# NYPD Shooting Incident Analysis

Cassandra Jones

2025-01-23

## Methodology

- Introduction
- Data Resource
- Data Visualization
- Predictive Modeling
- Potential Bias
- Conclusion

### Introduction

In this project, I analyzed a dataset containing shooting incidents that occurred in New York City during the year 2006. The goal was to explore the characteristics of these incidents and create a predictive model to determine whether an incident was a murder. Through my analysis, I examined various factors such as victim demographics, the safety of different precincts, and the distribution of incidents across different boroughs. Additionally, I developed a logistic regression model to predict whether a shooting incident was likely to be a statistical murder. The findings from this project aim to provide insights into patterns within violent incidents in New York City, with a particular focus on identifying potentially unsafe areas and understanding the distribution of incidents across different demographic groups.

### Data Resource

List of every shooting incident that occurred in NYC going back to 2006 through the end of the previous calendar year. This is a breakdown of every shooting incident that occurred in NYC…

Data Resource: https://catalog.data.gov/dataset/nypd-shooting-incident-data-historic

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

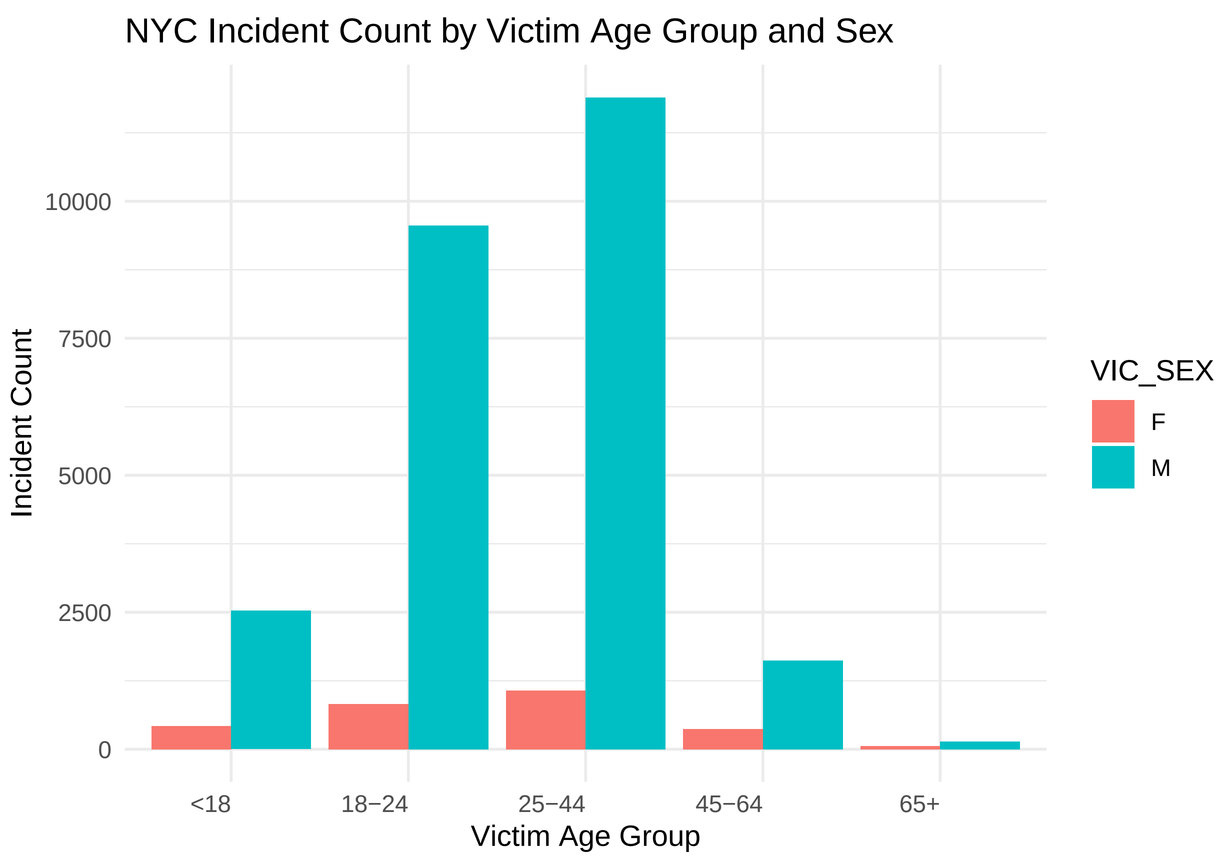
```r
df <- read.csv("NYPD_Shooting_Incident_Data__Historic_.csv")
summary(df)
```

```
##   INCIDENT_KEY        OCCUR_DATE         OCCUR_TIME            BORO          
##  Min.   :  9953245   Length:28562       Length:28562       Length:28562      
##  1st Qu.: 65439914   Class :character   Class :character   Class :character  
##  Median : 92711254   Mode  :character   Mode  :character   Mode  :character  
##  Mean   :127405824                                                           
##  3rd Qu.:203131993                                                           
##  Max.   :279758069                                                           
##                                                                              
##  LOC_OF_OCCUR_DESC     PRECINCT     JURISDICTION_CODE LOC_CLASSFCTN_DESC
##  Length:28562       Min.   :  1.0   Min.   :0.0000    Length:28562      
##  Class :character   1st Qu.: 44.0   1st Qu.:0.0000    Class :character  
##  Mode  :character   Median : 67.0   Median :0.0000    Mode  :character  
##                     Mean   : 65.5   Mean   :0.3219                      
##                     3rd Qu.: 81.0   3rd Qu.:0.0000                      
##                     Max.   :123.0   Max.   :2.0000                      
##                                     NA's   :2                           
##  LOCATION_DESC      STATISTICAL_MURDER_FLAG PERP_AGE_GROUP    
##  Length:28562       Length:28562            Length:28562      
##  Class :character   Class :character        Class :character  
##  Mode  :character   Mode  :character        Mode  :character  
##                                                               
##                                                               
##                                                               
##                                                               
##    PERP_SEX          PERP_RACE         VIC_AGE_GROUP        VIC_SEX         
##  Length:28562       Length:28562       Length:28562       Length:28562      
##  Class :character   Class :character   Class :character   Class :character  
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character  
##                                                                             
##                                                                             
##                                                                             
##                                                                             
##    VIC_RACE           X_COORD_CD        Y_COORD_CD        Latitude    
##  Length:28562       Min.   : 914928   Min.   :125757   Min.   :40.51  
##  Class :character   1st Qu.:1000068   1st Qu.:182912   1st Qu.:40.67  
##  Mode  :character   Median :1007772   Median :194901   Median :40.70  
##                     Mean   :1009424   Mean   :208380   Mean   :40.74  
##                     3rd Qu.:1016807   3rd Qu.:239814   3rd Qu.:40.82  
##                     Max.   :1066815   Max.   :271128   Max.   :40.91  
##                                                        NA's   :59     
##    Longitude         Lon_Lat         
##  Min.   :-74.25   Length:28562      
##  1st Qu.:-73.94   Class :character  
##  Median :-73.92   Mode  :character  
##  Mean   :-73.91                     
##  3rd Qu.:-73.88                     
##  Max.   :-73.70                     
##  NA's   :59
```

