hours, so heat warning messages should be timed to these periods.

2.2.2 Bike-Sharing as an Element of Integrated Urban Transport System (July 2017)

The academics that were working on this research [3] chose a different approach to examine, they focused on how much people use bike rental services in Poland and the Poles' attitude towards the use of the bike as an alternative means of transport in the city. They also pointed out the principles of how the bike-sharing systems work in the cities of Katowice and Cracow and they presented the idea of self-service public bike rental.

2.2.3 A review on bike-sharing: The factors affecting bike-sharing demand (March 2020)

The authors in this article [4] offer a lucid perspective on factors that affect the usage of bike-sharing systems. They study how weather conditions (such as temperature, precipitation, and humidity) and infrastructure (such as public transportation and land use) impact on the overall experience of users and the demand of bike-sharing systems. Some of the concepts proposed in this article are key when undertaking a study of this nature.

3 Framework

The Valenbisi network is a bike-sharing system comprising stations scattered throughout the city. Each station is equipped with racks where users can find bikes available for borrowing. The network typically spans various locations, allowing users to easily access bikes from one station and return them to another. Users can borrow a bike from any station, ride it to their destination, and return it to the nearest available rack. Figure 1 provides a visualization of the Valenbisi network.

The framework for the Valenbisi project encompasses several key aspects. In terms of needs, the project requires data on station occupancy, weather conditions, and event schedules.

For sensors, data about the station's occupancy is collected from physical sensors on each rack in the station. Meteorological stations provide weather data as is visible on figure 2.

Regarding sensor data, Valenbisi sensors provide updates on station occupancy every 15 minutes to the cloud, from where we collected the data. Meanwhile, weather sensors offer hourly updates to the cloud on temperature, precipitation, humidity, wind speed, and other meteorological variables.

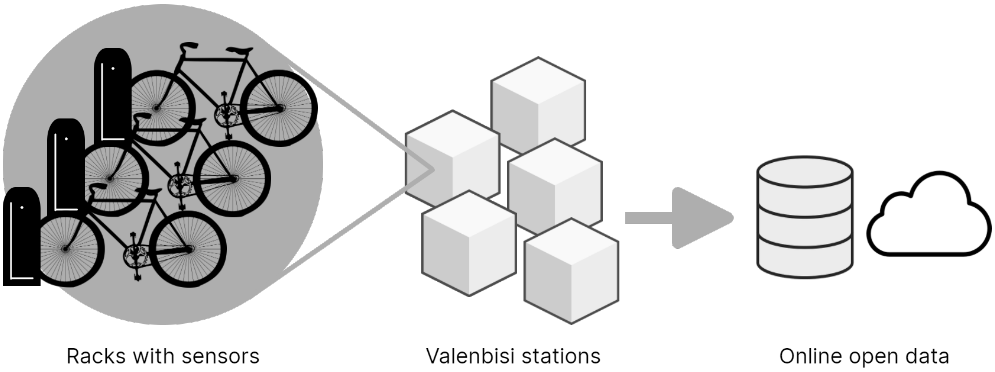

Fig. 1: Valenbisi network

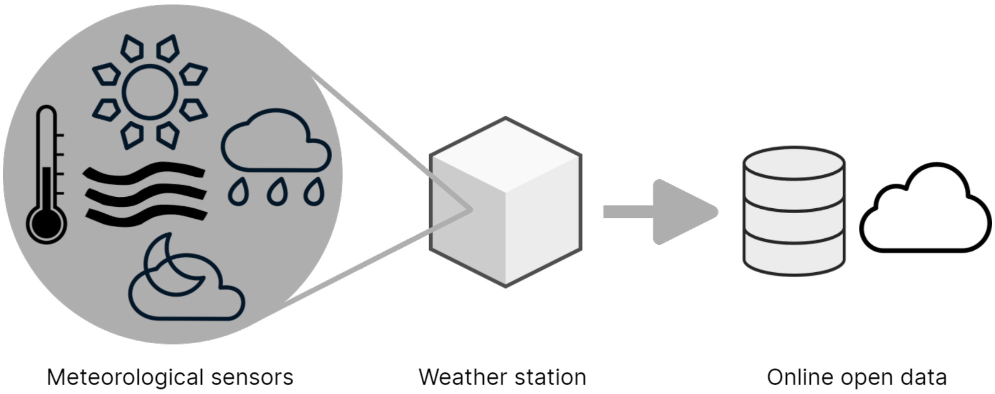

Fig. 2: Weather stations network

3.1 Data

In the project, we used three main datasets: Valenbisi Traffic Data, Weather Data and Festivities Data.

3.1.1 Valenbisi network data

We accessed real-time updates on station availability every 15 minutes from December 1st, 2022, to February 29th, 2024. This data includes the number of available bikes, free spaces, and station locations. Historical Valenbisi data from December 1st, 2022, up to the current date was obtained from Ceferra's Valenbisi GitHub repository[1] and web scraped from the Valencia City Council's open data webpage[2].

Variable name	Description	Data type	Example
id	ID of the station	int	901622
name	Name of the station	text	42_AVDA. DE LA PLATA
number	Building number in the street	int	42
address	Address of the nearest building	text	Av. de la Plata (Museo Fallero)
open	State of the station (T/F)	bool	T
available	Number of available bikes	int	18
free	Number of free spaces	int	2
total	Total number of spaces	int	20
geo_point_2d	X,Y coordinates	geo_point_2d	39.45878127329601,-0.3586583378805894

Table 1: Variables in the Valenbisi dataset

3.1.2 Weather Data

Weather conditions in Valencia were sourced from hourly updates spanning from December 1, 2022, to December 31, 2023. The dataset comprises information on temperature, humidity, sky conditions (e.g., rainy, sunny), wind speed, wind angle, percentage of clouds, and visibility. Weather data was scraped from Tiempo3[3].

Variable name	Description	Data type	Example
hour	Hour of the day	text	20:00
temp	Temperature (ºC)	float	11.1
sky	State of the sky	text	Llovizna
precip	Precipitation (mm)	float	0.1
humid	Humidity (%)	int	83
wind_speed	Speed of the wind (km/h)	float	3.7
wind_angle	Angle of the wind (º)	int	122
clouds	Clouds in the sky (%)	int	100
visib	Visibility (km)	float	7.3

Table 2: Variables in the weather dataset

[1] https://github.com/ceferra/valenbici
[2] https://valencia.opendatasoft.com/explore/dataset/valenbisi-disponibilitat-valenbisi-dsiponibilidad
[3] https://www.tiempo3.com/

3.1.3 Festivities Data

Information about days of the week, weekends, and public holidays from 2022 to 2024 was collected. This dataset aids in understanding the impact of holidays and weekends on bike-sharing patterns. We gathered these data from the annual holiday calendar publications in the Diario Oficial de la Comunidad Valenciana[4].

Variable name	**Description**	**Data type**	**Example**
day	Date	datetime	2023-17-04
dayofweek	Day of the week	text	Monday
category	Holiday/working day	text	Public holiday
motive	Motive for the holiday	text	San Vicente Ferrer

Table 3: Variables in the festivities dataset

4 Method

Following the collection of raw data, the subsequent necessity was to refine it into a final dataset. Subsequently, the dataset underwent clustering analysis.

4.1 Dataset

In this section, we are going to discuss the challenges connected to the final dataset creation, which were used to perform the analysis. The final dataset was concocted by merging the three previous datasets containing Valenbisi, weather, and festivities data.

