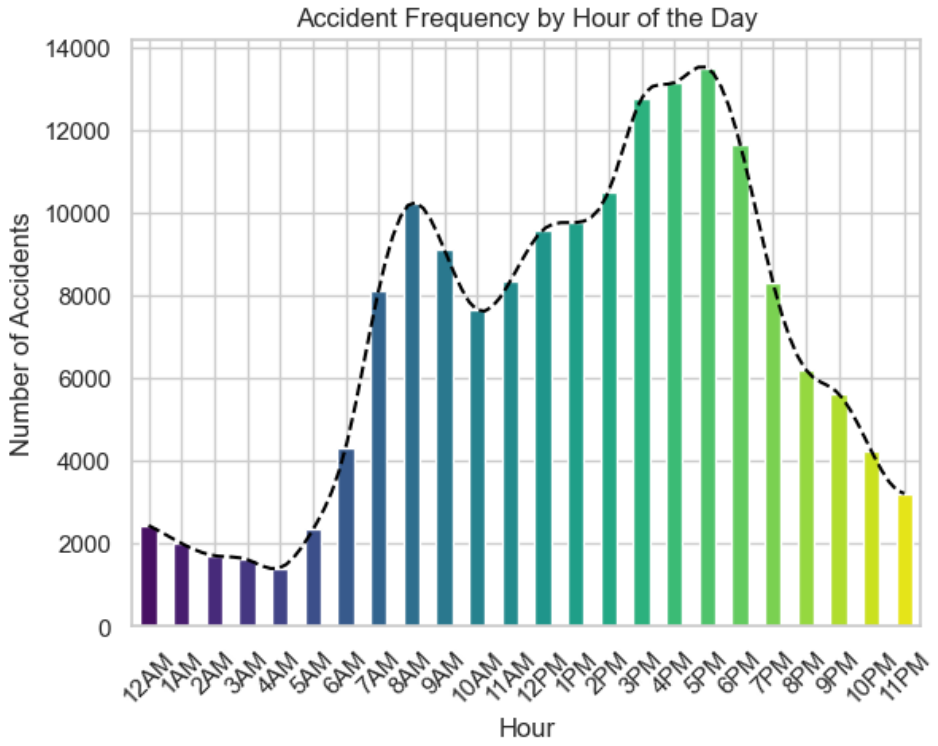
DSCC440 : Data Mining

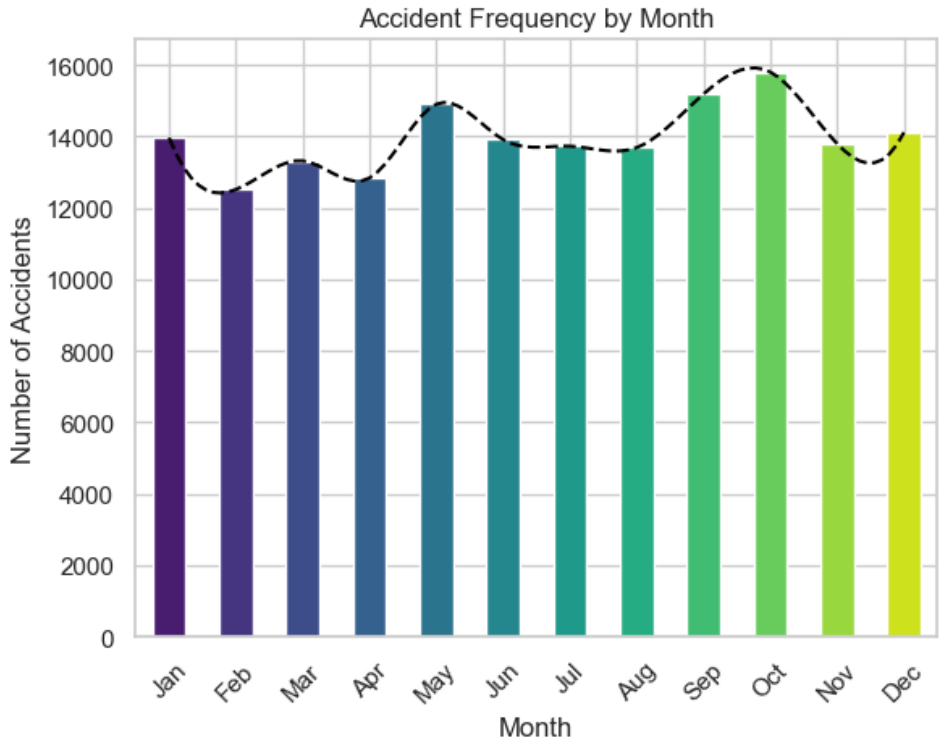
Under the supervision of
Professor J. Luo
University of Rochester

INSIGHTS INTO ROAD ACCIDENT DYNAMICS

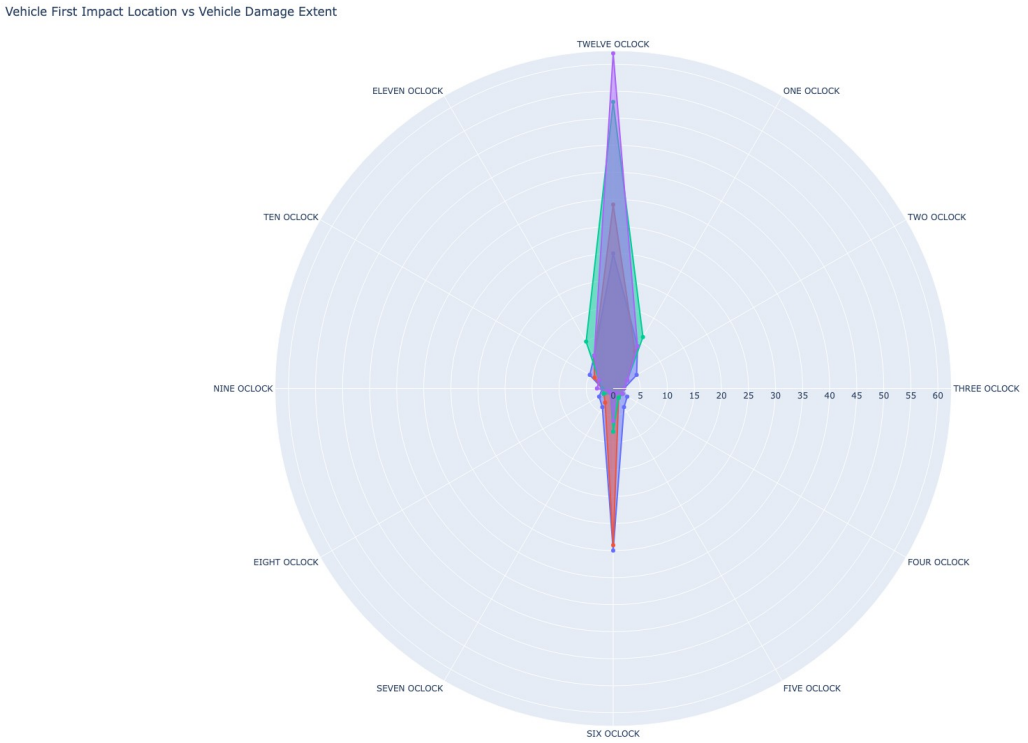
A MACHINE LEARNING APPROACH TO ACCIDENT SEVERITY PREDICTION AND FUTURE PREPAREDNESS

