IBM Applied Data Science Capstone Project

Coursera

Opening a new Indian Cuisine Restaurant in New York City, USA

Report

Submitted By:

Supriyo Chowdhury

March 2020

Introduction

New York City's demographics show that it is a large and ethnically diverse metropolis. It is the largest city in the United States with a long history of international immigration. Over the last decade the city has been growing faster than the region. The new York region continues to be by far the leading metropolitan gateway for legal immigrants admitted into the United States.

With its diverse culture, comes diverse food items. There are many restaurants in New York City, each belonging to different categories like Chinese, Indian, Italian and French etc.

For many Indians residing in the US or particularly in the New York City, its hectic to find Indian cuisine restaurant. Generally a lot of Indians look for their home cuisine in the foreign countries, and this brings to the development of a lot of Indian restaurants in the US. Restaurants are like one stop destination. For retailers, the central location and the large crowd, low neighbouring restaurants provides a great business. As a result, there are many Indian cuisine restaurants in the New York city, USA and many more are being built. Opening restaurants in perfect place allows restaurants managers to earn a lot. Particularly, the location of the restaurant is one of the most important factor that determines whether the mall will be success or a failure.

Business Problem

The objective of this capstone project is to analyse and select the best locations in the New York City to open a new Indian cuisine restaurant. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the main business question: **If a restaurant manager is looking for a new Indian cuisine restaurant in the New York City, US where would you recommend that they open it?**

Along with some sub-questions like:

1. Which areas have potential Indian Restaurants?
2. Which areas lacks Indian Restaurants?
3. Which is the best place to stay if you prefer Indian Cuisine?

Target Audience

This project is particularly useful to restaurant managers, property developers and investors looking to open or invest in new Indian cuisine restaurants in the New York City.

Data Section

To solve the problem we need the following data:

1. List of neighbourhoods in New York City.
2. Longitude and latitude of those neighbours. This is required in order to plot maps
3. Venue data, particularly data related to Indian Cuisine restaurants. This data will be used perform the clustering on the neighbours.

Data Source

For this project we need the following data:

1. New York City data that contains list Boroughs, Neighbourhoods along with their latitude and longitude.

    a. **Data Source:** *https://cocl.us/new_york_dataset*

    b. Description: This data set is the most important dataset that contains the information of the neighbourhoods. It also contains the latitude and longitude also.

2. List of Indian restaurants in each neighbourhood of New York City:

    a. Data Source: Foursqaure API

    b. Description: This data provides all the venues in each neighbourhood.

3. GeoSpace Dataset:

    a. Data Source: *https://data.cityofnewyork.us/CityGovernment/Borough-Boundaries/tqmj-j8zm*

    b. Description: This will help us to visualize the choropleth map.

