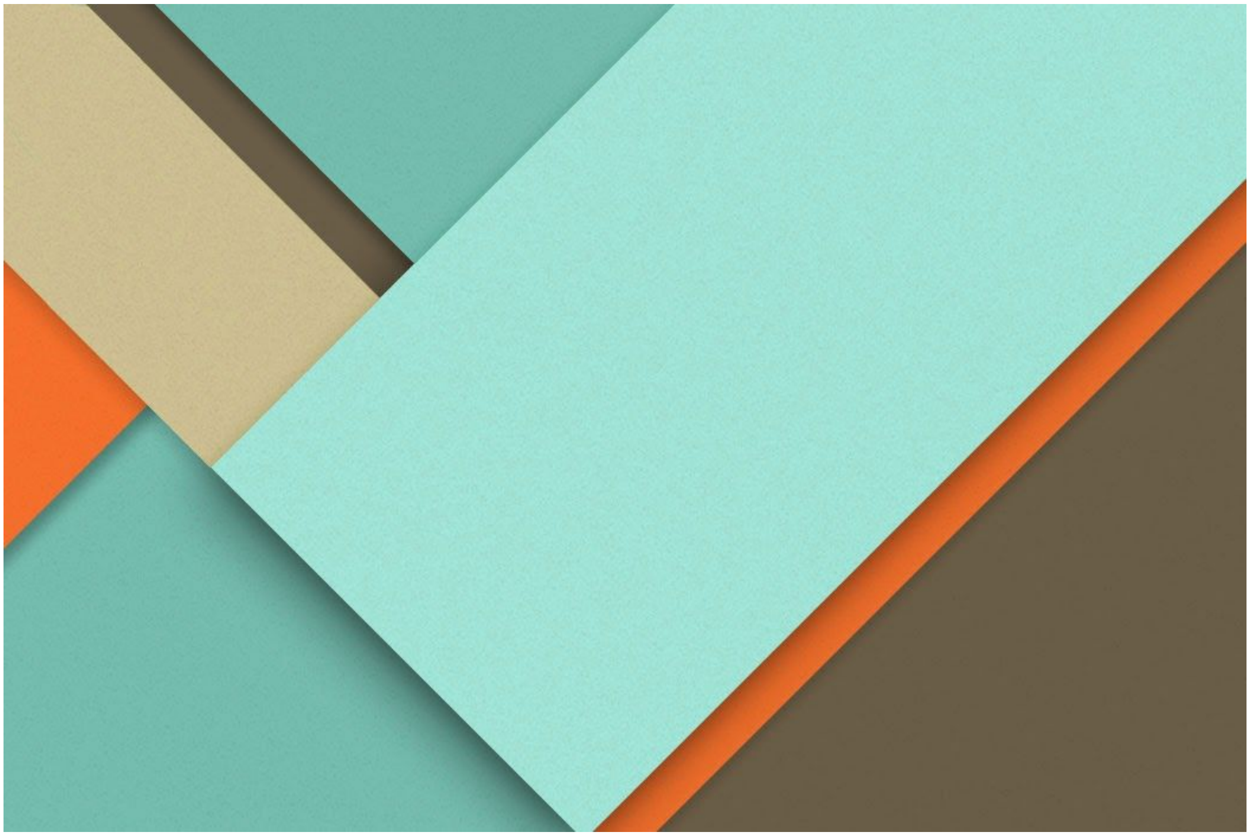


Coursera Capstone Project - Fashion Shops in Milan

By Annick Mekolle

May 2020

1. Description of the problem and a discussion of the background

A. INTRODUCTION / BUSINESS PROBLEM:

ANALYSIS OF DATA VENUES IN MILAN FOR THE FASHION INDUSTRY

Milan, in Italian, Milano, is a city in northern Italy, capital of Lombardy, and the second-most populous city in Italy after Rome. Milan served as the capital of the Western Roman Empire, the Duchy of Milan and the Kingdom of Lombardy–Venetia. The city proper has a population of about 1.4 million while its metropolitan city has 3.26 million inhabitants. It is a continuously built-up urban area, that stretches well beyond the boundaries of the administrative metropolitan city, is the fourth largest in the EU with 5.27 million inhabitants. The population within the wider Milan metropolitan area, also known as Greater Milan, is estimated at 8.2 million, making it by far the largest metropolitan area in Italy and the 4th largest in the EU.

Milan is considered a leading alpha global city, with strengths in the field of the art, commerce, design, education, entertainment, fashion, finance, healthcare, media, services, research and tourism. Its business district hosts Italy's stock exchange, and the headquarters of national and international banks and companies. In terms of GDP, it has the second-largest economy among EU cities after Paris, and is the wealthiest among EU non-capital cities. Milan is considered part of the Blue Banana and one of the "Four Motors for Europe".

Fashion and design

The city has been recognized as one of the world's four fashion capitals thanks to several international events and fairs, including Milan Fashion Week and the Milan Furniture Fair, which are currently among the world's biggest in terms of revenue, visitors and growth. It hosted the Universal Exposition in 1906 and 2015.

The city hosts numerous cultural institutions, academies and universities, with 11% of the national total enrolled students. Milan is the destination of *8 million overseas visitors every year*, attracted by its museums and art galleries that include some of the most important collections in the world, including major works by Leonardo da Vinci. The city is served by many luxury hotels and is the fifth-most starred in the world by Michelin Guide. The city is home to two of Europe's most successful football teams, A.C. Milan and F.C. Internazionale, and one of Europe's main basketball teams, Olimpia Milano. Milan will host the 2026 Winter Olympics together with Cortina d'Ampezzo.

Milan is widely regarded as a global capital in industrial design, fashion and architecture. In the 1950s and 60s, as the main industrial centre of Italy and one of Europe's most dynamic cities, Milan became a world capital of design and architecture. There was such a revolutionary change that Milan's fashion exports accounted for 726 million US dollars in 1952, and by 1955 that number grew to US$72.5 billion.

Modern skyscrapers, such as the Pirelli Tower and the Torre Velasca were built, and artists such as Bruno Munari, Lucio Fontana, Enrico Castellani and Piero Manzoni gathered in the city.

