Capstone Project - Rome Caput Mundi...

Introduction

Rome! That's a wonderful city with a lot of... air pollution

This project aims to select the neighborhood with the less concentration of air pollution in Rome where to do an **"aperitivo"** and share a nice talk with your friends. To do this will be taken into account the real air pollution data of Rome thank to different main control units located on the ground.

Problem

Unfortunately or fortunately air **pollution varies day-by-day and zone-by-zone** depending of the period of the year, of the number of automobiles, houses heating, farm animals (it sound strange but is's true...), and so on. For these reasons, the major is sometimes obligated to take some "extreme measures" such us the circulation ban of all diesel cars (up to Euro 6D-Temp) and so on. At the same time, people don't know where is the cleanest air in the city, so they prefer to remain home because the don't know where to go.

Data

Following data sources will be needed to extract/generate the required informations:

- **Part 1.a** - **PM10 values**: Real-time experimental measures of the PM10 particles scraped from the official ARPALAZIO website
- **Part 1.b** - **Latitude and Longitude values 1**: Geographical coordinates (lat and lon) of the control units extracted from an official pdf document on the ARPALAZIO website
- **Part 1.c** - **Neighborhoods**: Getting the main neighborhoods of Rome scraping a wikipedia page
- **Part 1.d** - **Latitude and Longitude values 2**: Geographical coordinates of neighborhoods obtained through geocoder

    o **Analysis 1**

- **Part 2** - **Bar**: Names and statistical data of bar and activities obtained from Foursquare services

    o **Analysis 2**

Methodology

There are a lot of data and a lot of possibilities so, we decided that only the first three control units will be choosen, those with less **PM10** values. The geographical coordinates of those control units will be crossed with the geographical coordinates of the neighborhoods in order to get all bar and choose the best three.

At that point it will be possible to choose where to have fun.

From the Part 1.a we obtained the real-time experimental measures of the PM10 particles scraped from the official ARPALAZIO website:

	Provincia	Stazione	Media Giornaliera02.02.2020 - (µg/m3)	Standard di leggeDal 01.01 al 01.02.2020 - Media di Periodo(µg/m3)	Standard di leggeDal 01.01 al 01.02.2020 - Sup. Giorni(50 µg/m3)
0	Frosinone	Alatri	29.0	52.0	21.0
1	Frosinone	Anagni	13.0	28.0	1.0
2	Frosinone	Cassino	76.0	99.0	27.0
3	Frosinone	Ceccano	NaN	110.0	29.0
4	Frosinone	Ferentino	30.0	58.0	22.0

In the same way we were able to retrieve geographical coordinates (lat and lon) of the control units extracted from an official pdf document on the same website:

	Rete	Stazione	Latitudine	Longitudine	QuotasIm
0	Roma	Arenula	41.894020	12.475368	31.0
1	Roma	Preneste	41.886018	12.541614	37.0
2	Roma	Francia	41.947447	12.469588	43.0
3	Roma	Magna Grecia	41.883064	12.508939	49.0
4	Roma	Cinecitta'	41.857720	12.568665	53.0
5	Roma	Colleferro oberdan	41.730840	13.004350	219.0
6	Roma	Colleferro europa	41.725010	13.009575	223.0
7	Roma	Allumiere	42.157741	11.908744	542.0
8	Roma	Civitavecchia	42.091629	11.802466	26.0
9	Roma	Guidonia	41.995679	12.726371	89.0
10	Roma	Villa Ada	41.932874	12.506971	50.0
11	Roma	Guido	41.889438	12.266300	61.0
12	Roma	Cavaliere	41.929383	12.658363	48.0
13	Roma	Ciampino	41.797880	12.607033	134.0
14	Roma	Fermi	41.864194	12.469531	26.0
15	Roma	Bufalotta	41.947649	12.533682	41.0
16	Roma	Cipro	41.906358	12.447596	31.0
17	Roma	Tiburtina	41.910257	12.548870	32.0
18	Roma	Malagrotta	41.874894	12.345598	55.0
19	Roma	Civitavecchia Porto	42.097053	11.788354	6.0

After that we get the main neighborhoods of Rome scraping a Wikipedia page:

	Sigla	Quartiere	Latitudine	Longitudine
0	Q. I	Flaminio	NaN	NaN
1	Q. II	Parioli	NaN	NaN
2	Q. III	Pinciano	NaN	NaN
3	Q. IV	Salario	NaN	NaN
4	Q. V	Nomentano	NaN	NaN

Roma
Quartieri
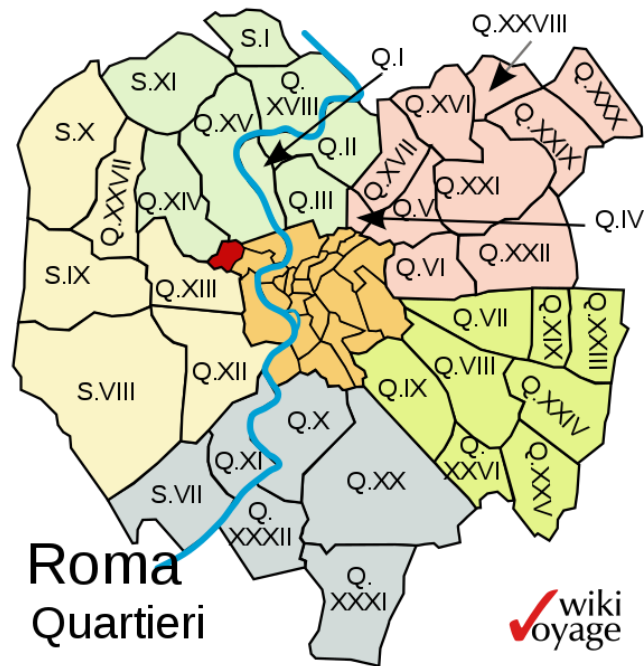
For the last part of section 1 geographical coordinates of neighborhoods were obtained through geocoder:

	Sigla	Quartiere	Latitudine	Longitudine
0	Q. I	Flaminio	41.92999	12.46451
1	Q. II	Parioli	41.93177	12.48622
2	Q. III	Pinciano	41.91886	12.48410
3	Q. IV	Salario	41.91331	12.50217
4	Q. V	Nomentano	41.91447	12.52210

At this point all control unit where filtered for place interest and analyzed in order to get the ones who measures the three lowest values. A lower value means less pollution and so a cleaner air.
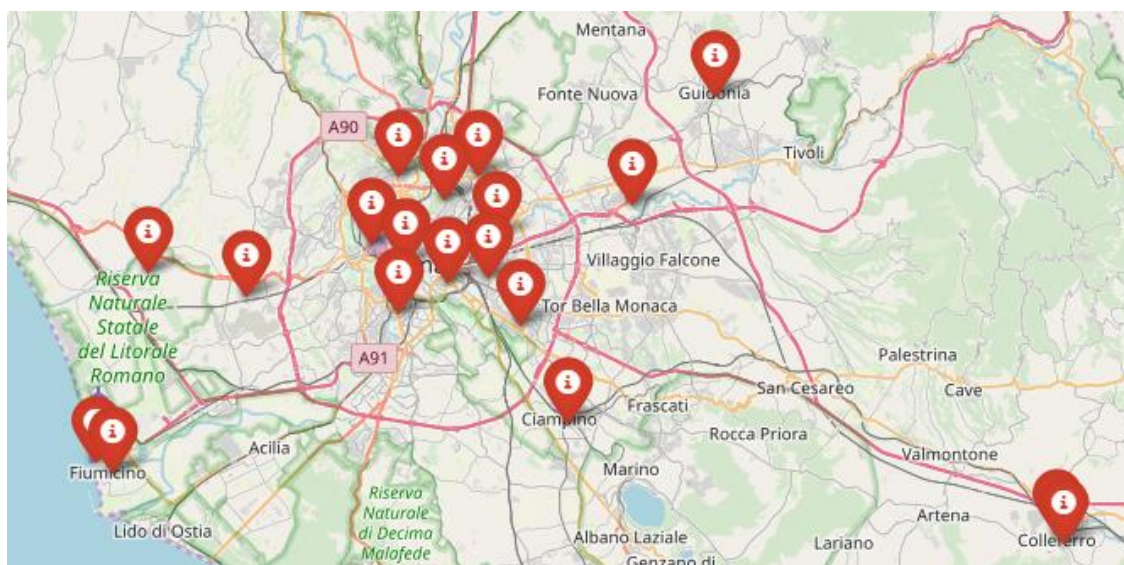

