EV Charging Station Network Analysis – the city of Milan

IMB Data Science Professional Certificate – Capstone Project

Author: Riccardo Donega

Abstract

My capstone project is focused on the analysis of the Electric Vehicle (EV) charging station network within the city of Milan. In order to consider the attractiveness of a charging station I will consider the proximity between them and other POI like bars, restaurant and cafe. Since the average EV owner will need to make quick charge stop during the day the best clusters will be the ones with a good concentration of bar and other venues near the charging point where the EVs owners could spend their time during the charging process. I will use a clusterization model (K-Means) to find some cluster with different characteristic and I will identify the best ones following my assumptions.

Introduction: Business Problem

The Mobility and Automotive sector are facing crucial changes both in term of business and technologies. These changes are led by the introduction of the Electric Vehicles (EV). Since these vehicles require a public and private charging network the institutions and governments need to enlarge and create efficient charging points networks.

In particular, to gives to the people the chance of charging the EV during the day enabling them to make long travel without coming back home, it is important to have the right number of charging point in the cities. Moreover, these infrastructures must be localized inside strategic area. In the future, people who travel from a city to another or even in the same city will need to charge the EV for short time in order to store the energy needed to come back home or to another charging point. To do that, the charging point must be localized near infrastructures that allows the people to stay for less than one hours working or enjoying a meal or drink.

In my project, I will analyze the actual charging network in the city of Milan. I will divide the network in different clusters based on the proximity of the charging stations to interesting POI like bars and restaurants. I will also take in consideration the reputation of that POI and the number of available charging point in that cluster. In this way, I will be able to find the best location (and cluster) to charge the EV during the day.

Therefore, I am interested to find the clusters where the charging infrastructures (and the charging points) are close to a high number of popular bars and restaurant.

This analysis has a different scope for different stakeholders:

- It could help people to find the best charging spot
- It could help the municipality to orientate the charging network development
- It could orientate the opening of new bars and POI near the more isolated charging stations

The Data - sources

For my project I used three main kind of data. The first are the data related to the charging point distribution in the city of Milan. The second are the Venues data in the city, while the third are the Suburbs' data of Milan

In the first case I chose to use the open source website opencharge map.org which offer an open source API service, to download the charging station data and localization

In the second case, I used the Foursquare API service to download the data related to the Venues which are close to each charging stations, including also the 'Likes' data for each venue.

To obtain the Suburbs' coordinate I used the API service of OpenChage. In this way I downloaded the suburbs related to each charging station.

The Data – feature selection and database creation

In order to build the dataset used for the analysis I started from the charging station. I download the data of all the Italian charging stations then I selected only the ones related to the city of Milan using the postcodes.

The database that I obtained includes the ID for each charging station, the number of charging point and the glocalization.

ChargePointID	NumbOfPoints	PostCode	Longitude	Latitude
150098	6.0	20139	9.237619	45.433683
149126	6.0	20123	9.172799	45.455730
148120	2.0	20158	9.164095	45.503988

Table 1: Charging Stations DB

Afterwards I moved to the Venues. First, I used the geodata of each charging station to download the Venues which are localized around them (radius of 200m). Once I obtained the closest Venues, I eliminated the ones which are not in the Restaurant, Bars od Café categories. Then I grouped the remained records in four macro categories: bars, restaurant, café and food courts. Then, I used the Venue ID filed to download the 'likes' data for each Venues. In this way I created a new column including the number of likes that the client have assigned to each venue. This allow to consider the popularity of each Venues.

ChargePointID	VenueID	Venue Latitude	Venue Longitude	Venue Category	Likes
149126	4f36d248e4b0e313d2e4eca7	45.455556	9.171689	Restaurant	142
23314	4f36d248e4b0e313d2e4eca7	45.455556	9.171689	Restaurant	142
149126	4b05887ff964a520b4c922e3	45.455482	9.173333	Restaurant	52

Table 2: Venues DB

Then, I merged the Venue data and the charging station data using the Charging Point ID as key. In this way I obtained a database where each record represents a Venues which is close to a specific charged station. The last step was to obtain the dichotomic variable for the Venue category and to group the data by the charging point ID.

