Capstone Project : The Battle of Neighbourhoods

**Finding Indian Restaurants in New York City**

**Rohan Bhatia**

August 7 , 2019

**1. Introduction**

**1.1 Background**

New York is the largest city in the United States with a long history of immigration. It has an extremely diverse culture with people from regions and races making it their home. With it's diverse culture, comes diverse cuisine.The City of New York is famous for its excellent cuisine. It's food culture includes an array of international cuisines influenced by the city's immigrant history.

Apart from all the business in the world , this city is known for it's diverse range of food.The cuisine of New York City comprises many cuisines belonging to various ethnic groups that have entered the United States through the city. Almost all ethnic cuisines are well represented in New York City, both within and outside the various ethnic neighborhoods.

Diverse culture of New York is what makes it's food so unique and popular because ultimately the more the better. People have so many options to chose from.

**1.2 Business Problem**

An Organization has approached me for with the objective to locate and recommend to the management which major parts and neighbourhoods of New York City will be best choice to find an Indian restaurant. The Management also expects to understand the rationale of the recommendations made.

**1.3 Target Audience**

To Recommend the correct location of the Indian Restaurants to the organization.

This would interest anyone who wants to look up great Indian Restaurants in the New York City i.e the residents of the city or anyone out there who crave for Indian Food.

**2. Data Acquisition and Cleaning**

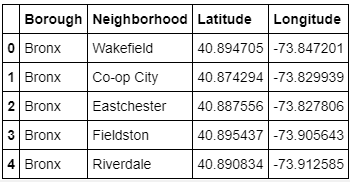
**2.1 Data Sources**

One City will be analysed in this Project : New York

We will be using the below datasets for analysing and finding required results.

### Data 1 : New York City data that contains list Boroughs, Neighborhoods along with their latitude and longitude.

* **Data source :** <https://cocl.us/new_york_dataset>
* **Description :** This data set contains the required information. And we will use this data set to explore various neighborhoods of new york city.



### Data 2 : Indian resturants in each neighborhood of new york city.

* **Data source :** Fousquare API

**Description :** By using this api we will get all the venues in each neighborhood. We can filter these venues to get only indian resturants.

### Data 3 : GeoSpace data

* **Data source :** <https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm>
* **Description :** By using this geo space data we will get the New york Borough boundaries that will help us visualize choropleth map.

**2.1 Data Cleaning**

Data is downloaded from multiple sources and combined into common tables. There were a lot of missing values because of lack of record keeping. I decided to only use data well specified and with complete fields so that analysis could be efficiently.

After cleansing, a data set should be consistent with other similar data sets in the system. The inconsistencies detected or removed may have been originally caused by user entry errors, by corruption in transmission or storage, or by different data dictionary definitions of similar entities in different stores.

Data cleaning differs from data validation in that validation almost invariably means data is rejected from the system at entry and is performed at the time of entry, rather than on batches of data.

**3. Methodology**

**3.1 Business Understanding**

Our main goal here is to find accurately the location of Indian Restaurants for an organization in the New York City.

**3.2 Analytic Approach**

New York City has a total of 5 boroughs and 306 neighbourhoods. In this project, this data is divided on the basis of restaurants location and their rating.

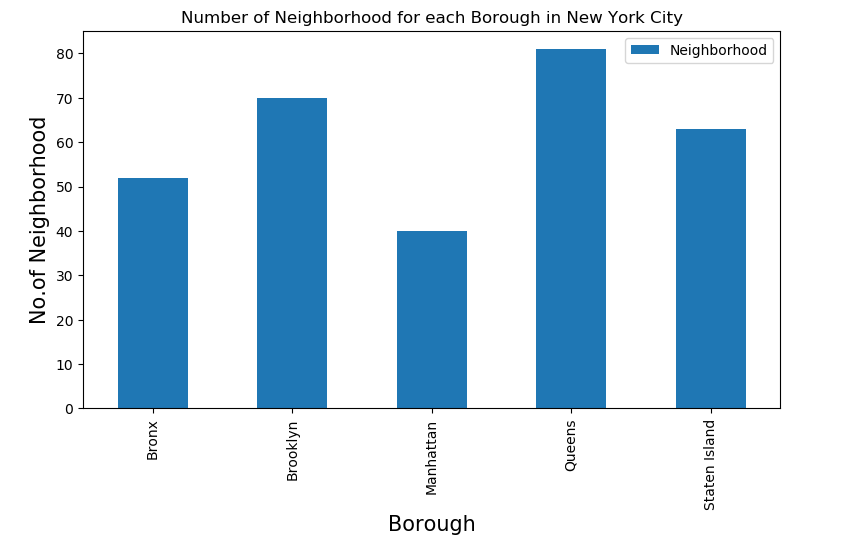
Then using foursquare Api , Indian Restaurants are found in each borough and neighbourhood.

