NYPD SHOOTING INCIDENT DATA REPORT

Presented by: Ayush Jain

INTRODUCTION AND DATA OVERVIEW

This report analyzes NYPD shooting incidents from 2006 to 2024, covering 29,744 records with 21 variables including incident details, location, and demographics of suspects and victims. The dataset is reviewed quarterly and published by the NYPD, providing insights into gun violence patterns across New York City boroughs.

DATA CLEANING AND TRANSFORMATION

1

Removal of Unused Columns

Coordinate columns X_COORD_CD and Y_COORD_CD were removed to streamline the dataset.

2

Conversion of Key Variables to Factors

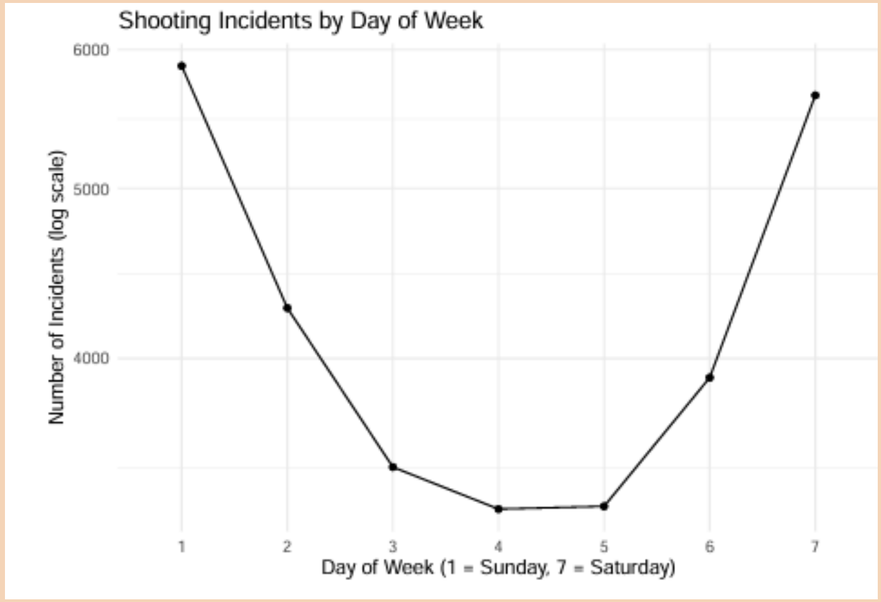
Categorical variables including borough, perpetrator and victim demographics, and location classification were converted to factor types.

3

Creation of New Temporal Variables

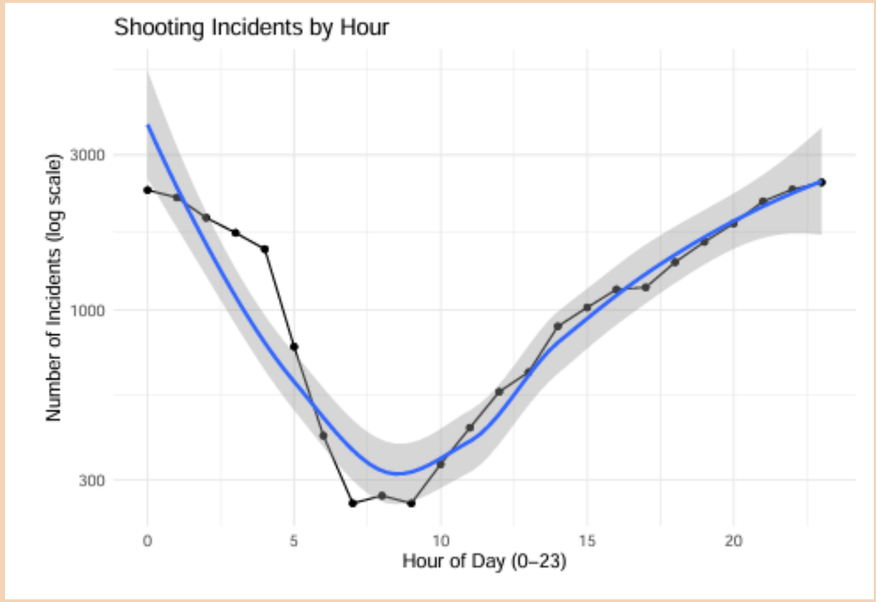
New variables such as OCCUR_HOUR, OCCUR_DATE, OCCUR_YEAR, and OCCUR_WEEKDAY were created to capture temporal aspects of incidents.

