

Battle of Manhattan Neighborhoods

This project is final step towards completing IBM Data Science Professional Certificate set of courses.

Introduction/Business Problem

New York is metropolitan area with over 8 millions residents and with over 65 millions visitors per year. It is a great place for restaurant business. As expected, New York is one of the greatest food cities in the world. There is such diversity of cuisines, like no where else. If you can imagine of food, New York probably has it.

John is a businessman and he wants to invest in a fine restaurant in Manhattan. He wants an analyse of neighborhoods and restaurants in this prosperous spot. The main goal is to make suggestion of the place and the type of a restaurant with the highest chance of success.

Data

For solving this problem we will need data from this URL: https://geo.nyu.edu/catalog/nyu_2451_34572. This data contain names of neighborhoods in New York and its longitudes and latitudes.

During inquiry, we will use data obtained via FourSquare API.

The search request in Foursquare API returns a list of venues near the current location in a JSON file, which will be converted into the dataframe.

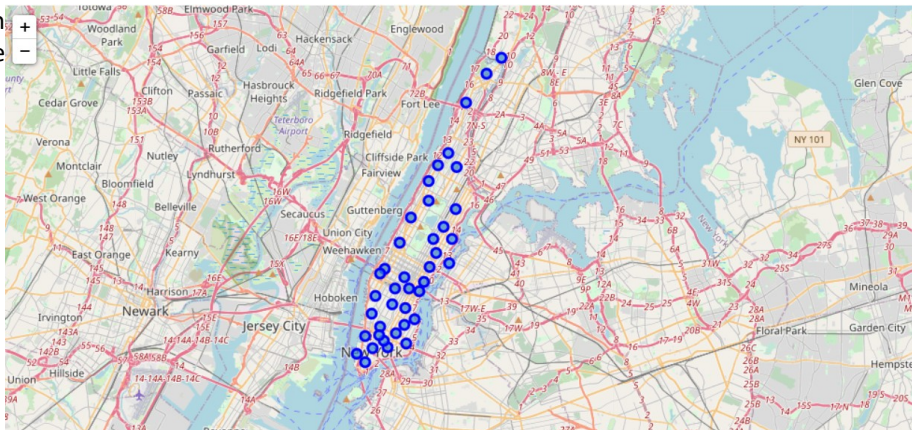
Analyzing the data, we will find Top 10 desirable locations for restaurant opening and, by clustering neighborhoods on similarity in restaurants we will recommend both the location and the type of restaurant with the highest chance of success - finding the neighborhood in a cluster which is similar to others, but having differences in distribution of restaurant types.

Methodology

Firstly, we download the New York data and convert it to dataframe. The second step is to select Manhattan Neighborhoods.

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Manhattan	Chinatown	40.715618	-73.994279
2	Manhattan	Washington Heights	40.851903	-73.936900
3	Manhattan	Inwood	40.867684	-73.921210
4	Manhattan	Hamilton Heights	40.823604	-73.949688

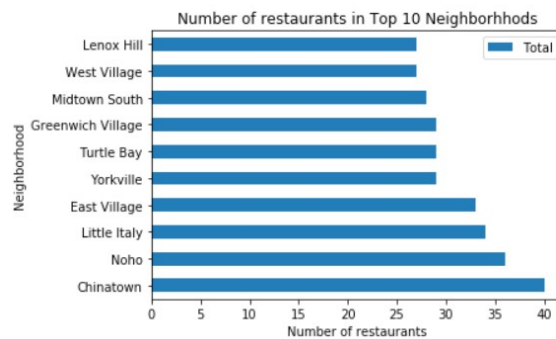
Then, we create map of Manhattan in folium using latitude and longitude values.



We obtain data via FourSquare API with venues in Manhattan and form a dataframe with restaurant venues only grouped by Neighborhood.

