Housing Sales Prices & Business Opportunities in Rome

IBM Data Science Professional Capstone Project

Battle of Neighborhoods

Daneyal Anis

March 28, 2020

Table of Contents

[1. Introduction 3](#_Toc36309831)

[2. Data Description 3](#_Toc36309832)

[3. Methodology 3](#_Toc36309833)

[3.1 Getting Geocodes and Creating Data-frame 3](#_Toc36309834)

[3.2 Visualizing Geographic Details and Boroughs – Using Folium and Geopy 4](#_Toc36309835)

[3.3 Exploring Boroughs in Rome and Common Venues – Using FourSquare API 4](#_Toc36309836)

[3.4 Analyzing Each Borough 6](#_Toc36309837)

[3.5 Clustering of Boroughs based on Venue Types Using K-Means 6](#_Toc36309838)

[3.6 Analyzing House Prices by Borough 9](#_Toc36309839)

[4. Results 11](#_Toc36309840)

[4.1 Consolidated Table with Analysis Results 11](#_Toc36309841)

[4.2 Visualizing the Clusters Using Folium 13](#_Toc36309842)

[4.3 Overlaying House Sales Prices as Choropleth Map 13](#_Toc36309843)

[5. Summary of Analysis 14](#_Toc36309844)

[6. Conclusion 14](#_Toc36309845)

[7. References 15](#_Toc36309846)

# 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

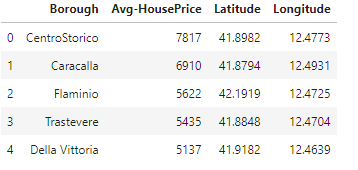
For this analysis, I used the following data sources:

* I obtained 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 obtained 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

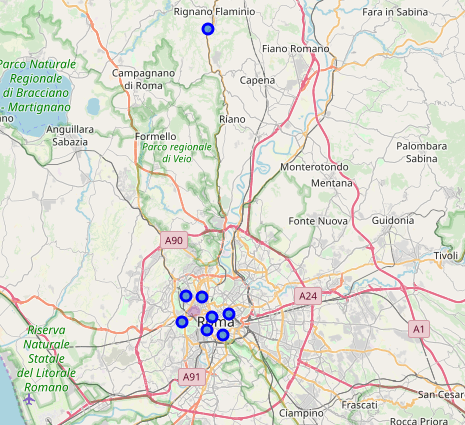
## Getting Geocodes and Creating Data-frame

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\*.



## 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.



## Exploring Boroughs in Rome and Common Venues – Using FourSquare API

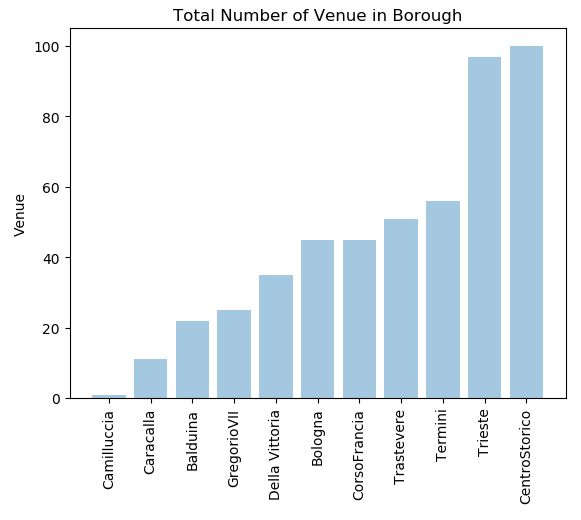
I then used FourSquare API to get details of the nearby venues in each borough of Rome.

Overall FourSquare API returned the following results.

1. 100 different categories of venues in Rome
2. Counts of venues by borough – snapshot below for reference:



1. I then visualized these results as a bar chart



The above 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.

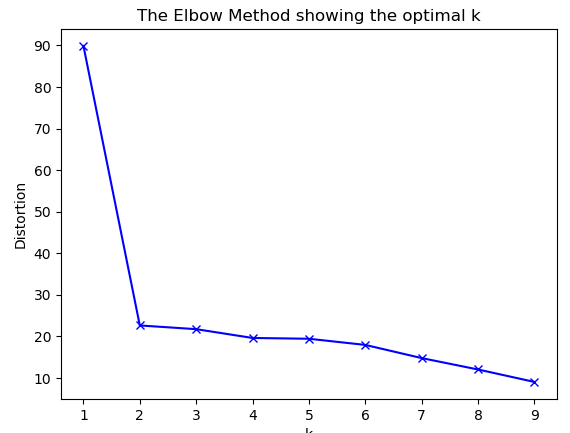
## Analyzing Each Borough

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.



