Identifying Suitable Locations in Manhattan for Opening up a New Branch of a Shop or Business using K Means Clustering

-Rohan PS. Pissurlencar

22nd July 2021

1. Introduction

Using Clustering to find similar neighborhoods to the current location in order to find the most suitable place for opening up another branch of your business.

2. Background

Imagine you have a shop or any business in the outskirts of New York city (In our case **Little Neck, Queens**) and the business does well in your current neighbourhood due to several geospatial features in proximity such as other business, parks, offices, etc.

Now you wish to open another branch of your business, in Manhattan for instance. But how do you decide which neighborhood inside Manhattan would be most suitable for your business and would ensure that your new branch continues to thrive as much as your current branch? You solve this problem why finding out all neighborhoods in Manhattan that are similar to your current neighborhood.

3. Data

So how does one decide which neighborhood is similar to your current neighborhood? This is where data science comes in.

We will use the New York city JSON data set which is available on the IBM Developer Skills Network (https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json). This dataset contains the different neighborhoods in New York city along with their Latitude and Longitude.

With the help of this datasets latitude and longitude co ordinates we will leverage the Foursquare API to explore each neighborhood and find the most prominent and commonly occurring venues in that neighborhoods vicinity. Once we have these details we will use these as our feature vector in order to fit this data in clustering machine learning algorithm such as K means clustering.

These clustering algorithms will group neighborhoods of similar type based on the feature set (in our case most common venues in the vicinity information) and label them in different clusters.

After this point it becomes a simple problem of identifying the neighborhoods in Manhattan which belong to the same cluster as our current neighborhood. These are Neighborhoods which are most suitable to open up our new branch which will see favorable market conditions similar to your current branch.

4. Methodology

As already stated in the above Data section we will leverage the data from the New York city JSON file which contains the names of the different neighborhoods within New York city and also Latitude and Longitude of each of the Neighborhood. We will use these Geospatial data along with the Foursquare API in order to explore each neighborhood find the most common and prominent type of venues for that Neighborhood. We will use this prominence of venue types for our feature set.

We will then feed this feature set to a clustering machine learning algorithm which will cluster the location together based on the prominence of similar types of venues in the neighborhood vicinity.

4.1 Data Exploratory Analysis

After a having a look at the dataset we can see the following counts for the number of data points which we will use after cleaning up the raw dataset:

Borough = 5

Neighborhood = 306

Therefore, there are 5 Boroughs and a total of 306 Neighborhoods in New York city.

As we have already stated in the Background section, our current Neighborhood lies in Little Neck in Queens. Let us visualize this location along with the other 305 locations on the map of New York which will be plotted as a dot which will be located on the neighborhood latitude and longitude.

(We will implement this using the Folium library)