4.1.1 Data Cleaning

We encountered several challenges during the data cleaning process, the primary one being the need to standardize the format of the Valenbisi data. Files from January 12, 2022, to May 22, 2023, were in a different format compared to those from May 22, 2023, to February 29, 2024. Although the variables' information was similar, there were substantial changes in variable names and the handling of unique IDs.

Additionally, we encountered missing files for certain days and hours within the same day, resulting in 8.57% of missing files (39 out of 455 days). This anomaly arose due to irregular updates on the official open data website or the GitHub repository from which we sourced the data. Fortunately, these missing files were uniformly distributed and did not significantly impact the quality of our analysis.

In contrast, the weather and festivities datasets exhibited minimal flaws and did not pose significant challenges during the data cleaning process.

[4]2022: https://dogv.gva.es/datos/2021/10/15/pdf/2021_10336.pdf
2023: https://dogv.gva.es/datos/2023/10/04/pdf/2023_9993.pdf
2024: https://dogv.gva.es/datos/2023/10/04/pdf/2023_9993.pdf

4.1.2 Data Preprocessing

In this section, we are going to concentrate on data preprocessing, focusing on extracting crucial information regarding station positions, occupancy, and overall usage patterns.

Latitude and Longitude of the Stations

The first step in data preprocessing involves extracting information about the location of each station. Latitude and longitude coordinates, stored in the column named geo_point_2d, are utilized for this purpose. The latitude represents the north-south position, while the longitude represents the east-west position.

Information about Stations' Occupancy and Total Number of Spaces

To gain insights into the size and utilization of each station, we calculate two key metrics: occupancy and vacancy. Occupancy refers to the percentage of racks currently in use at a station, while vacancy represents the percentage of racks that are currently available. These metrics provide valuable information about station usage patterns.

Addition of Increase or Decrease of Previous State

In order to track changes in station usage over time, we create additional features: bikes_diff and free_diff. These metrics measure the difference in the number of available bikes and free spaces, respectively, compared to the previous measurement. This allows us to monitor fluctuations in station activity and identify trends over time.

4.1.3 Final Features

In this section, we present an overview of the dataset used for further analyses. Table 4 describes each of the final features, providing essential information for analysis and modeling.

4.2 Clustering

In this section, we will discuss the clustering techniques employed to analyze the Valenbisi bike rental system in Valencia. Clustering is a fundamental unsupervised learning technique used to group similar data points together based on certain characteristics. In our analysis, we utilized two distinct approaches for clustering: one based on geographical location and the other based on the behavior of bike stations.

4.2.1 KMeans

KMeans is a popular clustering algorithm that partitions a dataset into a predetermined number of clusters, where each data point belongs to the cluster with the nearest mean (centroid). The algorithm iteratively assigns data points to the nearest centroid and recalculates the centroids until convergence [5]. How the algorithm works is visible in figure 3. By employing KMeans, we aimed to identify clusters of stations exhibiting similar usage patterns and operational dynamics.

Feature	Description
gid	The unique identifier assigned to each bike station.
available	The current count of available bikes at the station.
free	The current count of available docking spaces for bikes at the station.
total	The total number of bike docking spaces at the station.
w_temp	The temperature at the time of observation.
w_sky	A description of the sky condition (e.g., clear, cloudy).
w_precip	The amount of precipitation (in millimeters) at the time of observation.
w_humid	The relative humidity at the time of observation.
w_wind_speed	The wind speed at the time of observation.
w_clouds	The cloud coverage at the time of observation.
w_visib	The percentage of visibility at the time of observation.
dayofweek	The day of the week corresponding to the observation.
day_category	A categorization of the day (working day, weekend or holiday).
day_motive	If day_category is a holiday, day_motive is the name of the holiday, if day_category is a weekend day_motive is Saturday or Sunday.
lat	The latitude coordinate of the bike station location.
long	The longitude coordinate of the bike station location.
datetime	The date and time of the observation.
occupancy	percentage of used racks in a station
vacancy	percentage of free racks in a station
bikes_diff	how differs the number of bikes from the previous measurement
free_diff	how differs the number of free spaces from the previous measurement

Table 4: Dataset's features with their description

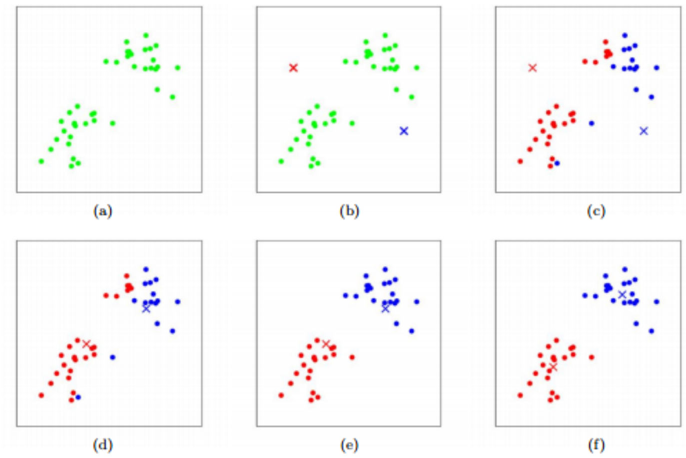

Fig. 3: KMeans algorithm. Training examples are shown as dots, and cluster centroids are shown as crosses. (a) Original dataset. (b) Random initial cluster centroids. (c-f) Illustration of running two iterations of KMeans. In each iteration, we assign each training example to the closest cluster centroid, then each cluster centroid is moved to the mean of the points assigned to it.[6]

4.2.2 Clustering based on Location

The first clustering approach we employed focused on grouping bike stations based on their geographical coordinates (latitude and longitude). This spatial clustering technique aims to identify clusters of stations that are geographically close to each other. By clustering stations based on their proximity, we aimed to uncover spatial patterns in the distribution of bike rental facilities across the city of Valencia.

Figure 4 shows the results. The size of the stations, represented as circles on the map shows the size of each station - stations with a bigger total number of racks in a station are bigger, and vice versa. We decided to cluster the stations into 8 clusters. Colors represent the particular clusters.

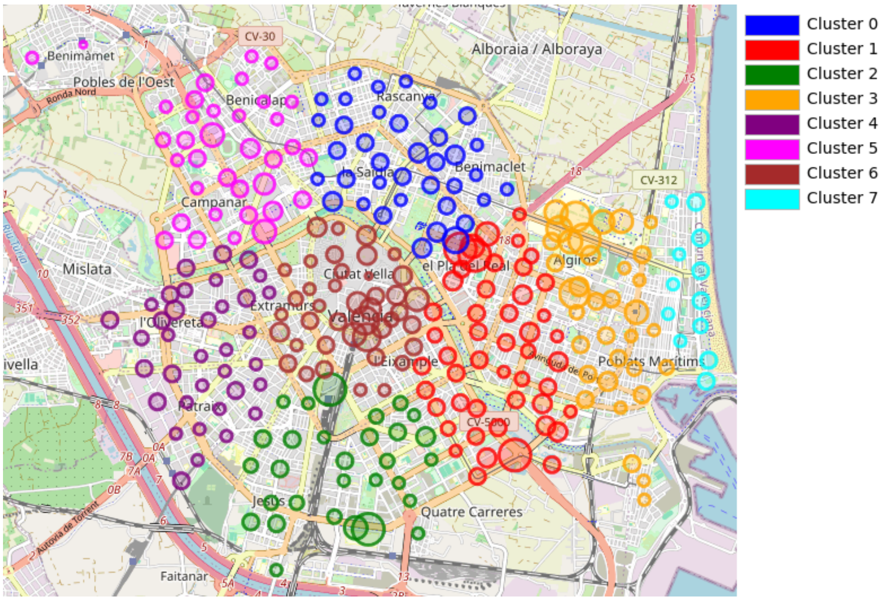

Fig. 4: Created clusters of stations based on their location with labels

Clustering based on location helps us understand how bike rental stations are used in different parts of Valencia. This lets us group stations in neighborhoods or city areas. For instance, knowing that city center stations are busy on weekdays and beach stations are popular in summer helps in managing resources efficiently. Additionally, it helps in identifying regional differences, like areas with high tourist activity, thus we can tailor services accordingly.

