**A Chinese restaurant in Rome**

Contents

[**A.1. Description & Discussion of the Background** 1](#_Toc38109846)

[**A.2. Data description & methodology** 1](#_Toc38109847)

[**A.2.1 Data description** 1](#_Toc38109848)

[**A.2 Methodology** 3](#_Toc38109849)

[**A.3. Results and Discussion** 5](#_Toc38109850)

[**A.4. Conclusions** 7](#_Toc38109851)

[**References** 7](#_Toc38109852)

# **A.1. Description & Discussion of the Background**

Rome is one of the most attractive metropolis in the world. Founded according to the tradition in the 753 BC, during its three millennia of history, it has been the beating heart of one of the most ancient civilizations, which influenced society, culture, language, literature, art, architecture, philosophy, religion, law and customs of the following centuries. It is a city full of history that attracts millions of tourists every year.

I was able to discover various aspects of this city having lived here for 4 years, for this reason I have decided to use Rome for my project. The city has a population of about 3 million people divided in 15 Boroughs called “*Municipi*”. With a surface of 1 287,36km^2, it is the fifth largest city in Europe. The city, therefore, has a varied urban structure with the presence of different ethnic groups.

Rome can be the ideal place to open a shop like a restaurant but it can hide many pitfalls at the same time. Indeed, the splendor and monumentality of the old town is opposed to the suburbs characterized by a very different social condition. In such a complex scenario, in the present work I have tried to answer to the following question: *“****where could a typical Chinese restaurant be located?****”*

Many factors can guide this analysis. The location of the restaurant in a specific neighborhood with respect to another may depend on the cost of rentals, the presence of a large social density, the influx of tourists, the proximity to other activities of the same type or historic sites, the general situation of the neighborhood (comfort, reliability).

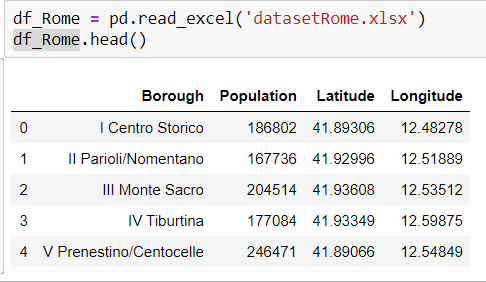
In the following a simple analysis using data science techniques have been carried out in order to find the best place to locate a Chinese restaurant. This work could therefore affect all people who intend to open a restaurant business of this kind in Rome.

# **A.2. Data description & methodology**

## **A.2.1 Data description**

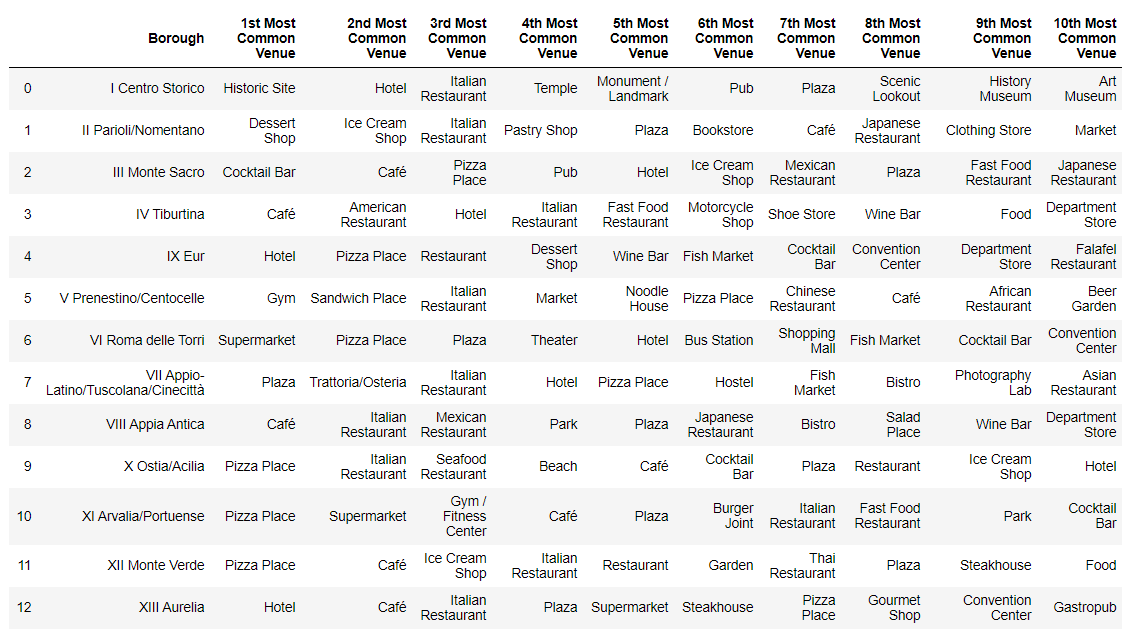
In the project the following data source have been considered:

* **Wikipedia.** From here it is possible to obtain general information on Rome in a reliable way **[1]** and at the same time, the administrative subdivision of the city **[2].** These data have been obtained manually from Wikipedia and disposed into a .xlsx datset. For simplicity, only the first entries of this dataframe have been displayed.

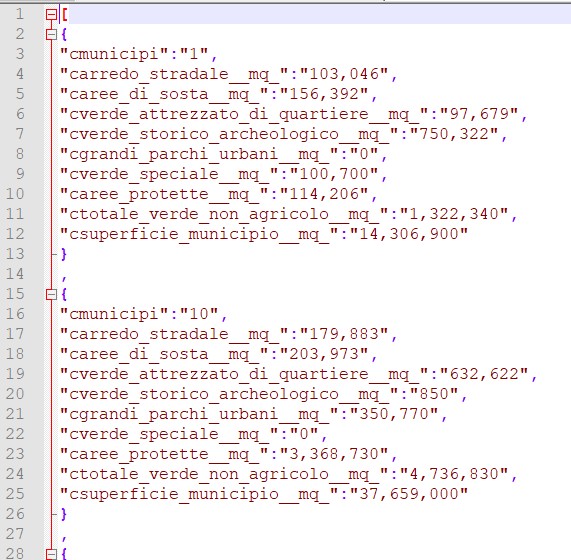


**FIG 1**. An image showing a dataframe with the name of each borough, the coordinates and the population. (source: own elaboration)

* **Forsquare API** allows to get the most common venues of a given Borough or Neighborhood of Rome **[3],** as shown in figure 2.



**FIG 2.** The image shows the top ten venue for each borough in the city of Rome, for simplicity only some lines of the dataframe are displayed (source: own elaboration).



* **DatiOpen.it.** It is a site that provides data about the city of Rome through multiple formats (csv, json ...). These data relate to social information, green areas, historical sites and points of interest such as libraries **[4]**. An exemple of dataframe is shown in figure 3.

