The Battle of Neighborhoods -IBM Capstone Project

Opening a new wine bar in Rome, Italy

Student:

Lucia Casucci

Date:

April 2020

Course:

IBM Capstone Project



# Background Discussion

Rome is the capital city and a special jurisdiction of Italy with 2,9 Million residents. Vatican City (the smallest country in the world) is an independent country inside the city boundaries of Rome. While Roman mythology dates the founding of Rome at around 753 BC, the city successively became the capital of the Roman Kingdom, the Roman Republic and the Roman Empire. After the fall of the Empire in the west, Rome slowly fell under the political control of the Papacy, which lasted until 1870. Beginning with the Renaissance, the Papacy pursued a coherent architectural and urban renovation over four hundred years, aimed at making the city the artistic and cultural center of the world. For this reason, Rome became first one of the major centers of the Italian Renaissance, and then the birthplace of both the Baroque style and Neoclassicism. Famous artists, painters, sculptors and architects made Rome the center of their activity, creating masterpieces throughout the city.

In 2019, Rome was the 11th most visited city in the world, third most visited in the European Union, and the most popular tourist destination in Italy. Rome is also the seat of several specialized agencies of the United Nations, such as the Food and Agriculture Organization (FAO), the World Food Program (WFP) and the International Fund for Agricultural Development (IFAD). The city also hosts the headquarters of many international business companies such as telecommications, aerospace, energy, pharmaceutical, and finance. The presence of renowned international brands in the city have made Rome an important center of fashion and design, and the Cinecittà Studios have been the set of many Academy Award–winning movies.

# Business Problem Description

In this assignment, I will explore, segment, and cluster the neighborhoods in the city of Rome to solve a business problem. I am interested in finding out what is the best location to open a modern wine bar.

The project will benefit any business personnel or freelancer that is interested in opening or investing in a wine bar in Rome. This analysis will be the first step to point out the best area of the city with the biggest potential for success in opening that business activity.

# Methodology

## Data sources

For the Rome neighborhood data, a Wikipedia page exists that has all the information we need to explore and cluster the neighborhoods (that named in Italian Municipi) in the capital of Italy (<https://it.wikipedia.org/wiki/Municipi_di_Roma>).

I will scrape the Wikipedia page and wrangle the data, clean it, and then read it into a pandas data frame so that it is in a structured format.

I will convert addresses into their equivalent latitude and longitude values using the geocoder library. I will use the Foursquare API (https://developer.foursquare.com/docs) to explore neighborhoods in Rome. I will use the explore function to get the most common venue categories in each neighborhood, check out their ratings, and the overall count of venues to understand the traffic.

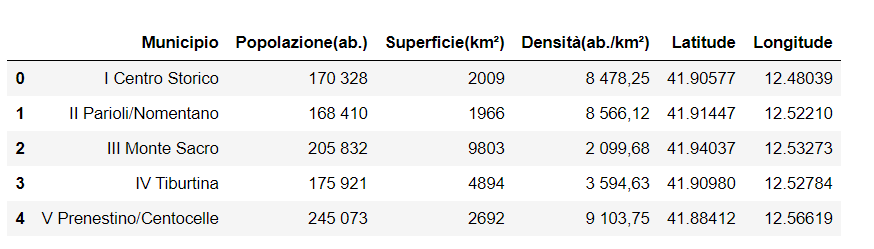
I will then use this feature to group the neighborhoods into clusters. I will use the k-means clustering algorithm to complete this task.

Finally, I will use the Folium library to visualize the neighborhoods in Rome and their emerging clusters, and identify the best one to open a modern wine bar.