Today, Milan is still particularly well known for its high-quality furniture and interior design industry. The city is home to FieraMilano, Europe's largest permanent trade exhibition, and Salone Internazionale del Mobile, one of the most prestigious international furniture and design fairs.

Milan is also regarded as one of the fashion capitals of the world, along with New York City, Paris, and London. Milan is synonymous with the Italian prêt-à-porter industry, as many of the most famous Italian fashion brands, such as ***Valentino, Gucci, Versace, Prada, Armani and Dolce & Gabbana,*** are headquartered in the city.

Numerous international fashion labels also operate shops in Milan.

Furthermore, the city hosts the Milan Fashion Week twice a year, one of the most important events in the international fashion system.

Target Audience

The fashion industry: Anyone wishing to be recognized on a large scale in fashion, Milan would be one of the best places to consider.

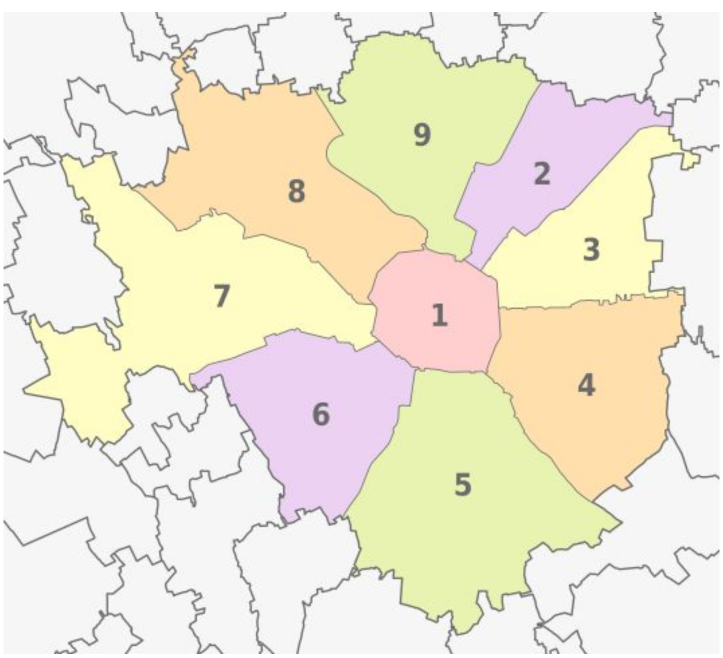
I have lived in Italy for close to two years and I have been very fascinated by the fashion industry all along and more especially after visiting Milan, whose energy is bustling high in this domain.

In this project, I will be working on Milan data and exploiting the neighborhoods in Milan, known as Districts, to find the most popular areas that would be good places to consider setting up a fashion shop or fashion industry, for maximum output.

B. SOURCE AND DESCRIPTION OF THE DATA

The data and tools I worked to solve the problem:

- Data of the municipalities (or boroughs) in Milan from Wikipedia. This data includes all nine municipalities, officially numbered from 1 to 9, their names, the area they cover in kilometre squared (km2), their population as of 2014 and the population density (inhabitants/km2) and their respective districts. Milan has nine (09) official Boroughs (in Italian, *Municipi*), numbered from 1 to 9, that is Borough 1, Borough 2, etc. The province of Milan covers an area of about 181.76 km2.



Original data extracted:

	Borough	Name	Area (km2)	Population(2014)	Population Density (inhabitants/km2)	Districts
0	1	Centro storico	9.67	96,315	11,074	Brera
1	2	Stazione Centrale	12.58	153.109	13,031	Adriano
2	3	Città Studi	14.23	141,229	10,785	Casoretto
3	4	Porta Vittoria	20.95	156.369	8,069	Acquabella
4	5	Vigentino	29.87	123,779	4,487	Basmetto
5	6	Barona	18.28	149,000	8,998	Arzaga
6	7	Baggio	31.34	170,814	6,093	Assiano
7	8	Fiera	23.72	181,669	8,326	Boldinasco
8	9	Porta Garibaldi	21.12	181,598	9,204	Affori

- Geopy's Nominatim, which a tool to search OSM data by name and address (geocoding) and to generate synthetic addresses of OSM points (reverse geocoding), in other words, to retrieve the coordinates of each district.

- Foursquare API to get the most common venues in the respective districts of Milan, enabling me to know what spots are most popular for our business. I used these popular spots to build clusters and from exploratory analysis was able to determine the best clusters and ultimately the perfect boroughs and districts that will be perfect for our business.

- Folium for visualization of my data on an interactive leaflet map, visualization of the clusters and the popular spots.

- Pandas and numpy libraries for exploratory analysis, as well as others like BeautifulSoup, Matplotlib, and Scikit learn.

2. METHODOLOGY

A. Exploratory Data Analysis: Machine Learning and Inferential Statistical Testing

I used Geopy Nominatim to get the latitudes and longitudes of each of the districts, then Foursquare API to retrieve the popular venues. The limit I used for the number of venues was 100 within a radius of 500. Milan has geographical coordinates with *latitude 45.4668* and *longitude 9.1905*.

Bar chart of the Number of Districts per Borough. 145 districts were extracted from the original data:

```
Text(0.5, 0, 'Borough')
```

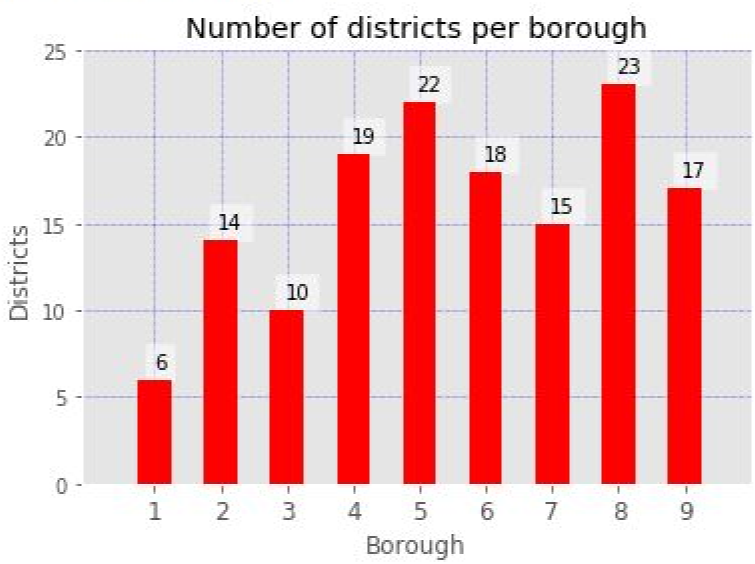


Figure 1

So, we see in Borough 1 there are 6 Districts; for Borough 2, 14 Districts and so on.

