MORINGA

Discover · Grow · Transform

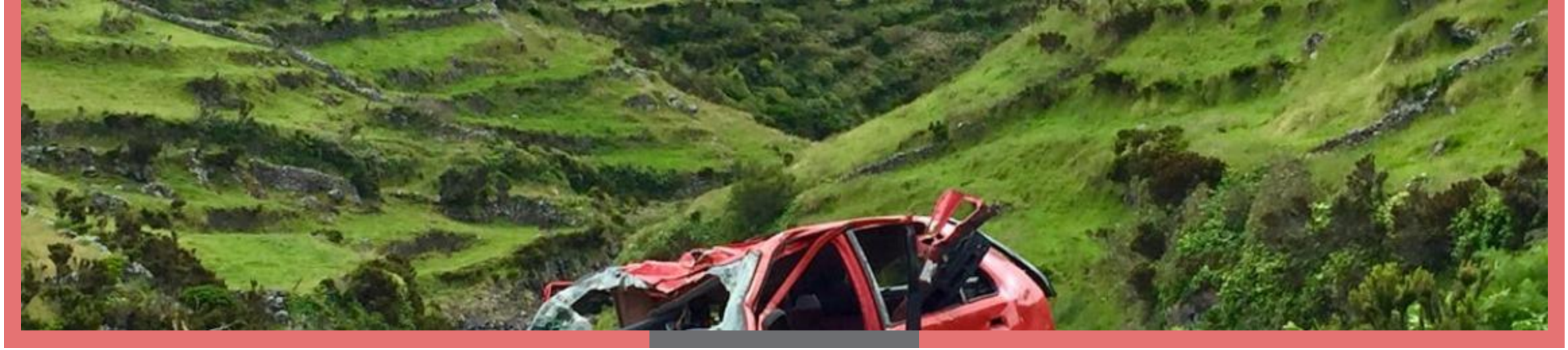
Predictive Modeling for Vehicle Crash Causes

Rahab Gachie
Benson Ouma
Franciscar Ndanu

June - 2025

Agenda

- Introduction
- Business Understanding
- Objectives
- Data
- Methodology
- Results
- Conclusion
- Recommendations



Introduction

Motor vehicle crashes are a major public safety issue in many global cities. Understanding the causes of these accidents is vital for enhancing road safety, optimizing traffic laws, and guiding policy. With open data access, machine learning tools can identify patterns and predict outcomes for actionable insights.

This project utilizes public crash, vehicle, and people datasets from Chicago to create a predictive model.

Business Understanding

Traffic accident Statistics

Motor vehicle accidents lead to thousands of injuries and fatalities each year, significantly impacting urban safety.

Economic Impact

These accidents result in substantial economic losses, prompting the need for measures to enhance safety in urban areas.

Contributory Causes

Identifying the main causes of crashes can guide effective interventions by policymakers and urban planners.

Objective

Build a model that can predict the primary contributory cause of a car accident, given information about the car, the people in the car, the road conditions etc.

Key Questions

1. What are the most significant factors contributing to vehicle crashes in Chicago?
2. How do weather and lighting conditions impact accident causation patterns?
3. How do driver demographics (age, gender) correlate with accident causes?

Data

Vehicle Data

Contains information about vehicles involved in a traffic crash

Driver/Passenger Data

Contains information about people involved in a crash and if any injuries were sustained

Crashes

Shows information about each traffic crash on city streets within the City of Chicago

Methodology

Data Collection

EDA

Preprocessing

Model Selection & Building

Model Evaluation

Evaluation Metrics

Accuracy

Measures the frequency of correct identifications of crash causes.

