NEIGHBOURHOOD SAFETY IN NEW YORK CITY

NEIGBHOURHOOD SAFETY

INTRODUCTION

1. **When you choose to live in New York, NY, you are choosing a lot more than a place to call home. You are choosing a lifestyle, defined by a city teeming with possibilities**
2. It is important to be proactive about one's safety, especially when it comes to choosing a place to live. After all, the safety of your neighborhood affects everything from one's happiness and sense of security to one's home resale value.

OBJECTIVE

- The objective of this capstone project is to analyze and profile neighborhoods across all boroughs in New York City based on safety. Using data science methodologies and machine learning techniques like clustering, this project aims to provide solutions to answer the question: If a homebuyer or renter were looking for the safest place to stay in NYC, where would you recommend?

DATA

To solve the problem, we will need the following data:

- Neighborhoods across New York. This defines the scope of the project that is confined to the city of New York and its five boroughs.
- Latitude and longitude coordinates of those neighborhoods. This is required in other to plot the map and give venue data
- Crime data related to arrests for some specific offences and venue data related to police departments. We will use this data to perform clustering on the Neighborhoods.

DATA SOURCES

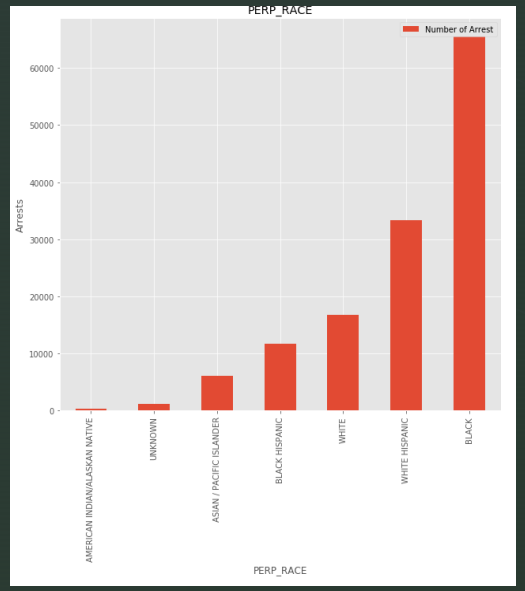
- The Crime Data is a List of every arrest in NYC in 2017. This data is manually extracted every quarter and reviewed by the Office of Management Analysis and Planning before being posted on the NYPD website. Each record represents an arrest effected in NYC by the NYPD and includes information about the type of crime, the location and time of enforcement. The data can be found here :

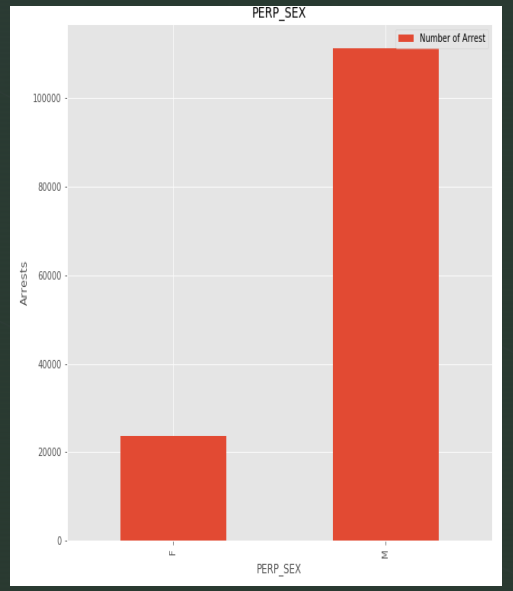
https://data.cityofnewyork.us/Public-Safety/NYPD-Arrests-Data-Historic-/8h9b-rp9u