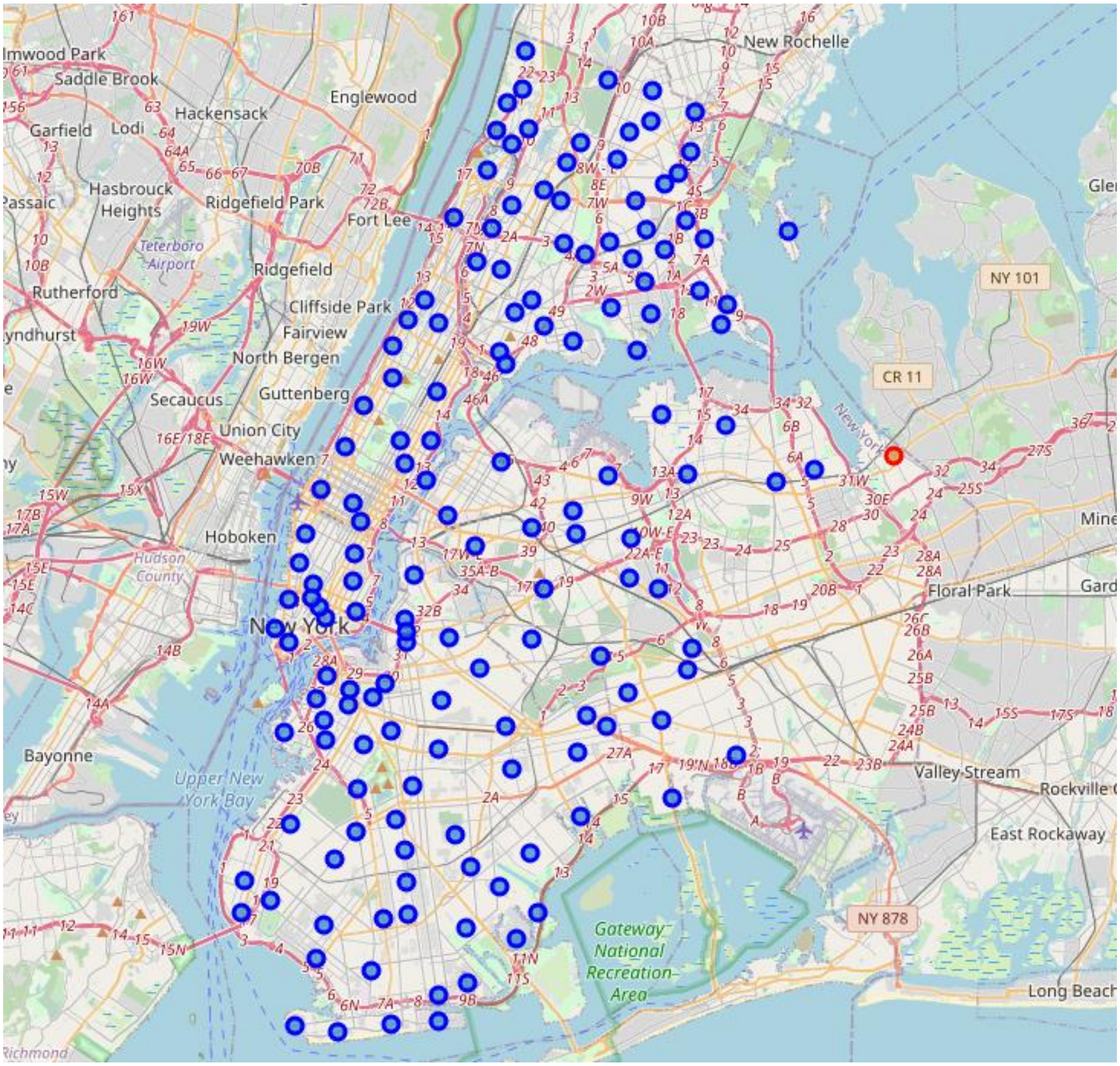

Figure 1: Map of New York city where all the neighborhoods are displayed as dots centered at the neighborhood latitude and longitude. The red dot identifies our Neighborhood of Little Neck while the Blue points identify all the other Neighborhoods inside New York city.

We have already stated that we wish to open our new branch in Manhattan. There we do not need to go through all the possible 305 locations in New York city but rather only through the neighborhoods inside of Manhattan. This will not only be simpler but will also save us precious time and computation resources.

After we filter the dataset for the Manhattan borough. After we do this, we find that there are 40 neighborhoods inside Manhattan. These are the Neighborhoods from which we need to select the most suitable location that is most similar to our current location based on proximity to other venues in the neighborhood (which was one of the reason our current business thrived).

Let us have a look at the map of the all the possible Manhattan neighborhoods:

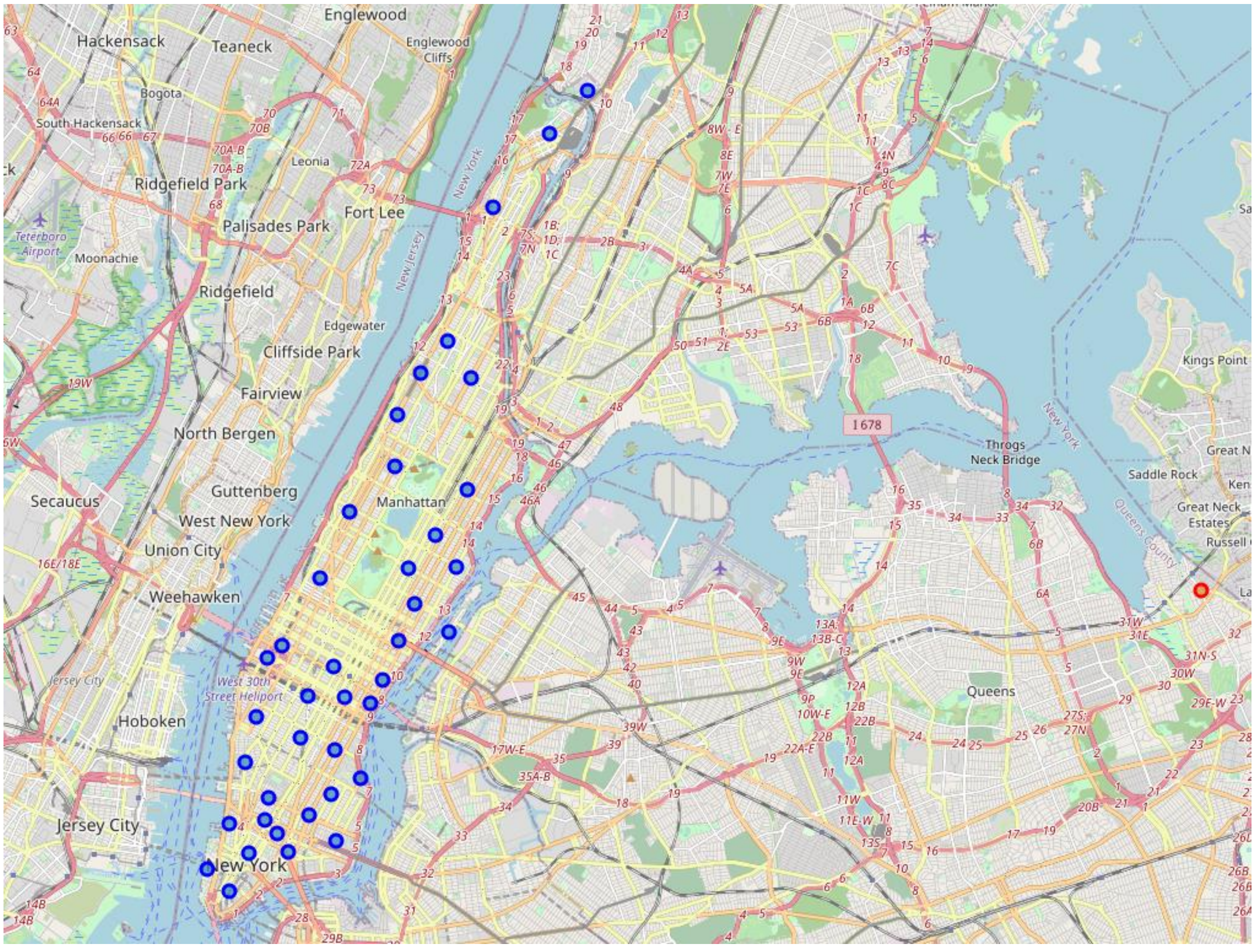

Figure 2: Map of New York city where all the neighborhoods in Manhattan are displayed as dots centered at the neighborhood latitude and longitude. The red dot identifies our Neighborhood of Little Neck while the Blue points identify all the other Neighborhoods inside Manhattan.

Now let us find the most prominent venues in the neighborhoods which will be the basis of the feature set which will be used for grouping the neighborhoods together into clusters which will help us identify the Manhattan neighborhoods which are the most identical to our current Neighborhood of Little Neck.

For the purpose of the feature set construction let us combine the 40 Manhattan neighborhoods and our Neighborhood of Little Neck in one data frame which will be the basis for the feature set construction and use for the clustering algorithm.

The dataframe looks like this :

	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
5	Manhattan	Manhattanville	40.816934	-73.957385
6	Manhattan	Central Harlem	40.815976	-73.943211
7	Manhattan	East Harlem	40.792249	-73.944182
8	Manhattan	Upper East Side	40.775639	-73.960508
9	Manhattan	Yorkville	40.775930	-73.947118
10	Manhattan	Lenox Hill	40.768113	-73.958860
11	Manhattan	Roosevelt Island	40.762160	-73.949168
12	Manhattan	Upper West Side	40.787658	-73.977059
13	Manhattan	Lincoln Square	40.773529	-73.985338
14	Manhattan	Clinton	40.759101	-73.996119
15	Manhattan	Midtown	40.754691	-73.981669
16	Manhattan	Murray Hill	40.748303	-73.978332
17	Manhattan	Chelsea	40.744035	-74.003116
18	Manhattan	Greenwich Village	40.726933	-73.999914
19	Manhattan	East Village	40.727847	-73.982226
20	Manhattan	Lower East Side	40.717807	-73.980890
21	Manhattan	Tribeca	40.721522	-74.010683
22	Manhattan	Little Italy	40.719324	-73.997305
23	Manhattan	Soho	40.722184	-74.000657
24	Manhattan	West Village	40.734434	-74.006180
25	Manhattan	Manhattan Valley	40.797307	-73.964286
26	Manhattan	Morningside Heights	40.808000	-73.963896

	Borough	Neighborhood	Latitude	Longitude
27	Manhattan	Gramercy	40.737210	-73.981376
28	Manhattan	Battery Park City	40.711932	-74.016869
29	Manhattan	Financial District	40.707107	-74.010665
30	Manhattan	Carnegie Hill	40.782683	-73.953256
31	Manhattan	Noho	40.723259	-73.988434
32	Manhattan	Civic Center	40.715229	-74.005415
33	Manhattan	Midtown South	40.748510	-73.988713
34	Manhattan	Sutton Place	40.760280	-73.963556
35	Manhattan	Turtle Bay	40.752042	-73.967708
36	Manhattan	Tudor City	40.746917	-73.971219
37	Manhattan	Stuyvesant Town	40.731000	-73.974052
38	Manhattan	Flatiron	40.739673	-73.990947
39	Manhattan	Hudson Yards	40.756658	-74.000111
40	Queens	Little Neck	40.770826	-73.738898

4.2 Feature Set Construction

We will use the Foursquare API on the above dataframe which will use the latitude and longitude for each neighborhood and find the top 10 most commonly occurring venues (shops, restaurants, land marks, etc.)

