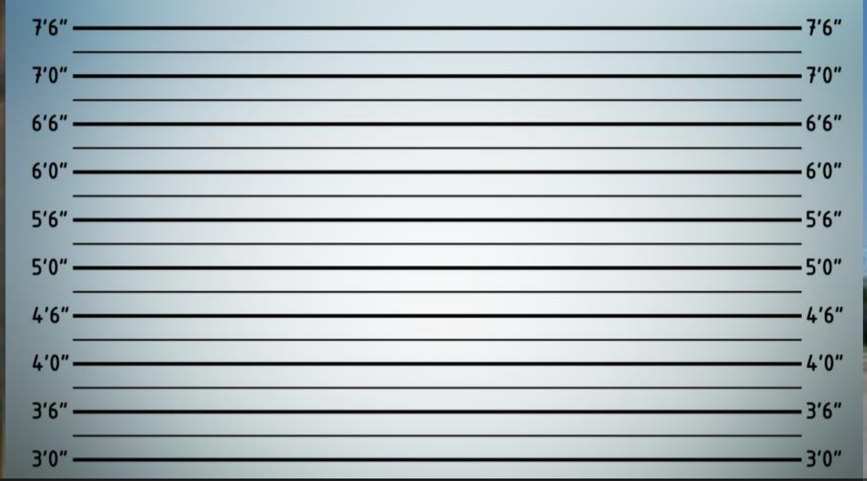


Terry Stops Analysis: Predicting Arrest Outcomes

A Data-Driven Analysis of Seattle Police Department Interventions

BUSINESS UNDERSTANDING

Objective

Analyze Terry Stops data to predict whether a stop will result in a formal arrest.

Goal

Uncover key drivers of police decision-making (e.g., time, location, demographics, weapons).

Value Proposition

- **Operational:** Insights into patterns and resource allocation.
- **Policy:** Evaluate transparency and consistency in interventions.
- **Impact:** Move beyond simple accuracy to identify actual arrest drivers in highly imbalanced data.

DATA UNDERSTANDING

Dataset

65,931 records from Seattle Police Department

Key Feature Groups

Subject Demographics: Age, Race, Gender.

Officer Demographics: ID, Gender, Race, Year of Birth.

Event Context: Date, Precinct, Sector, Beat.

Operational Factors: Call Type, Frisk Flag, Weapon Type

Target Variable

Arrest Flag (Y/N).

DATA QUALITY & CHALLENGES

Class Imbalance

No Arrest: ~89% (58,368 stops)

Arrest: ~11% (7,563 stops)

Challenge:
A "lazy" model predicting "No Arrest" every time would still be 89% accurate.

Data Issues

Missing Values: Imputed Officer Squad and Weapon Type.

