Road Accident in UK (2018)- Accident Severity Prediction

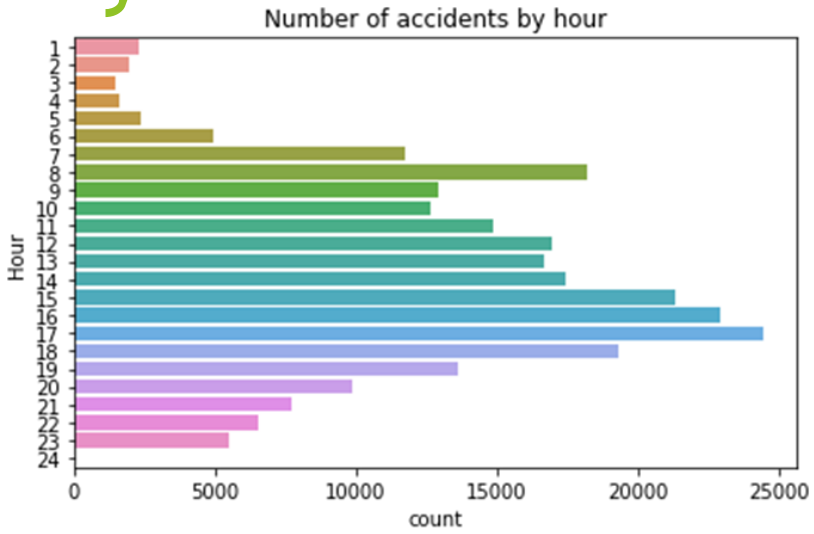
Introduction

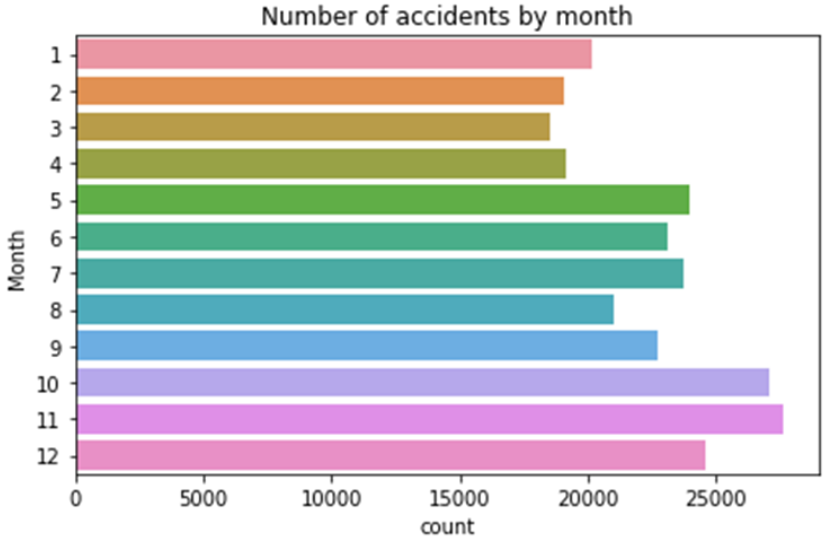
- About 1.35 million people died in road accident in the year 2016.
- It is currently eighth highest cause of death and WHO has predicted it to be seventh highest cause of death by the year 2030.
- This project is performed to predict the road accident and its severity of the United Kingdom in the year 2018
- different features such as location, weather condition, Light condition, age, sex, etc, wete used to predict the severity of the accident

