**Introduction**

There have been many traffic accidents causing injuries or property damage in Seattle between 2004 and 2020. The data set being used in this capstone recorded the severity of accidents as either property damage or injury. I am going to create a predictive model to inform drivers of the important factors that contribute to injuries caused by traffic accidents. The objective of the model is to inform drivers so they can make safer driving decisions and reduce their chances of being involved in a traffic accident causing an injury.

**Data Description**

The data set is a .csv file, the target variable to predict is an attribute with labeled data indicating the severity of the traffic crash as either property damage or injury. There is location indicating the latitude and longitude of the accidents as wells as collision type, the number of people involved, pedestrians, bicycles and vehicles involved as well as the date and time of the crash. The environmental factors could be important in predicting accident severity by including the road conditions for example was the road icy or wet or were the light conditions dark. There are environmental information available such as the weather, road and light conditions and if any of the drivers involved were under the influence.

There will be some cleaning and attribute selection required, for example the date attribute some have just the date while many others have the data and time. The time of the accidents could be an important factor so that attribute needs to be tested to confirm if including improves accident severity prediction. There are also entries that are blank in many attributes, which may have to be averaged, induced or removed. The size of the data set is 73 attributes with 194,673 rows.