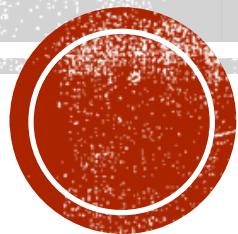


MOVING HOUSE AFTER COVID19

Applied Data Science IBM Capstone project
Fabrizio Ruffini



INTRO

After the COVID-19 outbreak, the working habits will likely change from an office-centered perspective to a more home-based culture with a prevalence of smart-working approach.

New houses and cities will be chosen looking at life-safety and health parameters

Here, we compare the four main Italian cities in terms of:

- pollution with
- appealing neighbourhood in terms of life quality



DATA ACQUISITION AND DESCRIPTION

Pollution data:

taken from Copernicus Atmosphere, whole 2019 up to June 2020:

- NO₂, O₃, PM_{2.5}, PM₁₀.
 - -> will be used to calculate Air quality index

Neighbourhoods data:

- Venue data from Foursquare Places Database
- Popular times data from google
- Traffic: google traffic layer was used to get info on traffic
- Historical monuments: openstreetmap was used to get info on historical monuments
- Parking lots: openstreetmap was used to get info on parking
- Green parks: openstreetmap was used to get info on green parks



METHODOLOGY

The analysis is composed by two steps.

First step, compare pollution for four main Italian cities namely Milan, Naples, Rome, and Turin is done, with the goal of understanding which is the less polluted.

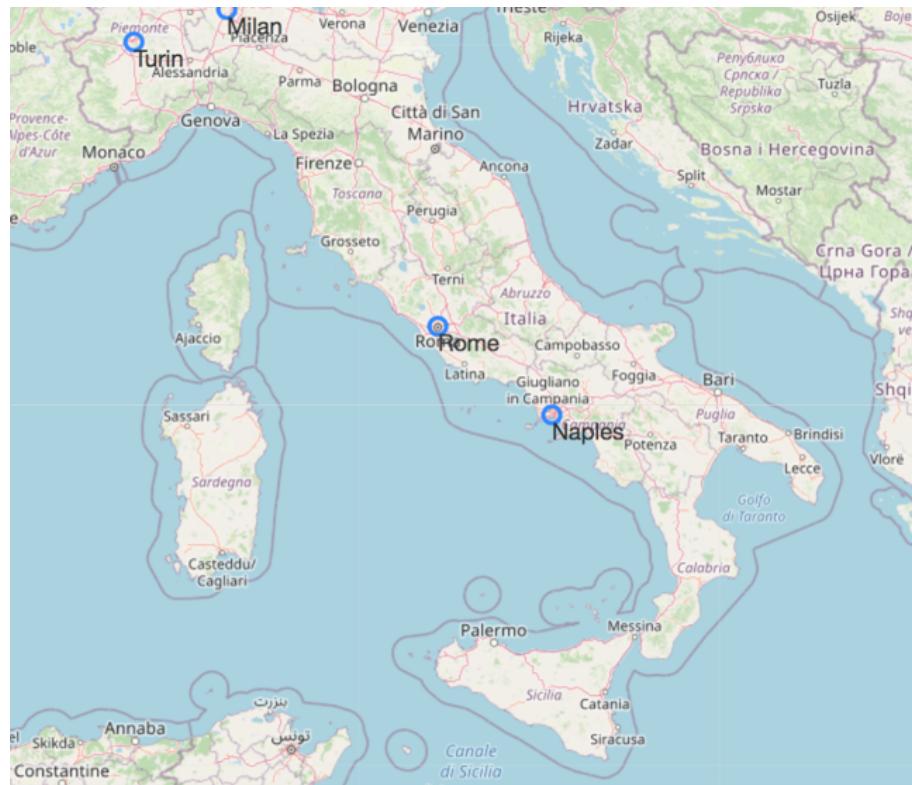
- an exploratory analysis is applied

Second step, focus on the neighbourhoods's characteristic with the goal of providing information on where to go to live based on life quality parameters.

