A Rome Case study: neighborhoods and venues analysis to locate a new business activity

A. Introduction

A.1. Description & Discussion of the Background

The City of Rome, is the most populous city of Italy. It is also a very popular city in the world for its culture and history, attracting thousands of tourists every year for its attractions and monuments. In addition, it is also a global hub of restaurants, food and wine. I decided to choose this city for my analysis because it’s where I grow up and where my roots belong to.

As you can see from the following figures, Rome is a city with a quite high population density. Being such a crowded city, leads the owners of shops and social sharing places to locate their activities where the population is dense. This also means that the market is highly competitive. Thus, any new business venture in the food market needs to be analyzed carefully. The insights derived from analysis will give good understanding of the business environment, which help in strategically targeting the market. This will help in reduction of risk. However, it is difficult to obtain information that will guide investors in this direction.

If some investor wants to open a new business activity, for instance a restaurant, we expect them to prefer the districts where there is a lower real estate cost and the type of business they want to install is less intense. At the same time, they may want to choose the district according to the social places density, then it’s better to include the places of interest close by.

When we consider all these problems, we can create a map and information chart where the real estate index is placed on Rome and each district is clustered according to the venue density. Then we will simulate two different scenarios: one in which our investor has some budget constraint and a second without any ecoonomical limit, with the final goal to find the best neighborhood to invest on.

A.2. Data Description

To consider the problem we can list the datas as below:

- Neighborhood list clustered with the average renting/selling costs for sqm from a famous Real Estate agency[1].
- Forsquare API to get the most common venues and the places of interest of all the given Borough of Rome [2].
- Geocoding API from Google Map, to get the center coordinates of the each Borough. [3].

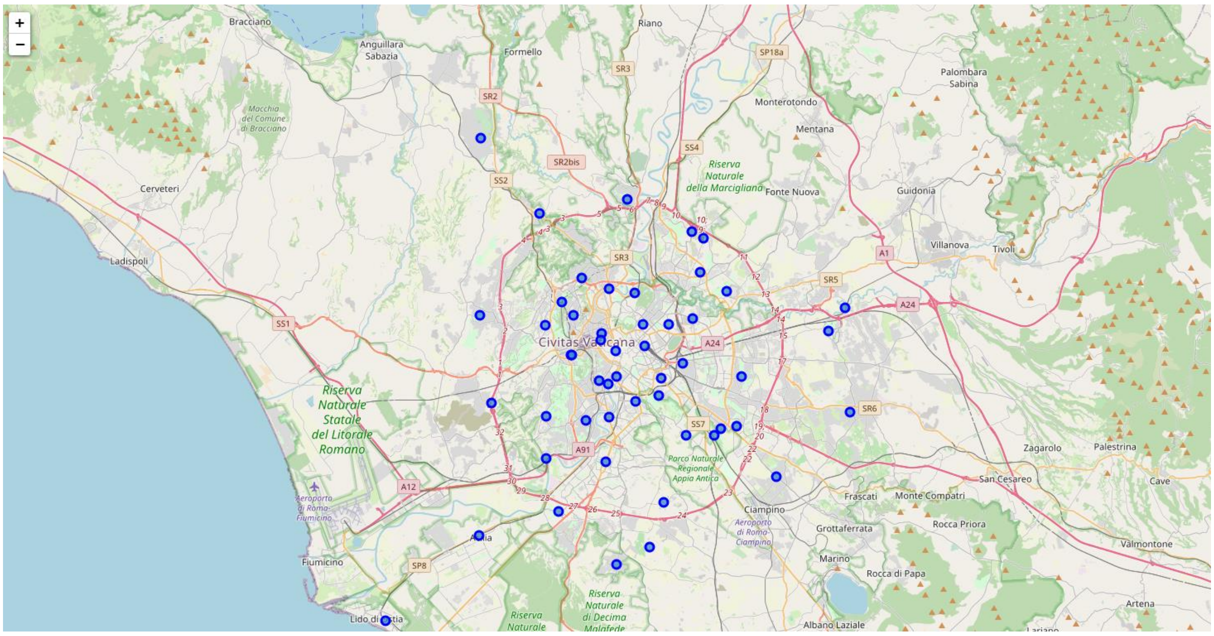
B. Methodology

As a database, we started from list containing the neighborhood grouped for the average selling and renting prices for sqm. Then I set the prices for a 150 sqm local.

