**Predict Accident Severity**

**Introduction/ Business Problem**

Road traffic accidents have a high cost of loss of life and economy of a city, state or country. Accidents may cause one or the other negative impacts in terms of property damage, medical treatment costs, insurance cost, legal issues etc. Identification of factors causing road accidents by studying past trends are of importance as it would provide guidance in setting up preventive measures and issuing guidelines to drivers/ general public which can benefit in reducing the occurrence of road accidents.

In this report, study is conducted on the dataset provided by Seattle Department of Transportation Traffic Management Division. The aim is to find the correlation between weather, location and road condition to predict severity of road accidents. In this report machine learning models are applied on past accident dataset to predict the accident severity which is divided into two levels 1-Property damage only collision and 2-Injury collision.

Overall, output of this study would help in better town-planning (by installation of traffic light, street lights etc at critical junctions/ roads) as well as reduce loss of life by setting up critical levels of emergency services responders across the city of Seattle.

**Data Understanding**

The dataset provided by the Seattle Department of Transportation Traffic Management Division covers accidents in from Janauary’2004 to May’2020. Overall dataset covers 194,673 accidents capturing 37 features and each accident is given a severity code of 1 or 2 by the authorities. Dataset is unbalanced as 70% of accidents are marked with severity code = 1 (property damage only collision) while remaining 30% of accidents are marked with severity code = 2 (injuries collision). This would require balancing of dataset to achieve accurate predictive model.

While the dataset is unbalanced it covers key features such as location of accident, weather conditions, road conditions, whether driver was under influence, whether accident was due to inattention etc. All these features/ parameters would play a pivotal role in predicting accident severity by using machine learning models. However, dataset comprises of features which have missing or unknown values. All such accidents/ data points would be excluded as part of data preparation for predictive modeling.