

# Car Accident Severity Report

---

## • Understanding the Problem

The Seattle government is going to prevent avoidable car accidents by employing methods that alert drivers, health system, and police to remind them to be more careful in critical situations. In most cases, not paying enough attention during driving, abusing drugs and alcohol or driving at very high speed are the main causes of occurring accidents that can be prevented by enacting harsher regulations. Besides the aforementioned reasons, weather, visibility, or road conditions are the major uncontrollable factors that can be prevented by revealing hidden patterns in the data and announcing warning to the local government, police and drivers on the targeted roads.

The target audience of the project is local Seattle government, police, rescue groups, and last but not least, car insurance institutes. The model and its results are going to provide some advice for the target audience to make insightful decisions for reducing the number of accidents and injuries for the city.

## • Data

The data was collected by the Seattle Police Department and Accident Traffic Records Department from 2004 to present.

The data consists of 37 independent variables and 194,673 rows. The dependent variable, "SEVERITYCODE", contains numbers that correspond to different levels of severity caused by an accident from 0 to 4.

Severity codes are as follows:

- 0: Little to no Probability (Clear Conditions)
- 1: Very Low Probability - Chance of Property Damage
- 2: Low Probability - Chance of Injury
- 3: Mild Probability - Chance of Serious Injury
- 4: High Probability - Chance of Fatality

Furthermore, because of the existence of null values in some records, the data needs to be preprocessed before any further processing.

## • Data Preprocessing

The dataset in the original form is not ready for data analysis. First we will check the datatype of every column, then we will drop the columns which are not required in the further process. Also most of the attributes are of object datatype, we will convert them into numeric datatypes. After analysing the data, I have decided to use these four attributes for this problem ( weather, severitycode, roadcond, lightcond ).

After understanding the data I found out that the target feature is imbalanced in this dataset. We will use simple statistical techniques to balance it.