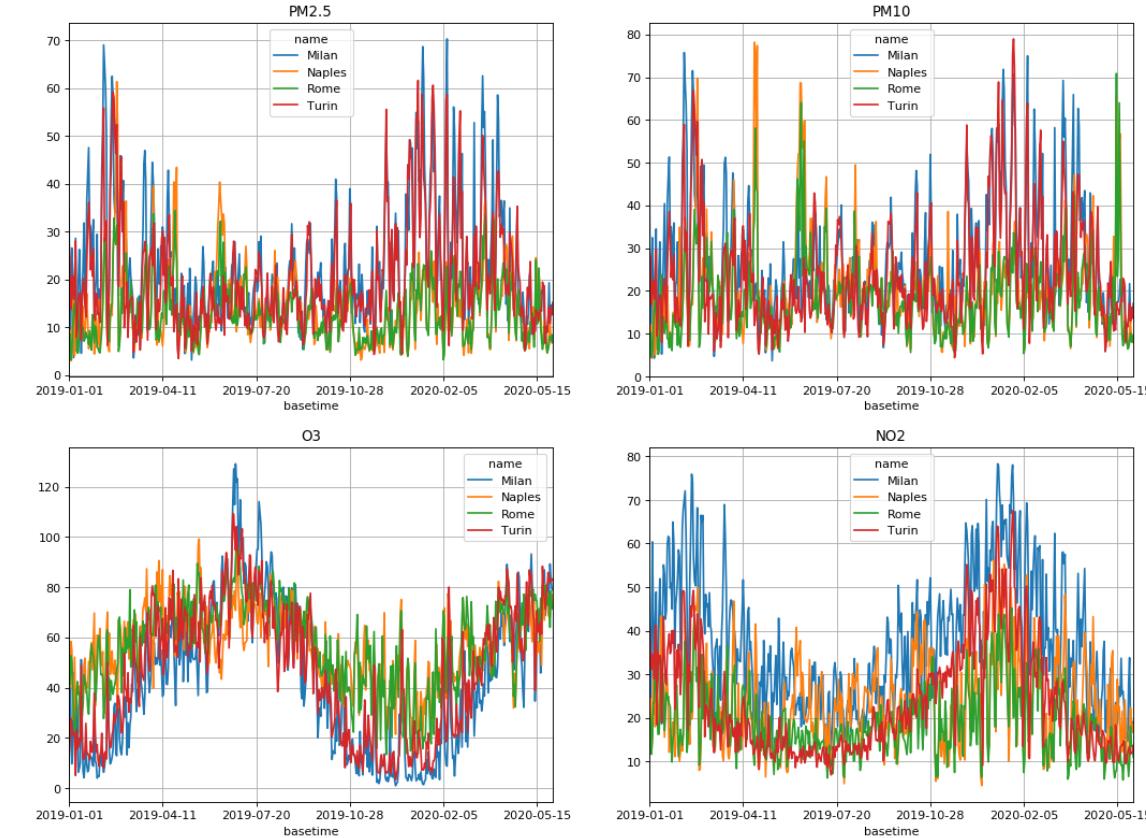
- Clustering techniques will be applied



DATA ANALYSIS I: FINDING BEST CITY

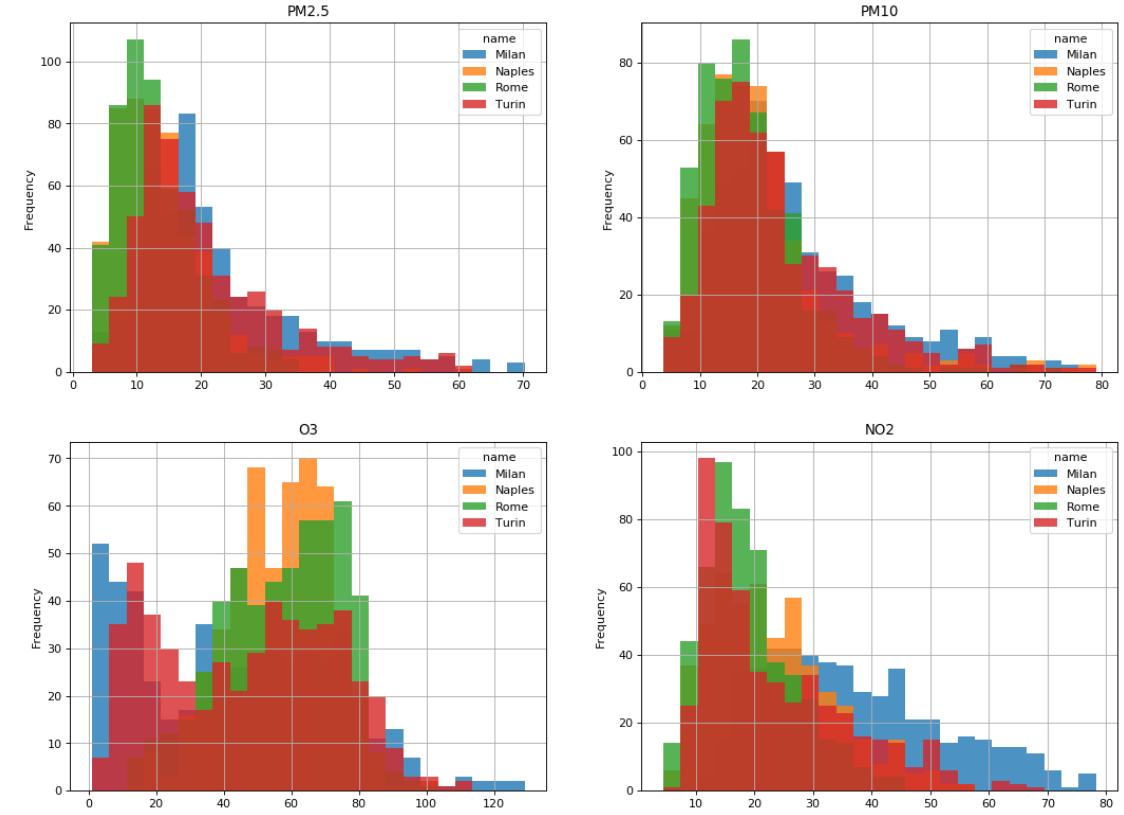


DATA EXPLORATION



Time trends

A seasonal pattern is clearly visible for O3 and NO₂, while it is less evident in PM2.5 and PM10. Generally, Milan and Turin show peaks and most pronounced behaviours with respect to Naples and Rome



Histograms:

Rome PM2.5 and PM10 histograms are clearly the ones with smaller median value



DATA PROCESSING

