Prediction of Severity of car accidents

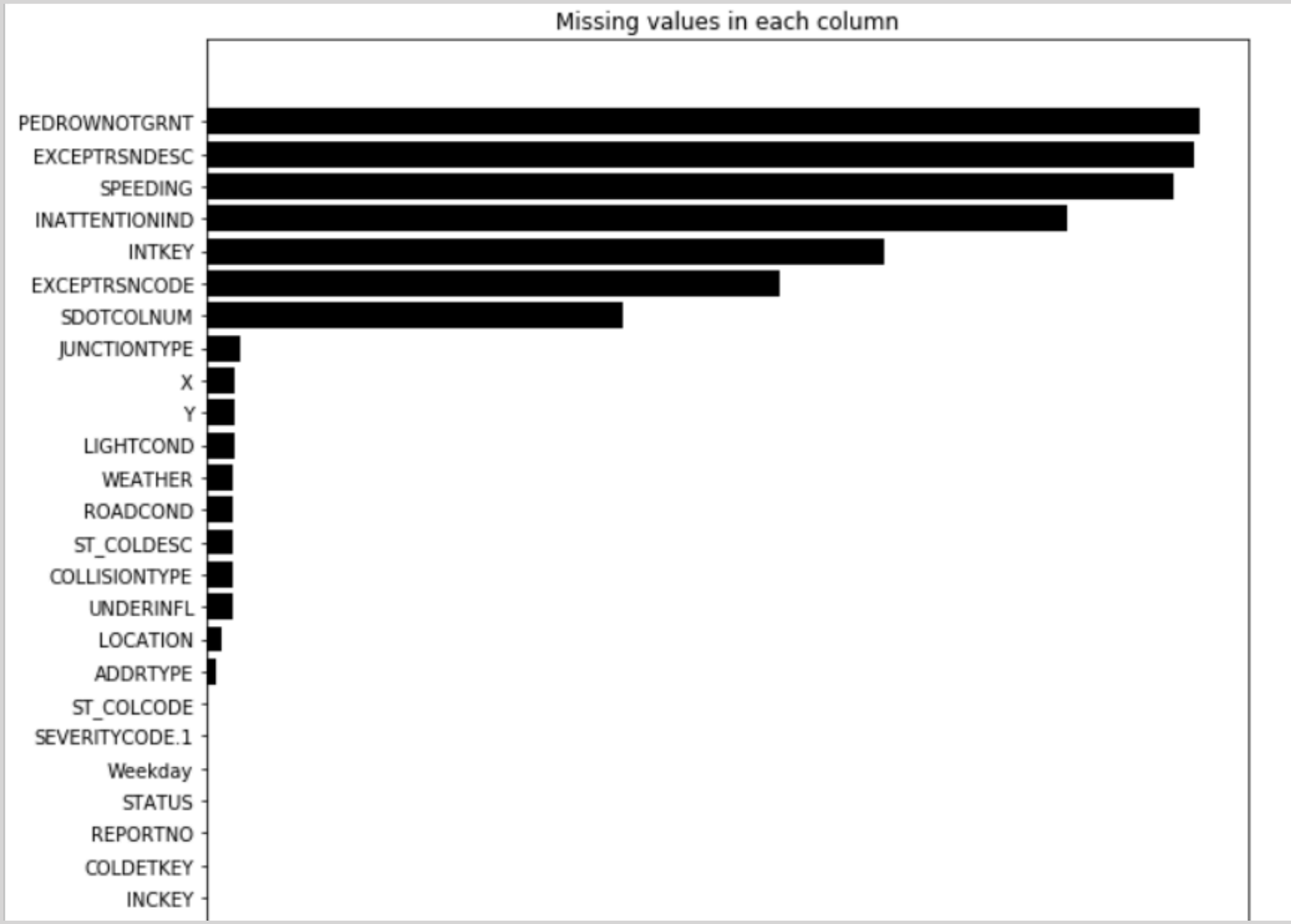
Introduction

Traffic accidents are a common occurrence that are due most of the time to a driver's actions but environment conditions and others factors related to road can have an impact of the occurrence. Some of the most common causes of car accidents include: **Speeding, Using a Device, Driver Fatigue, Drunk Driving, Defective Auto Parts, Rubbernecking, Poor Weather Conditions.**

About the weather it states that Weather conditions that leave the roadway wet or icy, or reduce visibility, pose a danger to vehicles on the road, and require drivers to pay extra attention, and to slow down. High winds, blowing dust, fog, and torrential downpours are common causes of accidents.

