Applied Data Science Capstone by IBM/Coursera
Avani Bhatnagar
August 11, 2021

Introduction

Location is very important when it comes to opening a restaurant. If a family wants to open up a restaurant in a city, they need to know different things about the location to be able to strategically select the correct neighborhood which will attract the most crowd, is affordable to open up a restaurant at, has lesser competition, population etc. In this situation, a family living in New York decides to open an Indian restaurant in the city. They are in the very initial stages of planning this business and one of their first points of focus is deciding the location of the restaurant. They know a lot of people enjoy Indian food, but for the most profit they need to able to find something that will also consider factors such as population, relative competitiveness in that area, etc.

Their goal is find maximum success while putting in least risks since they are using a lot of their family savings into this!

Data Acquisition and Cleaning

The Data is the most important part of any data science/analysis project. We need to make sure we have all the relevant data before we can dig into the "science" part of the process.

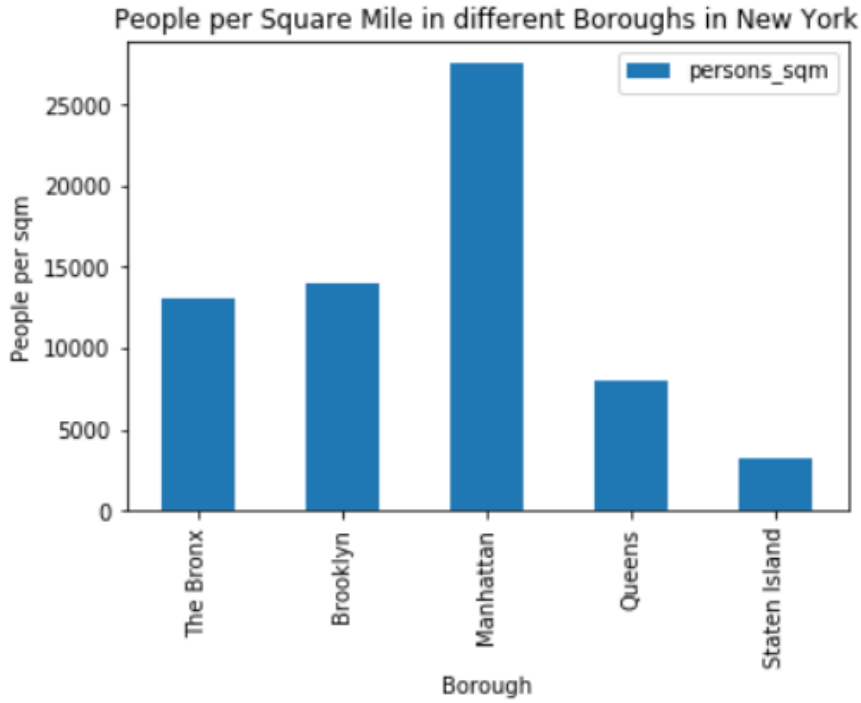
- New York Data: We use this dataset: https://en.wikipedia.org/wiki/New_York_City to get information about the various New York neighborhoods which will help us explore and decide the best location for an Indian restaurant.
- Foursquare API: We use Foursquare API to get various restaurants in New York such as Indian Cuisine. This will help the business get an idea of what's the best location to open their restaurant in terms on relative competition in the same business.
- Geospatial data - We use https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm to get the location details of New York - it gives us the Latitude and Longitude values for each Borough in New York. This will be greatly helpful in generating maps and visualizing the overall results.
- Demographic data - We use https://en.wikipedia.org/wiki/Demographics_of_New_York_City to get the demographic data for NYC's jurisdiction.

Exploratory Data Analysis

We start by looking at all boroughs of NYC with their population, gdp per capita and persons per square mile.

	borough	county	population	gdp_per_capita	persons_sqm
0	The Bronx	Bronx	1418207.0	42.10	13006.0
1	Brooklyn	Kings	2559903.0	70.82	13957.0
2	Manhattan	New York	1628706.0	22.83	27544.0
3	Queens	Queens	2253858.0	108.53	8018.0
4	Staten Island	Richmond	476143.0	58.37	3150.0

