Applied Data Science Capstone Project
IBM/Coursera

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. New York City was home to nearly 8.5 million people in 2014, accounting for over 40% of the population of New York State and a slightly lower percentage of the New York metropolitan area, home to approximately 23.6 million. 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.

New York City has also been a major point of entry for immigrants; the term "melting pot" was coined to describe densely populated immigrant neighborhoods on the Lower East Side. As many as 800 languages are spoken in New York, making it the most linguistically diverse city in the world. English remains the most widely spoken language, although there are areas in the outer boroughs in which up to 25% of people speak English as an alternate language, and/or have limited or no English language fluency. English is least spoken in neighborhoods such as Flushing, Sunset Park, and Corona.   With its diverse culture, comes diverse food items. There are many restaurants in New York City, each belonging to different categories like Chinese, Indian, and French etc.

For many Indian residing in the US or particularly in the New York City, its hectic to find Indian cuisine restaurant. As a result, a lot of Indian students and workers look for their home cuisine in the foreign countries, and this brings to the development of a lot of Indian restaurants in the US. For retailers, the central location and the large crowd, low neighboring 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 factors that determines whether the mall will be success or a failure.

Problems

Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the main business question: What is best location in New York City for an Indian Cuisine?

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 a new Indian cuisine restaurant in the New York City.

Data Section

For this project we need the following data:

- ❖ New York City data that contains list Boroughs, Neighborhoods along with their latitude and longitude.
    1. Data Source: https://cocl.us/new_york_dataset
    2. Description: This data set is the most important dataset that contains the information of the neighborhoods and the latitude & longitude as well.

- ❖ List of Chinese restaurants in each neighborhood of New York City.
    1. Data Source: Foursquare API
    2. Description: This data provides all the venues in each neighborhood. Filter the venues to get only Chinese restaurants.

- ❖ GeoSpace Dataset:
    1. Data Source: https://data.cityofnewyork.us/CityGovernment/Borough-Boundaries/tqmj-j8zm
    2. Description: Get the NY Borough boundaries that 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"

2. We will find all venues for each neighbourhood using Foursquare API

```
new_york_data=get_new_york_data()
```
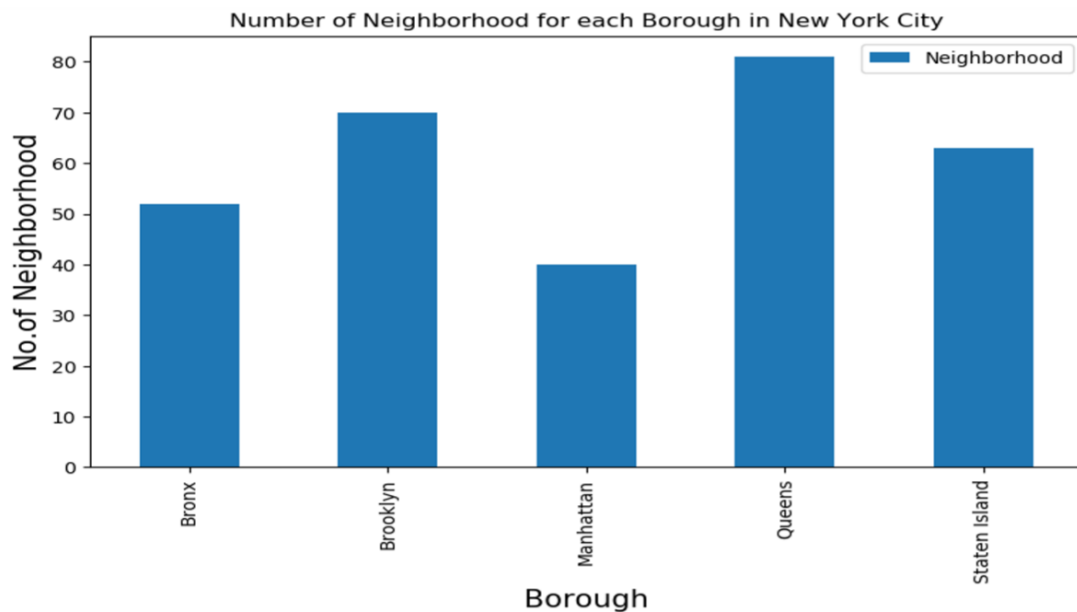
```
new_york_data.head()
```

	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

```
new_york_data.shape
```

(306, 4)

The above result shows that there are 306 different Neighborhoods in New York.

