

REPORT

Contents

1. Introduction

- 1.1. Background
- 1.2. Problem Statement

2. Dataset Information

3. Methodology

4. Results and Discussion

5. Conclusion

1. Introduction

1.1 Background

In this project, I am trying to find an ideal place to open an Indian restaurant in New York City. For this to be accomplished, I first found out the region with more number of Indians. Queens has around 6.2% of American Indians out its total population. Next step is to find out the neighbourhoods in Queens where people prefer to eat Indian food. After examining that data we can find out the best place to open an Indian Restaurant.

1.2 Business Problem

The basic idea of this project is to find the best place in New York City to open an Indian Restaurant. This is achieved by finding out the most Indian populated area in New York and find out the neighbourhoods where people tend to eat Indian food. This is performed using Data Science and Clustering methods.

2. Dataset Information

1. Data for Indian population of New York city is taken from
https://en.wikipedia.org/wiki/Indians_in_the_New_York_City_metropolitan_region

The data for Indian population in different regions of New York City was taken from this source. It was found out that Queens has around 145k Indians which is more than 6 percent of its total population.

2. The New York City data set is taken from
https://geo.nyu.edu/catalog/nyu_2451_34572

This repository contains data of different neighbourhoods in New York City. The data set consists of a total of 5 boroughs and 306 neighbourhoods. It also contains the latitude and longitude data for each neighbourhood.

3. Foursquare location data is used
https://api.foursquare.com/v2/venues/explore?&client_id={} &client_secret={} &v={} &ll={},{} &radius={} &limit={}

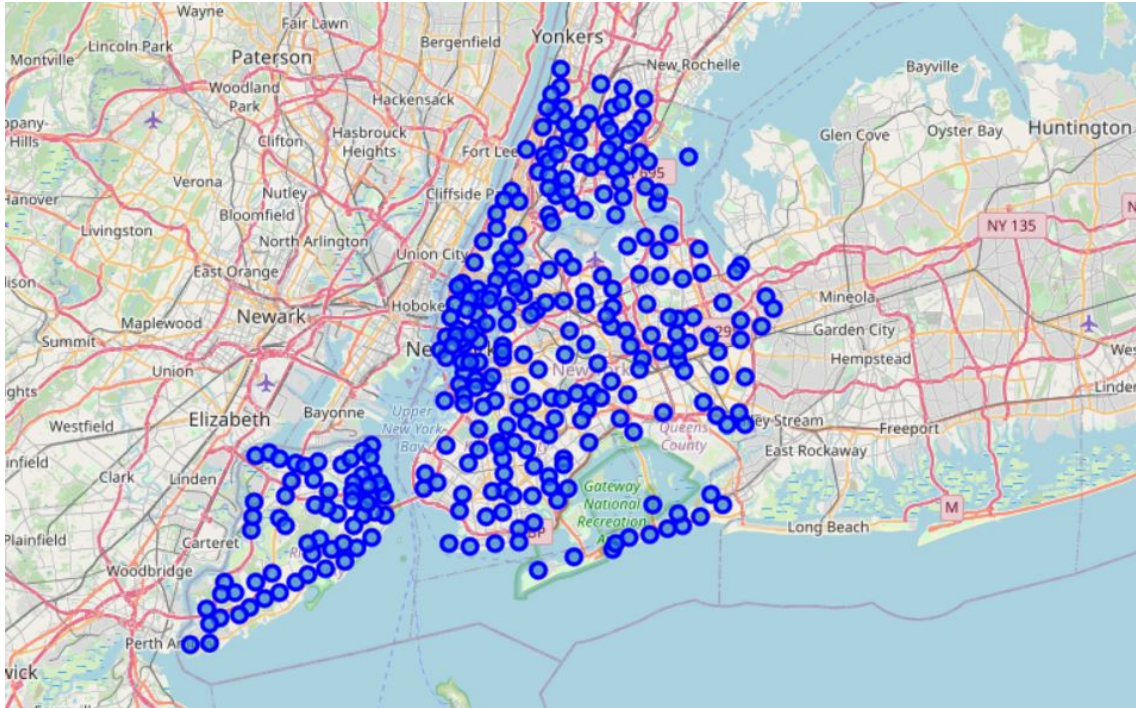
Foursquare location data was used to find out different venues around the neighbourhoods of Queens mainly locations of Indian Restaurants around the neighbourhood of Queens.

