

# The best place for an Italian Restaurant in Rome

An Applied Data Science Capstone Project



## 1. Introduction

Rome is a city rich in history and monuments. It is one of the cities with the greatest tourist attraction. Consequently, it has a large presence of tourists and therefore of potential customers during all the year. The opening of an Italian restaurant in Rome, therefore, represents a possible highly profitable activity. Now, the question is: where would be the best place they should open it? In this project we will try to find an optimal location for an italian restaurant. Specifically, this report will be targeted to stakeholders interested in opening an **Italian restaurant in Rome**, Italy. Since there are lots of restaurants in Rome we will try to detect **locations that are not already crowded with italian restaurants**. We are also particularly interested in **areas close to the city center**. The proximity to the center of the city could represent a determining factor given the greater number of tourists.

## 2. Data acquisition and Cleaning

### Data

Rome is divided into municipalities that have their own local administration ([https://it.wikipedia.org/wiki/Municipi\\_di\\_Roma](https://it.wikipedia.org/wiki/Municipi_di_Roma)). A list of all the municipalities and their geographical limits can be obtained (**municipi.geojson**) from the web page: <https://geonue.com/mappa-interattiva-municipi-di-roma/>. I decide to concentrate the study to the Central Municipality (Municipio I- Rome Center) because it is the historical heart of the city. In particular Municipio I is composed by 22 sub-regions called **rioni**.

A **rione of Rome** is a traditional administrative division of the city of Rome. "Rione" is an Italian term used since the 14th century to name a district of a town. The term was born in Rome, originating from the administrative divisions of the city. The word comes from the Latin word regio (pl. regiones, meaning region); during the Middle Ages the Latin word became rejones, from which rione comes. Currently, all the rioni are located in Municipio I of Rome.

A list of all the rioni of Rome can be obtained scraping the relative Wikipedia's page ([https://it.wikipedia.org/wiki/Rioni\\_di\\_Roma](https://it.wikipedia.org/wiki/Rioni_di_Roma)). The approximate positions of each rione of Rome can be derived from Google Maps.

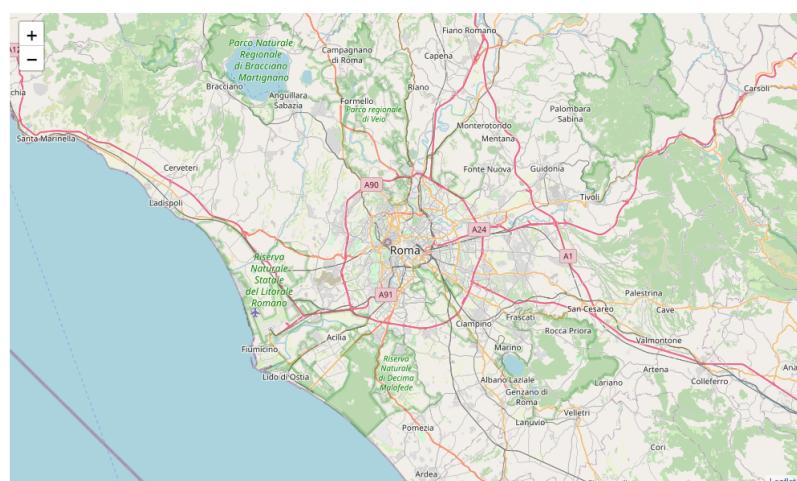
Based on definition of our problem, factors that will influence our decision are:

- number of existing restaurants in the area (any type of restaurant)
- number of Italian restaurants in the area
- distance from the city center

