Coursera Capstone Project: Week 4

Created by: Poulomi Saha

Introduction

New York City (NYC), is the most populous city in the United States. With an estimated 2019 population of 8,336,817, New York City is also the most densely populated major city in the United States. New York City is composed of five boroughs, each of which is a county of the State of New York. The city and its metropolitan area constitute the premier gateway for legal immigration to the United States, which has a larger immigrant population than any other country, with 47 million immigrants as of 2015. A published article claims that there are about 710,000 Indians residing in NYC, which is about 9% of the total NYC population. Among many other things, Indians are known for their love towards Indian flavoursome, spicy food. In this final project, the best locations for eating & ordering out Indian as well as opening an Indian restaurant is explored throughout NYC. Hence this will not only benefit the Indians and the diverse cultured people who are fond of Indian food, but also can be a guide to newcomers into the NYC.

Discussion Module

- How many Indian restaurants are there in NYC and how are the
- they distributed along different neighbourhoods/boroughs?
- Which are the best locations to find Indian food joints in NYC?
- In which areas do the newcomer Indians decide to reside?
- In which neighbourhood should a potential restaurant owner invest to have best chance of success?

Data/Sample

In order to answer above questions, the following data is needed:

- Geospatial information of NYC, including names of neighbourhoods, boroughs as well as their geodata(latitude, longitude).
- Details such as name,menu,ratings of Indian restaurants in NYC.

Above information can either be gathered from internet and/or requested from Foresquare API using python and geopy data.

Methodology

1. NYC geospatial data will be collected from https://cocl.us/new_york_dataset.
2. Data of Indian restaurants will be requested by using Foresquare API and Geopy map.
3. Collected data will be processed and exploratory analysis will be performed on the data.
4. Cleaned data will be visualised using different python libraries.

References

1. https://en.wikipedia.org/wiki/New_York_City
2. https://www1.nyc.gov/assets/immigrants/downloads/pdf/moia_annual_report_2018_final.pdf
3. https://www.novacredit.com/resources/desis-in-new-york-a-guide-for-indian-newcomers-in-new-york-city/#:~:text=New%20York%20City%20houses%20one,Indian%20residents%20as%20of%202017.

Libraries used to drive through the project

1. Pandas: To create and manipulate data frames.
2. Folium: Python visualization library would be used to visualize the neighbourhood cluster distribution of using an interactive leaflet map.
3. Scikit Learn: To import k-means clustering.
4. JSON: Library to handle JSON files.
5. XML: To separate data from presentation and XML stores data in plain text format.
6. Geopy/Geocoder: To retrieve Location Data.
7. Beautiful Soup and Requests: To scrap and library to handle HTTP requests.
8. Matplotlib: To Python Plotting Module.

Results of the Analysis performed

Distribution of NYC through geospatial data.