However, relying solely on location for clustering can overlook important variations. For example, places like hospitals or universities might have similar usage patterns even if they're not geographically close. This could lead to inefficient resource allocation, as stations with similar behaviors may end up in different clusters. Also, it might miss out on seasonal trends or cross-neighborhood influences, which are important for understanding bike rental dynamics across Valencia.

4.2.3 Clustering based on Similar Behavior

In addition to clustering based on geographical location, we also performed clustering based on the similar behavior of bike stations during the day. This approach involved analyzing the historical usage patterns of stations, and the hourly fluctuations in the number of available bikes. By clustering stations based on their behavioral similarities, we sought to identify groups of stations facing similar challenges or exhibiting similar usage patterns.

For this clustering, it was necessary to prepare the data. The average absolute difference in bike counts per station was computed for each hour. This operation created a dataset where each row represented a station and every feature represented the average absolute difference for a particular hour.

Figure 5 shows the results. Again the size of the stations, represented as circles on the map, shows the size of each station. For behavior clustering, we decided to cluster the stations into 8 clusters. Colors represent the particular clusters.

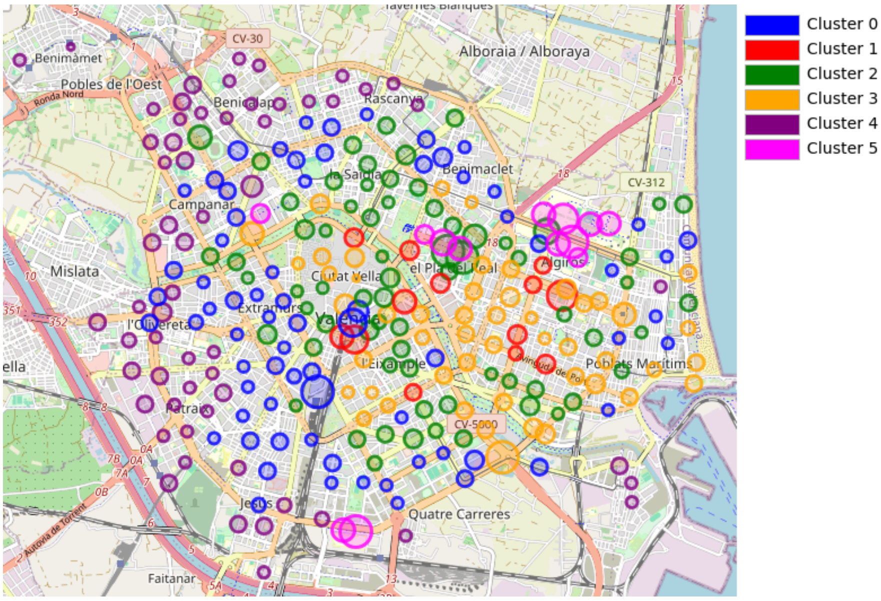

Fig. 5: Created clusters of stations based on their behavior with labels

Clustering based on behavior, where we analyze the usage patterns and operational dynamics of bike rental stations, offers advantages compared to clustering based on location. By considering how stations are used during the day, we can identify groups of stations that share similar usage patterns, regardless of their geographical proximity. For example, stations in different neighborhoods may exhibit similar usage patterns during certain times of the day or in response to specific events, such as festivals or holidays.

Additionally, clustering based on behavior enables us to better understand the underlying factors driving bike rental demand and to tailor strategies accordingly, leading to more effective resource allocation and service improvements

5 Experiments and Results

Data analysis and knowledge extraction from the data will be a theme of the following sections.

5.1 Exploratory Data Analysis

We will start with an analysis of the dataset and its properties to better understand the Valenbisi network situation.

Distribution of the stations' sizes

In this section, we conduct exploratory data analysis to gain insights into the data. Figure 6 illustrates the distribution of station sizes. From the graph is visible that most of the stations has a number of total racks between 15 and 25.

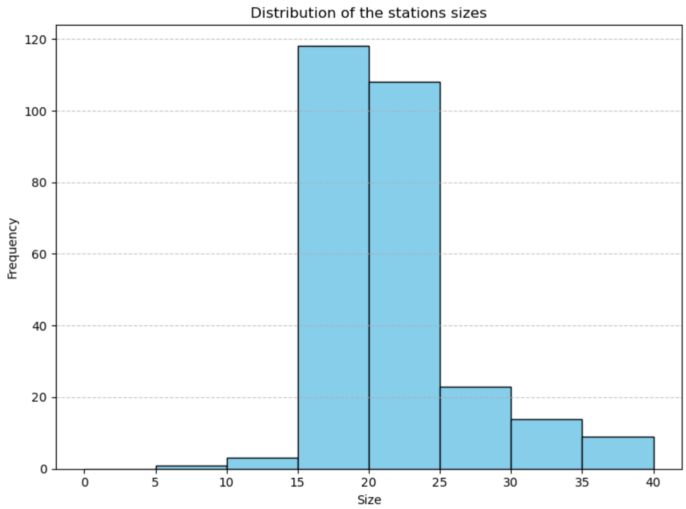

Fig. 6: Distribution of the stations sizes

The distribution of the stations in the city is visible in the graph 7. Each black circle on the map is a station, size of the circle differs according to the station's size. Stations with a bigger total number of racks are bigger and vice versa. The biggest stations are in the city center and in highly populated and important areas like universities, central hospital, and train stations. Main roads like Carrer d'Alfauir, Avinguda de Blasco Ibánez or Avinguda dels Tarongers are also surrounded by big stations. On the other hand for example the area Benimamet contains just two small stations (10 and 15 racks) with a high distance from other stations.

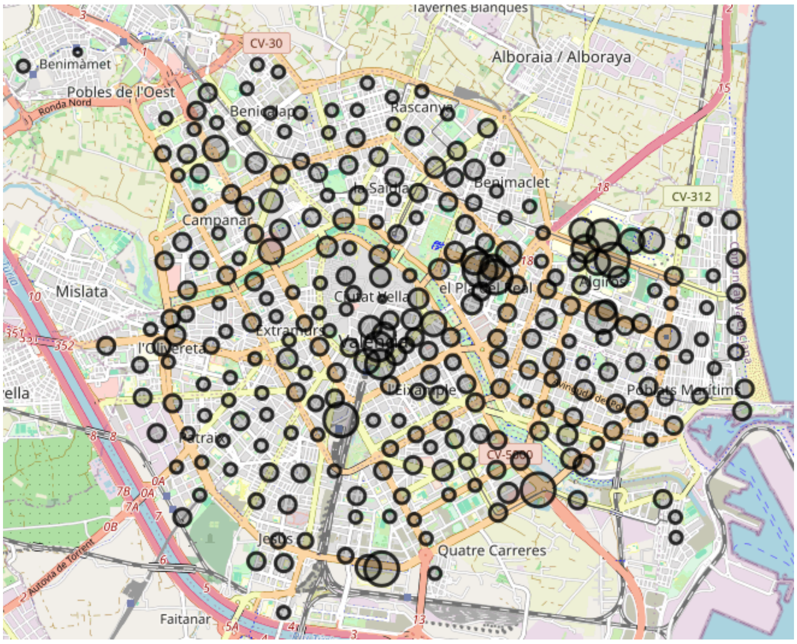

Fig. 7: Comparison of the stations' size and their average occupancy

Comparison of the stations' size and their average occupancy

The red rings in the black circles in the graph 8 correspond to the average number of available bikes in the station. Stations where the red ring is nearly the same size as the black circle are usually very full and it would be good to larger them or add another station near them.

