NYPD Shooting Incident Report

D.Parikh

2025-02-18

Overview

This R markdown file analyzed NYPD shooting incident data to provide results on:

-> Shootings per capita by Borough.

-> Proportion of Shootings resulting in Murder by Year.

-> Future Shooting incident forecast using time series (ARIMA).

The Libraries that we will be using today are

```
library(tidyverse)
library(forecast)
library(lubridate)
```

We are using shooting incident data from

```
NYPD_data <- read_csv("https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv")

NYPD_clean <- NYPD_data %>%
      mutate(OCCUR_DATE = mdy(OCCUR_DATE),
             OCCUR_TIME = hms(OCCUR_TIME))

NYPD_population_by_borough <- read_csv("https://data.cityofnewyork.us/resource/xywu-7bv9.csv")

NYPD_population_by_borough_clean <- NYPD_population_by_borough %>%
     mutate(average = (`_2000` + `_2010` + `_2020`) / 3)

NYPD_population_by_borough_clean_select <- NYPD_population_by_borough_clean %>% select(borough, average

NYPD_population_by_borough_clean_select <- NYPD_population_by_borough_clean_select %>% slice(-1)

borough_shootings <- NYPD_clean %>%
    group_by(`BORO`) %>%
    summarise(total_shootings = n())

borough_shootings <- borough_shootings %>%
      rename(borough = `BORO`)

NYPD_population_by_borough_clean_select = NYPD_population_by_borough_clean_select %>% mutate(borough =

NYPD_shootings_population_borough = left_join(borough_shootings,NYPD_population_by_borough_clean_select
```

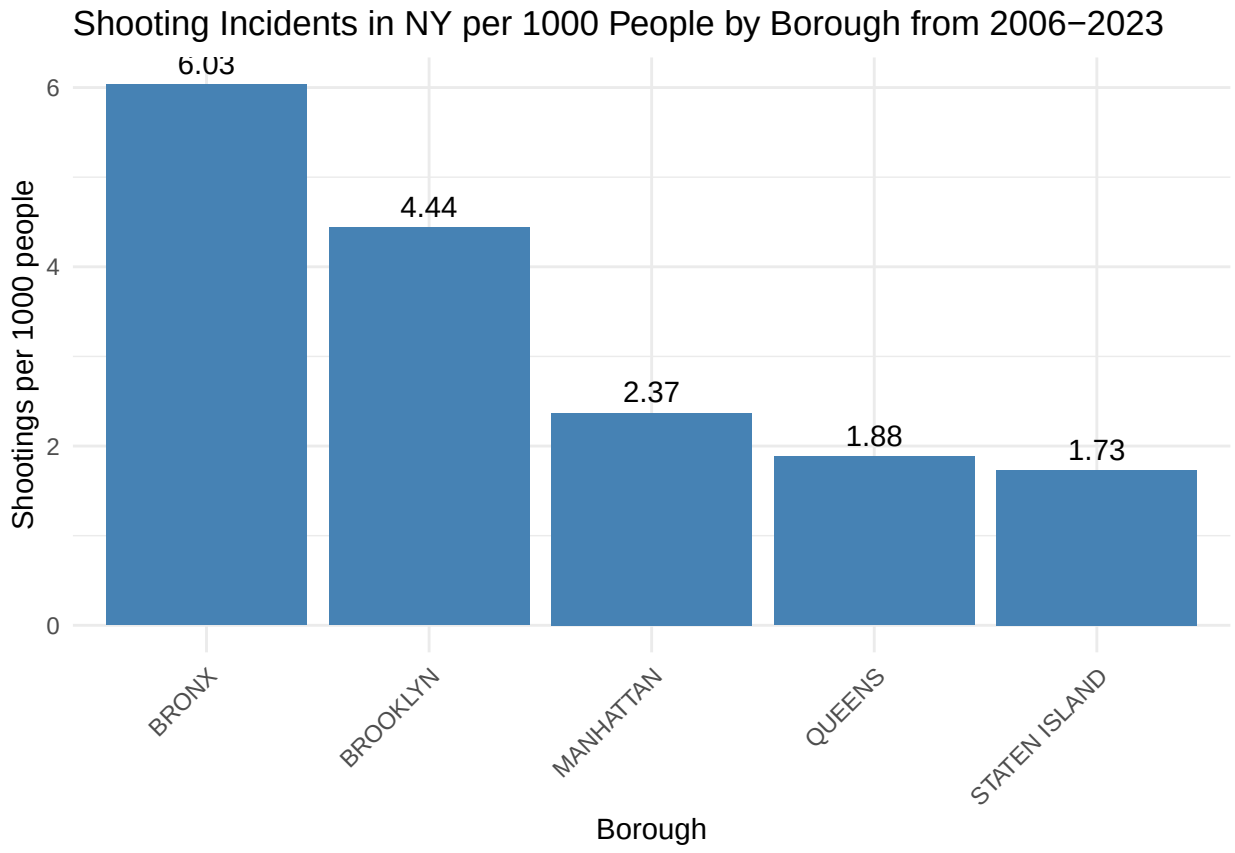
```
NYPD_shootings_population_borough = NYPD_shootings_population_borough %>% mutate(shootings_per_cap = to
```

