

Predicting the best neighborhoods for opening Korean restaurants in Manhattan, New York



A project report by Swapnil Srivastava
May 17th, 2020.

Introduction

The City of New York is famous for its wide variety of diverse cuisine. Its food culture includes an array of international cuisines influenced by the city's immigration history.

Korean restaurants have become very trendy these days in the United States, thanks to K-pop bands, K-dramas and appetizing food they eat (Kimchi and Bibimbap). My personal favorite is Korean BBQ.

Nowadays it seems that there is one Korean joint at every corner of NY, not only in major cities but also in smaller cities. Starting a Korean restaurant can be a great business opportunity, but you need to distinguish yourself from others to enjoy long-term success.

Business Problem

My client wants to open his business in Manhattan area, so I would focus on that borough during my analysis. We define potential neighborhood based on the number of Korean joints which are operating in each neighborhood.

Manhattan has full potential but also is a very challenging district to open a business in the sector because of high competition. New Korean restaurant should be open in an area if the neighborhood which has a smaller number of Korean restaurants. In this way the restaurant can attract more customers.

Therefore, this analysis is necessary to ensure that we have enough customers and that we are not so close to other places having large number of Korean restaurants.

Who would be interested?

My target audience would be stakeholders, who would want to open Korean restaurants in and around Manhattan, New York.

My job would be to accurately determine which places in Manhattan can one setup this business so that it would flourish, and one can make the most profit there.

Date Selection

We would first obtain the dataset of different localities of New York. Only then we can begin our data interpretation based upon the statistics of different locations such as Borough, Neighborhood, Latitude, Longitude etc.

For this we use data sources from Kaggle or from New York public dataset (https://cocl.us/new_york_dataset)

To identify the characteristics of our competitors' venues in Manhattan, we would first need to find out the number of Korean restaurants in Manhattan currently and their location.

We would use Foursquare key value for Korean restaurants to find out different such places operating in Manhattan, New York. (*#Korean Restaurant = 4bf58dd8d48988d113941735*)

Then we would use Google Map API to find their geographic coordinates based on their geographic location.

We would now group the Korean Restaurant venues by neighborhoods for easy analysis, to checkout which areas have maximum and minimum number of Korean restaurants.

Foursquare API is used to explore neighborhoods in Manhattan, New York.

After that, explore function to get different restaurant categories in each neighborhood.

```
#print the venue count of Korean restaurants in each neighborhood of Manhattan, NY.
manhattan_grouped
```