median and the standard deviation of each city subsample

	PM2.5	PM10	NO2	O3
Milan	18.955	23.005	34.545	42.345
Rome	11.610	17.105	17.435	59.815
Turin	17.030	20.790	19.090	49.750
Naples	12.290	17.925	21.650	58.220

median values

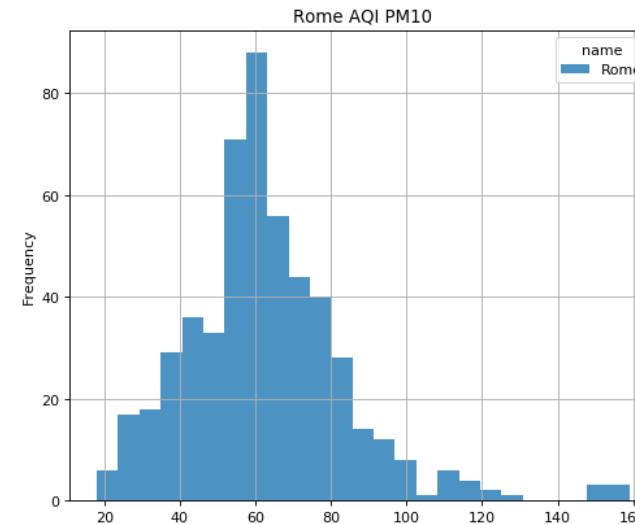
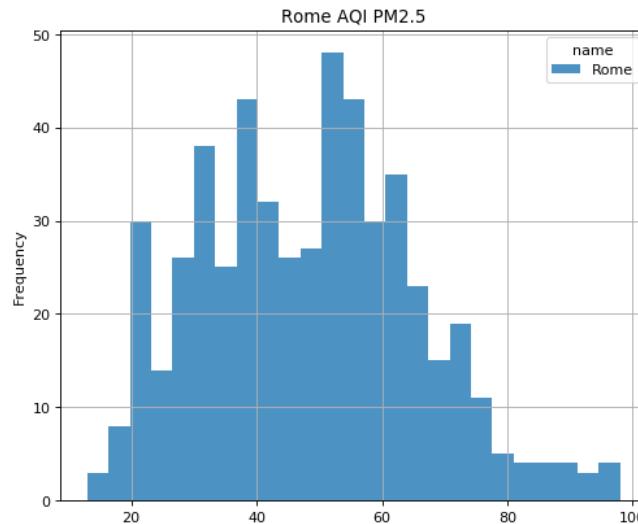
	PM2.5	PM10	NO2	O3
Milan	12.798048	14.049052	15.051572	28.922493
Rome	6.137670	9.056311	7.822407	17.653542
Turin	11.476924	12.764039	12.419786	25.850945
Naples	7.633185	11.412343	10.237374	14.930984

