CAPSTONE PROJECT - THE BATTLE OF NEIGHBOURHOODS – OPEN AN INDIAN RESTAURANT IN MANHATTAN (NYC)

NIKHIL KUMAR PUTTA

06 MAY 2020

1. Introduction & Business Problem

Background:

The City of New York is the most populous city in the United States. It is diverse and is the financial capital of USA. It is multicultural and provides lot of business opportunities with business-friendly environment. It has attracted many different players into the market. It is a global hub of business and commerce. The city is a major centre for banking and finance, retail, foreign trade, transportation, tourism, real estate, new & traditional media, advertising, legal services, accountancy, insurance, theatre, fashion, and the arts in the United States. This also means that the market is highly competitive. As it is highly developed city, so cost of doing business is also one of the highest. Thus, any new business venture or expansion needs to be analysed carefully. The insights derived from analysis will give good understanding of the business environment which help in strategically targeting the market. This will help in reduction of risk. And the Return on Investment will be reasonable.

Business Problem:

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

Indian restaurants have become so popular in the United States now it seems that there is one on every corner, not only in major cities but also in smaller cities. Starting an Indian restaurant can be a great business opportunity, as there are lot of Indians and Asians reside or travel to New York every day. Also, many Americans too fond of Indian cuisine. To succeed in this business, you need to distinguish yourself from others to enjoy long-term success.

Any restaurant will succeed only if it is strategically located. If you plan a real restaurant that can demand higher prices for Indian spices, focus on neighbourhoods and outlets that already attract a sophisticated Indian client. If you plan a cheap buffet restaurant, points to the masses looking for affordable high-traffic locations with large shopping centres and other local points of interest. So, it is evident that to survive in such competitive market it is very important to strategically plan. Various factors need to be studied in order to decide on the Location.

My Client wants to open his business in Manhattan area, so I only focus on that borough during my analysis. The objective is to locate and recommend to the management which neighbourhood of New York city will be best choice to start a restaurant. The Management also expects to understand the rationale of the recommendations made.

This would interest anyone who wants to start a new restaurant in New York City.

2. Data Acquisition & Cleaning

Source for New York City Data:

Neighbourhood has a total of 5 boroughs and 306 neighbourhoods. In order to segment the neighbourhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighbourhoods that exist in each borough as well as the latitude and longitude coordinates of each neighbourhood. This dataset exists for free on the web.

Link to the dataset is: https://geo.nyu.edu/catalog/nyu_2451_34572

Source for Venues Data:

New York city geographical coordinates data will be utilized as input for the Foursquare API, that will be leveraged to provision venues information for each neighbourhood. We will use the Foursquare API to explore neighbourhoods in New York City.

In addition, **Indian Restaurant category Id 4bf58dd8d48988d10f941735** is used for retrieving data from Foursquare API.

Data Cleaning & Feature Selection:

There were many unnecessary features in the data downloaded. These unnecessary features have to be dropped from the data and retain only Borough, Neighbourhood, Latitude and Longitude features.

Then we perform Data Wrangling to remove any missing values.

```
neighborhoods.head()
```