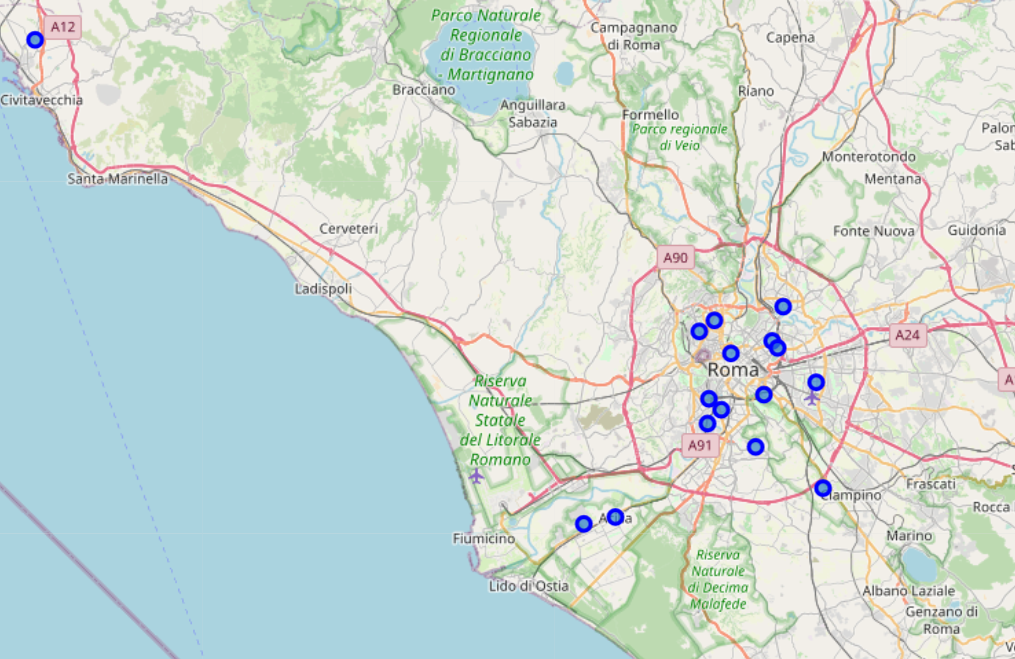
## Data Scraping and Cleaning

Using beautiful soup library for python, I was able to extract the Wikipedia page with the relevant data and information about the Neighborhoods in Rome.

The Wikipedia page had some information on the population, surface area, and director of each jurisdiction. After that, I used the python library geocoder to retrieve the latitude and longitude for each of these neighborhoods.

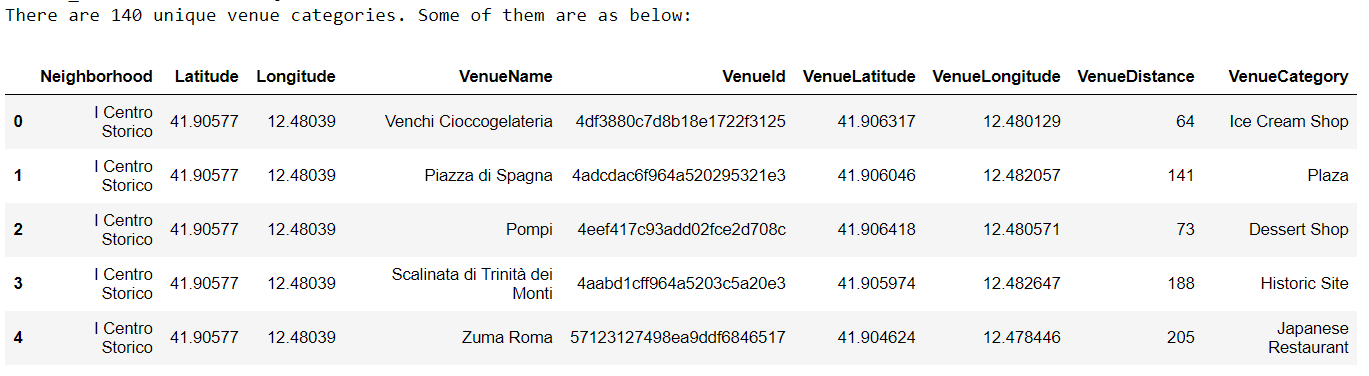


With the help of the folium library, I was able to plot each neighborhood as a mark to visualize the overall distribution in the city map.



