

IBM Applied Data Science Capstone Project

Opening Indian Restaurant in New York

Introduction

New York's demographics show that it is a large and ethnically diverse state. It is fast growing and developing state. Even though, Indian food is one of the popular food in New York, there are less number of Indian Restaurants in New York . Therefore, starting an Indian Restaurant in a perfect location in New York State will be a profitable business for people looking for an business opportunity. As increase in number of Indians in the state and the neighbouring states also accounts for the profitability of the Business.

This Final project explores the best location in the New York area to start a Indian Restaurant which can lead to success in the business. As location places an important role in the success or failure of the business, My project will try to find the best location where a person can open an Indian restaurant to obtain profits.

Business Problem

The objective of this project is to analyze the data and find the best location to open an Indian Restaurant in New York using Data Science methodology and instruments such as Data Analysis and Visualization along with machine learning algorithms, this project aims to find answer to the question "Which location in New York will be efficient to start a profitable Indian Restaurant?"

Target Audience

This project is useful for the people who are looking to start in a specific location(The type of business and location of the business can be changed accordingly).

Data

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

- Neighbourhood list of New York.
- Latitude and longitude coordinates of those neighborhoods.
- Venue data associated with Indian restaurants in the city. This will help us in further analysis of the neighborhoods.

Data Extraction

Data can be gathered from the following ways:

- Neighbourhood data can be obtained from scraping the Wikipedia.
- Geocoder Package helps to get the latitude and longitude data.
- Venue data can be obtained from FourSquare API.

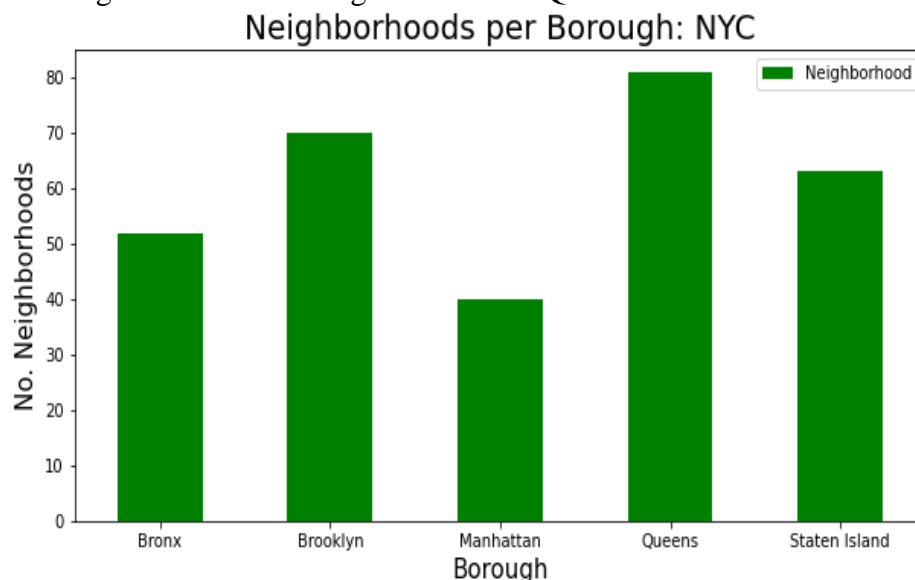
Methodology

- Data is collected from Wikipedia, scraped, cleaned and processed into a dataframe.
- Then FourSquare API is used to locate all venues and then filtered by Indian restaurants. From this venue data Ratings, tips, and likes by the users will be counted and made into a dataframe(Ratings, tips, and likes are premium calls in Four Square API. According to the API dated 1st November 2020, only 50 premium calls per day are allowed for free tier. Therefore it is advisable to use these calls carefully).
- These data will be visualized using different python libraries to get the required information.