TEMPORAL PATTERNS OF SHOOTING INCIDENTS



Higher Frequency of Shootings on Weekends

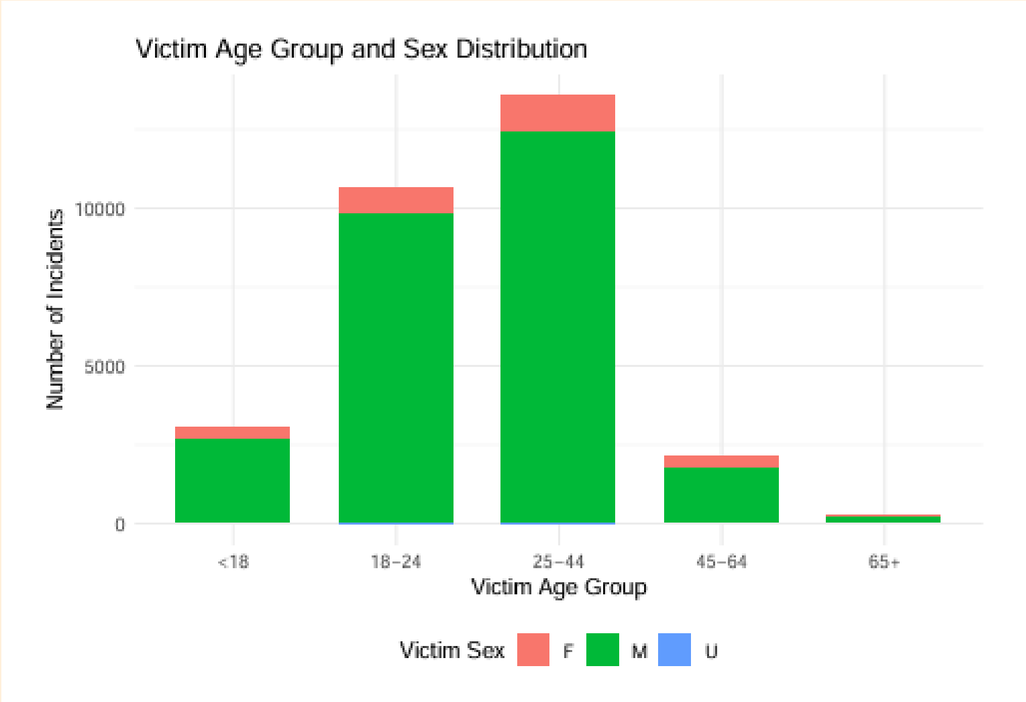
Shooting incidents occur more frequently on weekends, particularly on Saturday and Sunday, indicating increased risk during these days. Police patrols can be intensified accordingly.