## Problem solution

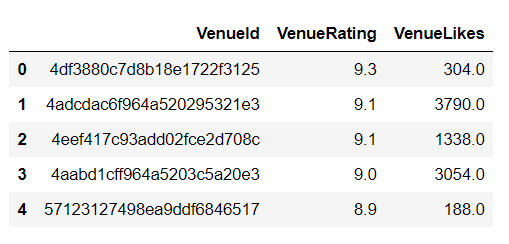
Foursquare API was necessary to get several important pieces of information to draw the conclusions of this analysis. With foursquare you can retrieve the venues of each neighborhood within a determinate radius, their ratings, and their likes by using a relatively short amount of code. After running the foursquare command to get at the most 100 venues per neighborhood, a dataframe listing their names, their distance by the neighborhood center, and their category was obtained.



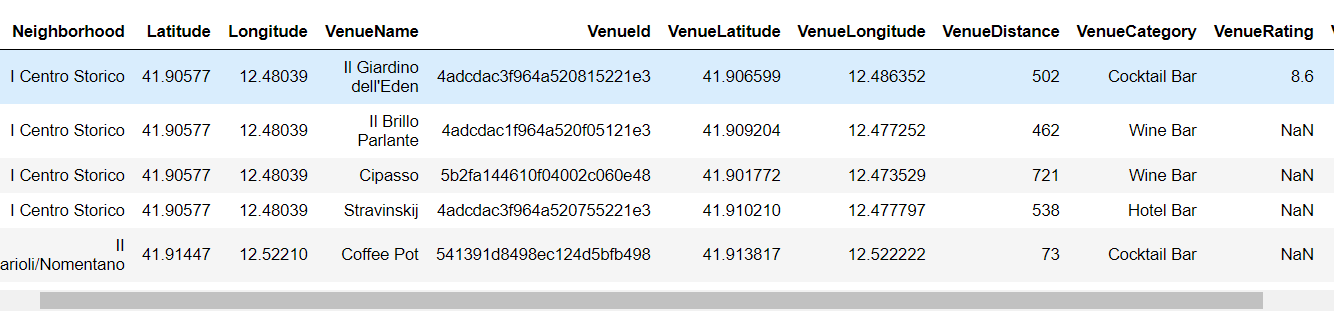
With the help of additional data wrangling, it was possible to see specifically which were the ten most common venues for each neighborhood, as shown in the snippet below for example.



Using the foursquare commands get rating and get likes, it was possible to find out exactly the number of likes and the ratings of the venues that we collected with the get command earlier, by associating them through each of their ID.

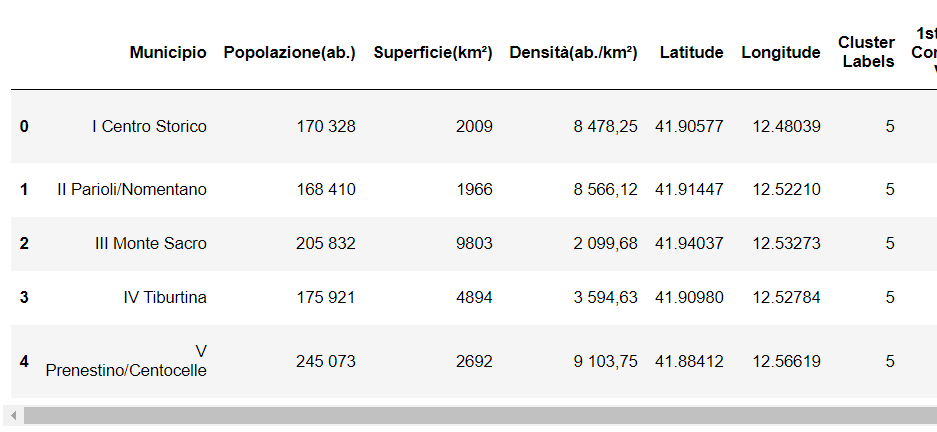


With a simple merge command, the dataframe above was joined to the first one that we obtained with the get venues command. A significant amount of venues did not have ratings over the dataset unfortunately, so it will not be possible to factor this point into our analysis as a crucial parameter.