Figure 1: All control units

Figure 2: Filtered control units

For these control units a bar chart is plotted in order to have, in a fast way, a visual feedback of the cleanest neighborhood. The problem is that every day PM10 value changes and the same for bar charts, maps and consequently results. It is possible to see some of results in Figure 3 and Figure 4 where the drop in terms of daily mean (dark gray) is due to a heavy nightly rain.
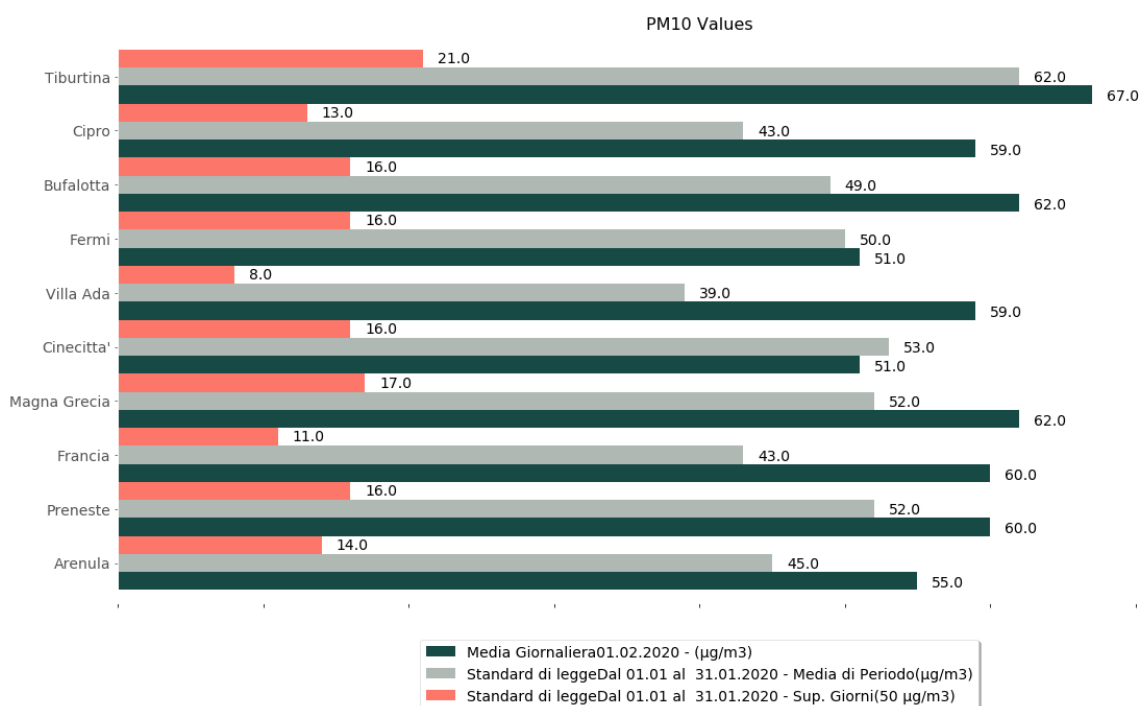

