New York City Collision Insights: Uncovering Traffic Safety Trends

Preeyonuj Boruah*, Prateek Srivastava*, Yogesh Sharma*
Department of Electrical and Computer Engineering
New York University
Tandon School of Engineering
Net-IDs: pb2715, ps4792, ys5270

Course Name: Big Data
Section: D
Semester: Fall 2023

Abstract

Following up on the success of the CompStat program in collecting rich and comprehensive data on homicides, NYPD started collecting data on traffic collisions as the second primary source of fatalities. Our project aims to leverage that data to understand the dynamics involved in a collision and if there exists a prevalent pattern. The dataset deals with several aspects of a crash, including date-time, exact locations, statistics of people involved, contributing factors, and vehicle details. We will work across every aspect of the data and create visualizations to identify critical components that result in a collision. Our analysis will include data processing using PySpark, Feature Engineering, and crafting visualizations to provide tangible insights. This document will state the Problem Statement, Objectives, Data Source information, Technologies used, Analysis, Visualization, and Future Scope.

Problem Statement

While New York City has made significant strides in implementing systems to collect and analyze data related to traffic accidents, the evolution from the Traffic Accident Management System (TAMS) to the Finest Online Records Management System (FORMS) has resulted in a rich repository of detailed accident data from MV-104AN forms. This comprehensive data, captured in the Motor Vehicle Collisions crash table, offers a unique opportunity to better understand the factors and patterns contributing to traffic accidents. However, there is a challenge in efficiently extracting actionable insights from this expansive dataset. The overarching problem is: "How can we utilize the comprehensive data from the Motor Vehicle Collisions crash table to derive meaningful insights and strategies to effectively reduce and eventually eliminate traffic-related fatalities in New York City?"

Objectives

Our analysis aims to achieve the following objectives:
1) Understanding how the contributing factors weigh in when related to the date-time aspect of the collision. Additionally, assessing the relationship between the location and contributing factors.
2) Generating borough-wise analysis of the collisions and cross-referencing it with different periods of a day.
3) Examining if seasonality plays a role in the frequency of collisions, such as holidays or actual seasons like Summer, Winter, etc.
4) Investigate the dynamics of collision between different types of vehicles, such as between two privately-owned vehicles or one public-owned vehicle (bus or taxi) with a privately-owned vehicle, etc.

Data Source

The Dataset is named "Motor Vehicle Collisions - Crashes" (New York Police Department (NYPD) - CompStat 2023). The link to the dataset is here. The data file size is 431.04 MB and has 2.04M rows. It consists of 29 features that focus on Crash date, Crash time, Borough, Zip codes, Accident contributing factors, People injured and other miscellaneous information.

Proposed Technologies

We can divide our project into 3 main sections:
1) Data Processing: Spark will be our central platform for processing and storing the data. Since it is open source and has extensive community support, it will be an excellent option for us. Apart from that, the MLlib module within Spark contains an optimal implementation of several statistical functions required in our analysis.
2) Analysis: Since we are using Spark as our data processing platform, we will use PySpark to run analysis queries. Even if Scala is an option, due to its familiarity with Python, PySpark has better functionality and comprehensive analytic-friendly modules.
3) Visualization: For visualization, we intend to use PlotLy through their Python APIs to create visualizations either on their platform or using Jupyter Notebooks.

References

New York Police Department (NYPD) - CompStat. 2023. Motor Vehicle Collisions - Crashes.

*These authors contributed equally.

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