ANALYZING FACTORS INFLUENCING NON-MOTORIST CRASH OCCURRENCE AND SEVERITY IN OKLAHOMA

Team #20
Tahsin Tabassum
Hananeh Omidi
Aqsa Rehman
Abrar Mohamed

INTRODUCTION

Context

Non-motorists — such as pedestrians and cyclists — are among the most vulnerable road users. Crashes involving non-motorists often result in severe injuries or fatalities, making it critical to understand and predict crash severity.

The Challange

- Can we accurately predict the severity of a crash based on conditions such as weather, road type, time of day, and environmental factors?
- The aim of this project is to develop predictive models and use tools from DSA 5103 to predict crash severity and identify the most significant risk factors for non-motorists.

ABOUT THE DATA

Data Source: Oklahoma Highway Safety Office website.

The Target variable is crash severity, and it has 5 levels: none, minor, possible, serious, and fatal.

Key Variables

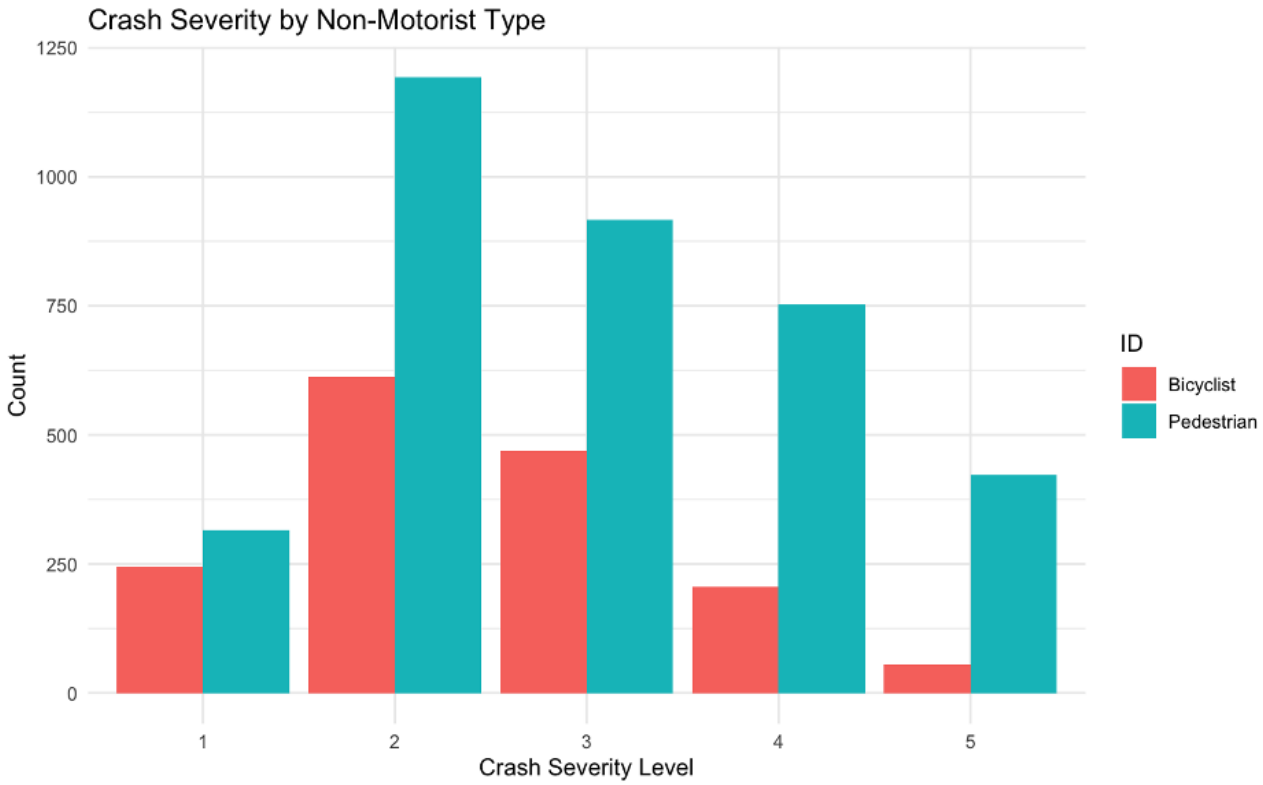
- Weather Conditions (e.g., clear, rainy, foggy)
- Time of Day (e.g., daytime, nighttime)
- Road Type (e.g., highway, urban street)
- Environmental Conditions (e.g., lighting, road surface)

