HOUSING SALES PRICES & BUSINESS OPPORTUNITIES IN ROME

Daneyal Anis

Introduction

- Rome is the eternal city and my favorite city in Europe. I have a lot of fond memories from it when I went back packing through it with my university friends back in 2003. Given that it has been so hard hit by COVID-19 pandemic, I want to pay tribute to it by using it as my project.
- Given that Rome is such a popular tourist destination and has so much history, it has really expensive real estate as well as population density. With the impacts to its economy due to COVID-19, investors can be looking at boroughs of Rome that have a high population and relatively lower real estate prices. In addition, using FourSquare data, I will also look at type of businesses in each borough to be able to recommend the best Rome neighborhood to start a business in and type of business to start based on the real estate prices and population density.

Data Description

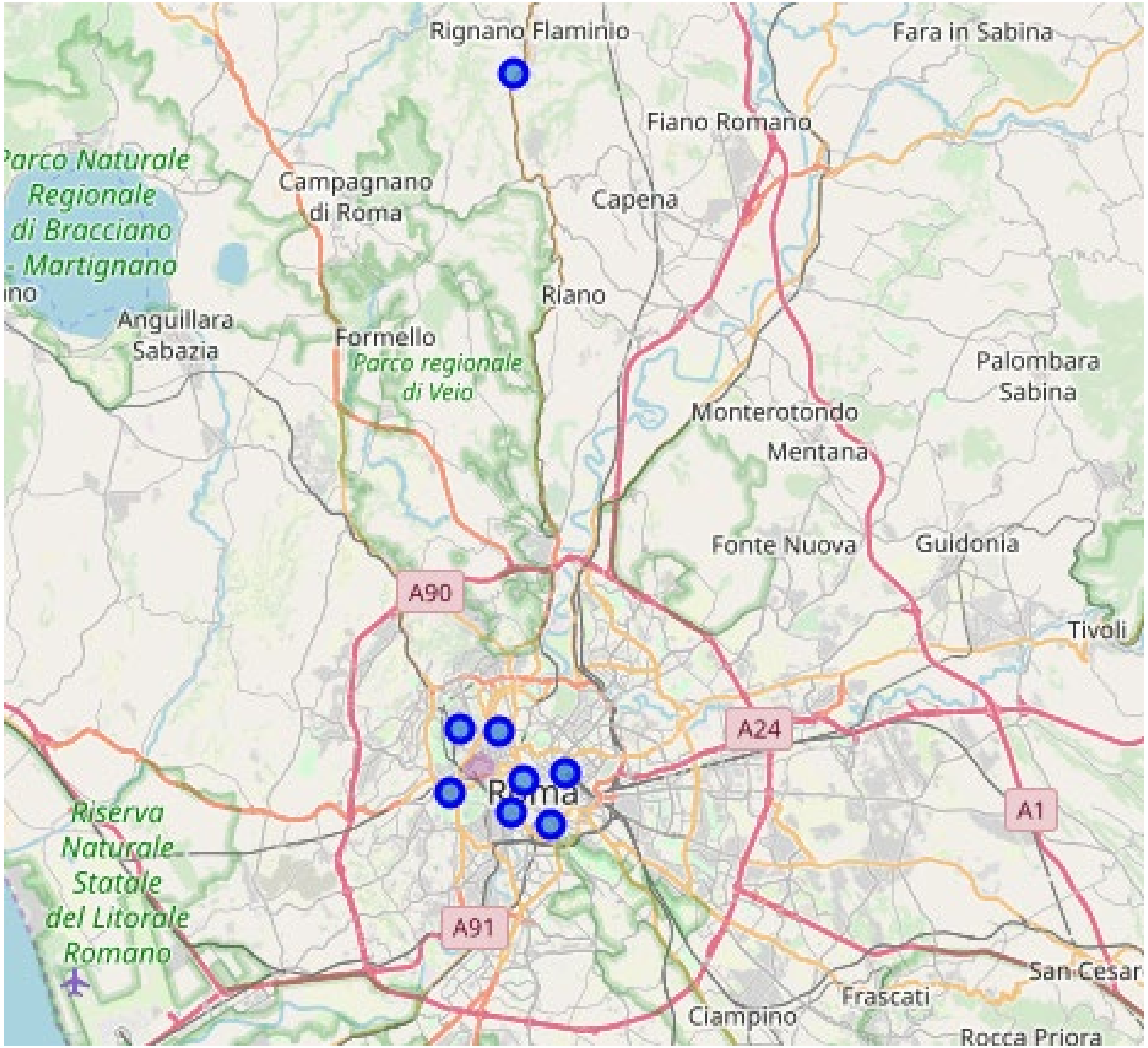
- For this analysis, I used the following data sources:
- I got the real estate prices by different neighborhoods of Rome using the statist.com site that contains data as of December, 2019 [https://www.statista.com/statistics/670698/asking-price-for-properties-for-sale-in-rome-by-area-italy/].
- I got the .json file for Rome from carto.com site that will help us create the choropleth map of it's neighborhoods [https://maurizioman.carto.com/tables/rome_admin/public/map].
- I used **Forsquare API** to get the most common venues of given Borough of Rome [https://foursquare.com/].

