COURSERA CAPSTONE PROJECT THE BATTLE OF NEIGHBORHOODS

AKSHAYA SATAM

INTRODUCTION

- Background: The average American moves about eleven times in their lifetime. We should always do proper research when planning our next move in life. Safety is a top concern when moving to a new area. If you don't feel safe in your own home, you're not going to be able to enjoy living there.

- Problem: The project aims to select the safest areas in each borough in New York based on the lowest crime rates, explore the neighborhoods of that borough to find the 5 most common venues in each neighborhood and finally cluster the neighborhoods using k-mean clustering.

- Interest: People who are considering relocating to New York or the travelers planning to visit this city will be interested to identify the safest borough in NYC and explore its neighborhoods and common venues around each neighborhood.

DATA DESCRIPTION

Based on definition of our problem, factors that will influence our decisions are:

- finding the safest borough based on crime statistics
- finding the 5 most common venues
- choosing the right neighborhood within the borough

Data required for this problem analysis is a combination of following sources:

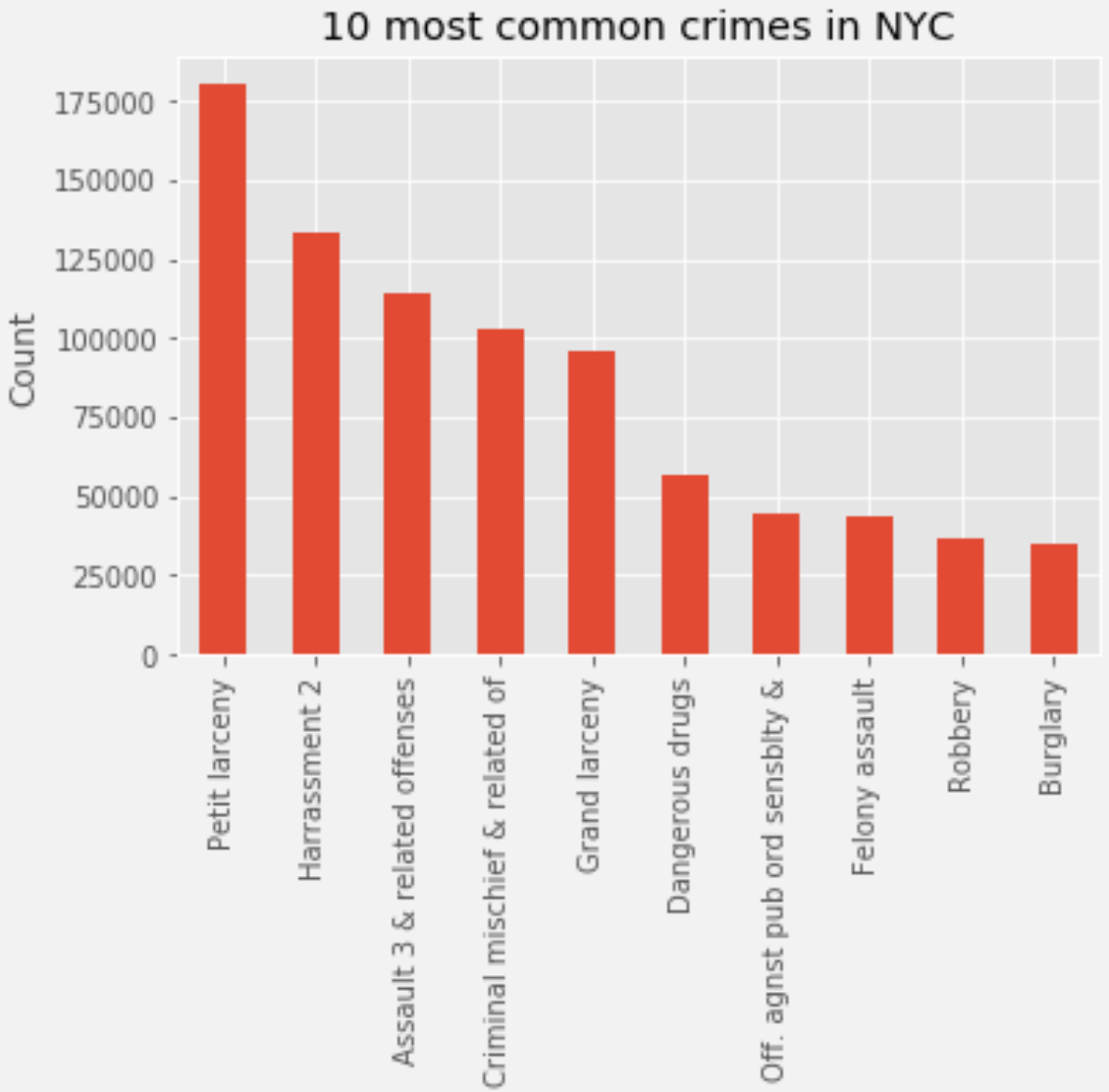
- New York City crimes reported from NYPD
- Population of NYC by Borough
- List of Neighborhoods in Staten Island acquired from Foursquare API