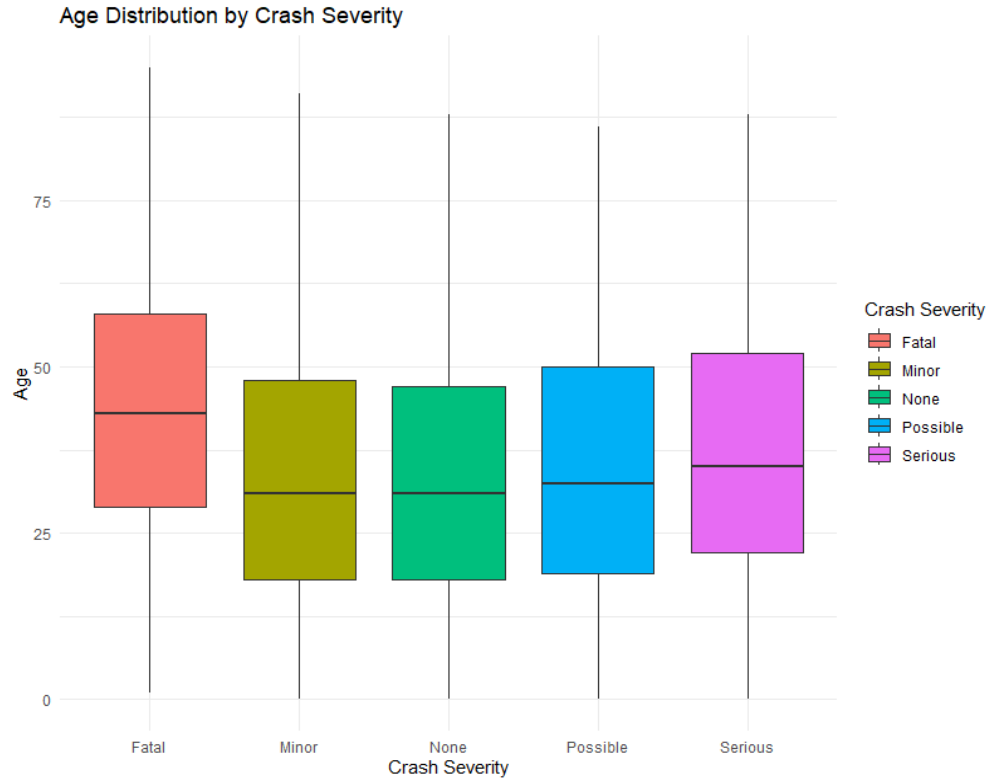
CODE	CRASH_SEVERITY Level	Description
1	**None**	No injuries or significant damage.
2	**Minor**	Minimal injuries or property damage.
3	**Possible**	Potential injuries that may or may not require medical attention.
4	**Serious**	Severe injuries necessitating urgent medical care.
5	**Fatal**	Incidents resulting in fatalities.

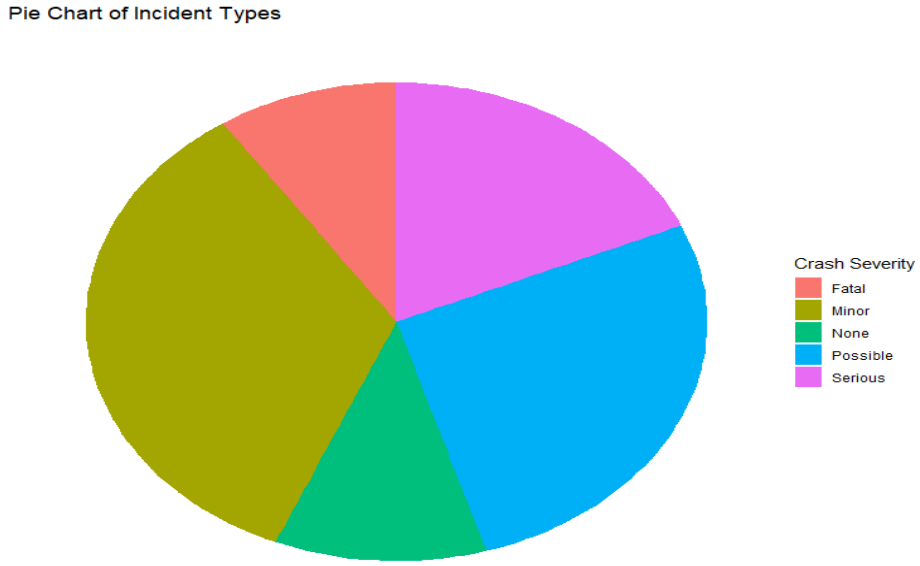
Data size

74122 observation , 67 variables (numerical and categorical with multiple levels)

