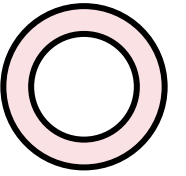




P R E D I C T I N G A C C I D E N T S S E V E R I T Y I N N E W Y O R K S T A T E

- M E E N A K S H I
A I S H W A R Y A



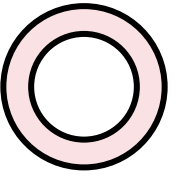
Introduction

The number vehicles on the road is increasing by the day.

This implies higher chances of road accidents.

Road accidents can be fatal and very dangerous; thus we must do our best to avoid them.

Predicting the road accident severity helps in avoiding them.



Data Acquisition and Cleaning

US Accidents - A Countrywide Traffic Accident Dataset (2016 - 2020), Kaggle dataset, data scrapped from <https://www.kaggle.com/sobhanmoosavi/us-accidents>

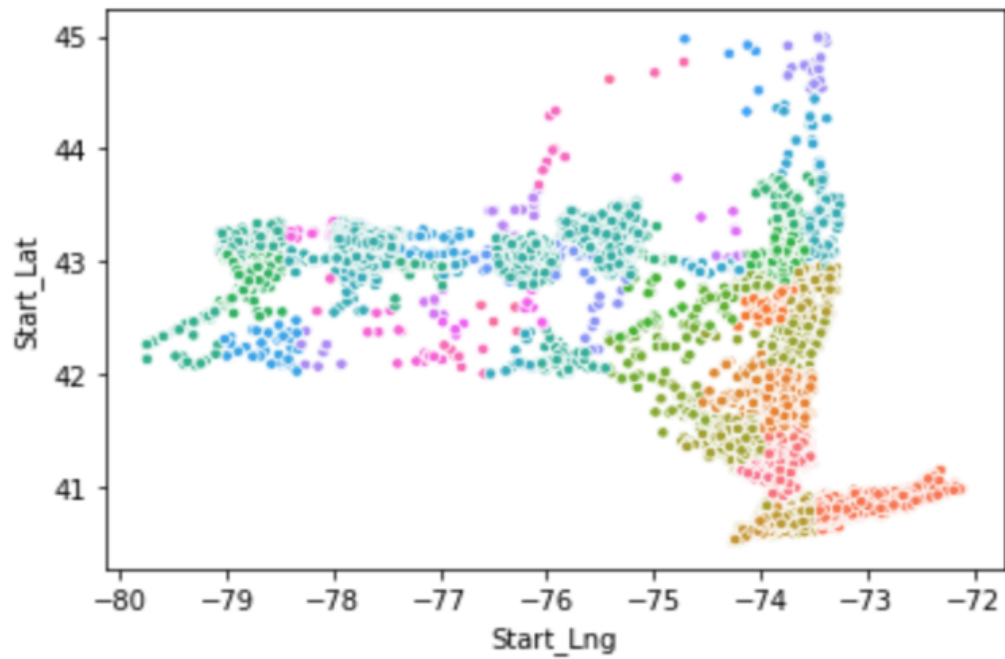
In total 35,13,617 rows and 49 features in raw dataset

