RUN AWAY FROM ELISA

DATA

We will use the following data:

- An .shp file, containing some information about municipi of Rome, downloaded by the official site of 'Comune di Roma'.
- An .shp file, containing some information about municipi of Milan, downloaded by the official site of 'Comune di Milano'.
- The wikipedia's page of Comune di Milano.
- A .csv file, containing the addresses of the municipi of Milan.
- Foursquare, that we will use to get the venues of all municipi.

  Municipio is the italian for borough.

DATA

The .shp file of Rome contains information about the population, the municipalities, the shape of the municipi (hence the area and the coordinates).

The .shp file of Milan contains information about the shape of the municipi.

The wikipedia's page contains information about the population.

DATA

We will use Foursquare get methods 'https://api.foursquare.com/v2/venues/explore' (to search all venues of each municipio) and 'https://api.foursquare.com/v2/venues/search' (to search all venues of each municipio in a fixed category).

We will focus on the following categories:

- Food
- Bars
- Transportation categories (metro, ferries, tram...)
- Fun venues (cinemas, theaters...)

DATA

The following is a sample of the dataframe that we will create and that we will work with.

	city	number	name	population	area	density	shape	munic_address
17	rome	9	Rome, Municipio 9	177032.0	198.125655	893.533954	POLYGON ((12.61272 41.70894, 12.61241 41.70848, 12.61216 41.70802, 12.61189 41.70775, 12.61117 4...	V.le Ignazio Silone, 100
20	rome	12	Rome, Municipio 12	142350.0	79.232385	1796.613840	POLYGON ((12.47519 41.88329, 12.47341 41.88207, 12.47237 41.88129, 12.47172 41.88079, 12.47116 4...	Via Fabiola, 14
13	rome	5	Rome, Municipio 5	245980.0	29.187286	8427.642199	POLYGON ((12.61131 41.92054, 12.61179 41.91998, 12.61284 41.91866, 12.61458 41.91644, 12.61508 4...	Via Torre Annunziata, 1
4	milan	5	Milan, Municipio 5	126089.0	34.467747	3658.173551	POLYGON ((9.18957 45.45197, 9.19036 45.45195, 9.19095 45.45195, 9.19271 45.45197, 9.19398 45.451...	Viale Tibaldi 41
11	rome	3	Rome, Municipio 3	204538.0	106.608384	1918.592070	POLYGON ((12.61873 41.99026, 12.61866 41.98983, 12.61856 41.98930, 12.61848 41.98866, 12.61842 4...	Via Umberto Fracchia, 45
3	milan	4	Milan, Municipio 4	161551.0	23.822123	6781.553532	POLYGON ((9.26759 45.47204, 9.26781 45.47140, 9.26894 45.47154, 9.27029 45.47148, 9.27028 45.470...	Via Oglio 18
21	rome	13	Rome, Municipio 13	137132.0	72.728879	1885.523345	POLYGON ((12.46128 41.88836, 12.46030 41.88807, 12.45937 41.88779, 12.45842 41.88753, 12.45785 4...	Via Aurelia, 470
9	rome	1	Rome, Municipio 1	197258.0	22.111084	8921.226975	POLYGON ((12.51578 41.89285, 12.51542 41.89242, 12.51521 41.89219, 12.51544 41.89179, 12.51516 4...	Via Luigi Petroselli, 50
8	milan	9	Milan, Municipio 9	187773.0	24.215001	7754.408232	POLYGON ((9.17572 45.53560, 9.17574 45.53537, 9.17608 45.53539, 9.17648 45.53542, 9.17680 45.535...	Via Guerzoni 38
12	rome	4	Rome, Municipio 4	177950.0	53.273109	3340.334456	POLYGON ((12.52850 41.92764, 12.52913 41.92764, 12.52924 41.92754, 12.52941 41.92730, 12.52966 4...	Via Tiburtina, 1163