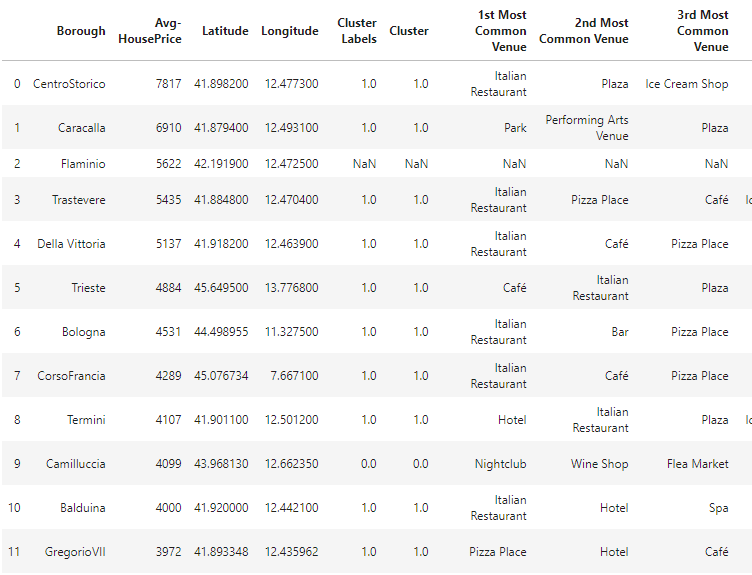
## Clustering of Boroughs based on Venue Types Using K-Means

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:



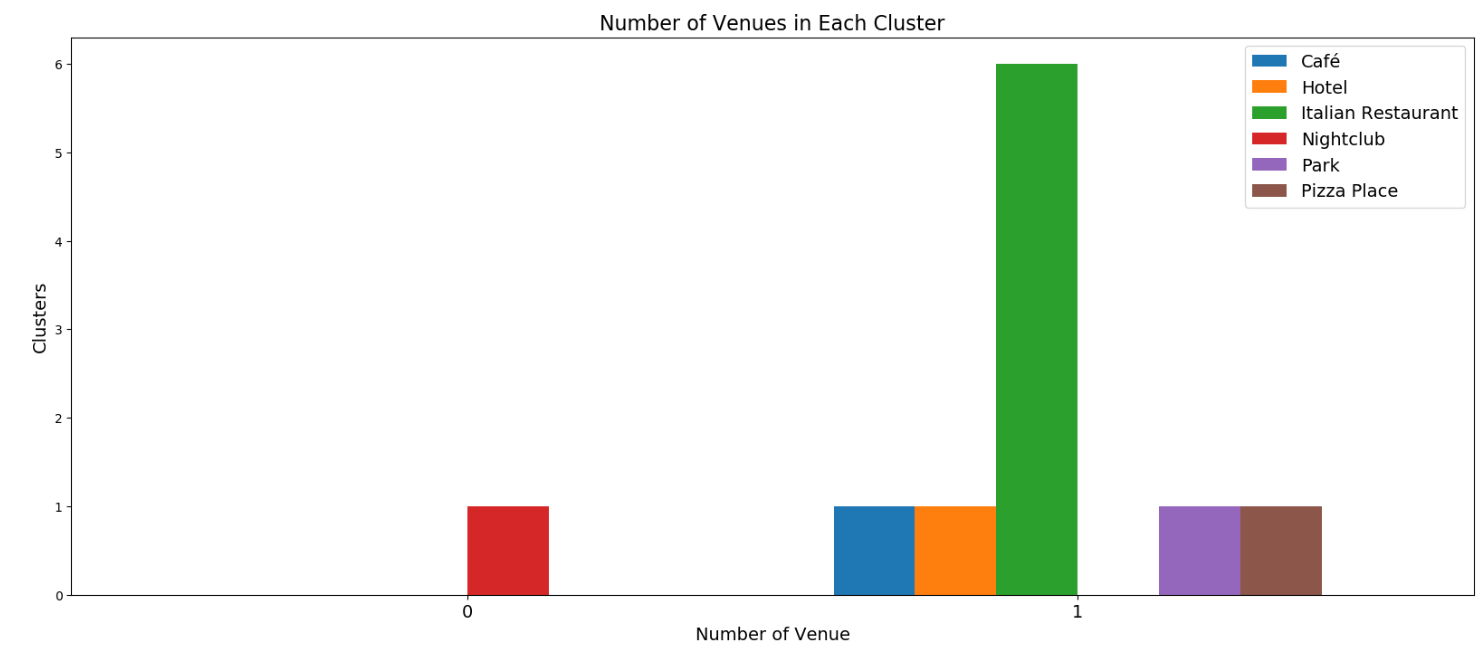
Elbow method above recommends the optimal number of clusters as 2 for the K-means analysis.