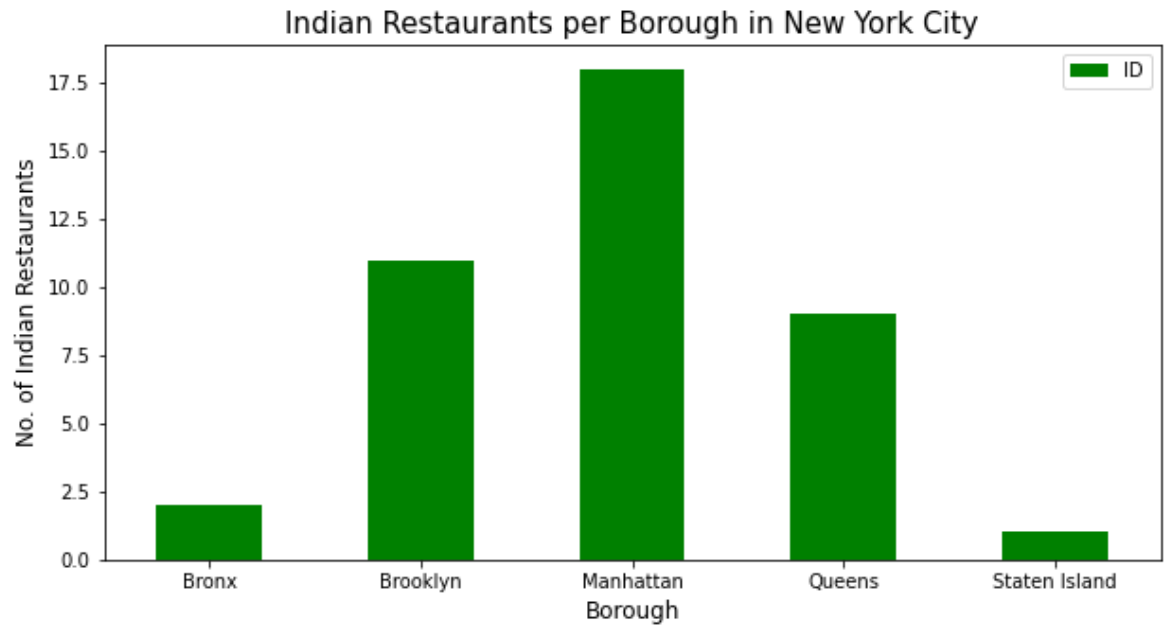
Results

The Results from our analysis are:

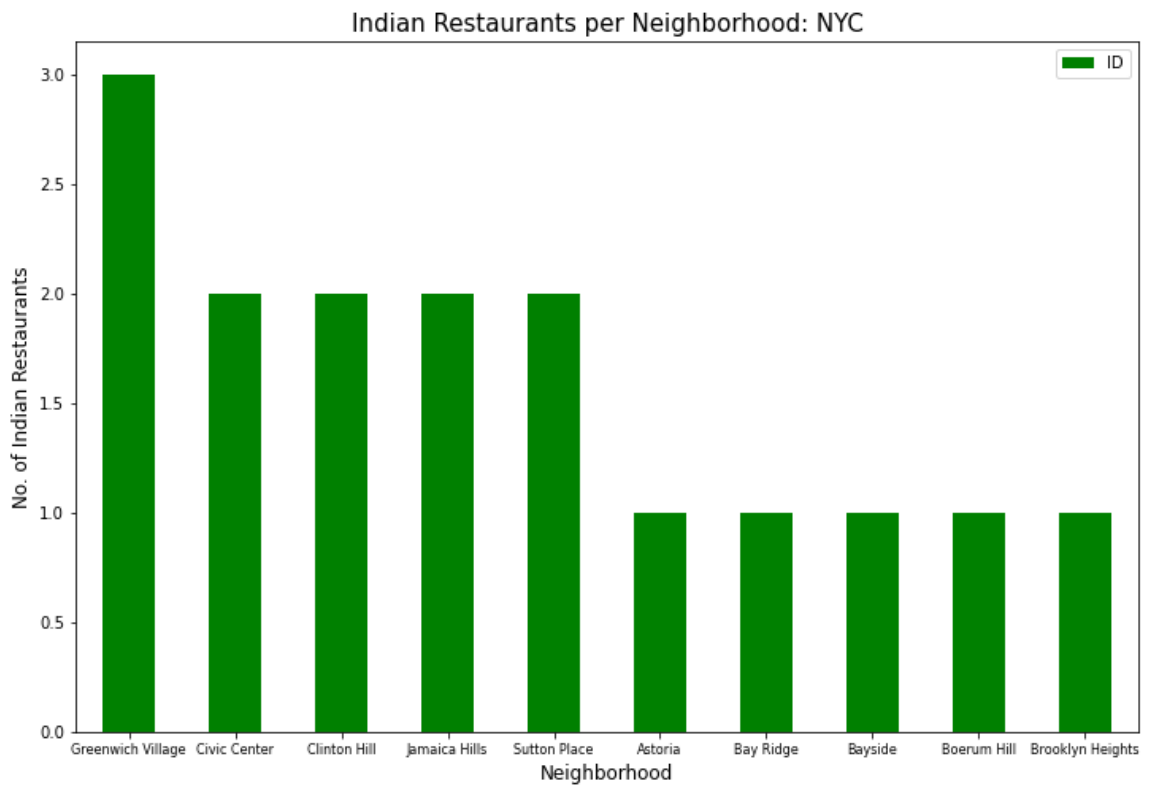
- Borough with highest number of neighbourhood is Queens.



