Introduction

Car accidents can occur all the time, however there are some conditions that may cause probabilities of having an accident arise due multiple variables. This report has a purpose to develop a model for Seattle government to predict the probabilities of having a car accident and severity, based on different conditions such as weather or road conditions.

The information was provided by Seattle Police Department from 2004 to 2020.

Business Problem

Identify the conditions that can cause future car accidents in order to alarm the people with anticipation to be aware and drive more carefully.

The avoidance of car accidents can result in many benefits:

- Save lives; or
- Reduce costs in damage infrastructure; or
- Reduce cost from police and paramedics to attend each accident

Data

The information comes from Seattle Police Department and recorded by Traffic Records and include Collisions at intersections or mid-blocks of a segment. The period information is from 2004 to May 2020.

The information is organized in a CSV file with 37 attributes and originally 194673 rows, the information is labelled and unbalanced. Additionally, a document with the description of each column were given.

Due to the labelled information, we can know the result of each record. We have selected the column SEVERITYCODE as Dependent variable. The possible values are:

- 1 – Property Damage Only Collision; and
- 2 – Injury Collision

For simplicity of the report, we shall consider the following:

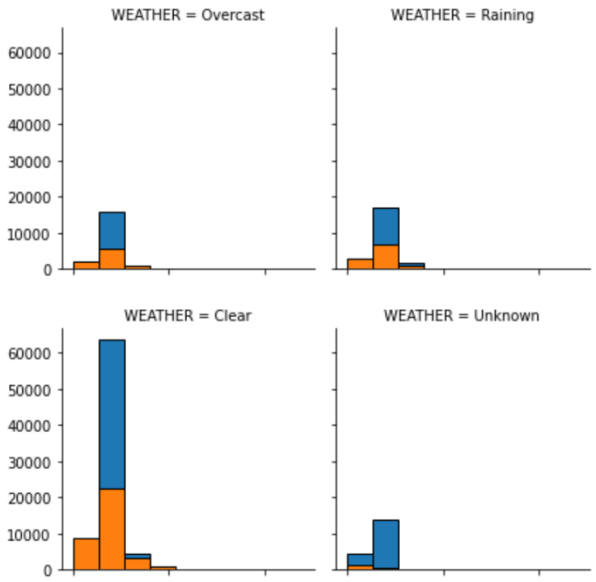
1. The weather conditions
2. The time of day

INCDTTM is the time of the accident, so is necessary convert that information to discrete values

- Midnight = 1 morning if hour is less than 6 and hour is more than 0;
- Morning = 2 if hour is more than or equal to 6 and hour is less than 12;
- Afternoon = 3 if hour more than or equal to 12 and hour is less than 18;
- Night = 4 if hour more than 24 and hour is less than or equal to 18

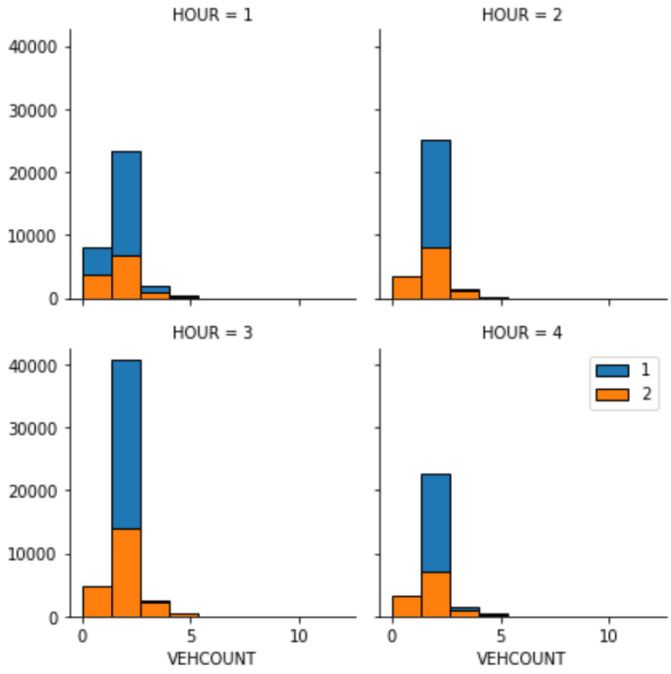
The weather conditions

Surprisingly, there are a greater number of accidents when the day is clear.



The time of day

In the next figure, we shall observe that a higher number of accidents was reported during the afternoon.



Methods of Classifications

We have compared the result using this methods.

- Linear regression
- Decision Tree

	linear regresion	decision tree
0	0.674835	0.66665

Conclusion

Research has shown that there is a higher probability of serious accident when ROADCOND is dry, and when the weather is overcast, raining or clear. A linear regression model is built to predict the severity of an traffic accident.