1. Cleaning and scraping of data:

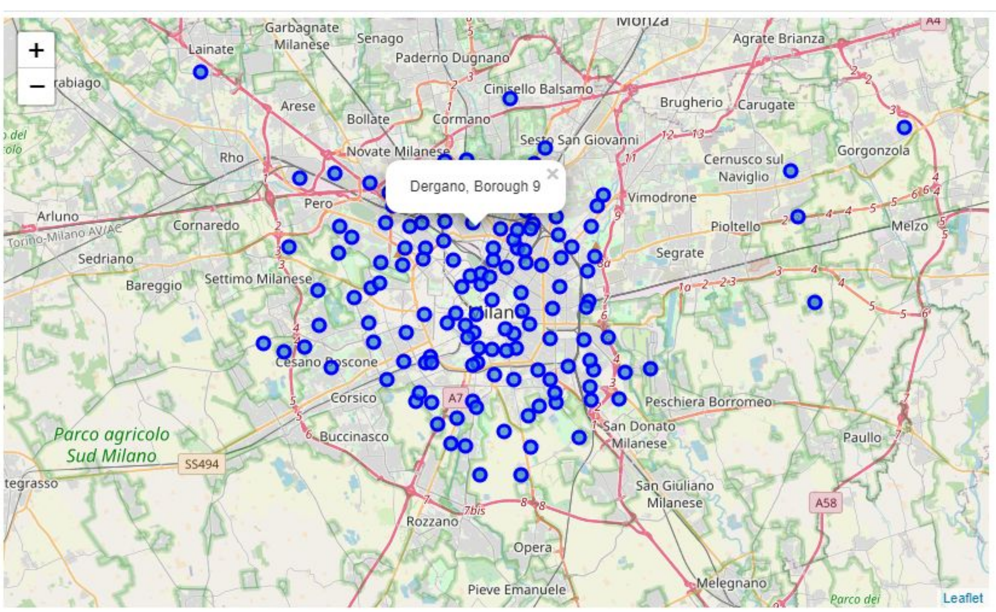
Cleaning and selection of data was done at every level, because the aim was to scale down the data as much as possible to have the best results. I made careful comparisons every time between the boroughs, the popular venues, the clusters, almost like a competition where candidates had to be disqualified at every stage.

Many of the districts returned were faulty in their names and others just did not exist in Milan. I scraped all of those redundancies and adjusted some of the district names, based on information from Wikipedia, while deleting the non-existent ones. At the end of that exercise I was left with 143 districts to work with.

2. Viusalizations

	Borough	Name	Districts	Latitude	Longitude
0	1	(Centro storico,)	Brera	45.471519	9.187735
1	1	(Centro storico,)	Centro Storico	45.444613	9.096351
2	1	(Centro storico,)	Conca del Naviglio	45.458560	9.177745
3	1	(Centro storico,)	Guastalla	45.458252	9.200023
4	1	(Centro storico,)	Porta Sempione	45.477128	9.170598
5	1	(Centro storico,)	Porta Tenaglia	45.477821	9.181593
6	2	(Stazione di Milano Centrale, Gorla, Turro, Gr...	Adriano	45.513572	9.251202
7	2	(Stazione di Milano Centrale, Gorla, Turro, Gr...	Crescenzago	45.509219	9.247484
8	2	(Stazione di Milano Centrale, Gorla, Turro, Gr...	Gorla	45.504945	9.224539
9	2	(Stazione di Milano Centrale, Gorla, Turro, Gr...	Greco	45.502184	9.211233

Tab 1: Dataframe of the first 10 districts with their geographical coordinates



Map 1

The geographical positions of the districts on the map are represented by the blue circles. The popup is an example of a district and the borough or municipality to which it belongs:

With Foursquare API, the venues were returned as follows (first 5 venues displayed here). Altogether, a total of 3682 venues which are of 290 distinct venue categories:

	Districts	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Borough
0	Brera	45.471519	9.187735	BVLGARI Hotel Milano	45.470621	9.189779	Hotel	1
1	Brera	45.471519	9.187735	Bulgari Lounge Bar	45.470014	9.188943	Cocktail Bar	1
2	Brera	45.471519	9.187735	Di Viole Di Liquirizia	45.471460	9.185336	Cupcake Shop	1
3	Brera	45.471519	9.187735	Pinacoteca di Brera	45.471979	9.188128	Art Museum	1
4	Brera	45.471519	9.187735	Palazzo di Brera	45.472019	9.188043	College Arts Building	1
5	Brera	45.471519	9.187735	Rigadritto	45.469721	9.188008	Gift Shop	1
6	Brera	45.471519	9.187735	Il Giardino	45.469989	9.188872	Lounge	1
7	Brera	45.471519	9.187735	SUSHI B	45.472153	9.186883	Japanese Restaurant	1
8	Brera	45.471519	9.187735	Bvlgari Ristorante	45.470005	9.188943	Restaurant	1
9	Brera	45.471519	9.187735	Piazza del Carmine	45.470102	9.185058	Plaza	1

Tab 2: Venues, their categories and their geographical coordinates

Below, we see as highlighted that the majority of popular venue categories consists of Italian Restaurants, Cafés, Pizza Places, Hotels and Ice Cream Shops, etc. Also, we have shops and stores that we will take into consideration, since that is our focus: popularity in terms of well-known boroughs and districts, that also have a large number of stores and shops. So it seems good that in our final choice of clusters to fit our analysis will be the ones that are characterized by these categories.