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

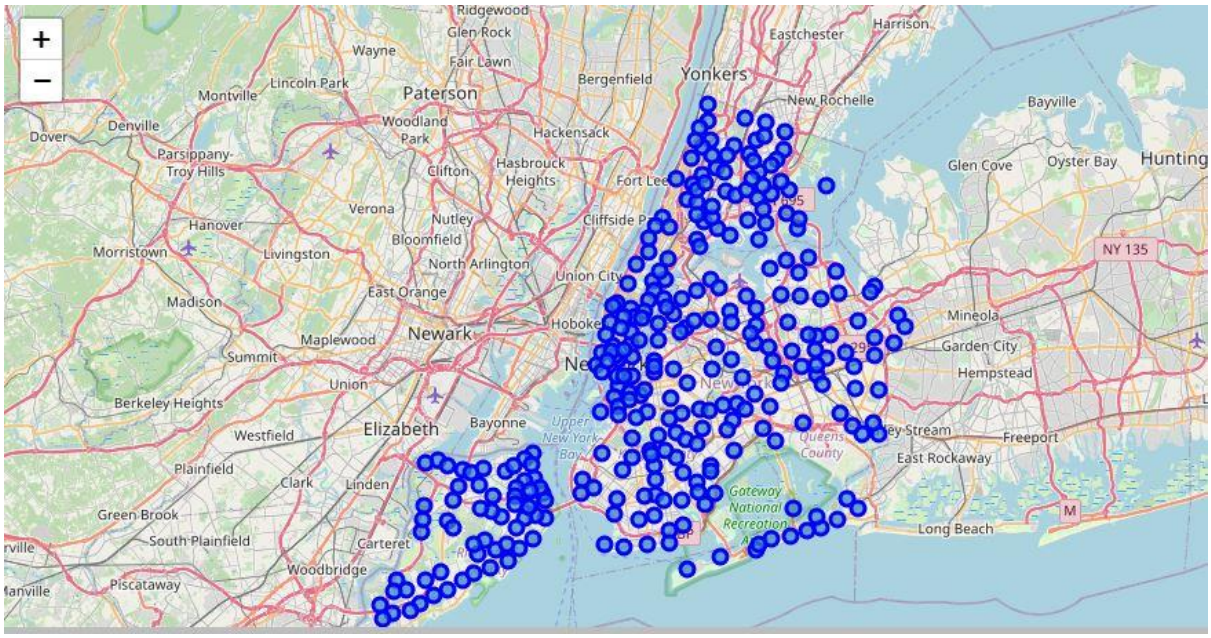
3. Methodology – Exploratory Data Analysis

In this project, I will use the basic methodology as taught in Week 3 lab of this course.

At first, we import all necessary libraries required for this project which includes NumPy, Pandas, Json, GeoPy, Matplotlib, KMeans, Folium and others.

Then I have downloaded the New York city data which is in Json format and then transformed it into Pandas Data frame by selecting only Borough, Neighbourhood, Latitude and Longitude features and dropping all other features.

Now, created a leaflet map of New York city neighbourhoods using latitude and longitude values.

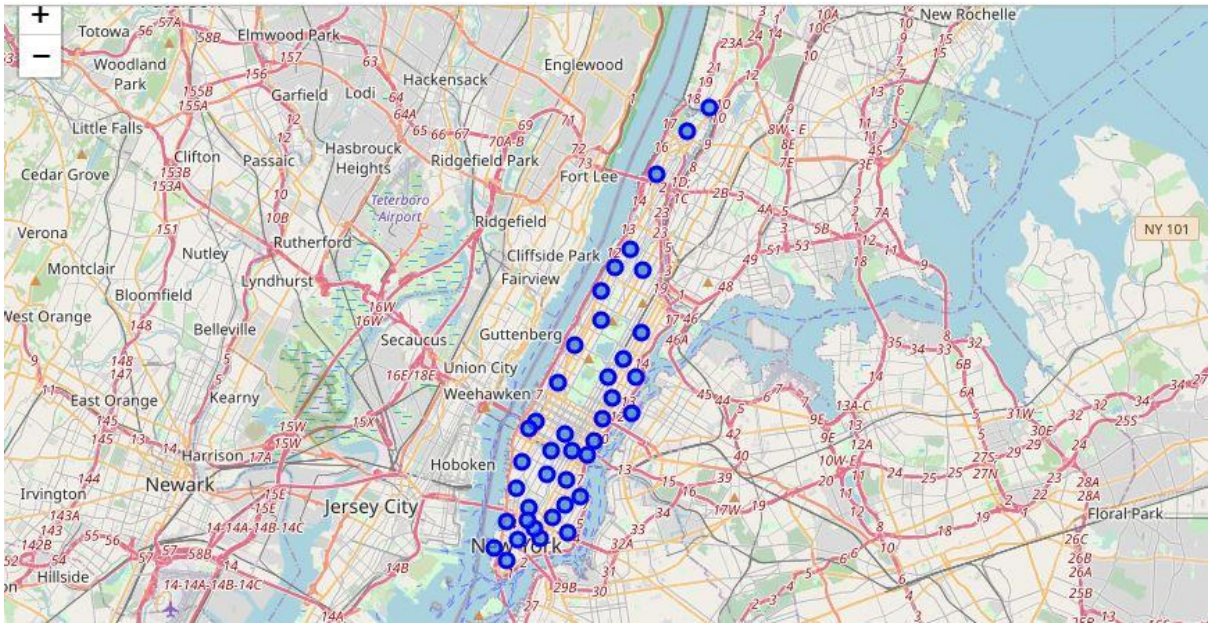


As my client is concerned only with Manhattan borough, I will create a new data frame with Manhattan neighbourhood data.

```
manhattan_data.head()
```