The database contains the number of venues divided in four categories for each charging station. Moreover, it contains the number of total charging point and the total number of likes of off the venues around the selected charging station.

ChargePointID	Longitude	Latitude	Likes	NumbOfPoints	Bar	Café	Food Court	Restaurant
21095	9.192996	45.489257	49.0	3.0	0	1	0	2
21138	9.209498	45.447949	53.0	10.0	0	1	1	3
21155	9.192835	45.459366	158.0	12.0	1	1	0	4

Table 3: Venues Likes

Finally, I associated each Charging point to the related city suburbs by using a reverse geocoding process. In this way the final database includes also the city suburb where each station is located

	Suburb	ChargePointID	Longitude	Latitude	Likes	NumbOfPoints	Bar	Café	Food Court	Restaurant
0	Baggio	47994	9.096476	45.449970	0.0	2.0	0	0	0	0
1	Bicocca	73182	18.423495	91.034368	12.0	6.0	1	0	0	1
2	Boldinasco	93161	9.137276	45.498186	2.0	2.0	0	1	0	0

Table 4: Final DB

How it is clear from the figures below, the charging points are located mostly in the center of the city. Only a couple are located near the border and, differently to the ones in the center, they do not have many Venues around them.

They seem to be not suggested for intra day charging stop. Probably this factor is mitigated from the proximity to the main roads however they are the first sign of the lack of an integrated plan able to offer leisure place around the charging station. This lack, how I will analysis, could be the reason for a lack of EV adoption and to the low use of public charging infrastructures.

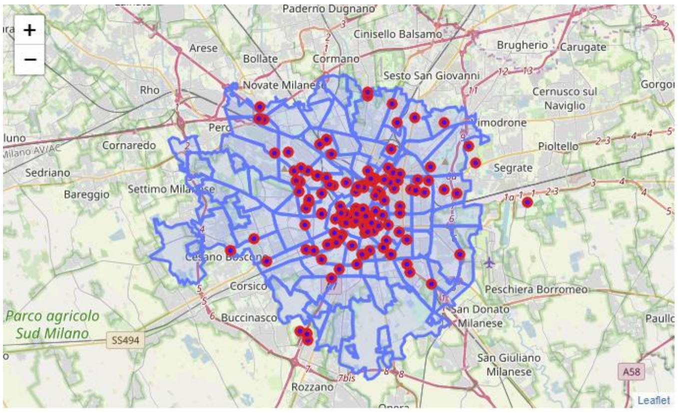

Figure 1: Charging Stations in Milan

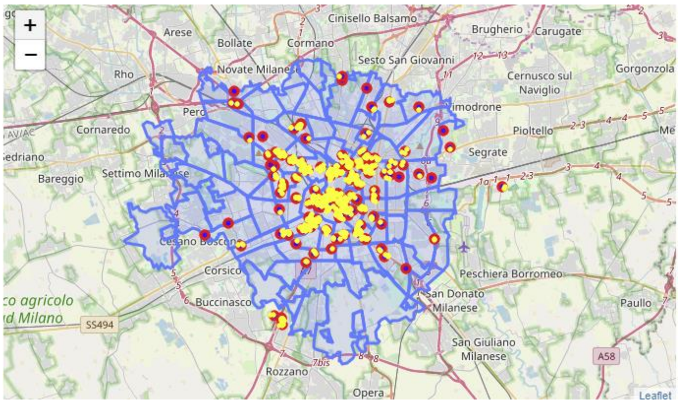

Figure 2: Charging Station (red) and selected Venues (yellow)

In the figure below we have a first look on the Venues categories. The grand part of them are Restaurant (of any kind) then there are the Bar and Café categories. Last there is the Food Court category. Theoretically, Bars and Café could be the best place to make quick stop (less than 1h) to charge the car. Moreover, they are usually provided with Wi-Fi which allow to continue working during the stop. However, also Restaurant and Food Court must be considered even if they could be more expensive.