The resultant dataset looks like this:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.910660	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.910660	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.876551	-73.910660	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.910660	Dunkin'	40.877136	-73.906666	Donut Shop
4	Marble Hill	40.876551	-73.910660	Starbucks	40.877531	-73.905582	Coffee Shop
...	...	...	...	...	...	...	...
3298	Little Neck	40.770826	-73.738898	Emily's Skin Care & Spa	40.772374	-73.734498	Spa
3299	Little Neck	40.770826	-73.738898	Allon Vision	40.766915	-73.738592	Doctor's Office
3300	Little Neck	40.770826	-73.738898	Deli & Grocery	40.773990	-73.742127	Deli / Bodega
3301	Little Neck	40.770826	-73.738898	Little Neck Cafe & Deli	40.774093	-73.742262	Deli / Bodega
3302	Little Neck	40.770826	-73.738898	City Line	40.772553	-73.733803	Outdoors & Recreation

3303 rows × 7 columns

There are a total of 3303 Venues belonging to several categories. For computational philosophy we will encode each type of venue in a neighborhood with the help of “One Hot” encoding and then grouping them together by Neighborhood:

	Neighborhood	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	Arepa Restaurant	Argentinian Restaurant	Art Gallery	...	Video Store	Vietnamese Restaurant	Volleyball Court	Waterfront	Whisky Bar	Wine Bar	Wine Shop	Wings Joint	Women's Store	Yoga Studio
0	Battery Park City	0.000000	0.000000	0.00	0.000000	0.010870	0.00	0.000000	0.000000	0.000000	...	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.010870	0.000000	0.021739	0.000000
1	Carnegie Hill	0.000000	0.000000	0.00	0.000000	0.011628	0.00	0.000000	0.011628	0.000000	...	0.00	0.011628	0.000000	0.000000	0.000000	0.011628	0.046512	0.000000	0.011628	0.034884
2	Central Harlem	0.000000	0.000000	0.00	0.065217	0.043478	0.00	0.000000	0.000000	0.043478	...	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	Chelsea	0.000000	0.000000	0.00	0.000000	0.030000	0.00	0.000000	0.000000	0.040000	...	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.030000	0.000000	0.010000	0.000000
4	Chinatown	0.000000	0.000000	0.00	0.000000	0.040000	0.00	0.000000	0.000000	0.000000	...	0.00	0.020000	0.000000	0.000000	0.000000	0.010000	0.000000	0.000000	0.000000	0.010000
5	Civic Center	0.000000	0.000000	0.00	0.000000	0.040000	0.01	0.000000	0.000000	0.010000	...	0.00	0.000000	0.000000	0.000000	0.010000	0.010000	0.020000	0.010000	0.000000	0.030000
6	Clinton	0.000000	0.000000	0.00	0.000000	0.040000	0.00	0.000000	0.000000	0.000000	...	0.00	0.000000	0.000000	0.000000	0.000000	0.010000	0.030000	0.000000	0.000000	0.000000
7	East Harlem	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	...	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
8	East Village	0.000000	0.000000	0.00	0.000000	0.020000	0.00	0.000000	0.010000	0.010000	...	0.00	0.030000	0.000000	0.000000	0.000000	0.030000	0.010000	0.000000	0.000000	0.000000
9	Financial District	0.000000	0.000000	0.00	0.000000	0.030000	0.00	0.000000	0.000000	0.000000	...	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.020000	0.000000	0.020000	0.000000
10	Flatiron	0.010000	0.000000	0.00	0.000000	0.040000	0.00	0.000000	0.000000	0.010000	...	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.030000	0.000000	0.010000	0.030000
11	Gramercy	0.000000	0.000000	0.00	0.000000	0.032258	0.00	0.010753	0.000000	0.010753	...	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.021505	0.000000	0.000000	0.010753
12	Greenwich Village	0.000000	0.000000	0.00	0.000000	0.020000	0.00	0.000000	0.000000	0.000000	...	0.00	0.020000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.010000

