

A.Introduction

A.1. Background

New York, is the most populous city in the United States. New York City is also the most densely populated major city in the United States. Located at the southern part of the U.S. it is one of the world's most populous megacities with 20 million people. New York City is composed of five boroughs, each of which is a county of the State of New York. The five boroughs—Brooklyn, Queens, Manhattan, the Bronx, and Staten Island. There are hundreds of distinct neighbourhoods throughout the boroughs, many with a definable history and character.

New York City is a global hub of business and commerce, as a center for banking and finance, retailing, world trade, transportation, tourism, real estate, new media, traditional media, advertising, legal services, accountancy, insurance, theater, fashion, and the arts in the United States. New York City's food culture includes an array of international cuisines influenced by the city's immigrant history.

A.2. Problem Description:

There are many restaurants in New York City, each belonging to different categories like Chinese, Indian, French etc. So as part of this project, we will list and visualize all major parts of New York City that has great Indian restaurants.

Now let me explain the context of this Capstone project through a scenario. Say you live on the New York city USA. **Queries that can be answered using above dataset.**

- What is best location in New York City for Indian Cuisine?
- Which areas have potential Indian Restaurant Market?
- Which all areas lack Indian Restaurants?
- Which is the best place to stay if I prefer Indian Cuisine?

A.3. Objective

The aim of this report is to study and analyze the neighborhoods Indian restaurants of New York city group them into similar clusters and, to analyze those clusters to gather meaningful information. That information can be used to find out neighborhoods that are same as your current neighborhood or at least similar.

A.4. Target Audience

This information provided by this report would be useful for people who are interested in searching for Indian restaurants that are highly similar to their existing neighborhood.

B. Data Description:

To consider the objective stated above, we can list the below data sources used for the analysis.

Approach

- Collect the new york city data from https://cocl.us/new_york_dataset
- Using FourSquare API we will find all venues for each neighborhood.
- Filter out all venues that are Indian Restaurants.
- Find rating , tips and like count for each Indian Restaurants using FourSquare API.
- Using rating for each restaurant , we will sort that data.
- Visualize the Ranking of neighborhoods using folium library(python)

Required Libraries

- pandas and numpy for handling data.
- request module for using FourSquare API.
- geopy to get co-ordinates of City of New York.
- folium to visualize the results on a map

C. Methodology:

Scrape the Wikipedia page and gathering data into a Pandas dataframe

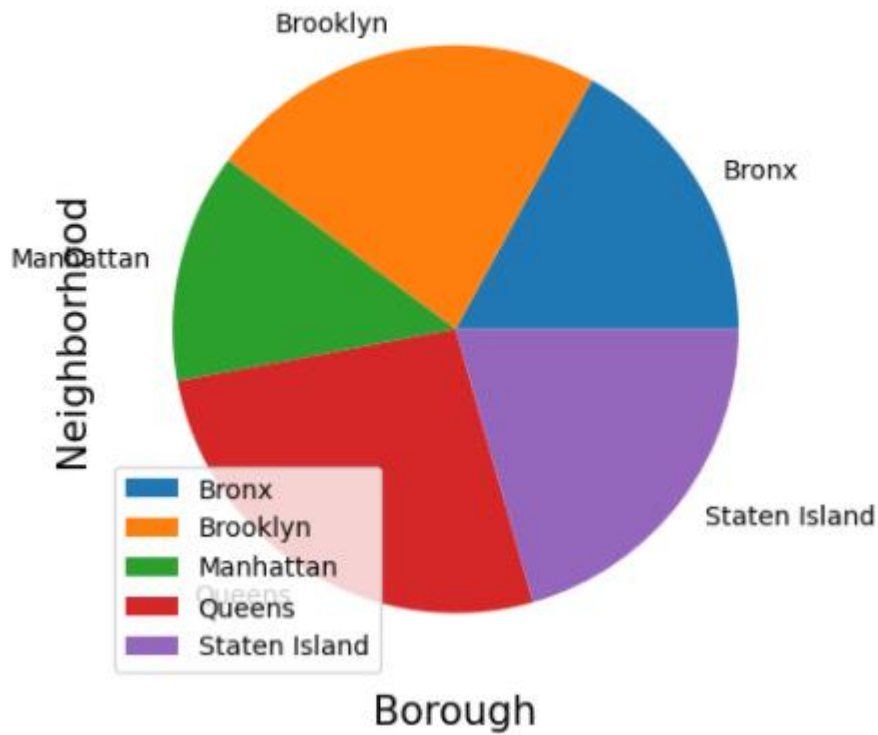
We define a function to interact with FourSquare API and get venues within a radius of 1000 metres for a given latitude and longitude. Below function will return us the venue id , venue name and category.

```
6]:
```

