Capstone Project - The Battle of Neighbourhoods

Shubham Singh

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.

Throughout its history, New York City has been a major point of entry for immigrants; the term "melting pot" was coined to describe densely populated immigrant neighbourhoods 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 neighbourhoods 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, French etc.

So as part of this project, we will list and visualize all major parts of New York City that has great Indian restaurants.

Data

For this project we need the following data:

- New York City data that contains list Boroughs, Neighbourhoods 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 neighbourhoods of New York city.
- Indian restaurants in each neighbourhood of new york city.
    - Data source: Foursquare API
    - Description: By using this api we will get all the venues in each neighbourhood. We can filter these venues to get only Indian restaurants.
- Geo Space 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.

Approach

- Collect the new york city data from https://cocl.us/new_york_dataset
- Using Four Square API we will find all venues for each neighbourhood.
- Filter out all venues that are Indian Restaurants.
- Find rating, tips and like count for each Indian Restaurants using Four Square API.
- Using rating for each restaurant, we will sort that data.
- Visualize the Ranking of neighbourhoods using folium library(python)

Business Problem:

Data that might contribute to determining -:

- What is best location in New York City for Indian Cuisine?
- Which areas have potential Indian Restaurant Market?
- Which all areas lack Indian Restaurants?
- Which is the best place to stay if I prefer Indian Cuisine?

Analysis

We will import the required libraries for python.

- pandas and numpy for handling data.
- request module for using FourSquare API.
- Geopy to get co-ordinates of City of New York.
- folium to visualize the results on a map

Data Cleaning

Now we define a function to get the geocodes i.e latitude and longitude of a given location using geopy.

We define a function to interact with Four Square API and get top 100 venues within a radius of 1000 metres for a given latitude and longitude. Below function will return us the venue id, venue name and category.
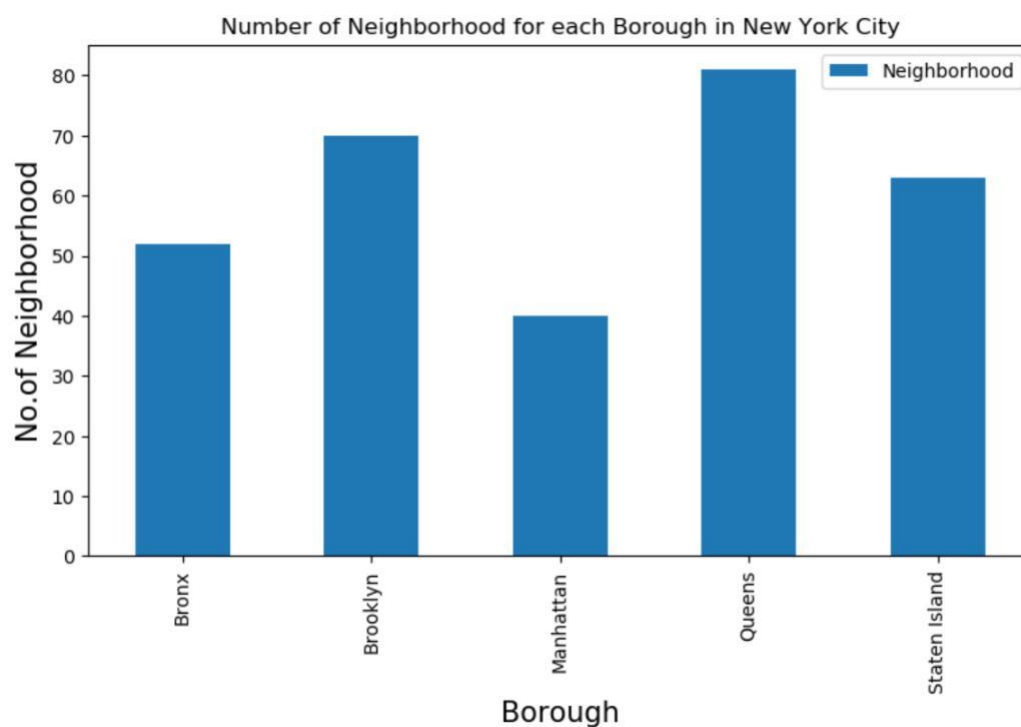
Now we will define a function to get venue details like like count , rating , tip counts for a given venue id. This will be used for ranking.