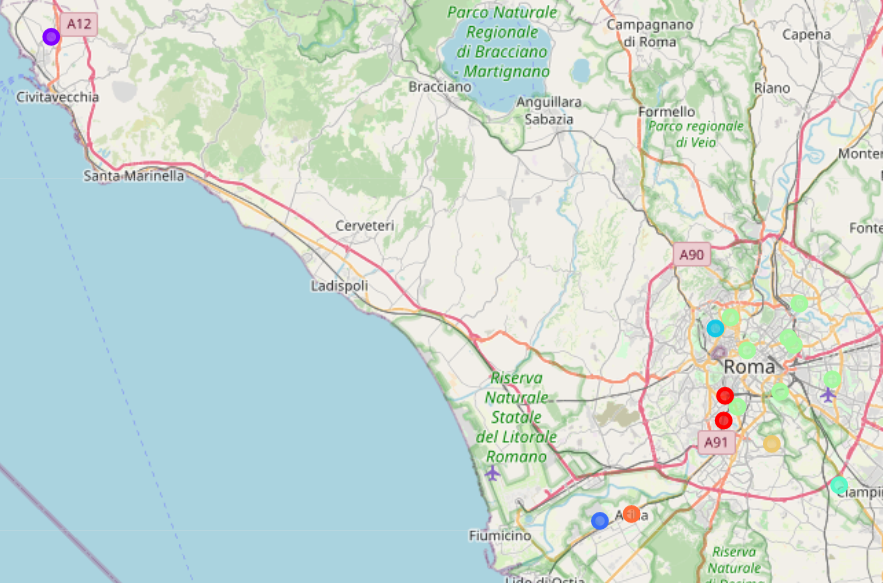


Bar charts were helpful to visualize the counts, mean ratings, and likes of the venues by Neighborhood. The goal of most of the visualizations listed in the discussion results are is to show how wine bars and similar joints are distributed in each neighborhood.

In order to group by the Neighborhood list in a more concise way and have a more general understanding of the areas of the city where it would be better to open a wine bar, a KNN algorithm was run to aggregate them in clusters. The number of optimal clusters to assign the neighborhoods to was found out to be 8.



With the help of folium, it was possible to see how these clusters are grouped on the map and how the algorithm aggregated the neighborhoods together.

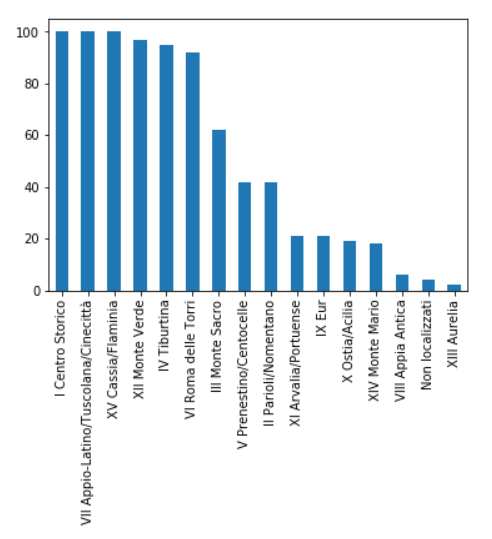


A similar investigation to the one done on the more granular dataset was performed, visualizing the counts, mean ratings, and likes of the venues at a cluster level. However, in this case, the goal of most of the visualizations listed in the discussion results are is to show how wine bars and similar joints are distributed in each neighborhood. The results of the analysis are shown in the following section.

# Discussion and Results

Analysis and discussion of the neighborhoods

The graph below shows the amount of venues returned by foursquare per neighborhood. We can see that almost 80% of the venues are concentrated in 6 neighborhoods. From this it is possible to infer that these neighborhoods have a higher traffic and flux of people than the others.

