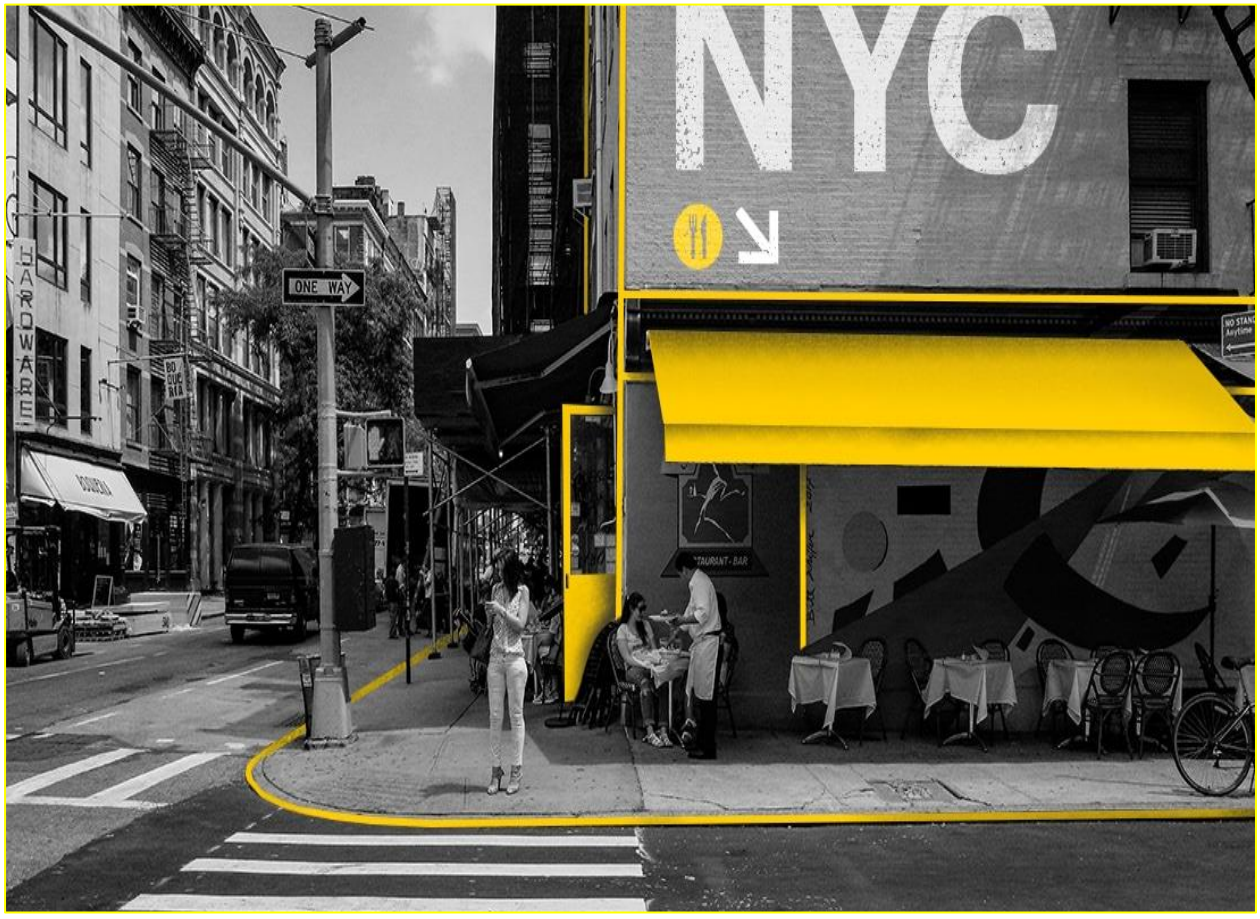
Coursera Capstone

IBM APPLIED DATA SCIENCE

AN INDIAN RESTAURANT IN NEW YORK CITY

By – Shrinath Rajeshirke



JULY 2021

INTRODUCTION

New York is the largest city in the United States with a long history of international immigration. It's a large and ethnically diverse metropolis. Indians in the New York city metropolitan area constitute one of the largest and fastest growing ethnicities in the New York city metropolitan area of the United States. he New York City region is home to the largest Indian American population among metropolitan areas by a significant margin, enumerating 711,174 uniracial individuals by the 2013–2017 U.S. Census American Community Survey estimates.

This final project explores the best locations for Indian restaurant in the city of New York. So that the owner of the new Indian restaurant can have a great success and consistent profit. In particular, the location of the restaurant is the most important factor for a successful business. So, our project will help investors to find out the best location for starting a Indian restaurant in the New York city.