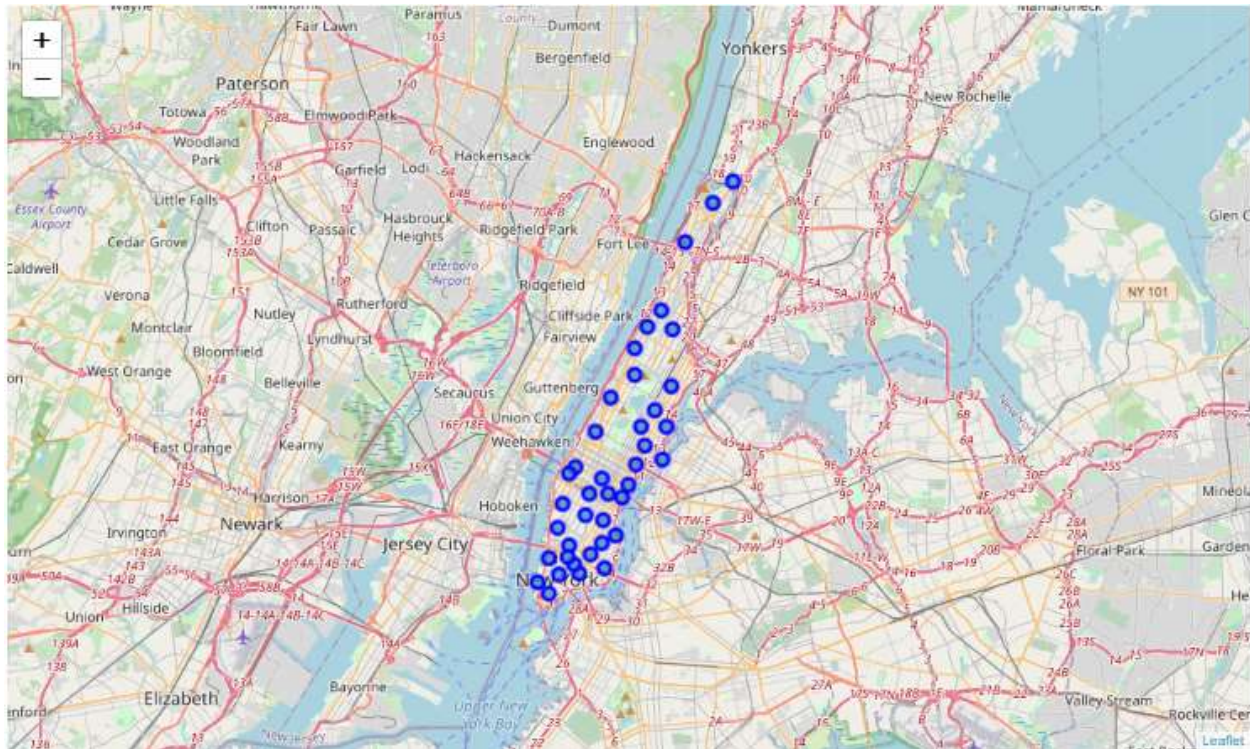
Out[19]:

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Battery Park City	7	7	7	7	7	7
Carnegie Hill	2	2	2	2	2	2
Chelsea	23	23	23	23	23	23
Chinatown	12	12	12	12	12	12
Civic Center	15	15	15	15	15	15
Clinton	26	26	26	26	26	26
East Village	24	24	24	24	24	24
Financial District	6	6	6	6	6	6
Flatiron	50	50	50	50	50	50
Gramercy	47	47	47	47	47	47
Greenwich Village	23	23	23	23	23	23
Hudson Yards	17	17	17	17	17	17
Lenox Hill	4	4	4	4	4	4
Lincoln Square	13	13	13	13	13	13
Little Italy	21	21	21	21	21	21
Lower East Side	14	14	14	14	14	14
Manhattan Valley	4	4	4	4	4	4
Manhattanville	2	2	2	2	2	2
Marble Hill	1	1	1	1	1	1
Midtown	50	50	50	50	50	50
Midtown South	50	50	50	50	50	50
Morningside Heights	6	6	6	6	6	6
Murray Hill	50	50	50	50	50	50
Noho	27	27	27	27	27	27
Roosevelt Island	1	1	1	1	1	1
Soho	21	21	21	21	21	21
Stuyvesant Town	7	7	7	7	7	7
Sutton Place	11	11	11	11	11	11
Tribeca	14	14	14	14	14	14
Tudor City	41	41	41	41	41	41
Turtle Bay	11	11	11	11	11	11
Upper East Side	2	2	2	2	2	2
West Village	15	15	15	15	15	15
Yorkville	2	2	2	2	2	2