	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

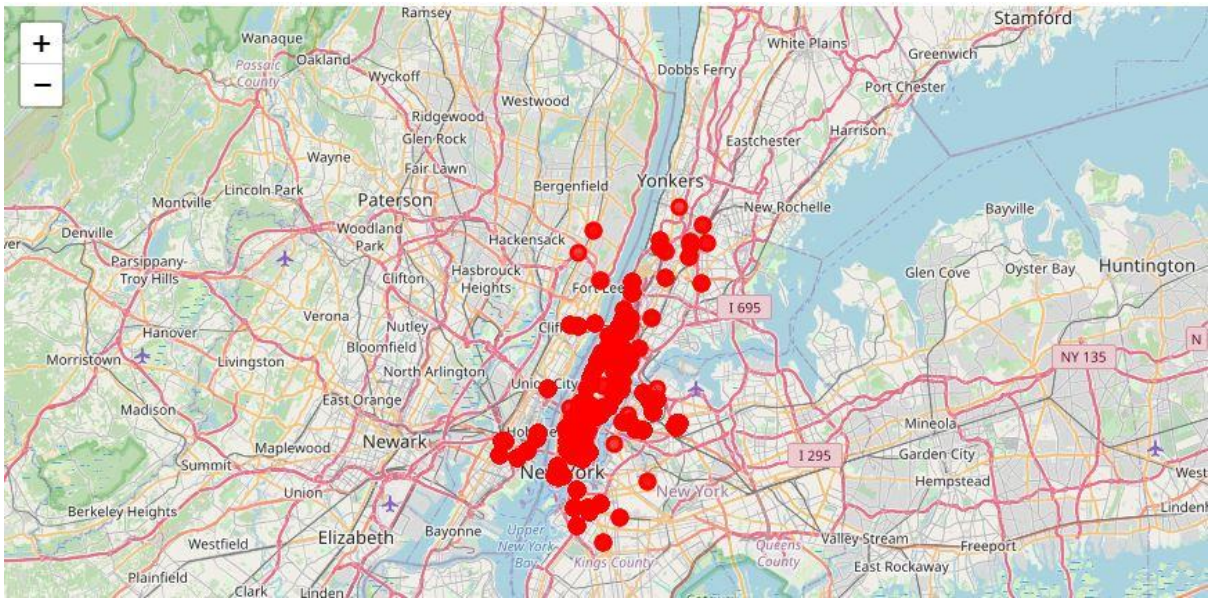
Now, created a leaflet map of Manhattan neighbourhoods using latitude and longitude values.



Then, I have proceeded with getting venues data through Foursquare API. Initially, we created an API request URL, then made a GET request to get venues data and finally filtered to get relevant data to get details of Indian restaurant categories in each neighbourhood of Manhattan.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Riverdale Indian Cuisine	40.880886	-73.908800	Indian Restaurant
1	Marble Hill	40.876551	-73.91066	Aman Restaurant	40.885174	-73.879550	Indian Restaurant
2	Marble Hill	40.876551	-73.91066	Agra Grill	40.884713	-73.862301	Indian Restaurant
3	Marble Hill	40.876551	-73.91066	Running Cool Restaurant	40.858264	-73.904341	Caribbean Restaurant
4	Marble Hill	40.876551	-73.91066	Delhi Masala Express	40.834512	-73.944967	Indian Restaurant

Now, I have created a leaflet map of all Indian restaurant categories using latitude and longitude values.



Then, I started with analysing each neighbourhood by exploring venues in each neighbourhood and filtering out 10 most common venues in each neighbourhood.

