# Coursera Applied Data Science Capstone project

## 1. Introduction:

When we think of deaths happening around the world for a wide number of reasons, we come across a hard fact that road accidents contribute to a major number and it's surprising to see that no technology is able to eradicate this problem.

There can be a wide number of reasons for this cause. The severity can vary from case to case but the level of physical, mental and economic damage that road accidents bring to the people and the government is quite alarming.

### 1.1 Business Problem:

In order to reduce the damage or loss of life, there has to be an intelligent system or an accurate prediction system that can predict the severity of an accident based on a certain number of factors.
With such an accurate model in hand, the rescue team can perform operations accordingly by prioritizing the highly severe ones first and then attending the least severe ones subsequently.

## 2. Data Narrative:

The source for the data is:
https://www.kaggle.com/sobhanmoosavi/us-accidents

### 2.1 Data description:

- This specific dataset contains car accidents data over 49 states in the USA. The data was curated through 2016 to 2020.
- Data was scraped from open source API's of the US state transportation and law enforcement agencies.
- There are 3.9 million records(rows) in the dataset with 49 features.
- All the features may or maynot be substantially useful in building a model for our business use case.
- A level of data understanding and preprocessing is required to be carried out to filter the most effective features for our prediction model we are trying to build.