EXPLORATORY DATA ANALYSIS

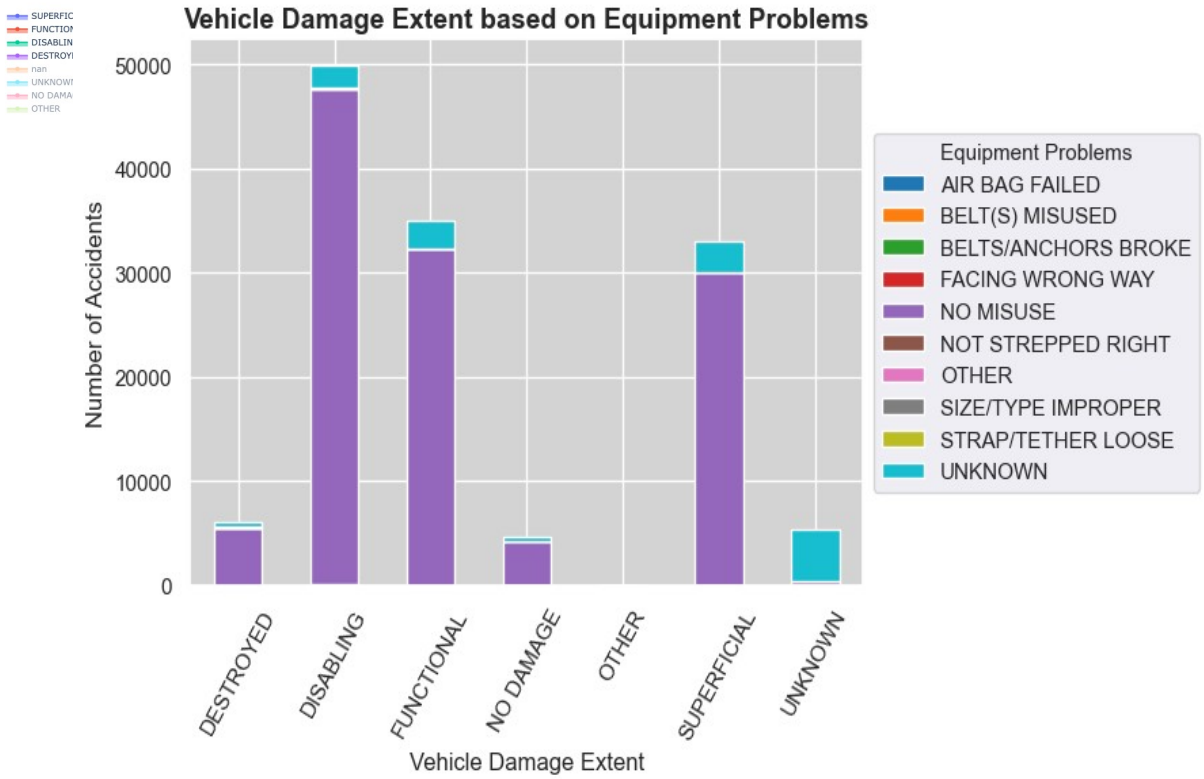




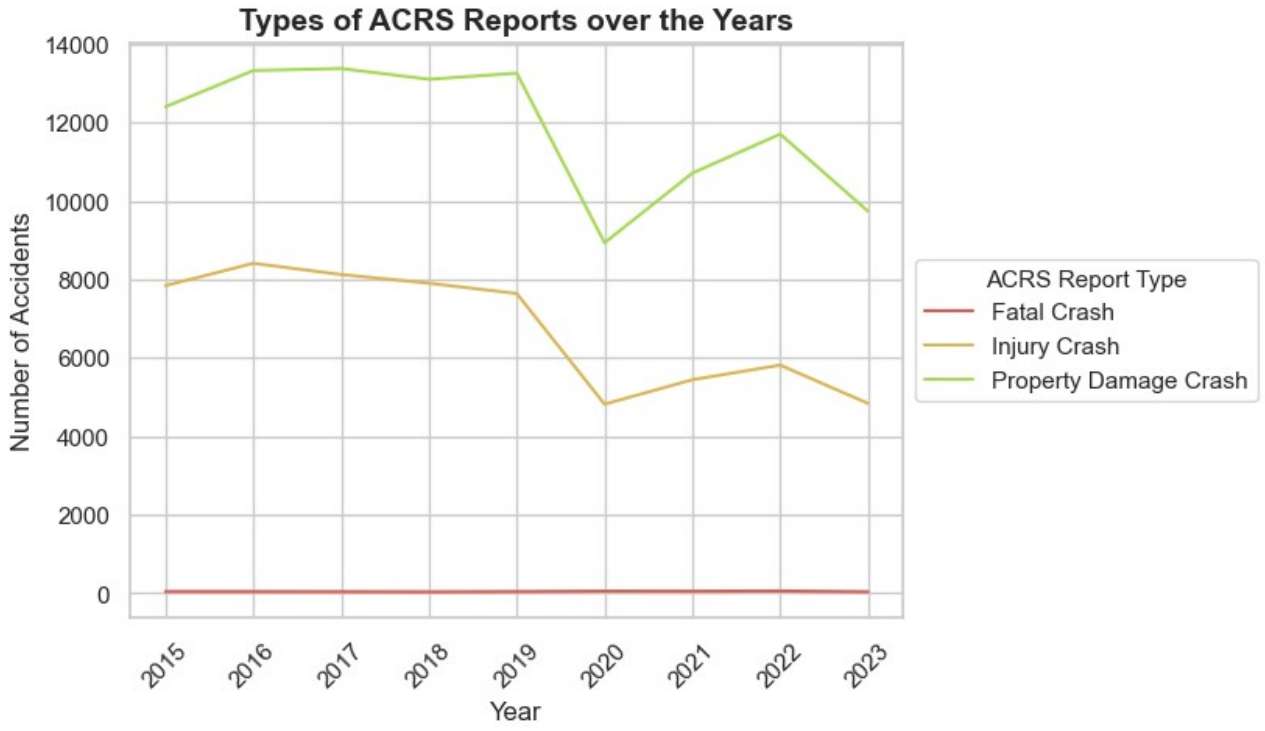
EXPLORATORY DATA ANALYSIS

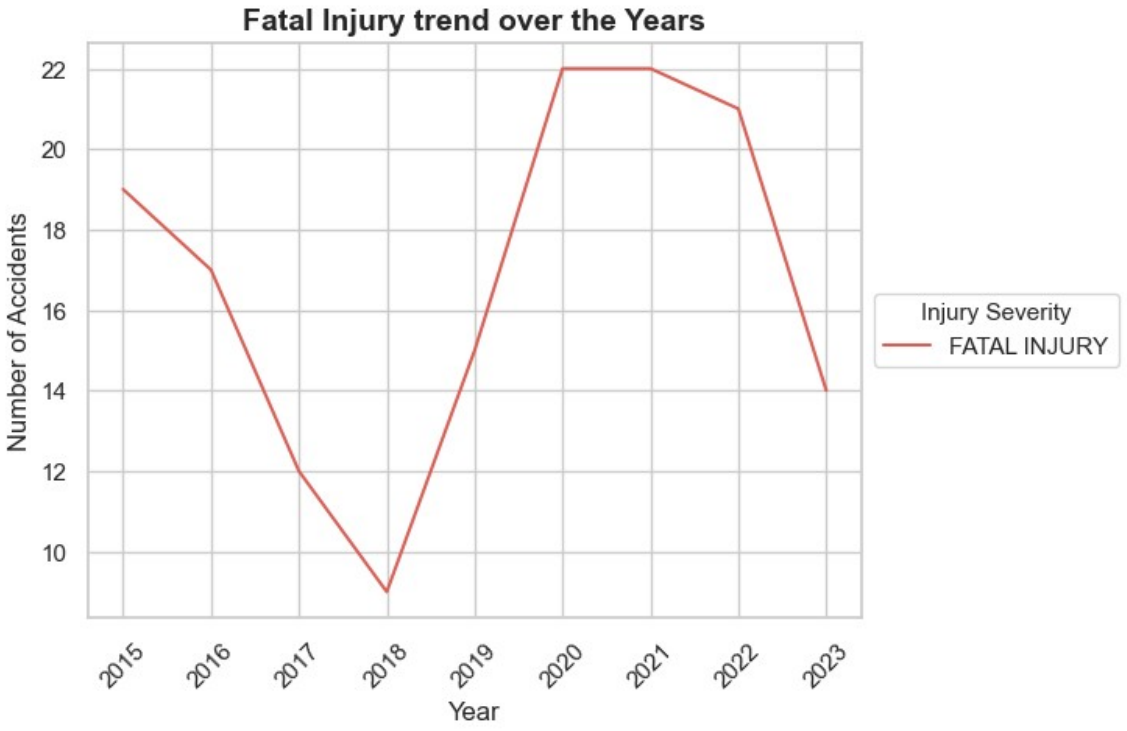


Vehicle First Impact Location vs Vehicle Damage Extent

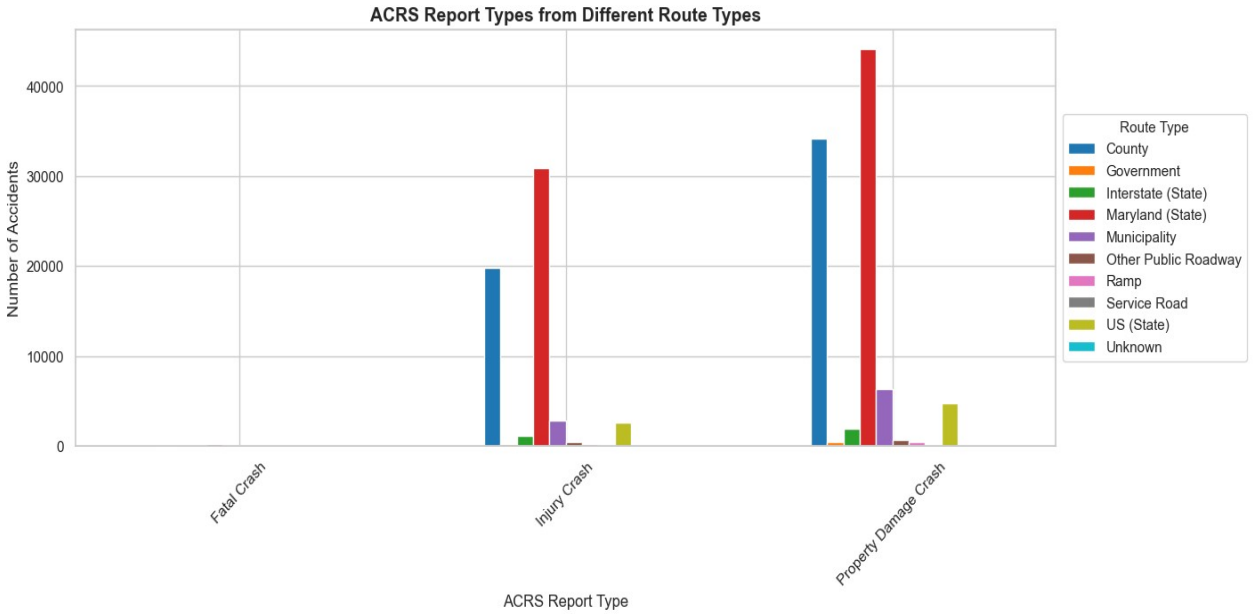


