Car sharing and neighborhoods in Milan, Italy

Paolo Cavadini, August 2020

Table of the contents:

1 Introduction

1.1 Background

Milan, Italy, radically evolved during the last decade, as a consequence of the Milan Expo 2015, from being the main industrial and financial hub of the country into the development of a sustainable touristic economy able to attract people looking for classic and contemporary arts, design, events, as well as nightlife and food. The transportation network in the city grew dramatically along the years, including surface and underground lines, high speed trains to international destinations, and sharing services for cars, bikes and scooters. Car sharing is quite popular. In 2019, Milan counts (source here) 2224 cars in sharing per million of inhabitants, Barcelona 249, Lyon 592, Munich 1158, Stuttgart 1715.

1.2 Problem

It is critical to build new parking areas for car sharing, conveniently located in the central neighborhoods. The Municipality counts only 113 parking areas according to the last update of the dataset about public parking for car sharing (source [4], see Data).

1.3 Audience and interest

The audience with interest in parking for car sharing will be real estate companies, as well as banks and financial institutions. Suppose that potential stakeholders are looking for an area which is not in the high end of the real estate value, is as close as possible to the city center but at the same time has low density of parking areas for car sharing, what neighborhoods should they consider?

I'll narrow down the neighborhoods in terms of distance from the city center (~3Km) and density of parking areas for car sharing (no more than 1 within 300m). I'll get the venues for the selected neighborhoods and I'll cluster them. I'll highlight on a map the clusters showing the most relevant information which will include their real estate value.

The stakeholders will have a short list of clustered neighborhoods to consider for further analysis.

2 Data

2.1 Data sources

Based on definition of the problem, factors that influenced my decision are:

- distance of neighborhood from the city center (~3Km).
- number of existing parking areas for car sharing in the range of 300m from the neighborhood (max 1).

The sources I had to extract/generate the required information:

[1] The website Immobiliare (www.immobiliare.it) publishes real estate values for Italy, aggregated from the regions through borough down to the street level (here). It was used to find the neighborhoods and the related real estate index.

[2] The Openstreetmap Nominatim API for geocoding (here) have been used to find the neighborhoods' latitude and longitude, and the coordinates of Milan center.

[3] The Foursquare Explore API have been used to retrieve the list of venues for the selected neighborhoods.

[4] The open dataset portal of the Municipality of Milan (https://dati.comune.milano.it/dataset) have been used to find the parking areas for car sharing (here).

2.2 Data collection and preparation

Neighborhoods

I obtained a list of neighborhoods in Milan, and their latitude and longitude. I first scrapped the data from a website [1] using **Pandas HTML**, and I used the **Nominatim API** to obtain the coordinates. I finally merged all the information into the neighborhoods dataframe which columns represent the neighborhood name, Zip code, latitude, longitude, and real estate Index.

[3]:

	Neighborhood	Zip	Latitude	Longitude	RE_index
0	Centro	10122	45.466800	9.190500	9.947
1	Arco della Pace	20154	45.475692	9.172428	8.335
2	Arena	20121	45.475264	9.181564	8.335
3	Pagano	20145	45.468285	9.161100	8.335
4	Genova	20144	45.452879	9.169715	7.534

