**Introduction/ Business Problem**

Traffic accidents cause fatalities and economic losses worldwide every year. Therefore, from the perspective of loss prevention, this is one of the main areas of ​​social concern. According to the preliminary estimates released from the National Safety Council, approximately 38,800 people lost their lives to car crashes in 2019. Around 4.4 million people were injured seriously enough to require medical attention in crashes in 2019. Although motor vehicle fatality and casualty has declined for the second consecutive year, the number of deaths is still unacceptable. Today, the majority of new vehicle models from manufacturers includes many advanced driver assistance and safety technology, such as automatic emergency braking, lane departure warning systems, driver monitoring system and so on, all of which are proven to be effective on reducing the severity of crashes or accidents. However, it is still necessary to study characteristics of the accident. Therefore, establishing an accident severity prediction model and improving the model are the keys to improve the safety performance of the road traffic system.

**Data Understanding**

In the accident severity modeling, the input vector is the characteristics of the accident, such as the attributes of driver behavior like speeding, and environmental characteristics like weather and road condition, and the output vector is the category corresponding to the severity of the accident.

A comprehensive dataset of 194,673 accidents occurring between 2004 to 2020 in Seattle was obtained for this analysis. the dataset consists of 39 columns describing the details of each accidents. The dependent variable, “SEVERITYCODE”, contains numbers that correspond to different severity level caused by the accident. “1” indicates property damage only collision, and “2” indicates injury collision. The existing data is not ready for data analysis; therefore it needs to be preprocessed before any further interpretation.

**Data Cleaning and Data Preprocessing**