Precision

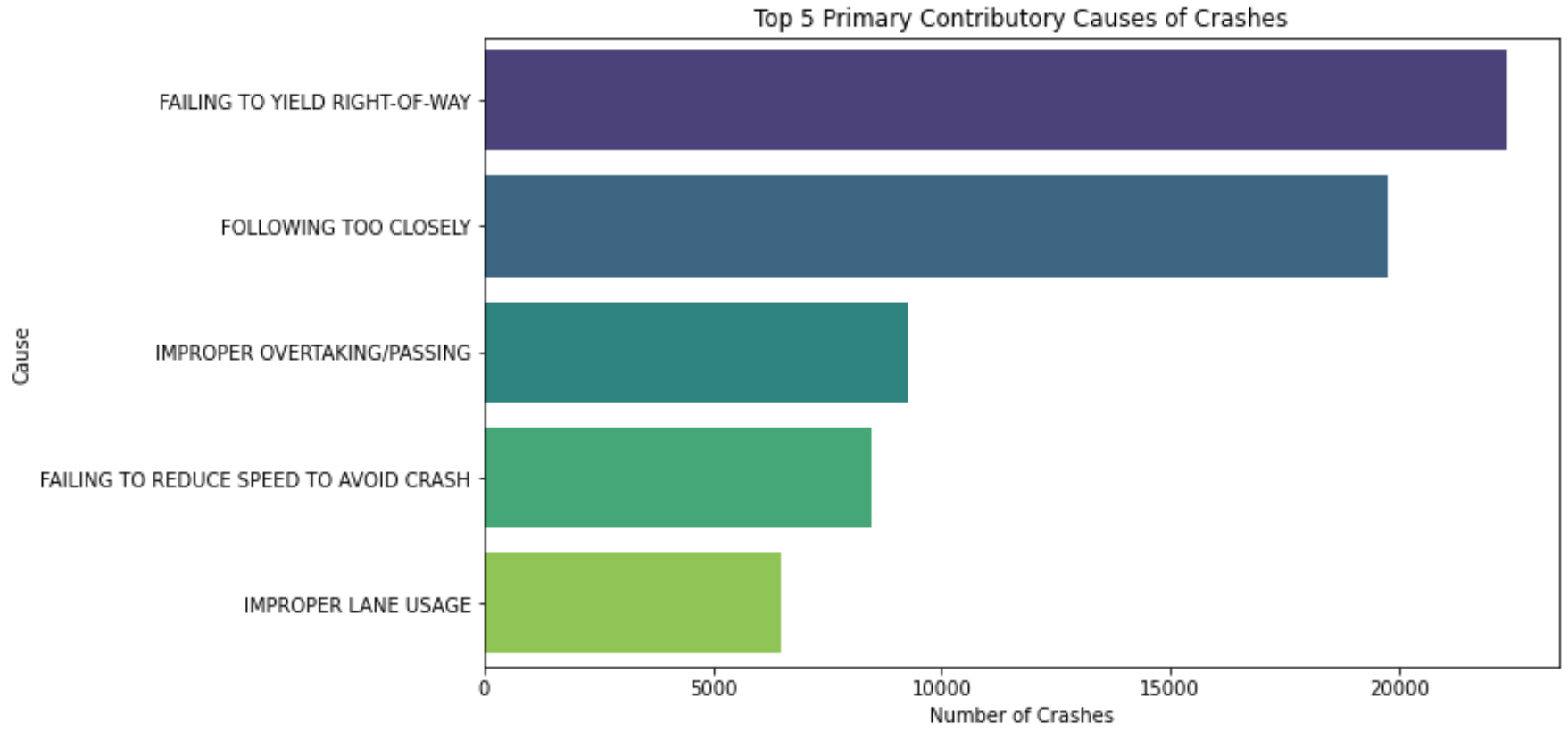
Calculates the proportion of true positive results among predicted positives, indicating the model's selectiveness in identifying vehicle crash causes.

Recall

Evaluates the model's efficacy in identifying all relevant crash instances, highlighting its capacity to capture widespread issues accurately.

