Analysing New York City Police Department (NYPD) Arrest Data For Crime Insights And Policing Strategies.

Group 2

Sai Manish Reddy Pannala - G01403888
Priyadarsini Devarapalli - G01354830
Sravani Pulusu - G01403752

Overview

01 Introduction

02 Description

03 Significance

04 Motivation

05 Visualization & Analytics

06 Related Work

07 Future Work

08 Conclusion

Introduction

- New York City, one of the world's busiest cities, is home to a diverse range of neighborhoods, cultures, and dynamics. As such, policing this city requires evidence-based strategies that are both effective and equitable.
- The NYPD Arrest dataset provides a wealth of information, including the type of crime, location, time of enforcement, and suspect demographics. The (NYPD) Arrest Data dataset, available on data.gov, is a comprehensive repository of information regarding arrests made within the city.
- The dataset offers a unique opportunity to investigate the particulars of law enforcement, crime, and community dynamics in New York City's diverse .
- This project aims to analyze the NYPD Arrest Data using data science and analytics to answer crucial questions regarding crime patterns, law enforcement practices, and community impact.

Challenges in Policing New York City: A Data-Driven Approach

Description of problem

- Crime is a significant problem in New York City, and understanding crime patterns and trends is crucial for developing successful crime prevention and reduction tactics.
- While this dataset is rich in information, it is also complex and large, making it difficult to extract useful insights manually. Although the NYPD Arrest Data dataset contains a large amount of arrest information, it is not used to guide crime prevention measures.
- The aim of this project is to analyze and predict crime patterns using machine learning techniques based on this dataset. Outline the project's specific objectives, such as providing valuable insights, enhancing public safety, and building stronger community relations.

Motivation

The motivation behind selecting the NYPD Arrest Data (Year to Date) dataset for this project is driven by several key factors:

- **Enhancing Public Safety**
- **Community Policing Enhancement**
- **Resource Allocation Optimization**
- **Analyzing Historical Crime Trends**
- **Data-Driven Decision-Making**

Bridging Data to Impact:

The Significance of Analyzing NYPD Arrest Data

01 Analyzing crime patterns provides insights into the types and locations of crimes, enabling the development of targeted prevention strategies.

02 Predictive analytics guide strategic resource allocation by identifying areas with higher crime rates, optimizing the deployment of NYPD resources.

03 Data-driven insights into crime trends support proactive community policing, enhancing positive interactions.

04 Data-driven decision-making promotes transparency by providing a clear basis for law enforcement actions, contributing to increased accountability and public trust.

Data Preprocessing

- We transformed an initial dataset comprising 112,507 rows into a refined dataset containing 3,104 entries. Column names such as ARREST_KEY, ARREST_DATE, ARREST_BORO, AGE_GROUP, PERP_SEX, PERP_RACE etc..

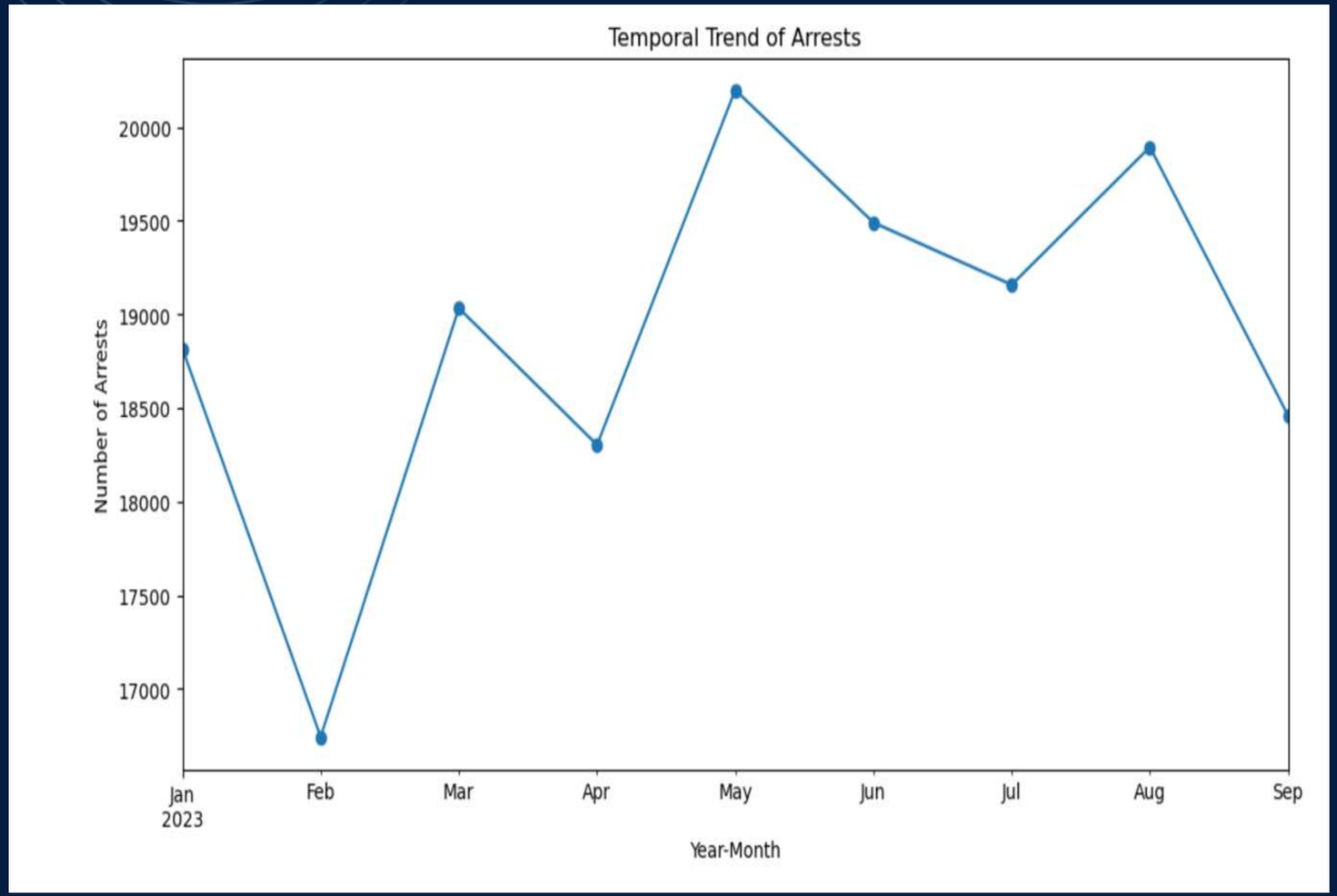
- Cleaning process involved addressing critical issues related to missing values and outliers to ensure that the dataset was in a suitable state for analysis.

