

Analysis of Car crashes and Metro traffic in NYC

Abstract

The goal of this project is to work with New York Police Department to try to reduce the number of car accidents in New York City by finding how can metro stations traffic impact the number of car accidents. I worked with data provided by new.mta.info and NYCOpenData to achieve the goal of this project. After cleaning the data, I started to explore the datasets to find answers that supports the goal I am trying to achieve. After visualizing my findings, I communicated my work using a PowerPoint presentation to better explain my findings to the police department in New York city.

Design

By finding the relation between the rate of car accidents and the traffic of the metro stations, we could help encourage more people to use the metro rather than using cars. Which will lower the rate of accidents and cut the costs and efforts of NYPD in handling the accidents.

Data

- **MTA Dataset:**

The New York subway MTA turnstile data is a series of data files containing cumulative number of entries and exits by station, turnstile, date and time. Data files are produced weekly. Data records are collected typically every 4 hours with some exceptions. Sample size is 3 months (from May to July 2019) to find the relation between car accidents and metro traffic without the impact of covid-19. The dataset has 2480677 rows and 11 columns.

- **Car Collisions Dataset:**

The Motor Vehicle Collisions crash table contains details on the crash event. Each row represents a crash event. The Motor Vehicle Collisions data tables contain information from all police reported motor vehicle collisions in NYC. The police report (MV104-AN) is required to be filled out for collisions where someone is injured or killed, or where there is at least \$1000 worth of damage. Sample size is 3 months (from May to July 2019) to find the relation between car accidents and metro traffic without the impact of covid-19. The dataset has 1815759 rows and 29 columns.

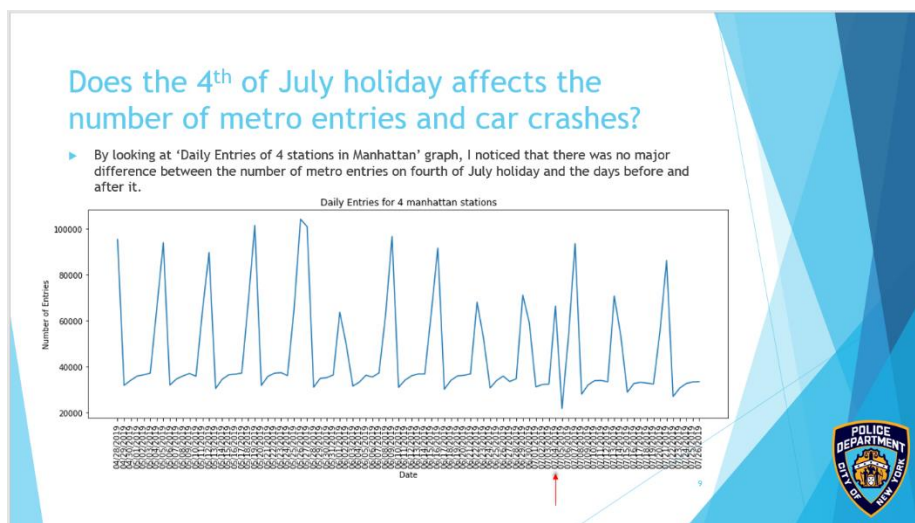
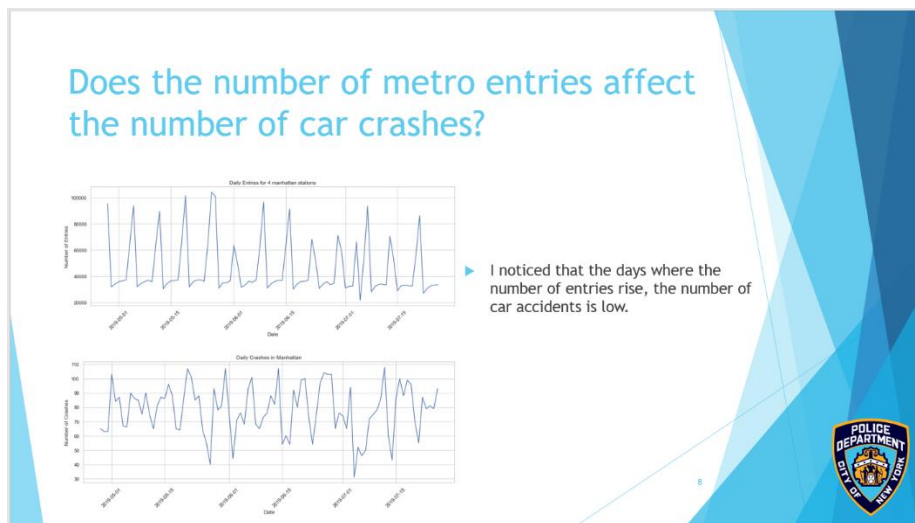
Algorithms

The methodology used in this project is: Problem understanding, data validation, Data exploration, data visualization.

Tools

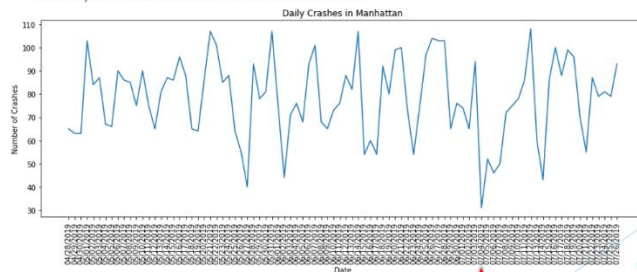
- Numpy and Pandas for data manipulation
- Matplotlib and Seaborn for plotting

Communication



Does the 4th of July holiday affects the number of metro entries and car crashes?

- ▶ However, by looking at 'Daily Crashes in Manhattan' graph, it is noticeable that the number of car crashes dropped that day.
- ▶ Which leads me to the conclusion that the 4th of July holiday does not affect the traffic in the metro stations, but it does affect the number of car crashes.



Does the number of stations in each borough affects the number of car crashes in that borough?

- ▶ By looking at the two graphs, it appears that the number of stations in each borough may have an impact on the number of car crashes in that borough.