Then their rating and tips are taken into consideration. This is done using Exploratory Data Analysis.

**3.3 Exploratory Data Analysis**

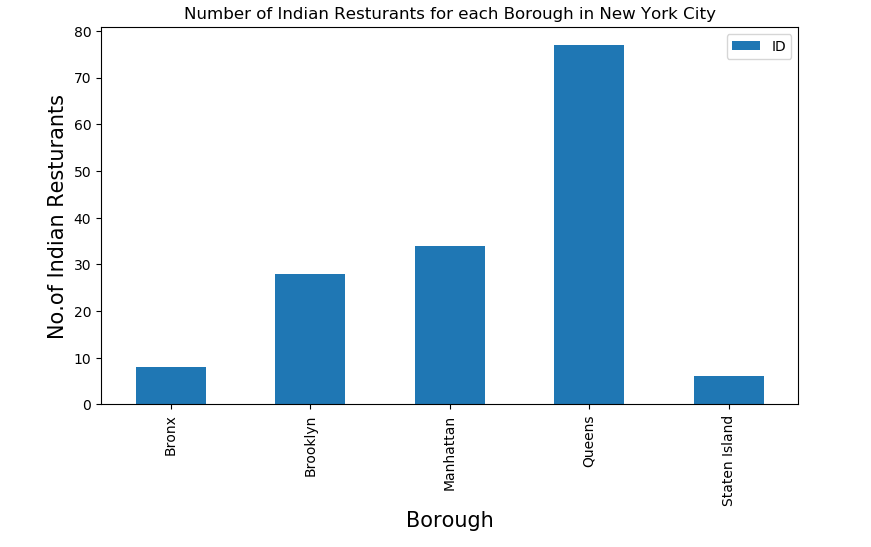
**Data 1 :** New York city data that contains boroughs and neighbourhoods.

1. In this we load the data from <https://cocl.us/new_york_dataset>
2. Transform the data of nested python dictionaries into a pandas dataframe.
3. This Dataframe contains the geographical coordinates of New York City neighbourhoods.
4. This data will be used with venues data from Foursquare.



**Data 2 :** Indian Restaurants in each neighbourhoods of new York city.

1. Foursquare api is used which gets us the venues present in each neighbourhood.
2. We then filter these venues to get only Indian Restaurants.
3. This api allows you to make limited no. of calls and Api key and secret is needed.