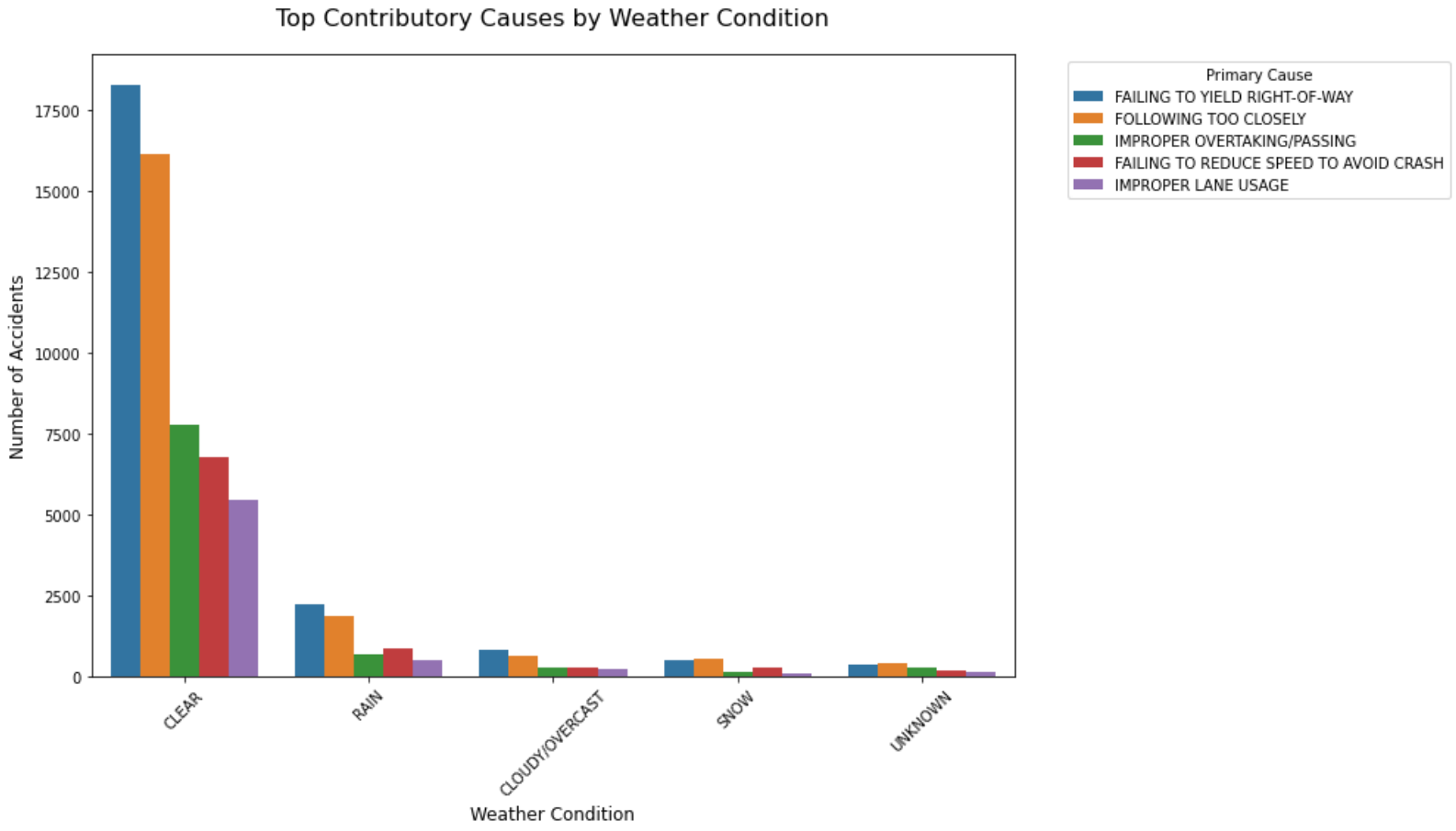
Results

Top 5 Primary Contributory Causes



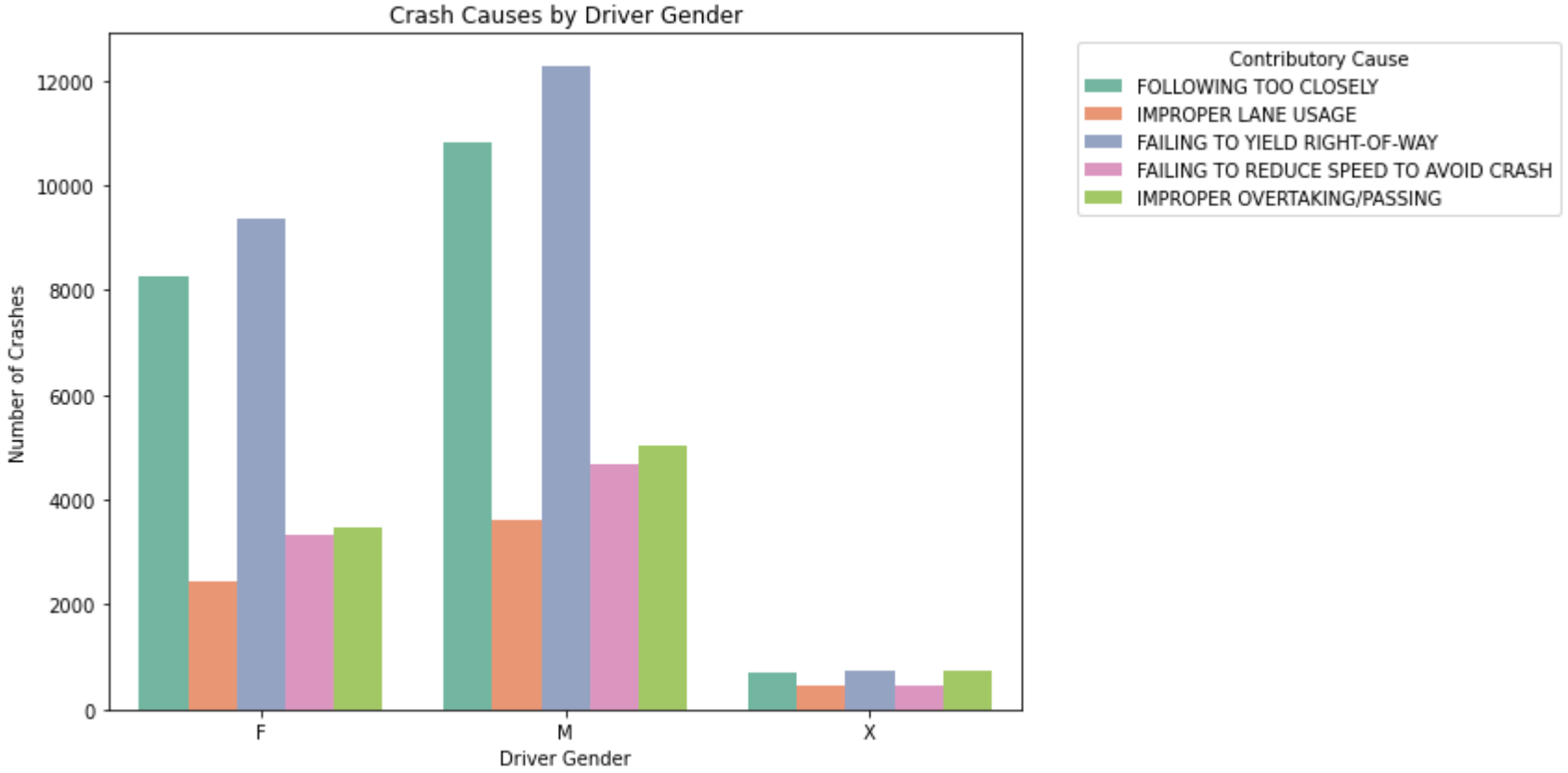
Results

Top Contributory Causes by Weather Condition



