Predictive Analysis of Traffic Accident Causes: Insights and Recommendations

A DATA-DRIVEN APPROACH TO IMPROVING ROAD SAFETY

Gladwell Chepkorir

Introduction

- **Problem Statement**: The City of Chicago has experienced a high incidence of traffic accidents, leading to significant concerns about road safety and the well-being of its residents. To address this issue, it is crucial to identify the primary contributory causes of these accidents. By understanding the factors that most frequently contribute to crashes, we can implement targeted interventions and policies aimed at reducing the number of traffic incidents. This project focuses on analyzing traffic crash data to uncover key patterns and contributory causes, with the ultimate goal of enhancing road safety and minimizing accidents in the city.

- **Objective**: Predicting the primary contributory cause of traffic accidents and providing actionable insights.

Dataset Overview

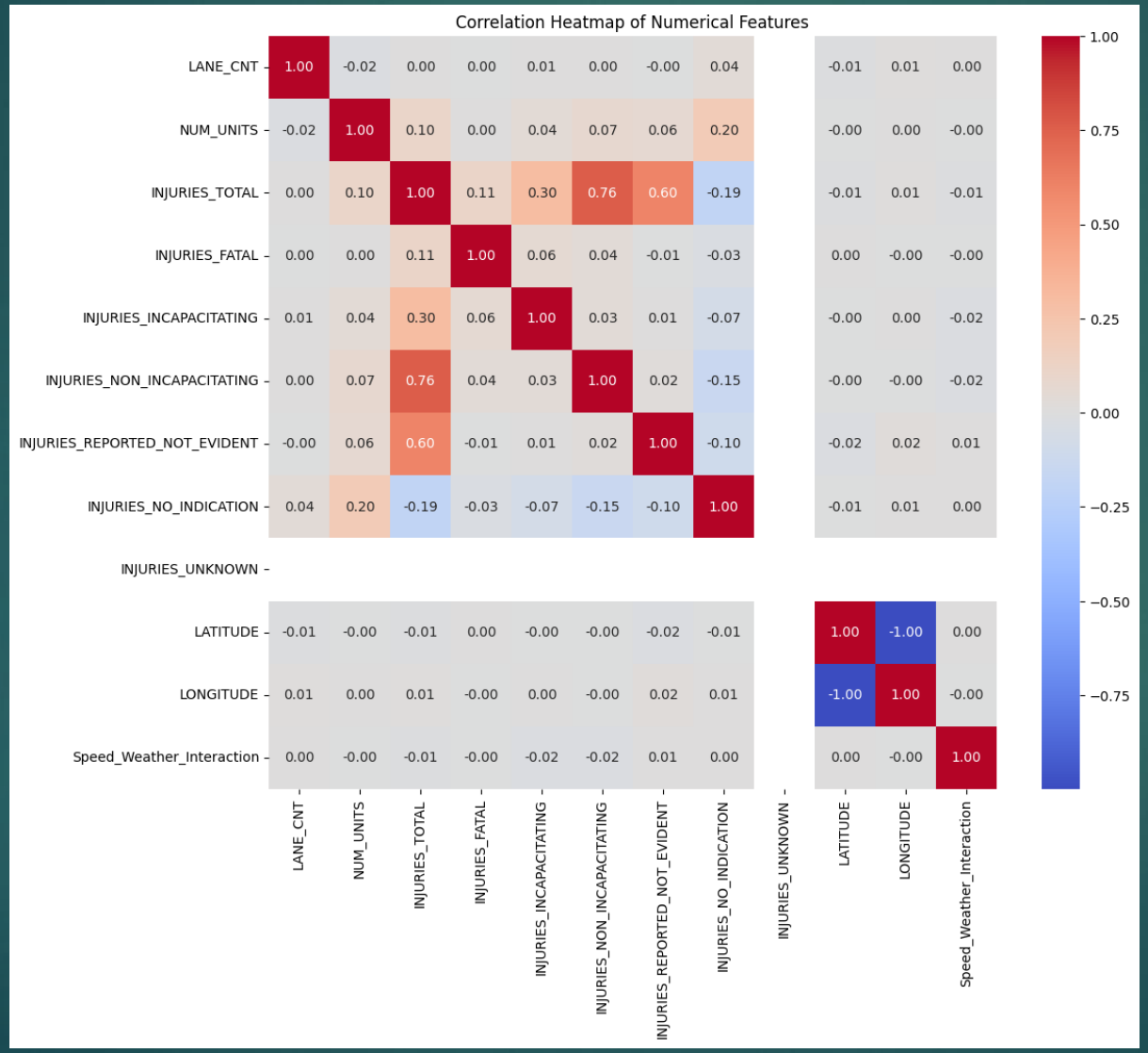
- **Dataset**: Chicago Traffic Crashes Dataset
- **Key Features -** Speed_Weather_Interaction, Is_Weekend, CRASH_HOUR_sin and CRASH_HOUR_cos, LANE_CNT, ROADWAY_SURFACE_COND, TRAFFIC_CONTROL_DEVICE, DEVICE_CONDITION, LIGHTING_CONDITION, FIRST_CRASH_TYPE.

