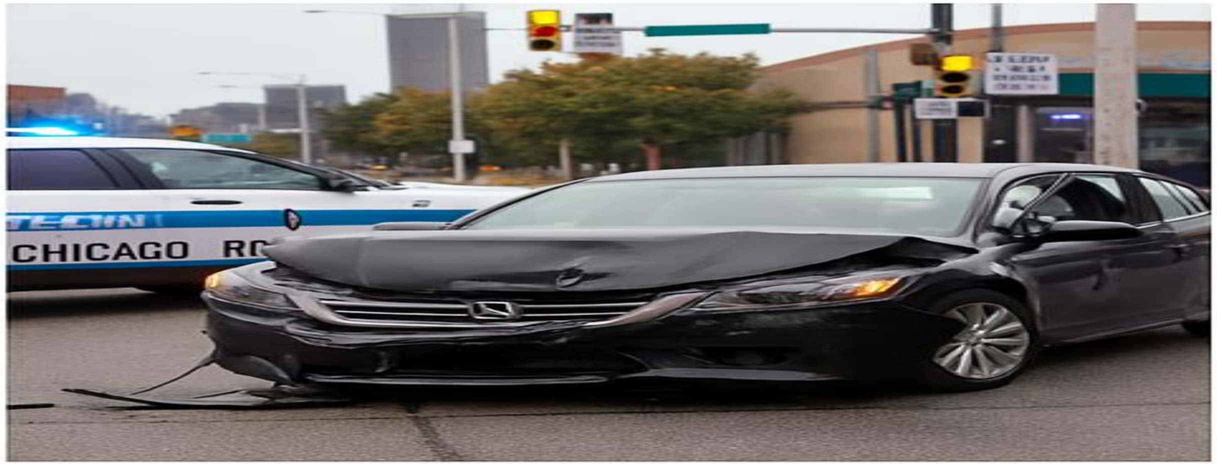
MORINGA

Phase 4 Project Modelling Chicago Accident Causes

Group 13

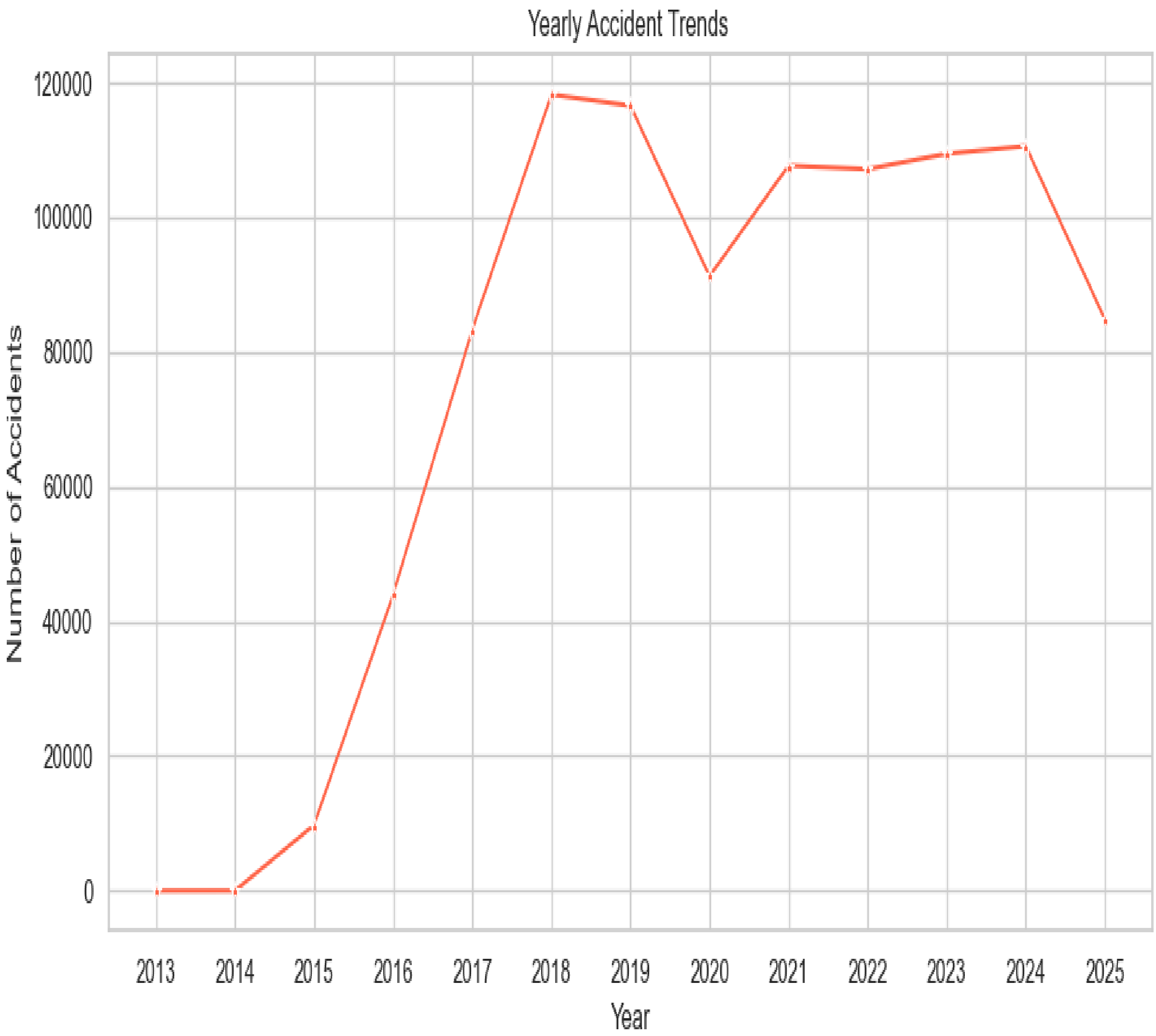


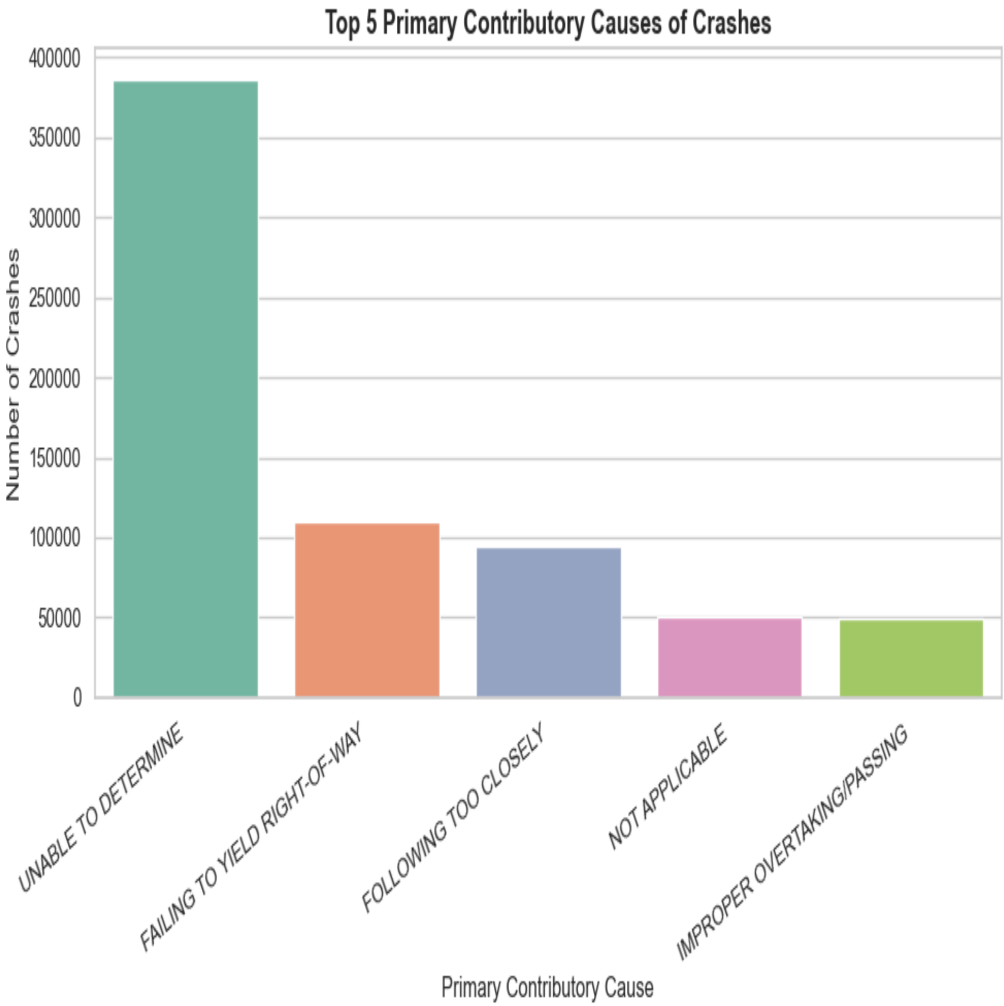
Introduction

The Chicago Accident Problem

- Chicago has experienced a sharp rise in road accidents, peaking at over 118,000 crashes in 2018, with consistently high levels in subsequent years.
- Analysis of crash causes reveals a **heavy concentration** in a few categories, with “Unable to Determine” and right-of-way failures accounting for a disproportionate share of incidents.
- The persistence of high-severity crashes highlights gaps in driver behavior, roadway conditions, and enforcement effectiveness.