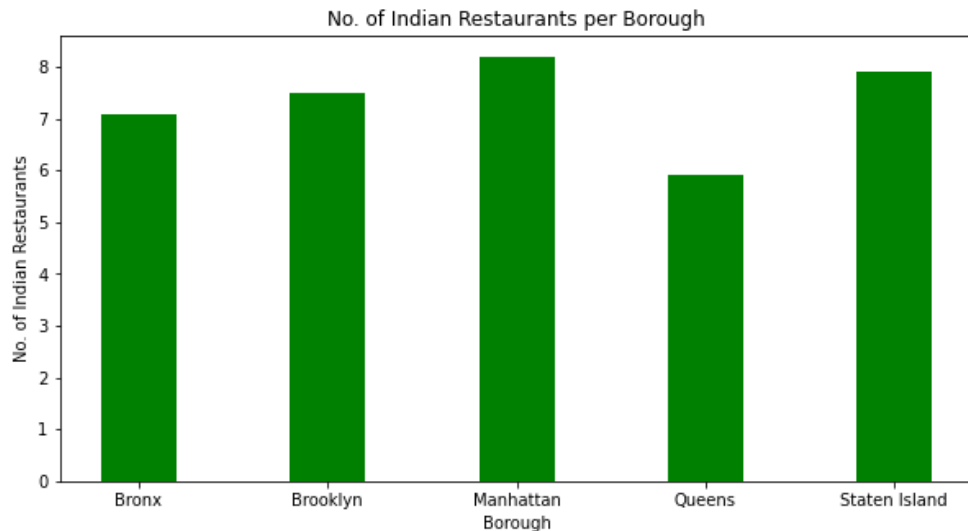
- Borough with highest Indian number of restaurants is Manhattan.



- Neighbourhood with most number of the Indian restaurants is Greenwich Village



- Manhattan has highest number of Indian Restaurants, followed by Staten Island. while Queens has lowest number of Indian Restaurants.



Discussion Section

To have a best Indian food I would prefer going to Tamarind TriBeCa or The Kati Roll Company, as they have best rating and most number of likes respectively. If you want to start an Indian Restaurant(considering the owner has adequate cost), we need to prefer starting an Indian Restaurant in Manhattan as we can see people are generous with tips and also people love eating Indian Food we can know this from number of likes and rating restaurants in Manhattan got, therefore, it is advisable to start an Indian Restaurant in Manhattan. If cost is an issue and want to start with less competition Brooklyn will be the perfect borough to start Indian Restaurants, as it has less competition and good amount of people who loves Indian Food.

Conclusions

This project goes through the process of finding the best place to start an Indian restaurant in New York state. This is done by acquiring necessary information from scrapping data from Wikipedia and Four Square API, then cleaning acquired data to make it useful for analysis, then perform data analysis on the cleaned data to extract information required to answer the question. The findings obtained by completing the whole process helped to answer the question “ where should we start Indian Restaurant in NY?”. The findings are it is advisable to start Indian Restaurant in Manhattan or Brooklyn based on financial state. This process can be further used by any investors to estimate best place to open any restaurant in any place by making small changes to the .