1. Maximum number of neighbourhoods are in **Queens.**

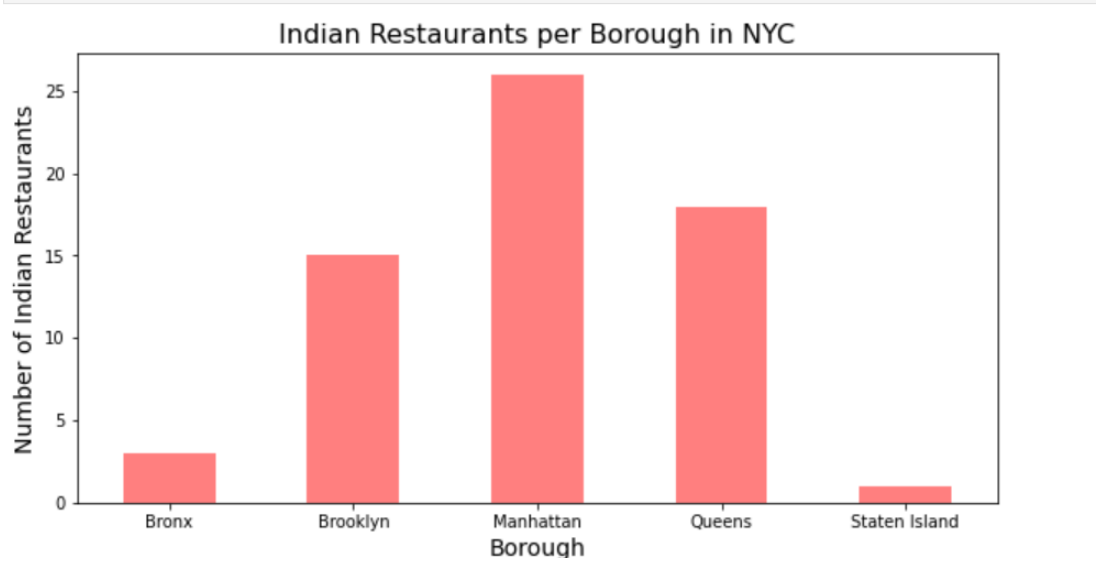
[7]:

Borough	Neighborhood	Latitude	Longitude
Bronx	52	52	52
Brooklyn	70	70	70
Manhattan	40	40	40
Queens	81	81	81
Staten Island	63	63	63

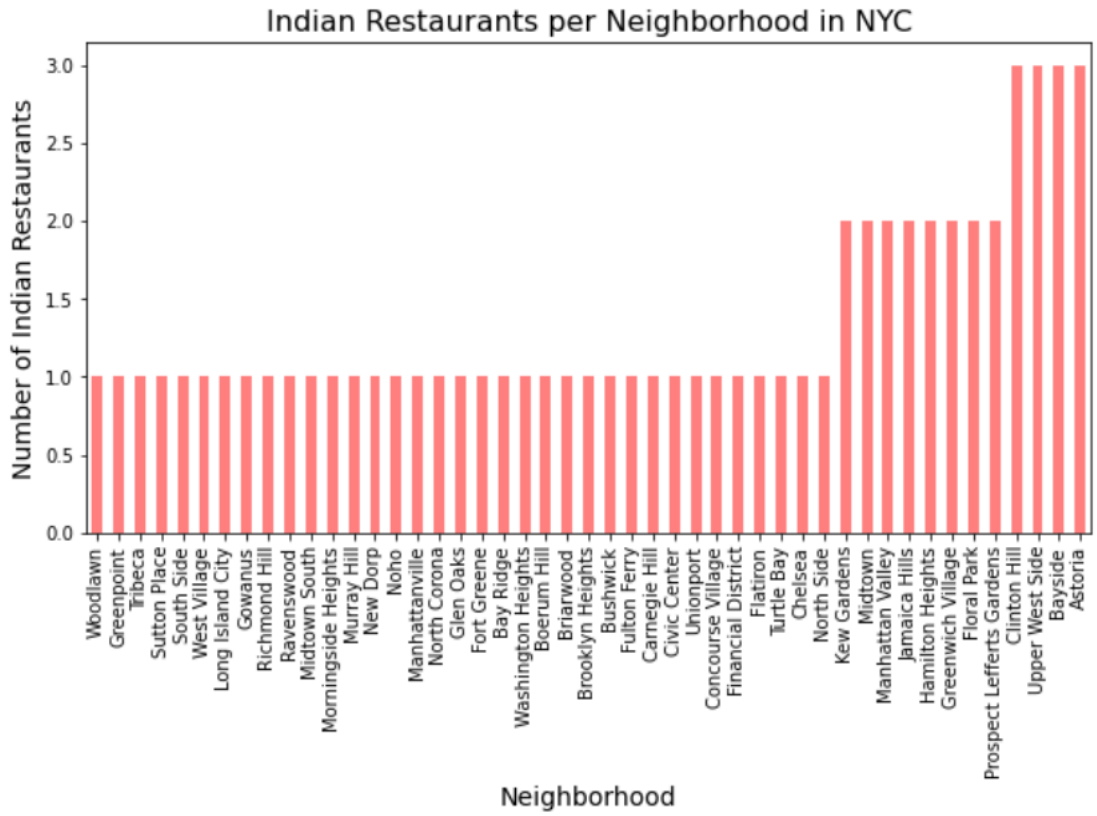
2. Maximum number of Indian restaurants are in **Manhattan** Borough.

