REPORT:

Battle of the Neighborhoods - Where to open Italian restaurant in New York City

A. Introduction

This final project explores the best locations for opening an Italian Restaurant throughout New York City and its five major Boroughs. New York City has long been considered as one of the Food Capitals of the World and it is also the place, I call home for the last 11 years. NYC is a major metropolitan area with more than 8 million people living within city limits. It is one of the most diverse cities worldwide and considered a safe haven for immigrants.

Little Italy's neighborhood in Manhattan around Mulberry Street is well known as the mecca of Italian cuisine at least when tourists are concerned but is it really the best place to open a new Italian restaurant. This report explores which neighborhoods and boroughs of New York City have the most and the best Italian restaurants. We will try to answer the questions:
- Where are the best places to open an Italian Restaurant? and
- Where should one go for a taste of the best Italian food in the city?"

B. Data

To answer those questions, we will use different datasets related to New York City neighborhoods as well FourSquare API data to obtain the required restaurant likes, ratings and tips.

The main data sources will be some URL open source dare available to us like:
- https://cocl.us/new_york_dataset , which will give us data about New York City neighborhoods and boroughs and their latitudes, and longitudes.
- https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm - This dataset contains neighborhood boundaries and other addition information related to the issues at hand.

as well as:
- FourSquare API requests that will be used to obtain categories, restaurant ratings, tips, likes etc.

C. Methodology

New York City has a total of 5 boroughs and 306 neighborhoods. In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood.

I used requests to get the .json file from the new_york_daraset URL after that the task I performed was transforming this data of nested Python dictionaries into a *pandas* dataframe and geopy to convert the addresses into latitude and longitude values. Here are the top 5 results from the table:

|	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

Using the matplotlib library we can also see how many Neighborhoods are there in each Borough. From the graph below its obvious that Queens contains the most and Manhattan the least number of Neighborhoods:
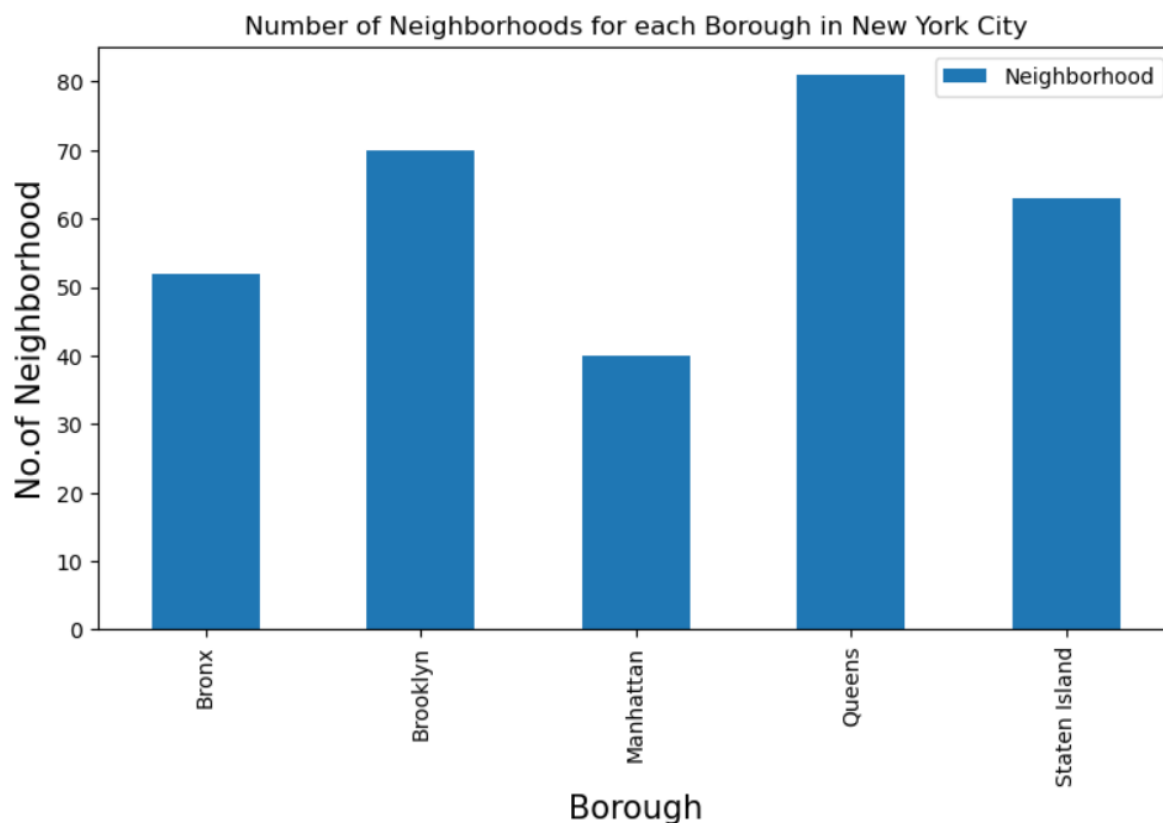
Now we can start utilizing the Foursquare API to explore the neighborhoods and segment them. What we are interested in this report more specifically is the Italian Restaurants – how many there are and where are they located in the city(which Neighborhood). FourSquare API **category** to Italian Restaurants and extracted the **venue_details ID and Name,** afer that combining it with the **new_york_data** and setting it in a **panda** dataframe:

	Unnamed: 0	Borough	Neighborhood	ID	Name
224	224	Manhattan	Flatiron	4b741ab4f964a5202cc82de3	Novità Cucina Creativa
225	225	Brooklyn	Fulton Ferry	593c0d2262420b7feccc3048	Cecconi's
226	226	Brooklyn	Dumbo	593c0d2262420b7feccc3048	Cecconi's
227	227	Staten Island	Prince's Bay	4c091311a1b32d7f5efb96f0	Certo's Gourmet Foods & Caterers
228	228	Staten Island	Lighthouse Hill	5079ccdce4b065f1dd8155ad	Casa verde

As we can see from the table there are 229 (panda dataframes start with 0) Italian Restaurants in New York city. Next I use **matplotlib** again to visualize where exactly are they situated by Borough and Neighborhood:
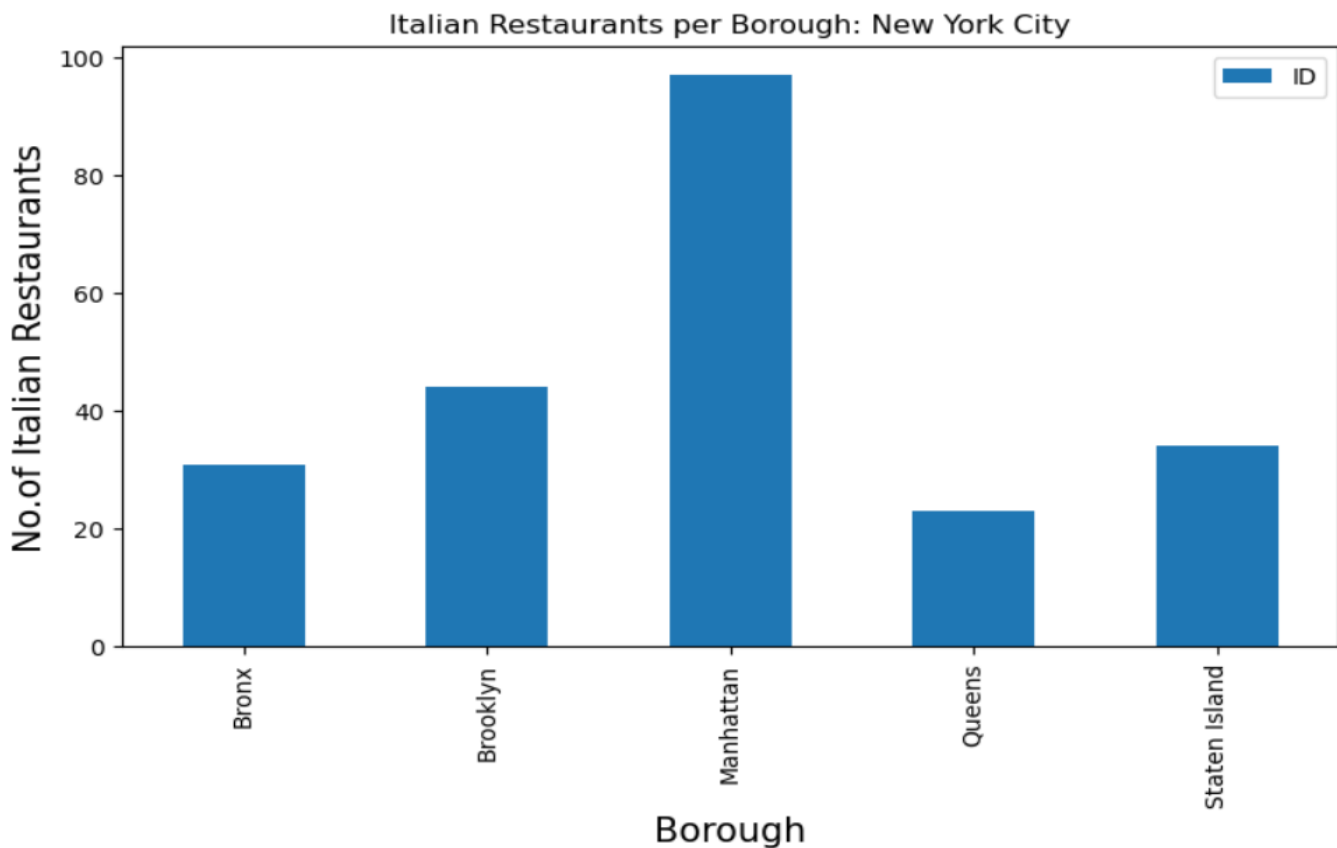
We can see that Manhattan has the biggest number of Italian Restaurants even though we established earlier that it has the least number of Neighborhoods. That means that the concentration of Italian Restaurants in Manhattan is much higher that any other Neighborhood.