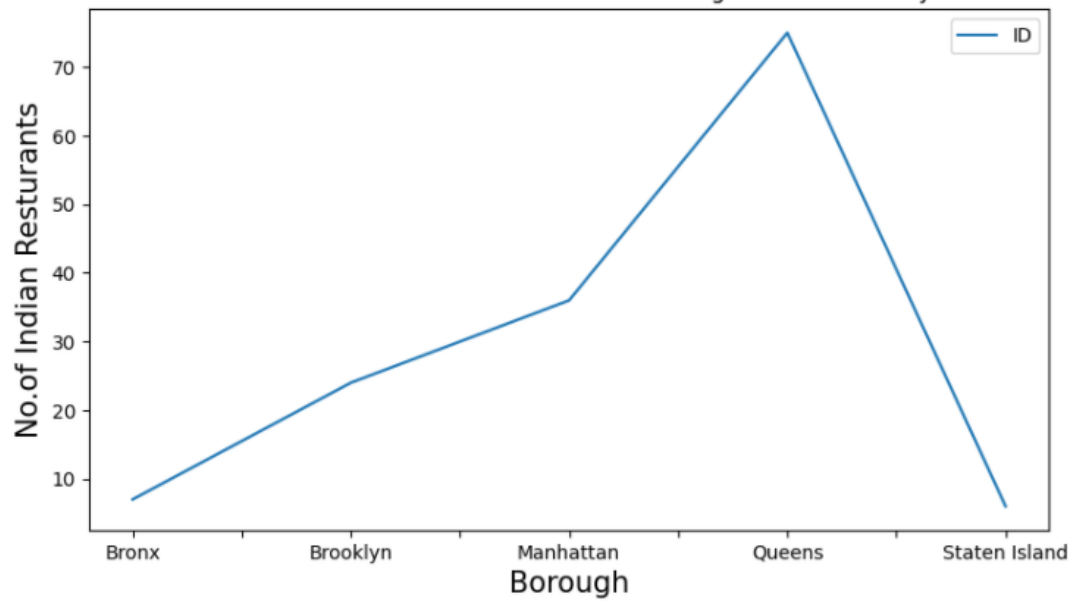
	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

