**PREDICTING ROAD ACCIDENT SEVERITY**

For the IBM Data Science Capstone, an analysis on all collisions provided by SPD and recorded by Traffic Records from 2004 to 2020. The original dataset is taken from SDOT Traffic Management Division, Traffic Records Group. The dataset has 38 columns consisting of 194673 observations without any duplicates. This dataset is a hybrid combination of numerical and categorical data types.

Road accidents is a major problem in the modern world – it is said more people are killed by road accidents than plane crashes. The objective of this analysis to investigate which are the factors driving the severity of a collisions and provide insights to the community on possible to reduce road accidents.

First, I will try to deal with the imbalanced dataset so that the accuracy of the analysis will not be affected by imbalanced dataset. After that, I will perform feature engineering to find the factors significant to severity of collisions. Finally, I will use several classification methods on the dataset and choose the best machine learning methods yielding the best output.