Venue Category	Districts	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude
Italian Restaurant	408	408	408	408	408	408
Pizza Place	259	259	259	259	259	259
Café	252	252	252	252	252	252
Hotel	130	130	130	130	130	130
Ice Cream Shop	122	122	122	122	122	122
Restaurant	110	110	110	110	110	110
Tram Station	95	95	95	95	95	95
Japanese Restaurant	90	90	90	90	90	90
Plaza	83	83	83	83	83	83
Cocktail Bar	82	82	82	82	82	82

Tab 3: Venue Categories returned and their quantity

Anything stores, shops, including shopping malls, clothing shops,etc, sorted by the number of venues.

	Venue Category	Borough	Number Venues
0	Ice Cream Shop	1	24
1	Ice Cream Shop	6	23
2	Ice Cream Shop	2	14
3	Ice Cream Shop	9	14
4	Coffee Shop	9	13
5	Ice Cream Shop	3	11
6	Ice Cream Shop	4	9
7	Dessert Shop	3	8
8	Clothing Store	8	7
9	Dessert Shop	4	7
10	Dessert Shop	1	7
11	Coffee Shop	2	6
12	Ice Cream Shop	8	6
13	Furniture / Home Store	9	6
14	Electronics Store	6	5
15	Sporting Goods Shop	1	5

Tab 4: Shops, stores and shopping malls

Bar chart of number of popular venues per borough:

```
Text(0.5, 0, 'Borough')
```

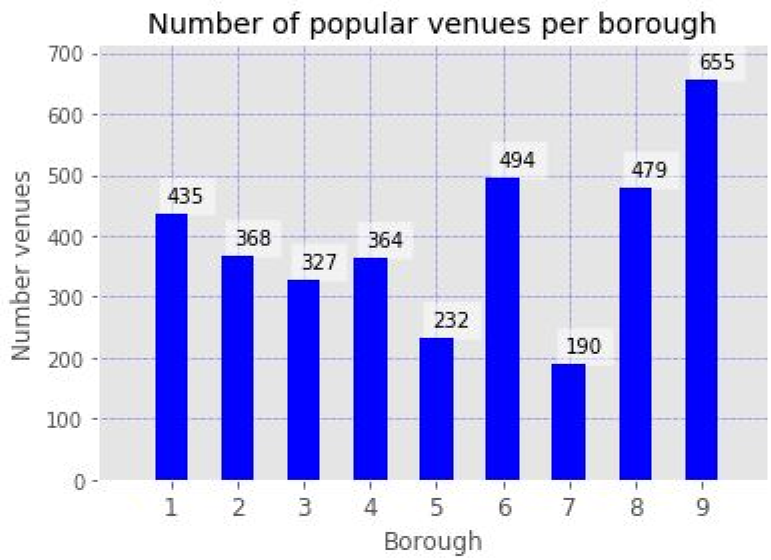


Figure 2

Since our aim here is to position our fashion industry in the Boroughs with high popularity, I decided to drop the least 2 popular boroughs, which are Boroughs 5 and 7, with 232 and 190 popular venues respectively. This will bring down our data to 3122 venues to cluster and 104 districts. Boroughs 1, 6, 8 and 9 have the most popular spots.

B. Cluster Segmentation

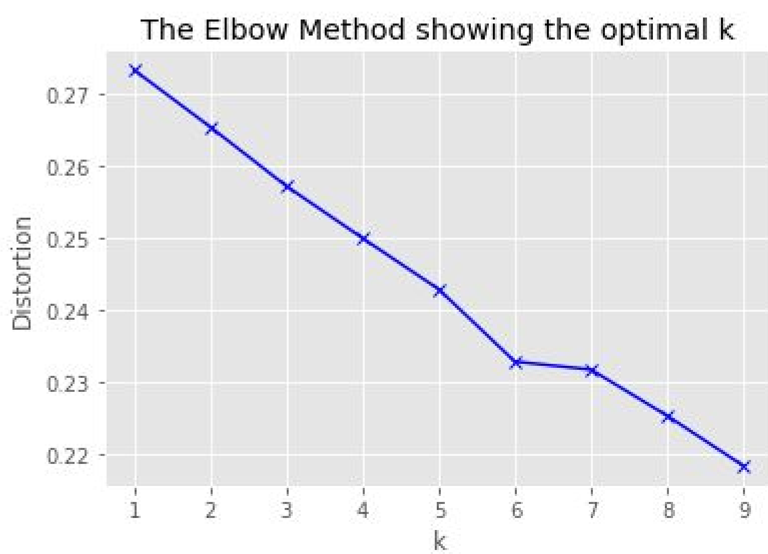
K-means clustering from Scikit Learn is an unsupervised machine learning method of vector quantization, originally from signal processing, that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean (cluster centers or cluster centroid), serving as a prototype of the cluster. It is popular for cluster analysis in data mining.

In view of clustering, I narrowed down my machine learning algorithm to take into consideration the 10 most common venues for each district. Hence, the dataset below showing the first few districts and their 10 top venues:

	Districts	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Acquabella	Hotel	Italian Restaurant	Plaza	Ice Cream Shop	Pizza Place	Café	Cocktail Bar	Japanese Restaurant	Seafood Restaurant	Bookstore
1	Adriano	Italian Restaurant	Trattoria/Osteria	Ice Cream Shop	Toy / Game Store	Filipino Restaurant	Emilia Restaurant	Event Space	Falafel Restaurant	Farm	Fast Food Restaurant
2	Affori	Italian Restaurant	Café	Pizza Place	Cocktail Bar	Fried Chicken Joint	Park	Health Food Store	Salon / Barbershop	Gym	Pool Hall
3	Arzaga	Pizza Place	Pub	Café	Pharmacy	Martial Arts Dojo	Athletics & Sports	Diner	Hobby Shop	Japanese Restaurant	Ice Cream Shop
4	Barona	Soccer Field	Trattoria/Osteria	Theater	Athletics & Sports	Café	Food	Flower Shop	Flea Market	Fish & Chips Shop	Emilia Restaurant

Tab 5: Districts and their top 10 venues

After trying out the elbow method on several occasions, which produced several distortions each time that were a little unclear, though showing the elbow method most of the time at k=6, I decided to choose k=6 as the optimum value.