Shooting per Capita by Borough

New York City is composed of five boroughs. Therefore, analyzing shooting incident data to identify the borough with the lowest shootings per capita can serve as a helpful metric for determining the safest borough in terms of shooting incidents.

Below is a bar chart showing shooting incidents per 1000 people by boroughs.

```
ggplot(NYPD_shootings_population_borough, aes(x = reorder(borough, -shootings_per_cap * 1000), y = shoo
  geom_bar(stat = "identity", fill = "steelblue") +
  geom_text(aes(label = round(shootings_per_cap * 1000, 2)), vjust = -0.5) +
  labs(title = "Shooting Incidents in NY per 1000 People by Borough from 2006-2023",
       x = "Borough",
       y = "Shootings per 1000 people") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



Proportion of Shootings resulting in Murder by Year

The NYPD shooting incident data also has a murder indicator. Thus, examining the data to determine a murder rate provided valuable insight on lethality of shooting incidents over time.

Below is a line chart showing proportion of shooting resulting in death over years.

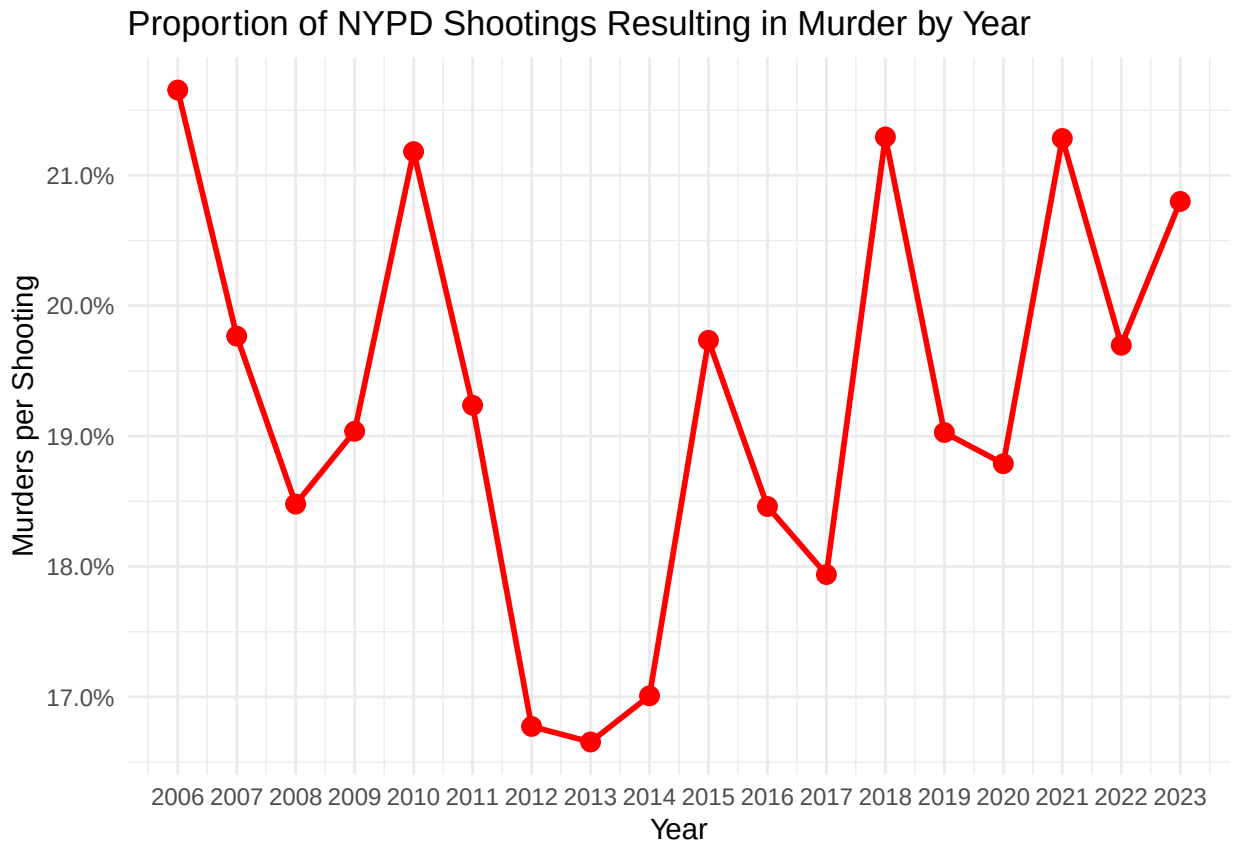
```
NYPD_clean <- NYPD_clean %>%
      mutate(Year = format(as.Date(OCCUR_DATE,format = "%Y-%m-%d"),"%Y"))

