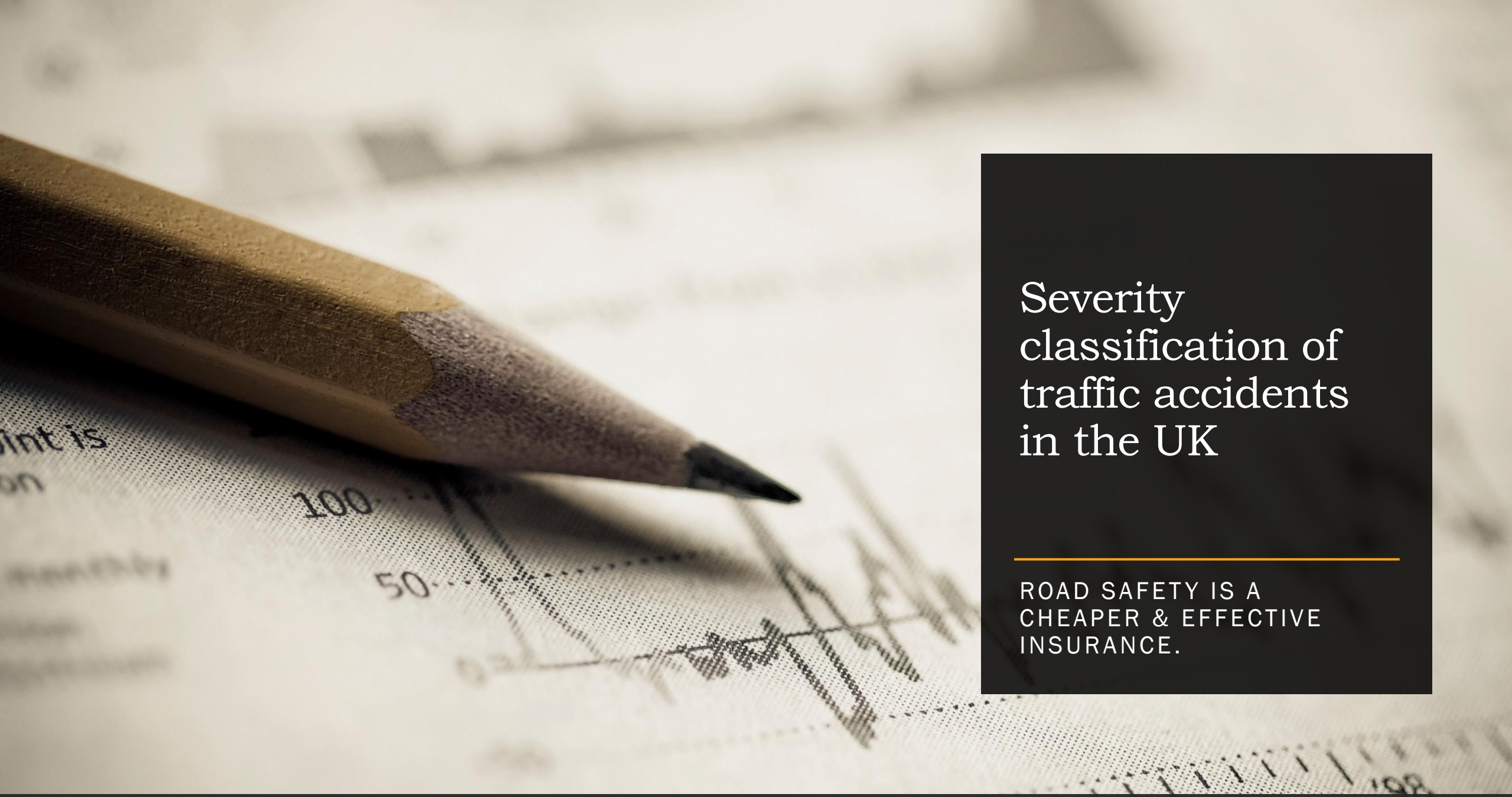


Severity classification of traffic accidents in the UK

ROAD SAFETY IS A CHEAPER & EFFECTIVE INSURANCE.