standard deviation values

Rome is the winner



FOCUS ON AIR QUALITY INDEX (AQI)



Air Quality Index (AQI)

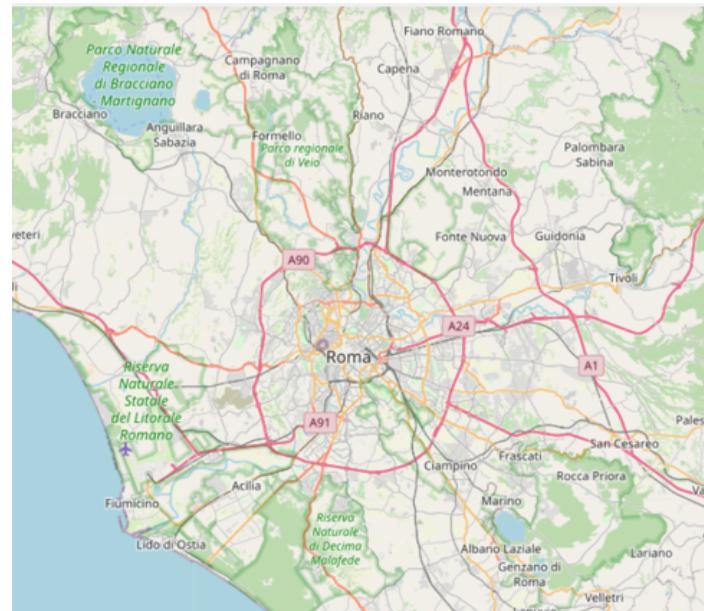
AQI Value	AQI Category	AQI Color
0 - 50	Good	Green
51 - 100	Moderate	Yellow
101 - 150	Unhealthy for Sensitive Groups	Orange
151 - 200	Unhealthy	Red
201 - 300	Very Unhealthy	Purple
301 - 500	Hazardous	Maroon

Air quality index calculated for PM2.5 and PM10.