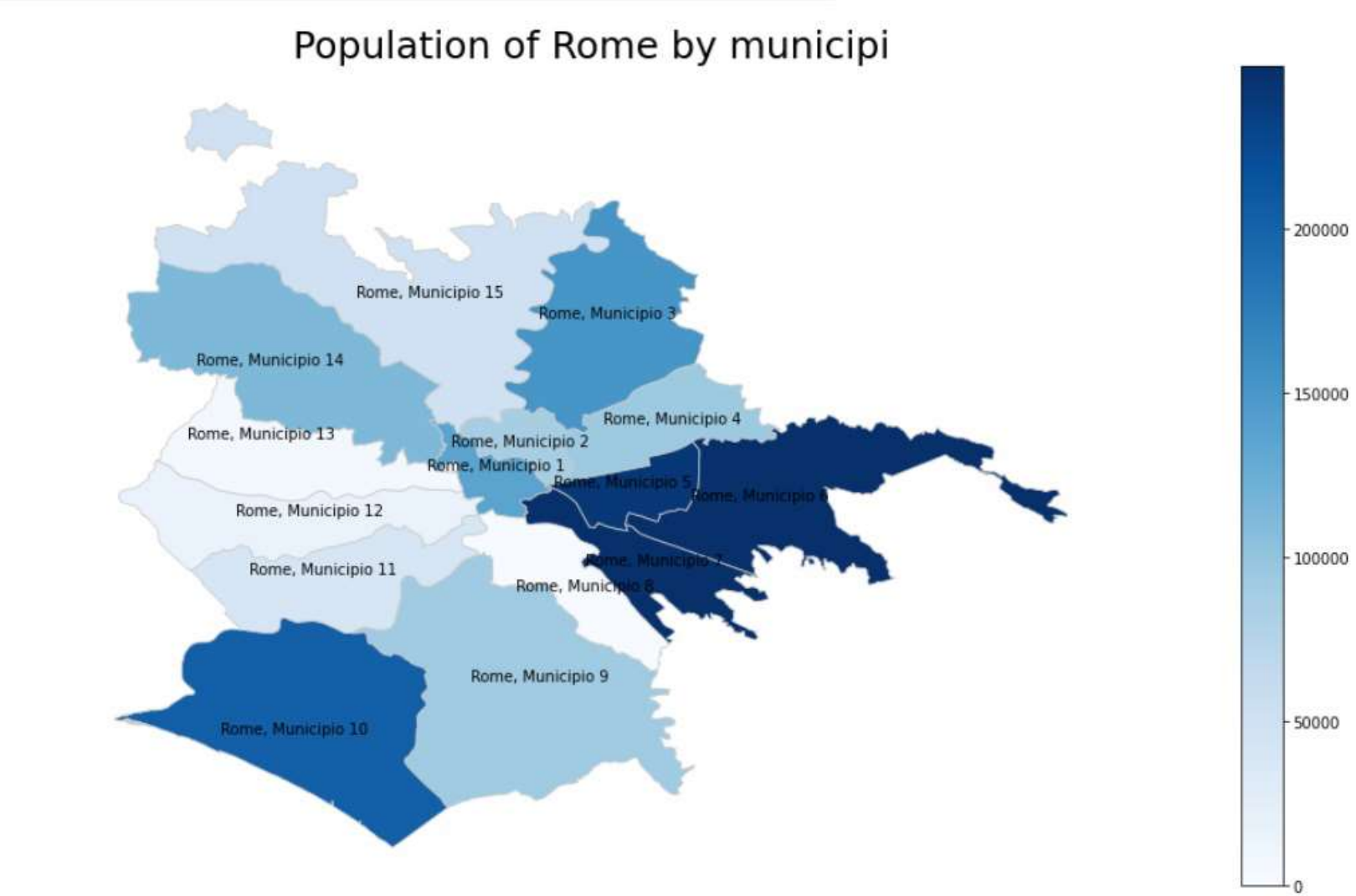
METHODOLOGY/RESULTS

In a first part of the work we will create a dataframe (df3) with all the information that we need. Every row will refer to a municipio and the columns contain information about:

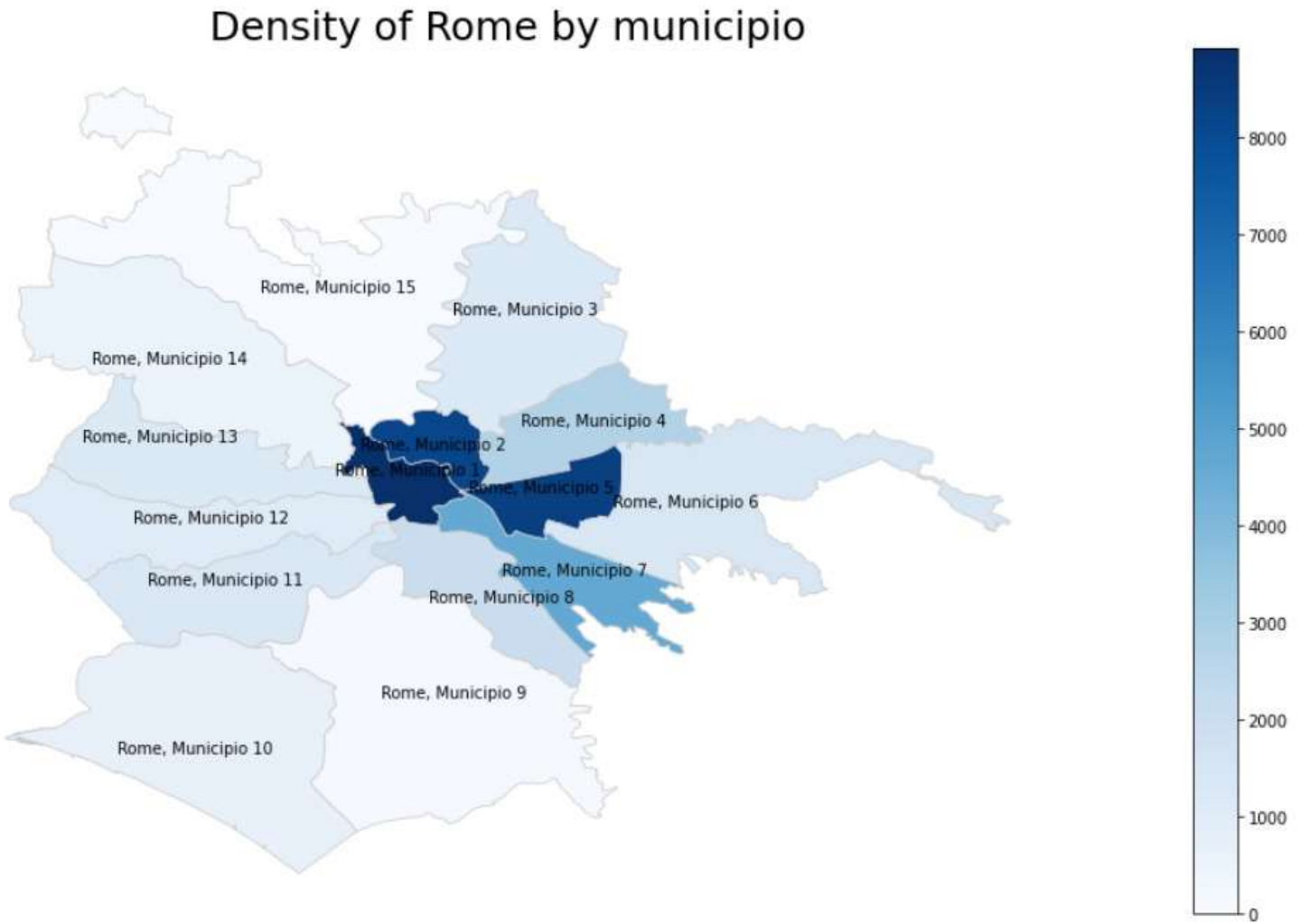
- Municipio's city
- Municipio's number
- Municipio's name
- Municipio's shape (the column geometry of the .shp file)
- Municipality latitude and longitude
- Representative latitude and longitude
- Centroid latitude and longitude
- Municipio's population
- Municipio's area
- Municipio's density of population