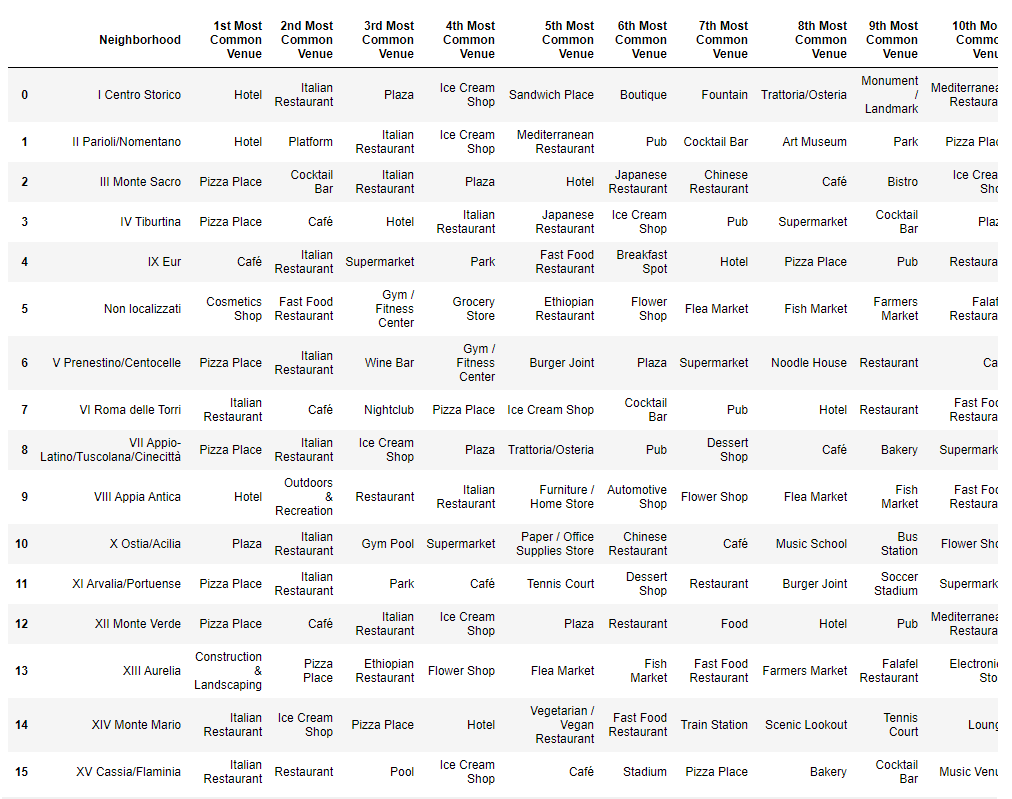


From the Wikipedia page it is possible to retrieve the demographics information of each neighborhood. Appio Latino/Tuscolana/Cinecitta’ has the highest count of habitants, followed by Roma delle Torri and Prenestino/Centocelle, which has the highest density of population as well. Centro Storico, Parioli and Appio/Latino/Tuscolana/Cinecitta’ have some of the highest population density of the dataset as well. A high population density suggests a higher potential for traffic in a newly opened venue.

|  |  |  |  |
| --- | --- | --- | --- |
| **Neighborhood** | **Population (in thousands)** | **Area (in km^2)** | **Density (k people/km^2)** |
| I Centro Storico | 170 | 20.00 | 8.50 |
| II Parioli/Nomentano | 168 | 20.00 | 8.40 |
| III Monte Sacro | 206 | 98.00 | 2.10 |
| IV Tiburtina | 176 | 49.00 | 3.59 |
| V Prenestino/Centocelle | 245 | 27.00 | 9.07 |
| VI Roma delle Torri | 258 | 114.00 | 2.26 |
| VII Appio-Latino/Tuscolana/Cinecittà | 307 | 46.00 | 6.67 |
| VIII Appia Antica | 131 | 47.00 | 2.79 |
| IX Eur | 184 | 183.00 | 1.01 |
| X Ostia/Acilia | 232 | 151.00 | 1.54 |
| XI Arvalia/Portuense | 156 | 72.00 | 2.17 |
| XII Monte Verde | 141 | 73.00 | 1.93 |
| XIII Aurelia | 134 | 64.00 | 2.09 |
| XIV Monte Mario | 192 | 134.00 | 1.43 |
| XV Cassia/Flaminia | 161 | 187.00 | 0.86 |