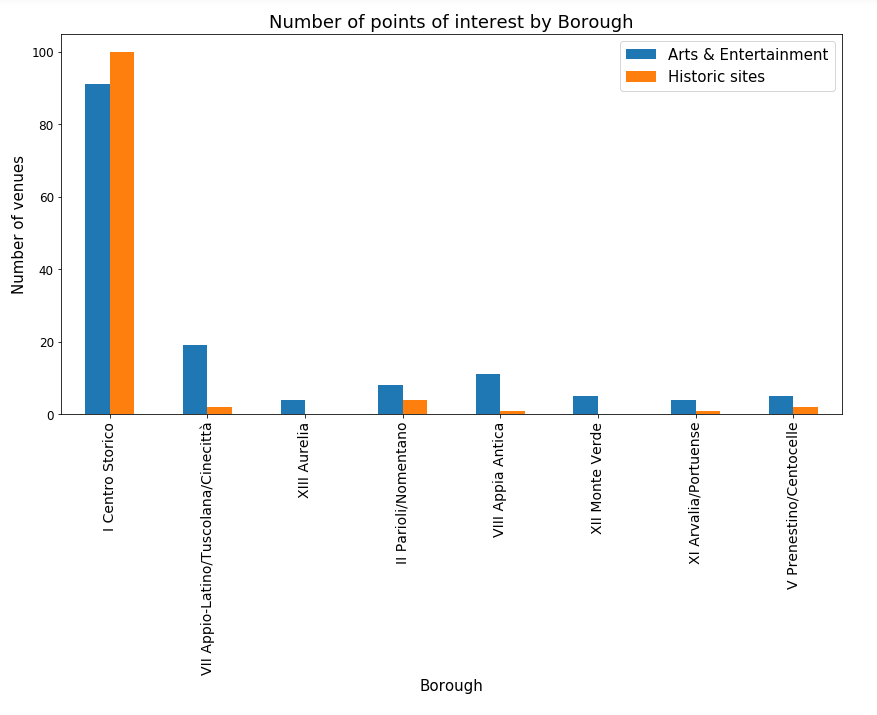
**FIG 3.** Example of Jason file containing data (source: own elaboration).

* **Google Maps [5].** Unfortunately, not all geographic data can be retrieved directly from the web. Sometimes I needed to use google maps and manually search for coordinates of specific points.

# **A.2 Methodology**

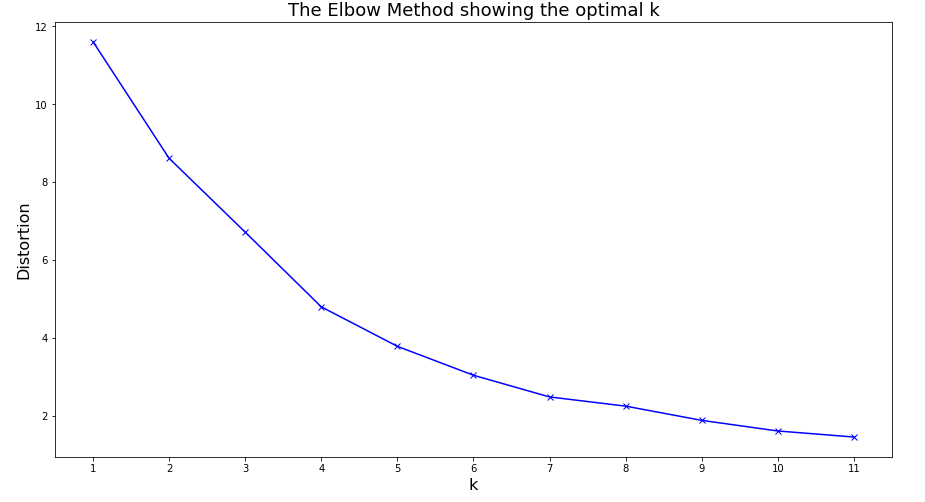
1. The first step has been to identify the Boroughs of Rome (as shown in image 4) and to find those of greatest interest. **Foursquare API** has allowed to get the most common venues of a given Borough. Those that were less attractive (lack of points of interest that can attract tourists as shown by bar chart in figure 5) and that were too far from the city center have been discarded. In this phase, the major problem was that there wasn’t a specific site that provides information about Rome’s borough in a way that could be exploited directly with python (for instance with libraries as beautiful soup). Therefore, this part took some offline work. A dataset with boroughs coordinates, their extensions and population has been created offline using data from **[1], [2], [4], [5], [6].**

**FIG 4**. An image showing Rome’s Boroughs. (source: own elaboration)



**FIG 5**. The image shows the number of arts & Entertainment venues and historic sites for each Borough. (source: own elaboration)