METHODOLOGY/RESULTS

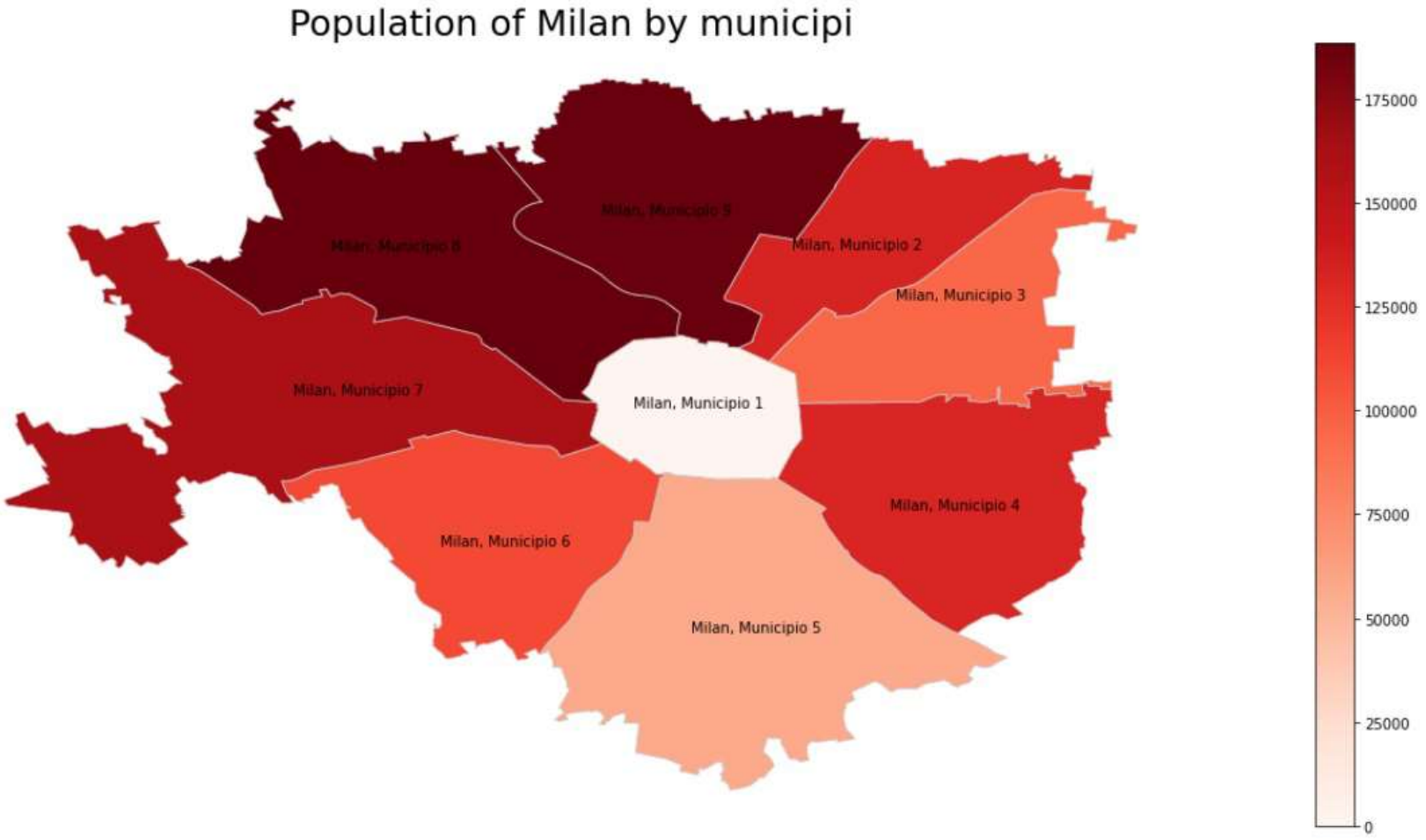
First we will visualize our data.