Then we visualize it to get an idea about the population density of each borough

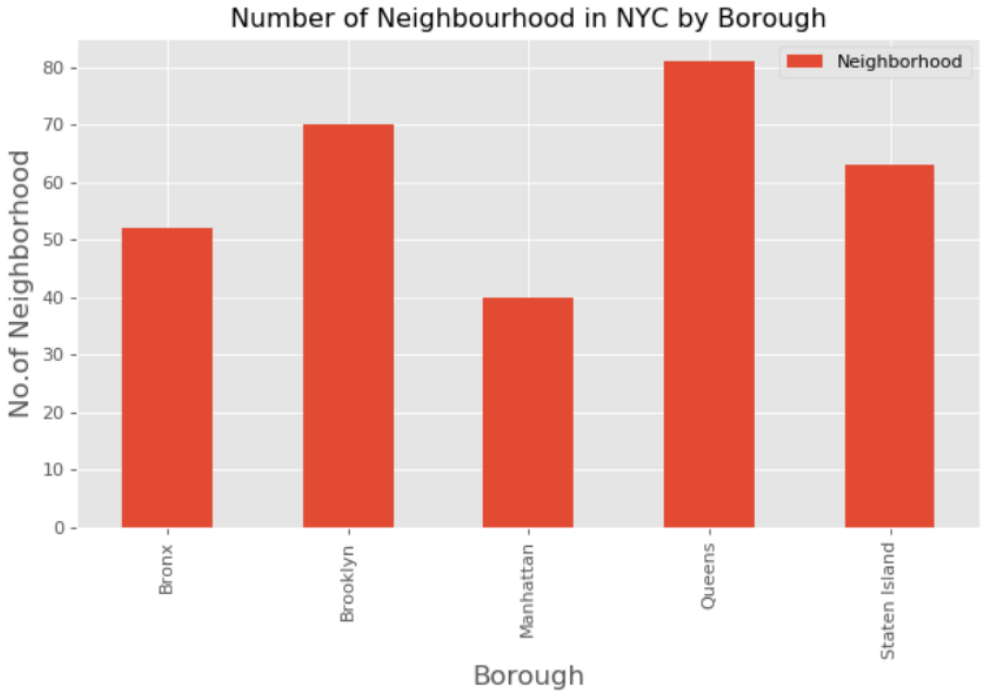


Then we look at the demographics of all the boroughs to get an idea of what % of people live there are Asian since we are concerned with Indian Cuisine.

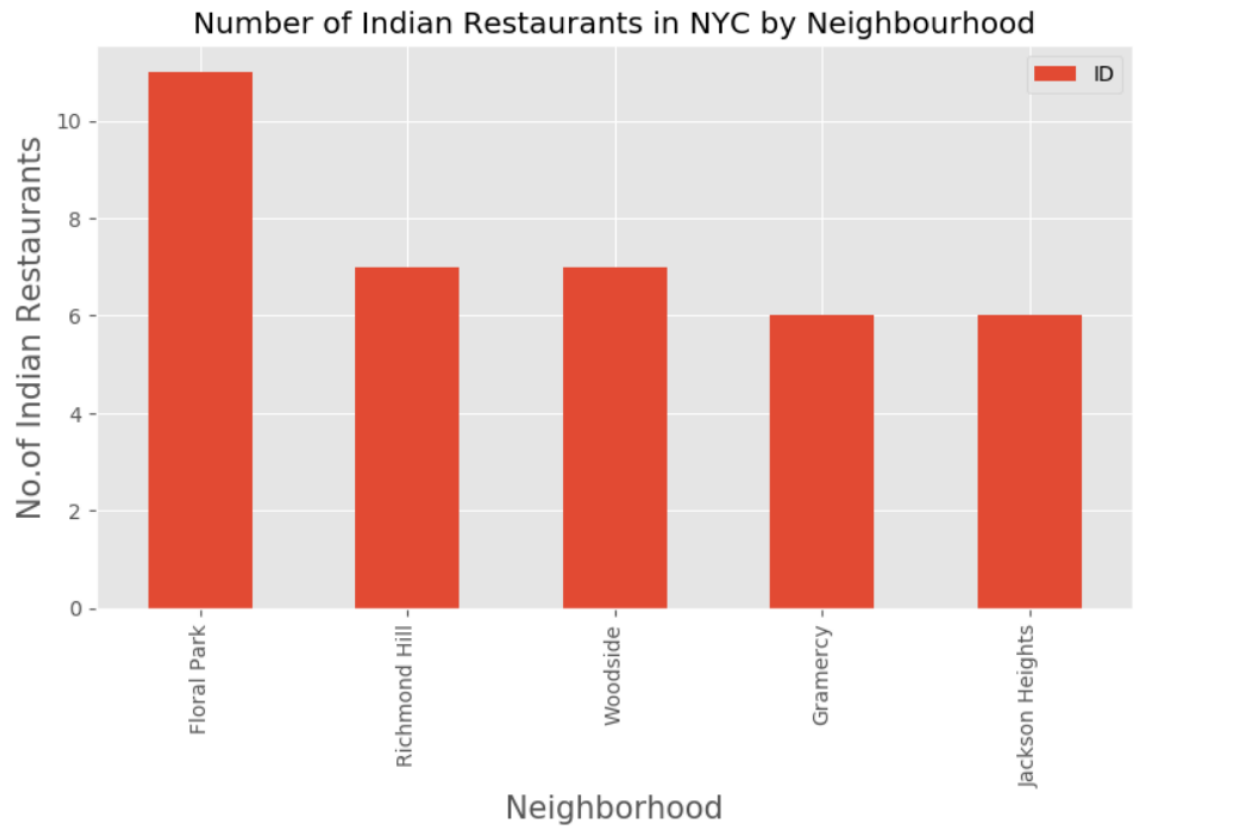
	jurisdiction	population_census	White	African_Amercian	Asian	Other	Mixed_Race	Hispanic_Latino_of_other_race
0	Queens	2,229,379	44.1	20.0	17.6	12.3	6.1	NaN
1	Manhattan	1,537,195	54.4	17.4	9.4	14.7	4.1	NaN
2	Bronx	1,332,650	29.9	35.6	3.0	25.7	5.8	NaN
3	Staten Island	443,728	77.6	9.7	5.7	4.3	2.7	NaN
4	NYC Total	8,008,278	44.7	26.6	9.8	14.0	4.9	NaN

