

Predicting Car Accident Severity

Introduction

Car accidents have a huge social, economic and environmental impact to our world. Approximately 54 million people sustained injuries from car accidents each year and more than 1.4 million resulted to deaths. Road traffic injuries are the leading cause of death for children and young adults aged 5-29 years. The road traffic crashes costs more than 3% of GDP.

It would be very important to predict the severity of a car accident and warn the drivers, in advance and on time, in order to drive more safely or even to adjust their journey.

Thus, the aim of this project is to apply quantitative methods and create a model that predicts the severity of a car accident based on weather and road conditions, traffic jam etc.

The further development of machine learning is critical and will play a crucial role on improving our lives in the following years.

The Data

The dataset is from Kaggle and contains details of traffic accidents in United States for the years of 2016 –2020. My focus was to analyse specifically the car accidents severity in Pennsylvania State

The features contain all sort of information about the time, the location, the road conditions, the weather conditions etc. Indicatively, there are information about the Severity, The Start-End time, Lat_Lng, the description, the city, the state, the street, the weather, the visibility, the wind, the road condition etc. There are 55 columns and my analysis concentrates on 33 columns.

The type of data is a combination of integers, floats, Boolean, objects etc. All the data has been cleaned and presented in an appropriate format. There were many adjustments to the columns related to the date and time data. All the nan or missing data and negative values has been removed.

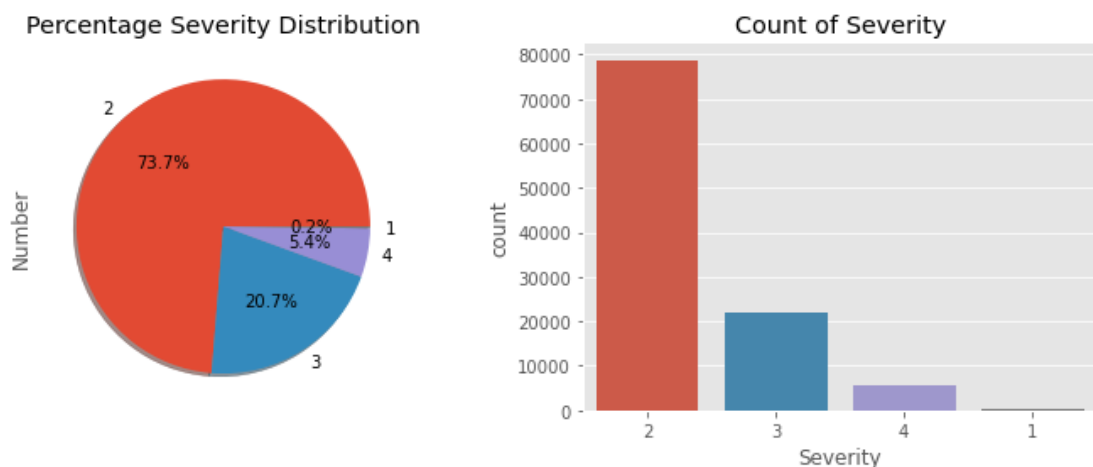
Some main component of my analysis was regarding the presence or not of amenity, crossing, junction, railway, roundabout, station, stop, traffic signal etc. Moreover,

Predicting Car Accident Severity

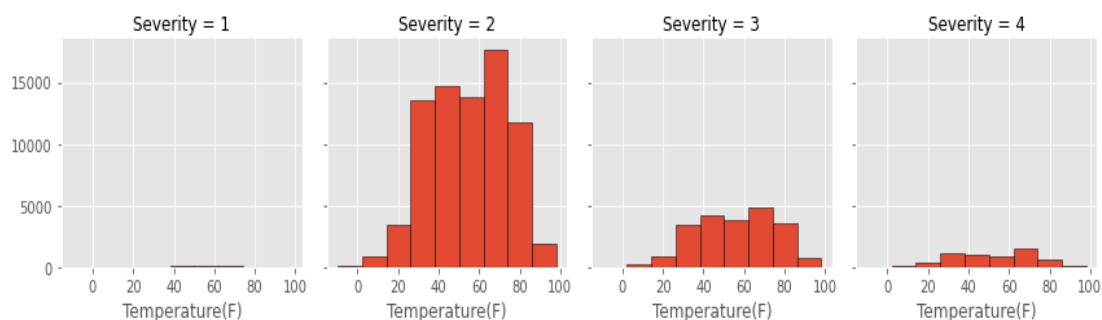
interesting information was about the temperature, the humidity, the visibility, the weather condition as well as features that shows the period of the day such as sunrise_sunset, civil_twilight etc. The target value, is the 'Severity' and is shown with numbers from 1 to 4, with 1 being the least significant and the 4 the most significant. Thus, the choice of feature is in three directions, the time, the weather conditions, and the road conditions.

