Analysis of the Neighborhoods in Manhattan

Ting Li
xjliting@gmail.com
2019-4-16

Introduction

Manhattan is the most densely populated of the five [boroughs](#) of [New York City](#) and its economic and administrative center, [cultural](#) identifier, and historical birthplace. It has been described as the cultural, financial, [media](#), and [entertainment](#) capital of the world, and it is one of the [highest-income places](#) in the United States with a population greater than one million.

Our client Lily Smith is travelling to Manhattan the next month. She will spend about 2 weeks there. This is her first trip to Manhattan and She really wants to enjoy her stay and spend an easy and relaxed two weeks there.

Lily mentioned that she is a big fan of nice food, and she likes to try different kinds of restaurants, so one of the basics for this trip is to enjoy the food and restaurants in the area she stays. Another key point is the hotel, Lily also would like to find a nice hotel in the neighborhood. Finally, the challenge part is that Lily is very interested in Anthropology, she is happy to see that different neighborhoods will have different lifestyle. She wants to spend the two weeks in two different hotels, so that she could really experience different lifestyles of the nearby neighborhoods in the two weeks.

Lily finally comes to us, hoping that with some location analysis we can help her find two neighborhoods including the hotels to stay so that she could achieve all her goals during the trip.

Methodology

Data

1. Manhattan neighborhood data (https://geo.nyu.edu/catalog/nyu_2451_34572)

   The New York Neighborhood has a total of 5 boroughs and 306 neighborhoods. For research purpose, we will only look at the Borough of Manhattan as the Manhattan neighborhood data.

   There are 40 neighborhoods in Manhattan in total. The data contains the name and latitude, longitude of each neighborhood in Manhattan. The distribution of the neighborhoods is shown below:
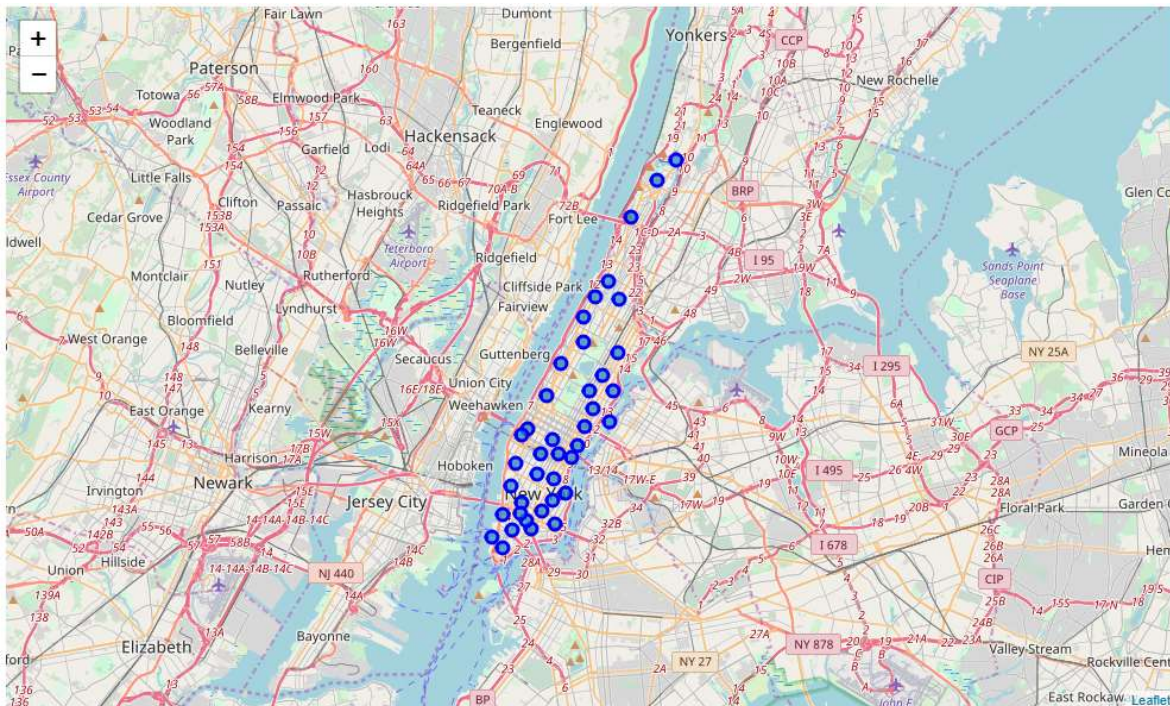
2. Location data from Foursquare API