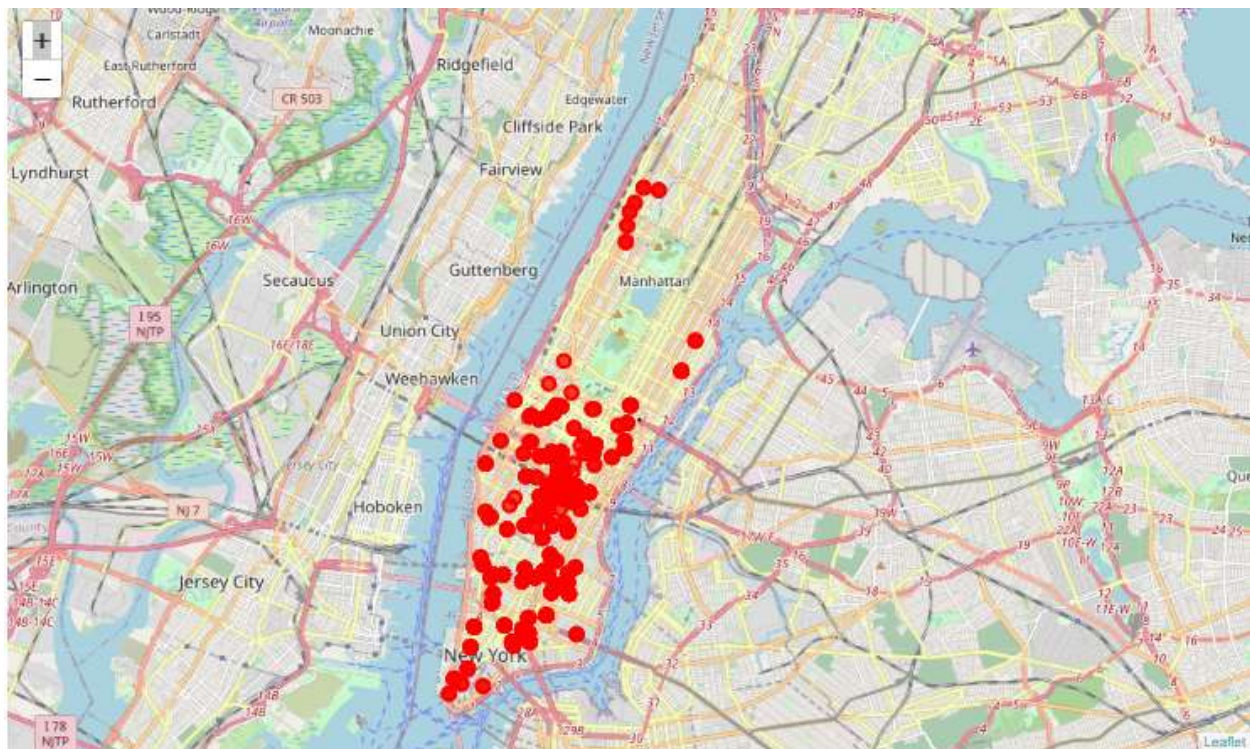
Methodology

Then using this feature to group the neighborhoods into clusters using K-means clustering algorithm will be used to complete this task. Also, the Folium library is used to visualize the neighborhoods in Manhattan and their emerging clusters.

Based on cluster analysis, we would be able to analyze the cluster areas which are the best places to open a new Korean restaurant.



Map of Manhattan, New York showing different neighborhoods.



Map showing various Korean restaurants in and around Manhattan, NY.

I have gone for K means algorithm for cluster analysis, particularly because:

Kmeans algorithm is an iterative algorithm that tries to partition the dataset into K pre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to **only one group**. It tries to make the intra-cluster data points as similar as possible while also keeping the clusters as different (far) as possible. It assigns data points to a cluster such that the sum of the squared distance between the data points and the cluster's centroid (arithmetic mean of all the data points that belong to that cluster) is at the minimum. The less variation we have within clusters, the more homogeneous (similar) the data points are within the same cluster.

The way Kmeans algorithm works is as follows:

1. Specify number of clusters K .
2. Initialize centroids by first shuffling the dataset and then randomly selecting K data points for the centroids without replacement.
3. Keep iterating until there is no change to the centroids. i.e. assignment of data points to clusters isn't changing.
 - Compute the sum of the squared distance between data points and all centroids.
 - Assign each data point to the closest cluster (centroid).
 - Compute the centroids for the clusters by taking the average of the all data points that belong to each cluster.