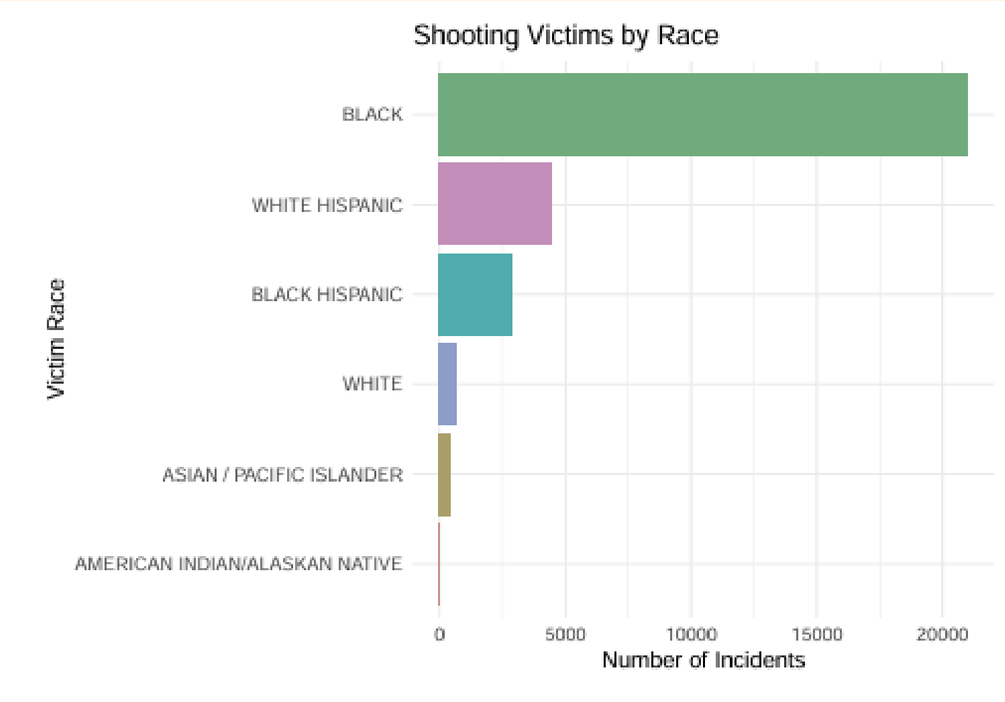


Peak Shooting Incidents During Nighttime

Shootings peak during nighttime hours especially from late evening to early morning (around 20:00 to 3:00). Enhanced patrolling during these hours is recommended.