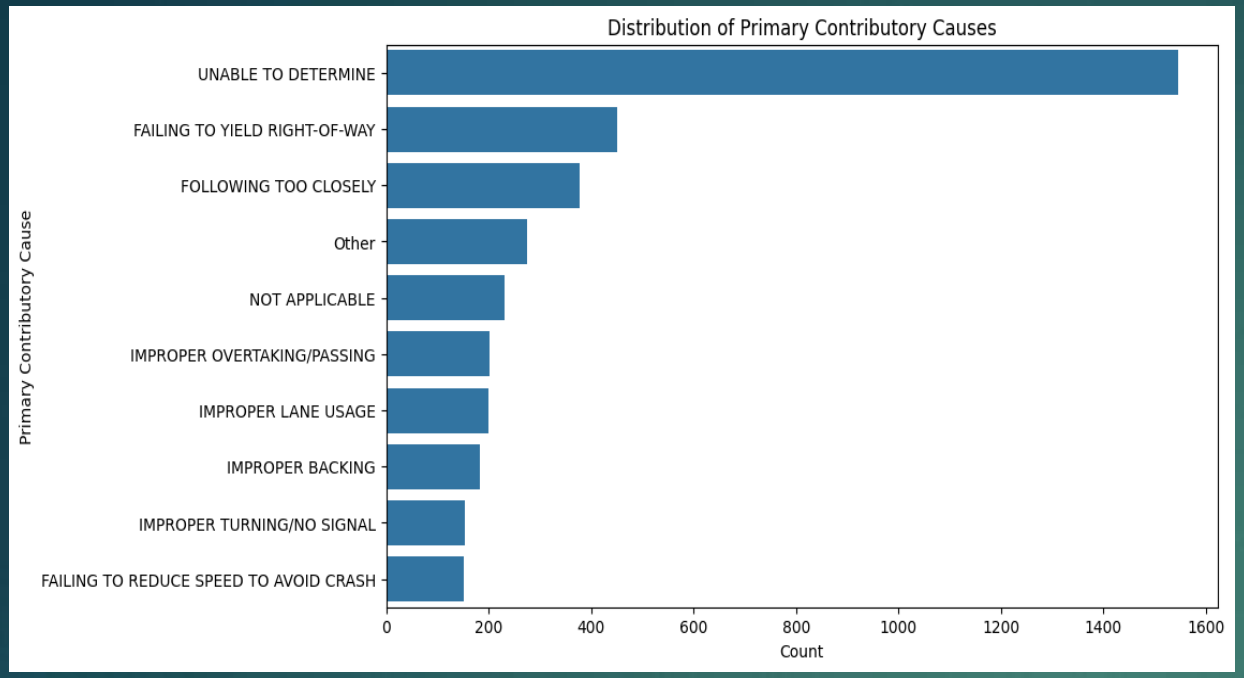
Sample Size: 48 columns

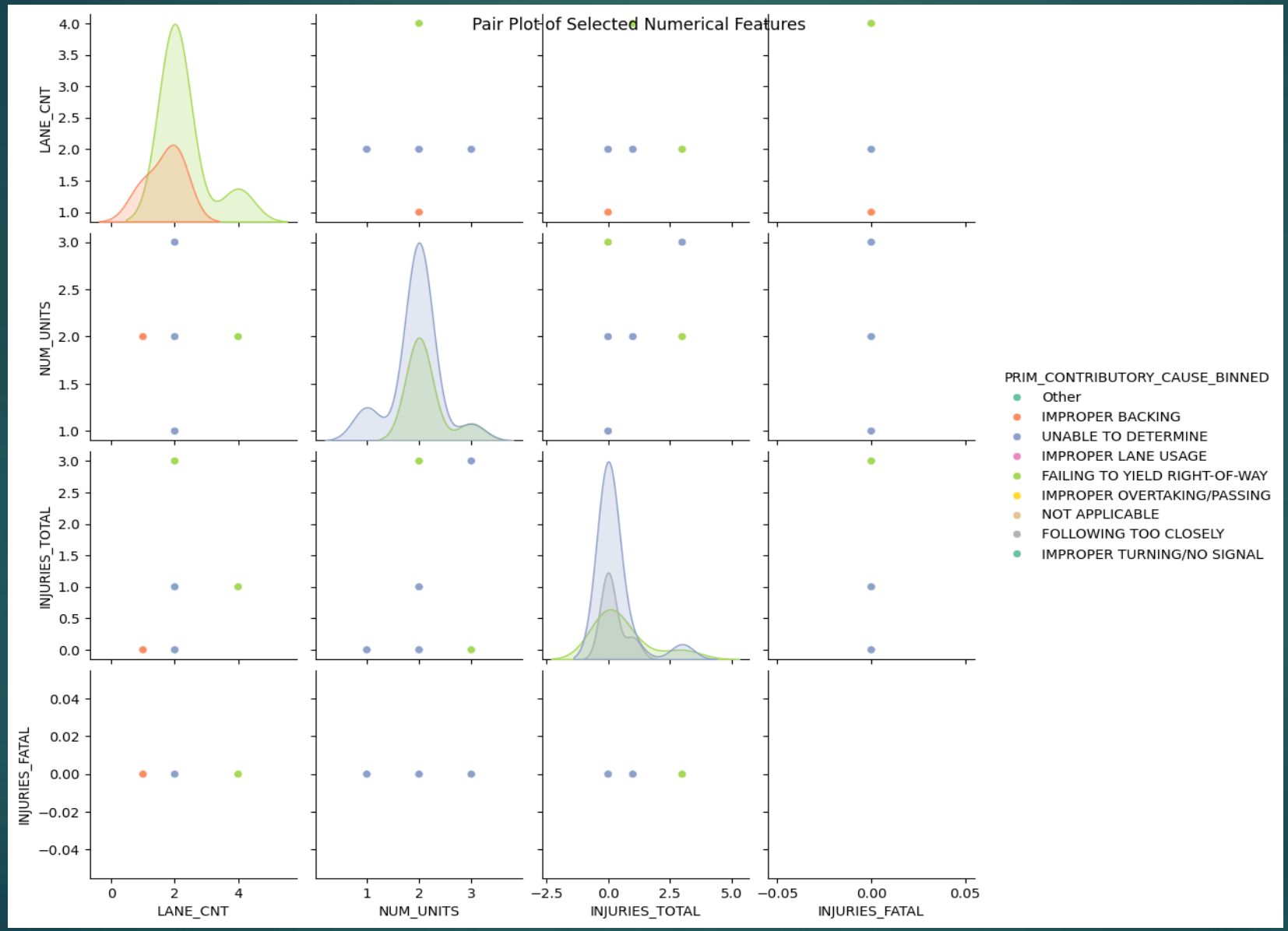
Data Preparation

- **Handling Missing Data:** Missing values were addressed using imputation strategies, filling in with mode values or dropping columns with excessive missing data.
- **Feature Engineering:** Cyclical features were created for