Our master data head has the main components Neighborhood, Average Local Price and Rent, Latitude and Longitude information of the city:

	Neighborhood	Sell	Rent	Latitude	Longitude
0	Centro Storico	1170300.0	3301.5	41.898226	12.477325
1	Prati, Borgo, Mazzini, Delle Vittorie	763200.0	2581.5	41.908703	12.465287
2	Parioli, Flaminio	833250.0	2683.5	41.933805	12.492523
3	Salario, Trieste	735150.0	2379.0	41.914692	12.499691
4	Testaccio, Trastevere	813000.0	2982.0	41.877725	12.470725

I used python folium library to visualize geographic details of Rome and its boroughs, then I created a map of Rome with boroughs superimposed on top. I used latitude and longitude values to get the visual as below:



Next step is to find out which is the level of concurrency: to do that I need to know which is the amount of restaurant already present in the territory.

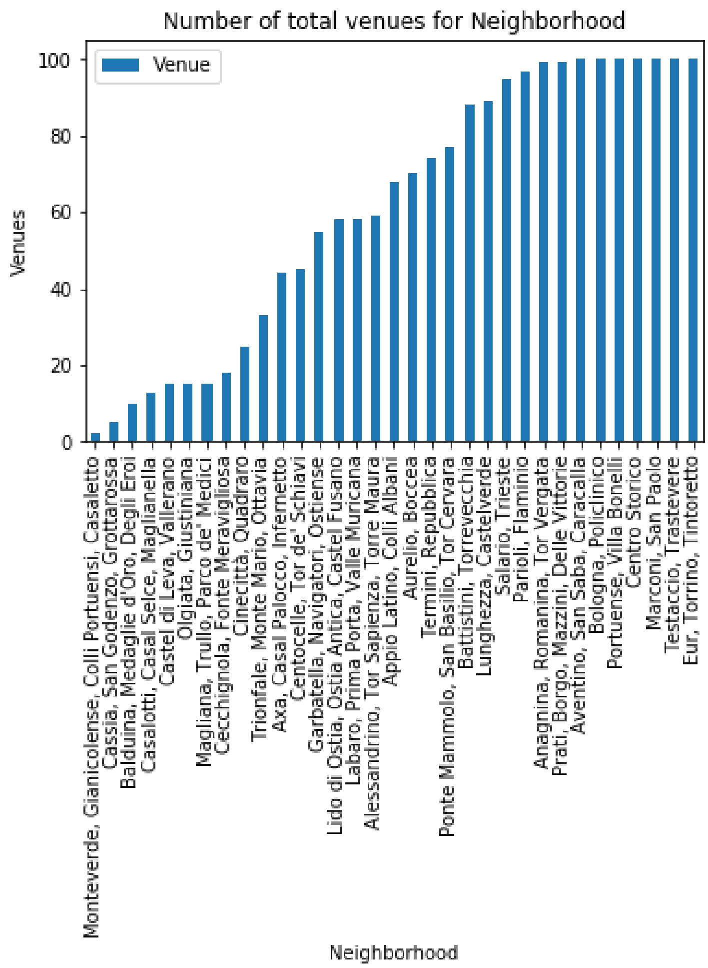
We utilized the Foursquare API to explore the restaurants and segment them: we set limit of results for 100 venues and 500 meter radius for each borough from their given latitude and longitude. Here is a head of the list Venues name, category, latitude and longitude from Foursquare API:

	Neighborhood	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Centro Storico	Pizza e Mozzarella	41.897598	12.479097	Pizza Place
1	Centro Storico	La Cabana	41.897251	12.482557	Italian Restaurant
2	Centro Storico	Scholars Lounge	41.896286	12.480130	Irish Pub
3	Centro Storico	Locanda del Prosciutto	41.897230	12.478640	Italian Restaurant
4	Centro Storico	Sapore di Mare	41.897371	12.478758	Seafood Restaurant

In summary of this data 2044 restaurants and cafe were returned by Foursquare. Here is a merged table of boroughs and the associated venues previously obtained:

	Neighborhood	Sell	Rent	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Centro Storico	1170300.0	3301.5	41.898226	12.477325	Pizza e Mozzarella	41.897598	12.479097	Pizza Place
1	Centro Storico	1170300.0	3301.5	41.898226	12.477325	La Cabana	41.897251	12.482557	Italian Restaurant
2	Centro Storico	1170300.0	3301.5	41.898226	12.477325	Scholars Lounge	41.896286	12.480130	Irish Pub
3	Centro Storico	1170300.0	3301.5	41.898226	12.477325	Locanda del Prosciutto	41.897230	12.478640	Italian Restaurant
4	Centro Storico	1170300.0	3301.5	41.898226	12.477325	Sapore di Mare	41.897371	12.478758	Seafood Restaurant

We can see that Eur, Testaccio, Marconi, the historical centrum, Policlinico and Parioli reached the 100 limit of venues. On the other hand; Quadraro, Magliana, Portuensi, Medaglie d'Oro, Olgiata, Castel di Leva, Cassia and Cecchignola boroughs are below 20 venues in our given coordinates with Latitude and Longitude, in the graph below:



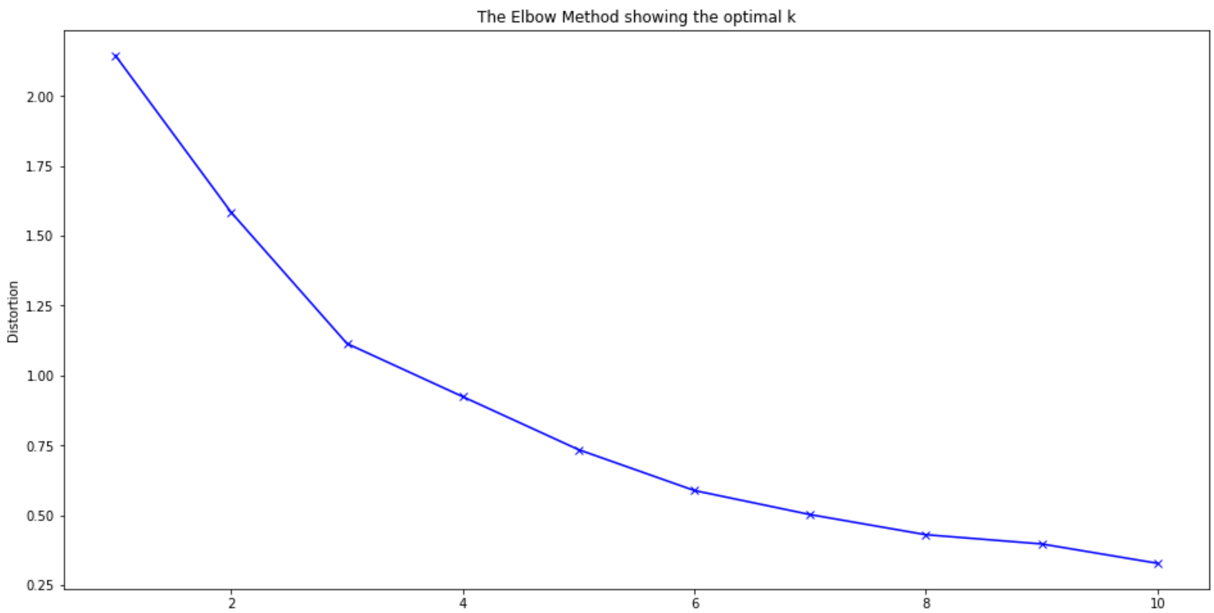
The result doesn’t mean that inquiry run all the possible results in boroughs. Actually, it depends on given Latitude and Longitude informations and in our case we just run single one for each borough. This can be increased giving more Latitude and Longitude coordinates.