Text(0.5, 0, 'Cluster')

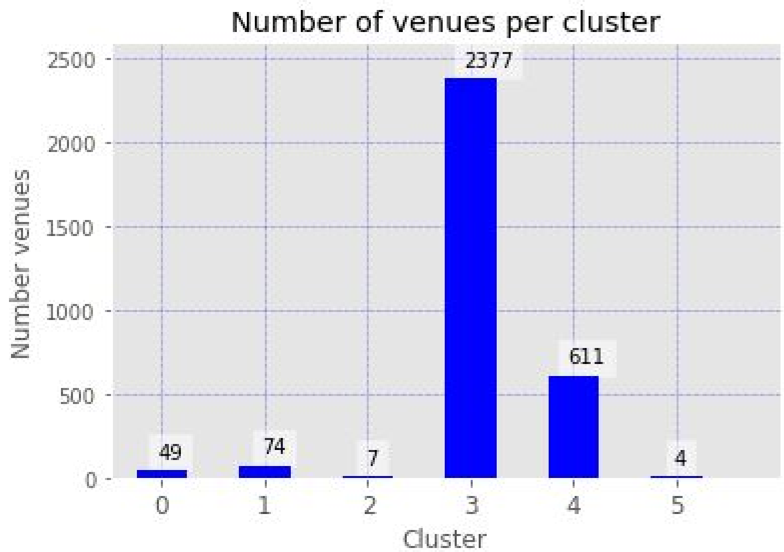


Figure 3

After applying k-Means to the data, the result displayed 6 clusters with popularity as shown below:

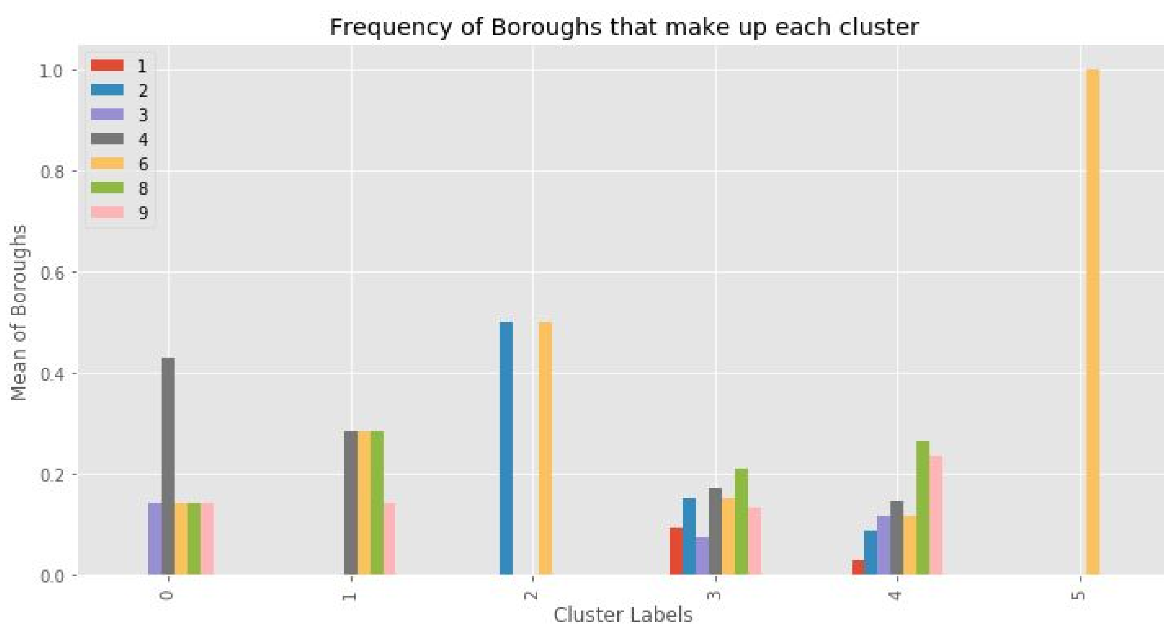


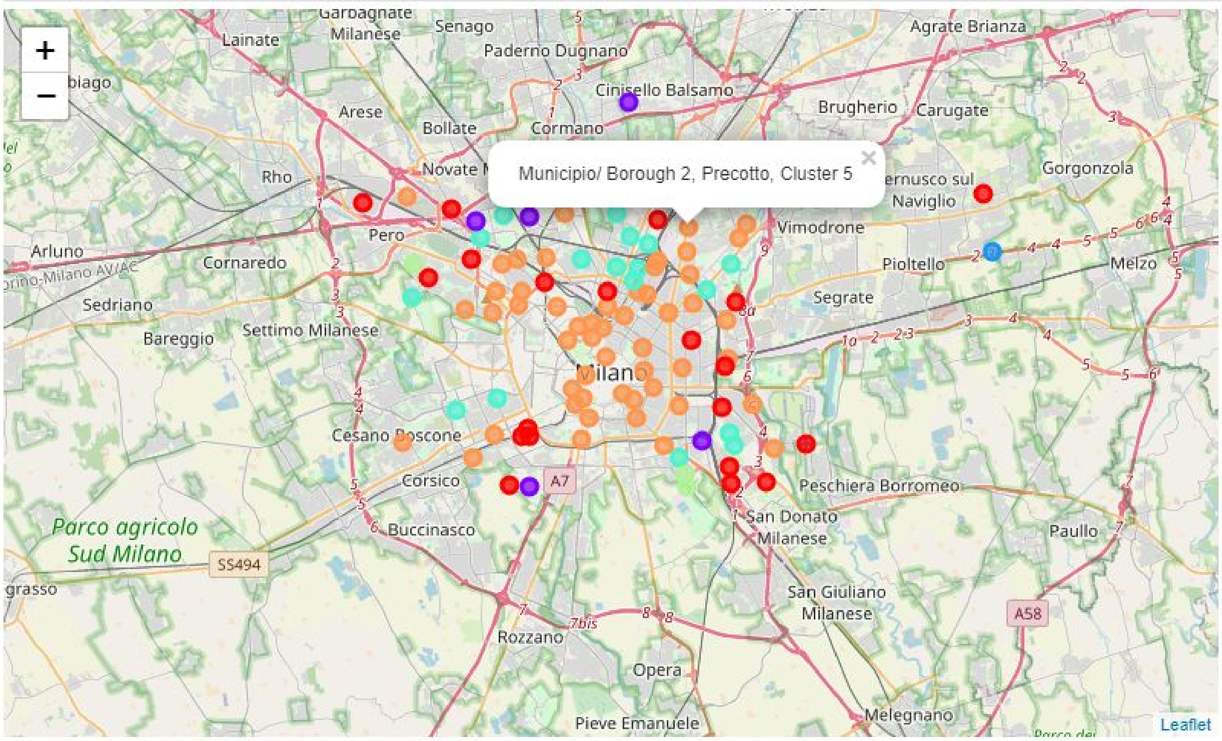
Figure 4

Figure 4 shows the mean of each borough per cluster, which signifies the quantity, like the measurement of each ingredient to make up a cake. The cake in this case is the cluster. For example, for cluster label 5 in *Figure 3*, it has just 4 venues, of which *Figure 4* shows as venues belonging only to Borough 6. Borough 6 is the orange bar as explained in the legend on the top left corner of *Figure 4*. So, cluster 5 is made up of only one borough.

We also observe that clusters 0, 1, 2 and 5 have very low popularity, with 49, 74, 7 and 4. So we will drop them. We will focus on clusters 3 and 4. I will now go further to identify the 3 best boroughs for our project.

Cluster 3 is an excellent source of information for our project, having 2377 venues of the 3122 venues retrieved from Foursquare.

Let's display the clusters on a folium map. We observe that there are several circles in several colors. Each of the colors represents a different cluster.



Map 2

Here are a few examples of the clusters, numbers 0 and 3 respectively with a few of their venues shown:

1. Cluster 0

```
milan_merged.loc[milan_merged['Cluster Labels'] == 0, milan_merged.columns[[0,2] + list(range(5, milan_merged.shape
```

	Borough	Districts	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th M Comi Ve
27	3	Quartiere Feltre	0	Football Stadium	Ice Cream Shop	Sports Club	Skate Park	Fast Food Restaurant	Supermarket	Asian Restaurant	Soccer Field	Fi C S
39	4	Omero	0	Soccer Field	Bakery	Italian Restaurant	Hotel	Pizza Place	Sports Club	Dance Studio	Deli / Bodega	Foo Stac
46	4	Taliedo	0	Pizza Place	Garden	Soccer Field	Bus Stop	Yoga Studio	Fish & Chips Shop	Farmers Market	Fast Food Restaurant	Fili Restau
47	4	Triulzo Superiore	0	Breakfast Spot	Bus Station	Fast Food Restaurant	Soccer Field	Filipino Restaurant	Fish & Chips Shop	Flea Market	Flower Shop	F
82	6	Ronchetto sul Naviglio	0	Italian Restaurant	Electronics Store	Park	Soccer Field	Bus Stop	Noodle House	Tram Station	Train Station	De S
121	8	San Leonardo	0	Soccer Field	Ice Cream Shop	Hotel	Athletics & Sports	Park	Metro Station	Yoga Studio	Farmers Market	Fast F Restau
129	9	Bovisasca	0	Soccer Field	Italian Restaurant	Supermarket	Chinese Restaurant	Bus Stop	Park	Art Gallery	Restaurant	Shop F

Tab 6: Cluster 0

4. Cluster 3

```
milan_merged.loc[milan_merged['Cluster Labels'] == 3, milan_merged.columns[[0,2] + list(range(5, milan_merged.shape
```

	Borough	Districts	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th M Comr Ve
0	1	Brera	3	Italian Restaurant	Hotel	Ice Cream Shop	Japanese Restaurant	Café	Art Museum	Plaza	Restau
2	1	Conca del Naviglio	3	Italian Restaurant	Café	Ice Cream Shop	Pizza Place	Bistro	Seafood Restaurant	Restaurant	Cocktail
3	1	Guastalla	3	Italian Restaurant	Pizza Place	Tram Station	Ice Cream Shop	Bakery	Restaurant	Japanese Restaurant	
4	1	Porta Sempione	3	Italian Restaurant	Pizza Place	Cocktail Bar	Japanese Restaurant	Bakery	Ice Cream Shop	Gastropub	Burger J
5	1	Porta Tenaglia	3	Italian Restaurant	Japanese Restaurant	Chinese Restaurant	Café	Sushi Restaurant	Wine Bar	Ice Cream Shop	Seaf Restau

Tab 7: Cluster 3

Cluster Labels	Borough	Districts
3	9	467
3	1	422
3	8	366
3	6	347
3	2	318
3	4	259
3	3	198
4	9	144
4	6	124
4	3	121
4	8	89
4	4	75
4	2	45
1	9	34
1	8	17
0	4	16
1	4	14
4	1	13
0	9	10
1	6	9
0	3	8
0	6	8
0	8	7
2	2	5
5	6	4
2	6	2

Tab 8: Number of venues returned for each borough in each of the clusters

We see from this dataframe that Boroughs 1, 6, 8 and 9 have very high popularity, which provides the results to our project.

Cluster Labels	Venue Category	
3	Italian Restaurant	305
	Café	142
	Pizza Place	120
	Hotel	96
4	Pizza Place	94
3	Ice Cream Shop	86
	Restaurant	76
	Japanese Restaurant	67
	Cocktail Bar	64
	Plaza	59

Tab 9: Venue Categories and Popularity

Popularity of venue categories according to the clusters, sorted according to number of venues per category, for example, cluster 3 has 305 restaurants, 142 cafés, etc.

After dropping unpopular clusters, we are left with 2 clusters. For each of them, the dataset below displays the boroughs contained and the numbers of popular venues returned for each of the boroughs. For example, borough 9 has 467 venues in cluster 3 and 144 venues in cluster 4 and so on.