time-related data, and interaction terms like speed-weather interactions were introduced to capture complex relationships.
- **Binning Rare Causes:** Rare contributory causes were binned into an "Other" category to simplify the target variable and improve model performance.

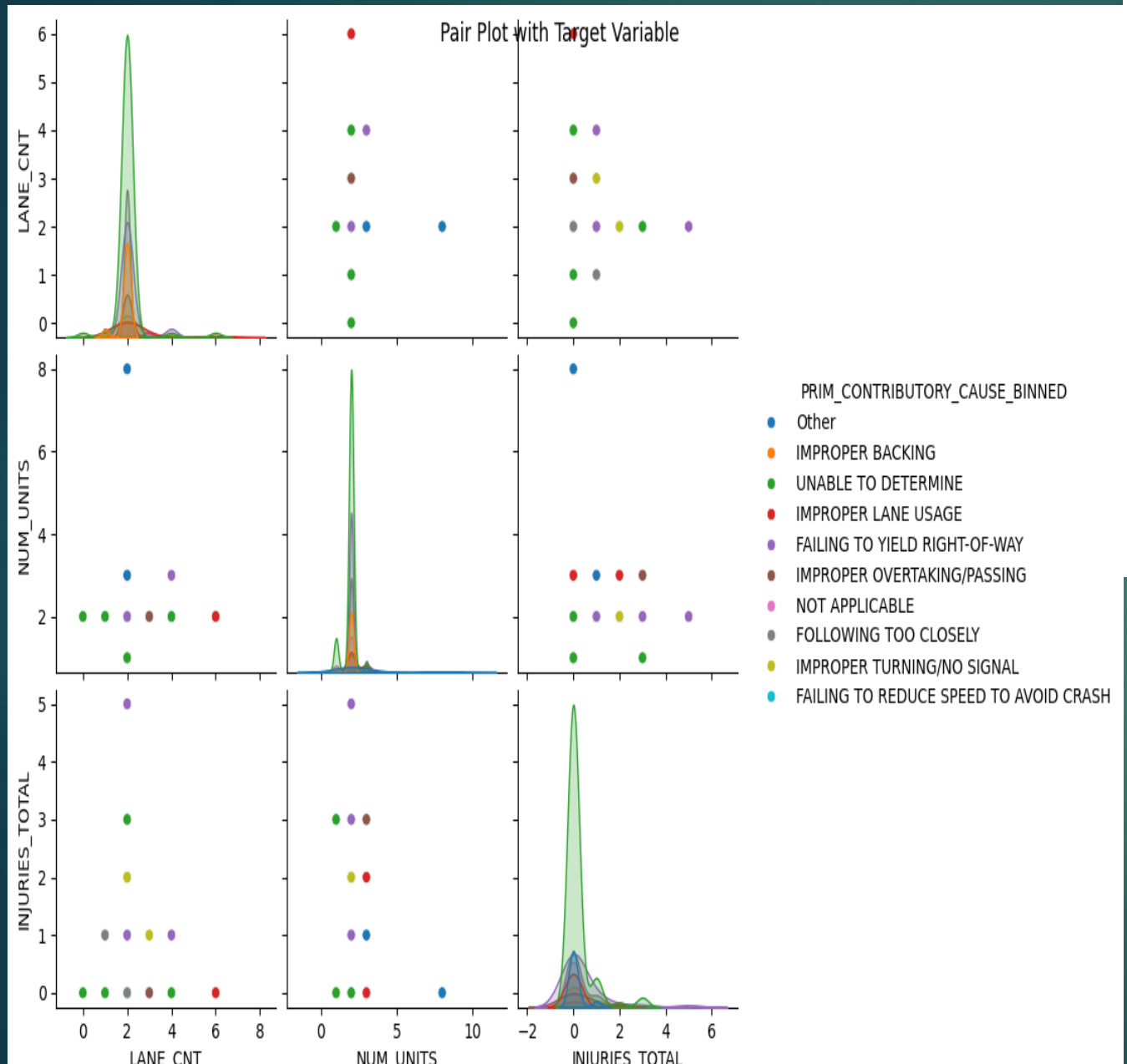
Exploratory Data Analysis (EDA)