	Neighborhood	Asian Restaurant	Caribbean Restaurant	Chinese Restaurant	Coffee Shop	Deli / Bodega	Food Truck	Grocery Store	Indian Restaurant	North Indian Restaurant	Snack Place	South Indian Restaurant	Tibetan Restaurant
0	Marble Hill	0	0	0	0	0	0	0	1	0	0	0	0
1	Marble Hill	0	0	0	0	0	0	0	1	0	0	0	0
2	Marble Hill	0	0	0	0	0	0	0	1	0	0	0	0
3	Marble Hill	0	1	0	0	0	0	0	0	0	0	0	0
4	Marble Hill	0	0	0	0	0	0	0	1	0	0	0	0

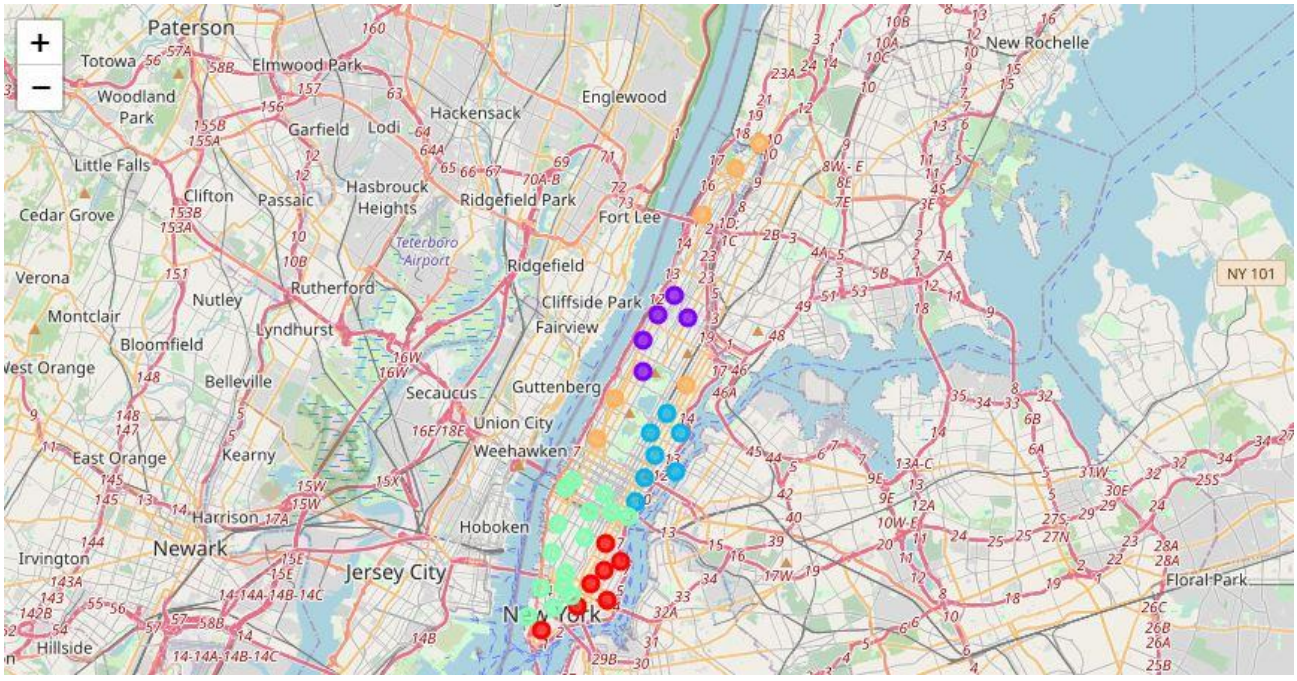
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Battery Park City	Indian Restaurant	South Indian Restaurant	Food Truck	North Indian Restaurant	Grocery Store	Deli / Bodega	Chinese Restaurant	Tibetan Restaurant	Snack Place	Coffee Shop
1	Carnegie Hill	Indian Restaurant	Tibetan Restaurant	South Indian Restaurant	North Indian Restaurant	Deli / Bodega	Chinese Restaurant	Asian Restaurant	Snack Place	Grocery Store	Food Truck
2	Central Harlem	Indian Restaurant	Tibetan Restaurant	South Indian Restaurant	Snack Place	North Indian Restaurant	Grocery Store	Food Truck	Deli / Bodega	Coffee Shop	Chinese Restaurant
3	Chelsea	Indian Restaurant	South Indian Restaurant	Food Truck	North Indian Restaurant	Grocery Store	Deli / Bodega	Tibetan Restaurant	Snack Place	Coffee Shop	Chinese Restaurant
4	Chinatown	Indian Restaurant	South Indian Restaurant	Food Truck	North Indian Restaurant	Grocery Store	Deli / Bodega	Coffee Shop	Chinese Restaurant	Tibetan Restaurant	Snack Place