Methodology

Getting Geocodes and Creating Data-frame

- As a database, I created the dataset of Rome boroughs by populating the neighborhood names, real estate prices and geographic coordinates and saved it in my github repository. I then saved it in the pandas dataframe which has the following columns *Borough, Average House Price, Latitude* and *Longitude*.

	Borough	Avg-HousePrice	Latitude	Longitude
0	CentroStorico	7817	41.8982	12.4773
1	Caracalla	6910	41.8794	12.4931
2	Flaminio	5622	42.1919	12.4725
3	Trastevere	5435	41.8848	12.4704
4	Della Vittoria	5137	41.9182	12.4639



Methodology (cont'd)

Visualizing Geographic Details and Boroughs – Using Folium and Geopy

Once the data-frame was ready, I used Folium library to visualize geographic details of Rome and its boroughs and further augmented that by using Geopy library to get the longitude and latitude values of Rome.

Methodology (cont'd)

Exploring Boroughs in Rome and Common Venues – Using FourSquare API

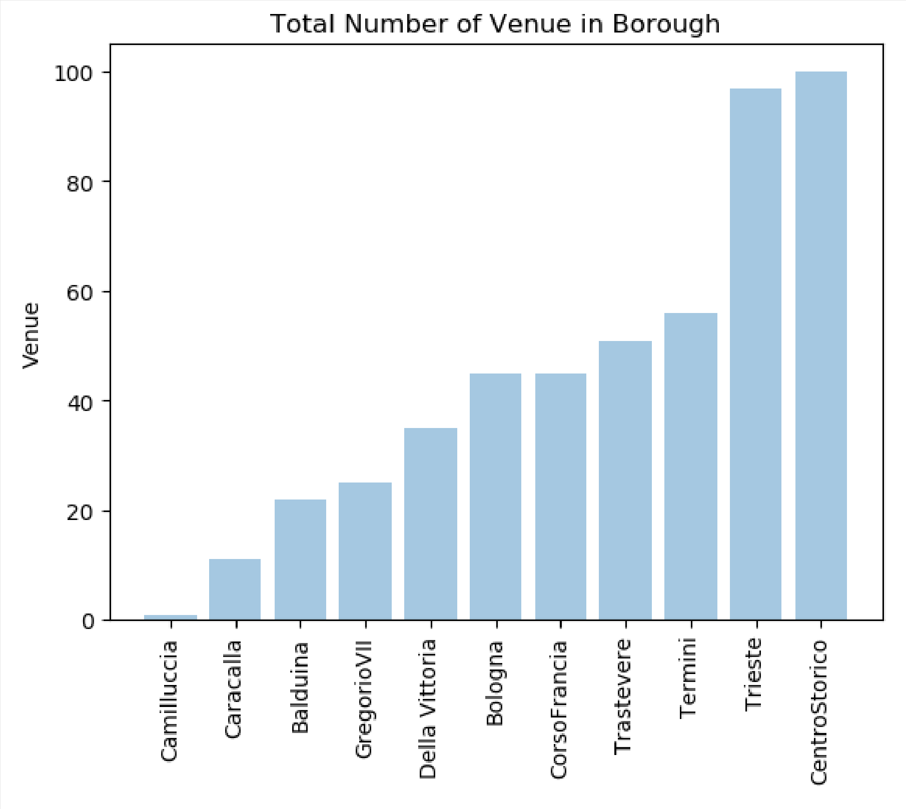
- I then used FourSquare API to get details of the nearby venues in each borough of Rome.
- Overall FourSquare API returned the following results.
- 100 different categories of venues in Rome
- Counts of venues by borough – snapshot below for reference:

	Borough	Count
0	Camilluccia	1
1	Caracalla	11
2	Balduina	22
3	GregorioVII	25
4	Della Vittoria	35
5	Bologna	45
6	CorsoFrancia	45
7	Trastevere	51
8	Termini	56
9	Trieste	97
10	CentroStorico	100

Methodology (cont'd)

I then visualized these results as a bar chart

The bar chart shows us that Centro Storico and Trieste have close to 100 venues, followed by Tremini, Trastavere, Corso Francia, Della Vittoria and Bologna that have venues in 40-60 range. Remaining boroughs are less venue rich like Georgio VII, Balduina, Caracalla and Camillucia. Camillucia specially seems really low in venues and potentially ripe for further investment.



Methodology (cont'd)

Analyzing Each Borough

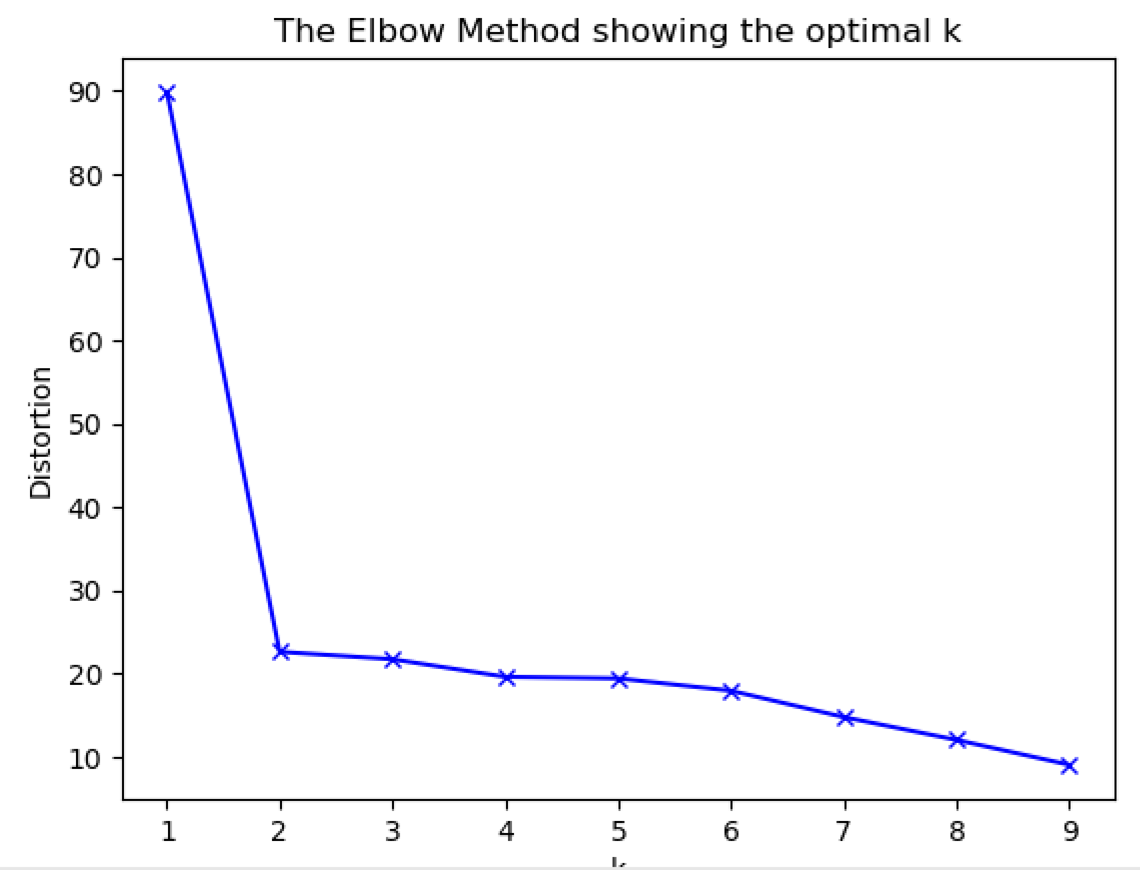
- Using one hot encoding, I categorized the venue types further and sorted them based on occurrence by each borough. That gave me the following results showing the top 3 most common venues by each borough. This will help me further determine the best investment opportunity in each borough depending on the types of venues that exist currently.

