LIVING IN NY

Luca Di Narzo

1. Introduction

1.1 Description of the problem and a discussion of the background.

My name is Luca and i come from Milan, Italy. I've always lived in Italy and for us a good espresso coffee is fundamental. In order to find the best espresso in Milan I tried out many coffees and once I've chosen the best one i stick with it. Now I'm going to live in New York. The culture of coffee is different in other countries and I'm afraid I won't be able to find enough coffee shops in NY. That's why I want to apply what I've learned so far to discover a neighbor where there're enough coffee shops in NY like Milan.

1.2 Business model

This work will be very helpful for other Italians who want to live abroad and apply the same methodology to other cities.

2. Data

2.1 Description of the data and how it will be used to solve the problem

The following data will be used:

- List of Boroughs and neighborhoods of Milan with their geodata;
- List of Boroughs and neighborhoods of Manhattan with their geodata;
- List of coffee shops of Milan with their geodata;
- List of coffee shops of Manhattan with their geodata.

2.2 Data Sources

Boroughs and neighborhoods of Milan from Wikipedia
(https://en.wikipedia.org/wiki/Municipalities_of_Milan);

Boroughs and neighborhoods of Manhattan from Wikipedia (https://en.wikipedia.org/wiki/Neighborhoods_in_New_York_City);

Geocode information from Geopy;

Coffee shops in Milan and Manhattan from Foursquare

3. Methodology

3.1 Data Processing.

The data taken from Wikipedia both for Milan and New York had to be cleaned.

In particular:

Milan:

For our purpose we needed just the name of the bourogh and the district. The initial table was:

	Borough	Name	Area(km2)	Population(2014)	Population density(inhabitants/km2)	Quartieri (districts)
0	1.0	Centro storico	9.67	96315.000	11074	Brera, Centro Storico, Conca del Naviglio, Gua...
1	2.0	Stazione Centrale, Gorla, Turro, Greco, Cresce...	12.58	153.109	13031	Adriano, Crescenzago, Gorla, Greco, Loreto, Ma...
2	3.0	Città Studi, Lambrate, Porta Venezia	14.23	141229.000	10785	Casoretto, Cimiano, Città Studi, Dosso, Lambra...
3	4.0	Porta Vittoria, Forlanini	20.95	156.369	8069	Acquabella, Calvairate, Castagnedo, Cavriano, ...
4	5.0	Vigentino, Chiaravalle, Gratosoglio	29.87	123779.000	4487	Basmetto, Cantalupa, Case Nuove, Chiaravalle, ...

Figure 1: Raw data

We cleaned the data selecting just the desired columns and splitting the district names obtaining a row for each district name. The final data obtained is:

	Borough	District
0	Centro storico	Brera
1	Centro storico	Centro Storico
2	Centro storico	Conca del Naviglio
3	Centro storico	Guastalla
4	Centro storico	Porta Sempione

Figure 2: Milan districts after the preprocessing step

New York

As for the Milan case we had to select the columns we needed and choose just Manhattan.

The raw data were:

	Community Board(CB)	Areakm2	Pop.Census2010	Pop./km2	Neighborhoods
0	Bronx CB 1	7.17	91497	12761	Melrose, Mott Haven, Port Morris
1	Bronx CB 2	5.54	52246	9792	Hunts Point, Longwood
2	Bronx CB 3	4.07	79762	19598	Claremont, Concourse Village, Crotona Park, Mo...
3	Bronx CB 4	5.28	146441	27735	Concourse, Highbridge
4	Bronx CB 5	3.55	128200	36145	Fordham, Morris Heights, Mount Hope, Universit...

Figure 3: NY raw data

The final data is:

	Borough	District
0	Manhattan CB 1	Battery Park City
1	Manhattan CB 1	Financial District
2	Manhattan CB 1	Tribeca
3	Manhattan CB 2	Chinatown
4	Manhattan CB 2	Greenwich Village

Figure 4: NY final data

3.2 Location Extraction

Both for Milan and New York just the tables with the Borough and District were available. To perform this analysis though the Latitude and Longitude of every district had to be found. To do this the package GEOPY has been used. The values of Latitude and Longitude have been added to the data frame obtained from the preprocessing step.

3.3 Number of cafes

Number of cafes in my district area.

First the number of cafes in Milan, in my district area have been found using the Foursquare package. In particular a radius of 500m has been selected for our purpose and we searched directly in the cafes category provided by Foursquare

Cafes in NY

In order to find the most suitable district area for me an analysis on all the Manhattan district had to be performed. For this purpose an iterator has been used to extract the number of cafes for every single district in Manhattan.