Cluster Label 0

```
In [60]: manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 0, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]
```

Out[60]:

	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
28	Battery Park City	Korean Restaurant	Poke Place	Deli / Bodega	Food Truck	Wings Joint	Food Stand	BBQ Joint	Cocktail Bar	Coffee Shop	Food & Drink Shop
29	Financial District	Korean Restaurant	Food Truck	Poke Place	Deli / Bodega	Wings Joint	Food Stand	BBQ Joint	Cocktail Bar	Coffee Shop	Food & Drink Shop

Cluster Label 1

```
In [61]: manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 1, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]
```

Out[61]:

	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	Marble Hill	Korean Restaurant	Wings Joint	Food Truck	BBQ Joint	Cocktail Bar	Coffee Shop	Deli / Bodega	Food & Drink Shop	Food Court	Food Stand
1	Chinatown	Korean Restaurant	Wings Joint	Food Truck	BBQ Joint	Cocktail Bar	Coffee Shop	Deli / Bodega	Food & Drink Shop	Food Court	Food Stand
5	Manhattanville	Korean Restaurant	Wings Joint	Food Truck	BBQ Joint	Cocktail Bar	Coffee Shop	Deli / Bodega	Food & Drink Shop	Food Court	Food Stand
8	Upper East Side	Korean Restaurant	Wings Joint	Food Truck	BBQ Joint	Cocktail Bar	Coffee Shop	Deli / Bodega	Food & Drink Shop	Food Court	Food Stand
9	Yorkville	Korean Restaurant	Wings Joint	Food Truck	BBQ Joint	Cocktail Bar	Coffee Shop	Deli / Bodega	Food & Drink Shop	Food Court	Food Stand
10	Lenox Hill	Korean Restaurant	Wings Joint	Food Truck	BBQ Joint	Cocktail Bar	Coffee Shop	Deli / Bodega	Food & Drink Shop	Food Court	Food Stand
11	Roosevelt Island	Korean Restaurant	Wings Joint	Food Truck	BBQ Joint	Cocktail Bar	Coffee Shop	Deli / Bodega	Food & Drink Shop	Food Court	Food Stand
17	Chelsea	Korean Restaurant	Cocktail Bar	Food Stand	Wings Joint	Food Truck	BBQ Joint	Coffee Shop	Deli / Bodega	Food & Drink Shop	Food Court
19	East Village	Korean Restaurant	Asian Restaurant	Food Truck	BBQ Joint	Cocktail Bar	Coffee Shop	Deli / Bodega	Food & Drink Shop	Food Court	Food Stand
20	Lower East Side	Korean Restaurant	Wings Joint	Food Truck	BBQ Joint	Cocktail Bar	Coffee Shop	Deli / Bodega	Food & Drink Shop	Food Court	Food Stand
30	Carnegie Hill	Korean Restaurant	Wings Joint	Food Truck	BBQ Joint	Cocktail Bar	Coffee Shop	Deli / Bodega	Food & Drink Shop	Food Court	Food Stand
31	Noho	Korean Restaurant	Asian Restaurant	Food Truck	BBQ Joint	Cocktail Bar	Coffee Shop	Deli / Bodega	Food & Drink Shop	Food Court	Food Stand
34	Sutton Place	Korean Restaurant	Wings Joint	Food Truck	BBQ Joint	Cocktail Bar	Coffee Shop	Deli / Bodega	Food & Drink Shop	Food Court	Food Stand
37	Stuyvesant Town	Korean Restaurant	Wings Joint	Food Truck	BBQ Joint	Cocktail Bar	Coffee Shop	Deli / Bodega	Food & Drink Shop	Food Court	Food Stand
38	Flatiron	Korean Restaurant	Fried Chicken Joint	Lounge	Food Court	Food Stand	BBQ Joint	Cocktail Bar	Coffee Shop	Deli / Bodega	Food & Drink Shop

Cluster Label 2

```
In [62]: manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 2, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]
```

Out[62]:

	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
15	Midtown	Korean Restaurant	Food Truck	Fried Chicken Joint	Food Court	Salad Place	BBQ Joint	Lounge	Coffee Shop	Deli / Bodega	Vegetarian / Vegan Restaurant
16	Murray Hill	Korean Restaurant	Food Truck	Fried Chicken Joint	Food Court	Salad Place	Lounge	Vegetarian / Vegan Restaurant	BBQ Joint	Cocktail Bar	Coffee Shop
18	Greenwich Village	Korean Restaurant	Asian Restaurant	Japanese Restaurant	Food Truck	Food Stand	BBQ Joint	Cocktail Bar	Coffee Shop	Deli / Bodega	Food & Drink Shop
21	Tribeca	Korean Restaurant	Asian Restaurant	Japanese Restaurant	Food Truck	Food Stand	BBQ Joint	Cocktail Bar	Coffee Shop	Deli / Bodega	Food & Drink Shop
22	Little Italy	Korean Restaurant	Asian Restaurant	Food Truck	BBQ Joint	Cocktail Bar	Coffee Shop	Deli / Bodega	Food & Drink Shop	Food Court	Food Stand
23	Soho	Korean Restaurant	Asian Restaurant	Japanese Restaurant	Food Truck	Food Stand	BBQ Joint	Cocktail Bar	Coffee Shop	Deli / Bodega	Food & Drink Shop
24	West Village	Korean Restaurant	Asian Restaurant	Japanese Restaurant	Food Truck	Food Stand	BBQ Joint	Cocktail Bar	Coffee Shop	Deli / Bodega	Food & Drink Shop
27	Gramercy	Korean Restaurant	Asian Restaurant	Miscellaneous Shop	Lounge	Vegetarian / Vegan Restaurant	Food Stand	Food Court	BBQ Joint	Cocktail Bar	Coffee Shop
32	Civic Center	Korean Restaurant	Asian Restaurant	Poke Place	Food Truck	Food Stand	BBQ Joint	Cocktail Bar	Coffee Shop	Deli / Bodega	Food & Drink Shop
33	Midtown South	Korean Restaurant	Fried Chicken Joint	Food Court	Salad Place	Lounge	Vegetarian / Vegan Restaurant	Food Truck	BBQ Joint	Cocktail Bar	Coffee Shop
35	Turtle Bay	Korean Restaurant	Food & Drink Shop	Vegetarian / Vegan Restaurant	Wings Joint	Food Stand	BBQ Joint	Cocktail Bar	Coffee Shop	Deli / Bodega	Food Court
36	Tudor City	Korean Restaurant	Vegetarian / Vegan Restaurant	Miscellaneous Shop	Lounge	Food & Drink Shop	Wings Joint	Food Court	BBQ Joint	Cocktail Bar	Coffee Shop

Cluster Label 3

```
In [63]: manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 3, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]
```

Out[63]:

	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
25	Manhattan Valley	Korean Restaurant	Wings Joint	Grocery Store	Food Stand	BBQ Joint	Cocktail Bar	Coffee Shop	Deli / Bodega	Food & Drink Shop	Food Court
26	Morningside Heights	Korean Restaurant	Wings Joint	Grocery Store	Food Stand	BBQ Joint	Cocktail Bar	Coffee Shop	Deli / Bodega	Food & Drink Shop	Food Court