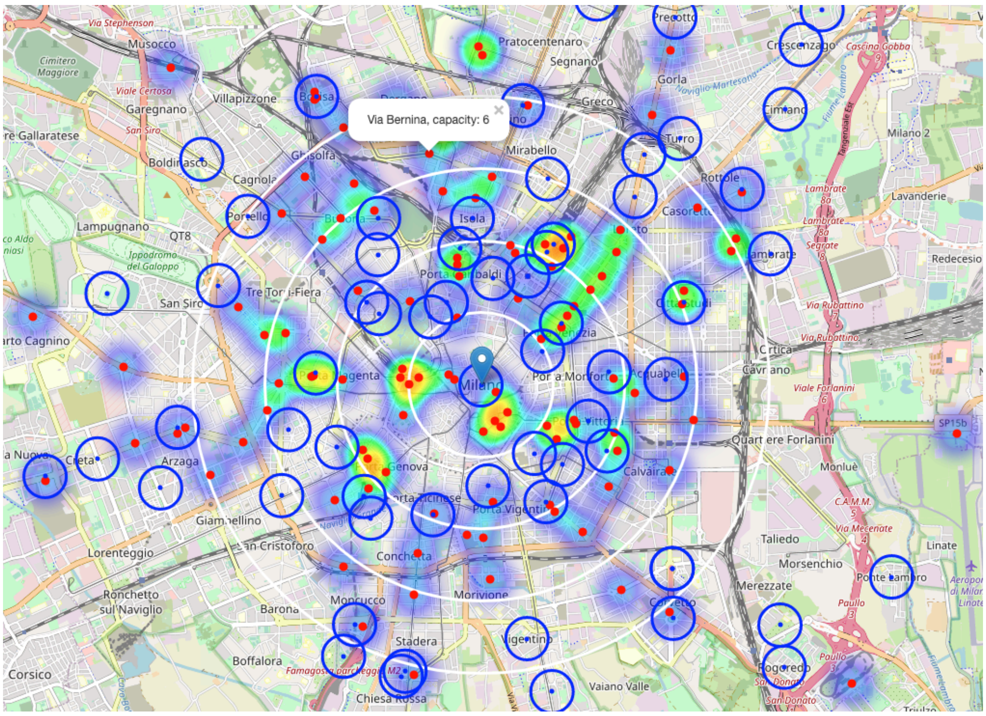
Parking areas for car sharing

I read into a dataframe the csv file about the parking areas for car sharing obtained from an open dataset [4]. I sliced the dataframe and I assigned English meaningful names to the columns.

[6]:

	Parking	Capacity	Address	Latitude	Longitude
0	Loreto/Mercadante	8	Via Mercadante davanti uscita M1 Argentina	45.483351	9.214544
1	Pta Venezia	8	P.zza Oberdan uscita M1 ex casello (fronte Bas...	45.473929	9.204847
2	Pagano	6	Via del Burchiello fronte parcheggio	45.467845	9.160451
3	Cadorna	8	P.le Cadorna fronte civico 1/3	45.466924	9.177717
4	Centrale Stazione	8	P.zza Duca d'Aosta davanti civico 16	45.483864	9.205185

The map below, generated using **folium,** shows the parking areas for car sharing (red dots), and the neighborhoods with a blue circle (300m radius) and a blue dot. Each popup for a parking area displays the parking name and capacity which is a synthetic indicator (it doesn't represent the actual number of parking spots). Each popup for neighborhoods displays the neighborhood name. The heatmap shows the density of the parking areas.



Foursquare

I used Foursquare to obtain **147 unique venue categories for 11 neighborhoods**, which are the result of the specifications agreed with the stakeholders (less than ~3km from the city center, no more than 1 parking area for car sharing in the range of 300m) that I will discuss later in the methodology section.

3 Methodology

In this project I am directing our efforts to detect neighborhoods of Milan that have low density of parking areas for car sharing, and lower real estate value.

In first step I have collected the required data: the neighborhoods obtained scarping data from a website and completing them using the **Nominatim API**, the parking areas for car sharing obtained from an open dataset published by the Municipality of Milan.

In a second step, in agreement with the stakeholders I will **limit my analysis to neighborhoods located ~3km from the city center, and with no more than 1 parking area for car sharing in the range of 300m.** For the selected neighborhoods I will use the **Fourquare API** to obtain the venues.

In the last step I will create clusters (using **k-means**) of neighborhoods, I will present them on a map to identify general zones which should be a starting point for further detailed exploration for the stakeholders. The map will show the clusters together with a heatmap which will represent the density of the parking areas for car sharing.

Selection of the neighborhoods.

In order to calculate the distances between the venues and the city center, and between the venues and the parking areas, I created functions to convert between WGS84 spherical coordinate system (latitude/longitude degrees) and UTM Cartesian coordinate system (X/Y coordinates in meters).

I calculated the distance to the city center for each neighborhood, and **I finally found 29 neighborhoods located within 3Km from the center.** I saved the result to a dataframe which includes neighborhood information and the distance to the city center:

[13]:

	Neighborhood	Latitude	Longitude	Distance from center	Zip	RE_index
0	Centro	45.466800	9.190500	0.000000	10122	9.947
1	Arco della Pace	45.475692	9.172428	1728.171587	20154	8.335
2	Arena	45.475264	9.181564	1174.372421	20121	8.335
3	Pagano	45.468285	9.161100	2310.051655	20145	8.335
4	Genova	45.452879	9.169715	2249.054111	20144	7.534