METHODOLOGY

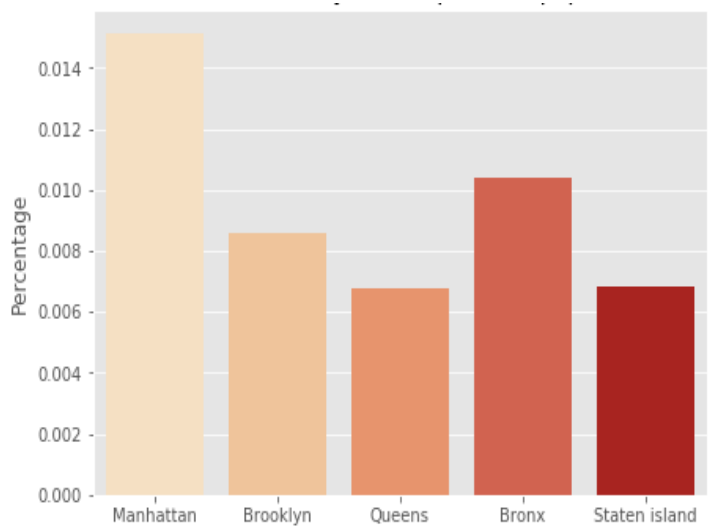
- Exploratory Data Analysis

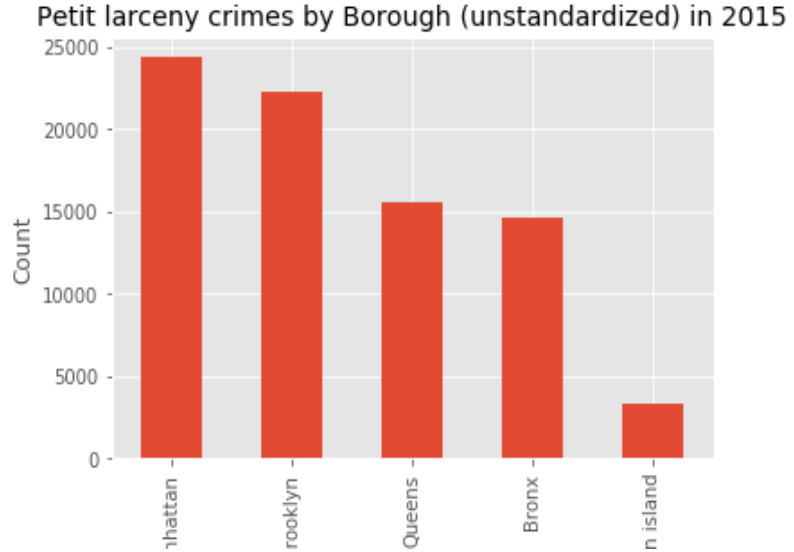
10 Most Common Crimes in NYC

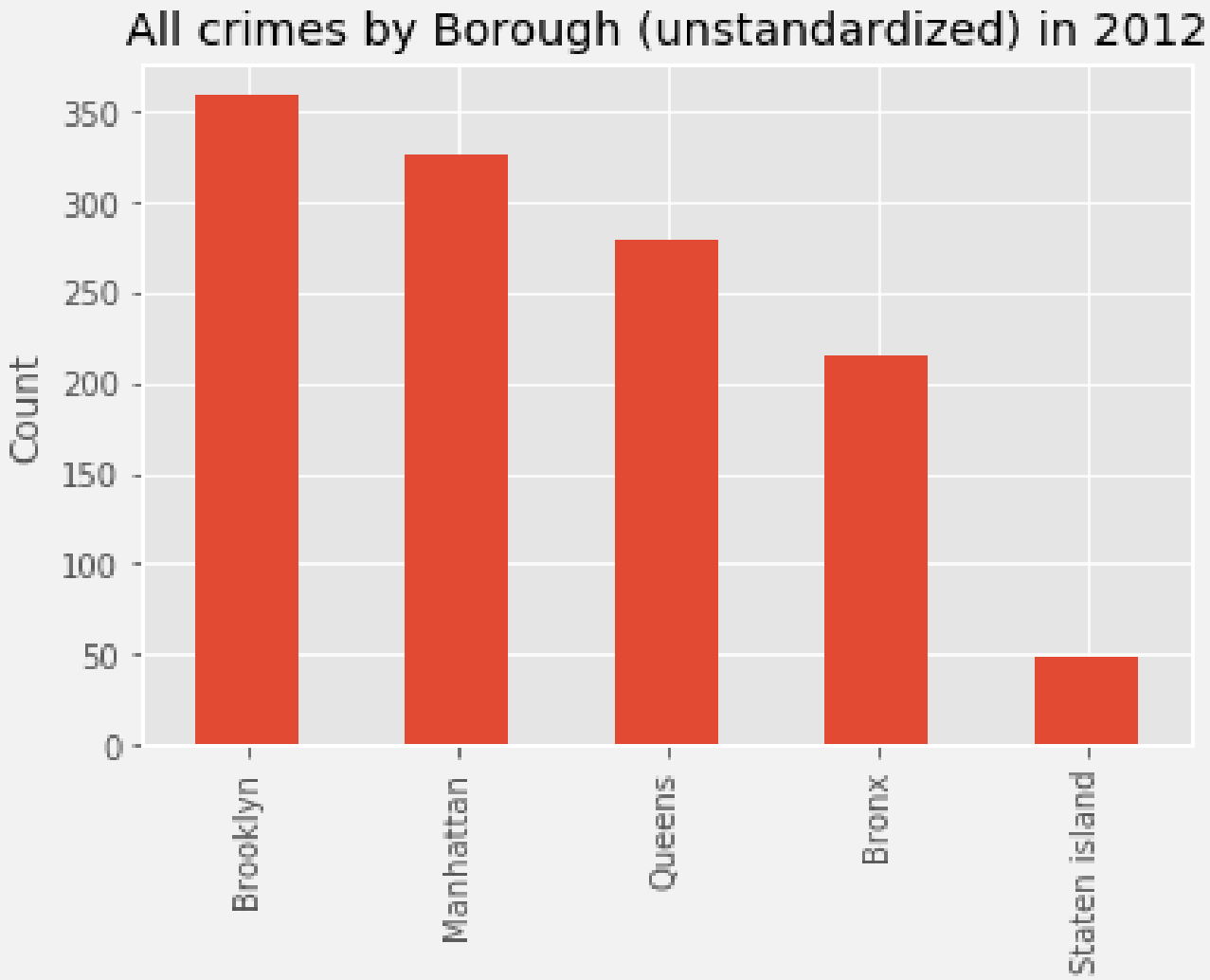
'Petit larceny' is the highest reported crime and 'Burglary' is the lowest recorded Crime. This helps the target audience know which crime has higher chance occurrence.



PETIT LARCENY CRIMES PER BOROUGH IN THE YEAR 2015 (STANDARDIZED AND UNSTANDARDIZED)