Methodology

Following the cleaning and preparation process, a graphical representation of data is taking place, in order to gain useful information, in an easy and quick way. For the representation of our target value we can conclude than most of the accidents can be categorized in the severity level 2, as shown below.

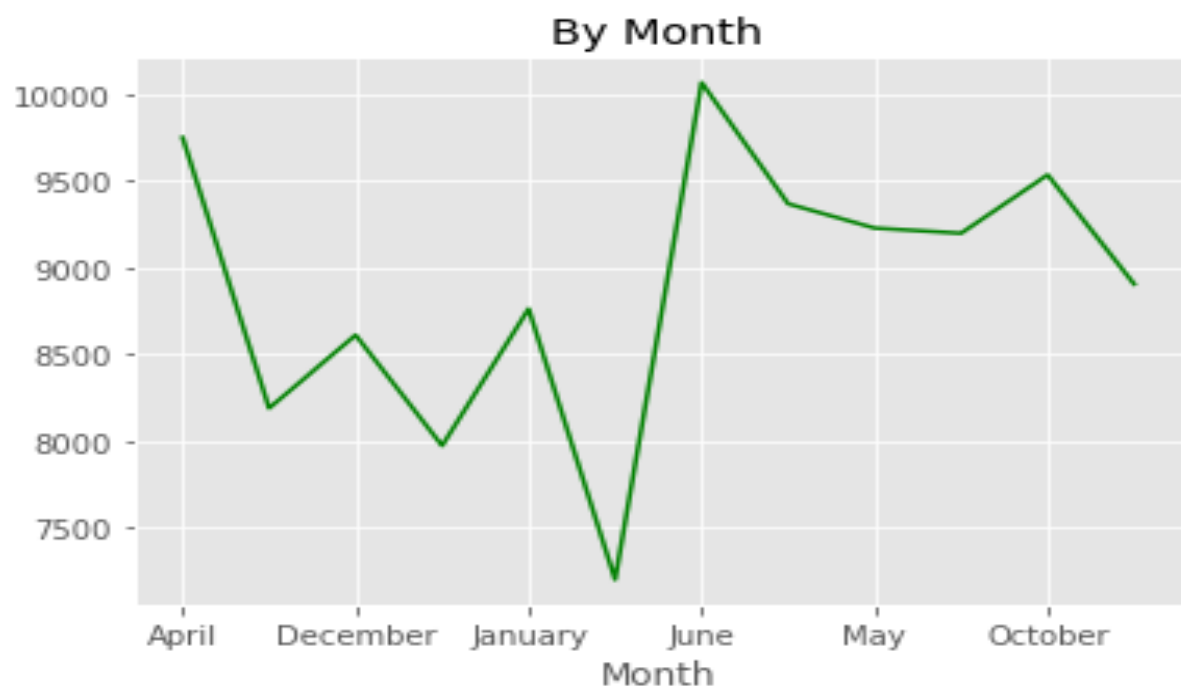
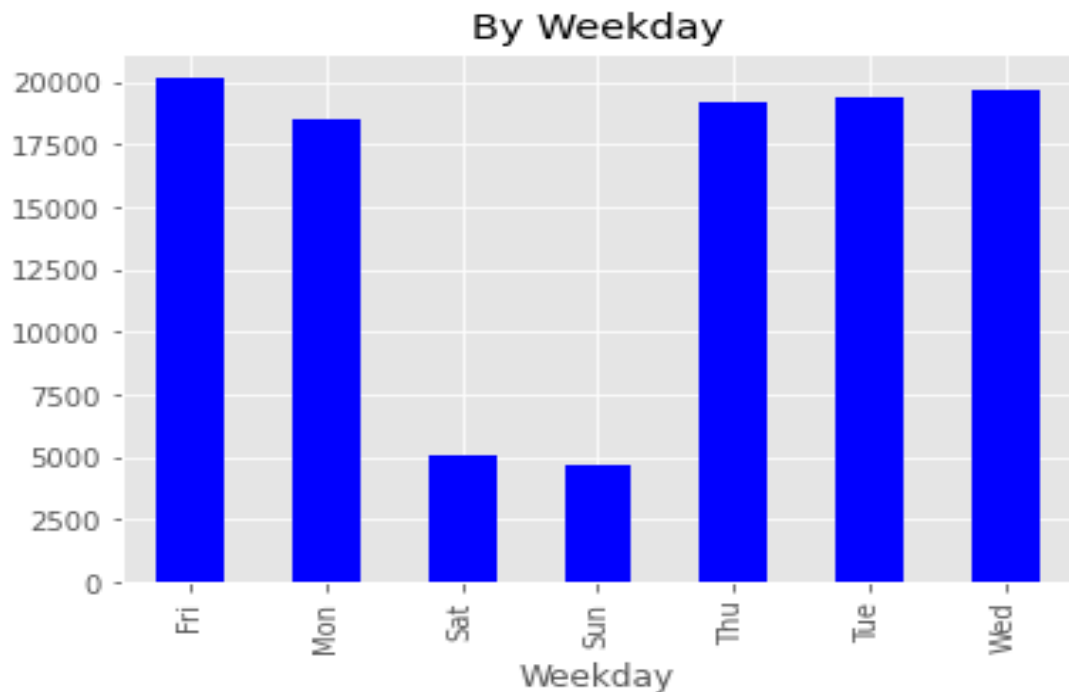