Now, **my goal is eventually to find the neighborhoods which have no more than 1 parking area within 300m from their center**. To comply with this constraint, I created a dataframe which merges neighborhood information, parking area information and the distance (<=300m) between each neighborhood and each matching parking area.:

[21]:

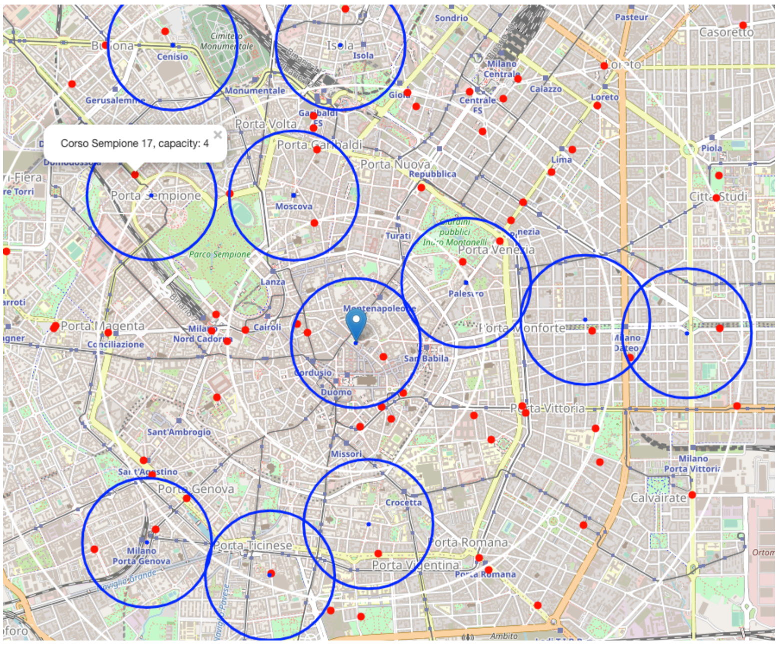
	Neighborhood	Distance	Parking	Capacity
0	Porta Venezia	63.480675	Pta Venezia	8
1	Pagano	70.604732	Pagano	6
2	City Life	239.854356	Centrale Stazione	8
3	Centrale	152.119034	Centrale Stazione	8
4	Cadore	139.887494	P.ta Vittoria/Largo Marinai d'Italia	8

I found **19 neighborhoods which are located in less than 300m from a parking area.** I filtered out the neighborhoods with more than 1 parking area within 300m and I finally found **11 neighborhoods candidate for clustering.** I joined the 11 selected neighborhoods to the neighborhoods dataframe.

[24]:

	Neighborhood	Parking	Zip	Latitude	Longitude	RE_index
0	Susa	1	20133	45.467484	9.223430	4.792
1	Sempione	1	20154	45.477104	9.170177	5.984
2	Quadronno	1	20136	45.454182	9.191743	8.465
3	Moscova	1	20121	45.477091	9.184342	8.745
4	Palestro	1	20121	45.471006	9.201477	8.465
5	Cenisio	1	20154	45.487565	9.172265	5.647
6	Isola	1	20159	45.487565	9.188972	5.647

I represented on a map the finally selected neighborhoods (blue circles) and the parking areas (red dots):



Foursquare.

After finding the **neighborhoods candidates**, I used the **Foursquare API** to get the venues for each of the 11 selected neighborhoods:

The **Foursquare API** found 705 venues.

```
[48]: Milan_venues
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Susa	45.467484	9.223430	La Ciribiciaccola	45.469540	9.223775	Dessert Shop
1	Susa	45.467484	9.223430	Locanda del Menarost	45.465184	9.221471	Italian Restaurant
2	Susa	45.467484	9.223430	Pizzeria Ciak	45.467878	9.225039	Pizza Place
3	Susa	45.467484	9.223430	Fujiyama	45.469193	9.223766	Sushi Restaurant
4	Susa	45.467484	9.223430	Ristorante Il Grissino	45.471248	9.221944	Italian Restaurant
...	...	...	...	...	...	...	...
700	Ticinese	45.450596	9.181951	Ristorante Da Teresa	45.447448	9.178658	Italian Restaurant
701	Ticinese	45.450596	9.181951	Parco Via Tabacchi	45.447163	9.182943	Park
702	Ticinese	45.450596	9.181951	Loop	45.448168	9.186667	Café
703	Ticinese	45.450596	9.181951	Prince Cafè	45.448768	9.177262	Cocktail Bar
704	Ticinese	45.450596	9.181951	Q.P. Chiùpizza	45.447493	9.178139	Pizza Place

705 rows × 7 columns

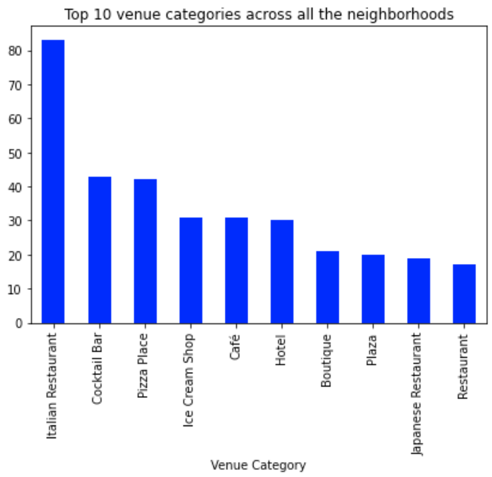
```
[49]: print('There are {} uniques categories.'.format(len(Milan_venues['Venue Category'].unique())))
```

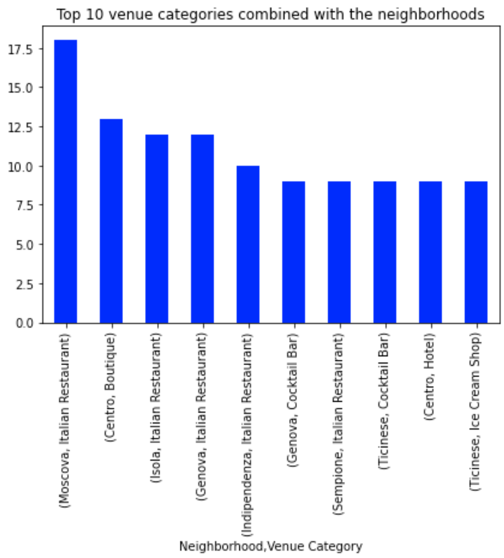
```
There are 147 uniques categories.
```

I finally found **147 unique venue categories**, which looks ok for clustering.

Analysis

