**IBM Data Science Capstone:**

Car Accident Severity Report

* Background:

Imagine you have to travel for a Business meeting and you take your car off and start driving. The weather is not good leading to poor visibility conditions. On the way, you come across a huge traffic - cars lined behind one another and there is hardly any chance for the traffic to move swiftly.

You come to know that a car has met with a terrible accident being the cause for the traffic jam and the passengers are severely injured.

We come across such situations or read about it almost every day thinking of a solution where we can avoid such accidents or receive an alert of the severity forehand.

* Introduction | Business Understanding:

In an effort to reduce the frequency of car collisions in a community, an algorithm must be developed to predict the severity of an accident given the current weather, road and visibility conditions. When conditions are bad, this model will alert drivers to remind them to be more careful predicting the severity.

* Data Understanding | Data Section:

Our predictor or target(dependent) variable will be 'SEVERITYCODE' as per the dataset since it is the measure of the severity of an accident ranged from 0 to 4. The attributes highly contributing to the severity of an accident in our dataset are 'WEATHER', 'ROADCOND', 'LIGHTCOND' (weather, road conditions and light conditions) represented as independent variables.

The severity code representation is as follows:

0 - unknown

1 - prop damage or property damage

2 - injury

2b - serious injury

3 - fatality

* **Feature Engineering:**

By looking at our dataset we come to know that our data is not fit for analysis. We will need to remove some columns which will not serve any purpose for our model and also we need to balance and normalize our dataset.

Also, some columns are of type object which are need to be converted to numeric type for the ease in our model building.