Demographics of Manhattan Based on the Distribution of Restaurants

Ke Guan

July 23, 2019

1.Introduction

1.1 Background

The United States possesses a long history of absorbing people worldwide successfully, and it prides itself to representing a nation of immigrants. Americans have offered opportunities to immigrants and their children and non-immigrants to better themselves and be thoroughly incorporated into U.S. Society. The successful integration of immigrants and non-immigrants contributes to economic vitality and a vibrant and ever-changing culture. This day, the 41 million immigrants in the United States represent 13.1 percent of the U.S. population. The U.S.-born children of immigrants, the second generation, represent another 37.1 million people or 12 percent of the population. Thus, together the first and second generations account for one out of four members of the U.S. population. The immigrants and non-immigrants integrate highly dependent on the racial and ethnic groups, the legal status, the social class, and the geographic area. This project mainly focused on Neighborhoods of Manhattan in New York City to detective geographic areas the immigrants and non-immigrants integrate based on the restaurant's distribution.

1.2 Problem

Data that might contribute to determining the distributions of restaurants with different food cultures around those neighborhoods in Manhattan.

1.3 Interest

People who want to launch a restaurant would be interested in the clustering of regions with various restaurants. Others who would be curious tourists about different food cultures.

1.4 Data acquisition

Data is requested from the Foursquare, then leverage the location data to explore venues around neighborhoods of Manhattan. Extract venues with "Restaurant" and according to the food cultures of these restaurants to cluster these neighborhoods and locate the clusters onto the map.

2.Methodology

2.1 Data Cleaning and Feature Selection

Data downloaded from the Foursquare were combined into one table. There were five boroughs and 306 neighbors in New York. As I primarily focused on Manhattan, I retrieved all neighbors for it. Thoroughly, there were 40 Neighbors involved, including Marble Hill, Chinatown, Washington Heights, etc. I designed

the limit as 200 venues and the radius 500 meters for each borough from their given latitude and longitude information. Hither remain five of the list Venues name, category, latitude, and longitude information from Foursquare API.

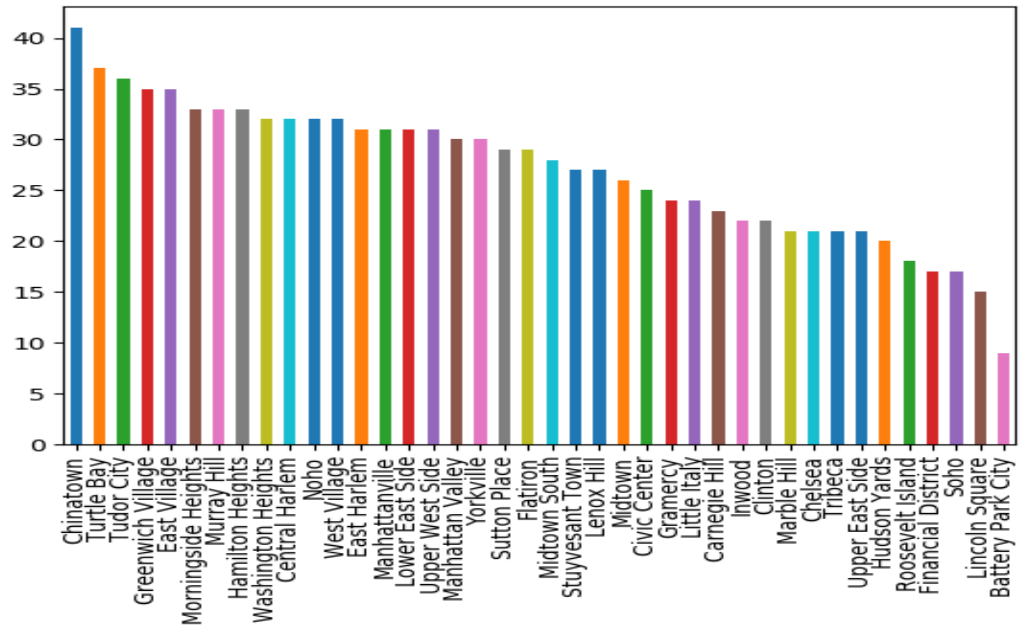
	name	categories	lat	lng
0	Hotel 50 Bowery NYC	Hotel	40.715936	-73.996789
1	Kiki's	Greek Restaurant	40.714476	-73.992036
2	Bar Belly	Cocktail Bar	40.715135	-73.991802
3	Spicy Village	Chinese Restaurant	40.717010	-73.993530
4	Mission Escape Games	General Entertainment	40.716505	-73.994720

