
IBM DATA SCIENCE CAPSTONE PROJECT

CAR ACCIDENT SEVERITY

Business problem - Introduction

1. A description of the problem and a discussion of the background

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, etc. 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:

What are the factors that have a high impact on road accidents?

Is there a pattern to them?

Correlation?



Tip

We will have to analyze the data to get a clearer picture and draw conclusions.

Introduction

Note that this work represents the final project of the IBM certification course, for which we have provided the data with which we will develop the project.

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

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

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 any organism, company or enterprise interested in reducing these figures, can focus the resources in points where these conditions converge.

In order to provide greater clarity, I will try to analyze the data, see if there are relationships or patterns, especially in high impact accidents, so that preventive measures can focus on these points as a first prevention strategy.

2. A description of the data and how it will be used to solve the problem

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:

speed information information on road conditions and visibility type of collision affected persons, etc 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.

Data Location: Coursera_Capstone/Data assets

Data set name: Data-Collisions (1)_shaped.csv





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. Obtaining and cleaning data Importing libraries and packages Uploading the data Training the model Selecting the methods: Tree model, Logistic Regression and KNN methodology

Results

Comparing the results obtained

According to the results of this method, the two attributes of the table have a determining influence on the increase of accidents, however more factors must be taken into account



Discussion and Conclusion

- Knowing that speed and road conditions are factors that influence traffic accidents and their severity, recommendations for reducing accidents could include
- Limiting speed in areas where road conditions are worse
- Surveillance and radar cameras -Maintain traffic control routes on these roads
- Enable emergency exit areas and rescue stations
- Improve the maintenance of these streets, signaling and lighting

—

THANKS!