	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	Balduina	Italian Restaurant	Hotel	Spa
1	Bologna	Italian Restaurant	Bar	Pizza Place
2	Camilluccia	Nightclub	Wine Shop	Flea Market
3	Caracalla	Park	Performing Arts Venue	Plaza
4	CentroStorico	Italian Restaurant	Plaza	Ice Cream Shop
5	CorsoFrancia	Italian Restaurant	Café	Pizza Place
6	Della Vittoria	Italian Restaurant	Café	Pizza Place
7	GregorioVII	Pizza Place	Hotel	Café
8	Termini	Hotel	Italian Restaurant	Plaza
9	Trastevere	Italian Restaurant	Pizza Place	Café
10	Trieste	Café	Italian Restaurant	Plaza

Methodology (cont'd)

Clustering of Boroughs based on Venue Types Using K-Means

- I used K-means analysis for creating cluster neighborhoods based on venue types. I first applied the elbow method to determine the optimal number of clusters to use in the K-means analysis
- Elbow method recommends the optimal number of clusters as 2 for the K-means analysis.



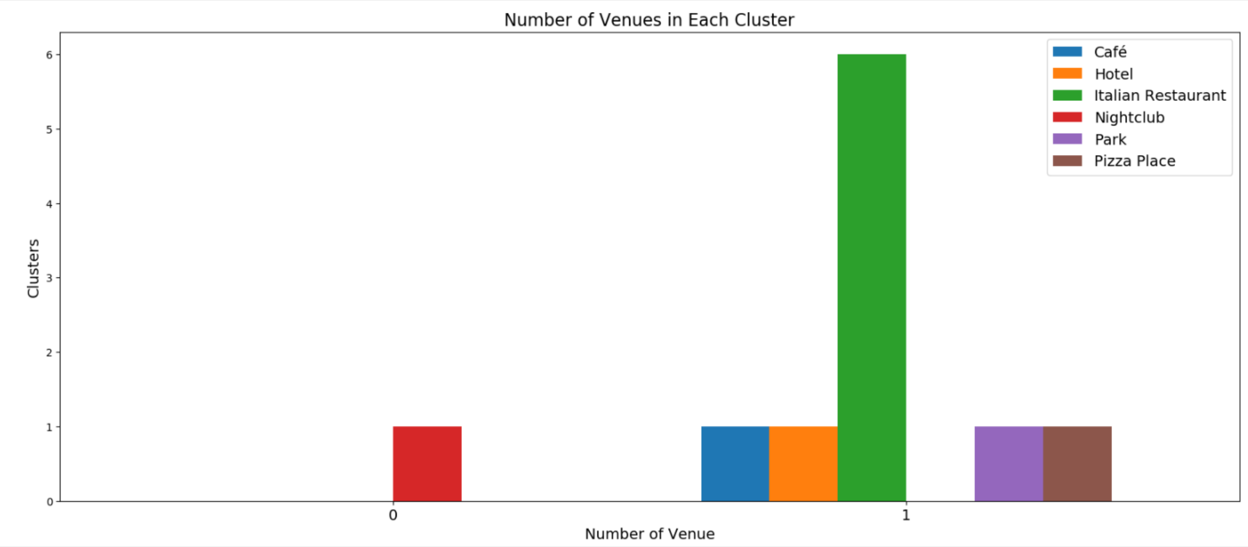
	Borough	Avg-HousePrice	Latitude	Longitude	Cluster Labels	Cluster	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	
0	CentroStorico	7817	41.898200	12.477300	1.0	1.0	Italian Restaurant	Plaza	Ice Cream Shop	
1	Caracalla	6910	41.879400	12.493100	1.0	1.0	Park	Performing Arts Venue	Plaza	
2	Flaminio	5622	42.191900	12.472500	NaN	NaN	NaN	NaN	NaN	
3	Trastevere	5435	41.884800	12.470400	1.0	1.0	Italian Restaurant	Pizza Place	Café	Ic
4	Della Vittoria	5137	41.918200	12.463900	1.0	1.0	Italian Restaurant	Café	Pizza Place	
5	Trieste	4884	45.649500	13.776800	1.0	1.0	Café	Italian Restaurant	Plaza	
6	Bologna	4531	44.498955	11.327500	1.0	1.0	Italian Restaurant	Bar	Pizza Place	
7	CorsoFrancia	4289	45.076734	7.667100	1.0	1.0	Italian Restaurant	Café	Pizza Place	
8	Termini	4107	41.901100	12.501200	1.0	1.0	Hotel	Italian Restaurant	Plaza	Ic
9	Camilluccia	4099	43.968130	12.662350	0.0	0.0	Nightclub	Wine Shop	Flea Market	
10	Balduina	4000	41.920000	12.442100	1.0	1.0	Italian Restaurant	Hotel	Spa	
11	GregorioVII	3972	41.893348	12.435962	1.0	1.0	Pizza Place	Hotel	Café	