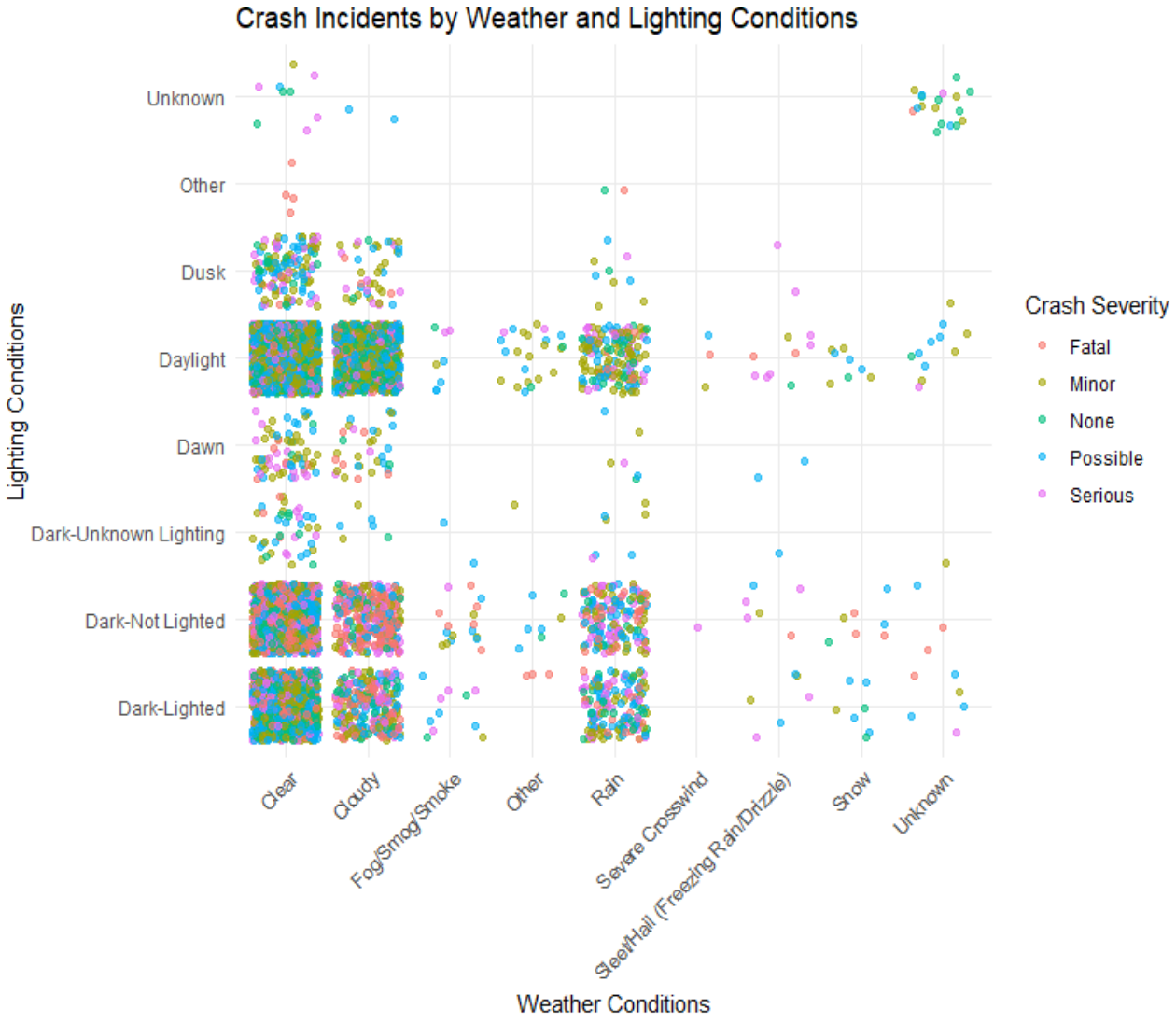
EXPLORATORY DATA ANALYSIS (EDA)