- Missing data, when present, can significantly impact the reliability, Outliers, which can skew statistical analyses and visualizations, were identified and removed

Visualization

Are certain crimes more common during specific seasons as socioeconomic conditions ?

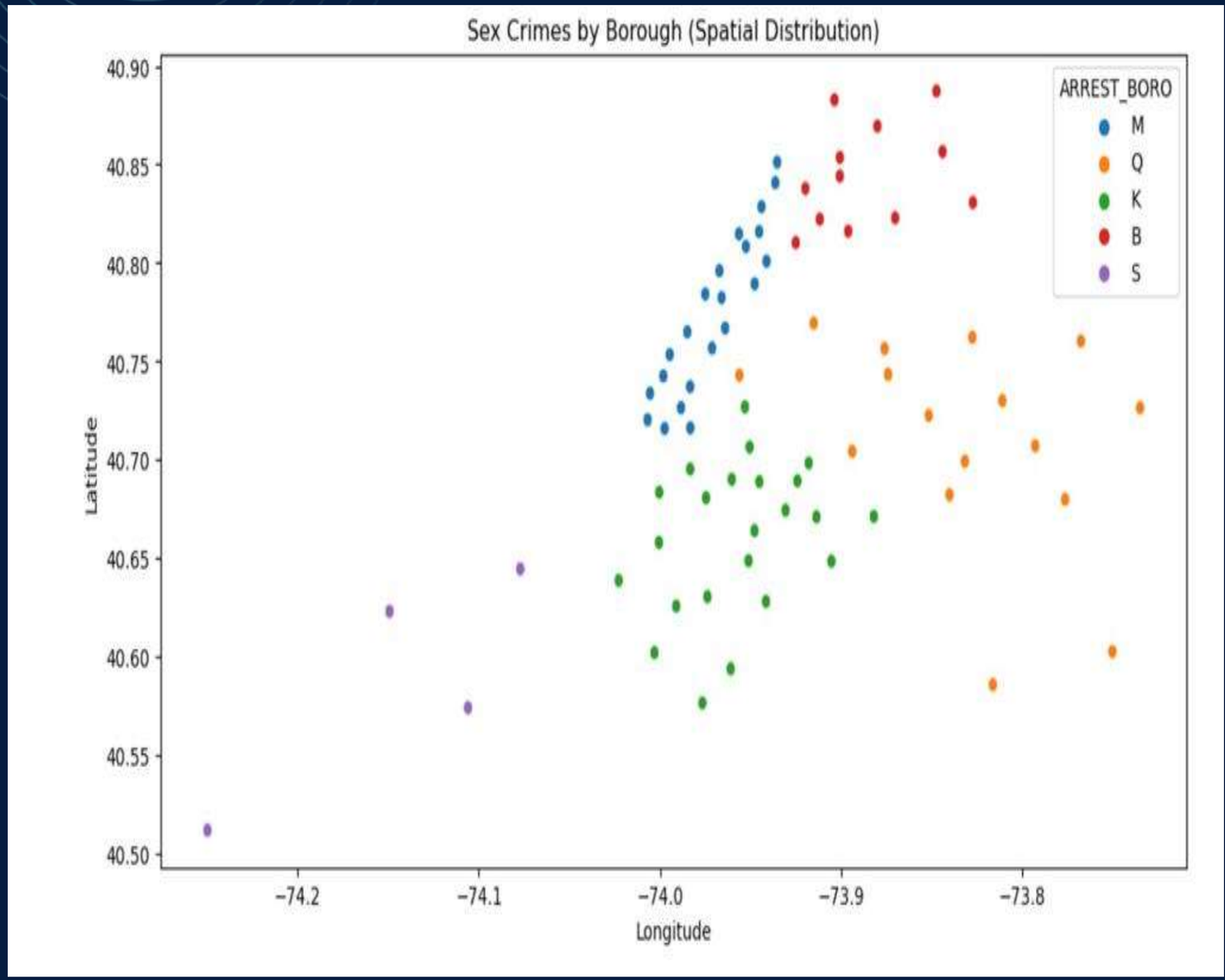
- The line chart reveals crime patterns throughout the year, peaking in May (20,000) and August (19,500), possibly influenced by warmer weather and increased social activities. February records the lowest crime rates, likely due to colder weather and decreased outdoor interactions.