	Neighborhood	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	Arepa Restaurant	Argentinian Restaurant	Art Gallery	...	Video Store	Vietnamese Restaurant	Volleyball Court	Waterfront	Whisky Bar	Wine Bar	Wine Shop	Wings Joint	Women's Store	Yoga Studio
13	Hamilton Heights	0.000000	0.015873	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	...	0.00	0.000000	0.000000	0.000000	0.000000	0.015873	0.000000	0.000000	0.000000	0.031746
14	Hudson Yards	0.000000	0.000000	0.00	0.000000	0.053333	0.00	0.000000	0.000000	0.013333	...	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
15	Inwood	0.000000	0.000000	0.00	0.000000	0.017857	0.00	0.000000	0.000000	0.000000	...	0.00	0.000000	0.000000	0.000000	0.000000	0.035714	0.017857	0.000000	0.000000	0.017857
16	Lenox Hill	0.000000	0.000000	0.01	0.000000	0.000000	0.00	0.000000	0.000000	0.010000	...	0.00	0.000000	0.000000	0.000000	0.000000	0.010000	0.020000	0.000000	0.010000	0.000000
17	Lincoln Square	0.000000	0.000000	0.00	0.000000	0.020202	0.00	0.000000	0.000000	0.000000	...	0.00	0.000000	0.000000	0.000000	0.000000	0.010101	0.030303	0.000000	0.000000	0.010101
18	Little Italy	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	...	0.00	0.010000	0.000000	0.000000	0.000000	0.020000	0.010000	0.000000	0.010000	0.000000
19	Little Neck	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	...	0.00	0.018868	0.000000	0.000000	0.000000	0.000000	0.000000	0.018868	0.000000	0.000000
20	Lower East Side	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.020833	0.041667	...	0.00	0.020833	0.000000	0.000000	0.000000	0.000000	0.020833	0.000000	0.020833	0.020833
21	Manhattan Valley	0.000000	0.000000	0.00	0.000000	0.020408	0.00	0.000000	0.000000	0.000000	...	0.00	0.040816	0.000000	0.000000	0.000000	0.000000	0.020408	0.020408	0.000000	0.020408
22	Manhattanville	0.000000	0.000000	0.00	0.000000	0.022727	0.00	0.000000	0.000000	0.000000	...	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
23	Marble Hill	0.000000	0.000000	0.00	0.000000	0.041667	0.00	0.000000	0.000000	0.000000	...	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.041667
24	Midtown	0.000000	0.000000	0.00	0.000000	0.030000	0.00	0.000000	0.000000	0.010000	...	0.00	0.010000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.010000	0.010000
25	Midtown South	0.000000	0.000000	0.00	0.000000	0.030000	0.00	0.000000	0.000000	0.000000	...	0.00	0.000000	0.000000	0.000000	0.000000	0.010000	0.000000	0.000000	0.000000	0.010000
26	Morningside Heights	0.000000	0.000000	0.00	0.000000	0.075000	0.00	0.000000	0.000000	0.000000	...	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

	Neighborhood	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	Arepa Restaurant	Argentinian Restaurant	Art Gallery	...	Video Store	Vietnamese Restaurant	Volleyball Court	Waterfront	Whisky Bar	Wine Bar	Wine Shop	Wings Joint	Women's Store	Yoga Studio
27	Murray Hill	0.000000	0.000000	0.00	0.000000	0.040000	0.00	0.000000	0.000000	0.000000	...	0.00	0.010000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
28	Noho	0.000000	0.000000	0.00	0.000000	0.010000	0.00	0.000000	0.010000	0.030000	...	0.00	0.010000	0.000000	0.000000	0.010000	0.030000	0.020000	0.010000	0.000000	0.010000
29	Roosevelt Island	0.000000	0.000000	0.00	0.000000	0.033333	0.00	0.000000	0.000000	0.000000	...	0.00	0.000000	0.000000	0.033333	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
30	Soho	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.020000	...	0.00	0.000000	0.000000	0.000000	0.000000	0.010000	0.000000	0.000000	0.020000	0.000000
31	Stuyvesant Town	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	...	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
32	Sutton Place	0.000000	0.010000	0.00	0.000000	0.020000	0.00	0.000000	0.000000	0.000000	...	0.00	0.010000	0.000000	0.000000	0.000000	0.010000	0.020000	0.000000	0.000000	0.010000
33	Tribeca	0.000000	0.000000	0.00	0.000000	0.066667	0.00	0.000000	0.011111	0.011111	...	0.00	0.000000	0.011111	0.000000	0.011111	0.044444	0.022222	0.000000	0.000000	0.011111
34	Tudor City	0.000000	0.000000	0.00	0.000000	0.012658	0.00	0.000000	0.000000	0.000000	...	0.00	0.025316	0.000000	0.000000	0.000000	0.000000	0.025316	0.000000	0.000000	0.012658
35	Turtle Bay	0.000000	0.000000	0.00	0.000000	0.010000	0.00	0.000000	0.000000	0.000000	...	0.00	0.000000	0.000000	0.000000	0.000000	0.030000	0.000000	0.000000	0.000000	0.000000
36	Upper East Side	0.000000	0.000000	0.00	0.000000	0.030000	0.00	0.000000	0.000000	0.040000	...	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.020000	0.000000	0.010000	0.030000
37	Upper West Side	0.010417	0.000000	0.00	0.000000	0.020833	0.00	0.000000	0.000000	0.000000	...	0.00	0.010417	0.000000	0.000000	0.000000	0.031250	0.010417	0.000000	0.000000	0.010417
38	Washington Heights	0.012195	0.000000	0.00	0.000000	0.012195	0.00	0.000000	0.000000	0.000000	...	0.00	0.000000	0.000000	0.000000	0.000000	0.012195	0.024390	0.000000	0.012195	0.000000
39	West Village	0.010000	0.000000	0.00	0.000000	0.030000	0.00	0.000000	0.000000	0.010000	...	0.00	0.000000	0.000000	0.000000	0.000000	0.030000	0.010000	0.000000	0.000000	0.000000
40	Yorkville	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	...	0.01	0.020000	0.000000	0.000000	0.000000	0.010000	0.030000	0.000000	0.000000	0.000000