	Neighborhood	Afghan Restaurant	African Restaurant	American Restaurant	Arepa Restaurant	Argentinian Restaurant	Asian Restaurant	Australian Restaurant	Austrian Restaurant	Brazilian Restaurant	Cajun / Creole Restaurant	Cambodian Restaurant	Canto Restaurant
0	Battery Park City	0	0	0	0	0	0	0	0	0	0	0	0
1	Carnegie Hill	0	0	1	0	1	0	0	0	0	0	0	0
2	Central Harlem	0	3	2	0	0	0	0	0	0	0	0	0
3	Chelsea	0	0	3	0	0	0	0	0	0	0	0	0
4	Chinatown	0	0	3	0	0	2	1	1	0	0	0	0

We add new 'Total' column with the total number of restaurants in each neighborhood, sort the data by that column and visualize the Top 10 neighborhoods with the highest number of restaurants.



We form a list of popular restaurant types. Firstly, we add a row 'Total_type' calculating the sums of each column. Popular restaurant type is considered the type having total count of restaurant type greater or equal to the average number of restaurants by type.

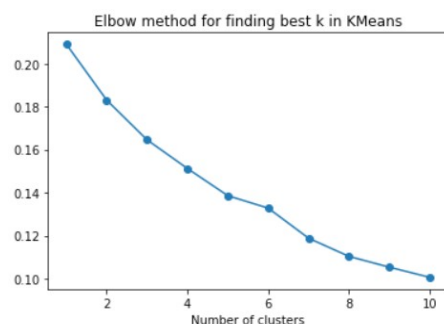
Popular types of restaurants are:

['American Restaurant', 'Chinese Restaurant', 'French Restaurant', 'Italian Restaurant', 'Japanese Restaurant', 'Mediterranean Restaurant', 'Mexican Restaurant', 'Restaurant', 'Seafood Restaurant', 'Sushi Restaurant', 'Thai Restaurant']

We form dataframe with the Neighborhoods and its 5 the most common venues.

	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	Chinese Restaurant	Mediterranean Restaurant	Mexican Restaurant	Vietnamese Restaurant	French Restaurant
1	Carnegie Hill	Japanese Restaurant	Vietnamese Restaurant	French Restaurant	Italian Restaurant	Thai Restaurant
2	Central Harlem	African Restaurant	French Restaurant	American Restaurant	Chinese Restaurant	Seafood Restaurant
3	Chelsea	American Restaurant	Italian Restaurant	Seafood Restaurant	Restaurant	Ramen Restaurant
4	Chinatown	Chinese Restaurant	Vietnamese Restaurant	American Restaurant	Mexican Restaurant	Malay Restaurant

The neighborhoods needs to be clustered based on there similarity/dissimilarity of restaurants venues. Firstly, we find best k for K - Means algorithm and use Elbow method.