[19]:

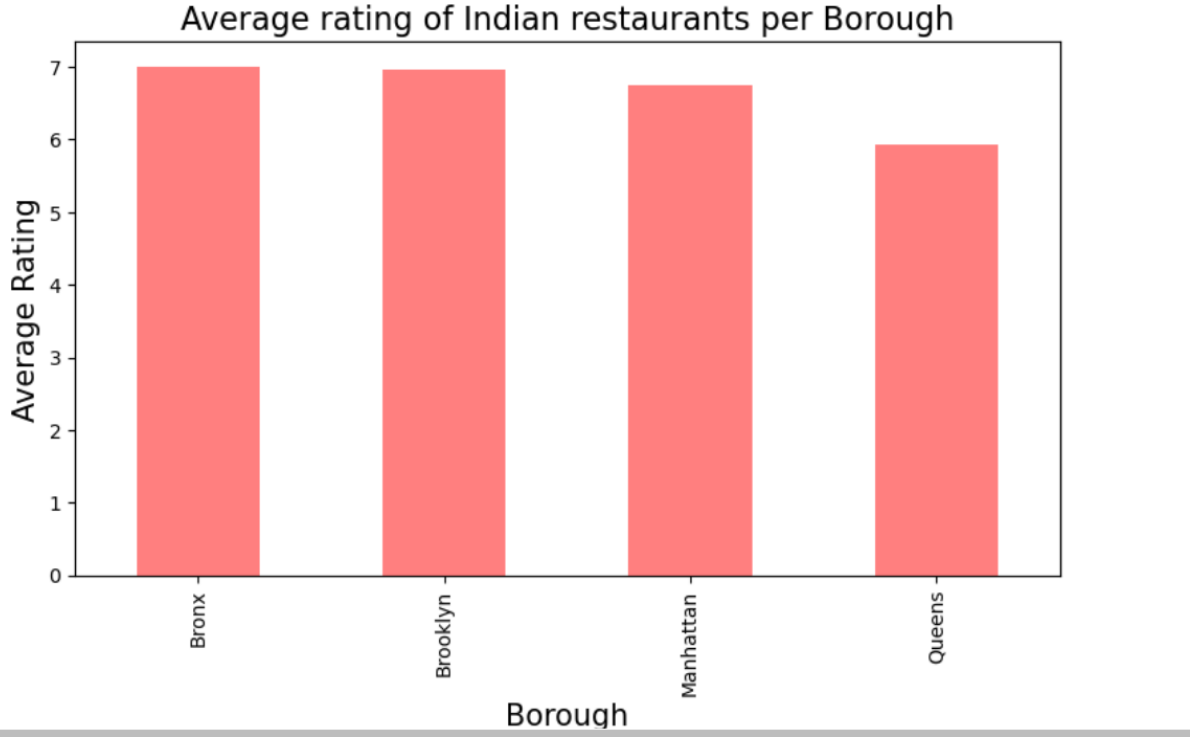
```
Borough
Bronx            3
Brooklyn        15
Manhattan       26
Queens          18
Staten Island    1
Name: ID, dtype: int64
```