Clustering Neighbourhoods using K-means clustering Algorithm:

Then, I have used these features to group the neighbourhoods into clusters by applying K-means clustering algorithm to complete this task. And also, the Folium library to visualize the neighbourhoods in Manhattan and its emerging clusters.

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	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	Manhattan	Marble Hill	40.876551	-73.910660	4	Indian Restaurant	Caribbean Restaurant	Tibetan Restaurant	South Indian Restaurant	Snack Place	North Indian Restaurant	Grocery Store	Food Truck	Deli / Bodega	Coffee Shop
1	Manhattan	Chinatown	40.715618	-73.994279	0	Indian Restaurant	South Indian Restaurant	Food Truck	North Indian Restaurant	Grocery Store	Deli / Bodega	Coffee Shop	Chinese Restaurant	Tibetan Restaurant	Snack Place
2	Manhattan	Washington Heights	40.851903	-73.936900	4	Indian Restaurant	Food Truck	Caribbean Restaurant	Tibetan Restaurant	South Indian Restaurant	Snack Place	North Indian Restaurant	Grocery Store	Deli / Bodega	Coffee Shop
3	Manhattan	Inwood	40.867684	-73.921210	4	Indian Restaurant	Caribbean Restaurant	Tibetan Restaurant	South Indian Restaurant	Snack Place	North Indian Restaurant	Grocery Store	Food Truck	Deli / Bodega	Coffee Shop
4	Manhattan	Hamilton Heights	40.823604	-73.949688	1	Indian Restaurant	Tibetan Restaurant	South Indian Restaurant	Snack Place	North Indian Restaurant	Grocery Store	Food Truck	Deli / Bodega	Coffee Shop	Chinese Restaurant

Now, I have created a leaflet map adding different markers to the data to show each cluster with different colour on the map.



4. Results Section

All neighbourhoods in Manhattan borough are segregated into five clusters using K-Means Clustering algorithm. Below images highlight ten most common venues for each neighbourhood in each cluster.

Cluster 1:

	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
1	Chinatown	Indian Restaurant	South Indian Restaurant	Food Truck	North Indian Restaurant	Grocery Store	Deli / Bodega	Coffee Shop	Chinese Restaurant	Tibetan Restaurant	Snack Place
19	East Village	Indian Restaurant	South Indian Restaurant	Food Truck	Chinese Restaurant	North Indian Restaurant	Grocery Store	Deli / Bodega	Tibetan Restaurant	Snack Place	Coffee Shop
20	Lower East Side	Indian Restaurant	South Indian Restaurant	Food Truck	Chinese Restaurant	North Indian Restaurant	Grocery Store	Deli / Bodega	Coffee Shop	Tibetan Restaurant	Snack Place
27	Gramercy	Indian Restaurant	South Indian Restaurant	Food Truck	Chinese Restaurant	North Indian Restaurant	Grocery Store	Deli / Bodega	Tibetan Restaurant	Snack Place	Coffee Shop
29	Financial District	Indian Restaurant	South Indian Restaurant	Food Truck	North Indian Restaurant	Grocery Store	Deli / Bodega	Coffee Shop	Chinese Restaurant	Tibetan Restaurant	Snack Place
31	Noho	Indian Restaurant	South Indian Restaurant	Food Truck	Chinese Restaurant	North Indian Restaurant	Grocery Store	Deli / Bodega	Tibetan Restaurant	Snack Place	Coffee Shop
37	Stuyvesant Town	Indian Restaurant	South Indian Restaurant	Food Truck	Chinese Restaurant	North Indian Restaurant	Grocery Store	Deli / Bodega	Tibetan Restaurant	Snack Place	Coffee Shop