In summary of this data **3991** venues and **337** unique categories were returned by Foursquare. Reframed these items into a dataset with selected features. Here is a merged table of boroughs and venues.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Latitude	Longitude	Category
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop
4	Marble Hill	40.876551	-73.91066	Dunkin'	40.877136	-73.906666	Donut Shop

76 unique categories were returned by extracted categories with “Restaurant.’ Some items have names containing specific Country and regional label, but some categories are not, then I combine categories together based on the name of restaurants and primary food the restaurants sell. For example, “Sushi Restaurant” typically sells diverse kinds of sushi, hence I classify the category as “Japanese Restaurant.” Another example is I classify “Shanghai Restaurant” as “Chinese Restaurant.”

After reforming the data frame, the total number of unique categories for these restaurants with diversity countries’ food cultures is 66. We can perceive that the number of restaurants in diverse neighborhoods. For instance, in Chinatown, there are 41 restaurants and 37 restaurants in Turtle Bay.



2.2 Data Exploratory and Model

Utilize one-hot encoding to weigh the restaurants within the neighborhoods and extract the top 5 kinds of restaurants for each Neighborhood. For example, in Chinatown, the highest frequency of categories is Chinese Restaurant (0.38), and the second highest is Vietnamese Restaurant (0.10).

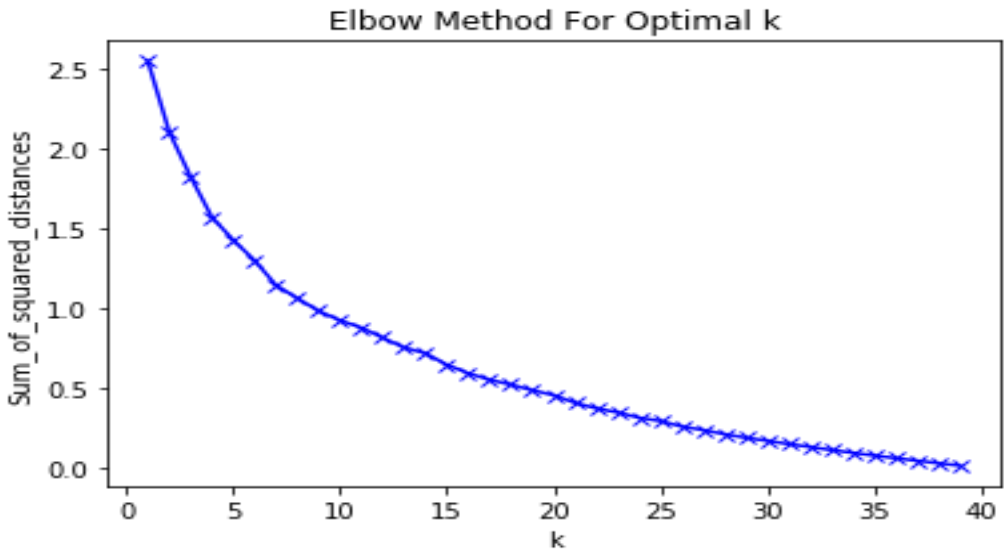
```
----Chinatown----
                venue  freq
0     Chinese Restaurant  0.51
1    American Restaurant  0.10
2    Japanese Restaurant  0.05
3       Malay Restaurant  0.05
4  Vietnamese Restaurant  0.05
5     Mexican Restaurant  0.02
6    Austrian Restaurant  0.02
7      Greek Restaurant  0.02
8     Korean Restaurant  0.02
9    English Restaurant  0.02
```

```
----Hudson Yards----
                venue  freq
0    Italian Restaurant  0.30
1   American Restaurant  0.25
2   Spanish Restaurant  0.10
3      Thai Restaurant  0.10
4  Caucasian Restaurant  0.05
5  Pakistani Restaurant  0.05
6   Japanese Restaurant  0.05
7     Greek Restaurant  0.05
8   Chinese Restaurant  0.05
9    Paella Restaurant  0.00
```

Then I created a table which shows list of top 10 categorical restaurants for each neighborhood in below table.

