# Car Accident Severity

DATA

The dataset used for this project is focused on automobile accidents that took place from 2004 to 2020 within the city of Seattle, Washington. This knowledge refers to the seriousness of each car accident, as well as the time and circumstances in which each accident took place. The model is intended to predict the seriousness of an accident, taking into account that the Seriousness Code variable was encoded in the form of 1 (Property Damage Only) and 2 (Physical Injury) in the form of 0 (Property Damage Only) and 1 (Physical Injury). After that, 0 was allocated to the element of each variable that could be the least likely cause of a serious accident, while a high number indicated an adverse condition that could lead to a higher severity of the accident. Whereas, for every attribute that was either Other or Unknown, there were specific values, removing those rows completely would have contributed to a lot of data loss that is not preferred.

Number of entries in data for each variable - Seattle, Washington
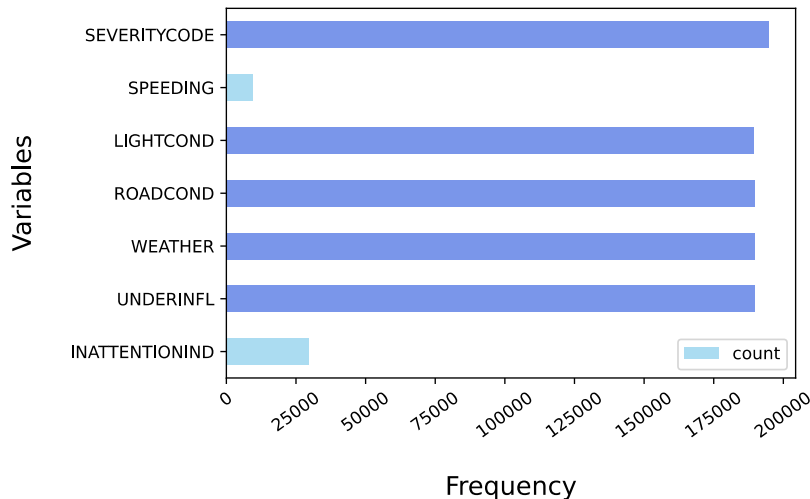
In order to deal with the issue of columns having a variation in frequency, arrays were made for each column which were encoded according to the original column and had equal proportion of elements as the original column. Then the arrays were imposed on the original columns in the positions which had *Other* and *Unknown* in them. This entire process of cleaning data led to a loss of almost 5000 rows which had redundant data, whereas other rows with unknown values were filled earlier.

Feature Selection

Feature Variables	Description
INATTENTIONIND	Whether or not the driver was inattentive (Y/N)
UNDERINFL	Whether or not the driver was under influence (Y/N)
WEATHER	Weather condition during the time of collision (Overcast/Rain/Clear)
ROADCOND	Road condition during the collision (Wet/Dry)
LIGHTCOND	Light conditions during the collision (Lights on/ Dark with light on)
SPEEDING	Whether the car wasabove the speed limit at the time of collision (Y/N)