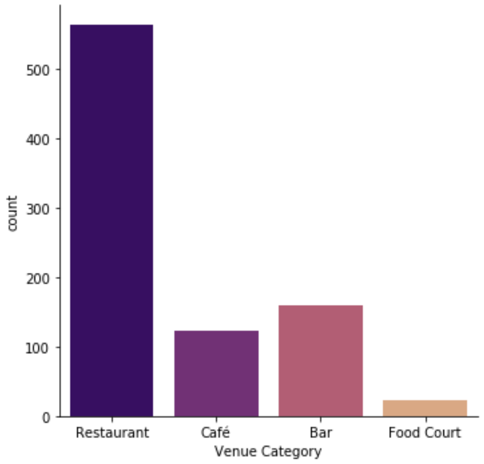

Figure 3: Venues Category

Methodology

In order to pursue the analysis phase I decided to apply a K-means clustering model. I used the Final Database by grouping all the records by the Suburbs. In this way each row was related to a single suburb and they contain the sum of number of charging point the sum of venues for each category and the sum of Likes of the venues associated to the single suburb.

Obviously before to run the model I exclude the rows containing useless information for the clusterization process like the geocoordinates, the suburb name and the charging station ID.

By applying a clusterization model on the suburbs' information my aim was to divide the suburbs in some cluster characterized by the same features. In this way it was possible to determine which suburbs cluster was the best in term of Charging Station network in relation to the closest venue and their reputation.

I chose to run the model for an incremental number of clusters starting from 2. I used the average Silhouette score like index to choose the best number of clusters.

The silhouette value is a measure of how similar an object is to its own cluster (cohesion) compared to other clusters (separation). The silhouette ranges from −1 to +1, where a high value indicates that the object is well matched to its own cluster and poorly matched to neighboring clusters. If most objects have a high value, then the clustering configuration is appropriate. If many points have a low or negative value, then the clustering configuration may have too many or too few clusters. By computing the average for each object in a cluster it is possible to estimate a general index of fitness of the model.

```
The silhouette_score for 2 clusters is 0.850481021382247
The silhouette_score for 3 clusters is 0.5999900796423698
The silhouette_score for 4 clusters is 0.5514813689085527
The silhouette_score for 5 clusters is 0.46112297462140006
The silhouette_score for 6 clusters is 0.5031524834398747
The silhouette_score for 7 clusters is 0.5060960597819064
The silhouette_score for 8 clusters is 0.5084961977744283
The silhouette_score for 9 clusters is 0.4999661023621272
```

Figure 4: Silhouette Score

Like it is possible to see in the figure above the Silhouette score it is pretty good. I chose to use 3 clusters even if the score for two clusters is better. That's why I would like to capture a more difference between the data.

Results

By running the model with 3 cluster I obtained the distribution on the figure below

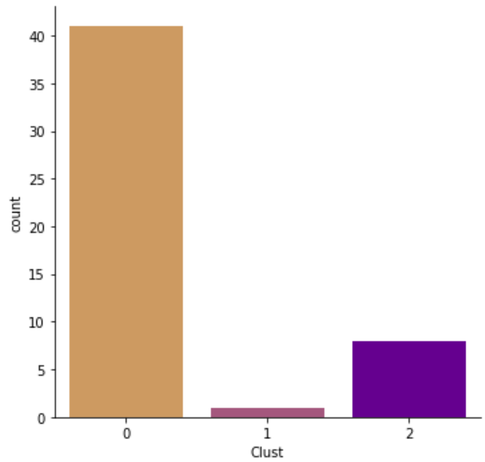

Figure 4: Clusters

Most of the suburbs are included in the cluster 0 (41). Then the cluster 2 (8) and finally the cluster 1 (1). If we look to the Cluster average value, we see:

Drivers	Cluster0	Cluster1	Cluster2
Likes	207.3	11598.0	2128.8
NumbOfPoints	20.1	398.0	126.5
Bar	1.0	43.0	9.4
Food Court	0.2	4.0	1.5
Restaurant	4.8	114.0	31.6
Cafè	1.3	35.0	4.1

It is clear that the best cluster is the number 1 with many charging points and with many venues with high rate. Moreover, there are more Bar and Café respect to the other cluster. The second best is the cluster 2 with a good number of venues and charging points.

At the opposite, the cluster 0 includes suburbs with a smaller number of charging points and most important with not many venues around them (also with less reputation!). This mean that this cluster repress the suburbs where the charging network is not well developed and ready to support a dynamic charging trend of the EV owner.

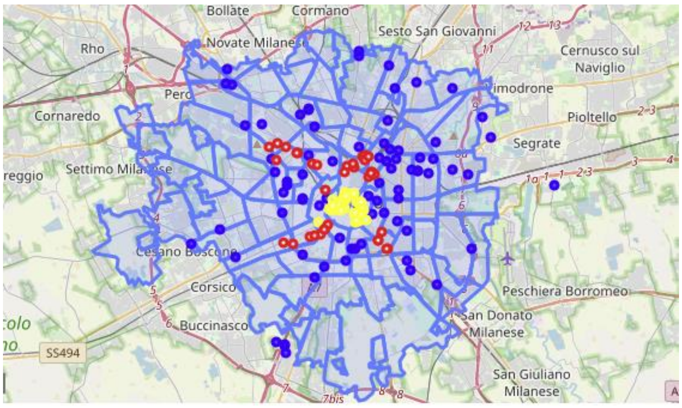

Figure 5: Charging point by cluster

If we take a look to the singles charging station, we can see how the ones in the cluster 1 (in yellow) are in the Historic Centre suburb. While, the station in the cluster 2 (in red) are in the suburbs around the center and on the main road converging to the heart of the city. These suburbs are Porta Garibaldi, Porta Genova, Porta Nuova, Porta Romana, Porta Sempione, Pioltello, Municipio 6 and Municipio 7.

The cluster 0 (in blue) even if includes most of the suburbs like we have seen before it has few charging stations. It means that the peripheric suburbs other that not having many venues around the charging station that have less charging points in comparison with the suburbs in the other clusters.

Conclusion

In conclusion the clusterization model show an interesting trend. The fact that the suburbs near the center are the only one ready in terms of charging stations and available venues around them while the suburbs in the peripheric zone are dramatically less equipped is the prove of the lack of a infrastructural plan.

Indeed, the charging points are located mostly in the city center suburbs because of the popularity of the EV vehicles in these zones and because of the large amount of people using these vehicles within these areas. Moreover, the availability of interesting Venues like Bar and Café is an effect of the proximity o the touristic and most popular area of the city. For this reason, the charging station in this area could take advantage of this infrastructures.

On the other sides far from the city center the charging infrastructure is far less developed. There are not many charging points, and these are located in isolated zone without any facilities around them. So, it is clear that there is not any plan to promote a uniform development of the infrastructure by the municipality. Furthermore, the lack of Venues around the charging point in the suburbs could limit their use because the drivers travelling through Milan to another city may avoid charging their car because the lack of place to wait during the charging process. This could reduce the EV penetration in the market because by limiting the range of use of these vehicles. The idea of buying an EV without having the real change of charging them during the day unless going in the center of the city, make the EV choice far less attractive.

The model shows clearly the limit of a charging infrastructure development plan which do not follow the potential clients needs but it only aims to make easy and sure profit where their EV are already popular.

The municipality should encourage the development of the charging network also outside the city center. This will start a virtuous cycle by spreading the use of the EV and convincing the investor to build new venues around the charging stations.