**Data Visualization**

Below is a bar plot showing incident counts by victim age group and sex. Interestingly, I found that male victims outnumbered female victims across all age groups, with the 18-24 and 25-44 age groups experiencing the highest number of incidents.

```
# summarise by age group and sex
age_group <- df%>%
  group_by(VIC_AGE_GROUP, VIC_SEX) %>%
  summarise(incident_count = n(), .groups = "drop")

# exclude unkown rows under both age group and sex columns
age_group <- age_group %>%
  filter(!(VIC_AGE_GROUP %in% c("1022", "UNKNOWN"))) %>%
  filter(!(VIC_SEX %in% c("U")))

ggplot(age_group, aes(x = VIC_AGE_GROUP, y = incident_count, fill = VIC_SEX)) +
  geom_col(position = "dodge") +  # Position = dodge creates clusters
  labs(title = "NYC Incident Count by Victim Age Group and Sex",
       x = "Victim Age Group",
       y = "Incident Count") +
  theme_minimal() +
  theme(axis.text.x = element_text(hjust = 1))
```




Below is a box plot to assess precinct safety. By counting incidents by borough and precinct, I calculated the mean number of incidents to be 370.93, with a standard deviation of 374.15. Using the safety threshold of the mean plus one standard deviation, I identified precincts above 745 incidents as potentially unsafe.

```r
library(ggplot2)
safety <- df %>%
  group_by(BORO, PRECINCT) %>%
  summarise(incident_count=n(), .groups="drop")

mean_incidents <- mean(safety$incident_count)
sd_incidents <- sd(safety$incident_count)
min_incident <- min(safety$incident_count)
max_incident <- max(safety$incident_count)
threshold <- mean_incidents + sd_incidents
cat("Incident Counts by BORO by Precinct", "\n")
```

```
## Incident Counts by BORO by Precinct
```

```r
cat("minimum: ", min_incident, "\n")
```

```
## minimum:  1
```

```r
cat("maximum: ", max_incident, "\n")
```

```
## maximum:  1628
```

```r
cat("mean: ", mean_incidents, "\n")
```

```
## mean:  370.9351
```

```r
cat("standard deviation: ", sd_incidents, "\n")
```

