1. Data

* All collisions provided by SPD and recorded by Traffic Records. This includes all types of collisions.
* Collisions will display at the intersection or mid-block of a segment
* Timeframe: 2004 to Present

1. How will it be used to solve the probelm Use the offered metadata on accidents and severity in this course to execute the ML model. It is widely available on the Canadian government website and can also be found from Week 1 link in the course. In order to implement the solution, the following steps will be executed:

* Data Pre-processing
* Clean the data to remove excess columns
* Replace NaN data values, or clear empty columns.
* Ensure all data types of each data is correct
* Create multiple regression to see which variables most impact the accidence severity code
* Use KNN or Decision Tree or Logistic Regression to classify new conditions as either 1 or 2 in severity code. As an example, ROADCOND is likely to be important. When road conditions are wet, the risk of accident increases to 2. So I will analyze it as a part of classification process while also doing accuracy evaluations depending on the machine learning technique deployed. If it is a decision tree, I will compute entropy
* In order to determine which ML algorithm will give us the best output, use evaluation metrics Jaccard, confusion matrix and F1 Score to determine the accuracy