Introduction

Across the US, crashes each year result in thousands of lives lost, injured victims, and billions of dollars in property damage. The National Highway Traffic Safety Administration (NHTSA) uses data from many sources, and one of their goals is to reduce both human and property damage.

Many different factors are provided to the NHTSA, and not all factors are important in determining the severity of accidents. The goal is to identify these important factors and take a proactive approach in minimizing both the cost and the severity of accidents.

Business

Not all information provided to dispatch personnel is relevant to the severity of the accident. The goal is to identify and predict factors that contribute to injuries to people and the severity of the injuries. By identifying these factors, we aim to predict and minimize further serious accidents as well as giving emergency personnel the ability to prepare for the severity of the accident based on the information provided by dispatch personnel using our model.

Target Audience

The intended audience for this project will be the Department of Transportation of Seattle, Washington. Due to the danger of vehicle collisions, providing solutions that may reduce the amount of accidents can significantly improve the quality of life of pedestrians & overall ensure public safety.

Data

The data that will be used to conduct this analysis is the compiled collision dataset from 2004 to Present within Seattle, Washington. This data source includes 194,673 rows and 38 columns was taken from the Seattle Department of Transportation that is continuously updated weekly. This will include the severity of each accident, the type of vehicle involved if any, the location of where the collision took place, as well as weather conditions that may have had an adverse effects on the event.

Specifically, weather conditions, road conditions, and types of junctions can collectively be used to determine the most dangerous parts of the road where severe accidents are most likely to occur. Most importantly, it contains a severity code that ranges from 0 (unknown) to 3 (fatality). Being able to use the various features within the dataset to better predict this the level of severity of the collision can allow proper safety features to be placed in certain junctions where the probability is higher.