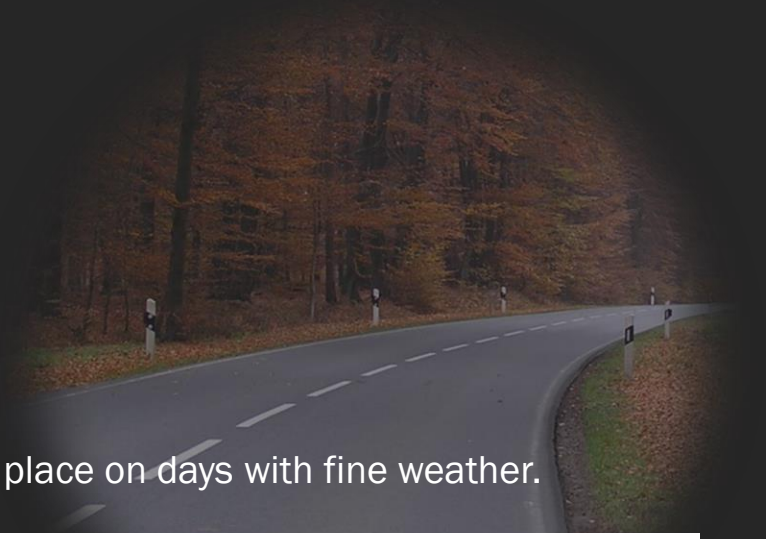
Severity classification of traffic accidents in the UK

- Severity of a traffic accidents can be classified into 3 categories, Slight, Serious and Fatal.
- Understanding the reasons for an accident to fall into one of the three categories can greatly improve road safety.
- Traffic accidents may be caused by weather, road conditions, inattentive drivers, road design and numerous other contributing factors.
- Proper interpretation of such contributing factors could lead to safer road design, safer cars and/or better driving habits leading to safer roads for drivers and pedestrians.