	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	Italian Restaurant	Japanese Restaurant	Vegetarian / Vegan Restaurant	American Restaurant	Mexican Restaurant	Mediterranean Restaurant	French Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fast Food Restaurant
1	Carnegie Hill	Italian Restaurant	Japanese Restaurant	Vietnamese Restaurant	Indian Restaurant	American Restaurant	French Restaurant	German Restaurant	Mexican Restaurant	Mediterranean Restaurant	Chinese Restaurant
2	Central Harlem	Southern / Soul Food Restaurant	African Restaurant	American Restaurant	Japanese Restaurant	French Restaurant	Chinese Restaurant	Tapas Restaurant	Mexican Restaurant	Caribbean Restaurant	Middle Eastern Restaurant
3	Chelsea	Italian Restaurant	American Restaurant	Chinese Restaurant	Japanese Restaurant	Tapas Restaurant	French Restaurant	Israeli Restaurant	Vegetarian / Vegan Restaurant	Paella Restaurant	Indian Restaurant
4	Chinatown	Chinese Restaurant	American Restaurant	Vietnamese Restaurant	Japanese Restaurant	Malay Restaurant	Dim Sum Restaurant	Thai Restaurant	Vegetarian / Vegan Restaurant	New American Restaurant	Korean Restaurant

There are some common venue categories in Neighborhoods. In this reason, I used unsupervised learning K-Means algorithm to cluster the neighborhoods. I analyzed the K-Means with elbow method. It shows the sum of squared error is consistently decreasing.



I ran K-Means to cluster the neighborhoods into five clusters because after k=5, the slope of the curve slows down.

At this point, I prepare the table with labels of the clusters for each neighborhood.

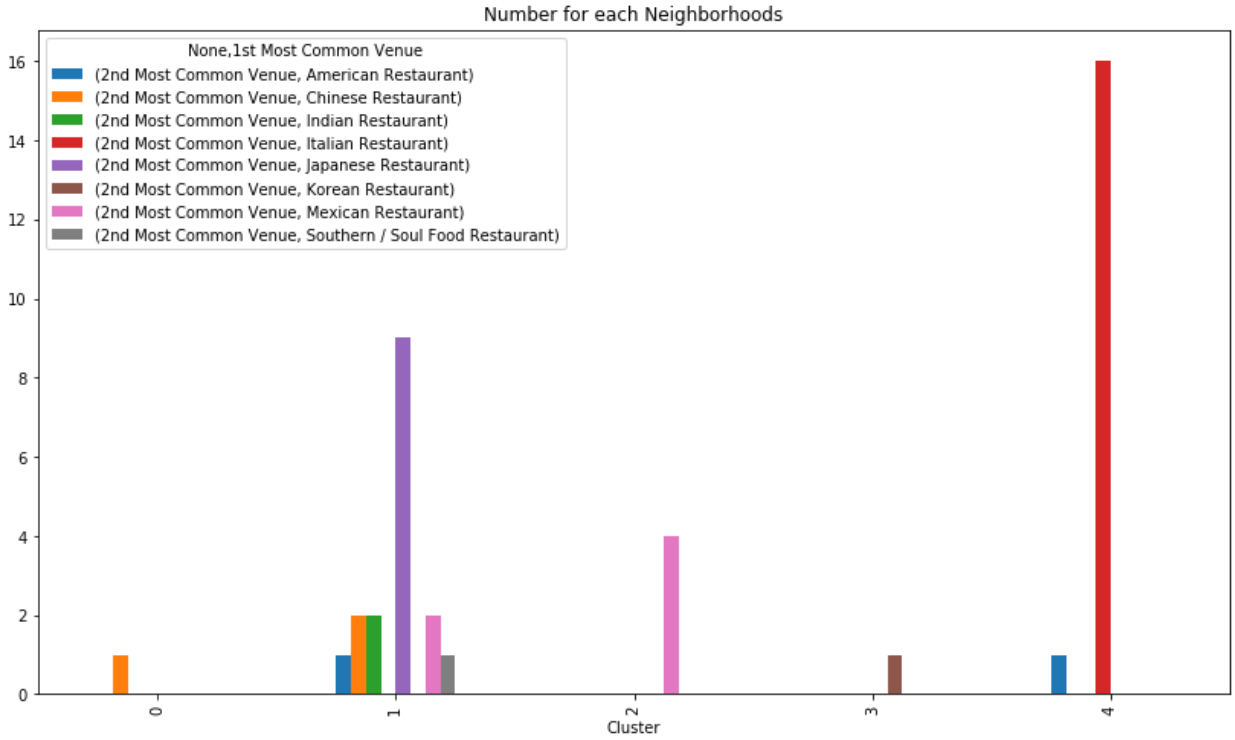
	Borough	Neighborhood	Latitude	Longitude	Cluster	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Manhattan	Marble Hill	40.876551	-73.910660	4	Mexican Restaurant	Japanese Restaurant	Thai Restaurant	Cuban Restaurant	Spanish Restaurant
1	Manhattan	Chinatown	40.715618	-73.994279	3	Chinese Restaurant	American Restaurant	Vietnamese Restaurant	Japanese Restaurant	Malay Restaurant
2	Manhattan	Washington Heights	40.851903	-73.936900	4	Mexican Restaurant	Spanish Restaurant	Tapas Restaurant	Chinese Restaurant	Japanese Restaurant
3	Manhattan	Inwood	40.867684	-73.921210	4	Mexican Restaurant	Spanish Restaurant	Japanese Restaurant	Chinese Restaurant	Mediterranean Restaurant
4	Manhattan	Hamilton Heights	40.823604	-73.949688	2	Mexican Restaurant	Japanese Restaurant	Ethiopian Restaurant	Spanish Restaurant	Caribbean Restaurant