Cluster 2:

	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
4	Hamilton Heights	Indian Restaurant	Tibetan Restaurant	South Indian Restaurant	Snack Place	North Indian Restaurant	Grocery Store	Food Truck	Deli / Bodega	Coffee Shop	Chinese Restaurant
5	Manhattanville	Indian Restaurant	Food Truck	Tibetan Restaurant	South Indian Restaurant	Snack Place	North Indian Restaurant	Grocery Store	Deli / Bodega	Coffee Shop	Chinese Restaurant
6	Central Harlem	Indian Restaurant	Tibetan Restaurant	South Indian Restaurant	Snack Place	North Indian Restaurant	Grocery Store	Food Truck	Deli / Bodega	Coffee Shop	Chinese Restaurant
25	Manhattan Valley	Indian Restaurant	South Indian Restaurant	North Indian Restaurant	Tibetan Restaurant	Snack Place	Grocery Store	Food Truck	Deli / Bodega	Coffee Shop	Chinese Restaurant
26	Morningside Heights	Indian Restaurant	Food Truck	Tibetan Restaurant	South Indian Restaurant	Snack Place	North Indian Restaurant	Grocery Store	Deli / Bodega	Coffee Shop	Chinese Restaurant

Cluster 3:

	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
8	Upper East Side	Indian Restaurant	South Indian Restaurant	Tibetan Restaurant	North Indian Restaurant	Deli / Bodega	Chinese Restaurant	Asian Restaurant	Snack Place	Grocery Store	Food Truck
9	Yorkville	Indian Restaurant	Tibetan Restaurant	South Indian Restaurant	Snack Place	North Indian Restaurant	Deli / Bodega	Chinese Restaurant	Asian Restaurant	Grocery Store	Food Truck
10	Lenox Hill	Indian Restaurant	South Indian Restaurant	Tibetan Restaurant	North Indian Restaurant	Deli / Bodega	Chinese Restaurant	Asian Restaurant	Snack Place	Grocery Store	Food Truck
11	Roosevelt Island	Indian Restaurant	South Indian Restaurant	Tibetan Restaurant	North Indian Restaurant	Deli / Bodega	Chinese Restaurant	Asian Restaurant	Snack Place	Grocery Store	Food Truck
30	Carnegie Hill	Indian Restaurant	Tibetan Restaurant	South Indian Restaurant	North Indian Restaurant	Deli / Bodega	Chinese Restaurant	Asian Restaurant	Snack Place	Grocery Store	Food Truck
34	Sutton Place	Indian Restaurant	South Indian Restaurant	Tibetan Restaurant	North Indian Restaurant	Deli / Bodega	Chinese Restaurant	Asian Restaurant	Snack Place	Grocery Store	Food Truck
35	Turtle Bay	Indian Restaurant	South Indian Restaurant	Tibetan Restaurant	North Indian Restaurant	Deli / Bodega	Chinese Restaurant	Asian Restaurant	Snack Place	Grocery Store	Food Truck