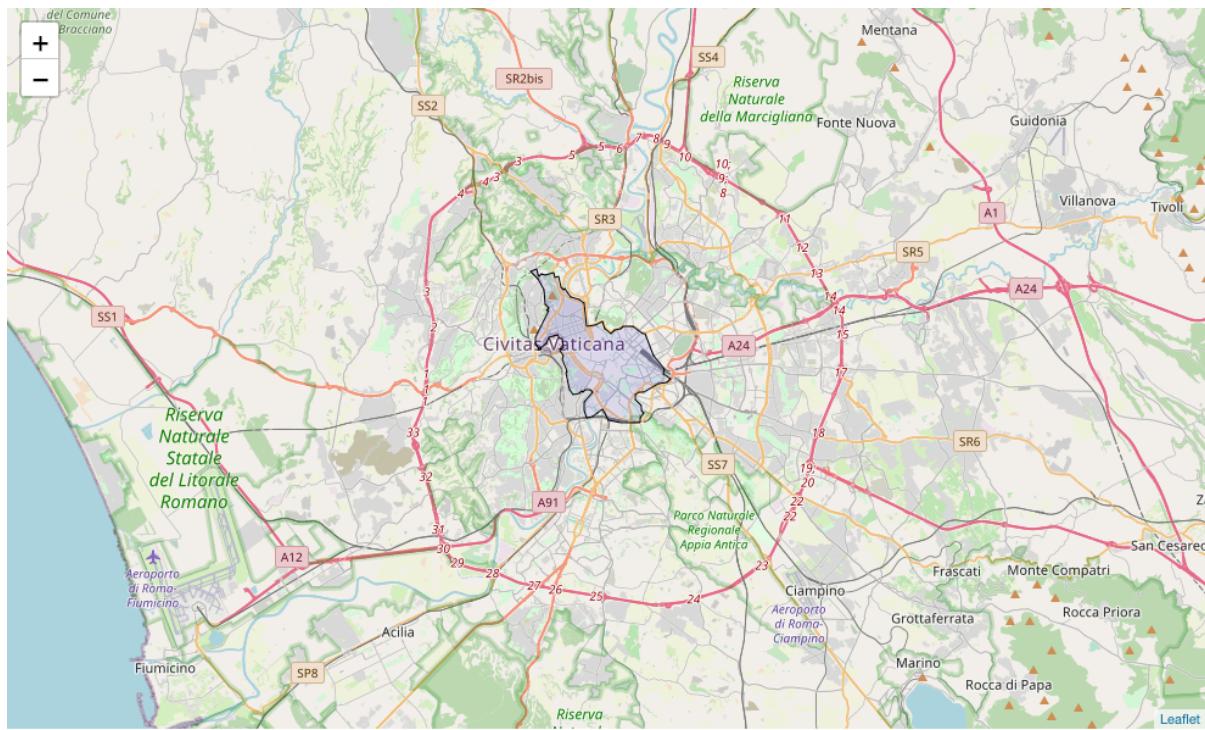
Number of restaurants and their type and location in every rione were obtained using **Foursquare API**.

### Neighborhood Candidates

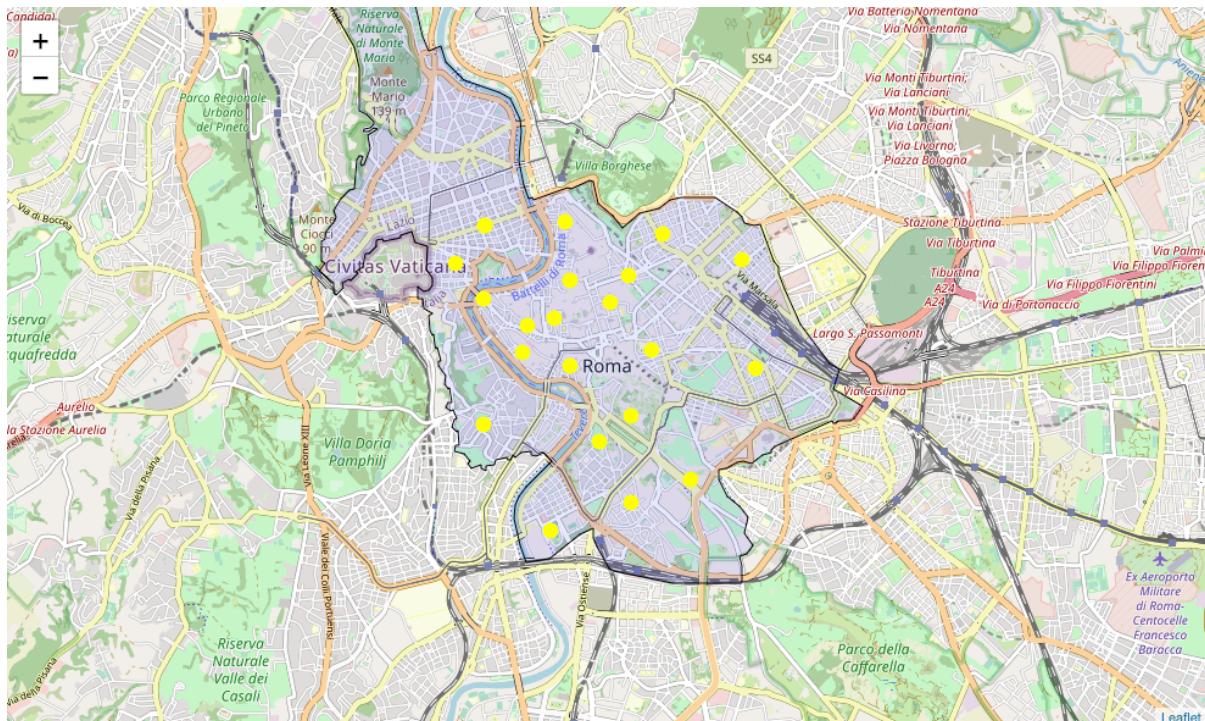
Let's first find the latitude and longitude of Rome city center, using geopy and plot a map of the city.