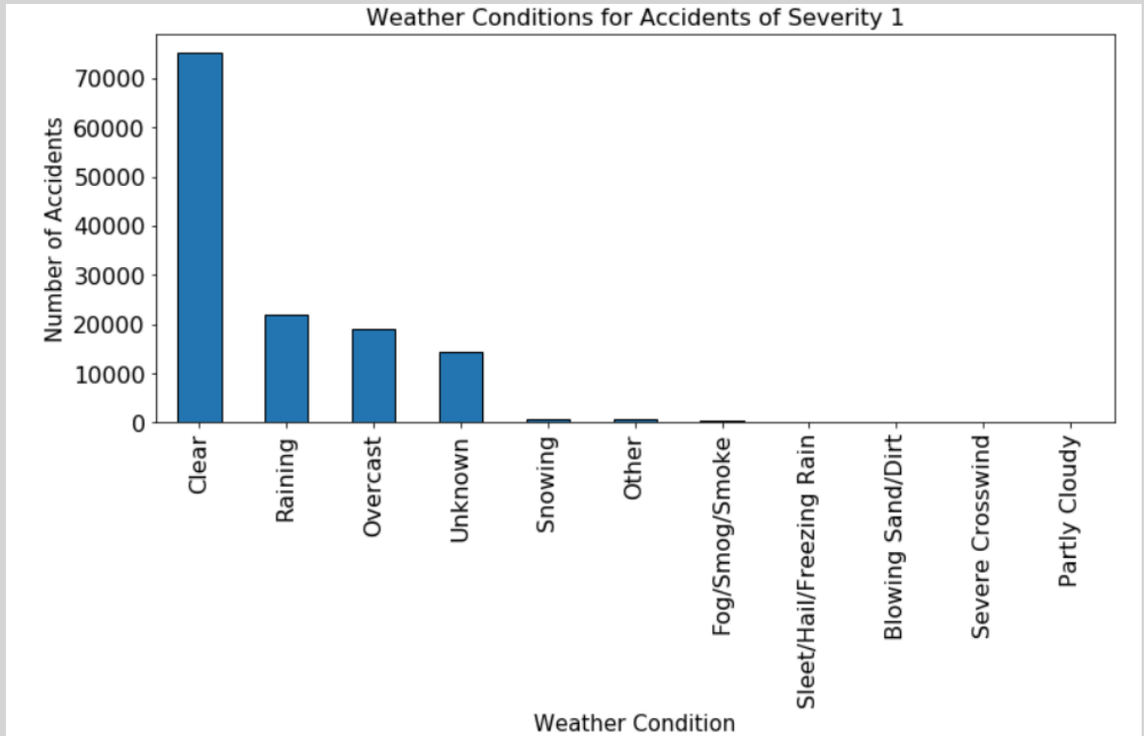
Data Preparation

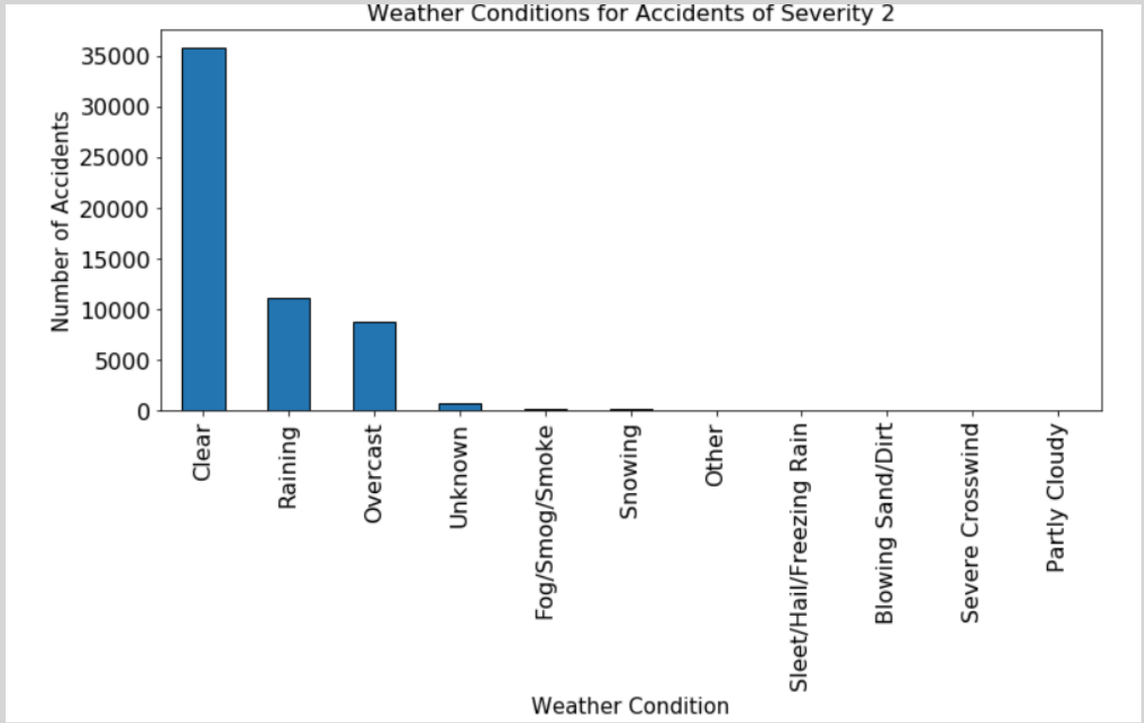
- In the dataset we did not use feature such as **SPEEDING** because it has most of the value missing
- 194,673 observations and 38 variables
- For the missing value we did not drop them but instead replace with “Unknown”