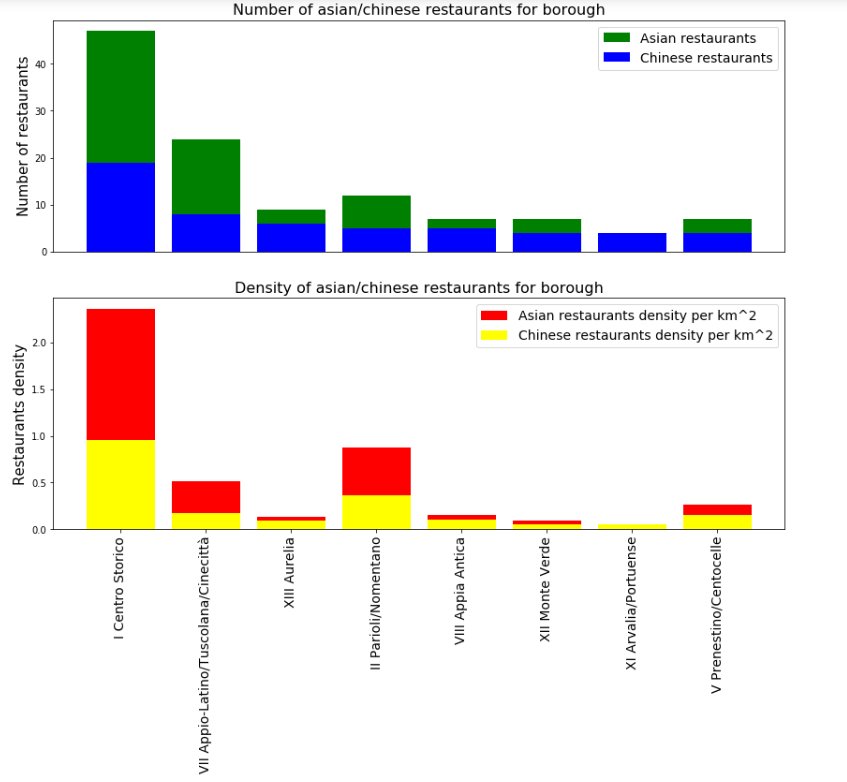
1. Once the most interesting boroughs have been identified, each single Borough have been analyzed individually. A list of Neighborhoods with their coordinates have been produced for each Borough (using a dataset created offline). Foursquare API has allowed again to get the most common venues of a given Neighborhood.
2. Thanks to the unsupervised machine learning technique of the **K-means algorithm**, it has been possible to identify similarities between areas and further reduce uncertainty about the future location of the restaurant. The optimal value of the k-means algorithm has been obtained using the elbow method as shown in figure 6.Three aspects have been considered, in particular: the distance of the selected area to the city center, the proximity to points of interest and the presence of similar restaurants in the neighborhood. The method allowed to consider a specific class. Finally, within the same class, the area that best matched the proposed criteria has been chosen.
3. At the end it has been possible to identify a candidate area to launch our business.



**FIG 6**. The figure shows the number of distortions with respect to the k value of the k-means clustering technique. The optimal value is k=7 (source: own elaboration)

