

GAURAV AMRUTKAR

CAR ACCIDENT SEVERITY PREDICTION

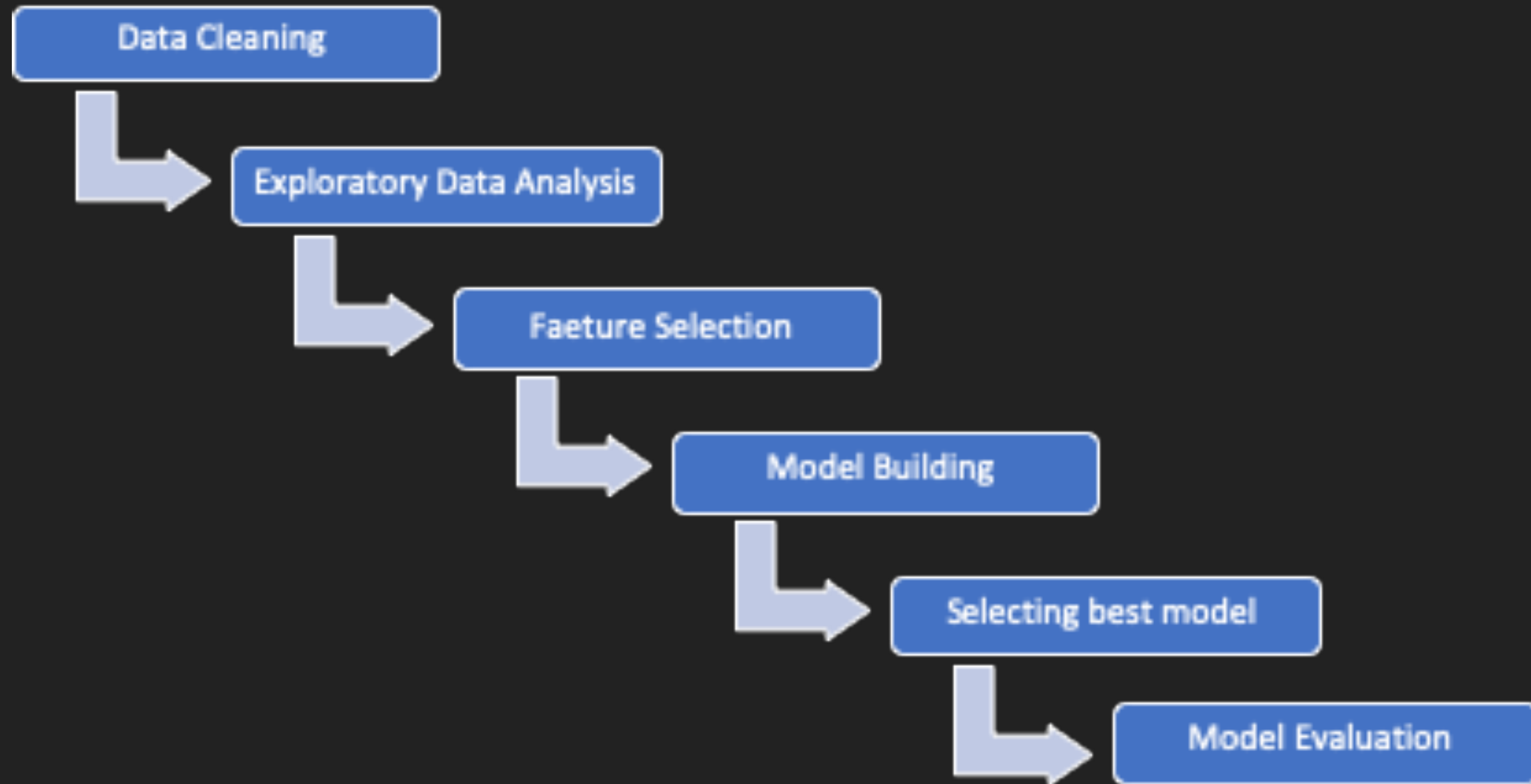
INTRODUCTION

- ▶ The objective is to reduce the frequency of car collisions in a community, a model must be developed to predict the severity of an accident given the current weather, road and visibility conditions.
- ▶ When conditions are bad, this model will alert drivers to remind them to be more careful.

