

# **IBM Data Science Capstone Project**

## **Car accident severity**

### **Background**

IBM applied data science specialization capstone project to analyse Seattle city accidents data. Auto crashes can happen constantly, but there are a few conditions where the probabilities of a mishap emerge due different factors. To complete capstone, its required to working on a case study which is to predict the severity of an accident based on the available dataset.

This report has as reason build up a model for Seattle government to foresee the probabilities of accidents seriousness, in view of various conditions as climate or street conditions.

The idea behind the project to have a data model in place to warn people, given the weather and the road conditions about the possibility of getting into a car accident and how severe it would be.

The data was given via Seattle Police Department from 2004 to 2020.

### **Problem Statement**

Recognize the conditions that can cause future accidents in Seattle city as an example so as to caution the individuals with expectation to know and drive more cautiously.

- Accidents predictability
- The keep away from of mishaps will bring about products benefits:
- Spare lives as principle advantage
- Lessen costs in harm framework
- Lessen cost from police and paramedics to go to every mishap

### **Data**

The data originates from Seattle Police Department and recorded by Traffic Records and incorporate Collisions at convergence or mid-square of a section. The period data is from 2004 to May 2020.

The data is composed in a CSV record with 37 credits and initially 194673 lines, the data is named and lopsided. Moreover an archive with the depiction of every segment were given. Also we need to check what type of data we have.

Due our data is named we know the outcome for each record, we select SEVERITYCODE as Dependent variable. The potential qualities are:

**0 — unknown**

**1 — prop damage**

**2 — injury**

**2b — serious injury**

**3 — fatality**

The information is unbalanced by the difference in samples for each accident type. In our case there are only two types of accidents from the five possible options.