- Stakeholders currently lack **predictive tools** to identify which conditions most strongly influence crash severity.
- This project responds to that gap by **modelling crash causes** using machine learning to support proactive safety interventions.

Accident trends and primary contributors





Objectives & Significance

The project is anchored on the following objectives

---

i. Develop a predictive model capable of classifying the primary contributory cause of a crash.

ii. Identify major behavioral, environmental, and roadway factors influencing crash severity.

iii. Analyze spatial patterns to detect high-risk zones and accident hotspots.

iv. Assess temporal patterns such as peak hours, days, and seasonal trends.

v. Provide actionable, evidence-based recommendations to support Chicago’s road-safety planning.

Why This Analysis Matters for Policy and Stakeholders?

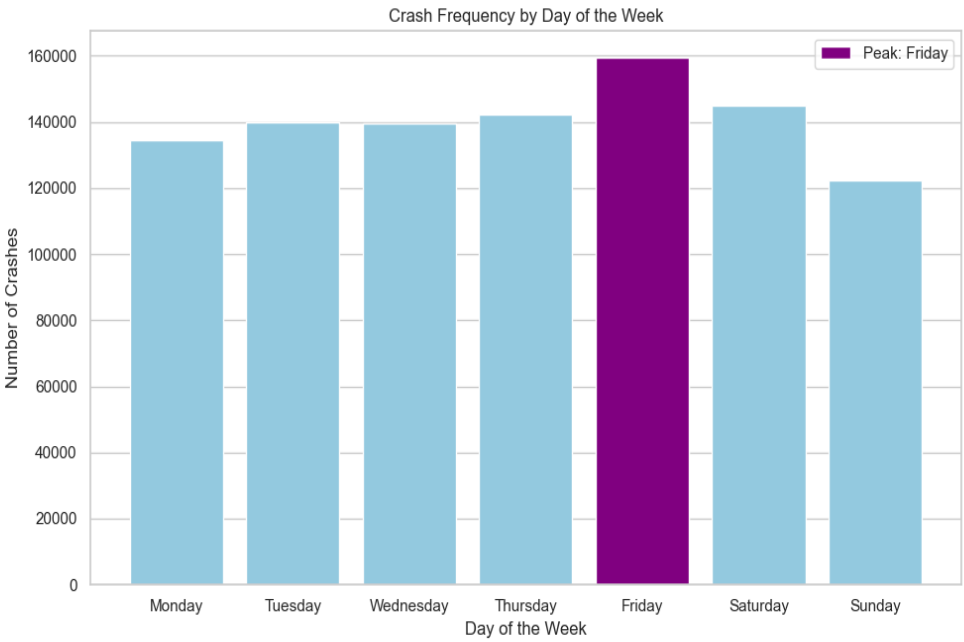
- ❑Enables the City of Chicago to anticipate high-risk conditions rather than respond reactively.
- ❑Supports data-driven resource allocation, especially for law enforcement and roadway maintenance teams.
- ❑Helps transportation planners prioritize infrastructure improvements in accident-prone corridors
- ❑Equips policymakers with evidence for targeted safety campaigns (e.g., speeding, right-of-way compliance).
- ❑Strengthens long-term planning by quantifying how environmental and behavioral factors contribute to crashes.