Foursquare is a social location service that allows users to explore the world around them. By connecting to the Foursquare API, users/developers can access the Foursquare database and get the information they need for various location data. In this study, we utilize the Foursquare API and get the venue information for each neighborhood in Manhattan.

K-means clustering

K-means clustering is one of the simplest and popular unsupervised machine learning algorithms. A cluster refers to a collection of data points aggregated together because of certain similarities. The methodology identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. The K-means algorithm tries to minimize the intra-cluster distances and maximize the inter-cluster distances, so that the items within the cluster is very similar to each other and the items from different clusters will be very different.

Data process

With the Manhattan neighborhood data, and the Foursquare API, we are able to find all the venues (limit 100 per neighborhood) for each neighborhood. At the meantime, since we are interested in the restaurants and hotels, we also find the number of restaurants and hotels for each neighborhood.

Starting from the client's needs, we first filter out the neighborhoods that don't have a hotel—eventually, our client is looking for hotels. After this step, there are 21 neighborhoods in total for us to choose from. Based on the venues in each neighborhood, we choose the top 10 common venue categories in each neighborhood and group them into 6 clusters. We mark the clusters on the map with different colors as below:
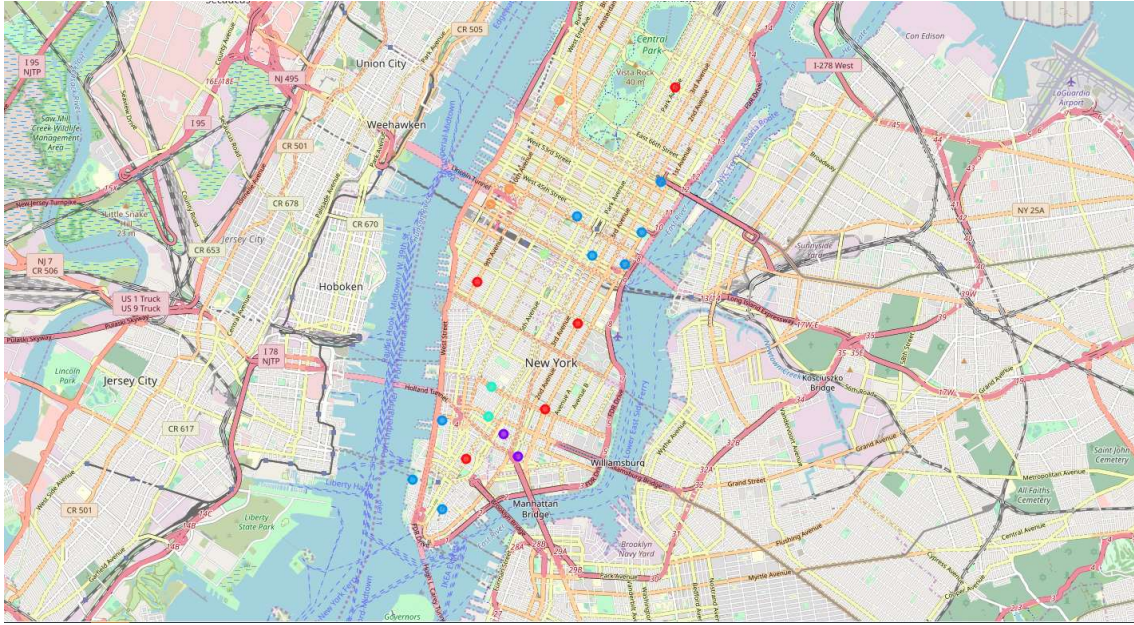
The average value for each cluster is listed below. The table shows that the cluster 2 has 8 neighborhoods, which is the most, and cluster 4 has the least-only 1 neighborhood. On average, cluster 5 has the least restaurants. And cluster 1 has 1 hotel per neighborhood on average.