Then based on the **1st Most Common Venue and 2nd Most Common Venue** in each cluster, created table to find the main food culture for the neighborhoods.

		2nd Most Common Venue
Cluster	1st Most Common Venue	
0	Chinese Restaurant	1
1	American Restaurant	1
	Chinese Restaurant	2
	Indian Restaurant	2
	Japanese Restaurant	9
	Mexican Restaurant	2
	Southern / Soul Food Restaurant	1
2	Mexican Restaurant	4
3	Korean Restaurant	1
4	American Restaurant	1
	Italian Restaurant	16

Cluster 0 includes one neighborhood, and the most typical restaurant in the neighborhood is Chinese Restaurant. Cluster 4 involves 17 neighborhoods, 16 out of 17 neighborhoods take Italian Restaurants as the 1st most common Restaurant.

From the below graph, we can label cluster 0 as “Chinese food culture” region, which means in this neighborhood, probably, a considerable number of Chinese immigrants and non-immigrants live here. Cluster 1 includes 16 neighborhoods, even 9 neighborhoods regard Japanese food as 1st most common restaurant, this Cluster contains other countries’ food. We can mark it as “mixed” which means immigrants and non-immigrants living here come from different countries. There are four neighborhoods take Mexican Restaurant as the 1st most common restaurants in Cluster 2. And Cluster 3 merely includes one neighborhood whose restaurants primarily sell Korean food. Cluster 4 involves 17 neighborhoods, restaurants here mainly are American and Italian Restaurants, hence, the neighborhoods’ population are mainly white.



Therefore, labels for these 5 clusters:

Cluster 0: “Chinese”

Cluster 1: “Mixed”

Cluster 2: “Mexican”

Cluster 3: “Korean”

Cluster 4: “White”

3. Results

Let’s detect which neighborhoods are included in each cluster.

Cluster 0: “Chinese”

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1	Chinatown	Chinese Restaurant	American Restaurant	Vietnamese Restaurant	Japanese Restaurant	Malay Restaurant

There is no surprise that Chinatown has Chinese restaurant as the 1t most common restaurant.

Cluster 1: “Mixed”

In the neighborhoods contained in Cluster 1, Japanese food culture spreads mostly. We see Stuyvesant town has the American and Italian restaurants as the most common restaurant, and there are 74.1%

