IBM Applied Data Science Capstone Report Week 2

Reducing The Severity of Car Accidents in Seattle

Azizi Omar

September 2, 2020

1 Introduction

1.1 Background

Car accident, also called as a traffic collision, occurs when a vehicle collides with another vehicle of any size, animal, pedestrian, road debris or any other stationary objects, such as a tree, building or pole. Traffic collisions often result in property damage, injury, permanent disability, death and financial costs to all the individuals involved. In the United States, the yearly average number of car accident is 6 million [1]. 3 million people in the United States are injured every year in car accidents [1].

1.2 Business Problem

The severity of car accident is highly linked to the type and the number of involved parties. In general, the most severe car accidents are those which cause human injury. This is the aspect of car accident which we are to focus in this capstone project. Factors which contribute to this type of injury need to be determined. In this capstone project, after identifying the factors of car accident contributing to the human injury, we will investigate the methods to reduce the severity of car accidents in Seattle related to human injury.

1.3 Interest

Seattle Police Department (SPD) and the government of Seattle would be interested with the outcomes of this capstone. They would be able to use them to propose and implement adequate measures to reduce the human injury caused by car accident.

2 Data

2.1 Data Sources

For this project, we will use the shared data, which is the collision data provided by Seattle Police Department (SPD). The timeframe of the data based on the provided description is from 2004 to present. The dataset includes all types of collisions. For our purposes, we will only use collision data involving **vehicles with other parties**.

2.2 Data Description

A pandas profiling of the dataset has been generated and accessible via this link https://zyron92.github.io/coursera/ibm_applied_ds_capstone_profiling_final.html.

There are 38 attributes in this dataset, excluding the index. The attribute which is going to be our label or **target variable is the severity code ("SEVERITYCODE"),** as this variable will determine the severity of the car accident. The attribute of vehicles count ("VEHCOUNT") will help us to determine traffic collisions involving vehicles (cars). The rest of the attributes will be further investigated, in order to determine the important features which contribute to human injury caused by car accident.

There is a total of **189,588 observations** of traffic collisions involving minimum one vehicle (car). Percentage of missing cells is fairly high, 14.1% (1,042,065). The missing data and the outliers will be handled in the data preparation part of the next section.