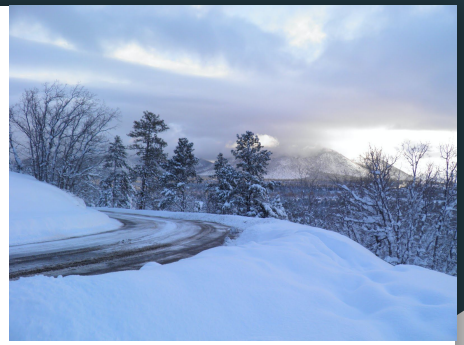
Capstone Project

Car accident severity in Seattle

Introduction

The purpose of this project was to use environmental factors to predict the severity of a car accident, allowing the driver or AI controlling the vehicle to prepare in advance for these likely areas or conditions in which the most severe incidents occur.

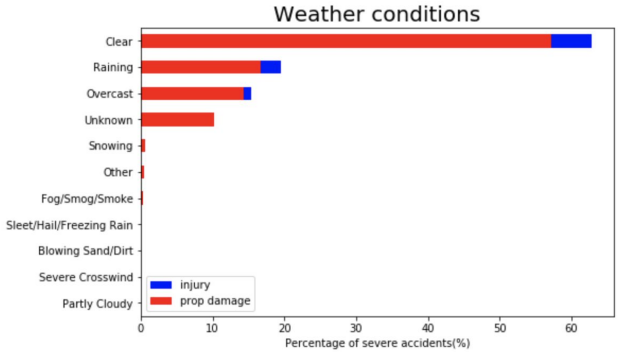
Data



The data used in this project relies on previous statistics recorded by Seattle City, which records the variables within historic accidents, the vehicle or vehicles involved and the conditions during the incident.

The severity of the incident will be the dependent variable whereas environmental factors such as weather conditions, light conditions and road conditions will be our independent variables.

Methodology



By splitting the master dataset into columns we care about and removing blank rows we can make this bulk data set into a more refined set.

We can then split this data into severe incidents and minor incidents, and then compare the conditions precursing each event.

By graphically displaying these datasets as normalised percentages of incidents we can visually determine the findings.

Results & Discussion

The visual representation of the statistics in a graphical form was a really useful way to display the results and make sound analytical judgement based on the data available

The plots clearly displayed no evidence to support the conclusion that severity of incident was dependent on the weather, light or road conditions.

Unfortunately the dependent variable did not have the scale of events desired to gain a greater spectrum of incidents and their likely causes, and therefore I am not content with completely ruling out severity of incident based on conditions. It would be interesting to investigate the likelihood of an incident based on these conditions - for example, during a bad snow storm not many people will go out in their cars, but of those that do, I predict the likelihood and risk factor is increased for severe incidents. This would be the ideal dataset to approach this problem, but is naturally very hard to gain statistical data on the number of drivers at any given hour.

Conclusion

From the data available there is no correlation between road, weather or light conditions and severity of incident. Most incidents occured in optimum conditions, however I predict this is due to the number of vehicles and pedestrians/cyclists on the road during these times.