DATA SCIENCE CAPSTONE

ANALYSIS OF NEIGHBOURHOODS IN QUEENS, NEW YORK

FOR SETTING UP OF A RESTAURANT

What is the purpose of this project?

Using the location data that we have acquired about Queens, NY, we can analyse neighbourhoods using different Machine Learning algorithms and Data Science techniques to visualise the neighbourhoods, find the list of areas with heavy population and diversity, find other restaurants in close proximity who are potential competitors, etc.,

By using these algorithms, we can arrive at a conclusion and finalise on the most appropriate location(s).

Who will it benefit?

This project will be of use to aspiring entrepreneurs who wish to set-up restaurants in Queens, that will ensure good profits and popularity. Using this project, one can determine the safest neighborhood and one that can minimise competition and thrive in terms of profits.

Other data science enthusiasts with a passion for analysis might also find this project interesting.

Data acquisition and cleaning

DATA SOURCES:

For this project, Foursquare data will be used. Foursquare is a company providing location data of different places in the world. By creating an account, we have access to their API with our unique credentials. Using those credentials, we can call their API for accessing their data at any point of time during analysis.

For analysing the safety of Queens, we have to acquire the dataset recording the daily crimes in the borough. A free dataset for the whole of New York from the web was acquired from NYC crime govt website.

Data cleaning:

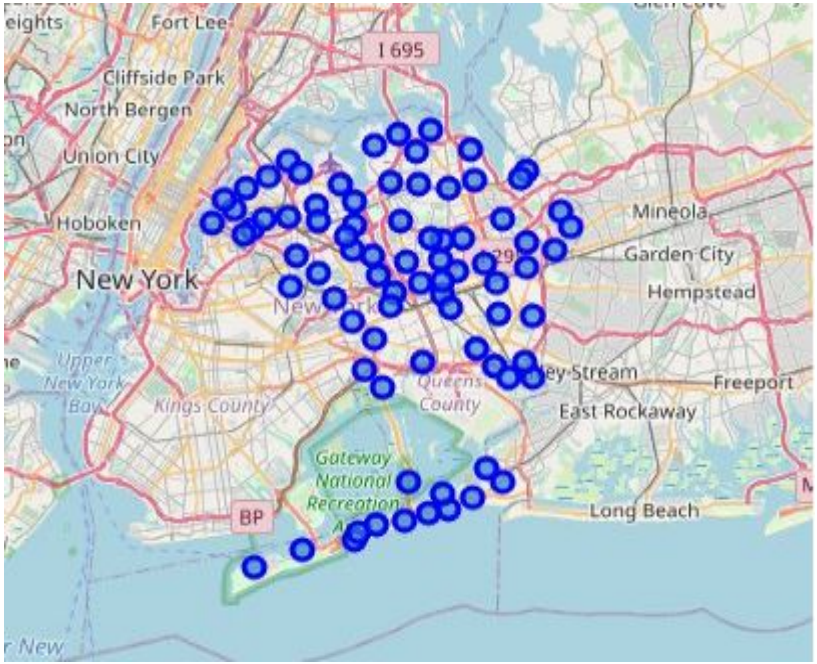
All unnecessary data such as those that contain no value in the analysis were removed.

Fields having null values were removed.

All time fields were removed.

All dates were changed to the conventional datetime format.