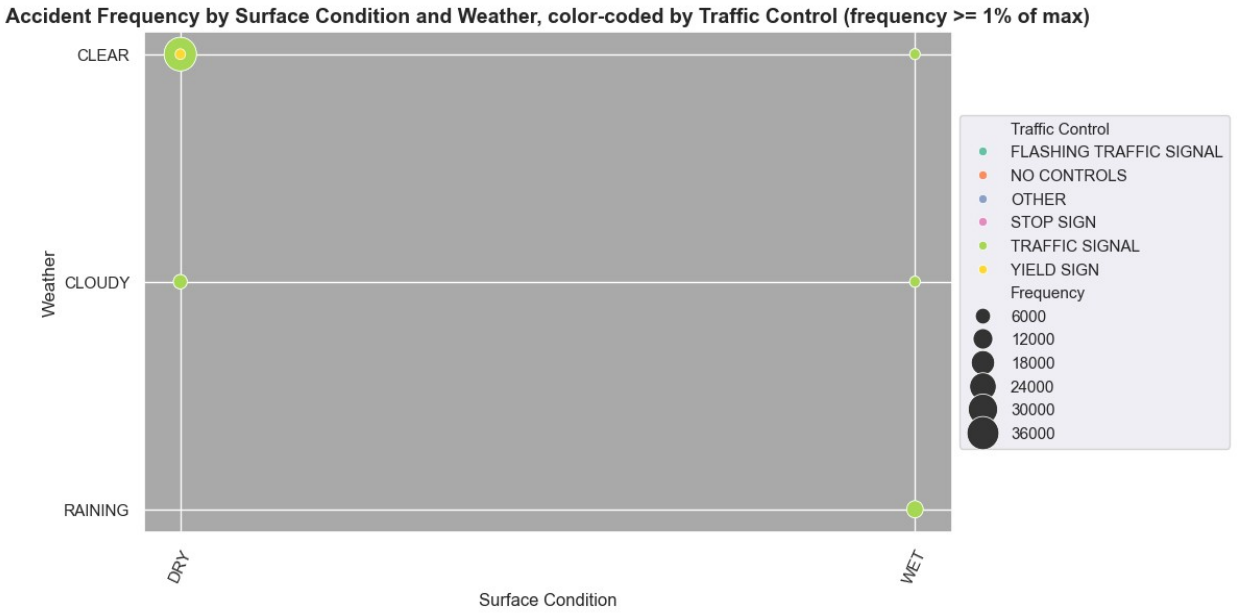
EXPLORATORY DATA ANALYSIS



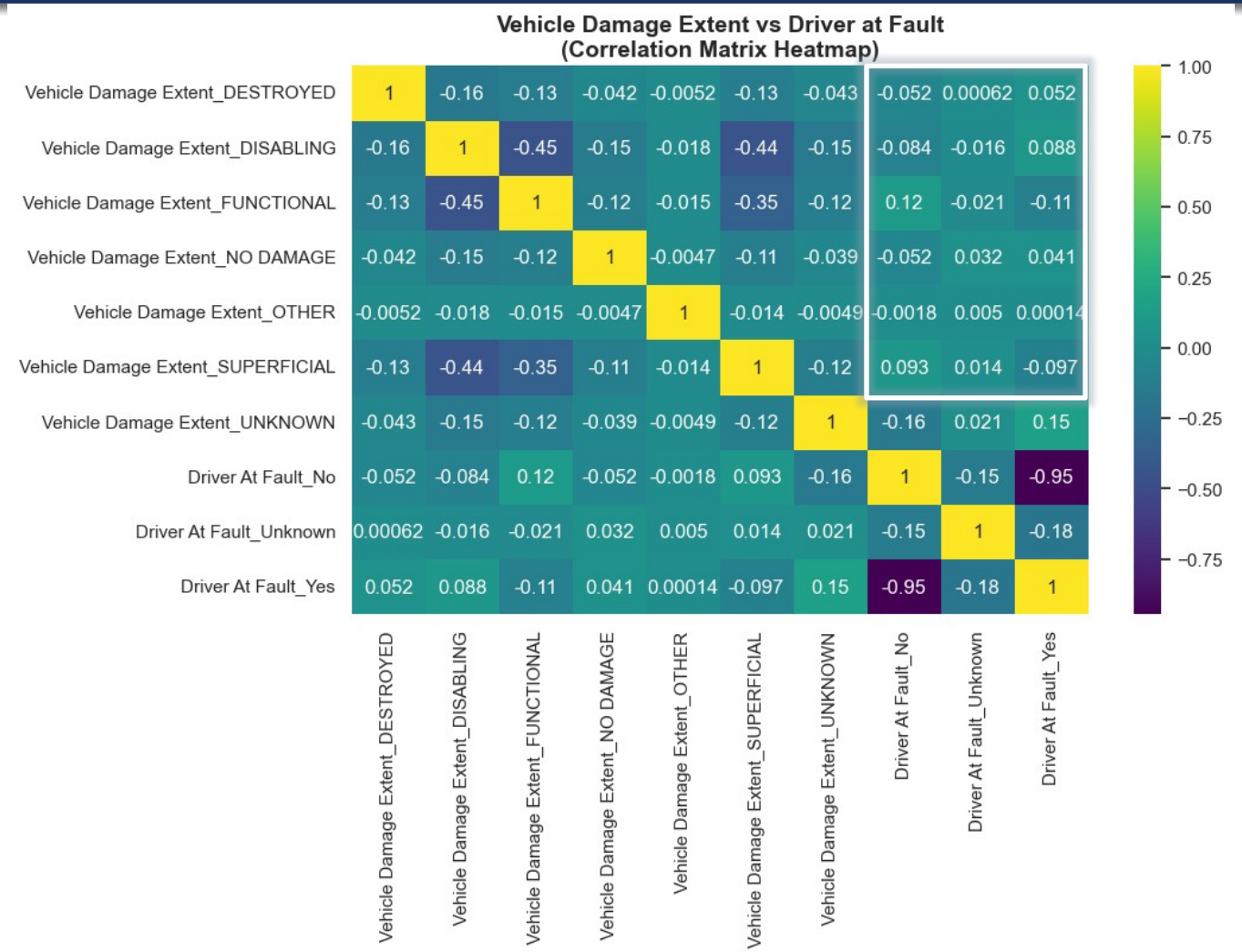


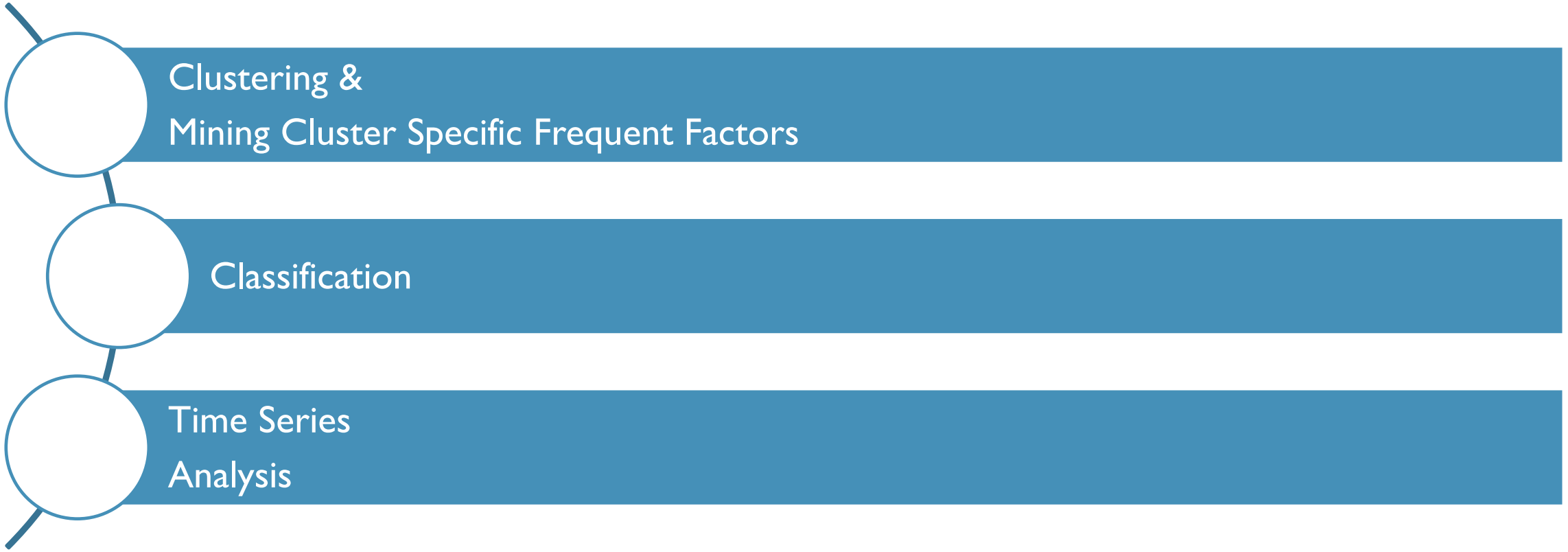
EXPLORATORY DATA ANALYSIS



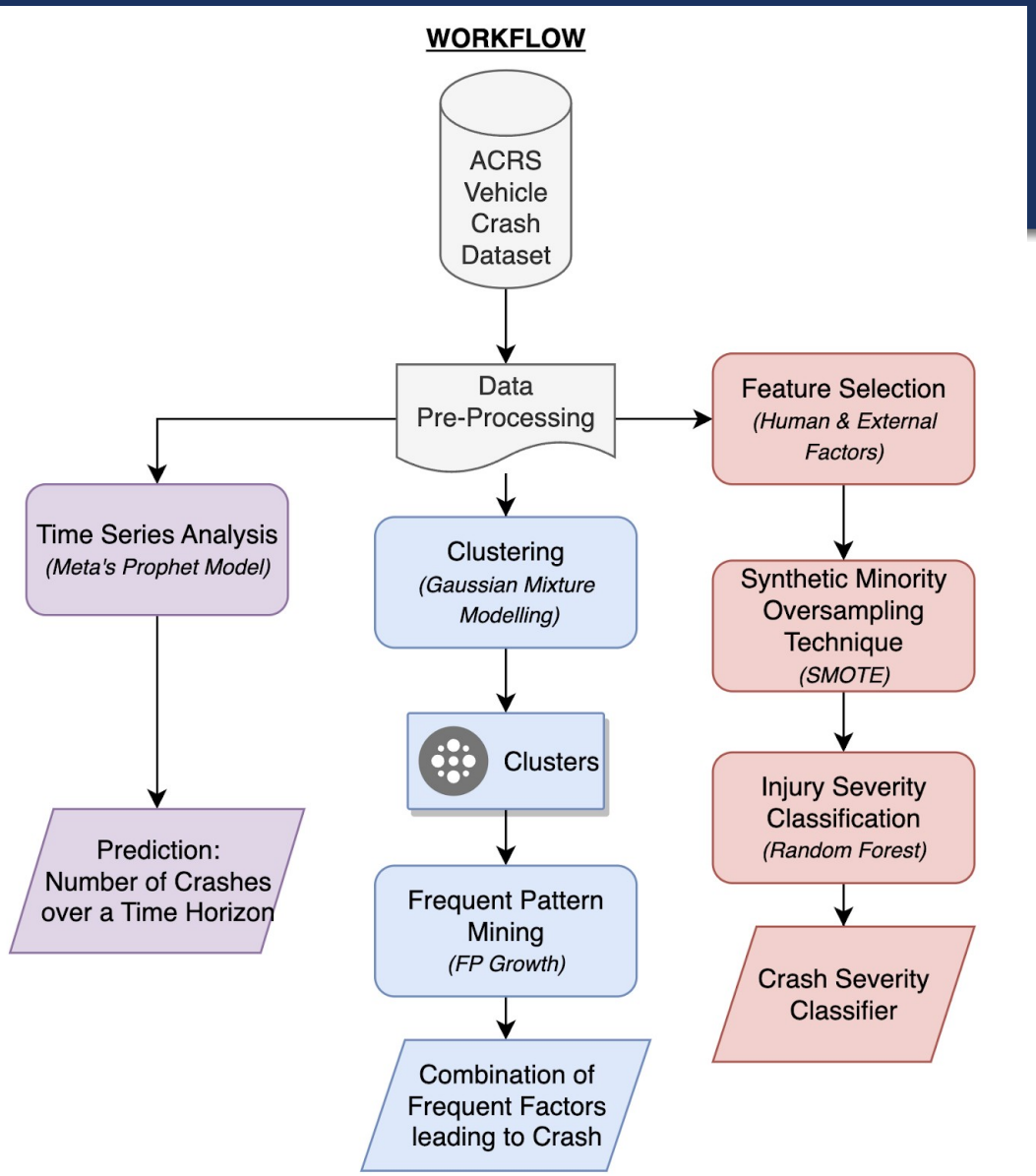


EXPLORATORY DATA ANALYSIS



