THE BATTLE OF NEIGHBORHOODS

APPLIED DATA SCIENCE CAPSTONE PROJECT

INTRODUCTION/BUSINESS PROBLEM

The clients, a couple of vegan chefs, are looking for an ideal location to open a vegan restaurant in New York City, particularly in the boroughs of Manhattan and Brooklyn

WHAT IS A GOOD LOCATION FOR A VEGAN RESTAURANT?

Characteristics of Neighborhood that would be a good location for a vegan restaurant

- Has other vegan/vegetarian restaurants
- Has health food stores
- Has yoga studios
- Has gym/fitness centers
- Has farmer's market
- Has hiking/bike trails

DATA SOURCES

Neighborhoods Data

New York city neighborhoods data

https://geo.nyu.edu/catalog/nyu_2451_34572

Foursquare API

METHODOLOGY

Data Acquisition

- json file containing New York City neighborhoods data is downloaded
- Location-based venues data is obtained by making calls to the Foursquare API

Data Wrangling

The neighborhoods data and venues data in json format are parsed and cleaned up and loaded into data frames.

METHODOLOGY(cont.)

Exploratory Data Analysis

The data frame was quickly examined to make sure that the data was loaded properly and data for all five boroughs and three hundred and six neighborhoods were present.

The venues data for all the neighborhoods in each borough is examined. The shape of the data frames is examined.

METHODOLOGY(cont.)

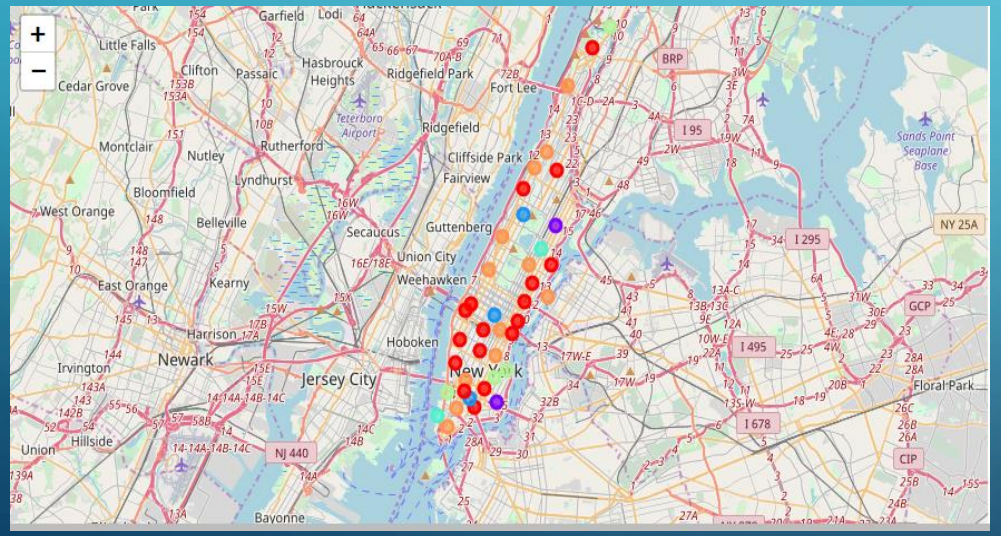
Data Analysis/Machine Learning

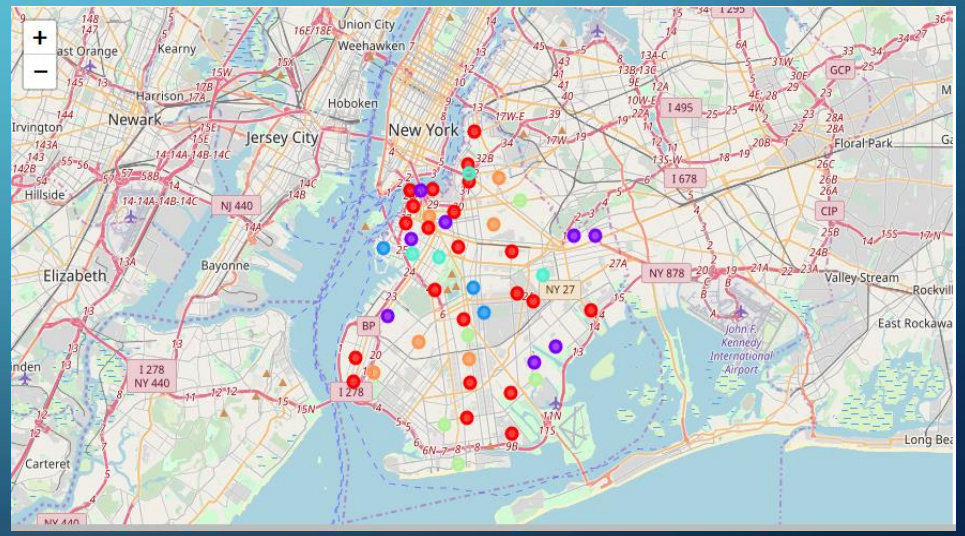
- The venues data for each neighborhood is grouped by taking the mean of the frequency of occurrence of each category of venue.
- The venues are then sorted in descending order of the mean.
- A new data frame is then created with the top 5 venues for each neighborhood.
- The neighborhoods in each borough are clustered using k-means clustering.

METHODOLOGY(cont.)

Data Visualization

The clusters in each boroughs are plotted on a map using folium map rendering library.





RESULTS

The clusters for Manhattan and Brooklyn neighborhoods were examined. Cluster 0 and cluster 5 in the Manhattan clusters, have neighborhoods that have a good mix of the type of businesses and facilities that we were looking for. So, we can pick a neighborhood in these 2 clusters as location for the vegan restaurant in Manhattan. Similarly, cluster 0 and cluster 1 in the Brooklyn clusters, have neighborhoods that have a good mix of the type of businesses and facilities that we were looking for.

DISCUSSION

At the start of the project, the assumption was made that by segmenting and clustering the neighborhoods based on the types of venues in those neighborhoods, we can select a neighborhood for the location of a vegan restaurant. After going through the analysis, we were able to identify a set of neighborhoods that would be suitable locations for a vegan restaurant. Additional demographic data for the neighborhoods would be useful in narrowing it down further.

CONCLUSION

The business problem that was addressed in this project was finding a location for a vegan restaurant in the boroughs of Manhattan and Brooklyn in New York city. The New York city neighborhoods data and foursquare location-based venues data were used to analyze the neighborhoods. K-means clustering was used to divide the neighborhoods into clusters. From these clusters, the clusters that had the neighborhoods that would be suitable for a vegan restaurant were selected.