



Opening a gym in Rome

A territory analysis

Contents

- Introduction/business problem
- Data
- Data preparation
- Methodology
- Results
- Discussion
- Conclusion

Introduction/Business problem

- We live in a time when physical well-being and health care is highly developed
- For this reason there is a strong demand for structures where to practice physical activity
- Finding the right place to open an activity is one of the key factors that make the difference between being successful or lose a lot of money
- Using data science methodology and machine learning techniques this project aims to help decision making for investors to find the optimal location where to open a new gym in Rome

Introduction/Business problem

- During the daytime, specially early in the morning and during lunch break, office areas provide an excellent opportunity for gyms
- Our study, leveraging demographics, Foursquare and economical data tries to understand which is the best location to open a new gym in Rome

Data

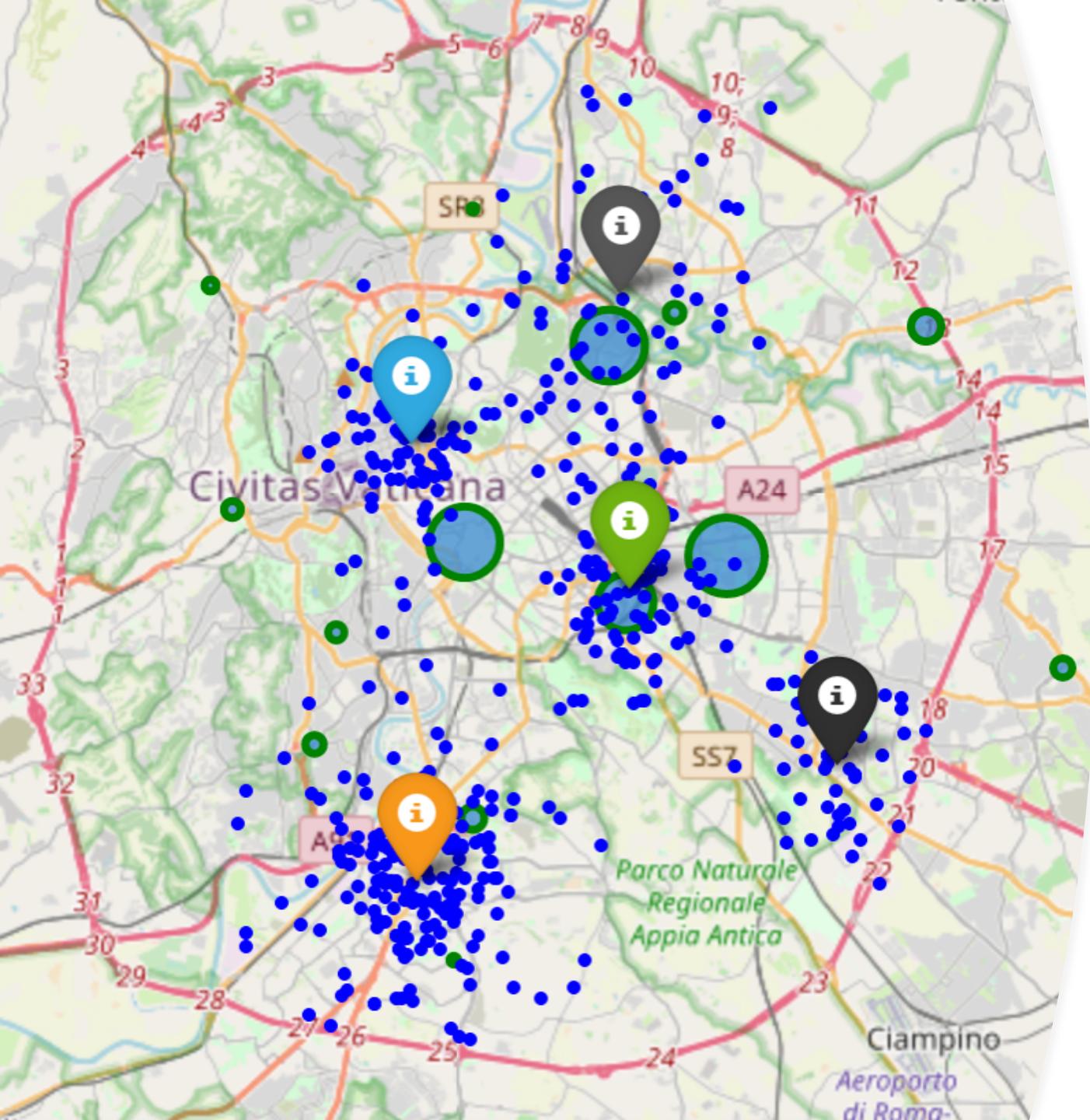
- List of Neighborhoods of Rome from Wikipedia
- Latitude and longitude coordinates of Neighborhoods from Google
- Rent location prices to inform investors about the investment required
- Foursquare data to explore boroughs
- Opendata databases to complement Foursquare information