Let us convert this massive dataframe into the top 10 venues for each neighborhood for the sake of visualization like this as a sample:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Battery Park City	Park	Coffee Shop	Clothing Store	Hotel	Gym	Boat or Ferry	Playground	Memorial Site	Shopping Mall	Pizza Place
1	Carnegie Hill	Coffee Shop	Café	Wine Shop	Yoga Studio	Cosmetics Shop	Gym / Fitness Center	Bookstore	French Restaurant	Gym	Bar
2	Central Harlem	African Restaurant	Seafood Restaurant	American Restaurant	Gym / Fitness Center	Chinese Restaurant	French Restaurant	Art Gallery	Bar	Public Art	Music Venue
3	Chelsea	Coffee Shop	Bakery	Art Gallery	Ice Cream Shop	Hotel	American Restaurant	Wine Shop	French Restaurant	Seafood Restaurant	Market
4	Chinatown	Chinese Restaurant	Bakery	Cocktail Bar	American Restaurant	Salon / Barbershop	Dessert Shop	Mexican Restaurant	Bubble Tea Shop	Hotpot Restaurant	Ice Cream Shop

5. Machine Learning

Now that we have formed our featureset which contains the information about the most prominent venues in their vicinity. We will use this feature set to fit a clustering Machine Learning which will group our neighbourhoods together based on the prominent venues that are nearby them.

5.1 K Means Clustering

As the name suggests we will use this unsupervised clustering algorithm to group together the dataset into K clusters.

For us we will select 5 clusters and apply the algorithm. The algorithm will then classify each neighborhood with a label on the basis of the most prominent venue in that neighborhood and then group them together.

The resultant dataframe looks like this:

	Cluster Labels	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Borough	Latitude	Longitude
0	1	Battery Park City	Park	Coffee Shop	Clothing Store	Hotel	Gym	Boat or Ferry	Playground	Memorial Site	Shopping Mall	Pizza Place	Manhattan	40.711932	-74.016869
1	1	Carnegie Hill	Coffee Shop	Café	Wine Shop	Yoga Studio	Cosmetics Shop	Gym / Fitness Center	Bookstore	French Restaurant	Gym	Bar	Manhattan	40.782683	-73.953256
2	1	Central Harlem	African Restaurant	Seafood Restaurant	American Restaurant	Gym / Fitness Center	Chinese Restaurant	French Restaurant	Art Gallery	Bar	Public Art	Music Venue	Manhattan	40.815976	-73.943211
3	1	Chelsea	Coffee Shop	Bakery	Art Gallery	Ice Cream Shop	Hotel	American Restaurant	Wine Shop	French Restaurant	Seafood Restaurant	Market	Manhattan	40.744035	-74.003116
4	1	Chinatown	Chinese Restaurant	Bakery	Cocktail Bar	American Restaurant	Salon / Barbershop	Dessert Shop	Mexican Restaurant	Bubble Tea Shop	Hotpot Restaurant	Ice Cream Shop	Manhattan	40.715618	-73.994279

Now let us plot the clusters on the map to visualize the grouping:

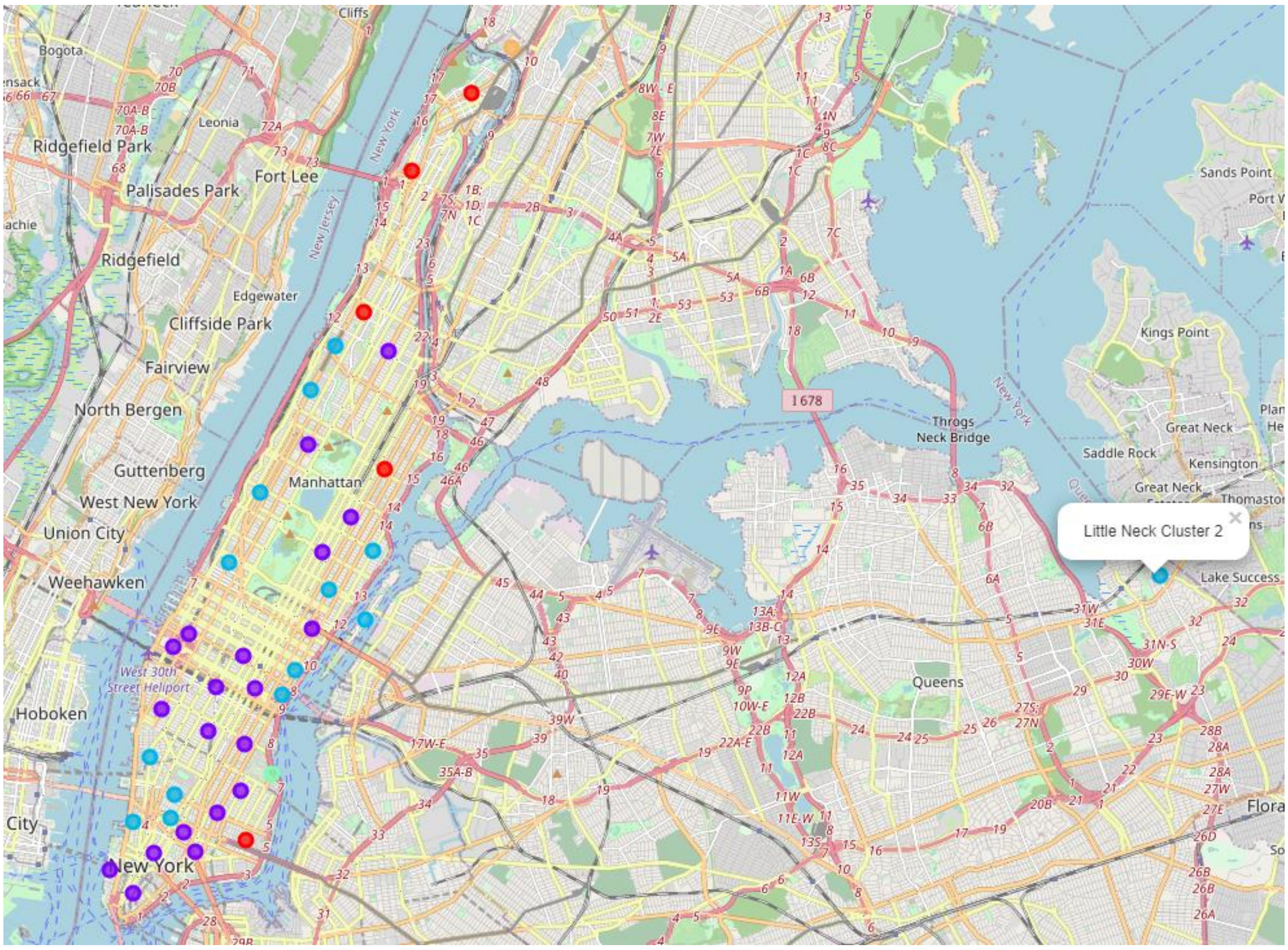

Figure 3: The Manhattan Neighborhoods clustered in 5 groups (Red, Blue, Purple, Light Green and Orange)

6. Results

From the above map it can be seen that K means has divided all the possible neighborhoods in Manhattan into 5 predefined clusters or group (Red, Purple, Blue, Lime Green and Orange). We also can identify that our Neighborhood (Little Neck) belongs to the Blue cluster. Therefore out of all the 40 possible locations in Manhattan, the neighborhoods highlighted in Blue are most similar to our neighborhood and therefore these are the best locations to open up new branch which will face similar geographical conditions which our current shop or business in which resulted it to thrive.

7. Discussion

K means identified that all the blue points in the same cluster has our neighborhood. These blue neighborhoods are the most suitable neighborhoods where we can open up a new branch of our shop or business. Let us have a look which are those neighborhoods.

Now let us have a look at all the possible neighbourhoods which belong to the blue cluster which are the best possible locations to open up our new business or shop branch as it will ensure that your new branch is located in an environment which is similar to your current business or shop.