I decide to study the Central Municipality (**Municipio I - Rome Center**) because it is the historical heart of the city.

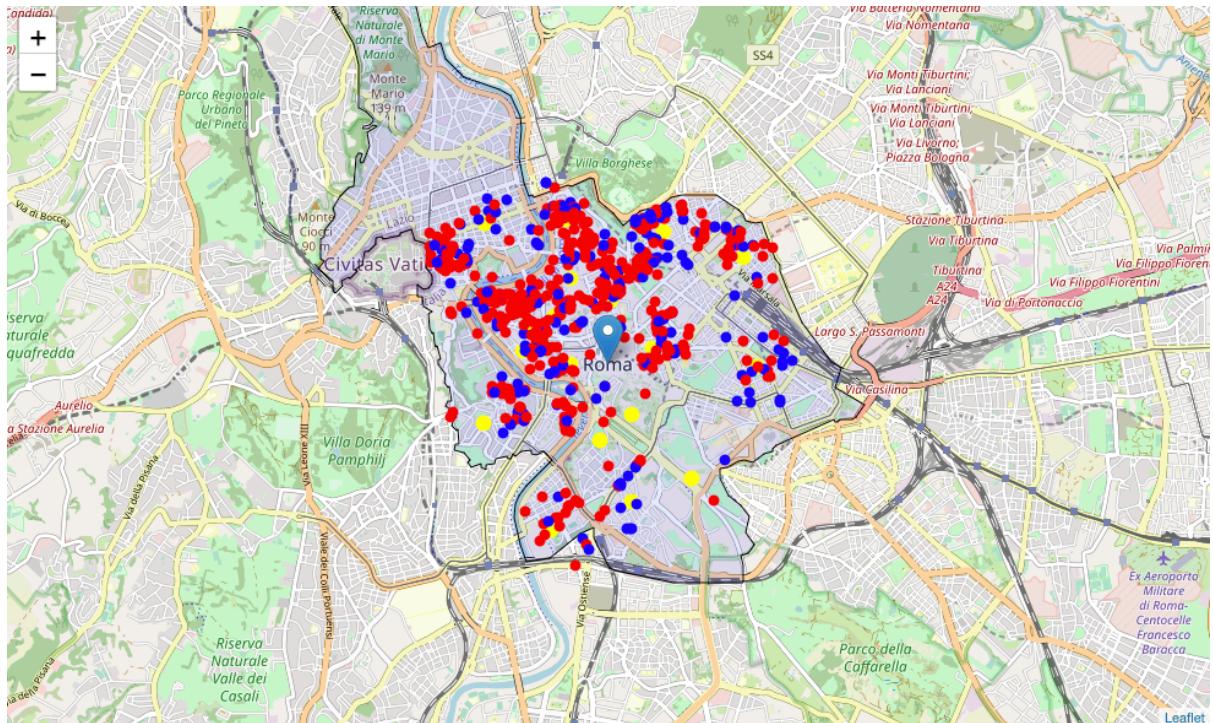


**Municipio I** can be divided in **22** sub-regions called **rioni**. I have created a list of all **rioni** of Rome from Wikipedia using **BeautifulSoup**. The latitude and longitude of each **rione** was obtained from Google Maps. Finally, I have plotted all the locations candidates in the Municipal I of Rome.