Negative time durations, Null value features are dropped

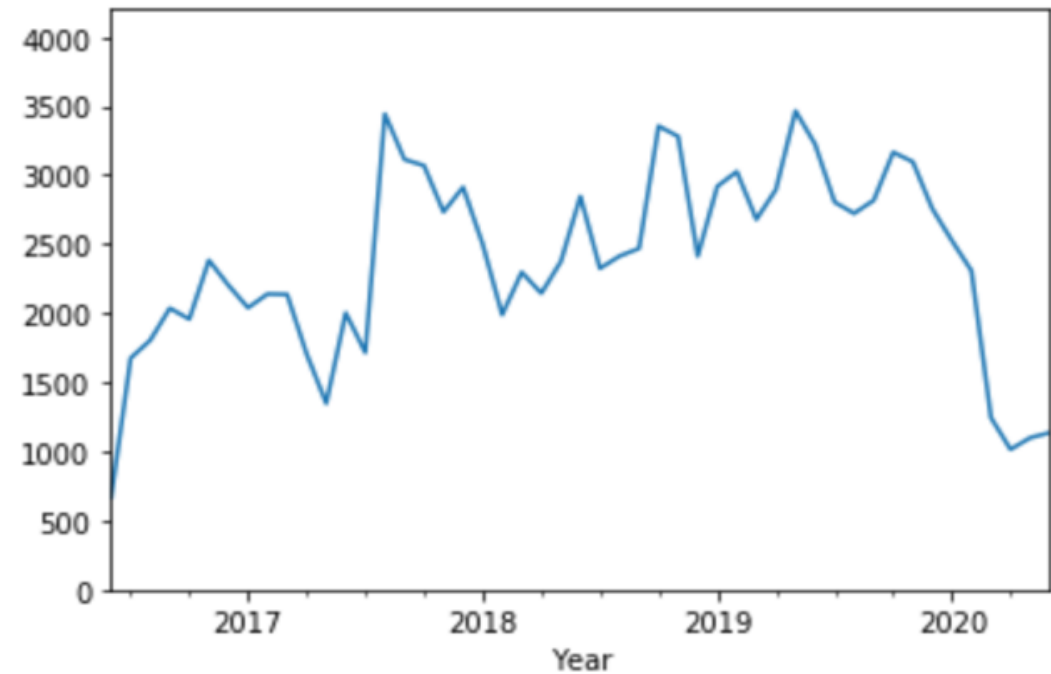
Cleaned data contains 34 features

○ Exploratory Data Analysis