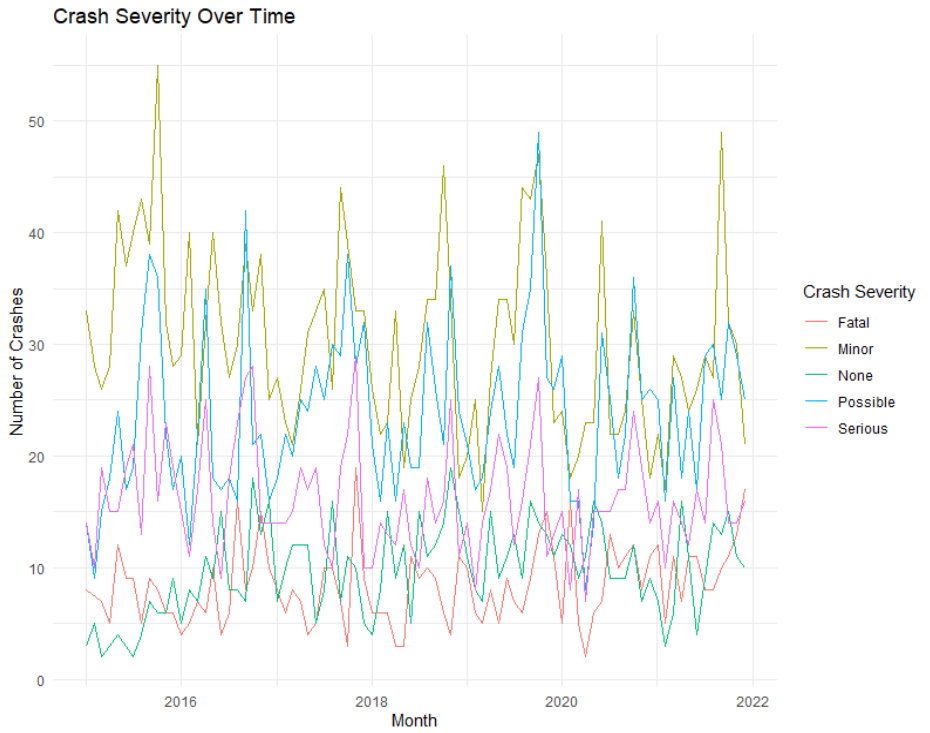


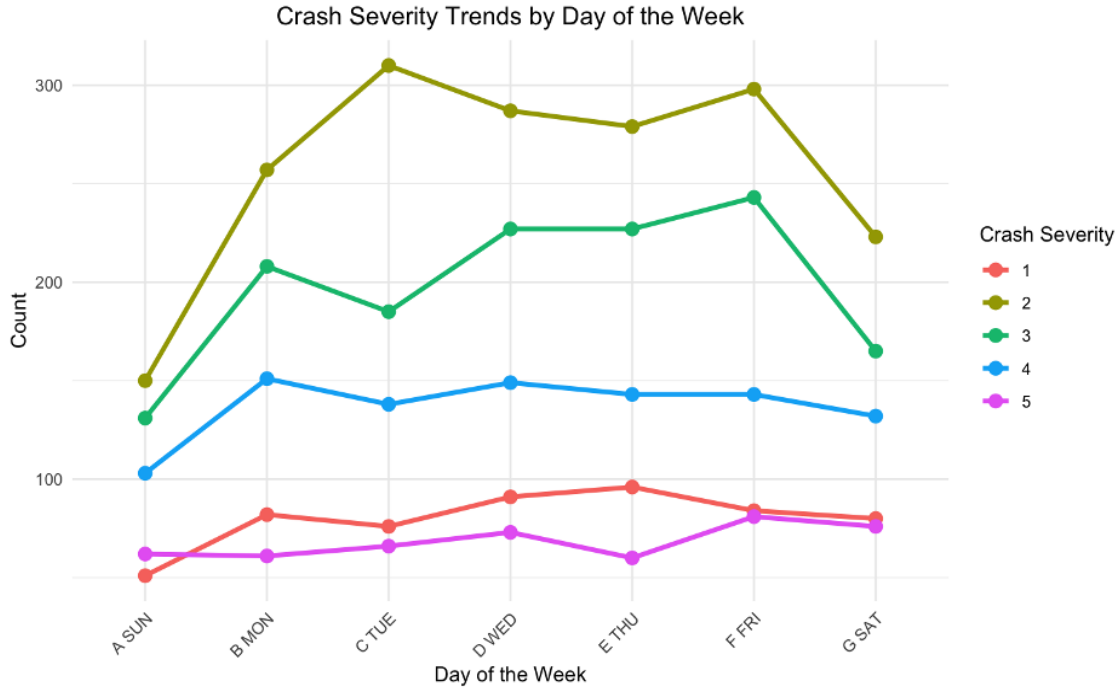




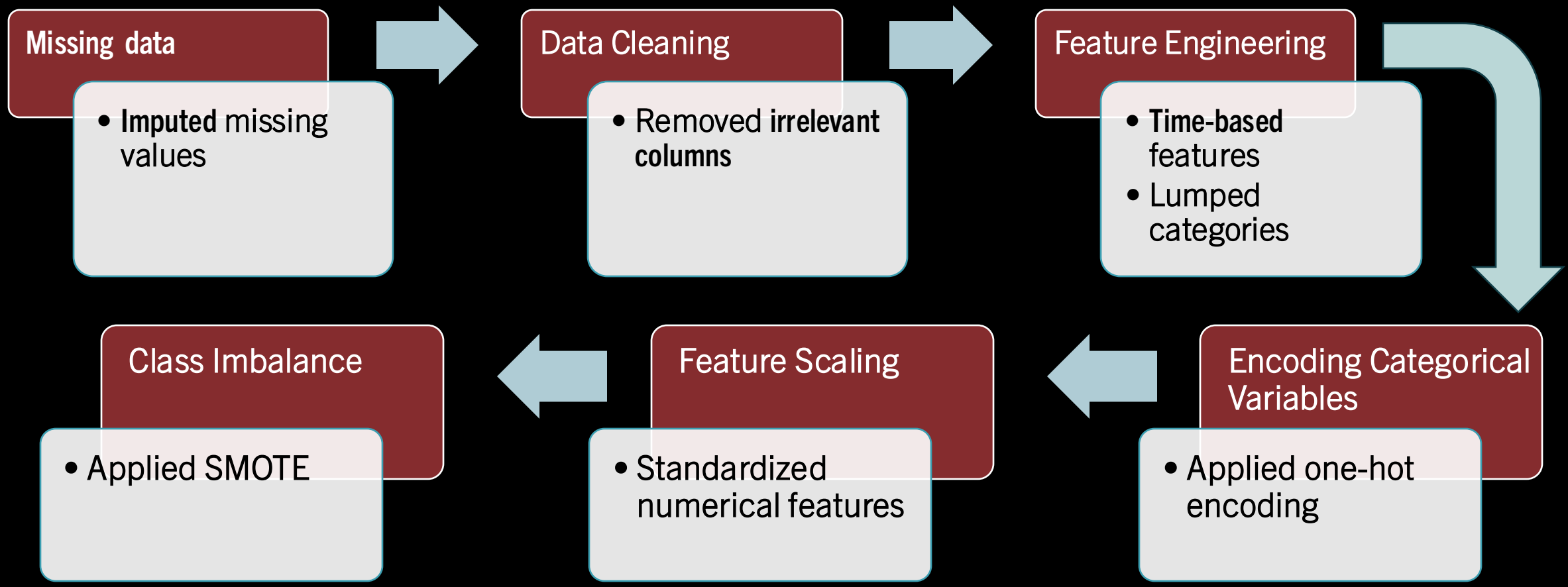
EXPLORATORY DATA ANALYSIS (EDA)







