**Data Science Capstone Project**

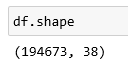
*Description of the Data:*

In this project I am going to deal with a data set called ‘Data-Collisions’ which is take from

<https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv>

This data consists of ‘Car Collisions’ details for different conditions of weather, road, light conditions etc. in a community.

The data we are going to use in this project is a huge data.



The data has to be analysed properly and the missing values, misleading values has to be filled. Feature Engineering must be carried out on the data to identify the proper variables which affect the SEVERITY.

As the Severity i.e. the target variable is a categorical variable the best way is to find the features which are highly correlated to target variable. As data is very large considerable data can be used as test data.