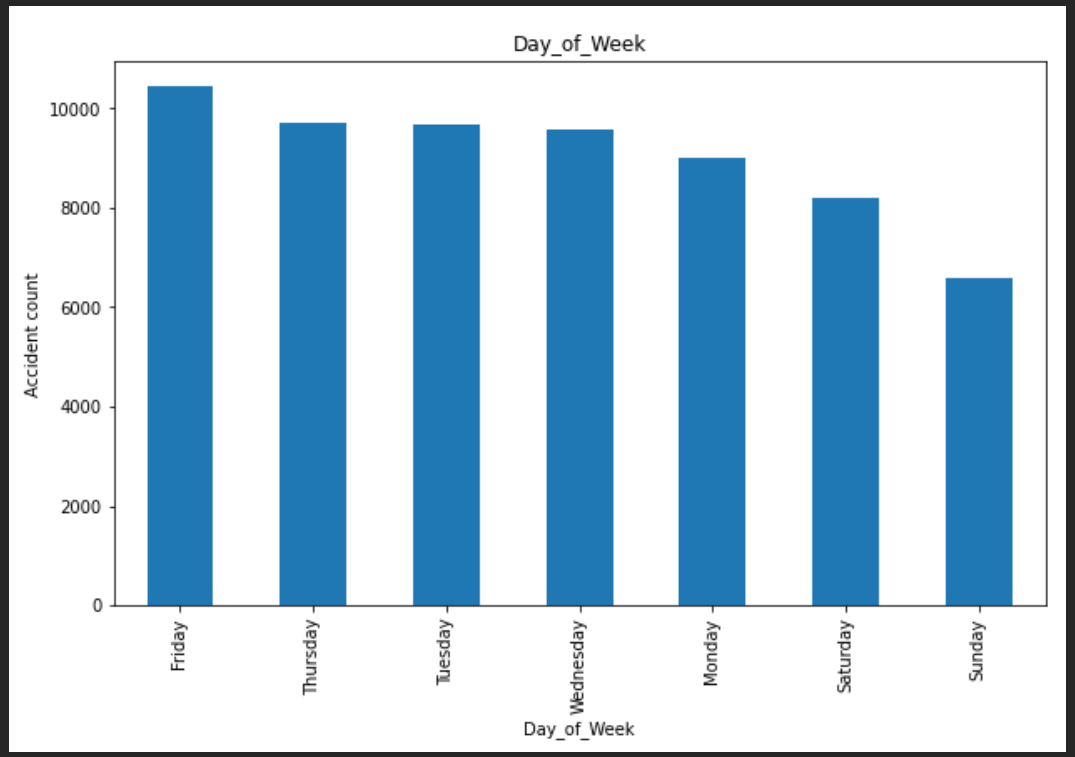
Data acquisition and cleaning

- The dataset used for this study, made available by the government of UK, contains records for over 2 million accidents with 34 attributes to each of them.
- The dataset is bound to the geographical boundaries of the UK (including Scotland) and time frame from 2005 to 2017.
- Out of 34 attributes, 11 attributes were selected to create a feature set.
- The data set was imbalanced, the three categories (Slight, Serious, Fatal) are distributed at 84.7%, 14% and 1.3% respectively.