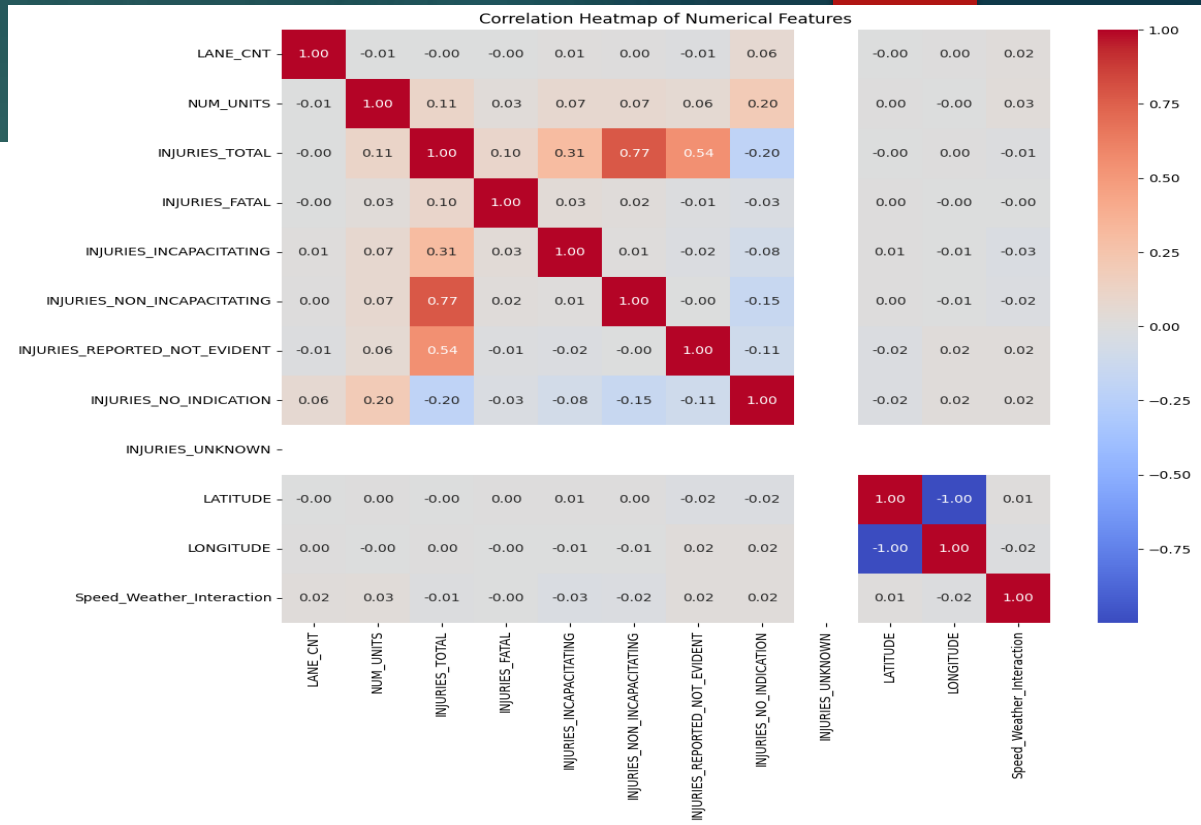


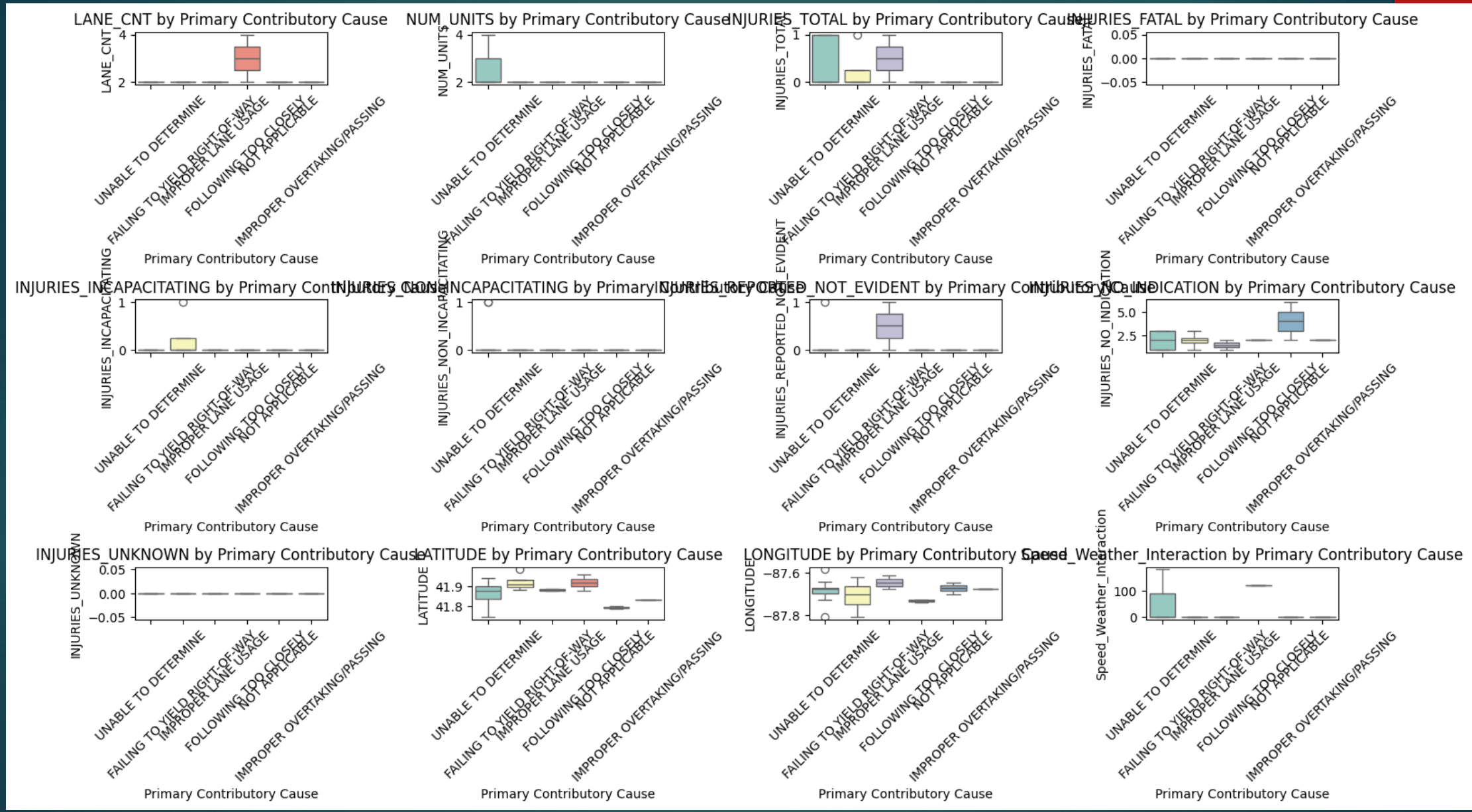

Distribution of Primary Contributory Causes
UNABLE TO DETERMINE
FAILING TO YIELD RIGHT-OF-WAY
FOLLOWING TOO CLOSELY
Other
NOT APPLICABLE
IMPROPER OVERTAKING/PASSING
IMPROPER LANE USAGE
IMPROPER BACKING
IMPROPER TURNING/NO SIGNAL
FAILING TO REDUCE SPEED TO AVOID CRASH
Primary Contributory Cause
0
200
400
600
800
1000
1200
1400
1600
Count


Pair Plot of Selected Numerical Features
LANE_CNT
NUM_UNITS
INJURIES_TOTAL
INJURIES_FATAL
PRIM_CONTRIBUTORY_CAUSE_BINNED
Other
IMPROPER BACKING
UNABLE TO DETERMINE
IMPROPER LANE USAGE
FAILING TO YIELD RIGHT-OF-WAY
IMPROPER OVERTAKING/PASSING
NOT APPLICABLE
FOLLOWING TOO CLOSELY
IMPROPER TURNING/NO SIGNAL