Italian Restaurants per Neighborhood: New York City
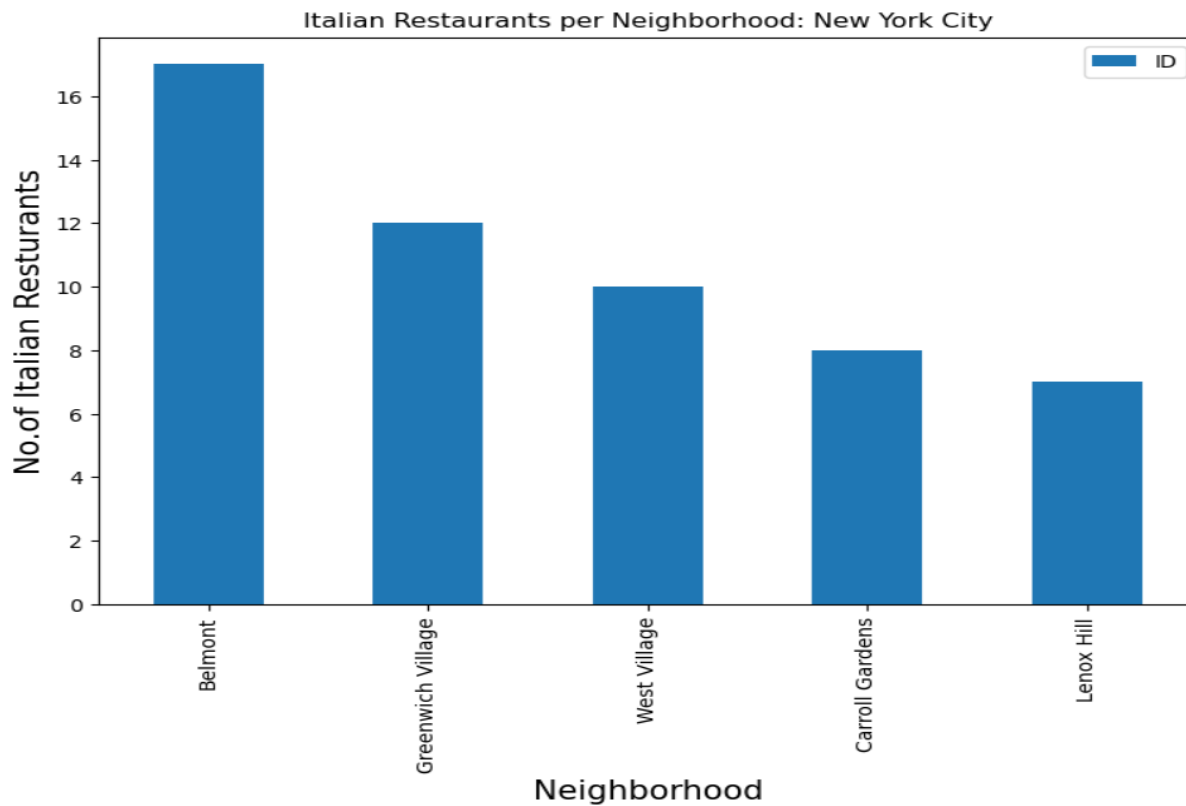
The top 3 Neighborhoods with the bigges concentration of Italian Reataurans are Belmont, Greenwich Village and West Village which are in the Boroughs of The Bronx and Manhattan. From them Greenwich Village and West Village are located in an area of the city that is popular with both tourists and locals alike withe Greenwich Village having the most restaurants.