I sorted out **the top 10 venue categories**, and the **top 10 neighborhood/venue category combinations**. Food dominates with Italian restaurants and cocktail bars, but we also have hotels and boutiques. We should be able to cluster **neighborhoods characterized by lifestyle, shopping, and hospitality districts.**





I used the **onehot encoding** to obtain the dataframe for clustering, and I calculated the 5 top common venues for each of the 11 selected neighborhoods.

[36]:

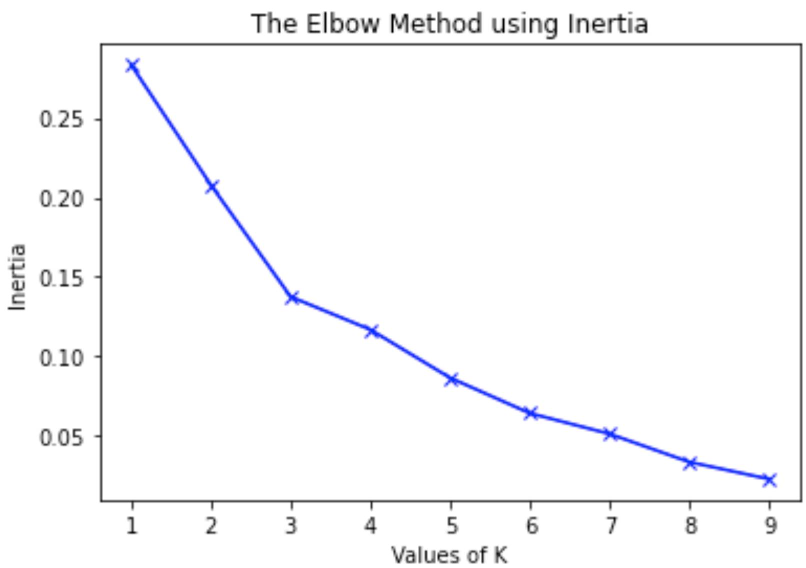
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Cenisio	Pizza Place	Hotel	Gym / Fitness Center	Thai Restaurant	Seafood Restaurant
1	Centro	Boutique	Hotel	Italian Restaurant	Plaza	Ice Cream Shop
2	Genova	Italian Restaurant	Cocktail Bar	Pizza Place	Seafood Restaurant	Bar
3	Indipendenza	Italian Restaurant	Ice Cream Shop	Pizza Place	Café	Dessert Shop
4	Isola	Italian Restaurant	Pizza Place	Café	Bistro	Cocktail Bar
5	Moscova	Italian Restaurant	Ice Cream Shop	Wine Bar	Café	Restaurant
6	Palestro	Boutique	Cocktail Bar	Women's Store	Plaza	Art Gallery
7	Quadronno	Wine Bar	Hotel	Pizza Place	Salad Place	Café
8	Sempione	Italian Restaurant	Pizza Place	Cocktail Bar	Japanese Restaurant	Ice Cream Shop
9	Susa	Italian Restaurant	Hotel	Ice Cream Shop	Café	Convenience Store
10	Ticinese	Ice Cream Shop	Cocktail Bar	Italian Restaurant	Café	Pizza Place

Clustering

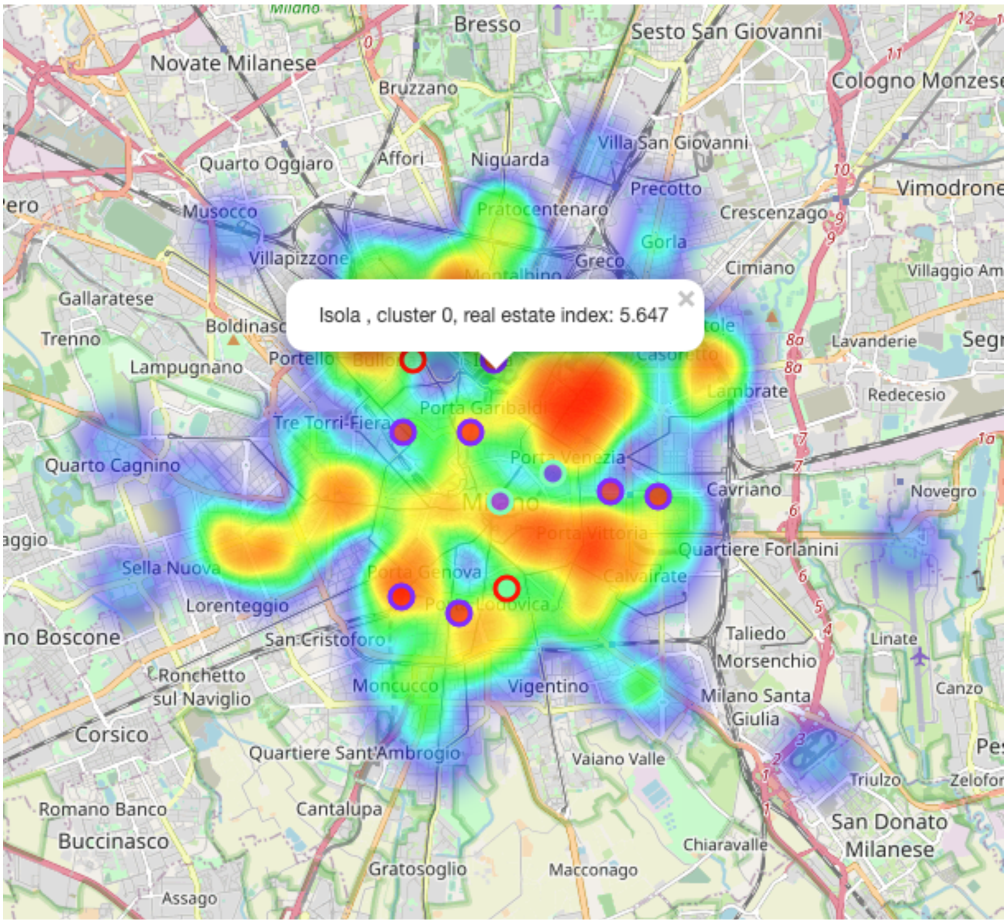
I used **GridSearchCV** to find the optinal parameters for K-means.