AQI is typically under the health warning orange-red thresholds



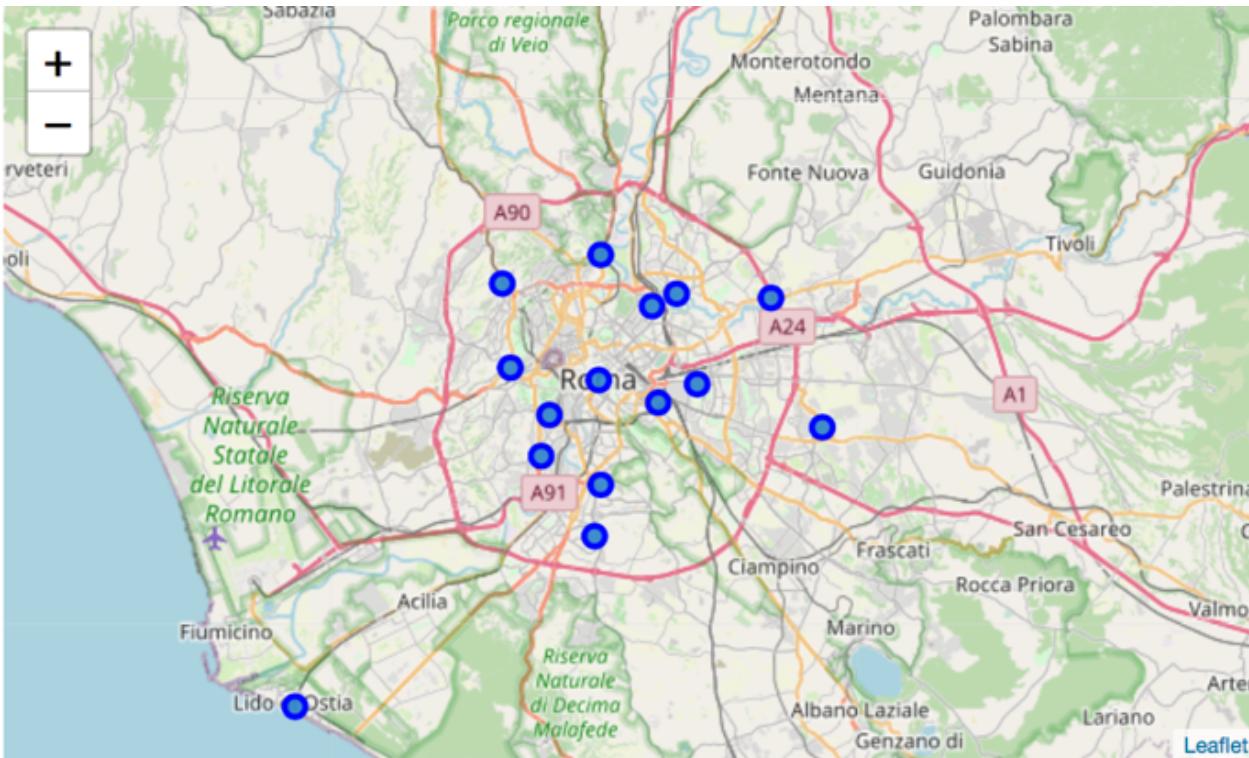
DATA ANALYSIS II: FINDING LOVELY NEIGHBOURHOOD IN ROME



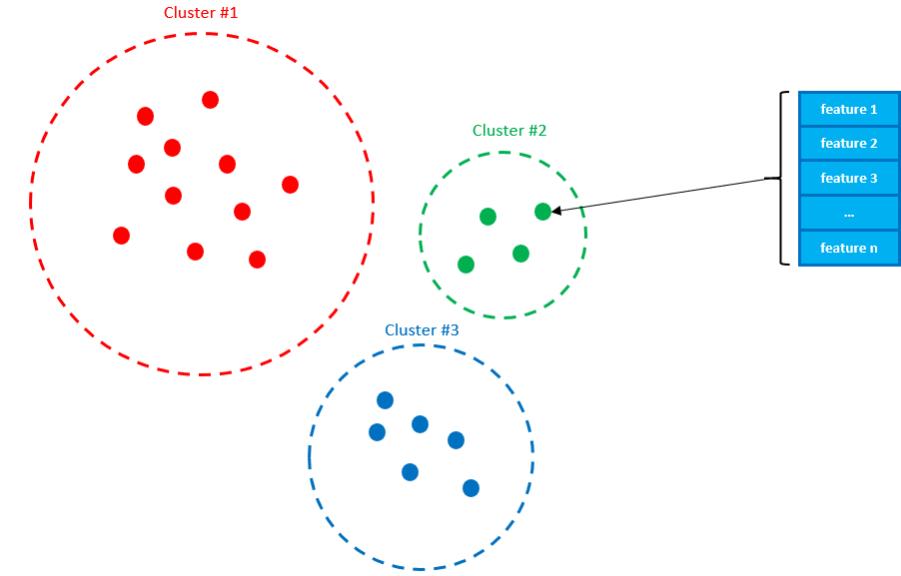
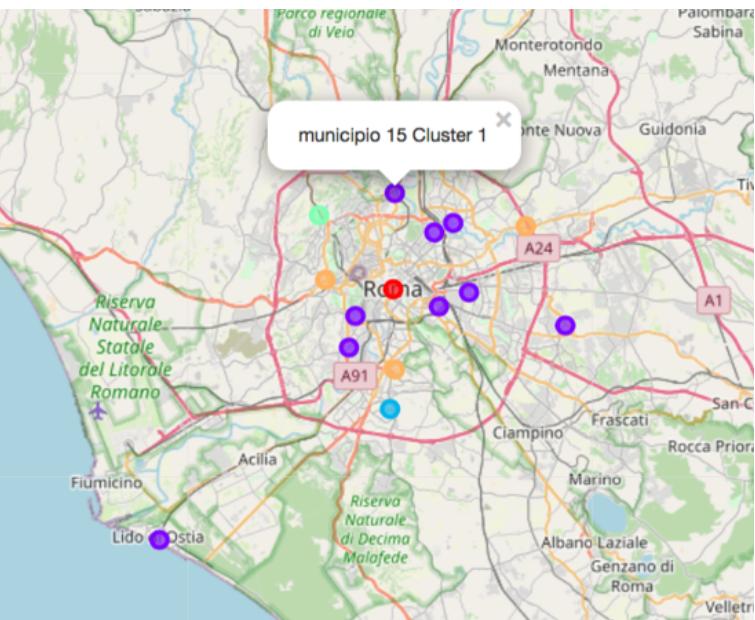
DATA EXPLORATION