BUSINESS PROBLEM

The main objective of the capstone project is to analyse & select the best location in the city of New York to open a new Indian restaurant. We will use required data science methodologies such as Data Analysis & Data visualization. This project will provide the solutions to answer the business problem of starting a new Indian restaurant in the city. So that the business keep growing in the upcoming years.

TARGET AUDIENCE

This project is useful to investors and developers looking to open an Indian restaurant in the city of New York. Overall, New York is a great place to open a new Indian restaurant. The objective is to recommend the location which neighbourhood of New York city a best choice to start a restaurant. As the 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 the New York city, each belonging to different categories like French, Indian, Japanese, etc. Why did we decide to focus on Indian cuisine in our project? Now when the idea of a dynamic variety of food, the spiciness & healthy lifestyle conquered the minds of people all over the country, Indian restaurants became extremely popular, as they offer a great alternative to regular American eating habits.

DATA

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

- New York City data containing the neighbourhoods and boroughs.
- Latitude and longitude co-ordinates of those neighbourhoods. This is required to plot the map & get the venue data.
- Venue data, particularly data related to restaurants. We are going to use this data to perform further analysis of the neighbourhoods.

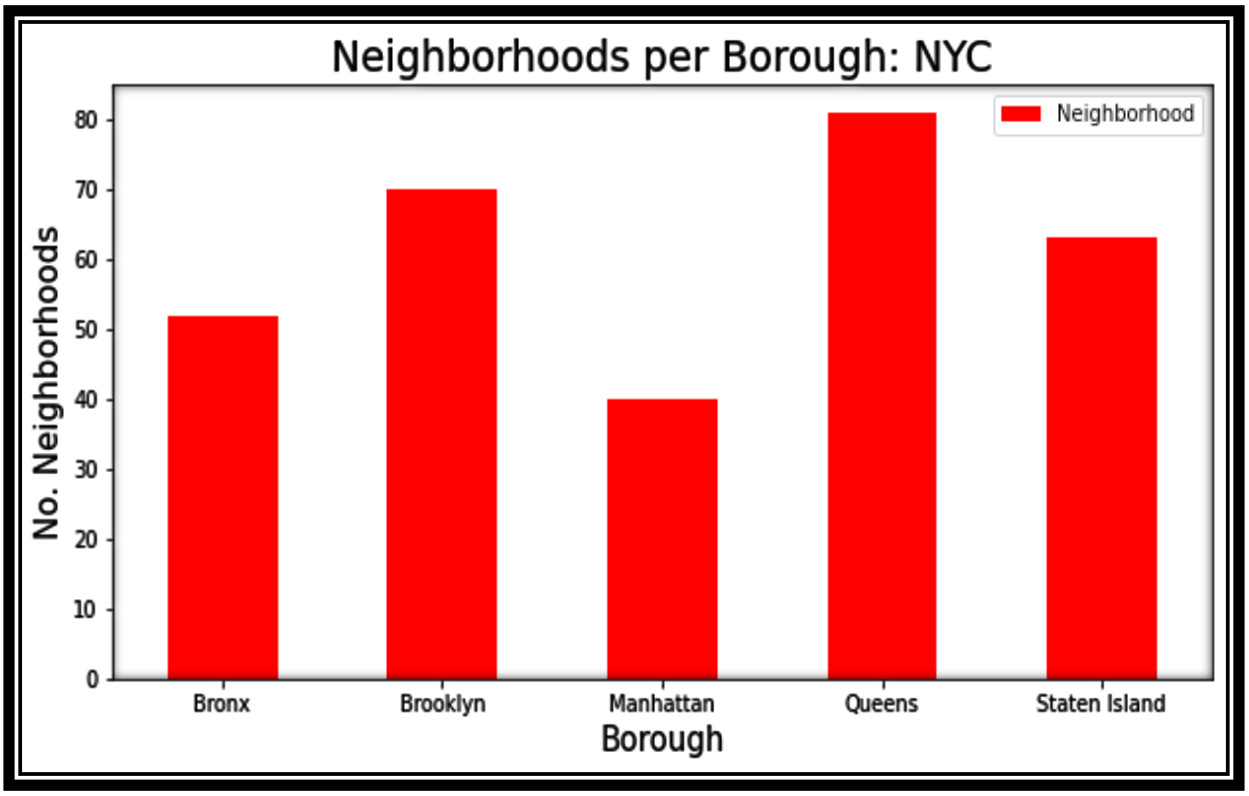
This project will require using of many data science skills, from web scrapping (opensource dataset), working with API (Foursquare), data cleaning, data wrangling, to map visualization (Folium). In the Methodology section, we will discuss and describe any exploratory data analysis that we did, any inferential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

