**Data Description :**

The data set to be used to build the model will be collected from the Seattle Traffic Management Division, Traffic Records Group.

Data has been provided in the form of comma separated version (csv) file “**Data-Collisions.csv**”

The data file has 38 columns with one column repeating hence effectively 37 columns and 65691 rows.

Each row provides the detail on one particular accident occurrence .

It has the details on the description and Severity of the Accidents happened on a particular day at a given Location and Junction. It has the details on the Weather condition , Light condition and the Road condition when the accident happened. It has record whether the vehicle was over speeding or not. It also has the details how many vehicles and people were involved or whether pedestrians were also involved. Data set also records kind of collision and whether the parked car was hit in the accident.

All these details have been recorded in the 37 columns.

We also refer to the columns as Features while building Machine Learning (ML) model.

Not all features will be used to build the model.

We will take the subset of all the columns (features) which would have more impact in predicting the Accident Severity . To find it we will visualize the data , refer the metadata , apply statistical methods to narrow down on the columns we need to build the model.

If needed we will convert the categorical data into numerical data and change the datatypes of the columns. We will then split the data into Training data set and the Testing data set.

We will build the Model using the Training data set and evaluate the model using the Testing data set.