These stations are for example around the street Avinguda d'Ausias March, or Avinguda de Peris i Velero. Highly occupied are usually also the stations in the districts Benimaclet, Natzaret or Camins al Grau, and also stations around the beach. Solitery highly occupied stations in their areas are for example stations in the Plaça de César Orquín Serra, next to the la Fonteta de Sant Lluís, in Carrer de l'Esparreguera, or in Carrer de l'Arquitecte Segura de Lago.

Comparison of the stations' sizes and their average vacancy

Graph 9 shows the comparison of the stations' sizes (black circle) and their average vacancy. The blue rings in black circles correspond to the average number of free racks in the station. Stations where the blue ring is nearly the same size as the black

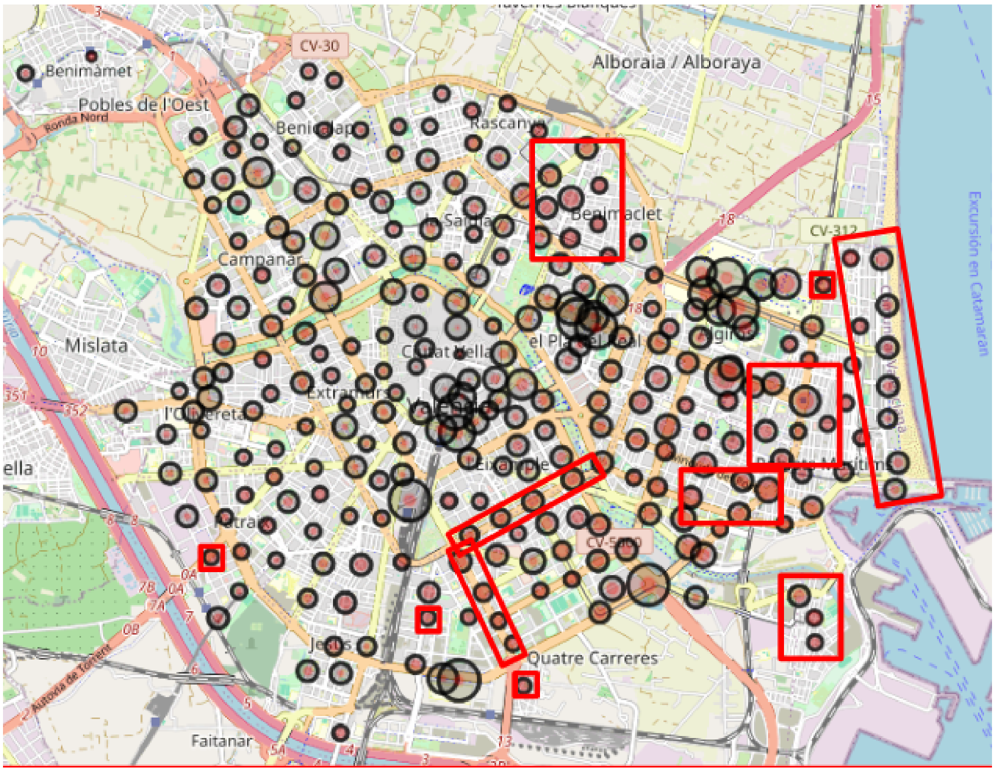

Fig. 8: Comparison of the stations' size and their average occupancy

circle are usually very empty and users may have a problem borrowing here a bike. A possible solution would be to transport more bicycles to these areas.

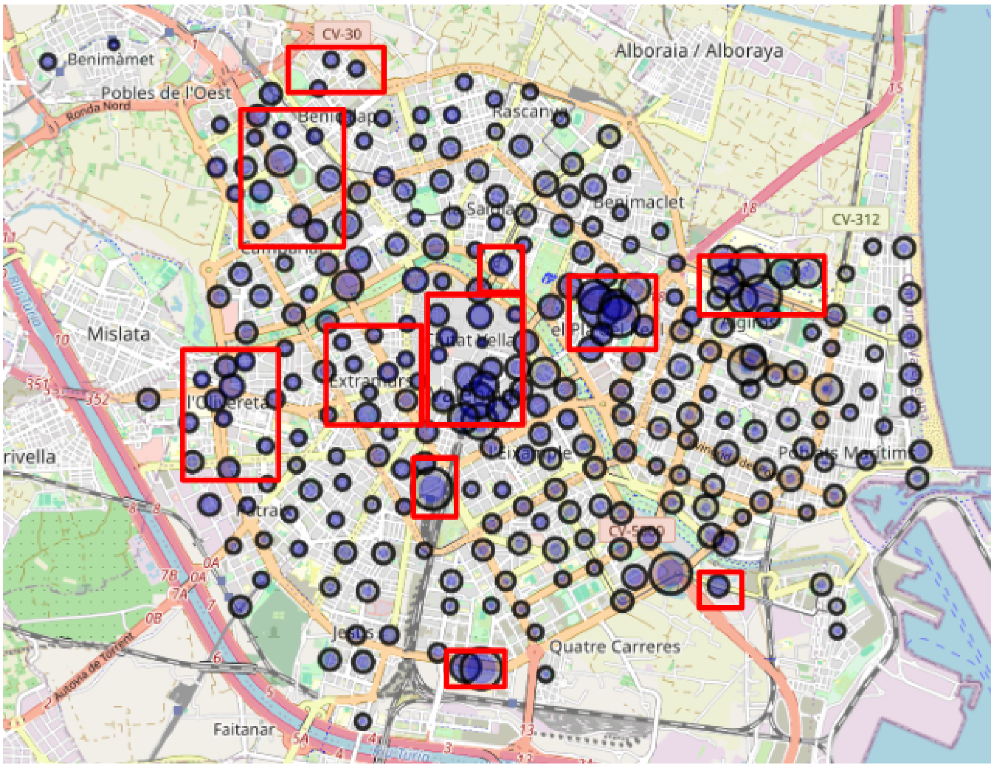

Fig. 9: Comparison of the stations' sizes and their average vacancy

Stations with high vacancy are in the historical city center Ciutat Vella, especially around Colón. The university campuses also have problems with high vacancies. Other heavily affected parts of the city are l'Olivereta, Beniferri, Benicalap, and Extramurs. Solitary highly vacant stations in their areas are for example stations in Carrer del Poeta Bodria, the station next to the Oceanographic Museum, and the stations around the Hospital Universitari i Politècnic La Fe, or around the train station Valencia Joaquín Sorolla.

5.2 Data analyses of the clusters

As we have previously explained, both location-based and behavior-based clustering have their advantages and disadvantages when it comes to explaining bike traffic in the network. However, clustering based on behavior is of greater interest as it associates stations facing similar problems and dysfunctions when providing their services. In this section, we will focus on analyzing the data using solely the results of the clustering based on behavior.