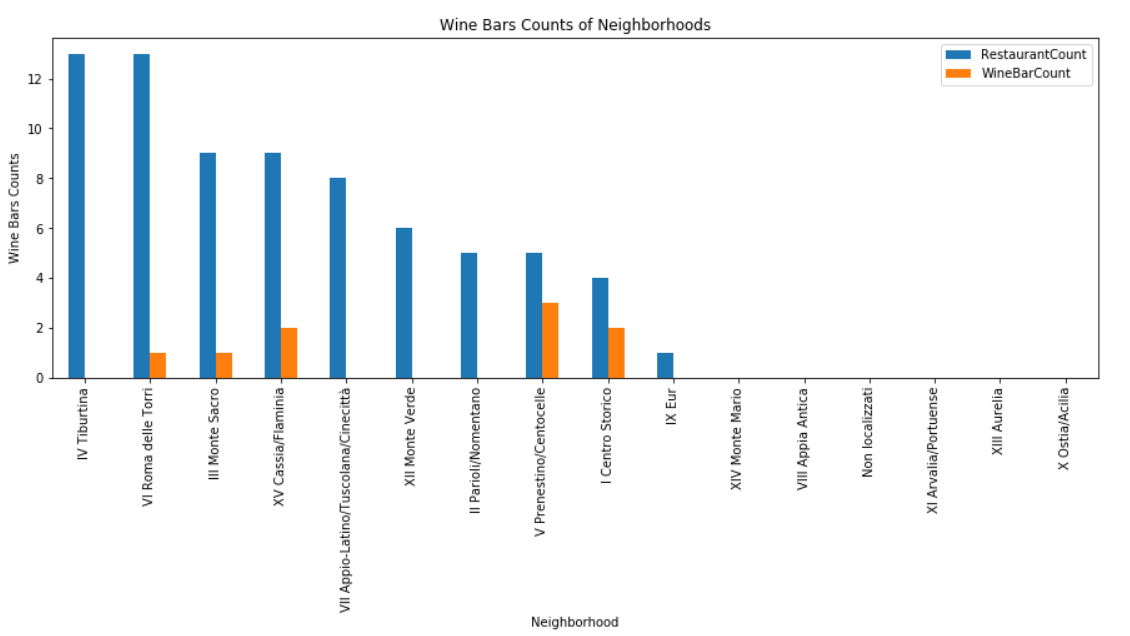
Since Rome is the capital of Italy, it was not surprising to see how Italian restaurants, pizza places, cafes, and ice cream shops are the most common venues across the whole city. However, at a neighborhood level, it is possible to notice how there plenty of other different type of venues that come up in the top 10 (art museum, supermarket, Park, and other types of ethnic restaurants).





Since the goal of this study is to identify what are the top 3 places to open a wine bar in the city of Rome, it is imperative to check how many wine bars and late night drink venues (like cocktail bars, wine bars, happy hour venues, pubs) are located in each neighborhood and in the city as a whole. People that are looking for grabbing a drink might be targeting this type of places, and it is important to make sure that our venue will not be in an area where a lot of other type of late-night drinks are already up and running.

In the city of Rome there are in total 68 venues that serve late night drinks between wine bars, pubs, cocktail bars, sports bars, and breweries. 9 of those 68 venues are the wine bars that are currently open. It is possible to see how there are wine bars only in 5 neighborhoods, while 10 of them have venues to grab late-night drinks. A lot of high traffic neighborhoods do not have many wine bars and late night joints where people can go grab for a drink. Appia/Tuscolana/Cinecitta’ and Monte Verde have a lot of venues and high traffic, do not have any wine bar yet and they have relatively few options for late night drinks. Parioli could be another location, as well as Tiburtina since even if there are other options, there is not a single wine bar yet.