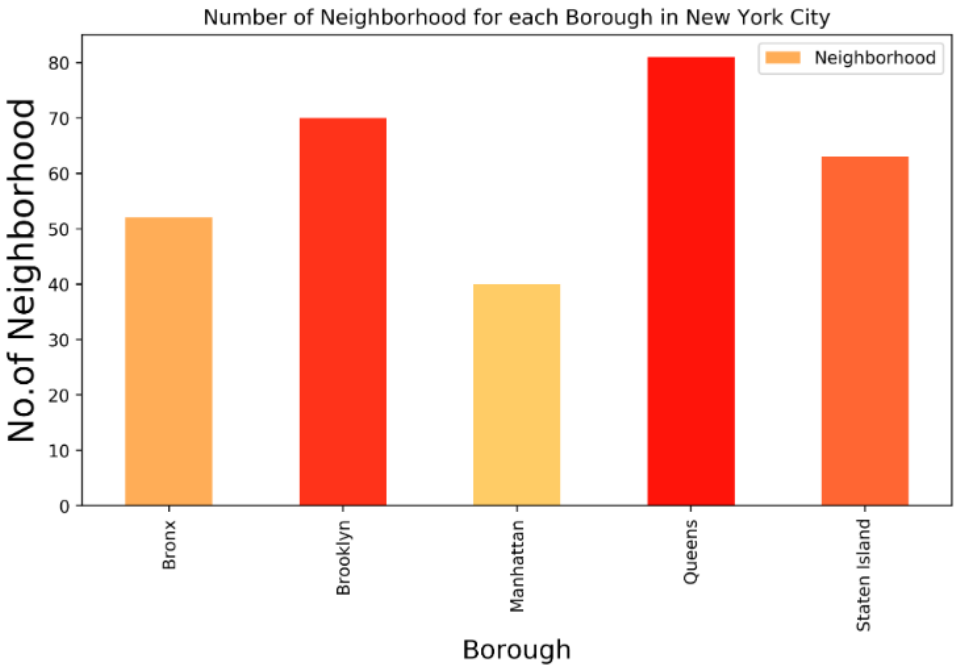
Methodology

1. We begin by collecting the New York city data from the following link "*https://cocl.us/new_york_dataset*"

Out[120]:

	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
5	Bronx	Kingsbridge	40.881687	-73.902818
6	Manhattan	Marble Hill	40.876551	-73.910660
7	Bronx	Woodlawn	40.898273	-73.867315
8	Bronx	Norwood	40.877224	-73.879391
9	Bronx	Williamsbridge	40.881039	-73.857446

2. We visualized the data into different neighbors



3. We will find all venues for each neighborhood using Foursquare API

8.We will collect the Indian Cuisine Restaurants for each neighbourhood

In [12]:
```
# prepare neighborhood list that contains indian resturants
column_names=['Borough', 'Neighborhood', 'ID','Name']
indian_rest_ny1=pd.DataFrame(columns=column_names)
count=1
for row in new_york_data.values.tolist():
    Borough, Neighborhood, Latitude, Longitude=row
    venues = get_venues(Latitude,Longitude)
    indian_resturants=venues[venues['Category']=='Indian Restaurant']
    print('(',count,'/',len(new_york_data),')','Indian Resturants in '+Neighborhood+', '+Borough+':'+str(len(indian_resturants)))
    for resturant_detail in indian_resturants.values.tolist():
        id, name , category=resturant_detail
        indian_rest_ny1 = indian_rest_ny.append({'Borough': Borough,
                                                 'Neighborhood': Neighborhood,
                                                 'ID': id,
                                                 'Name' : name
                                                }, ignore_index=True)
    count+=1
```

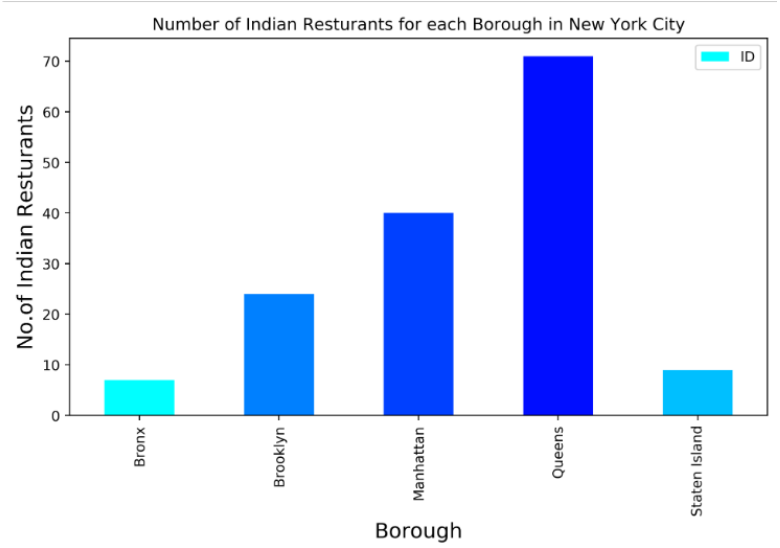
```
( 1 / 306 ) Indian Resturants in Wakefield, Bronx:0
( 2 / 306 ) Indian Resturants in Co-op City, Bronx:0
( 3 / 306 ) Indian Resturants in Eastchester, Bronx:0
( 4 / 306 ) Indian Resturants in Fieldston, Bronx:0
( 5 / 306 ) Indian Resturants in Riverdale, Bronx:0
( 6 / 306 ) Indian Resturants in Kingsbridge, Bronx:0
( 7 / 306 ) Indian Resturants in Marble Hill, Manhattan:0
( 8 / 306 ) Indian Resturants in Woodlawn, Bronx:0
( 9 / 306 ) Indian Resturants in Norwood, Bronx:0
( 10 / 306 ) Indian Resturants in Williamsbridge, Bronx:0
( 11 / 306 ) Indian Resturants in Baychester, Bronx:0
( 12 / 306 ) Indian Resturants in Pelham Parkway, Bronx:0
( 13 / 306 ) Indian Resturants in City Island, Bronx:0
( 14 / 306 ) Indian Resturants in Bedford Park, Bronx:0
( 15 / 306 ) Indian Resturants in University Heights, Bronx:0
( 16 / 306 ) Indian Resturants in Morris Heights, Bronx:0
( 17 / 306 ) Indian Resturants in Fordham, Bronx:0
( 18 / 306 ) Indian Resturants in East Tremont, Bronx:0
( 19 / 306 ) Indian Resturants in West Farms, Bronx:0
( 20 / 306 ) Indian Resturants in High  Bridge, Bronx:0
( 21 / 306 ) Indian Resturants in Melrose  Bronx:0
```