Map of Accidents

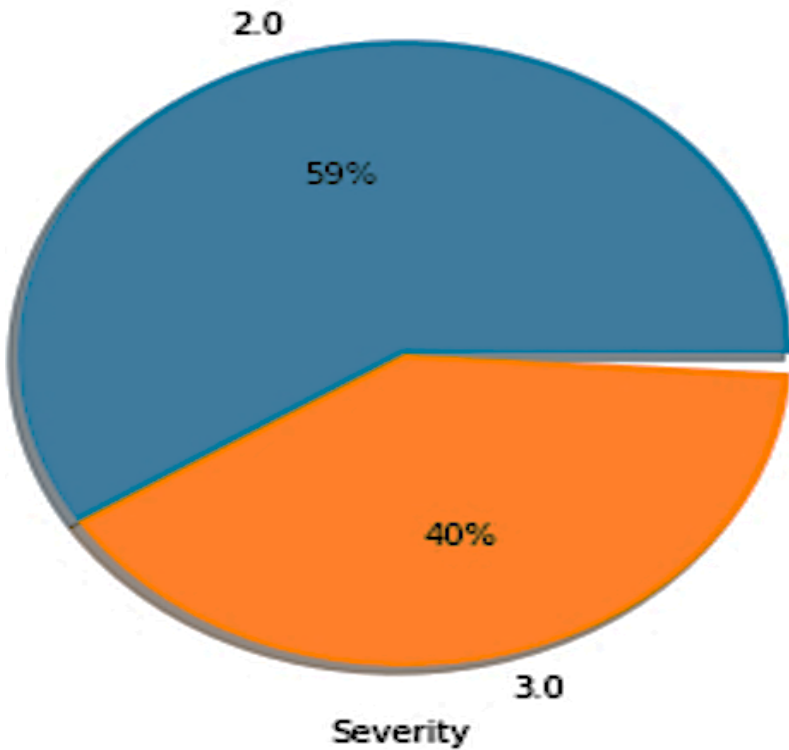


Time Series Analysis

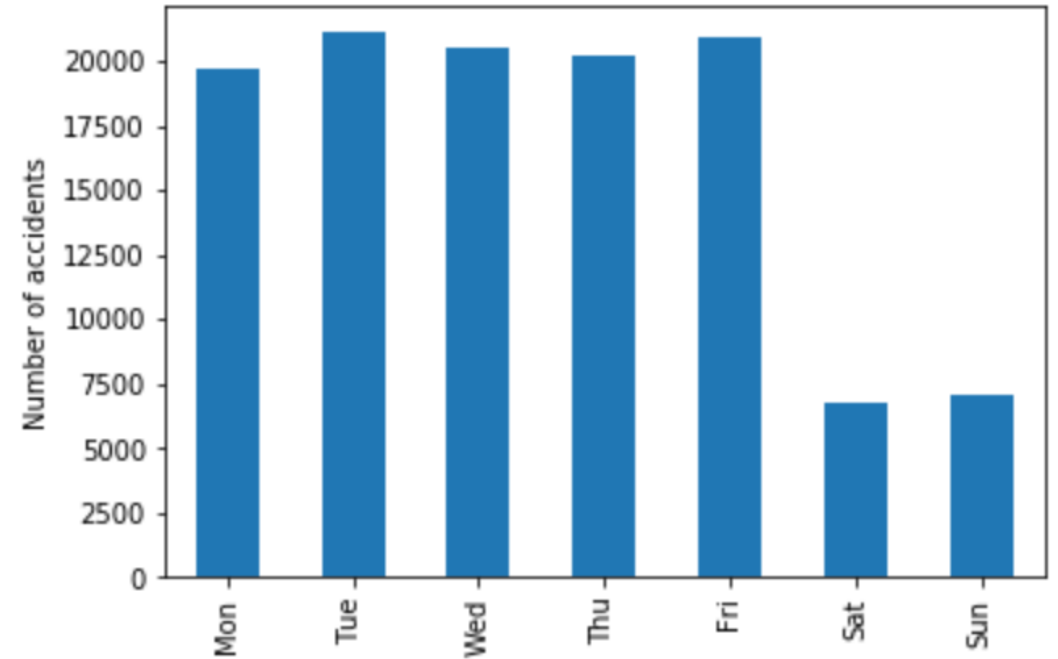
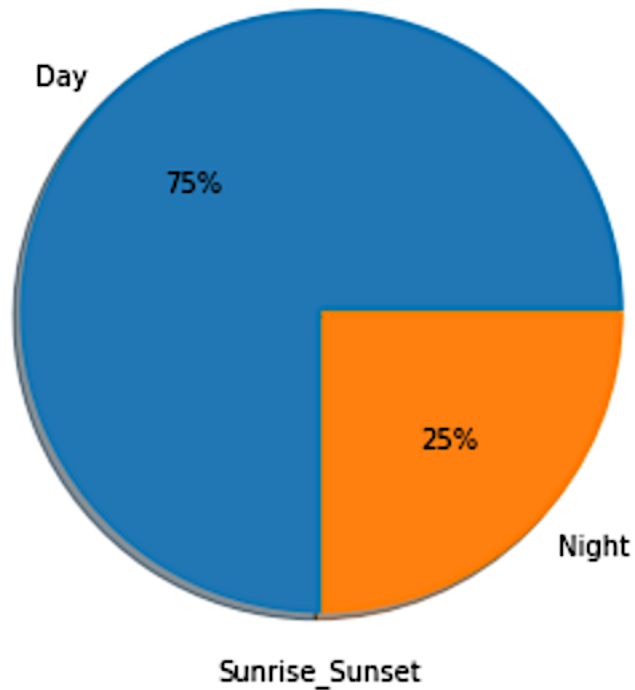