NYPD_murder_per_shooting <- NYPD_clean %>%
      group_by(Year) %>%
      summarise( total_shootings = n(), murders = sum(STATISTICAL_MURDER_FLAG == TRUE, na.rm = TRUE))

NYPD_murder_per_shooting <- NYPD_murder_per_shooting %>% mutate(murder_rate = murders/total_shootings)

NYPD_murder_per_shooting$Year <- as.integer(NYPD_murder_per_shooting$Year)

ggplot(NYPD_murder_per_shooting, aes(x = Year, y = murder_rate)) +
    geom_line(color = "red", size = 1) +
    geom_point(color = "red", size = 3) +
    scale_y_continuous(labels = scales::percent_format(accuracy = 0.1)) +
    scale_x_continuous(breaks = NYPD_murder_per_shooting$Year) +
    labs(title = "Proportion of NYPD Shootings Resulting in Murder by Year",
         x = "Year",
         y = "Murders per Shooting") +
    theme_minimal()
```



Future Shooting Incident Forecast

Looking at this data, it is very important to be able to forecast the amount of shootings based on the trend followed since 2006.

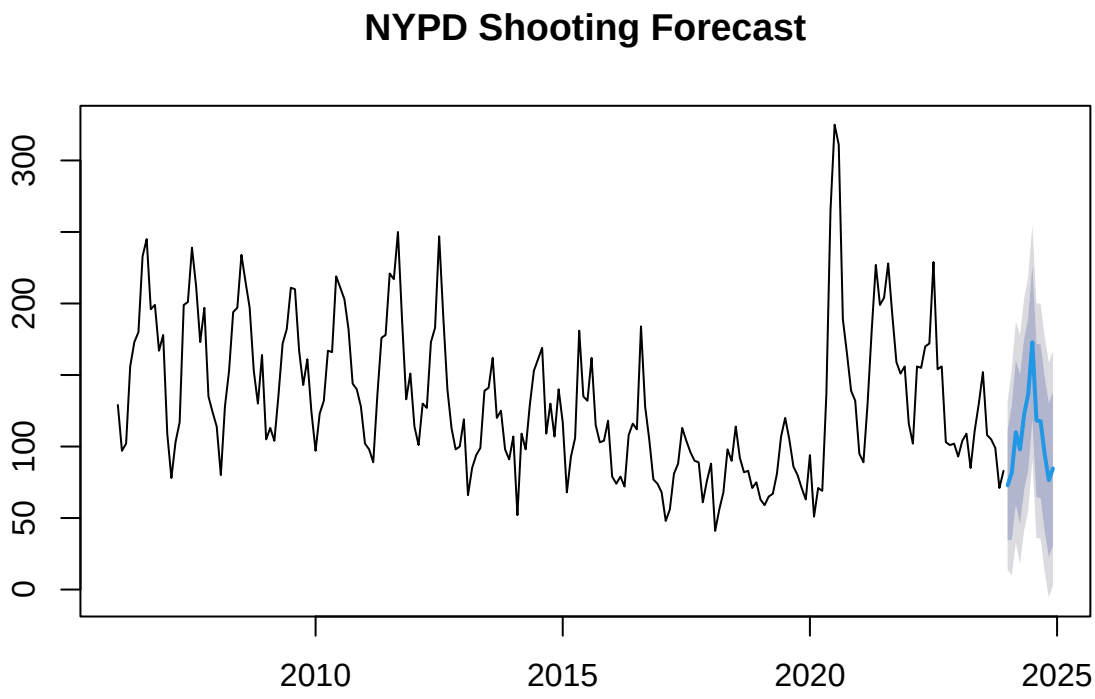
The end goal is to use this data to provide some actionable insight that the police can use to reduce the amount of shootings. Though the current analysis in this document is not enough to derive such an insight, it provides a good foundation for a deeper analysis.

Below is a forecast plot using time series analysis through ARIMA model.

```
NYPD_clean <- NYPD_clean %>%
    mutate(Month = format(as.Date(OCCUR_DATE,format = "%Y-%m-%d"),"%m"))

NYPD_shooting_fr_data <- NYPD_clean %>%
    group_by(Year, Month) %>%
    summarise(shootings = n()) %>%
    arrange(Year,Month) %>%
    mutate(date_occured = make_date(Year,Month)) %>%
    select(date_occured, shootings)

NYPD_roll_data <- ts(NYPD_shooting_fr_data$shootings,frequency = 12, start = c(2006,1))
NYPD_arima_model = auto.arima(NYPD_roll_data)
NYPD_shooting_forecast <- forecast(NYPD_arima_model, h =12)
