DATA SCIENCE CAPSTONE PROJECT

Introduction:

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. It provides lot of business opportunities and 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 center for banking and finance, retailing, world trade, transportation, tourism, real estate, new media, traditional media, advertising, legal services, accountancy, insurance, cinema hall, 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. The food culture includes an array of international cuisines influenced by the city's immigrant history. Indian restaurant 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 a Indian restaurant can be a great business opportunity, but you need to distinguish yourself from others to enjoy long-term success.

If you plan a real restaurant that can demand higher prices for fresh Indian Curry and fresh Indian spices delivered daily from India, 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 centers and other local points of interest.

Client Requirement :

My client wants to open his business in Manhattan area, so I focus on that borough during my analysis. We define potential neighbourhood based on the number of Indian restaurant which are operating right in each neighbourhood. Manhattan has full potential but also is a very challenging district to open a business because of high competition.
New Indian restaurant can be open in an area with inadequate number of Indian restaurants in the neighbourhood.
In this way the restaurant can attract more customers. Therefore, this analysis necessary to ensure that we have enough customers and that we are not so close to other Indian restaurant.

Data

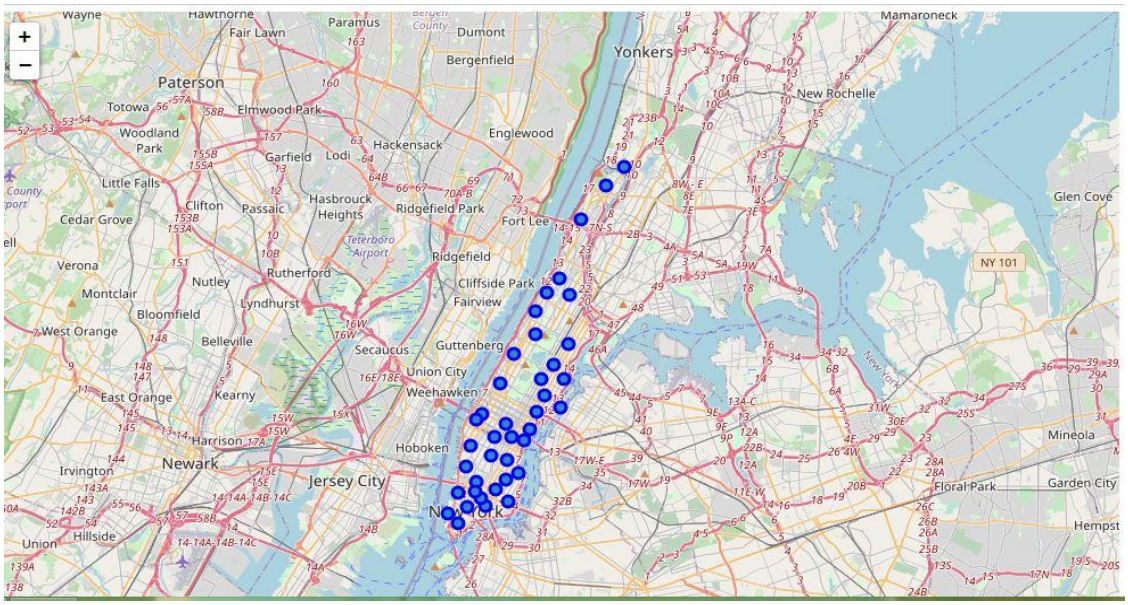
Neighbourhood has a total of 5 boroughs and 306 neighbourhoods. In order to segment the neighbourhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighbourhoods that exist in each borough as well as the latitude and longitude coordinates of each neighbourhood. This dataset exists for free on the web. Link to the dataset is: https://geo.nyu.edu/catalog/nyu_2451_34572

	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

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. The below is image of the Foursquare API data for the Manhattan city.

