Capstone Project - The Battle of Neighborhoods (Week 2)

1. Introduction

1.1. Background

A friend of mine is specialized in Italian cuisine and he is looking to open a restaurant in New York. He is not sure yet about the exact location and for this he asked me to help him. My friend currently lives in Brooklyn but he likes very much Manhattan.

Manhattan is the most densely populated of New York City's 5 boroughs. It's mostly made up of Manhattan Island, bounded by the Hudson, East and Harlem rivers. Among the world's major commercial, financial and cultural centers, it's the heart of "the Big Apple." Its iconic sites include skyscrapers such as the Empire State Building, neon-lit Times Square and the theaters of Broadway. — Google

1.2. Problem description

In order to find the most suitable location for the restaurant, we need to have a clear picture of all restaurants in Manhattan. We have to focus in the areas where most of Italian restaurants are located.

1.3 Interested Audience

This project may be relevant for other people who consider opening a restaurant or coffee shop in New Your or other areas. This logic can be applied to other venues like Gyms, barbershops, Boutiques, dessert shops, etc.

2. Data description

2.1 Data description

First, we need to retrieve the data for all neighborhoods located in our target area, Manhattan. After we retrieve the needed data, we will use the Foursquare API to identify the restaurants in the target area.

2.2 Required Data to resolve the problem

I will use the New York data in a json format in order to retrieve the coordinated of all neighborhoods in Manhattan. Data source: https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json

Using the Foursquare API, I retrieved the venues from all Manhattan neighborhoods.

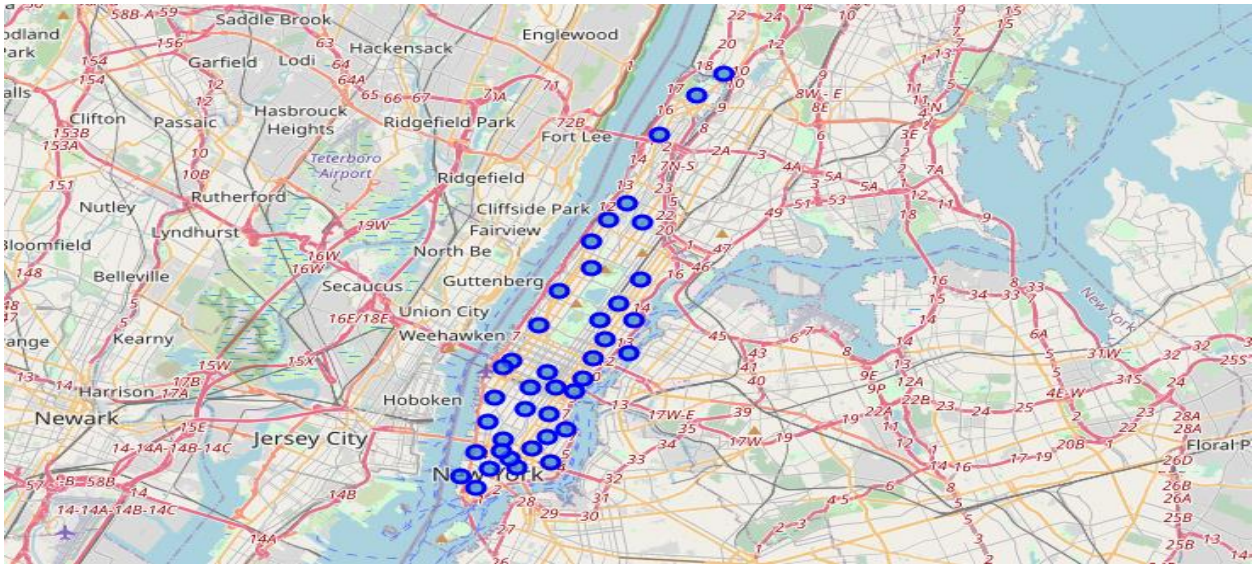
3. Methodology

The objective of this project is to identify Neighborhoods where Italian Restaruants are 1st most common venues.

In the first step, we have collected the New York data that includes Borough, Neighborhood, Latitude and Longitude.

