

# Car Accident Severity

## Business Understanding

In order to prevent car accidents from happening in Seattle, relevant authorities may be interested in using the data, such as road and light condition and the weather, to predict the likelihood of a car accident in a given situation. Historical data has been collected over a period of time, indicating factors which contribute to the occurrence of car accidents. Each accident leads to different levels of severity. The report can be used to alert drivers to be especially mindful while driving and to notify relevant authorities to take action to reduce accident rates by improving road conditions and/or implementing new signs along the road.

## Data Understanding

All collisions in the dataset are provided by SPD and recorded by Traffic Records. It consists of 38 columns and 194673 rows in the CSV file. For the purpose of this report, 'SEVERITYCODE' will be used as a dependent/target variable, while attributes or the independent variables include 'WEATHER', 'ROADCOND', 'LIGHTCOND'.

## Methodology

The available dataset will have to be preprocessed and cleaned first. Methods involve the use of KNN, Decision Tree, and Regression.

## Results & Discussion

From the dataset, there are various ways the data can be interpreted and used. The raw data cannot be used to compare in Python because they are mostly objects, or at least the columns chosen to be used in this report. The object type is converted into the “categorical” type, so that a number can be assigned to each value for calculation. Class “1” of “SEVERITYCODE” has been downsized as to be equal to Class “2”, rendering the pair suitable for comparison.

Then KNN, Decision Tree and Log Regression are applied to the cleaned dataset. The results are shown below.

Model	Jaccard Score	F1 Score
KNN	0.52	0.54
Decision Tree	0.56	0.54
Log Regression	0.52	0.51

## Conclusion

Based on the analysis, it suggests that certain conditions of the road, the light, and the weather may serve as contributing factors to different levels of damage or severity.