Rome is divided into 15 municipalities

Municipio	lat	lon
municipio 1	41.893056	12.482778
municipio 2	41.929958	12.518931
municipio 3	41.936080	12.535116
municipio 4	41.933491	12.598746
municipio 5	41.890665	12.548488
municipio 6	41.869657	12.632731
municipio 7	41.881700	12.522800
municipio 8	41.841228	12.484290
municipio 9	41.814879	12.479980
municipio 10	41.730660	12.280531
municipio 11	41.855282	12.444762
municipio 12	41.876099	12.450100
municipio 13	41.899141	12.424158
municipio 14	41.940965	12.418627
municipio 15	41.955435	12.484851



DATA PROCESSING



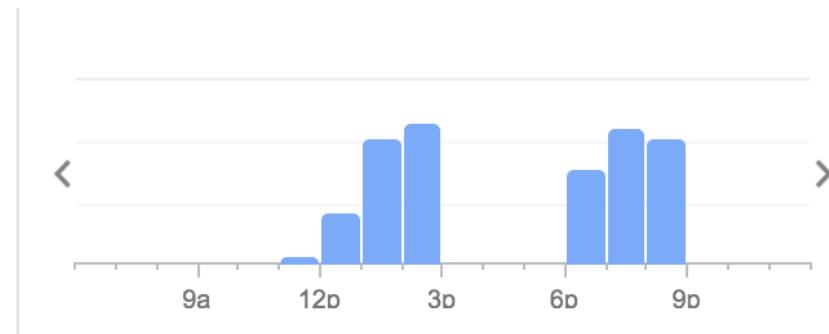
Clustering technique applied to Foursquare queries to create 5 clusters.

- Cluster 0: historical places
- Cluster 1: touristic places
- Cluster 2: convention area
- Cluster 3: parks places
- Cluster 4: not-touristic place



NEIGHBORHOOD EXPLORATION

Assuming we want to live in cluster 0 area, that is the historical ones, let's have a look at surrounding interesting information and places

