**Introduction:**

The problem statement is to identify the probability and severity of an accident on a particular day and in a particular weather. Various factors like weather, time, day, road conditions etc will be analysed to predict what type of accidents can occur during that time. Moreover, severity of an accident will also be predicted.

**Target Audience:** This problem will be of interest to residents of Seattle and people who are travelling to/from Seattle. Solution to this problem would help these people to travel safely. It would help people in understanding the risks of travelling on a particular day depending upon the weather conditions. For ex: If a person who wants to travel to Seattle on a rainy day at night, he/she can check chances of an accident and its severity. Consequently, making the decision of travelling or not.

**Data**:

“Collisions—All Years” dataset was used for this problem. Metadata of this dataset can be found on this link : <https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Metadata.pdf>

Dataset can be downloaded from this link : <https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv>

This data set has around 200k observations and 37 attributes. SEVERITYCODE of the accident is used as a label and this will be used in the end to check the severity of the accident that can occur given other conditions.

**Understating data:** Some attributes in the dataset like LOCATION, EXCEPTRSNCODE, EXCEPTRSNDESC, SEVERITYCODE, PEDROWNOTGRNT or SPEEDING cannot be used directly. Some of these columns have missing values that need to addressed properly and some columns like REPORTNO and LOCATION cannot be used entirely.

Moreover, few columns like LIGHTCOND, ROADCOND, JUNCTIONTYPE, ADDRTYPE, WEATHER will have to be hot encoded for proper data modelling.

Columns like OBJECTID, INCKEY, COLDETKEY, REPORTNO cannot be used in model creation as these columns do not have the information required to predict accident severity.