# Data analysis for predicting car accidents severity

# The aim of the project

- 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 millions 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 hugly important to predict the severity of a car accident and warn the drivers,in advance and on time, in order to drive more savely or even to adjust their journey.

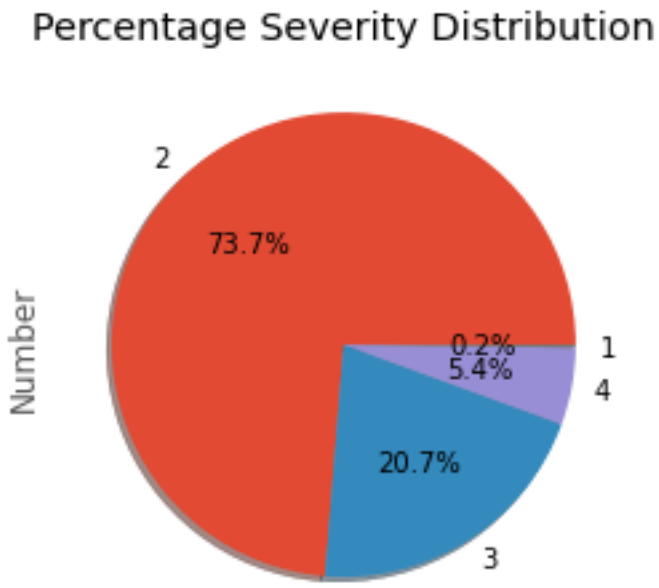
- Thus, the aim of this project is to apply quantitative methods and create a model that predicts the severity of a car accident based on weather and road conditions, traffic jam etc.

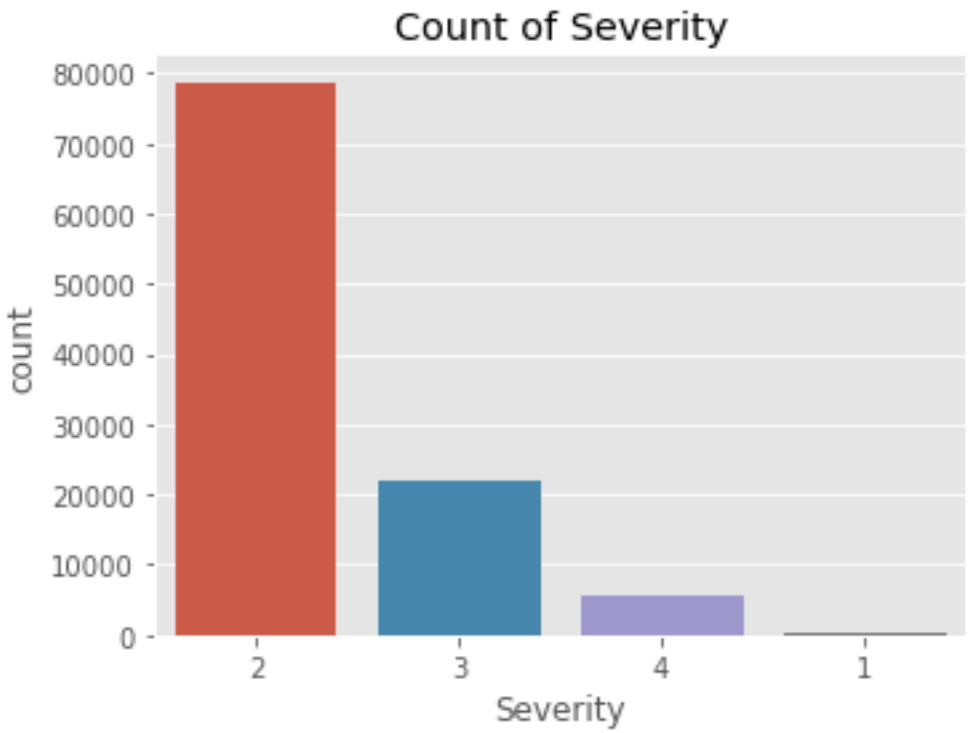
# The Dataset

- The dataset is from Kaggle and contains details of traffic accidents in United States for the years of 2016 – 2020
- My focus was to analyze specifically the car accidents severity in Pennsylvania State
- The features contains all sort of information about the time, the location, the road conditions, the weather conditions etc.
- From 55 columns, that analysis concentrates on 33 columns
- All the data has been cleaned and presented in an appropriate format

