**Introduction/Business**

Road safety procedures are an import aspect of road management and driving as they save lives, reduce injuries and keep roads running in an optimal way. Road safety procedures are constantly evolving as new methodologies to keep the number of incidents and the severity of incidents to a minimum. However, with road incidents still occurring there is still improvements to be made in road safety procedures, therefore it is necessary for predict if there are certain road conditions and factors that lead to more incidents with higher severity. If it is possible to predict these road conditions that lead to incidents then it will allow authorities to notify drivers when to drive more carefully in ‘higher risk’ conditions which may reduce the likelihood of an incident, save lives and reduce the severity of crashes. This project will aim to predict the severity of an accident in different road conditions.

**Data**

Data that may contribute to the increased number of incidents and severity may include weather conditions, light conditions, junction type, speed of car and metrics to describe the kind of crash. The data comes from Seattle and comprises traffic collisions from 2004 to the present and is updated weekly. The data contains key information on time, location, severity, number of persons involved, incident type, injuries, environmental conditions and human wrongdoing.

The data must be cleaned to ensure than missing data and anomalies do not skew the model results. The data is already combined into a table, however some data is missing. Where missing data was needed but was not available it was removed. Features that will not be required, such as REPORTNO will be removed. Data will be visualised to determine if there are any outliers/anomalies that may skew the model results and should be removed. The data must be explored to determine any relationships between the data. For example we can plot the severity of incident against the road conditions to determine if there is a relationship.

Here are some charts showing the split of different feature types that might be useful. We will look at these relationships to determine if they can be modelled together to provide

A screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generated

*Chart displaying the number incidents per address type*

*Chart displaying the number incidents per road condition class*

A picture containing drawing, clock

Description automatically generatedA screenshot of a cell phone

Description automatically generated

*Chart displaying the number incidents per light condition class*

*Chart displaying the proportion of class 2 vs class 1 incidents*