Predicting Traffic Accident Severity

Applied Data Science Capstone

Vipin Wakodikar

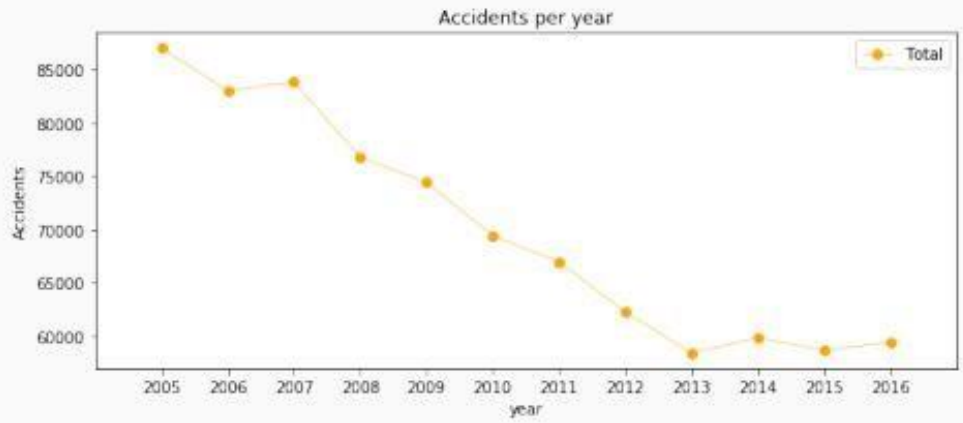
Traffic accidents are...

- Cause of 1.35 million deaths globally in 2016.
- Main cause of death among those aged 15–29 years.
- Predicted to become the 7th leading cause of death by 2030.

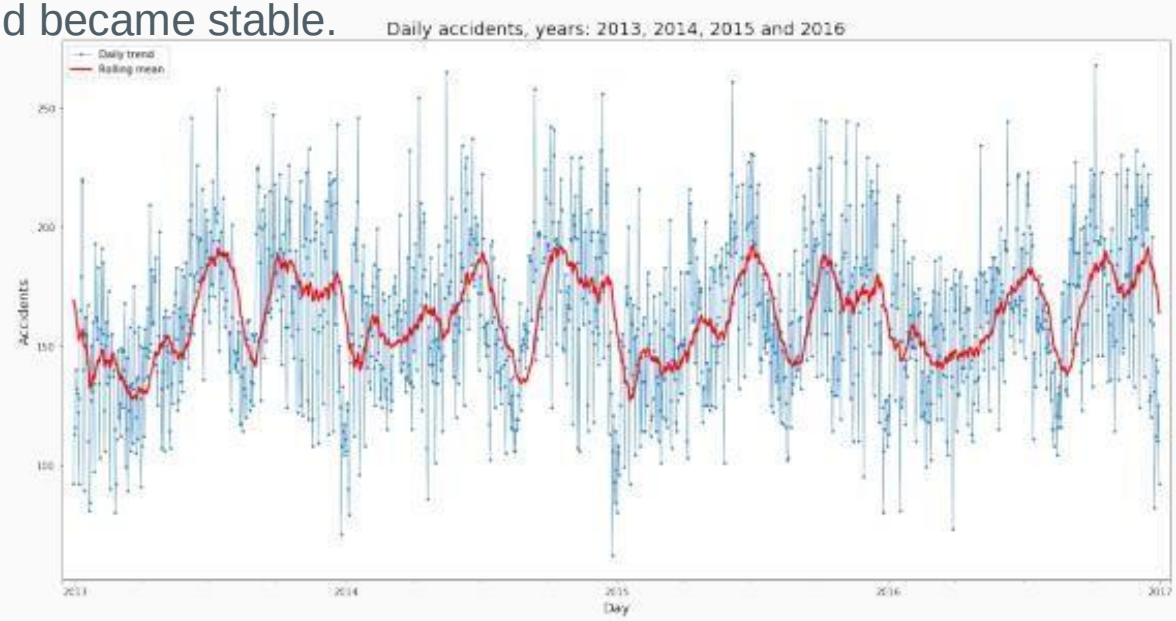
Predicting the accident severity in advance could be used to send the exact required staff and equipment to the place of the accident, thus saving a significant amount of lives each year.

Road safety should be a prior interest for governments, local authorities and private companies investing in technologies that can help reduce accidents and improve overall driver safety.