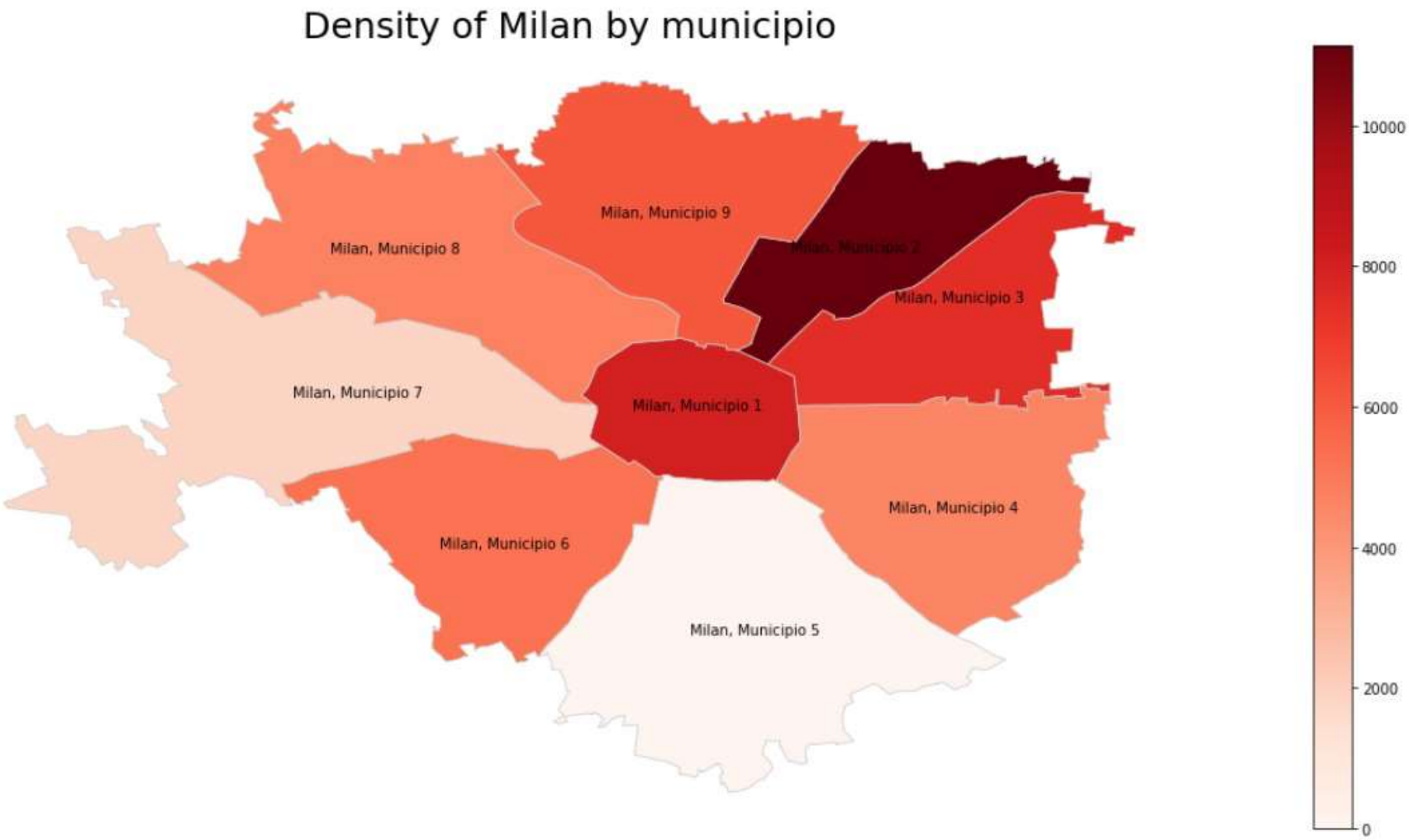
METHODOLOGY/RESULTS



METHODOLOGY/RESULTS



METHODOLOGY/RESULTS



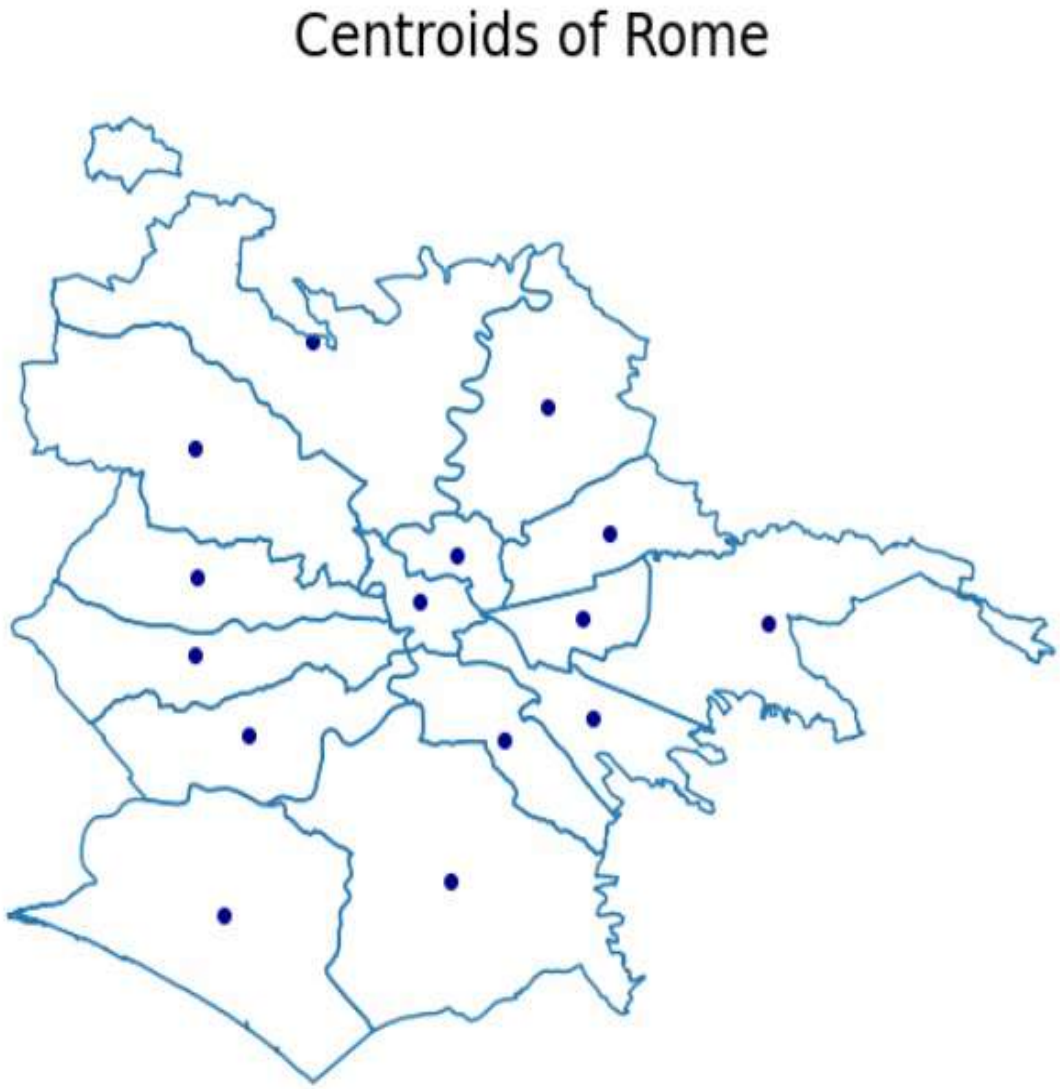
METHODOLOGY/RESULTS

Next we will use Foursquare.

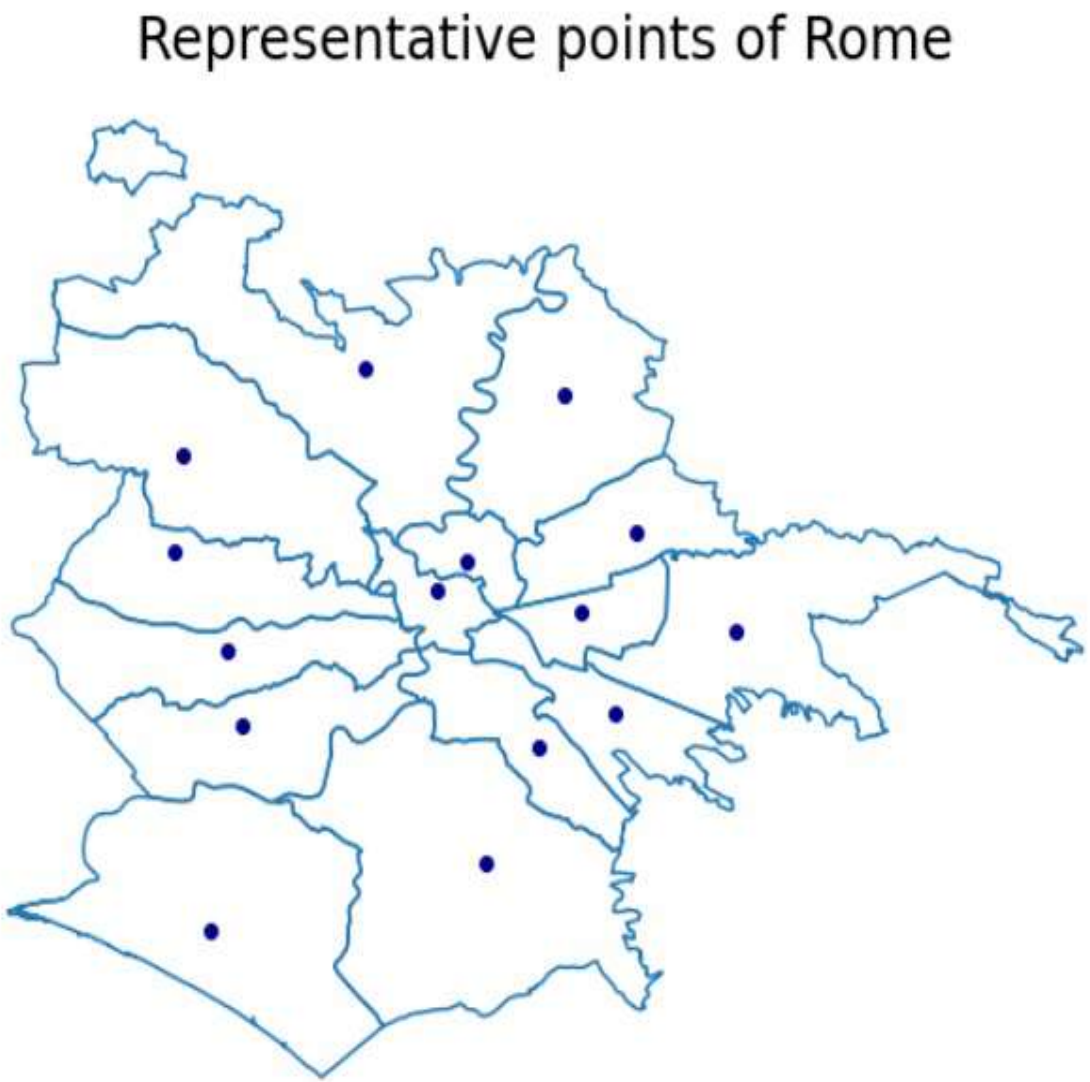
In order to use Foursquare we will assign to each municipio a circle. We will choose the center of the circle between the municipality, the representation point and the centroid (our choice will be the representative point) and we will calculate the radius of each circle (see the notebook).