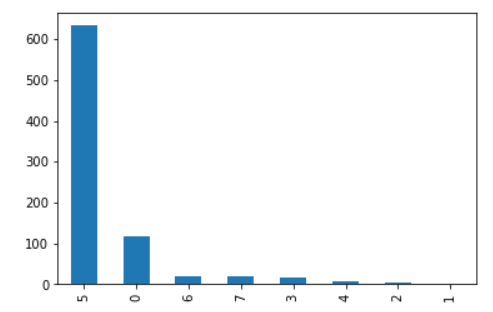
## Analysis and discussion of the clusters

The KNN algorithm suggested that 8 was the optimal number of clusters. The tables below summarize how the neighborhoods were assigned to the clusters, the population size, surface area, and density for each cluster.

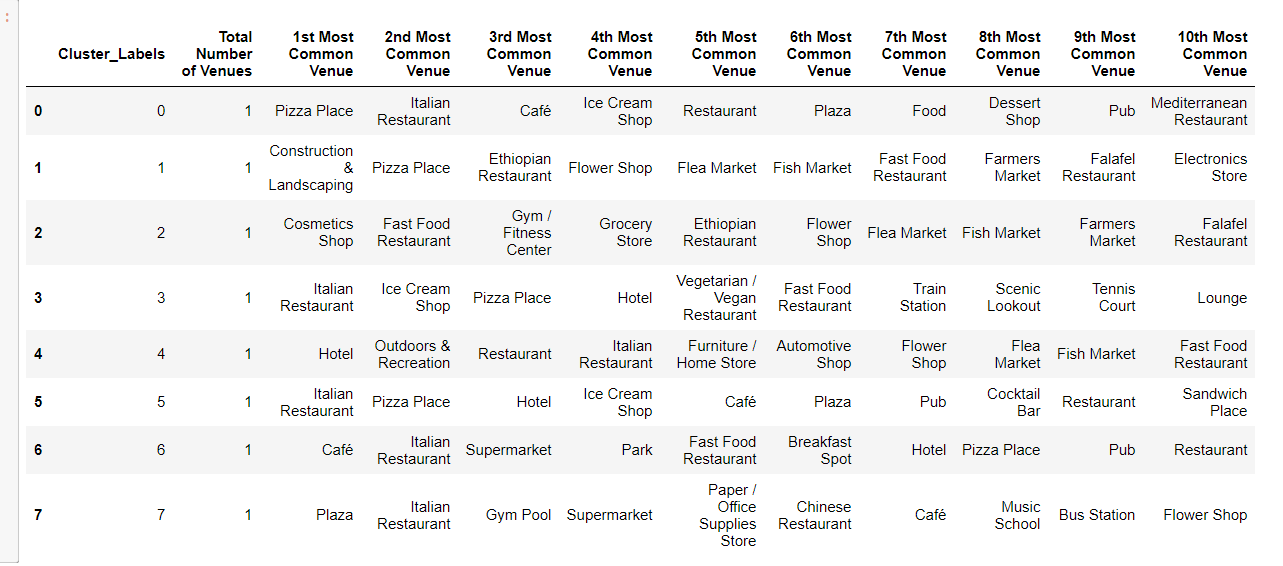
|  |  |
| --- | --- |
| **Neighborhood** | **Cluster** |
| I Centro Storico | 5 |
| II Parioli/Nomentano | 5 |
| III Monte Sacro | 5 |
| IV Tiburtina | 5 |
| V Prenestino/Centocelle | 5 |
| VI Roma delle Torri | 5 |
| VII Appio-Latino/Tuscolana/Cinecittà | 5 |
| VIII Appia Antica | 4 |
| IX Eur | 6 |
| X Ostia/Acilia | 7 |
| XI Arvalia/Portuense | 1 |
| XII Monte Verde | 0 |
| XIII Aurelia | 3 |
| XIV Monte Mario | 5 |
| XV Cassia/Flaminia | 2 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Cluster** | **Population in thousands** | **Area in km^2** | **Density (k people/km^2)** |
| 0 | 141 | 73 | 0.52 |
| 2 | 161 | 187 | 1.16 |
| 3 | 134 | 64 | 0.48 |
| 4 | 131 | 47 | 0.36 |
| 5 | 1722 | 508 | 0.30 |
| 6 | 184 | 183 | 0.99 |
| 7 | 232 | 151 | 0.65 |
| 1 | 156 | 72 | 0.46 |