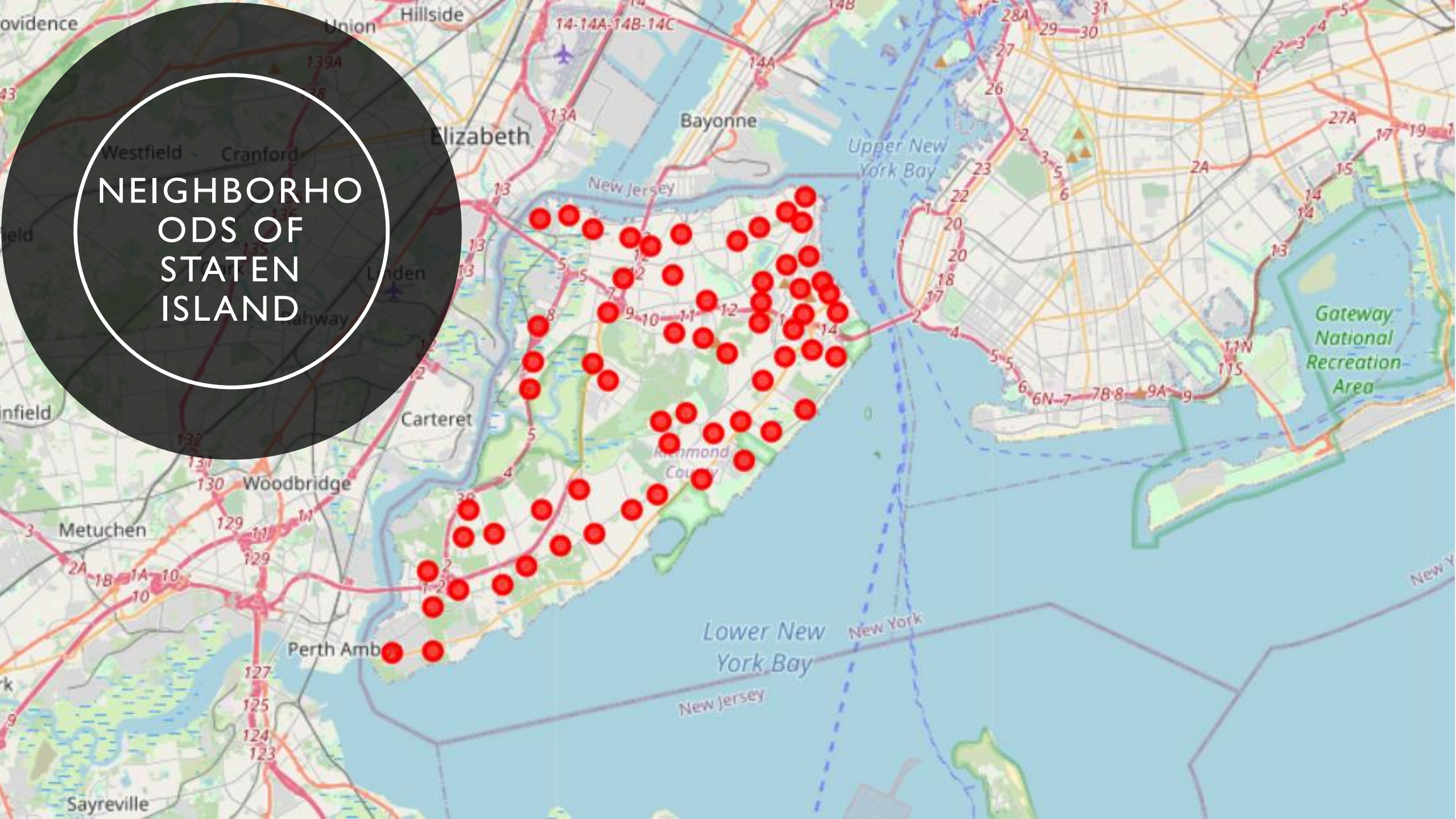






BOROUGHS WITH HIGHEST CRIME RATES

- Comparing five Boroughs with the highest crime rate during 2012 it is evident that Brooklyn has the highest crimes recorded and Staten Island has the lowest crime rate.
- All these plots gives us a relative analysis showing that Staten Island is the safest borough when compared to the crime rates.