Another, interesting thing is to see the distribution of Severity in relation to the Temperature, and as we can see in the chart, the most accidents are happening on 60-80 (F).



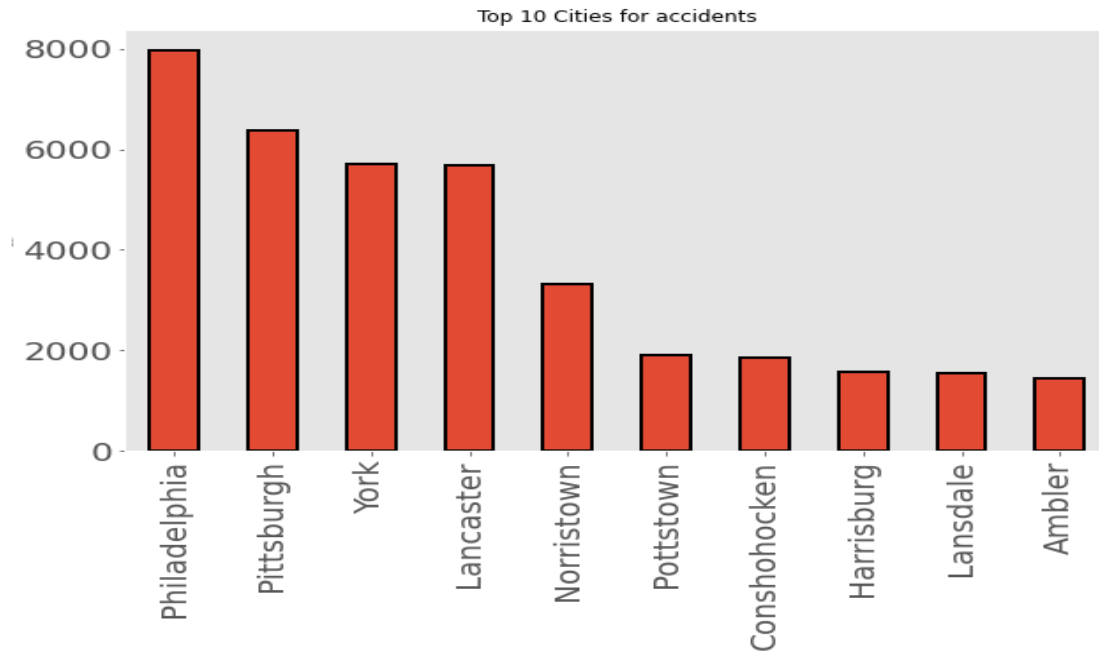
Predicting Car Accident Severity

Now let's have a look at how the weekdays and the months affect the number and severity of car accidents. Based on the data we can easily notice that on Saturdays and Sundays have significantly lower accidents than the other weekdays and that most of the accidents are taking place in June, April and October.



Predicting Car Accident Severity

Regarding the location, cities that are most to have accidents are the Philadelphia, Pittsburgh, York etc.



Regarding the prediction methods, there have been used 3 different machine learning algorithms, a linear regression, the decision tree and the k-nearest neighbours. The algorithm with the highest accuracy was the linear regression 0.9, followed by the decision tree with 0.87, and the knn with the lower 0.812. The accuracy of the current models was satisfactory and it has room for lot of improvements. It is very important to enrich the dataset with more records, information and variables and gain all sort of information about the accidents conditions and use additional machine learning techniques. Machine learning models can be very useful on predicting the severity of car accidents and can be beneficial in many ways. The further development of machine learning is critical and will play a crucial role on improving our lives in the following years.