Solution/Finding the best Neighborhood

In the process of finding the best neighborhood we visualize all boroughs by number of neighborhoods to get a population density of each.



Then we use the Foursquare API to get the Indian Restaurants in each neighborhood.



Then we combine data sets to get a list of Neighborhoods with their boroughs and Indian Restaurant name, likes, and ratings.

	Borough	Neighborhood	ID	Name	Likes	Rating	Tips
0	Bronx	Woodlawn	4c0448d9310fc9b6bf1dc761	Curry Spot	5	7.6	10
1	Bronx	Williamsbridge	0	0	0	0.0	0
2	Bronx	Parkchester	4c194631838020a13e78e561	Melanies Roti Bar And Grill	3	6.3	2
3	Bronx	Spuyten Duyvil	4c04544df423a593ac83d116	Cumin Indian Cuisine	13	5.8	9
4	Bronx	Concourse	551b7f75498e86c00a0ed2e1	Hungry Bird	8	7.1	4

We then find the restaurants with the highest ratings.

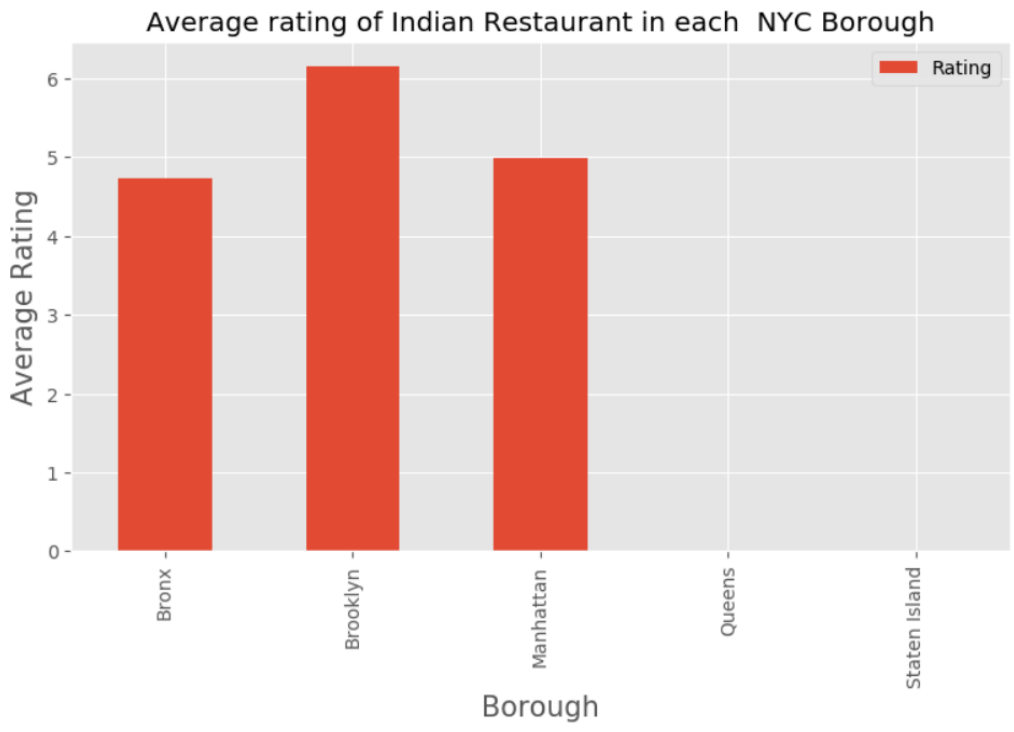
We found neighborhoods with the highest ratings

```
[81]: ny_neighborhood_stats.sort_values(['Average Rating'],ascending=False).head(10)
```