## Foursquare

Now that we have our location candidates, let's use Foursquare API to get info on restaurants in each area. We're interested in venues in 'food' category, but only those that are proper restaurants - coffee shops, pizza places, bakeries etc. are not direct competitors so we don't care about those. So we will include in our list only venues that have 'restaurant' in category name, and we'll make sure to detect and include all the subcategories of specific 'Italian restaurant' category, as we need info on Italian restaurants in the area.



We can see all the collected restaurants in our areas of interest on map. Italian restaurants are depicted in red and Not-Italian restaurants are depicted in blue. Now we have all the restaurants in area, and we know which ones are Italian restaurants. We also know which restaurants exactly are in vicinity of every candidate location (yellow). This concludes the data gathering phase - we're now ready to use this data for analysis to produce the report on optimal locations for a new Italian restaurant.

## 3. Methodology

In this project we will direct our efforts on detecting areas of Rome that have low restaurant density, particularly those with low number of Italian restaurants. We will limit our analysis to area around city center.

In first step we have collected the required data: location and type (category) of every restaurant. We have also identified Italian restaurants (according to Foursquare categorization).

Second step in our analysis will be calculation and exploration of '**restaurant density**' across different areas of Rome - we will use **heatmaps** to identify a few promising areas close to center with low number of restaurants in general (*and* low number of Italian restaurants in vicinity) and focus our attention on those areas.

In third and final step we will focus on most promising areas and within those create **clusters of locations that meet some basic requirements**: we will take into consideration locations with no more than 50 Italian restaurants in radius of 500 meters, with a distance less than 1000 meters from the city Center.

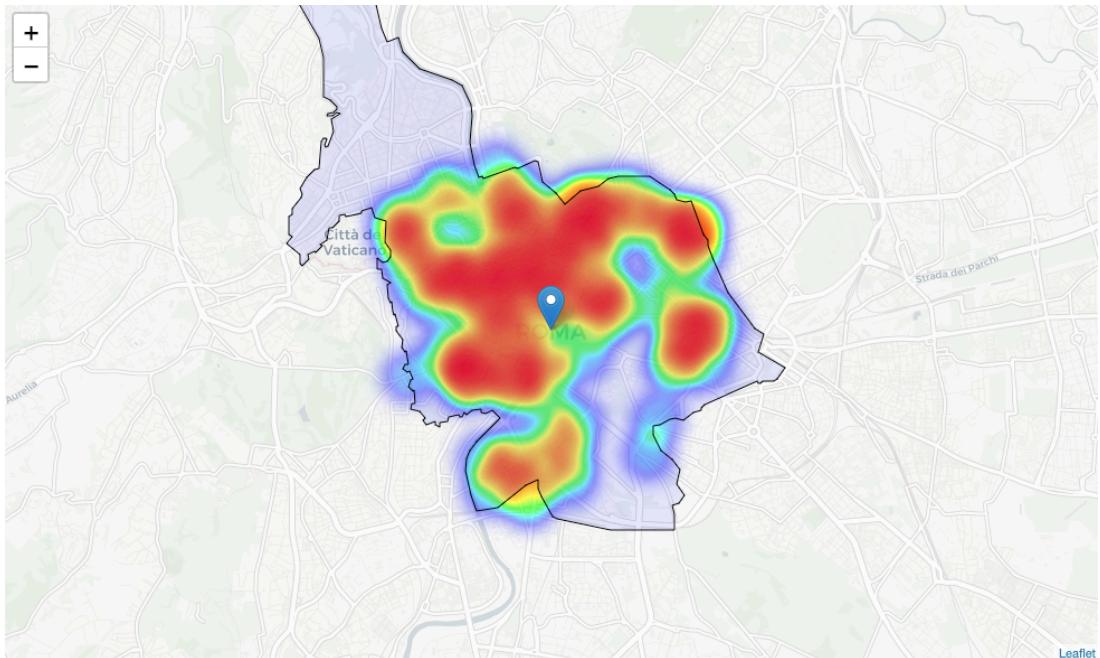
We will present map of all such locations but also create clusters (using **k-means clustering**) of those locations to identify general zones which should be a starting point for final search for optimal venue location by stakeholders.