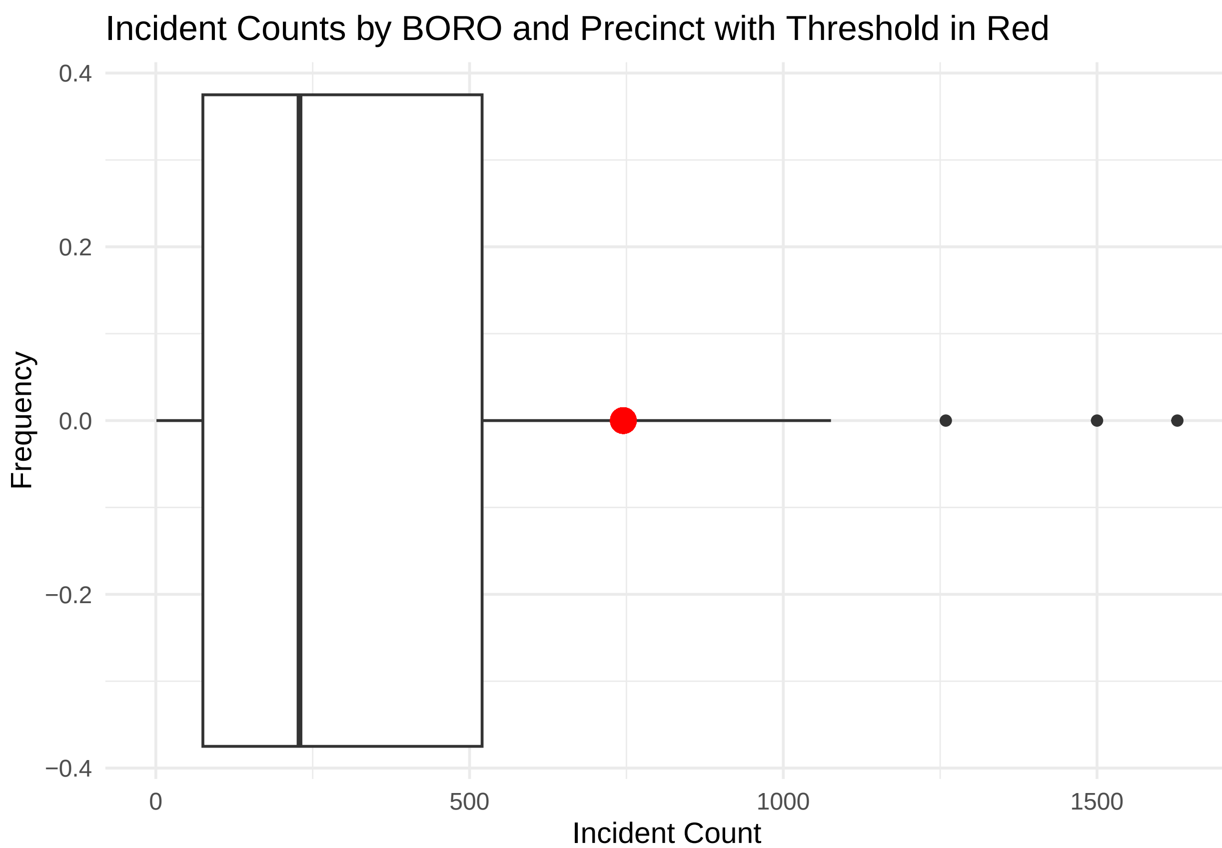
```
## standard deviation:  374.1505
```

```r
cat("Threshold: ", threshold, "\n")
```

```
## Threshold:  745.0855
```

```r
ggplot(safety, aes(x = incident_count)) +
  geom_boxplot() +
  # Add red dot at threshold
  geom_point(aes(x = threshold, y = 0), color ="red", size = 4) +
  labs(title = "Incident Counts by BORO and Precinct with Threshold in Red",
       x = "Incident Count",
       y = "Frequency") +
  theme_minimal()
```

```
## Warning in geom_point(aes(x = threshold, y = 0), color = "red", size = 4): All aesthetics have lengt
## i Please consider using `annotate()` or provide this layer with data containing
##   a single row.
```




