Mohammed Saad Khan 2019250 CE

Husnain Ali 2019178 CE

Exploring Traffic Violations Data to Identify Trends and Insights

(CS-439) Data Science

Project Proposal

Introduction:
This project's goal is to investigate and evaluate the Kaggle dataset on traffic infractions. The dataset includes a thorough history of traffic offences in Maryland, United States, from 2012 to 2018. The dataset comprises of nearly 4 million observations and 35 variables, including data on the demographics of the drivers, the features of the vehicles, and the specifics of the violations. Through this study, we hope to understand the causes that contribute to traffic violations, identify the most frequent offences, and acquire insights into the patterns and trends in traffic violations.

Objectives:
To make the data on traffic offences ready for study, clean and preprocess it. Use descriptive analysis to determine the most frequent traffic infractions, the characteristics of the drivers who commit them most frequently, and the places where they occur most frequently. Look for any seasonal patterns or trends by analyzing the temporal patterns of traffic violations. Investigate any relationships between the type of infraction, the characteristics of the driver, and the time of day, day of the week, or season. To determine any characteristics that lead to increased rates of traffic offences in specific areas or for particular demographics, use machine learning approaches like clustering or regression.

Methodology:
To ensure that the dataset is prepared for analysis, the project will start with data cleaning and preprocessing. The next step is to find patterns and trends in the data using descriptive analysis techniques. To find any seasonal or temporal trends in traffic violations, temporal analysis will be done utilizing time-series analysis and visualization tools. Machine learning techniques will be utilized to uncover characteristics contributing to high rates of traffic violations, and exploratory analysis will be undertaken to identify relationships between driver demographics and violation categories.

Expected Outcomes:
This research will shed light on the patterns and trends in traffic offences, the characteristics of the drivers who commit the most offences, and the hot spots for offences. Also, we hope to find any temporal or seasonal patterns in traffic offences and investigate links between different types of violations and driver characteristics. Also, by identifying any reasons that may be behind high traffic violation rates, machine learning approaches will assist in informing both policy and enforcement activities.

Conclusion:
To find trends and insights into the patterns of traffic offences, this project will examine and analyze the dataset of traffic violations. The findings of this project will be invaluable in helping law enforcement organizations and decision-makers develop efficient strategies for preventing traffic infractions and ensuring road safety.