

Capstone Project - The Battle of Neighborhoods

Applied Data Science Capstone Project

Indian Restaurant in New York



Kanika Sehgal

(May 26, 2021)

Introduction

New York City (NYC), often simply called New York, is the most populous city in the United States. With an estimated 2019 population of 8,336,817 distributed over about 302.6 square miles (784 km²), New York City is also the most densely populated major city in the United States.

The restaurant industry is essential to New York City's social and economic fabric. From small, family-owned restaurants and food carts to four-star world-famous establishments, the City abounds with restaurants that offer cuisines from every corner of the globe. Since March 2020, the restaurant industry has been hit very hard by the COVID-19 pandemic. Mandatory closures, stay-at-home and social distancing orders, the onset of a severe economic recession, and travel restrictions have resulted in unprecedented upheaval for the industry. Due to these situations the Indian population living in New York is neither able to get food from their culture nor they could travel back to their country. As part of this project we need to explore the best locations for Indian restaurant throughout the city of New York for the owner such that it could reach to the target customer and could have great success and consistent profit despite the pandemic situation.

However, as with any business, opening a new restaurant will require serious considerations specifically due to pandemic situation.

Business Problem

The objective of this Capstone project is to analyze and select the best locations in the city of New York to open a new Indian restaurant. Using Data Science methodology and instruments such as Data Analysis and Visualization, this project aims to provide solutions to answer the business question:

- Where in the city of New York, should the investor open an Indian Restaurant?
- Which areas have large number of Indian Restaurant Market?
- Which all areas have a smaller number of Restaurants?
- Which areas have highest population of Indian customers?
- Which is the best place to stay if I prefer Indian Cuisine?
- What places are having best restaurant in New York?

Target Audience of this project and some demographic facts

This project is particularly useful to developers and investors looking to open or invest in an Indian restaurant in the city of New York. Overall, New York is a great place to open a restaurant with an ethnical cuisine. As New York is the most diverse city in the world (800 languages are spoken in New York). With its diverse culture, comes diversity in the food items. There are many restaurants in New York City, each belonging to different categories like Chinese, Indian, French, etc. Why did we decide to focus on Indian cuisine in our project? The Indian population living in New York is neither able to get food from their culture nor they could travel back to their country due to pandemic situation therefore having more number of Indian restaurants would serve as a blessing in this tough time.

Data

To solve the problem, we will need the following data:

- New York City data containing the neighborhoods and boroughs.
- Latitude and longitude coordinates of those neighborhoods. This is required to plot the map and get the venue data.
- Venue data, particularly data related to restaurants. We are going to use this data to perform further analysis of the neighborhoods.
- Python packages and Dependencies:
 - Pandas - Library for Data Analysis
 - NumPy - Library to handle data in a vectorized manner
 - JSON - Library to handle JSON files
 - Geopy - To retrieve Location Data
 - Requests - Library to handle http requests
 - Matplotlib - Python Plotting Module
 - Folium - Map rendering Library

Methodology

New York City data containing the neighborhoods and boroughs will be obtained from the open data source: https://cocl.us/new_york_dataset.

Geographical coordinates of the neighborhoods (latitude and longitude) using Python Geocoder package.

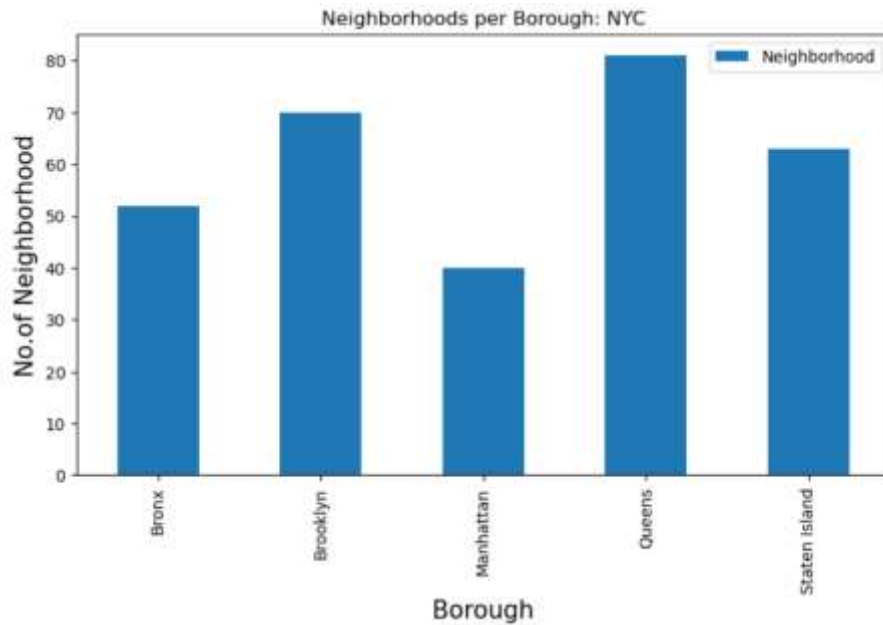
Finally, we will use Foursquare API to get the venue data for the neighborhoods defined at the previous step. This API has a database of more than 105 million places. This project would use Four-square API as its prime data gathering source. Many organizations are using to geo-tag their photos with detailed info about a destination, while also serving up contextually relevant locations for those who are searching for a place to eat, drink or explore. This API provides the ability to perform location search, location sharing and details about a business. Foursquare users can also use photos, tips and reviews in many productive ways to add value to the results.