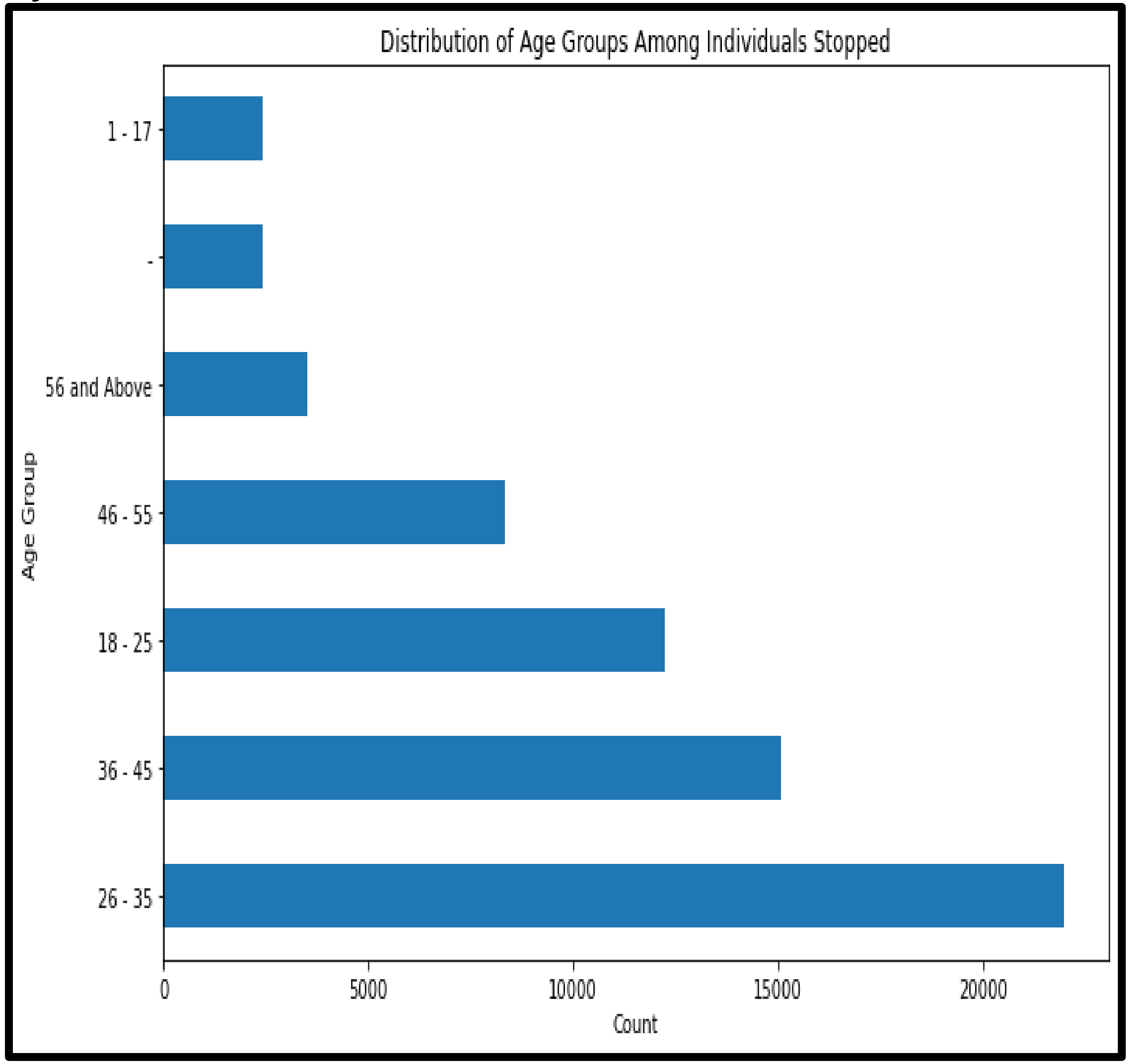
Placeholders: Standardized messy entries like '-', 'None/Not Applicable', and 'Unknown'.

Cardinality: Call Type and Sector have many unique values requiring special encoding.

EXPLORATORY DATA ANALYSIS (EDA)