Visualization

What specific neighborhood or borough in New York City exhibits a higher incidence of sex crimes based on the analysis of the NYPD Arrest Data?,

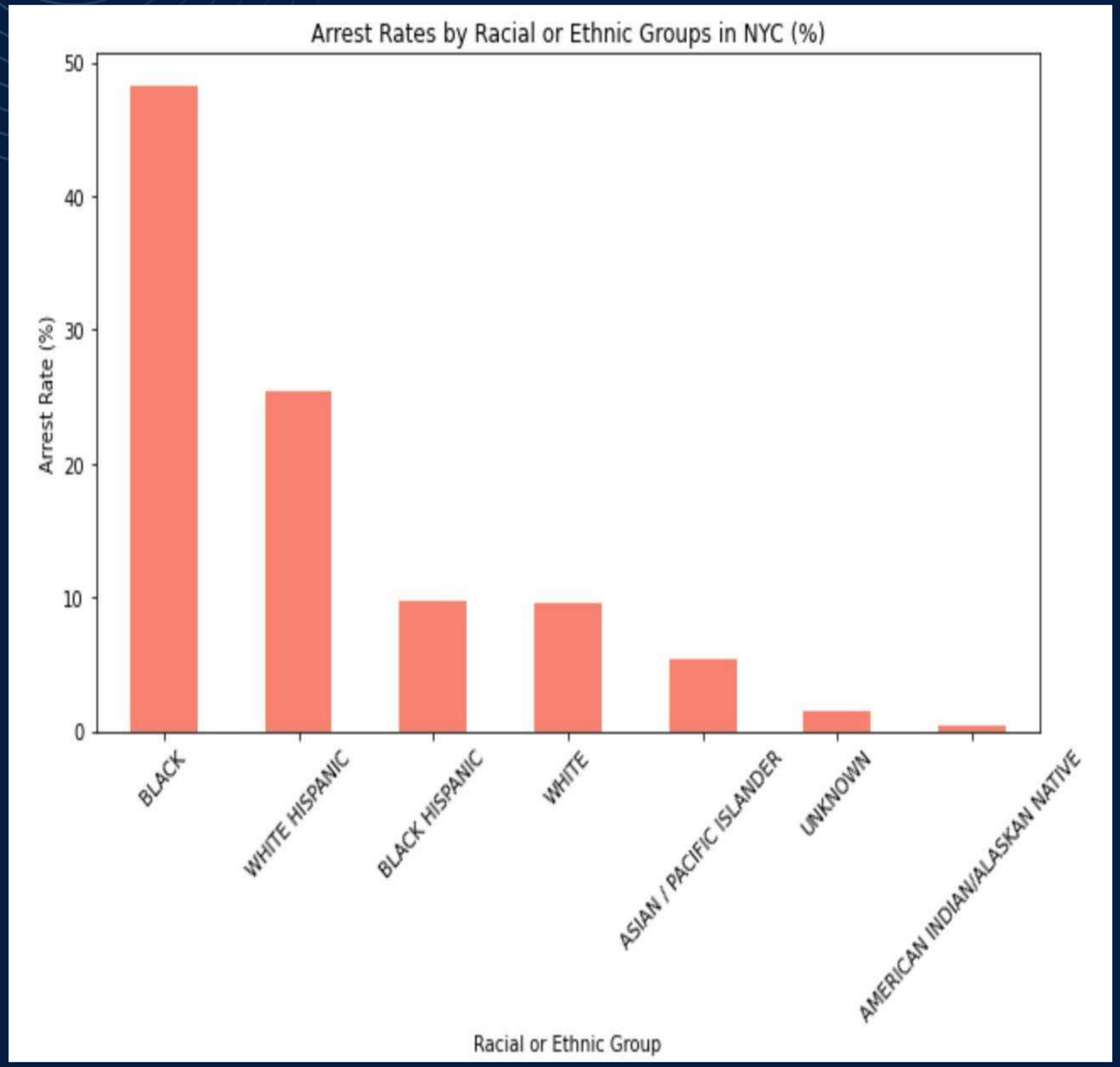
- The scatter plot reveals concentrated sex crime occurrences in the Bronx and Manhattan, suggesting potential influences from factors like population density and socio-economic conditions. In contrast, Staten Island records the fewest instances,