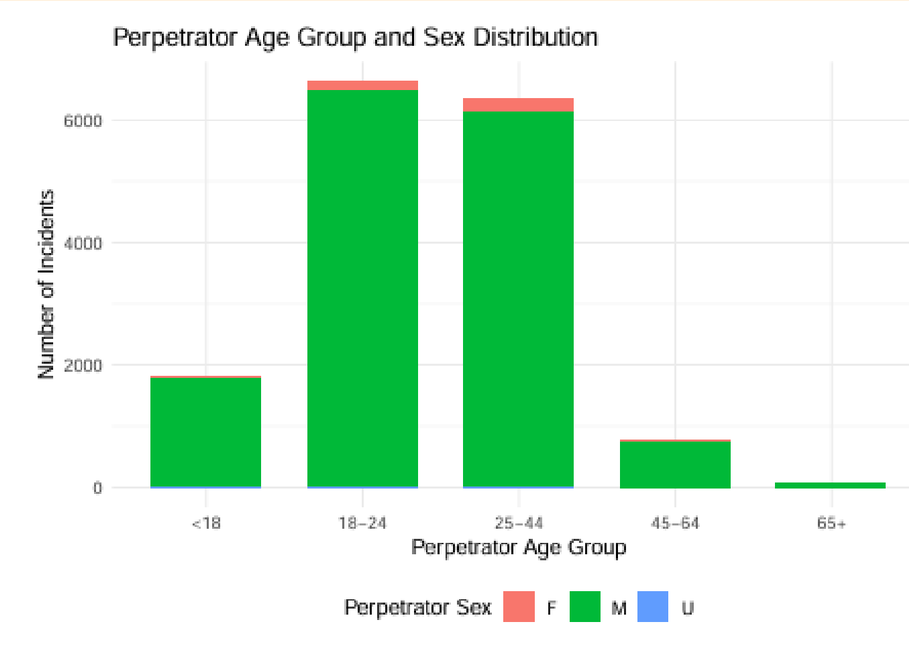
DEMOGRAPHIC ANALYSIS OF VICTIMS

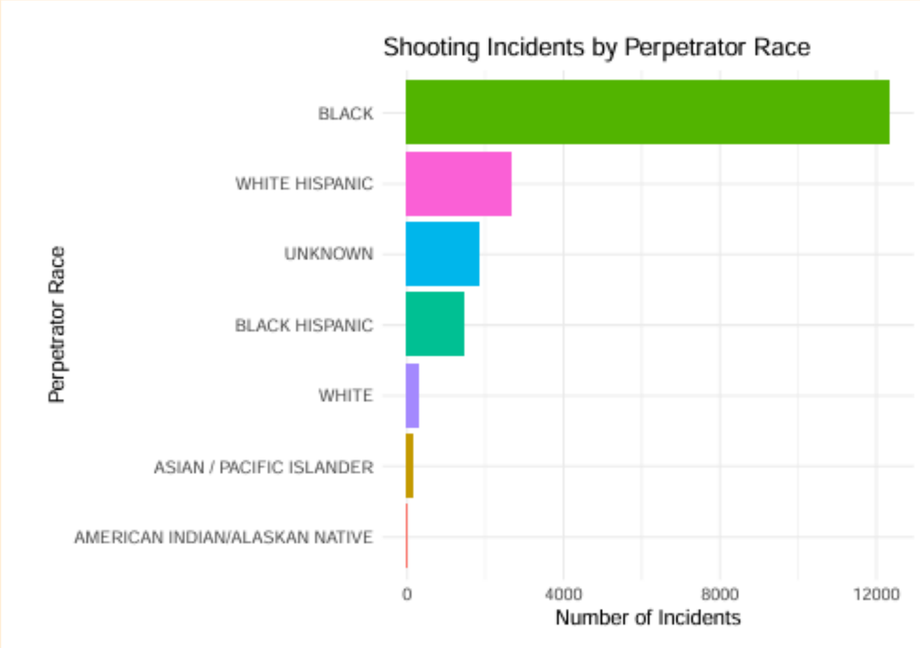




Males aged 25–44 are the most frequent victims, with Black individuals constituting the majority of victims.