Data preparation

- Some data wrangling needed to properly fit demographics data of neighborhoods with their position
- Gyms position found using Foursquare APIs, integrated with data from
http://www.datopen.it/it/opendata/Roma_Capitale_Impianti_sportivi
- Price of rental found in <https://www.immobiliare.it/mercato-immobiliare/lazio/roma/>
- BAU operations to select, add, clean etc data

Methodology

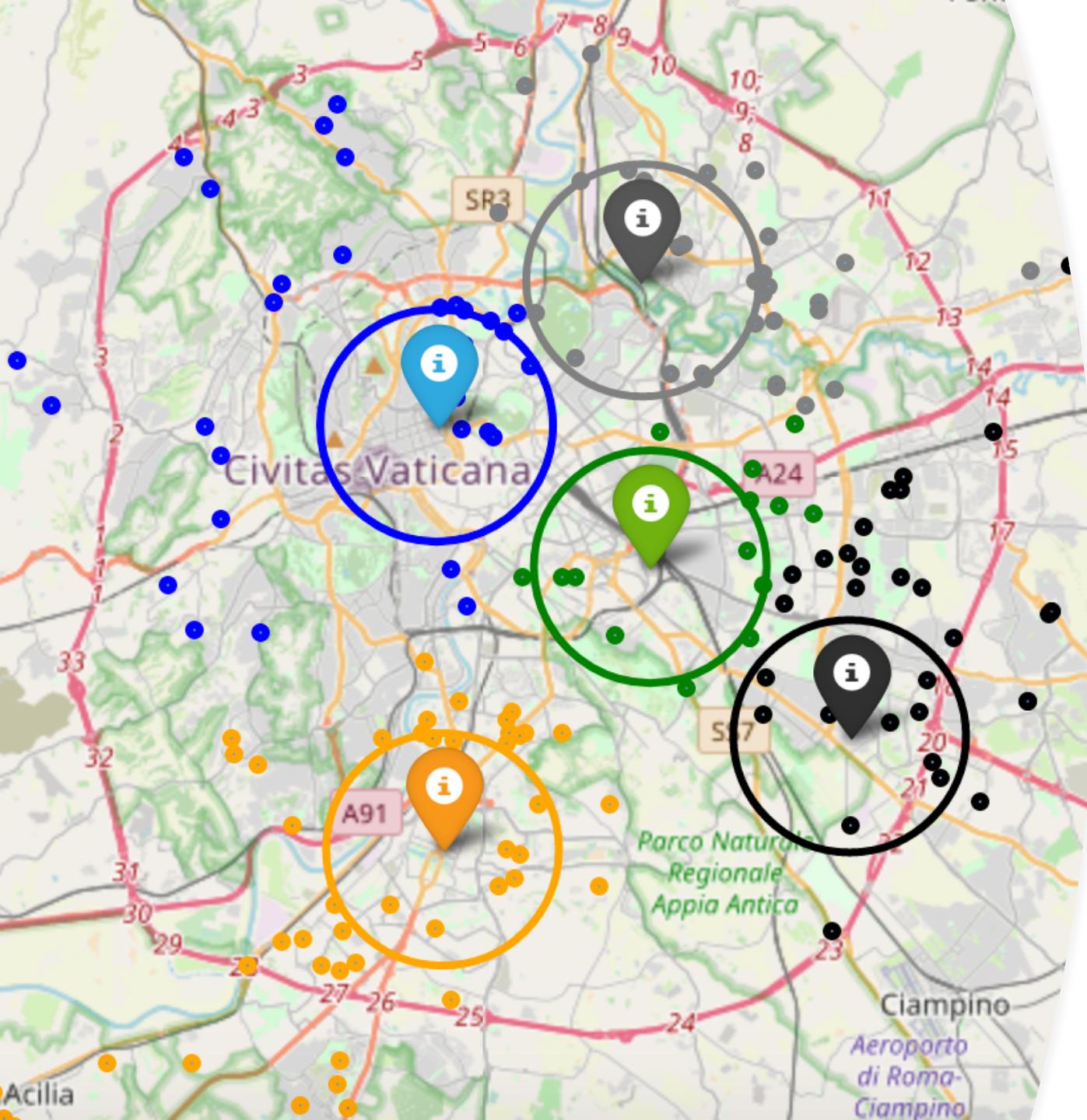
- Studing how offices are spread on the territory, we discovered that there are 5 areas of greater concentration. By using a simple kmeans algorithm, these areas were clustered and for each cluster we used its center to properly represent it.
- Then each gym has been assigned to one of the clusters (again by using kmeans)
- Distances and mean distance of all the gyms belonging to the same cluster from its center were calculated
- This gave us an idea of how far gyms are from areas were offices are concentrated
- With a similar technique mean home rental price are associated to each cluster



Offices position

Here the position of the main offices is represented by a blue point. It is easy to see that they are not evenly spread, but there are areas of major concentration. In order to find how they are distributed, and the center of each cluster, kmeans algorithm is used. Centers are represented by a colored icon.