In [14]:
```
indian_rest_ny.head()
```

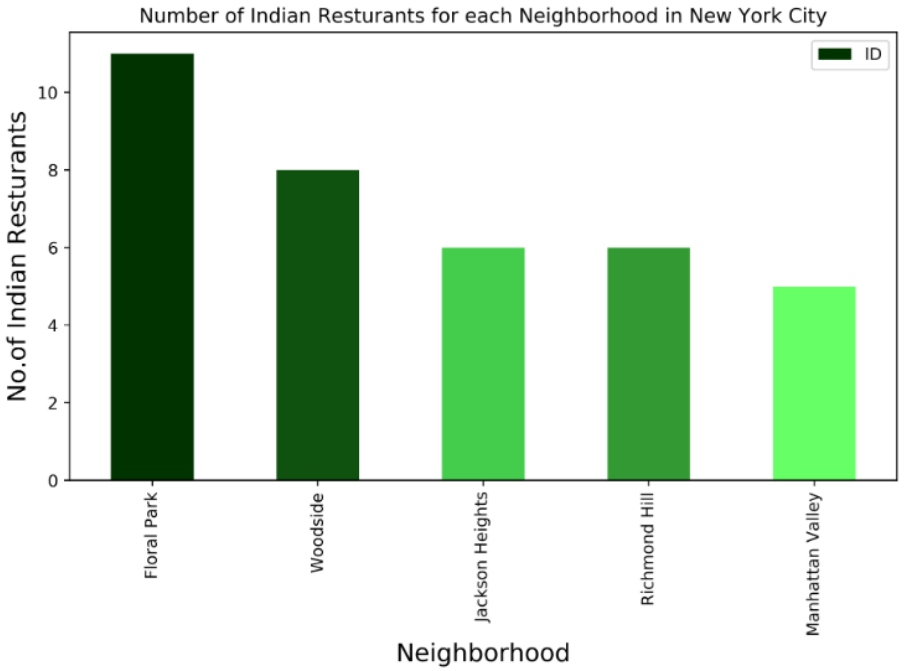
Out[14]:

	Borough	Neighborhood	ID	Name
0	Bronx	Woodlawn	4c0448d9310fc9b6bf1dc761	Curry Spot
1	Bronx	Parkchester	4c194631838020a13e78e561	Melanies Roti Bar And Grill
2	Bronx	Spuyten Duyvil	4c04544df423a593ac83d116	Cumin Indian Cuisine
3	Bronx	Concourse	551b7f75498e86c00a0ed2e1	Hungry Bird
4	Bronx	Unionport	4c194631838020a13e78e561	Melanies Roti Bar And Grill