### Predictive Modeling

Below is a logistic regression model to predict whether an incident was a murder, achieving a training score of 0.8063 and a test score of 0.8062.

```
library(lubridate)
# set date and time format
df$OCCUR_DATE <- mdy(df$OCCUR_DATE)
df$OCCUR_TIME <- hms(df$OCCUR_TIME)

# create day of week, month, and hour columns
df$DAY_OF_WEEK <- wday(df$OCCUR_DATE)
df$MONTH <- month(df$OCCUR_DATE)
df$HOUR <- hour(df$OCCUR_TIME)

# select useful columns and delete rows with NAs for modeling
df_model <- df %>% select(BORO, PRECINCT, STATISTICAL_MURDER_FLAG,
                          VIC_AGE_GROUP, VIC_RACE, VIC_SEX, Latitude, Longitude,
                          MONTH, HOUR, DAY_OF_WEEK)
df_model <- na.omit(df_model)
df_model$PRECINCT <- as.factor(df_model$PRECINCT)
df_model$STATISTICAL_MURDER_FLAG[df_model$STATISTICAL_MURDER_FLAG == "true"] <- 1
df_model$STATISTICAL_MURDER_FLAG[df_model$STATISTICAL_MURDER_FLAG == "false"] <- 0
df_model$STATISTICAL_MURDER_FLAG <- as.integer(df_model$STATISTICAL_MURDER_FLAG)
head(df_model)
```

```
##        BORO PRECINCT STATISTICAL_MURDER_FLAG VIC_AGE_GROUP VIC_RACE VIC_SEX
## 1 MANHATTAN       14                       1         25-44    BLACK       M
## 2     BRONX       48                       1         18-24    BLACK       M
## 3    QUEENS      103                       0         18-24    BLACK       M
## 4     BRONX       42                       0         25-44    BLACK       M
## 5  BROOKLYN       83                       0         25-44    BLACK       M
## 6 MANHATTAN       23                       0         25-44    BLACK       M
##   Latitude Longitude MONTH HOUR DAY_OF_WEEK
## 1 40.75469 -73.99350     5    0           5
## 2 40.85440 -73.88233     7   22           2
## 3 40.71063 -73.76777     5   19           1
## 4 40.83242 -73.89071     9   21           3
## 5 40.68844 -73.91022     2   21           1
## 6 40.79773 -73.94651     7   23           5
```

```r
library(caTools)
```

```
## Warning: package 'caTools' was built under R version 4.4.2
```

```r
# Set a seed for reproducibility
set.seed(123)

# Split the data: 70% for training, 30% for testing
split <- sample.split(df_model$STATISTICAL_MURDER_FLAG, SplitRatio = 0.7)

# Create training and testing datasets
train_data <- subset(df_model, split == TRUE)
test_data <- subset(df_model, split == FALSE)

# Set X_train, X_test (features) and y_train, y_test (target variable)
X_train <- train_data[, c("BORO", "PRECINCT", "VIC_AGE_GROUP", "VIC_RACE", "VIC_SEX", "Latitude", "Long
y_train <- train_data$STATISTICAL_MURDER_FLAG

X_test <- test_data[, c("BORO", "PRECINCT", "VIC_AGE_GROUP", "VIC_RACE", "VIC_SEX", "Latitude", "Longitu
y_test <- test_data$STATISTICAL_MURDER_FLAG
```

```r
# Create dummy variables for categorical features
X_train_dummies <- model.matrix(~ BORO + PRECINCT + VIC_AGE_GROUP + VIC_RACE + VIC_SEX - 1, data = X_tra
X_test_dummies <- model.matrix(~ BORO + PRECINCT + VIC_AGE_GROUP + VIC_RACE + VIC_SEX - 1, data = X_tes


# Combine numeric features with dummy variables for both train and test datasets
X_train_final <- cbind(X_train_dummies, X_train[, c("Latitude", "Longitude", "MONTH", "HOUR", "DAY_OF_WE
X_test_final <- cbind(X_test_dummies, X_test[, c("Latitude", "Longitude", "MONTH", "HOUR", "DAY_OF_WEEK

# View the final data
head(X_train_final)
```

```
##   BOROBRONX BOROBROOKLYN BOROMANHATTAN BOROQUEENS BOROSTATEN ISLAND PRECINCT5
## 1         0            0             1          0                 0         0
## 3         0            0             0          1                 0         0
## 5         0            1             0          0                 0         0
```

```
## 7         0            0             0          1                 0         0
## 8         0            1             0          0                 0         0
## 9         1            0             0          0                 0         0
##   PRECINCT6 PRECINCT7 PRECINCT9 PRECINCT10 PRECINCT13 PRECINCT14 PRECINCT17
## 1         0         0         0          0          0          1          0
## 3         0         0         0          0          0          0          0
## 5         0         0         0          0          0          0          0
## 7         0         0         0          0          0          0          0
## 8         0         0         0          0          0          0          0
## 9         0         0         0          0          0          0          0
##   PRECINCT18 PRECINCT19 PRECINCT20 PRECINCT22 PRECINCT23 PRECINCT24 PRECINCT25
## 1          0          0          0          0          0          0          0
## 3          0          0          0          0          0          0          0
## 5          0          0          0          0          0          0          0
## 7          0          0          0          0          0          0          0
## 8          0          0          0          0          0          0          0
## 9          0          0          0          0          0          0          0
##   PRECINCT26 PRECINCT28 PRECINCT30 PRECINCT32 PRECINCT33 PRECINCT34 PRECINCT40
## 1          0          0          0          0          0          0          0
## 3          0          0          0          0          0          0          0
## 5          0          0          0          0          0          0          0
## 7          0          0          0          0          0          0          0
## 8          0          0          0          0          0          0          0
## 9          0          0          0          0          0          0          0
##   PRECINCT41 PRECINCT42 PRECINCT43 PRECINCT44 PRECINCT45 PRECINCT46 PRECINCT47
## 1          0          0          0          0          0          0          0
## 3          0          0          0          0          0          0          0
## 5          0          0          0          0          0          0          0
## 7          0          0          0          0          0          0          0
## 8          0          0          0          0          0          0          0
## 9          0          0          0          0          0          0          0
##   PRECINCT48 PRECINCT49 PRECINCT50 PRECINCT52 PRECINCT60 PRECINCT61 PRECINCT62
## 1          0          0          0          0          0          0          0
## 3          0          0          0          0          0          0          0
## 5          0          0          0          0          0          0          0
## 7          0          0          0          0          0          0          0
## 8          0          0          0          0          0          0          0
## 9          1          0          0          0          0          0          0
##   PRECINCT63 PRECINCT66 PRECINCT67 PRECINCT68 PRECINCT69 PRECINCT70 PRECINCT71
## 1          0          0          0          0          0          0          0
## 3          0          0          0          0          0          0          0
## 5          0          0          0          0          0          0          0
## 7          0          0          0          0          0          0          0
## 8          0          0          0          0          0          0          0
## 9          0          0          0          0          0          0          0
##   PRECINCT72 PRECINCT73 PRECINCT75 PRECINCT76 PRECINCT77 PRECINCT78 PRECINCT79
## 1          0          0          0          0          0          0          0
## 3          0          0          0          0          0          0          0
## 5          0          0          0          0          0          0          0
## 7          0          0          0          0          0          0          0
## 8          0          0          0          0          1          0          0
## 9          0          0          0          0          0          0          0
##   PRECINCT81 PRECINCT83 PRECINCT84 PRECINCT88 PRECINCT90 PRECINCT94 PRECINCT100
## 1          0          0          0          0          0          0           0
```