Methodology (cont'd)

I then used Scipy spatial distance library to apply the K-means algorithm and that divided up Rome's neighborhoods into clusters as follows:

Methodology (cont'd)

- Based on above, Rome neighborhoods have been divided up into clusters 0 and 1 with one outlier (Flaminio) that did not fit either of the clusters.
- This can be visualized as follows based on bar chart and table as follows:



	Clusters	Labels
0	0	Night Club
1	1	Multiple Social Venues

	Borough	Join
0	Balduina	6 Italian Restaurant, 2 Hotel, 1 Asian Restaurant
1	Bologna	7 Italian Restaurant, 4 Bar, 3 Pizza Place
2	Camilluccia	1 Nightclub
3	Caracalla	2 Park, 1 Festival, 1 Garden
4	CentroStorico	16 Italian Restaurant, 14 Plaza, 9 Ice Cream Shop
5	CorsoFrancia	6 Italian Restaurant, 4 Café, 4 Pizza Place
6	Della Vittoria	7 Italian Restaurant, 4 Café, 2 Breakfast Spot
7	GregorioVII	3 Café, 3 Hotel, 3 Pizza Place
8	Termini	16 Hotel, 12 Italian Restaurant, 4 Plaza
9	Trastevere	12 Italian Restaurant, 5 Pizza Place, 4 Café
10	Trieste	11 Café, 8 Italian Restaurant, 8 Plaza

Methodology (cont'd)

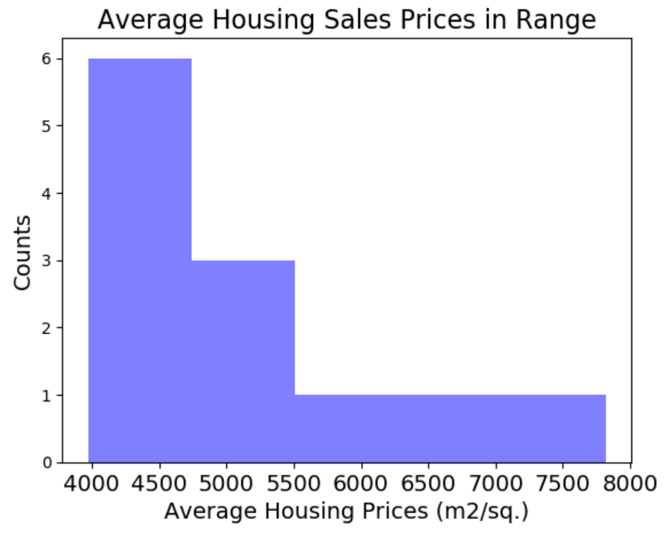
We can also present counts of different venue types in each neighborhood

Methodology (cont'd)

Analyzing House Prices by Borough

- I then analyzed house sales prices by borough for per square meter and divided up the prices in ranges using a histogram to further group the boroughs.

