IBM Applied Data Science Capstone

Opening an Indian Restaurant in New York

Padmaja Tutika

4/13/2021

Introduction: Business Problem

The objective of this Capstone project is to analyze and find best possible location in New York City to open an Indian cuisine restaurant with western twist. New York is the largest city in US with a long history of international immigrants that has population of diverse culture, cuisine and magnificent people. New York is home to an array of outstanding Indian restaurants that serve cuisine akin to that found in Indian households or India itself, including amazing regional specialties, takeout tikka masala, and handheld snacks.As with any new business, it is important to analyze the location, neighborhood, and other factors that would lead the new Indian restaurant to great success and profit.

Interest

Obviously, developers and investors looking to open or invest in an Indian restaurant in the city of New York will be interested. New York is a great place to open a restaurant with an authentic but still with a modern twist menu that can attract people from diverse culture and ethnicity. Indian restaurant has provided variety of choices from authentic to fusion food.

Data acquisition and cleaning

To solve the business problem, following data are required:

- data on New York City neighborhoods
- boroughs to include boundaries, latitude, longitude, restaurants
- restaurant ratings and tips

Methodology

New York City data containing the neighborhoods and boroughs, latitudes, and longitudes will be obtained from the data source: https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json

All data related to locations and quality of Indian restaurants will be obtained via the FourSquare API utilized via the Request library in Python.

- Foursquare API: This project would use Four-square API as its prime data gathering source as it has a database of millions of places, especially their places API which provides the ability to perform location search, location sharing and details about a business.
- Work Flow: Using credentials of Foursquare API features of near-by places of the neighborhoods would be mined. Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 500.
- Clustering Approach: k-means clustering algorithm

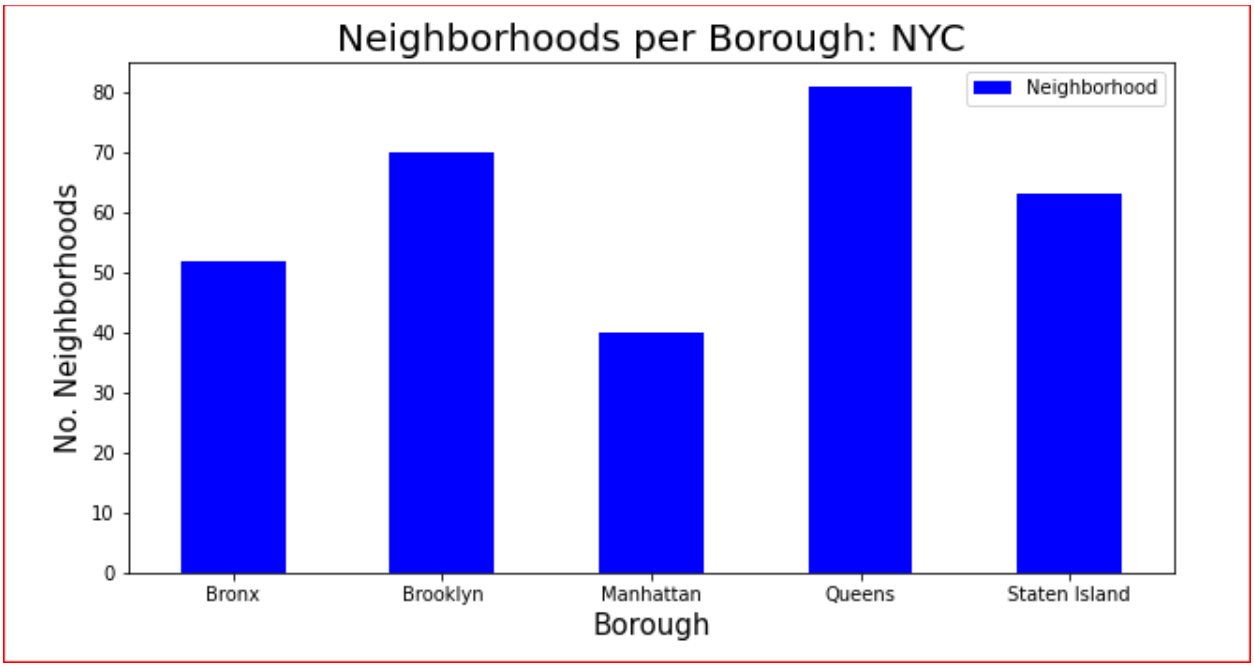
Libraries which are used to develop the Project:

- Pandas: For creating and manipulating dataframes.
- Folium: Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.
- Scikit Learn: For importing k-means clustering.
- JSON: Library to handle JSON files.
- XML: To separate data from presentation and XML stores data in plain text format.
- Geocoder: To retrieve Location Data.
- Beautiful Soup and Requests: To scrap and library to handle http requests.
- Matplotlib: Python Plotting Module.

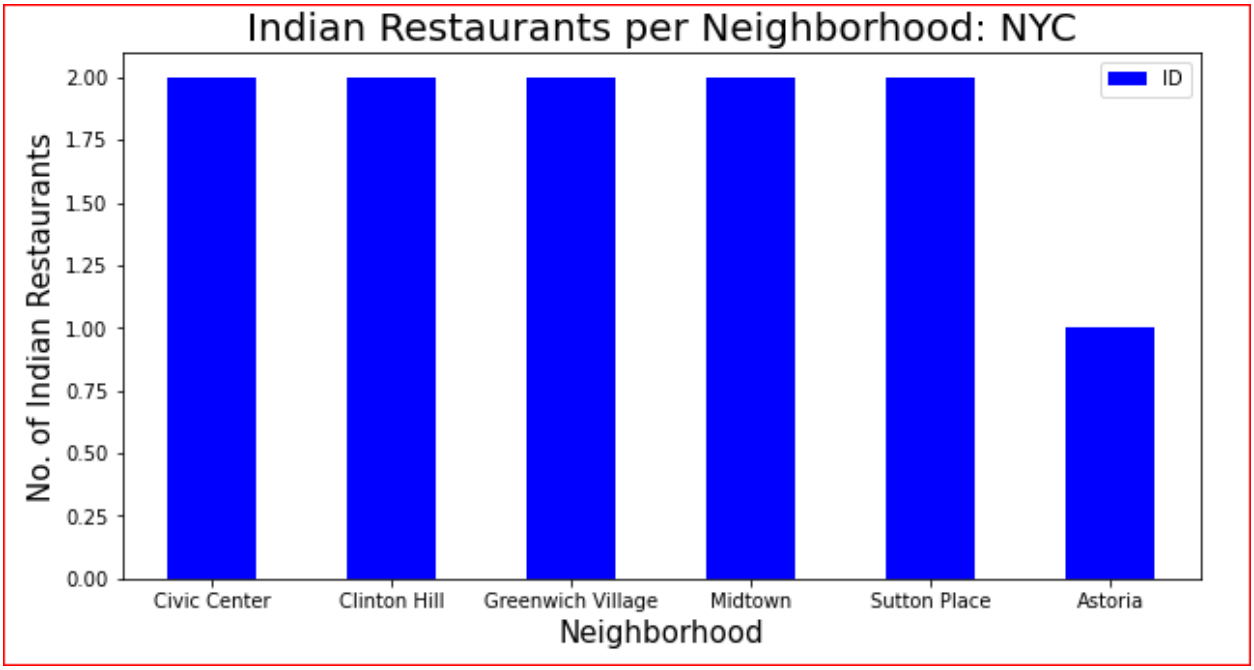
Results/Conclusions:

The results of our analysis shown below:

- We see that Queens has the highest number of Neighborhoods.



