

Predicting Traffic Accident Severity

Applied Data Science Capstone

Jyotsana Rathore

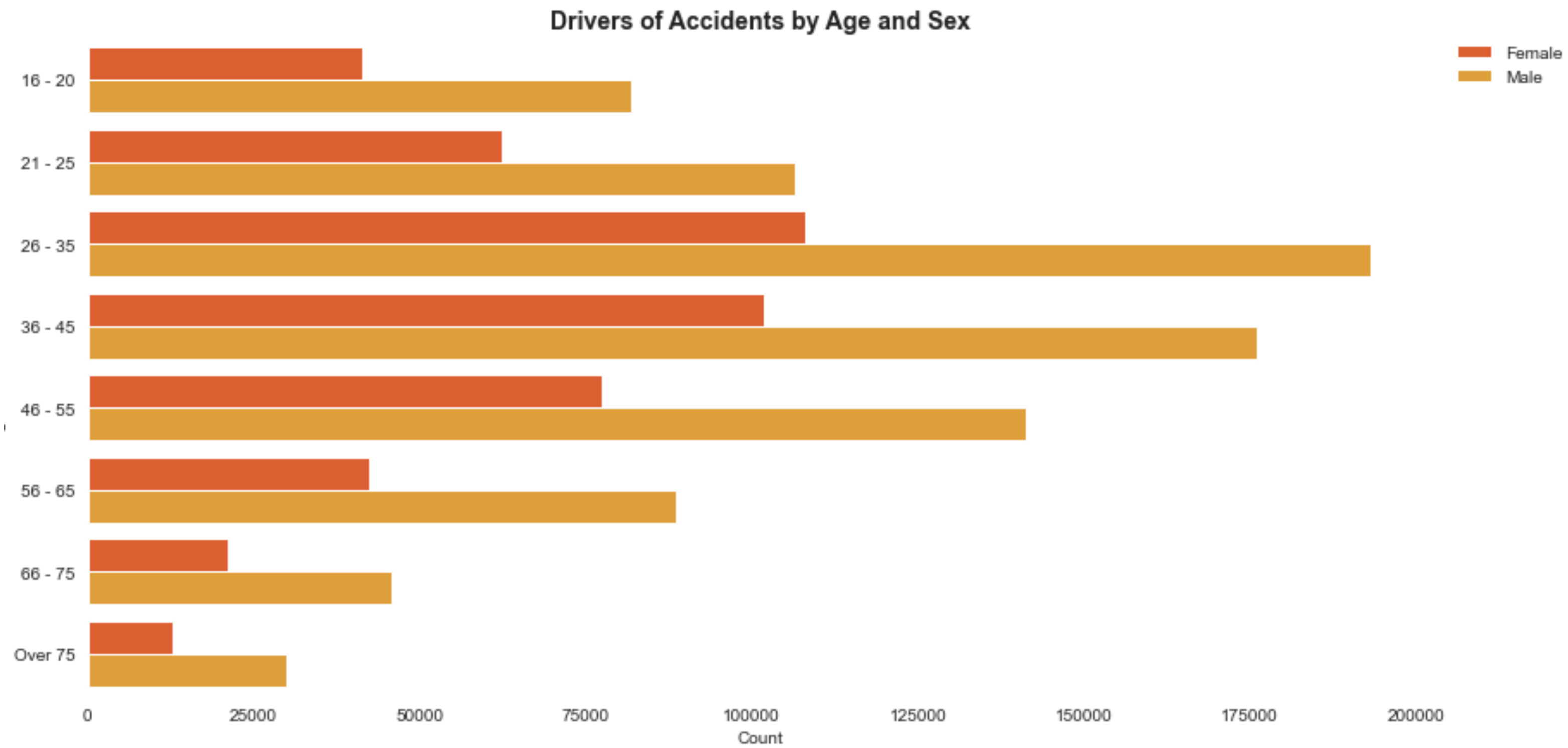
INTRODUCTION:

- Traffic accidents are extremely common, and it is more frequent in the sprawling metropolis. Because of their frequency, traffic accidents are a major cause of death globally, cutting short millions of lives per year
- By 2017, the total number of cars in the UK is 31,200,182, indicating that per 1000 people have 471 motor vehicles, rank 35th in 160 countries. However, road accidents happen every day
- In 2018, there were a total of 160,597 casualties of all severities in road traffic crashes.
- The potential to predict the accident severity (e.g. What will cause accidents? What will make accidents worse?) is therefore useful not only to public safety stakeholders but also transportation administrators and individual travellers
- A system that can predict the cause of traffic accidents and predict the accident severity can potentially save lives and aid in developing better roadway designs.

