# **Coursera Capstone Report**

## ****Introduction****

**In the context of collision problem, is very important to know the influence of weather, road and light conditions to predict the severity of collisions, so the main goal of this study is predicting this severity considering environmental factors.**

**The audience of this study are Government workers that are concerned in evaluate and improve conditions in public roads to reduce severity of collisions in these roads.**

## ****Data****

The data I will be using in Cousera Final Capstone is the Data Collision data of the Final Capstone, in which data the severity of collisions is related to conditions like location, weather conditions, amount of people involved, type of collision, among other factors.

This data will be the main data to predict severity of collisions using weather conditions, light conditions and road condition.

For instance, the first assumption is that wet floor will increase severity of collisions, and weather conditions like rain, snow and fog will increase severity of collisions.

Darker conditions will increase severity of collisions too.

## Procedure

The data I use has many Nan values, so I deleted them. The next step was selecting only data related to read conditions, light conditions (as independent columns) and severity code as dependent column.

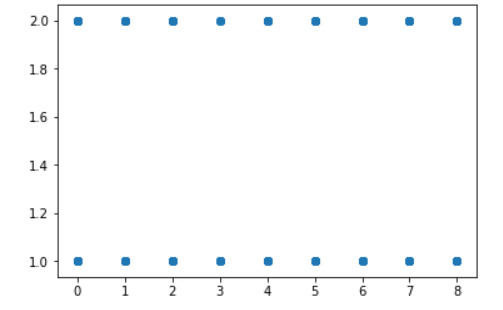
Then, I used multiple linear regression, applying label encoder to redefine categorical values to numerical values.

The next step was applying train test split method to use in the multiple linear regression algorithm.

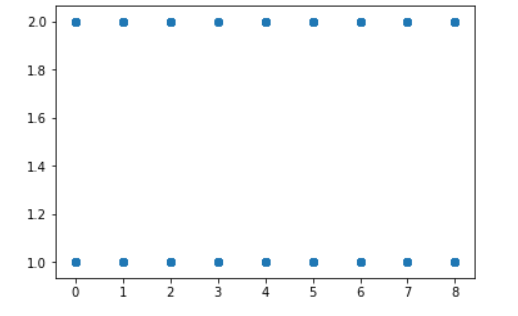
## Results

Next images show the scatter plots.

Severity code vs. Road conditions



Severity code vs. light conditions



The final results are:

Residual sum of squares: 0.21

Variance score: 0.01

So the model is predicting correctly the severity code of light and road conditions using multiple linear regression after making a label encoding of several light and road conditions.