Introduction/Business Problem

New York is the most populous city in the United States. With an estimated 2018 population of 8.5 million. New York is also considered as one of the most multi-cultural city where the city's population in 2010 was 44% white (33.3% non-Hispanic white), 25.5% black (23% non-Hispanic black), 0.7% Native American, and 12.7% Asian.

Asian Americans in New York City, according to the 2010 Census, number more than one million, which is ever rising. The mentioned population and race/ethnicity figures do not include hundreds of thousands of tourists (Specifically, Asian origin).

Born to immigrant parents in United States and having developed strong Asian (Indian/Pakistani/Bengal) taste, liking and disliking – many residents in New York have problems finding Asian restaurants, if and when they find one, how to figure out if they are worth the effort of travelling across New York city in humongous traffic and busy routes.

New York City is often referred to collectively as the five boroughs. There are hundreds of distinct neighborhoods throughout the boroughs

This project will determine which boroughs have the highest number of Asian restaurants, how many have good ratings and how can the results of this project help the residents and of course, the visitors decide and choose their next dinner/lunch.

Data Collection and Cleaning

New York Latitude and Longitude data along with the details including population density of each Borough/Neighborhood will be extracted from the below data source link.

https://cocl.us/new_york_dataset

I will be using the Foursquare API to get the restaurant data in each borough which will be cleansed/shortlisted to Asian restaurants and then their respective ratings.

I will use Google Map, 'Search Nearby' option to get the center coordinates of the each Borough.

The following data is precisely used including list of Boroughs and neighborhoods of New York with their geodata (latitude and longitude), list of Asian (Indian) restaurants in each neighborhood and ratings for each restaurant.

Methodology

Once we have the required data, we will be comparing the restaurants in every Borough and then among the ones in each Borough top ratings (above 8.0) will be used to graph for better visualization.

Then the strategy is to map this information on the map in order to make it much more simple and easy to visualize so that decision making is quick and efficient when selecting venues.

The consolidated information on a map using the map package will facilitate choice making among residents and visitors who have no knowledge of the City. We will use the following tools to generate the results for our project/report:

CSV reader for mapping/cleansing of data
Matlab for Plotting
Geo location to retrieve the Longitude and Latitude information
Foursquare account and use the credentials

Exploratory Data Analysis

The data source helped us retrieve all neighborhoods in the Boroughs in New York. We extracted four columns to have the geo location for each neighborhood to help us explore and visualize the data conveniently.

```
In [15]: new_york_data=get_new_york_data()
         new_york_data.head()
```

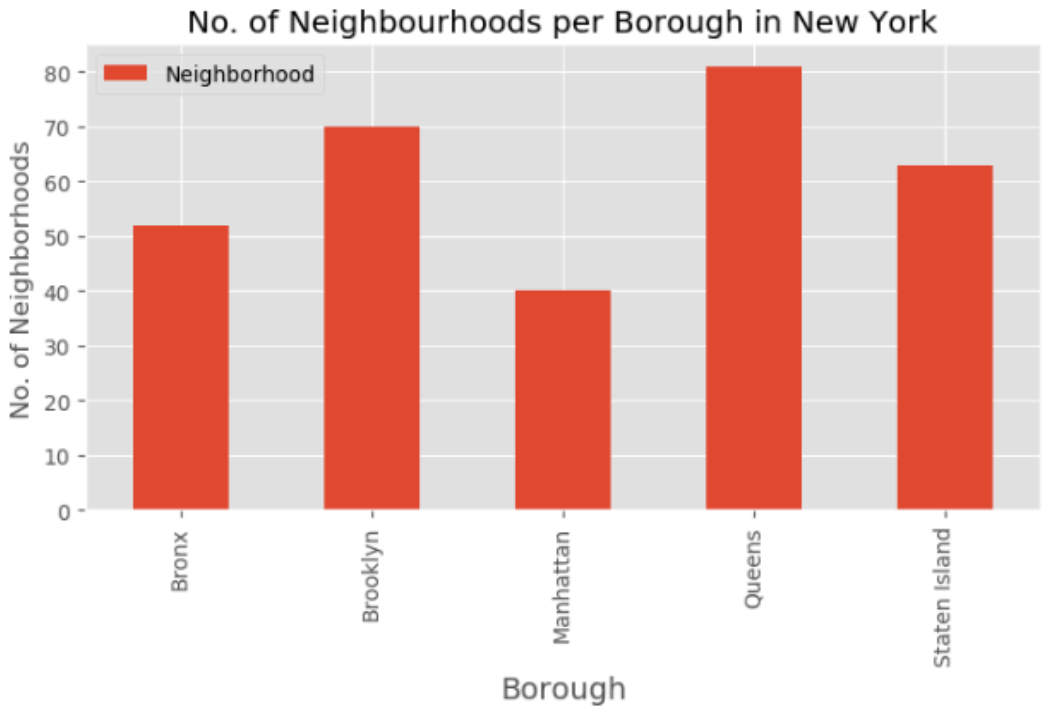
Out[15]:

	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

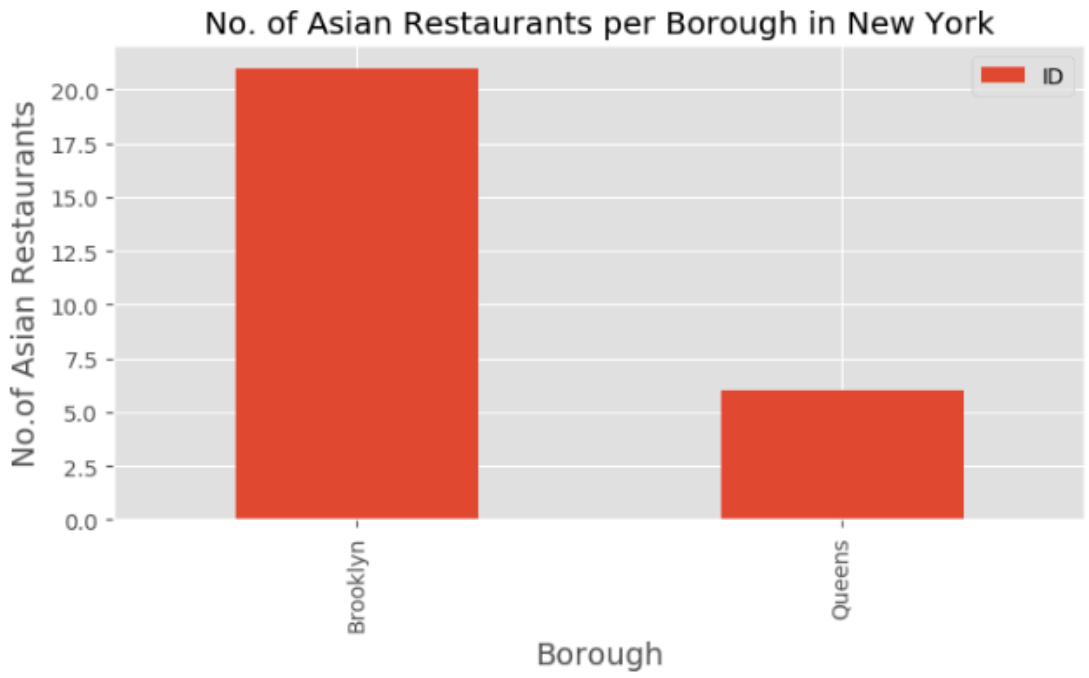
```
In [16]: new_york_data.shape
```

Out[16]: (306, 4)

Once the data was available, it was used to plot a chart to compare the neighborhoods in each Borough.



Next step was to search for the Asian restaurants in each Borough. During the checks, we only found Asian cafes only in two Boroughs of New York. They were also plotted to help visualize the data easily. Please see below:

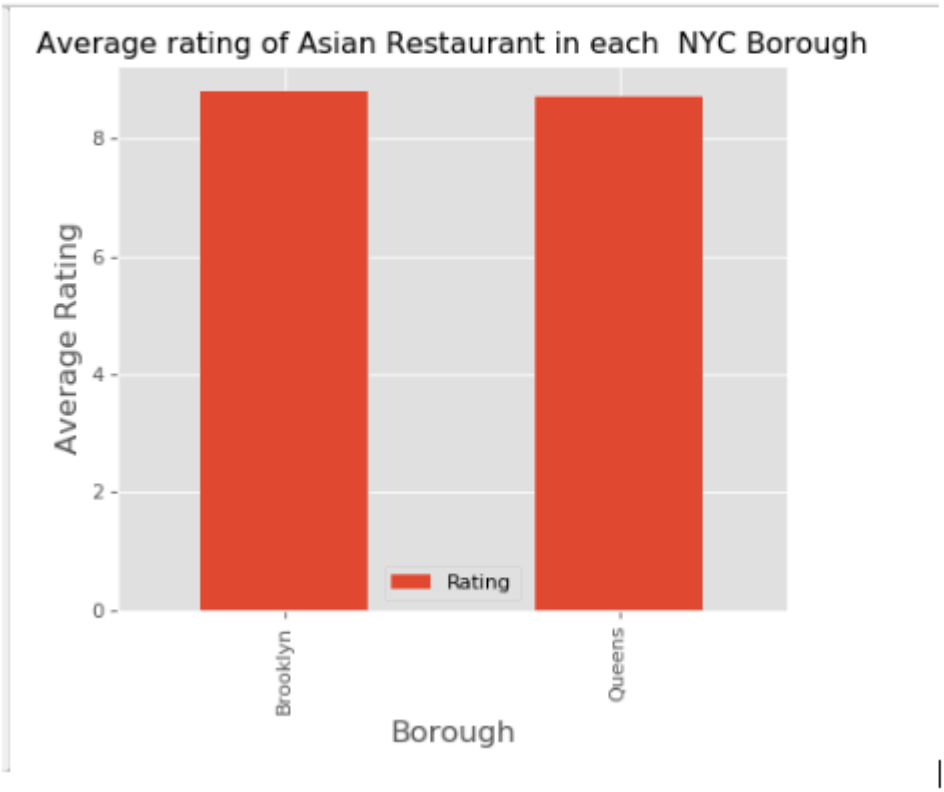


The data shows that Brooklyn has the highest number of Asian restaurants and could be a potential place to visit, relocate or attend for food lovers.