METHODOLOGY

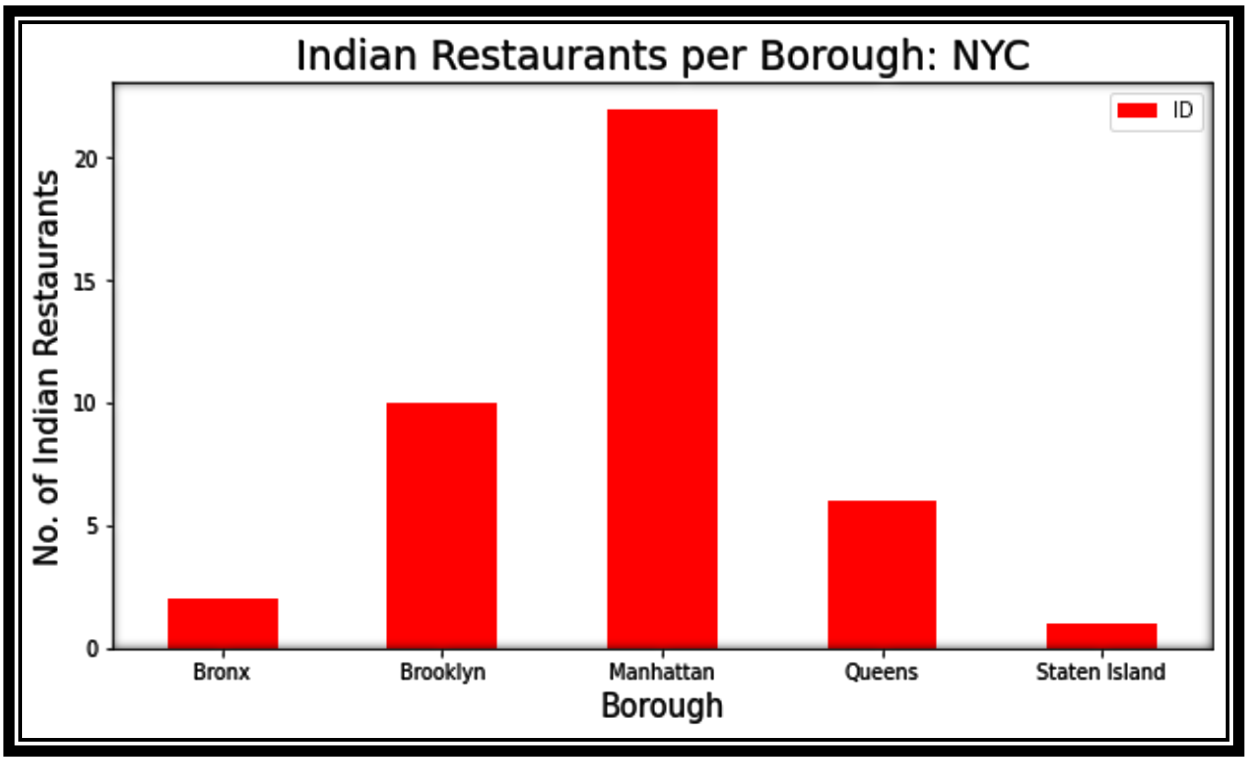
- Data will be collected from https://cocl.us/new_york_dataset and cleaned & processed into a data frame.
- Foursquare be used to locate all venues and then filtered by Indian restaurants. Ratings, tips and likes by users will be counted and added to the data frame.
- Data will be sorted based on rankings.
- Finally, the data will be visually assessed using graphing from Python libraries.

