

# Capstone Project - Car accident severity (Week 1)

## Data Section:

All collisions are provided by SPD and recorded by Traffic Records. This includes all types of collisions happening in Seattle City, which was updated weekly. Collisions will display at the intersection or mid-block of a segment.

Timeframe: 2004 to Present.

Contact Person: SDOT GIS Analyst

Contact Organization: SDOT Traffic Management Division, Traffic Records Group

Contact email: DOT\_IT\_GIS@seattle.gov

## Data Information:

- 1) The dataset includes 37 attributes. Some of them are much related to the severity of an accident, such as weather condition, car speeding and light conditions. Note that not all the attributes are useful.
- 2) There are many empty inputs in the accident data. For instance, the column "ROADCOND" includes 5012 empty inputs. There also exist both numerical and categorical types of data, which leads to the utilization of feature engineering.
- 3) The dataset has labels so that supervised machine learning technologies are suitable to predict the severity of an accident; nevertheless, the dataset is not label balanced. We should balance the data firstly before we create a model.