Cluster Label 4

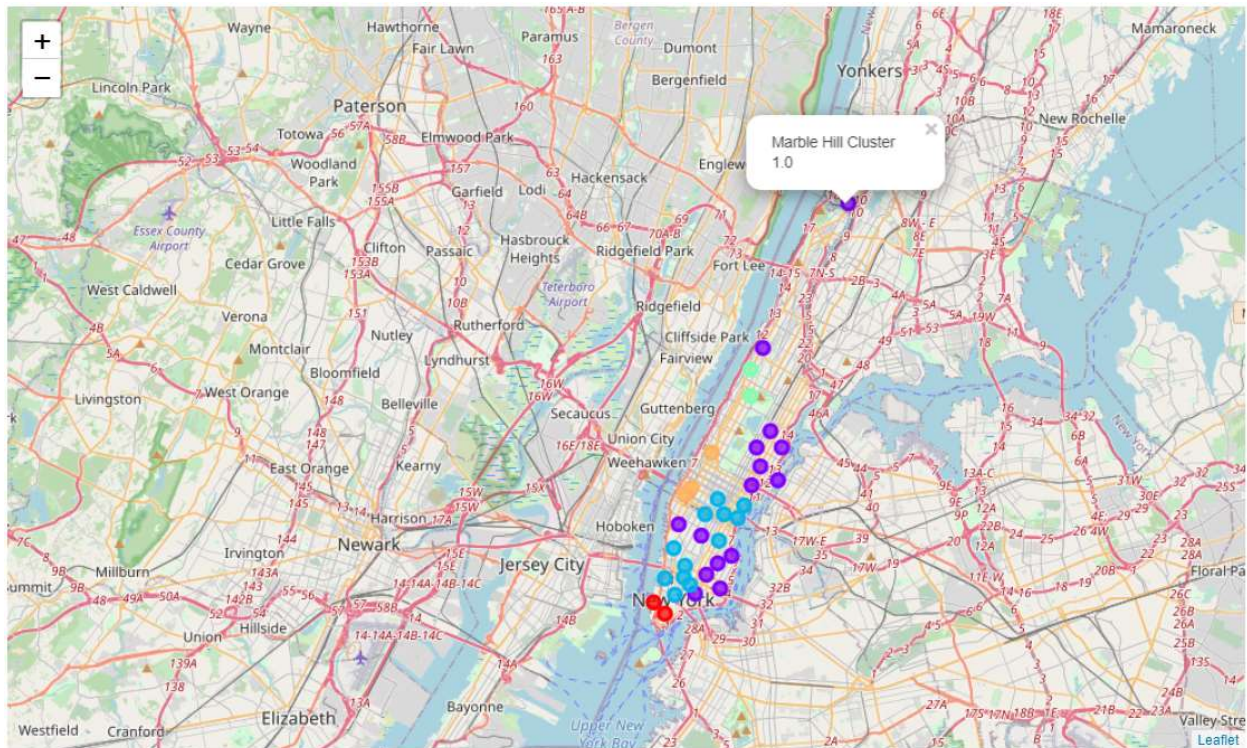
```
In [64]: manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 4, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]
```