Multivariate analysis






LANE_CNT by Primary Contributory Cause
NUM_UNITS by Primary Contributory Cause
INJURIES_TOTAL by Primary Contributory Cause
INJURIES_FATAL by Primary Contributory Cause
INJURIES_INCAPACITATING by Primary Contributory Cause
INJURIES_NON_INCAPACITATING by Primary Contributory Cause
INJURIES_REPORTED_NOT_EVIDENT by Primary Contributory Cause
INJURIES_NO_INDICATION by Primary Contributory Cause
INJURIES_UNKNOWN by Primary Contributory Cause
LATITUDE by Primary Contributory Cause
LONGITUDE by Primary Contributory Cause
Speed_Weather_Interaction by Primary Contributory Cause
Primary Contributory Cause
UNABLE TO DETERMINE
FAILING TO YIELD RIGHT-OF-WAY
IMPROPER LANE USAGE
FOLLOWING TOO CLOSELY
NOT APPLICABLE
IMPROPER OVERTAKING/PASSING

Modeling Approach

- **Models Tested**:
- Models tested were: Logistic Regression, Ridge Classifier, Lasso Logistic Regression, Decision Tree, Random Forest, Gradient Boosting.

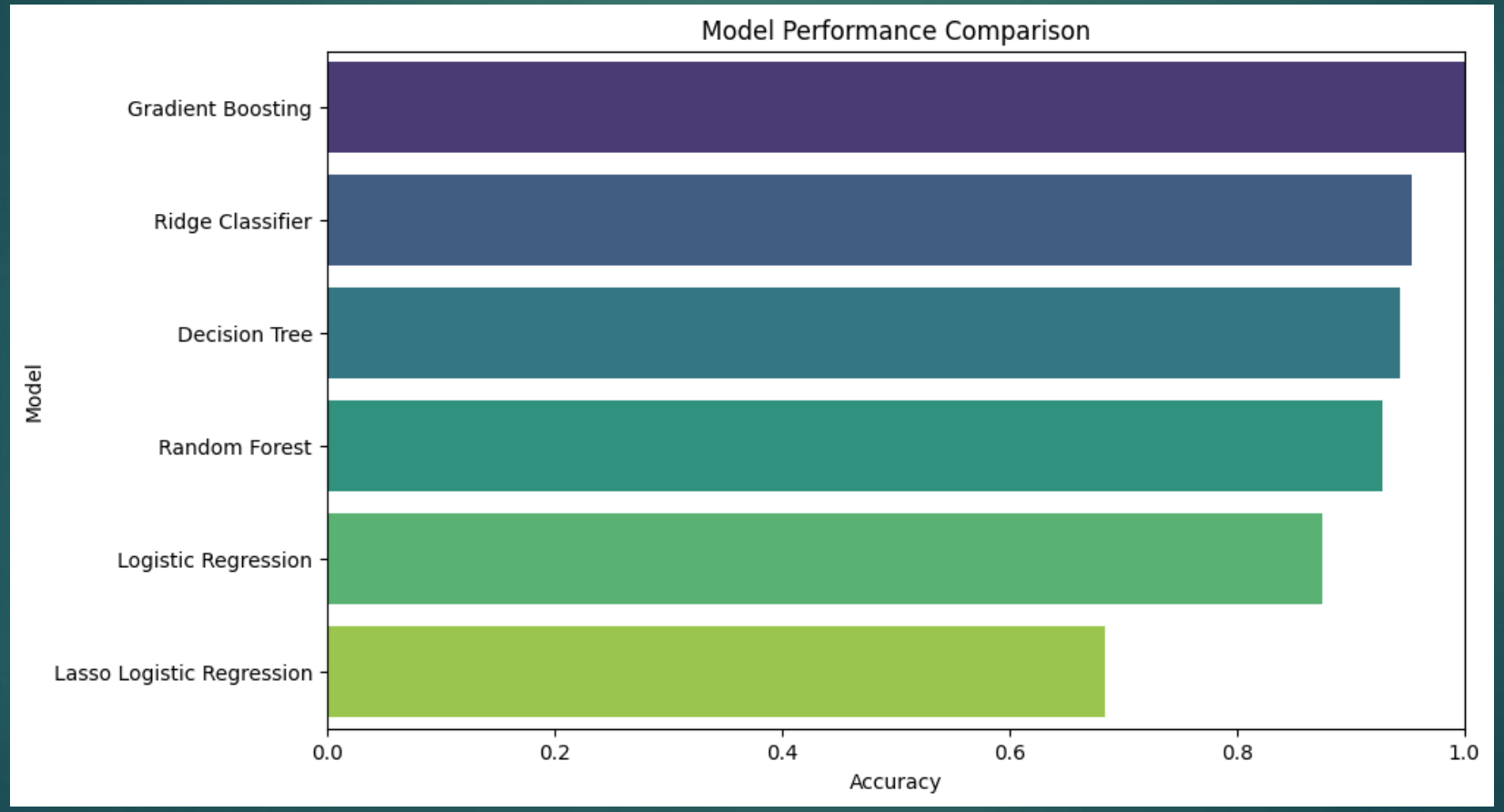
- **Evaluation Metrics**:
- - Accuracy: The proportion of correct predictions made by the model.
- - Precision: The ability of the model to correctly identify positive instances.
- - Recall: The ability of the model to capture all positive instances.
- - F1-Score: The harmonic mean of precision and recall, balancing the two.
- - Confusion Matrix: A detailed breakdown of the model's performance across different classes.