Now we define a function to get the New York city data such as Boroughs, Neighbourhoods along with their latitude and longitude.

We will call the above function to get the New York city data.

Exploratory Data Analysis

So there are total of 306 different Neighbourhoods in New York

Number of Neighborhood for each Borough in New York City

We see that Queens has highest number of neighbourhoods

Now we will collect Indian Restaurants for each Neighbourhood

Now that we have got all the Indian Restaurants in New York city , we will analyze it
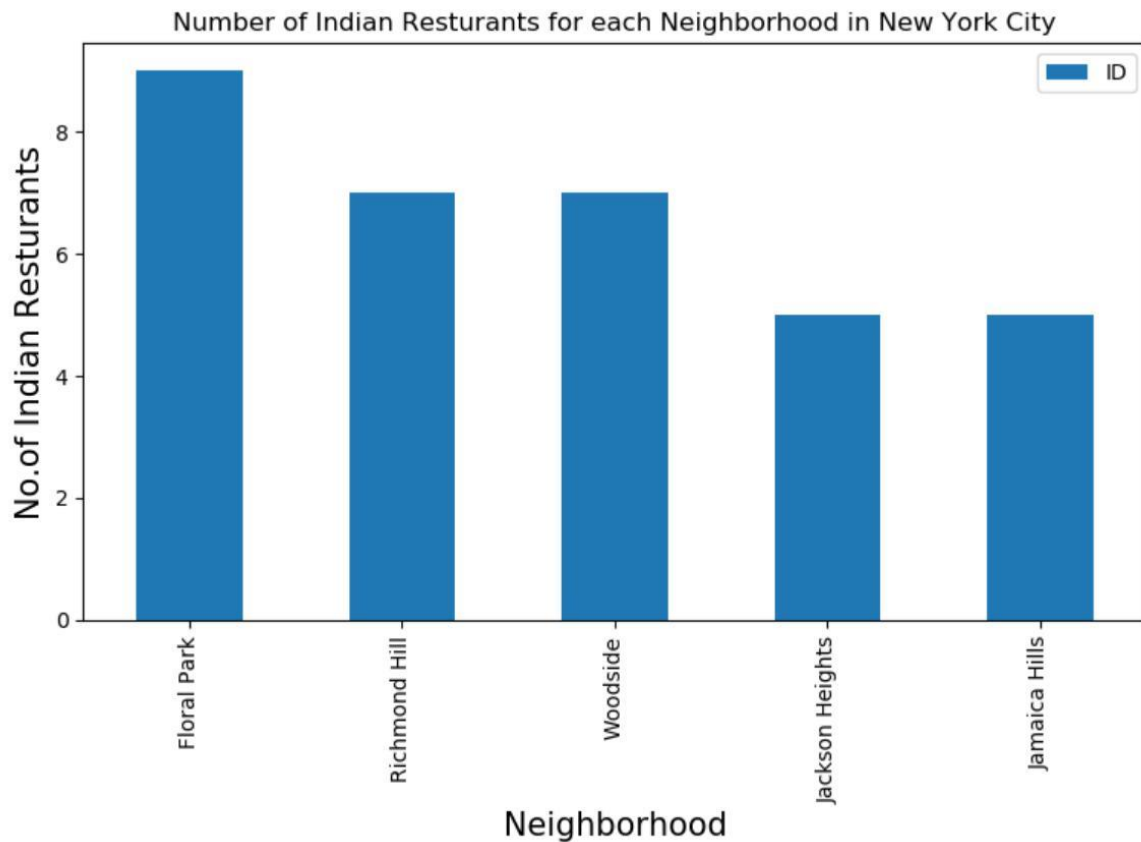
	Borough	Neighborhood	ID	Name
0	Bronx	Riverdale	4c04544df423a593ac83d116	Cumin Indian Cuisine
1	Bronx	Kingsbridge	4c04544df423a593ac83d116	Cumin Indian Cuisine
2	Bronx	Woodlawn	4c0448d9310fc9b6bf1dc761	Curry Spot
3	Bronx	Parkchester	4c194631838020a13e78e561	Melanies Roti Bar And Grill
4	Bronx	Spuyten Duyvil	4c04544df423a593ac83d116	Cumin Indian Cuisine

We got 153 Indian Restaurants across New York City

We see that Queens has the largest number of indian Restaurants

So Floral Park in Queens has the highest number of Indian Restaurants with a total count of 9.