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



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 below:





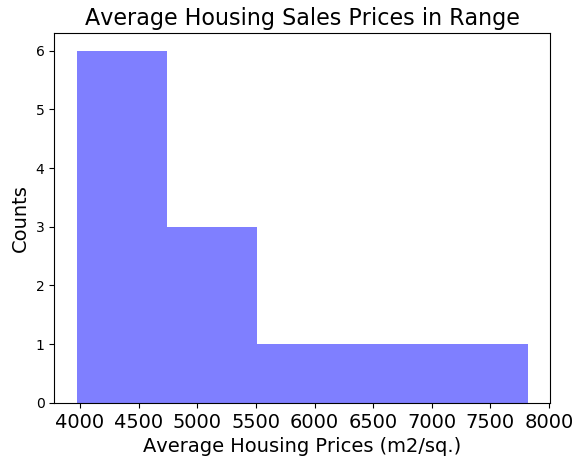
We can also present counts of different venue types in each neighborhood as follows:



## 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.





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:



# Results

## Consolidated Table with Analysis Results

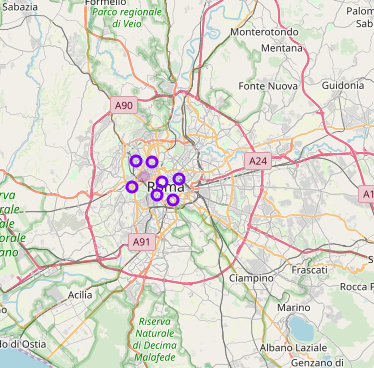
We can now add house sales price details to the cluster table that also include the top venue list by neighborhood:





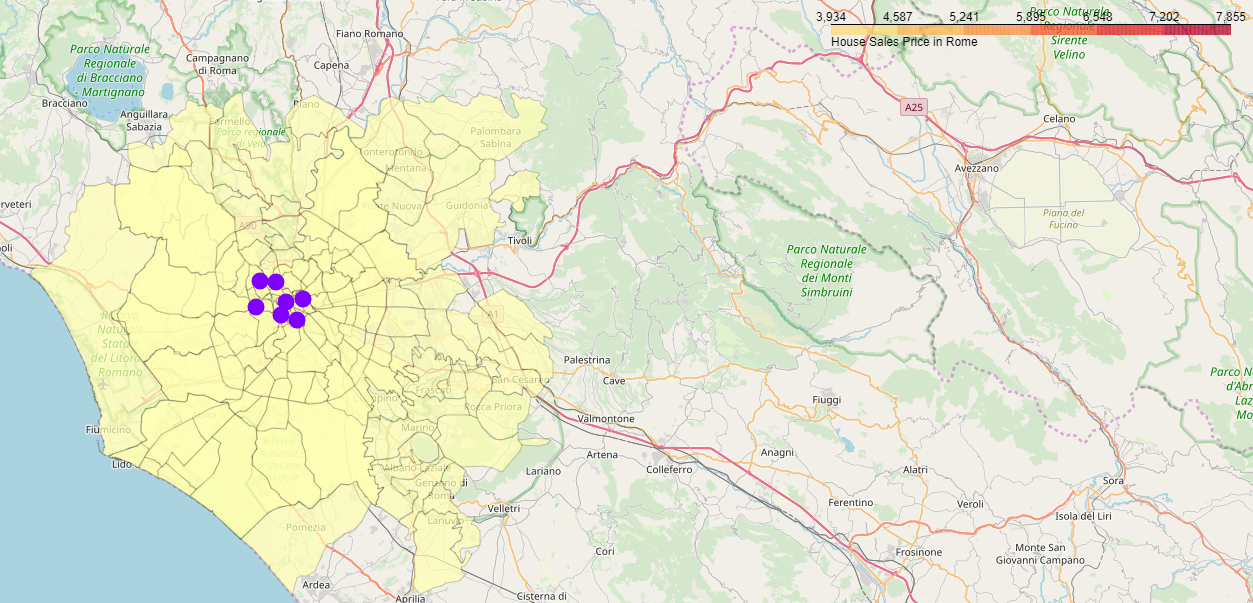
## Visualizing the Clusters Using Folium

We can now visualize the clusters created earlier using Folium:



## 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 <https://maurizioman.carto.com/tables/rome_admin/public/map>



# 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/>].