

Coursera Capstone Project

IBM Applied Data Science Capstone

Analysing to open a Indian Restaurant in New York City

Assignment: Week 1

By Vasanthan Veerasingham

April 2020

Table of Contents

Contents	Page
1. Executive Summary	3
2. Introduction	4
3. Methodology	5
4. Results	8
5. Discussion	10
6. Conclusion	10

Executive Summary

This data science report explores the best locations for Indian restaurants throughout the city of New York. Potentially the owner of the new Indian restaurant can have great success and consistent profit. We use our data science knowledge to find the facts on two basic questions to open an restaurant in New York How well New York people like Indian foods, Which neighborhood place is best to open a Indian restaurant in New York? And, Which Indian restaurant their favorite now?

This data science report is particularly useful to developers and investors looking to open or invest in a Indian restaurant in the city of New York. In consequence, as a reader of this data science report, common people of New York can find where they have Indian restaurants to eat healthy foods for better life.

1.Introduction

1. 1 Background

New York is the most populous city in USA, with a population of 8 million. Foreign-born people account for nearly half of the population of New York. New York is recognized as one of the most multicultural and cosmopolitan cities in the world. New York has no dominant culture or nationality, which also makes it one of the world's most diverse cities. New York is an international Centre of business, finance, art and culture. New York offers a very wide commercial variety, especially in the area of gastronomy. The trend that comes to stay, are Asian restaurants, particularly Indian restaurants. This data science report explores the best locations for Indian restaurants throughout the city of New York. Potentially the owner of the new Indian restaurant can have great success and consistent profit.

1. 2 Problem

In this data science report we have to find the solution for a real business problem. When a business man want to open a Indian restaurant in New York, we use our data science knowledge to find the facts on two basic questions to open an restaurant in New York How well New York people like Indian foods, Which neighborhood place is best to open a Indian restaurant in New York? And Which Indian restaurant their favorite now?

6.3 Interest

This data science report is particularly useful to developers and investors looking to open or invest in a Indian restaurant in the city of New York. Broadly, New York is a great place to open a restaurant with an ethnical cuisine. New York is home to many ethnic neighborhoods such as Little India, Little Italy, Chinatown and Little Jamaica and comes with diversity in the food items too.

Now when the idea of a healthy lifestyle conquered the minds of people all over the country, Indian restaurants became extremely popular, as they offer a healthy alternative to regular American eating habits. In consequence, as a reader of this data science report, common people of New York can find where they have Indian restaurants to eat healthy foods for better life.

2. Methodology - Data acquisition and Data wrangling

2.1 Data Acquisition

In this data science project we collect the data from three various data sources:

- New York City dataset will contains detail of the information about neighborhoods and boroughs. We can get it from the open data source: https://cocl.us/new_york_dataset.
- We need Latitude and longitude coordinates of those neighborhoods. This is required to plot the map and get the venue data. We will get the geographical coordinates of the neighborhoods (latitude and longitude) using Python Geocoder package.
- We acquired Venue data, particularly data related to restaurants. We are going to use this data to perform further analysis of the neighborhoods and boroughs. We will get this data by working with Foursquare API.

The data preparation and wrangling for each of the three sources of data is done separately in the project.

Open source New York City dataset contains the neighborhoods and boroughs. But in the project we should for visually map. We can get the geographical coordinates of the neighborhoods (latitude and longitude) by using Python Geocoder package.

Finally, we will use Foursquare API to get the venue data for the neighborhoods defined at the previous step. Foursquare has one of the largest databases of 105+ million places and over 125,000 developers use this application. Foursquare API provides many categories of the venue data; we are particularly interested in the restaurant data to solve the business problem defined above.

This project will require to using of many data science skills, such like web scrapping from open source dataset, working knowledge of Foursquare API, data cleaning, data wrangling on the many format of dataset, understanding the folium package for map visualization.

In the next 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 to get the solutions for the our data science project problem.

