Predicting
Car Accident Severity
with Machine Learning Models

Mustafa Bal

10 October 2020

Problem Statement

- 1,35 million people die every year as result of road traffic accidents (WHO).
- 20 - 50 million people get injured every year in traffic accidents.
- Car accidents are the number one cause of death for teenagers in the USA.
- Significant economic losses.
- Using Machine Learning algorithms, Data Scientists can develop models that predict severity of traffic accidents.

Data

- Dataset is taken from Kaggle and it is originally hosted by the City of Seattle.

- It includes accident records with all types of collisions and covers accidents between the years of 2004 and 2020.

- The copy of the data that is used for this study has 38 columns and 194673 entries.

Exploratory Data Analysis

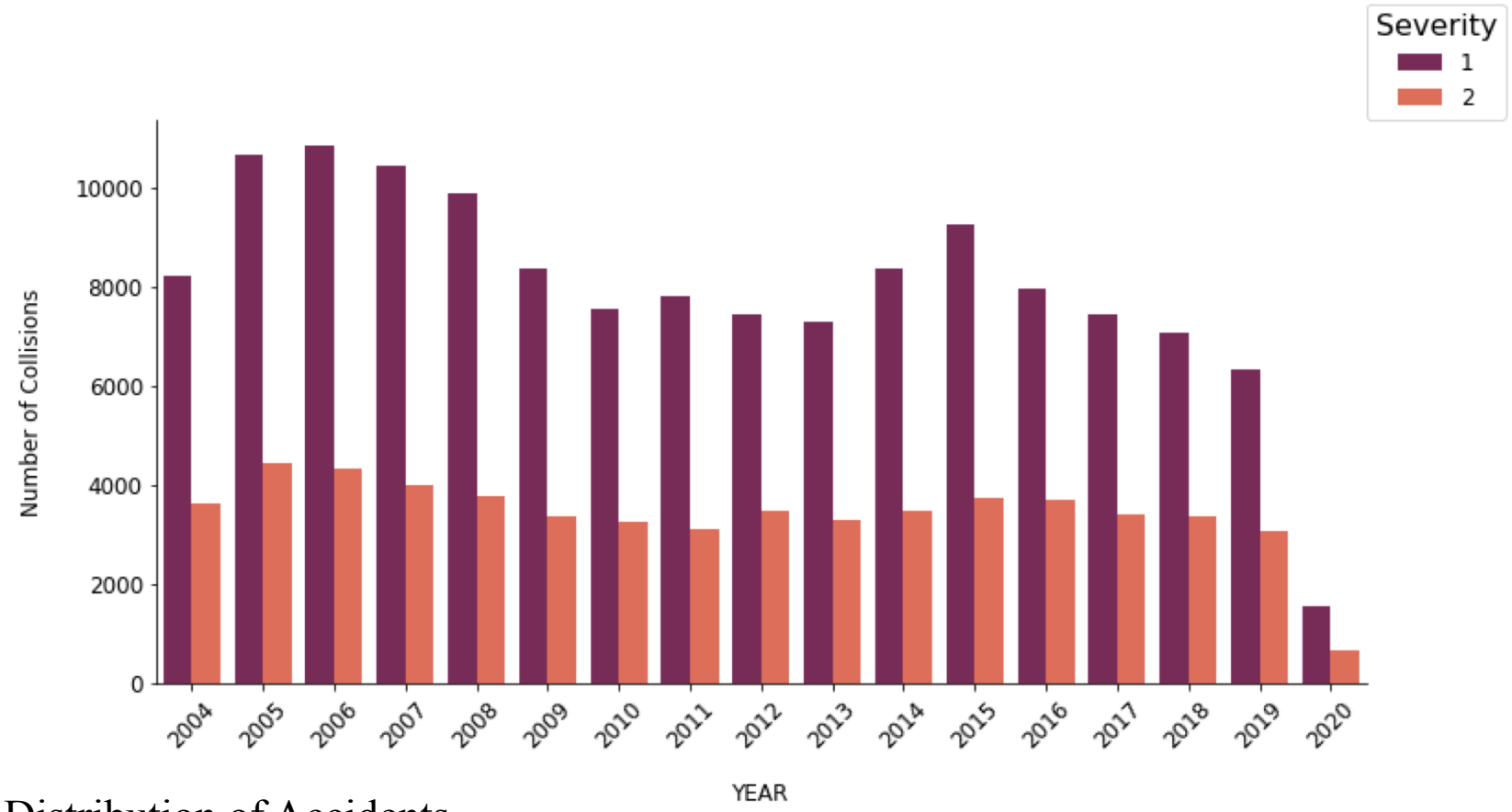


Figure 1: Yearly Distribution of Accidents

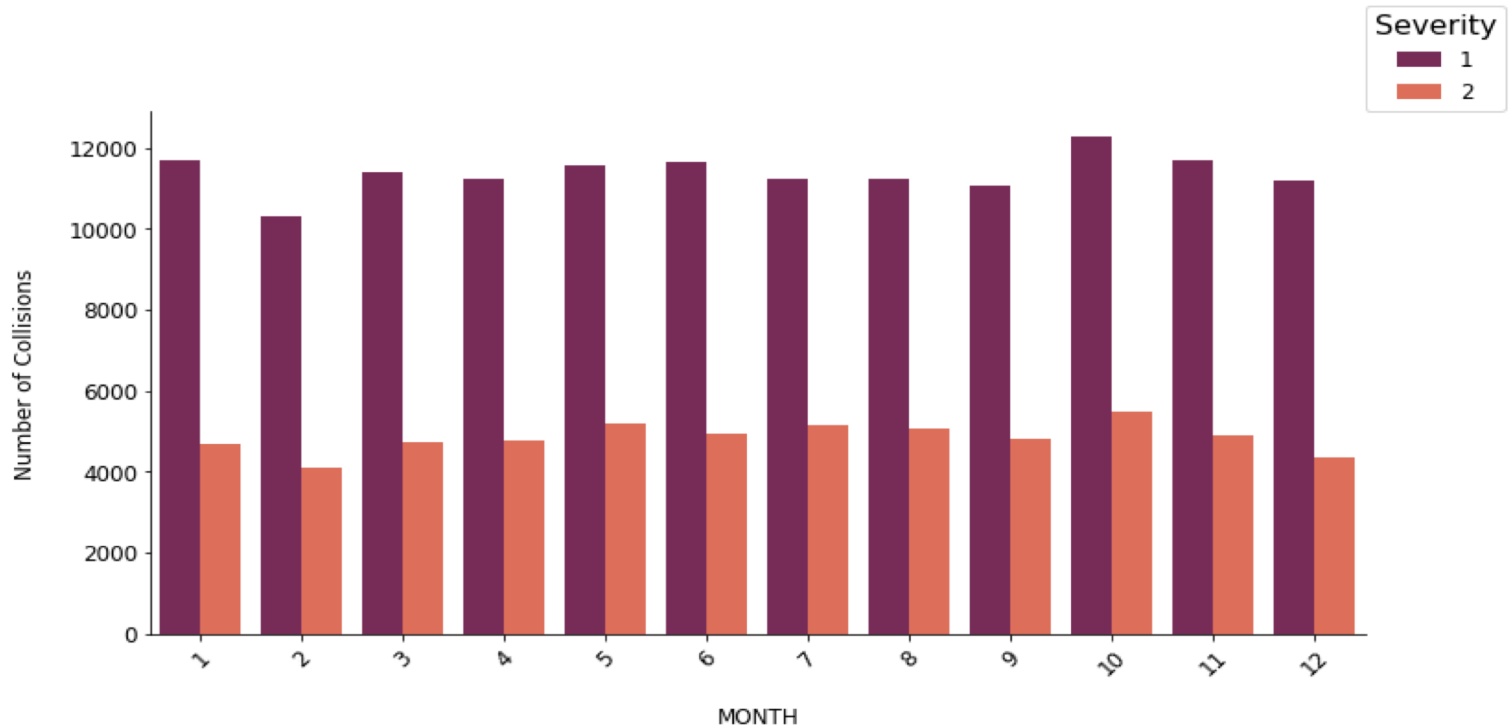


Figure 2: Monthly Distribution of Accidents

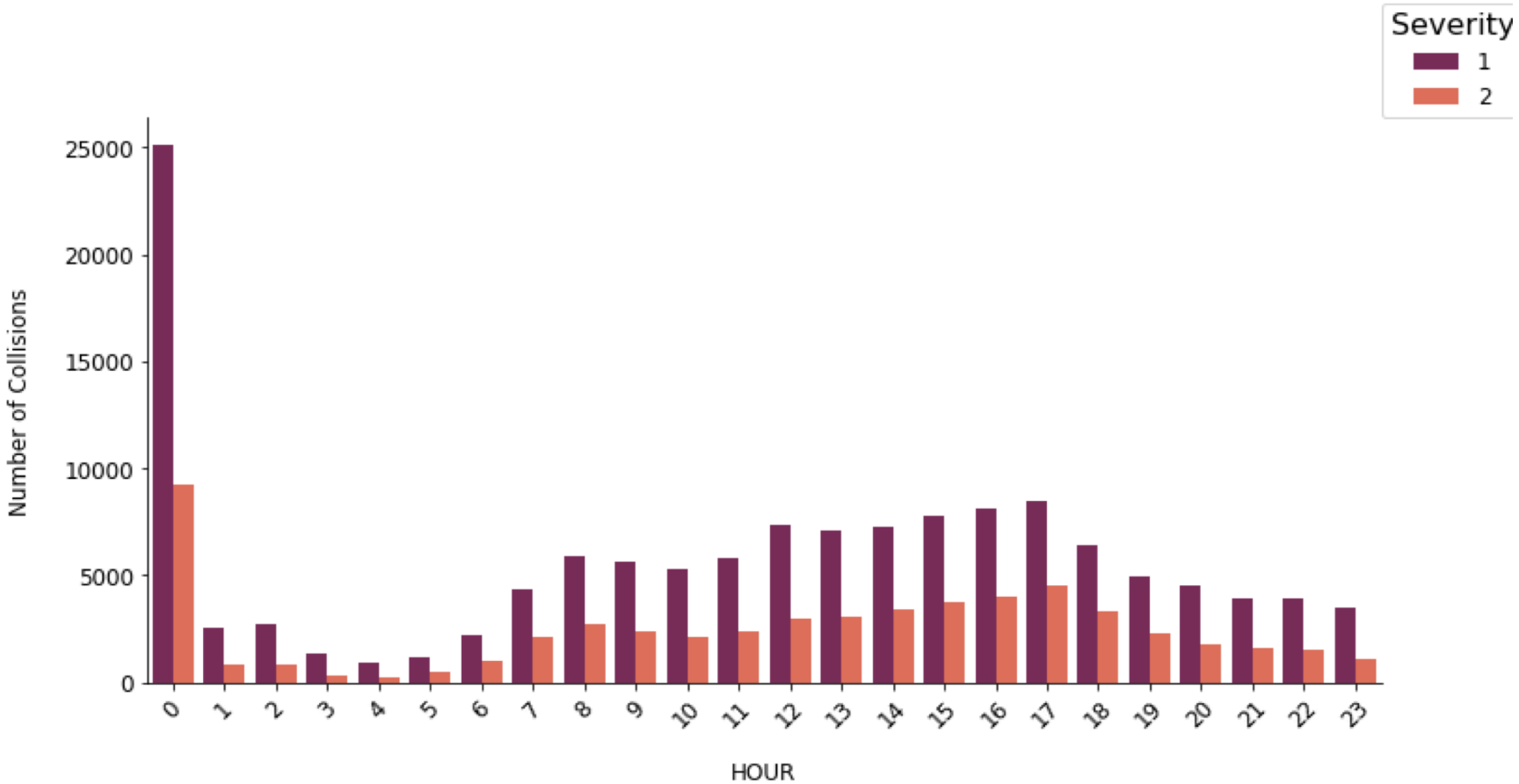