Methodology

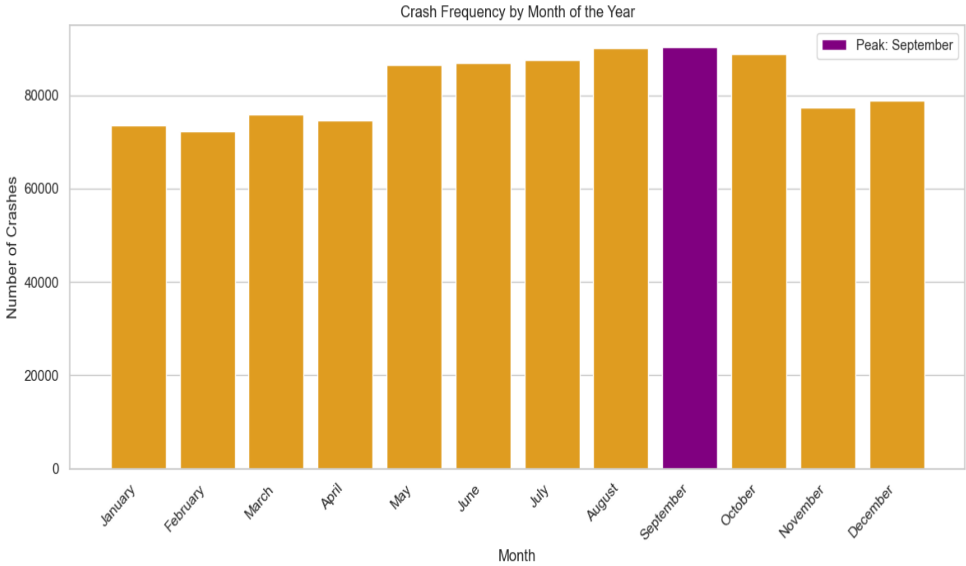
- **Data Cleaning & Integration**: Combined crash, vehicle, and people datasets; removed missing and inconsistent records; standardized time, location, and categorical fields.
- **Exploratory Analysis**: Identified temporal, seasonal, and spatial patterns using yearly trends, monthly peaks, hourly distributions, and weekday–hour heatmaps.
- **Feature Engineering:** Aggregated crash causes into broader groups, encoded categorical variables, scaled numeric features, and constructed meaningful predictors from weather, road conditions, and time-of-day data.
- **Class Balancing & Correlation Assessment:** Applied Random Oversampling (ROS) to address class imbalance and used correlation analysis to reduce noise and improve model signal quality.
- **Model Training & Evaluation:** Developed and compared **Decision Tree, Random Forest, XGBoost**, and **Neural Network models**, evaluating accuracy, ROC performance, and interpretability to select the best classifier.

Preliminary findings

- **CRASH INTENSITY PEAKS ON FRIDAYS AND DURING SEPTEMBER,** INDICATING THE NEED FOR TARGETED ENFORCEMENT AND TRAFFIC-FLOW MANAGEMENT DURING END-OF-WEEK AND LATE-SUMMER HIGH-RISK PERIODS.