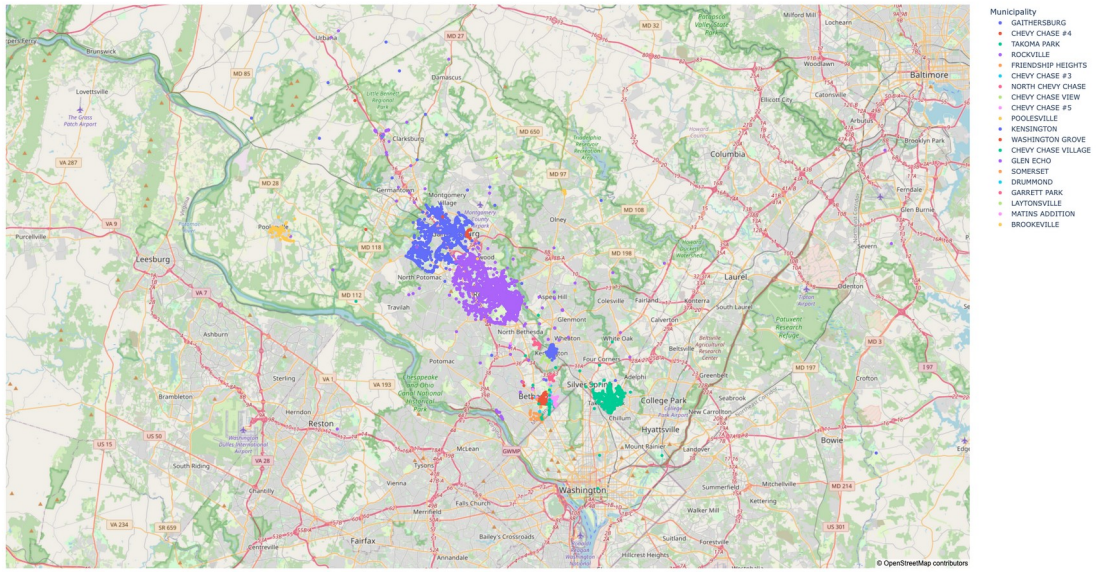


METHODOLOGY

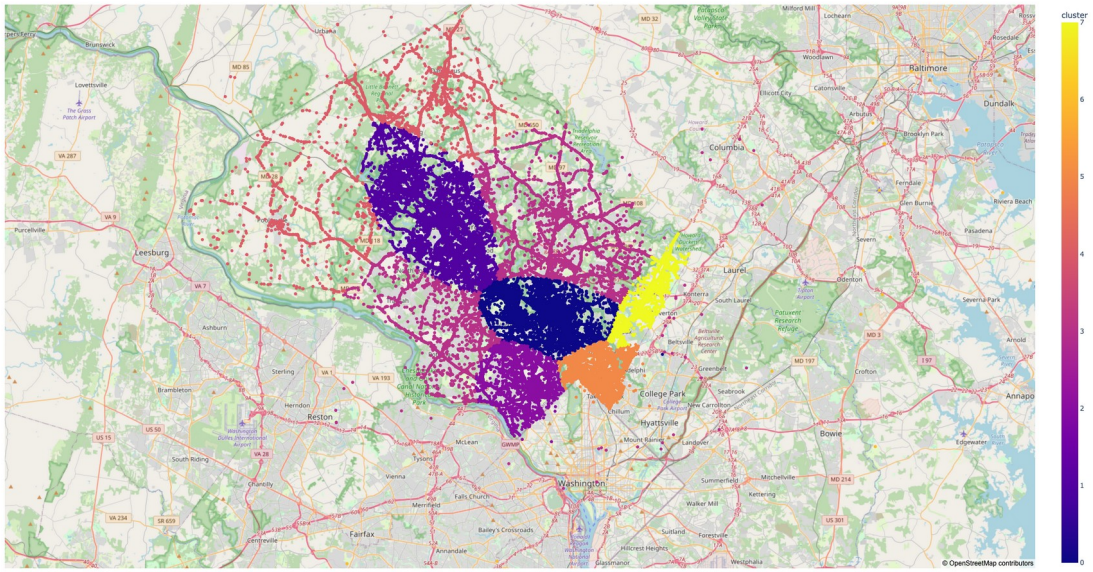

WORKFLOW
ACRS Vehicle Crash Dataset
Data Pre-Processing
Time Series Analysis (Meta's Prophet Model)
Prediction: Number of Crashes over a Time Horizon
Clustering (Gaussian Mixture Modelling)
Clusters
Frequent Pattern Mining (FP Growth)
Combination of Frequent Factors leading to Crash
Feature Selection (Human & External Factors)
Synthetic Minority Oversampling Technique (SMOTE)
Injury Severity Classification (Random Forest)
Crash Severity Classifier