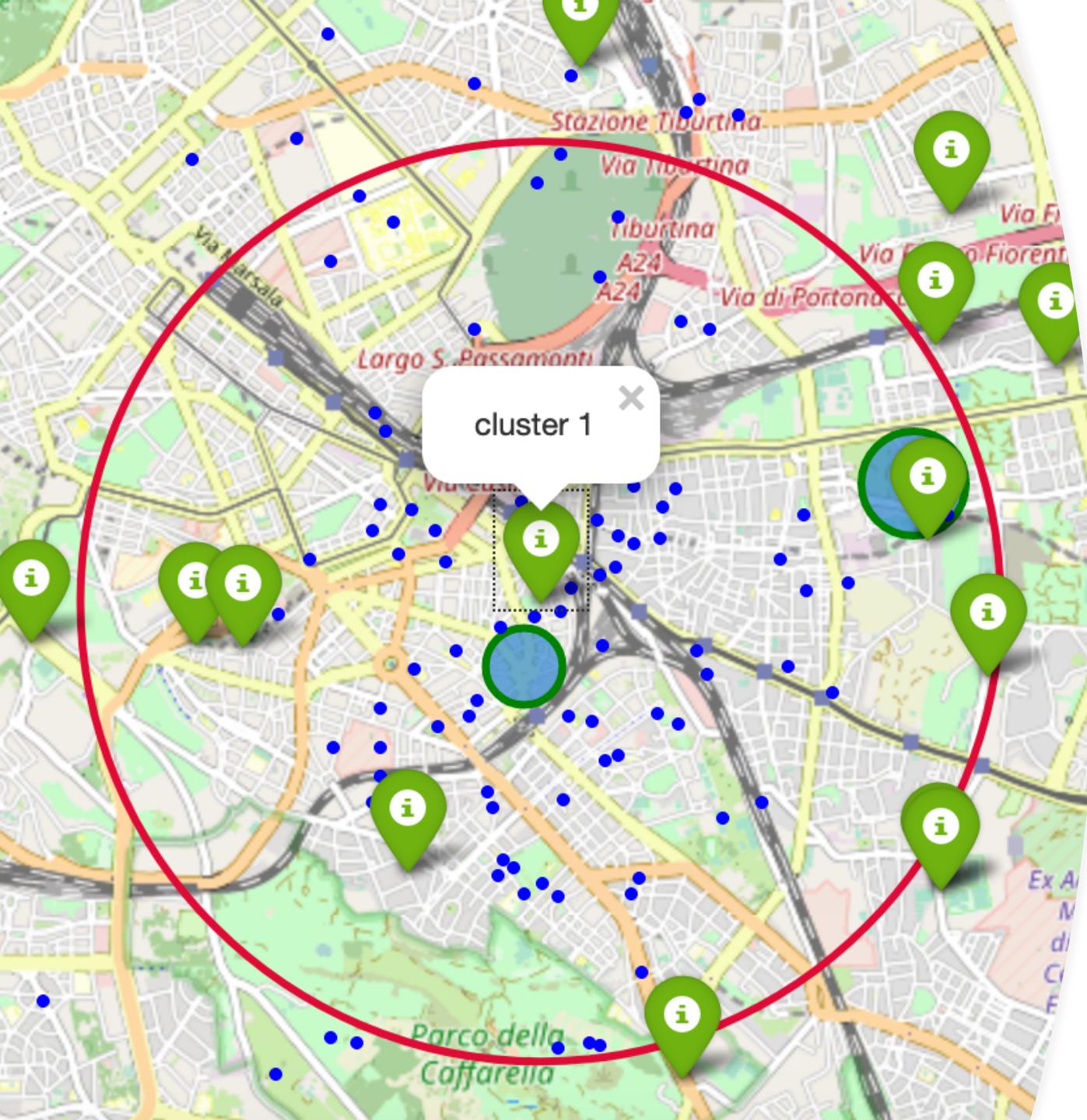
On the same map, the green circles represent the center of the neighborhoods and their dimension is proportional to the population density in each of them.



Gym position

In this map the position of the gyms already present in Rome is represented by a colored point. Each gym is associated to an office cluster. Circles are 2500 m radius centered in the center of the clusters.

This plot gives us a visual information on where offices are concentrated (center of clusters), with respect to gyms position.



Best area found

This area has several advantages:

- It has a high number of offices
- And a low number of gyms
- It also has the second highest mean distance of gyms from the center of the cluster
- The office/gym ratio is the highest
- The population density of the two neighborhoods is among the highest

Results

Cluster	Gyms mean distance from cluster center	# gyms	# offices	Office/gym ration	Population density	Rental price €/m ²
0	1511 m	10	76	7.6	8500	1555
1	1951 m	5	85	17.0	7500	1524
2	1758 m	14	156	11.1	2800	1214
3	2075 m	10	40	4.0	8600	1199
4	1477 m	12	49	4.1	6700	1124

Discussion

- Thanks to this study, we gave to the investor interested in opening a gym in Rome several information to help him finding the best area where to start with this activity.
- Now, complementing this with economic information, he can decide whether to start or not and which is the ROI.