Car Severity Prediction

APPLIED DATA SCIENCE CAPSTONE/ DATA SCIENCE PROFESSIONAL CERTIFICATE

Submitted by: Anishka Lunawat

Statistics – Need for this analysis

Road traffic crash result in 1.35 million people with half of these resulting in death or major injuries

Young adults and children are more vulnerable to road traffic accidents

Heavy economic burden on victims apart from mental and physical trauma of those affected

Road traffic injuries impact national economies costing countires 3% of the GDP

Data

Recorded for accidents recorded in France from the year 2005 to 16 was sourced

Data for the various parameters was sourced from Kaggle

Data was cleaned and relevant data is only used for this project
◦ Some data had to be dropped
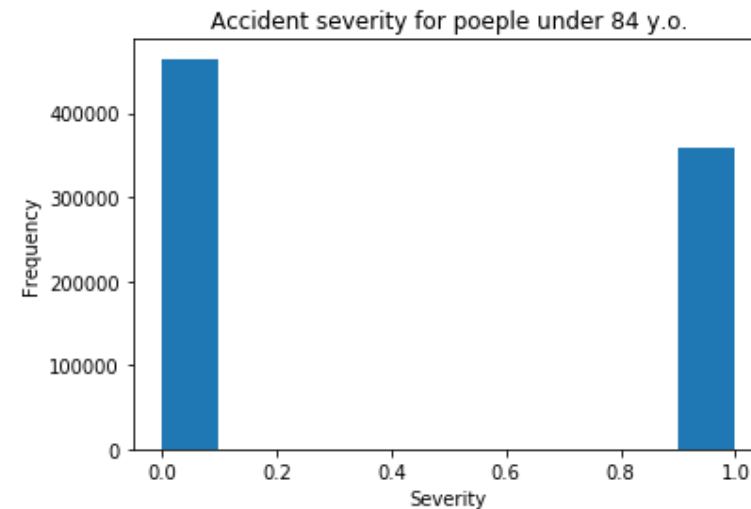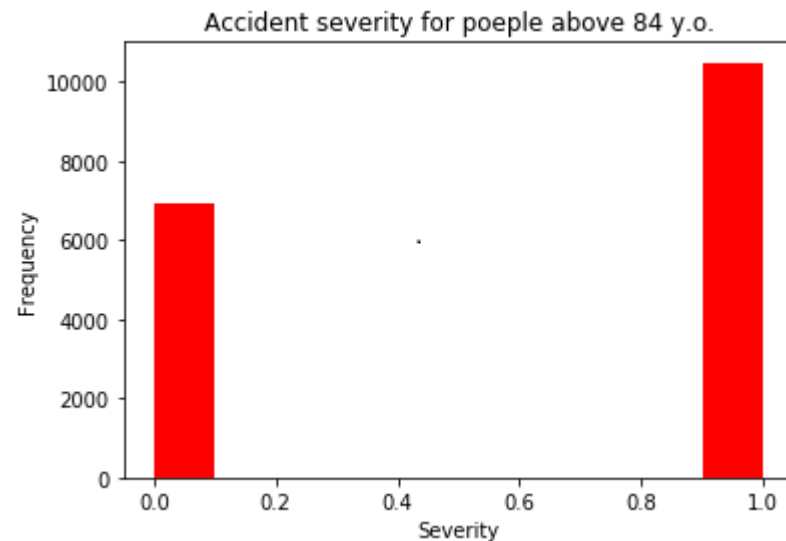◦ Outliers and missing data was replaced