CLUSTERING CRASH SITE LOCATIONS

Clusters of Crashes based on Municipality in Montgomery County, MD



Clusters of Crashes in Montgomery County, MD



FREQUENT PATTERN MINING

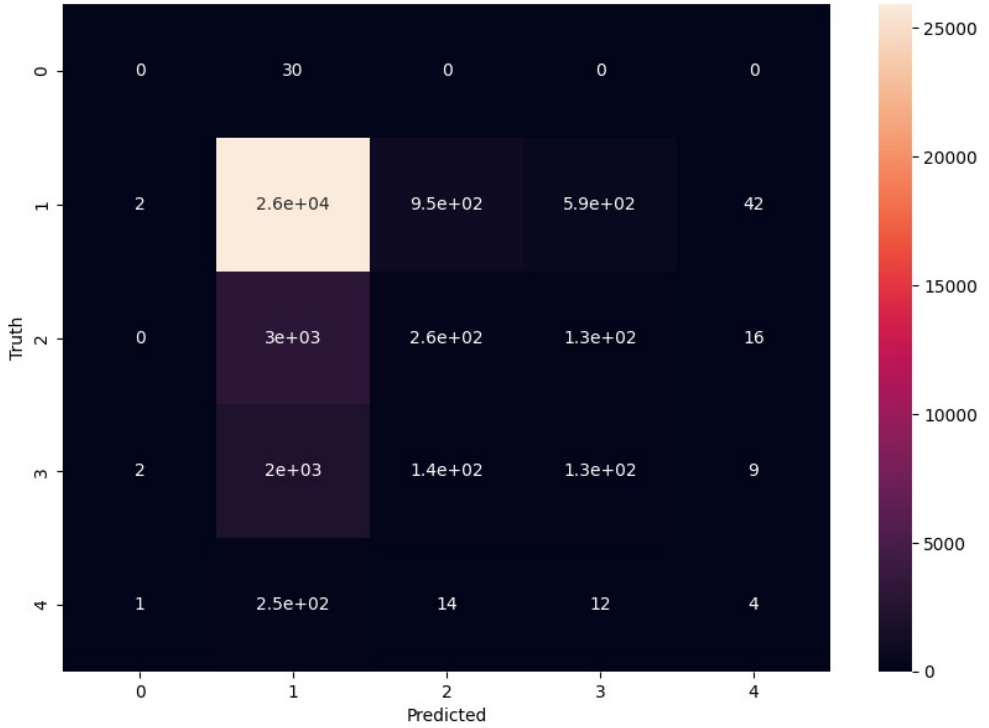
- Constraints for pattern mining using FPGrowth
- `'ASSOCIATION_NUM_ATTRIBUTES': 4,`
- `'SUPPORT_NUM_ATTRIBUTES': 5,`
- `'ASSOCIATION_RULE_THRESHOLD': 0.8,`
- `'MIN_LENGTH_OF_PATTERN': 3,`
- `'SUPPORT_THRESHOLD_PERCENTAGE': 40,`
- `'MAX_NUM_OF_FREQUENT_PATTERNS': 30`

Cluster & Size	Frequent itemsets	Itemsets with minimum 3 frequent patterns
Cluster 0 (18731)	14	4
Cluster 1 (9349)	17	5
Cluster 2 (5643)	21	7
Cluster 3 (38431)	15	5
Cluster 4 (29414)	18	6
Cluster 5 (50997)	10	1
Cluster 6 (15234)	22	8