This project will require using of many data science skills, from web scrapping (open source dataset), working with API (Foursquare), data cleaning, data wrangling, to map visualization (Folium). In the next Methodology section, we will discuss and describe any exploratory data analysis that we did, any inferential statistical testing that we performed, and what machine learning techniques were used.

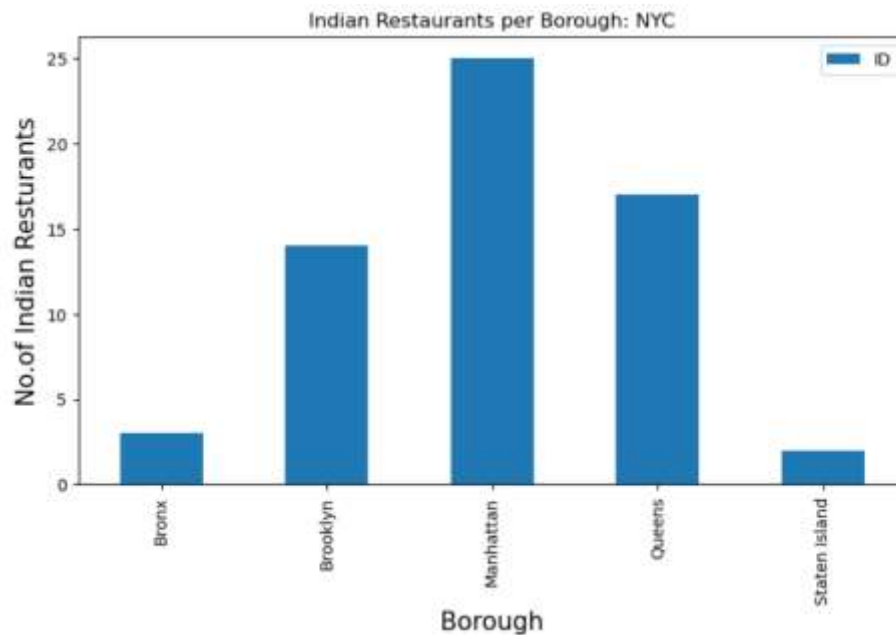
Analysis and Results

The results of our analysis showed below:

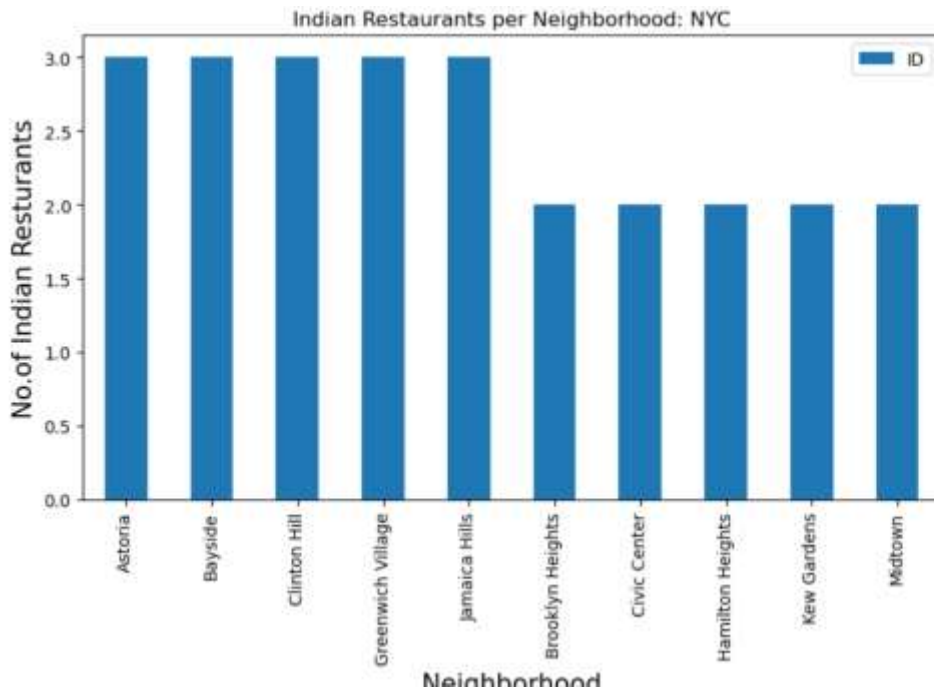
1. We see that Queens has the highest number of Neighborhoods in New York City.