The clustering algorithm has separated the stations into six clusters. Their associated numbers and colors are as follows: 0-blue, 1-red, 2-green, 3-yellow, 4-purple, and 5-pink. As depicted in Figure 10, the 1-red cluster is composed of highly frequented stations located centrally within the Valenbisi network, such as Blasco Ibáñez stations and those near the train station. The 3-yellow cluster is associated with central stations, whereas clusters 2-green, 0-blue, and 4-purple progressively move away from the network center. This directly influences the average usage that users allocate to each station, as Valenbisi usage tends to be more appealing for shorter distances, making it advantageous to be located deeper within the network. This substantial difference in average usage among clusters is evident in other graphs such as 14 and 17, which will be further analyzed later.

Finally, the 5-pink cluster exhibits the most distinct behavioral differences, as will be analyzed in this section. This cluster comprises stations from three markedly different areas: the campus of the Universitat Politècnica de Valéncia, the campus of the Universitat de Valéncia, and the Hospital La Fe. It also includes the Valenbisi station nearest to the Nou Centre shopping mall, one of the largest in the city.

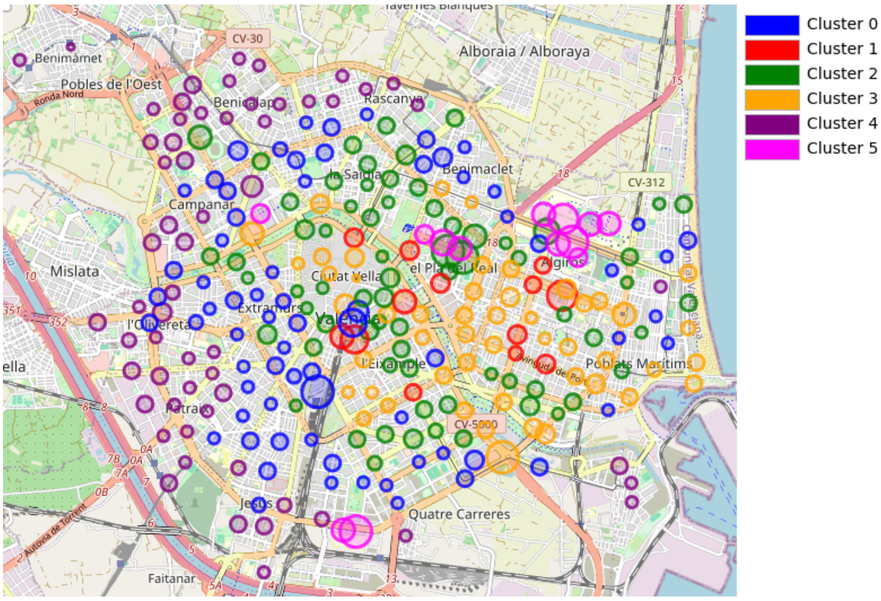

Fig. 10: Created clusters of stations based on their behavior with labels

5.3 Differences by date

Another interesting behavior pattern to analyze is connected to changes in network usage based on time. We will analyze how the system acts differently according to an hour of the day, day in the month, during the year, or day category.

5.3.1 Hourly changes

The first thing that catches the eye when looking at 11 is how all lines in the graph plummet during the nighttime hours. While this behavior may be trivial, as logically bike usage decreases at night, it is worth highlighting the remarkably different behavior of the stations in cluster 5, namely those close to the universities and the Hospital Universitari i Politècnic La Fe. The variation in the usage of these stations is clearly evident: it peaks during rush hours and slowly declines throughout the late afternoon.

An even more differentiated behavior is shown in 12, where the average hourly occupancy is depicted. During nighttime hours, the average percentage of racks occupied by a bike in cluster 5 stations stabilizes around 20% while in the rest of clusters, it hovers between 40% to 50%. Furthermore, cluster 5 stations' occupancy soars rapidly from hours 7 to 10 and gradually descends during the rest of the day, except for a minor peak at 16.

This gives us a clear notion of the nature of each group of clusters. While cluster 5 relates to stations that follow a "business hours" behavior, the rest of the clusters behave as expected for residential areas.