Figure 3: Distribution of Accidents on Hourly Basis

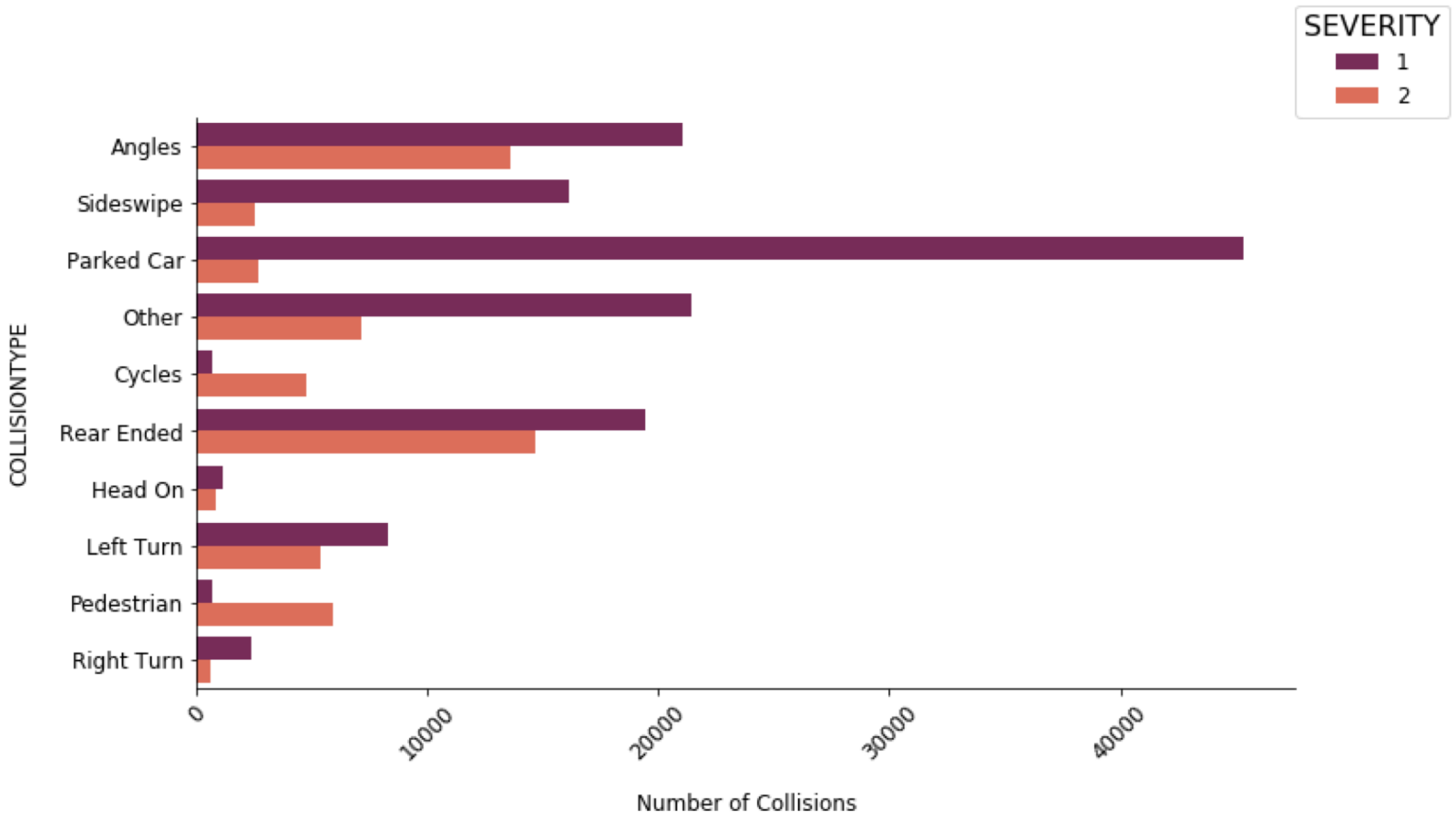


Figure 4: Severity Levels by Collision Types

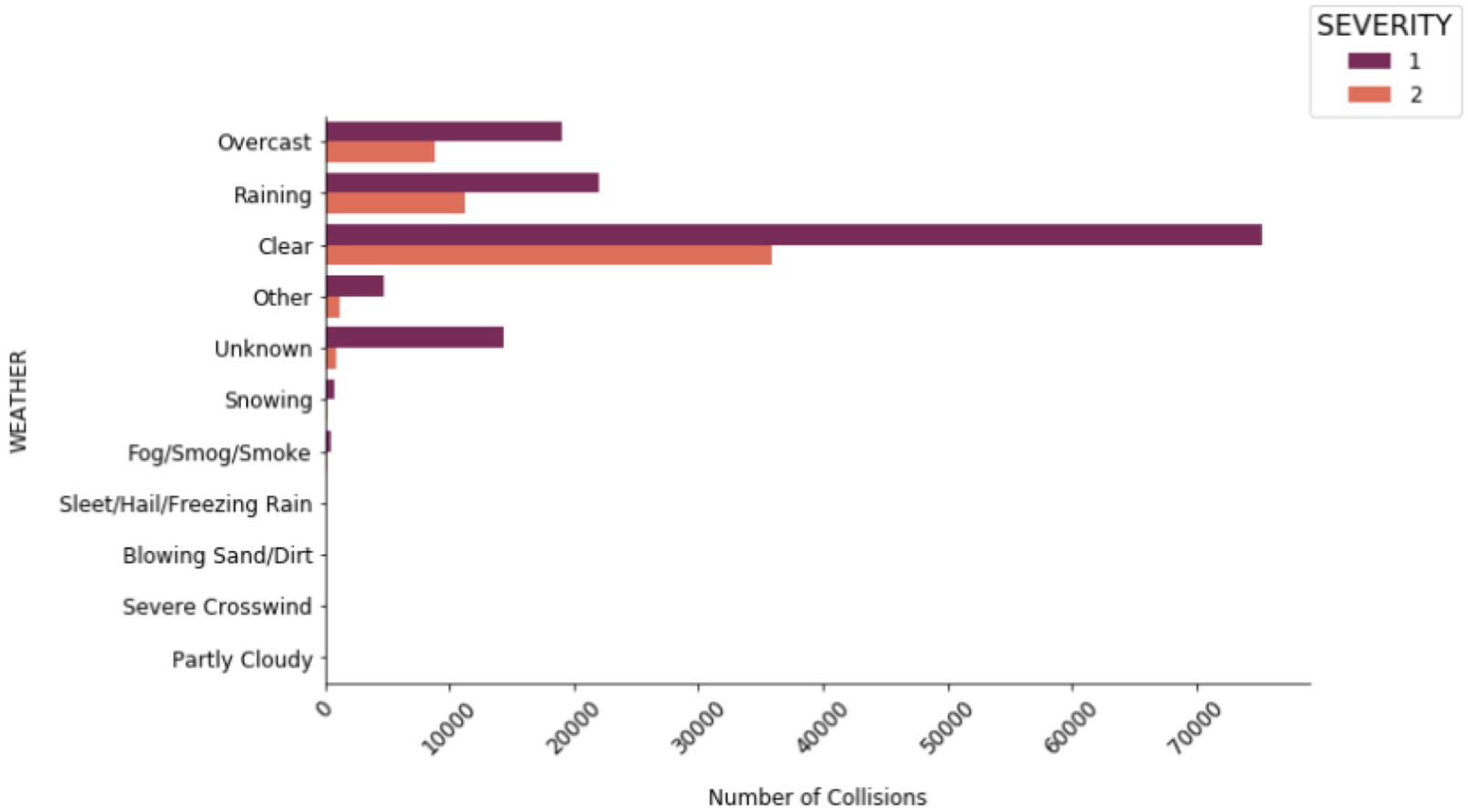