DATA DESCRIPTION

The dataset on road traffic accident analysis has about 2 million rows and 63 columns.

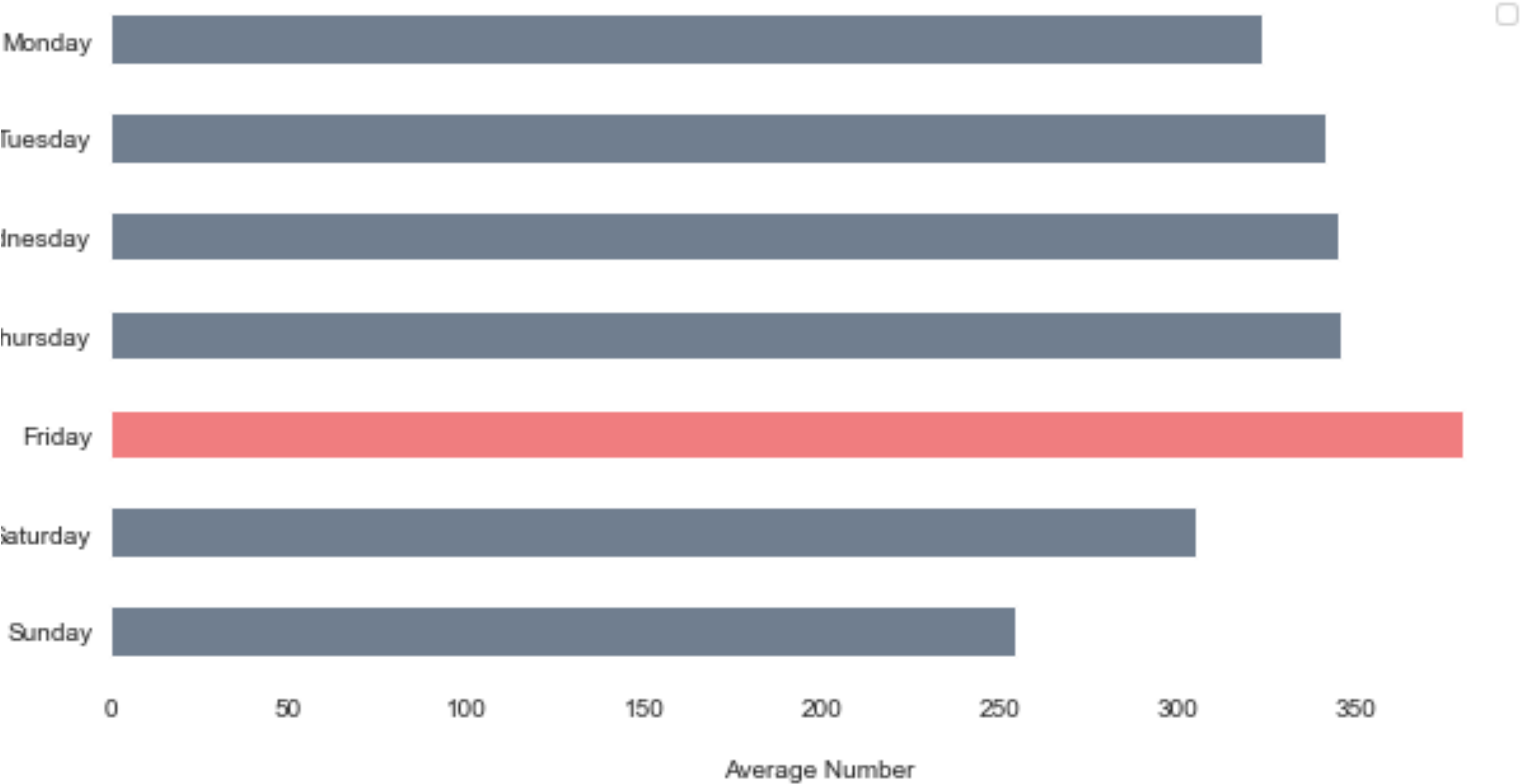
The predictor variables in the dataset had information geographical locations, weather conditions, type of vehicle, number of casualties and vehicle manoeuvres. The dataset had missing values which were dropped.