White, 3.6% African American, 0.1% Native American, 12.6% Asian, 0.1% from other races, and 2.6% from two or more races. Hispanic or Latino of any race were 6.7% of the population. Therefore, this neighborhood should be clustered into Cluster 4. From the K-Means elbow graph, when taking K=5, there is a sum of squared error exist. Next, we can regard this as a misclassification.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
4	Hamilton Heights	Mexican Restaurant	Japanese Restaurant	Ethiopian Restaurant	Spanish Restaurant	Caribbean Restaurant
5	Manhattanville	Mexican Restaurant	Italian Restaurant	American Restaurant	Japanese Restaurant	Chinese Restaurant
6	Central Harlem	Southern / Soul Food Restaurant	African Restaurant	American Restaurant	Japanese Restaurant	French Restaurant
11	Roosevelt Island	Japanese Restaurant	Greek Restaurant	Chinese Restaurant	Mexican Restaurant	Caribbean Restaurant
15	Midtown	Japanese Restaurant	American Restaurant	Cuban Restaurant	Indian Restaurant	Mediterranean Restaurant
16	Murray Hill	Japanese Restaurant	Chinese Restaurant	French Restaurant	Italian Restaurant	Jewish Restaurant
19	East Village	Japanese Restaurant	Chinese Restaurant	Mexican Restaurant	Vegetarian / Vegan Restaurant	Korean Restaurant
20	Lower East Side	Japanese Restaurant	Mexican Restaurant	Chinese Restaurant	Italian Restaurant	Southern / Soul Food Restaurant
22	Little Italy	Chinese Restaurant	Italian Restaurant	Thai Restaurant	Mediterranean Restaurant	Japanese Restaurant
25	Manhattan Valley	Indian Restaurant	Mexican Restaurant	Chinese Restaurant	Italian Restaurant	Thai Restaurant
26	Morningside Heights	Chinese Restaurant	Italian Restaurant	Mexican Restaurant	American Restaurant	Indian Restaurant
29	Financial District	Japanese Restaurant	Italian Restaurant	American Restaurant	Falafel Restaurant	Mediterranean Restaurant
34	Sutton Place	Indian Restaurant	Italian Restaurant	Mexican Restaurant	American Restaurant	Japanese Restaurant
35	Turtle Bay	Japanese Restaurant	French Restaurant	Italian Restaurant	Greek Restaurant	Indian Restaurant
36	Tudor City	Japanese Restaurant	Chinese Restaurant	Thai Restaurant	Mexican Restaurant	Jewish Restaurant
37	Stuyvesant Town	American Restaurant	Italian Restaurant	Mexican Restaurant	Thai Restaurant	Japanese Restaurant
38	Flatiron	Japanese Restaurant	American Restaurant	New American Restaurant	Italian Restaurant	Mexican Restaurant

Cluster 2: “Mexican”

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Marble Hill	Mexican Restaurant	Japanese Restaurant	Thai Restaurant	Cuban Restaurant	Spanish Restaurant
2	Washington Heights	Mexican Restaurant	Spanish Restaurant	Tapas Restaurant	Chinese Restaurant	Japanese Restaurant
3	Inwood	Mexican Restaurant	Spanish Restaurant	Japanese Restaurant	Chinese Restaurant	Mediterranean Restaurant
7	East Harlem	Mexican Restaurant	Thai Restaurant	Chinese Restaurant	Moroccan Restaurant	Spanish Restaurant

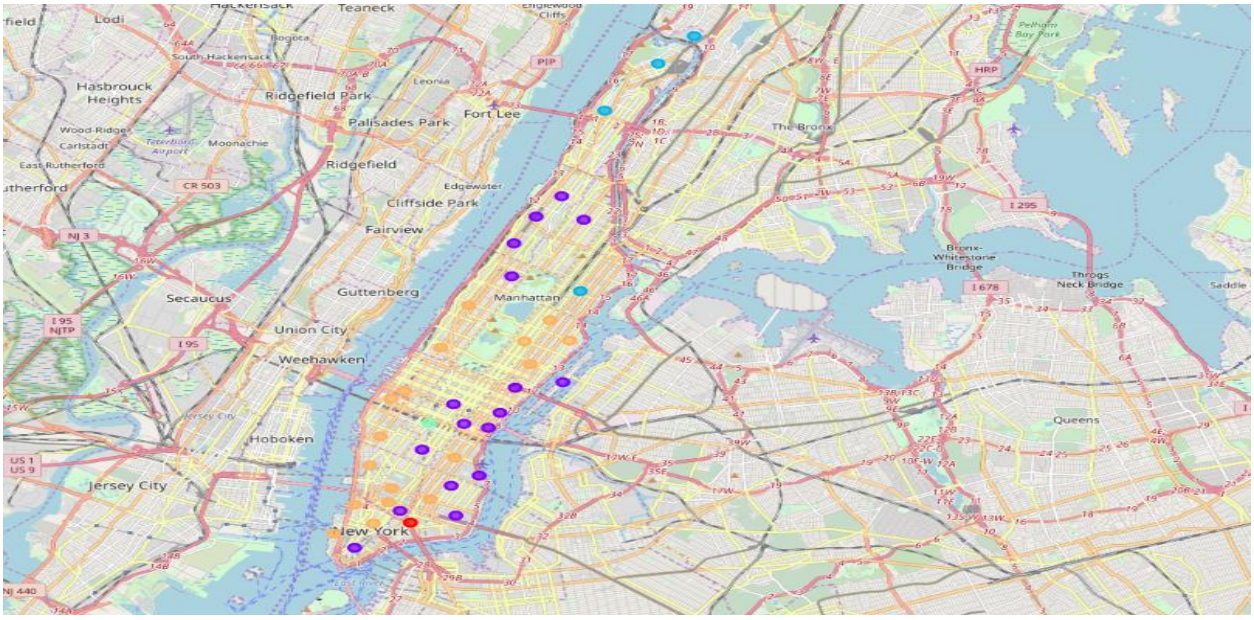
Cluster 3: “Korean”

