Analysis of NYPD Shooting Incident Data: Predicting Fatal Outcomes Based on Victim Demographics

Introduction:

The objective of this project is to explore the NYPD Shooting Incident Data in a bid to gauge whether some of the demographic attributes of victims, including age, sex, and race, will dictate whether or not a shooting incident results in a fatality. This is within a larger project to look at what factors determine the outcomes of shooting incidents in New York City.

Research Question

The guiding question for this analysis is: Can a victim's demographics (race, sex, or age) predetermine whether the shooting will be fatal?

Dataset Overview

The dataset used for conducting this analysis is the NYPD Shooting Incident Data (Historic), a comprehensive data on all shooting incidents in New York City from 2006 through the end of last calendar year. The data is pulled and analyzed manually by the Office of Management Analysis and Planning of the NYPD quarterly and subsequently made publicly available.

Dataset Structure

Rows: Each row represents a single shooting incident.

Columns: The dataset has 19 columns, some of the key variables being:

INCIDENT_KEY: A distinct key for each incident.

OCCUR_DATE and OCCUR_TIME: Date and time of the incident.

BORO: Borough of occurrence.

STATISTICAL_MURDER_FLAG: Flag indicating whether the incident was a fatality (TRUE) or not (FALSE).

VIC_AGE_GROUP, VIC_SEX, VIC_RACE: Victim demographic information, age group, sex, and race.

Data Import

```
url <- "https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv"

shootings <- read_csv(url)
```

The dataset contains 23,585 rows and 19 columns, providing a comprehensive view of shooting incidents in NYC over a 15-year period.

To streamline the analysis, several columns that were not relevant to the research question were removed. These included:

PRECINCT

JURISDICTION_COD

LOCATION_DESC

X_COORD_CD

Y_COORD_CD

Lon_Lat

The following R code was used to remove these columns:
shootings <- shootings %>% select(-c(PRECINCT, JURISDICTION_CODE, LOCATION_DESC, X_COORD_CD, Y_COORD_CD, Lon_Lat))
Converting Data Types

Several variables were converted to factors to facilitate analysis, as they represent categorical data:

BORO

PERP_AGE_GROUP

PERP_SEX

PERP_RACE

VIC_AGE_GROUP

VIC_SEX

VIC_RACE

STATISTICAL_MURDER_FLAG

The conversion was done using the following code:

shootings$BORO <- factor(shootings$BORO)

shootings$PERP_AGE_GROUP <- factor(shootings$PERP_AGE_GROUP)

shootings$PERP_SEX <- factor(shootings$PERP_SEX)

shootings$PERP_RACE <- factor(shootings$PERP_RACE)

shootings$VIC_AGE_GROUP <- factor(shootings$VIC_AGE_GROUP)

shootings$VIC_SEX <- factor(shootings$VIC_SEX)

shootings$VIC_RACE <-
factor(shootings$VIC_RACE)shootings$STATISTICAL_MURDER_FLAG <-
factor(shootings$STATISTICAL_MURDER_FLAG)

Fatal vs Non-Fatal Shootings visualizations
The dataset includes a flag (STATISTICAL_MURDER_FLAG) that indicates whether a shooting resulted in a fatality. The distribution of fatal and non-fatal shootings is as follows:

table(shootings$STATISTICAL_MURDER_FLAG)

Non-Fatal Shootings: 19,085

Fatal Shootings: 4,500

This indicates that approximately 19% of shooting incidents in the dataset resulted in fatalities.

By Age group

The age group of the victim appears to be a significant factor in determining the outcome of a shooting. The following table shows the distribution of fatal and non-fatal shootings by victim age group:

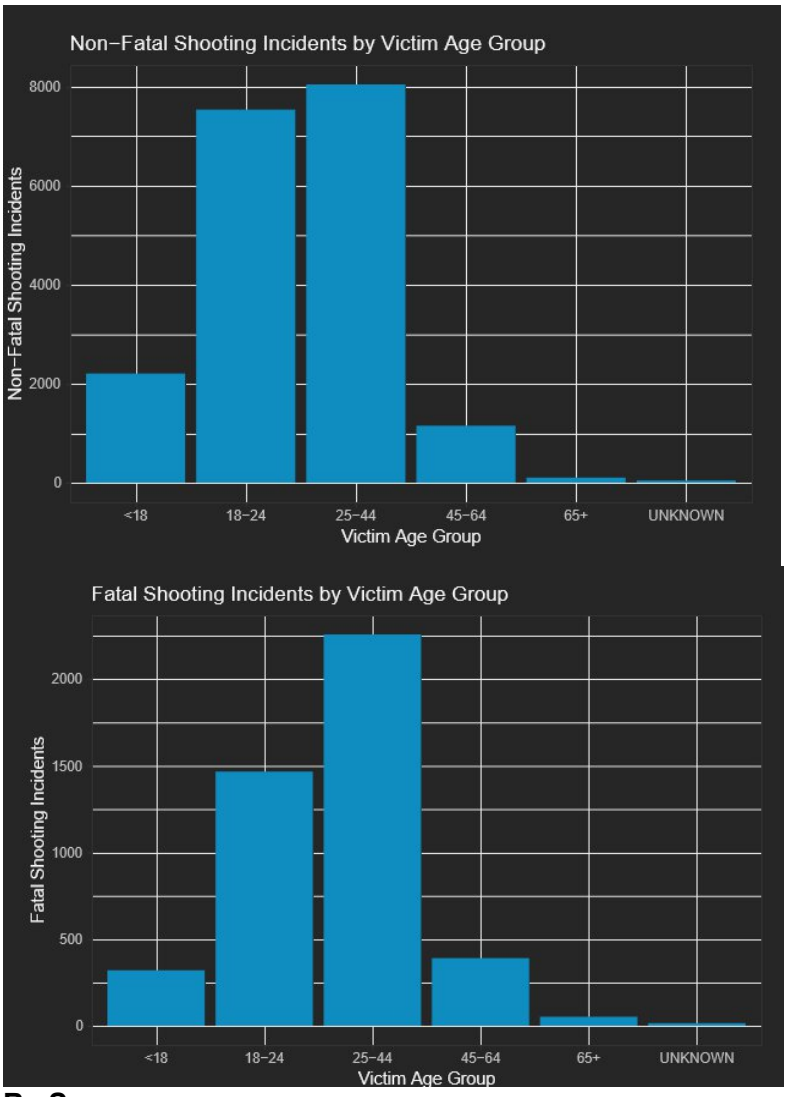
table(shootings$STATISTICAL_MURDER_FLAG, shootings$VIC_AGE_GROUP)

Non-Fatal Shootings: The majority of victims were in the 18-24 and 25-44 age groups.

Fatal Shootings: Similarly, the majority of fatal shootings involved victims in the 18-24 and 25-44 age groups, but the proportion of fatal shootings increased with age.

Visualizations were created to compare the distribution of age groups for fatal and non-fatal shootings:

```
# Non-Fatal Shootings by Age Group

shootings %>%

  filter(STATISTICAL_MURDER_FLAG == FALSE) %>%

  ggplot(aes(x = VIC_AGE_GROUP)) +

  geom_bar(fill = "#0F8DC0") +

  theme_bw() +

  labs(x = "Victim Age Group", y = "Non-Fatal Shooting Incidents", title = "Non-Fatal Shooting Incidents by Victim Age Group")

# Fatal Shootings by Age Group

shootings %>%

  filter(STATISTICAL_MURDER_FLAG == TRUE) %>%

  ggplot(aes(x = VIC_AGE_GROUP)) +

  geom_bar(fill = "#0F8DC0") +

  theme_bw() + labs(x = "Victim Age Group", y = "Fatal Shooting Incidents", title = "Fatal Shooting Incidents by Victim Age Group")
```