Cluster Labels	Borough	Total Popular Venues per Borough on each cluster
3	9	467
3	1	422
3	8	366
3	6	347
3	2	318
3	4	259
3	3	198
4	9	144
4	6	124
4	3	121
4	8	89
4	4	75
4	2	45
4	1	13

Tab 10: Number of venues per borough on each cluster

So again, we have our beautiful bar chart, displaying the frequency of the boroughs in the 3 best clusters:

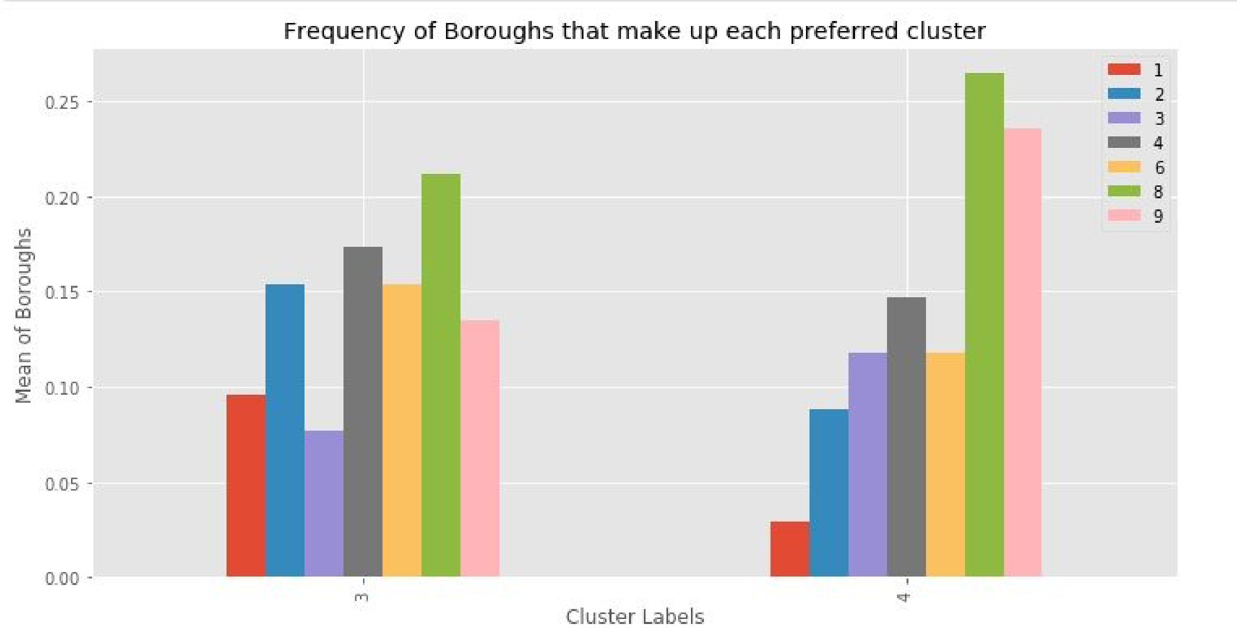


Figure 5

3. RESULT SECTION

Renaming the clusters

- **Cluster 3** : "*1st Preferred Cluster*"
- **Cluster 4** : "*2nd Preferred Cluster*"

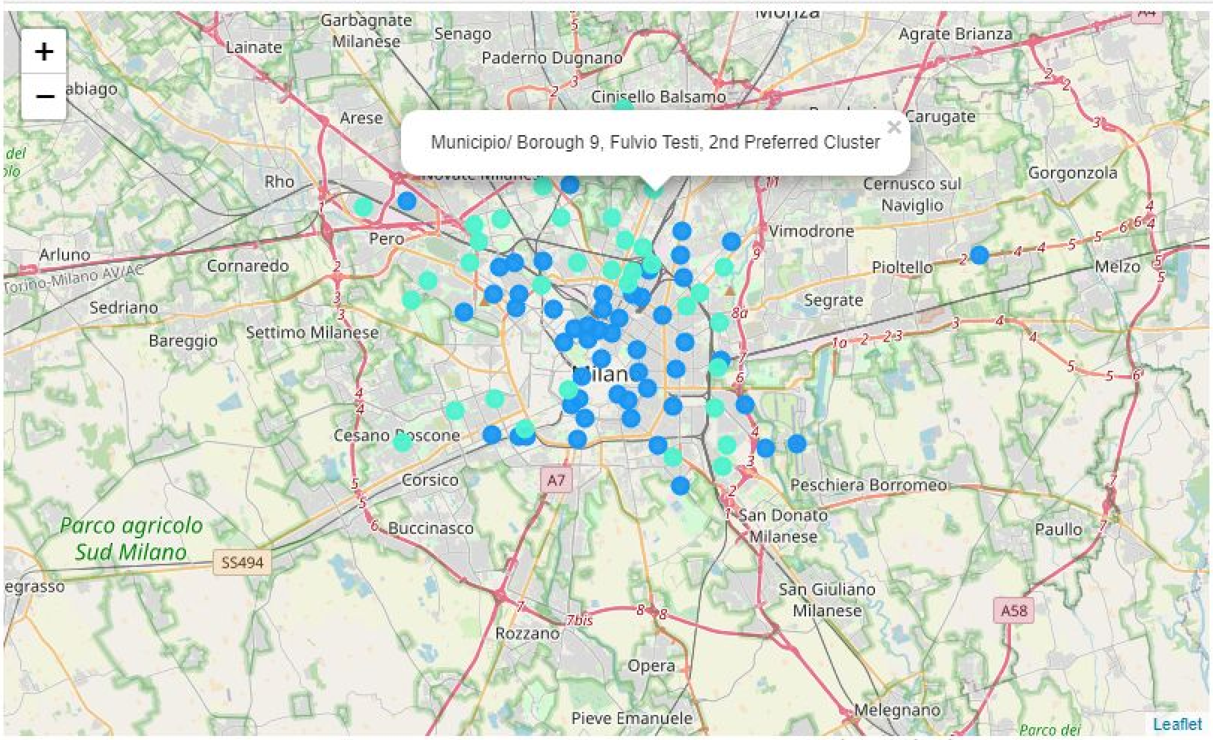
	Borough	Name	Districts	Cluster Labels	Cluster Names	Latitude	Longitude
0	1	(Centro storico,)	Brera	3	1st Preferred Cluster	45.471519	9.187735
1	1	(Centro storico,)	Conca del Naviglio	3	1st Preferred Cluster	45.458560	9.177745
2	1	(Centro storico,)	Guastalla	3	1st Preferred Cluster	45.458252	9.200023
3	1	(Centro storico,)	Porta Sempione	3	1st Preferred Cluster	45.477128	9.170598
4	1	(Centro storico,)	Porta Tenaglia	3	1st Preferred Cluster	45.477821	9.181593
5	2	(Stazione di Milano Centrale, Gorla, Turro, Gr...	Crescenzago	3	1st Preferred Cluster	45.509219	9.247484
6	2	(Stazione di Milano Centrale, Gorla, Turro, Gr...	Gorla	3	1st Preferred Cluster	45.504945	9.224539
7	2	(Stazione di Milano Centrale, Gorla, Turro, Gr...	Piazzale Loreto	3	1st Preferred Cluster	45.485859	9.216384
8	2	(Stazione di Milano Centrale, Gorla, Turro, Gr...	Maggiolina	3	1st Preferred Cluster	45.492483	9.202369
9	2	(Stazione di Milano Centrale, Gorla, Turro, Gr...	Ponte Seveso	3	1st Preferred Cluster	45.491607	9.206681