Cluster Labels	Neighborhood
0	5
1	2
2	8
3	2
4	1
5	3

Cluster Labels	Latitude	Longitude	restaurant_count	hotel_count
0	40.739074	-73.987770	25.800000	2.400000
1	40.717471	-73.995792	35.500000	1.000000
2	40.737849	-73.987588	26.500000	4.000000
3	40.724558	-74.000285	32.000000	1.500000
4	40.748510	-73.988713	30.000000	6.000000
5	40.763096	-73.993856	18.333333	3.333333

The average number of restaurants in all the neighborhoods in Manhattan is 26.7. Since our client would like a neighborhood with many restaurants, we will proceed to limit our options to cluster 1, 3 and 4, due to their larger number of restaurants in their neighborhood. From the table above, we can see that there are 2,2 and 1 neighborhoods within these 3 clusters respectively.
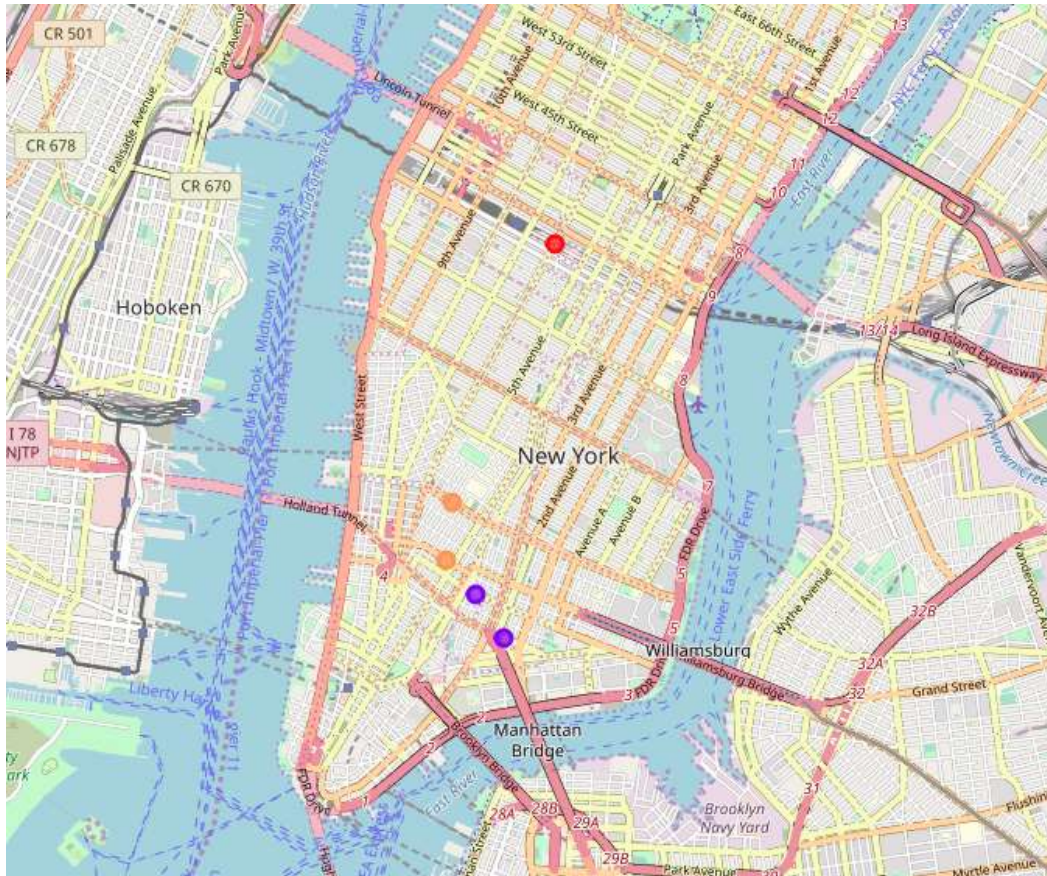
After looking into the cluster detail, we noticed that the neighborhood of "Soho" has 19 restaurants, so we remove it from our option list because it's below the average. Now we have 4 neighborhoods from 3 clusters to choose from.