Model Comparison

- **Performance Overview**: The bar chart below represents the performance of each model.

- **Best Model**: Gradient Boosting in general became the best for this problem because of how large the dataset is.
- It combines the strengths of ensemble learning with the flexibility of Decision Trees, allowing it to model complex non-linear relationships, handle interactions between features and generalize well on unseen data. Its performance were further enhanced through Hyperparameter Tuning, which likely led to it outperforming the other models tested.

Models performance

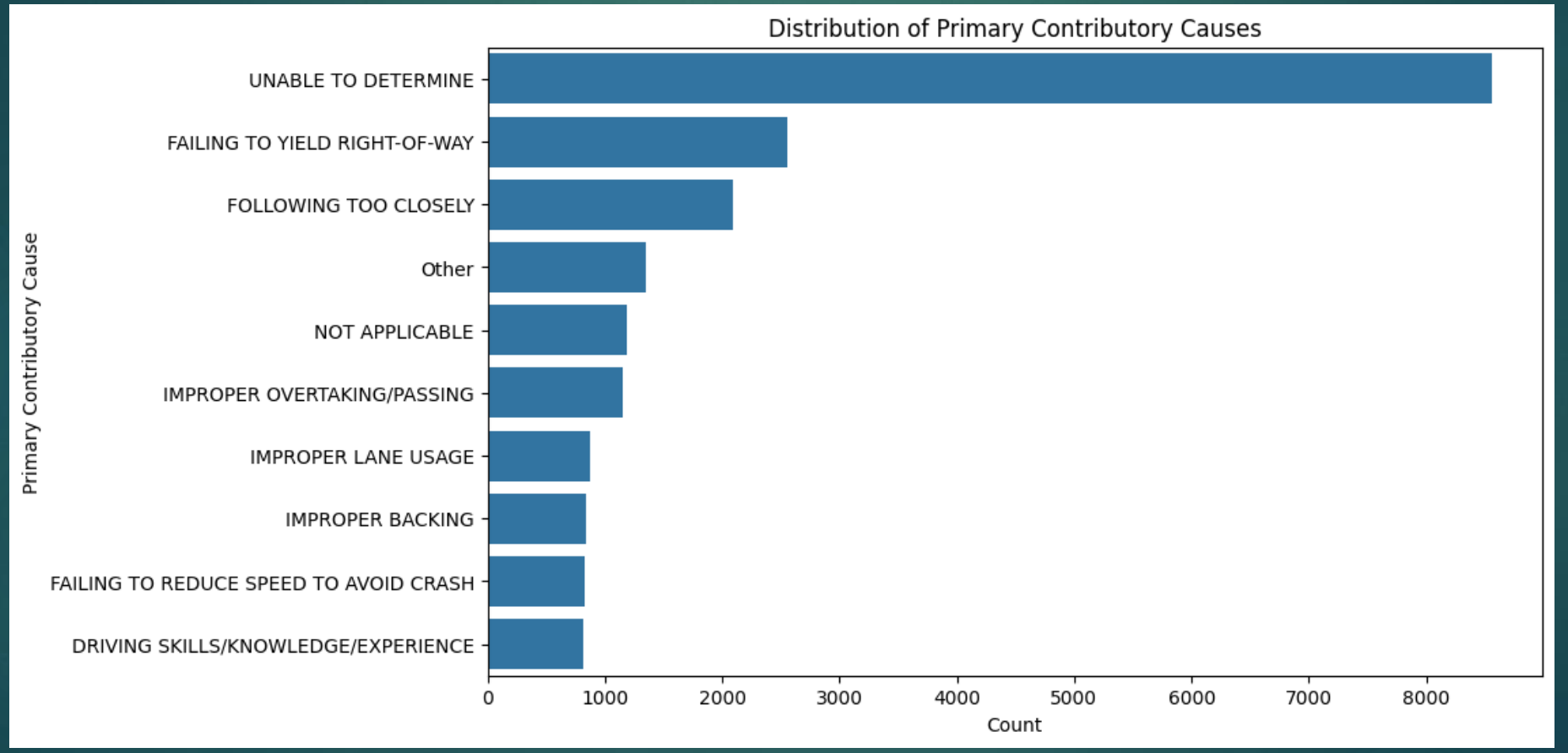


Top Features: Below is a bar chart showing the top 10 features by importance.

Interpretation:

- **High uncertainty:** Many accidents have unclear causes ("UNABLE TO DETERMINE").
- **Right-of-way issues:** Failing to yield is a major cause.
- **Following too closely:** Leads to frequent rear-end collisions.
- **Improper driving:** Unsafe behaviors like improper lane usage and overtaking are common.
- **Driver skills:** Lack of experience and knowledge also contribute to accidents.