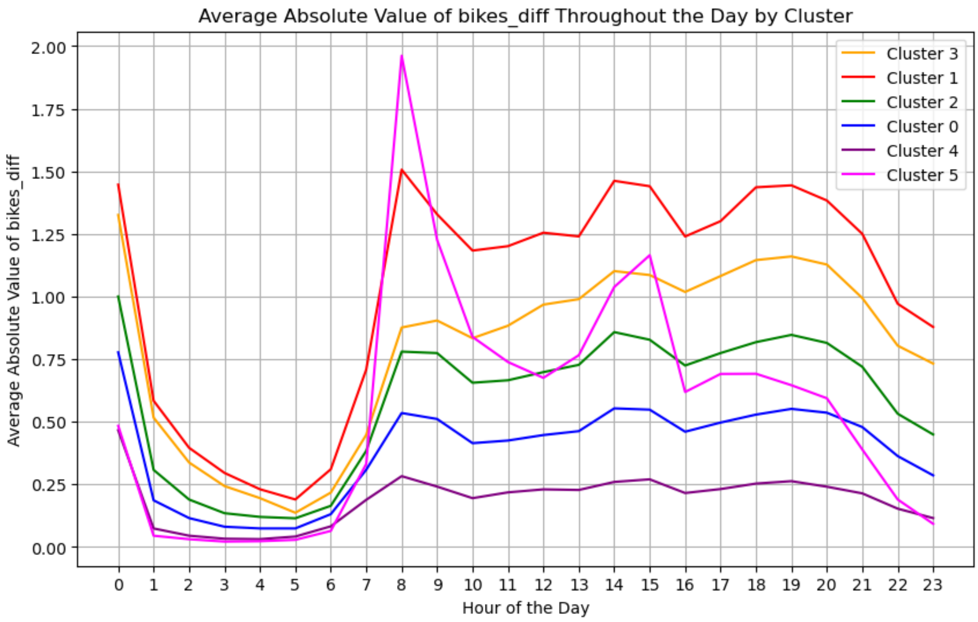

Fig. 11: Hourly Average Absolute Value of bikes_diff by Cluster

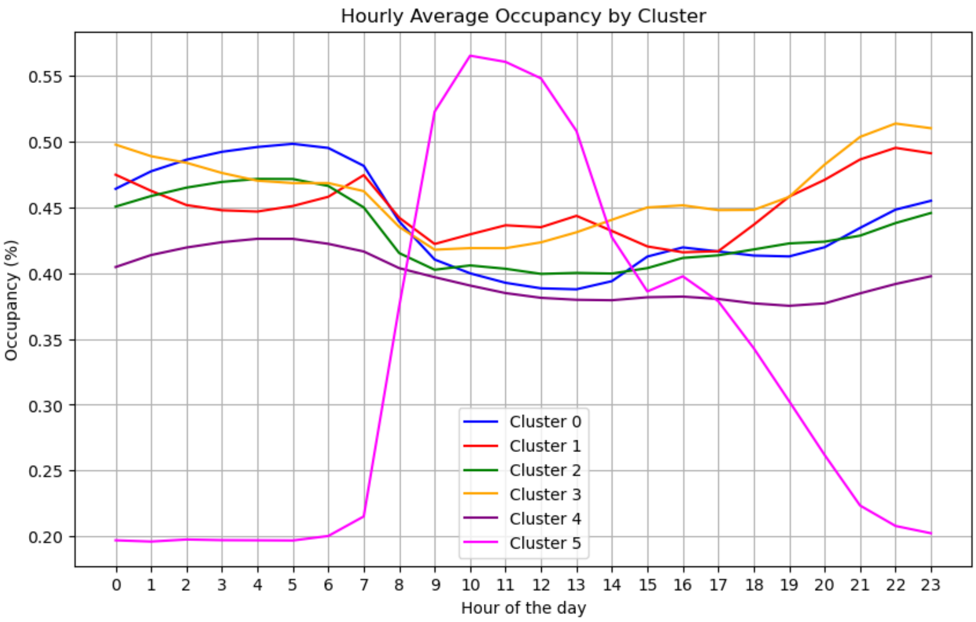

Fig. 12: Hourly Average Occupancy by Cluster

5.3.2 Monthly changes

The average usage of each station by month in 13 drops heavily during the winter months of December, January, and February. This relation between usage and temperature will be further explored later. Regarding cluster 5, we can observe the steep effect of the summer season in July and mainly August, when fewer students use Valenbisi to attend their classes. We can detect this same effect in 14, where average occupancy declines during summer and begins to ascend from September on. The remarkably high percentage of occupancy in the month of June is left to be more closely investigated in the future.

Clusters 0, 1, 2, 3, and 4 in 13 are clearly distinguished from each other by their average monthly usage. This distinction is closely related to the spatial arrangement of the stations in the network. In map 5, it can be clearly seen that red stations (belonging to cluster 1) are more central, followed closely by the yellow ones (cluster 3), then the green ones (cluster 2), then the blue ones (cluster 0), and finally the purple ones (cluster 4). The closer a station is to the center of the network, the higher its average usage. This way of grouping stations makes sense when we take into account that the proximity to other stations and key places in the city can affect the decision of biking or taking other means of transportation.

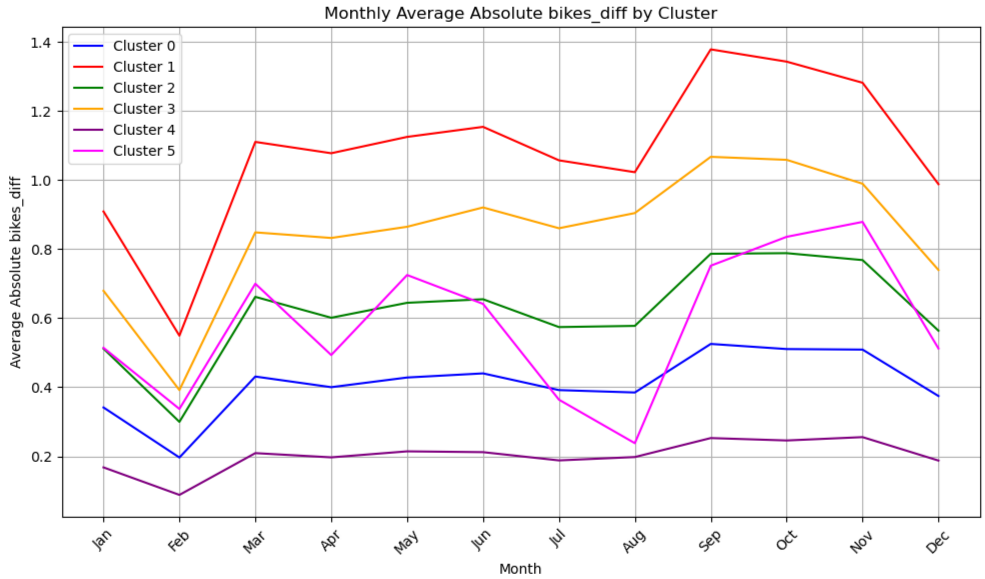

Fig. 13: Monthly Average Absolute Value of bikes_diff by Cluster

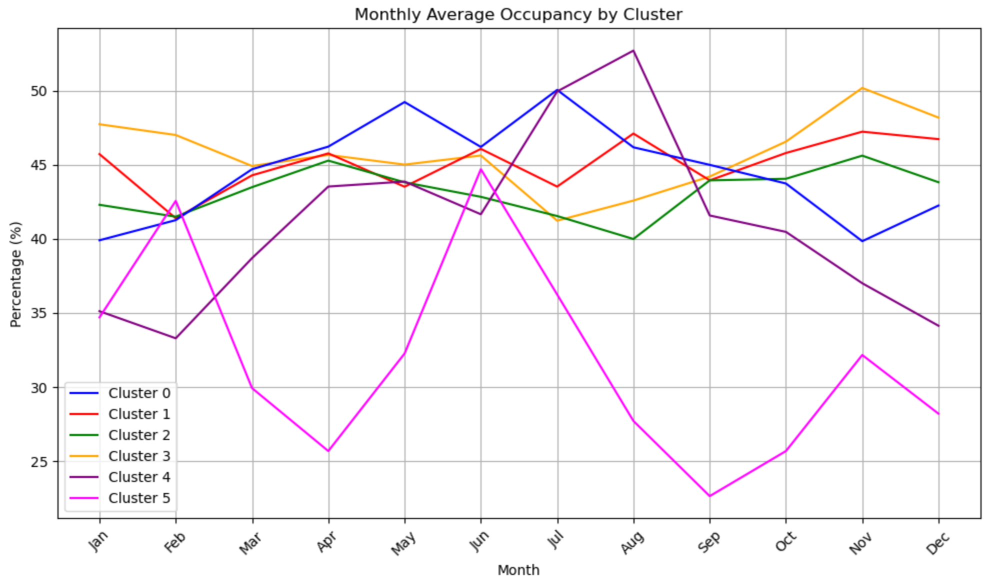

Fig. 14: Monthly Average Occupancy by Cluster

5.3.3 Differences based on day category

Another factor we decided to analyze in this study is the relationship between the usage of the Valenbisi network and the type of day, namely, weekdays, weekends, or holidays. We show the distribution of each category in 15 following the categorization previously explained in 3.

In 16, we find again the same distinct behavior of cluster 5. The average occupancy on non-teaching days (public holidays and weekends) is much lower than in working days. This effect would be even more pronounced if we didn't have into account those working days that are non-teaching days, such as summer, Easter, and Christmas. As expected from previous results, the occupancy percentage in the rest of the clusters is always lower on working days and slightly higher on public holidays and weekends.