2.2 Data Wrangling / Cleaning

From open data source: https://cocl.us/new_york_dataset

We extract the data in the tabular format by using geo location function of the geo coder package we can see will get in the form tabular format as shown in the in the following fig.

```
Out[5]:
```

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

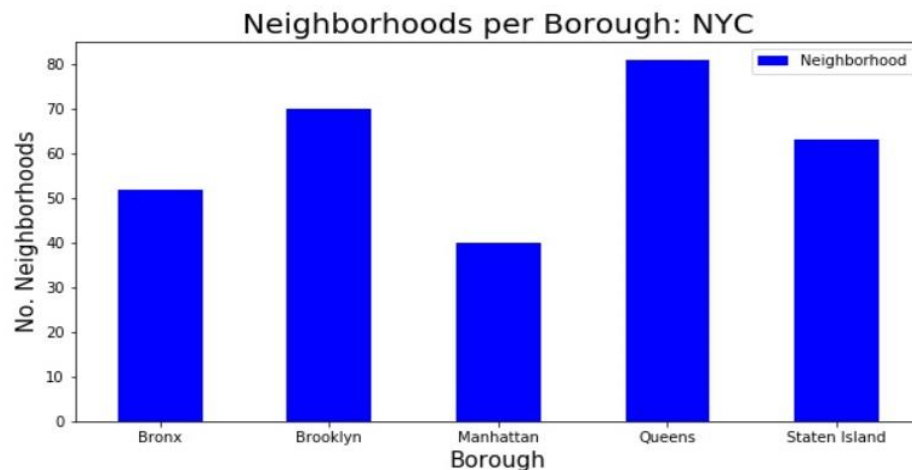
Initial Dataset get from sources

Dimension of Data

```
In [6]: new_york_data.shape
Out[6]: (306, 4)
```

There are total of 306 different Neighborhoods in New York.

Then we use the matplotlib package in the python language to plot the graph of data to visualize the dataset in the bar chart format so audience can easily what they see and what they can get idea about data science problem and the solutions. The following fig show the bar chart of the output



By using Foursquare API on the dataset we obtained from the open source, we can get the venues in the New York City neighborhoods such like Indian Restaurant information shown in the following fig.

Out[14]:

Unnamed: 0	Borough	Neighborhood	ID	Name	
0	0	Bronx	Woodlawn	4c0448d9310fc9b6bf1dc761	Curry Spot
1	1	Bronx	Unionport	4c194631838020a13e78e561	Melanies Roti Bar And Grill
2	2	Brooklyn	Bay Ridge	545835a1498e820edc6f636f	Bombay Grill
3	3	Brooklyn	Greenpoint	51a5445c498ee0f182370cb2	Agra Taj Mahal
4	4	Brooklyn	Bushwick	5169f5c4e4b0c7fcb77a0f3c	Agra Heights

In [15]: `indian_ny.shape`

Out[15]: (42, 5)

There are 42 Indian Restaurants in the New York City

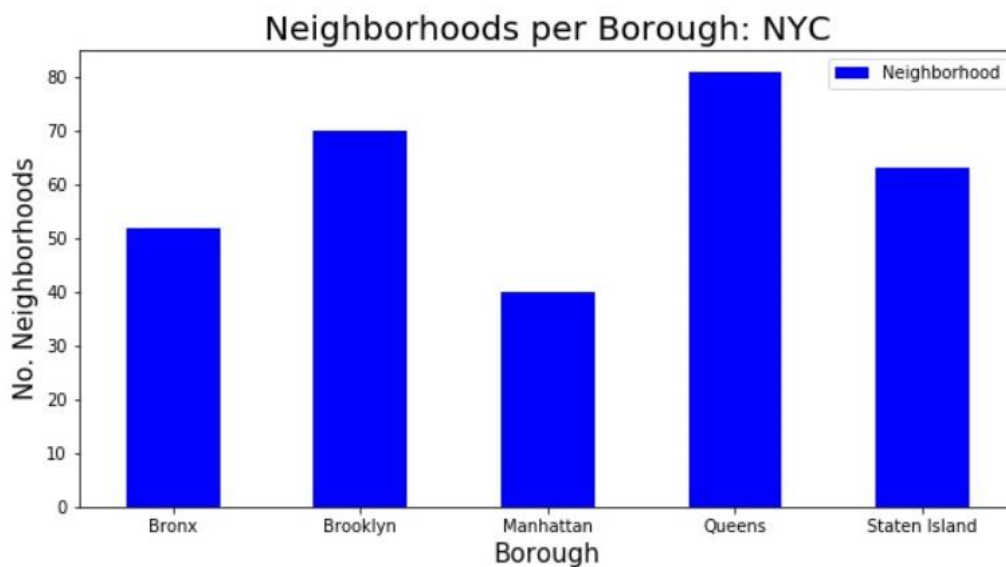
As like shown above we use the various data science tools to acquiring, cleaning, wrangling the datasets to preparation for the analyzing the data.