Tab 11: Boroughs in clusters 3 and 4

	Borough	Total Popular Venues per Borough on each cluster	Total Popular Venues	Percentage
0	1	435	435	100.00
1	2	363	368	98.64
2	3	319	327	97.55
3	4	334	364	91.76
4	6	471	494	95.34
5	8	455	479	94.99
6	9	611	655	93.28

Tab 12: Percentage of each borough found in the preferred clusters

All of Borough 1 is found in both clusters (100%), 98% of Borough 2 and so on.



Map 3: Map of the 2 preferred clusters (3 and 4)

Popular districts for each cluster:

1. Popular districts in Cluster 3

	Districts	Cluster Labels	Venue Category	Borough	Latitude	Longitude	Cluster Names
0	Porta Nuova	3	154	2	45.479858	9.192694	1st Preferred Cluster
1	Conca del Naviglio	3	100	1	45.458560	9.177745	1st Preferred Cluster
2	Porta Tenaglia	3	100	1	45.477821	9.181593	1st Preferred Cluster
3	Porta Garibaldi	3	98	9	45.480665	9.186888	1st Preferred Cluster
4	Brera	3	90	1	45.471519	9.187735	1st Preferred Cluster
5	Bullona	3	88	8	45.487549	9.166286	1st Preferred Cluster
6	Porta Venezia	3	87	3	45.474498	9.204801	1st Preferred Cluster
7	Porta Romana	3	82	4	45.452239	9.202056	1st Preferred Cluster
8	Conchetta	3	81	6	45.445473	9.176827	1st Preferred Cluster
9	Isola	3	81	9	45.487565	9.188972	1st Preferred Cluster
10	Porta Sempione	3	77	1	45.477128	9.170598	1st Preferred Cluster

Tab 13: Popular districts in cluster 3

2. *Popular districts in Cluster 4*

52	Foppette	4	53	6	45.449205	9.152853	2nd Preferred Cluster
53	Casoretto	4	44	3	45.488521	9.227395	2nd Preferred Cluster
54	Sant'Ambrogio	4	41	6	45.461391	9.172917	2nd Preferred Cluster
55	Lambrate	4	36	3	45.483148	9.241998	2nd Preferred Cluster
56	Montalbino	4	28	9	45.500007	9.193065	2nd Preferred Cluster
57	Gamboloita	4	28	4	45.439875	9.221240	2nd Preferred Cluster
58	Rottole	4	28	3	45.492661	9.233057	2nd Preferred Cluster
59	Prato Centenaro	4	25	9	45.509626	9.199271	2nd Preferred Cluster
60	Mirabello	4	22	2	45.495917	9.200479	2nd Preferred Cluster
61	Garegnano	4	21	8	45.502489	9.127848	2nd Preferred Cluster
62	Dergano	4	19	9	45.502513	9.176784	2nd Preferred Cluster

Tab 14: Popular districts in cluster 4

4. DISCUSSION SECTION

A. Observations

Selected Top 3 Boroughs:

From the final observations:

1. Borough 1 is a perfect candidate for the business
2. Borough 9 comes next and finally,
3. Borough 8 is 3rd preferred.

However, the districts could provide different results as shown in the dataframe Tab 13: Porta Nuova, belonging to Borough 2 seems to have a lot of reviews.

- Throughout the days I was working, each time I performed the k-Means algorithm, I had different alterations for the elbow method, but 6 usually was at the elbow.
- K-Means at 6 also mapped the clusters into different cluster labels each time I reran my codes, sometimes exchanging cluster 0 contents with that of cluster 4, or cluster 1 contents with cluster 5, etc. But what I noticed each time was that the same venues always made up the same clusters, which proves the accuracy of the algorithm.
- Venue popularity is based on information retrieved from Foursquare which in my opinion can be biased in some cases. Popular places are known based on reviews which means if a venue is not reviewed and is still popular, it may not appear in our data. If all existing venues had reviews I believe the results on most popular venues would have been slightly different. Nevertheless, the popularity of the boroughs in my opinion are very accurate, though for the districts, it may not be the case.
- Foursquare updates regularly, so the venues I retrieved each time were not always constant. Sometimes, fewer venues and other times more. But it did not affect the results on the clusters, just the total number of venues and their categories changed slightly.

B. Recommendations

It is important that after such studies, stakeholders should visit Ground 0 to have a better view and understanding about the districts. There is nothing that proves better than making a survey to determine what best suits at the time. Districts may differ in terms of popularity at a given time, this ultimately depends on the businesses that are functional.

However, I am very certain that my analysis is accurate enough and resolves a big part of the decision making, especially as the data I worked on is real-time.

5. CONCLUSION

It was interesting carrying out exploratory analysis on the data provided. I strongly believe that my work would be very helpful for Fashion start up businesses targeting Milan, especially with the millions of tourists that visit the city regularly.

6. References

- Municipalities of Milan, Wikipedia
- Foursquare API
- Geoportale Commune Milano