**Capstone Project -**

**Car Accident Severity in Seattle City**

**Project Objective**

Driving is an inseparable part of our daily life - It brings convenience to us, but at the same time we are always exposed to the risk of potential car accidents every time when we drive on the road. To protect our self from running into accidents which cause severe damages and the loss of life (the worst), we should be better off if we can know in advance about the possibility and severity of a car accident - which might occur given the weather, road condition and other factors. Especially important, if we are visiting a city or place which we are not familiar with, if we can predict the severity of a car accident, then we will drive more safely to avoid the big accidents as much as possible.

**Dataset** **Brief**

A CSV file called “Data-Collisions (1)” is used for this project. This dataset is comprised of a list of 37 attributes which some of them will be used to train and test the machine learning model. This dataset is about “Car severity”, so the first column coloured in yellow called “severity” will be treated as the labeled data. This dataset has unbalanced labels because some of them have missing data, and it has both numerical and categorical types of data. Therefore, we should balance the data and do feature engineering for a machine leaning model with less biases and good predictability.

Among the 37 attributes, 8 of them are selected to train and test the model, including: location, road condition, weather condition, junction type, car speeding, number of people involved, light conditions, number of vehicles involved.