# BUSINESS PROBLEM

This article outlines the procedures applied to analyse data on car accident severity. Globally, a car accident is a huge challenge. The study of (Deme, 2019) stipulates that road traffic deaths are the eighth leading cause of death more than tuberculosis and HIV/AIDS.

According to (Martinez, Sanchez, and Yañez-Pagans, 2019), an estimated 1.3 million people die in traffic accidents each year worldwide, and millions more are injured, with developing countries excessively affected. The number of global traffic deaths will be around 1.8 million annually by 2030, making it the eighth cause of death worldwide.

In the African Continent, South Africa has a high amount of road accidents than other African countries. The results (Ncube, Cheteni, and Sindiyandiya, 2016) suggest that drunken driving, potholes, and not using seat belts are determinants in the number of road related fatalities. Figure 1 below demonstrates some of the car accident causes on the South African roads.

Figure 1: Cause of accidents in South Africa roads

## Problem statement

A road accident is a fatality that institutes a significant cause of disability and untimely death. Therefore, methods to reduce accident severity are of great interest to traffic agencies and the public (Deb, R. and Liew, A.W.C., 2015). The following questions asked to understand the factors which cause car accidents:

* Does speeding cause accidents?
* Can road conditions cause accidents?

## Targeted audience

The assignment on car accident severity can contribute immensely to the body of knowledge since the study's finding can be used as a reference when conducting research. The target audience of the project is academics and non-academic researchers. Academics can use the information to educate learners. The private sector can use the outcome of car severity accidents to avoid driving on roads with high-risk areas.

Data is used by many insurance companies to help determine their pricing of premiums, underwrite policies, and to structure deductibles. Most information is used to assist companies in understanding the client's behaviours as well as client needs. Nowadays, some car insurance companies come with data applications that can be downloaded after an accident to give evidence on what happened, which resulted in an accident. They can even tell whether the driver was at fault or at what speed they were traveling. Other data that may be useful include age profile, gender, type of car, road conditions, weather conditions, time of date, distances, and GPS information on places travelled and distances covered.

The town planners will draft mitigation strategies that inform their road planning and calming traffic strategies. These can result in eradicating the number of accidents and damages for the city.

# DATA

Coursera supplied the CSV file and metadata deployed for this assignment. Coursera sourced this data from the Seattle Police Department (Traffic Records) from 2004. Coursera course video suggested the following before analysing data:

* Refer to the web link on the CSV file for a good description of each attribute
* The label for the data set is severity, which describes the fatality of an accident.
* The shared data has unbalanced titles; therefore, balance the data to avoid biased model.
* Feature engineering is required

In this assignment, the SEVERITYCODE is the dependent variable. The independent variables comprise information based on accident characteristics (accident occurrence time and accident location), vehicle characteristics (vehicle type involved and vehicle condition), environmental factors (weather condition and visibility distance), and road conditions (pavement condition and roadway surface condition).