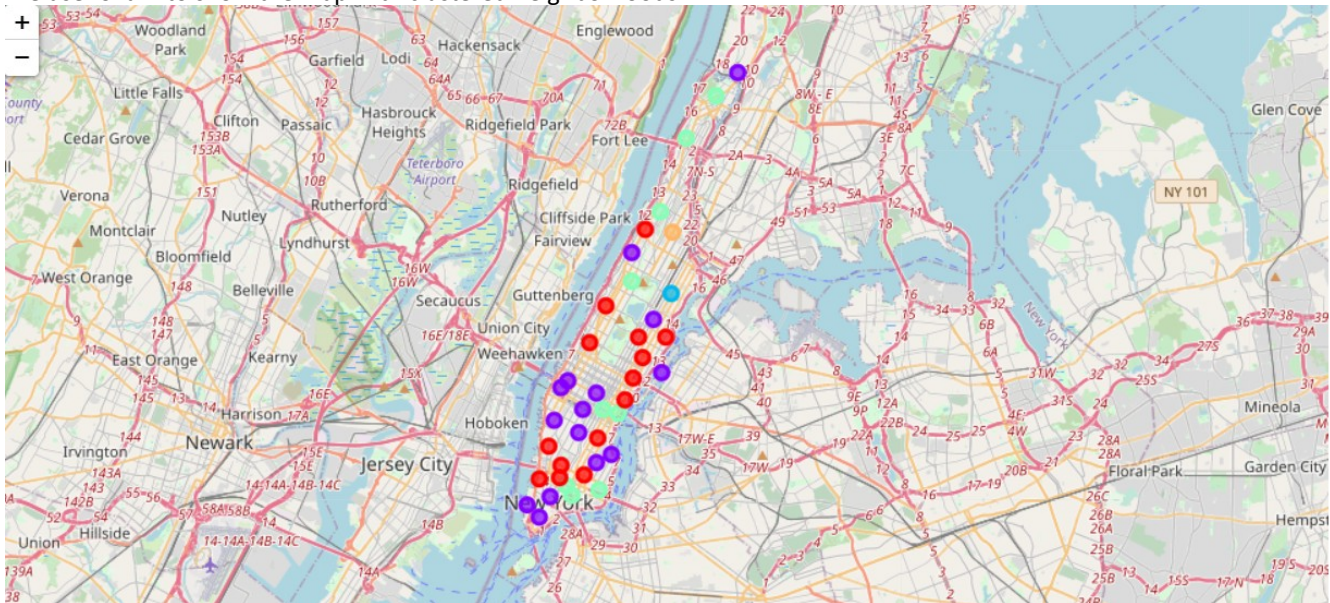


After clustering the neighborhoods having number of clusters $k = 5$, we have dataframe with clustered neighborhoods and Cluster Labels.

	Cluster Labels	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Latitude	Longitude
0	1	Battery Park City	Chinese Restaurant	Mediterranean Restaurant	Mexican Restaurant	Vietnamese Restaurant	French Restaurant	40.711932	-74.016869
1	1	Carnegie Hill	Japanese Restaurant	Vietnamese Restaurant	French Restaurant	Italian Restaurant	Thai Restaurant	40.782683	-73.953256
2	4	Central Harlem	African Restaurant	French Restaurant	American Restaurant	Chinese Restaurant	Seafood Restaurant	40.815976	-73.943211
3	1	Chelsea	American Restaurant	Italian Restaurant	Seafood Restaurant	Restaurant	Ramen Restaurant	40.744035	-74.003116
4	3	Chinatown	Chinese Restaurant	Vietnamese Restaurant	American Restaurant	Mexican Restaurant	Malay Restaurant	40.715618	-73.994279

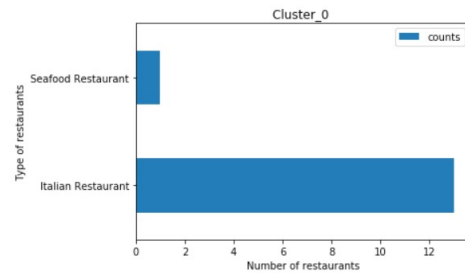
Results

We use folium to show the map with clustered neighborhoods.



- Cluster 0 and the plot with number of the 1st Most Common Venues are shown below:

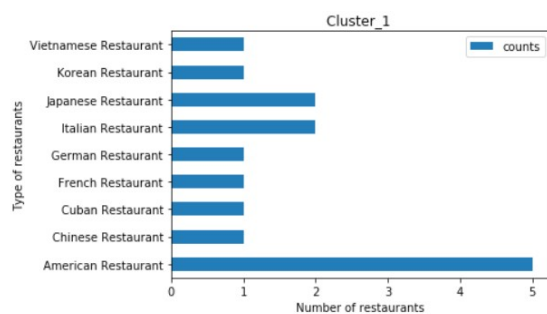
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
11	Gramercy	Italian Restaurant	Mexican Restaurant	American Restaurant	Thai Restaurant	Sushi Restaurant
12	Greenwich Village	Italian Restaurant	Vietnamese Restaurant	Sushi Restaurant	Chinese Restaurant	Indian Restaurant
16	Lenox Hill	Italian Restaurant	Sushi Restaurant	Turkish Restaurant	Thai Restaurant	Mexican Restaurant
17	Lincoln Square	Italian Restaurant	American Restaurant	Mediterranean Restaurant	French Restaurant	Seafood Restaurant
21	Manhattanville	Seafood Restaurant	Italian Restaurant	Chinese Restaurant	Mexican Restaurant	Japanese Curry Restaurant
27	Noho	Italian Restaurant	Japanese Restaurant	Mexican Restaurant	Vietnamese Restaurant	Seafood Restaurant
29	Soho	Italian Restaurant	Mediterranean Restaurant	French Restaurant	Sushi Restaurant	Falafel Restaurant
31	Sutton Place	Italian Restaurant	Restaurant	Mexican Restaurant	Vegetarian / Vegan Restaurant	Persian Restaurant
32	Tribeca	Italian Restaurant	American Restaurant	Greek Restaurant	French Restaurant	Korean Restaurant
34	Turtle Bay	Italian Restaurant	French Restaurant	Seafood Restaurant	Sushi Restaurant	Turkish Restaurant
35	Upper East Side	Italian Restaurant	Sushi Restaurant	Mexican Restaurant	French Restaurant	Thai Restaurant
36	Upper West Side	Italian Restaurant	Seafood Restaurant	Thai Restaurant	American Restaurant	Vegetarian / Vegan Restaurant
38	West Village	Italian Restaurant	American Restaurant	New American Restaurant	French Restaurant	Seafood Restaurant
39	Yorkville	Italian Restaurant	Sushi Restaurant	Japanese Restaurant	Mexican Restaurant	Chinese Restaurant



As we can see, there are mainly Italian restaurants in these Neighborhoods.

- Cluster 1 and the plot with number of the 1st Most Common Venues:

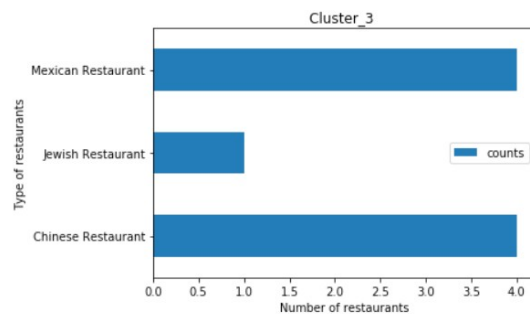
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Battery Park City	Chinese Restaurant	Mediterranean Restaurant	Mexican Restaurant	Vietnamese Restaurant	French Restaurant
1	Carnegie Hill	Japanese Restaurant	Vietnamese Restaurant	French Restaurant	Italian Restaurant	Thai Restaurant
3	Chelsea	American Restaurant	Italian Restaurant	Seafood Restaurant	Restaurant	Ramen Restaurant
5	Civic Center	French Restaurant	American Restaurant	Sushi Restaurant	Vietnamese Restaurant	Cuban Restaurant
6	Clinton	American Restaurant	Italian Restaurant	New American Restaurant	Thai Restaurant	Brazilian Restaurant
8	East Village	Vietnamese Restaurant	Japanese Restaurant	Mexican Restaurant	Ramen Restaurant	Seafood Restaurant
9	Financial District	American Restaurant	Falafel Restaurant	Japanese Restaurant	Restaurant	French Restaurant
10	Flatiron	Italian Restaurant	American Restaurant	Japanese Restaurant	Vegetarian / Vegan Restaurant	Mediterranean Restaurant
14	Hudson Yards	Italian Restaurant	American Restaurant	Restaurant	Caucasian Restaurant	Peruvian Restaurant
22	Marble Hill	American Restaurant	Seafood Restaurant	Vietnamese Restaurant	Filipino Restaurant	Egyptian Restaurant
23	Midtown	Cuban Restaurant	Japanese Restaurant	Mediterranean Restaurant	South American Restaurant	Vegetarian / Vegan Restaurant
24	Midtown South	Korean Restaurant	Japanese Restaurant	American Restaurant	New American Restaurant	Italian Restaurant
25	Morningside Heights	American Restaurant	Mexican Restaurant	Indian Restaurant	Seafood Restaurant	Greek Restaurant
28	Roosevelt Island	Japanese Restaurant	Greek Restaurant	French Restaurant	Egyptian Restaurant	Empanada Restaurant
30	Stuyvesant Town	German Restaurant	French Restaurant	Egyptian Restaurant	Empanada Restaurant	English Restaurant