Now we will get the ranking of each Restaurant for further analysis.

So we got data for all Restaurants Now let's save this data to a csv sheet. In case we by mistake modify it. As the number of calls to get details for venue are premium call and have limit of 500 per day, we will refer to saved data sheet csv if require.

Let's verify the data from saved csv file

We see that values like Likes, Tips are string values. We would need to convert them into float for further analysis

Now the data types looks correct

Now let's visualize neighbourhood with maximum average rating of Restaurants

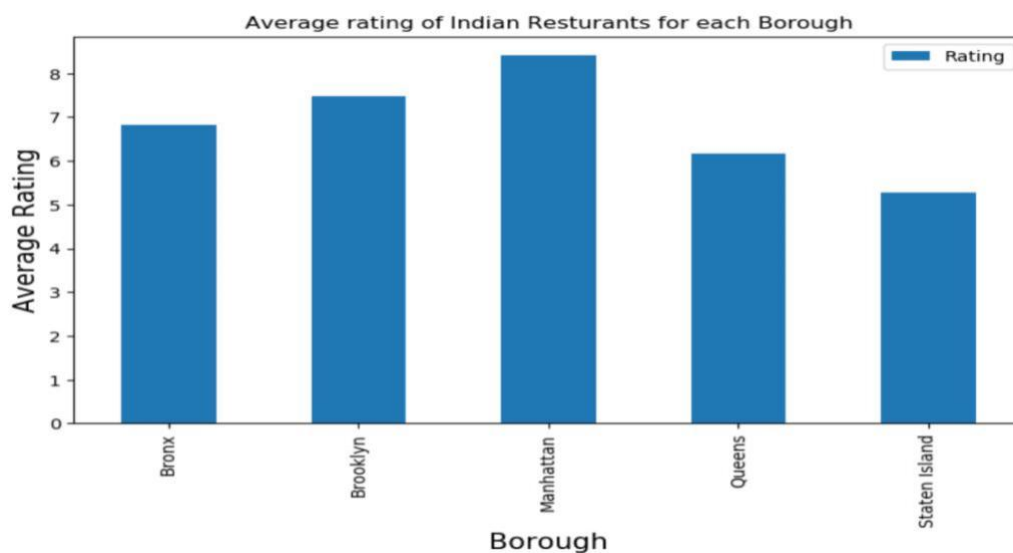|	Neighborhood	Average Rating
0	Astoria	9.200000
71	Sunnyside	9.200000
75	Tribeca	9.200000
5	Blissville	9.200000
11	Civic Center	9.200000
47	Midtown	9.000000
48	Midtown South	9.000000
30	Gramercy	8.866667
66	Roosevelt Island	8.700000
53	North Side	8.700000

Above are the top neighbourhoods with top average rating of Indian

Restaurants Similarly these are the average rating of Indian Restaurants for

each Borough

	Borough	Average Rating
2	Manhattan	8.414706
1	Brooklyn	7.478571
0	Bronx	6.812500
3	Queens	6.155844
4	Staten Island	5.266667

Let's visualize it

We will consider all the neighbourhoods with average rating greater or equal 9.0 to visualize on map