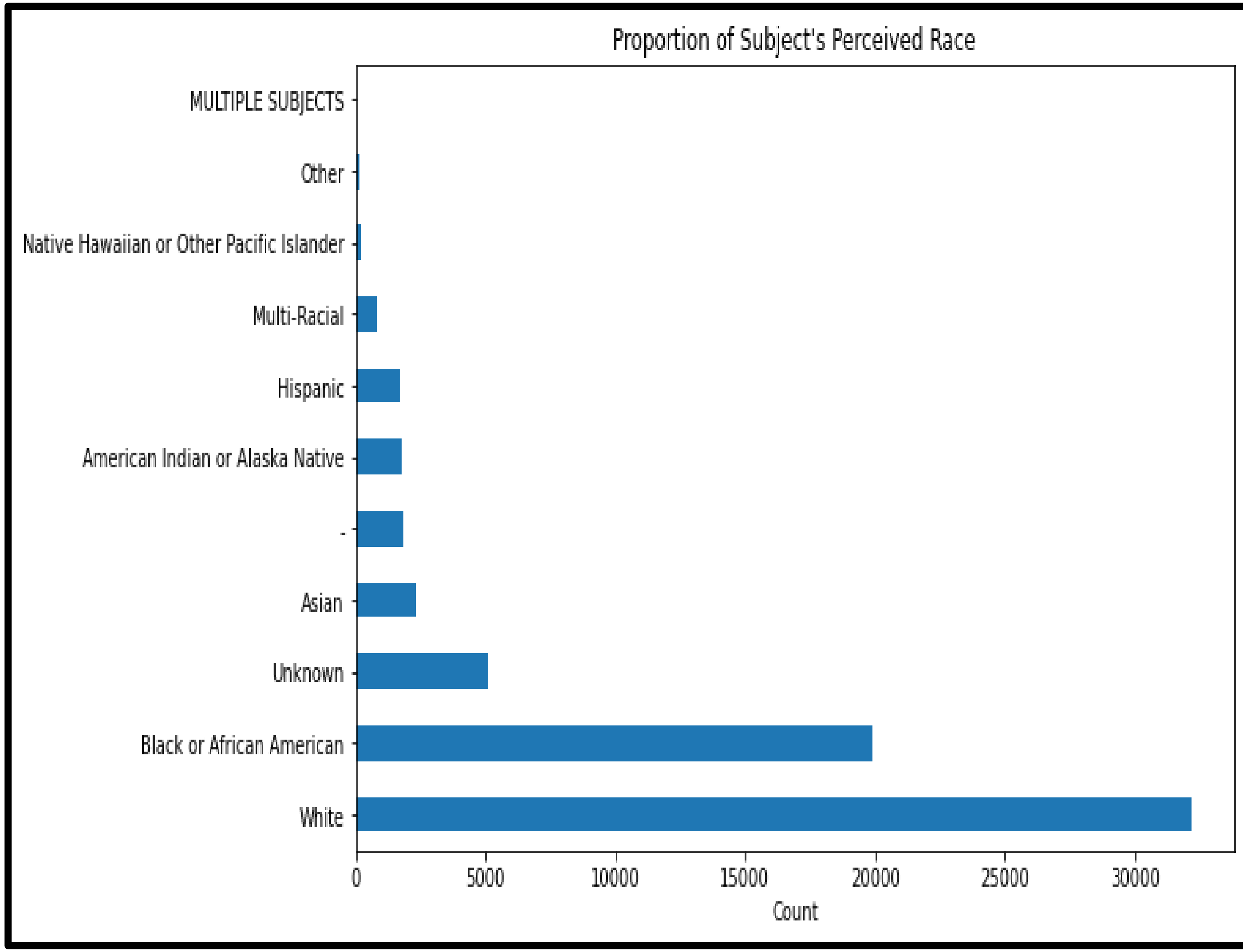
Subject Age

Most frequently stopped group: 26–35, followed by 36–45.



Subject Race:

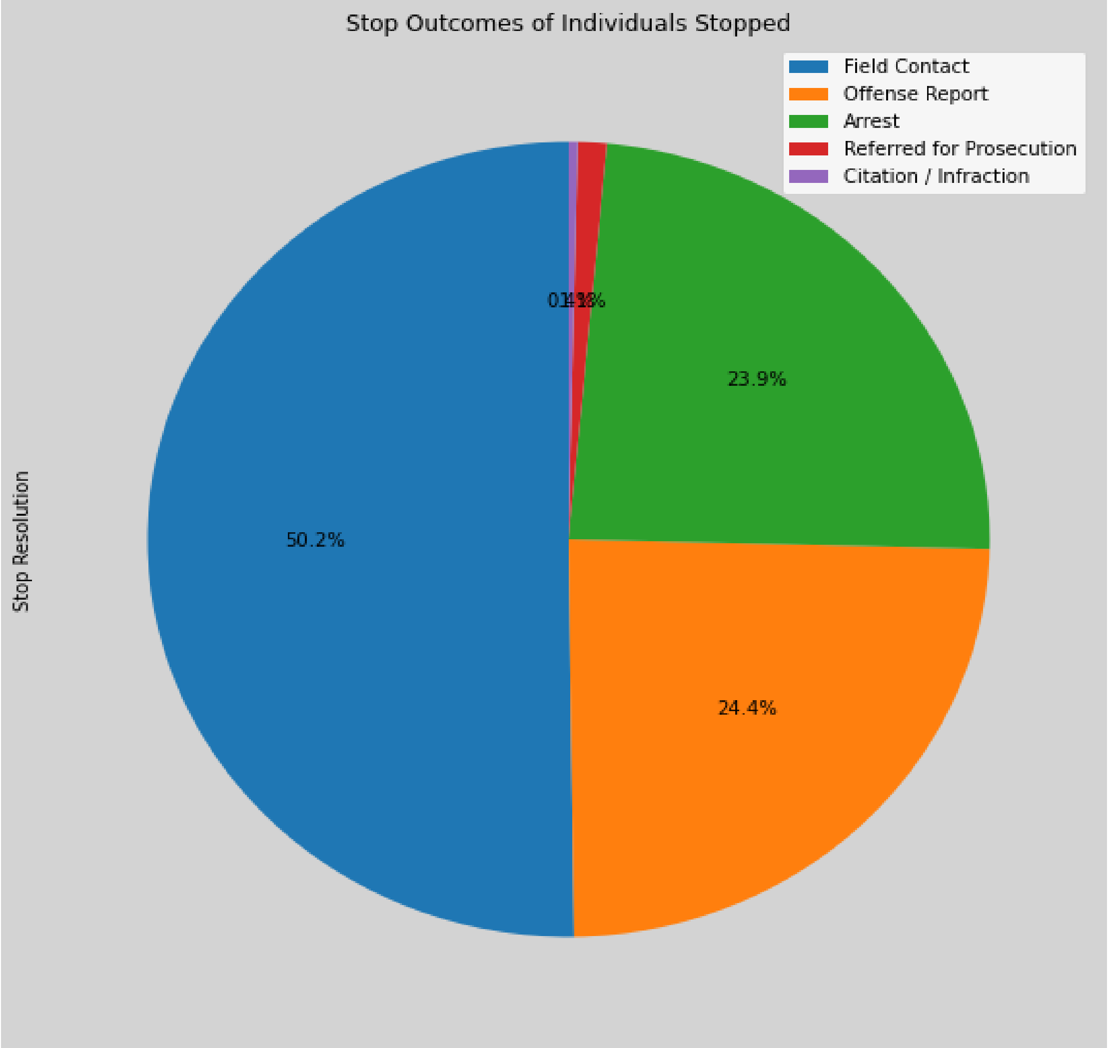
Majority of subjects perceived as White or Black/African American.



EXPLORATORY DATA ANALYSIS (EDA)