DATA PREPARATION



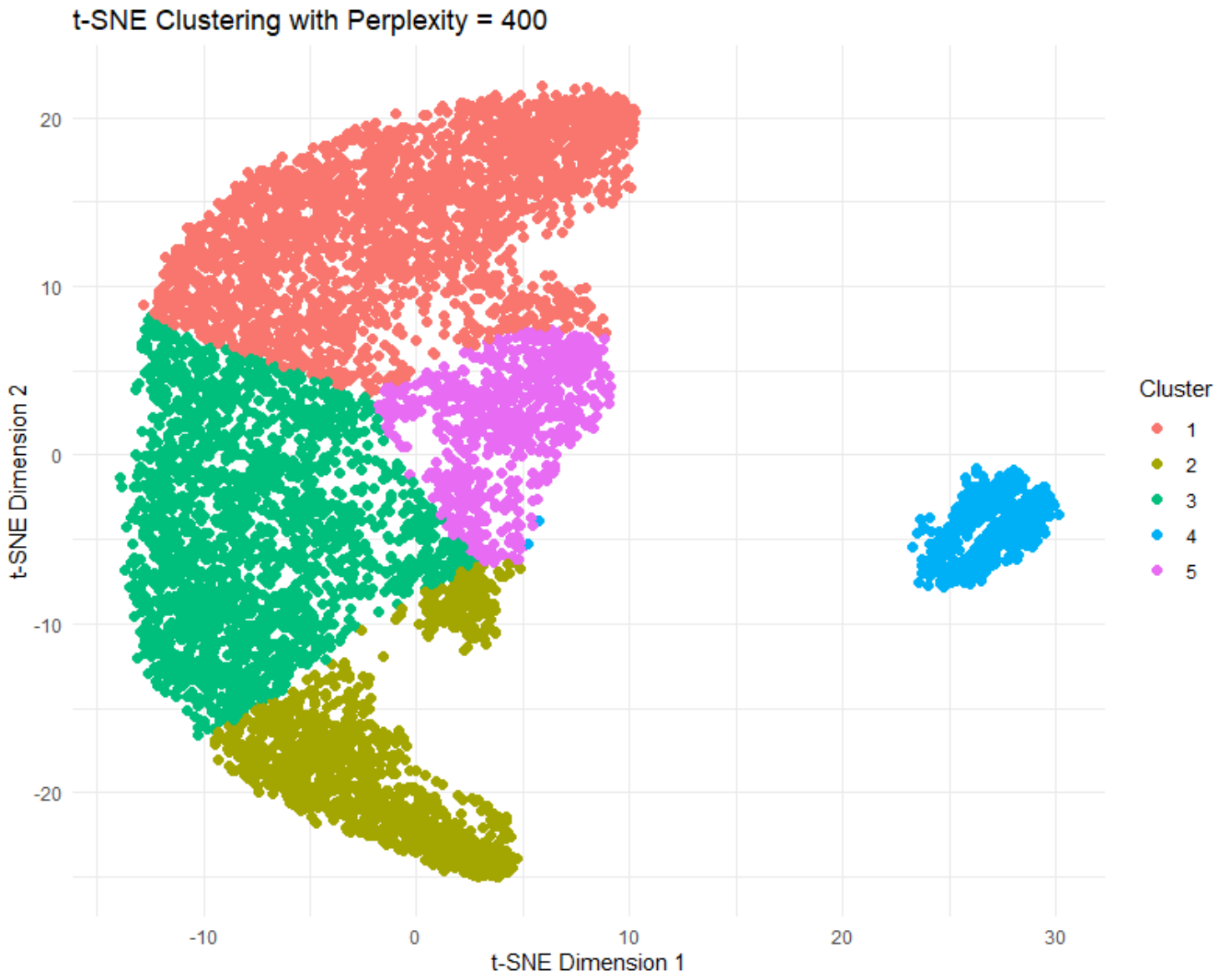
Split data into training and testing sets, 80% for training, 20% for testing
28 variables modeled

MODELING APPROACH

Type	Method	Detailed Classification
Unsupervised	K-Means Clustering	Clustering
Supervised	Multinomial Logistic Regression (MLR)	Statistical Model
	Support Vector Machine (SVM)	Classification and Regression
	Random Forest (RF)	Ensemble Learning
	Artificial Neural Networks (ANN)	Deep Learning
	Conditional Forest (CF)	Ensemble Learning (Tree-Based)
	Extreme Gradient Boosting (XGBoost)	Gradient Boosting Machine
	Stacking	Ensemble Learning (Meta-Modeling)

10-fold cross-validation was utilized to ensure the robustness and reliability of our model performance

CLUSTERING



Approach: K-means clustering applied on PCA-reduced data, visualized using t-SNE, grouped non-motorist crashes into five distinct clusters based on severity, environment, and temporal factors.

Silhouette Score (PAM): 0.561872

Cluster 1: Daytime crashes in clear weather, minor injuries.

Cluster 2: Urban intersection crashes, high occupancy, dense traffic risks.

Cluster 3: Night crashes, poor lighting, severe/fatal outcomes.

Cluster 4: Daytime crashes in clear weather, minor injuries.