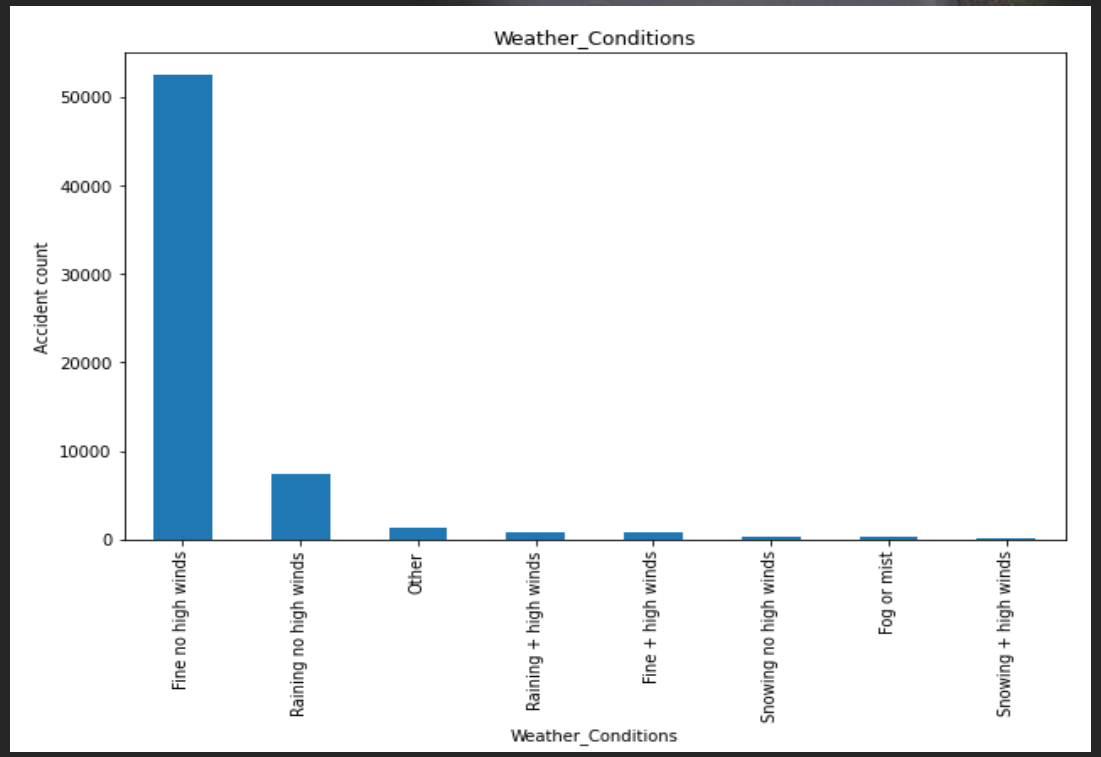
Data Exploratory Analysis

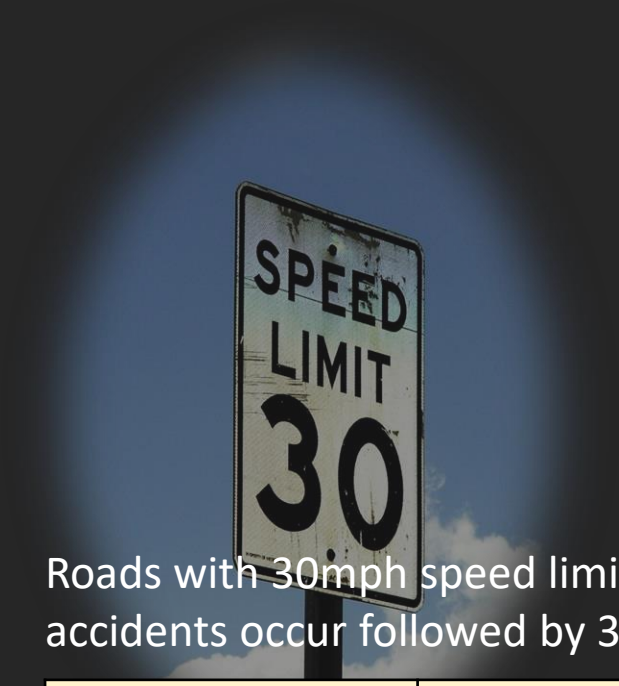


The highest number of accidents (335,183) took place on Fridays and the lowest being 225,327 on Sundays.



80% of accidents took place on days with fine weather.



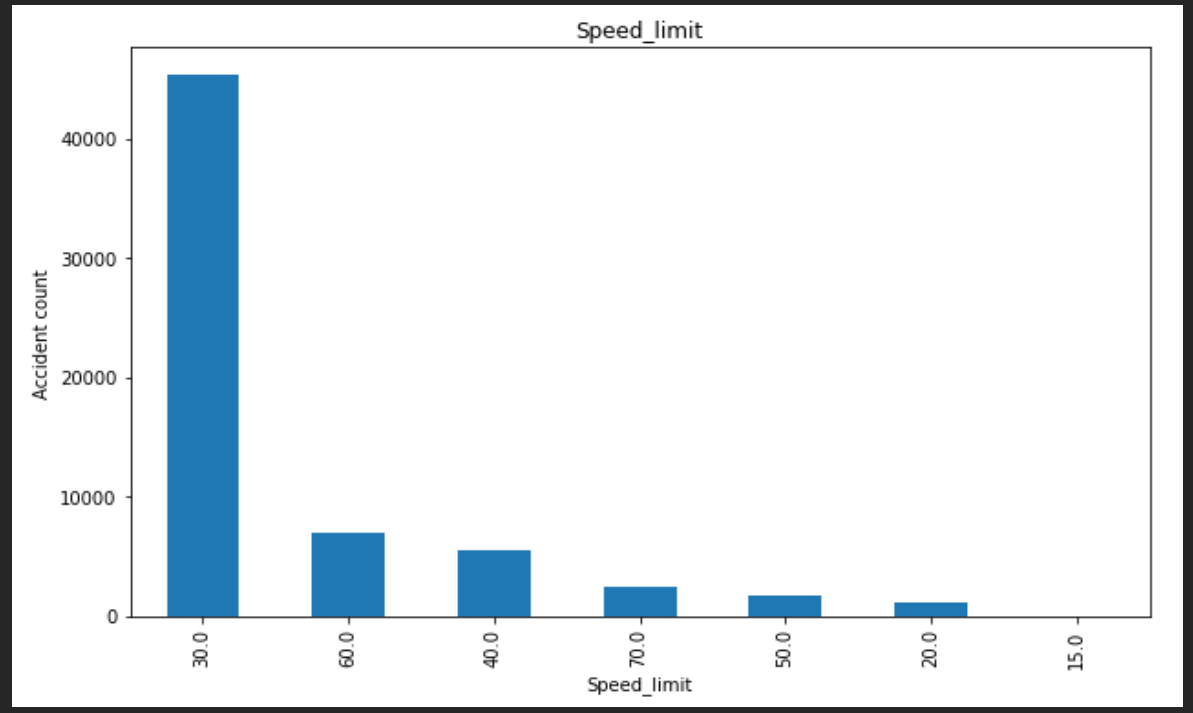


Data Exploratory Analysis

Roads with 30mph speed limits are where 42% of Fatal accidents occur followed by 31% on roads with 60 mph.