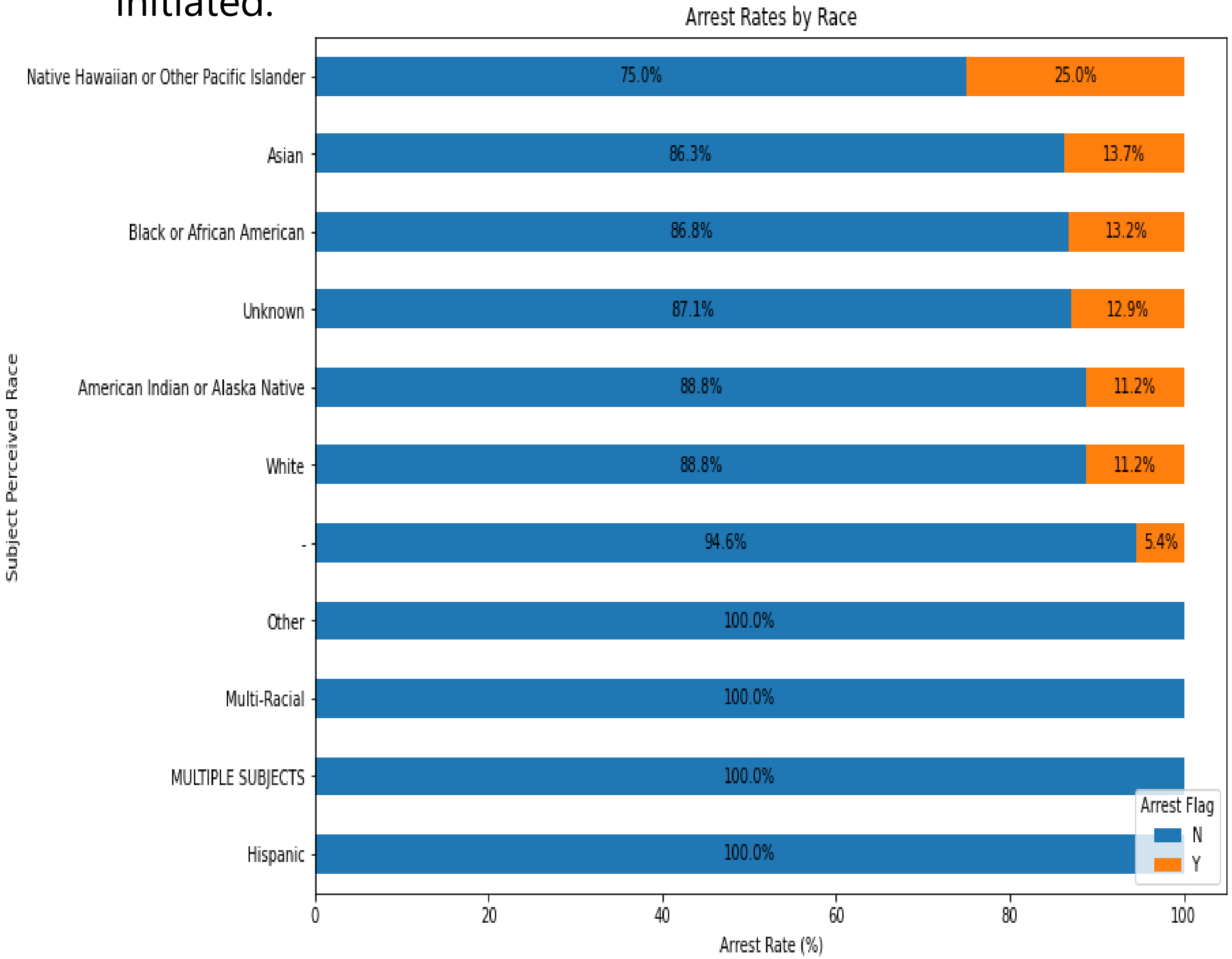
Stop Resolutions

Most common: "Field Contact" (No booking).
Followed by: "Offense Report" and "Arrest".



Arrest Rates by Race:

Visualized potential disparities in outcomes once a stop is initiated.



FEATURE ENGINEERING

Time & Date

Buckets
Morning, Afternoon, Evening, Night (to capture risk profiles).

Cyclical Encoding
Transformed Hour/Month/Day into Sine/Cosine waves (preserving time continuity, e.g., 23:00 is close to 00:00).

Demographic Interactions

Age Gap
Calculated Age Difference (Officer Age - Subject Age).

Bias Flags
Created Same_Race, Same_Gender, and Minority_Interaction (White Officer stops Non-White Subject) features.

FEATURE SELECTION & PREPROCESSING

Dropped Columns

Redundant IDs and raw date fields.

Encoding Strategy

Ordinal Encoding
Subject Age Group.

One-Hot Encoding
Gender, Race, Precinct, Time of Day.

Target Encoding
High-cardinality features like Initial Call Type, Sector, and Weapon Type.

Scaling

Applied Standard Scaler to the numerical features (Officer Age, Year, Age_Difference) to normalize ranges.

BASELINE MODEL PERFORMANCE

Model

Dummy Classifier (Strategy: Most Frequent).

Results

Accuracy: ~89% (Matches the majority class).

Recall (Arrests): 0.00% (Identifies zero arrests).

Verdict
High accuracy is deceptive.

The model provides no value for public safety or policy analysis since it doesn’t identify any arrests.

LINEAR MODEL - LOGISTIC REGRESSION

Improvement

Active learning of relationships.

Performance

Recall (Arrests): 77% (Huge improvement over baseline)

ROC-AUC: 0.82 (Strong predictive power)