The next step required to actually find the places (restaurants) having good ratings. For this the data was further used to plot another figure:

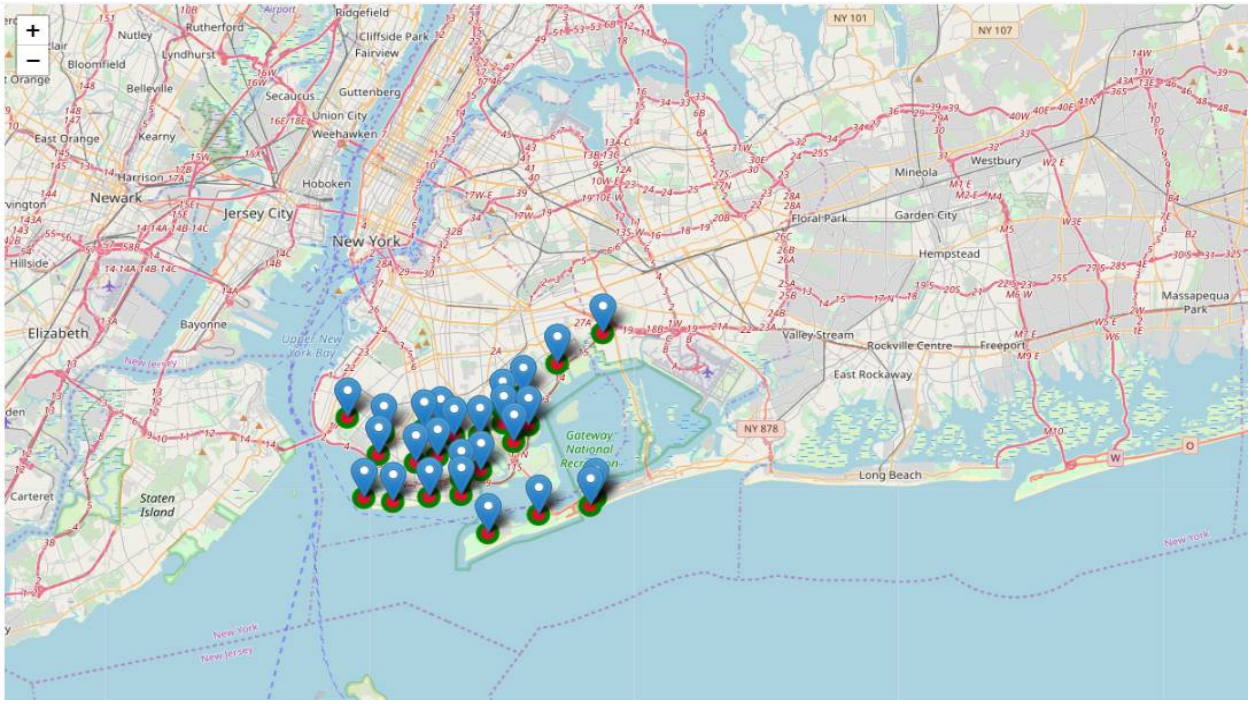
Out[96]:

	Borough	Average Rating
0	Brooklyn	8.809524
1	Queens	8.716667



Restaurants with highest ratings in each Borough were extracted using the above data exploratory techniques.

Now to make it convenient for the users searching for top rated Asian restaurants in a particular location we decided to use Folium to plot them on the map. This will be shown as markers and also giving the ratings for each.



Discussion Section

According to our reported exploration and data analysis, keeping in mind our ultimate goal, of recommending Asian restaurants in New York's each Borough, we were able to get some helpful information for the potential visitors/residents. The Brooklyn area has the maximum number of restaurants (Asian) and many have ratings over 8.0.

The study has some limitations as the data might not be the latest and also out basic Foursquare account has limitations in terms of getting ratings and number of searches we can do in each day.

We recommend to do some more research on data collection and the analysis to have a clearer picture of other factors which could lead potential customers to visit or not visit a particular location.

Distance of subways and crime rate are important factors to take in account. Also, the parking facilities for families/customers attending using their own transport instead of public transport.

This project can further be elaborated to explore the menu and price comparison to help the consumers understand and make better informed decisions before they actually decide or visit a venue.

Conclusion

We concluded our study and completed our first Data Science project. There was a question which was present at the start of the project, lead to a detailed study and come recommendations, along with results generated using data science techniques.

The study can also be used by the potential business owners to decide where the gaps are? What factors can be mitigated to increase the customer satisfaction and ratings.

Personally, concluding this project gives me some level of confidence and an opportunity to apply what I had learnt during this course. The data can be made easy to understand by using the tools such as plotting graphs, consolidating complex information on maps and make it easy to visualize and make informed decisions.