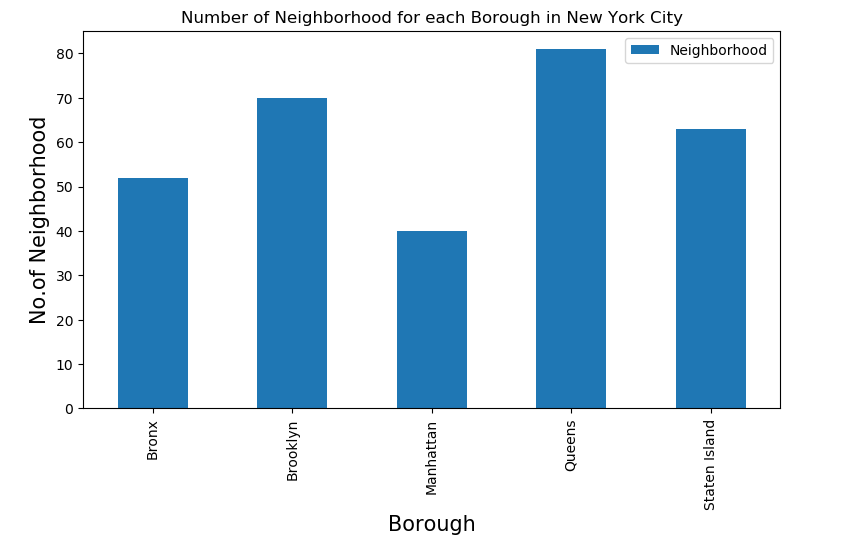


**Data 3 :** Geospacial Data

1. Here we load the data from <https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm>.
2. By using this geospace data we will get the New York Borough boundaries that will help us visualize choropleth map.
   1. **Relationship between No. of Neighbourhoods and Borough**

The New York City data can be used and inferences can be easily drawn in the relationship between No. of Neighbourhoods and Boroughs.

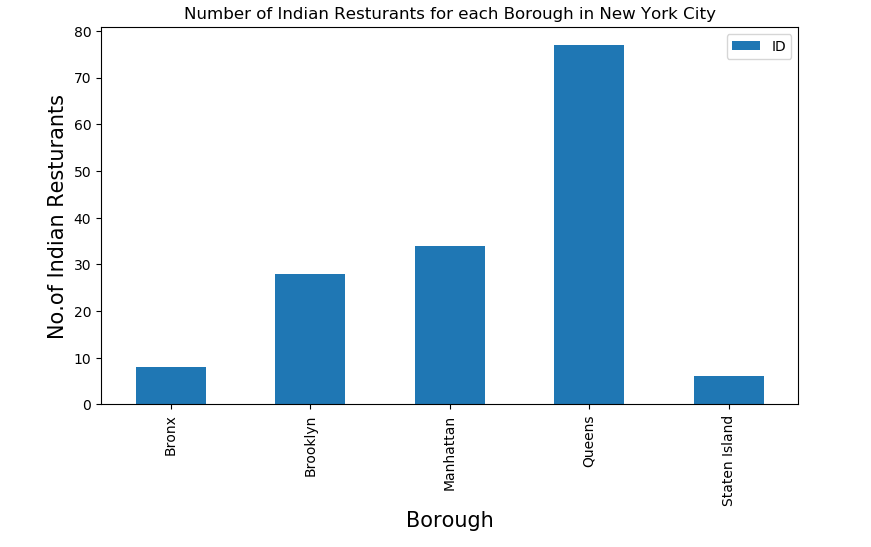
This relationship specifies how many Neighbourhoods are contained in each borough.



* 1. **Relationship between No. of Indian Restaurants and Borough**

This relationship can be formed using the data from Foursquare venues api which provides Data for Restaurants based on longitude and latitude.

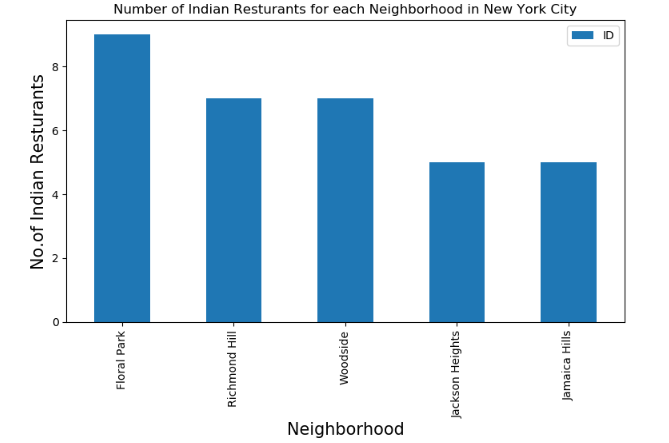
Here the goal is to specify how many Indian Restaurants are in each Borough.