The following pictures justify the choice of the representative points.

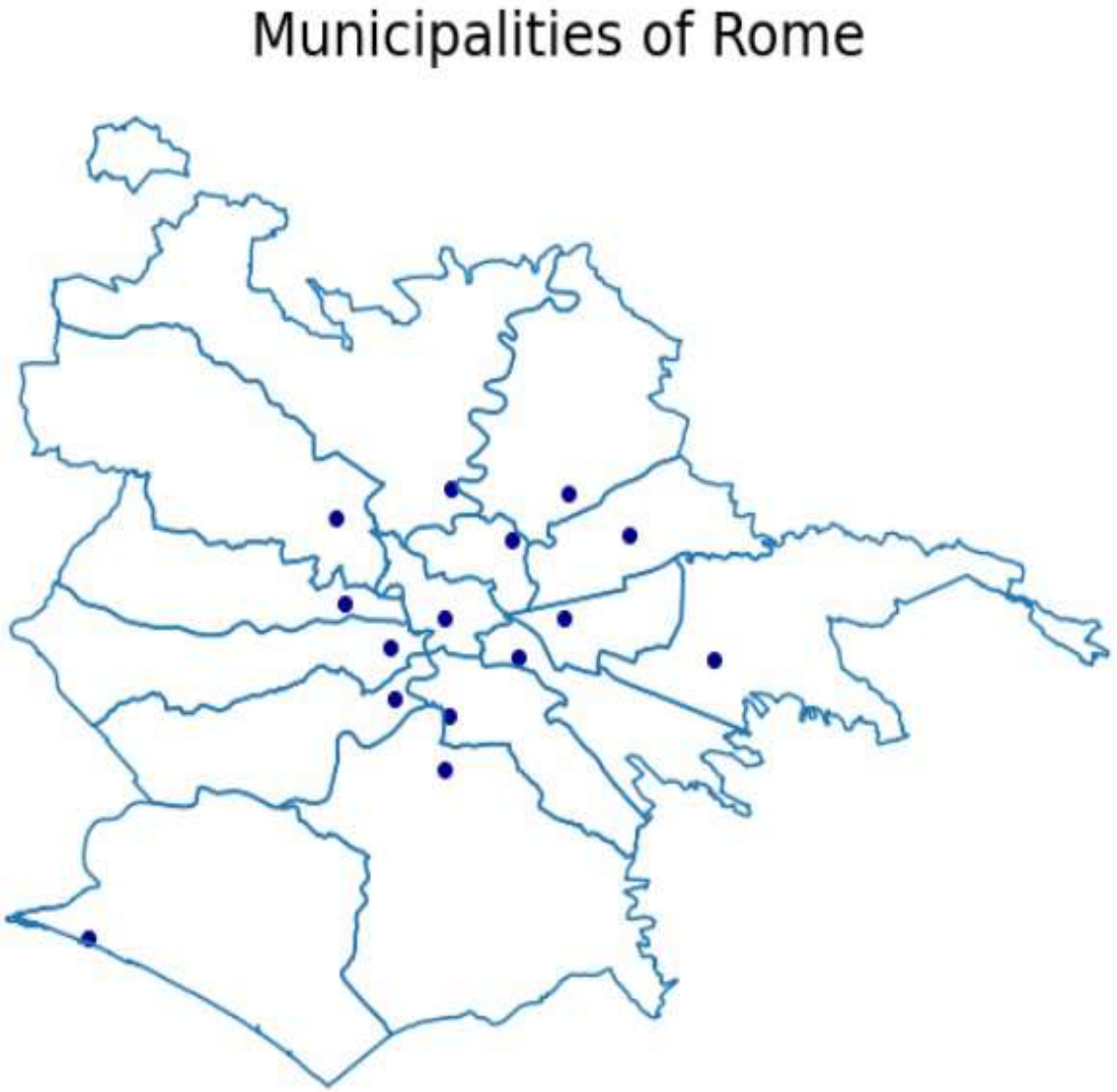
METHODOLOGY/RESULTS



METHODOLOGY/RESULTS



METHODOLOGY/RESULTS



METHODOLOGY/RESULTS

We will use the Foursquare get method 'https://api.foursquare.com/v2/venues/search' to get all the venues of each municipio in the following two categories:

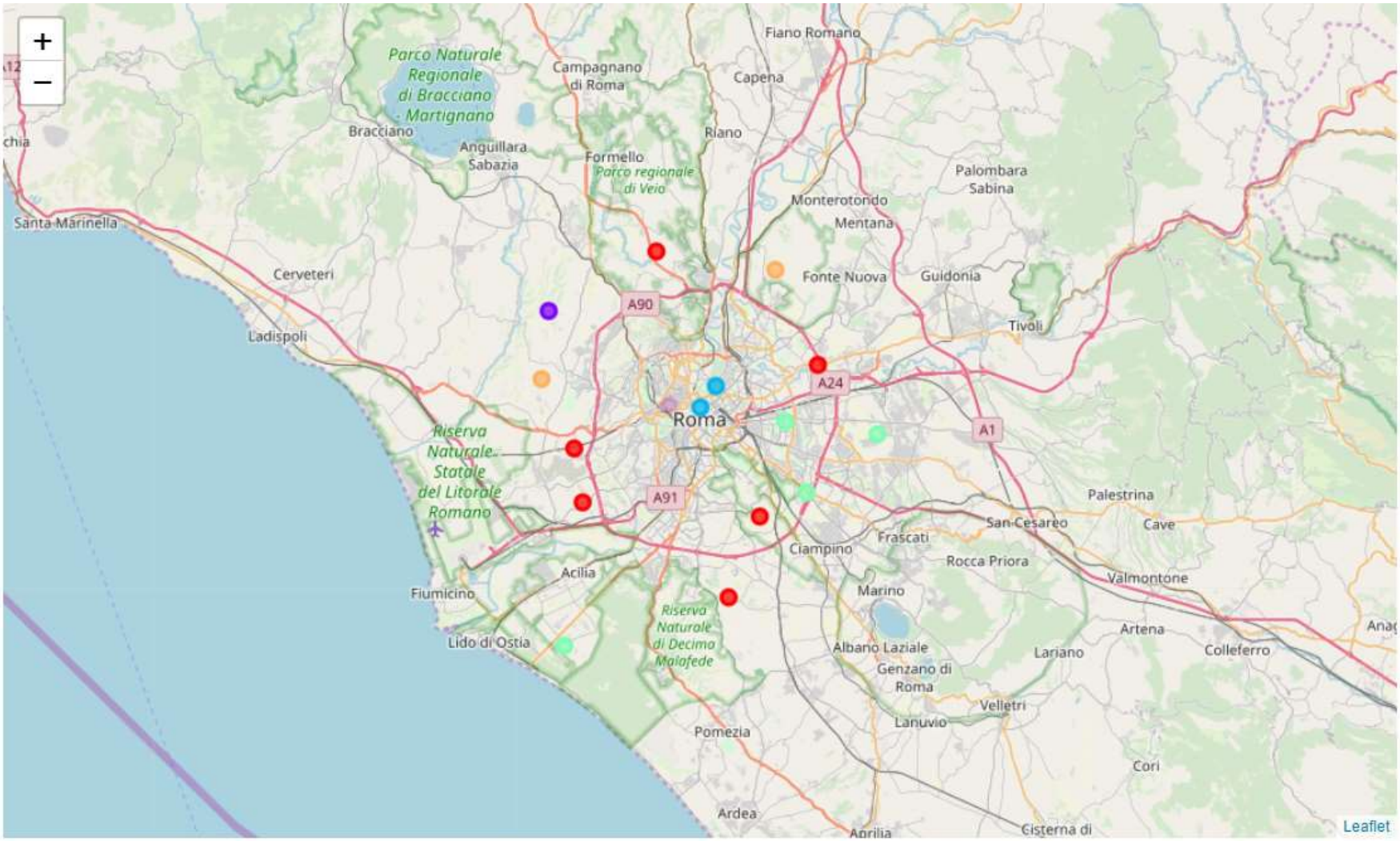
- Food venues
- Bar venues

Then we will analyze all the venues retrieved and we will get two different clusterizations of the municipi, using KMeans and choosing the best K with the elbow method (see the notebook).

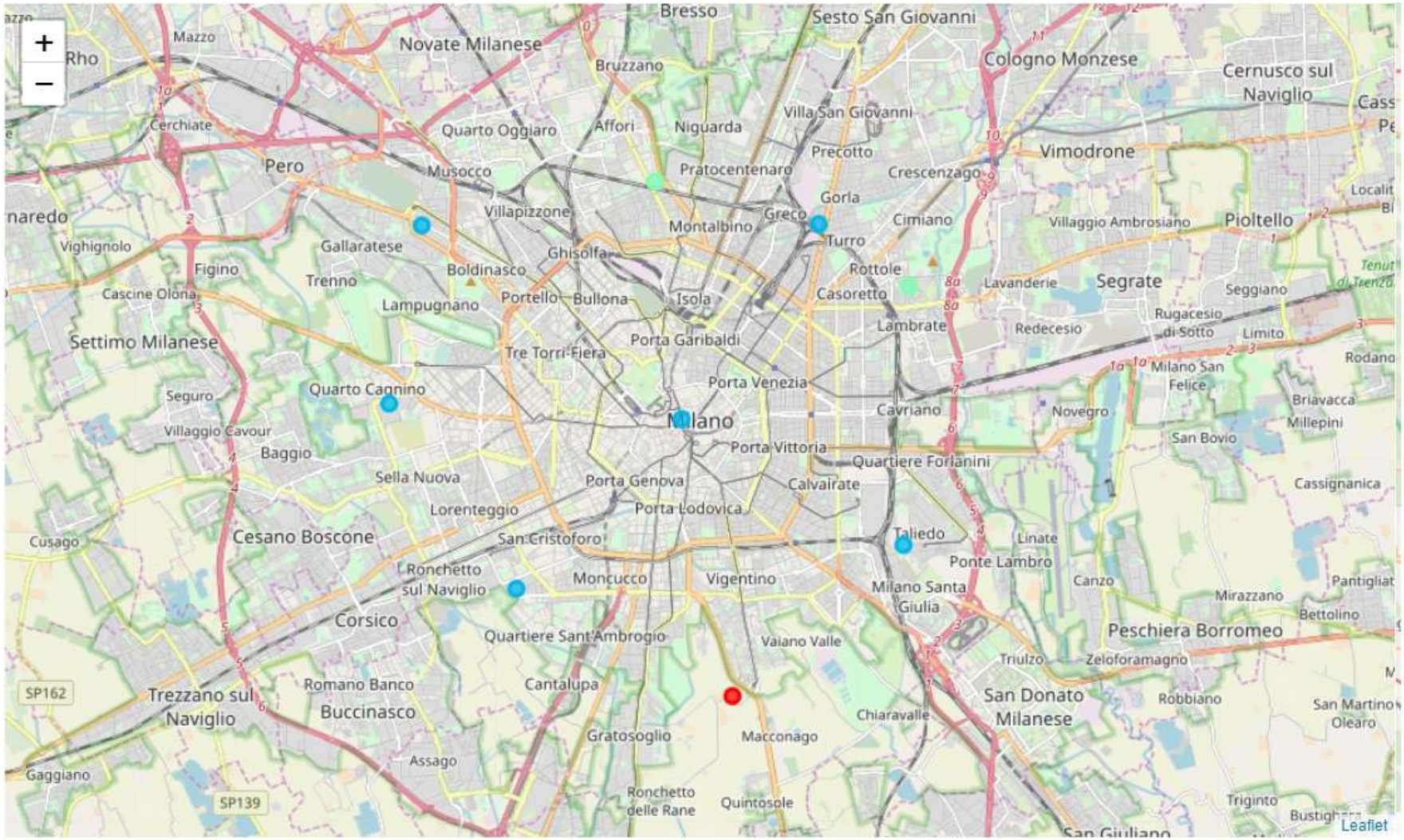
The following pictures show the result of the two clusterizations.

The first two images refer to the food clusterization, the section two images to the bar clusterization.