```
## 3          0          0          0          0          0          0           0
## 5          0          1          0          0          0          0           0
## 7          0          0          0          0          0          0           0
## 8          0          0          0          0          0          0           0
## 9          0          0          0          0          0          0           0
##   PRECINCT101 PRECINCT102 PRECINCT103 PRECINCT104 PRECINCT105 PRECINCT106
## 1           0           0           0           0           0           0
## 3           0           0           1           0           0           0
## 5           0           0           0           0           0           0
## 7           0           0           0           0           0           0
## 8           0           0           0           0           0           0
## 9           0           0           0           0           0           0
##   PRECINCT107 PRECINCT108 PRECINCT109 PRECINCT110 PRECINCT111 PRECINCT112
## 1           0           0           0           0           0           0
## 3           0           0           0           0           0           0
## 5           0           0           0           0           0           0
## 7           0           0           0           0           0           0
## 8           0           0           0           0           0           0
## 9           0           0           0           0           0           0
##   PRECINCT113 PRECINCT114 PRECINCT115 PRECINCT120 PRECINCT121 PRECINCT122
## 1           0           0           0           0           0           0
## 3           0           0           0           0           0           0
## 5           0           0           0           0           0           0
## 7           1           0           0           0           0           0
## 8           0           0           0           0           0           0
## 9           0           0           0           0           0           0
##   PRECINCT123 VIC_AGE_GROUP18-24 VIC_AGE_GROUP25-44 VIC_AGE_GROUP45-64
## 1           0                  0                  1                  0
## 3           0                  1                  0                  0
## 5           0                  0                  1                  0
## 7           0                  0                  0                  1
## 8           0                  0                  1                  0
## 9           0                  1                  0                  0
##   VIC_AGE_GROUP65+ VIC_AGE_GROUPUNKNOWN VIC_RACEASIAN / PACIFIC ISLANDER
## 1                0                    0                                0
## 3                0                    0                                0
## 5                0                    0                                0
## 7                0                    0                                0
## 8                0                    0                                0
## 9                0                    0                                0
##   VIC_RACEBLACK VIC_RACEBLACK HISPANIC VIC_RACEUNKNOWN VIC_RACEWHITE
## 1             1                      0               0             0
## 3             1                      0               0             0
## 5             1                      0               0             0
## 7             1                      0               0             0
## 8             1                      0               0             0
## 9             1                      0               0             0
##   VIC_RACEWHITE HISPANIC VIC_SEXM VIC_SEXU Latitude Longitude MONTH HOUR
## 1                      0        1        0 40.75469 -73.99350     5    0
## 3                      0        1        0 40.71063 -73.76777     5   19
## 5                      0        1        0 40.68844 -73.91022     2   21
## 7                      0        1        0 40.67331 -73.78989     6   19
## 8                      0        1        0 40.66858 -73.92698     7    1
## 9                      0        1        0 40.85151 -73.88382     5   18
```

```
##   DAY_OF_WEEK
## 1           5
## 3           1
## 5           1
## 7           2
## 8           5
## 9           7
```

```r
# Combine target and predictors into a single dataframe for modeling
train_data <- data.frame(X_train_final, STATISTICAL_MURDER_FLAG = y_train)

# Fit the logistic regression model
model <- glm(STATISTICAL_MURDER_FLAG ~ ., data = train_data, family = binomial)

# View model summary
summary(model)
```

```
##
## Call:
## glm(formula = STATISTICAL_MURDER_FLAG ~ ., family = binomial,
##     data = train_data)
##
## Coefficients: (5 not defined because of singularities)
##                                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)                        9.252e+01  2.644e+02   0.350 0.726433
## BOROBRONX                         -1.791e-01  9.960e-01  -0.180 0.857330
## BOROBROOKLYN                      -6.069e-01  7.964e-01  -0.762 0.446017
## BOROMANHATTAN                      4.316e-01  8.372e-01   0.515 0.606225
## BOROQUEENS                        -2.734e-01  8.671e-01  -0.315 0.752530
## BOROSTATEN.ISLAND                         NA         NA      NA       NA
## PRECINCT5                         -3.323e-01  5.940e-01  -0.559 0.575853
## PRECINCT6                         -5.782e-02  7.178e-01  -0.081 0.935792
## PRECINCT7                         -1.205e+00  6.185e-01  -1.948 0.051393 .
## PRECINCT9                         -6.961e-01  5.746e-01  -1.212 0.225685
## PRECINCT10                        -4.038e-01  6.095e-01  -0.662 0.507685
## PRECINCT13                        -6.127e-01  6.344e-01  -0.966 0.334198
## PRECINCT14                        -4.316e-01  6.294e-01  -0.686 0.492922
## PRECINCT17                        -1.063e+00  1.213e+00  -0.876 0.381102
## PRECINCT18                        -1.006e+00  7.131e-01  -1.411 0.158237
## PRECINCT19                        -4.418e-01  7.853e-01  -0.563 0.573727
## PRECINCT20                        -9.511e-01  7.232e-01  -1.315 0.188473
## PRECINCT22                        -1.286e+01  5.354e+02  -0.024 0.980837
## PRECINCT23                        -8.925e-01  5.601e-01  -1.594 0.111036
## PRECINCT24                        -5.882e-01  6.009e-01  -0.979 0.327572
## PRECINCT25                        -8.150e-01  5.684e-01  -1.434 0.151650
## PRECINCT26                        -1.690e+00  6.459e-01  -2.616 0.008902 **
## PRECINCT28                        -7.660e-01  5.729e-01  -1.337 0.181172
## PRECINCT30                        -5.630e-01  6.014e-01  -0.936 0.349184
## PRECINCT32                        -7.576e-01  5.750e-01  -1.318 0.187660
## PRECINCT33                        -7.766e-01  6.172e-01  -1.258 0.208348
## PRECINCT34                        -8.697e-01  6.379e-01  -1.363 0.172783
## PRECINCT40                        -3.013e-01  2.089e-01  -1.443 0.149101
## PRECINCT41                        -1.479e-01  2.172e-01  -0.681 0.495870
## PRECINCT42                        -1.259e-01  1.826e-01  -0.690 0.490338
```