	Borough	Neighborhood	Latitude	Longitude
6	Manhattan	Marble Hill	40.876551	-73.910660
100	Manhattan	Chinatown	40.715618	-73.994279
101	Manhattan	Washington Heights	40.851903	-73.936900
102	Manhattan	Inwood	40.867684	-73.921210
103	Manhattan	Hamilton Heights	40.823604	-73.949688

In the second step, we used the Foursquare API to identify the venues in Manhattan. We will focus only on Restaurants located in Manhattan. I used the coordinated from the above date to generate the map below:



Using the Foursquare API we get the data below:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
8	Marble Hill	40.876551	-73.910660	Land & Sea Restaurant	40.877885	-73.905873	Seafood Restaurant
17	Marble Hill	40.876551	-73.910660	Boston Market	40.877430	-73.905412	American Restaurant
28	Chinatown	40.715618	-73.994279	Kiki's	40.714476	-73.992036	Greek Restaurant
30	Chinatown	40.715618	-73.994279	Spicy Village	40.717010	-73.993530	Chinese Restaurant
32	Chinatown	40.715618	-73.994279	Wah Fung Number 1 Fast Food 華豐快餐店	40.717278	-73.994177	Chinese Restaurant
34	Chinatown	40.715618	-73.994279	Da Yu Hot Pot 大渝火锅	40.716735	-73.995752	Hotpot Restaurant
39	Chinatown	40.715618	-73.994279	Xi'an Famous Foods	40.715232	-73.997263	Chinese Restaurant
43	Chinatown	40.715618	-73.994279	Forgtmenot	40.714459	-73.991546	New American Restaurant
44	Chinatown	40.715618	-73.994279	Dimes	40.714830	-73.991719	American Restaurant
50	Chinatown	40.715618	-73.994279	Cervo's	40.714763	-73.991455	Spanish Restaurant
56	Chinatown	40.715618	-73.994279	Wayla	40.718291	-73.992584	Thai Restaurant

In the final step, we will cluster all restaurant locations using k-means clustering in order to identify the optimal Italian Restaurant location.

4. Analysis

We analyzed each Neighborhood by converting categorical variables into dummy/indicator variables:

	Neighborhood	Afghan Restaurant	African Restaurant	American Restaurant	Arepa Restaurant	Argentinian Restaurant	Asian Restaurant	Australian Restaurant	Austrian Restaurant	Braz Restau
8	Marble Hill	0	0	0	0	0	0	0	0	
17	Marble Hill	0	0	1	0	0	0	0	0	
28	Chinatown	0	0	0	0	0	0	0	0	
30	Chinatown	0	0	0	0	0	0	0	0	
32	Chinatown	0	0	0	0	0	0	0	0	

