**Introduction/Business Problem**

City/Town planners always have difficulty to determine where to put stop sign, stop light, extra precautions such as pedestrian crossings with extra indications(flashing lights, road signs or signals) for automotives /cyclist/pedestrians etc. Since Police collect detailed data of the accident locations and other various accident data information, can we use that data to identify the locations of severe accidents and weather is a factor in such accidents and then give a predictive tool for road safety planning. This project is to identify such locations from the accident data and identify how much a weather is a contributing factor in the severity of the accident.

**Data Source**

We will use the sample Seattle data set provided in the course for this problem solving. Some key information is necessary for this kind of problem prediction is to have latitude/longitude information or accident location information, weather conditions etc. are important. The data set provided has the required basic data for solving the problem.

The data set has the following features for data analysis, train and test the model.

|  |  |  |
| --- | --- | --- |
| Feature | Data Type | Comment |
| SEVERITYCODE | Int | Target |
| (X)Longitude | Float | Keep |
| (Y)Latitude | Float | Keep |
| OBJECTID | Int | Unique ID |
| INCKEY | Int | Drop |
| COLDETKEY | Int | Drop |
| REPORTNO | Int | Drop |
| STATUS | Text | Drop |
| ADDRTYPE | Text | Keep |
| INTKEY | Int | Evaluate |
| LOCATION | Text | Keep, Evaluate |
| EXCEPTRSNCODE | Text | Drop |
| EXCEPTRSNDESC | Text | Drop |
| SEVERITYDESC | Text | Keep, evaluate |
| COLLISIONTYPE | Text | Keep, evaluate |
| PERSONCOUNT | Int | Keep, evaluate |
| PEDCOUNT | Int | Keep, evaluate |
| PEDCYLCOUNT | Int | Keep, Evaluate |
| VEHCOUNT | Int | Keep, evaluate |
| INCDATE | Text | Keep, transform to Day of the week for analysis |
| INCDTTM | Text | Keep, & Transform to ranges |
| JUNCTIONTYPE | text | Keep, evaluate |
| SDOT\_COLCODE | Int | Drop |
| SDOT\_COLDESC | text | Drop |
| INATTENTIONIND | Text | Keep, evaluate |
| UNDERINFL | Text | Keep, Evaluate |
| WEATHER | Text | Keep |
| ROADCOND | Text | Keep |
| LIGHTCOND | Text | Keep |
| PEDROWNOTGRNT | Text | Keep, Evaluate |
| SDOTCOLNUM | Int | Drop |
| SPEEDING | Text | Keep, evaluate |
| ST\_COLCODE | Int | Drop |
| ST\_COLDESC | Int | Drop |
| SEGLANEKEY | Int | Drop |
| CROSSWALKKEY | Int | Drop |
| HITPARKEDCAR | Text | Drop |

Since our intention is to identify the severity of the collision, we will use **“severitycode”** as our target feature. Using Data preparatory analysis, we will reduce the number of features needed for this analysis.

Based on the Metadata given for the data set, some features can be dropped as they are mostly informational for the authorities for their tracking. These were indicated next to feature as **Drop.**

Other features will be evaluated based on information.

Since this is large data set, we will use only the most recent 3 years data for the analysis.

Further after data exploratory analysis, we will further determine which features to be dropped.

If any feature has many blanks then that feature will be dropped.