

Relationship between car features and severe accidents

Capstone Project - IBM

INTRODUCTION

High relation between the velocity of the car and car accident

Also → size of the car and the color

Data can confirm if these arguments are true or not → This project aims to predict if there are more traits apart from the driver's ones that can indirectly be inductors of car accident.

Public organizations like traffic ones would be very interested in seeing these correlations, as well as car producers, in order to know to which car model should they focus in developing new security structures and programs.

DATA ACQUISITION AND CLEANING

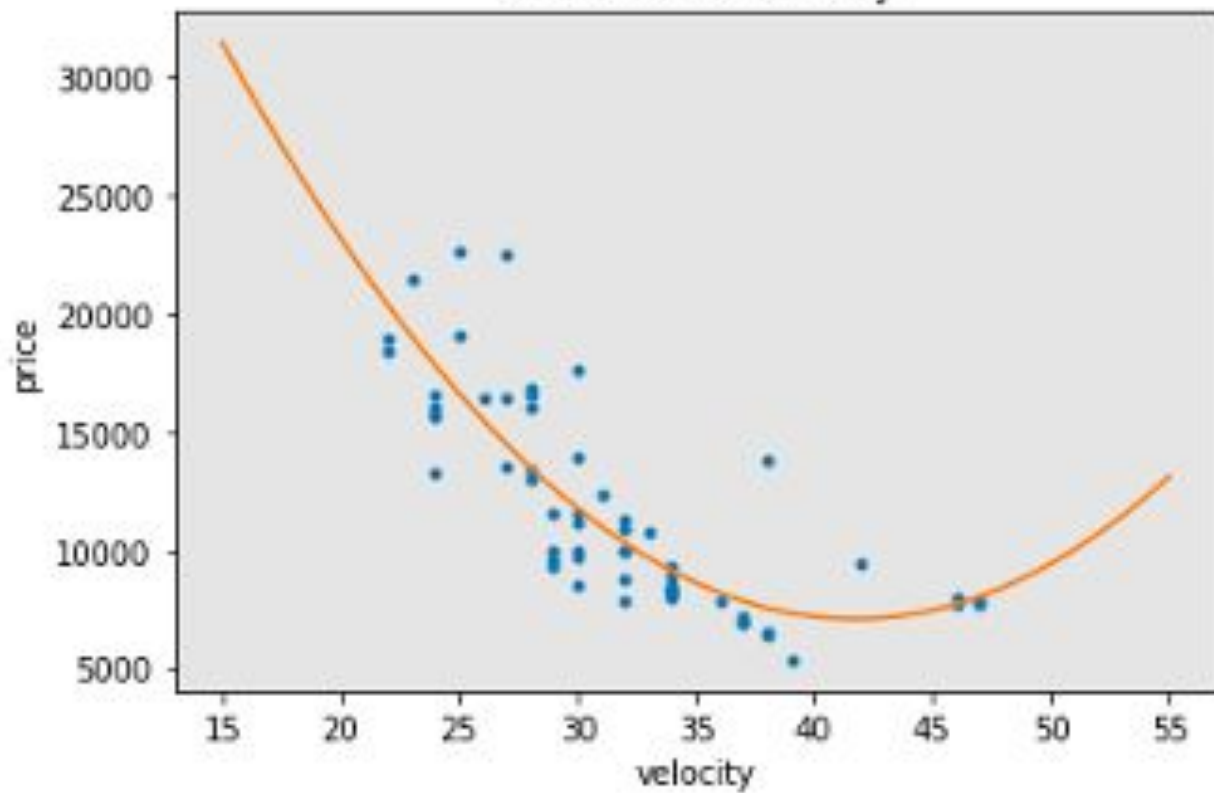
it has been decided to take the characteristics of 'color', 'size', 'velocity' , 'power', 'value' and 'victims' as traits to be analyzed from the year 2020.

