PREDICTING SEVERITY OF ACCIDENT

1 Data Acquisition and Screening

1.1 Data Source:

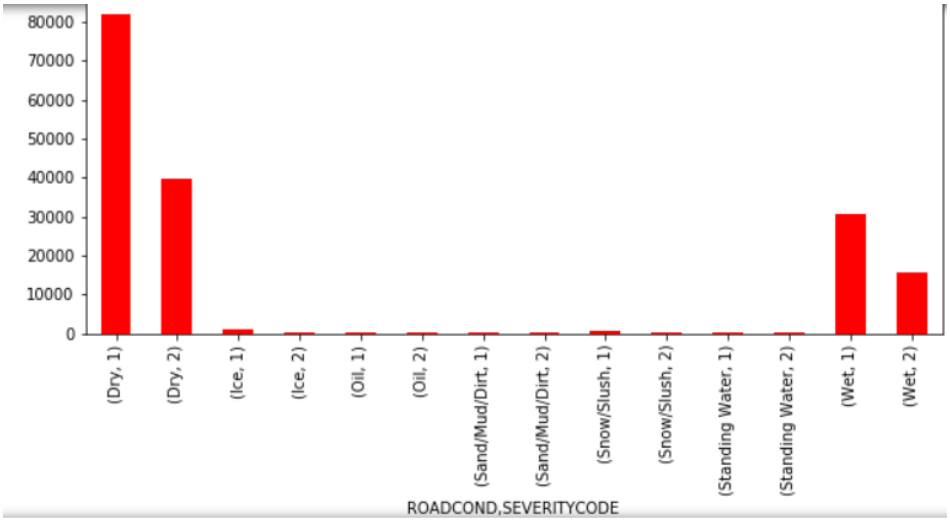
Data related to accidents and their severity are available through many online links but here we have chosen the data source that is shared by this course. I find it enough for my project, since it has lot of data against any feature.

1.2 Data Selection:

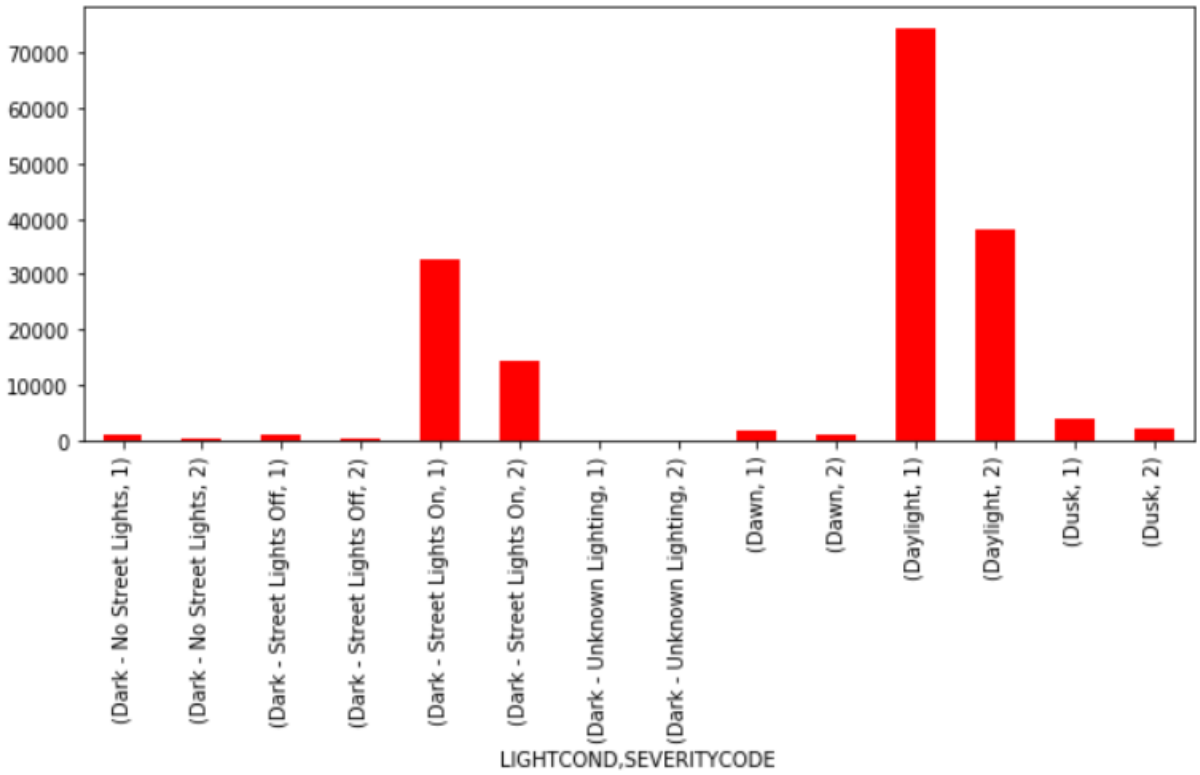
There is lot of information available in our data set. First, I focused the columns that are much related to provide us required result or which are easy to be segregated into binary information. I checked the values of incidents on those features and found reasonable high occurrence of our required label.

We have used SEVERITYCODE as our required label from our dataset. This column has two values, “1” and “2”. “1” means that there is an accident but no injury or just property damaged, while “2” means that injury occurred.

I used three features for the input to our model. They are: “ROADCOND”, which shows the condition of road, whether it is wet or dry. There is also another feature which has similar results, “WEATHER”. Means when there is rain, road condition will be wet and no rain means, either wet or dry.



I used “LIGHTCOND” as second input. Light condition shows some serious results where it is not good or dark with junction type feature.



Third input I used for modeling is “JUNCTIONTYPE”. This feature shows some reasonable values of severity at certain junctions.