|  |
| --- |
| Battle of neighborhoods |
|  |

|  |  |
| --- | --- |
| 24/06/2021 | IBM Capstone project |

**Table of contents**

|  |  |
| --- | --- |
| **Topic** | **Page number** |
| Introduction | 2 |
| Data | 3 |
| Methodology | 4 |
| Results & Discussions | 5 |
| Conclusion | 6 |

1. **Introduction :**

**Problem background:**

The City of New York, is the most populous city in the United States. It is diverse and is the financial capital of USA. It is multicultural. It provides lot of business opportunities and business friendly environment. It has attracted many different players into the market. It is a global hub of business and commerce. The city is a major 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. This also means that the market is highly competitive. As it is highly developed city so cost of doing business is also one of the highest. Thus, any new business venture or expansion needs to be analyzed carefully. The insights derived from analysis will give good understanding of the business environment which help in strategically targeting the market. This will help in reduction of risk. And the Return on Investment will be reasonable.

**Problem description:**

The City of New York is famous for its excellent cuisine. It's food culture includes an array of international cuisines influenced by the city's immigrant history.

Indian restaurants have become so popular in the United States now it seems that there is one on every corner, not only in major cities but also in smaller cities. Starting a Indian restaurant can be a great business opportunity, but you need to distinguish yourself from others to enjoy long-term success.

**Find a suitable location:**

If you plan a real restaurant that can demand higher prices for fresh fish, delivered daily from India, focus on neighborhoods and outlets that already attract a sophisticated Indian client. If you plan a cheap buffet restaurant, points to the masses looking for affordable high-traffic locations with large shopping centers and other local points of interest. So it is evident that to survive in such competitive market it is very important to strategically plan. Various factors need to be studied in order to decide on the Location

**Target Audience:**

My client wants to open his business in Queens area, so I only focus on that borough during my analysis. The objective is to locate and recommend to the management which neighborhood of Queens will be best choice to start a restaurant. The Management also expects to understand the rationale of the recommendations made.

1. **Data:**

To find the best location for our Indian restaurant, we will use the following sources of information:

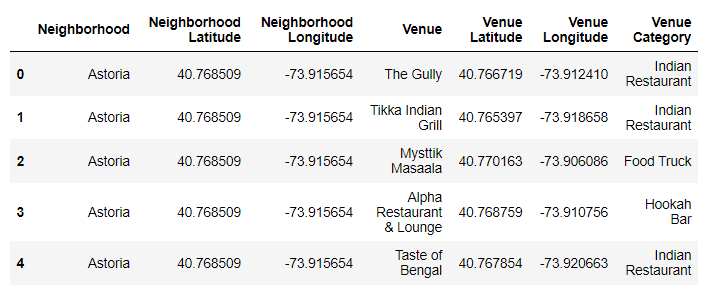
New York has a total of 5 boroughs and 306 neighborhoods. In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood. Luckily, this dataset exists for free on the web. Feel free to try to find this dataset on your own, but here is the link to the dataset: https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork\_data.json

**Data collection:**

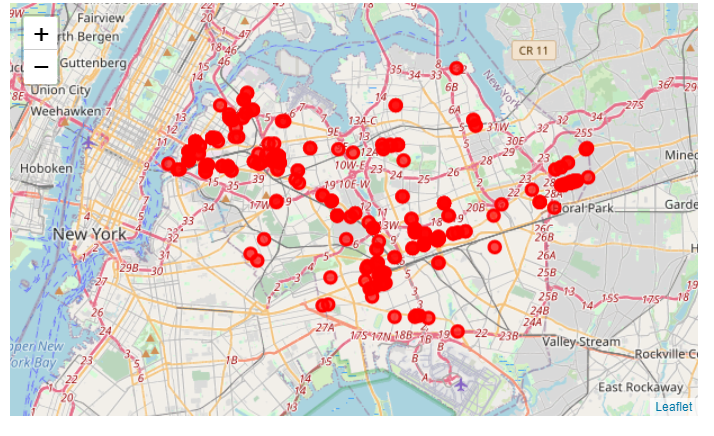
1. Use of foursquare is focused to fetch nearest venue locations so that we can use them to form a cluster. Foursquare API leverages the power of finding nearest venues in a radius (in my case: 500mts) and also corresponding coordinates, venue location and names. After calling, the following data frame is created:
2. From Foursquare Venues Categories - https://developer.foursquare.com/docs/resources/categories Indian restaurants category Id - 4bf58dd8d48988d10f941735
3. **Methodology:**