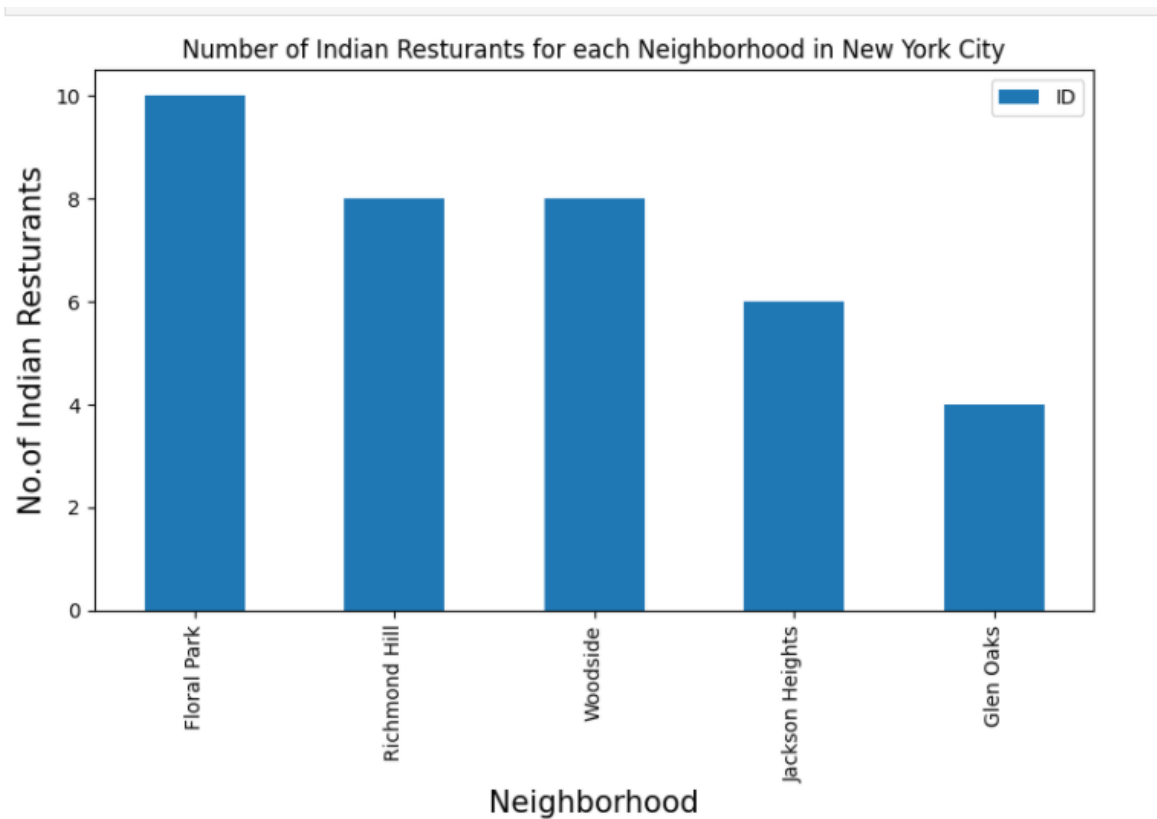
Print the number of neighbourhoods in New York City

Number of Neighborhood for each Borough in New York City



Number of Indian Resturants for each Borough in New York City





Analyze each neighborhood

```
.7]:
```

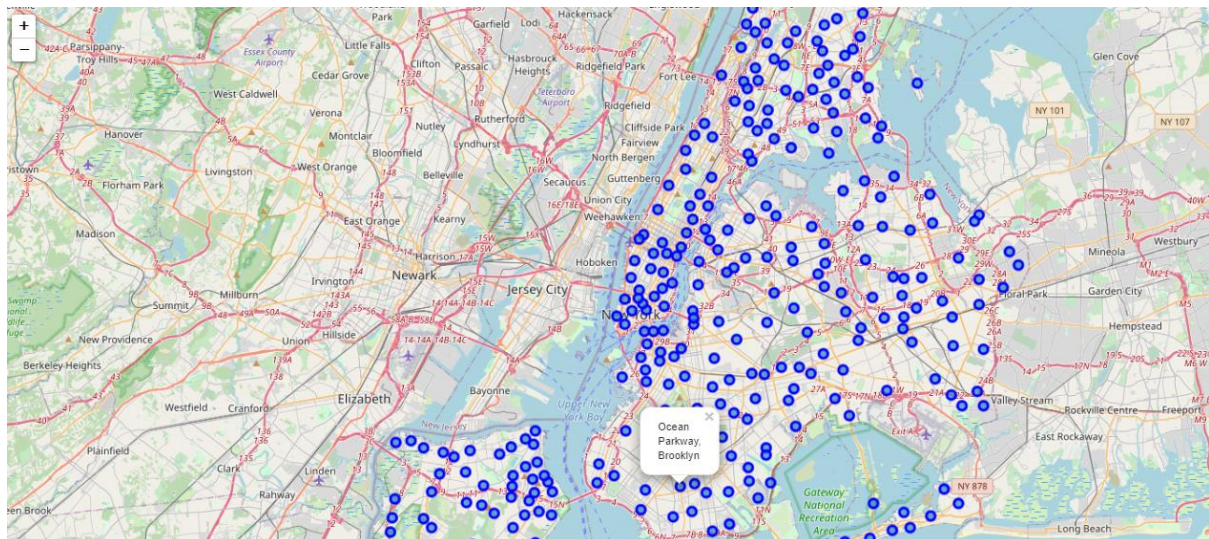
	Borough	Neighborhood	ID	Name	Likes	Rating	Tips
0	Bronx	Woodlawn	4c0448d9310fc9b6bf1dc761	Curry Spot	5	7.9	10
1	Bronx	Parkchester	4c194631838020a13e78e561	Melanies Roti Bar And Grill	3	6.1	2
2	Bronx	Spuyten Duyvil	4c04544df423a593ac83d116	Cumin Indian Cuisine	13	5.8	9
3	Bronx	Concourse	551b7f75498e86c00a0ed2e1	Hungry Bird	8	7.1	3
4	Bronx	Unionport	4c194631838020a13e78e561	Melanies Roti Bar And Grill	3	6.1	2