```
## PRECINCT43                          -2.716e-01  1.990e-01  -1.365 0.172296
## PRECINCT44                          -1.467e-01  1.771e-01  -0.829 0.407315
## PRECINCT45                          -2.966e-01  2.784e-01  -1.065 0.286682
## PRECINCT46                          -8.946e-03  1.603e-01  -0.056 0.955504
## PRECINCT47                          -2.078e-01  1.720e-01  -1.208 0.227019
## PRECINCT48                          -1.385e-01  1.675e-01  -0.827 0.408365
## PRECINCT49                          -3.534e-05  2.001e-01   0.000 0.999859
## PRECINCT50                           1.002e-01  2.586e-01   0.387 0.698510
## PRECINCT52                                  NA         NA      NA       NA
## PRECINCT60                           3.017e-01  4.961e-01   0.608 0.543104
## PRECINCT61                           6.529e-01  4.947e-01   1.320 0.186877
## PRECINCT62                           6.967e-01  5.344e-01   1.304 0.192346
## PRECINCT63                           2.327e-01  4.405e-01   0.528 0.597251
## PRECINCT66                           2.175e-01  5.545e-01   0.392 0.694834
## PRECINCT67                           3.699e-01  3.840e-01   0.963 0.335520
## PRECINCT68                           1.465e-01  6.522e-01   0.225 0.822219
## PRECINCT69                           3.386e-01  4.202e-01   0.806 0.420326
## PRECINCT70                           2.432e-01  4.054e-01   0.600 0.548520
## PRECINCT71                           1.697e-02  3.839e-01   0.044 0.964743
## PRECINCT72                           3.556e-01  4.720e-01   0.753 0.451182
## PRECINCT73                           1.795e-01  3.730e-01   0.481 0.630329
## PRECINCT75                           2.649e-01  3.847e-01   0.689 0.491034
## PRECINCT76                           3.143e-01  4.203e-01   0.748 0.454623
## PRECINCT77                           4.242e-01  3.660e-01   1.159 0.246438
## PRECINCT78                          -3.460e-01  5.940e-01  -0.583 0.560219
## PRECINCT79                           1.994e-01  3.541e-01   0.563 0.573331
## PRECINCT81                           2.428e-01  3.613e-01   0.672 0.501607
## PRECINCT83                           1.856e-02  3.715e-01   0.050 0.960155
## PRECINCT84                           4.010e-02  4.436e-01   0.090 0.927984
## PRECINCT88                           2.855e-01  3.855e-01   0.741 0.458900
## PRECINCT90                           2.445e-01  3.731e-01   0.655 0.512312
## PRECINCT94                                  NA         NA      NA       NA
## PRECINCT100                         -1.150e+00  5.812e-01  -1.978 0.047886 *
## PRECINCT101                         -5.608e-01  5.137e-01  -1.092 0.274980
## PRECINCT102                         -9.121e-02  3.355e-01  -0.272 0.785717
## PRECINCT103                         -3.677e-01  3.260e-01  -1.128 0.259337
## PRECINCT104                         -5.108e-01  4.111e-01  -1.243 0.214052
## PRECINCT105                         -6.810e-02  3.897e-01  -0.175 0.861303
## PRECINCT106                          2.384e-01  3.504e-01   0.680 0.496263
## PRECINCT107                         -9.565e-03  3.664e-01  -0.026 0.979173
## PRECINCT108                         -9.591e-01  5.263e-01  -1.822 0.068393 .
## PRECINCT109                          2.683e-01  3.342e-01   0.803 0.422177
## PRECINCT110                          2.619e-01  2.985e-01   0.877 0.380323
## PRECINCT111                         -1.252e+01  2.381e+02  -0.053 0.958047
## PRECINCT112                          1.930e-01  6.516e-01   0.296 0.767029
## PRECINCT113                         -2.680e-01  3.531e-01  -0.759 0.447786
## PRECINCT114                         -2.114e-01  2.832e-01  -0.746 0.455492
## PRECINCT115                                 NA         NA      NA       NA
## PRECINCT120                         -1.916e-01  5.315e-01  -0.360 0.718495
## PRECINCT121                          2.824e-02  5.706e-01   0.050 0.960520
## PRECINCT122                          2.136e-01  6.025e-01   0.355 0.722893
## PRECINCT123                                 NA         NA      NA       NA
## VIC_AGE_GROUP18.24                   2.689e-01  7.292e-02   3.687 0.000227 ***
## VIC_AGE_GROUP25.44                   6.334e-01  7.043e-02   8.993  < 2e-16 ***
```

```
## VIC_AGE_GROUP45.64                8.233e-01  9.033e-02   9.114  < 2e-16 ***
## VIC_AGE_GROUP65.                  1.116e+00  1.867e-01   5.974 2.31e-09 ***
## VIC_AGE_GROUPUNKNOWN              8.520e-01  3.639e-01   2.341 0.019218 *
## VIC_RACEASIAN...PACIFIC.ISLANDER  1.210e+01  1.996e+02   0.061 0.951663
## VIC_RACEBLACK                     1.206e+01  1.996e+02   0.060 0.951836
## VIC_RACEBLACK.HISPANIC            1.192e+01  1.996e+02   0.060 0.952397
## VIC_RACEUNKNOWN                   1.156e+01  1.996e+02   0.058 0.953823
## VIC_RACEWHITE                     1.236e+01  1.996e+02   0.062 0.950654
## VIC_RACEWHITE.HISPANIC            1.221e+01  1.996e+02   0.061 0.951228
## VIC_SEXM                         -2.258e-02  6.134e-02  -0.368 0.712778
## VIC_SEXU                         -1.193e+01  1.771e+02  -0.067 0.946260
## Latitude                         -3.012e-01  2.374e+00  -0.127 0.899031
## Longitude                         1.269e+00  1.756e+00   0.723 0.469701
## MONTH                            -3.163e-03  5.778e-03  -0.547 0.584086
## HOUR                              2.287e-03  2.161e-03   1.059 0.289796
## DAY_OF_WEEK                      -7.397e-03  8.299e-03  -0.891 0.372768
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 19615  on 19951  degrees of freedom
## Residual deviance: 19281  on 19857  degrees of freedom
## AIC: 19471
##
## Number of Fisher Scoring iterations: 12
```

```r
# Predict on the test set
y_pred <- predict(model, newdata = data.frame(X_test_final), type = "response")

# Convert probabilities to binary outcome (0 or 1)
y_pred_class <- ifelse(y_pred > 0.5, 1, 0)

# Evaluate model performance (confusion matrix)
table(y_test, y_pred_class)
```

```
##       y_pred_class
## y_test    0
##      0 6894
##      1 1657
```