4. Results

4.1 Location extraction results.

In this section the maps showing my district area in Milan and all the districts area in Manhattan can be seen in the following figures:

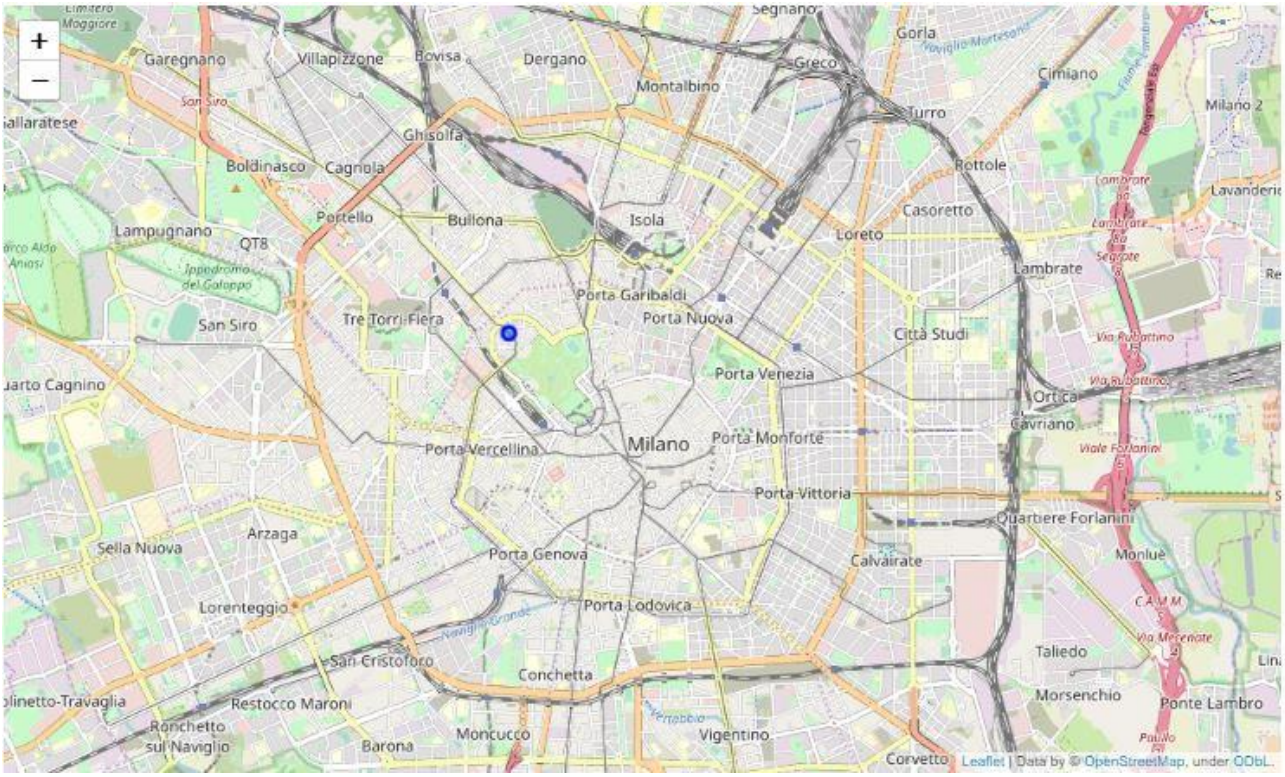

Figure 5 My district in Milan

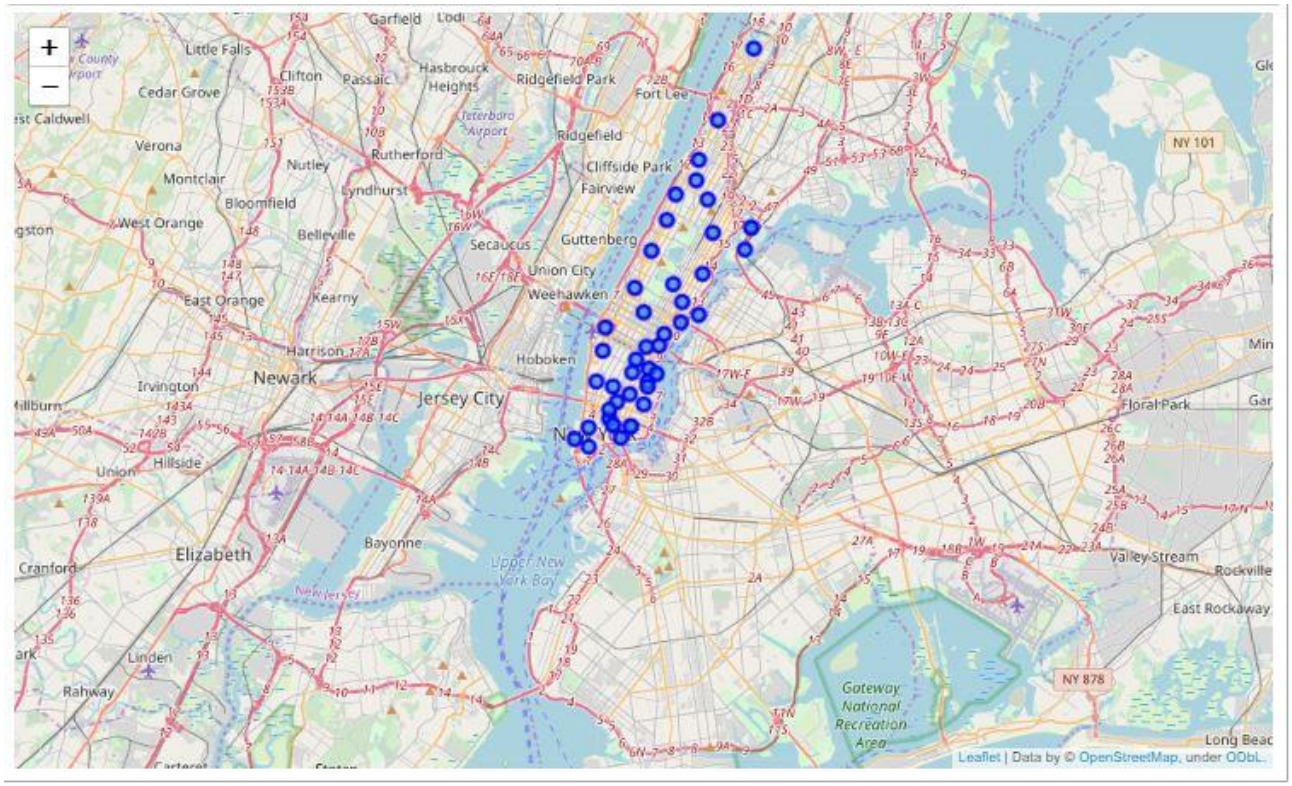

Figure 6: Districts in Manhattan

4.2 Number of cafes results.