# **A.3. Results and Discussion**

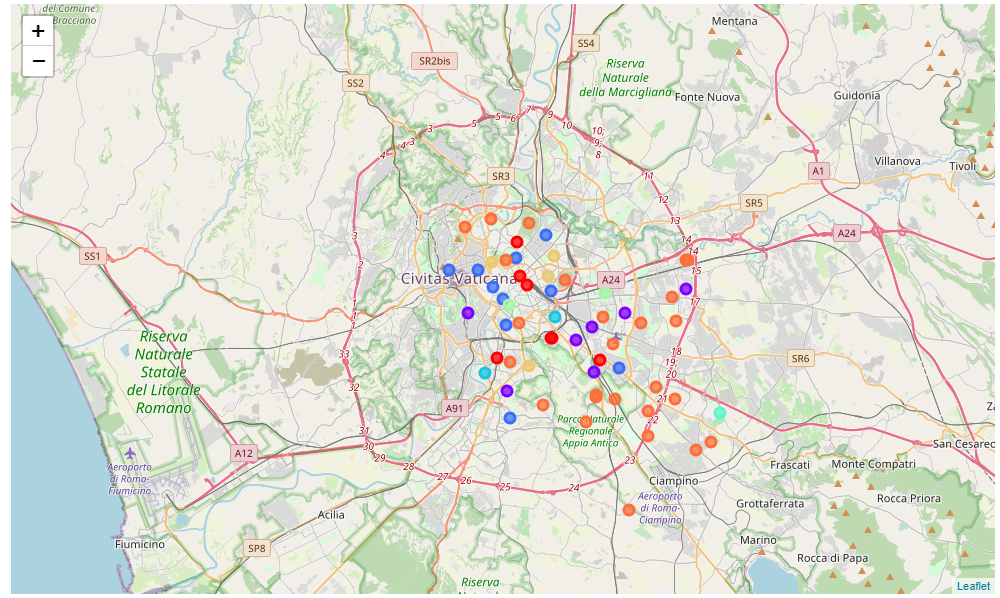
Our analysis has permitted to see how decisions can be made in the presence of numerous variables and factors. In particular, it has been observed that thanks to data science techniques it has been possible to identify an optimal position for a Chinese restaurant in a city as vast as that of Rome. Different boroughs of Rome have been first studied through bar charts that showed the number and density of Asian and Chinese restaurants in each area as shown in figure 7.



**FIG 8**. The figure shows the number of Asian/Chinese restaurants for specific boroughs. The second bar chart refers to density of Asian and Chinese restaurants. clusters on Rome Map. (source: own elaboration)

Once considered the most attractive Borough, unsupervised machine learning techniques was used and have proven very useful in managing a large number of data. By using the K-means algorithm we have identified 7 clusters (figure 8).

In particular, we have noticed that some neighborhoods did not fall into any of the clusters as they did not have any Asian/Chinese restaurants within the specified distance. We have therefore created a new class with all the neighborhoods that did not contain any Asian and in particular Chinese restaurants. We have studied this class and we have chosen the neighborhood closest to the historic center of the city.



**FIG 8**. The figure shows the identified clusters on Rome Map. (source: own elaboration)

The following analysis allowed to identify an optimal area for stakeholders. Indeed, the selected neighborhood that is the Neighborhood *“1G Celio”* is close to the historic city center and far from the competition of restaurant of the same type.

# **A.4. Conclusions**

Purpose of this project was to identify Rome areas close to center with a low number of Asian restaurants (in particular Chinese restaurant) in order to help stakeholders in choosing the optimal location for a new Chinese restaurant.

At first, we looked for the Rome's Boroughs that could be more interesting, looking at the number of historical sites, museums and art galleries for each Borough through Foursquare. Secondly the number of Asian restaurants, in particular Chinese and their density have been studied for the selected areas.

Once the most interesting boroughs have been identified, a more detailed geographical analysis has been carried out. We have considered neighborhoods for the selected boroughs and through Foursquare we have obtained the kind of asian restaurants that was present in those areas. Through the Kmeans algorithm it has been possible to identify similarities between different areas. In particular, it emerged that in some central areas there were no Chinese restaurants. Finally, we have considered these neighborhoods and we have chosen the closest one to the historic center.

For next works, the analysis may be improved by considering other data such as the average rental prices. Furthermore, some of the datasets used in the following project have been created offline by taking information from different internet web sites, in the future it will be possible to think of automating this process.

# **References**

**[1]** <https://en.wikipedia.org/wiki/Rome>

**[2]** <https://en.wikipedia.org/wiki/Administrative_subdivision_of_Rome>

**[3]** <https://developer.foursquare.com/>

**[4]** <http://www.datiopen.it/it/catalogo-opendata/daticomuneromait>

**[5]** <https://www.google.fr/maps>