It is the cluster with miscellaneous restaurants.

- Cluster 3 and the plot with number of the 1st Most Common Venues:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
4	Chinatown	Chinese Restaurant	Vietnamese Restaurant	American Restaurant	Mexican Restaurant	Malay Restaurant
13	Hamilton Heights	Mexican Restaurant	Caribbean Restaurant	Indian Restaurant	Sushi Restaurant	Chinese Restaurant
15	Inwood	Mexican Restaurant	Restaurant	Chinese Restaurant	Caribbean Restaurant	Spanish Restaurant
18	Little Italy	Chinese Restaurant	Mediterranean Restaurant	Italian Restaurant	Thai Restaurant	Seafood Restaurant
19	Lower East Side	Chinese Restaurant	Vietnamese Restaurant	French Restaurant	Italian Restaurant	American Restaurant
20	Manhattan Valley	Mexican Restaurant	Vietnamese Restaurant	Chinese Restaurant	Japanese Restaurant	American Restaurant
26	Murray Hill	Jewish Restaurant	Japanese Restaurant	Mediterranean Restaurant	Indian Restaurant	Cuban Restaurant
33	Tudor City	Mexican Restaurant	Vietnamese Restaurant	Asian Restaurant	Thai Restaurant	Greek Restaurant
37	Washington Heights	Chinese Restaurant	Spanish Restaurant	Latin American Restaurant	Mexican Restaurant	Tapas Restaurant



It is Chinese - Mexican cluster.

Discussion

The neighborhoods are clustered in 5 clusters based on their similarity/dissimilarity of restaurant venues. The best investment is in the best spots, so further discussion will be made for the Top 10 Neighborhoods by the number of restaurants.

In cluster 0, there are 5 neighborhoods from the Top 10 list of neighborhoods having the largest number of restaurants in Manhattan. These are: Lenox Hill, Greenwich Village, Turtle Bay, West Village, Yorkville and Noho. They are the great locations for opening Italian restaurants, because Italian restaurant is the most frequent type in these areas. On the second thought, these areas are recommended for every single type of popular restaurants, determined prior in the notebook. There is great diversity of restaurant types, and these are areas known for a big number of restaurants, so it could be successful. Manhattanville could be a good choice for opening an Italian Restaurant. It is in Cluster 0, so similar to other neighborhoods in that cluster, but not having Italian Restaurant as a dominant type.

In cluster 1, there are 2 neighborhoods from the Top 10 list. These are: Midtown South and East Village. It is the cluster with miscellaneous restaurants, so the recommendation is to open a restaurant of a popular type of your choice.

In cluster 3, there are 2 neighborhoods from the Top 10 list. These are Chinatown and Little Italy. Dominant types of restaurants are Chinese and Mexican, so it is recommended to open a Chinese or a Mexican restaurant in these neighborhoods.

In clusters 3 and 4 there are no neighborhoods from Top 10, so we will not discuss them.

Conclusion

After the analysis, we have a deeper understanding of restaurant distribution in Manhattan, New York. However, it is not a simple task to make the final recommendation of the best location and the type of restaurant to open. Further analysis is necessary to make the final decision, but we've suggested some ideas!