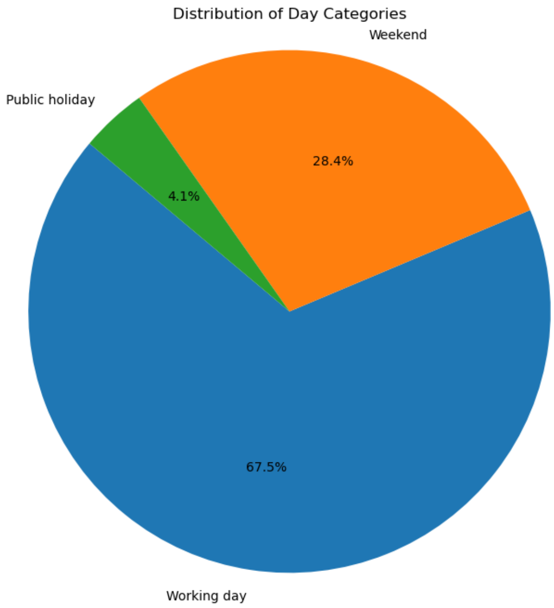

Fig. 15: Distribution of Day Categories

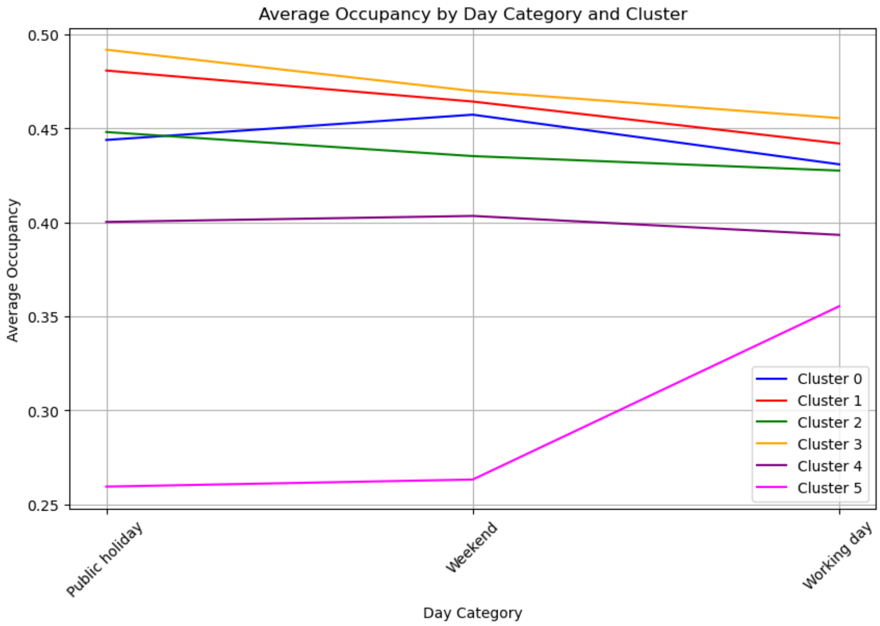

Fig. 16: Average Occupancy by Day Category and Cluster

5.4 Differences by weather

Weather can highly affect the usage of the network Examining how weather influences the system will be the purpose of the following chapter.

5.4.1 Temperature

The effect of temperature on the utilization of the Valenbisi network exhibits similarity across clusters 0, 1, 2, 3, and 4. As depicted in Figure 17, a peak in usage is observed within the temperature range of 25 to 30 degrees Celsius across all clusters. Also, a marginal decline in usage is noted with higher temperatures, although maintaining relative stability. Conversely, usage progressively diminishes with lower temperatures, reaching its minimum within the minimal temperature range of 0 to 5 ºC.

Cluster 5 follows a very similar pattern for temperatures of 25 to 30 ºC and less, but the decline in usage when reaching higher temperatures is more pronounced than in the other clusters. This is most likely due to the decrease in students attending university classes during the hottest months of the year, namely July and August.

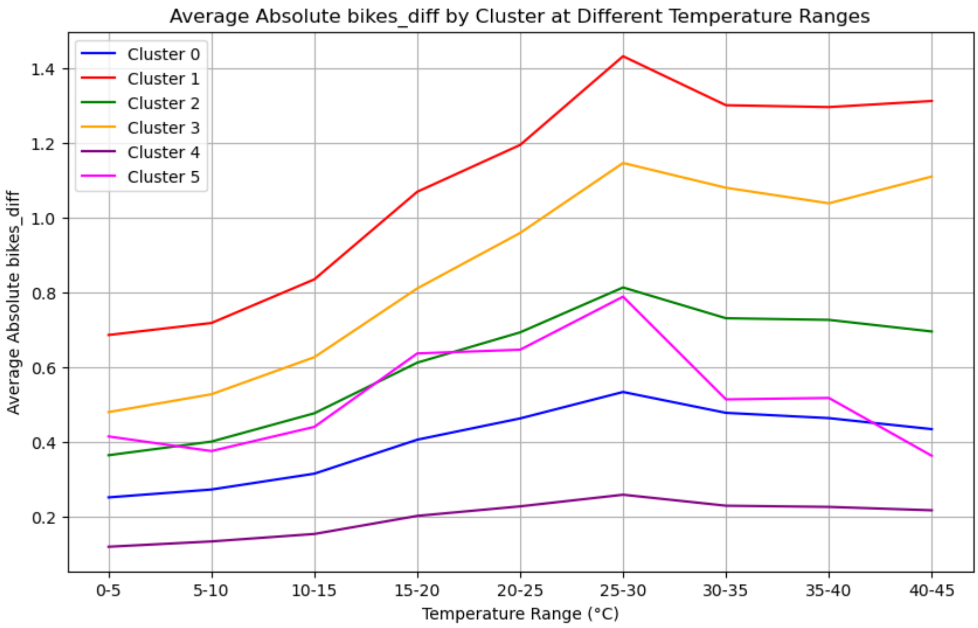

Fig. 17: Average Absolute bikes_diff by Cluster at Different Temperature Ranges

5.4.2 Precipitation

Regarding the effect of precipitation on network utilization shown in 18, it should be noted that all clusters follow a very similar pattern: usage peaks at 0.5-1 mm and 1.5-2 mm ranges and descends abruptly with all other ranges without ever reaching true minimal values. Although cluster 5 follows this trend too, it peaks more notably at the 0.5-1 mm range and recovers slightly at the 3-4 mm range only to decline steeply at the >4 mm range.

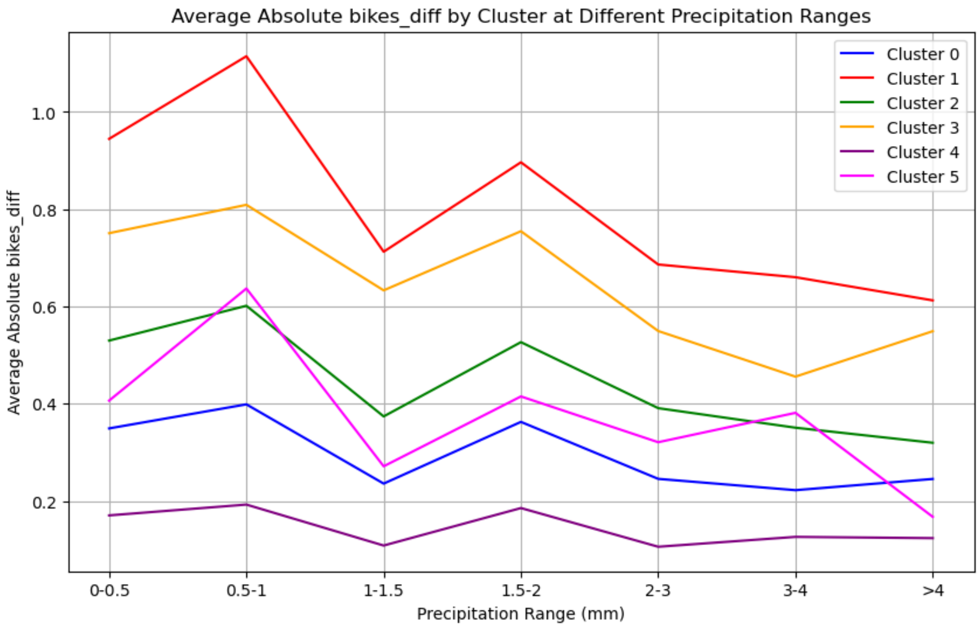

Fig. 18: Average Absolute Value of bikes_diff by Cluster at Different Precipitation Ranges

5.4.3 Sky condition

The sky condition attribute of the dataset gives us information about the overall weather conditions. The attribute itself consists of a high number of different very detailed descriptions of the current weather situation.