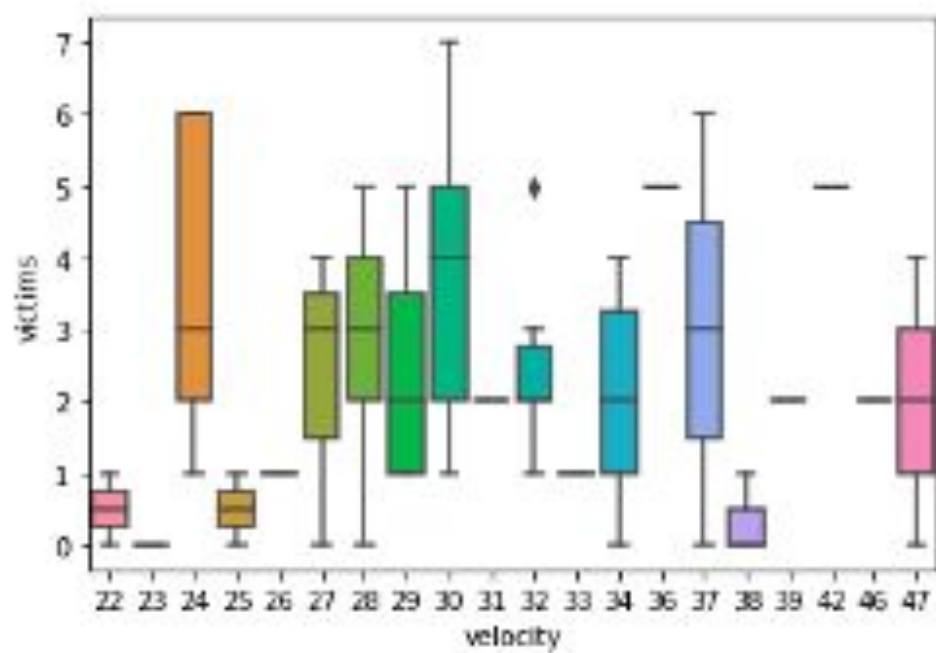
All the redundant information or similar one has been discarded. Cases of cars that had equal velocity, color or size, as well as if it had victims have been joined to one set, as it was not given any reliable added information to the project.

EXPLORATORY DATA ANALYSIS

All the information that was needed regarding the objective of the project was set in the table created in the previous step. Target was not needed to be calculated but to be found. For example, the aim of the study was to find if there was a correlation between a feature and the severity of the accident, in this case victims. Relations were needed, but no other calculation or other new columns.