EDA-Seasonality

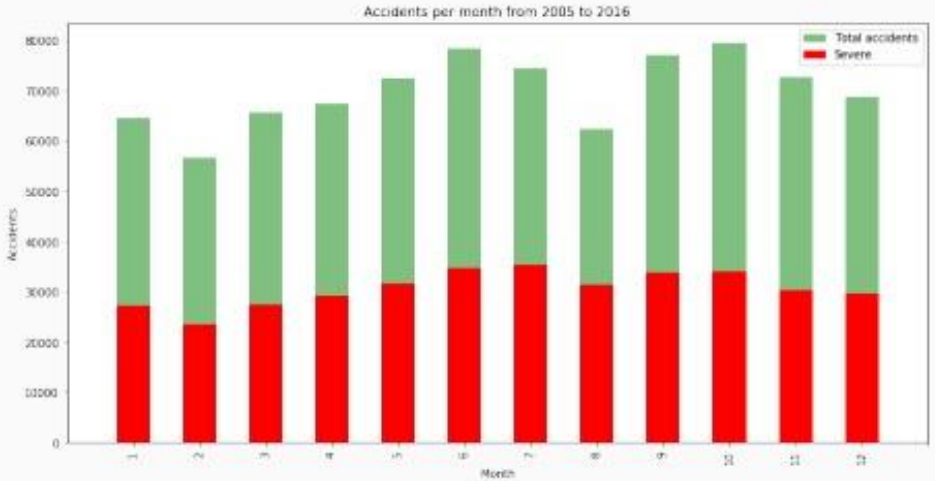


The number of traffic accidents decreased over the years from 2005 to 2013, after which the trend became stable.

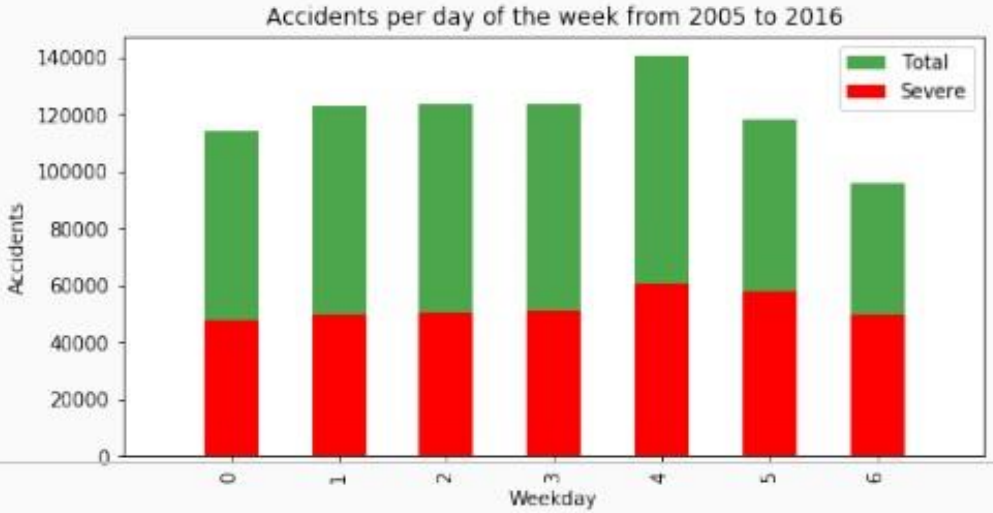


EDA-Seasonality

Accidents increase from March to June and then again in September, decreasing at the end of the year.

