

Introduction

For many travelers, having the freedom to explore the city to experience and enjoy various aspects is a great way to relax and enjoy weekends. Similarly, traveling via a vehicle, cycling, or walking are the most prominent ways for one to journey throughout the city for your everyday errands. Due to the large population that uses public roads, safety has become a top priority for the municipal government to ensure they are able to reduce the number of accidents that occur. Therefore, analyzing the various factors that could help predict accident severity can guide the board to implement changes in a timely manner that may reduce the number of fatalities & serious injuries. Consequently, this will ultimately lower the economic costs that would trickle into other fields within the economy.

Business Problem

The objective of this capstone project is to analyze the collision data set for Seattle, WA and determine the most pertinent factors including weather, road conditions, visibility, and various other factors that best predict accident severity. Using various analytical techniques and machine learning algorithms such as K nearest neighbors. This project will be used to answer the business question: How can the city of Seattle, Washington best predict the severity of collisions that occur and what avenues can be explored to remedy this issue?

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.