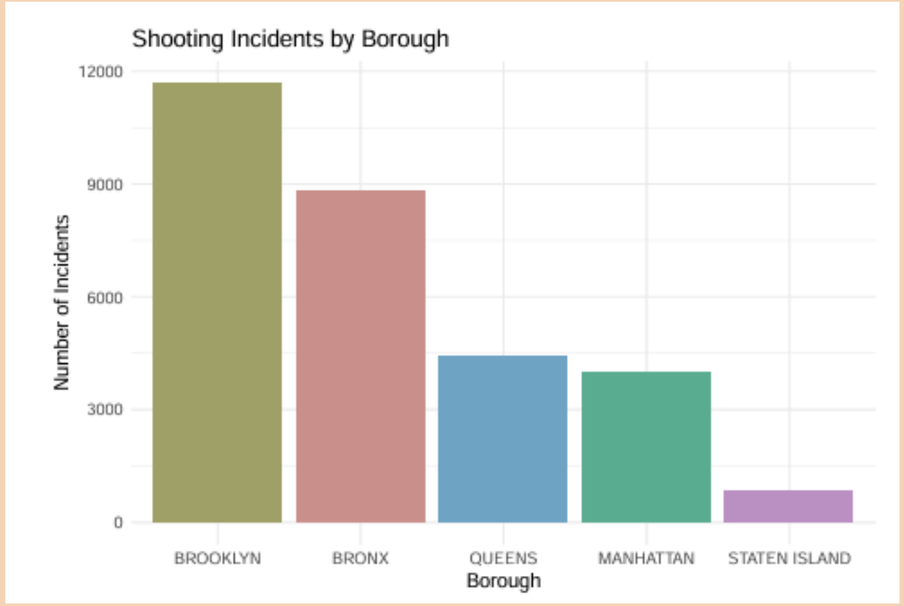
DEMOGRAPHIC ANALYSIS OF PERPETRATORS

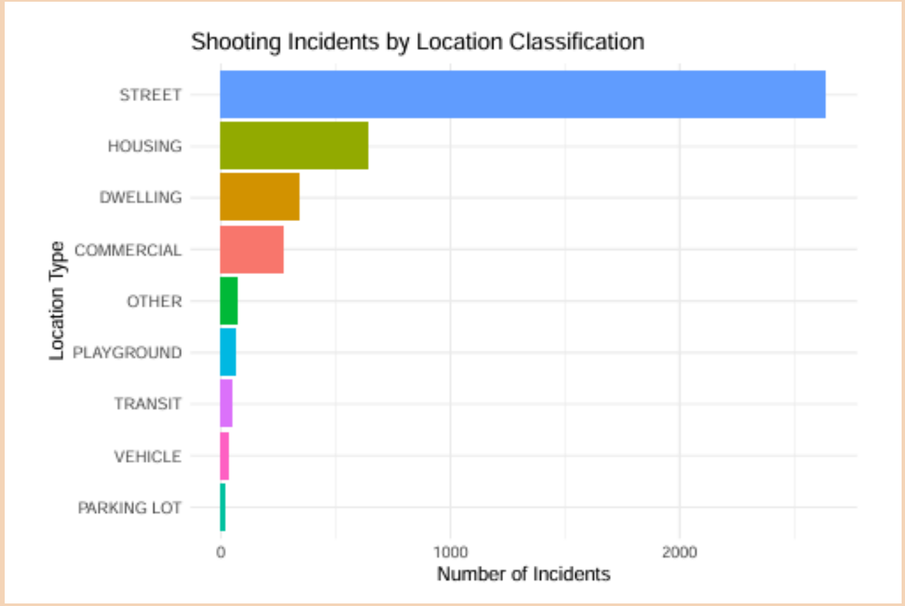




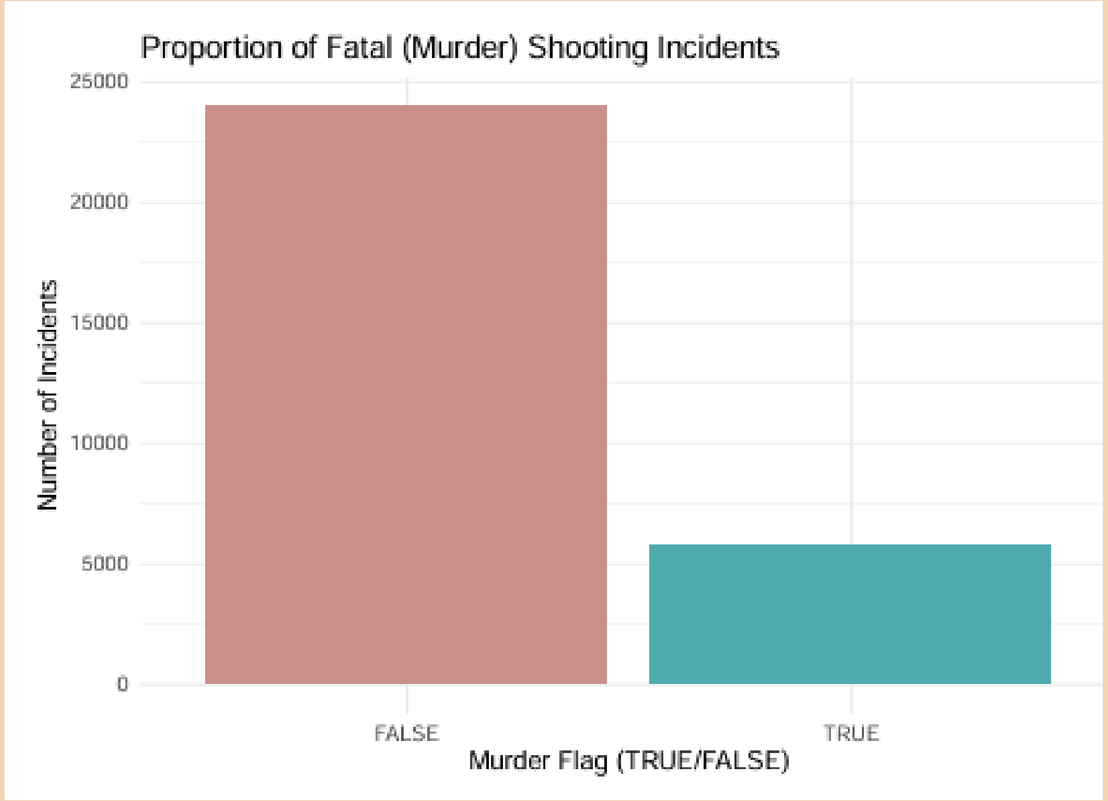
Males dominate as perpetrators, especially in the 18–24 age group, with Black individuals representing the largest racial group.

GEOGRAPHIC AND LOCATION-BASED INSIGHTS





MURDER RATES



PREDICTIVE MODELING OF MURDER OUTCOMES

Objective of Predictive Modeling

The goal is to classify shooting incidents as resulting in murder or not, using complete data records.