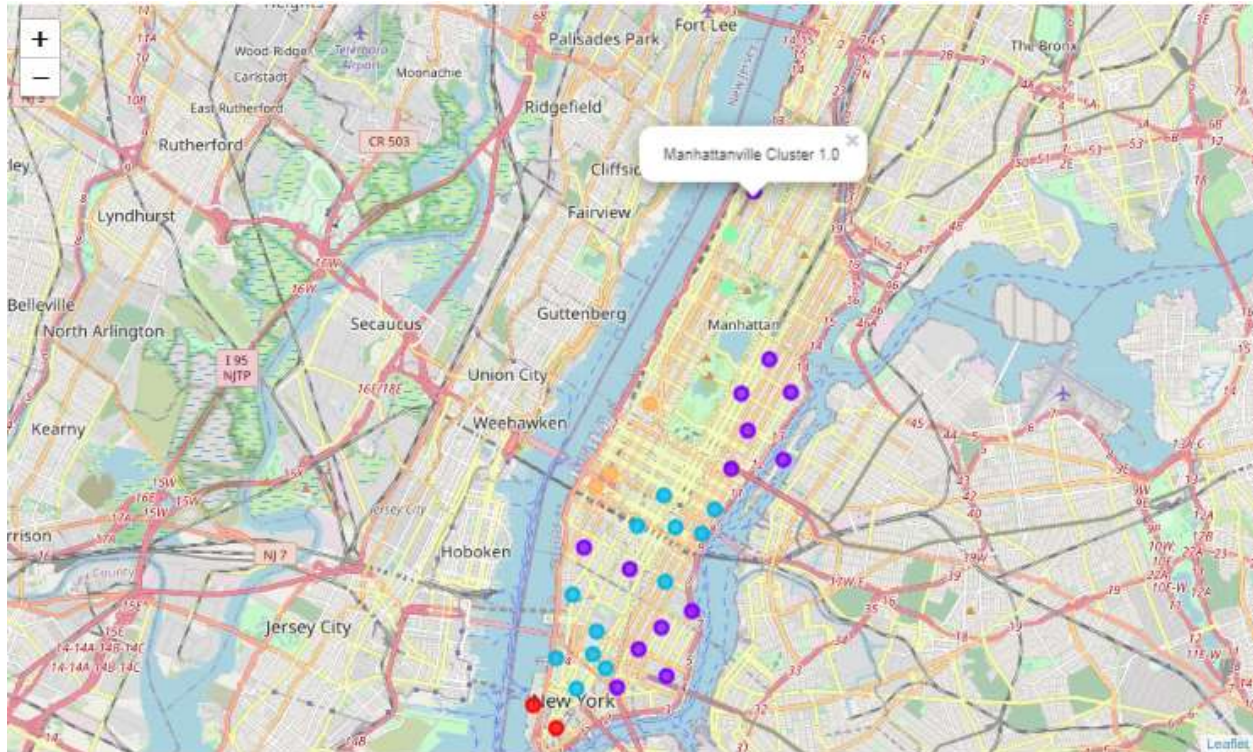
Out[64]:

	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
13	Lincoln Square	Korean Restaurant	Fried Chicken Joint	Coffee Shop	Food Truck	BBQ Joint	Cocktail Bar	Deli / Bodega	Food & Drink Shop	Food Court	Food Stand
14	Clinton	Korean Restaurant	Fried Chicken Joint	Food Stand	Salad Place	BBQ Joint	Coffee Shop	Food Truck	Cocktail Bar	Deli / Bodega	Food & Drink Shop
39	Hudson Yards	Korean Restaurant	Fried Chicken Joint	Food Stand	Salad Place	BBQ Joint	Coffee Shop	Cocktail Bar	Deli / Bodega	Food & Drink Shop	Food Court

Results

Maps showing clusters of neighborhoods having Korean Restaurants, using K-means algorithm with K=5.





As per the analysis, we see that the best places to open Korean restaurants would be Marble Hill and Manhattanville (Cluster 1.0), also Manhattan Valley (Cluster 3.0) could also be ideal neighborhood since these are away from the neighborhoods with large number of Korean Restaurants, hence lesser competition from these counterparts.

Discussion

- This analysis is performed on limited data. The results may vary and can depend on several factors such as population density in the area, transportation, demographics, etc. But if good amount of data is available there is scope to come up with better and more accurate results.
- There is high competition in Chinatown and Soho so it is very risky to open business in these areas.
- More detailed analysis by adding other factors such as population density, transportation, demographics of inhabitants.
- It can be inferred that Korean food is very popular cuisine in Manhattan, could be due to high influx of Korean population and that people appreciate the cuisine.
- So Korean food businesses are good business ventures and can generate good revenue if opened in best neighborhoods favoring the cuisine.

Conclusion

Through data science capstone project for **predicting the best neighborhoods for opening Korean restaurants in Manhattan, New York**, the required outcome is achieved and demonstrated. We have been successfully able to incorporate data science concepts such as K-

means algorithm for clustering, maps and Foursquare API to analyze and predict the neighborhoods which can be considered to open up the Korean restaurant business. Although all of the goals of this project were met there is definitely room for improvement and development.

Using this idea, one can determine the suitable locations for setting up any business model by analyzing the demographical, geographical and other various different data sources of the location.