Out[81]:

	Neighborhood	Average Rating
69	Tribeca	9.10
52	Prospect Heights	9.00
31	Greenwich Village	9.00
74	West Village	8.85
20	East Village	8.70
26	Fort Greene	8.70
11	Chelsea	8.70
44	Midtown	8.70
13	Clinton Hill	8.70
53	Prospect Lefferts Gardens	8.70

We group them by boroughs to get average rating for each borough.

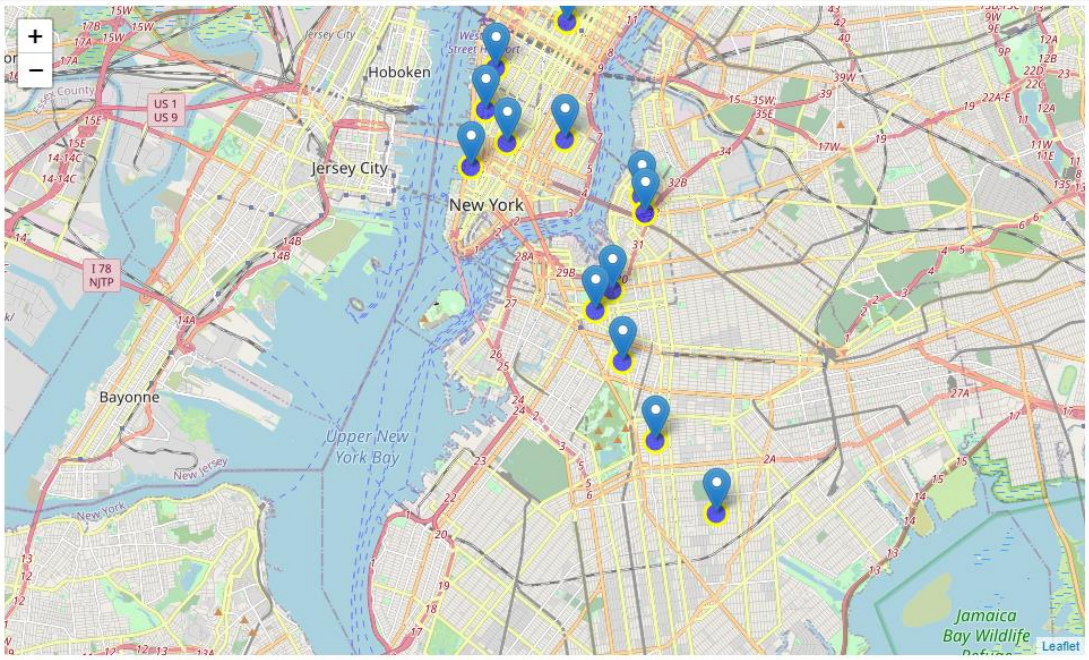


We visualize the neighborhoods with highest Indian restaurant ratings to help our user visualize the likely spots to open an Indian restaurant.

Neighborhoods with highest ratings

In [91]:
```
# add pop-up text to each marker on the map
for lat, lng, label in ny_neighborhood_stats[['Latitude','Longitude','Label']].values:
    folium.Marker([lat, lng], popup=label).add_to(ny_map)
# add incidents to map
ny_map.add_child(incidents)
```

Out[91]:



Conclusion

From all our analysis and results & findings, we can come to a conclusion that Manhattan would be a great place to open an indian restaurant since it has the highest population density and has the 2nd highest Asian population which will attract a large crowd. It also has some of the highest rated Indian restaurants so people probably stay in that area when looking for Indian food.

Another conclusion is that Queens can also be a good option if they want to be close to other Indian restaurants since it has the highest Asian population and it also has highest number of Indian restaurants.

This is a basic data science analysis. We can do further analysis using other algorithms in the future