Logistic Regression and SVM Models

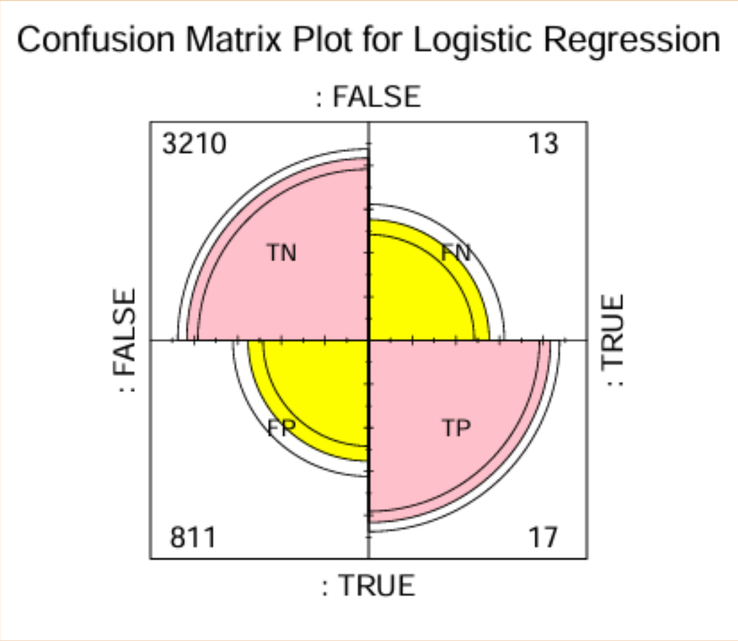
A logistic regression and SVM models predict murder outcomes based on borough, location type, time, and geographic coordinates.

Model Insights and Limitations

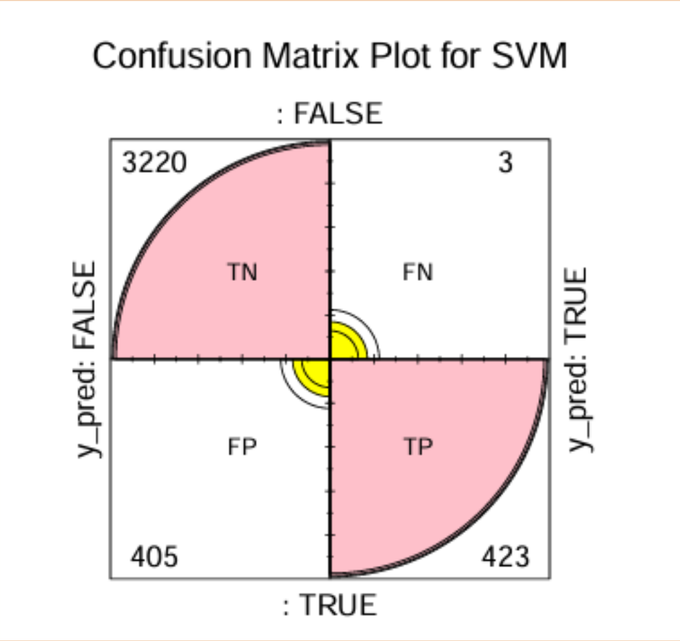
Many predictors show non-significant effects, indicating limited predictive power and the need for further model refinement.

CONFIDENCE MATRIX

LOGISTIC REGRESSION



SVM



MODEL COMPARISON

LOGISTIC REGRESSION

```
##
##          FALSE TRUE
##    FALSE  3210   13
##    TRUE    811   17
```

```
(3210+17)/nrow(model_data)
```

```
## [1] 0.7965934
```

SVM

```
##         y_pred
##          FALSE TRUE
##    FALSE  3220    3
##    TRUE    405  423
```

```
(3220+423)/nrow(model_data)
```

```
## [1] 0.8992841
```

CONCLUSIONS AND RECOMMENDATIONS

- Increase Police Presence on Weekends : Shooting incidents are more frequent on weekends, especially Saturday and Sunday, suggesting enhanced weekend patrolling could reduce violence.
- Focus on Young Males in Brooklyn : Young males aged 18-24 are the majority of perpetrators, with Brooklyn having the highest shooting incidents, indicating targeted interventions are needed.
- Strategic Policing Recommendations : Combining temporal and demographic data supports focused law enforcement efforts to mitigate gun violence effectively.

THANK YOU