	Borough	Avg-HousePrice
0	GregorioVII	3972
1	Balduina	4000
2	Camilluccia	4099
3	Termini	4107
4	CorsoFrancia	4289
5	Bologna	4531
6	Trieste	4884
7	Della Vittoria	5137
8	Trastevere	5435
9	Flaminio	5622
10	Caracalla	6910
11	CentroStorico	7817



	Borough	Avg-HousePrice	Level_labels
0	GregorioVII	3972	Low Level HSP
1	Balduina	4000	Mid Level HSP
2	Camilluccia	4099	Mid Level HSP
3	Termini	4107	Mid Level HSP
4	CorsoFrancia	4289	Mid Level HSP
5	Bologna	4531	Mid Level HSP
6	Trieste	4884	Mid Level HSP
7	Della Vittoria	5137	Mid Level HSP
8	Trastevere	5435	Mid Level HSP
9	Flaminio	5622	Mid Level HSP
10	Caracalla	6910	High Level HSP
11	CentroStorico	7817	High Level HSP

Methodology (cont'd)

Analyzing House Prices by Borough

Based on the above histogram, house sales price (HSP) ranges can be defined as follows:

4000 AHP : "Low Level HSP"

4000-6000 AHP : "Mid Level HSP"

6000-8000 AHP : "High Level HSP"

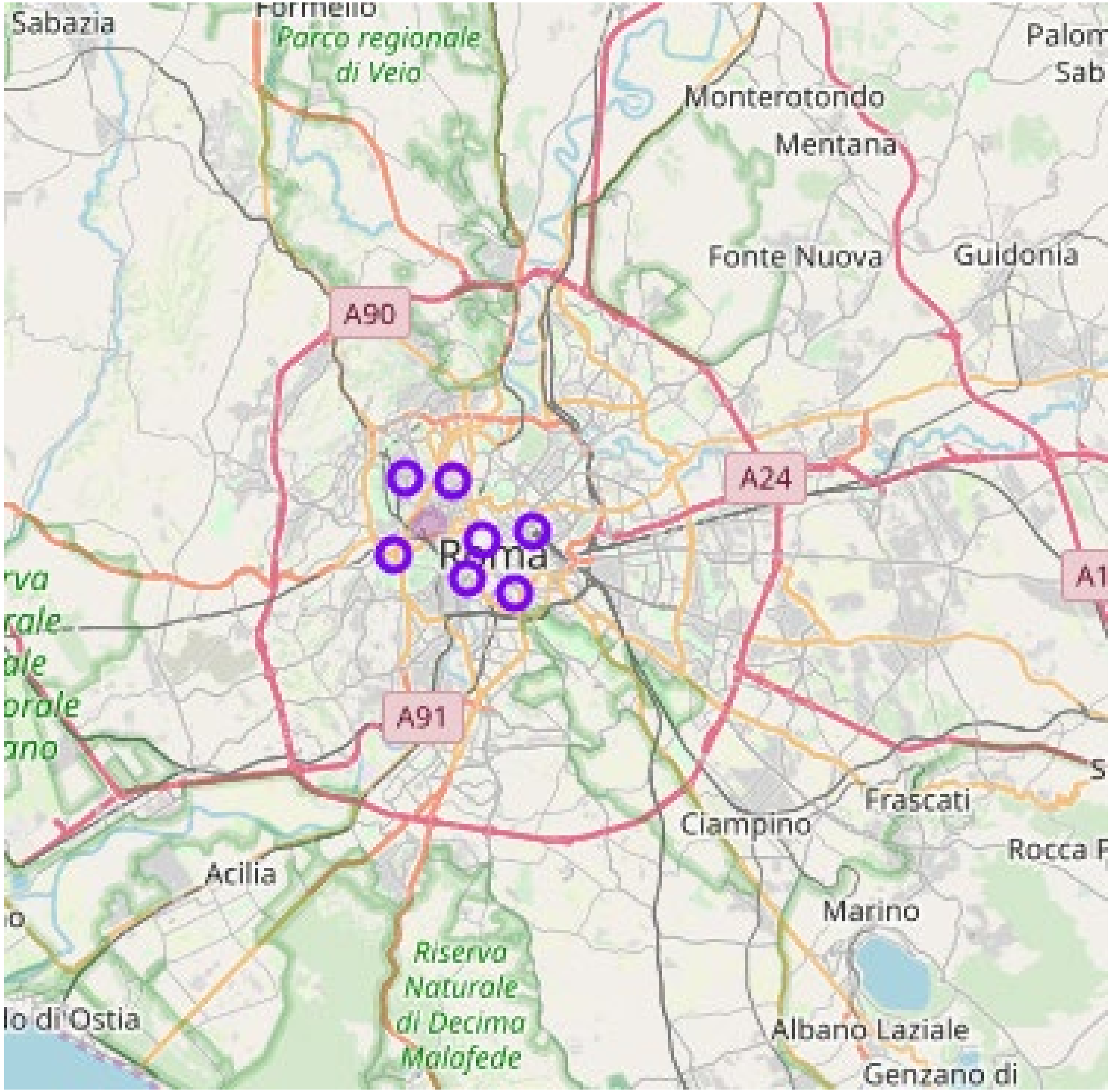
We can now put the boroughs in the following price range labels