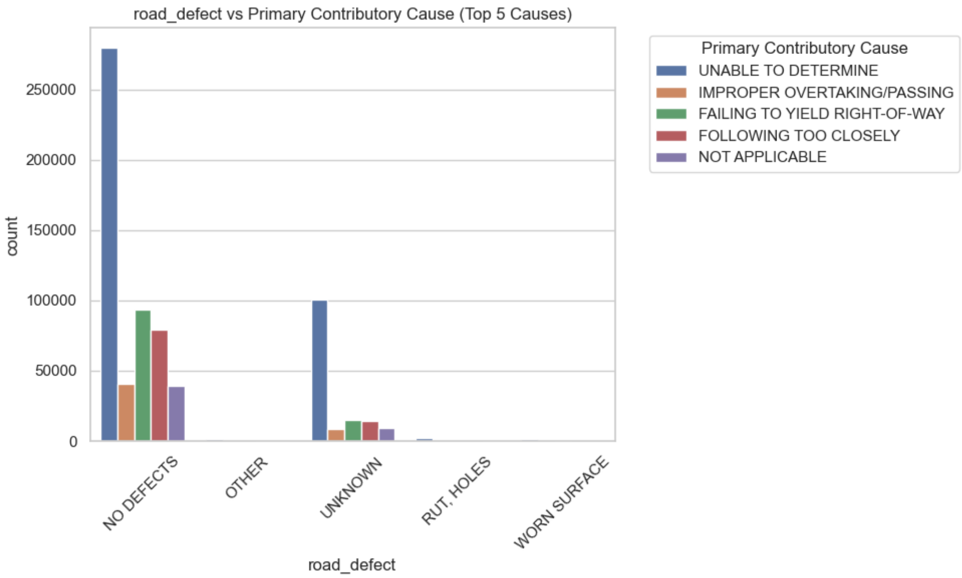


- THESE TEMPORAL PATTERNS SUPPORT **STRATEGIC DEPLOYMENT OF POLICE, SIGNAGE, AND CONGESTION CONTROLS** TO REDUCE ACCIDENT LIKELIHOOD DURING PREDICTABLE SURGE WINDOWS.

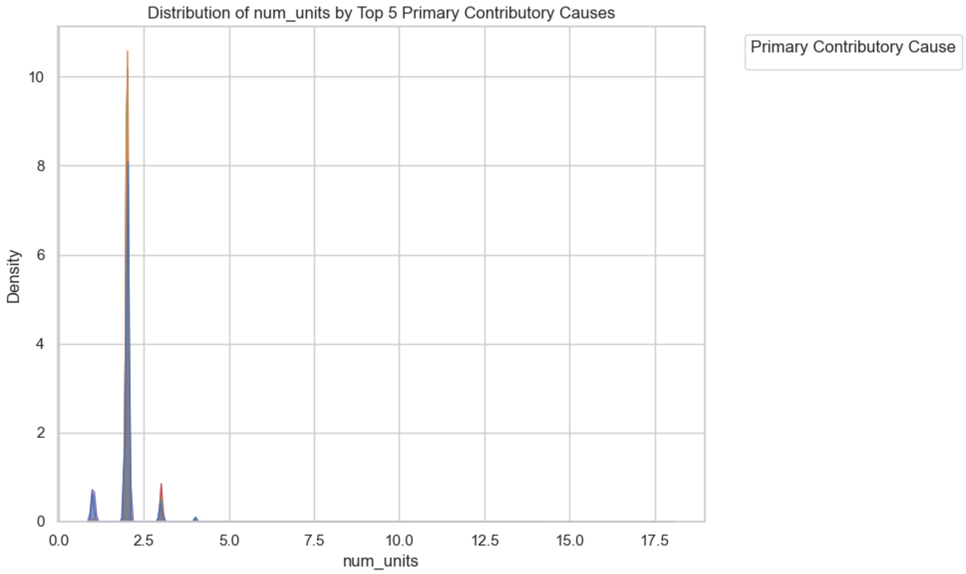


Preliminary findings

THE DOMINANCE OF CRASHES ON ROADS MARKED "NO DEFECTS" SUGGESTS THAT INTERVENTIONS SHOULD FOCUS ON DRIVER BEHAVIOR POLICIES, INCLUDING STRICTER ENFORCEMENT OF OVERTAKING, RIGHT-OF-WAY, AND SAFE-DISTANCE RULES.

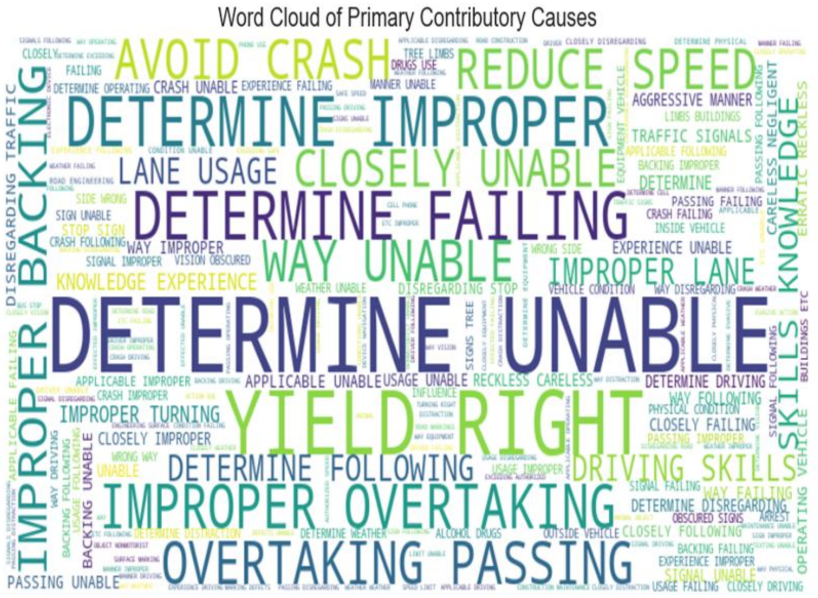


SINCE SEVERE CAUSES FREQUENTLY INVOLVE MULTI-VEHICLE COLLISIONS, POLICYMAKERS SHOULD PRIORITIZE LANE-MANAGEMENT STRATEGIES, CONGESTION CONTROL, AND PUBLIC AWARENESS CAMPAIGNS TARGETING TAILGATING AND UNSAFE MERGING.

