**PREDICTING CAR ACCIDENT SEVERITY IN SEATTLE**

Submitted for the Coursera Data Science Capstone Project

**Abstract**

This report analyses attributes which may affect car accident severity in Seattle, Washington State US, by using a Machine Learning Model

**Submitted by Portia Maurice**

**October 2020**

1. **INTRODUCTION**
   1. **BUSINESS UNDERSTANDING**

Road accidents are a global problem which, according to the World Health Organisation, results in the deaths of 135 million people annually. However, while mortality is a key indicator of the scale of the problem, non-fatal outcomes should also be measured to reflect fully the burden of disease due to road traffic collisions and understanding the attributes that lead to these is important data to inform policy interventions.

Injuries and deaths are preventable and, while there are regulations in place to curb these – such as those for wearing seat belts, driving without distractions and driving under the influence of alcohol or drugs – there is always the opportunity to gain greater insight through understanding the causes of such accidents.

This study presents us with the opportunity to predict the severity of an accident using key attributes such as weather and road conditions, based on the 2004-2020 dataset on car accidents in the city of Seattle in Washington, USA. The predictive model can be used to enhance safety and help mitigate the incidence or severity of road traffic accidents.

The study would be of benefit to the following key stakeholders:

* Government – in particular the Seattle State Department of Transportation and the US National Highway Traffic Safety Administration (NHTSA), as well as urban planners
* Car manufacturers – safety is a key driver in the industry, and
* Citizens of the city of Seattle

In the United States broadly, the road traffic fatality rate was 15.2 per 100 000 inhabitants in 2000, according to the World Health Organisation. This is among the highest for developed countries but has been significantly reduced over the past 20 years and now stands at 11.2 deaths per 100 000 people. Interestingly, although the trend is tracing downward, during Q2 of 2020, when there were far fewer vehicles out on the road due to Covid19 stay-at-home orders, fatalities per 100 million vehicle miles travelled increased to 1.42, up from 1.06 for the full year 2019 – this even though the overall fatality rate declined.

The seaport city of Seattle is the largest city in the state of Washington. It is in the top 10 most dangerous cities for car accidents in the US, but City officials are proud that the trend has been a downward one for more than 11 years. According to data from the Seattle State Department of Transportation, the city sees more than 10 000 crashes a year, resulting in an average of 20 people losing their lives and over 150 seriously injured. Seattle has set itself a hoal to end traffic deaths and serious injuries by 2030 and a study such as this would assist them to anticipate issues, invest in smarter street design and other safety enhancements, informed by the data.

The key objective of the model is to isolate those features that best predict the severity of an accident. The output of the model could inform plans for infrastructure upgrades, safety campaigns and traffic control. In addition, as the municipality plans for a future which includes self-driving cars, it becomes even more relevant to understand the causes of accidents and mitigate the risks.

1. **DATA UNDERSTANDING**

The dataset to be used is a publicly accessible record of accidents in Seattle between January 2004 and May 2020. The full labelled dataset contains data on 190 000 accidents during the period. Each incident has been assigned a unique identifier, and a total of 37 data attributes which could be used to describe the accident, have been included. Among the attributes included are weather conditions, location data, collision type, light and road conditions, and whether or not the driver was under the influence or distracted. Not all of these will be significant predictors of accident severity, which is the target variable.

A group of attributes to input into the machine learning model will be selected based on their level of correlation with the target variable. Regression and other methods will be used to determine the set of multiple data attributes which may affect accident severity. In turn, the machine learning model will be built to predict the severity of an accident based on these attributes. Some feature engineering will be required to improve the predictability of the model. Data visualisation, using Seattle road maps in tandem with the data, will be used to obtain a view of accident density across the city. Data preprocessing and standardisation methods will be employed as part of the exploratory data analysis, to prepare the dataset before providing it as input to the machine learning algorithm.

References:

<https://www.who.int/violence_injury_prevention/road_safety_status/2018/en/>

<https://www.caranddriver.com/news/a340145/2019-2020-traffic-deaths-coronavirus/>

<https://www.colburnlaw.com/category/pedestrian-safety/>

<https://www.seattle.gov/visionzero>

<https://www.seattletimes.com/seattle-news/transportation/seattle-traffic-deaths-and-injuries-down-slightly-last-year-most-of-the-fatalities-were-pedestrians/>

<https://cdc.gov/injury/features/global-road-safety/index.html>