Let's analyze the data further by ranking each restaurant. I set the FourSquare API **category** to Italian Restaurants and extracted the **venue_details(ID) of Likes, Rating and Tips** for each **ID:**

	Borough	Neighborhood	ID	Name	Likes	Rating	Tips
219	Manhattan	Turtle Bay	4acfb632f964a52076d520e3	Ristorante Grifone	15	7.1	6
220	Manhattan	Flatiron	5cc3adaa838e59002cb4629d	Rezdôra	86	8.8	22
221	Manhattan	Flatiron	4e72a19945dd1ac617629a71	Il Pesce	125	8.6	49
222	Manhattan	Flatiron	3fd66200f964a520bce51ee3	Basta Pasta	435	8.3	149
223	Manhattan	Flatiron	53370455498e18d7e6607a1d	La Pizza & La Pasta	199	8.5	57
224	Manhattan	Flatiron	4b741ab4f964a5202cc82de3	Novità Cucina Creativa	110	8.7	40
225	Brooklyn	Fulton Ferry	593c0d2262420b7feccc3048	Cecconi's	339	8.8	73
226	Brooklyn	Dumbo	593c0d2262420b7feccc3048	Cecconi's	339	8.8	73
227	Staten Island	Prince's Bay	0	0	0	0.0	0
228	Staten Island	Lighthouse Hill	5079ccdce4b065f1dd8155ad	Casa verde	8	7.8	3

4

From the table above we can see that there is a correlation between the number of Likes and the Rating – the Restaurants with more Likes also have higher Rating. In regards of that information I decided continuing ahead to concentrate solely on the Ratings.

Now that we have that information let's see which are the Boroughs and Neighborhoods with the highest Average Rating using the **.groupby** function:

	Borough	Average Rating
2	Manhattan	8.312371
1	Brooklyn	7.445455
0	Bronx	7.377419
3	Queens	6.569565
4	Staten Island	4.661765

	Neighborhood	Average Rating
13	Bushwick	9.200
11	Boerum Hill	9.100
45	Hamilton Heights	9.100
87	Tribeca	9.100
8	Bedford Stuyvesant	9.000
36	Fort Greene	8.800
21	Clinton Hill	8.800
38	Fulton Ferry	8.800
28	Dumbo	8.800
44	Greenwich Village	8.725

Average rating of Italian Restaurants for each Borough
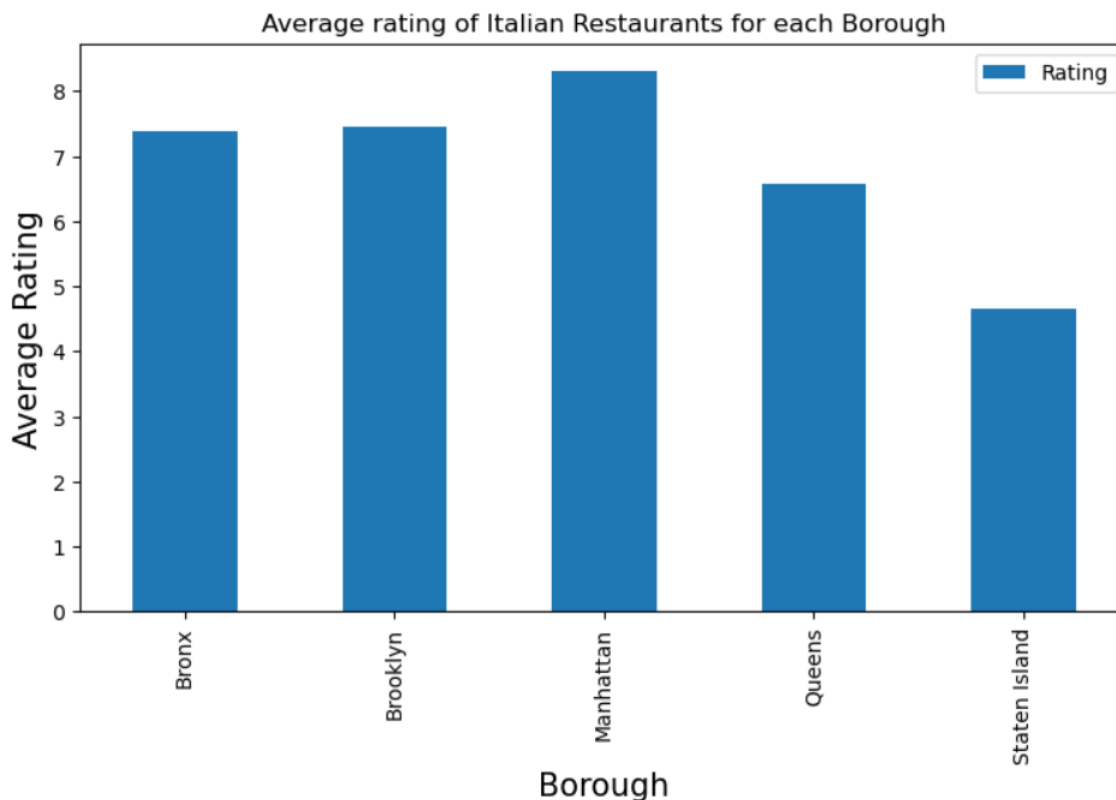
The Borough with the highest Average Rating is Manhattan. The Neighborhoods of Tribeca (Manhattan), Hamilton Heights (Manhattan)and Bushwick (Brooklyn) are some of the best neighborhoods for Italian cuisine with Average Rating of 9.1 and 9.2.