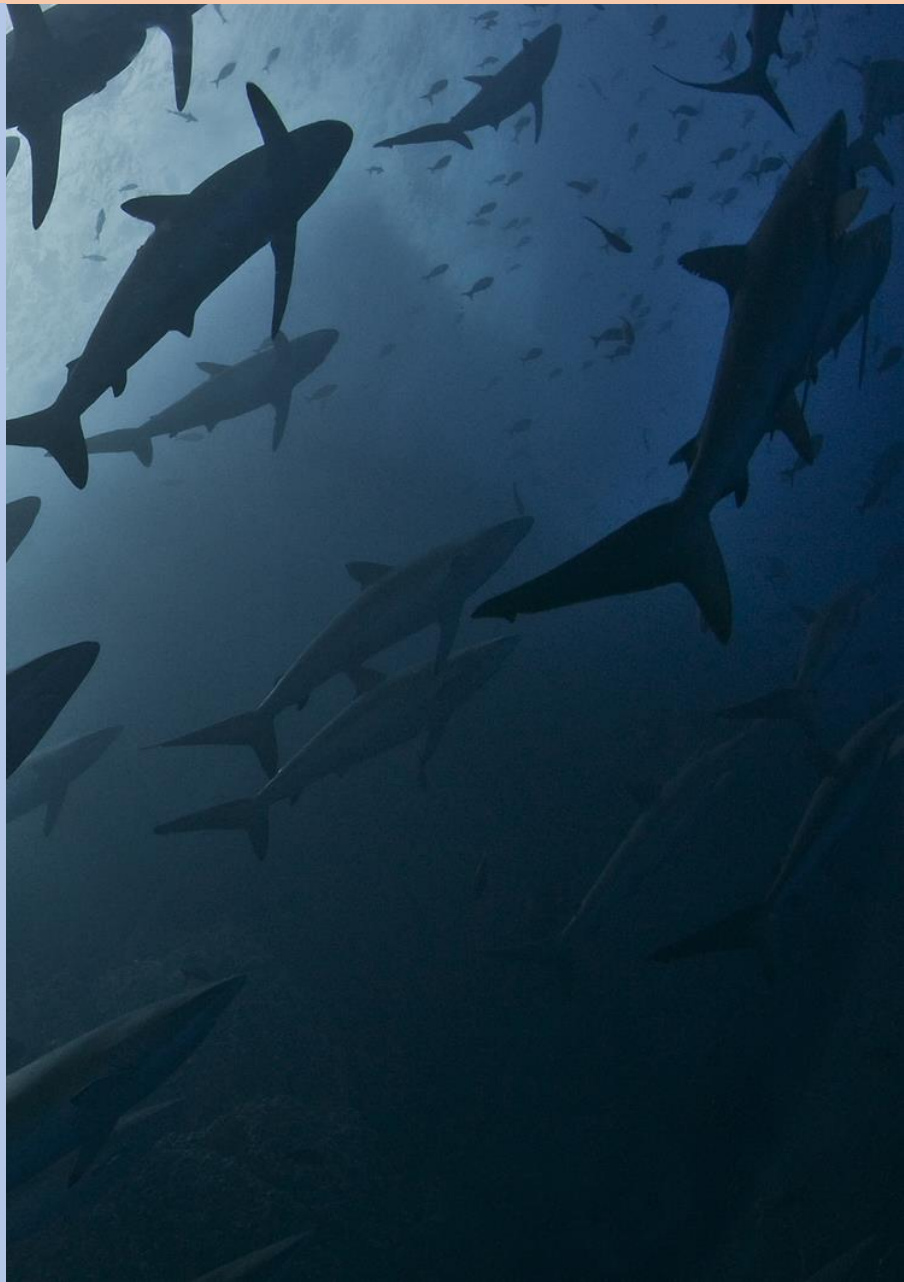
Key Drivers Identified

Initial Call Type (Dominant factor).

Sector and Weapon Type are strong secondary predictors.

Insight

Situational factors outweigh demographics in this linear model.



TREE-BASED MODELS

1. Random Forest

ROC-AUC: 0.868.

Recall: 86%.

Top Feature: Year (Suggests policy/temporal shifts are critical).