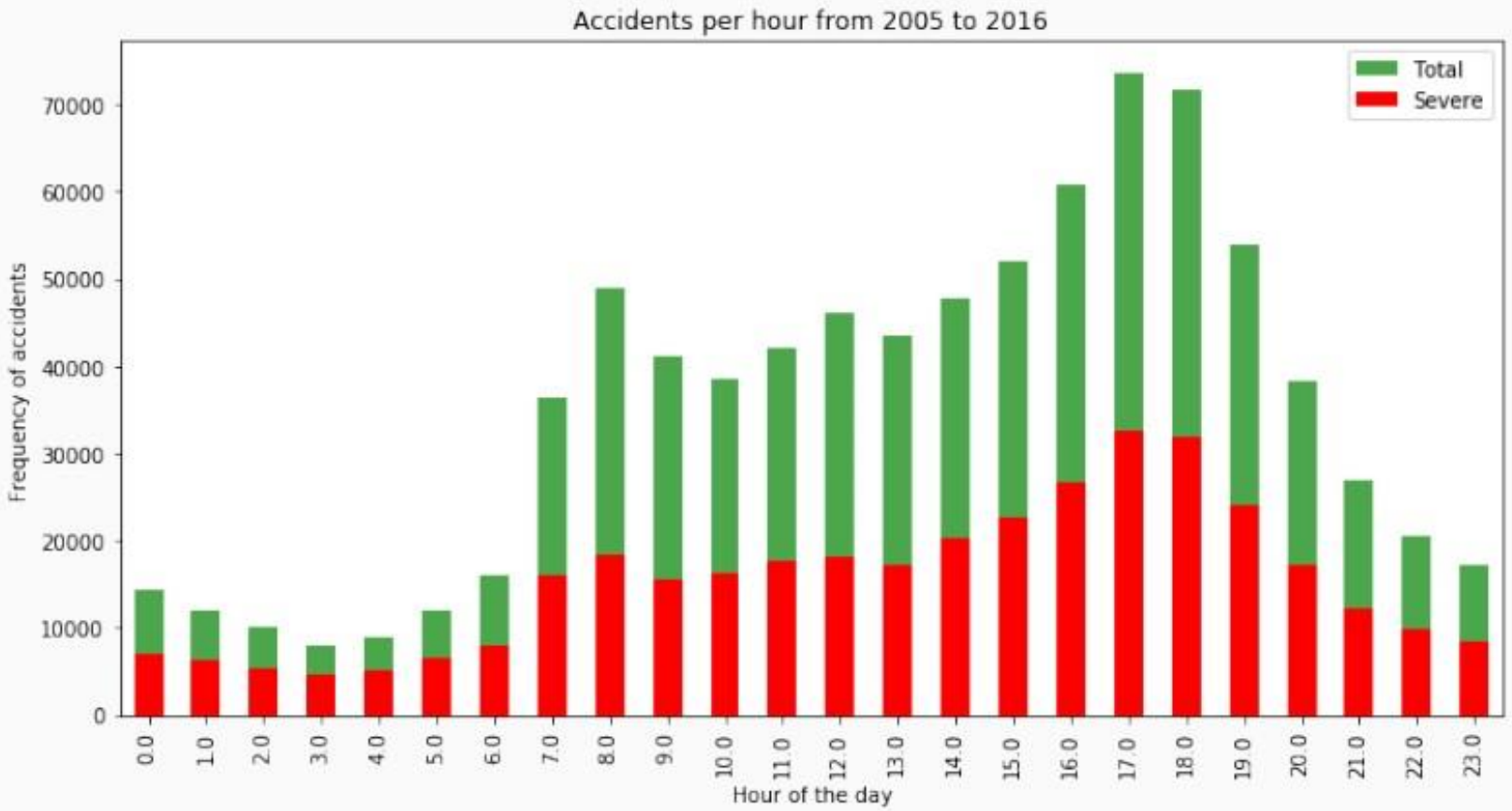


Steady trend during the **week**. More accidents on Friday and less on Sunday



EDA-Seasonality

The trend of highly severe accidents is proportional to the global trend.



Spikes:

8am: people go to work
5-6pm: people return home.

Classification Models

Random Forest:

10 decision trees

maximum depth of 12 features

Logistic Regression

c=0.001

K-Nearest Neighbor

K=16

Supervised Vector Machine

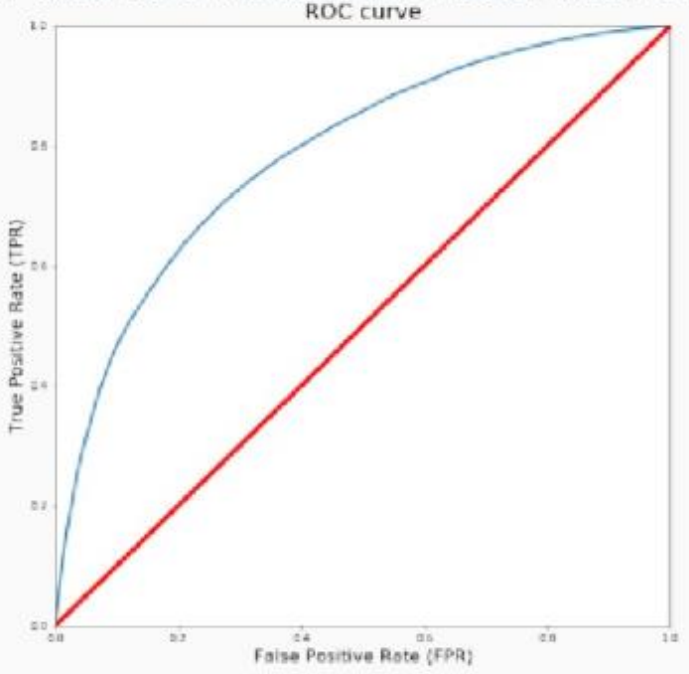
Due to computation inefficiency, training size was reduced to 75,000 samples.

Results

This table reports the results of the evaluation of each model.

Algorithm	Jaccard	f1-score	Precision	Recall	Time(s)
Random Forest	0.722	0.72	0.724	0.591	6.588
Logistic Regression	0.661	0.65	0.667	0.456	6.530
KNN	0.664	0.66	0.652	0.506	200.58
SVM	0.659	0.65	0.630	0.528	403.92

With no doubt the *Random Forest* is the best model, in the same time as the *log. res.* it improves the accuracy from 0.66 to 0.72 and the recall from 0.45 to 0.59.



Conclusion and future projects

Built useful models to predict the severity of a traffic accident.

Accuracy of the models has room for improvement.

Future projects:

Add features such as vehicle speed and time of uninterrupted traveling.

Prediction of potential accident, critical spots and time.