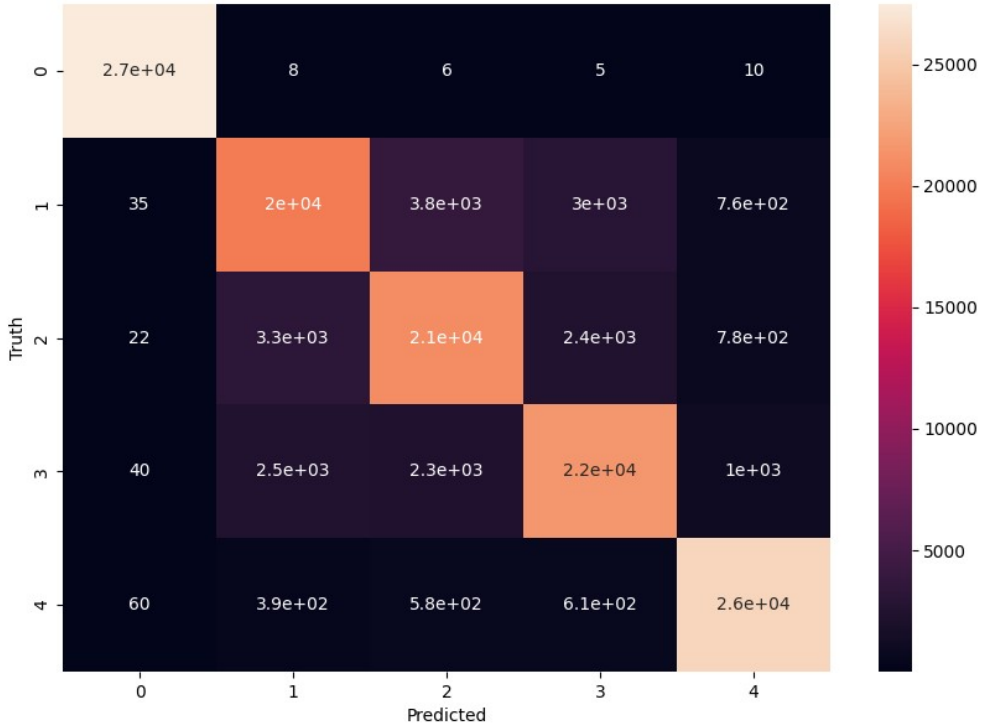
Cluster-wise Frequent Patterns

CLASSIFICATION

TACKLING CLASS IMBALANCE WITH **SMOTE**

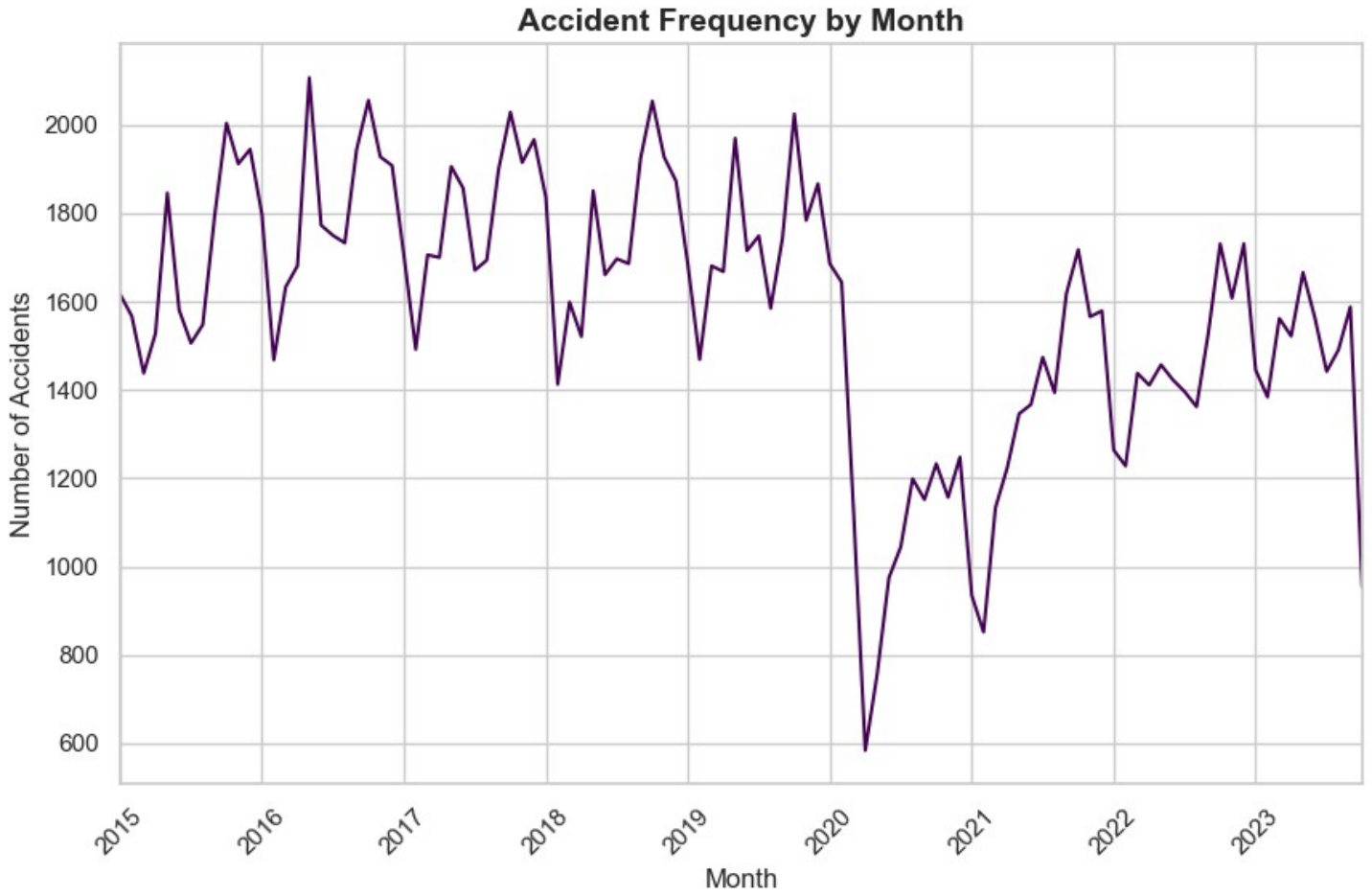


Confusion matrix for original data



Confusion matrix after over-sampling with SMOTE

TIME SERIES ANALYSIS

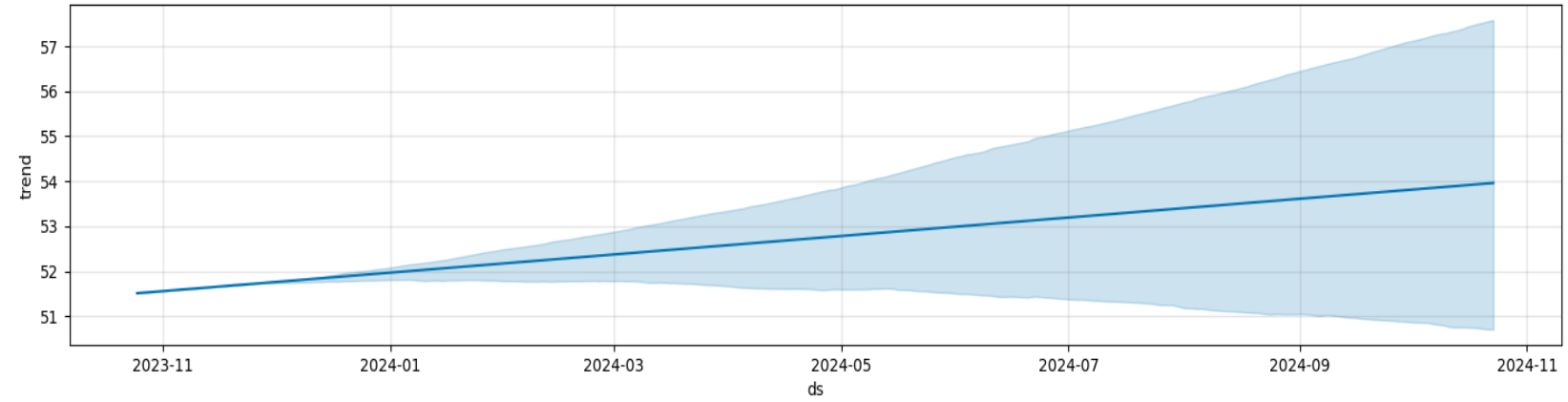


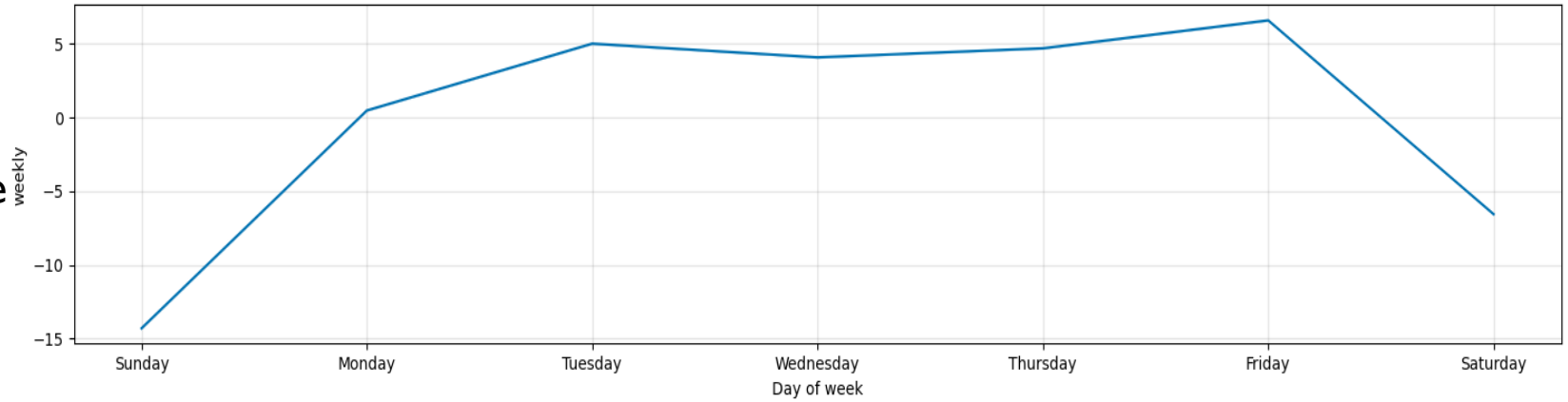
TIME SERIES ANALYSIS

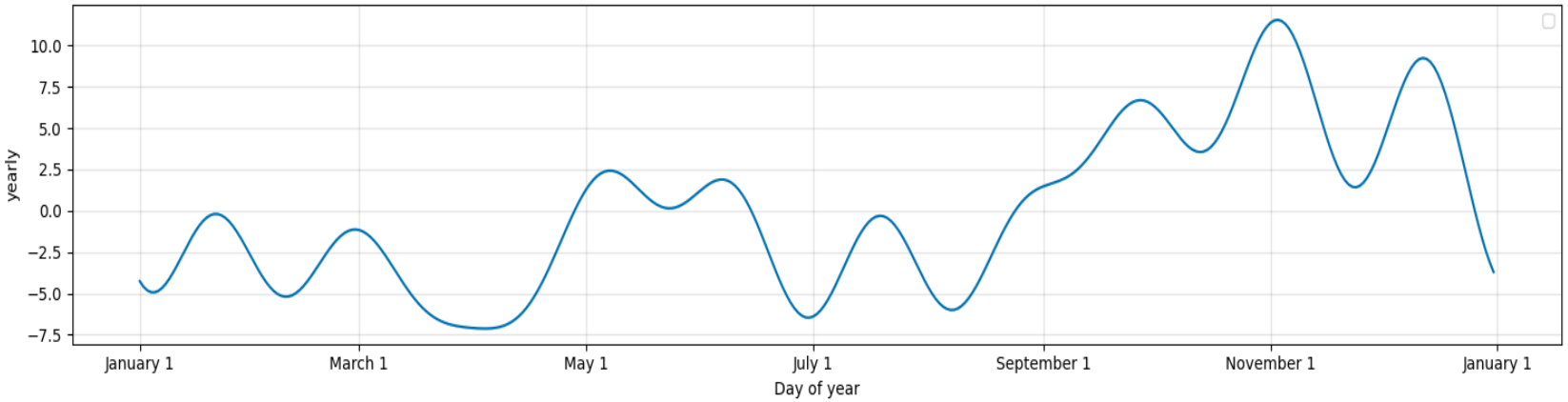
Model Components