	Borough	Avg-HousePrice	Latitude	Longitude	Cluster Labels	Cluster	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	CentroStorico	7817	41.898200	12.477300	1.0	1.0	Italian Restaurant	Plaza	Ice Cream Shop
1	Caracalla	6910	41.879400	12.493100	1.0	1.0	Park	Performing Arts Venue	Plaza
2	Flaminio	5622	42.191900	12.472500	NaN	NaN	NaN	NaN	NaN
3	Trastevere	5435	41.884800	12.470400	1.0	1.0	Italian Restaurant	Pizza Place	Café
4	Della Vittoria	5137	41.918200	12.463900	1.0	1.0	Italian Restaurant	Café	Pizza Place
5	Trieste	4884	45.649500	13.776800	1.0	1.0	Café	Italian Restaurant	Plaza
6	Bologna	4531	44.498955	11.327500	1.0	1.0	Italian Restaurant	Bar	Pizza Place
7	CorsoFrancia	4289	45.076734	7.667100	1.0	1.0	Italian Restaurant	Café	Pizza Place
8	Termini	4107	41.901100	12.501200	1.0	1.0	Hotel	Italian Restaurant	Plaza
9	Camilluccia	4099	43.968130	12.662350	0.0	0.0	Nightclub	Wine Shop	Flea Market	Coffee Shop
10	Balduina	4000	41.920000	12.442100	1.0	1.0	Italian Restaurant	Hotel	Spa	Coffee Shop
11	GregorioVII	3972	41.893348	12.435962	1.0	1.0	Pizza Place	Hotel	Café	Breakfast Spot