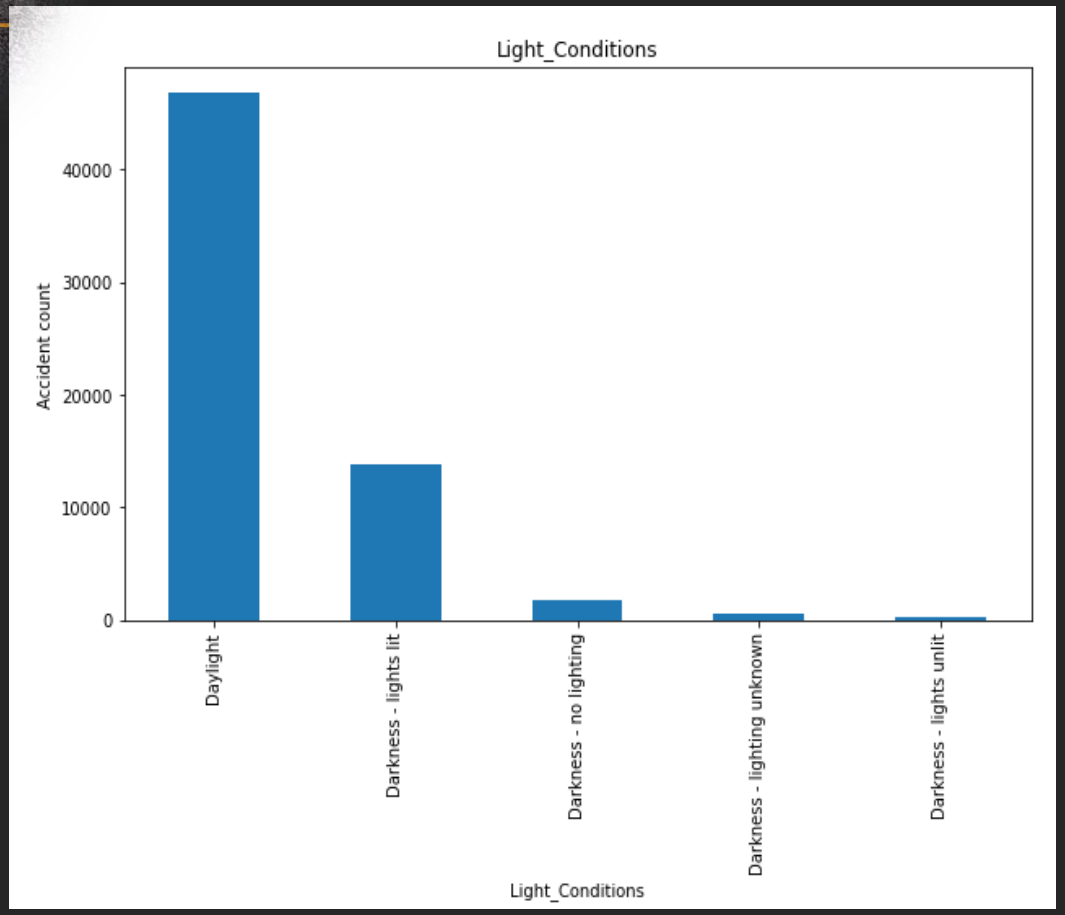
Road speed limit	Accident Severity		
	Fatal	Serious	Slight
15	N/A	0.0001	N/A
20	1.61%	1.99%	1.85%
30	41.86%	68.61%	72.52%
40	10.20%	8.48%	8.82%
50	4.65%	2.90%	2.73%
60	31.31%	14.70%	10.27%
70	10.38%	3.30%	3.81%

Vast majority of all accidents take place on roads with 30 mph speed limits.

