



IBM applied data science capstone project

# Recommender system to build an Indian restaurant in New York

BY KAVITA PATIL

AUGUST 2020

# BUSINESS PROBLEM

The objective of this Capstone project is to analyse 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?



# DATA

## **Data required**

- List of borough and neighborhoods of New York
- Latitude and longitude coordinates of those neighborhoods
- Venue data, particularly data related to restaurants

## **Sources of data**

- Open source dataset of New York: [https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset)
- Geocoder package for latitude and longitude coordinates

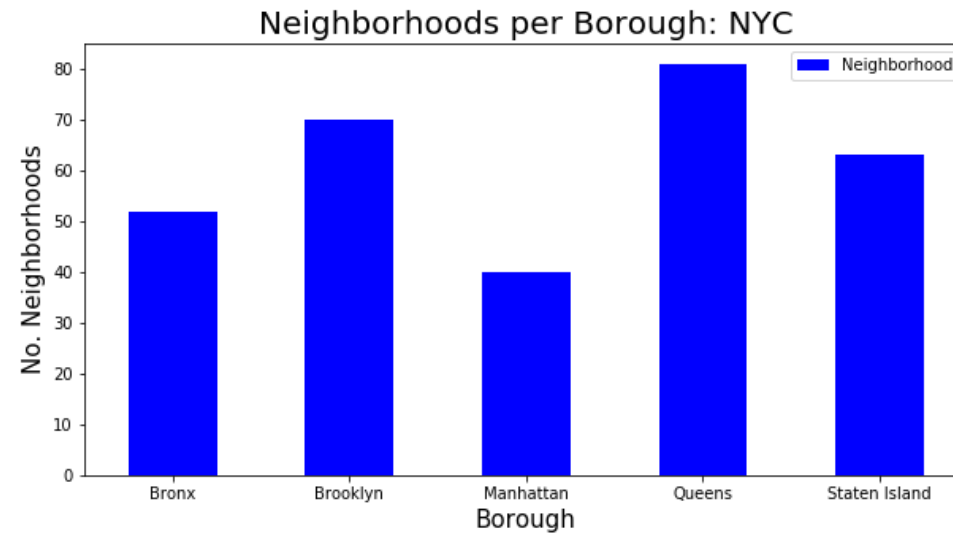
# METHODOLOGY

- Collect the data from [https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset) . Clean and process the data into a dataframe.
- Using FourSquare, locate all venues and then filter by Indian restaurants. Ratings, tips, and likes by users will be counted and added to the dataframe.
- Sort the data based on rankings.
- Visually assess the data be will be using graphing from Python libraries.

# RESULTS

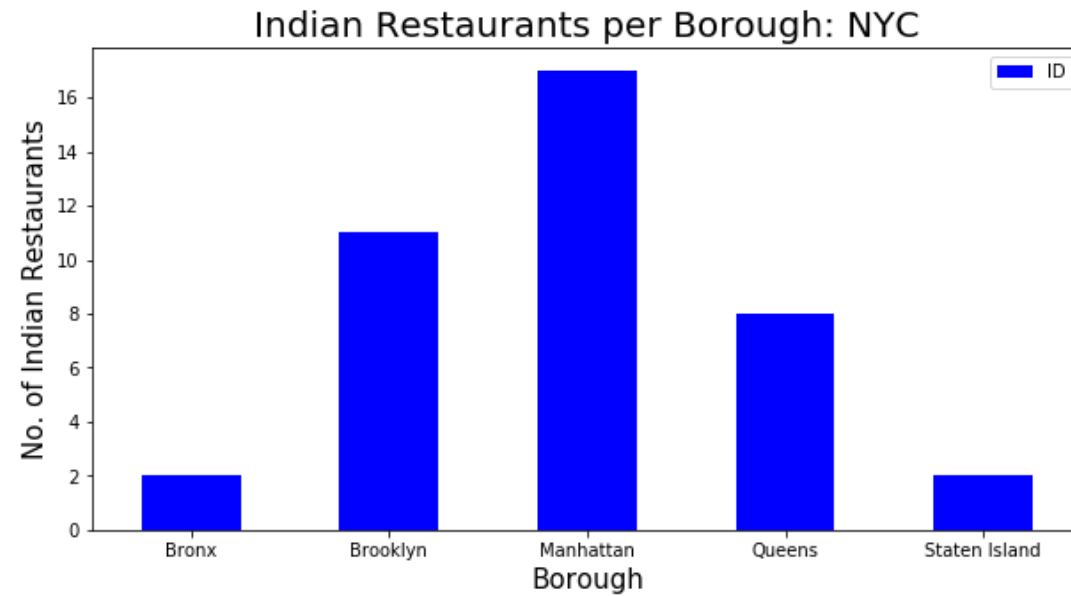
Below chart shows the count of neighbourhoods per borough