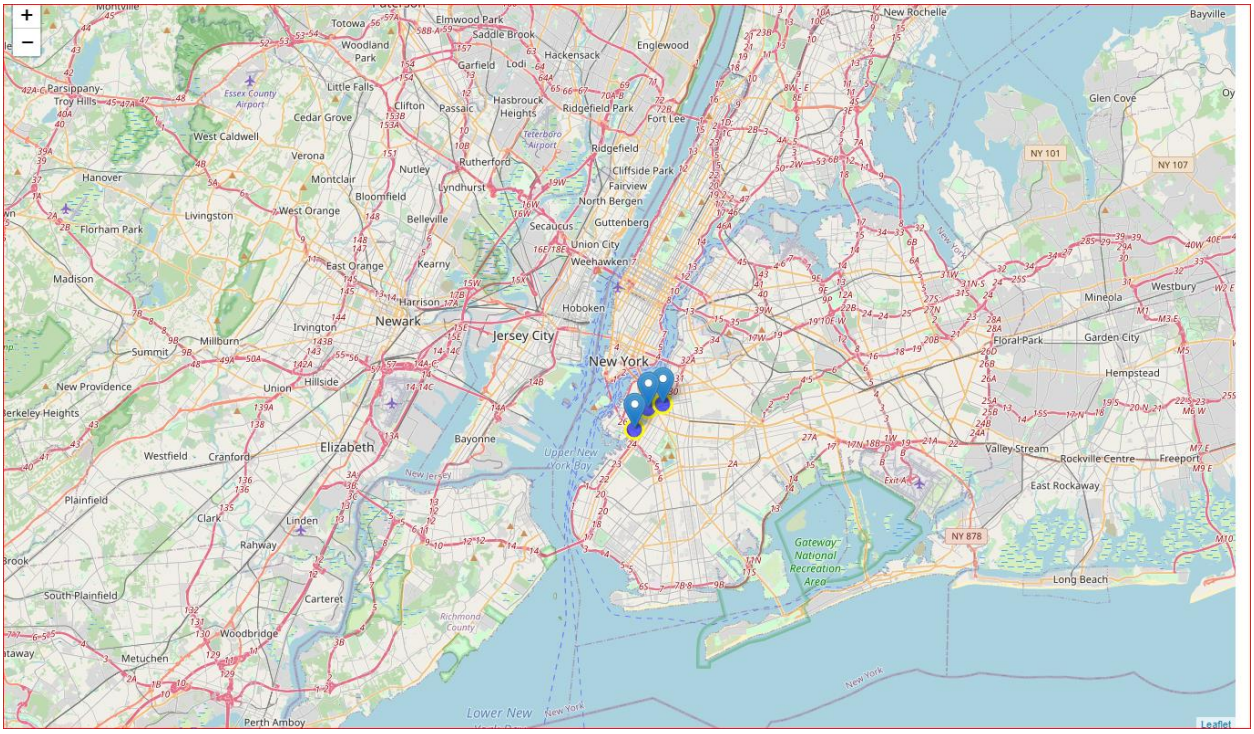
- Most of the neighborhoods in New York has the highest number of Indian restaurants. Manhattan is one of them.



- Manhattan and Staten Island has the highest average rating for Indian restaurants. Brooklyn and Queens goes right behind with bit lower rating.

	Neighborhood	Average Rating
0	Astoria	8.20
2	Bayside	8.00
3	Boerum Hill	8.10
6	Chelsea	8.80
10	Fort Greene	8.60
11	Gowanus	8.00
13	Greenwich Village	8.65
20	Midtown	8.40
21	New Dorp	8.20
22	Noho	8.80
23	North Side	8.30
24	Prospect Lefferts Gardens	8.80
26	Sutton Place	8.25
27	Tribeca	9.00
31	West Village	8.50

Some top neighborhoods with best average rating ranging from 8.0 till 9.0.



Limitations and recommendations for future research

All the above analysis were done based on the data collected at Foursquare. There were limitations on API calls with the free account. For better analysis and research, the source of data can be expanded to other external databases and the limitation on API calls can also be waived.

Discussion / Conclusion

Manhattan and Brooklyn have the best rated Indian restaurants. Based on the above information, I would state that Manhattan and Brooklyn are the best locations for Indian cuisine in NYC. To have the best shot of success, I would open a Indian restaurant in Manhattan. Manhattan has multiple neighborhoods with average ratings exceeding 8.0. Midtown in Manhattan has about 834 likes which is the highest among other neighborhood.