Capstone Project - Car accident severity (Week 1)

**Introduction:**

In this capstone I will explore and analyze a dataset collected about traffic accidents across the UK. The UK government collects and publishes (usually on an annual basis) detailed information about traffic accidents across the country. This information includes, but is not limited to, geographical locations, weather conditions, type of vehicles, number of casualties and vehicle manoeuvres, making this a very interesting and comprehensive dataset for analysis and research.

The data for this project is available on Kaggle as UK Road Safety: Traffic Accidents and Vehicles (<https://www.kaggle.com/tsiaras/uk-road-safety-accidents-and-vehicles>)

The traffic accidents have a countless effect on the society due to mortality rate and damages. The traffic accidents are one of the leading causes of death due to injury. In this capstone attention will be paid to identify factors affecting the road accident and severity of these factors. Applying machine learning model traffic accident data records can help to understand the characteristics of driver's behaviour, roadway condition and weather condition that were causally connected with different injury severity.

**Project Objective:**

The objective of our project is as follows:

* Find the factors affecting the road accident severity in UK.
* Through visualizations and machine learning algorithms, build a model to predict the seriousness of road accident based on Weather Conditions and other important variables as accurately as possible.

**Data Description:**

* The data for this project is obtained from a user on Kaggle and was composed from information on the United Kingdom’s Government Open Data website.
* It consists of two different datasets that contain information from 2005 2017 that were combined on a common field (accident\_index).
  + Vehicle\_Information.csv: A file containing information about the vehicles, point of impact, maneuvers made, driver information, etc.
  + Accident\_Information.csv: A file containing details about the accident that include location, junction details, date

**Data understanding:**

In this phase, dataset will be collected and extracted from a source’s csv file. Then, attributes (columns) will be seclected that will use to train your machine learning model. Also, assess the condition of chosen attributes by looking for trends, certain patterns, skewed information, correlations, and so on.

**Data Preparation:**

The data preparation includes all the required activities to construct the final dataset which will be fed into the modeling tools. Data preparation can be performed multiple times and it includes balancing the labeled data, transformation, filling missing data, and cleaning the dataset.