In my district in Milan for a radius of 500m a total of 45 cafes have been found. The results can be seen in the following figure:

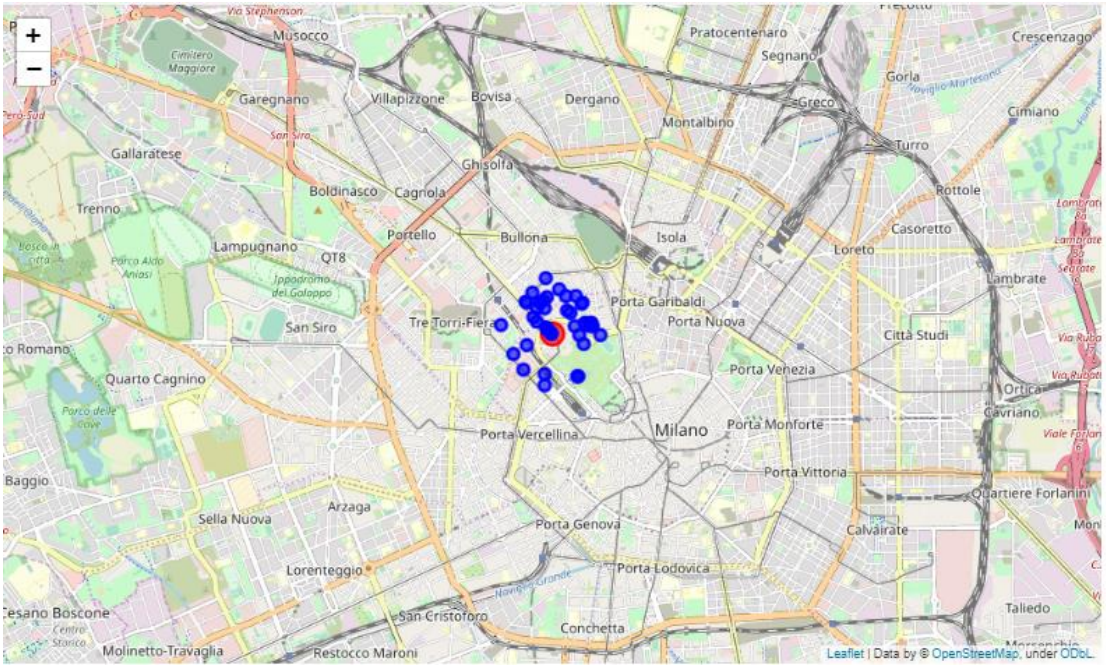

Figure 7: Cafes in my district

The iteration for the New York district produced very different results. In some districts the number of cafes were less than 5, in some cases they were 1.