**Exploratory Data analysis:**

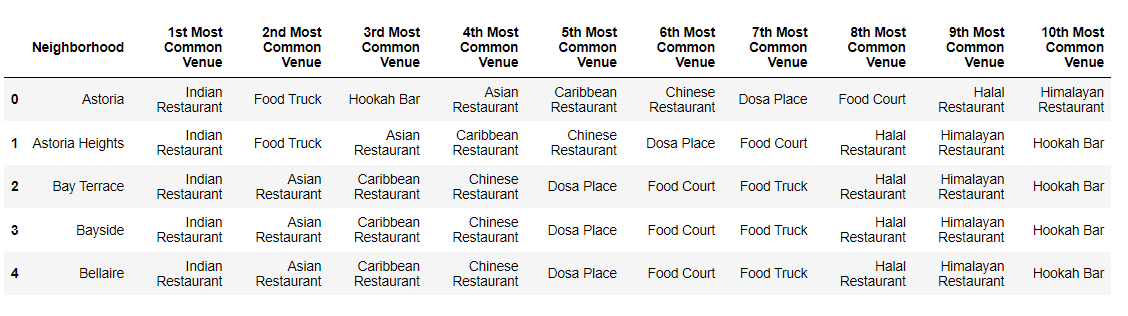
* Venues with borough as Queens are filtered into a data frame



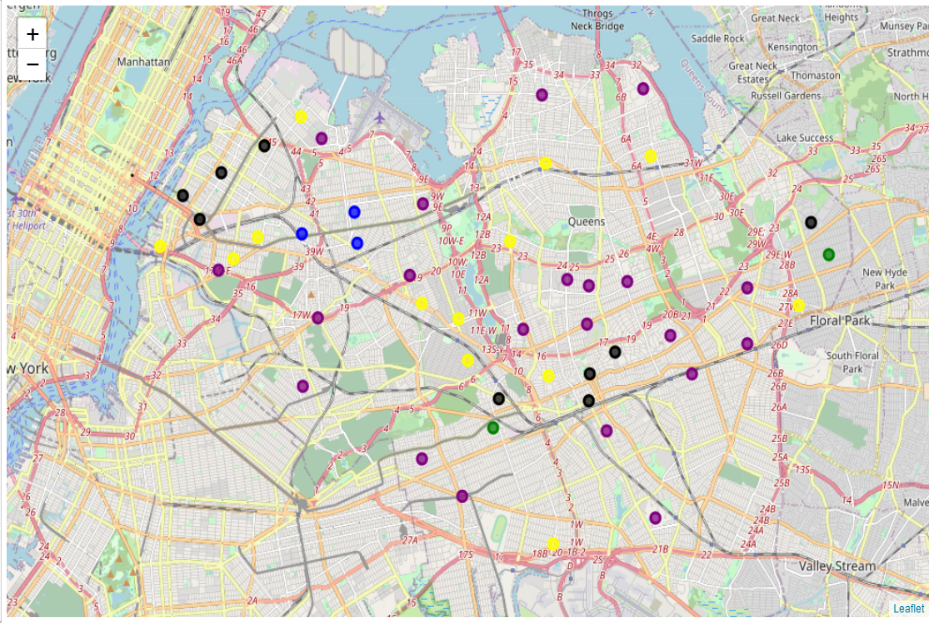
* Map is created to visualize the Indian restaurants in Queens



* Top 10 common venues in each neighborhood are analyzed



* The data is preprocessed by one hot encoding and clustered by Kmeans clustering.
* Clustered data is visualized using Folium



1. **Results & Discussions:**

* The clustering model has clustered the data into 5 clusters based on the count of each cuisine
* All the clusters have the 1st most common venue as Indian restaurants and secondly Indian Chinese restaurants
* We also have to consider that there are neighborhoods with no data of any restaurants situated near them.
* clusters 1 and 2 has a total of 5 neighborhoods and the common venues are related to Asian cuisines.
* clusters 0,3 and 4 has the remaining neighborhoods where the Asian cuisines dominate the top 10 common venues
* Based on this, we can recommend the stakeholder to open a restaurant in neighborhoods on clusters 1 and 2.
* The stakeholder can open an Indian restaurant in the following neighborhoods
  + Woodside
  + Jackson Heights
  + Elmhurst
  + Richmond Hill
  + Floral Park

1. **Conclusion:**

Although the final goal of the project is met, there is definitely room for further improvement and development where more features can be used to fit the model.