Globally, **80 unique categories** were returned by Foursquare. So we decided to focus on the top 10 venues category for each borough, as represented in the table below:

	Neighborhood	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
32	Monteverde, Gianicolense, Colli Portuensi, Cas...	Pizza Place	Café	Italian Restaurant	Chinese Restaurant	Restaurant	Bakery	Fast Food Restaurant	Japanese Restaurant	Seafood Restaurant	Kebab Restaurant
31	Cassia, San Godenzo, Grottarossa	Italian Restaurant	Café	Restaurant	BBQ Joint	Pizza Place	Trattoria/Osteria	Fast Food Restaurant	Steakhouse	Bakery	Burger Joint
30	Balduina, Medaglie d'Oro, Degli Eroi	Italian Restaurant	Pizza Place	Trattoria/Osteria	Fast Food Restaurant	Restaurant	Bakery	Steakhouse	Gastropub	Chinese Restaurant	Fried Chicken Joint
29	Casalotti, Casal Selce, Maglianella	Italian Restaurant	Pizza Place	Café	Chinese Restaurant	Steakhouse	Gastropub	Fast Food Restaurant	Diner	Restaurant	Breakfast Spot
28	Castel di Leva, Vallerano	Italian Restaurant	Restaurant	Café	Pizza Place	Roman Restaurant	Trattoria/Osteria	Gastropub	Bakery	Sandwich Place	Greek Restaurant

Because we had some common venue categories in boroughs, we clustered them by using unsupervised machine learning **K-means algorithm,** as one of the most common cluster method of unsupervised learning.

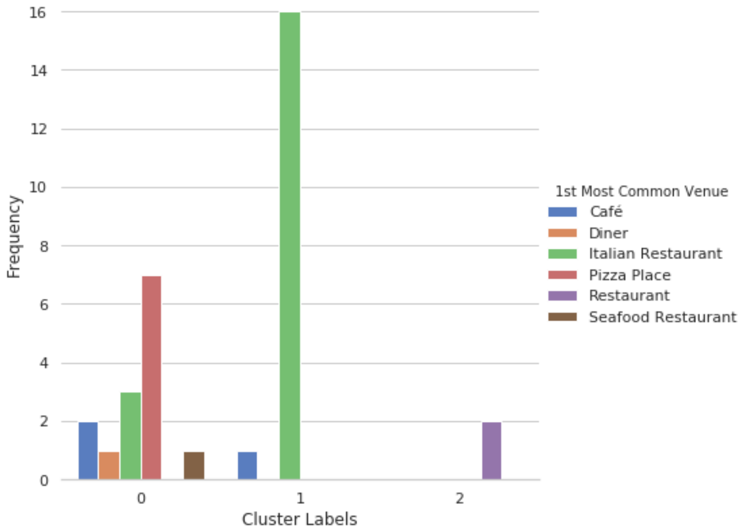
We set **3** as the optimum parameter of k to analyze the K-Means, as a result of the elbow method:



In the table below, you can see the cluster labels for each borough assigned by the algorithm:

	Cluster Labels	Neighborhood	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
32	0	Monteverde, Gianicolense, Colli Portuensi, Cas...	Pizza Place	Café	Italian Restaurant	Chinese Restaurant	Restaurant	Bakery	Fast Food Restaurant	Japanese Restaurant	Seafood Restaurant	Kebab Restaurant
31	1	Cassia, San Godenzo, Grottarossa	Italian Restaurant	Café	Restaurant	BBQ Joint	Pizza Place	Trattoria/Osteria	Fast Food Restaurant	Steakhouse	Bakery	Burger Joint
30	0	Balduina, Medaglie d'Oro, Degli Eroi	Italian Restaurant	Pizza Place	Trattoria/Osteria	Fast Food Restaurant	Restaurant	Bakery	Steakhouse	Gastropub	Chinese Restaurant	Fried Chicken Joint
29	1	Casalotti, Casal Selce, Maglianella	Italian Restaurant	Pizza Place	Café	Chinese Restaurant	Steakhouse	Gastropub	Fast Food Restaurant	Diner	Restaurant	Breakfast Spot
28	1	Castel di Leva, Vallerano	Italian Restaurant	Restaurant	Café	Pizza Place	Roman Restaurant	Trattoria/Osteria	Gastropub	Bakery	Sandwich Place	Greek Restaurant

Then we grouped each of these clusters by the 1st Most Common Venue and build a bar chart which helped us to find proper labels for each cluster:



After examining the above graph, we labeled each cluster as follows:

- Cluster 0 : "Cafe & Pizzeria"
- Cluster 1 : "Italian Restaurant"
- Cluster 2 : "Other Restaurant"

Along these results, we added some other insight in our analysis: the number of places of interest (monuments, touristic attractions, leisure places) inside each neighborhood. It's an important factor to take into account, since it should guarantees us a minimal influx of people passing by. To obtain it, we made a new query with the Foursquare API and add the amount of restaurants already present for each neighborhood in our master table:

	Neighborhood	Number of Restaurant	Sell	Rent	Latitude	Longitude	Number of places of interest
0	Monteverde, Gianicolense, Colli Portuensi, Cas...	2	573000.0	2172.0	41.880041	12.463194	27
1	Cassia, San Godenzo, Grottarossa	5	464400.0	1888.5	41.982783	12.413697	4
2	Balduina, Medaglie d'Oro, Degli Eroi	10	584550.0	2196.0	41.919967	12.442128	44
3	Casalotti, Casal Selce, Maglianella	13	331350.0	1471.5	41.920130	12.364608	26
4	Olgiata, Giustiniana	15	329850.0	1440.0	42.029616	12.364968	30

Because we wanted to offer a quick and detailed view using the map, we include as additional info the concrete amount of the top 3 venues :

	Neighborhood	Top 3 Venues
0	Alessandrino, Tor Sapienza, Torre Maura	Pizza Place 13 Café 11 Italian Restaurant 10
1	Anagnina, Romanina, Tor Vergata	Italian Restaurant 49 Café 14 Restaurant 9
2	Appio Latino, Colli Albani	Italian Restaurant 15 Pizza Place 14 Trattoria...
3	Aurelio, Boccea	Italian Restaurant 20 Pizza Place 12 Café 11
4	Aventino, San Saba, Caracalla	Italian Restaurant 43 Café 8 Pizza Place 7

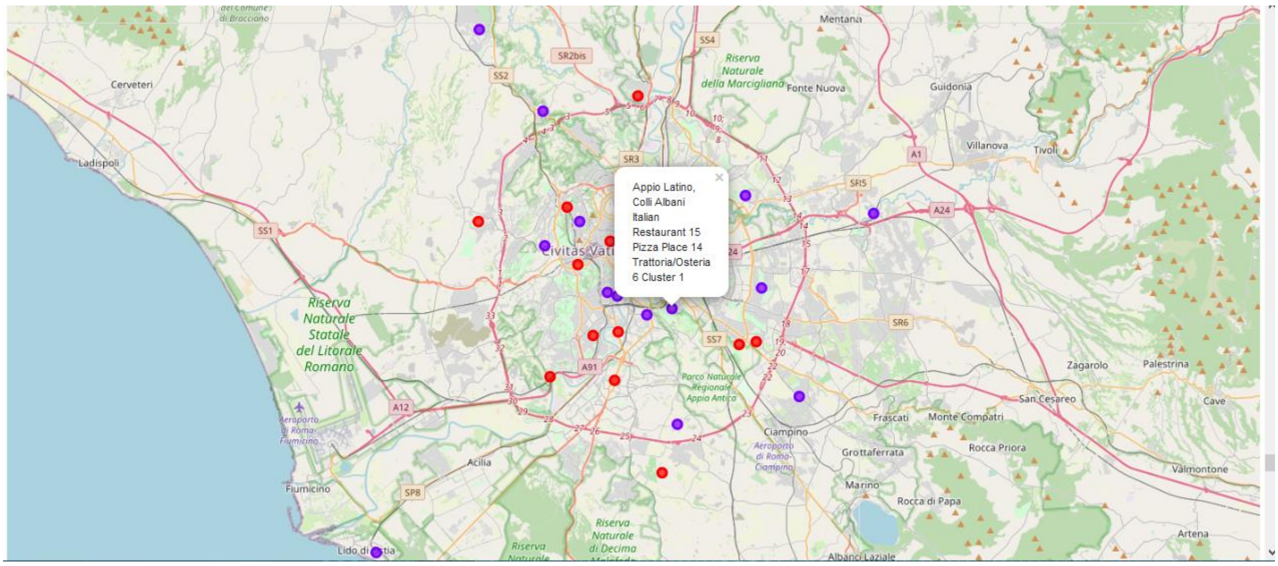
C. Results

As final step, we merged those new variables with related cluster informations in our main master table:

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 Most Common Venue	10th Most Common Venue	Top 3 Venues
0	Pizza Place	Café	Italian Restaurant	Chinese Restaurant	Restaurant	Bakery	Fast Food Restaurant	Japanese Restaurant	Seafood Restaurant	Kebab Restaurant	Restaurant 1 Café 1 African Restaurant 0
1	Italian Restaurant	Café	Restaurant	BBQ Joint	Pizza Place	Trattoria/Osteria	Fast Food Restaurant	Steakhouse	Bakery	Burger Joint	Restaurant 2 Diner 1 BBQ Joint 1
0	Italian Restaurant	Pizza Place	Trattoria/Osteria	Fast Food Restaurant	Restaurant	Bakery	Steakhouse	Gastropub	Chinese Restaurant	Fried Chicken Joint	Pizza Place 6 Italian Restaurant 3 Café 1
0	Italian Restaurant	Pizza Place	Café	Chinese Restaurant	Steakhouse	Gastropub	Fast Food Restaurant	Diner	Restaurant	Breakfast Spot	Italian Restaurant 7 Agriturismo 3 Restaurant 2
0	Italian Restaurant	Café	Pizza Place	Restaurant	Fast Food Restaurant	Chinese Restaurant	Seafood Restaurant	Bakery	Steakhouse	Burger Joint	Italian Restaurant 4 Café 4 Restaurant 1

The following is a clustered map boroughs of Rome that shows the following information:

- Borough name
- Cluster name
- Top 3 number of venue



Now, let’s say for example our investor has a budget constraint and he’s willing to pay not more than 2000 euro monthly for his activity, then we need to create a subset of candidates:

	Neighborhood	Number of Restaurant	Sell	Rent	Latitude	Longitude	Number of places of interest
1	Castel di Leva, Vallerano	4	430050.0	1798.5	41.777052	12.505426	23
2	Cassia, San Godenzo, Grottarossa	5	464400.0	1888.5	41.982783	12.413697	4
3	Casalotti, Casal Selce, Maglianella	9	331350.0	1471.5	41.920130	12.364608	26
5	Magliana, Trullo, Parco de' Medici	15	372750.0	1753.5	41.831738	12.419640	26
6	Olgiata, Giustiniana	16	329850.0	1440.0	42.029616	12.364968	30
7	Cecchignola, Fonte Meravigliosa	18	529350.0	1800.0	41.804497	12.516652	23
8	Cinecittà, Quadraro	25	411300.0	1836.0	41.851790	12.577179	24
9	Trionfale, Monte Mario, Ottavia	34	397050.0	1827.0	41.928204	12.432327	18
10	Axa, Casal Palocco, Infernetto	46	329700.0	1599.0	41.902347	12.491022	25
11	Centocelle, Tor de' Schiavi	47	367650.0	1734.0	41.850110	12.564226	24
13	Lido di Ostia, Ostia Antica, Castel Fusano	57	351900.0	1597.5	41.731118	12.286269	27
14	Labaro, Prima Porta, Valle Muricana	58	301950.0	1570.5	41.991468	12.486836	33
15	Alessandrino, Tor Sapienza, Torre Maura	66	332100.0	1738.5	41.882435	12.581442	25
17	Aurelio, Boccea	71	488550.0	1965.0	41.895489	12.440924	22
19	Ponte Mammolo, San Basilio, Tor Cervara	77	376200.0	1807.5	41.935173	12.568815	30
20	Battistini, Torrevecchia	84	427800.0	1846.5	41.906268	12.414909	42
21	Lunghezza, Castelverde	87	244050.0	1461.0	41.924698	12.667409	31
22	Anagnina, Romanina, Tor Vergata	95	341250.0	1629.0	41.820559	12.610515	27
25	Portuense, Villa Bonelli	100	485550.0	1894.5	41.855102	12.452388	21

In our subset, we calculated the mean for each variable included in our analysis:

	Number of Restaurant	Sell	Rent	Latitude	Longitude	Number of places of interest
count	20.000000	20.000000	20.000000	20.000000	20.000000	20.000000
mean	50.700000	390060.000000	1732.800000	41.883893	12.481529	25.150000
std	33.413045	74314.643312	165.937956	0.073497	0.094584	7.191186
min	4.000000	244050.000000	1440.000000	41.731118	12.286269	4.000000
25%	17.500000	331912.500000	1598.625000	41.845517	12.418457	22.750000
50%	52.000000	374475.000000	1776.000000	41.888962	12.479085	25.000000
75%	78.750000	438637.500000	1838.625000	41.925574	12.565373	27.750000
max	100.000000	529350.000000	1998.000000	42.029616	12.667409	42.000000

Then to find our optimal combination, we should pick the zone where the rent and the number of restaurant are under the average and the places of interest above. Then we’ll get the result:

	Neighborhood	Number of Restaurant	Sell	Rent	Latitude	Longitude	Number of places of interest	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5 Mo Comm Ven
6	Olgiata, Giustiniana	16	329850.0	1440.0	42.029616	12.364968	30	1	Pizza Place	Italian Restaurant	Café	Food	East Europe Restaur

If we don’t have any budget constraint we can do the same with our master table and we can have more choices:

	Neighborhood	Number of Restaurant	Sell	Rent	Latitude	Longitude	Number of places of interest	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	Co V
3	Casalotti, Casal Selce, Maglianella	9	331350.0	1471.5	41.920130	12.364608	26	0	Italian Restaurant	Café	Restaurant	Pizza Place	Roman Restaurant	Trattoria/Osteria	Gastropub	B
5	Magliana, Trullo, Parco de' Medici	15	372750.0	1753.5	41.831738	12.419640	26	0	Italian Restaurant	Pizza Place	Café	Steakhouse	Fast Food Restaurant	Chinese Restaurant	Bakery	Se Resta
6	Olgiata, Giustiniana	16	329850.0	1440.0	42.029616	12.364968	30	1	Pizza Place	Italian Restaurant	Café	Food	Eastern European Restaurant	Emilia Restaurant	Ethiopian Restaurant	F Resta
10	Axa, Casal Palocco, Infernetto	46	329700.0	1599.0	41.902347	12.491022	25	2	Restaurant	BBQ Joint	Bagel Shop	Diner	Food	Eastern European Restaurant	Emilia Restaurant	Eth Resta
13	Lido di Ostia, Ostia Antica, Castel Fusano	57	351900.0	1597.5	41.731118	12.286269	27	1	Italian Restaurant	Pizza Place	Café	Chinese Restaurant	Restaurant	Japanese Restaurant	Bakery	Se Resta
14	Labaro, Prima Porta, Valle Muricana	58	301950.0	1570.5	41.991468	12.486836	33	0	Italian Restaurant	Pizza Place	Sandwich Place	Restaurant	Café	Trattoria/Osteria	Bakery	Se Resta

D. Discussion

As I mentioned before, Rome is a big city with a high population density in a narrow area. The total number of measurements and population densities of the 33 districts in total can vary. As there is such a complexity, very different approaches can be tried in clustering and classification studies. Moreover, it is obvious that not every classification method can yield the same high quality results for this metropolis.

I used the Kmeans algorithm as part of this clustering study. When I tested the Elbow method, I set the optimum k value to 3. However, only 33 district coordinates were used. For more detailed and accurate guidance, the data set can be expanded and the details of the neighborhood or street can also be drilled.

I also performed data analysis through this information by adding the coordinates of districts and home sales price averages as static data on Real State Agency Website.

I ended the study by visualizing the data and clustering information on the Rome map, choosing the best location with the optimal combination for each of the two different scenarios.

F. Conclusion

As a result, people are turning to big cities to start a business or work. For this reason, people can achieve better outcomes through their access to the platforms where such information is provided.

For not only investors but also city managers can manage the city more regularly by using similar data analysis types or platforms to strategically locate a new place or activity.

Neverthless, this study could be enriched very easiy adding other meaningful factors.

Please, feel free to share your thoughts: any advice or tips would be appreciated.

G. References:

- [1] Real state agency- Roma Real estate market
- [2] API Foursquare
- [3] Google Geocoding API