3. We will then filter out all venues with Indian restaurant for further analysis.

```
( 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:1
( 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
```

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

	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

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

	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

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

6. Next we will consider all the neighborhoods with average rating greater or equal 9.0 to visualize on map.
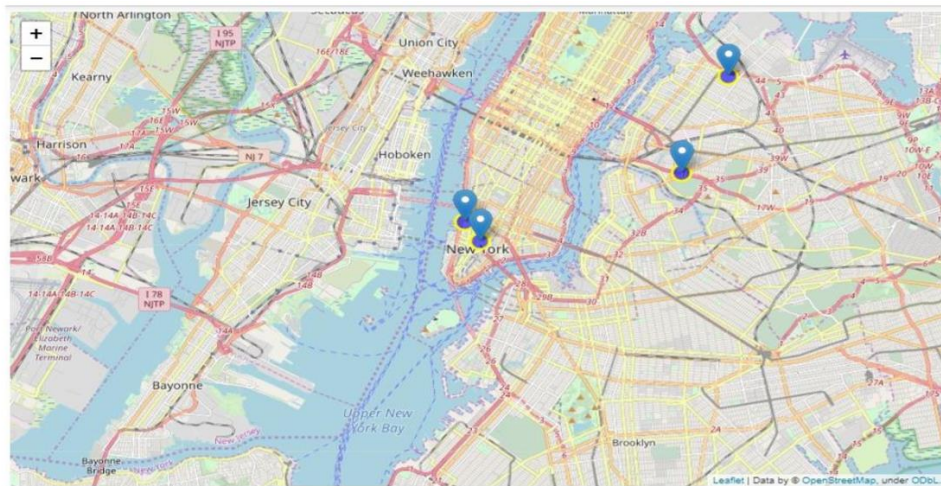
	Neighborhood	Average Rating
0	Astoria	9.0
5	Blissville	9.0
12	Civic Center	9.1
69	Tribeca	9.1

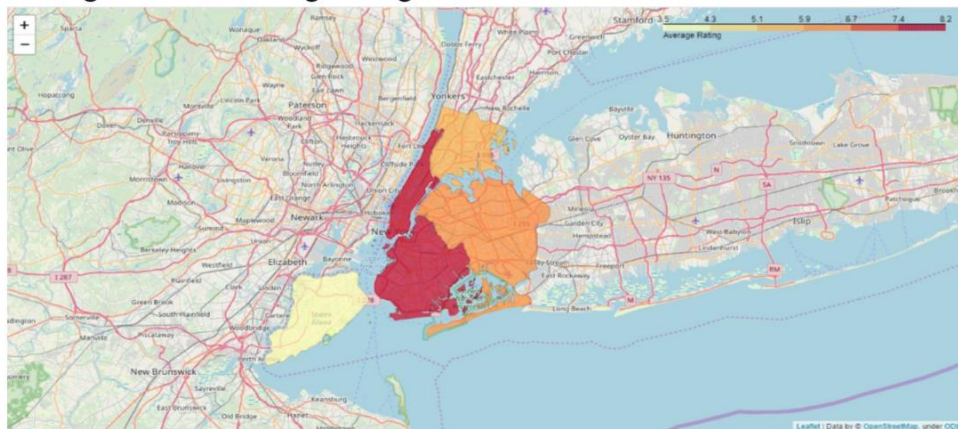7. We will join this dataset to original New York data to get longitude and latitude.

	Borough	Neighborhood	Latitude	Longitude	Average Rating
0	Queens	Astoria	40.768509	-73.915654	9.0
1	Queens	Blissville	40.737251	-73.932442	9.0
2	Manhattan	Civic Center	40.715229	-74.005415	9.1
3	Manhattan	Tribeca	40.721522	-74.010683	9.1

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

Neighborhoods based on average rating:

Borough based on average rating:

Result

So now we can answer the questions asked above in the Questions section:

Answers:

For Main Question: Astoria (Queens), Blissville (Queens), Civic Center (Manhattan) are some of the best neighborhoods for Indian cuisine.

1. Manhattan have potential Indian Restaurant Market.
2. Staten Island ranks last in average rating of Indian Restaurants where lacks Indian food service.
3. Manhattan is the best place to stay if you prefer Indian Cuisine.

Conclusion

In this report, I explore the Indian restaurants in New York City and successfully choose three best locations for an Indian Cuisine. The improvement lies in the preciseness of coding and more convenient logic to process.