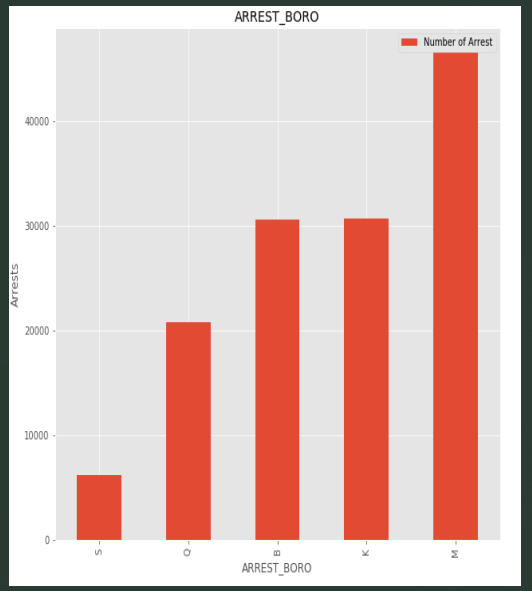
METHODOLOGY

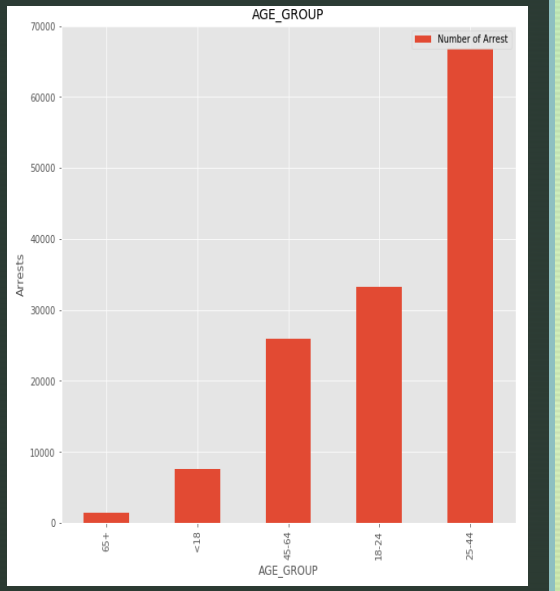
- The data contains 18 attributes or columns that describe every arrest in New York documented in the year 2017.
- However there isn't any attribute indicating which neighborhood in New York city each arrest coordinate belongs to. To extract the coordinates, the geo-names reverse geo-coding API was used to import neighborhoods for each arrest coordinate. The API provided neighborhood attribute for 134,832 arrest instances while other arrest instances with unknown neighborhoods were removed from the final data.
- The data consists of 265 unique crime categories; therefore, one cannot simply choose a neighborhood to relocate to just because it is in a borough with lower overall crimes committed. Hence, the top 5 crime categories in each neighborhood was also highlighted.

