

## **Data**

The data that will be used to conduct this analysis is the compiled collision dataset from 2004 to Present within Seattle, Washington. This data source includes 194,673 rows and 38 columns was taken from the Seattle Department of Transportation that is continuously updated weekly. This will include the severity of each accident, the type of vehicle involved if any, the location of where the collision took place, as well as weather conditions that may have had an adverse effects on the event.

Specifically, weather conditions, light conditions, and types of junctions can collectively be used to determine the most dangerous parts of the road where severe accidents are most likely to occur. Most importantly, it contains a severity code that ranges from 0 (unknown) to 3 (fatality) that will be the main focus of this analysis. Being able to use the various features within the dataset to better predict this the level of severity of the collision can allow proper safety features to be placed in certain junctions where the probability is higher.

## **Methodology**

I performed several analyses including:

- Created a visual heat map of all of the Nan values present within the database to see how much is missing and what pieces of the data is useful.
- Used bar graphs to visualize the difference between the relationships of various factors to the severity of the corresponding collisions.
- Gave each categorical value a numerical placeholder to facilitate the function of further analysis.
- Devised a regression analysis to predict the severity of future collisions based on the given data.