Visualization

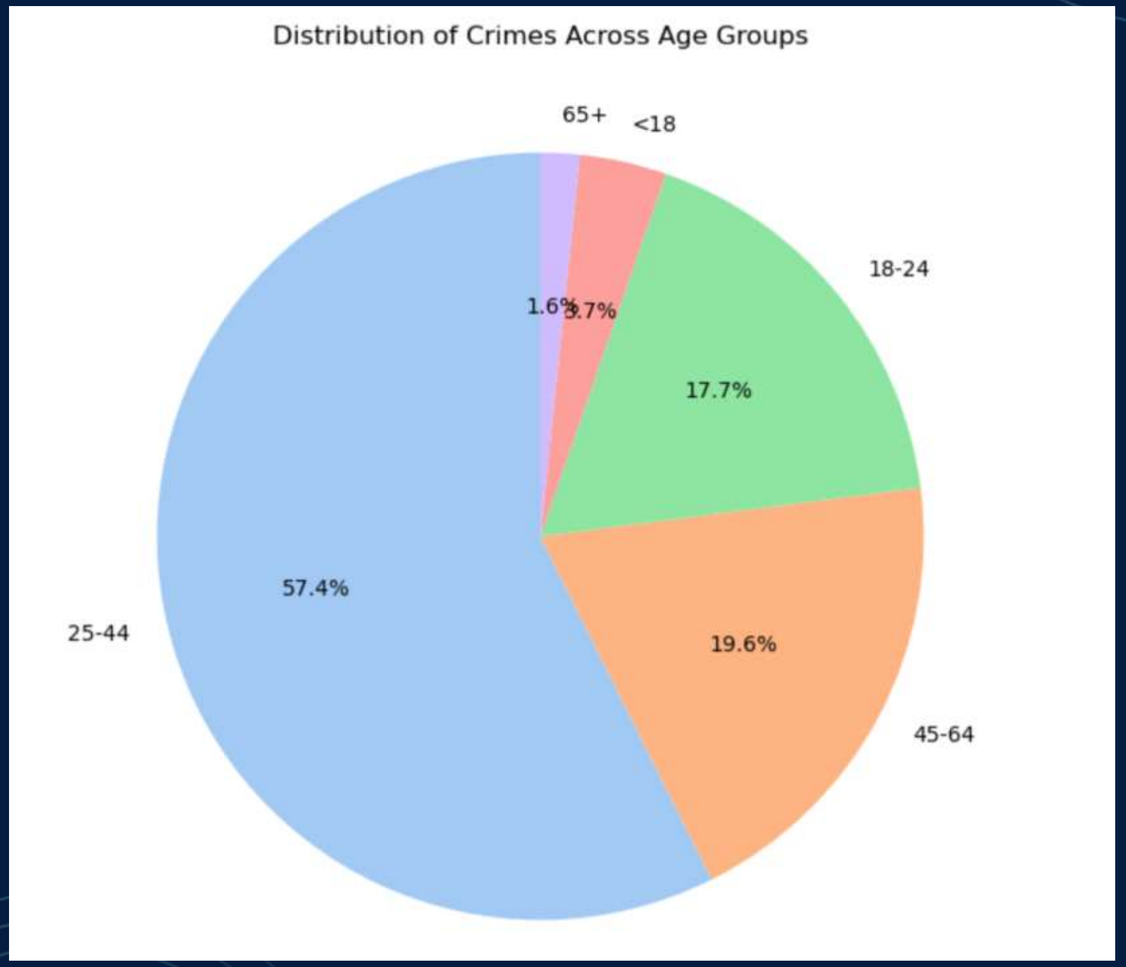
Is there a discernible variation in arrest rates among different racial or ethnic groups based on the analysis of NYPD Arrest Data?

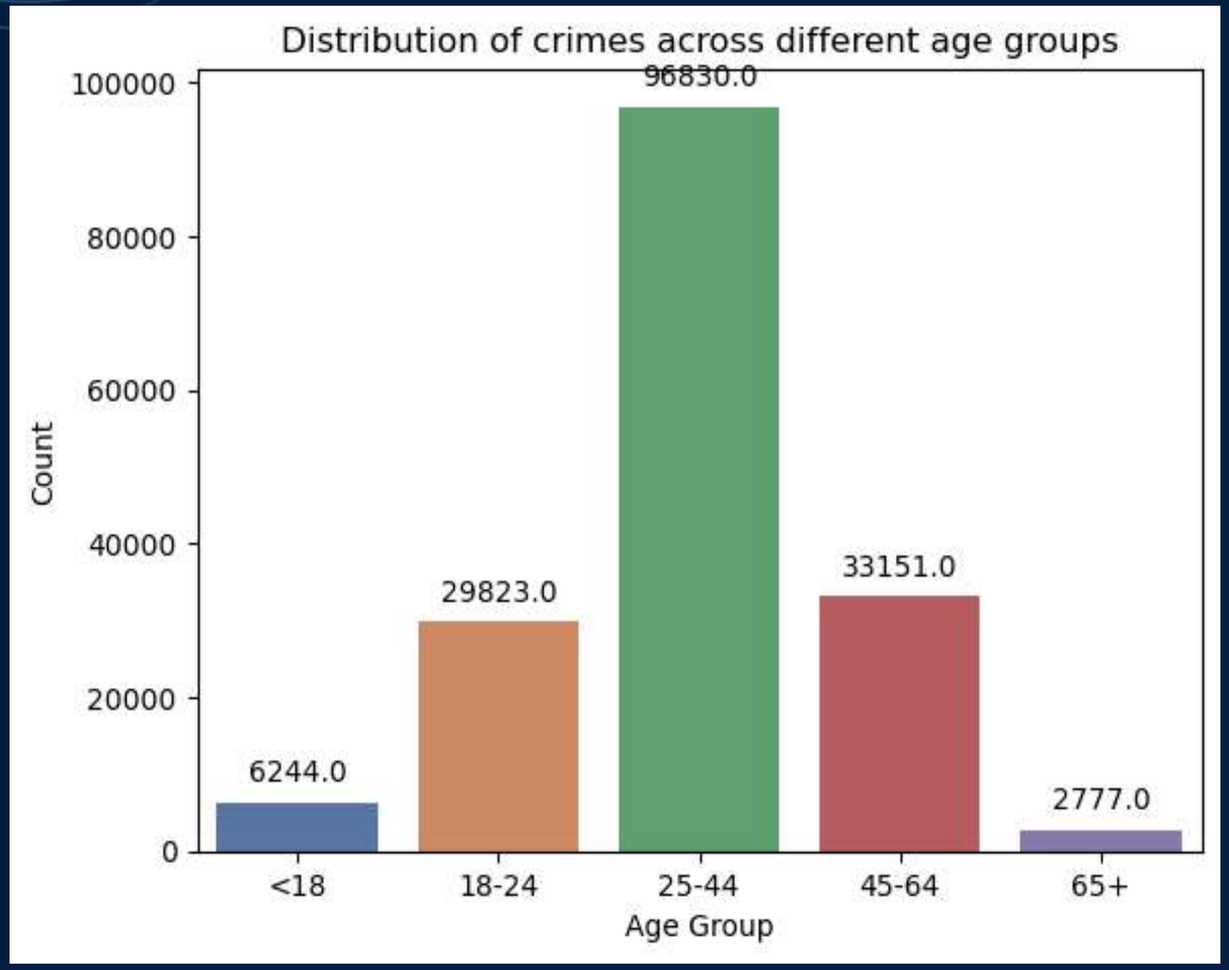
- The bar graph visually represents arrest rates across various racial or ethnic groups, highlighting disparities in law enforcement interactions. Each bar corresponds to a specific group, offering a comparative view of arrest frequencies.