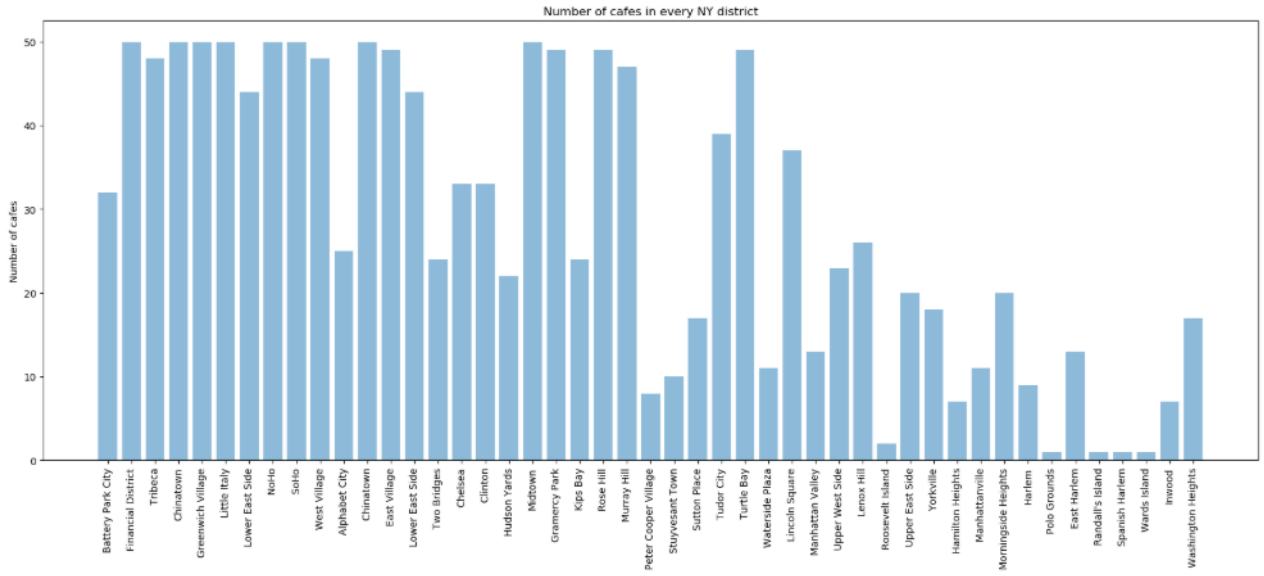

Figure 8: Number of cafes in NY

The final number of suitable districts with a number of cafes higher than 45 can be seen in the following table.

	Borough	District	latitude	longitude	Number of cafes
1	Manhattan CB 1	Financial District	40.707612	-74.009378	50
2	Manhattan CB 1	Tribeca	40.715380	-74.009306	48
3	Manhattan CB 2	Chinatown	40.716491	-73.996250	50
4	Manhattan CB 2	Greenwich Village	40.731980	-73.996566	50
5	Manhattan CB 2	Little Italy	40.719273	-73.998215	50
7	Manhattan CB 2	NoHo	40.725875	-73.993957	50
8	Manhattan CB 2	SoHo	40.722880	-73.998750	50
9	Manhattan CB 2	West Village	40.734186	-74.005580	48
11	Manhattan CB 3	Chinatown	40.716491	-73.996250	50
12	Manhattan CB 3	East Village	40.729269	-73.987361	49
18	Manhattan CB 5	Midtown	40.762268	-73.979544	50
19	Manhattan CB 6	Gramercy Park	40.737925	-73.985932	49
21	Manhattan CB 6	Rose Hill	40.743338	-73.984159	49
22	Manhattan CB 6	Murray Hill	40.748157	-73.978750	47
27	Manhattan CB 6	Turtle Bay	40.753467	-73.968866	49

Figure 9: Number of cafes for every district with an amount higher than Milan

5. Discussions

It turns out from this analysis that there's an high number of districts in Manhattan with an high concentration of cafes. This is what i expected from a cosmopolitan city like NY.

In this way now i know that i can go to NY and enjoy a good coffee everywhere.

6. Conclusions

This methodology has been very helpful to choose the district in NY with the highest concentration of cafes. The advantage is that it can be applied to every city and for different categories like restaurants, museum or gyms thank to the flexibility provided by Foursquare.

In the future this could be strengthen providing also a review rate of the various cafes present in the district.