○ Accident Severity Level

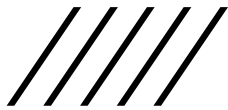
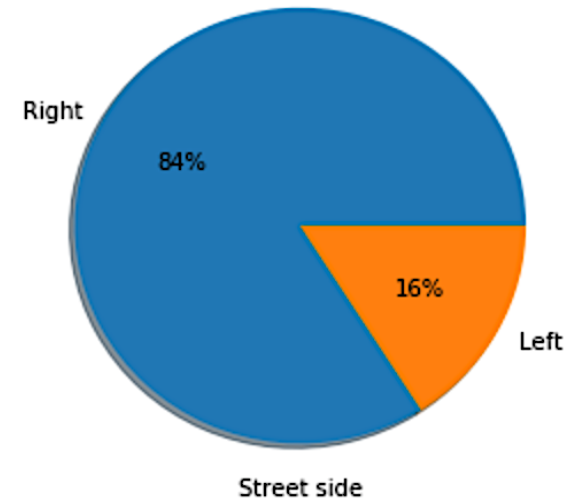
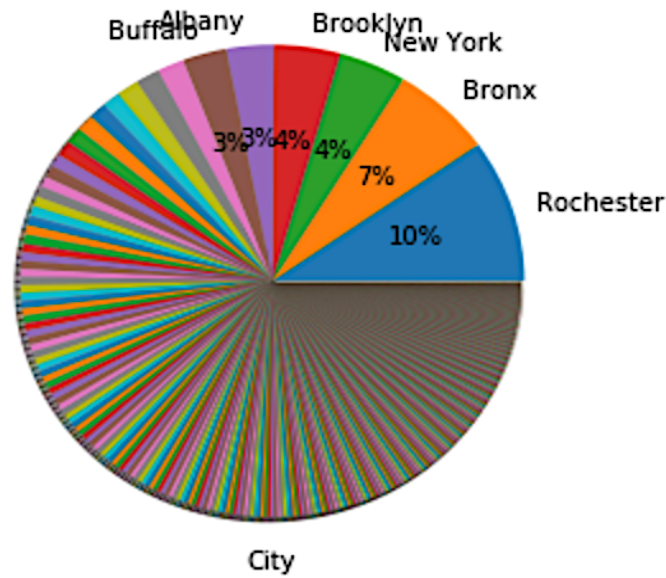
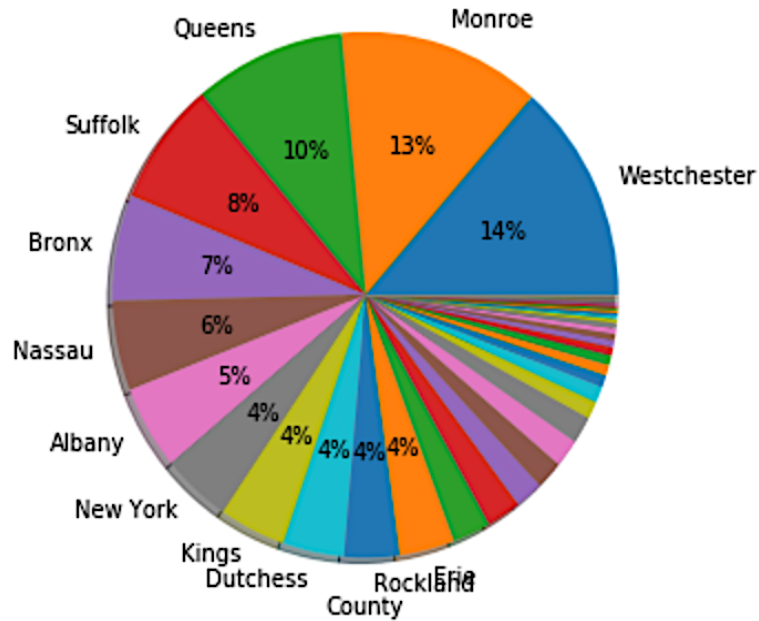
Level of Severity	Number of accidents
2	69142
3	47085
4	162
1	25



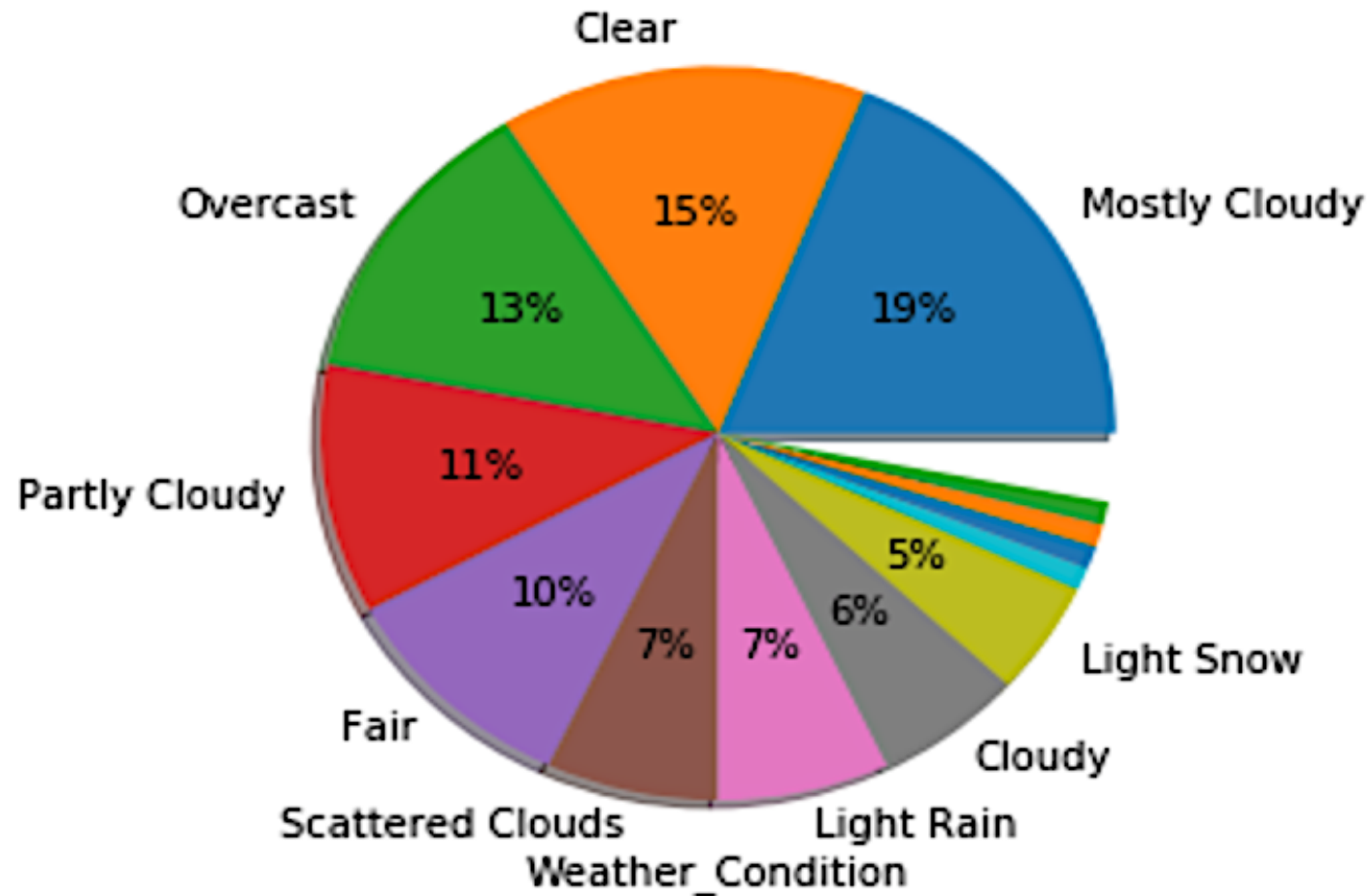
○ When do most accidents takes place?