The weightage the model would provide to predict the crash frequency based on the date, day and year of timestamp.

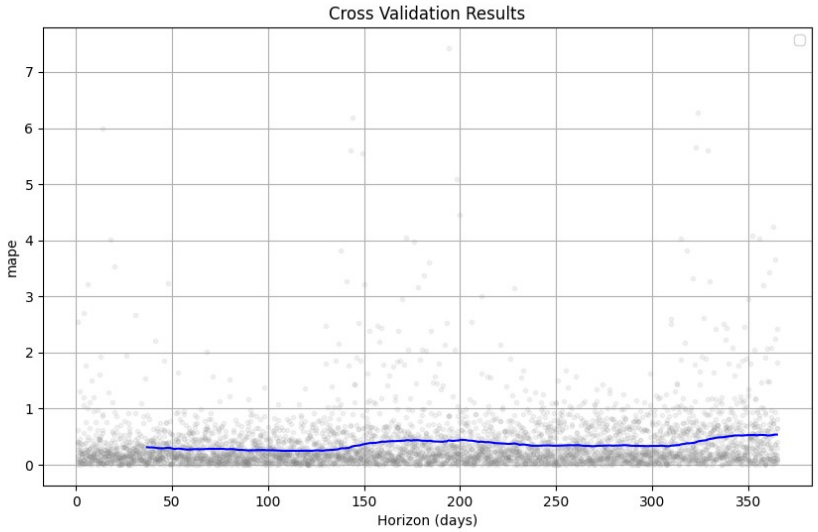
Eg: Accidents are more likely from Tuesday to Friday and in November



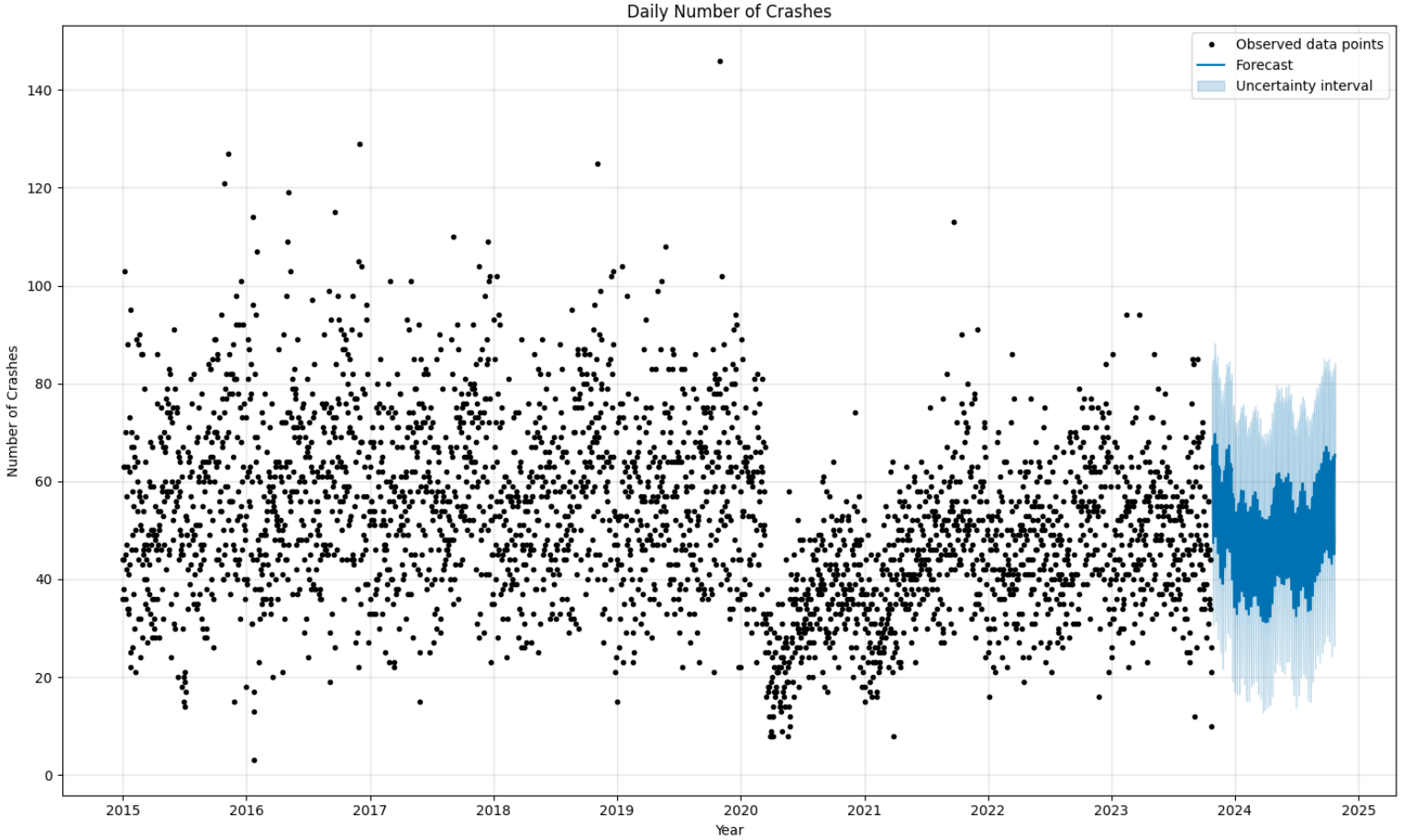




TIME SERIES ANALYSIS



Mean Average Percent Error vs Time(days)



Forecast for the year 2024

DSCC 440 FINAL PROJECT [FALL 2023]

By

- Neel Agarwal
- Jayant Patil

THANK YOU