**Data**

The data was collected by the Seattle Police Department and Accident Traffic Records Department from 2004 to present.

The data consists of 37 independent variables and 194,673 rows. The dependent variable, “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

After analyzing the data set, I have decided to focus on only four features, severity, weather conditions, road conditions, and light conditions.

There are some null values in some columns and we need to fill those values using with certain methods like mean, median etc

There is imbalance in the column called severity code. So, we can resample the given data to get balanced.