Applied Data Science - Capstone

# **Introduction/Business Problem**

The case study is predicting the severity of the accident. The Outcome of the case study is mainly targeted for Traffic departments, Travellers & Tourist, local residents and others those who use the roadway service for the day to day work.

This will would greatly help the Traffic to monitor the traffic due to accident and take necessary to action to control for hassle free transportation.

For the travellers & tourist, Local residents will be aware of severity and take necessary actions to be safe and avoid accidents.

# Data Section

The Data Set is the Collision details happened over the 2004 to present. It holds all the attributes to the

* Key – Unique Identification
* Location of Collision
* Address Type – Alley, Block, Intersection
* Collision type
* Weather, Road Conditions, Light Conditions
* Fatalities
* No of Persons injured,
* No of pedestrians injured

These attributes will guide in predicting the Target variable the Severity of the Accident.

As the Target variable provided It’s a supervised algorithm with Categorical type (1,2). Also the data set is highly imbalance with below details so we need to balance the dataset before training the models.

1. 136485
2. 58188

There are total of 38 attributes we need to understand the correlation with the target variable and identify the necessary attributes to train the model.

**Models Used:**

* Multinominal NaïveBayes
* Gradient Boosting Classification
* Random Forest Classifier