DATA ANALYSIS

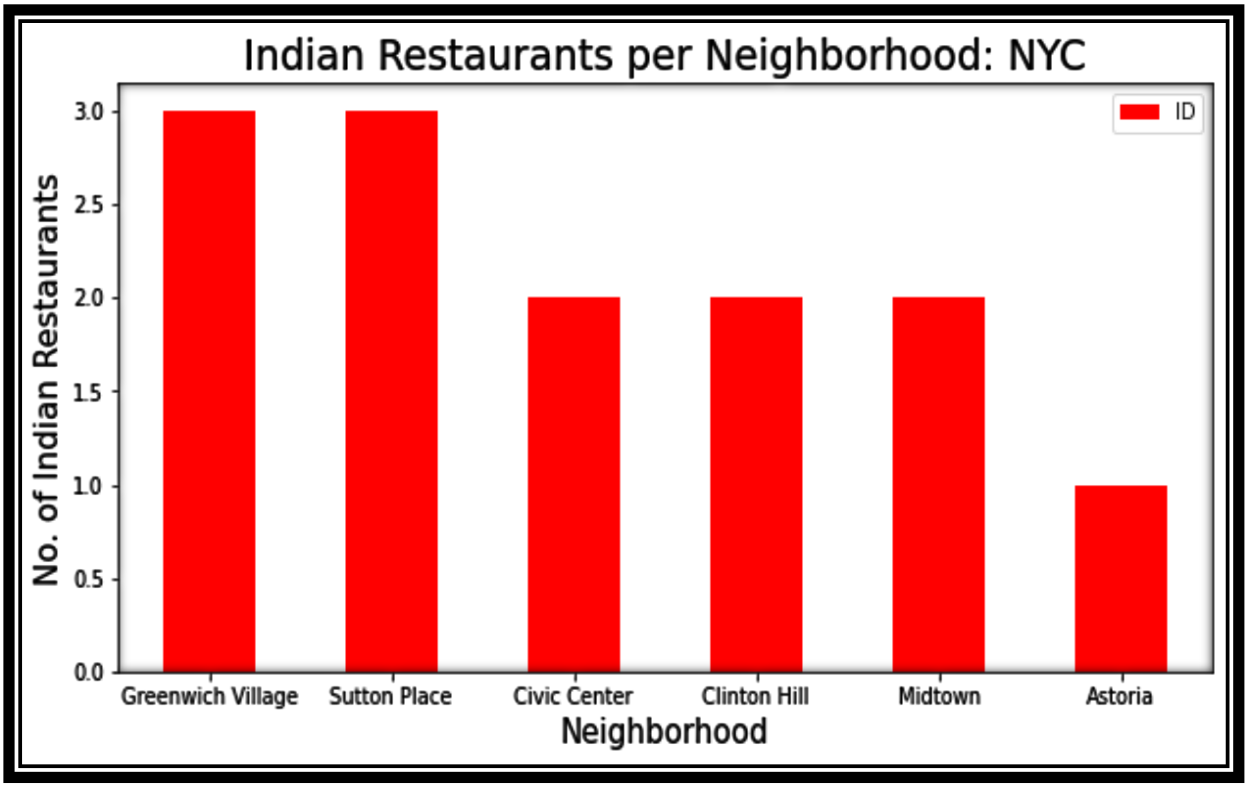
The data analysis we have done is as follows:



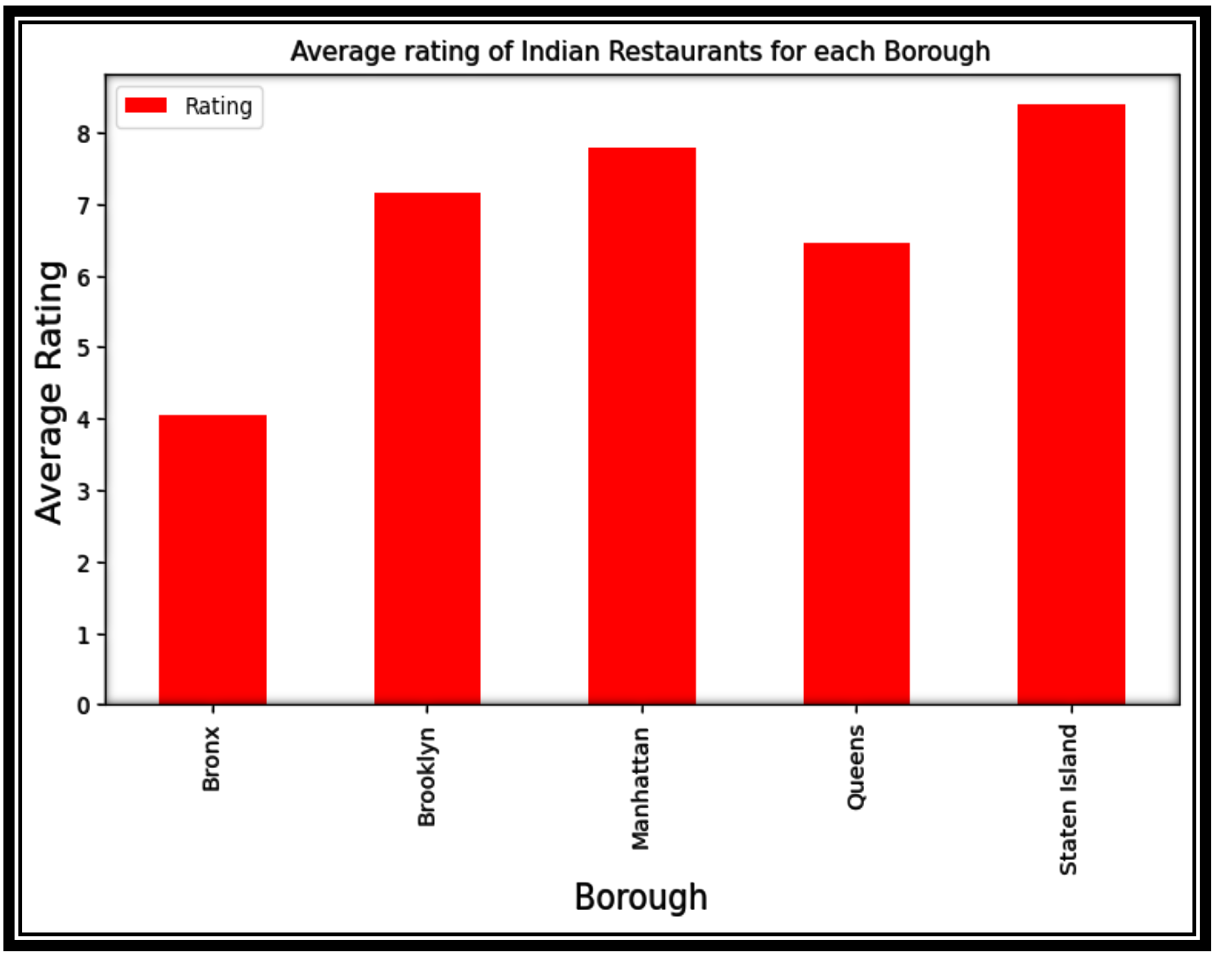
From the above bar diagram, we can see that the Queens has the highest number of Neighbourhoods & Manhattan has the least number of neighbourhoods.



Although Manhattan has the least number of neighbourhoods but, it has the highest number of Indian restaurants. Manhattan has more than 20 Indian restaurants.



We can see that the above bar plot shows that the Greenwich Village and Sutton Place has the highest number of Indian restaurants with a total count of 3 each.



Staten Island has the highest average rating for Indian restaurants. Manhattan goes right behind the Staten Island with high average rating of Indian restaurants.

Index Number	Neighbourhood	Average Rating
28	Tribeca	8.7
0	Astoria	8.6
7	Chelsea	8.6
25	Prospect Lefferts Gardens	8.5
24	Noho	8.4
23	New Dorp	8.4
15	Greenwich Village	8.2333
32	Woodlawn	8.1
31	West Village	8.1
22	Midtown	8.1

Table: Top 10 neighbourhoods with highest average rating

Two neighbourhoods with the highest average rating for Indian Restaurants are Tribeca located in Manhattan and Astoria located in Queens.

DISCUSSION SECTION

Manhattan and Staten Island have the best rated Indian restaurants on average. Queens and The Bronx have the least amount of Indian restaurants per borough. However, of note, Greenwich Village & Sutton Palace in Manhattan has the highest number of Indian Restaurants in all of NY. Despite Manhattan having the least number of neighbourhoods in all five boroughs, it has most number of Indian restaurants. Based on the above information, I would state that Manhattan and Staten Island are the best locations for Japanese cuisine in NYC. To have the best shot of success, I would open a Indian restaurant in Brooklyn. Brooklyn has multiple neighbourhoods with average ratings exceeding 8.0 on a scale of 1.0 to 10.0 and has less amount of Indian restaurants than Manhattan and Staten island, making competition easier. Also, we should keep in mind, that real estate prices in Brooklyn are much cheaper than in Manhattan. Finally, I would go to The Kati Roll Company in Manhattan for the best Indian food based on 835 likes. As a final note, all of the above analysis is depended on the accuracy of Foursquare data. A more comprehensive analysis and future work would need to incorporate data from other external databases.

LIMITATIONS AND SUGGESTIONS FOR FUTURE RESEARCHES

All of 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.

CONCLUSIONS

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 this 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 neighbourhoods/ boroughs in terms of opening an Indian restaurant.