D. Results:

Adding the Cluster Labels to the Venue Data

11]:	Neighborhood	Average Rating
14	Prospect Heights	8.9
15	Prospect Lefferts Gardens	8.7
11	North Side	8.5
2	Clinton Hill	8.5
17	South Side	8.5
6	Fort Greene	8.5
4	East Flatbush	8.4
20	Washington Heights	8.1
10	Manhattanville	8.0
12	Ocean Hill	7.9

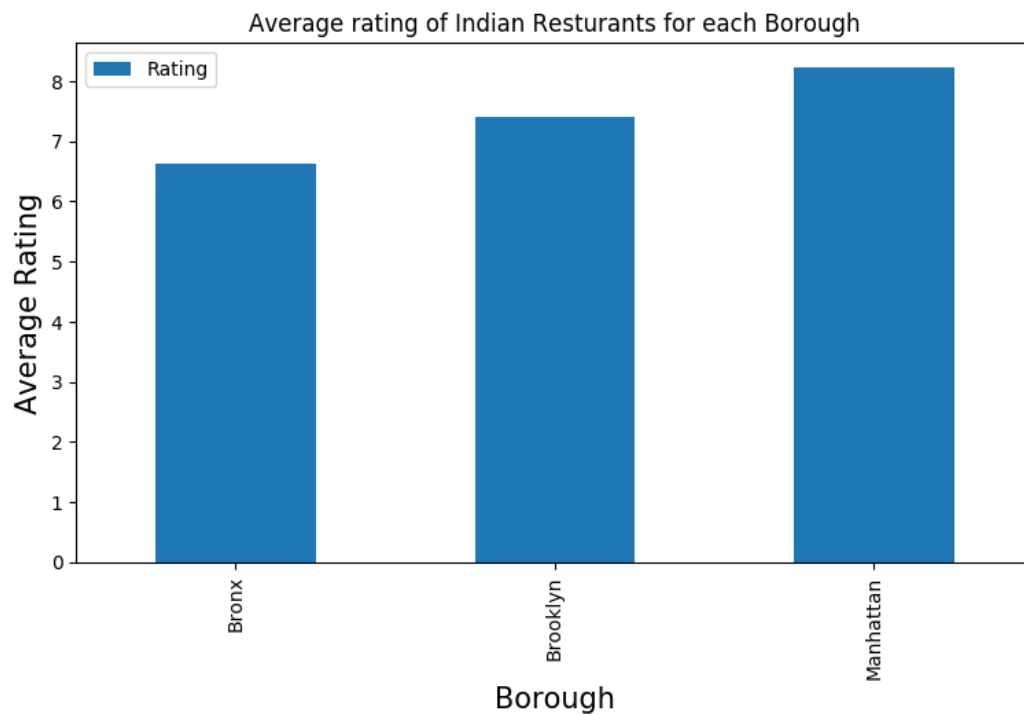
We use the matplotlib and folium packages to visualize the clusters on a map of New York.



E. Discussion:

The intent with which analysis was carried out was to find out similar neighborhoods for a searching for Indian restaurants

	Borough	Average Rating
2	Manhattan	8.227273
1	Brooklyn	7.408696
0	Bronx	6.620000



F. Conclusion:

In a fast moving world, there are many real life problems or scenarios where data can be used to find solutions to those problems. Like seen in the example above, data was used to cluster neighborhoods in New York based on the most Indian restaurants in those neighborhoods.

- Wikipedia content: https://en.wikipedia.org/wiki/New_York_City
- Foursquare API