**CAR ACCIDENT SEVERITY**

**Business Problem**

Traffic accidents represent one of the leading causes of death worldwide and of economic expenditure. Despite the numerous measures and campaigns that are deployed every year to raise awareness of the seriousness of the problem, it still occurs quite frequently. The impact of road accidents on society and the economy is high, and human losses are compounded by large expenditures on health care, awareness campaigns, mobilization of specialized personnel. The WHO sets the economic impact of road accidents in a developed country at 2 to 3 % of GDP, a significant figure for any country Collaboration to reduce these losses has become an important issue of general interest

Defining the problem:

1. What are the factors that have high impact on road accidents?
2. Is there a pattern for them?
3. Correlation?

We have to analyse the data to get a clearer picture and draw conclusion

**INTRODUCTION**

The data has been collected and shared by the Seattle police department and are provided by Coursera for downloading through a link.

It takes into account the period of time from 2004, recording information related to the severity of traffic accident, location, type of collision, weather and road conditions, visibility and number of people involved.

The objective is to define the problem, to find the factors that can have a relevant weight in the quantity and seriousness of the accidents so that they can organism, company interested in reducing these figures.

**DATA TO BE USED**

For an accurate prediction of the magnitude of damage caused by accidents, they require a large number of reports on traffic accidents with accurate data to train prediction models. The data set provided for this work allows the analysis of a record of 200,000 accidents in the state of Seattle, from 2004 to the date it is issued, in which 37 attributes or variables are recorded and the codification of the type of accident is allowed, grouped according to 84 codes. The information can be extracted from it:

1. Speed information
2. Information and road conditions and visibility
3. Type of collision
4. Affected persons

The data will be used so that we can determine which attributes are most common in traffic accidents in order to target prevention at these high-incidence points.

**DATA SOURCE**

*Data Source: These data have been collected and shared by the Seattle Police Department (Traffic Records) and are provided by Coursera for downloading through a link.*

**METHODOLOGY**

Objective: The objective of this project is to predict the severity of a traffic accident based on the other characteristics contained in the report.

Packages and libraries: We will use libraries and packages for both data manipulation and data visualization. PANDA, NUMPY, SCIPY, Matplotlib, Seaborn

A data analysis will be performed in order to determine what type of methodology and learning of the machine will be the most appropriate, in addition to obtaining a first contact with the data that we find more relevant to use in this project.