Capstone Project - Car accident severity

**A.1. Description & Discussion of the Background**

Road Accident is the most undesirable and unexpected thing to occur to a road user, though they happen quite often. Unfortunately, we can see a minatory rise of road accidents in United Kingdom, conspicuously highroad accidents over the past few years. It has a massive impact on society as well as in the economy of our country as there is an immense cost of fatalities and injuries. According to a recent report, annually on an average 9,000 lives have been taken by road accidents and lead to almost 28,000 injuries. This record indicates that every day, approximately 28 people were killed by road accidents and it is quite devastating. Besides this, according to WHO, the economic cost of road accidents to a developing country like us is 2-3% of GDP, which is a significant loss for a country like ours. Moreover, reducing this loss has become a great matter of concern for our country now.

**A.2. Data Description**

For the accurate prediction of the severity of accidents, a considerable number of traffic accident records with full information is required to train by using the proposed approaches. In this research work, the authors have collected a dataset from the Traffic Bureau that consists of total 37,885 traffic accidents record from the year 2007-2017. The entire dataset will split into two parts- Training Dataset and Test Dataset. 70% of the whole dataset has been chosen randomly by using a python library as a training data set and the remaining 30% has been used as our test dataset. We have used the 70-30 ratio for splitting dataset because of its proven accuracy.

**Audience**

People who travel on a daily basis or even if someone new commuting to the highway will have an idea of the road condition and how time, weather and other scenarios while driving may cause collision. They will get awareness to take precautions

**Benefits**

it will reduce the number of collision since they will have better idea of what to expect

**Data**

The data we will be using is the collision data from the example, as well as the pdf describing the data, for reference

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

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

**Methodology**

#### Initial data will be cleaned in Excel and then using Python for additional cleaning, analyzing and visualizing state traffic accident data. The idea is to extract data to evaluate its finding, combine factors leading to collision together, and building linear regression models to evaluate the chances of collision.

To do this, we will be using the following modules/libraries

* Pandas
* Numpy
* Folium
* Geocode
* Matplotlib
* Scikit-learn