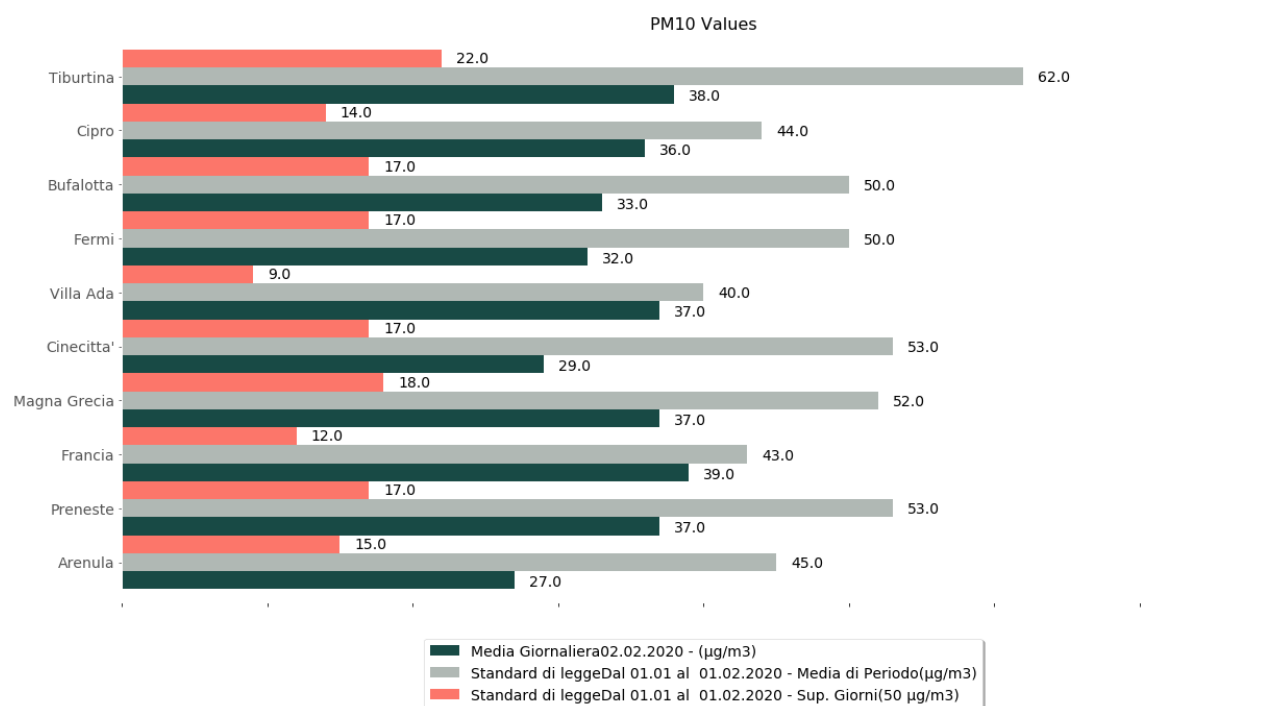*Figure 3: PM10 values on feb 01, 2020*

Figure 4: PM10 values on feb 02, 2020

	index	Rete	Stazione	Latitudine	Longitudine	QuotasImm	Media Giornaliera02.02.2020 - (µg/m3)	Standard di leggeDal 01.01 al 01.02.2020 - Media di Periodo(µg/m3)	Standard di leggeDal 01.01 al 01.02.2020 - Sup. Giorni(50 µg/m3)
0	0	Roma	Arenula	41.894020	12.475368	31.0	27.0	45.0	15.0
1	4	Roma	Cinecitta'	41.857720	12.568665	53.0	29.0	53.0	17.0
2	14	Roma	Fermi	41.864194	12.469531	26.0	32.0	50.0	17.0

Figure 5: The best three control units

The distance between the main neighborhoods and the best three control units were calculated.

	Sigla	Quartiere	Latitudine	Longitudine	Distance:0 (Arenula)	Distance:1 (Cinecitta')	Distance:2 (Fermi)
0	Q. I	Flaminio	41.929990	12.464510	4099.361283	11785.650891	7327.984804
1	Q. II	Parioli	41.931770	12.486220	4292.588701	10694.200713	7640.022079
2	Q. III	Pinciano	41.918860	12.484100	2855.047850	9758.174626	6197.060894
3	Q. IV	Salario	41.913310	12.502170	3085.583379	8277.064254	6093.213908
4	Q. V	Nomentano	41.914470	12.522100	4486.429075	7394.533312	7084.414132
5	Q. VI	Tiburtino	41.897850	12.521010	3801.674406	5956.388178	5671.762671
6	Q. VII	Prenestino-Labicano	41.885700	12.535210	5039.187799	4165.716573	5940.553343
7	Q. VIII	Tuscolano	41.871640	12.540550	5942.338888	2795.758364	5938.751706
8	Q. IX	Appio-Latino	41.874610	12.513330	3812.413183	4952.109546	3807.157903
9	Q. X	Ostiense	41.860010	12.480470	3805.262849	7308.488870	1018.373886
10	Q. XI	Portuense	41.852830	12.457240	4819.797409	9244.414922	1622.614894
11	Q. XII	Gianicolense	41.874220	12.457660	2645.054661	9373.401735	1486.299986
12	Q. XIII	Aurelio	41.896100	12.439270	2996.776532	11531.837080	4343.191159
13	Q. XIV	Trionfale	41.921080	12.438060	4311.109509	12904.170233	6840.836521