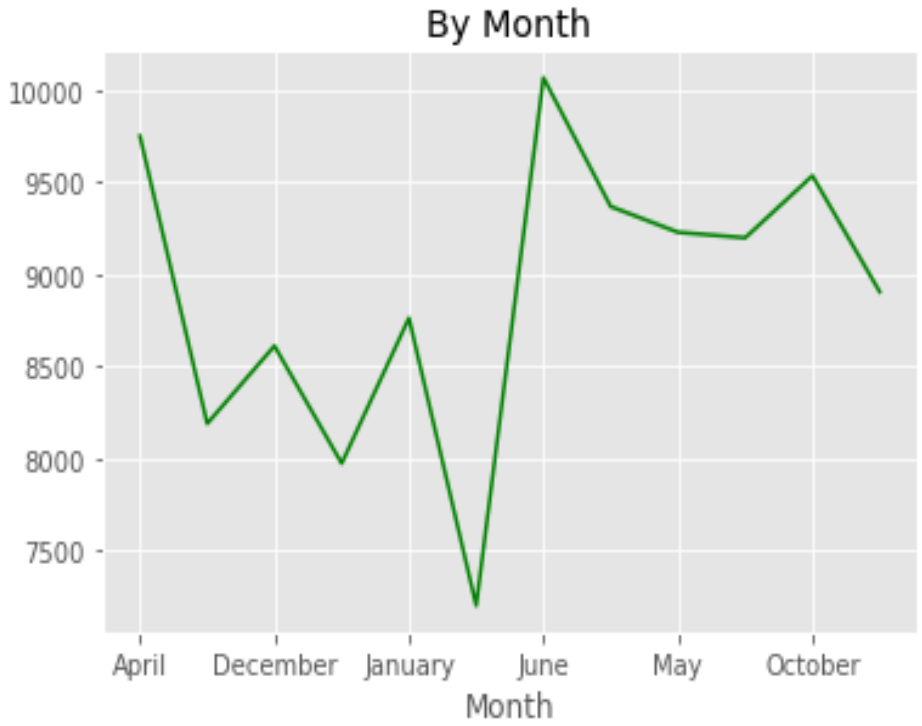
# Key Statistics

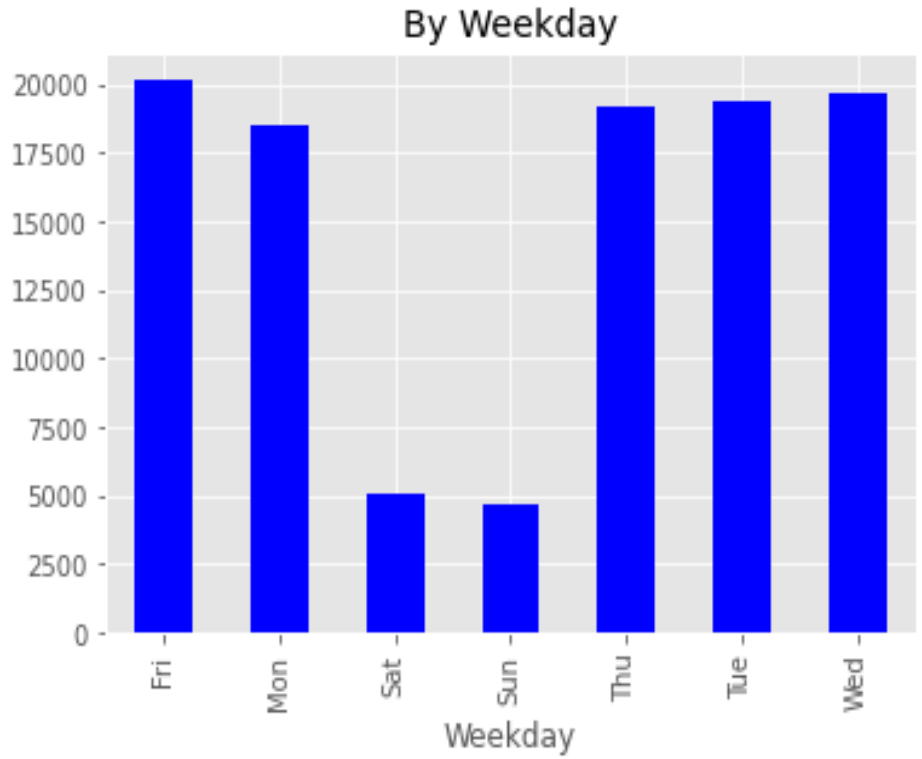
## The severity of accidents



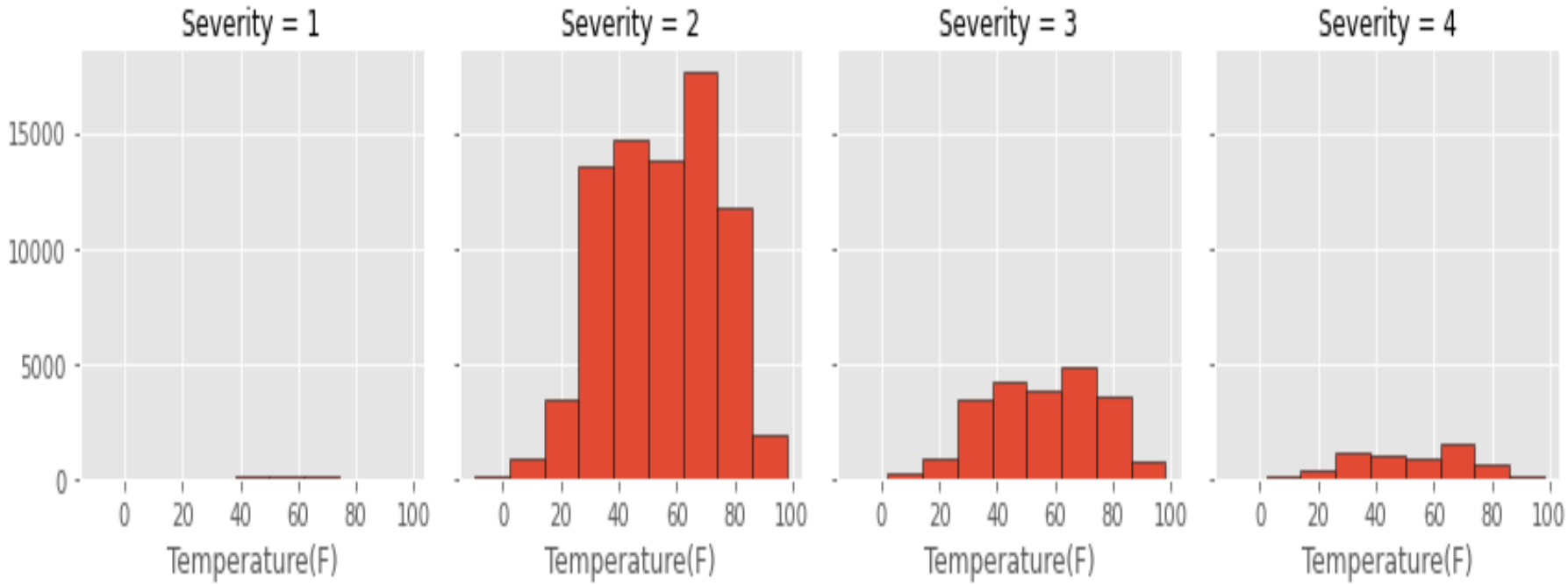



# Number of accidents per day and month

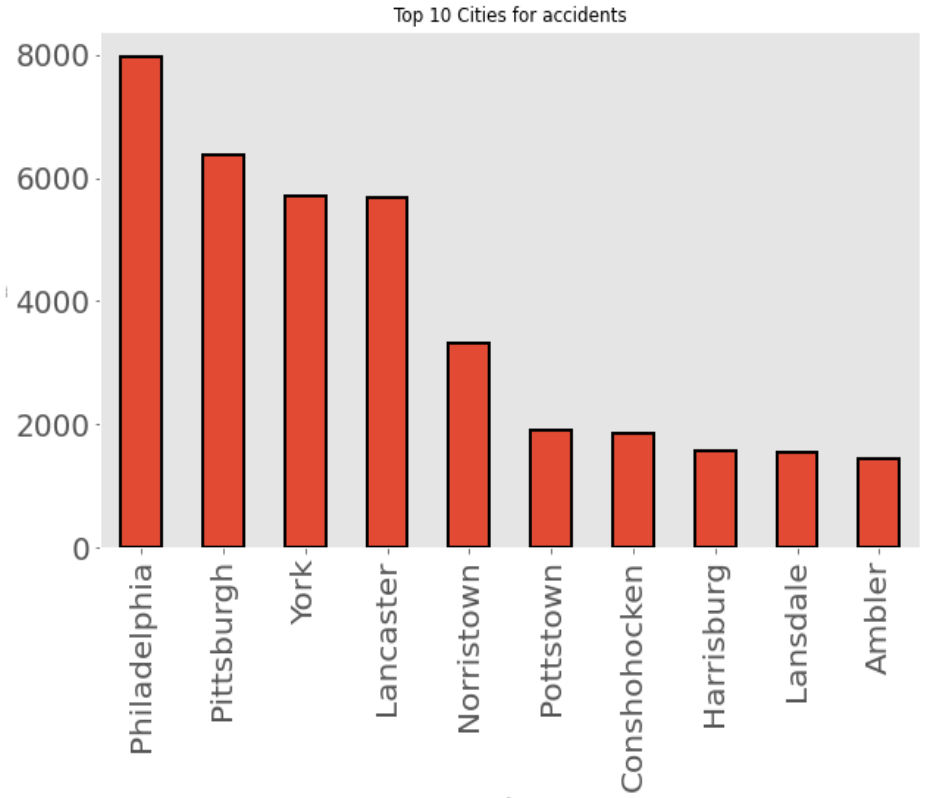


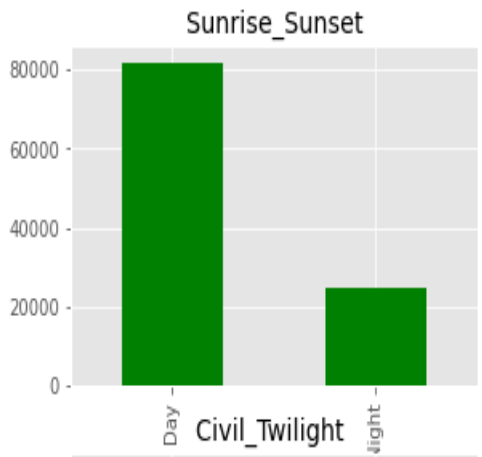



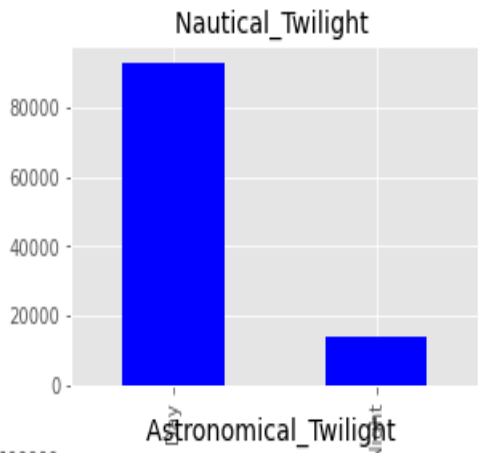
# The severity under Temperature

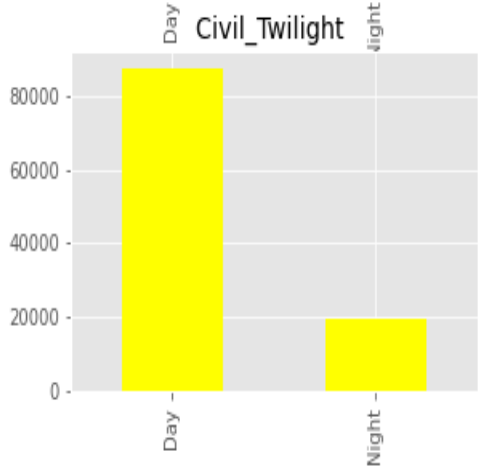