Figure 5: Severity Levels by Weather

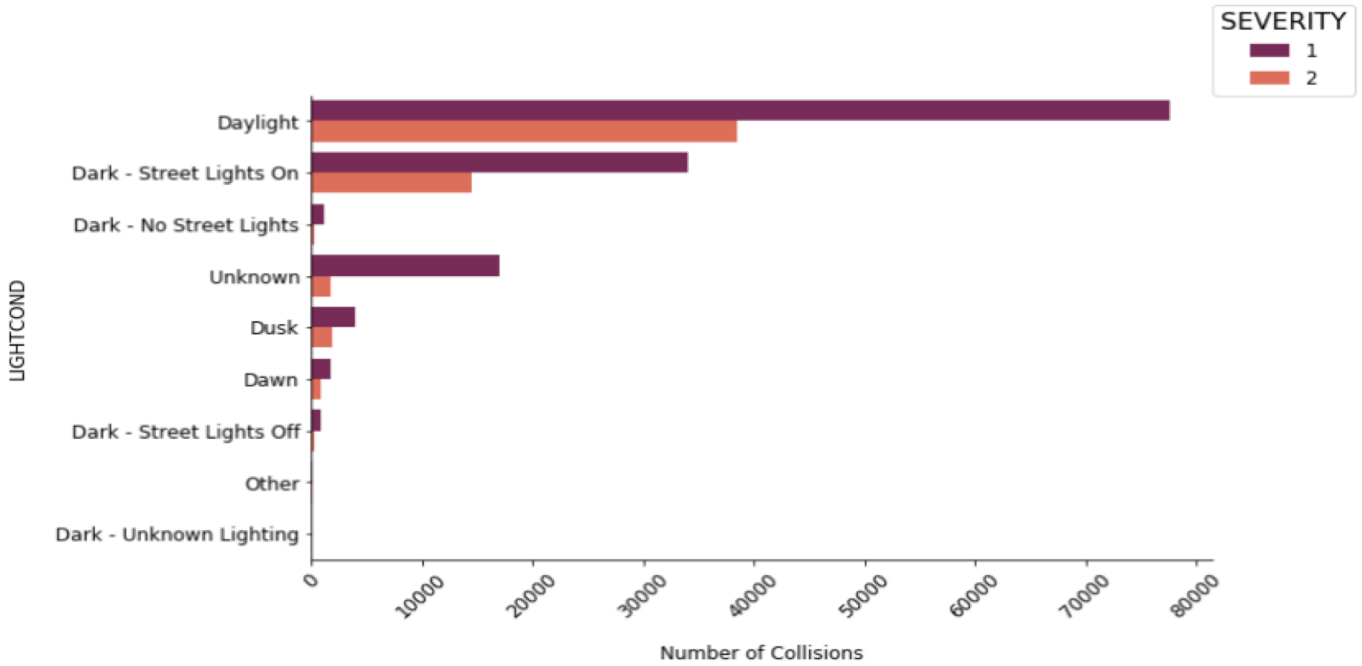


Figure 6: Number of Collisions by Light Conditions

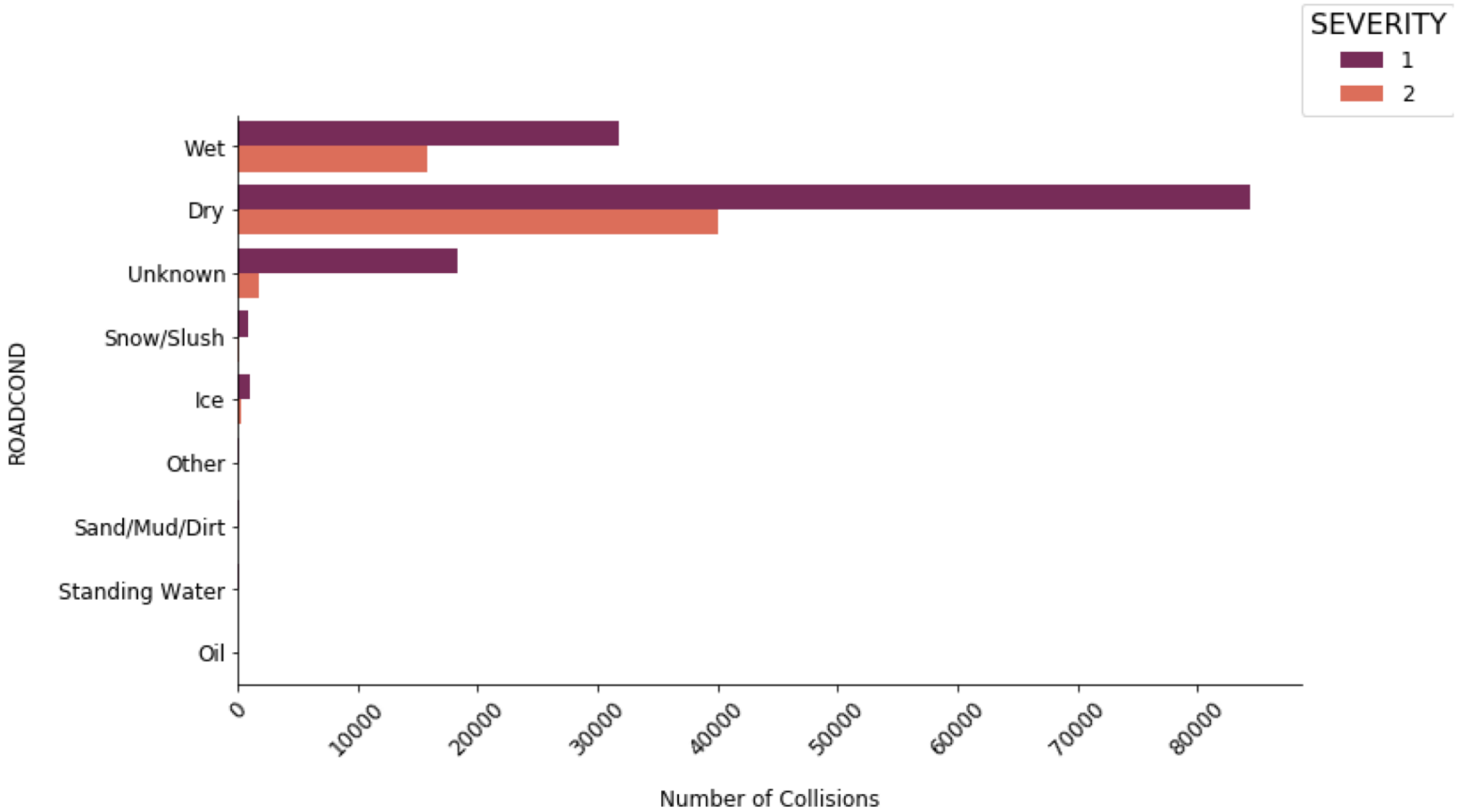
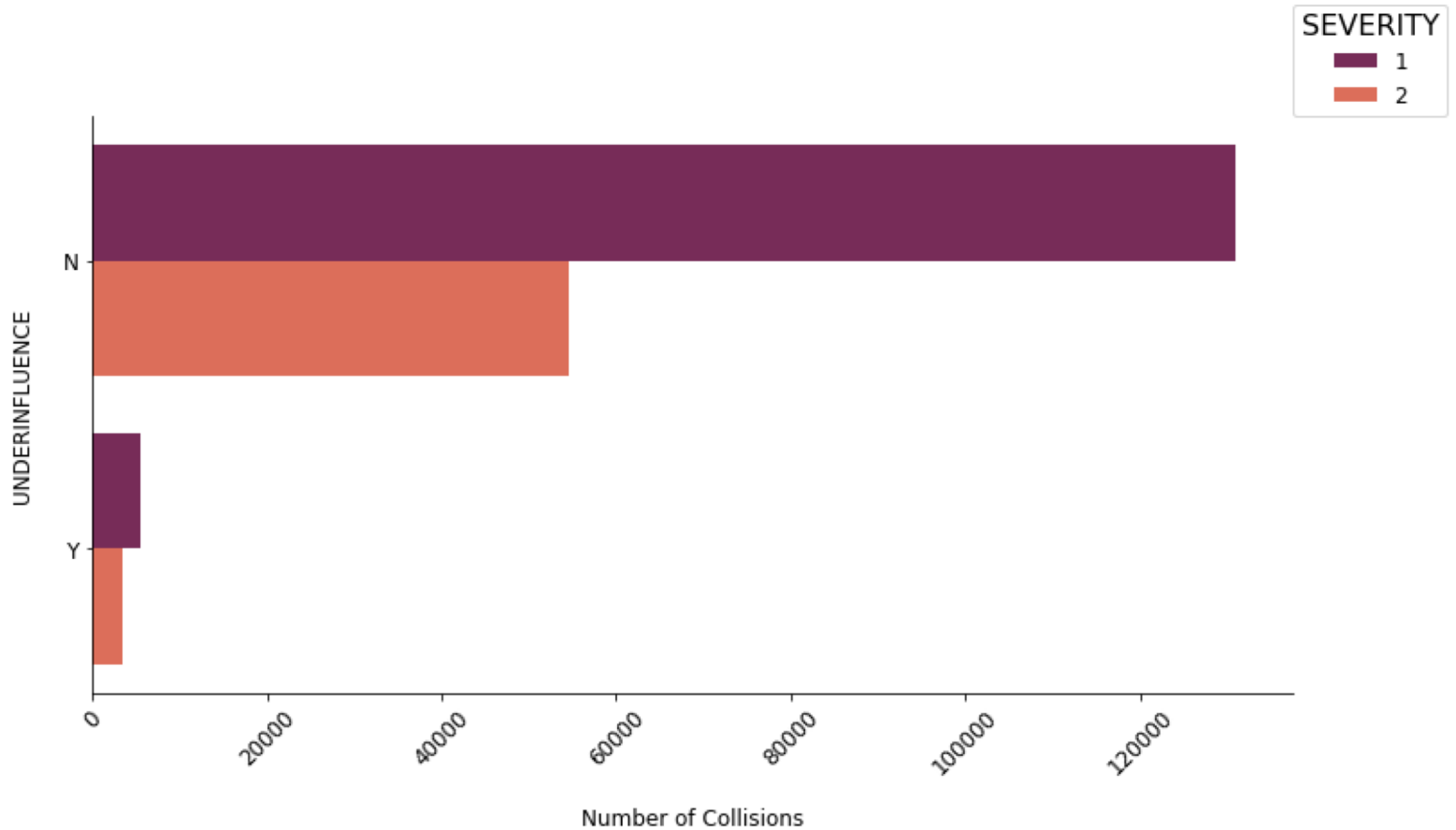