	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

These are the Neighbourhoods in Manhattan city on the Folium Map

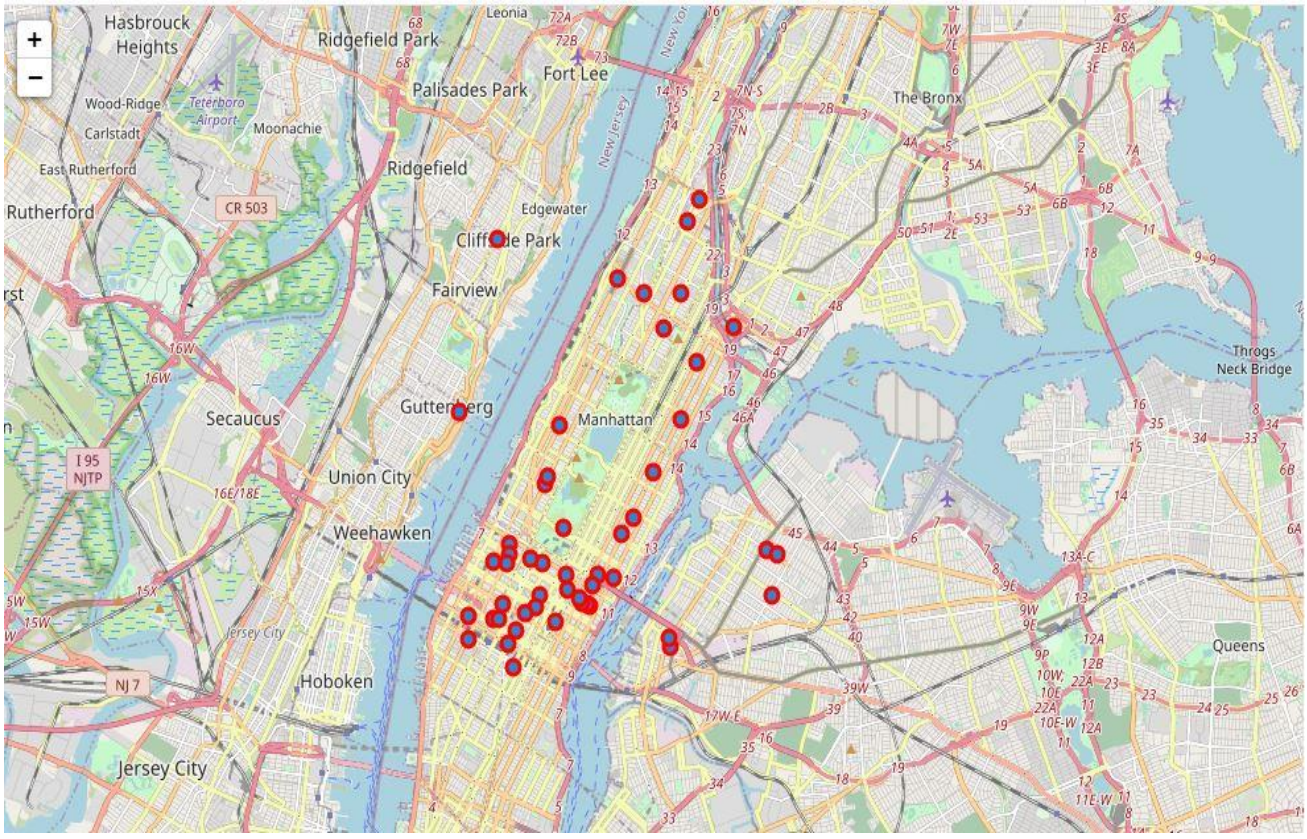


Methodology:

Convert addresses into their equivalent latitude and longitude values. Then we will use the Foursquare API to explore neighbourhoods in Manhattan, New York. After that, explore function to get Indian Restaurant categories in each neighbourhood.

	Venue	Latitude	Longitude
0	Indian Restaurant	40.750702	-73.998100
1	Indian fast food cart	40.759469	-73.972256
2	Indian Project - Broadway Bites	40.749850	-73.987908
3	Indian Summer	40.805740	-73.947320
4	Indian Tanpura	40.788515	-73.974405

These are Indian Restaurants in the Manhattan city.



Model Used:

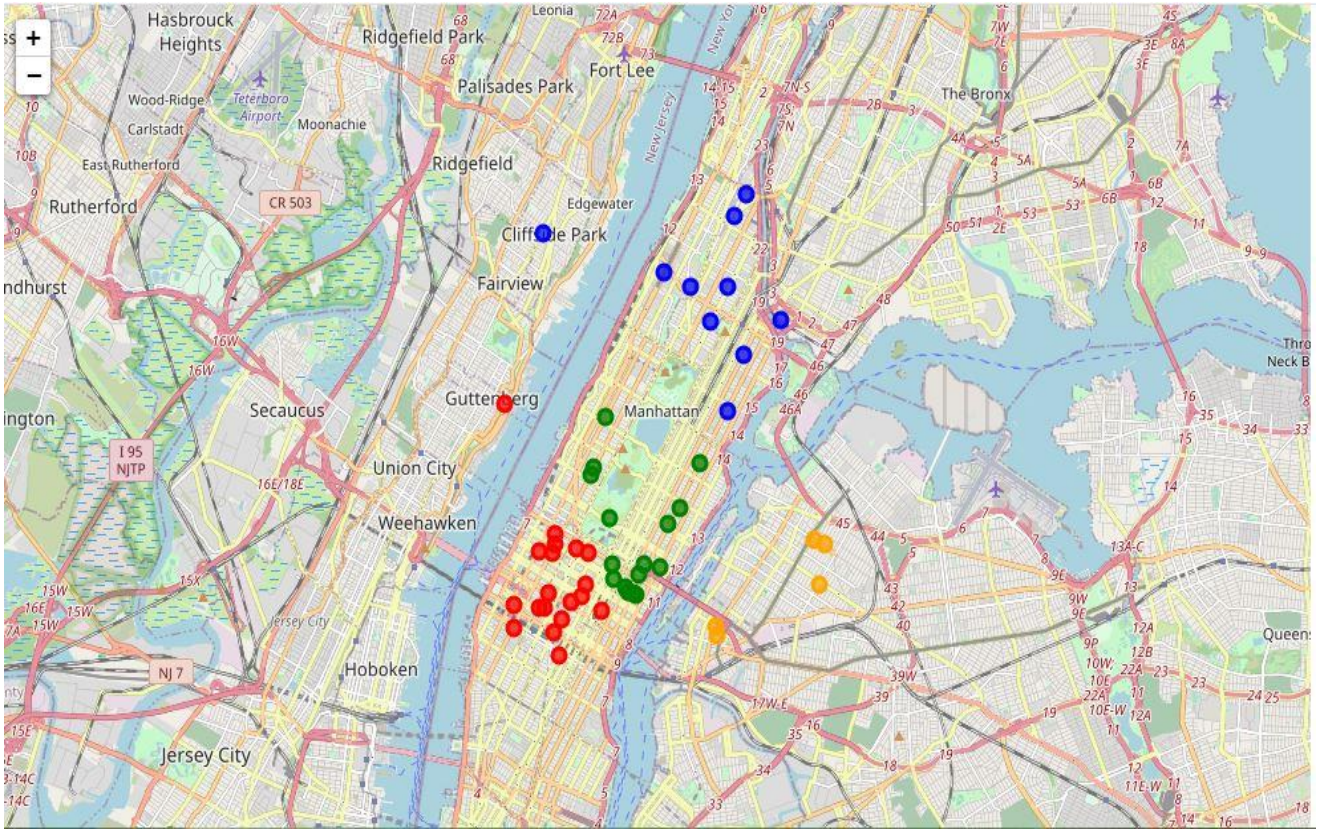
We used K Means machine learning algorithm to cluster the restaurants.