4. We visualized the number of Indian restaurants for each borough.



5. We visualized the number of Indian restaurants for each neighborhood in New York.



6. Created another dataframe with Borough and Number of Restaurants in each borough.

Out[106]:

	Borough	No. of Restaurants
0	Queens	71
1	Manhattan	40
2	Brooklyn	24
3	Bronx	7
4	Staten Island	9

7. Next using Foursquare API, we will find the Ratings, Tips, and Number of Likes for all the Indian Restaurants.

Out[20]:

	Borough	Neighborhood	ID	Name	Likes	Rating	Tips
0	Bronx	Woodlawn	4c0448d9310fc9b6bf1dc761	Curry Spot	5	7.6	10
1	Bronx	Parkchester	4c194631838020a13e78e561	Melanies Roti Bar And Grill	3	5.8	2
2	Bronx	Spuyten Duyvil	4c04544df423a593ac83d116	Cumin Indian Cuisine	13	6.1	9
3	Bronx	Concourse	551b7f75498e86c00a0ed2e1	Hungry Bird	8	6.9	3
4	Bronx	Unionport	4c194631838020a13e78e561	Melanies Roti Bar And Grill	3	5.8	2

8. We will then sort Neighborhoods and Borough the data keeping Ratings as the constraint.

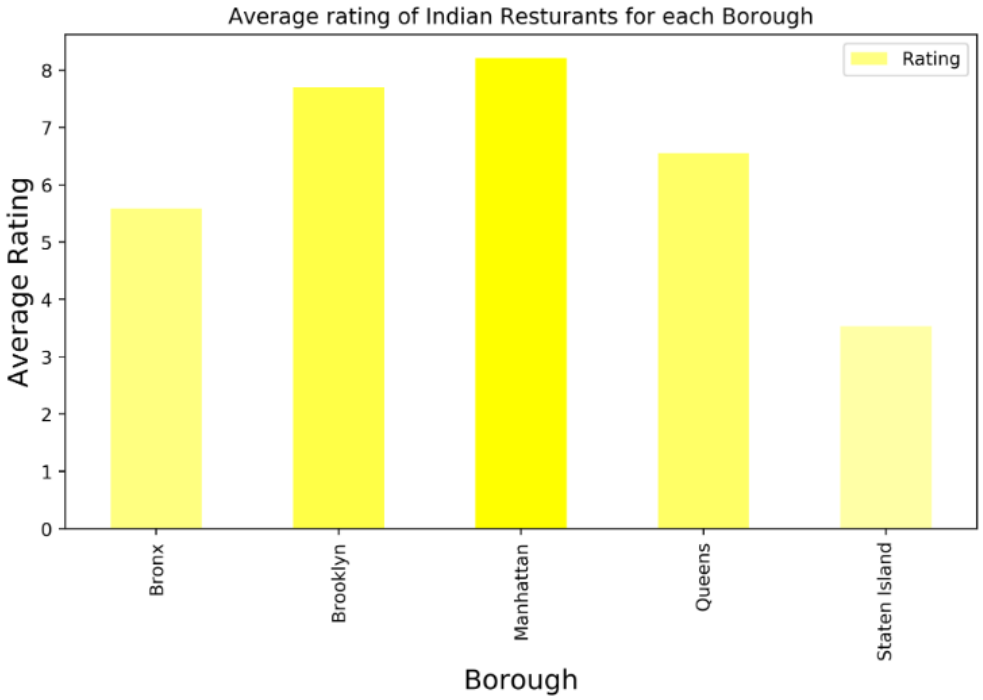
Out[79]:

	Neighborhood	Average Rating
12	Civic Center	9.100000
69	Tribeca	9.100000
0	Astoria	9.000000
5	Blissville	9.000000
75	West Village	8.800000
44	Midtown South	8.800000
43	Midtown	8.800000
29	Gramercy	8.733333
25	Fort Greene	8.700000
11	Chelsea	8.700000

ut[38]:

	Borough	Average Rating
2	Manhattan	8.210000
1	Brooklyn	7.700000
3	Queens	6.552113
0	Bronx	5.585714
4	Staten Island	3.533333

9. We visualize average rating Indian Restaurants for each borough



10. Next we will consider all the neighborhoods with average rating greater or equal 8.0 to visualize on map.

ut[81]:

	Neighborhood	Average Rating
0	Astoria	9.000000
1	Bay Ridge	8.400000
5	Blissville	9.000000
9	Carnegie Hill	8.066667
10	Central Harlem	8.050000
11	Chelsea	8.700000
12	Civic Center	9.100000
13	Clinton Hill	8.150000
18	East Flatbush	8.100000
20	Elmhurst	8.300000
25	Fort Greene	8.700000
29	Gramercy	8.733333
30	Grant City	8.100000
32	Holliswood	8.300000

11. We will join this dataset to original New York data to get longitude and latitude.

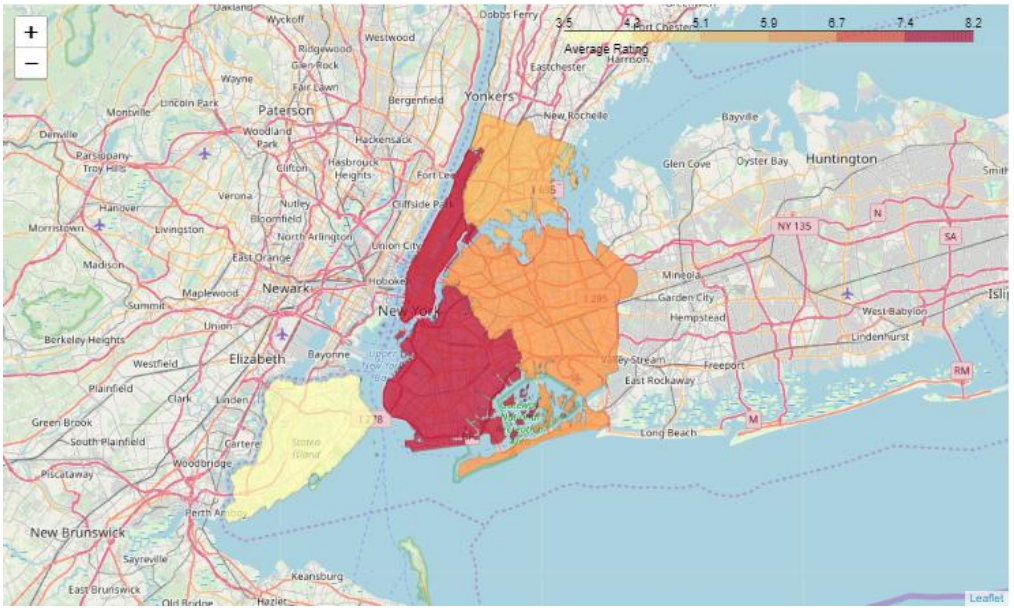
Out[83]:

	Borough	Neighborhood	Latitude	Longitude	Average Rating
0	Queens	Astoria	40.768509	-73.915654	9.000000
1	Brooklyn	Bay Ridge	40.625801	-74.030621	8.400000
2	Queens	Blissville	40.737251	-73.932442	9.000000
3	Manhattan	Carnegie Hill	40.782683	-73.953256	8.066667
4	Manhattan	Central Harlem	40.815976	-73.943211	8.050000
5	Manhattan	Chelsea	40.744035	-74.003116	8.700000
6	Staten Island	Chelsea	40.594726	-74.189560	8.700000
7	Manhattan	Civic Center	40.715229	-74.005415	9.100000
8	Brooklyn	Clinton Hill	40.693229	-73.967843	8.150000
9	Brooklyn	East Flatbush	40.641718	-73.936103	8.100000
10	Queens	Elmhurst	40.744049	-73.881656	8.300000
11	Brooklyn	Fort Greene	40.688527	-73.972906	8.700000
12	Manhattan	Gramercy	40.737210	-73.981376	8.733333
13	Staten Island	Grant City	40.576216	-74.105856	8.100000

