|  |  |
| --- | --- |
| A picture of a winding road and trees  **Capstone Project: Battle of Neighborhoods**  Peer Graded Assignment | Abstract  The Report compares two Boroughs of New York city according to the needs of the client.  Abdul Hanan Muhammad Siddique  IBM Data Science Professional Certificate |

# Purpose

The report is a part of final project, that is Peer Graded Assignment, in the course IBM Data Science Professional Certificate. The report will compare two Boroughs of New York city according to our clients’ need and propose the best possible neighbourhood for our client to move.

# Introduction

New York is a multi-cultural city and also the most popular city of United States of America. New York City has been described as the cultural, financial, and media capital of the world, significantly influencing commerce, entertainment, research, technology, education, politics, tourism, art, fashion, and sports. As the headquarters of United Nations in located in New York, it is an important centre for international diplomacy. The city holds people from all ethnicities and as many as 800 languages are spoken in New York, making it the most linguistically diverse city in the world. New York is home to more than 3.2 million residents born outside the United States, the largest foreign-born population of any city in the world as of 2016. New York consists of five Boroughs which are Brooklyn, Queens, Manhattan, the Bronx, and Staten Island.

This diversity is the reason our client wants to move to New York from their home city of Austin, Texas. Our client has decided to move to New York to look for better job opportunities and has did some research of their own. They have identified that they would love to move somewhere in Manhattan or Queens. The conditions that are important for our clients are the easy availability of “Indian Stores” and “Nightclubs”. Hence our discussion would go about exploring all the three boroughs using these metrics. We will use Foursquare API to get information on these boroughs and then compare it with each other to find our result of borough with most number of “Indian Stores” and “Nightclubs”.

# Data

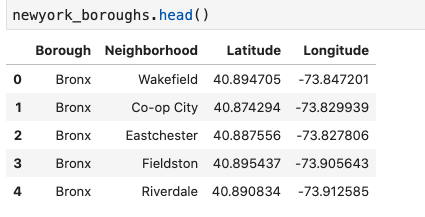
New York boroughs and neighbourhood data would be obtained from

<https://geo.nyu.edu/catalog/nyu-2451-34572>

The data has already been exported into a .json file and will be used directly for the purposes of this report. Google API would be used to get locations of venues at times as well. Moreover, Foursquare API would be used to explore various boroughs of New York and find data on the Neighborhoods. The Foursquare explore function will be used to get the most common venue categories in each neighborhood, and then use this feature to group the neighborhoods into clusters. The following information are retrieved on the first query:

* Venue ID
* Venue Name
* Coordinates : Latitude and Longitude
* Category Name

# Methodology

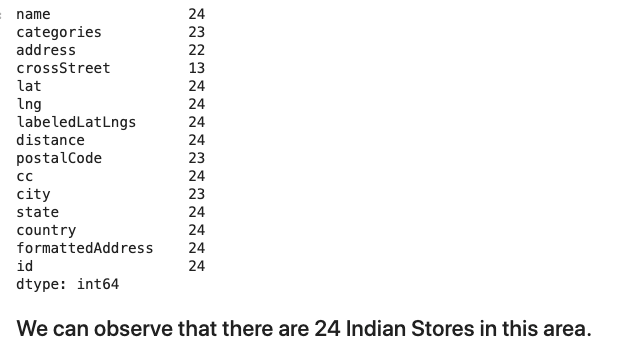
First I loaded all the necessary libraries that would be used during the assignment. Then I imported the New York dataset from the link and download it into the terminal using the *wget* command. Then I explored the dataset to get a better understanding of the dataset. I first created an empty dataframe with the necessary columns and then transformed the New York dataset into a pandas dataframe. I loaded the dataset into the newly formed dataframe. 

I received the latitude and longitude of the New York city using the Geopy library and used Folium library to create a map of five boroughs of New York with circle markers on all neighborhoods.