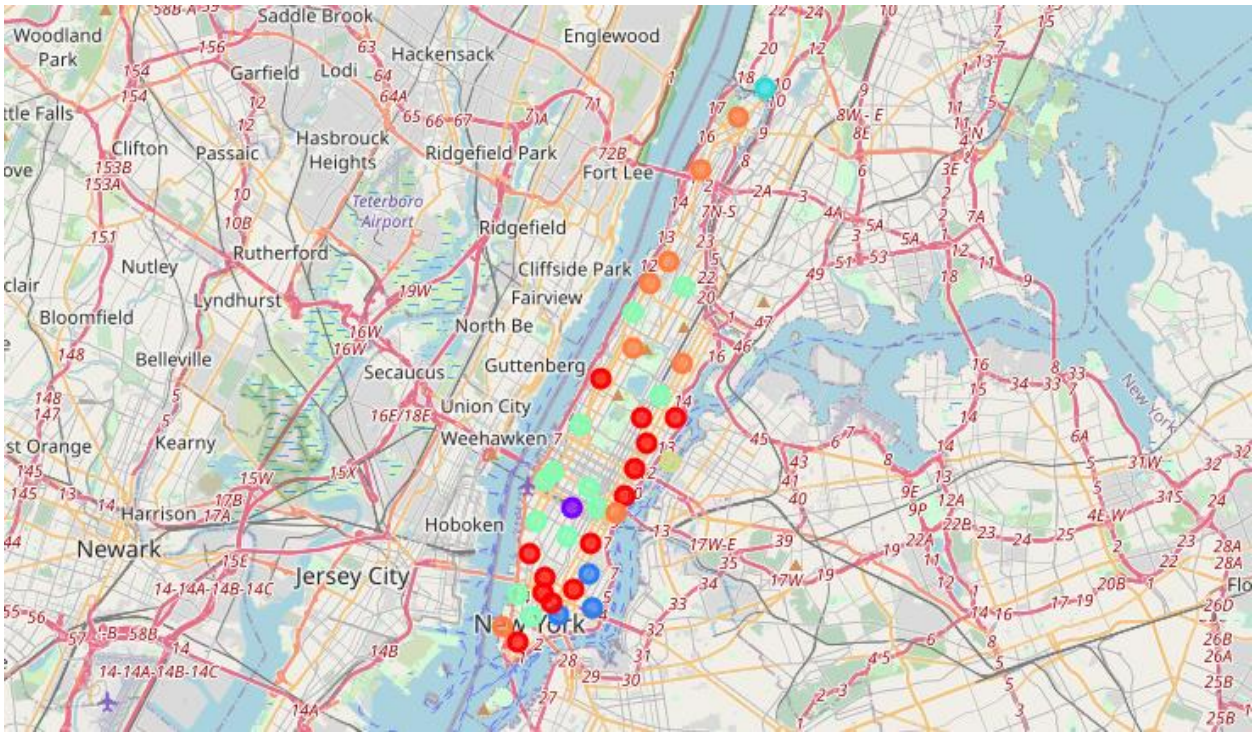
We also grouped rows by neighborhood and by taking the mean of the frequency of occurrence of each category:

	Neighborhood	Afghan Restaurant	African Restaurant	American Restaurant	Arepa Restaurant	Argentinian Restaurant	Asian Restaurant	Australian Restaurant	Austrian Restaurant	Brazilian Restaurant	Cajun / Creole Restaurant
0	Battery Park City	0.0	0.000000	0.000000	0.000000	0.0000	0.000000	0.0	0.000000	0.0	0.0
1	Carnegie Hill	0.0	0.000000	0.062500	0.000000	0.0625	0.000000	0.0	0.000000	0.0	0.0
2	Central Harlem	0.0	0.142857	0.142857	0.000000	0.0000	0.000000	0.0	0.000000	0.0	0.0
3	Chelsea	0.0	0.000000	0.157895	0.052632	0.0000	0.052632	0.0	0.000000	0.0	0.0
4	Chinatown	0.0	0.000000	0.076923	0.000000	0.0000	0.051282	0.0	0.025641	0.0	0.0

Then we created a new dataframe to display the top 10 Restaurants for each neighborhood:

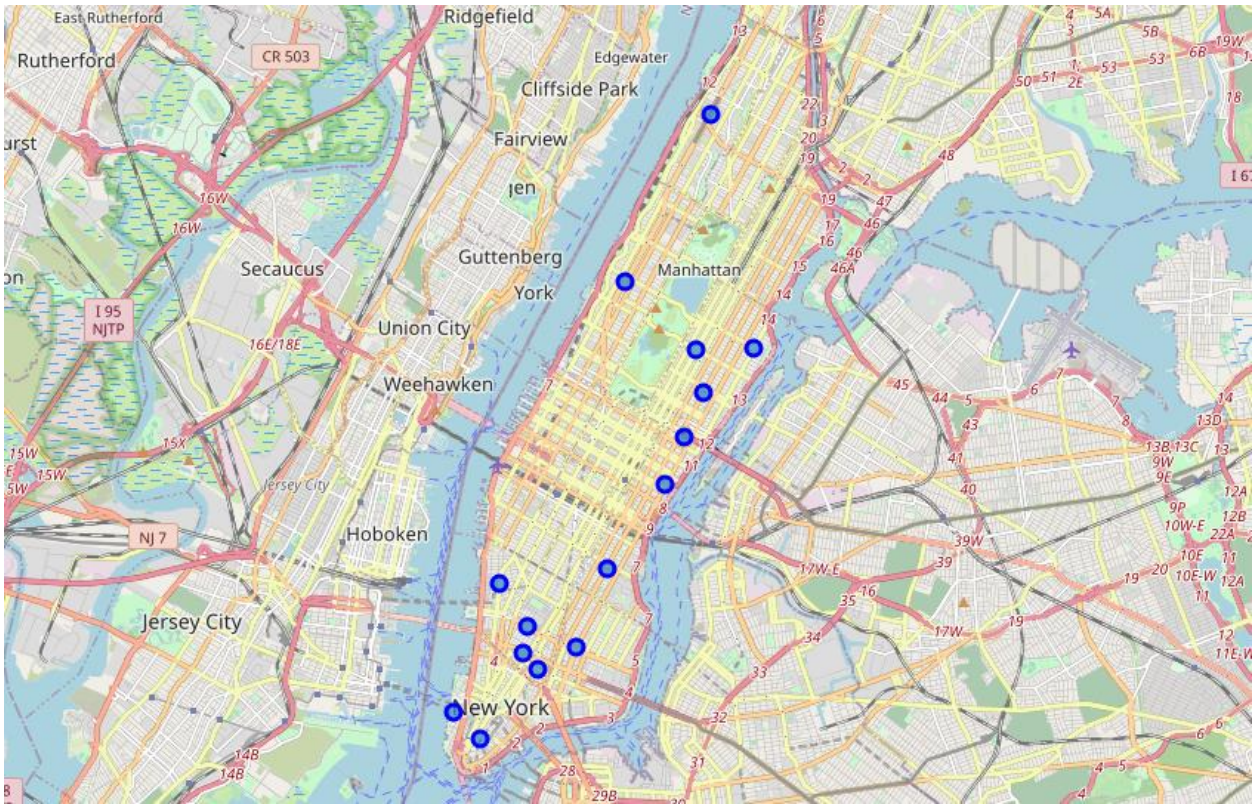
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Battery Park City	Italian Restaurant	Japanese Restaurant	Chinese Restaurant	Mexican Restaurant	Sushi Restaurant
1	Carnegie Hill	French Restaurant	Vietnamese Restaurant	Fast Food Restaurant	Japanese Restaurant	American Restaurant
2	Central Harlem	French Restaurant	African Restaurant	American Restaurant	Seafood Restaurant	Chinese Restaurant
3	Chelsea	American Restaurant	French Restaurant	Italian Restaurant	Seafood Restaurant	Japanese Restaurant
4	Chinatown	Chinese Restaurant	Vietnamese Restaurant	American Restaurant	Shanghai Restaurant	Greek Restaurant

We clustered the Neighborhoods using k-means clustering and we plotted the coordinated to the map:



After Examination, we concluded that cluster # 0 (red dot on the map) should be the recommended area as in this cluster Italian Restaurants are mostly first most common venues.

We focus only on the area where Italian Restaurants are first most common venue and using the coordinates, we generate a new map with the recommended Neighborhoods:



5. Results and Discussion

We were able to identify and cluster in seven groups the location of Restaurants in Manhattan. The results from our analysis shows that Italian Restaurants are first most common venues in many Neighborhoods in Manhattan.
After the clustering, we focused only on Italian Restaurants that are first most common venues and we noticed that they are located in “Lower Manhattan” and near to the “East River” area. The areas mentioned above will be considered as a starting point for a more detailed analysis like competition, available properties and renting price.

.

6. Conclusion

The purpose of this project was to help e friend to find Neighborhoods in Manhattan where Italian Restaurants are very popular and people living and working there like Italian cuisine. By using Foursquare API we were able to identify and to sort Restaurants in Manhattan and their respective Neighborhood. We clustered the locations in seven groups in order to identify the

areas where Italian Restaurants are first most common venues. We concluded that most prominent Italian Restaurants are located in Neighborhoods like Little Italy, Financial District, Noho, Greenwich Village, etc.
My friend will make the final decision after he further analyses other characteristics of the above Neighborhoods.