2. Although Manhattan had the least number of neighborhoods, it does have the highest number of Indian restaurants.



3. The Astoria, Bayside, Clinton Hill, Greenwich Village and Jamaica Hills have the highest number of Indian Restaurants with a total count of 3



4. Restaurants in Greenwich Village:

	Borough	Neighborhood	ID	Name
24	Manhattan	Greenwich Village	56c931b2cd10407d3ecbefda	Ananda
25	Manhattan	Greenwich Village	4593ed04f964a52050401fe3	The Kati Roll Company
26	Manhattan	Greenwich Village	4d992c1e942ba093ebb5968c	THELEwala

5. Restaurant with maximum likes:

```

Borough      Manhattan
Neighborhood  Midtown
ID            49d91c12f964a520015e1fe3
Name          The Kati Roll Company
Likes         834.0
Rating        8.7
Tips          256.0
Name: 21, dtype: object

```

6. Restaurant with maximum tips:

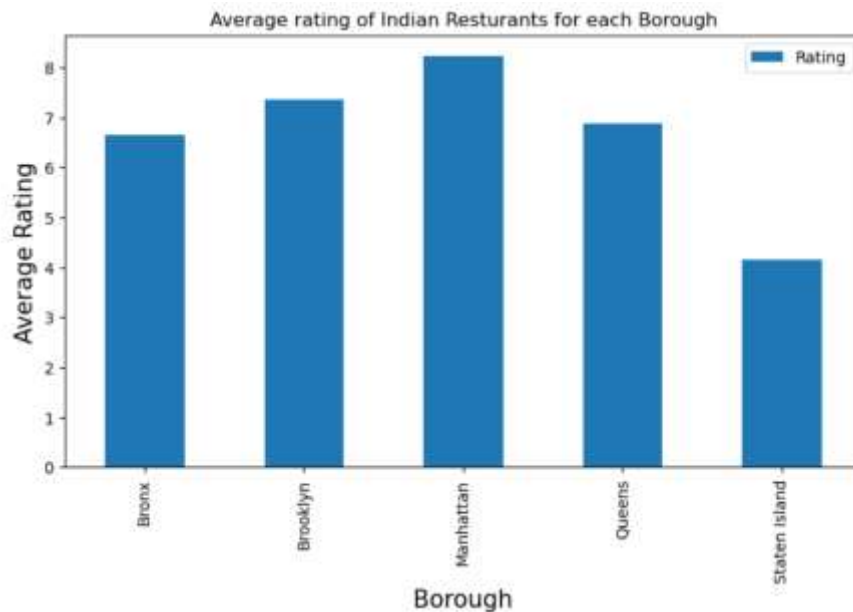
```
Borough      Manhattan
Neighborhood  Midtown
ID            49d91c12f964a520015e1fe3
Name          The Kati Roll Company
Likes         834.0
Rating        8.7
Tips          256.0
Name: 21, dtype: object
```

7. Restaurant with maximum ratings:

```
Borough      Manhattan
Neighborhood  Tribeca
ID            4bbb9dbded7776b0e1ad3e51
Name          Tamarind TriBeCa
Likes         602.0
Rating        9.1
Tips          152.0
Name: 27, dtype: object
```

8. Average rating of Indian restaurants per Borough.

Manhattan has the highest average rating for Indian Restaurants. Brooklyn goes right behind the Manhattan with the average rating of its Indian restaurants only a little lower than Manhattan.