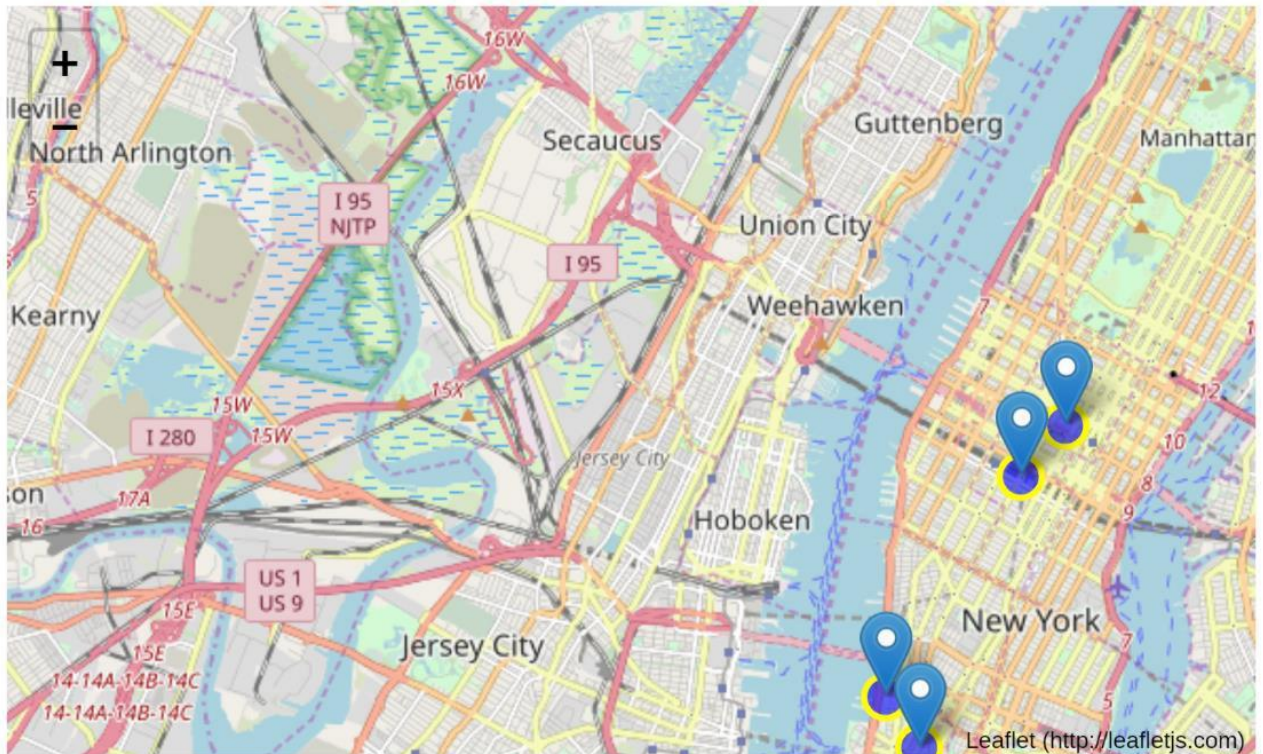|	Borough_x	Neighborhood	Latitude_x	Longitude_x	Average Rating	Label	Borough_y	Latitude_y	Longitude_y
0	Queens	Astoria	40.768509	-73.915654	9.2	Astoria, Queens(9.2)	Queens	40.768509	-73.915654
1	Queens	Blissville	40.737251	-73.932442	9.2	Blissville, Queens(9.2)	Queens	40.737251	-73.932442
2	Manhattan	Civic Center	40.715229	-74.005415	9.2	Civic Center, Manhattan(9.2)	Manhattan	40.715229	-74.005415
3	Manhattan	Midtown	40.754691	-73.981669	9.0	Midtown, Manhattan(9.0)	Manhattan	40.754691	-73.981669
4	Manhattan	Midtown South	40.748510	-73.988713	9.0	Midtown South, Manhattan(9.0)	Manhattan	40.748510	-73.988713
5	Queens	Sunnyside	40.740176	-73.926916	9.2	Sunnyside, Queens(9.2)	Queens	40.740176	-73.926916
6	Queens	Sunnyside	40.740176	-73.926916	9.2	Sunnyside, Queens(9.2)	Staten Island	40.612760	-74.097126
7	Staten Island	Sunnyside	40.612760	-74.097126	9.2	Sunnyside, Staten Island(9.2)	Queens	40.740176	-73.926916
8	Staten Island	Sunnyside	40.612760	-74.097126	9.2	Sunnyside, Staten Island(9.2)	Staten Island	40.612760	-74.097126