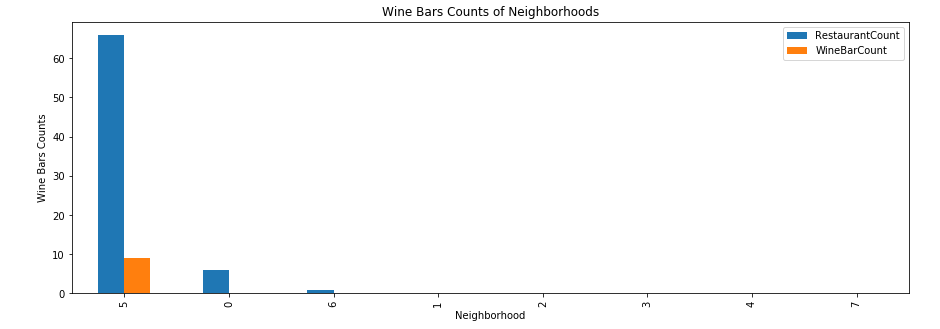
Cluster 5 Is by far the biggest by population count and by surface area, and it corresponds to majority of the city center. Observing how the venues are allocated in the eight clusters, it is possible to see how cluster 5 has slightly more than 2/3 of the venues of the whole dataset. However, cluster 6 and cluster 2 have by far the highest population density.



Each cluster has slightly different top 10 venue types as well, even though Italian restaurants, cafes, and pizza places are always the most recurring ones. Most of the likes and ratings are for venues located in cluster 5.



It was expected to see cluster 5 having the highest number of late-night drink venues. However, it is the only clusters where there are wine bars. Wine bars are only 7% of total late-night venues. Cluster 0 has about 10 late-night drink venues and no wine bar at all, resulting as a potential good option to explore to open a new wine bar.



# Conclusion

After reviewing both the granular and the overall analyis results, I would recommend three areas as the best ones to open a wine bar in Rome: Cluster 0/ Neighborhood Monte Verde, Cluster 5/ Neighborhood Tiburtina, and Cluster 5/Neighborhood Appia/Tuscolana/Cinecitta’.

Cluster 5’s Appia/Tuscolana/Cinecitta’ is the top option to open a wine bar in Rome. It has the highest population count and one of the highest densities of the whole dataset. Therefore, there is a higher potential for locals to visit the wine bar compared to the other 2 choices below. It has the highest count of venues and majority of the top 10 are food joints, indicating a high traffic. There is no wine bar yet, even though there are some late-night drink options.

Monte Verde is one of the neighborhoods with the biggest count of venues and therefore high traffic. Since its population density and count is within the average of the dataset, most likely a lot of people gather or go there to visit its restaurants and venue. It is possible to see that its top venues are restaurants and food joints, and It would be good to add to an already high traffic neighborhood for food joints a wine bar venue. It has no wine bars yet, and there are relatively little other late-night drink options compared to the other neighborhoods. This makes Monte Verde the second top option to open the wine bar.

Cluster 5’s neighborhood Tiburtina is the third top option to open a wine bar. It has the one of highest counts of venues with a lot of the top 10 venues being food joints. It does not have a wine bar yet. Like Monte Verde, its population density and count are within the average of the dataset, so most likely a lot of people gather or go there to visit its restaurants and venue. However, it has 12 late-night drinks venues, double the ones of Monte Verde.

The next steps to select the top place to open our venue would be considering some socio-economic factors of these three neighborhoods, such as the cost of renting a space to convert a wine bar in those areas, as well as the age and income distribution of the people that usually visit their venues.

## Sources used: <https://en.wikipedia.org/wiki/Rome>

Pictures : Crawfordcreations.com,sacurrent.com