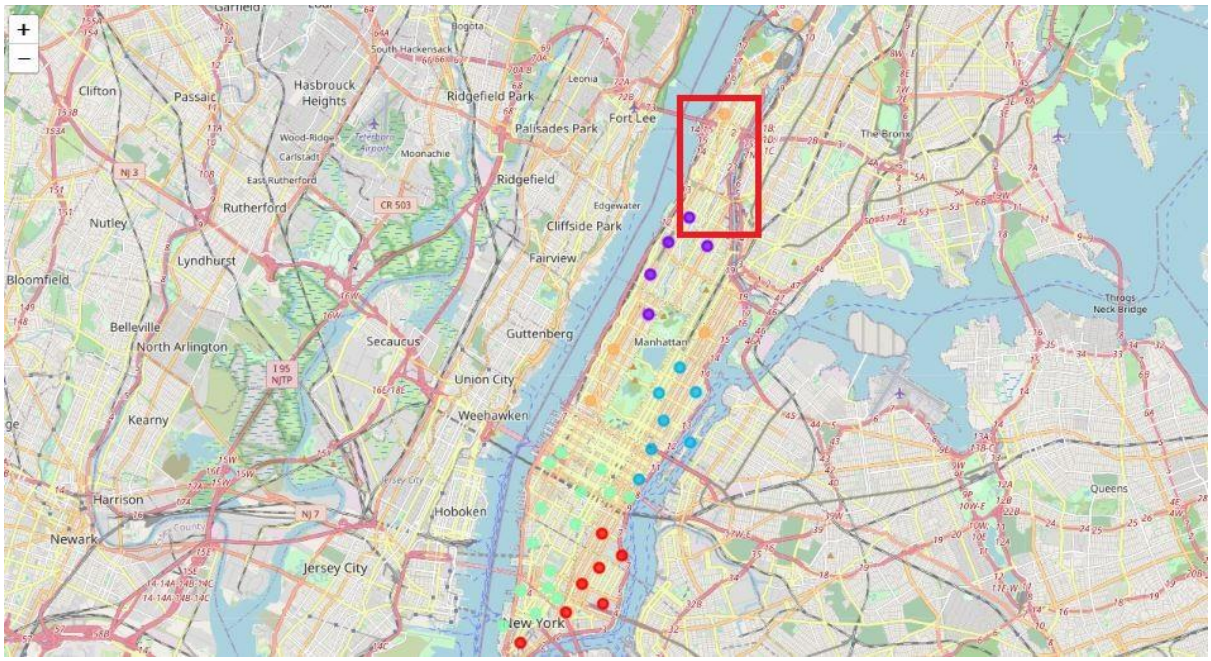
Cluster 4:

	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
14	Clinton	Indian Restaurant	South Indian Restaurant	Food Truck	North Indian Restaurant	Grocery Store	Deli / Bodega	Tibetan Restaurant	Snack Place	Coffee Shop	Chinese Restaurant
15	Midtown	Indian Restaurant	South Indian Restaurant	Food Truck	North Indian Restaurant	Grocery Store	Deli / Bodega	Chinese Restaurant	Tibetan Restaurant	Snack Place	Coffee Shop
16	Murray Hill	Indian Restaurant	South Indian Restaurant	Food Truck	North Indian Restaurant	Grocery Store	Deli / Bodega	Chinese Restaurant	Tibetan Restaurant	Snack Place	Coffee Shop
17	Chelsea	Indian Restaurant	South Indian Restaurant	Food Truck	North Indian Restaurant	Grocery Store	Deli / Bodega	Tibetan Restaurant	Snack Place	Coffee Shop	Chinese Restaurant
18	Greenwich Village	Indian Restaurant	South Indian Restaurant	Food Truck	North Indian Restaurant	Grocery Store	Deli / Bodega	Chinese Restaurant	Tibetan Restaurant	Snack Place	Coffee Shop
21	Tribeca	Indian Restaurant	South Indian Restaurant	Food Truck	North Indian Restaurant	Grocery Store	Deli / Bodega	Chinese Restaurant	Tibetan Restaurant	Snack Place	Coffee Shop
22	Little Italy	Indian Restaurant	South Indian Restaurant	Food Truck	North Indian Restaurant	Grocery Store	Deli / Bodega	Chinese Restaurant	Tibetan Restaurant	Snack Place	Coffee Shop
23	Soho	Indian Restaurant	South Indian Restaurant	Food Truck	North Indian Restaurant	Grocery Store	Deli / Bodega	Chinese Restaurant	Tibetan Restaurant	Snack Place	Coffee Shop
24	West Village	Indian Restaurant	South Indian Restaurant	Food Truck	North Indian Restaurant	Grocery Store	Deli / Bodega	Tibetan Restaurant	Snack Place	Coffee Shop	Chinese Restaurant
28	Battery Park City	Indian Restaurant	South Indian Restaurant	Food Truck	North Indian Restaurant	Grocery Store	Deli / Bodega	Chinese Restaurant	Tibetan Restaurant	Snack Place	Coffee Shop
32	Civic Center	Indian Restaurant	South Indian Restaurant	Food Truck	North Indian Restaurant	Grocery Store	Deli / Bodega	Chinese Restaurant	Tibetan Restaurant	Snack Place	Coffee Shop
33	Midtown South	Indian Restaurant	South Indian Restaurant	Food Truck	North Indian Restaurant	Grocery Store	Deli / Bodega	Chinese Restaurant	Tibetan Restaurant	Snack Place	Coffee Shop
36	Tudor City	Indian Restaurant	South Indian Restaurant	Food Truck	North Indian Restaurant	Deli / Bodega	Chinese Restaurant	Tibetan Restaurant	Snack Place	Grocery Store	Coffee Shop
38	Flatiron	Indian Restaurant	South Indian Restaurant	Food Truck	North Indian Restaurant	Grocery Store	Deli / Bodega	Chinese Restaurant	Tibetan Restaurant	Snack Place	Coffee Shop
39	Hudson Yards	Indian Restaurant	South Indian Restaurant	Food Truck	North Indian Restaurant	Grocery Store	Deli / Bodega	Tibetan Restaurant	Snack Place	Coffee Shop	Chinese Restaurant