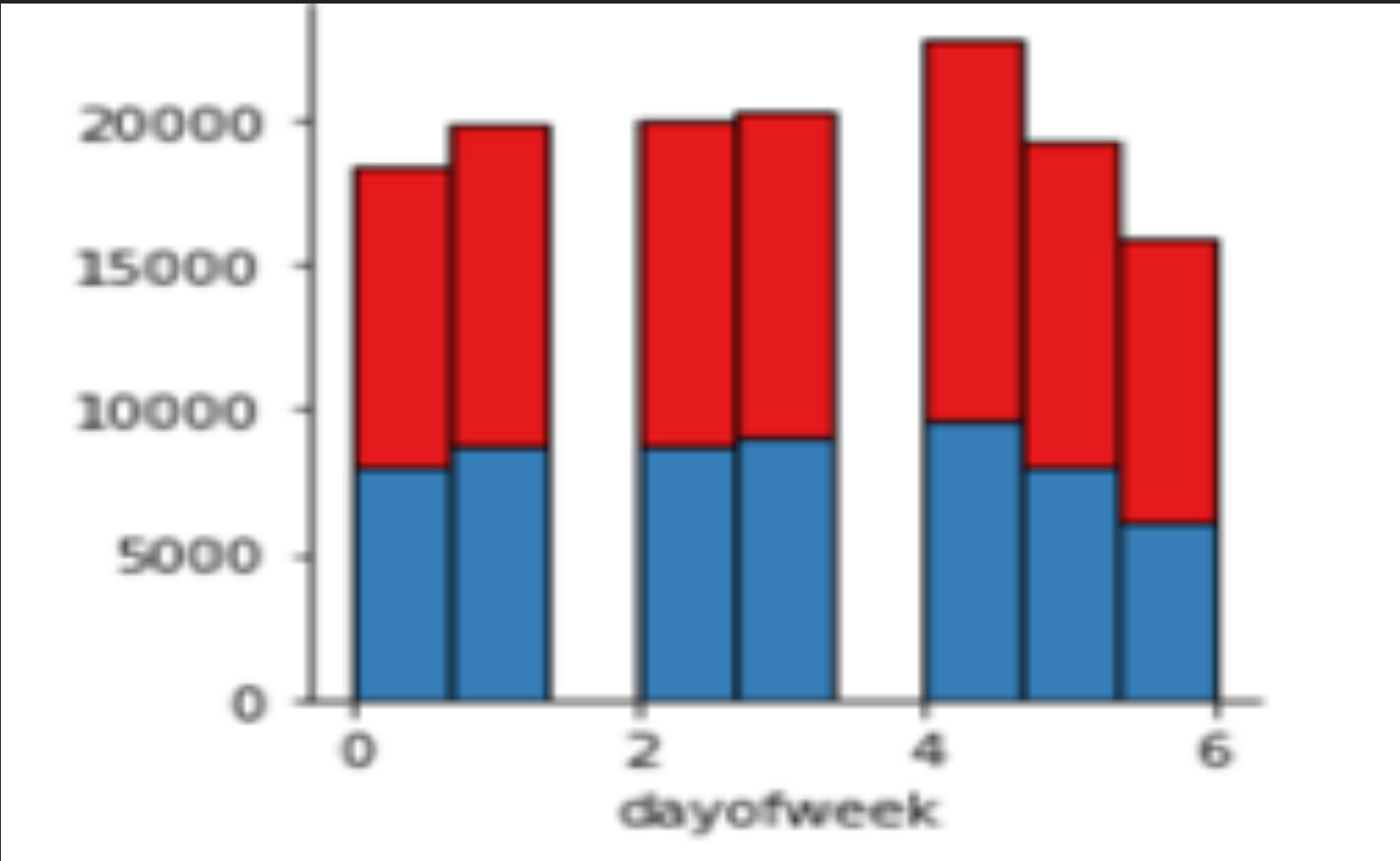
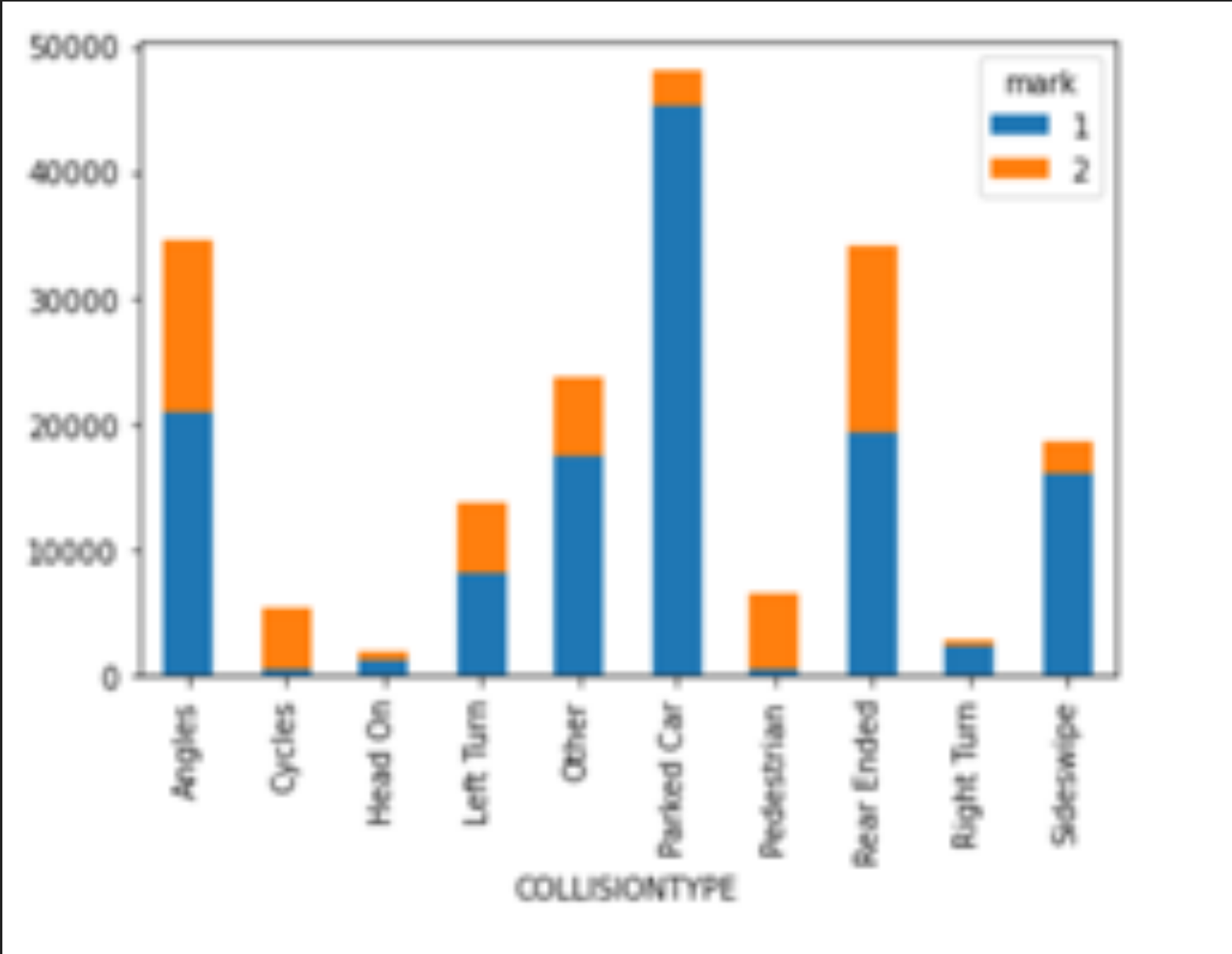
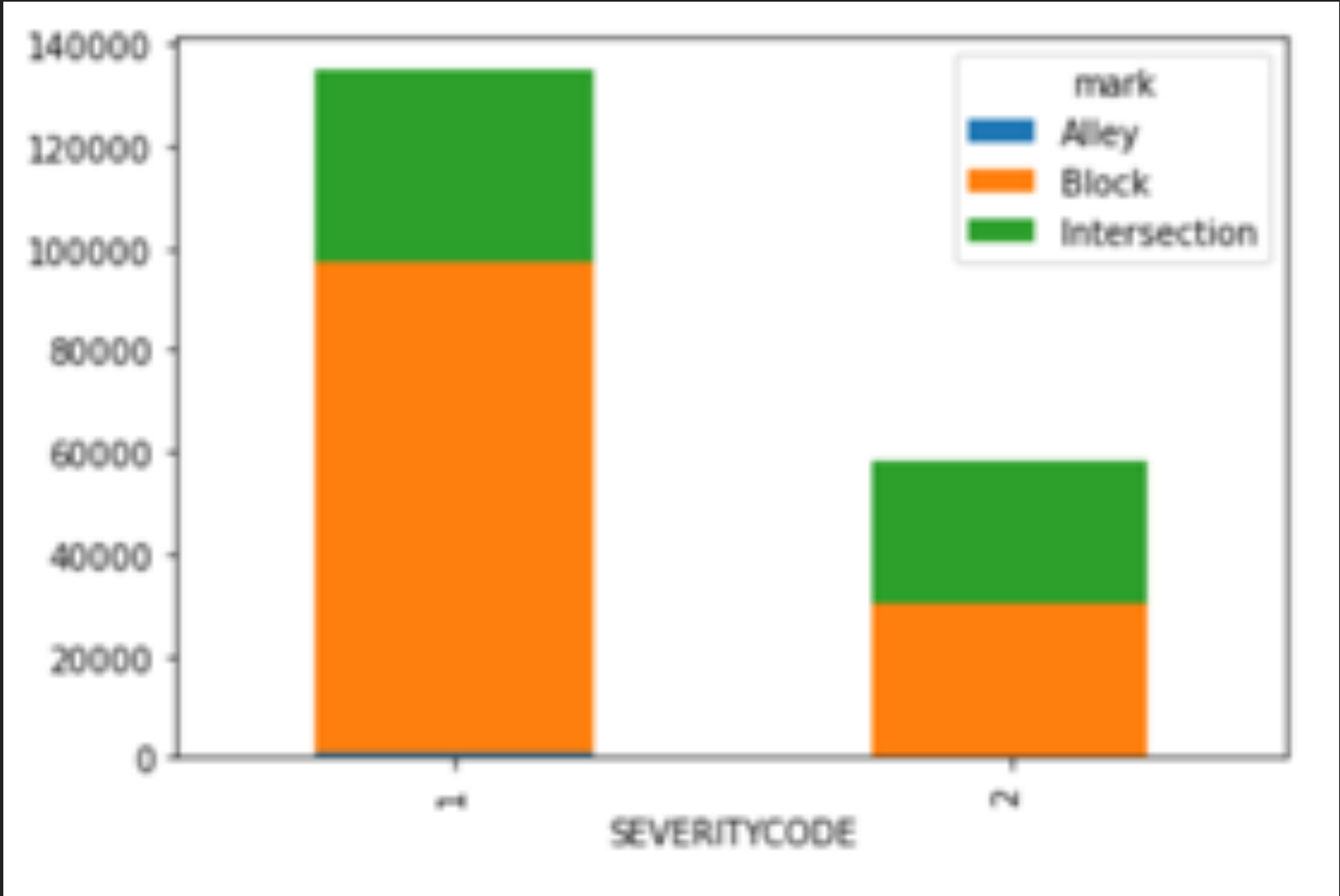
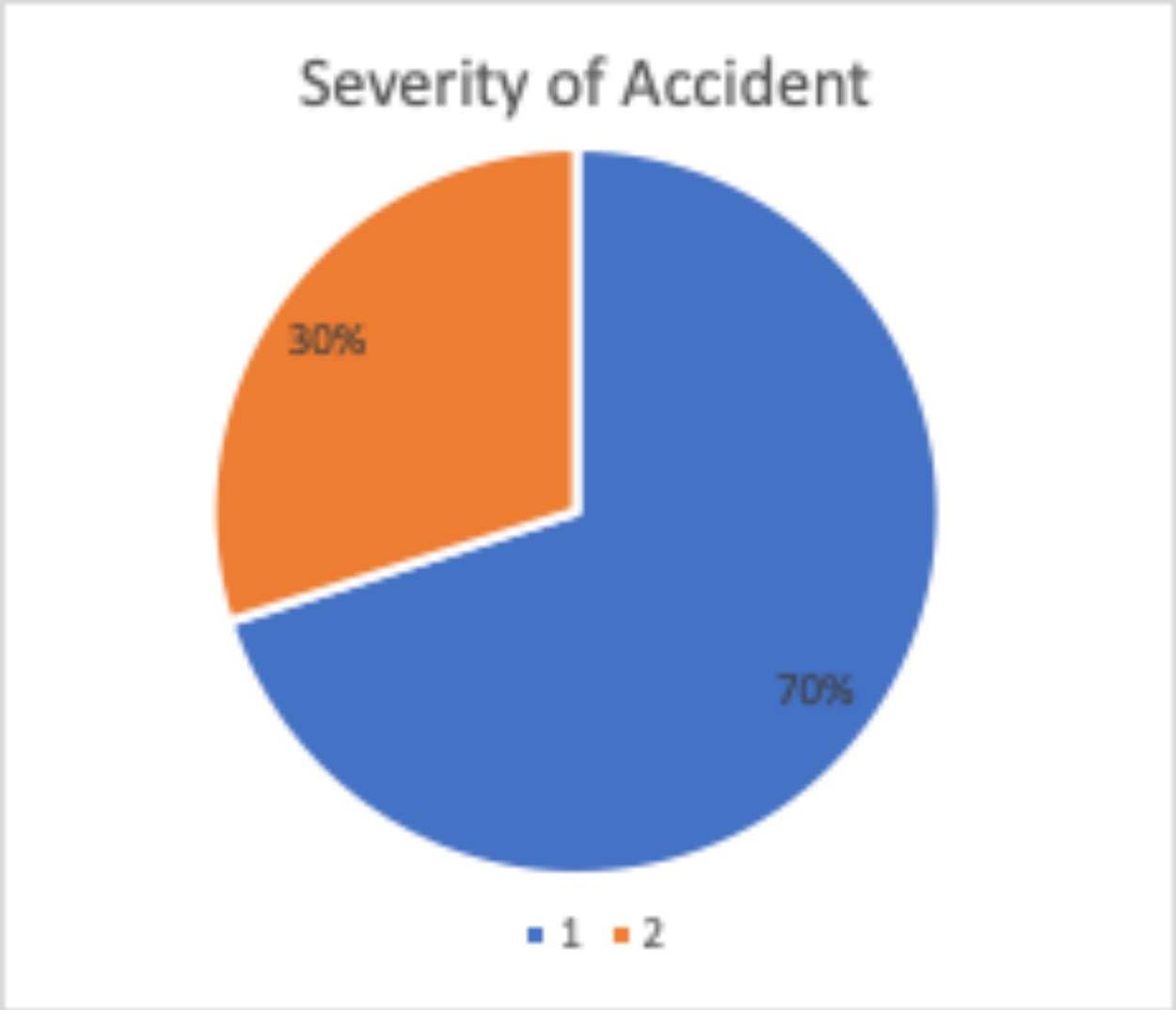
DATA CLEANING

- ▶ The data is open source in which the target variable is the severity of accident
- ▶ There are around 37 independent variables in the data which can be used for the prediction of severity
- ▶ For the categorical independent variables, we have seen its distribution with the severity of the accident and grouped the values of high cardinality categorical variables. The details of this distribution are mentioned in the next section of Exploratory data analysis.
- ▶ The box plot of the numeric variables was created and the values of this variables were restricted to 1% to 99% of the total distribution in order to deal with the outliers.

METHODOLOGY



EXPLORATORY DATA ANALYSIS



PREDICTIVE MODELLING

Classifier	F1 Score
KNN	0.70
Decision Tree	0.70
SVC	0.69
Logistic regression	0.70
Random Forest	0.71

- ▶ Out of all the models random forest has the best performance and hence is finally selected for the prediction.

CONCLUSION

- ▶ We were able to achieve ~76% AUC which is a quite better performance index. Hence by implementing this model we can reduce the number of accidents and of course significantly reduce their severity.