For these days the same control units are the best so we obtain the same map. The radius of green circles is the distance between the selected control units and the nearest neighborhoods.
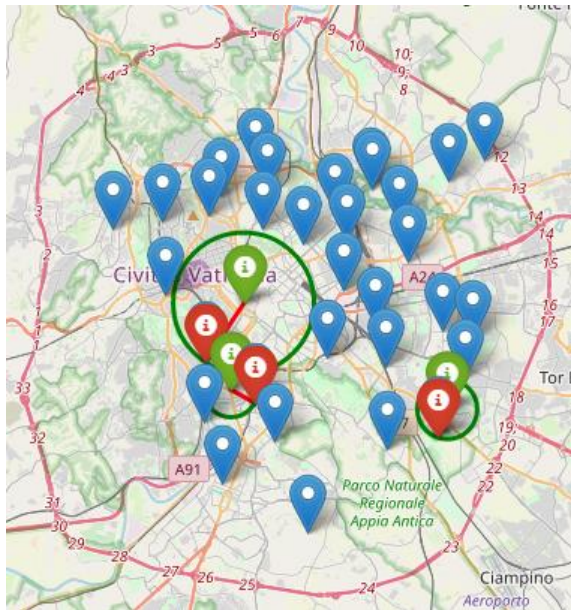

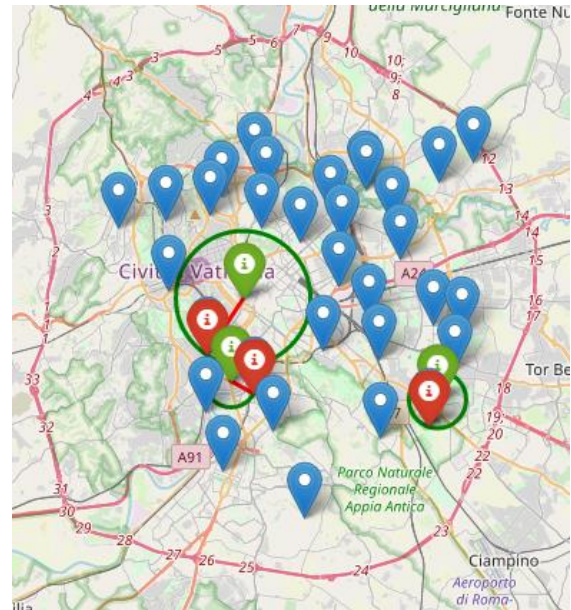Figure 6: Best control units on feb 01, 2020

Figure 7: Best control units on feb 02, 2020

Rome is full of commercial venues, so we limited to 30 the number of venues returned by Foursquare in a radius of 1000 meters of the geographical coordinates of the identified neighborhoods.

	Quartiere	Latitudine quartiere	Longitudine quartiere	Attivita'	Latitudine attiv.	Longitudine attiv.	Categoria
0	Gianicolense	41.87422	12.45766	L'Osteria di Monteverde	41.872937	12.455286	Mediterranean Restaurant
1	Gianicolense	41.87422	12.45766	Bruno ai Quattro Venti	41.877092	12.458166	Pizza Place
2	Gianicolense	41.87422	12.45766	Your Music	41.876792	12.457864	Music Store
3	Gianicolense	41.87422	12.45766	La Gourmandise	41.878868	12.462239	Ice Cream Shop
4	Gianicolense	41.87422	12.45766	Bar Vitali dal 1922	41.875480	12.461995	Restaurant

Figure 8: Returned venues

Quartiere	Latitudine quartiere	Longitudine quartiere	Attivita'	Latitudine attiv.	Longitudine attiv.	Categoria
Appio Claudio	30	30	30	30	30	30
Gianicolense	30	30	30	30	30	30
Ostiense	30	30	30	30	30	30

Figure 9: Returned venues grouped by neighborhoods and then counted

A frequency analysis has done in order to retrieve the most common venues for the selected neighborhoods along with a clustering operation conducted to localize the center of those venues.

