

Habits

John Doe

March 22, 2005

Abstract

In this project, I conducted a thorough analysis of reported crimes from the New York Police Department dataset (2006-2022), focusing on offenses related to dangerous drugs and weapons. After filtering the data to include incidents reported between January 1, 2021, and December 31, 2022, and refining it to only include "DANGEROUS DRUGS" and "DANGEROUS WEAPONS," the dataset was made publicly available on GitHub.

The analysis addressed two key research questions. Firstly, it investigated whether there is a significant difference in average response times for incidents involving 'DANGEROUS WEAPONS' compared to 'DANGEROUS DRUGS.' The results, supported by summary statistics and a two-sample t-test, revealed a statistically significant difference, emphasizing the importance of tailored law enforcement strategies for distinct offense categories.

The second question explored potential variations in the average response time for reported 'DANGEROUS WEAPONS' incidents between Queens and Manhattan. Despite visualizations and summary statistics suggesting differences, the statistical test did

Import libraries

```
library(tidyverse)
```

Overview slide

In the overview slide, I'll go over the context of the data collection, the description of the independent and dependent variables and state the research question for this project.

Context on the data collection

The dataset represents reported felony, misdemeanor, and violation crimes documented by the New York Police Department (NYPD) from 2006 to 2022.

The original dataset has over 8.3 millions complaints. To narrow the focus, the data was filtered to only include incidents reported between January 1, 2021 to December 31, 2022.

Further the data was filtered to only include offenses falling under the categories of "DANGEROUS DRUGS" and "DANGEROUS WEAPONS," as denoted by the "OFNS_DESC" (Offense Description) variable.

The aim is to concentrate the analysis on crimes related to *dangerous drugs* and *dangerous weapons* during the specified time frame.

The final dataset was saved as a CSV file and made publicly

Import the data

Import the data from GitHub

```
Complaint_data <- read.csv("https://raw.githubusercontent.com/
```

```
Complaint_data %>%  
  summarize(row = nrow(.),  
            col = ncol(.))
```

```
##      row col  
## 1 28504  40
```

The dataset has 28504 complains and 40 columns

Answer Research question 1

This part of the project will answer the first research question by using the appropriate test statistic.

Summary statistics

Create a new column called "response_time" which is the difference between the time NYPD came vs Exact time of occurrence for the reported event

```
# Format the variable into time format
Complaint_data <- Complaint_data %>%
  mutate(CMPLNT_FR_TM = as.POSIXlt(CMPLNT_FR_TM, format = "%m/%d/%Y %H:%M"),
         CMPLNT_TO_TM = as.POSIXlt(CMPLNT_TO_TM, format = "%m/%d/%Y %H:%M"))
# Take the difference between the two column to find the response time
Complaint_data <- Complaint_data %>%
  mutate(response_time = difftime(CMPLNT_TO_TM, CMPLNT_FR_TM, units = "mins"))
#Print the first 5 response times
head(Complaint_data$response_time)
```

```
## Time differences in mins
```

Answer Research question 2

This part of the project will answer the second research question by using the appropriate test statistic.

Summary statistics

Select the columns we are going to answer the second research question.

```
Complaint_df2 <- Complaint_data %>%  
  select(response_time,  
         OFNS_DESC,  
         PD_DESC,  
         BORO_NM,  
         SUSP_RACE,  
         VIC_RACE)
```

Filter the data to only include offenses related to dangerous weapons in Queens and Manhattan

```
Complaint_df2 <- Complaint_df2 %>%
```

Why is this analysis important?

Analyzing response times for offenses involving dangerous drugs and weapons is crucial for law enforcement resource allocation, public safety, and policy development.

Limitations of the analysis?

Limitations include potential data bias from external factors like weather impacting results, challenges in inferring causation, and a limited temporal scope (2021-2022) restricting broader applicability to evolving crime trends or law enforcement practices.