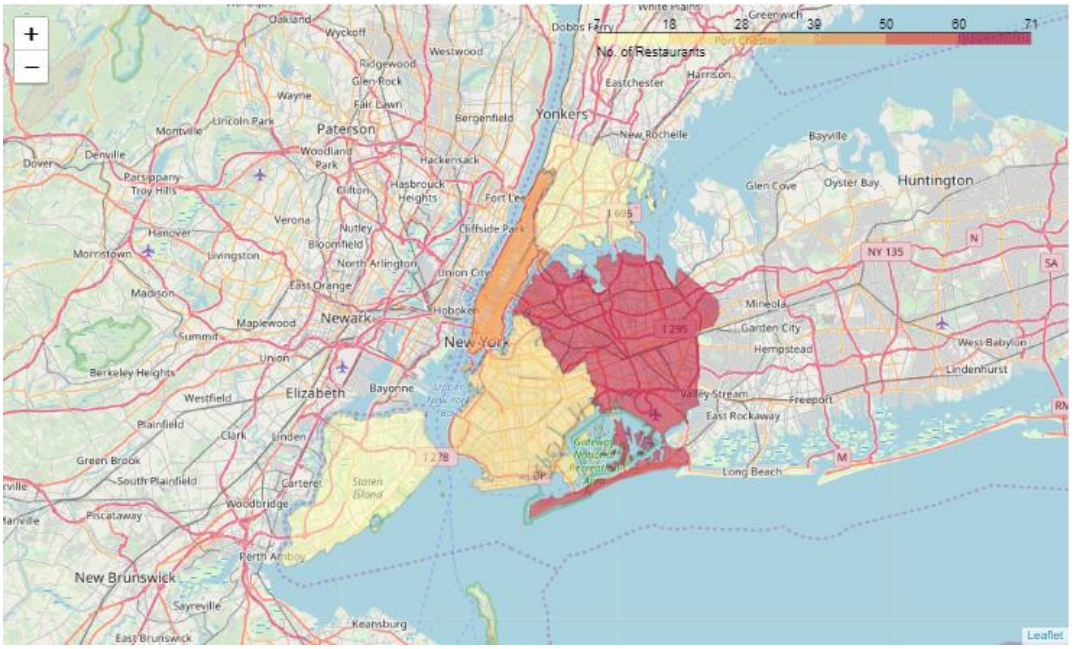
12. Finally, we will visualize the Neighborhoods and Borough based on average Rating using python's Folium library.

Neighborhoods based on average rating:

Boroughs based on average rating:



13. Finally, we will visualize the Borough based on Number of Indian cuisine restaurants using python's Folium library.



Results

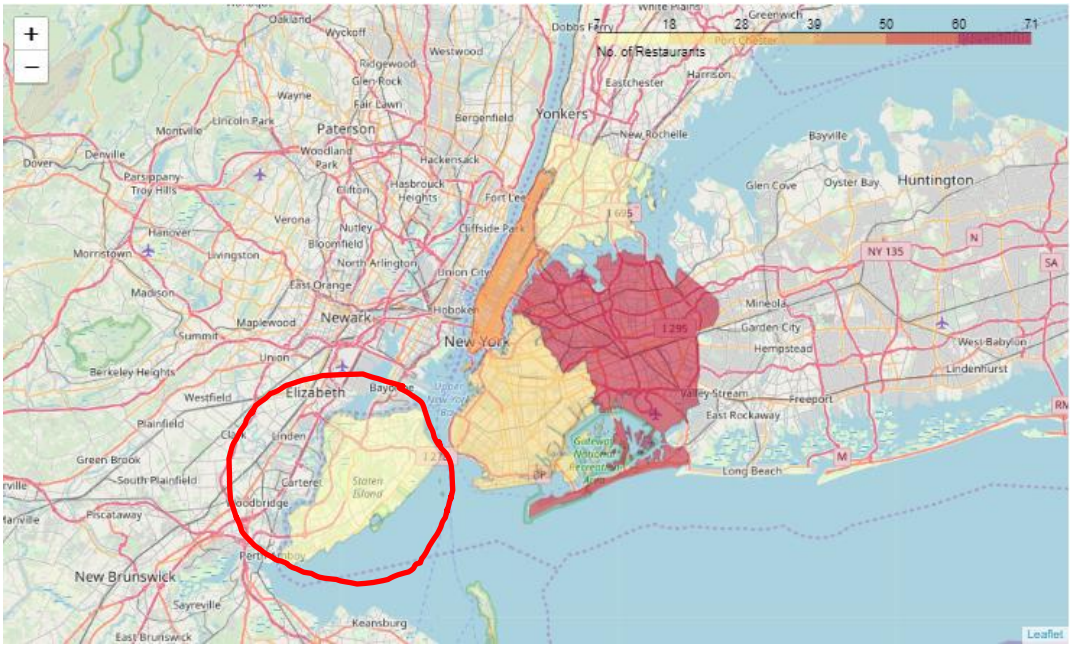
Hence we got our answer to the question.

The best place to make an Indian Cuisine Restaurant in New York is **Staten Island** as this island has the least number of restaurants and the least rating restaurants, hence with least competition. Hence there is a huge possibility of great business is a new restaurant is opened.

The sub-questions can be answered as:

1. **Which areas have potential Indian Restaurants?**
   Manhattan has the potential Indian restaurants
2. **Which areas lacks Indian Restaurants?**
   Bronx lacks Indian Restaurants
3. **Which is the best place to stay if you prefer Indian Cuisine?**
   Manhattan is the best place to stay in New York if you prefer Indian Cuisine

Other than this many other observations are being made: Astoria(Queens), Blissville(Queens), Civic Center(Manhattan) are some of the best neighborhoods for Indian cuisine.



Discussion

As observations noted from the map in the Results section, most of the Indian cuisine restaurants are concentrated in the Queens borough in the New York city. On the other hand, Bronx has least number of Indian cuisine restaurants in the neighborhoods. This represents a great opportunity and high potential areas to open new Indian cuisine restaurants as there is very little to no competition from existing restaurants, but the relative rating of the restaurants are more than that of the Staten islands. From another perspective, the results also show that the oversupply of Indian cuisine restaurants mostly happened in the Queens but also have a low average rating than that of the Manhattan. Therefore, this project recommends Manhattan as the best place to stay if you prefer Indian cuisine. And as the main outcome of this project it is stated that restaurant owners to capitalize on these findings to open new Indian Cuisine Restaurants in neighborhoods in Staten Islands with little to no competition.

Limitations and Suggestions for Future Research

In this project, we only consider two factors i.e. frequency of occurrence of Indian cuisine restaurants and the frequency of good rating restaurants, there are other factors such as population and income of residents that could influence the location decision of a new Indian cuisine restaurants. However, to the best knowledge of mine such data are not available to the neighborhood level required by this project. Future research could devise a methodology to estimate such data to be used in the clustering algorithm to determine the preferred locations to open a new Indian Cuisine Restaurants. In addition, this project made use of the Developer Account of Foursquare API that came with limitations as to the number of API calls and results returned. Future research could make use of paid account to bypass these limitations and obtain more results.

Conclusion

In this project, I have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing required data analysis, and lastly providing recommendations to the relevant restaurants owners and investors regarding the best locations to open a new Indian Cuisine Restaurants. To answer the business question that was raised in the introduction section, the answer proposed by this project is: **The neighborhoods in Staten Island in New York City are the most preferred locations to open a new Indian Cuisine Restaurants.** The findings of this project will help the relevant restaurant owner and investors to capitalize on the opportunities on high potential locations in their decisions to open a new Indian Cuisine Restaurants.