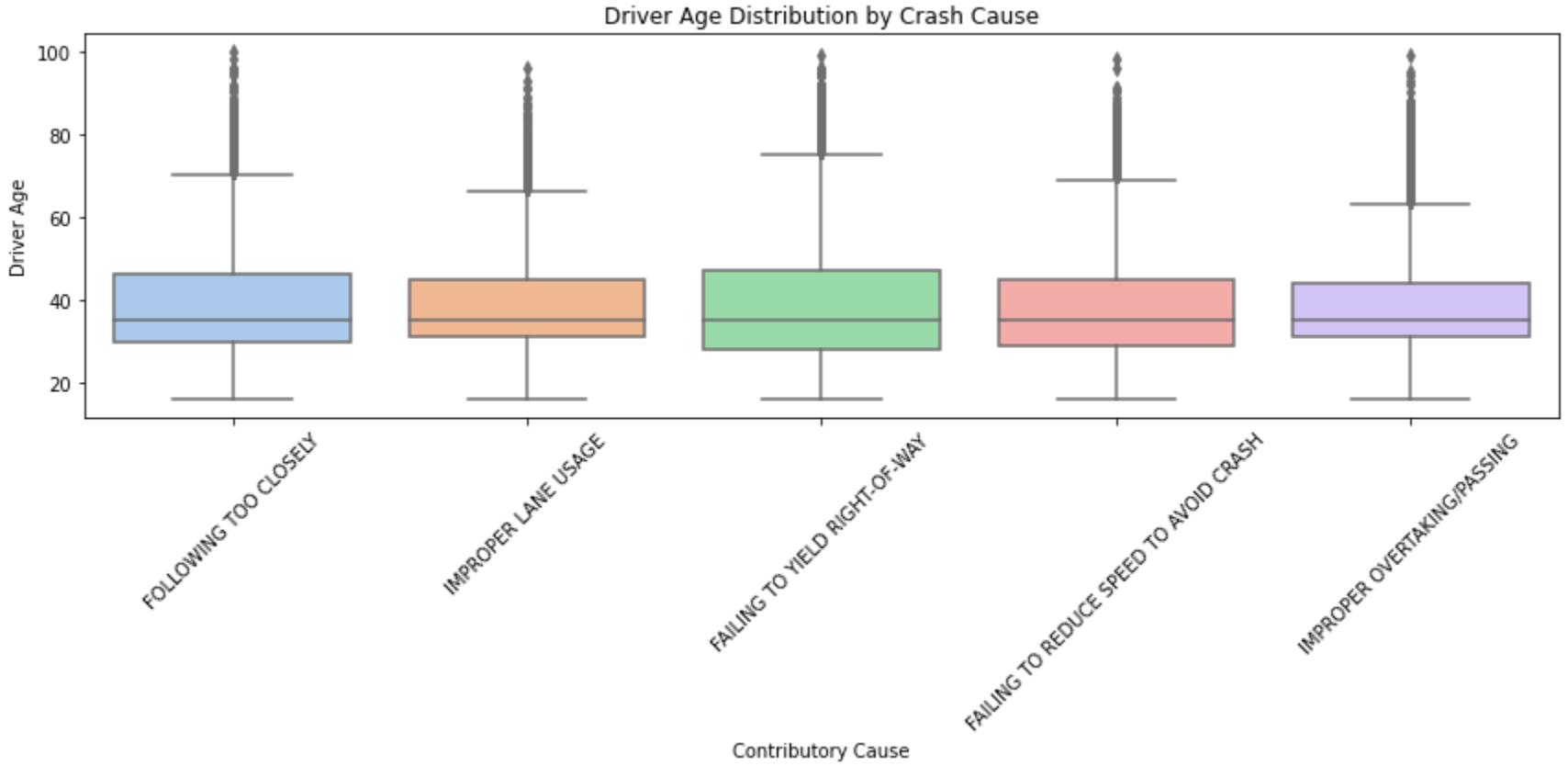
Results

Crash Causes by Driver Gender



Results

Driver Age Distribution by Crash Cause

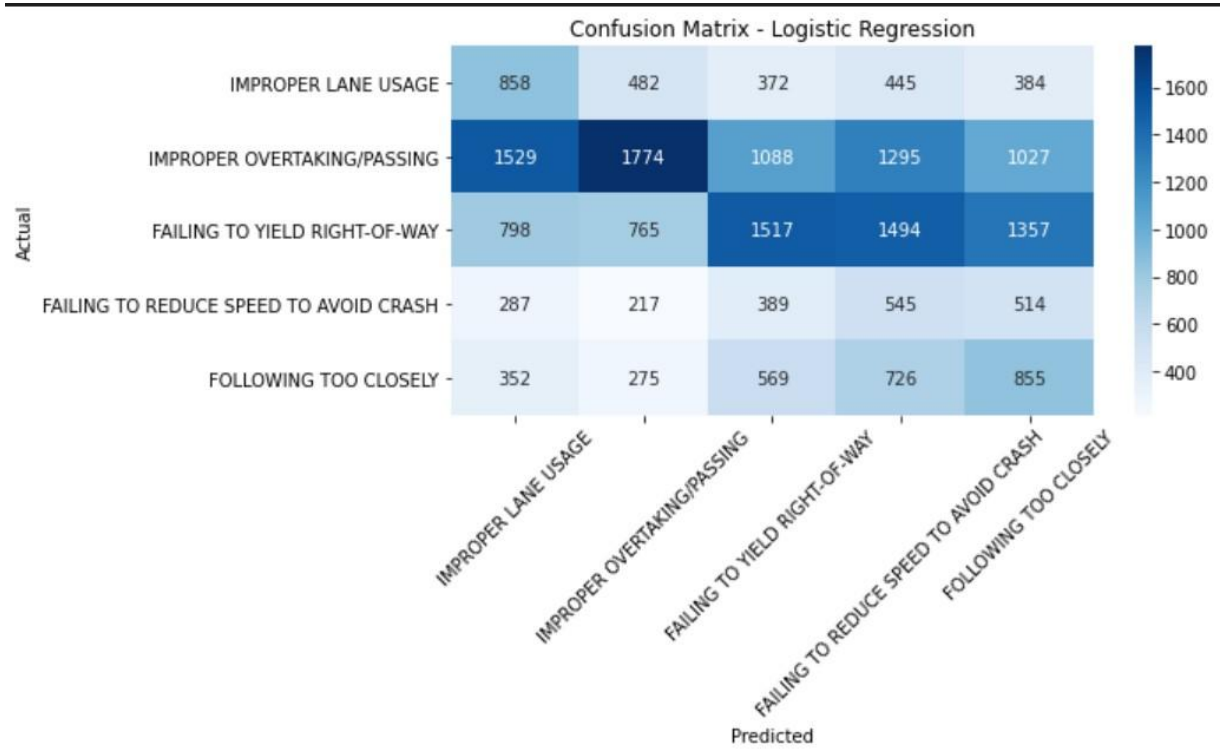


Results

Logistic Regression Evaluation

Accuracy: 0.279
Classification Report:

	precision	recall	f1-score	support
FAILING TO REDUCE SPEED TO AVOID CRASH	0.22	0.34	0.27	2541
FAILING TO YIELD RIGHT-OF-WAY	0.50	0.26	0.35	6713
FOLLOWING TOO CLOSELY	0.39	0.26	0.31	5931
IMPROPER LANE USAGE	0.12	0.28	0.17	1952
IMPROPER OVERTAKING/PASSING	0.21	0.31	0.25	2777
accuracy			0.28	19914
macro avg	0.29	0.29	0.27	19914
weighted avg	0.35	0.28	0.29	19914