Results

Consolidated Table with Analysis Results

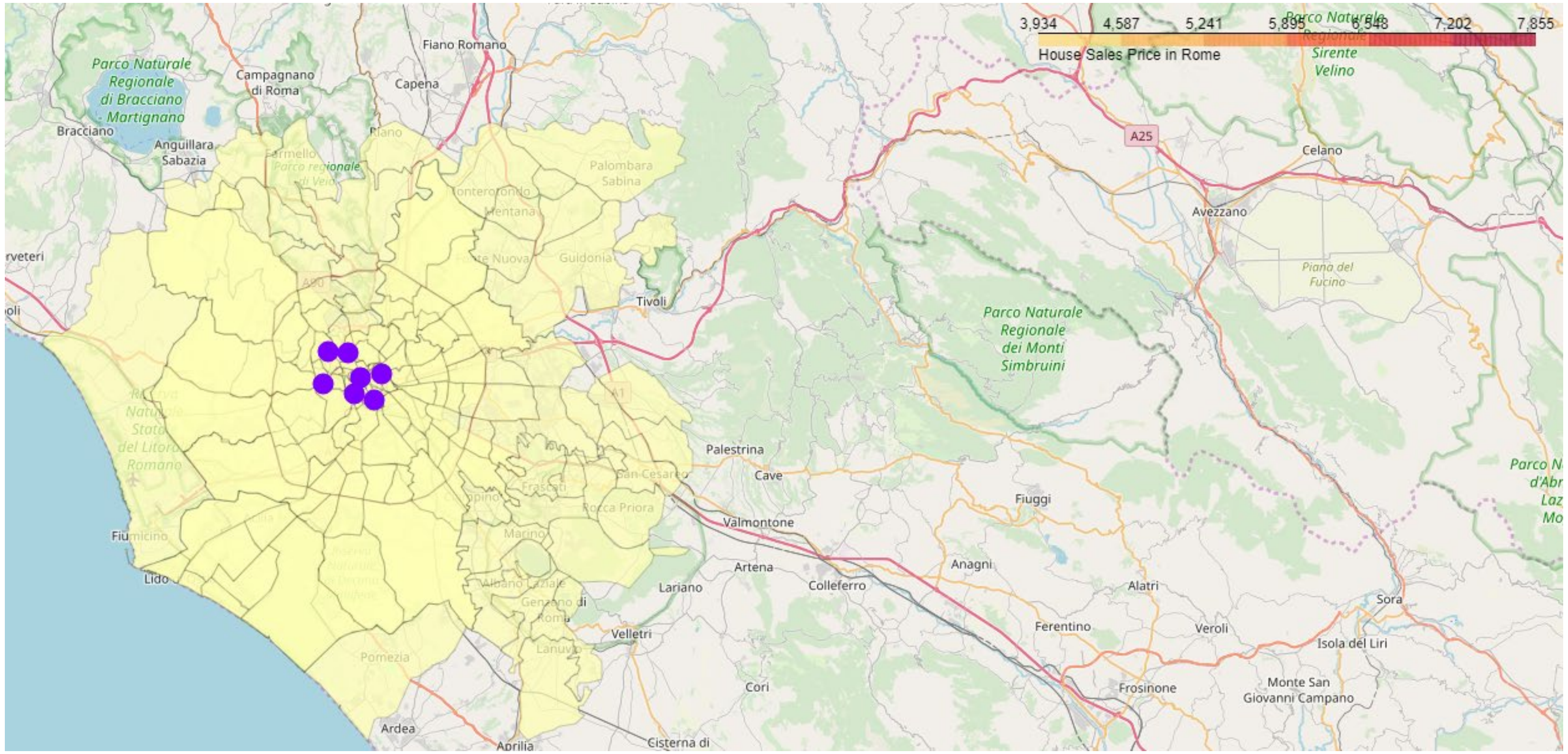
We can now add house sales price details to the cluster table that also include the top venue list by neighborhood



Results

Visualizing the Clusters Using Folium

We can now visualize the clusters created earlier using Folium:



Results

Overlaying House Sales Prices as Choropleth Map

We then overlay house sales prices on the cluster map using choropleth and json file containing coordinates of Rome that I obtained from

Summary of Analysis

- As a summary of my analysis, I used a data set containing the names of different neighborhoods for Rome as well as the average house price and longitude and latitude coordinates of those neighborhoods. I further augmented this dataset by using Foursquare API, to bring in details of the most common venues in each neighborhood.
- I used K-mean algorithm and elbow method that recommended segmenting the data into 2 clusters overall where Camillucia neighborhood ended up in Cluster 0 and remaining neighborhoods ended up in Cluster 1. When I analyzed the most common venues in each neighborhood, it also became quite apparent that while in Cluster 1, there are a number of italian restaurants and other socialization venues like cafes and hotels, Cluster 0 mainly had night clubs and wine bars.
- When I further visualized the data by overlaying the real estate prices, Cluster 0 looks even more attractive from a business investment perspective as the real estate prices fall in the mid level sales price range.

Conclusion

- Based on the above analysis, I recommend Cluster 0 (Camillucia neighborhood) as a good option for business investment and more specifically opening up an Italian restaurant or Pizza joint as there is minimum to no competition for food locations for night club goers and the real estate price is in the mid-range.

References

- **[1] Rome – Statista:** Real estate prices by different neighborhoods of Rome as of December, 2019 [https://www.statista.com/statistics/670698/asking-price-for-properties-for-sale-in-rome-by-area-italy/].
-
- **[2] Carto Rome:** .json file for Rome neighborhoods [https://maurizioman.carto.com/tables/rome_admin/public/map].
-
- **[3] Foursquare API** for most common venues of by boroughs in Rome [https://foursquare.com/].