**Data**

The data was collected by the Seattle traffic police department from 2004 to present.

The data is raw data set which has inconsistencies and imbalanced , which needs to be cleaned and processed.

The dependent variable is “SEVERITYCODE”, contains numbers that correspond to different levels of severity caused by an accident from 0 to 4.Severity codes are as follows:

0: Little to no Probability (Clear Conditions)

1: Very Low Probability — Chance or Property Damage

2: Low Probability — Chance of Injury

3: Mild Probability — Chance of Serious Injury

4: High Probability — Chance of Fatality

The data consists of 37 independent variables and records of 194,673 accident cases occurred in Seattle.

Not all 37 attributes might be used for drawing an outcome of severity, some of them are optional information for the model like –

* Date of the incident – which logically doesn’t effect the severity of the accident
* Codes used for collision – which is just an identifier given for each collision

Some of the attributes which may play a key role in modelling the outcome –

* Location- certain areas have heavy traffic usually and lets us know the possibility for occurrence of accident
* Weather condition - Bad weather conditions like during rainfall and snowfall , there are more chances of vehicle to be out of control
* Type of collision – It tells us ways in which collision occurs
* State of the driver – Speed driving or drink and drive can cause a fatality, as the vehicle is more uncontrollable by the driver
* Road condition – standing water, mud, dirt on roads have chances for property damage
* Light condition – Dark night without lights have chances of fatality
* Junction type – certain junction types are risky and there are high chances for injury