# Severity of car accidents

1. Introduction
2. Data
3. Model
4. Testing
5. Conclusions

**1. INTRODUCTION**

Traffic accidents, little instants in which the life can change, and in some occasions, end; year by year thousands of lives are lost due this terrible incidents which can be caused for the most insignificant reasons, as animals on the way or a distraction while driving.

There is not a unique reason for the traffic accidents, the causes are too diverse and there are so many factors intervening at the same time that every case needs to be evaluated in detail in order to find all the hidden answers, answers which can be founded on the driver, the speed, the weather, etc.

On this project we’ll find the correlation between the specific factors and the severity of the accidents, understanding severity as the amount of human lives lost on the collision, for this we’ll need a dataset in which all the different conditions present on the moment of the incident are signed.

**2. DATA**

For this project we will use the dataset *"Accidents de trànsit amb morts o ferits greus a Catalunya"* wich reports all the traffic incidets reported on Catalonya on the period 2010-2020 with the details about weather conditions, road conditions and number of casualties, the data set can be found on the link “<https://analisi.transparenciacatalunya.cat/Transport/Accidents-de-tr-nsit-amb-morts-o-ferits-greus-a-Ca/rmgc-ncpb>”.

**3. METHODOLOGY**

The methodology used on this project will be based on a multiple linear regression model, there will be 4 targets, the main target that is the total number of victims on the accident (F\_VICTIMS on the dataset), and the secondary targets that will be the number of deceases on the accident (F\_MORTS on the dataset) and wounded (F\_FERITS\_GREUS and F\_FERITS\_LLEUS on the dataset).

For helping us on the analysis, we’ll use the columns in which each variable (weather conditions, status of the road, velocity limit, etc.) is selected as influence on the incident, this columns are filled with values of “yes” and “no” easily transformable to Boolean values [1, 0]; the missing values on this columns are signed as “sense specificar”.

First we’ll evaluate each one of the secondary targets in order to find the correlation between the external factors and the amount of deceases and wounded victims, then we evaluate the total amount of victims in relationship with the first evaluated targets.