## 4. Analysis

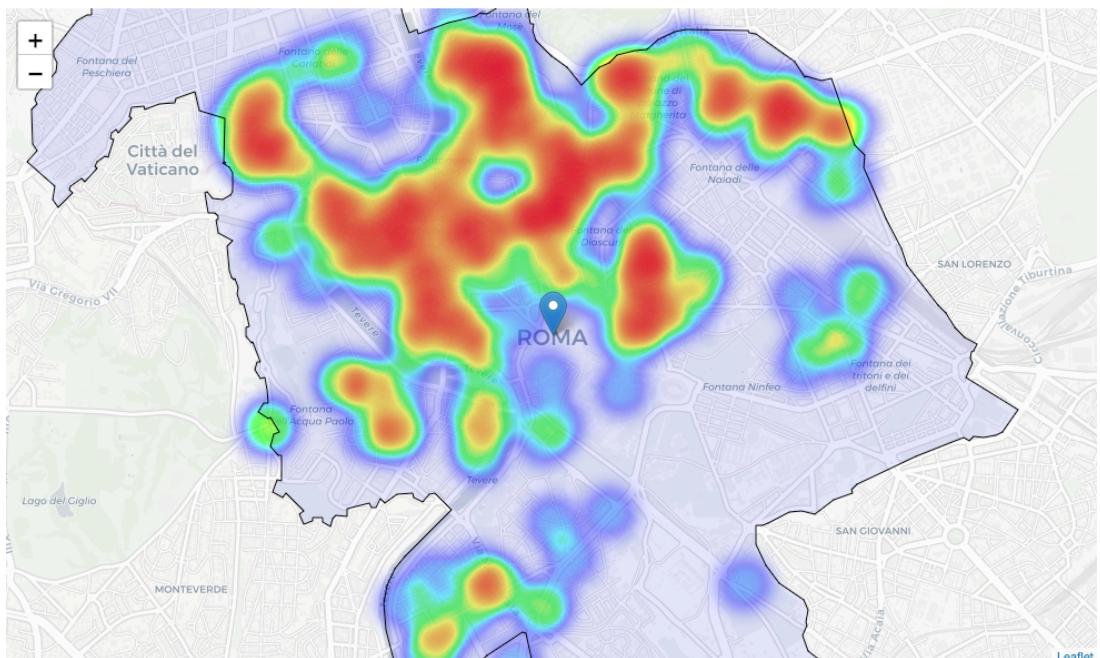
First, we performed some basic explanatory data analysis and derive some additional info from our raw data. We counted the **number of restaurants in every area candidate**.

	Rioni	Latitude	Longitude	Restaurants in area	Italian Restaurants in area
0	Monti	41.894962	12.489225	47	38
1	Trevi	41.902773	12.485952	65	51
2	Colonna	41.902222	12.477778	36	29
3	Campo Marzio	41.908458	12.477030	63	48
4	Ponte	41.900310	12.465617	63	51
5	Parione	41.897599	12.471746	74	57
6	Regola	41.894722	12.471111	55	37
7	Sant'Eustachio	41.898333	12.475556	68	53
8	Pigna	41.900000	12.483333	72	57
9	Campitelli	41.888030	12.486391	6	4
10	Sant'Angelo	41.893333	12.477778	29	19
11	Ripa	41.885346	12.481799	16	14
12	Trastevere	41.887222	12.465556	33	24
13	Borgo	41.904066	12.461629	39	30
14	Esquilino	41.893056	12.503889	25	10
15	Ludovisi	41.907164	12.490854	68	44
16	Sallustiano	41.908078	12.465706	22	12
17	Castro Pretorio	41.904389	12.501916	28	18
18	Celio	41.881450	12.494751	2	1
19	Testaccio	41.876111	12.475000	20	16
20	San Saba	41.879017	12.486293	18	7
21	Prati	41.908078	12.465706	22	12

We analysed the **density of restaurants** and we tried to extract some meaningful info from that.