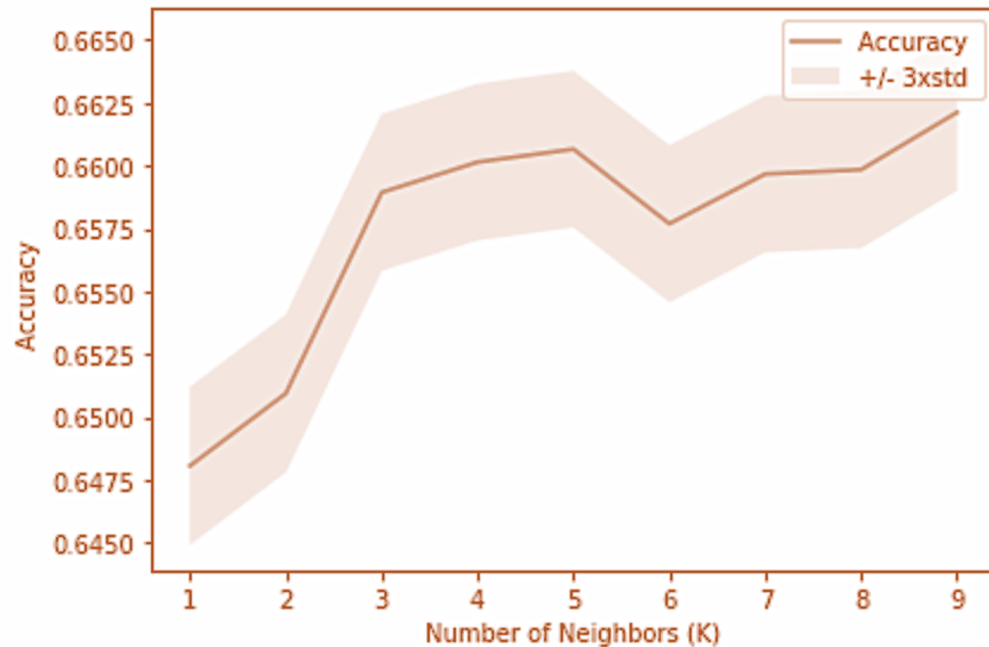
Where do most accidents happen?



- With what weather condition do most accidents happen?



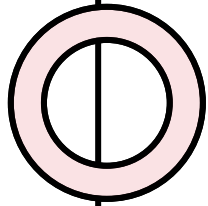
Machine Learning Model



- The highest accuracy is achieved at K=9 with accuracy of 0.6621.

Dataset	Accuracy
Train Set	0.7362854473805714
Test Set	0.6621139887471545





Conclusion

Built model to predict accident severity



Accuracy of model has room for improvement



Can capture more features

State of driver of the vehicle

Road condition

Vehicle performance characteristics