Results

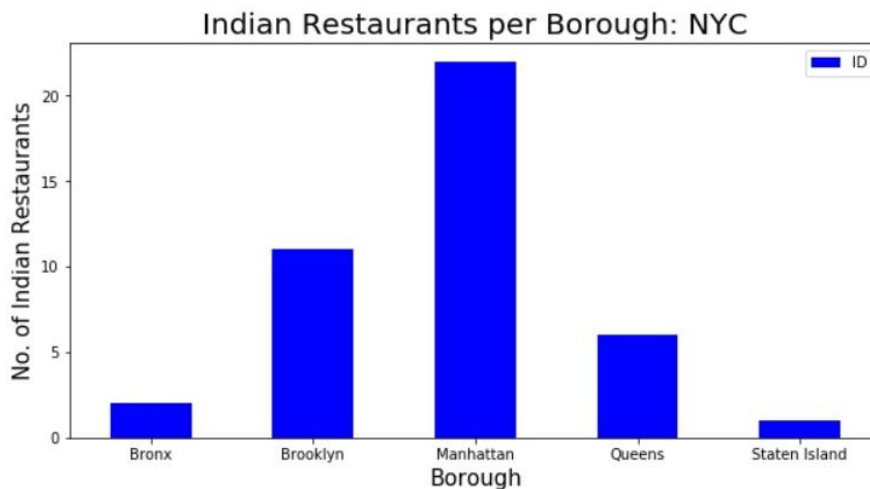
In this section we show the visually show the results of our analysis and how we get the solutions from the results for our data science problems.

The results of our analysis showed below:

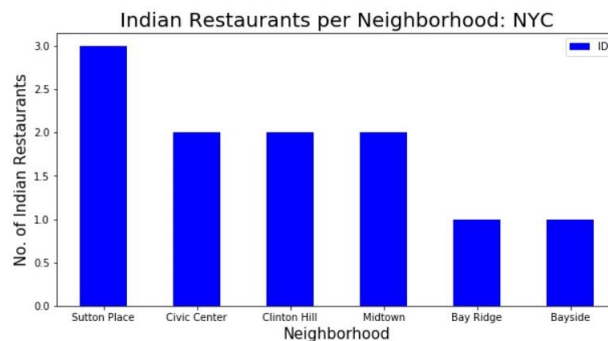
1. We see that Queens has the highest number of Neighborhoods.



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



The Sutton Place in Manhattan has the highest number of Indian Restaurants with a total count of 3.