3. Methodology

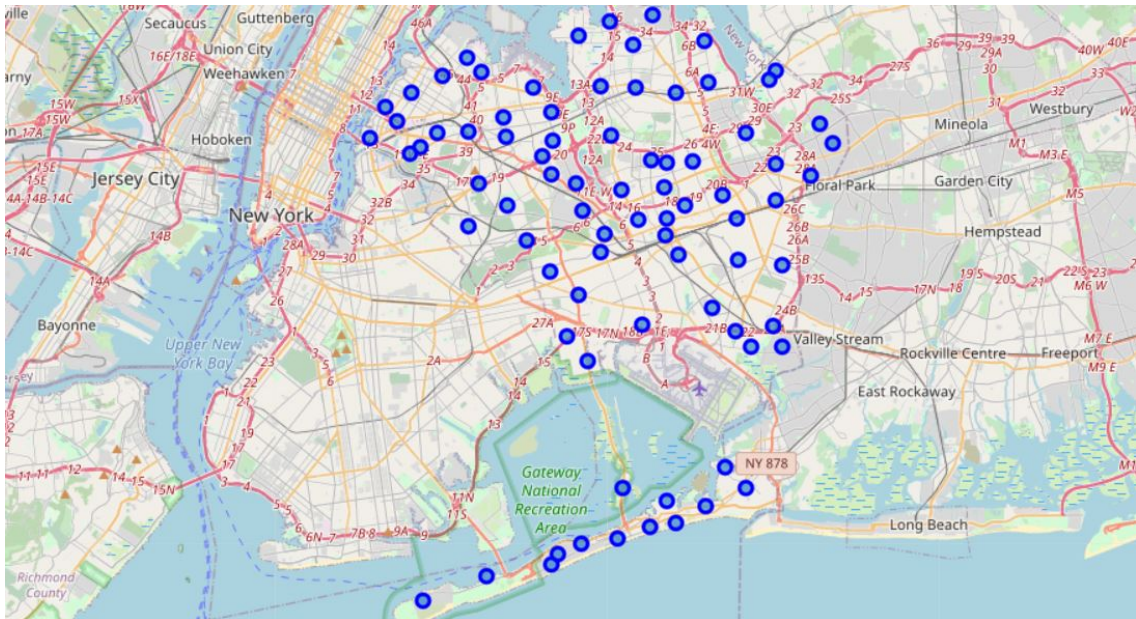
- The first step is to obtain the New York City data set. The data set consists information about 5 boroughs and 306 neighbourhoods. It also contains latitude, longitude information about the neighbourhoods. I then transform edthe data into data frames using pandas library to perform further operations on it.
- Using the New York population data set , I then found out the borough where more number of Indians reside. A new data frame is created using pandas which includes the population of all 5 boroughs. Queens was found out to be the most populated area with around 145000 people which is more than 6 percent of its total population.
- I then plotted the neighbourhoods of New York using the latitude and longitude data. This was done using the folium package. Neighbourhoods of Queens was also plotted using the folium package.
- I then gather information about all the neighbourhoods in queens and also find out different venues around queens using the foursquare location data. Using the foursquare api, I found out the 10 most common venues for each neighbourhood within 500m radius. Other information like venue latitude, longitude , number of unique venues was also analysed.
- After performing the one hot encoding on the data, I filtered out the mean of frequency of Indian Restaurants nearby Queens. I then created a new data frame which also includes the latitude and longitude information of the neighbourhoods
- Now the data is clustered into 4 different clusters using K-means clustering. I then plotted different clusters using the folium package. Later the clusters were examined.

4. Results

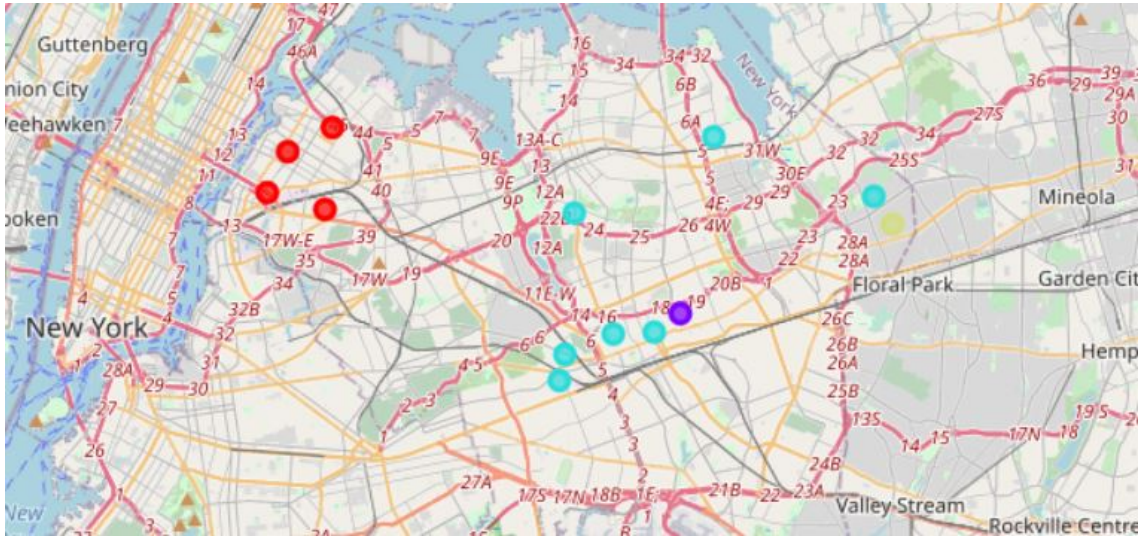
Map of neighbourhoods in New York



Map of neighbourhoods in Queens



Clusters of different neighbourhoods close to Indian Restaurants



From examining the above clusters, it can be said that people in neighbourhoods of Astoria, Long Island City, Ravenswood, Sunnyside gardens(all red cluster) and Briarwood, Richmond Hill, Jamaica Hills, Kew Garden(parts of blue cluster) prefer to eat Indian food. So I would suggest opening an Indian Restaurant in any of these mentioned neighbourhoods.

5. Conclusion

In this project, we successfully found out the best place to open an Indian Restaurant in the city of New York using Data Science and Machine learning methods. I would also like to mention that other factors like population density of Indians in neighbourhoods, number of people visiting the existing restaurants and other factors can be considered to make the model better.