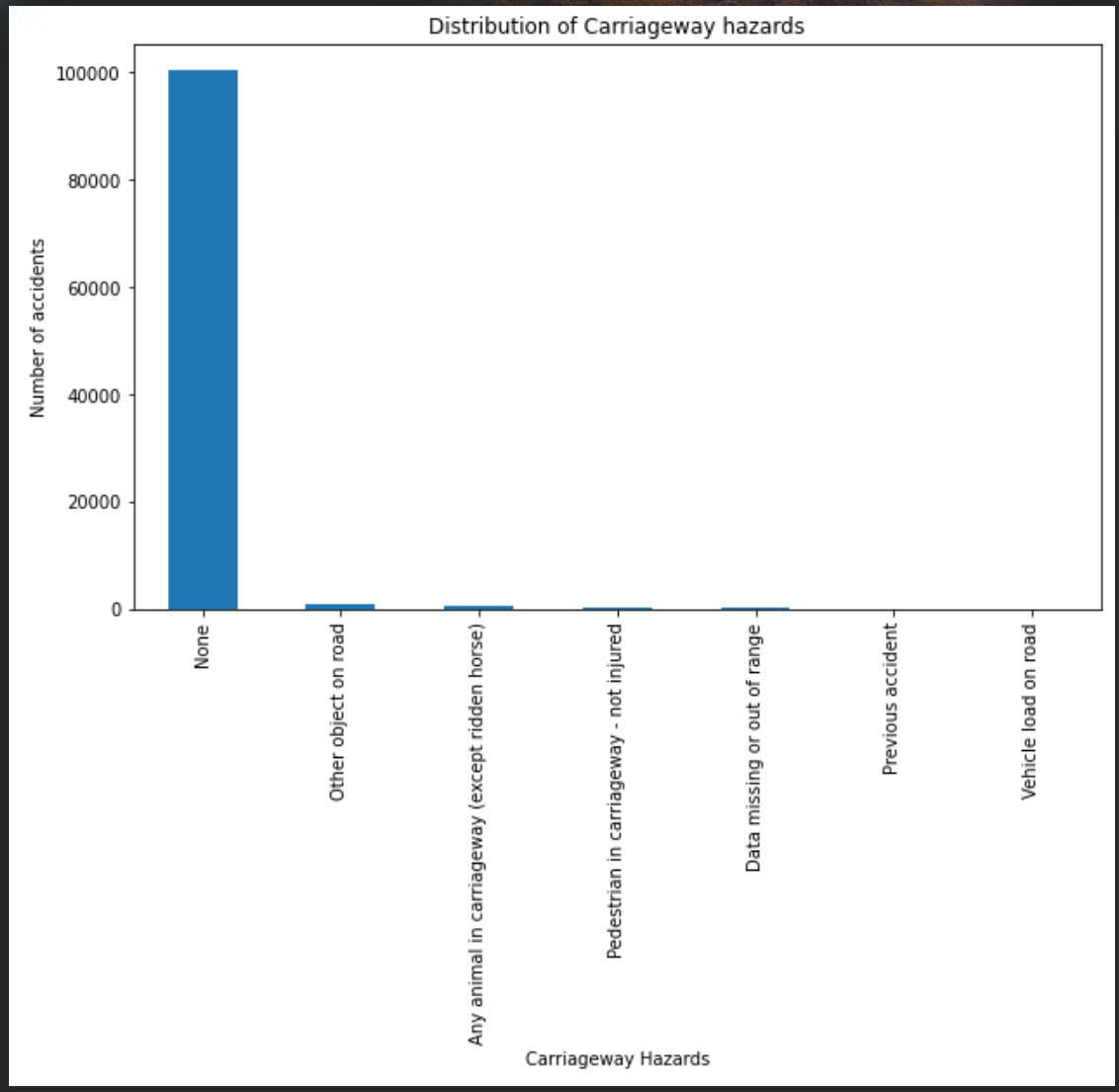


Data Exploratory Analysis

73% of all accidents take place in broad daylight



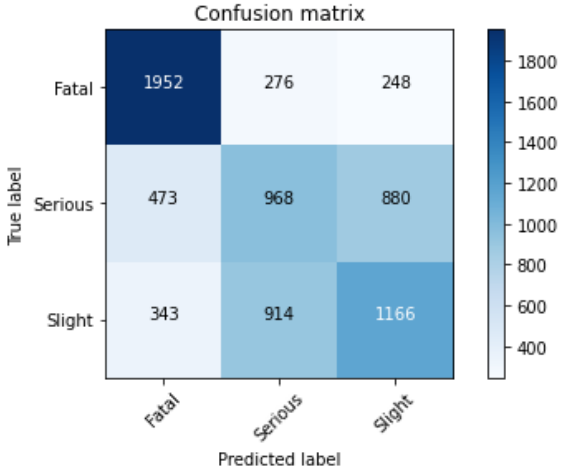
Almost all accidents take place in the absence of any apparent road hazards



Modelling

	precision	recall	f1-score	support
Fatal	0.71	0.79	0.74	2476
Serious	0.45	0.42	0.43	2321
Slight	0.51	0.48	0.49	2423
accuracy			0.57	7220
macro avg	0.55	0.56	0.56	7220
weighted avg	0.56	0.57	0.56	7220