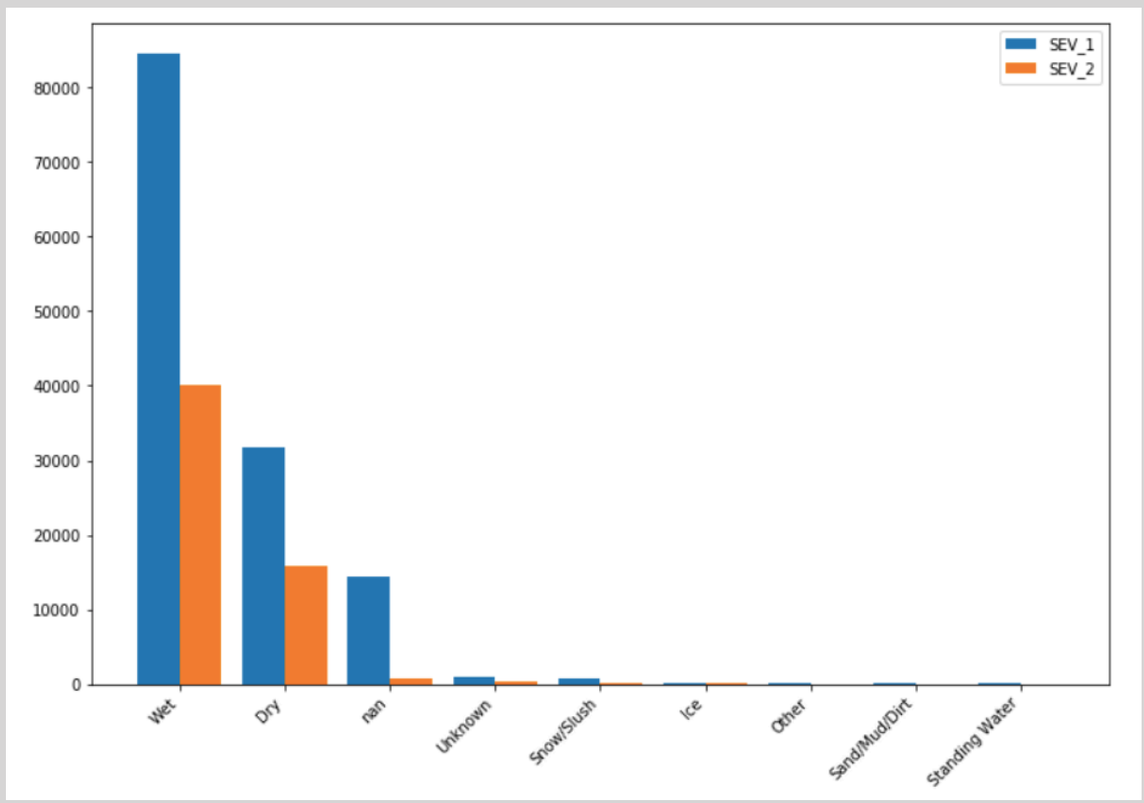