* 1. **Relationship between No. of Indian Restaurants and Neighbourhoods**

Here we form the relationship between No. of Indian Restaurants and Neighbourhoods.

The plot below shows the no of Indian Restaurants in different neighbourhoods using box plot.

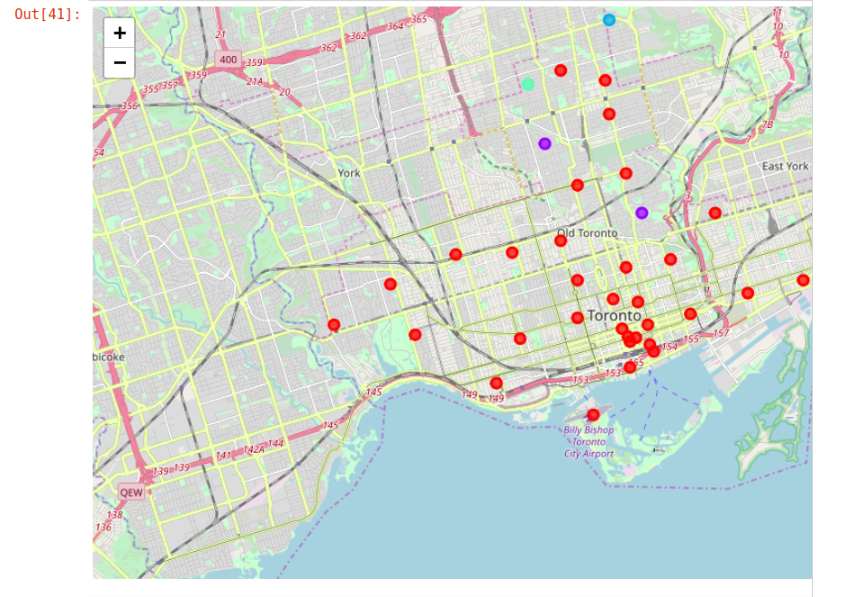


* 1. **Modelling**

**K-means clustering based on mean occurrence of venue categories :**

To cluster the neighbourhoods into clusters based on the location of Indian Restaurants, we have used k-means clustering which provides better efficiency and f –score and create map clusters based on presence of Indian Restaurants in various neighbourhoods.

K –means clustering aims to partition n observations into K clusters in which each observation belongs with the nearest mean. It uses iterative refinement approach.



Here Clusters of Indian Restaurants can be found in various neighbourhoods based on their rating, tips data from venues Foursquare.

**5.0 Results**

Our analysis shows that there are a great number of Indian restaurants in the city of New York. These analysis is based on various neighbourhoods of the city. Indian Cuisine is one of favourite of the city.

Some cities such as Manhattan is the best place to be if someone prefers Indian Cuisine.

This analysis is based on boroughs which is further divided into neighbourhoods based on the location of Indian Restaurants, once a list of Indian Restaurants is accumulated, further analysis is done of the basis of Likes , Tips and Rating.

Eventually as far as the results are concerned , these are one of the best to places to visit for Best Indian Restaurants

Astoria, Queens Blissville , Queens Civic Centre , Manhattan

Staten Island ranks last in the average rating of Indian Restaurants.

**6.0 Discussion**

1. There is a scope of finding more restaurants since the analysis is purely based on Foursquare data.
2. There is a scope ratings of some restaurants change over time with customer review .
3. Places where the there is decrease in the presence of Indian Cuisine can show increase in coming times.

**7.0 Conclusion**

Purpose of this project was to identify areas in NYC containing Indian Restaurants in order to aid Stakeholders in narrowing down the search for Indian cuisine in New York City. By calculating Restaurant Density Distribution we have calculated we have first identified the boroughs . Then Every Neighbourhoods in Borough which satisfy our basic requirements of the venues in Foursquare. Clustering of those locations was then done in order to create major locations which contain Indian Restaurants with high ratings.

Final Decision will be made by the Stakeholders in the organization that how this analyis will help them since we are able to find the Indian Restaurants based on a lot of factors faviorable in that particular area.