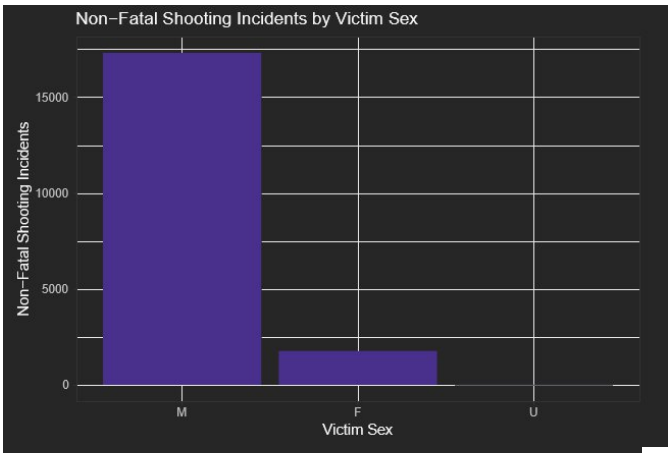
By Sex

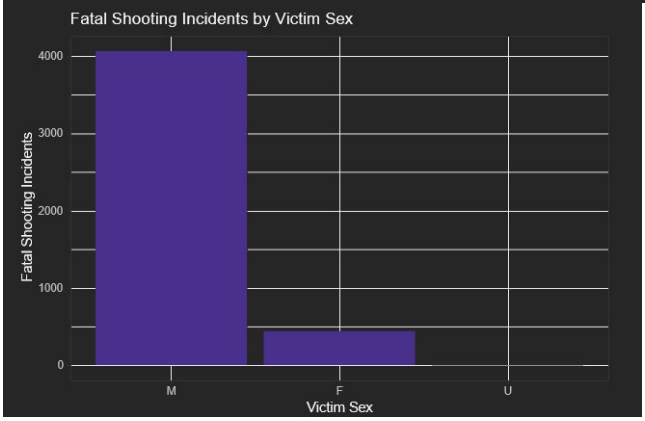
The sex of the victim was also analyzed to determine if it played a role in the likelihood of a shooting being fatal. The majority of victims were male, but the distribution of fatal and non-fatal shootings by sex was relatively consistent:
table(shootings$STATISTICAL_MURDER_FLAG, shootings$VIC_SEX)

Male Victims: 17,309 non-fatal, 4,061 fatal

Female Victims: 1,766 non-fatal, 438 fatal

Visualizations were created to compare the distribution of sex for fatal and non-fatal shootings:





The analysis suggests that while male victims are more frequently involved in shootings, the likelihood of a shooting being fatal does not significantly differ between male and female victims.

By race

The race of the victim was also examined to determine if it influenced the likelihood of a shooting being fatal. The majority of victims were Black, followed by White Hispanic and Black Hispanic individuals:

table(shootings$STATISTICAL_MURDER_FLAG, shootings$VIC_RACE)

Black Victims: 13,714 non-fatal, 3,155 fatal

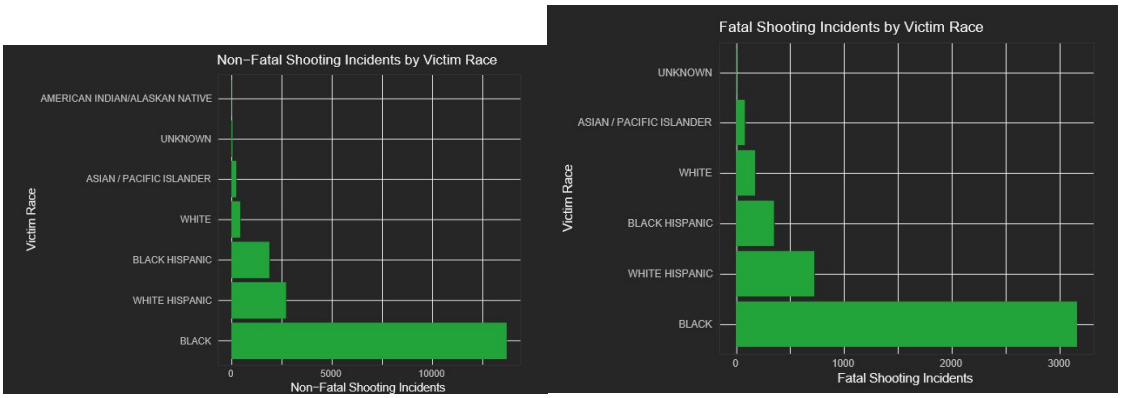
White Hispanic Victims: 1,893 non-fatal, 352 fatal

Black Hispanic Victims: 2,245 non-fatal, 531 fatal

Visualizations were created to compare the distribution of race for fatal and non-fatal shootings:

```
# Non-Fatal Shootings by Race

shootings %>%

  filter(STATISTICAL_MURDER_FLAG == FALSE) %>%

  ggplot(aes(x = fct_infreq(VIC_RACE))) +

  geom_bar(stat = 'count') +
```

```
  geom_bar(fill = "#23A43B") +
  coord_flip() +
  theme_bw() +
  labs(x = "Victim Race", y = "Non-Fatal Shooting Incidents", title = "Non-Fatal Shooting Incidents by Victim Race")

# Fatal Shootings by Race
shootings %>%
  filter(STATISTICAL_MURDER_FLAG == TRUE) %>%
  ggplot(aes(x = fct_infreq(VIC_RACE))) +
  geom_bar(stat = 'count') +
  geom_bar(fill = "#23A43B") +
  coord_flip() +
  theme_bw() +
  labs(x = "Victim Race", y = "Fatal Shooting Incidents", title = "Fatal Shooting Incidents by Victim Race")
```



The analysis indicates that while Black victims are disproportionately affected by shootings, the likelihood of a shooting being fatal does not significantly differ across racial groups.

Conclusion

The analysis of the NYPD Shooting Incident Data reveals that the victim's age is a good predictor of whether a shooting is fatal or not. Specifically, older victims (65+) are more likely to

die from the incident, while younger victims (18-24 and 25-44) are more likely to survive. Sex and race, on the other hand, do not appear to be significant variables in whether a shooting is fatal or not.