Visualization

- **What age demographic exhibits a higher prevalence of criminal activity based on the analysis of NYPD Arrest Data?"**





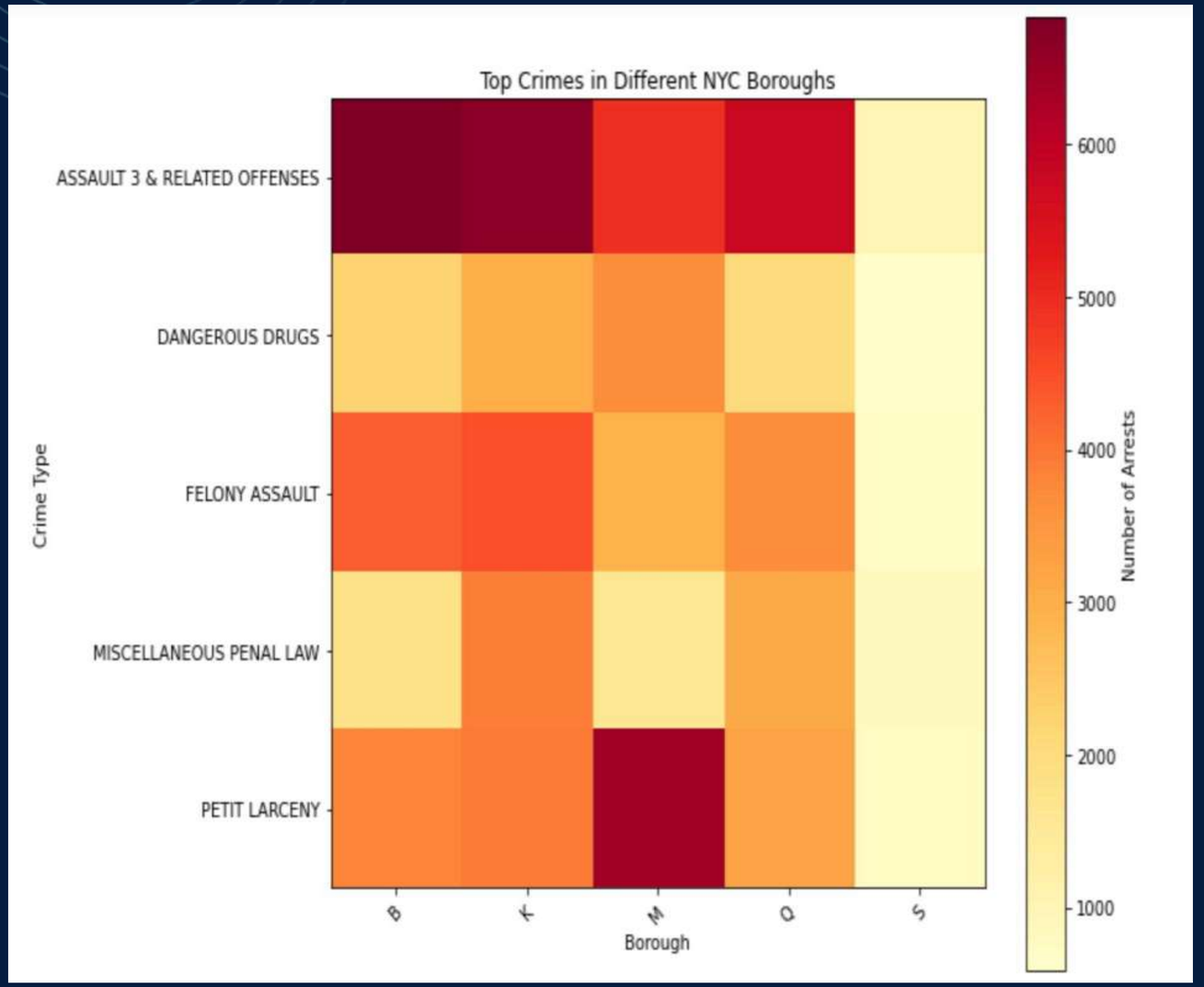
- The pie chart visually represents the distribution of criminal activity across different age groups

- The bar graph illustrates the distribution of criminal activities among age groups, revealing peaks in certain age brackets.