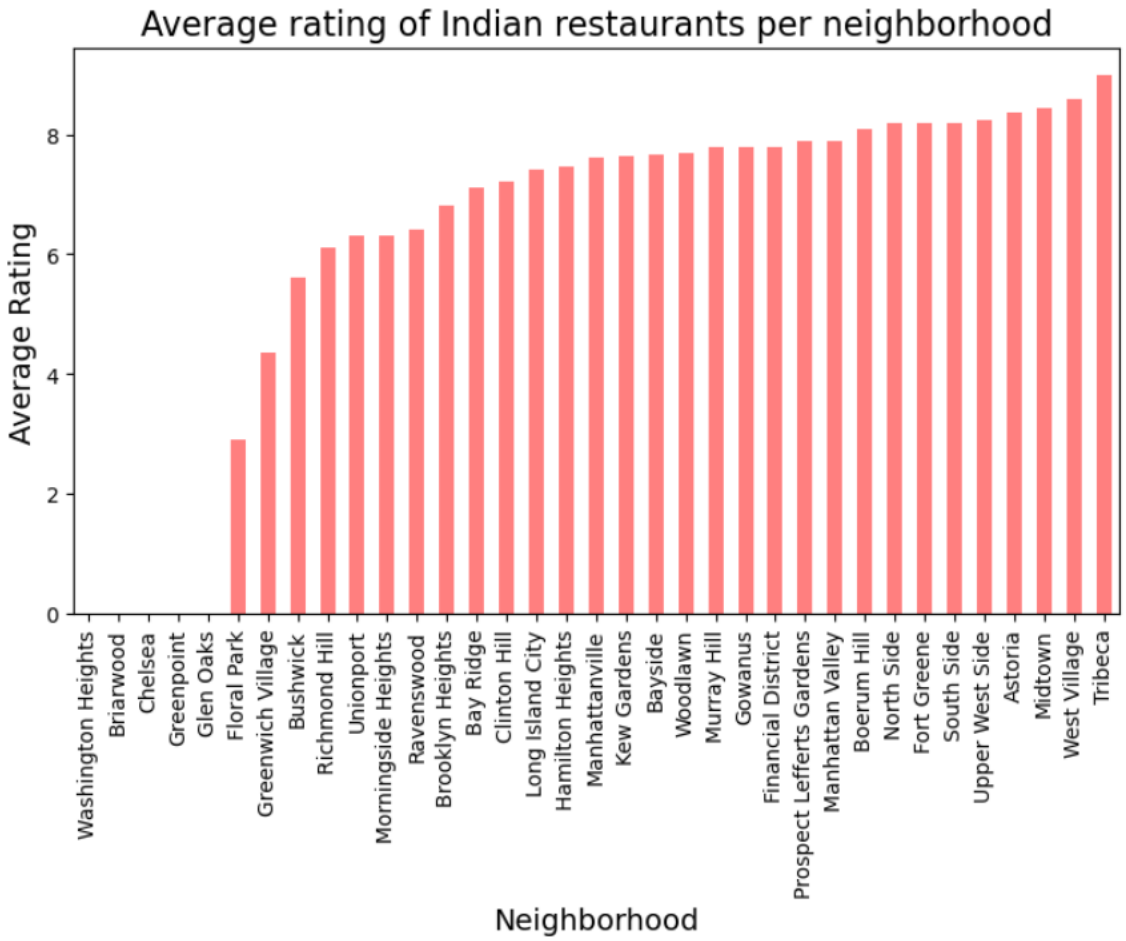
3. Maximum number of Indian restaurants in NYC neighbourhoods.