Check accuracy below

```r
# Predict on the training set
train_pred_prob <- predict(model, newdata = data.frame(X_train_final), type = "response")

# Convert predicted probabilities to binary outcomes (0 or 1)
train_pred_class <- ifelse(train_pred_prob > 0.5, 1, 0)

# Calculate accuracy on the training set
train_accuracy <- mean(train_pred_class == y_train)
train_accuracy
```

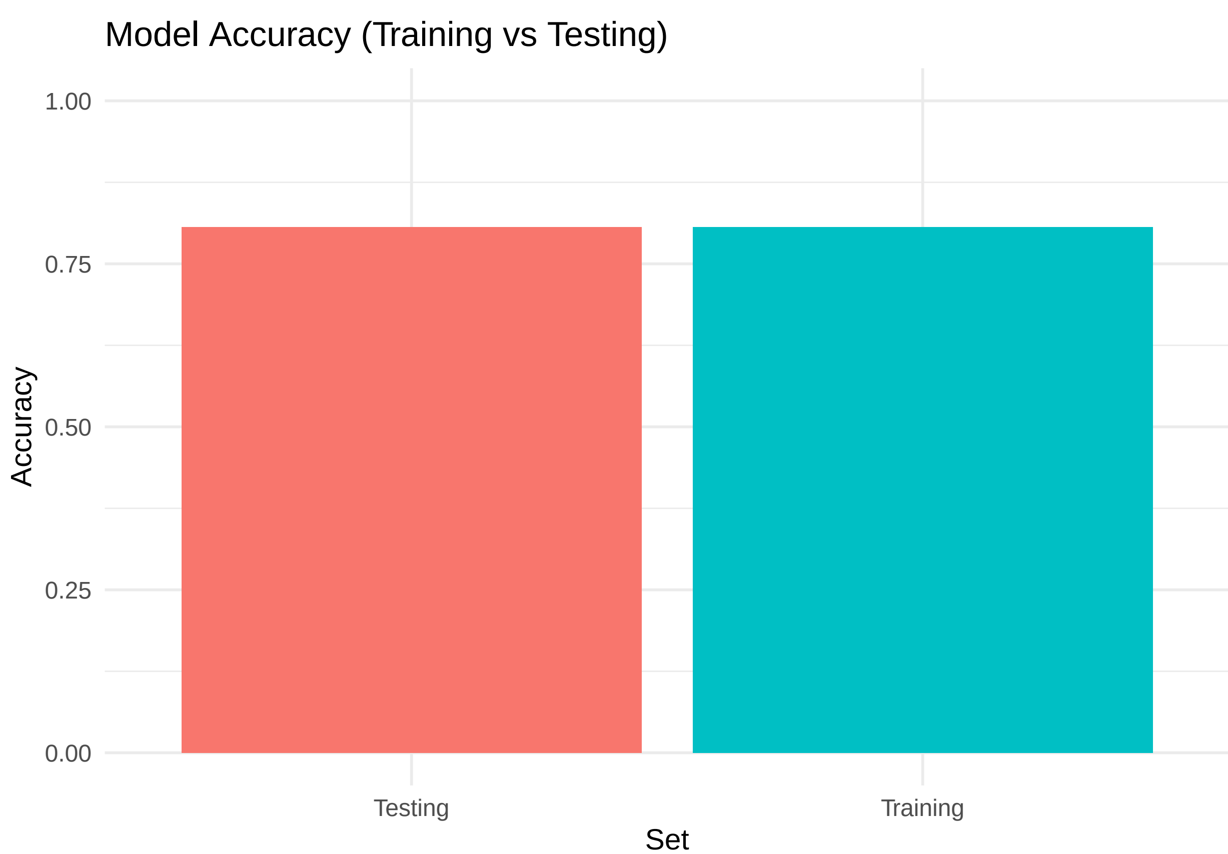
```
## [1] 0.8062851
```

```r
# Predict on the test set
test_pred_prob <- predict(model, newdata = data.frame(X_test_final), type = "response")

# Convert predicted probabilities to binary outcomes (0 or 1)
test_pred_class <- ifelse(test_pred_prob > 0.5, 1, 0)

# Calculate accuracy on the test set
test_accuracy <- mean(test_pred_class == y_test)
test_accuracy
```

```
## [1] 0.8062215
```

