**Statistics for Business Analytics (BANA-6610)**

**Final Project Proposal**

**Created on: 25th November 2021**

**Project Topic: Accidents in USA, a Countrywide Traffic Accident analysis.**

**Division of work:**

I will be working on cleaning of dataset and generating EDA for the data which consists of 1,516,064 rows and 47 columns.

I will be generating multilinear regression models, which will tell us what factors affect the severity of accidents on roads.

I will be working on database for getting the following outcomes:

1. The state that has the highest number of accidents.
2. At what time do accidents usually occur in the US
3. Factors Affecting Accident Severity
4. Predicting the location of the accident
5. Predicting the accident prone zone in each state.
6. Generating the EDA for the data.
7. Predicting the number of accidents might occur in next years.

I will also work on developing logistic regression model for columns which have values true and false.

This Analysis will be useful to the Road safety department of the government, hospitals and response teams as they will be prepared for such accidents and will be knowing the measures to take beforehand the accident occur in an accident prone zone ( after analysis ). Will tell us what step must be taken to avoid such accidents. Will be recommending the measures to take.

**Methodologies we will be using:**

1. Regression models (linear/ multilinear/ logistic)
2. Time series forecasting
3. Various libraries for less coding.
4. Data visualisation with great UI
5. Line of best fit.
6. K-means clustering.
7. Exploratory Data Analysis