Traffic Accidents: The where and the why.

Kapilan Mahalingam, Business Project Proposal, August 2022

Introduction

This project was completed in August of 2022 as part of the Metis Data Science and Machine Learning Bootcamp's business module. The goal of this project is to quickly analyse a large dataset, finding interesting questions and patterns that merit further analysis, and create an interactive dashboard to allow interested stakeholders to do the same. Specifically, the project focuses on identifying business needs quickly rather than deep analysis or complex techniques---it's more focused on communication and ideas.

Question/Need:

What are the areas/conditions that lead to the most traffic accidents? How do we optimise roadway resource allocation to minimize accidents, and provide input to design decisions going forward?

Data:

For this project, I'm using a dataset from Kaggle, on US traffic accidents. Due to the size of the dataset I'll only be using a portion of it, with the goal of identifying solutions that could quickly help decrease accident severity.

The impact hypothesis for this project is that identification of high accident zones/conditions will let implement strategies to improve saftey.

Solution Path

Firstly, a classification model to try and sort locations (streets, or counties) into high/low accident zones. Then looking at the realtionship between weather conditions/accident sites with frequency of accidents, and identify anomalously accident-prone areas. Allowing the stakeholders to play with and do the above themselves using an interactive dashboard is the end goal.

Tools:

I'll use Excel and Tableau, as these are the course requirements. Correlation plots, scatterplots, histograms sould be able to identify the most accident-prone locations and weather conditions. Following that, using a ML model to identify predictors of accidents might be a productive direction to take.

MVP Goal:

Identifying the most dangerous street/county/city in the data would be a good starting point.