

## Description of the data and how it will be used to solve the problem

The data that are used by the project have been generalized by the Traffic Records Group of the SDOT Traffic Management Division. They involve information related to the severity of the accident (ranked from 1 to 5) and information related to objective dimensions of the accident such as time, place as well as light, road and weather conditions.

After a preliminary exploration of the database, it has been found that it involves only two categories for the severity column which are the following:

*2:injury*

*1:property damage*

An exploratory analysis will take place in order to decide which data are more relevant and adequate to perform a solid analysis. The exploratory analysis will involve the generalization of bar graphs and maps.

After the independent variables are selected, an implementation of prediction models such as Decision Tree and Logistic Regression will follow and finally the implemented models will be evaluated and compared through metrics such as the Jaccard similarity score and the f1 score.