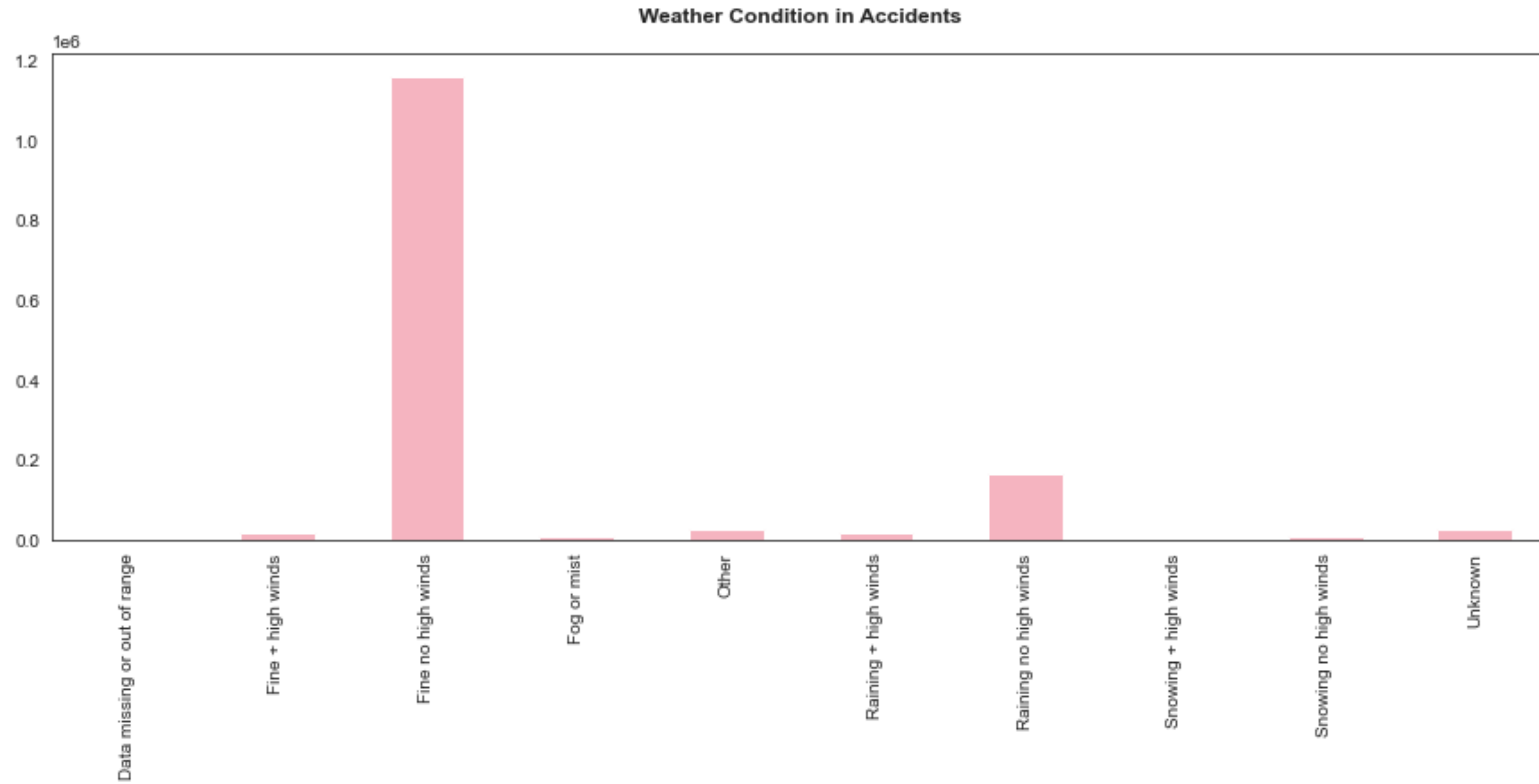
During the exploratory data analysis part, it was found some compelling details like the age group between 26 years to 35 years was more frequent with accidents than any other age group.