plot(NYPD_shooting_forecast, main = "NYPD Shooting Forecast")
```



Bias

While analyzing the NYPD Shooting incident dataset, several biases are to be considered:

-> **Reporting**: Changes in reporting methodology through the years as well as through different precints and officers. To deal with Reporting bias, I have accepted the bias and made people aware about it.

-> **Missing data**: There is some missing information in different data fields such as **PERP_SEX** or **PERP_AGE_GROUP** amongst others. Furthermore, this data only constitutes recorded incidents. There might be much more incidents than that. To deal with Missing data, I have not analyzed any fields with missing data.

->**Social Bias**: There might be social bias against a particular race or ethnicity skewing the data. To deal with Social Bias, I have not analyzed any fields like race or sex.

->**Population Bias**: When looking at incidents by an area, the results might get skewed as the population of each areas will be different. To deal with Population Bias, I have used data from NYC OPEN DATA - "https://data.cityofnewyork.us/resource/xywu-7bv9.csv". This data provides population information in New York by Boroughs through census. For my Shootings by Borough analysis, I have taken average of population in 2000, 2010 & 2020 Census information to calculate a proportion of shootings to population.

Conclusion

This analysis explored NYPD shooting incident data (2006-2024) from NYC Open Data. We determined shootings per capita by borough, tracked the proportion of shootings resulting in murder over time, and generated a future shooting incident forecast using an ARIMA model.The analysis done provides a great opportunity to do further analysis on, and figure out the key indicators that will help us strategize to fight against these shootings and protect the NYC & its people.