**Description:**

The selected data provided by SPD and recorded by Traffic Records includes all types of collisions from 2004 to present. This is obtained through the organization SDOT Traffic Management Division, Traffic Records Group.

An ideal data to solve this kind of problems should contain adequate number of attributes and observations. The selected data contains about 37 attributes out of which few are unbalanced labels which should be balanced in order to create an unbiased Machine learning model.

Examining the data, I believe that a supervised Machine learning algorithm (a classification algorithm to be more specific) can yield good results for this problem.

**Explaining the data:**

The first column named SEVERITY is the labelled data indicating the fatality of an accident, Which states weather the accident is severe or not. The remaining columns have different types of attributes. Out of which few or all can be used to train the model.

Selection of attributes to train the model is one of the most important and initial steps while building a model. Correctly selected attributes i.e. the attributes which effect the predicted output significantly should be identified and weighted accurately. In the similar way the attributes which does not have a significant effect on the output should be dropped to reduce the complexity of the model. The target is to achieve a generalised model with maximum accuracy.

The attributes like LOCATION, WEATHER, LIGHTCOND are some of the examples for good features that can be used to build the model.