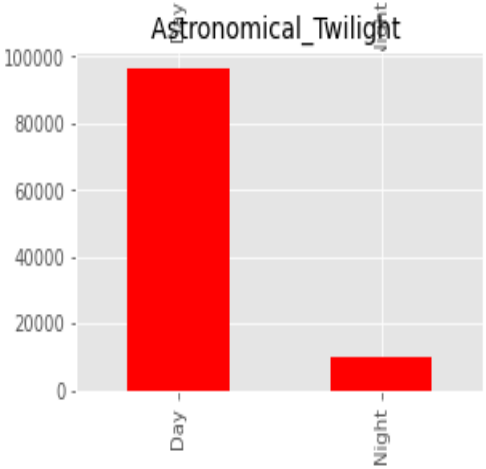
# Other statistics



# Machine Learning algorithms performance

- **Logistic regression**:

  Accuracy score: 0.900

- **Decision tree**:

  Tree entropy accuracy : 0.869

  Tree gini accuracy : 0. 874

- **K-Nearest Neighbors**:

  Knn.score: 0.812

  Accuracy score: 0.812

# Conclusions

- Machine learning models can be very useful on predicting the severity of car accidents
- The accuracy of the current model was good and it has room for lot of improvements
- It is important to enrich the dataset with more records, information and variables and gain all sort of information about the accidents conditions
- The further development of machine learning is critical and will play a crucial role on improving our lives in the following years