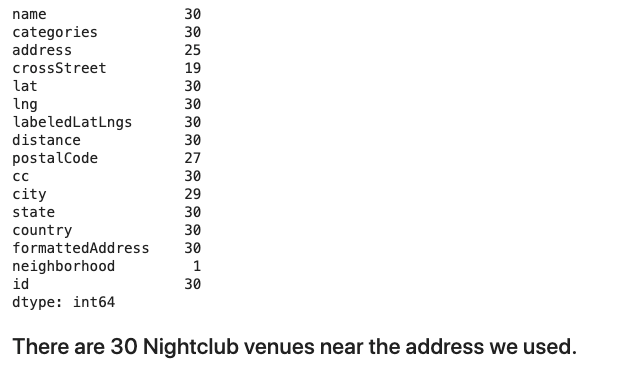


Then I explored the Manhattan Borough and found out the number of Indian stores and nightclubs in the area desired by ours client. I performed the same analysis on an area in Queens Borough to get a better result for our recommendation.

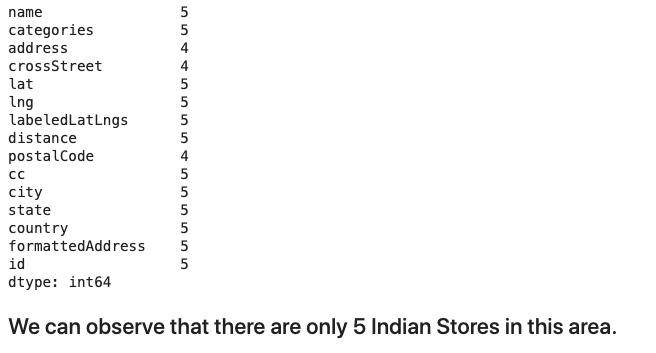
# Results

The result of the analysis return that there are 30 Indian stores near the area selected in Manhattan. 

The results also show that there are 30 Nightclubs near area selected in Manhattan.



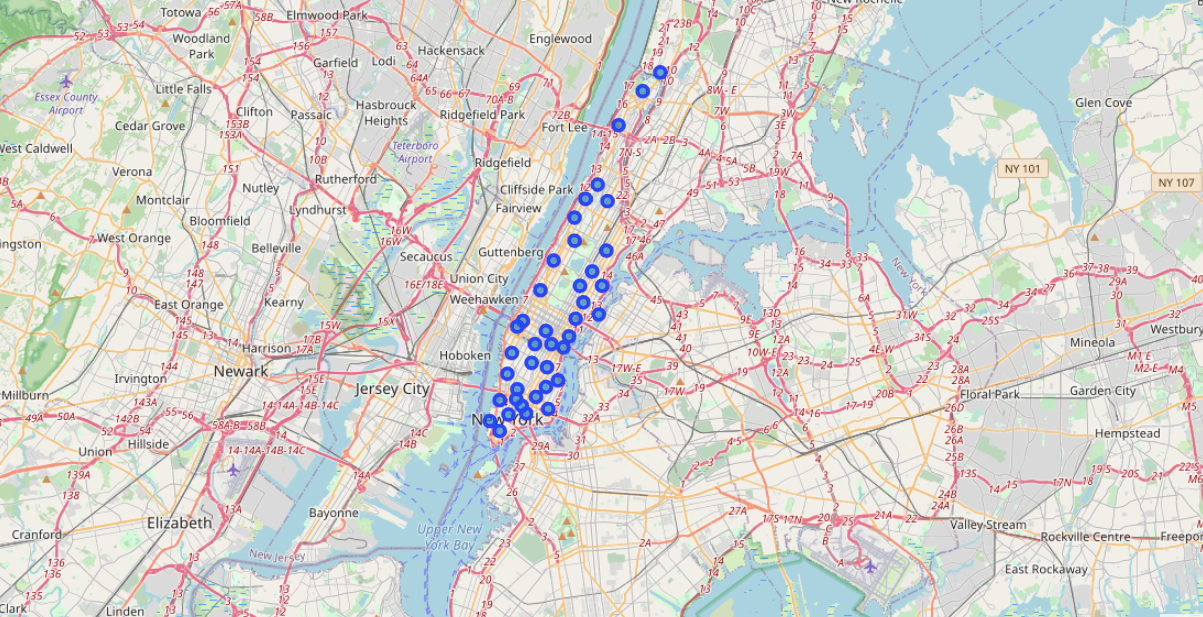
Exploring the address in Queens dataset, I observed there are only 5 Indian restaurants near the selected area.