```
DeprecationWarning)
[38]: {'init': 'random', 'n_clusters': 3, 'random_state': 0}
```

To confirm the number of clusters, I used the **Elbow method and inertia**:



I represented the clustered neighborhoods on a **folium** map and a **heatmap** to show the density of parking areas for car sharing. Each neighborhood shows a popup to indicate the name of the neighborhood, the cluster to which it belongs, and the real estate index.



This is the detail of the three clusters:

```
[54]: Cluster0=df_final_venues.loc[df_final_venues['Cluster Labels'] == 0, df_final_venues.columns[[0]+ list(range(3, df_final_venues.shape[1]))]]
Cluster0
```

[54]:

	Neighborhood	Latitude	Longitude	RE_index	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Susa	45.467484	9.223430	4.792	0	Italian Restaurant	Hotel	Ice Cream Shop	Café	Convenience Store
1	Sempione	45.477104	9.170177	5.984	0	Italian Restaurant	Pizza Place	Cocktail Bar	Japanese Restaurant	Ice Cream Shop
3	Moscova	45.477091	9.184342	8.745	0	Italian Restaurant	Ice Cream Shop	Wine Bar	Café	Restaurant
6	Isola	45.487565	9.188972	5.647	0	Italian Restaurant	Pizza Place	Café	Bistro	Cocktail Bar
7	Indipendenza	45.468429	9.213323	6.622	0	Italian Restaurant	Ice Cream Shop	Pizza Place	Café	Dessert Shop
8	Genova	45.452879	9.169715	7.534	0	Italian Restaurant	Cocktail Bar	Pizza Place	Seafood Restaurant	Bar
10	Ticinese	45.450596	9.181951	7.534	0	Ice Cream Shop	Cocktail Bar	Italian Restaurant	Café	Pizza Place

```
[55]: Cluster1=df_final_venues.loc[df_final_venues['Cluster Labels'] == 1, df_final_venues.columns[[0]+ list(range(3, df_final_venues.shape[1]))]]
Cluster1
```

[55]:

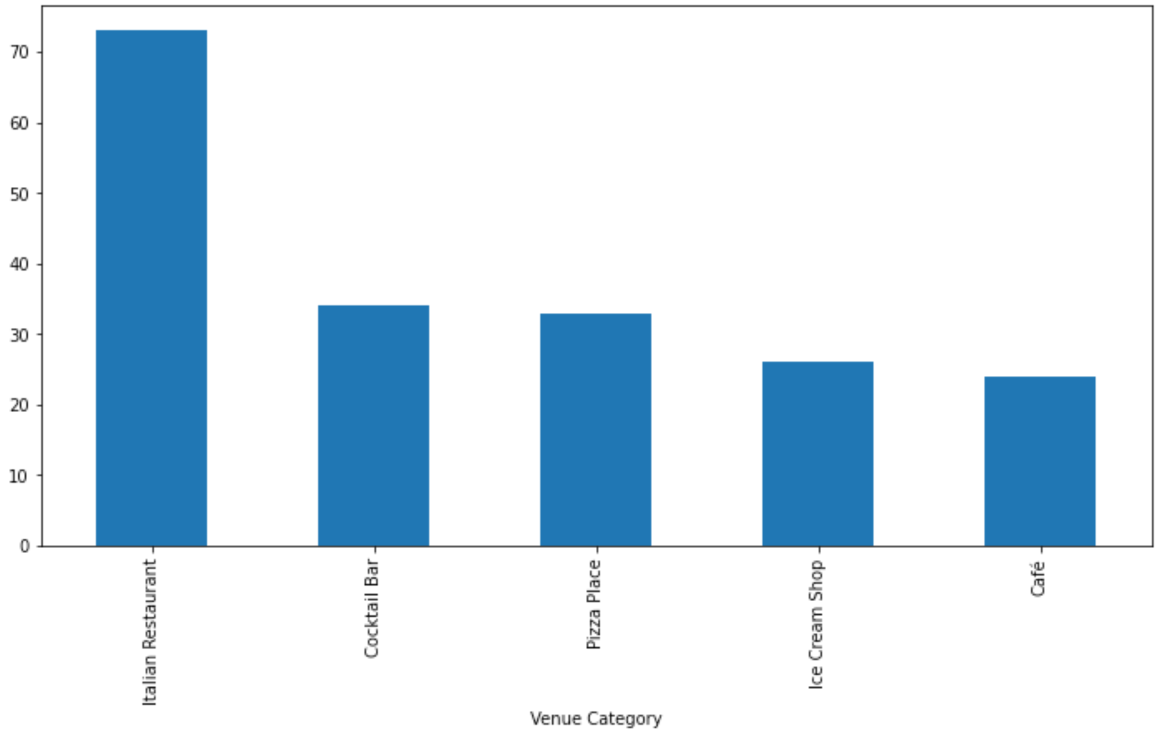
	Neighborhood	Latitude	Longitude	RE_index	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
4	Palestro	45.471006	9.201477	8.465	1	Boutique	Cocktail Bar	Women's Store	Plaza	Art Gallery
9	Centro	45.466800	9.190500	9.947	1	Boutique	Hotel	Italian Restaurant	Plaza	Ice Cream Shop