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
33	Midtown South	Korean Restaurant	Japanese Restaurant	American Restaurant	Vegetarian / Vegan Restaurant	Persian Restaurant

Cluster 4: “White”

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
8	Upper East Side	Italian Restaurant	French Restaurant	Vegetarian / Vegan Restaurant	American Restaurant	Japanese Restaurant
9	Yorkville	Italian Restaurant	Japanese Restaurant	Mexican Restaurant	Thai Restaurant	Chinese Restaurant
10	Lenox Hill	Italian Restaurant	Japanese Restaurant	Thai Restaurant	Mexican Restaurant	Afghan Restaurant
12	Upper West Side	Italian Restaurant	Vegetarian / Vegan Restaurant	Indian Restaurant	American Restaurant	Japanese Restaurant
13	Lincoln Square	Italian Restaurant	French Restaurant	American Restaurant	Mediterranean Restaurant	Mexican Restaurant
14	Clinton	Italian Restaurant	American Restaurant	Chinese Restaurant	New American Restaurant	Japanese Restaurant
17	Chelsea	Italian Restaurant	American Restaurant	Chinese Restaurant	Japanese Restaurant	Tapas Restaurant
18	Greenwich Village	Italian Restaurant	Japanese Restaurant	American Restaurant	Indian Restaurant	Caribbean Restaurant
21	Tribeca	American Restaurant	Italian Restaurant	Greek Restaurant	French Restaurant	Chinese Restaurant
23	Soho	Italian Restaurant	Mediterranean Restaurant	Vegetarian / Vegan Restaurant	Japanese Restaurant	French Restaurant
24	West Village	Italian Restaurant	New American Restaurant	American Restaurant	Japanese Restaurant	French Restaurant
27	Gramercy	Italian Restaurant	American Restaurant	Japanese Restaurant	Thai Restaurant	Indian Restaurant
28	Battery Park City	Italian Restaurant	Mediterranean Restaurant	Vegetarian / Vegan Restaurant	Japanese Restaurant	American Restaurant
30	Carnegie Hill	Italian Restaurant	Japanese Restaurant	Vietnamese Restaurant	Indian Restaurant	American Restaurant
31	Noho	Italian Restaurant	French Restaurant	Japanese Restaurant	Mexican Restaurant	American Restaurant
32	Civic Center	Italian Restaurant	French Restaurant	American Restaurant	Japanese Restaurant	Falafel Restaurant
39	Hudson Yards	Italian Restaurant	American Restaurant	Spanish Restaurant	Thai Restaurant	Pakistani Restaurant

Created a clustered map of neighborhoods of Manhattan in below.



4. Discussion

Manhattan has been described as the cultural, financial, media, and entertainment capital of the world. Detecting the distribution of local ethnicity based on the distribution of restaurants is only one way of observing the population distribution in Manhattan. According to the demographic data obtained from the census, it may be more intuitive to classify the neighborhoods in Manhattan by its local ethnicity. There are many other methods for classifying neighborhoods in addition to k-means based on food cultures, like Hierarchical Clustering and decision tree, Random Forest.

When k=5 is used, from the elbow diagram, it can be seen that there is an error when K=5. When K=40 the error rate is 0, however, it does not make much sense to cluster 40 neighborhoods as 40 clusters.

I ended the study by visualizing the data and clustering information on the map. People who in the further willing to invest to open restaurant can take it a reference.

5. Conclusion

According to 2012 Census estimates, 65.2% of the population was White, 18.4% Black or African American, 1.2% American Indian and Alaska Native, 12.0% Asian, and 3.1% of two or more races. 25.8% of Manhattan's population was of Hispanic or Latino origin, of any race. As results of my analysis, the distribution of population can be clustered. It agrees that people integrate highly dependent on the racial and ethnic groups.