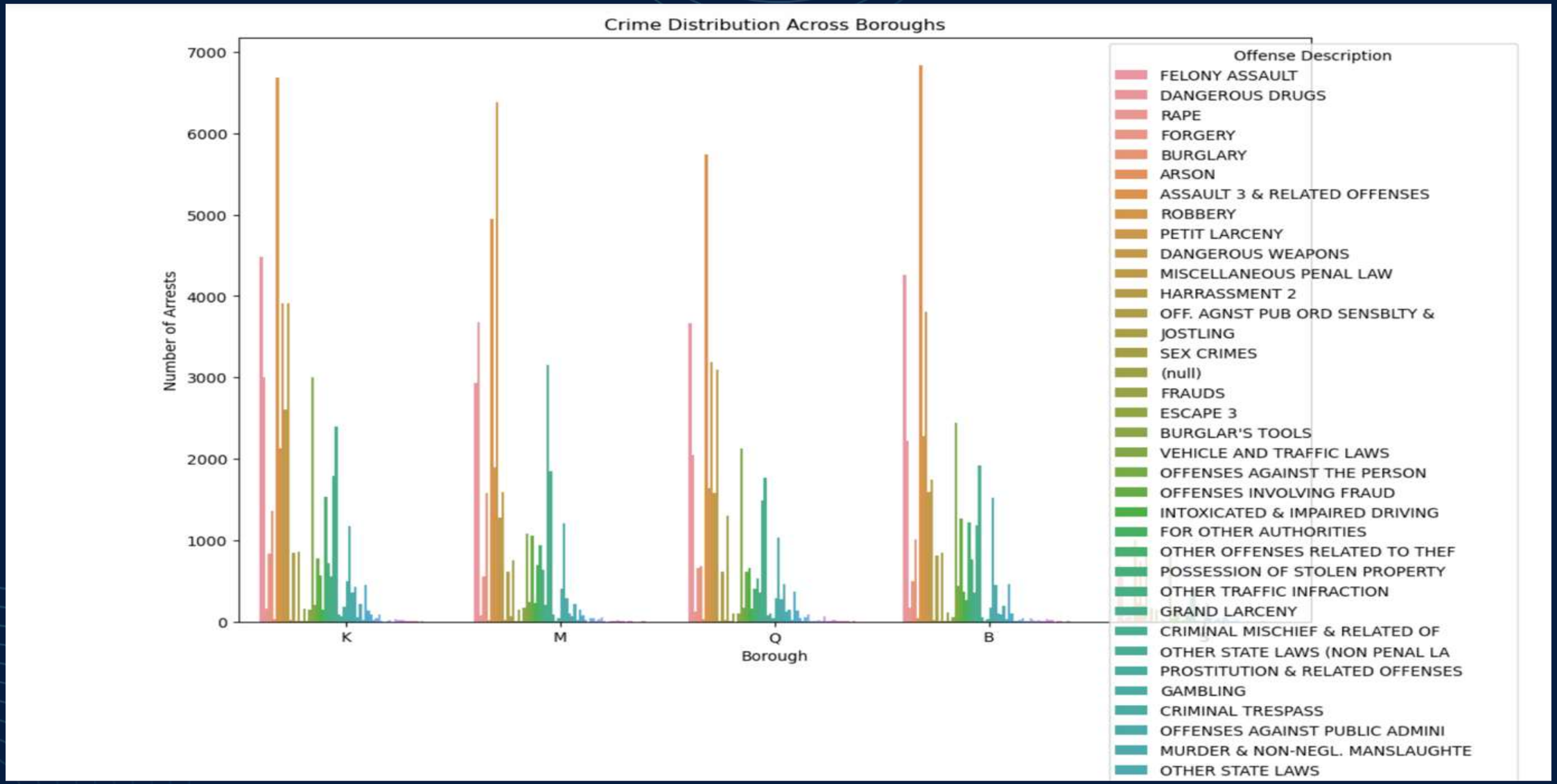
Visualization

In New York City, which specific neighborhoods or boroughs exhibit the highest arrest rates for various types of crimes, as identified through an analysis of NYPD Arrest Data?

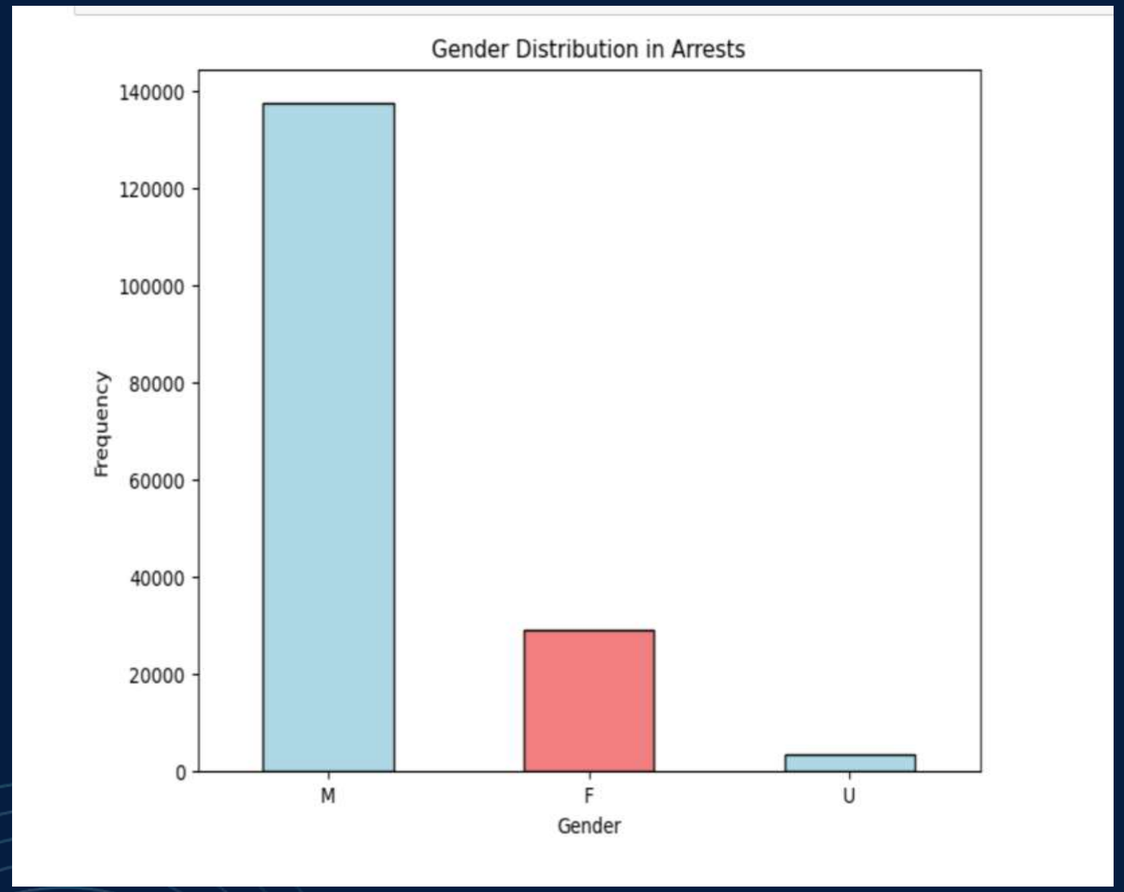
- The heat map visually represents the spatial distribution of arrests for different crimes across City neighborhoods. Intensity variations in color indicate the concentration of arrests, providing a nuanced understanding of crime hotspots.