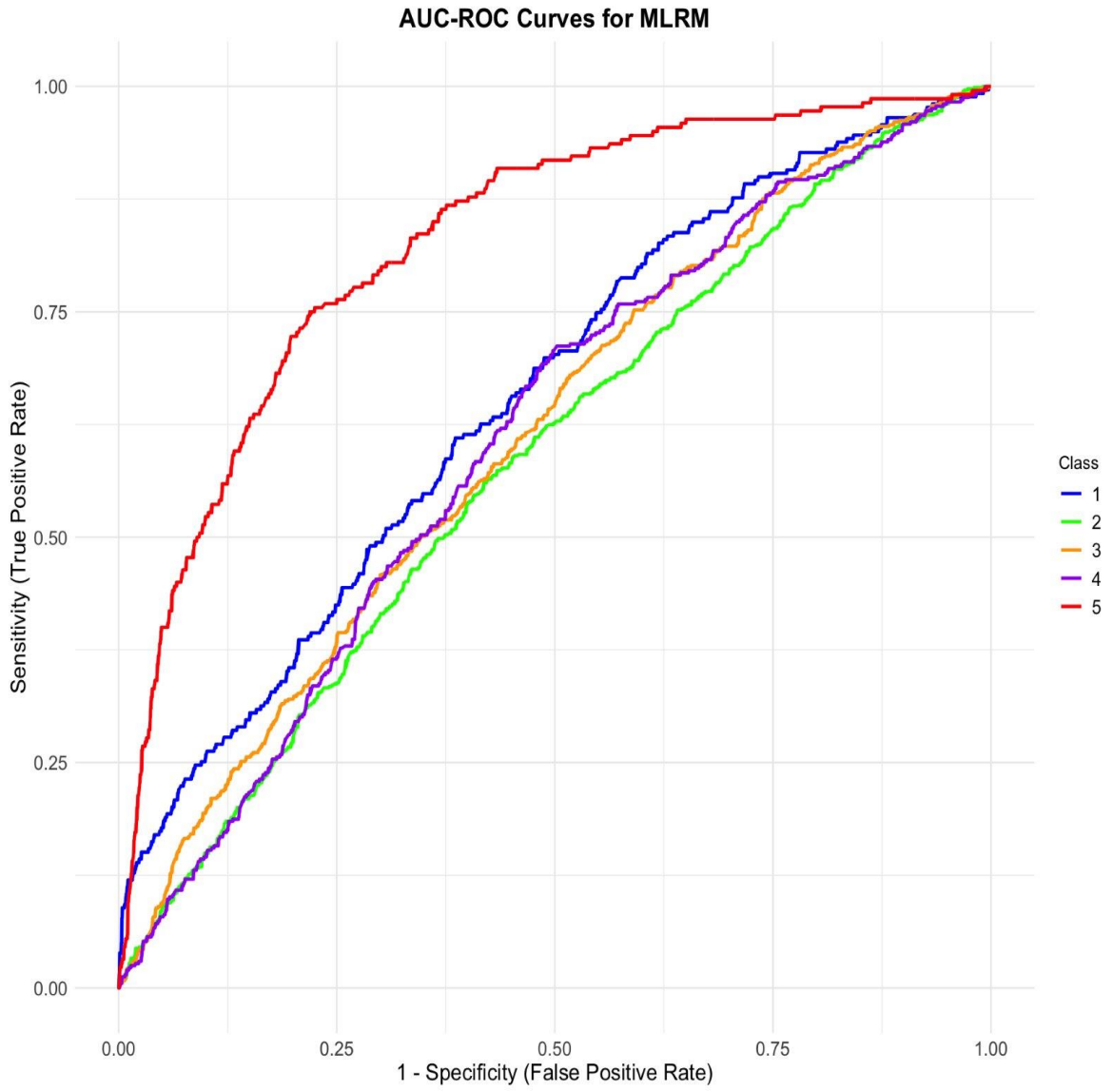
Cluster 5: Hit-and-run crashes, varied severity, unpredictable risks.

STATISTICAL MODEL

Multinomial Logistic Regression

Metric	None	Possible	Minor	Serious	Fatal
Sensitivity	0.11969	0.74930	0.17570	0.09606	0.40455
Specificity	0.98830	0.33730	0.90093	0.93344	0.94162
Pos Pred Value	0.57407	0.35590	0.40075	0.24375	0.43204
Neg Pred Value	0.89493	0.73360	0.74348	0.82219	0.93508
Prevalence	0.11646	0.32820	0.27383	0.18255	0.09892
Detection Rate	0.01394	0.24600	0.04811	0.01754	0.04002
Detection Prevalence	0.02428	0.69110	0.12005	0.07194	0.09263

Metric	None	Possible	Minor	Serious	Fatal	Overall
F1-Score	0.4578	0.7586	0.8089	0.8618	0.9480	**0.77 (Macro Avg)**
AUC	0.6531	0.5861	0.6130	0.6113	0.8283	**0.66 (Macro Avg)**
G-Mean	0.7528	0.5243	0.3348	0.3235	0.2967	**0.446 (Avg)**



ML MODELS

Suport Vector Machine

Class	Precision	Recall	F1_Score	Specificity	G_Mean
None	**0.6206897**	0.1104294	0.1875000	0.9916477	0.3309186
Possible	0.3586498	0.2130326	0.2672956	0.8593895	0.4278761
Minor	0.3903743	0.5770751	0.4657097	0.5318275	0.5539895
Serious	0.2727273	**0.0219780**	0.0406780	0.9867440	0.1472640
Fatal	0.2297297	0.7338129	0.3499142	0.7449664	0.7393687
AVG	0.3744342	0.3312656	0.2622195	0.8229150	0.4398834

Artificial Neural Network

Class	Precision	Recall	F1_Score	Specificity	G_Mean
None	0.4594595	0.1042945	0.1700000	0.9848140	0.3204850
Possible	0.3603604	**0.1002506**	0.1568628	0.9343201	0.3060493
Minor	0.3747045	0.6264822	0.4689349	0.4568789	0.5350014
Serious	1.0000000	**0.0073260**	0.0145455	1.0000000	0.0855921
Fatal	0.2004132	0.6978417	0.3113965	0.7114094	0.7045929
AVG	**0.4789875**	0.3072390	0.2243479	0.8174845	0.3903441