```
[56]: Cluster2=df_final_venues.loc[df_final_venues['Cluster Labels'] == 2, df_final_venues.columns[[0]+ list(range(3, df_final_venues.shape[1]))]]
Cluster2
```

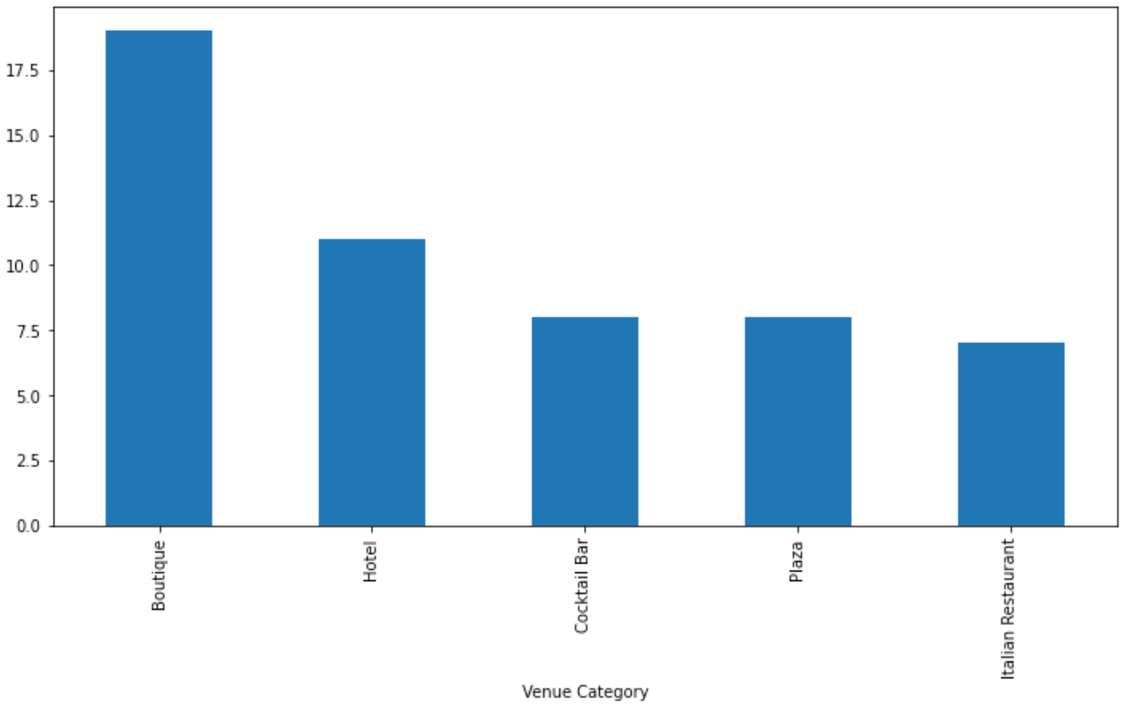
[56]:

	Neighborhood	Latitude	Longitude	RE_index	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
2	Quadronno	45.454182	9.191743	8.465	2	Wine Bar	Hotel	Pizza Place	Salad Place	Café
5	Cenisio	45.487565	9.172265	5.647	2	Pizza Place	Hotel	Gym / Fitness Center	Thai Restaurant	Seafood Restaurant

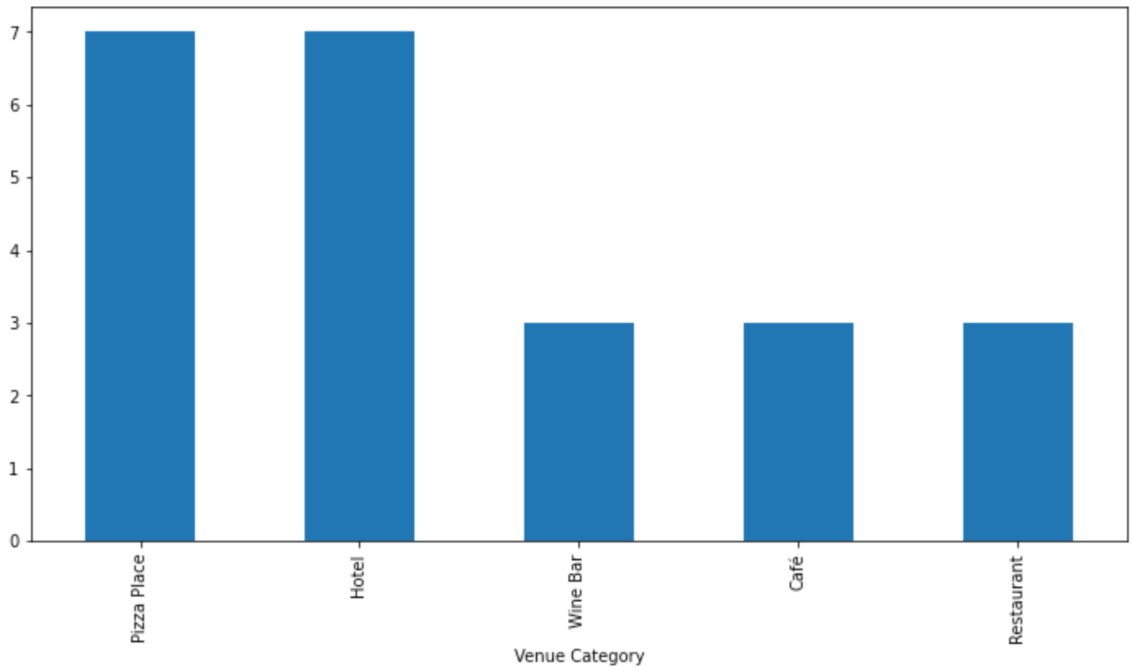
Cluster 0 is primarily a **lifestyle area.** Looking at the map, the neighborhoods *Sempione*, *Isola*, *Indipendenza* and *Susa* are suburbs and show a lower real estate index. *Moscova* is central and expensive, *Porta Genova* and *Ticinese* are still central and lifestyle.



Cluster 1 is primarily a **shopping district**, it's smaller in terms of venues that cluster Zero. The neighborhoods are clearly central and have a high real estate index.



Cluster 2 is primarily a **suburb hotels area**, it's smaller in terms of venues than cluster Zero and One. The neighborhoods are different because *Cenisio* is suburbs and has a low real estate index, while *Quadronno* is still central and has a high real estate index.



4 Results and discussion

The analysis shows that although parking areas for car sharing are present in the center of Milan (~3Km), there are zones of low parking spot density, no more than 1 in range of 300m, fairly close to city center. The Highest concentration of parking areas was detected East and close to the city center. I focused my attention to areas **East and less close to the city center (*Indipendenza, Susa*)** because they also **combine a lower real estate value**. Similarly, other neighborhoods were identified as potentially interesting **North from the city center (*Cenisio, Sempione, Isola*).**

