Capstone Project
Car accident severity in Seattle city

By Sam Joseph

2. DATA

a. Data Source

To proceed with this project, we are using the data available from below repository
https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv
The dataset has information gathered on the road traffic accidents of Seattle City.
From the data extracted, the key attributes are:

a. Severity Code – The values are 1 & 2 which denotes property damage & human injury respectively. Code 2 is valued as more severe
b. Weather – Sample values: "Raining", "Clear", "Overcast", …
c. Road Condition – Sample values: "Wet", "Dry", "Ice", …
d. Light Condition – Sample values: "Daylight", "Dawn", …

b. Approach to solve the problem

There are many columns that we will not use for this model. After downloading the data we will convert the data in panda data frame and remove the columns that is not required. The initial dataset consists of 38 columns (features/attributes) and 194673 rows. The dataset will be cleaned according to the requirements of this project. The data will be analysed to identify the set of criteria on which high severity accidents happen based on the attributes evaluates.