Cluster 5:

	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	Indian Restaurant	Caribbean Restaurant	Tibetan Restaurant	South Indian Restaurant	Snack Place	North Indian Restaurant	Grocery Store	Food Truck	Deli / Bodega	Coffee Shop
2	Washington Heights	Indian Restaurant	Food Truck	Caribbean Restaurant	Tibetan Restaurant	South Indian Restaurant	Snack Place	North Indian Restaurant	Grocery Store	Deli / Bodega	Coffee Shop
3	Inwood	Indian Restaurant	Caribbean Restaurant	Tibetan Restaurant	South Indian Restaurant	Snack Place	North Indian Restaurant	Grocery Store	Food Truck	Deli / Bodega	Coffee Shop
7	East Harlem	Indian Restaurant	Tibetan Restaurant	Snack Place	Asian Restaurant	South Indian Restaurant	North Indian Restaurant	Grocery Store	Food Truck	Deli / Bodega	Coffee Shop
12	Upper West Side	Indian Restaurant	South Indian Restaurant	North Indian Restaurant	Deli / Bodega	Tibetan Restaurant	Snack Place	Grocery Store	Food Truck	Coffee Shop	Chinese Restaurant
13	Lincoln Square	Indian Restaurant	South Indian Restaurant	North Indian Restaurant	Deli / Bodega	Tibetan Restaurant	Snack Place	Grocery Store	Food Truck	Coffee Shop	Chinese Restaurant

All Clusters Manhattan Indian Restaurants Map:



Two points in highlighted region are as below.

- Indian restaurant at Hamilton Heights (violet marker)
- Indian restaurant at Washington Heights (orange marker)

5. Discussion Section

In this section, I would be discussing the observations that I have noted and the recommendation that I can make based on the results.

This analysis is performed on limited data (considering client is interested only in Manhattan borough). The reason behind using K-Means Clustering algorithm is to form clusters in such a way that similar samples go into a cluster, and dissimilar samples fall into different clusters.

- From the above clusters data and Indian Restaurants map in Manhattan borough, we can say that, in Cluster 2 & Cluster 5, we have very few Indian restaurants in 10 most common venues when compared to other clusters.
- South Indian restaurant as 2nd most common venue both in Cluster 1 & Cluster 3.
- Competition in very less between Morningside Heights and East Harlem in Central Manhattan, Sutton Place and Hell's Kitchen in Midtown Manhattan.

Recommendation:

As there are no Indian restaurants situated between Hamilton Heights in Cluster 2 and Washington Heights in Cluster 5, I recommend opening a new Indian Restaurant in these surroundings. Also, the other reason behind this recommendation is these neighbourhoods are less costly than Central Manhattan or Midtown Manhattan.

Although, this data analysis gives definite insights for the client, it would have been even better if we do more detailed analysis by adding other factors such as availability of transportation hubs nearby, demographics of inhabitants, cost of establishment and others.

6. Conclusion

In this study, I performed data wrangling, feature selection, exploratory data analysis and applied unsupervised machine learning (K-Means) technique on data which I have learned in IBM Data Science Professional Certificate course to solve a real-world problem of finding the best place to set-up an Indian restaurant in New York City.

Although all of the goals of this project were met through my analysis, definitely there is a scope for further improvement.