**Introduction/Business Problem**

The Seattle Department of Traffic along with the Seattle Police Department has made the collision data in the city of Seattle publicly available. It provides the severity label for each collision along with a series of other attributes that are relevant to each collision (e.g. ADDRTYPE (Address Type), JUNCTIONTYPE (Category of Junction), UNDERINFL (Under Influence or not) etc.) . The data set has a total of 194674 instances of raw data. While these are records of past collisions, this data can be used to build a predictive model to predict the severity of a collision. Upon building a successful prediction model, we can apply the multiple combinations of attributes onto the prediction model to have an idea of the severity that would result if a collision were to occur given that set of attributes. Having this prediction, relevant authorities like the SDOT or SPD can work to put in place preventive barriers in the collision area. These barriers would include police check points, speed humps, possible redesigning of roads, placing of traffic lights to slow traffic in regions prone to have higher severity levels. Furthermore, through visualization techniques such as Folium, we can see which areas in Seattle are highly prone to more collisions. Hence the business objective of this project will be to build a severity prediction model which allows the user such as SDOT or SPD, to do prior predictions of severity, given a particular combination of attributes so that prior measures can be taken to reduce the severity of such collisions and prevent such collisions if possible.