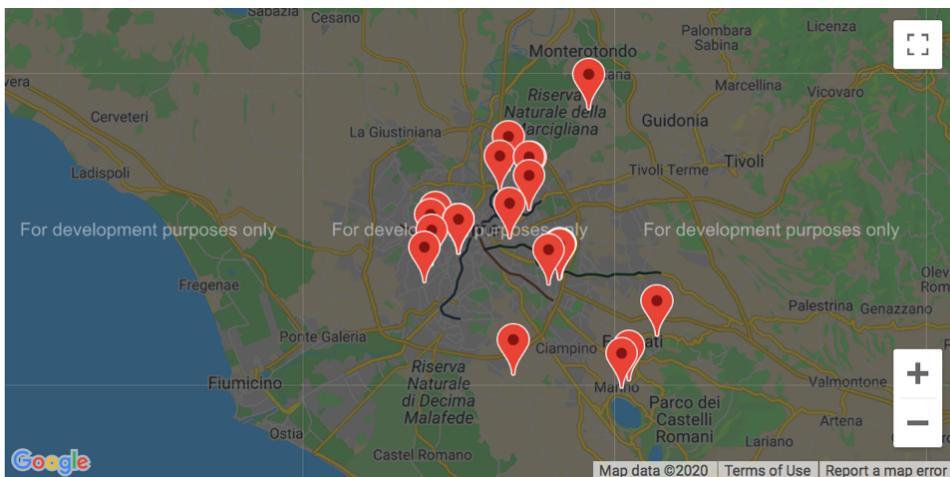


FOOD: Restaurant “Pizzicheria salentina”
most frequent hours on Friday

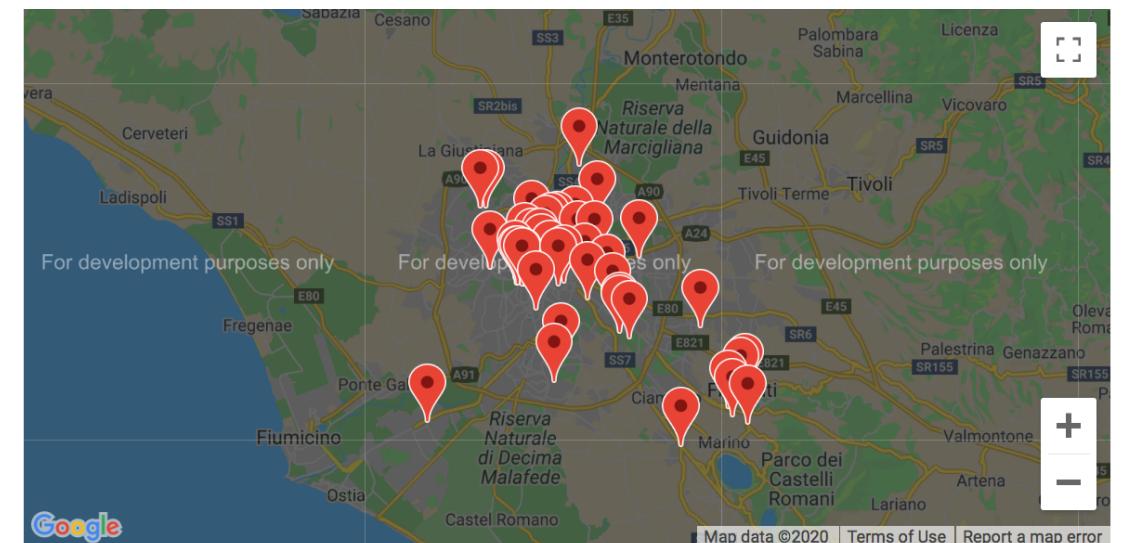


NEIGHBORHOOD EXPLORATION (2)

Assuming we want to live in cluster 0 area, that is the historical ones, let's have a look at surrounding interesting information and places



LEISURE: parks



Historical monuments

RESULTS

- Interesting results (and it was also very funny):
 - Acquire different information from different points of view,
 - very beneficial for estate agency to acquire a data science-powered mind-set representing a fantastic competitive advantage.
- very important in this period to face new crisis and opportunities that may come after the COVID19.
 - We do not have a detailed understanding on how the world will change, so having additional knowledge and expertise with respect to your competitors will be probably essential.



DISCUSSION

To improve our results:

- 1. For the air quality index, calculate NO₂ and O₃ values
- 2. Try different clustering approach (eg, density based tech)
- 3. Perform deeper analysis on neighborhood lifestyle data



CONCLUSION

In conclusion, the capability of using data analytics is very important and can be applied where numbers are available.

In this uncertain times, it's good to enlarge your expertise and polish your skills, and this is the main reason I took this capstone project activity

Thanks for reading!

F Ruffini



BIBLIOGRAPHY

- Info on air quality:
 - - Aq1: <https://towardsdatascience.com/sensing-the-air-quality-5ed5320f7a56>
 - - Aq2: <https://www.epa.gov/sites/production/files/2014-05/documents/zell-aqi.pdf>
- Info on NO₂ <https://www.epa.gov/no2-pollution/basic-information-about-no2>
- Info on O₃ <https://www.epa.gov/ground-level-ozone-pollution/ground-level-ozone-basics>
- Info on PM <https://www.epa.gov/pm-pollution/particulate-matter-pm-basics>
- Foursquare data: <https://developer.foursquare.com/docs/places-database/details/>
- Google maps data:
<https://developers.google.com/maps/documentation/javascript/tutorial>