EDA target

The target feature a binary classifier, describing the accident severity
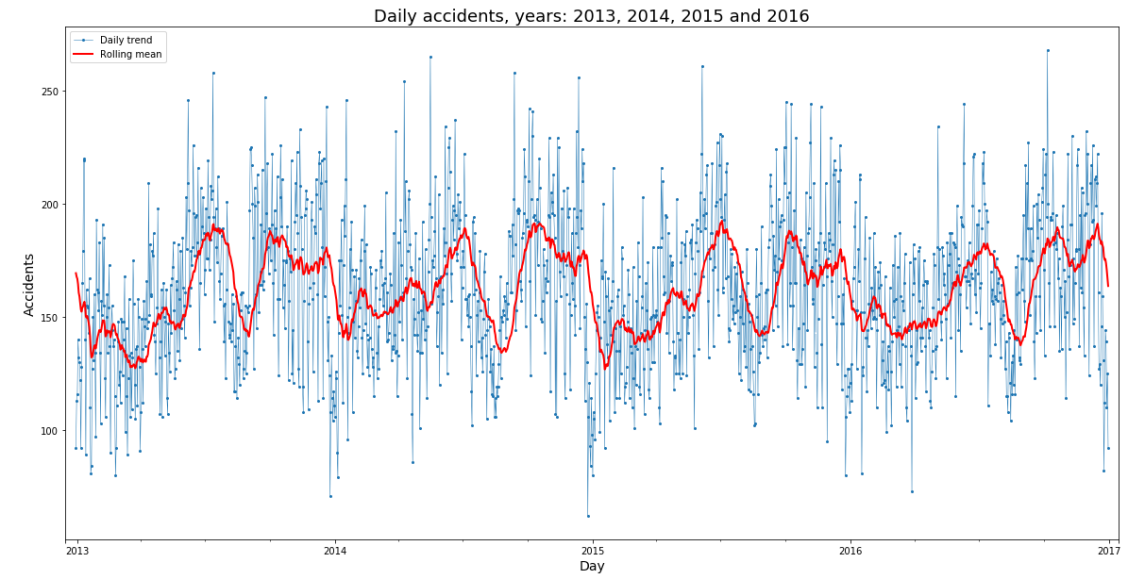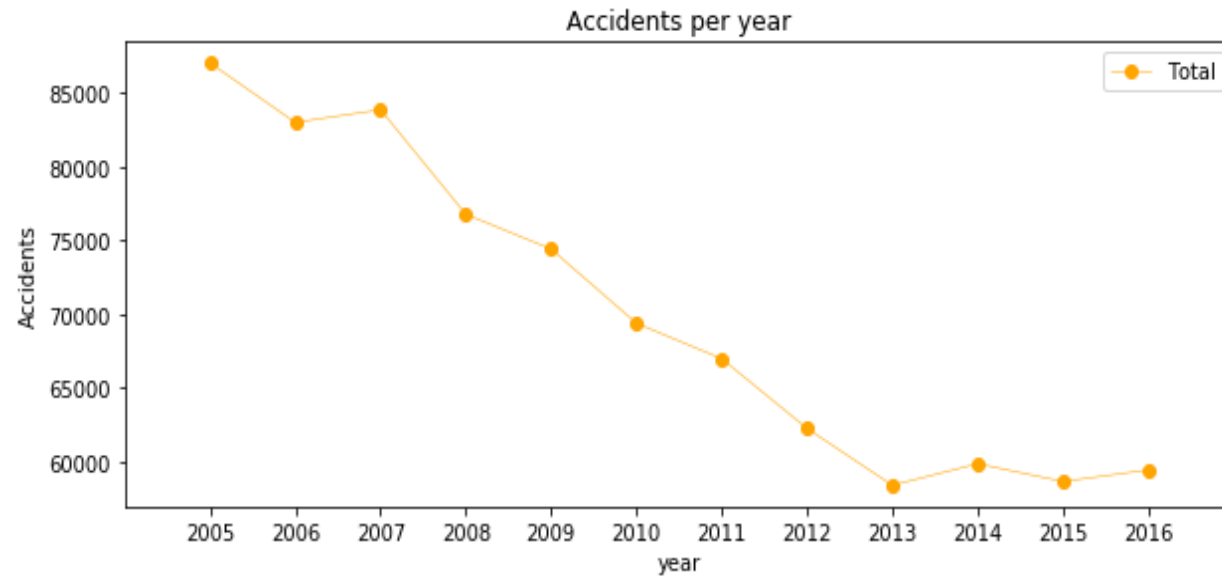
0 – Low