	Quartiere	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Appio Claudio	Pizza Place	Ice Cream Shop	Italian Restaurant	Café	Plaza
1	Gianicolense	Italian Restaurant	Pizza Place	Café	Ice Cream Shop	Restaurant
2	Ostiense	Italian Restaurant	Café	Ice Cream Shop	Pub	Diner

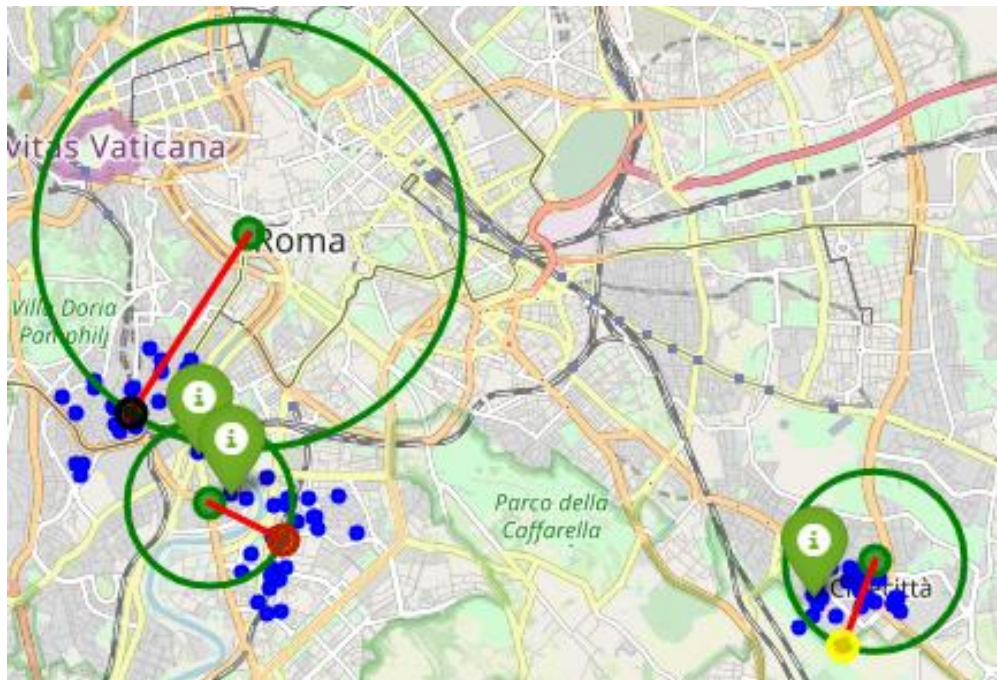Figure 10: Most common venues

Results and conclusions

Figure 11: Commercial returned venues and their clusters

From the map shown in Figure 11 it is possible to see that the cluster position corresponds to the exact position of the chosen neighborhoods (the less polluted). This means that when Foursquare returns those commercial venues, it does using a cluster logic maybe based on a less/mean distance from the specified point of interest.

	index	Quartiere	Latitudine quartiere	Longitudine quartiere	Attivita'	Latitudine attiv.	Longitudine attiv.	Categoria	Distance:0 (Arenula)	Distance:1 (Cinecitta')	Distance:2 (Fermi)
0	59	Appio Claudio	41.84835	12.56391	La Dolce Vita	41.853951	12.559284	Lounge	8253.855618	882.775751	7519.797217
1	22	Gianicolense	41.87422	12.45766	Tiger	41.869838	12.467817	Furniture / Home Store	2760.663606	8459.307621	643.404463
2	80	Ostiense	41.86001	12.48047	Osteria Mavi	41.865330	12.472500	Italian Restaurant	3199.010922	8008.602520	276.412515

Figure 12: Selected venues

CN