4. Average rating of Indian restaurants per Borough.



5. Average ratings of Indian restaurants per neighbourhood.



Best Location(s) to enjoy Indian cuisine in NYC

- Indian Restaurant in NYC with maximum Likes.

```
[33]: Borough                        Manhattan
      Neighborhood                     Midtown
      ID              49d91c12f964a520015e1fe3
      Name                The Kati Roll Company
      Likes                                 830
      Rating                                8.7
      Tips                                  255
      Name: 23, dtype: object
```

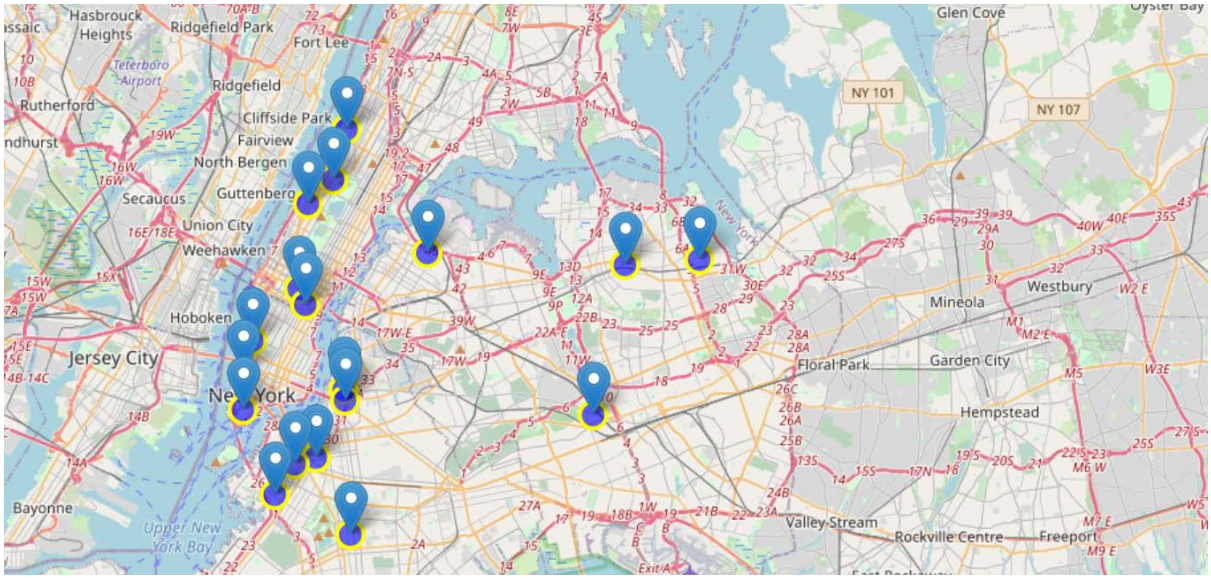
- Indian Restaurant in NYC with maximum Ratings.

```
[34]: Borough                        Manhattan
      Neighborhood                      Tribeca
      ID              4bbb9dbded7776b0e1ad3e51
      Name                     Tamarind TriBeCa
      Likes                                 601
      Rating                                  9
      Tips                                  150
      Name: 29, dtype: object
```

- Indian Restaurant in NYC with maximum Tips/Comments.

```
[35]: Borough                        Manhattan
      Neighborhood                     Midtown
      ID              49d91c12f964a520015e1fe3
      Name                The Kati Roll Company
      Likes                                 830
      Rating                                8.7
      Tips                                  255
      Name: 23, dtype: object
```

Data Visualization of Neighbourhoods with Indian restaurants having Average rating of >7.5



Conclusion

1. **Manhattan** has the best rated Indian restaurants.
2. Although Manhattan has the least number of neighbourhoods among other boroughs, it has the greatest number of Indian restaurants.
3. The **Bronx** has the least number of Indian restaurants per borough
4. Bronx and Staten island together have only 4 Indian restaurants
5. Based on the above analysis, in order to successfully venture into opening an Indian restaurant, it is recommended to invest in a restaurant in **Bronx**, especially on the border of Manhattan and Bronx, reason being:
a) Bronx is right next to Manhattan thus easily reachable by public transport.
b) Limited presence of Indian eateries, although average likes and ratings of the existing Indian joints.
6. However, it must be noted that this analysis is based on limited data.
Hence, a comprehensive analysis should be performed to reconfirm these findings.