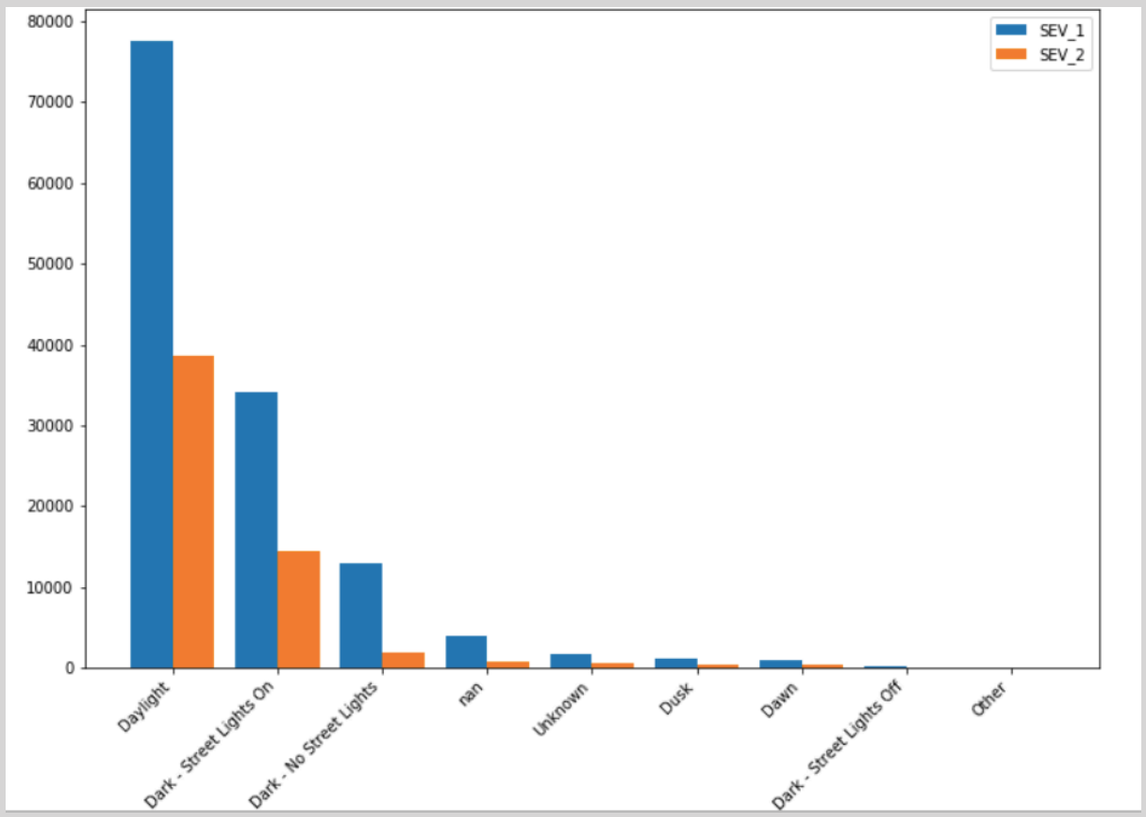
Data Visualization:

Weather Condition , Road Condition, Light Condition versus severity



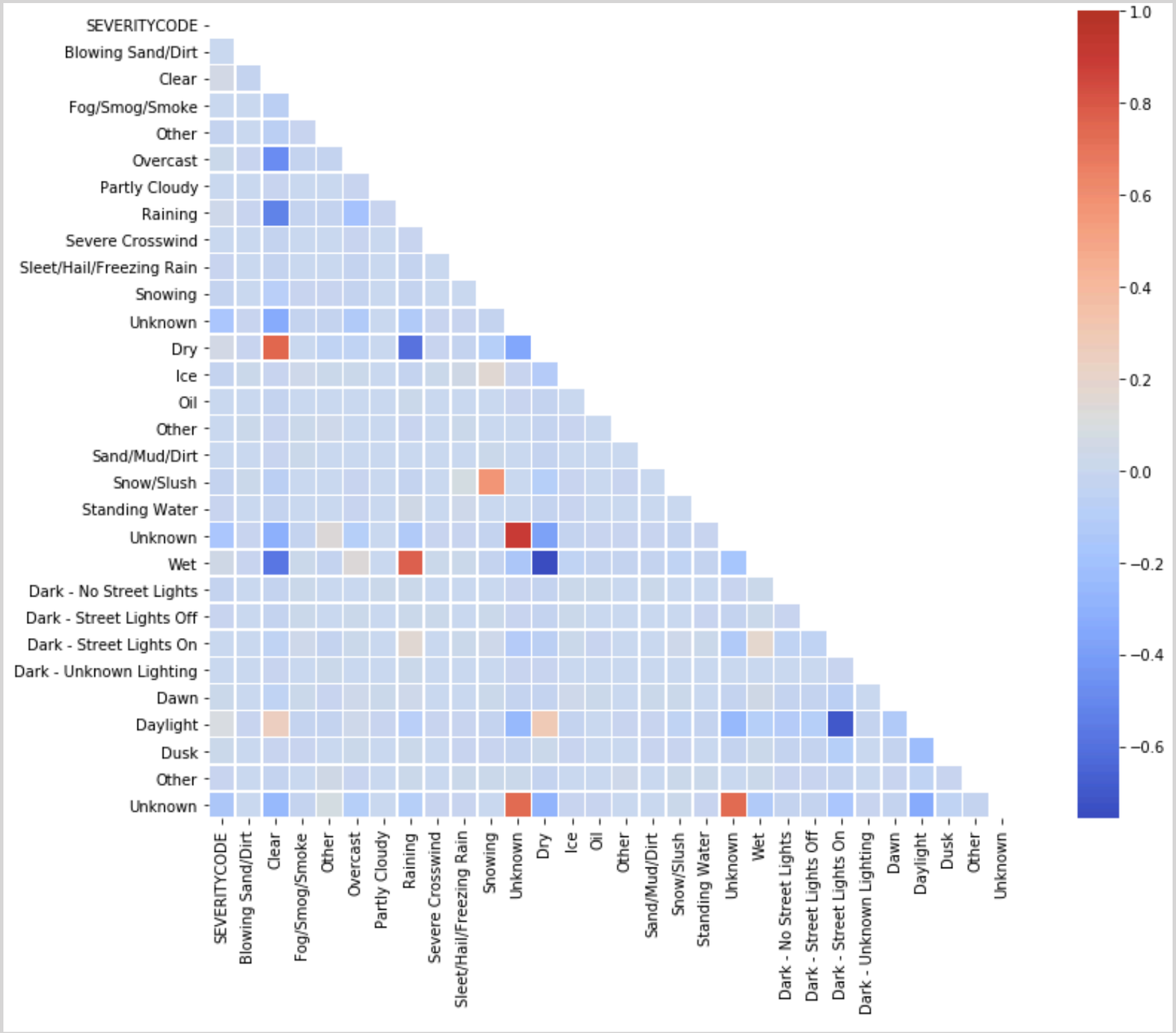






Correlation between Features

we perform a correlation matrix to see the relationship among the severity code, weather, road condition and light condition. we notice a small relation between severity code and others value but very strong relation between weather and road cold condition



Results and Conclusion

Based on the results of various algorithm (KNN; Decision Tree, Logistic Regression) it seems that the prediction of the Severity using Road condition, light condition and weather is not perfectly what we expected. This conclusion is based on the accuracy percentage being around mid-60%.

Further analysis for improvement can be made on others feature such as the junction type, the location

Algorithm	KNN	Decision Tree	Logistic Regression
Accuracy	0,558	0,556	0,559
Jaccard similality score	0,56	0,56	0,56
F1-score	0,55	0,48	0,49
Log loss	n.a	n.a	0,67