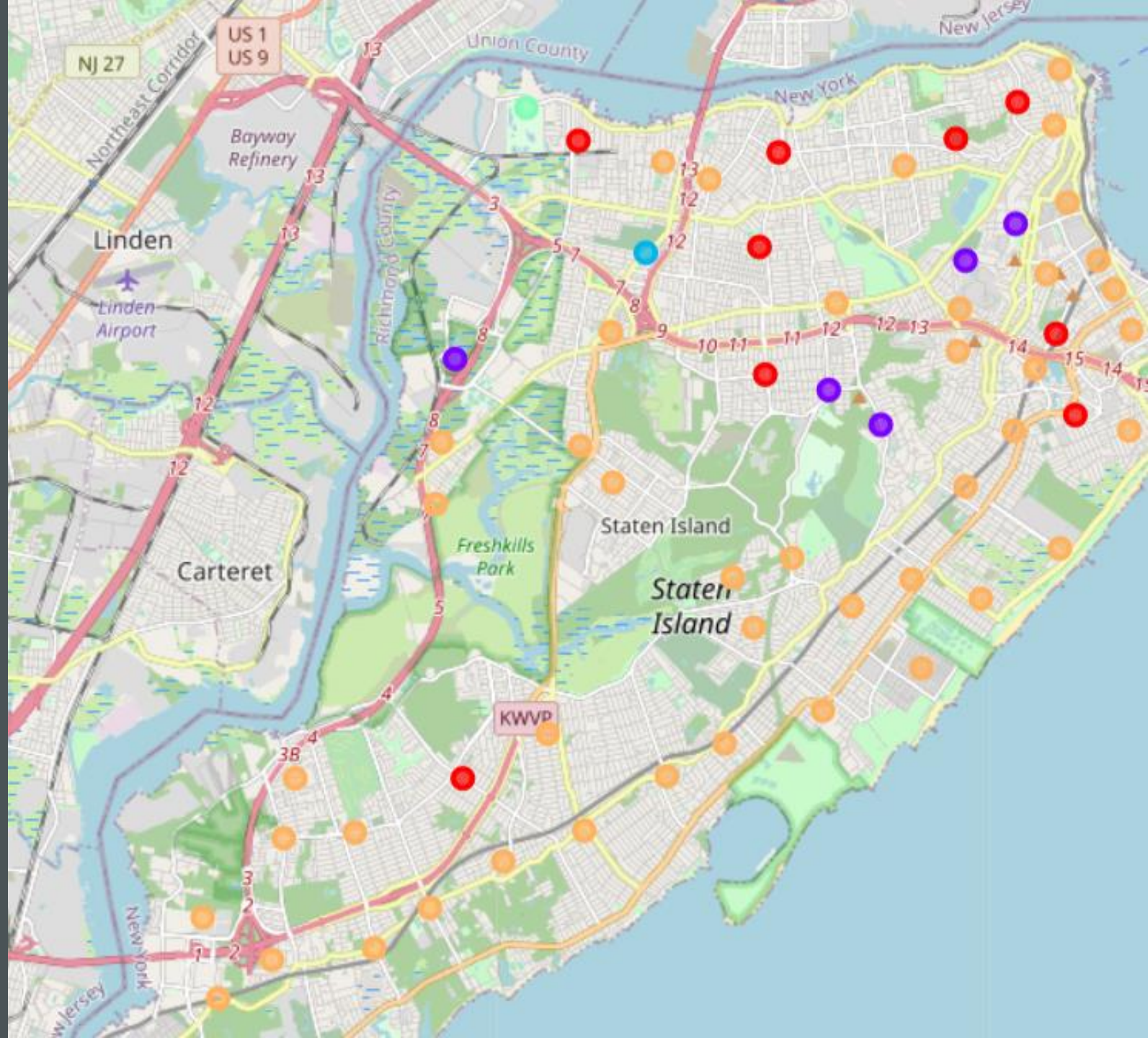

NEIGHBORHOODS OF STATEN ISLAND
Hillside
Union
Elizabeth
Bayonne
Westfield
Cranford
Carteret
Woodbridge
Metuchen
Perth Amboy
Sayreville
New Jersey
Upper New York Bay
Lower New York Bay
New York
New Jersey
Gateway National Recreation Area
Richmond County

DATA MODELLING

- Using the final dataset containing the neighborhoods in Staten along with the latitude and longitude, we can find all the venues within a 1000-meter radius of each neighborhood by connecting to the Foursquare API.
- One hot encoding is done on the venues data.
- One hot encoding is a process by which categorical variables are converted into a form that could be provided to ML algorithms to do a better job in prediction
- To help people find similar neighborhoods in the safest borough we will be clustering similar neighborhoods using K - means clustering which is a form of unsupervised machine learning algorithm that clusters data based on predefined cluster size.

RESULTS

- Applying K-means clustering we can access each cluster created to see which neighborhoods were assigned to each of the five clusters
- Visualizing the clustered neighborhoods on a map using the folium library, each cluster is color coded for the ease of presentation, we can see that majority of the neighborhood falls in the orange cluster which is the fourth cluster.



DISCUSSION

- The aim of this project is to help people who want to relocate to the safest borough in New York, expats can choose the neighborhoods to which they want to relocate based on the most common venues in it.
- For example, if a person is looking for a neighborhood with good connectivity and public transportation, we can see that Clusters 4 and 5 have Train stations and Bus stops as the most common venues.
- If a person is looking for a neighborhood with stores and restaurants in a close proximity, then the neighborhoods in the third cluster is suitable.
- For a family I feel that the neighborhoods in Cluster 4 are more suitable dues to the common venues in that cluster, these neighborhoods have common venues such as Parks, Gym/Fitness centers, Bus Stops, Restaurants, Electronics Stores and Soccer fields which is ideal for a family.
- The choices of neighborhoods may vary from person to person.

CONCLUSION

- This project helps a person get a better understanding of the neighborhoods with respect to the most common venues in that neighborhood.
- It is always helpful to make use of technology to stay one step ahead i.e. finding out more about places before moving into a neighborhood.
- We have just taken safety as a primary concern to shortlist the safest borough of New York.
- The future of this project includes taking other factors such as cost of living in the areas into consideration to shortlist the borough, such as filtering areas based on a predefined budget.