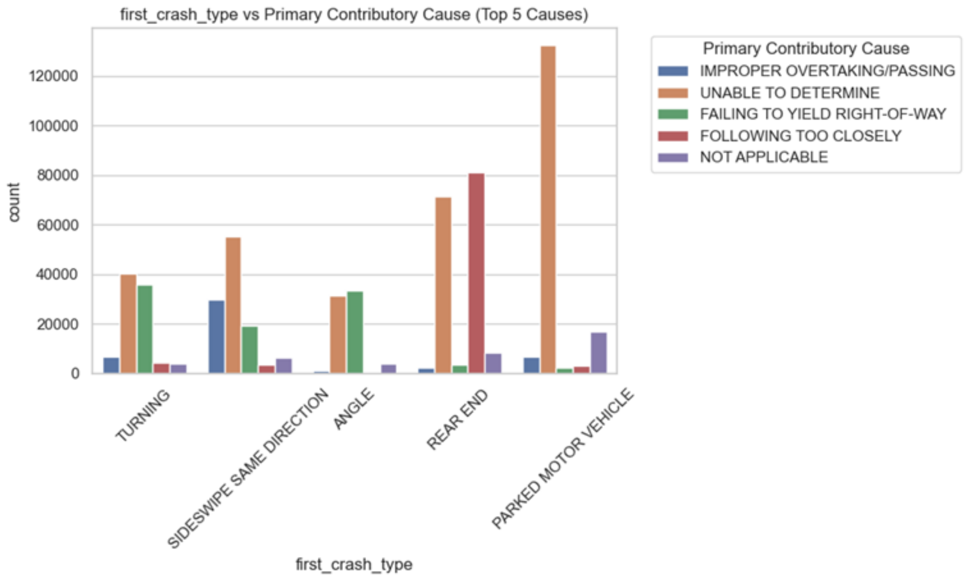


Preliminary findings

THE DOMINANCE OF HUMAN-ERROR TERMS SUCH AS *IMPROPER*, *FAILING*, *CLOSELY*, AND *YIELD* CONFIRMS THAT POLICY SHOULD PRIORITIZE **BEHAVIORAL ENFORCEMENT AND DRIVER-EDUCATION REFORMS** RATHER THAN SOLELY INFRASTRUCTURE REMEDIES.

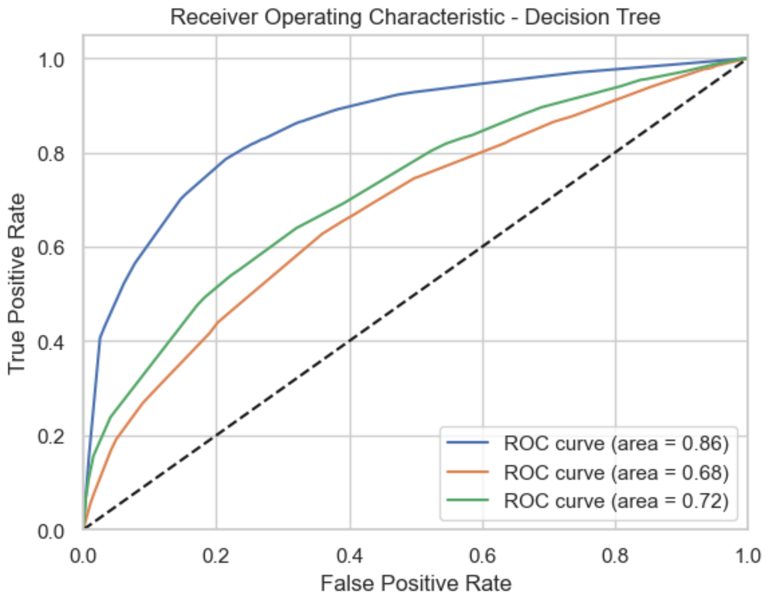


REAR-END AND SIDESWIPE COLLISIONS ARE STRONGLY ASSOCIATED WITH FOLLOWING-DISTANCE AND LANE-DISCIPLINE FAILURES, HIGHLIGHTING THE NEED FOR **TARGETED ENFORCEMENT CAMPAIGNS AND ADAPTIVE TRAFFIC-FLOW POLICIES** TO REDUCE THESE PREVENTABLE CRASH TYPES.