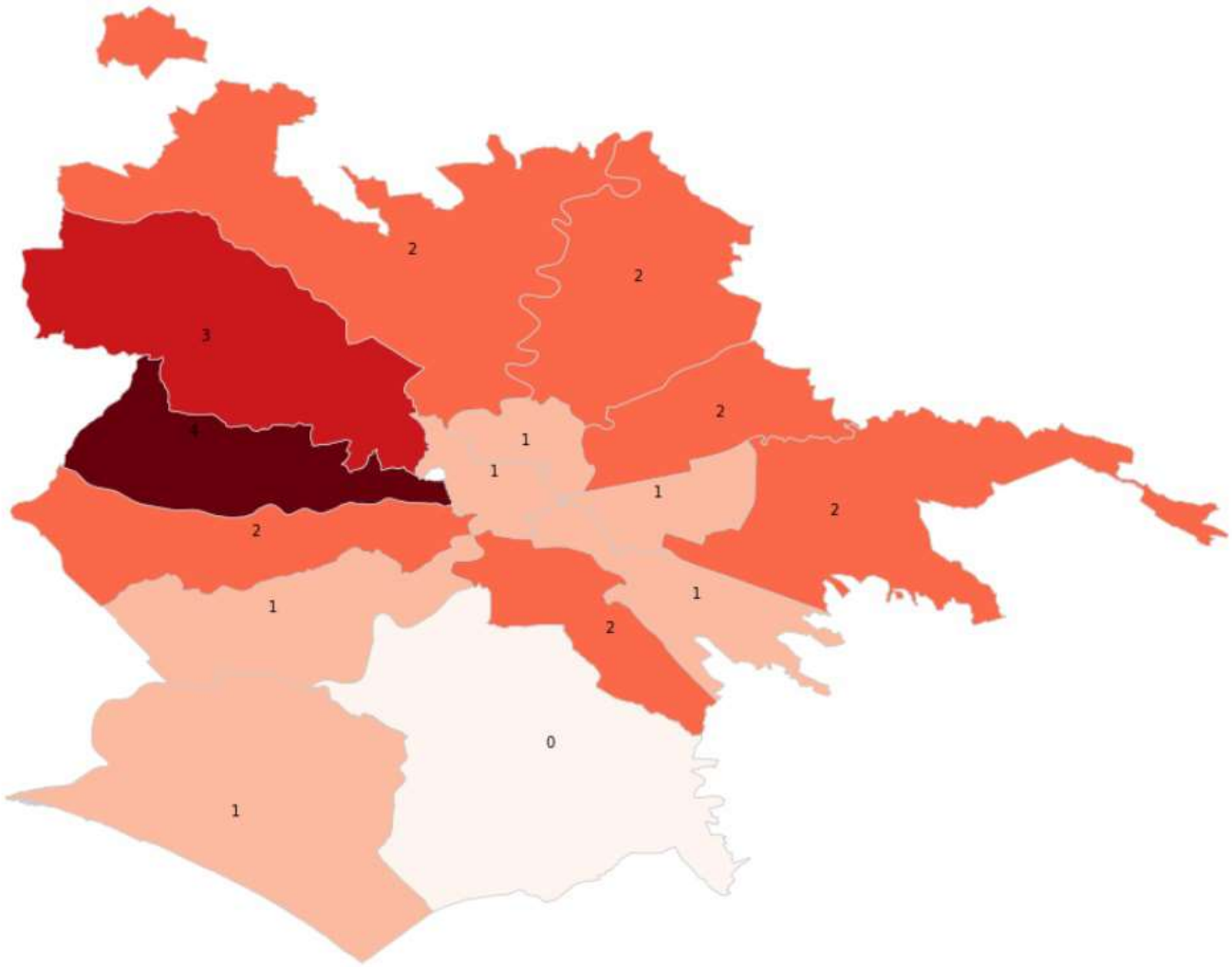
METHODOLOGY/RESULTS



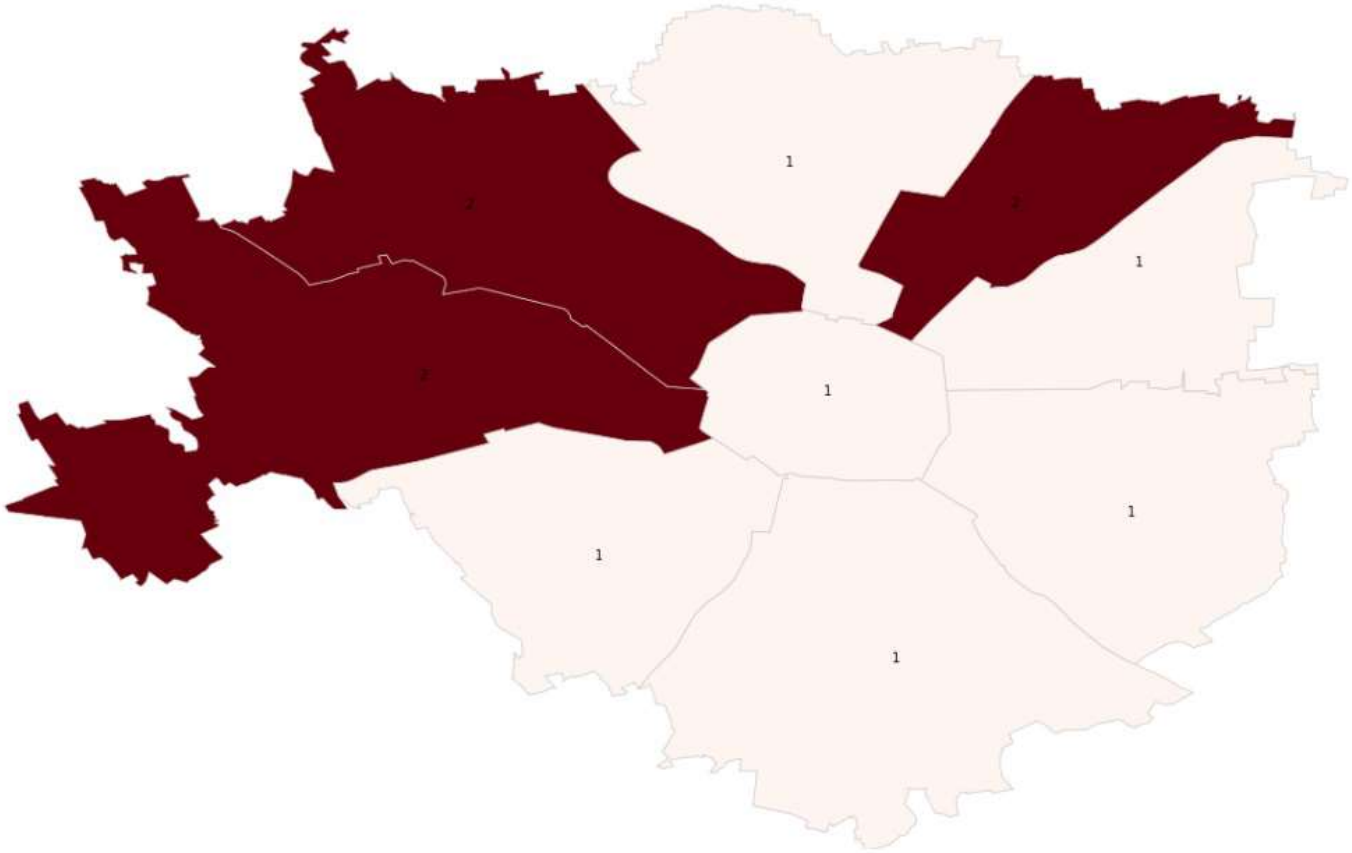
METHODOLOGY/RESULTS



METHODOLOGY/RESULTS



METHODOLOGY/RESULTS



METHODOLOGY RESULTS

In the last part of the work, we will use the Foursquare get method 'https://api.foursquare.com/v2/venues/search' to get all the venues of each municipio in the following categories:

- Transportation venues (metro, ferries, tram...)
- Fun venues (cinemas, theaters...)

Then we will assign a transportation score and a fun score to each municipio, dividing all municipi in four quartiles (see the notebook).

We will get the following dataframe. Foursquare does not work so well in Italy, that's why NaN values.

METHODOLOGY/RESULTS

	city	name	transp_venues_score	transp_venues_cut	fun_venues_score	fun_venues_cut
0	milan	Milan, Municipio 1	111271.0	3rd_quartile	76765.0	4th_quartile
1	milan	Milan, Municipio 2	33808.0	1st_quartile	10428.0	2nd_quartile
2	milan	Milan, Municipio 3	555958.0	4th_quartile	329237.0	4th_quartile
3	milan	Milan, Municipio 4	284349.0	4th_quartile	46821.0	3rd_quartile
4	milan	Milan, Municipio 5	242426.0	4th_quartile	43675.0	3rd_quartile
5	milan	Milan, Municipio 6	401615.0	4th_quartile	232047.0	4th_quartile
6	milan	Milan, Municipio 7	128147.0	3rd_quartile	8481.0	2nd_quartile
7	milan	Milan, Municipio 8	136748.0	3rd_quartile	169838.0	4th_quartile
8	milan	Milan, Municipio 9	164716.0	4th_quartile	76305.0	4th_quartile
9	rome	Rome, Municipio 1	79912.0	2nd_quartile	19732.0	2nd_quartile
10	rome	Rome, Municipio 2	32955.0	1st_quartile	6376.0	1st_quartile
11	rome	Rome, Municipio 3	105638.0	3rd_quartile	40868.0	3rd_quartile
12	rome	Rome, Municipio 4	71016.0	2nd_quartile	22031.0	3rd_quartile
13	rome	Rome, Municipio 5	186886.0	4th_quartile	15864.0	2nd_quartile
14	rome	Rome, Municipio 6	89746.0	2nd_quartile	5850.0	1st_quartile
15	rome	Rome, Municipio 7	39888.0	2nd_quartile	64896.0	3rd_quartile
16	rome	Rome, Municipio 8	7494.0	1st_quartile	1128.0	1st_quartile
17	rome	Rome, Municipio 9	3723.0	1st_quartile	18927.0	2nd_quartile
18	rome	Rome, Municipio 10	136693.0	3rd_quartile	5053.0	1st_quartile
19	rome	Rome, Municipio 11	3855.0	1st_quartile	5274.0	1st_quartile
20	rome	Rome, Municipio 12	5474.0	1st_quartile	NaN	NaN
21	rome	Rome, Municipio 13	67092.0	2nd_quartile	NaN	NaN
22	rome	Rome, Municipio 14	NaN	NaN	NaN	NaN
23	rome	Rome, Municipio 15	NaN	NaN	NaN	NaN

METHODOLOGY/RESULTS

We will finally select all the municipi of Rome and Milan that satisfy the following criteria:

- They are not municipio 2 of Rome (where my ex-girlfriend live).
- They are in the same food cluster of municipio 2 of Rome.
