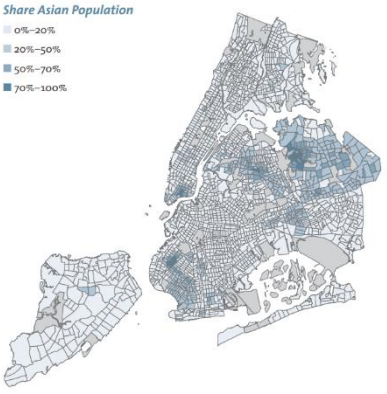
IBM Data Science
Applied Data Science Capstone

Bengali Restaurant in New York City

Introduction

New York City has an extraordinarily diverse population. It is the largest city in the United States with a long history of international immigration. It is one of the few cities in the country in which four different racial/ethnic groups each make up at least 10 percent of the population. According to the 2010 decennial census, 33 percent of New York City residents are white, 26 percent are Hispanic, 26 percent are black, and 13 percent are Asian. While the diversity of New York City's population is not reflected in many of the city's neighborhoods, over the past 20 years the racial and ethnic makeup of the city's neighborhoods (census tracts) has increasingly come to look more like that of the city itself. Between 2000 and 2010 the city as a whole became more Hispanic and Asian. While the population at large has declined the past few years, the city's Asian population has, meanwhile, become more concentrated.[1]



According to the 2000 US Census, there are approximately 57,000 Bangladeshis in the United States and the New York metropolitan area is home to the largest Bangladeshi population. From 1990 to 2000, New York City's Bangladeshi population increased by 471% (from 4,955 to 28,269). About 85% of Bangladeshi New Yorkers were foreign-born, and 77% of these immigrants came to the US from 1990 to 2000. Some 30% of all Bangladeshi immigrants in the city in 2000 were naturalized US citizens – compared with 45% of all foreign-born city residents.[2]

Business Problem

The location of food service business will impact its success nearly as much as the menu. If the restaurant is in the wrong place, we cannot attract the number of customers we will need in order to stay in business. The same is true if our location is inundated with competitors, or has poor visibility, or is hard to find. There are many things that must be considered as we look for a location in which to open our business.

The objective of the Capstone project is to analyze and select the best locations in the city of New York to open a new Bengali/Bangladeshi restaurant, using data science methodology and instruments such as data analysis and visualization.

[1] Racial and Ethnic Makeup of NYC Neighborhoods, The Furman Center for Real Estate & Urban Policy
[2] AAFNY's "Census Profile: New York City's Bangladeshi American Population"

Data

To solve the problem, we will need the following data:

- New York City data containing the neighborhoods and boroughs.
- Latitude and longitude coordinates of those neighborhoods.
- Venue data, particularly data related to restaurants.

New York City data containing the neighborhoods and boroughs will be obtained from the open data source: **https://cocl.us/new_york_dataset**. Next, we will get the geographical coordinates of the neighborhoods (latitude and longitude) using the **Python Geocoder** package. We will use the **Foursquare API** to get the venue data for the neighborhoods defined at the previous step. Foursquare has one of the largest databases of 105+ million places, 900+ venue categories, 30+ attribute fields, and over 150,000 developers use this application. Among the many Foursquare API provided categories of the venue data, we are particularly interested in the restaurant data to solve the business problem defined above.