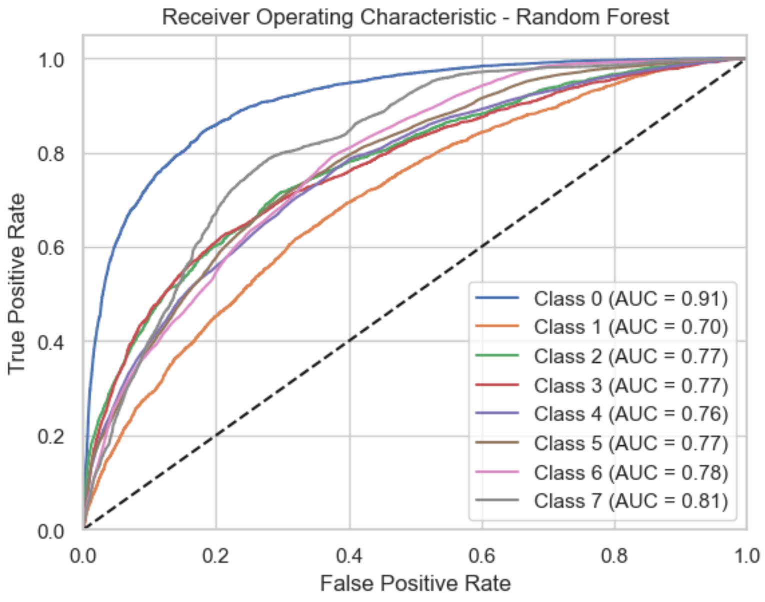


Inferential insights

THE ROC CURVES SHOW THAT BOTH THE DECISION TREE AND RANDOM FOREST MODELS ACHIEVE STRONG DISCRIMINATORY POWER, WITH SEVERAL CLASSES EXHIBITING AUC VALUES ABOVE 0.75–0.90, INDICATING THAT THE MODELS RELIABLY DISTINGUISH BETWEEN DIFFERENT CONTRIBUTORY-CAUSE CATEGORIES.

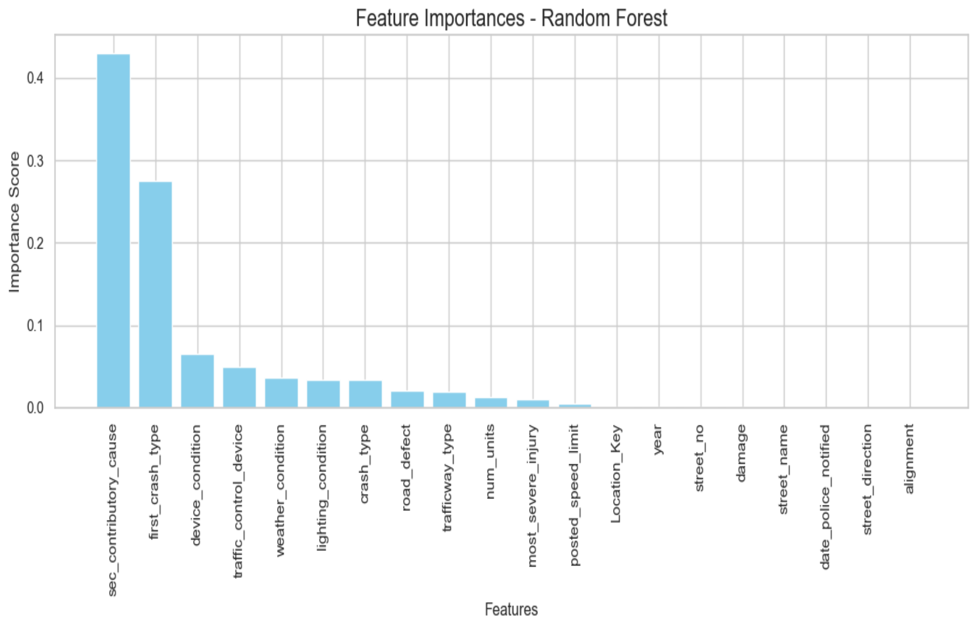


THE CONSISTENTLY HIGHER AUC SCORES FOR SOME CLASSES (E.G., CLASS 0) SUGGEST THAT CERTAIN CRASH-CAUSE PATTERNS ARE MORE PREDICTABLE, ENABLING **POLICYMAKERS** TO PRIORITIZE TARGETED INTERVENTIONS WHERE MODEL CONFIDENCE IS STRONGEST.