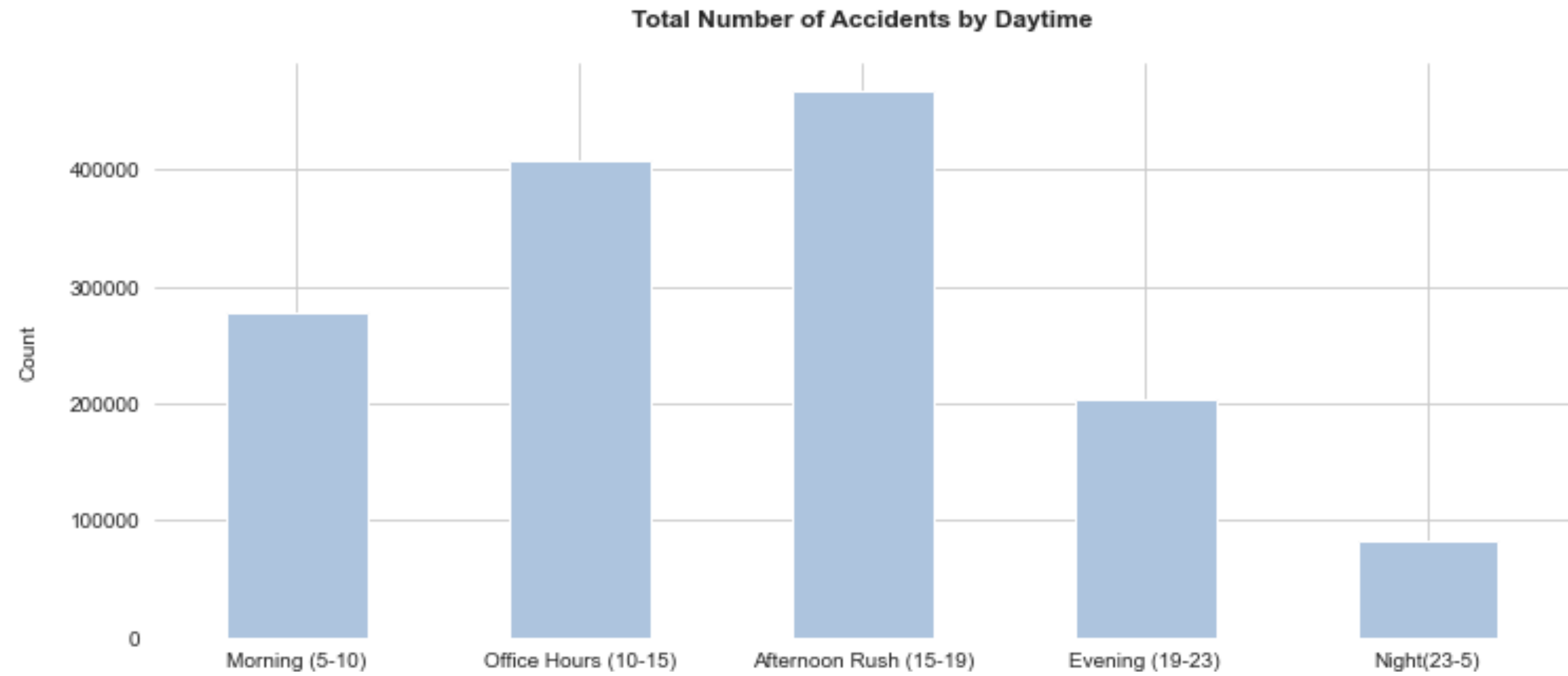
Average Accidents per Weekday



Most of the accidents happened on Friday



weather conditions did not play much role at the time of accident.



Most of the accidents happened during the afternoon rush that is between 3.00 pm to 7.00 pm