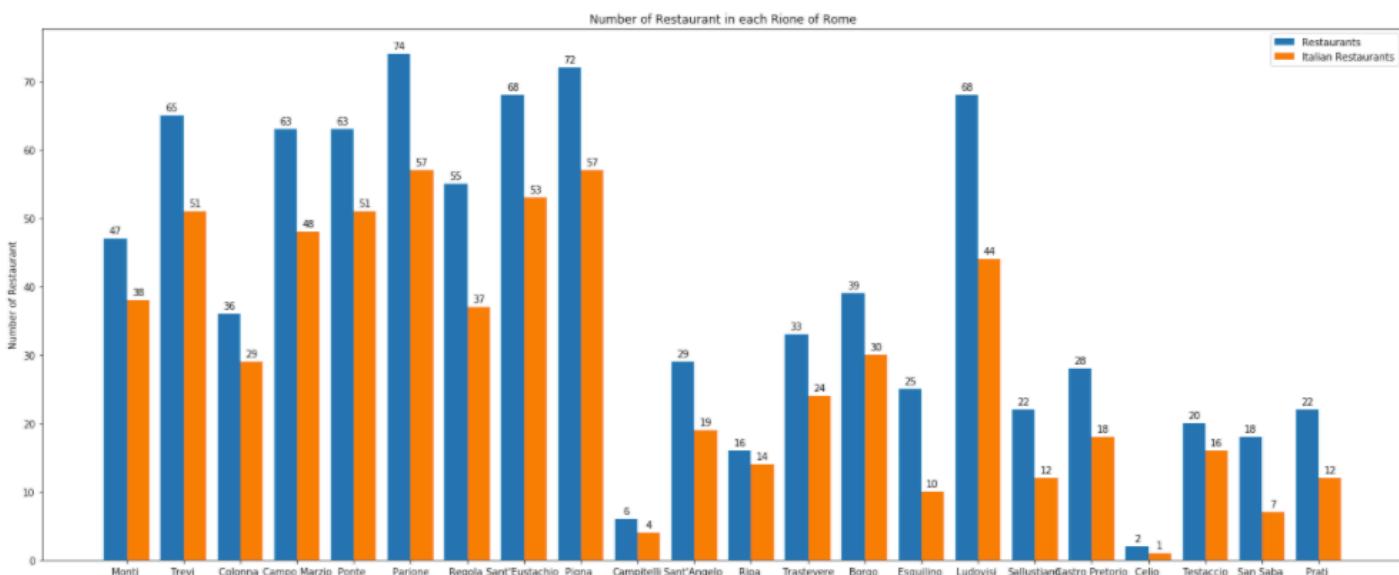


Looks like a few pockets of low restaurant density closest to city center can be found **south, south-east from city center**. We created another heatmap showing **density of Italian restaurants** only.



This map indicates higher density of existing Italian restaurants directly north, south and east from the center, with closest pockets of **low Italian restaurant density positioned south-east, south and south-west from city center**.

Based on this we will now focus our analysis on these areas.



We plotted the number of italian and not-italian restaurants for any candidate location in order to estimate which areas have the lowest number of restaurants.

The Average number of italian restaurants in every area with a radius of 500m is 28.72. The Average number of restaurants in every area with a radius of 500m is 39.59.

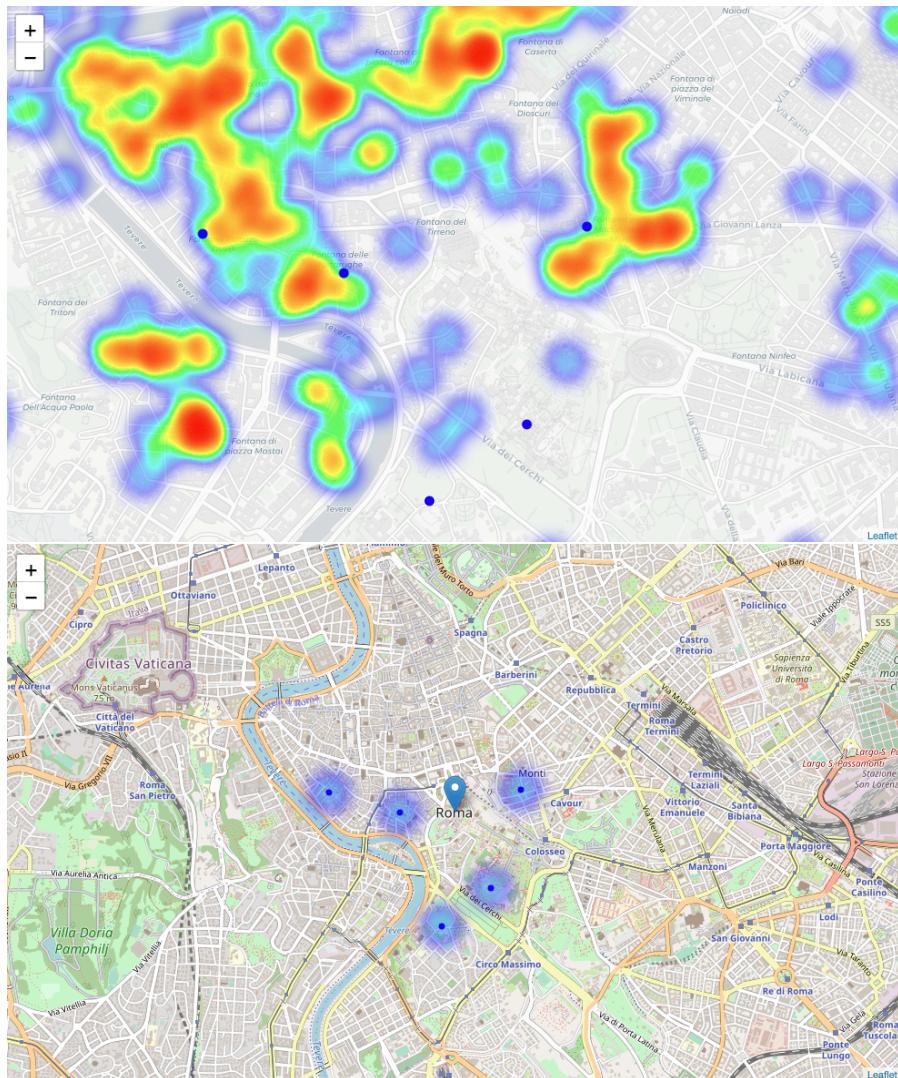
Then we calculated the distance of each **rione** from the center of the city.