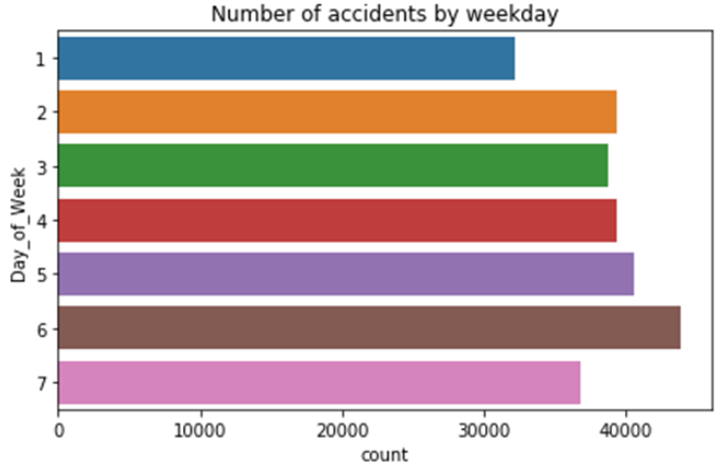
2. Data Acquisition and cleaning

- The data of the road accident was collected from https://data.gov.uk.
- Three different data namely, accident data, causalities data and vehicle data were used
- These data were merged in a single dataframe.
- The raw dataset contained 69 columns and 270941 rows
- The Location_Easting_OSGR, Location_Northing_OSGR, Longitude and Latitude had 120 missing values, for which the rows of missing values were dropped. LSOA_of_Accident_Location had 15762 missing values, for which entire row was dropped for the analysis.
- The final dataset contained 68 rows and 270787 rows

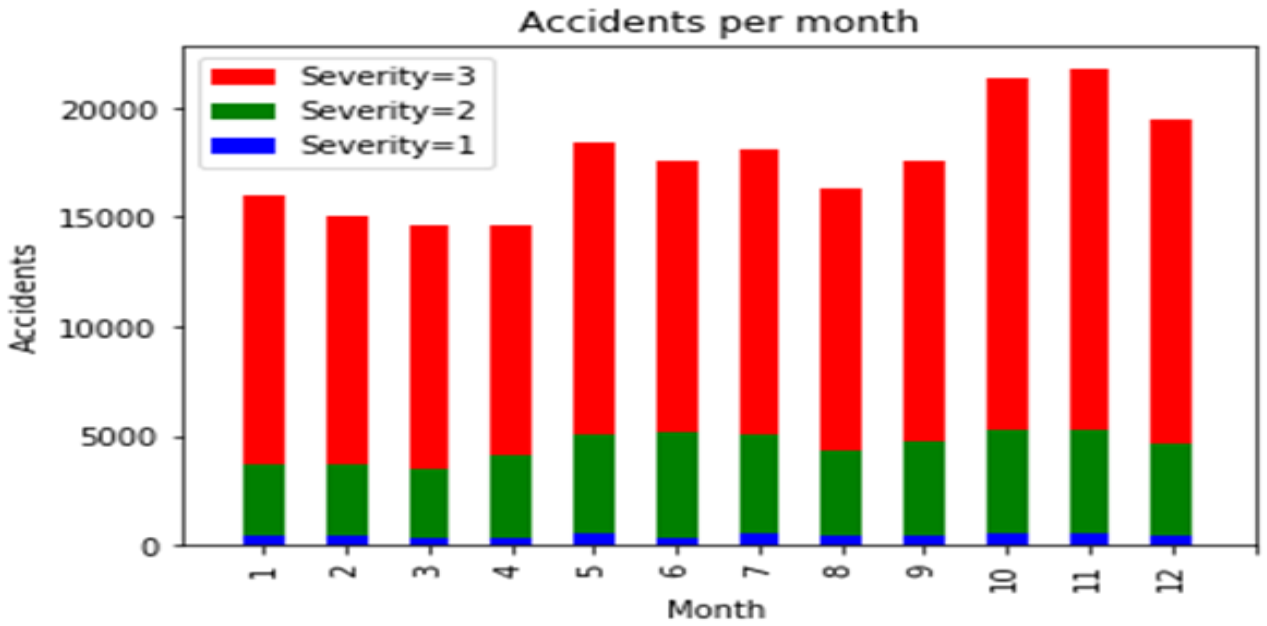
Number of accidents by hour, month and day of a week