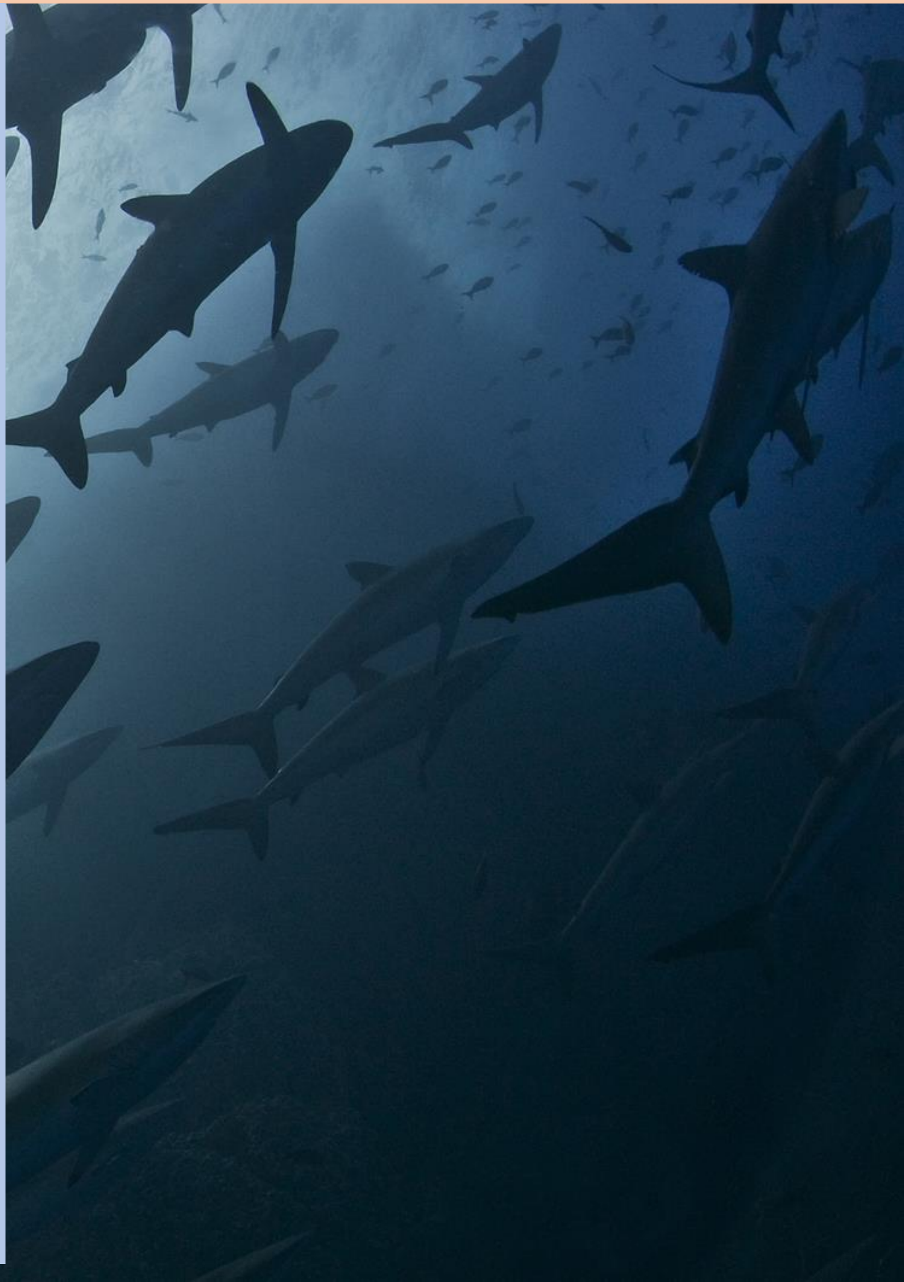
2. Gradient Boosting

Recall: Only 4% (Severe overfitting to majority class). It didn’t weight the classes.

3. AdaBoost

Recall: 89% (Good, but prone to overfitting)

ROC-AUC: 0.82 (Strong predictive power)



CHAMPION MODEL - XGBOOST

Configuration

Extreme Gradient Boosting with scale_pos_weight to handle class imbalance.

Performance

- **ROC-AUC: 0.878** (Highest discriminative ability for the classes).
- **Recall: 89%** (Correctly identifies nearly 9 out of 10 arrests).
- **F1-Score:** 0.45 (Balanced given that my dataset had extreme imbalance).

Why it Wins

Best balance of distinguishing power and sensitivity to the minority class (Arrests).

CONCLUSION

KEY FINDINGS

•Arrest outcomes are heavily influenced by situational context (Call Type, Sector) rather than just Officer or Subject demographics.

•Time(Year)is a critical predictor, implying evolving policing policies.

RECOMMENDATIONS

- Deploy the XGBoost model to flag high-probability arrest scenarios for review.
- Investigate the specific years that drive predictions to understand policy impacts.
- Monitor the Minority_Interaction flag in future data to track bias trends over time.

Name: Brian Chege
GitHub: https://github.com/BC-254
LinkedIn: www.linkedin.com/in/bchege