	Rioni	Latitude	Longitude	Restaurants in area	Italian Restaurants in area	Distance from Center
10	Sant'Angelo	41.893333	12.477778	29	19	0.426619
0	Monti	41.894962	12.489225	47	38	0.551934
9	Campitelli	41.888030	12.486391	6	4	0.654231
8	Pigna	41.900000	12.483333	72	57	0.743490
7	Sant'Eustachio	41.898333	12.475556	68	53	0.826684
11	Ripa	41.885346	12.481799	16	14	0.891649
6	Regola	41.894722	12.471111	55	37	0.990785
5	Parione	41.897599	12.471746	74	57	1.040956
2	Colonna	41.902222	12.477778	36	29	1.077835
1	Trevi	41.902773	12.485952	65	51	1.080402
12	Trastevere	41.887222	12.465556	33	24	1.590160
20	San Saba	41.879017	12.486293	18	7	1.614606
4	Ponte	41.900310	12.465617	63	51	1.630319
18	Celio	41.881450	12.494751	2	1	1.642981
15	Ludovisi	41.907164	12.490854	68	44	1.673160
14	Esquilino	41.893056	12.503889	25	10	1.734904
3	Campo Marzio	41.908458	12.477030	63	48	1.752680
17	Castro Pretorio	41.904389	12.501916	28	18	1.995854
19	Testaccio	41.876111	12.475000	20	16	2.023116
13	Borgo	41.904066	12.461629	39	30	2.129891
16	Sallustiano	41.908078	12.465706	22	12	2.173794
21	Prati	41.908078	12.465706	22	12	2.173794

Finally, we filtered those locations: we were interested only in locations with **no more than 50 italian restaurants in radius of 500 meters**, and locations with **a distance from the city center less then 1000 meters**.

