**IBM Capstone Project – Accident severity prediction**

## **Data Collection and Preprocessing**

The data was collected from Kaggle [and](https://www.kaggle.com/sobhanmoosavi/us-accidents) Soban Moosavi [[24].](https://smoosavi.org/datasets/us_accidents) The table below shows the original list of predictors used in the obtained data set. Predictors were filtered since most were redundant to the response ”Severity”. Moreover, 243 rows of NA values were removed as they contributed to less than 0.5% of the final data. The highlighted cells in the table shows the filtered list of final predictors used in modelling.

**Red**-Response Variable

**Purple**- Predictors used along with response

**Green**-Used to plot accident point in Exploratory Data Analysis

Table 1: List of Original Predictors.

|  |  |
| --- | --- |
| Column Name with Description | |
| ID | Unique identifier of the accident record |
| Source | Indicates source of the accident report |
| TMC | Traffic accident may have a Traffic Message Channel |
| Severity | Severity number between 1 and 4 |
| Start Time | Start time of the accident in local time zone |
| End Time | End time of the accident in local time zone |
| Start Lat | Latitude in GPS coordinate of the start point |

Table 1: List of Original Predictors.

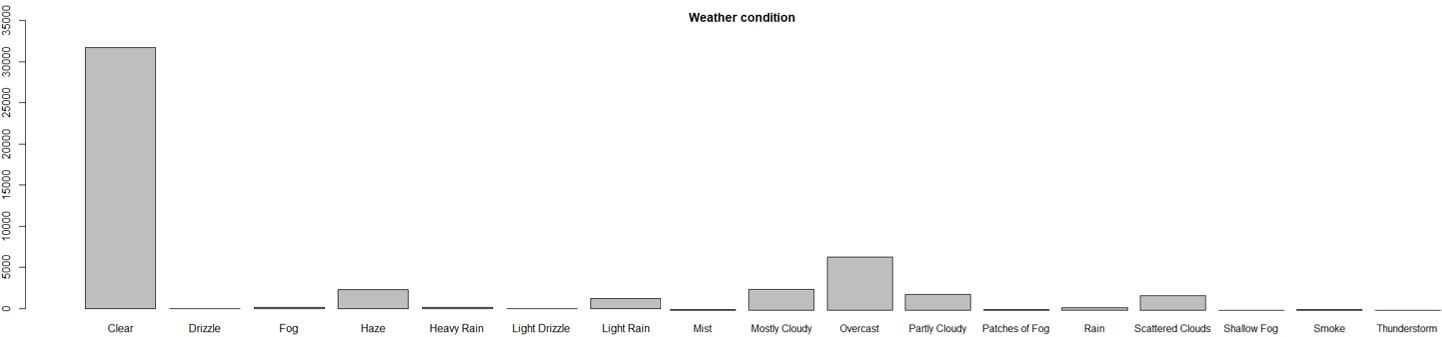
|  |  |
| --- | --- |
| Column Name with Description | |
| Start Lng | Longitude in GPS coordinate of the start point |
| End Lat | Latitude in GPS coordinate of the end point |
| End Lng | Longitude in GPS coordinate of the end point |
| Distance(mi) | The length of the road affected |
| Description | natural language description of the accident |
| Number | Street number in address field |
| Street | Street name in address field |
| Side | Relative side of the street |
| City | City in address field |
| County | County in address field |
| State | State in address field |
| Zipcode | Zipcode in address field |
| Country | Country in address field |
| TimeZone | Timezone based on the location |
| Airport Code | Denotes an airport-based weather station |
| Weather Timestamp | Time-stamp of weather observation |
| Temperature(F) | Temperature (in Fahrenheit) |
| Wind Chill(F) | Wind chill (in Fahrenheit) |
| Humidity | Humidity (in percentage) |
| Pressure(in) | Pressure (in inches) |
| Visibility(mi) | Visibility (in miles) |
| Wind Direction | Wind direction |
| Wind Speed(mph) | Wind Speed (in miles per hour) |
| Precipitation(in) | precipitation (in inches) |
| Weather Condition | Weather condition (rain, snow, thunderstorm, fog, etc.) |
| Amenity | Indicates presence of amenity |
| Bump | Indicates presence of speed bump |
| Crossing | Indicates presence of Crossing |
| Give Way | Indicates presence of Give Way |
| Junction | Indicates presence of Junction |
| No Exit | Indicates presence of No Exit |
| Railway | Indicates presence of Railway |
| Roundabout | Indicates presence of Roundabout |
| Station | Indicates presence of Station |
| Stop | Indicates presence of Stop |
| Traffic Calming | Indicates presence of Traffic Calming |

List of Original Predictors.

|  |  |
| --- | --- |
| Column Name with Description | |
| Traffic Signal | Indicates presence of Traffic Signal |
| Turning Loop | Indicates presence of Turning Loop |
| Sunrise Sunset | Period of day based on sunrise/sunset |
| Civil Twilight | Period of day based on Civil Twilight |
| Nautical Twilight | Period of day based on Nautical Twilight |
| Astronomical Twilight | Period of day based on Astronomical Twilight |

## Grouping similar Weather Conditions

The initial list of weather condition had 17 levels ranging from clear to thunderstorms and rain.



Spread of initial Weather Conditions

Similar weather conditions were grouped together to provide easy interpretability with 4 standard levels.

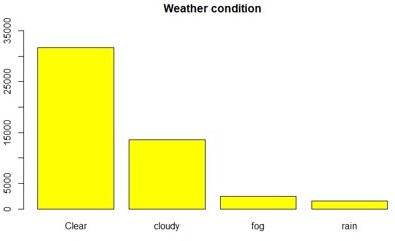
The grouping is as follows:

1.Clear

2.Rain - Drizzle, Heavy Rain, Light Drizzle, Light Rain, Thunderstorms and Rain

3.Cloudy - Mostly Cloudy, Overcast, Partly Cloudy, Scattered Clouds

4.Fog - Smoke, Fog, Haze,Mist, Patches of Fog, Shallow Fog



Grouped Weather Conditions