The neighborhood of *Cenisio* is close to hotels and intercepts traffic arriving from North, North West (Malpensa airport, directions arriving from the A8 and A9 highways -Switzerland). The neighborhood of *Sempione* and *Isola* (lifestyle) are near *Cenisio*, and *Sempione* is actually slightly closer to the city center. The neighborhoods of *Indipendenza* and *Susa*, both lifestyle, intercept traffic arriving from East (Linate airport, Bergamo airport and directions along the A4 highway -Venice).

Those neighborhoods candidates have been clustered to create zones of interest. They will become starting points for more detailed local analysis based on other factors. As I learned during the course, this does not imply that those zones are actually optimal locations for a new parking area. It is possible that there is a very good reason for small number of parking for car sharing in any of those areas, regardless of lack of competition in the area. Recommended zones should therefore be considered only as a starting point for more detailed analysis which could eventually result in locations which have not only no nearby competition, but also other factors taken into account and all other relevant conditions met.

5 Conclusions

Purpose of this project was to identify Milan **areas within 3Km from the city center with no more than 1 parking area for car sharing within a 300m radius**. This will help the stakeholders to narrow down the search for optimal location for new parking areas. Clustering of those neighborhoods was then performed in order to create major zones of interest which will be used as starting points for final exploration by stakeholders.

A final decision on optimal parking areas location will be made by stakeholders based on specific characteristics of neighborhoods in every recommended zone, taking into consideration additional factors.