```
Confusion matrix, without normalization
[[1952  276  248]
 [ 473  968  880]
 [ 343  914 1166]]
```



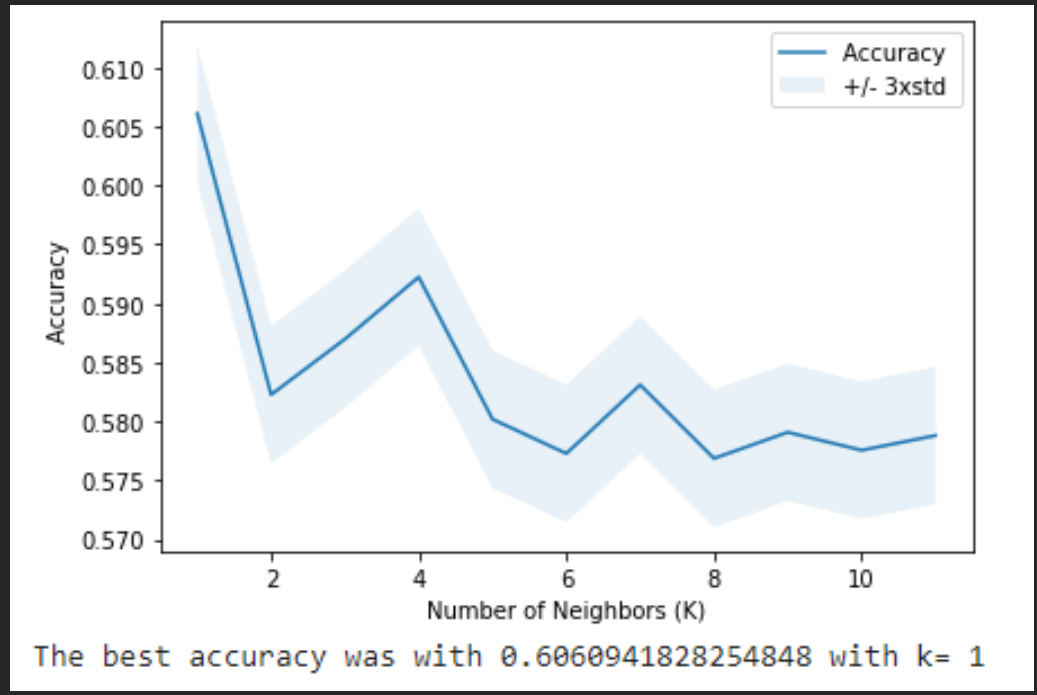
SVM model (confusion matrix to the left) was able to correctly predict 1952 Fatal severity cases out of 2476 with accuracy of 0.74. However, accuracy for predicting Slight and Serious severity cases were lower with accuracy of 0.49 and 0.43 respectively. The average accuracy of the model was at 0.57.