	Cluster Labels	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Borough	Latitude	Longitude
12	2	Greenwich Village	Italian Restaurant	Clothing Store	Sushi Restaurant	Boutique	Indian Restaurant	Coffee Shop	Cosmetics Shop	Dessert Shop	Bubble Tea Shop	Ice Cream Shop	Manhattan	40.726933	-73.999914
16	2	Lenox Hill	Italian Restaurant	Coffee Shop	Sushi Restaurant	Café	Cocktail Bar	Pizza Place	Gym / Fitness Center	Gym	Burger Joint	Steakhouse	Manhattan	40.768113	-73.958860
17	2	Lincoln Square	Plaza	Café	Theater	Concert Hall	Performing Arts Venue	Wine Shop	Park	Food Truck	Coffee Shop	Indie Movie Theater	Manhattan	40.773529	-73.985338
22	2	Manhattanville	Coffee Shop	Deli / Bodega	Mexican Restaurant	Bar	Italian Restaurant	Seafood Restaurant	Fried Chicken Joint	Bike Trail	Spanish Restaurant	Scenic Lookout	Manhattan	40.816934	-73.957385
26	2	Morningside Heights	Bookstore	American Restaurant	Coffee Shop	Park	Café	Sandwich Place	Deli / Bodega	Burger Joint	Food Truck	Seafood Restaurant	Manhattan	40.808000	-73.963896
29	2	Roosevelt Island	Coffee Shop	Deli / Bodega	Residential Building (Apartment / Condo)	Gym	Supermarket	Bus Line	Grocery Store	Greek Restaurant	Outdoors & Recreation	Soccer Field	Manhattan	40.762160	-73.949168
30	2	Soho	Clothing Store	Boutique	Italian Restaurant	Shoe Store	Mediterranean Restaurant	Salon / Barbershop	Hotel	Coffee Shop	Sporting Goods Shop	Bakery	Manhattan	40.722184	-74.000657
33	2	Tribeca	American Restaurant	Italian Restaurant	Park	Wine Bar	Café	Spa	Gym / Fitness Center	Greek Restaurant	French Restaurant	Basketball Court	Manhattan	40.721522	-74.010683

	Cluster Labels	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Borough	Latitude	Longitude
34	2	Tudor City	Park	Mexican Restaurant	Café	Pizza Place	Greek Restaurant	Gym	Diner	Coffee Shop	Garden	Seafood Restaurant	Manhattan	40.746917	-73.971219
35	2	Turtle Bay	Italian Restaurant	Coffee Shop	Japanese Restaurant	Sushi Restaurant	Café	Ramen Restaurant	Park	Deli / Bodega	Seafood Restaurant	Steakhouse	Manhattan	40.752042	-73.967708
37	2	Upper West Side	Italian Restaurant	Bakery	Wine Bar	Coffee Shop	Bar	Café	Bagel Shop	Sports Bar	Mediterranean Restaurant	Indian Restaurant	Manhattan	40.787658	-73.977059
39	2	West Village	Italian Restaurant	New American Restaurant	Cocktail Bar	Park	American Restaurant	Cosmetics Shop	Coffee Shop	Ice Cream Shop	Wine Bar	Theater	Manhattan	40.734434	-74.006180
40	2	Yorkville	Italian Restaurant	Gym	Coffee Shop	Bar	Deli / Bodega	Sushi Restaurant	Wine Shop	Japanese Restaurant	Mexican Restaurant	Ice Cream Shop	Manhattan	40.775930	-73.947118

There are 13 (out of a total 40) which have been identified to be the best suitable neighborhoods most similar to the our current neighborhood of Little Neck best for opening a new branch of our shop or business.

Let us look at the venue features of our current neighbourhood:

	Cluster Labels	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Borough	Latitude	Longitude
19	2	Little Neck	Chinese Restaurant	Deli / Bodega	Korean Restaurant	Italian Restaurant	Coffee Shop	Spa	Bank	Bakery	Bus Station	Peruvian Restaurant	Queens	40.770826	-73.738898

Looking at the above dataframes we see that the clustered neighbourhoods, 13 possible neighbourhoods in Manhattan where we can open an another branch od f our shop or business which will have similar geographic features similar to the ones in our current neighbourhood such as Italian Restaurants, Coffee shops, etc.

8. Conclusion

In conclusion we have successfully used K means Clustering to identify 13 possible neighbourhoods (out of a total of 40) in Manhattan which will be the most suitable to open up a new Branch of our shop or business as they are most similar to our current neighbourhood therefore will ensure similar customer base in these new locations enabling good business for the new venture.