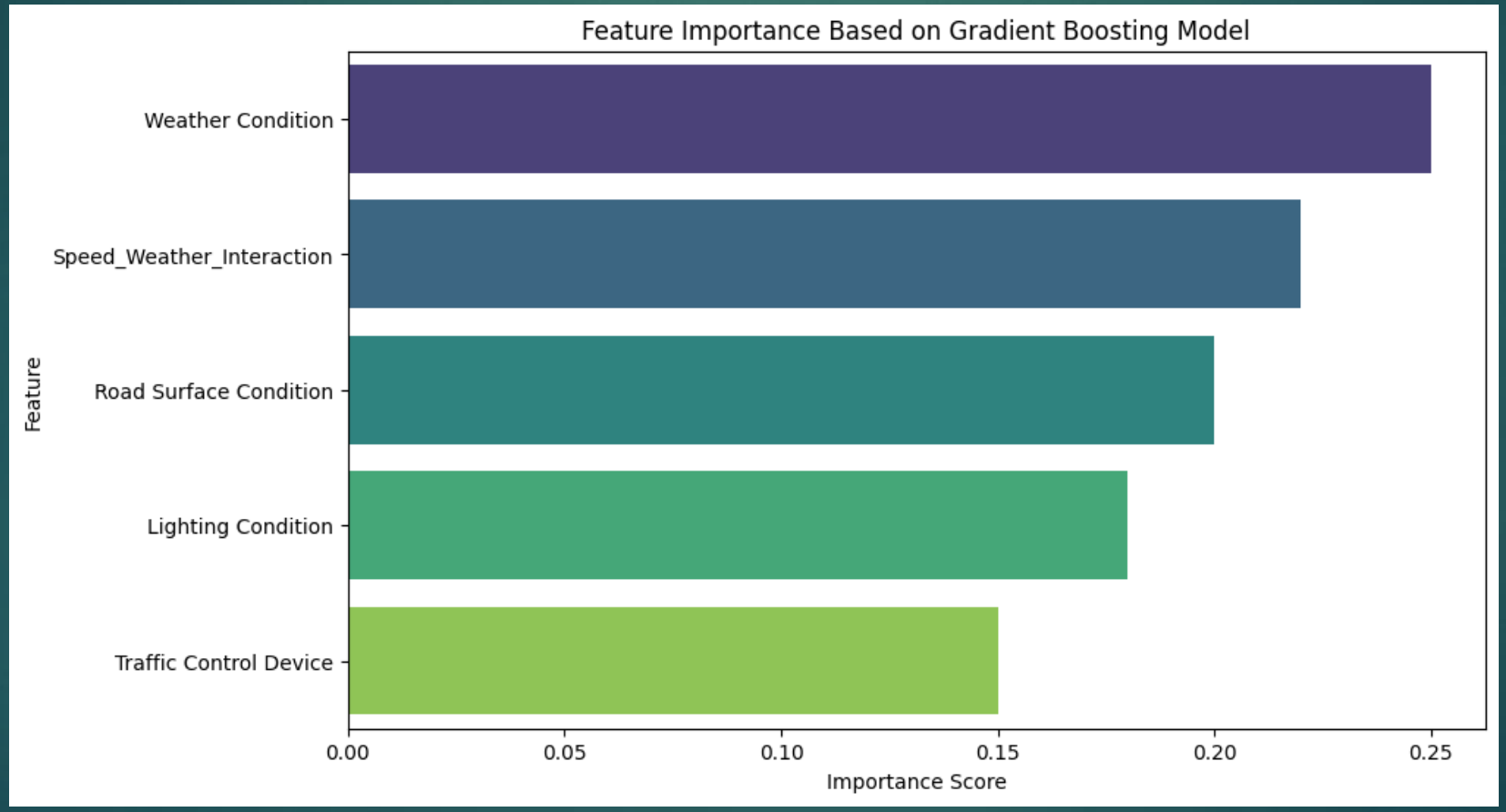
Top Features by Importance



Insights

- **Summary of Key Insights**: Key features that heavily influenced the likelihood of an accident included weather conditions, road surface conditions, lighting conditions, and the type of traffic control devices.
- The interaction between speed and weather conditions was particularly notable, indicating that accidents are more likely when adverse weather coincides with higher speeds.
- **Visuals**: Below are visuals that supports our insights.

Feature importance bar chart based on gradient boosting outputs model



Recommendations

- Targeted Safety Campaigns
- Improved Road Infrastructure
- Adaptive Speed Limits
- Enhanced Law Enforcement
- Further Research

Conclusion

- **Final Thoughts**: The analysis of the traffic crash data has provided valuable insights into the primary contributory causes of accidents and the critical factors influencing road safety. By identifying the top features and causes associated with accidents, we can make informed recommendations to enhance road safety in the City of Chicago

- By implementing these recommendations, the City of Chicago can reduce traffic accidents, enhance road safety, and ultimately save lives. The insights gained from this analysis provide a clear pathway to making the city's roads safer for everyone.

Next Steps

- Consider deploying the best-performing model in a real-time traffic monitoring system to provide proactive alerts and reduce accident rates.

Q&A