

Data

The data used in this capstone project is the collisions data collected from Seattle, US. It contains all the collisions happened since 2004 to present. The data set include 194673 rows and 38 columns. The attributes mainly include the location, severity, collision type, number of pedestrians, bicycles, and vehicles involved, injuries and fatalities, time, weather, road condition, light condition, and speed. The information is too much, and therefore, we should select only some of them which are highly related to the prediction.