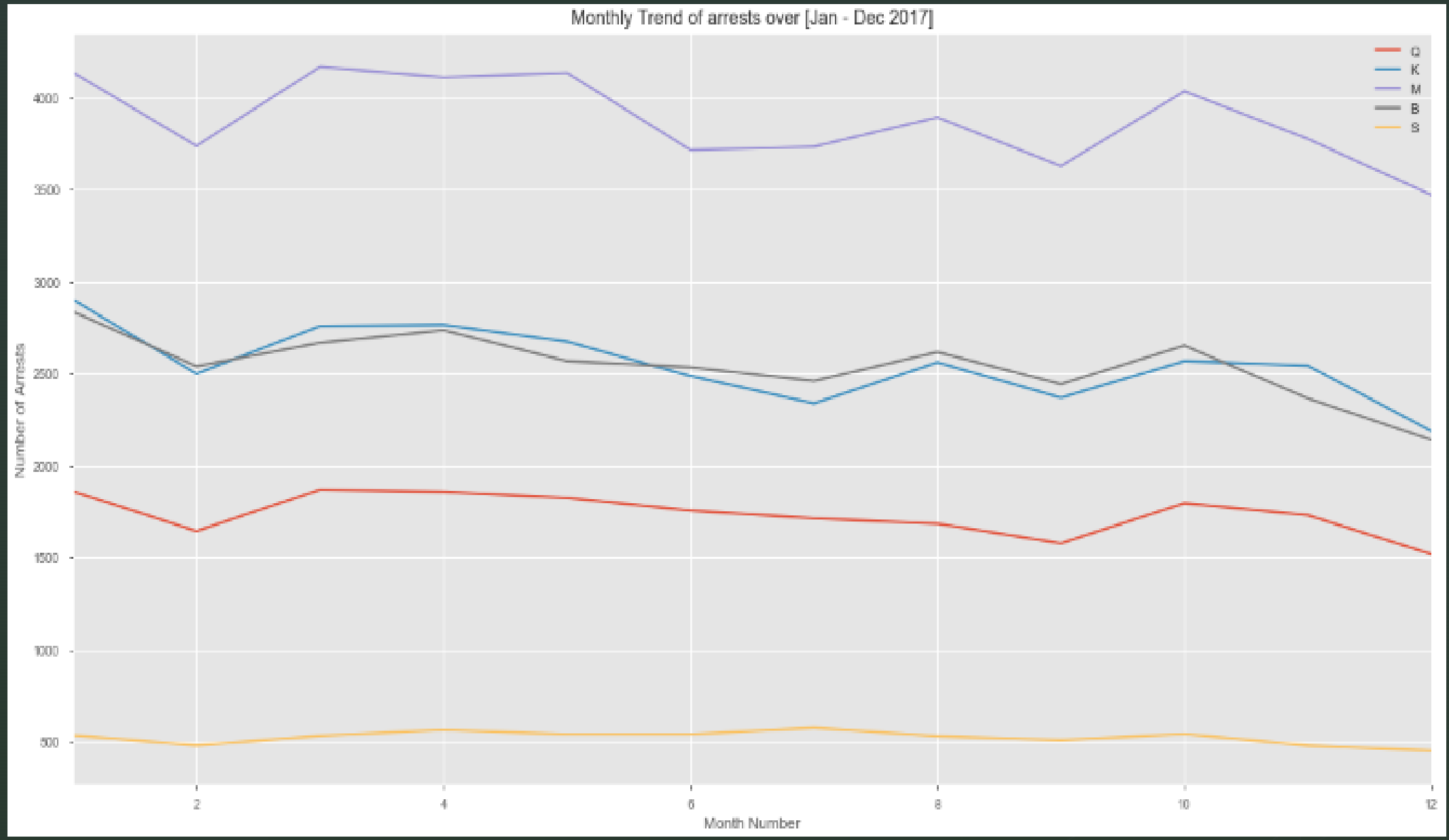
ANALYSIS


PERP_RACE
Number of Arrest
Arrests
60000
50000
40000
30000
20000
10000
0
AMERICAN INDIAN/ALASKAN NATIVE
UNKNOWN
ASIAN / PACIFIC ISLANDER
BLACK HISPANIC
WHITE
WHITE HISPANIC
BLACK
PERP_RACE