Random Forest

Class	Precision	Recall	F1_Score	Specificity	G_Mean
None	0.4600000	0.1411043	0.2159624	0.9794989	0.3717681
Possible	0.3668478	0.3383459	0.3520209	0.7844588	0.5151877
Minor	0.4228571	0.5849802	0.4908789	0.5852156	0.5850979
Serious	0.4032258	0.1831502	0.2518892	0.9386910	0.4146341
Fatal	0.3487395	0.5971223	0.4403183	0.8844146	0.7267074
AVG	**0.40033404**	**0.36894058**	**0.35021394**	**0.83445578**	**0.52267904**

XG Boost

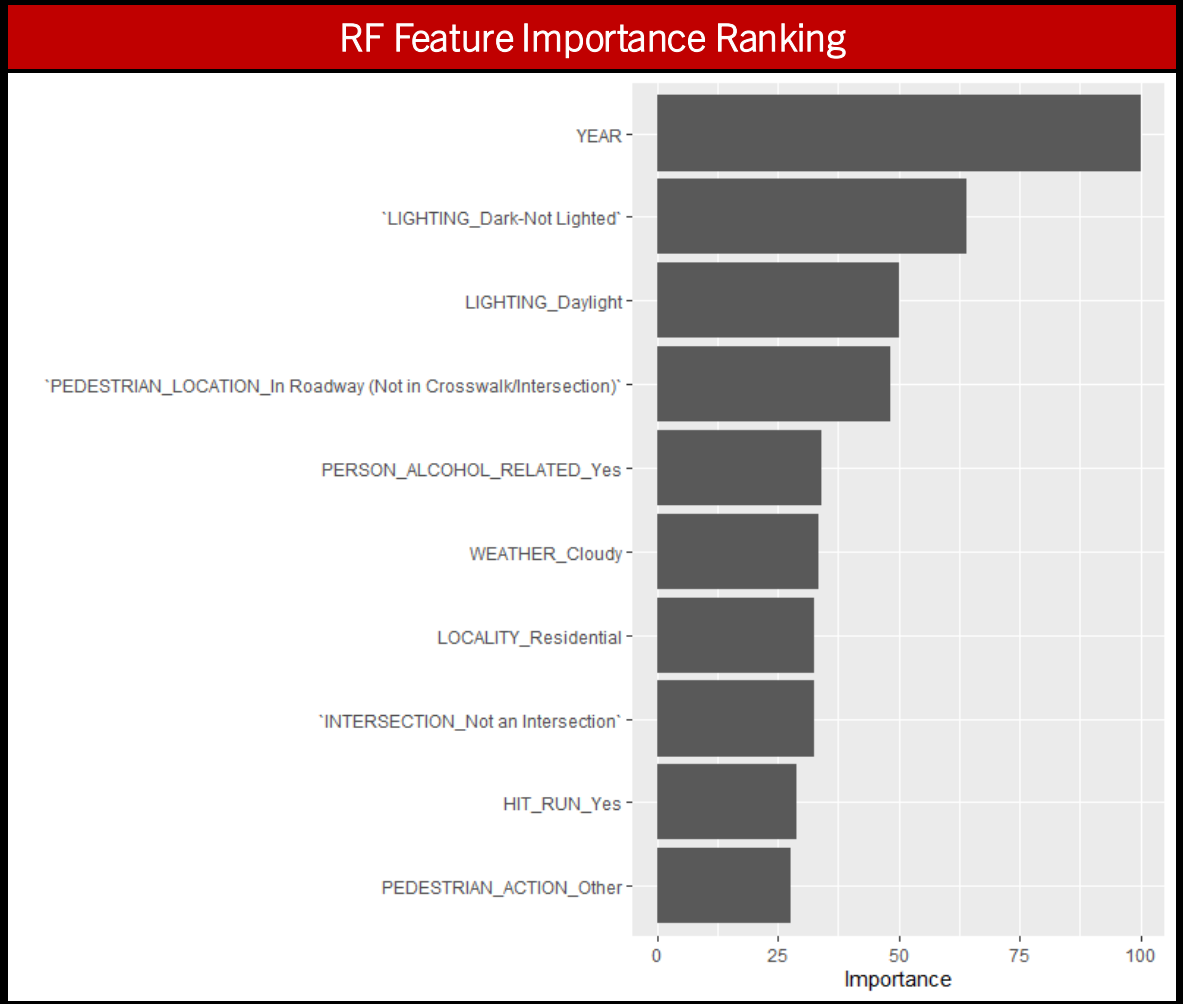
Class	Precision	Recall	F1_Score	Specificity	G_Mean
None	0.5227273	0.1411043	0.2222222	0.9840547	0.3726316
Possible	0.3738872	0.3157895	0.3423913	0.8048104	0.5041336
Minor	0.3975000	0.6284585	0.4869832	0.5051335	0.5634318
Serious	0.3120567	0.1611722	0.2125604	0.9196355	0.3849930
Fatal	0.4240506	**0.4820144**	0.4511785	0.9321402	**0.6703022**
AVG	**0.40604436**	**0.34570778**	**0.34306712**	**0.82915486**	**0.49909844**