Manhattan has the least number of neighbourhoods while Queens has the highest.



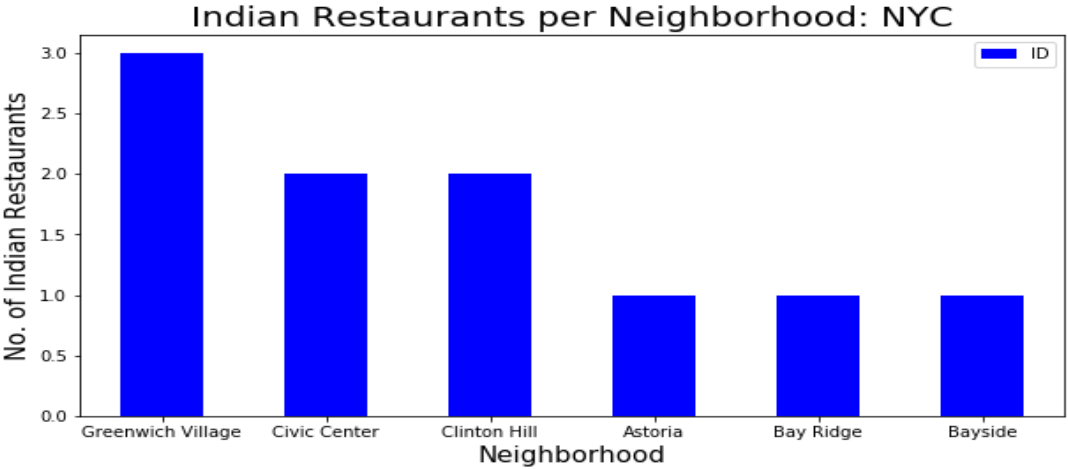
# CONTINUED..

Despite of Manhattan having least number of neighbourhoods, it has highest number of Indian restaurants.



# CONTINUED..

. Greenwich in Manhattan has the highest number of Indian Restaurants with a total count of 3.



:

	Borough	Neighborhood	ID	Name
19	Manhattan	Greenwich Village	56c931b2cd10407d3ecbefda	Ananda
20	Manhattan	Greenwich Village	4593ed04f964a52050401fe3	The Kati Roll Company
21	Manhattan	Greenwich Village	4d992c1e942ba093ebb5968c	THELEwala

# CONTINUED

	Neighborhood	Average Rating
30	Tribeca	9.1
23	Noho	8.8
11	Fort Greene	8.8
26	Prospect Lefferts Gardens	8.7
8	Chelsea	8.7
14	Greenwich Village	8.6
25	North Side	8.5
34	West Village	8.5
21	Midtown	8.4
0	Astoria	8.4

Above are the top neighborhoods with the highest average ratings of Indian restaurants.



# DISCUSSION

- Based on the information, Manhattan and Brooklyn are the best locations for Indian cuisine in NYC
- Brooklyn has multiple neighbourhoods with average ratings exceeding 8.0 on a scale of 1.0 to 10.0 and has a smaller number of Indian restaurants than Manhattan, making competition easier.
- Therefore, opening a restaurant in Brooklyn will be more profitable.
- The real estate prices in Brooklyn are cheaper than in Manhattan. Hence, it will be wiser to build a business there.
- The best rated company in Manhattan is Kati Roll company which has the maximum likes and ratings.
- All the analysis is depended on the accuracy of Four-Square data. Hence, a more comprehensive analysis and future work would need to incorporate 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 neighbourhoods/boroughs in terms of opening an Indian restaurant.