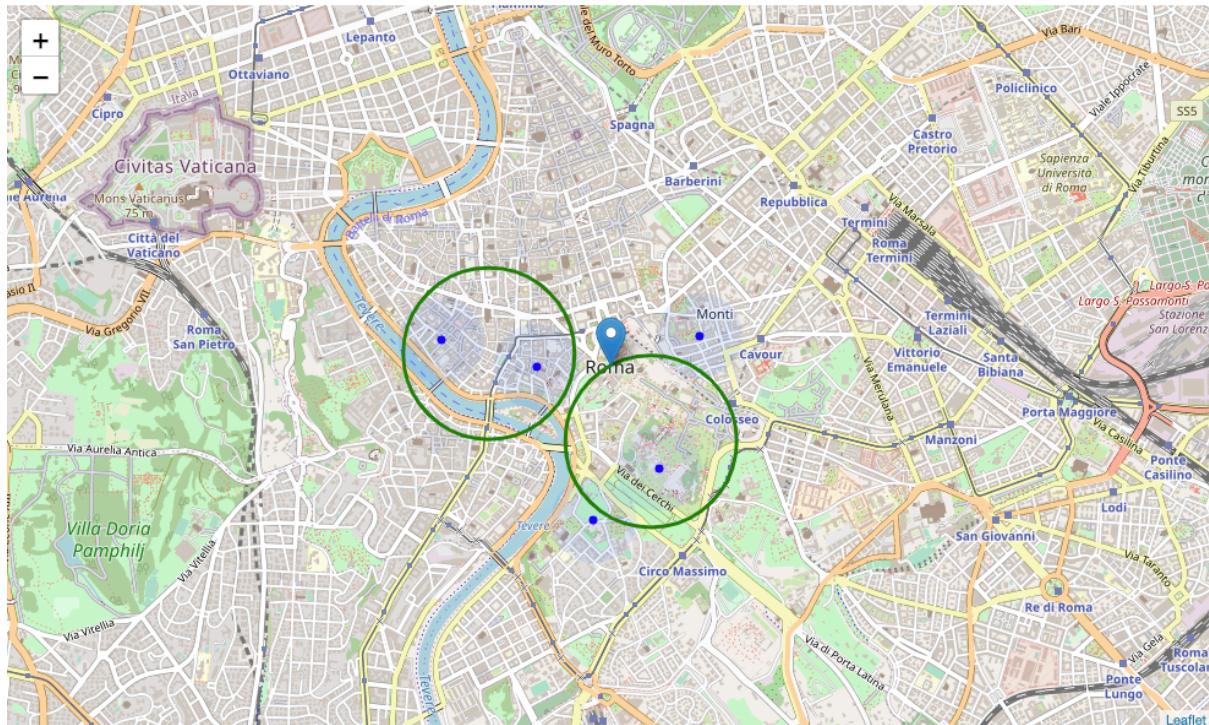
We obtained a bunch of locations fairly close to the center (less then 1km), and we know that each of those locations has no more than 50 italian restaurants in radius of 500m. Any of those locations is a potential candidate for a new Italian restaurant, at least based on nearby competition.

	Rioni	Latitude	Longitude	Restaurants in area	Italian Restaurants in area	Distance from Center
0	Monti	41.894962	12.489225	47	38	0.551934
6	Regola	41.894722	12.471111	55	37	0.990785
9	Campitelli	41.888030	12.486391	6	4	0.654231
10	Sant'Angelo	41.893333	12.477778	29	19	0.426619
11	Ripa	41.885346	12.481799	16	14	0.891649



What we have now is a clear indication of zones with low number of italian restaurants in vicinity.

Finally we clustered those locations to create centers of zones containing good locations. Those zones, their centers will be the final result of our analysis.



This concludes our analysis. We have created 2 main areas representing centers of zones containing locations with low number of italian restaurants and all zones being fairly close to city center (all less than 1km from Rome center).

Their centers/addresses should be considered only as a starting point for exploring area in search for potential restaurant locations. Most of the zones are located in rioni **Campitelli**, **Ripa**, **San'Angelo**, which we have identified as interesting due to being popular with tourists, fairly close to city center and well connected by public transport and very close to the main monuments.

## 5. Results and Discussion

Our analysis shows that although there is a great number of restaurants in Municipio I of Rome there are pockets of low restaurant density fairly close to city center. Highest concentration of restaurants was detected north and west from the city centre. After directing our attention to this more narrow area of interest we looked to the **rioni**; those locations were then filtered so that those with more than 50 italian restaurants in radius of 500m and those with a distance greater than 1Km from the Rome Center were removed. Those location candidates were then clustered to create zones of interest which contain greatest number of location candidates. Result of all this is 2 zones containing largest

number of potential new restaurant locations based on number of existing venues. This, of course, does not imply that those zones are actually optimal locations for a new restaurant!

Purpose of this analysis was to only provide info on areas close to Rome center but not crowded with existing restaurants (particularly Italian) - it is entirely possible that there is a very good reason for small number of restaurants in any of those areas, reasons which would make them unsuitable for a new restaurant regardless of lack of competition in the area. Recommended zones should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no nearby competition but also other factors taken into account and all other relevant conditions met.

## 6. Conclusion

Purpose of this project was to identify Rome areas close to center with low number of restaurants (particularly Italian restaurants) in order to aid stakeholders in narrowing down the search for optimal location for a new Italian restaurant.

By calculating restaurant density distribution from Foursquare data we generated extensive collection of locations which satisfy some basic requirements regarding existing nearby restaurants.

Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.