9. Neighborhoods with top 10 average rating:
Two neighborhoods with the highest average rating for Japanese Restaurants are Tribeca and NoHo (both located in Manhattan).

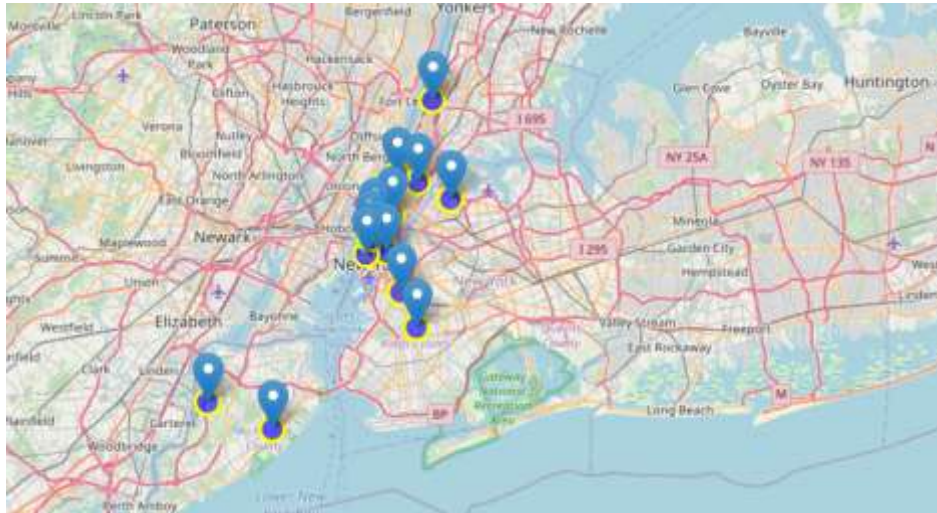
	Neighborhood	Average Rating
29	Tribeca	9.100000
25	Noho	8.800000
9	Chelsea	8.700000
17	Greenwich Village	8.666667
31	Upper West Side	8.600000
13	Fort Greene	8.600000
22	Midtown	8.550000
33	West Village	8.500000
0	Astoria	8.500000
24	New Dorp	8.300000

10. Neighborhoods with average rating greater than 8:

	Borough	Neighborhood	Latitude	Longitude	Average Rating
0	Queens	Astoria	40.768509	-73.915654	8.500000
1	Manhattan	Carnegie Hill	40.782683	-73.953256	8.000000
2	Manhattan	Chelsea	40.744035	-74.003116	8.700000
3	Staten Island	Chelsea	40.594726	-74.189560	8.700000
4	Brooklyn	Fort Greene	40.688527	-73.972906	8.600000
5	Manhattan	Greenwich Village	40.726933	-73.999914	8.666667
6	Manhattan	Midtown	40.754691	-73.981669	8.550000
7	Staten Island	New Dorp	40.572572	-74.116479	8.300000
8	Manhattan	Noho	40.723259	-73.988434	8.800000
9	Brooklyn	Prospect Lefferts Gardens	40.658420	-73.954899	8.100000
10	Manhattan	Tribeca	40.721522	-74.010683	9.100000
11	Manhattan	Upper West Side	40.787658	-73.977059	8.600000
12	Manhattan	Washington Heights	40.851903	-73.936900	8.000000
13	Manhattan	West Village	40.734434	-74.006180	8.500000

Discussion

Based on the results of our analysis, I would state that Manhattan is the best location for Indian cuisine in NYC. To have the best shot of success, I would open an Indian restaurant in Manhattan. Manhattan has multiple neighborhoods with average ratings exceeding 8.0 on a scale of 1.0 to 10.0. In Particular, I would recommend considering opening an Indian Restaurant either in Tribeca or NoHo, because both of these neighborhoods have the highest rating for Indian restaurants.



Limitations and suggestions for future researches

All the above analysis is depended on the accuracy of FourSquare data. Besides, during this project, we used a free Sandbox Tier Account of Foursquare API that goes with limitations as to the number of API calls and results returned. To get better results, future research work and more comprehensive analysis could consider using a paid account to bypass these limitations as well as incorporating data from other external databases.

Conclusion

In the project we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing data analysis, and lastly providing recommendations to the investors/developers. During the project, we applied different data science methods and instruments to get the answer to our main question: "Where in the City of New York should the investor open an Indian Restaurant?" The findings of this project will help the relevant investor better understand the advantages and disadvantages of different New York neighborhoods/boroughs in terms of opening an Indian restaurant.