1- High

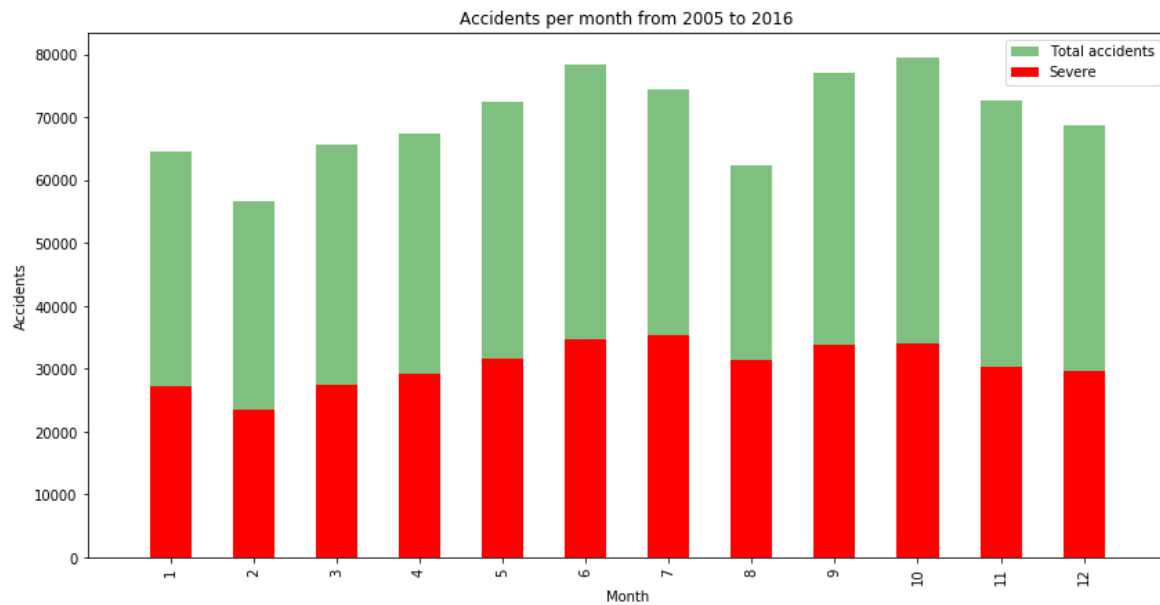It is a balanced dataset with more cases of severity =0

Accident severity for poeple above 84 y.o.

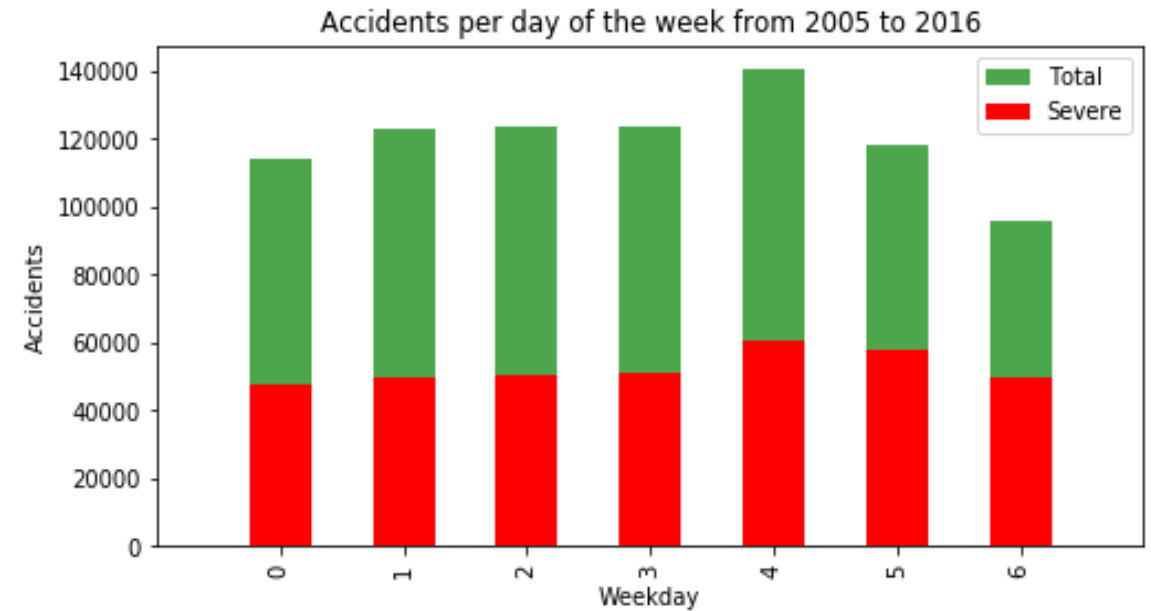Accident severity for poeple under 84 y.o.