We will join this dataset to original new york data to get lonitude and latitude

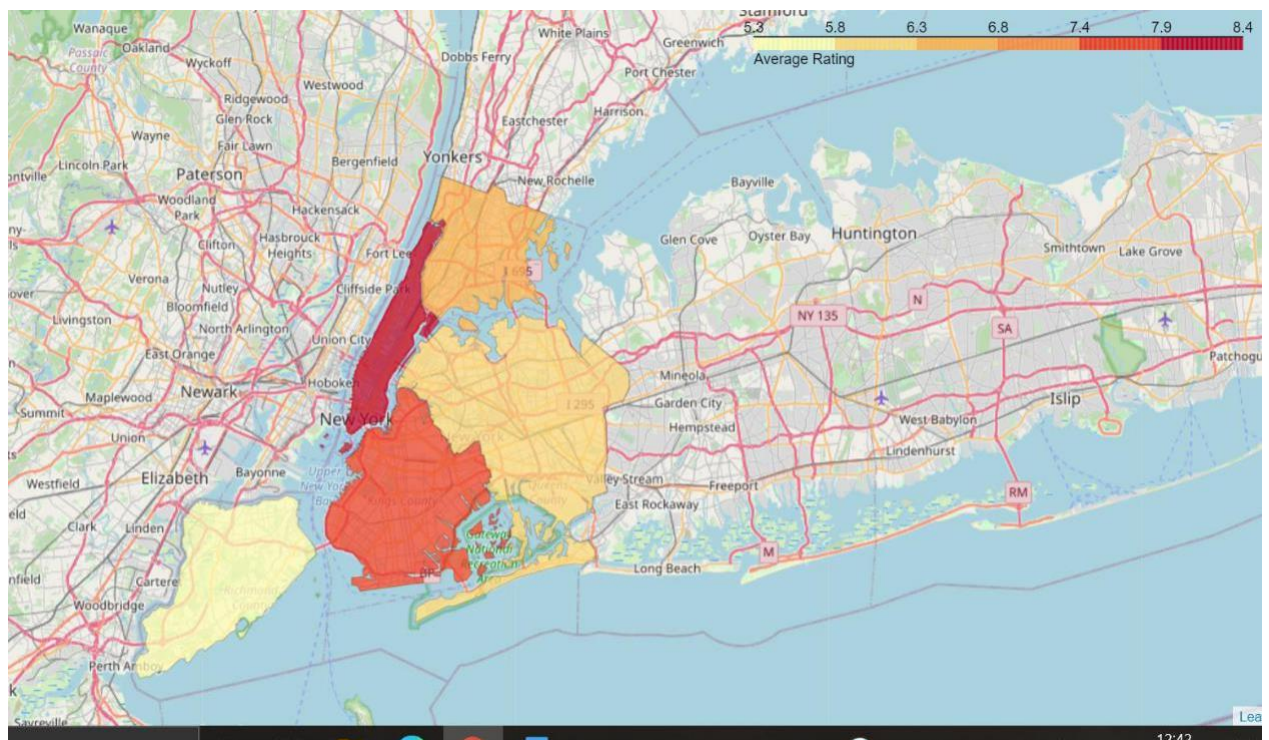|	Borough	Neighborhood	Latitude	Longitude	Average Rating
0	Queens	Astoria	40.768509	-73.915654	9.2
1	Queens	Blissville	40.737251	-73.932442	9.2
2	Manhattan	Civic Center	40.715229	-74.005415	9.2
3	Manhattan	Midtown	40.754691	-73.981669	9.0
4	Manhattan	Midtown South	40.748510	-73.988713	9.0
5	Queens	Sunnyside	40.740176	-73.926916	9.2
6	Staten Island	Sunnyside	40.612760	-74.097126	9.2
7	Queens	Sunnyside	40.740176	-73.926916	9.2
8	Staten Island	Sunnyside	40.612760	-74.097126	9.2
9	Queens	Sunnyside	40.740176	-73.926916	9.2
10	Staten Island	Sunnyside	40.612760	-74.097126	9.2
11	Queens	Sunnyside	40.740176	-73.926916	9.2
12	Staten Island	Sunnyside	40.612760	-74.097126	9.2
13	Manhattan	Tribeca	40.721522	-74.010683	9.2

Now we will show this data on a map

Let's add a new field to dataframe for labelling purpose

Now that we have visualized the Neighbourhoods.

Let's Visualize Boroughs based on average Rating

Conclusion

- Astoria(Queens), Blissville(Queens), Civic Center(Manhattan) are some of the best neighbourhoods for Indian cuisine.
- Manhattan have potential Indian Restaurant Market/
- Staten Island ranks last in average rating of Indian Restaurants.
- Manhattan is the best place to stay if you prefer Indian Cuisine.

Limitations

- The ranking is purely on basis of rating of Restaurants
- The accuracy of data depends purely depends on the data provided by FourSquare