Similarly, there are only 24 Nightclubs near the are selected in Queens.

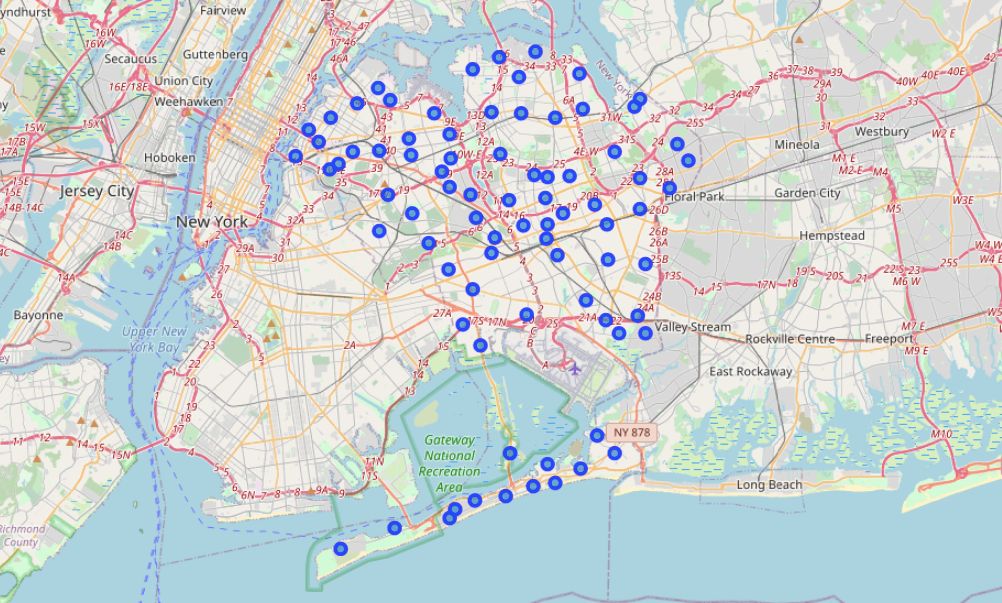
# Discussion

First I sliced the original dataframe and created a dataframe for all the neighborhoods in Manhattan. Then I obtained the longitude and latitude of the Manhattan area using the *ny\_explorer* user agent. Then I obtained a map of Manhattan with circle markers for all the neighborhoods.



Then I loaded the my Foursquare Credentials and selected a random address in Manhattan using Google Maps. I ran the query for the number of Indian stores in that area with a radius of 1km and sent a get request to get the details of the stores in the area. I created a dataframe with the information important for my analysis and filtered that dataframe according to the requirement. I found that there are 30 Indian stores near the selected address. I checked the rating of two random restaurants using foursquare API to get an idea about the quality of food in the area. Then I ran a query to get the number of nightclubs near the address. For the nightclubs, I increased the search radius to 5km to get a better results. Short radius was selected for stores because the client wanted the stores to be at a walking distance from the address. I made a new dataframe from the nightclub data and filtered this result into a new dataframe to get the desired data only. I found out that there are 30 nightclub venues near the address.

Then I sliced the Queens dataset from the New York dataset to explore the Queens dataset. I obtained the latitude and longitude of Queens and used Folium to obtained the map of Queens with circle markers at neighborhoods.



I selected an address from Google Maps and obtained the number of Indian Store near that area. I found that there are only 5 Indian stores near that area. I tried to find the rating of two random restaurants there and found that one of the restaurant has rating of 9.0 which is quite high. I can suggest it to the client even if he moves to Manhattan. Then I found the number of Nightclub venues near that area and found that there are 24 Nightclubs near the address which is close to my observation of 30 in Manhattan.

# Conclusion

From the above discussion, I can suggest my client to move in the Manhattan borough rather than Queens borough because it aligns closely with the requirement of my client. I will also suggest him to try the Restaurant with 9.0 rating.