EDA - Seasonality

The number of traffic accidents decreased from 2005-13 and the trend thereon looks stable

EDA - Seasonality

Accidents per month from 2005 to 2016
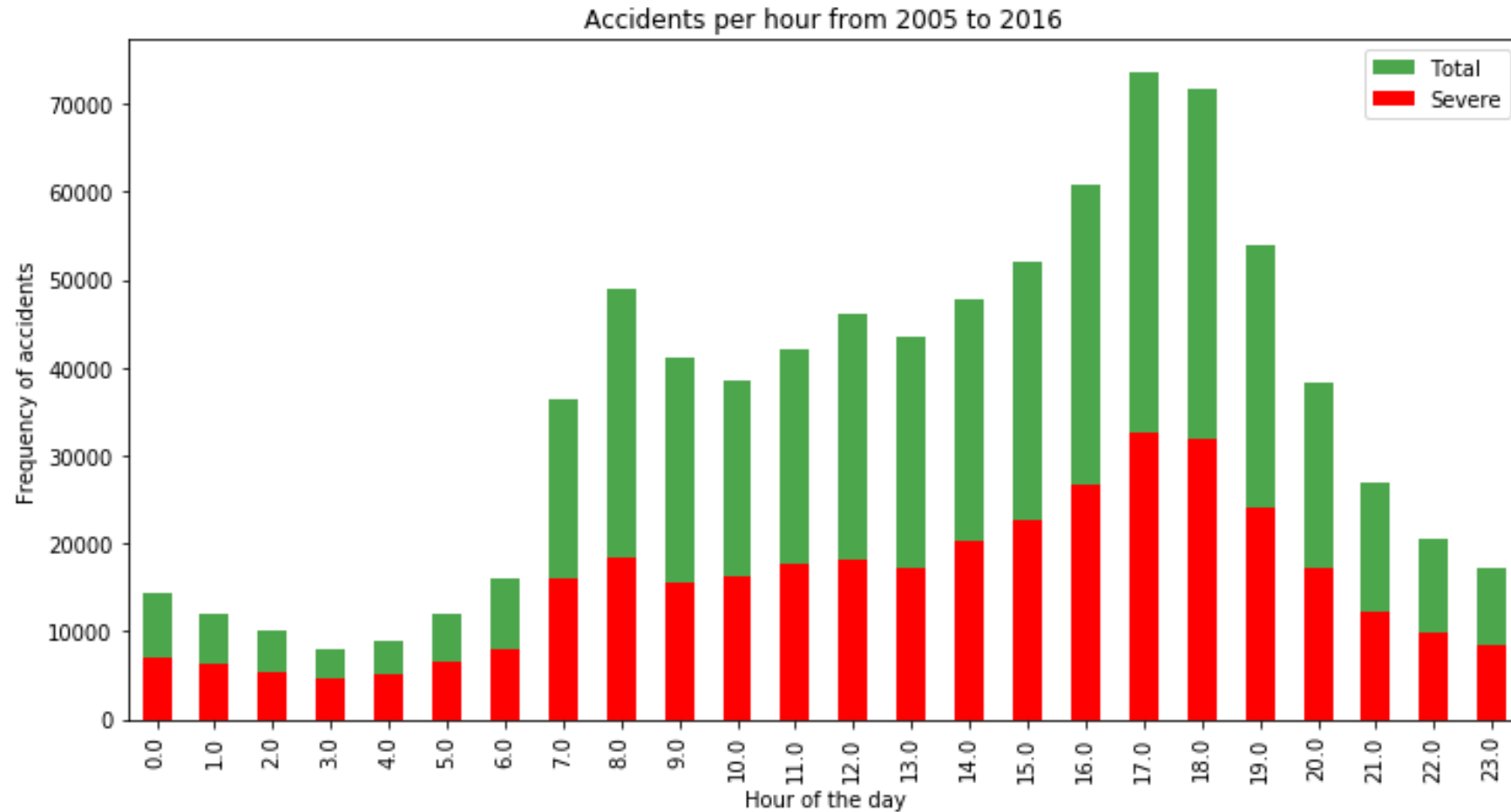

Accidents per day of the week from 2005 to 2016

Accidents were higher from March to June and September

Accidents trend is stead for the week with a spike on Fridays and dip on Sunday

EDA Seasonality

Accidents per hour from 2005 to 2016

Spikes:

Office commute hours

Classification model

Random forest (Decision tree)
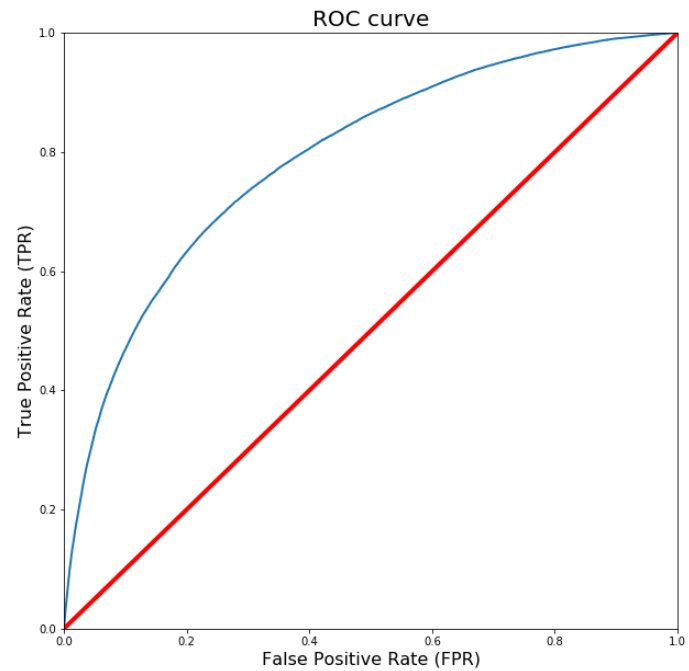
- ◦ 10 decision tree
- ◦ Maximum depth of 12 features

Logistic Regression

- ◦ C = 0.001

K-near neighbour

- ◦ k = 16

Result

Algorithm	Jaccard	F1 score	Precision	Recall	Time
RF (Decision Tree)	0.73	0.72	0.72	0.72	18.58
LR	0.56	0.43	0.53	0.56	4.48
KNN	0.75	0.75	0.76	0.76	20

The best results were noted with Decision tree (Random forest) method

Conclusion

Model can be used with improvements in reducing traffic incidents

Further analysis can help develop a robust model with higher accuracy
- More features if added can result in better prediction of severity
- Many more values can be predicted using such analysis