- They are in the same bar cluster of municipio 2 of Rome.
- They are in the 1st and 2nd quartile according to transportation score.
- They are in the 1st and 2nd quartile according to fun score.

The only municipio that satisfies those conditions is Rome, municipio 1.

METHODOLOGY/RESULTS

We select all the munciipi that have the same food cluster of municipio 2 of Rome, the same bar cluster of municipio 2 of Rome.

```
mask1 = (df3['bar_cluster'] == 1) & (df3['food_cluster'] == 1)
df3[mask1]['name']
```

```
0     Milan, Municipio 1
3     Milan, Municipio 4
5     Milan, Municipio 6
9      Rome, Municipio 1
10     Rome, Municipio 2
Name: name, dtype: object
```

We select all the munciipi that have the transportation score in the 1st and 2nd quartiles and fun score in the 1st and 2nd quartiles.

```
mask2 = (df3['transp_venues_cut'].isin(['1st_quartile', '2nd_quartile'])) & (df3['fun_venues_cut'].isin(['1st_quartile', '2nd_qu
df3[mask2]['name']
```

```
1     Milan, Municipio 2
9      Rome, Municipio 1
10     Rome, Municipio 2
14     Rome, Municipio 6
16     Rome, Municipio 8
17     Rome, Municipio 9
19    Rome, Municipio 11
Name: name, dtype: object
```

```
mask3 = mask1 & mask2
df3[mask3]['name']
```

```
9     Rome, Municipio 1
10    Rome, Municipio 2
Name: name, dtype: object
```

DISCUSSION

The analysis is not exhaustive. Clearly other criteria may be used to infer the best municipio. I choose those criteria according to my personal taste, other choices may be done.

I remark that Foursquare does not perform very well in Italy. That's why the results can appear limited.

However, the employed methods and all the functions that I wrote can work with every city in the world. You only need a .shp file of the city, with municipi's shape and information about population.

DISCUSSION

You can also change the criteria, just changing the foursquare categories (food, bar, transportations, fun).

The majority of data are not updated and incoherent. For example the information about population of Rome and Milan refer to different years. I focused on the consistence of my code and not on the reliability of the data that I used.

To be honest, if you download the .shp file from the official website of Comune di Roma, you have a problem, because the don't specify which CRS they used. They work with EPGS 3004, I converted it, with an external program, to the standard EPGS 4326 to use the same CRS of Milan data. I really don't now how to do it in my code.

CONCLUSIONS

The optimal municipio is municipio 1 in Rome.

I will move next to the Coliseum!!!

THANK YOU!