1. **A description of the data and how it will be used to solve the problem. (15 marks)**

I am using the provided Data-Collisions.csv dataset. The dataset has an initial total of 194,673 observations characterized by 38 different features. The data set contains data that might contribute to determining the severity of an accident which include: the number of vehicles involved, location where the accident occurred, the weather condition, light condition, road condition, junction, number of people etc.

Our goal is to predict the severity of an accident. , in the dataset the target variable is called “SEVERITYDESC“ And It takes two levels i.e. injury collisions and property damage only collision represented by codes 1& 2 respectively. To achieve our goal, we will go through the following steps;

* Feature exploration( with data cleaning)
* Dimensionality reduction
* Model building
* Optimization and final model selection