Below is to visualize model accuracy to compare the train and test scores.

```r
# Plot training vs test accuracy
accuracy_data <- data.frame(
  Set = c("Training", "Testing"),
  Accuracy = c(train_accuracy, test_accuracy)
)

library(ggplot2)
ggplot(accuracy_data, aes(x = Set, y = Accuracy, fill = Set)) +
  geom_bar(stat = "identity", show.legend = FALSE) +
  ggtitle("Model Accuracy (Training vs Testing)") +
  ylim(0, 1) +
  theme_minimal()
```



### Potential Bias

a key limitation of my analysis is the potential bias in the precinct-level analysis. Since I only considered incident counts without factoring in the population size of each precinct, the results may be skewed.

### Conclusion

To summarize, the analysis of the 2006 shooting incident data revealed interesting patterns, such as the predominance of male victims in all age groups and the higher frequency of incidents among individuals aged 18-44. My assessment of precinct safety, based on incident counts, identified certain areas with higher levels of violence. The predictive model for classifying incidents as murders showed a reasonably strong performance, with an accuracy of approximately 80.6%. However, it's important to recognize the potential bias in the analysis due to the lack of population data for each precinct, which may skew the results. Overall, while this project provides valuable insights, further improvements in model accuracy and a more nuanced understanding of precinct-level safety would require additional data and analysis.