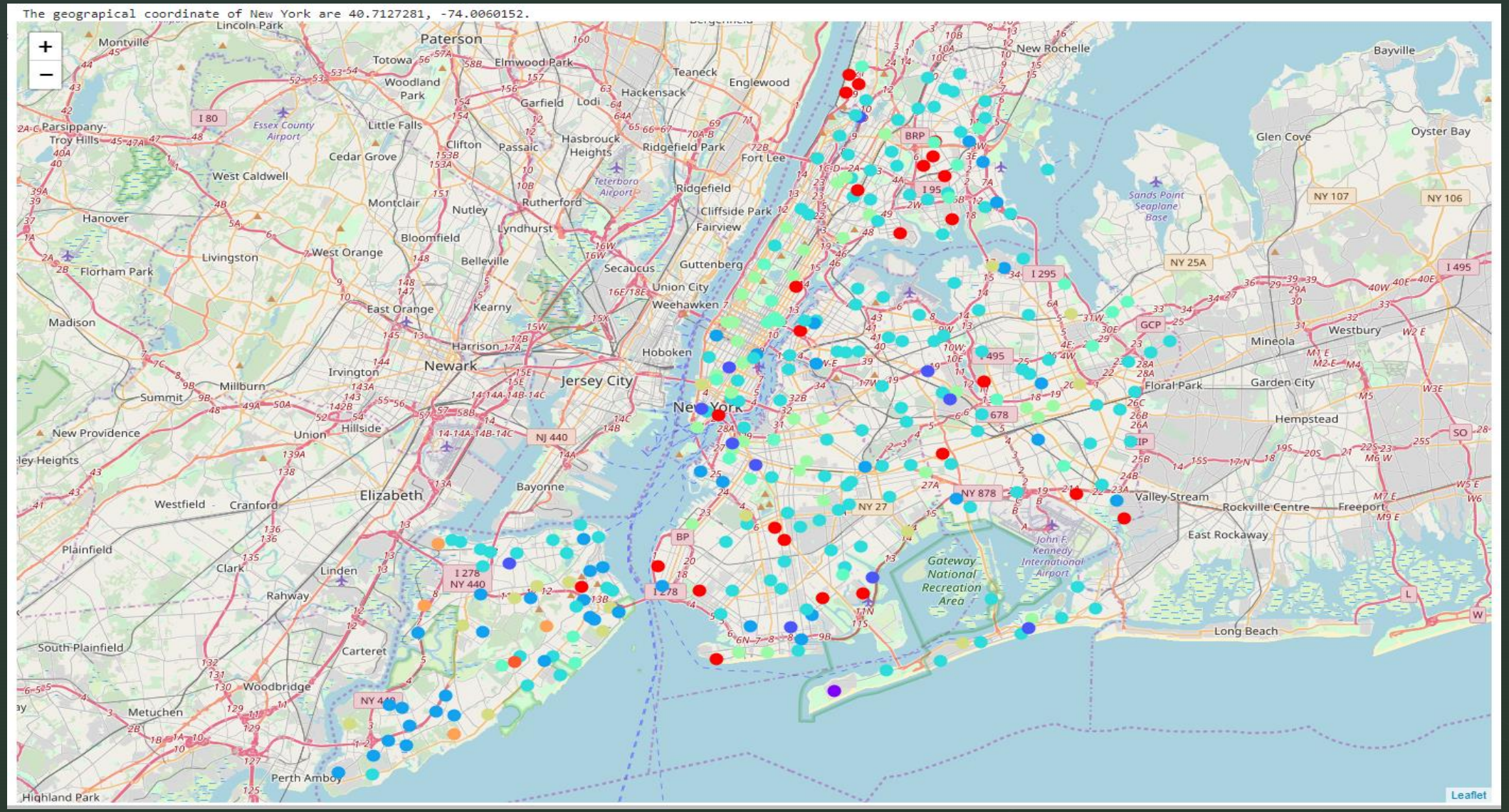

PERP_SEX
Number of Arrest
Arrests
100000
80000
60000
40000
20000
0
F
M
PERP_SEX


ARREST_BORO
Number of Arrest
Arrests
40000
30000
20000
10000
0
S
Q
B
K
M
ARREST_BORO


AGE_GROUP
Number of Arrest
Arrests
70000
60000
50000
40000
30000
20000
10000
0
65+
<18
45-64
18-24
25-44
AGE_GROUP


Monthly Trend of arrests over [Jan - Dec 2017]
Q
K
M
B
S
4000
3500
3000
2500
2000
1500
1000
500
Number of Arrests
2
4
6
8
10
12
Month Number

Cluster Map Plot



DISCUSSION & CONCLUSION

- A neighborhood in Staten highland might have crime characteristics that is similar to neighborhoods in boroughs with higher crime rates and vice versa. Therefore, to properly identify neighborhoods that share similar properties, it is important to reference the plot of clustered neighborhoods on the Map plot. This is helpful especially when trying to find a neighborhood in another borough that shares similar crime properties but has lower real-estate cost compared to another neighborhood of interest in a similar or different borough.
- This analysis will be very important for individuals who intend to relocate to New York City but would like to identify the safer neighborhoods in any borough of choice. In addition, tourists and real estate agents can benefit from this analysis. The results can be further improved on if neighborhood attribute information is available for all the data instance of arrests, data from other years is also analyzed and average cost of houses in each of the neighborhood is put into consideration as part of the analysis.