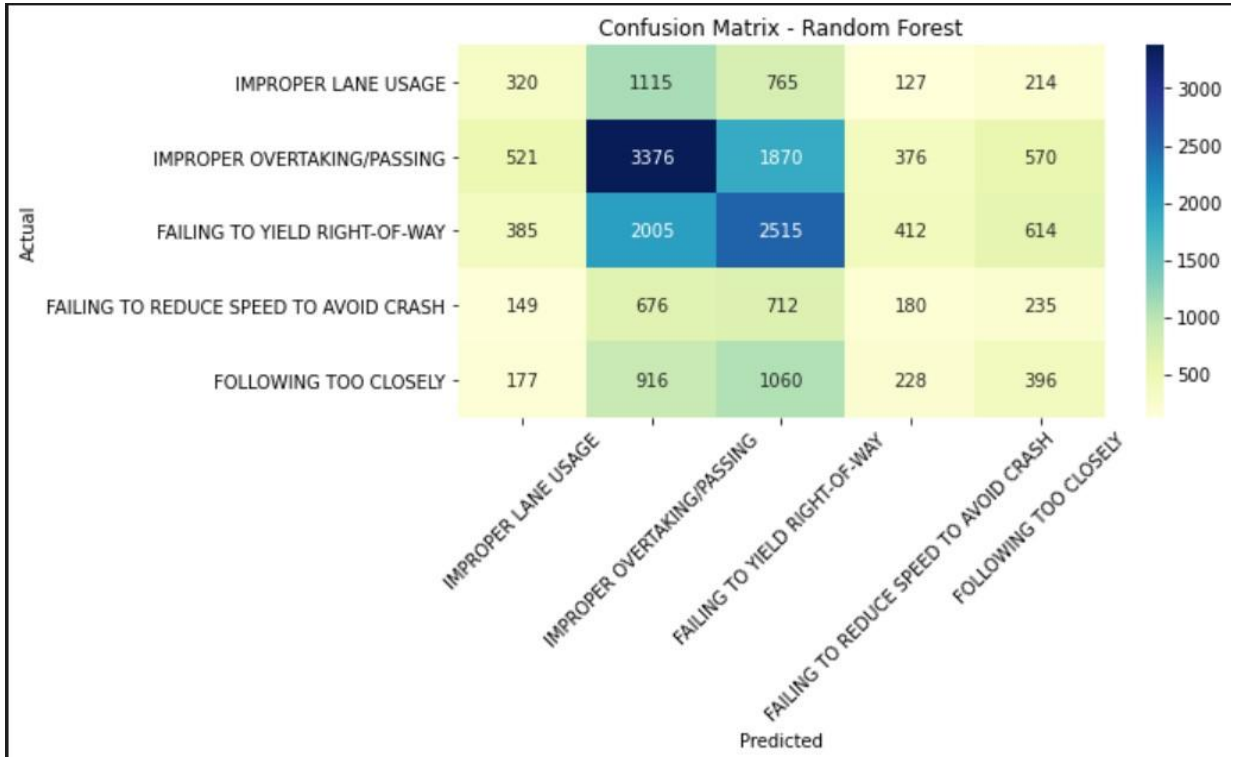


Results

Random Forest Evaluation

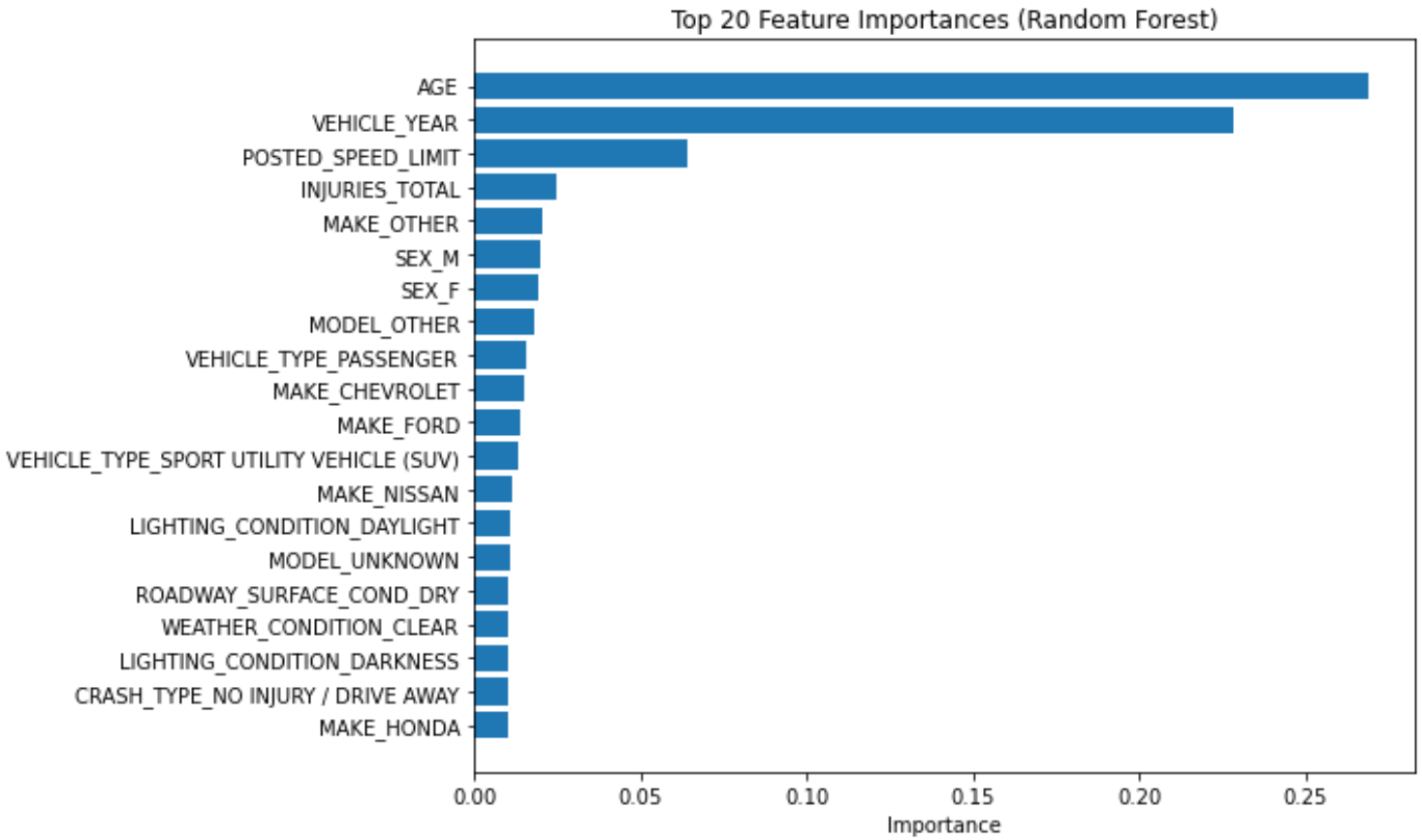
Accuracy: 0.3408155066787185

	precision	recall	f1-score	support
FAILING TO REDUCE SPEED TO AVOID CRASH	0.21	0.13	0.16	2541
FAILING TO YIELD RIGHT-OF-WAY	0.42	0.50	0.46	6713
FOLLOWING TOO CLOSELY	0.36	0.42	0.39	5931
IMPROPER LANE USAGE	0.14	0.09	0.11	1952
IMPROPER OVERTAKING/PASSING	0.20	0.14	0.16	2777
accuracy			0.34	19914
macro avg	0.26	0.26	0.26	19914
weighted avg	0.32	0.34	0.32	19914



Results

Feature Importance



Conclusions

- Most Common Crash Causes :Failing to yield right-of-way, following too closely, improper overtaking/passing, failing to reduce speed, and improper lane usage
- Logistic Regression offers transparency on which features influence crash causes. Random Forest provides better predictive accuracy for practical use
- The models can help identify likely causes of a crash based on input conditions, enabling targeted interventions

Recommendations

- Driver-Focused Programs: Develop public awareness and training campaigns addressing top causes.
- Targeted interventions; Focus enforcement and road safety improvements on road signage.
- - Evaluate using additional metrics (e.g., ROC AUC, confusion matrices)