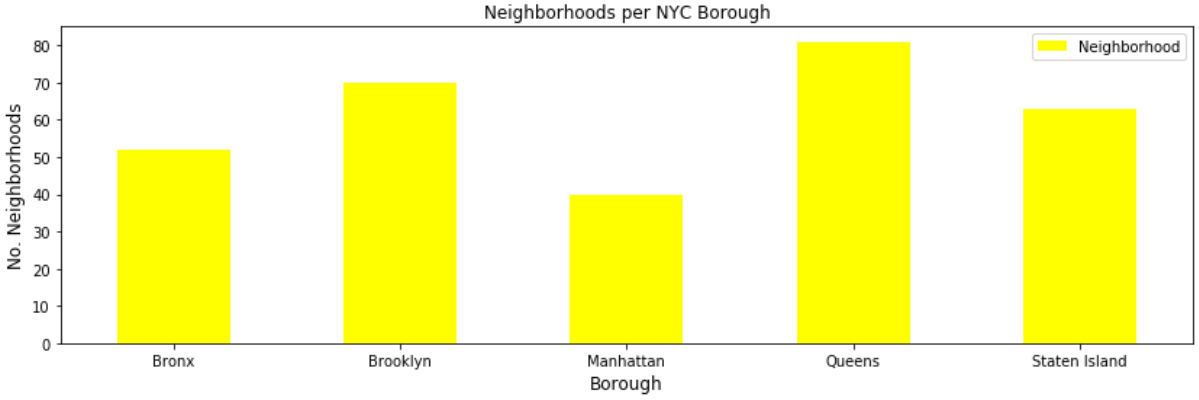
Methodology

- Collect data from https://cocl.us/new_york_dataset.
- Clean and process data into a dataframe.
- Use FourSquare to locate all venues.
- Filter by Bengali restaurants.
- Sort data based on rankings.
- Visually assess data using graphing from Python libraries.

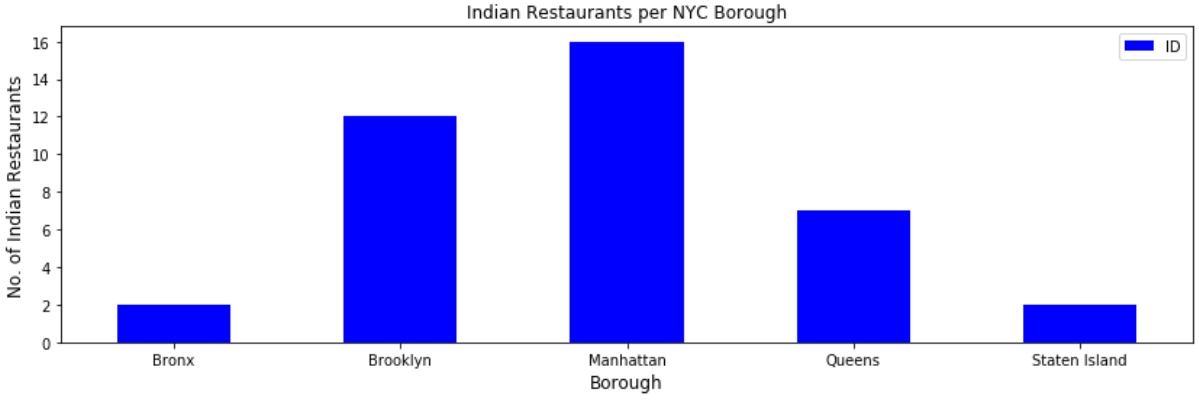
Results

1. In Foursquare database, we got 0 Bangladeshi Restaurants across NYC. This should be quite good for business as no pure Bangladeshi restaurant is available to cater to the increasing Bangladeshi population. For further analysis, we will check for Indian restaurants which are similar to the Bangladeshi cuisine.

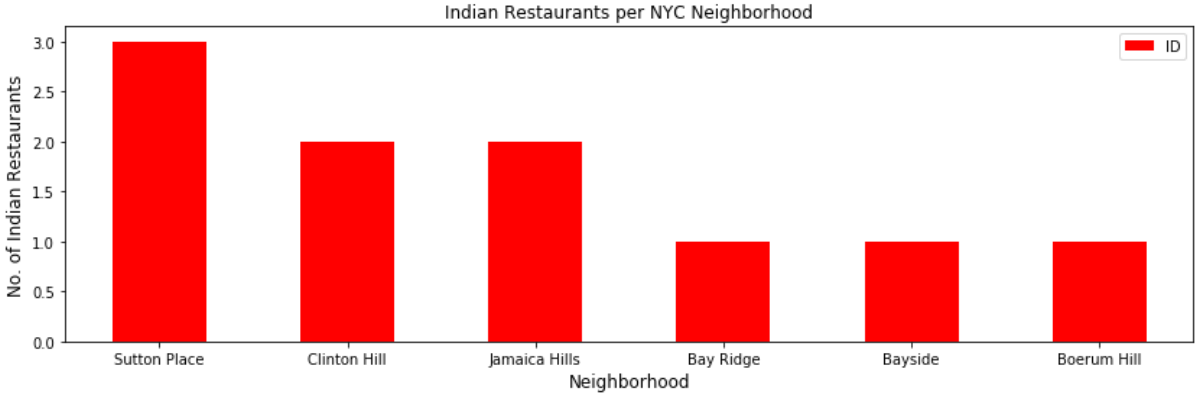
2. Queens has the highest number of Neighborhoods.