Now these neighborhoods all satisfy the restaurant checkpoint, it's time that we add in the "nice hotel" standard. Again, by applying the Foursquare API, we can find the hotels in these 4 neighborhoods, and their ratings as well. There are 9 hotels in these neighborhoods and their rating ranges from

Sorting by the rating, the hotels in these neighborhoods are listed in the table below.

Based on the results, we will choose the neighborhood 'Midtown South' with hotel 'The NoMad Hotel' (rating: 9.5) and neighborhood 'Chinatown' with hotel 'Hotel 50 Bowery NYC' (rating: 8.9) to recommend to our client for the trip.

Neighborhood	Borough	Cluster Labels	restaurant_count	Hotel	Hotel rating
Midtown South	Manhattan	4	31	The NoMad Hotel	9.5
Midtown South	Manhattan	4	31	Kimpton Hotel Eventi	9.1
Midtown South	Manhattan	4	31	Ace Hotel New York	9.0
Chinatown	Manhattan	3	43	Hotel 50 Bowery NYC	8.9
Battery Park City	Manhattan	1	9	Conrad New York	8.7
Midtown South	Manhattan	4	31	MADE Hotel	8.5
Financial District	Manhattan	1	22	AKA Wall Street	8.3
Financial District	Manhattan	1	22	Andaz Wall Street	8.3
Financial District	Manhattan	1	22	W New York - Downtown	8.3
Battery Park City	Manhattan	1	9	W New York - Downtown	8.3
Midtown South	Manhattan	4	31	Courtyard by Marriott New York Manhattan/Heral...	8.3
Midtown South	Manhattan	4	31	Hyatt House New York/Chelsea	8.2
Financial District	Manhattan	1	22	Q&A Residential Hotel	8.1
Financial District	Manhattan	1	22	Residence Inn by Marriott New York Downtown Ma...	8.1
Battery Park City	Manhattan	1	9	Holiday Inn Manhattan-Financial District	7.8
Battery Park City	Manhattan	1	9	New York Marriott Downtown	7.8
Battery Park City	Manhattan	1	9	Courtyard by Marriott New York Downtown Manhat...	7.6

Discussion

This study is only for preliminary research purpose, further analysis can be done in the future to get better suggestions.

1. In our study, we only considered the number of restaurants in each neighborhood yet didn't take into consideration of the quality/rating of each restaurant. For better client assistance, it'd be better to include the quality information for each restaurant and maybe phone number, hours, tips as well.
2. We only considered the location data in Manhattan, and we believe it would be better if we can consider other perspectives in these neighborhoods (such as demographic data of the people living in the neighborhood) when we do the clustering analysis.

Conclusion

In this project, we analyzed the neighborhood information using the Manhattan neighborhood data and the Foursquare API location data and recommended the two neighborhoods with two hotels for our client Lily to stay for her trip to Manhattan next month. These two neighborhoods satisfied her requirements on the following: nice food and restaurants to choose from and different neighborhoods which can have different lifestyle experiences.

With researching on the top venues in each neighborhood, we group all the neighborhoods into different clusters, so that when we choose neighborhoods from different clusters, the experiences of lifestyle would be different. Finally, we find the hotels with top ratings to recommend to our client.

As the results, 'NoMad Hotel' with a rating of 9.5 in neighborhood 'Midtown South'' and 'Hotel 50 Bowery NYC' with rating of 8.9 in the neighborhood 'Chinatown' will be our recommendation to the client.

Reference

1. Wikipedia, https://en.wikipedia.org/wiki/Manhattan
2. Understanding K-means Clustering in Machine Learning, https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1
3. https://foursquare.com/