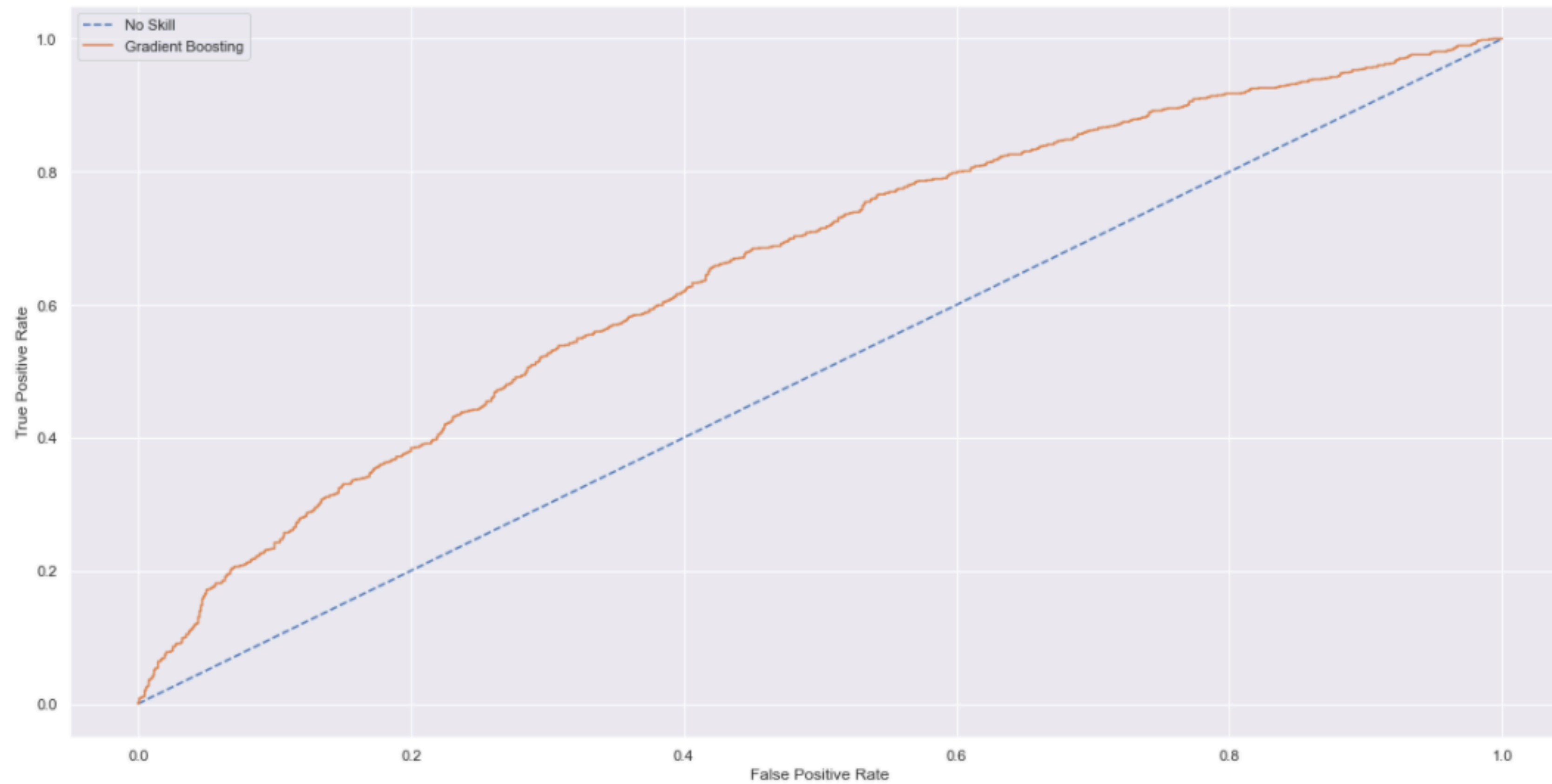
Models Description

The final modelling dataset of 381K rows and 62 columns was split into training data with 70% and test data with 30%

Spot-checking technique was implemented to determine the machine learning model which was best suited to predict the accident severity

Hyperparameter tuning was used to modify parameters in all the 6 models which are Logistic Regression, Random Forest classifier, Gradient Boosting Classifier, Linear Discriminant analysis, Extra Trees classifier and Bagging Classifier to improve the performance of the models.

No Skill: ROC AUC=0.500
Gradient Boosting: ROC AUC=0.655



Result

Conclusion

The problem of traffic accident severity prediction using Road traffic accident analysis data:

This was a challenging problem due to class imbalanced as the majority of the accident severity belonged to Slight accident severity. It was surprising to see that weather and light conditions did not have much impact on the fatal and severe accident severity. It was observed that during the permutation testing, features like engine capacity, number of vehicles, speed limit, age group of drivers played an important role for improving the accuracy of the Gradient boosting classifier model.