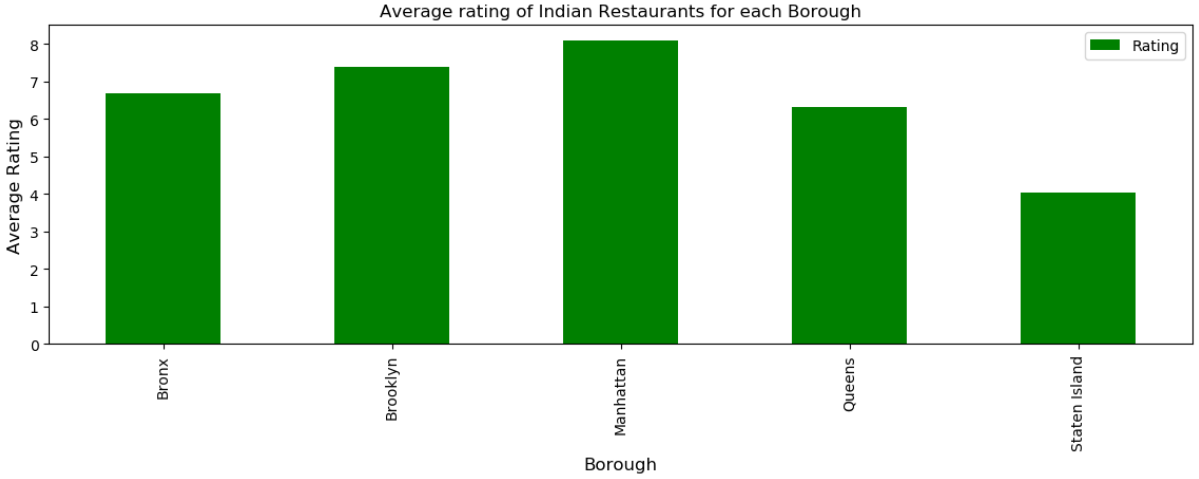
3. Manhattan has the least number of neighborhoods, but it does have the highest number of Indian restaurants.



4. Sutton Place in Manhattan has the highest number of Indian Restaurants with a total count of 3.



5. Manhattan has the highest average rating for Indian Restaurants. Brooklyn goes right behind the Manhattan with the average rating of its Indian restaurants only a little lower than Manhattan.



6. Two neighborhoods with the highest average rating for Indian Restaurants are Tribeca and Chelsea (both located in Manhattan).

	Neighborhood	Average Rating
1	Bayside	8.200000
2	Boerum Hill	8.100000
7	Chelsea	8.700000
8	Civic Center	8.300000
11	Fort Greene	8.600000
14	Greenwich Village	8.300000
21	Murray Hill	8.100000
22	New Dorp	8.100000
23	Noho	8.800000
24	North Side	8.400000
25	Prospect Lefferts Gardens	8.500000
29	Sutton Place	8.133333
30	Tribeca	9.100000

7. Neighborhoods with restaurants with average rating of 8 and above are mapped below.

Discussion

- Manhattan and Brooklyn have the best rated Indian restaurants on average.
- Staten Island and The Bronx have the least number of Indian restaurants per borough.
- Sutton Place Hill in Manhattan has the highest number of Indian Restaurants in all of NY.
- Despite Manhattan having the least number of neighborhoods in all five boroughs, it has the greatest number of Indian restaurants.

Based on the above information, we can state that Manhattan and Brooklyn are comparatively better locations for Bangladeshi cuisine in NYC. We can open a Bangladeshi restaurant in Brooklyn which has multiple neighborhoods with average ratings exceeding 8.0 on a scale of 1.0 to 10.0 and has a smaller number of Indian restaurants than Manhattan, making competition easier. Brooklyn is also close to Jackson Heights, Queens which the main hub for Bangladeshi population. Moreover, the real estate prices in Brooklyn are much cheaper than in Manhattan.

Conclusion

All of the above analysis is depended on the accuracy of Foursquare data. A more comprehensive analysis would need to incorporate data from other external databases like area-specific population demographics, real estate prices etc.