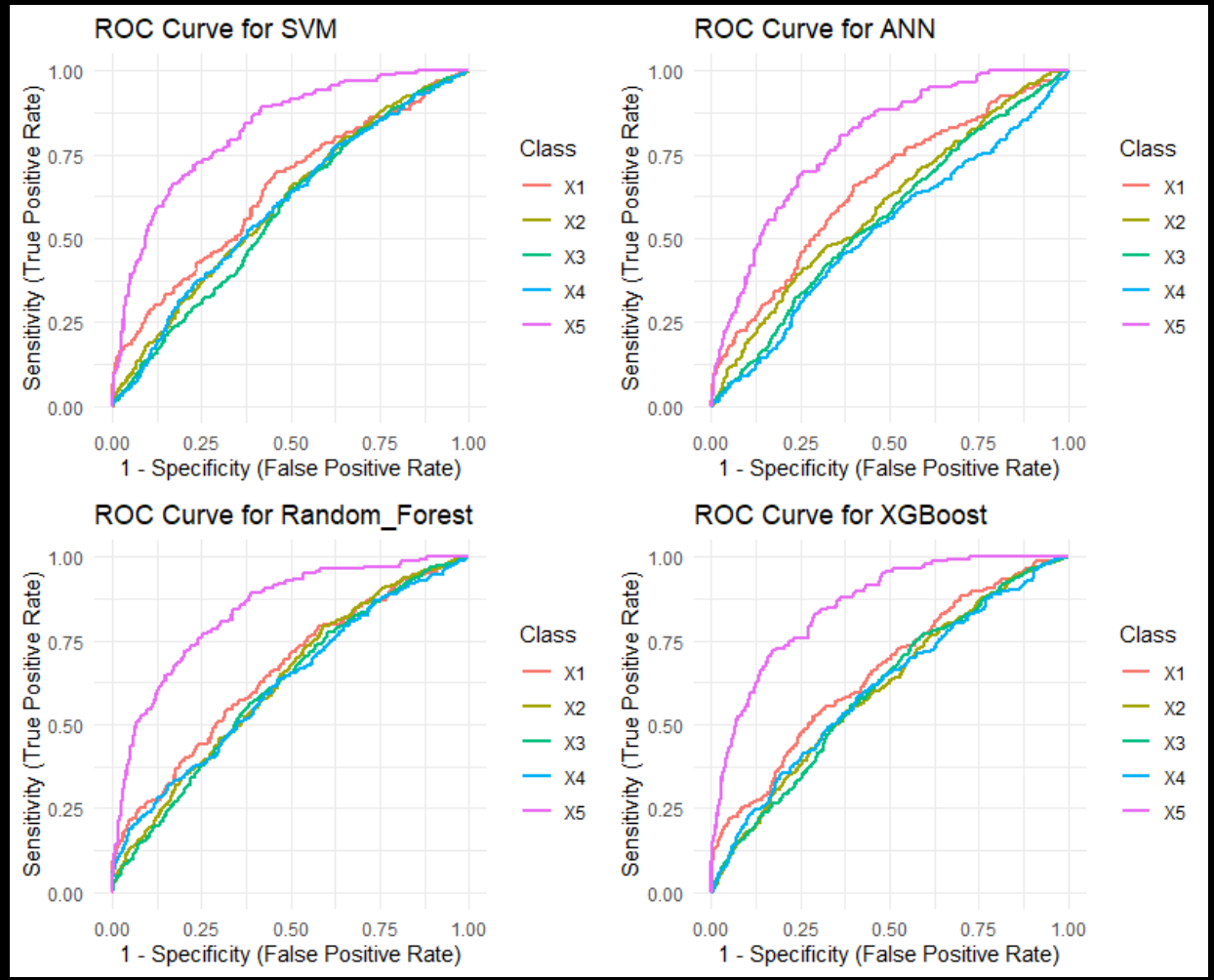
STACKED PERFORMANCE MEASURE

Stacking					
Class	Precision	Recall	F1_Score	Specificity	G_Mean
None	0.8695652	0.1226994	0.2150538	0.9977221	0.3498855
Possible	0.4420601	0.5162907	0.4763006	0.7594820	0.6261897
Minor	0.4575342	0.6600791	0.5404531	0.5934292	0.6258675
Serious	0.4500000	0.2307692	0.3050847	0.9362055	0.4648090
Fatal	0.5206612	0.4532374	0.4846154	0.9567487	0.6585091
AVG	**0.54796414**	**0.4532374**	**0.4846154**	**0.89636552**	**0.54505216**

GRAPHS OF RF FEATURE RANK AND ROC CURVES

RF Feature Importance Ranking





CONCLUSION, INSIGHTS AND RECOMMENDATIONS

Insights

Key Risk Factors

Time of Day and Lighting

- Nighttime conditions significantly increase crash severity

Weather Conditions

- Clear weather is associated with less severe outcomes, while adverse weather can increase crash severity.

Road Type and Urban Density

- Crashes at urban intersections are more frequent

Recommendations

Enhance Street Lighting

- Install or upgrade streetlights in high-risk areas, especially intersections and crosswalks

Weather-Responsive Traffic Systems

- Implement adaptive traffic control systems that adjust signal timings during adverse weather

Intersection Safety Measures

- Implement traffic-calming measures and dedicated non-motorist pathways at busy urban intersections
- Install smart signals with longer pedestrian crossing times and sensors that detect non-motorists to improve safety

THANK YOU, ANY QUESTION?