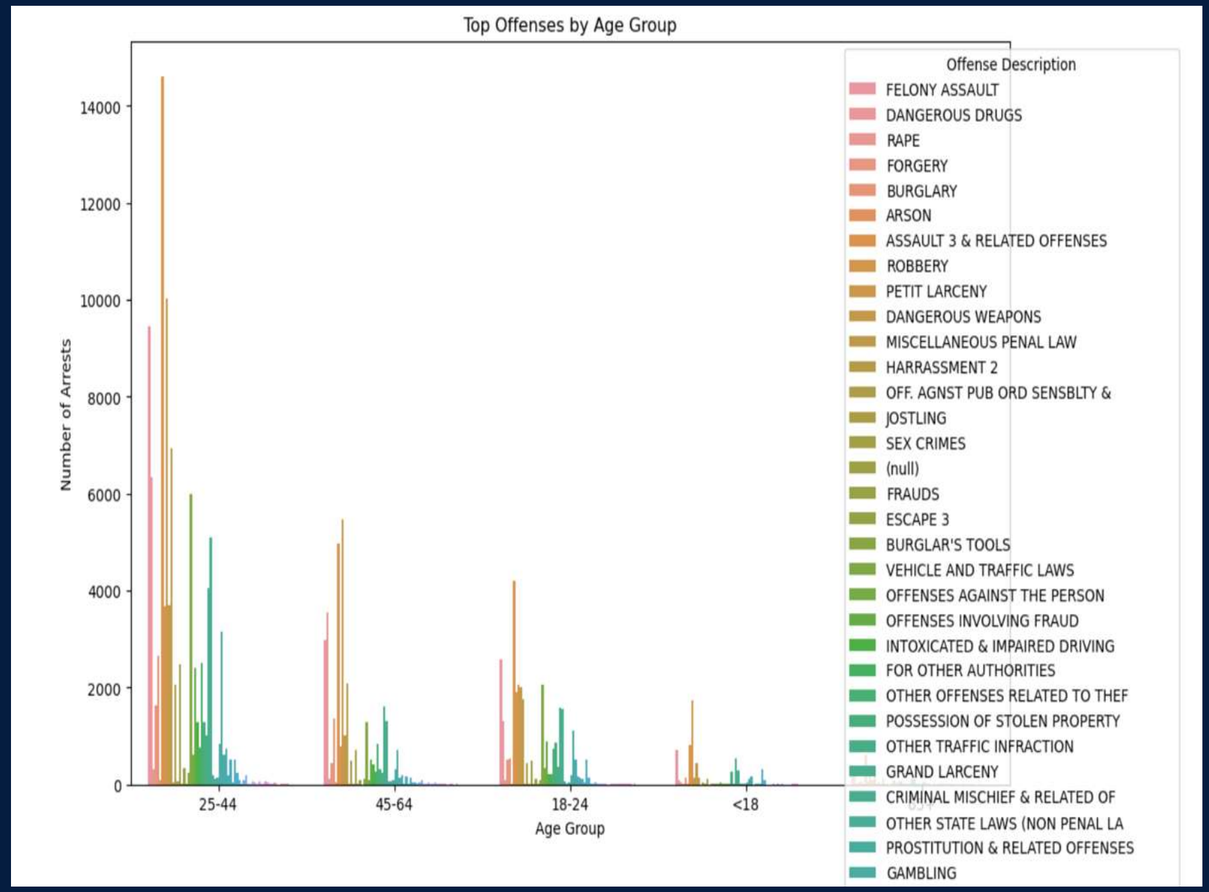


Visualization

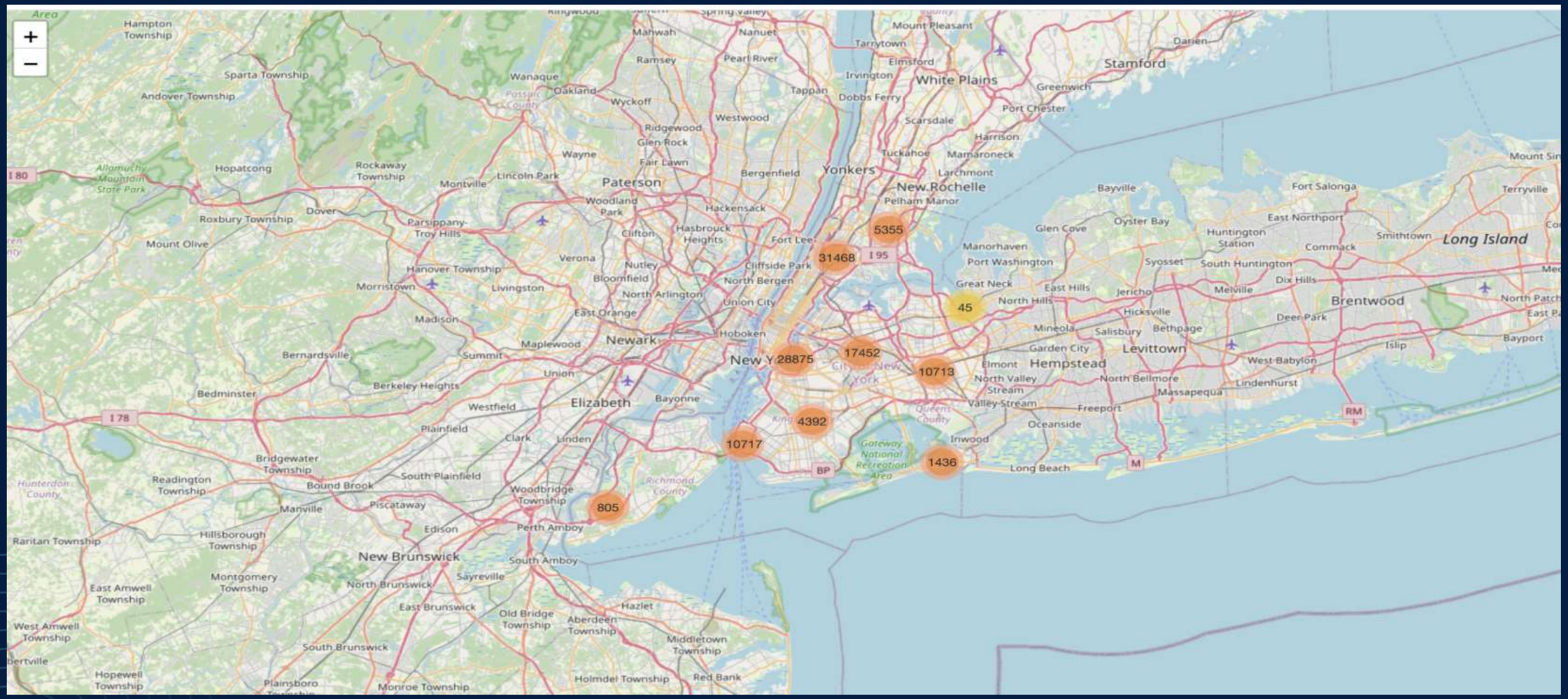


More Related Graphs





Geospatial Visualization of Areas with more crimes



Machine Learning Model Result

Gradient Boost Model

Accuracy (Gradient Boosting Classifier): 0.81

	precision	recall	f1-score	support
F	0.00	0.00	0.00	5775
M	0.81	1.00	0.89	27538
U	0.00	0.00	0.00	706
accuracy			0.81	34019
macro avg	0.27	0.33	0.30	34019
weighted avg	0.66	0.81	0.72	34019

Random Forest Model

Accuracy (Random Forest Classifier): 0.81

	precision	recall	f1-score	support
F	0.41	0.01	0.03	5775
M	0.81	0.99	0.89	27538
U	0.21	0.01	0.02	706
accuracy			0.81	34019
macro avg	0.47	0.34	0.31	34019
weighted avg	0.73	0.81	0.73	34019

Both the Random Forest and Gradient Boosting Classifier models have an overall accuracy of 0.81. However, they differ significantly in their performance across different classes

NYPD Crime Analytics

- May and August consistently exhibit higher arrest rates and February consistently records the lowest crime rates, suggesting a seasonal pattern influenced by warmer weather
- The Bronx and Manhattan show concentrated occurrences of sex crimes, Staten Island records the fewest instances of sex crimes, indicating a lower prevalence compared to other boroughs.
- The age group between (25-44) the highest contribution to overall crime rates. Highlights disparities in arrest rates among black and Hispanic groups.
- K, M, and Q boroughs have significantly higher crime rates than others.