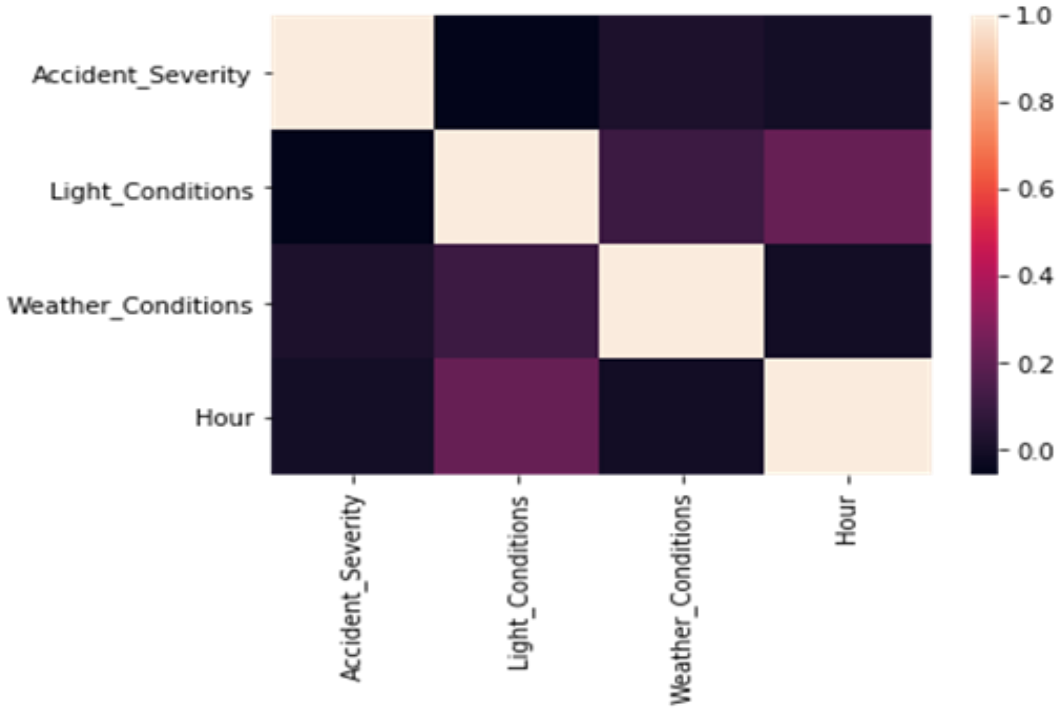


Accident Severity and Month



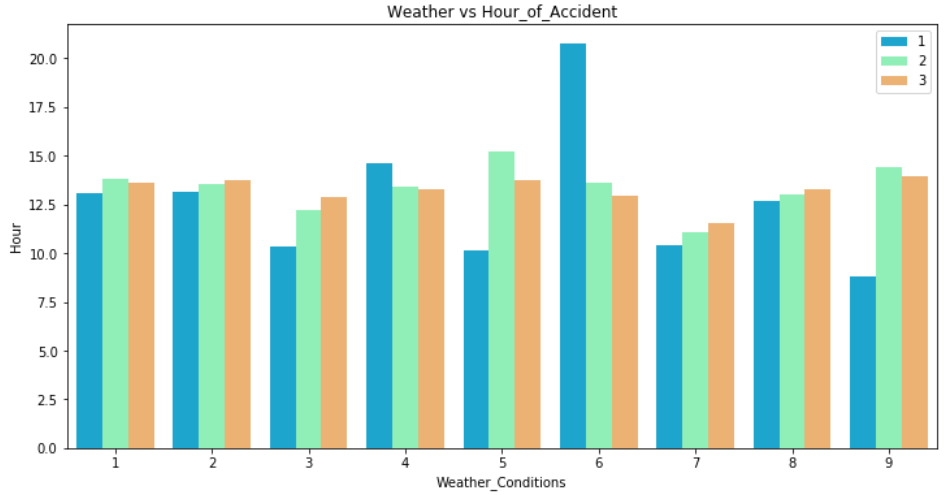
- The graph showed the accident severity = 3 were most in every month whereas severity = 2 was minimum in every month and severity = 3 was negligible compared to others
- Here severity = 1 is fatal, severity = 2 is serious and severity = 3 is slight

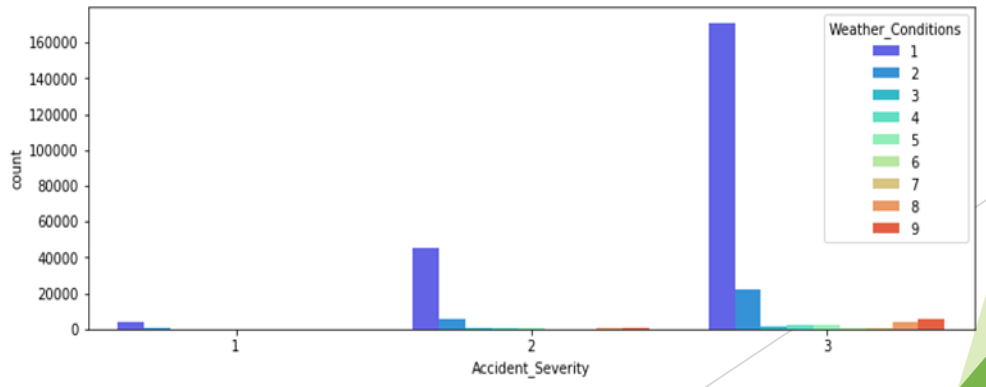
Correlations between accident severity, weather condition, light condition and hour.