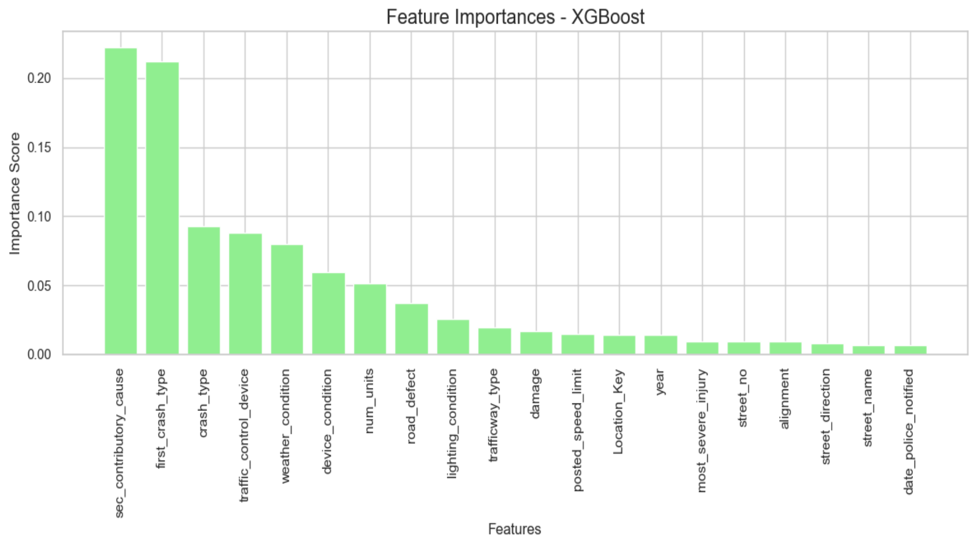


Inferential insights

THE FEATURE-IMPORTANCE PROFILES FROM RANDOM FOREST AND XGBOOST SHOW THAT **SECONDARY CONTRIBUTORY CAUSES, INITIAL CRASH TYPE, ROAD DEFECTS, AND TRAFFIC CONTROL CONDITIONS** ARE THE DOMINANT PREDICTORS OF CRASH CAUSATION.



THESE INSIGHTS PROVIDE A RELIABLE EMPIRICAL FOUNDATION FOR TARGETED SAFETY INTERVENTIONS, ESPECIALLY IN INFRASTRUCTURE MAINTENANCE AND TRAFFIC-FLOW REGULATION.



Recommendations

What should be done ?

- **Prioritize Infrastructure Repairs (Road Defects: ~0.04–0.06 weight):** Road-defect variables show meaningful predictive contribution, reinforcing the need for structured maintenance of worn surfaces, potholes, and degraded lanes in high-volume corridors.

- **Strengthen Traffic Control Compliance (Traffic-Control Devices: ~0.06–0.10 weight):** The influence of device condition and traffic-control indicators highlights the need to intensify enforcement at intersections, crossings, and signalized zones.

- **Enhance Behavior-Focused Interventions (Behavioral Predictors: ~0.10–0.15 combined weight):** Factors such as overtaking errors, following too closely, and right-of-way failures carry significant explanatory weight, justifying targeted driver-behaviour campaigns and automated enforcement.

- **Redesign High-Risk Crash Typologies (Crash Type / First Crash Type: ~0.20–0.26 weight):** High importance of “rear-end”, “sideswipe”, and related crash configurations indicates the need for lane-marking improvements, roadway redesigns, and speed-management systems.

- **Improve Supervisory Data Systems (Secondary Contributory Cause: ~0.20–0.22 weight):** The dominant influence of secondary contributory cause underscores reporting gaps; enforcing stricter and standardized crash documentation will improve analytic precision and policy response.