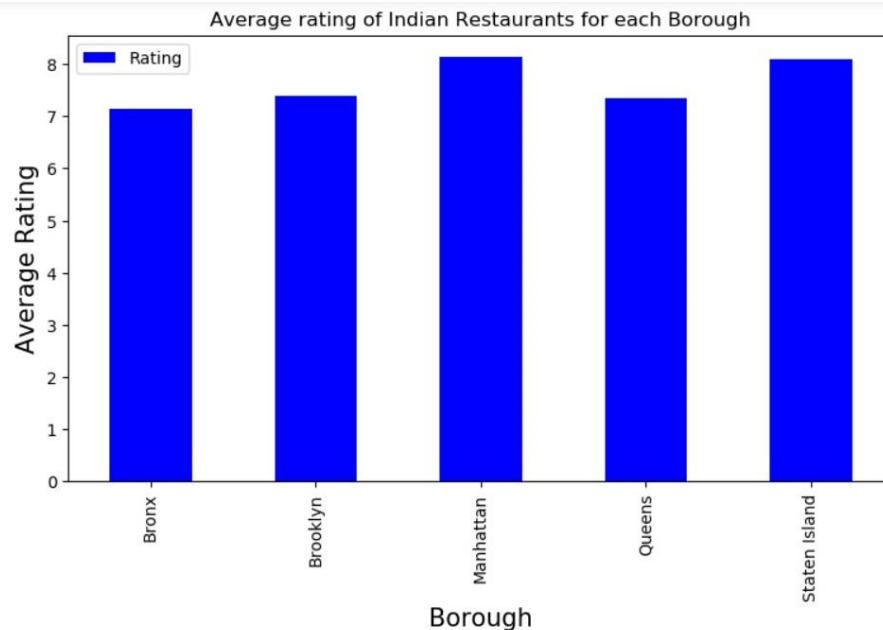


```
In [19]: indian_ny[indian_ny['Neighborhood']=='Sutton Place']
```

```
Out[19]:
```

Unnamed: 0	Borough	Neighborhood	ID	Name
37	Manhattan	Sutton Place	4a63bfb4f964a520b3c51fe3	Chola Eclectic Indian Cuisine
38	Manhattan	Sutton Place	3fd96200f964a52099e91ee3	Dawat
39	Manhattan	Sutton Place	4fe4fb50c2ee335e4fea69d	Moti Mahal Delux

Sutton Place neighborhood in Manhattan Borough have the highest no of Indian Restaurants = 3



Manhattan and Staten Island have the highest average rating for Indian Restaurants. Then Queens and Bronx go right behind the Manhattan and Staten Island with the average rating of its Indian restaurants only a little lower than Manhattan. Neighborhood with the highest average rating for Indian Restaurants is Tribeca which is located in Manhattan.

Out[35]:

	Borough	Neighborhood	Latitude	Longitude	Average Rating
0	Queens	Bayside	40.766041	-73.774274	8.20
1	Manhattan	Chelsea	40.744035	-74.003116	8.70
2	Staten Island	Chelsea	40.594726	-74.189560	8.70
3	Manhattan	East Village	40.727847	-73.982226	8.40
4	Manhattan	Flatiron	40.739673	-73.990947	8.40
5	Brooklyn	Fort Greene	40.688527	-73.972906	8.70
6	Manhattan	Greenwich Village	40.726933	-73.999914	8.40
7	Manhattan	Midtown	40.754691	-73.981669	8.55
8	Staten Island	New Dorp	40.572572	-74.116479	8.10
9	Manhattan	Noho	40.723259	-73.988434	8.80
10	Brooklyn	North Side	40.714823	-73.958809	8.50
11	Brooklyn	Prospect Lefferts Gardens	40.658420	-73.954899	8.60
12	Manhattan	Sutton Place	40.760280	-73.963556	8.20
13	Manhattan	Tribeca	40.721522	-74.010683	9.10
14	Manhattan	West Village	40.734434	-74.006180	8.50
15	Bronx	Woodlawn	40.898273	-73.867315	8.00

Discussion

Based on the results of our analysis, Manhattan and Staten Island are the best locations for Indian cuisine in New York City. Staten Island is a more promising place to open a new Indian restaurant. Neighborhoods Chelsea is the best place in Staten Island to open a new Indian restaurant.

Conclusion

Why did we decide to focus on Indian cuisine in our project? Now when the idea of a healthy lifestyle conquered the minds of people all over the World, Indian restaurants became extremely popular, as they offer a healthy alternative to regular American eating habits. In consequence, as a reader of this data science report, common people of New York can find where they have Indian restaurants to have healthy foods for better life.