- Significantly higher arrests for males (almost 140,000) compared to females (around 40,000).

Related Work

According to the article, by Carolyn Rebecca Block Statistical Analysis Centre, some crimes are more common during specific seasons or times of the day.

Bill Hutchinson says, there are significant differences in arrest rates for different racial and ethnic groups in the United States. According to the article you linked, Black Americans accounted for 52% of all arrests in New York City in 2020, despite only making up 24% of the city's population.

According to By Georgia Worrell and Tina Moore, Adults between 25 and 34 years old are arrested the most (32%), followed by those 18-24 (especially violent crimes).

Future Work

Advanced Predictive Analytics

Socio-Economic Landscape Study

Geospatial Analysis for Resource Optimization

Bias Mitigation in Predictive Models

Conclusion

- Our analysis of the NYPD arrest data provides valuable insights that could potentially help in reducing crime. Some of them include:
  - Identifying High-Risk Areas
  - Understanding Demographic Trends
  - Recognizing Common Offenses
  - Improving Predictive Policing
  - Identifying the season with highest number of crimes

- The analysis of NYPD Arrest Data reveals intricate patterns in temporal, spatial, demographic, and offense-related dimensions of arrests in New York City. The findings emphasize the need for targeted law enforcement strategies and policy interventions to address disparities and effectively allocate resources.

THANK YOU