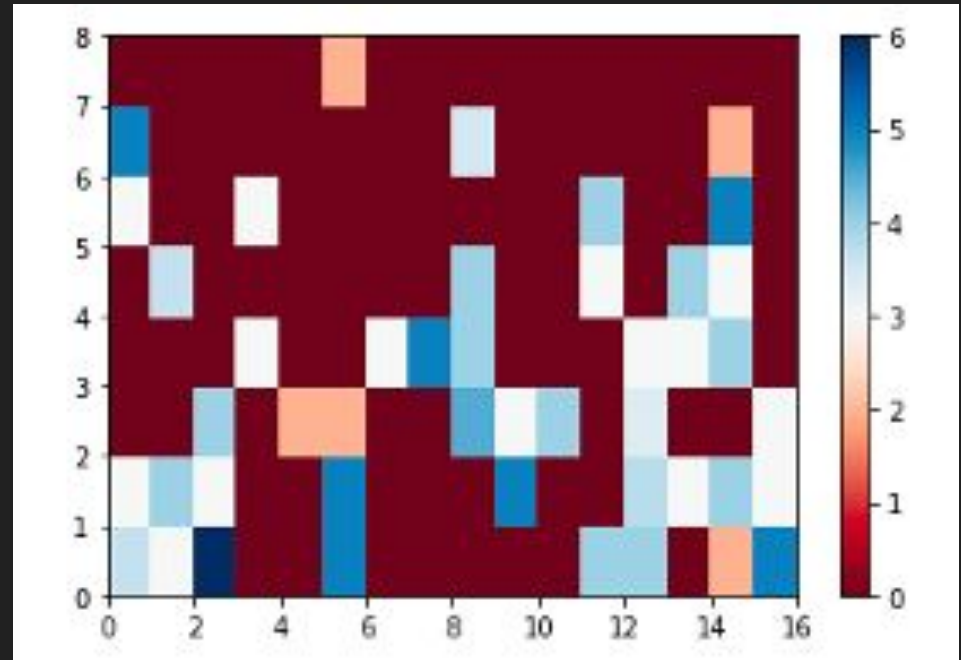
Price and car velocity





Other relationships

Relationship between the size of the car, color and the number of victims



PREDICTIVE MODELLING

Even though it was thought that some characteristics about the car were indirect causes of severe accidents, regression models have shown that is not possible

Predictions

As shown in the previous chapters it has been seen that the best characteristic to analyze a relationship between victims and one indirect feature is the velocity one. An indirect characteristic that becomes directly related to a severe accident cause.

Regarding the standard algorithms a number of 9 test samples were given, whereas another 50 samples were from the training set. The resulting score is 1.0, which might seem that the he predicted values had a narrower range than the previous analyzed ones resulting that the prediction errors were larger as the actual values deviated further from zero. In this case the result was given to 1. This is not desirable, as errors are high.

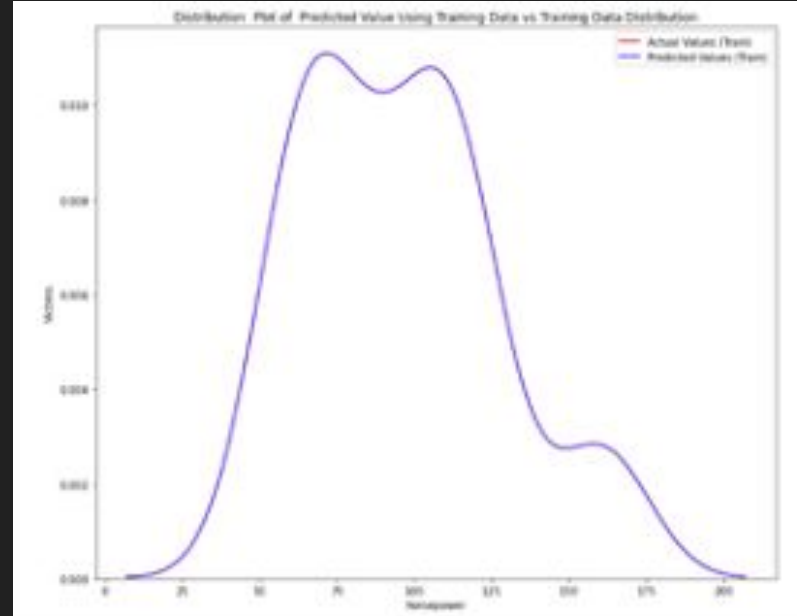
Problems

The problem was that it was very common sets with very little victims, whereas severe ones were rare. This made that the model analyzed first the accident with few victims rather the ones with more than three victims.

The solution was to give more importance to the sets that referred to severe accidents not only in the training set but also in the test one. With this solution, all models had similar range and distribution.

Other problems

Some errors with the data did not permit the analysis of cross validation score and prediction models, as shown in the next figure. This is due to error data or an error of formatting it, which implies not getting the expected results.



CONCLUSIONS

It was able to analyze that even though there are some features that people normally assign to an accident cause, in reality there are not. Regarding the schemas that have been prepared in this project, the only cause in severe accidents is velocity, and moreover the horsepower of a car, whereas the color and the size of it does not really affect.

An unexpected data has been found, and it is that less velocity produces more victims, which may be caused to the place where they occur, for instance in towns or cities. Accidents with high velocity impacts also create some severe accidents, but fewer.

FUTURE ANALYSIS

From this analysis it can be expected that even though these features are not indirectly implied in a cause of an accident, it might seem that not all the fault may be from the driver.

It is true that the person who drives the car is highly responsible of the cause of the accident, but there might be other indicators to take into account that would create a more severe accident and are not controlled by the driver itself, for example the size of the car or other similar features.