We will **merge** this dataset to the original **new_york_data** dataset to combine latitude and longitude data with the Average Rating by Neighborhood data to create the table containing only the Restaurants with Average Rating >=8.5 which we will after that use for our **Folium Map**.

	Borough	Neighborhood	Latitude	Longitude	Average Rating
0	Brooklyn	Bedford Stuyvesant	40.687232	-73.941785	9.000
1	Brooklyn	Boerum Hill	40.685683	-73.983748	9.100
2	Brooklyn	Bushwick	40.698116	-73.925258	9.200
3	Brooklyn	Clinton Hill	40.693229	-73.967843	8.800
4	Queens	Corona	40.742382	-73.856825	8.700
5	Brooklyn	Dumbo	40.703176	-73.988753	8.800
6	Manhattan	Flatiron	40.739673	-73.990947	8.580
7	Brooklyn	Fort Greene	40.688527	-73.972906	8.800
8	Brooklyn	Fulton Ferry	40.703281	-73.995508	8.800
9	Manhattan	Greenwich Village	40.726933	-73.999914	8.725
10	Manhattan	Hamilton Heights	40.823604	-73.949688	9.100
11	Manhattan	Little Italy	40.719324	-73.997305	8.600
12	Manhattan	Midtown South	40.748510	-73.988713	8.500
13	Bronx	Morris Park	40.847549	-73.850402	8.500
14	Bronx	Pelham Gardens	40.862966	-73.841612	8.600
15	Queens	Rego Park	40.728974	-73.857827	8.600
16	Manhattan	Soho	40.722184	-74.000657	8.725
17	Manhattan	Tribeca	40.721522	-74.010683	9.100
18	Manhattan	Upper West Side	40.787658	-73.977059	8.650
19	Manhattan	Washington Heights	40.851903	-73.936900	8.500
20	Manhattan	West Village	40.734434	-74.006180	8.650
21	Bronx	Woodlawn	40.898273	-73.867315	8.700

On the Map below we can see (especially if we zoom in) that the Italian Restaurants with Average Rating higher than 8.5 and mostly concentrate in the Neighborhoods of Tribeca, Greenwich Village and West Village:

D. Results / Conclusion

As I mentioned before, New York City has long been considered as one of the Food Capitals of the World and it is also the place. NYC is a major metropolitan area with more than 8 million people living within city limits so it is no surprise that we are working with are big set of data. To make the data more manageable I utilized the Foursquare API to explore the boroughs and segment them. I designed the limit as 100 venues and the radius 400 metesr for each borough from their given latitude and longitude information.

I performed data analysis extracting and using the New York Dataset and FourSquare API requests by adding the coordinates, calculating Average Ratings for all Italian Restaurants by Borough and specific Neighborhood and ended the study by visualizing the data on the New York City Folium map. The Average Rating Table and The Map gives us a clear picture where is to open our Italian Restaurant and be relatively sure of its success.

Here are the conclusions that I manage to extrapolate after reviewing the problem and analyzing the data:

- Manhattan and Brooklyn have the best rated Italian restaurants on average.
- Tribeca (Manhattan), Hamilton Heights (Manhattan)and Bushwick (Brooklyn) are some of the best neighborhoods for Italian cuisine with Average Rating of 9.1 and 9.2. However, from those 3 only Tribeca is located in an area of the city that is popular with both tourists and locals alike the other two Neighborhoods are mostly residential.

- Manhattan is also the Neighborhood with the biggest concentration of Italian restaurants and have booming Italian Cuisine Market
- In conclusion Tribeca (Manhattan) is the best place to open Italian Restaurant.
- If we are just looking for a spot to enjoy good Italian food, we should definitely check out and the Neighborhoods of Hamilton Heights, Bushwick and if my personal favorite Greenwich Village all with Average Rating over 8.5

Limitations of the data

As a final note, allof the above analysis is highly dependent on the adequacy and accuracy of FourSquare data.

Thank you for your attention,

Gergana Topalova