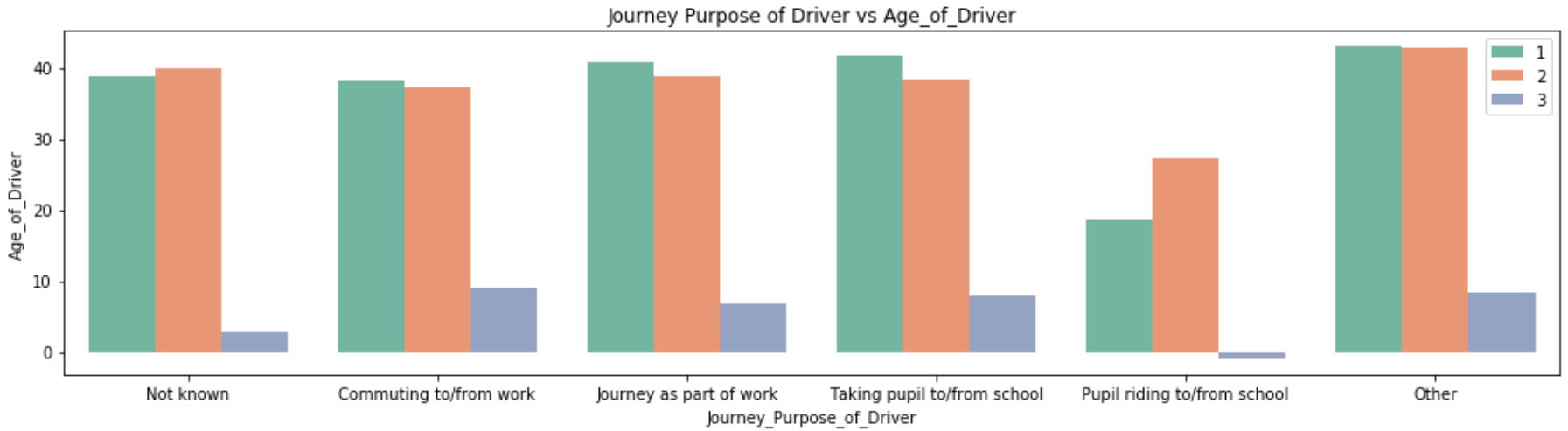
Accident Severity And Weather Conditions

- accident usually occurred during afternoon due to weather condition.
- The accidents with slight severity occurred the most
- Accidents usually took place when the Weather conditions were fine and also there weren't any high winds: meaning which the weather conditions didn't effectively contribute to occurrences of accidents.





relation between journey purpose of the driver and their age and sex



Predictive Modelling

- Month , Hour , Day of Week , Longitude , Latitude , Age of Driver , Sex of Driver , Weather Conditions and Light Conditions data were used to predict the accident severity
- Decision tree (Random Forest), Logistic Regression and K-Nearest Neighbour were used for predicting the severity
- random forest was done with 100 decision tree, logistic regression with c =0.001, solver as libilinear and for KNN, the value of k was chosen as 19

Results

- The metrics used to compare the accuracy of the model were jaccard score, f1-score, precision and recall.
- Their weighted average were taken.
- From the table, we can say that Random Forest is the best model to predict the severity of the accident.
- The logistic regression and KNN model showed similar accuracy.

Algorithm	Jaccard score	F1-score	precision	recall	time
Random Forest	0.89	0.87	0.90	0.88	41.34757
Logistic Regression	0.78	0.68	0.61	0.78	3.51829
KNN	0.78	0.71	0.71	0.78	3.63040

Conclusion

- The model was built to predict the accident severity of the road accident.
- The random forest was the best model with highest accuracy.
- The data used was the road traffic accident of United Kingdom
- This analysis can be used for future prospect