Then use the feature to group the neighbourhoods into clusters K-means clustering algorithm will be use to complete this task. And also, the Folium library to visualize the neighbourhoods in Manhattan and its emerging clusters.

We merged the cluster labels with the Venue and their respective coordinates in the single table.

We then applied the Folium Library to show the Clusters on the Folium Map

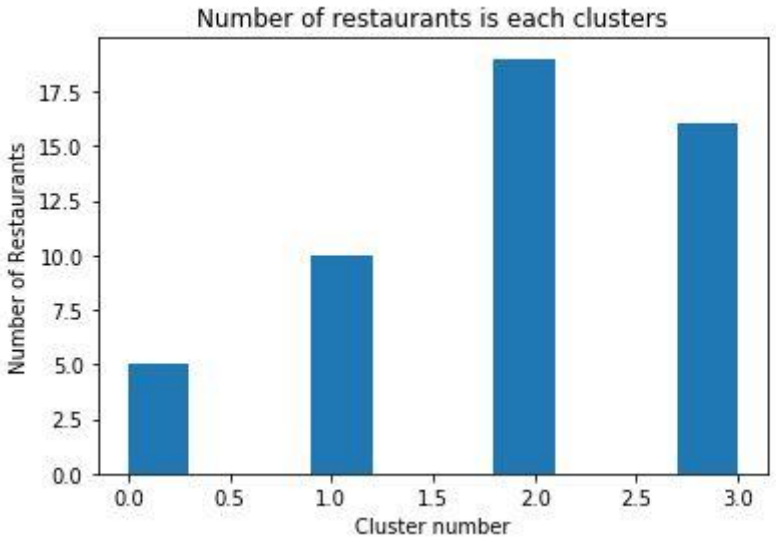
	Labels	Venue	Latitude	Longitude
0	2	Indian Restaurant	40.750702	-73.998100
1	3	Indian fast food cart	40.759469	-73.972256
2	2	Indian Project - Broadway Bites	40.749850	-73.987908
3	1	Indian Summer	40.805740	-73.947320
4	3	Indian Tanpura	40.788515	-73.974405



The clusters are represented on Folium Map.

Results:

From the clusters we visualized that most of the Indian Restaurants are located in Cluster 2.



Discussion

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

This analysis is performed on limited data. This may be right or may be wrong. But if good amount of data is available there is scope to come up with better results.

• There is high competition in 7th Avenue and Lexinton Avenue so it is very risky to open business in these areas.

• It can be done more detailed analysis by adding other factors such as transportation, demographics of inhabitants.

Finally, Four Square proved to be a good source of data but frustrating at times. Despite having a Developer account I regularly exceeded my hourly limit locking me out for the day.

Conclusion:

- Most of the Indian Restaurants are concentrated in the east of Manhattan city with the highest number in cluster 2 and moderate number in cluster 3.
- On the other hand, cluster 0 has very low number to totally no Indian restaurants in the neighbourhoods. This represents a great opportunity and high potential areas to open new Indian Restaurants as there is very little to no competition from existing restaurants.
- Meanwhile, the restaurants in cluster 2 are likely suffering from intense competition due to oversupply and high concentration of restaurants. From another perspective, this also shows that the oversupply of Indian restaurants mostly happened due to more Indians living in the premises.
- Therefore, this project recommends the client to capitalize on these findings to open new Indian Restaurant in neighbourhoods in cluster 0 with little to no competition.
- Clients with unique selling propositions to stand out from the competition can also open new Indian Restaurants in neighbourhoods in cluster 1 with moderate competition.
- Lastly, clients are advised to avoid neighbourhoods in cluster 2 which already have high concentration of Indian Restaurants and suffering from intense competition.