Logistic Regression model achieved around 0.47 accuracy.

Decision Tree model achieved around 0.64 accuracy.

K Nearest Neighbor(KNN)

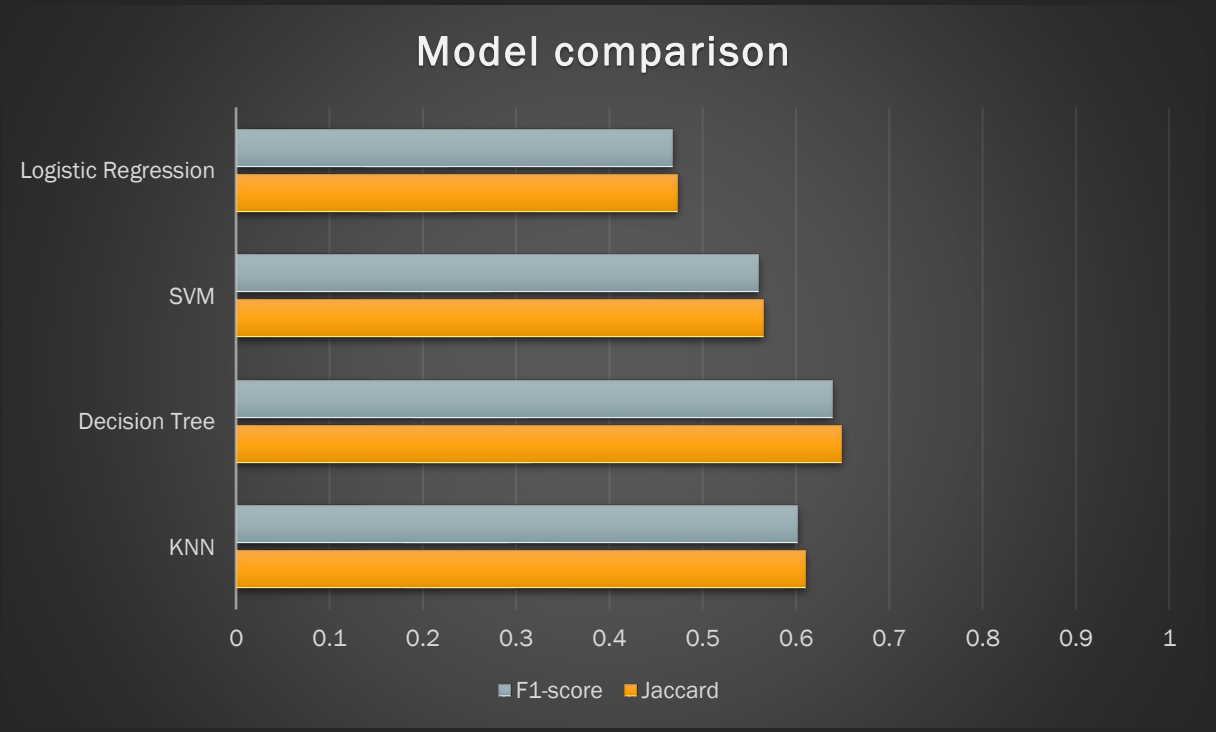
Best K value was identified to be 1 with an accuracy score of 0.6



Model performance comparison

The table and the graph show that the decision tree model has achieved the best accuracy, closely followed by KNN model.

Algorithm	Jaccard	F1-score	Log Loss
KNN	0.610942	0.601854	NA
Decision Tree	0.649169	0.639322	NA
SVM	0.565928	0.560169	NA
Logistic Regression	0.473546	0.468204	1.01672



Conclusion

- The decision tree model can classify accident severity with an accuracy of up to 64%.
- The SVM model can predict fatal accidents with an accuracy of up to 74%.
- The study has shown that most fatal accidents take place during daytime, fine weather, good road conditions and on low speed limit roads, this suggests that reckless driving could be the cause behind most accidents, specifically, the fatal ones.