Categorization of Sky Conditions

To enhance data visualization and analysis, we undertook a process to categorize the sky condition values into fewer, more manageable categories. This categorization simplifies the interpretation of weather patterns and facilitates clearer insights. We categorized the sky conditions as follows:

Sun:

- Despejado
- Soleado
- Parcialmente nublado

Clouds:

- Cielo cubierto
- Nublado
- Neblina

Rain:

- Lluvia moderada a intervalos
- Ligeras precipitaciones
- Llovizna, Ligeras lluvias
- Llovizna a intervalos
- Periodos de lluvia moderada
- Lluvia moderada
- Lluvias ligeras a intervalos
- Periodos de fuertes lluvias

Storm:

- Cielos tormentosos en las aproximaciones
- Intervalos de lluvias ligeras con tormenta en la región
- Lluvias fuertes o moderadas, Fuertes lluvias

From the graph 19 depicting the distribution of sky conditions in Valencia, we can observe that the main sky condition is sunny weather, followed by cloudy conditions, rain, and stormy weather. This uneven distribution complicates the study of the impact of sky conditions on the Valenbisi network due to the scarcity of non-sunny days, especially rainy ones.

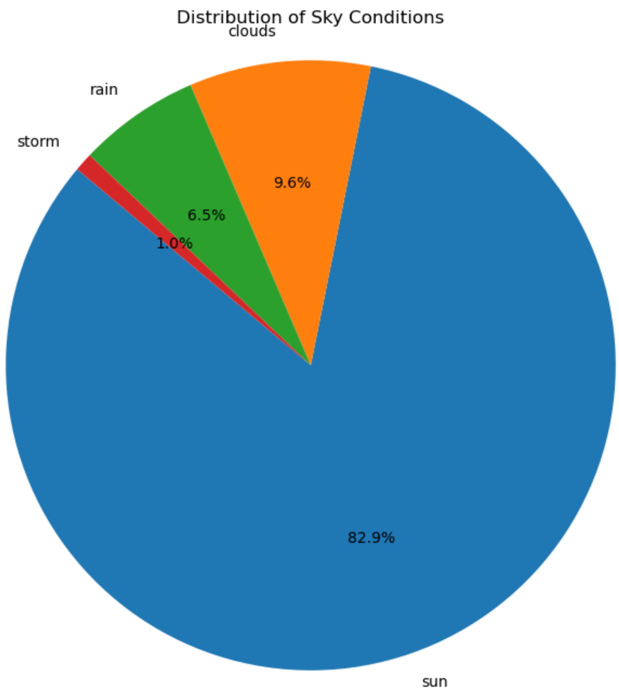

Fig. 19: Distribution of Sky Conditions

The average vacancy by sky condition, as illustrated in 20, provides a visually compelling insight into the inverse relationship between cluster 4 (stations located primarily in residential areas) and cluster 5 (stations located mainly in university areas). While the percentage of free racks in cluster 4 increases with clouds, decreases with rain, and continues to decrease with storm, in cluster 5 it decreases with clouds and continuously increases with rain and then with storm. This behavior aligns with the expected outcome that people tend to stay at home when it rains, especially during a storm.

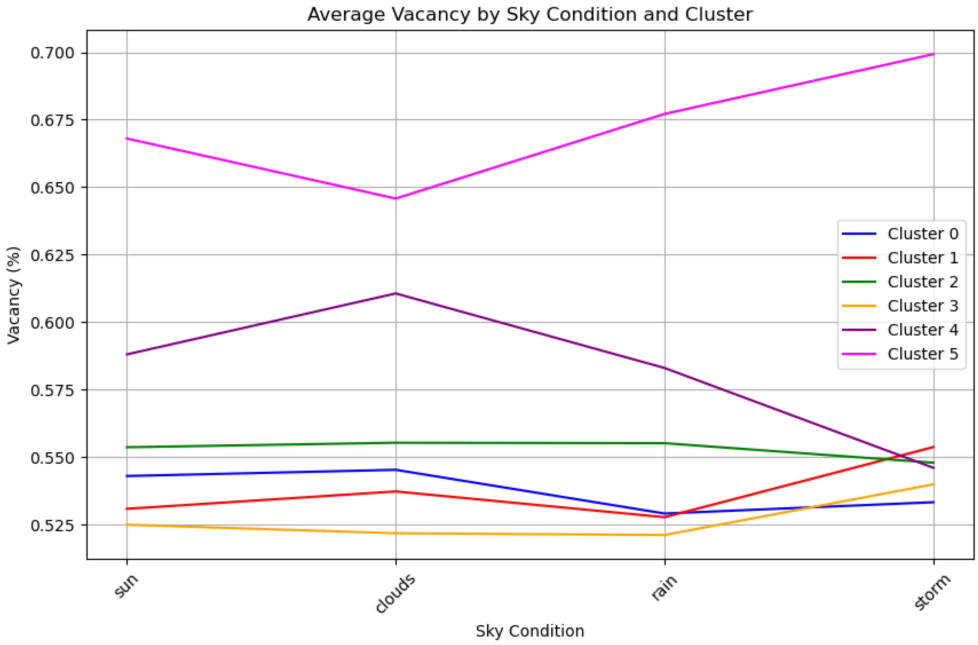

Fig. 20: Average Vacancy by Sky Condition and Cluster

6 Conclusion

In conclusion, our study delves into the analysis of Valencia's Valenbisi bike network, aiming to address challenges in bike availability and station occupancy across different parts of the city and during peak times. By analyzing historical data and considering factors like weather conditions, holidays, and user flow, we aimed to provide valuable insights to enhance the network's quality.

Our research uncovered distinct usage patterns and operational dynamics across various clusters of bike stations. The clustering process revealed spatial and behavioral differences. We identified critical factors influencing bike rental demand, such as temperature, precipitation, and sky conditions, offering perspectives to optimize the network.

Furthermore, our analysis highlighted the impact of day categories, with notable differences observed between working days, weekends, and holidays. Stations located in university areas exhibited unique usage patterns, particularly influenced by academic schedules.

Finally, future research could explore specific actionable interventions and their effectiveness in addressing the identified challenges, ultimately contributing to smarter, sustainable, and more efficient urban mobility networks.

References

[1] Balachandran, G., Kumar, S.D., Nageshwaran, S., Ram, R.N., ME, M.R.K.: Bike rental demand forecast using ml techniques

[2] Ahn, Y., Okamoto, D., Uejio, C.: Investigating city bike rental usage and wet-bulb globe temperature. International journal of biometeorology **66**(4), 679–690 (2022)

[3] Czech, P., Turoń, K., Urbańczyk, R.: Bike-sharing as an element of integrated urban transport system. In: Advanced Solutions of Transport Systems for Growing Mobility: 14th Scientific and Technical Conference" Transport Systems. Theory & Practice 2017" Selected Papers, pp. 103–111 (2018). Springer

[4] Eren, E., Uz, V.E.: A review on bike-sharing: The factors affecting bike-sharing demand. Sustainable Cities and Society **54**, 101882 (2020)

[5] Bock, H.-H.: In: Brito, P., Cucumel, G., Bertrand, P., Carvalho, F. (eds.) Clustering Methods: A History of k-Means Algorithms, pp. 161–172. Springer, Berlin, Heidelberg (2007)

[6] Jordan, M.: Kmeans algorithm. Stanford CS221 (2013)