Figure 7: Number of Collisions by Road Conditions


SEVERITY
1
2
UNDERINFLUENCE
N
Y
0
20000
40000
60000
80000
100000
120000
Number of Collisions

Feature Engineering and Data Preparation

- 7 features, that seemed most relevant for our inquiry have been selected.

	SEVERITYCODE	WEATHER	ROADCOND	LIGHTCOND	COLLISIONTYPE	UNDERINFLUENCE	SPEEDING
0	2	Overcast	Wet	Daylight	Angles	N	N
1	1	Raining	Wet	Dark - Street Lights On	Sideswipe	N	N
2	1	Overcast	Dry	Daylight	Parked Car	N	N
3	1	Clear	Dry	Daylight	Other	N	N
4	2	Raining	Wet	Daylight	Angles	N	N

- Missing Values
- Label Encoding

Model Development

- No Free Lunch in Data Science!

- Logistic Regression
- K-Nearest Neighbour (KNN)
- Decision Tree
- Random Forests
- Gradient Boosting (GBM)
- Support Vector Machine (SVM)

Results

Model	F1 Score	Accuracy
Decision Tree	0.685978	0.749634
KNN	0.697810	0.740979
LOGISTIC REGRESSION	0.696848	0.749994
GRADIENT BOOSTING	0.691827	0.750995
RANDOM FOREST	0.692660	0.750970
SVM	0.691684	0.751098

Conclusion

- Various models can predict the severity of an accident with an approximate accuracy of 75%.
- We can better understand and identify the contributing factors of road traffic accidents and their severity.
- Predicting the severity of an accident correctly could also be helpful for first responders to react and work more efficiently.