

Data 608 - Proposal Final Project

Gabriel Santos

2023-03-25

Motor Vehicle Collisions Analysis

The goal of my final project is to visualize the motor vehicle collisions crash in New York City.

I want to be able to show through graphs an analysis of the crashes of the last 5 years. Identify which days of the week there are more crashes. Identify which months have more crashes. Identify crash by time of day. Identify which boroughs have the most number of crashes.

I think this analysis is revealing since we can identify where more crashes are occurring and be able to suggest policies to reduce the number of crashes. We can also find out if we have improved in the policies implemented in recent years or if new policies should be proposed to reduce crashes. We can generate training to reduce crashes according to the factors that present more frequency in ls crash.

Data source

The dataset I plan to use for my final project are: Motor Vehicle Collisions - Crashes, from NYC Open Data.

Link: <https://data.cityofnewyork.us/Public-Safety/Motor-Vehicle-Collisions-Crashes/h9gi-nx95>

The Motor Vehicle Crash Data Tables contain information on all police-reported motor vehicle collisions in New York City. It has data on the date and time of the accident. In which borough the accident occurred. If they show up injured or dead. The factor that contributed to the accident. The type of vehicle. Among other data.

```
url_data <- "https://raw.githubusercontent.com/GabrielSantos33/DATA608_Final_Project/main/Data_Motor_Vel
crash_df <- read.csv(url_data)
crash_df <- as_tibble(crash_df)
```

```
colnames(crash_df)
```

```
## [1] "crash_date"      "crash_time"      "borough"
## [4] "number_injured"  "number_killed"   "contributing_factor"
## [7] "vehicle_type"
```

```
nrow(crash_df)
```

```
## [1] 328879
```

```
## # A tibble: 10 x 7
##   crash_date crash_time borough   number_injured number_k~1 contr~2 vehic~3
##   <date>      <chr>      <chr>         <int>         <int> <chr>   <chr>
## 1 2020-10-06 3:52      QUEENS             1             3 Unsafe~ Sedan
## 2 2020-07-11 20:18    BROOKLYN           4             3 Driver~ Sedan
```

```
## 3 2022-07-10 20:47      STATEN ISLAND      6      3 Unsafe~ Sedan
## 4 2019-06-08 0:07      BROOKLYN          0      2 Unsafe~ Sedan
## 5 2020-12-12 22:52     BROOKLYN          0      2 Traffi~ Sedan
## 6 2021-05-24 0:27      QUEENS            0      2 Unspec~ Sedan
## 7 2021-12-16 0:15      BRONX             0      2 Unsafe~ Statio~
## 8 2021-12-24 9:21      MANHATTAN         0      2 Unspec~ Box Tr~
## 9 2021-01-01 4:40      QUEENS            1      2 Unsafe~ Pick-u~
## 10 2022-07-19 14:25    BRONX             1      2 Unsafe~ Motorc~
## # ... with abbreviated variable names 1: number_killed, 2: contributing_factor,
## # 3: vehicle_type
```

Methodology

The final project will be done by RStudio(.rmd) and I will create a visualization in shinyApp, using Plotly. To see the frequencies and types of crashes that occur, also to be able to analyze in which borough more crashes occur.

The libraries I plan to use are:

- library(tidyverse)
- library(readr)
- library(plotly)
- library(gridExtra)
- library(kableExtra)
- library(lubridate)
- library(ggrepel)
- library(rgdal)
- library(scales)
- library(shiny)

Visualization

- View crash for the last 5 years.
- Amount of crash according to the day of the week.
- Which quarters have more crashes per year.
- crash by time of day.
- Which boroughs have more number of crashes.

Conclusion

Present the conclusions according to the graphs made and thus be able to identify the data proposed for this project.