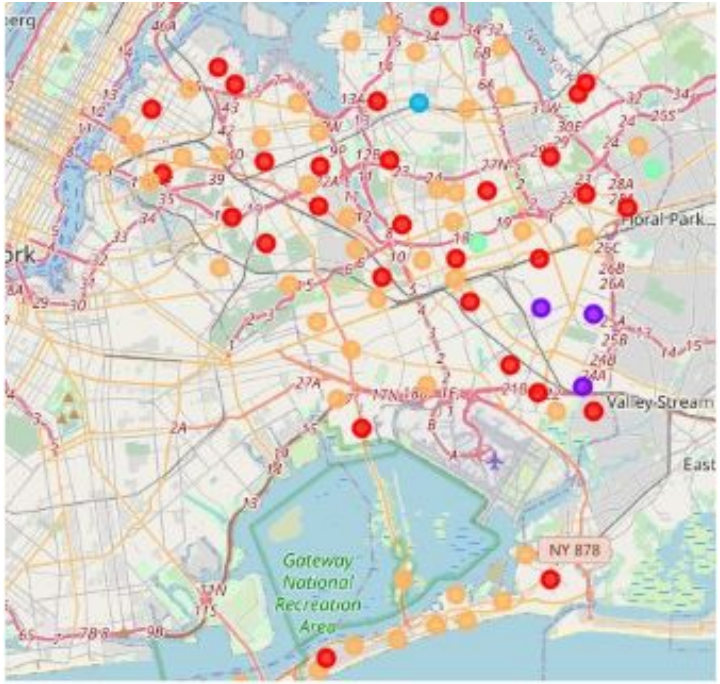
Visualising the map of Queens;



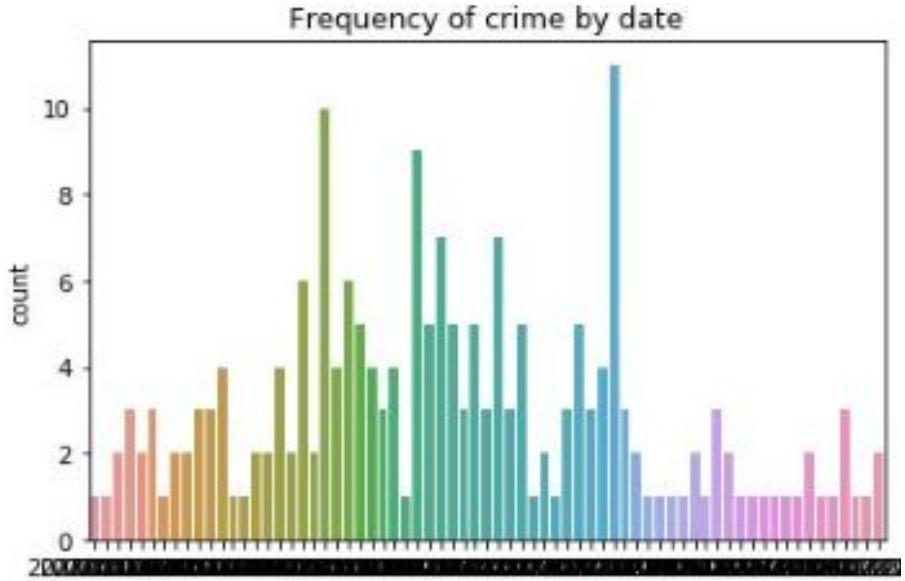
Modeling- Machine Learning Model

- For modeling the data, we are not going to perform any predictions.
- So a regression or classification type of modeling will prove inefficient for our analysis.
- By grouping similar restaurants together, we can find out what neighborhoods in Queens serve the same cuisine.
- We will use k-means clustering to group similar restaurants together as one.

Clustering


Floral Park
Valley Stream
East
NY 878
Gateway
National
Recreation
Area

Analysing crime data to see annual frequency



Crime plots.


Englewood
Ridgefield Park
Fort Lee
Ridgefield
Cliffside Park
Fairview
North Bergen
Guttenberg
Union City
Weehawken
Hoboken
New York

Conclusions

If the desired cuisine is Caribbean, then neighborhoods in cluster 1 would prove profitable. Those neighbourhoods are St Albans, Laurelton and Cambria Heights.

If the desired cuisine is Korean, then the neighborhood for that would be Murray Hill.

The neighborhood called Floral Park is famous for its Indian and vegetarian restaurants.

Conclusions (contd.)

If one wishes to set up a restaurant of Italian/South-east Asian cuisine, any neighborhood is preferable. Under confusion, one can decide using the crime data. Clearly the southern regions have a lesser crime rate. In that aspect, St Albans and Laurelton are the clear winners.

The aspiring owner can choose a cuisine which already has popularity in a neighborhood, but he/she should be wary of the competition it poses. My suggestion would be to choose a safe neighborhood with decent popularity of preferred cuisine and that should be the best way to go!