UNMASKING BIAS: LEVERAGING DATA ANALYTICS TO DRIVE EQUITY IN N.Y CRIMINAL JUSTICE SYSTEM

Utilizing Machine Learning to Identify and Mitigate Systemic Disparities

TABLE OF CONTENTS

EXECUTIVE SUMMARY

EXECUTIVE SUMMARY

Utilizing Machine Learning to Identify and Mitigate Systemic Disparities:

- **The Problem**: Systemic biases in the New York criminal justice system perpetuate racial, geographic, and socio-economic disparities.

- **The Approach**: Leverage advanced data analytics and machine learning to identify patterns of bias.

- **Key Findings**: **age** is the most crucial factor, **economic factors** and **race** influence police behavior, **self-initiated stops** by agents are more likely to lead to unjustified actions; **potential systemic biases in law enforcement practices.**

- **The Impact:** Data-driven insights to guide reforms in law enforcement practices and reduce disparities.

- **The Recommendations**: real need for **age-sensitive training**, implement **community-based policing** strategies, revise **protocols for self-initiated stops**, audit regularly to **increase accountability** and identify officers with **patterns**.

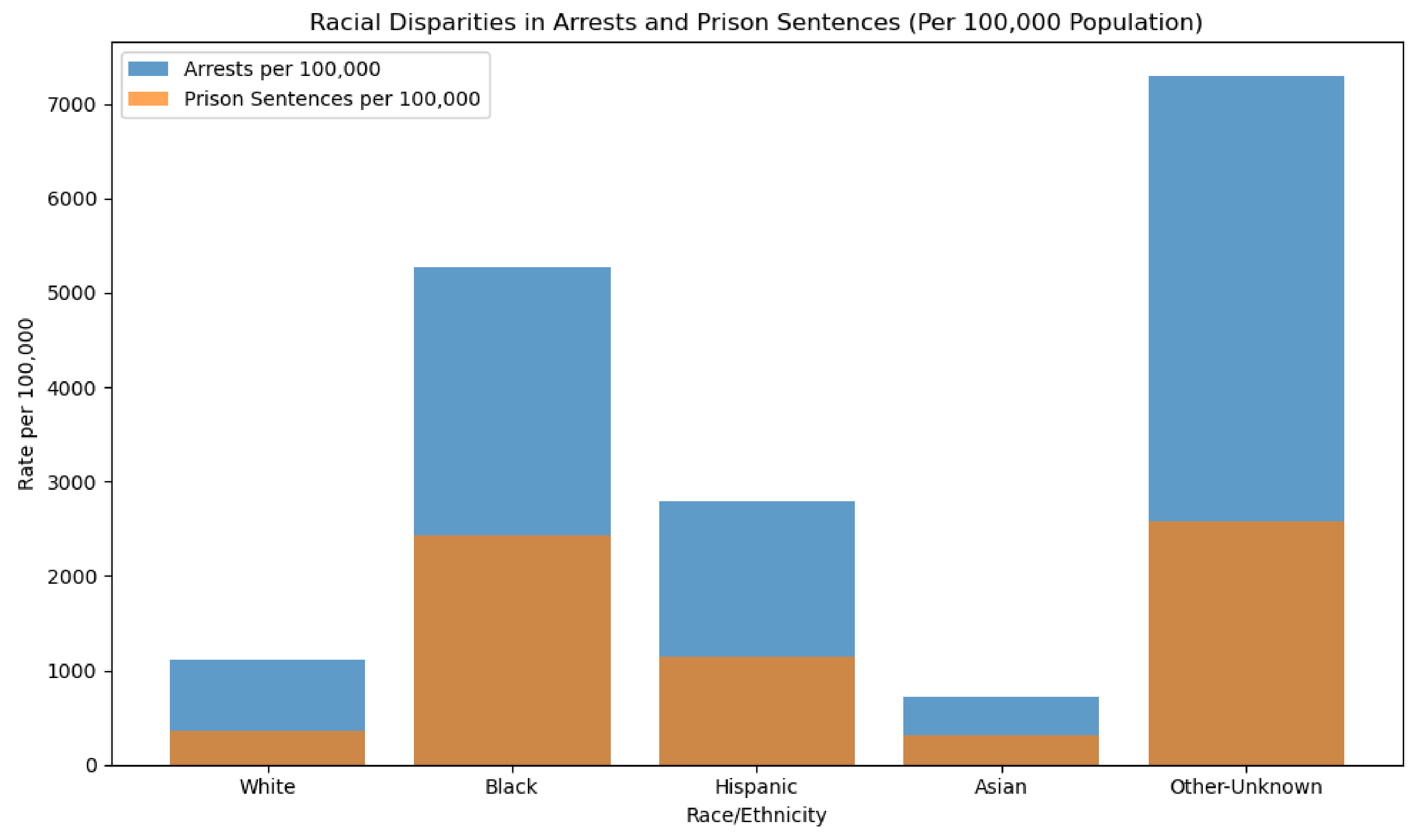
THE PROBLEM

THE CHALLENGE: BIAS IN THE NEW YORK CRIMINAL JUSTICE SYSTEM

"Systemic biases in the New York criminal justice system perpetuate racial, geographic, and socio-economic disparities, demanding innovative, data-driven solutions to foster equity and fairness in law enforcement practices."

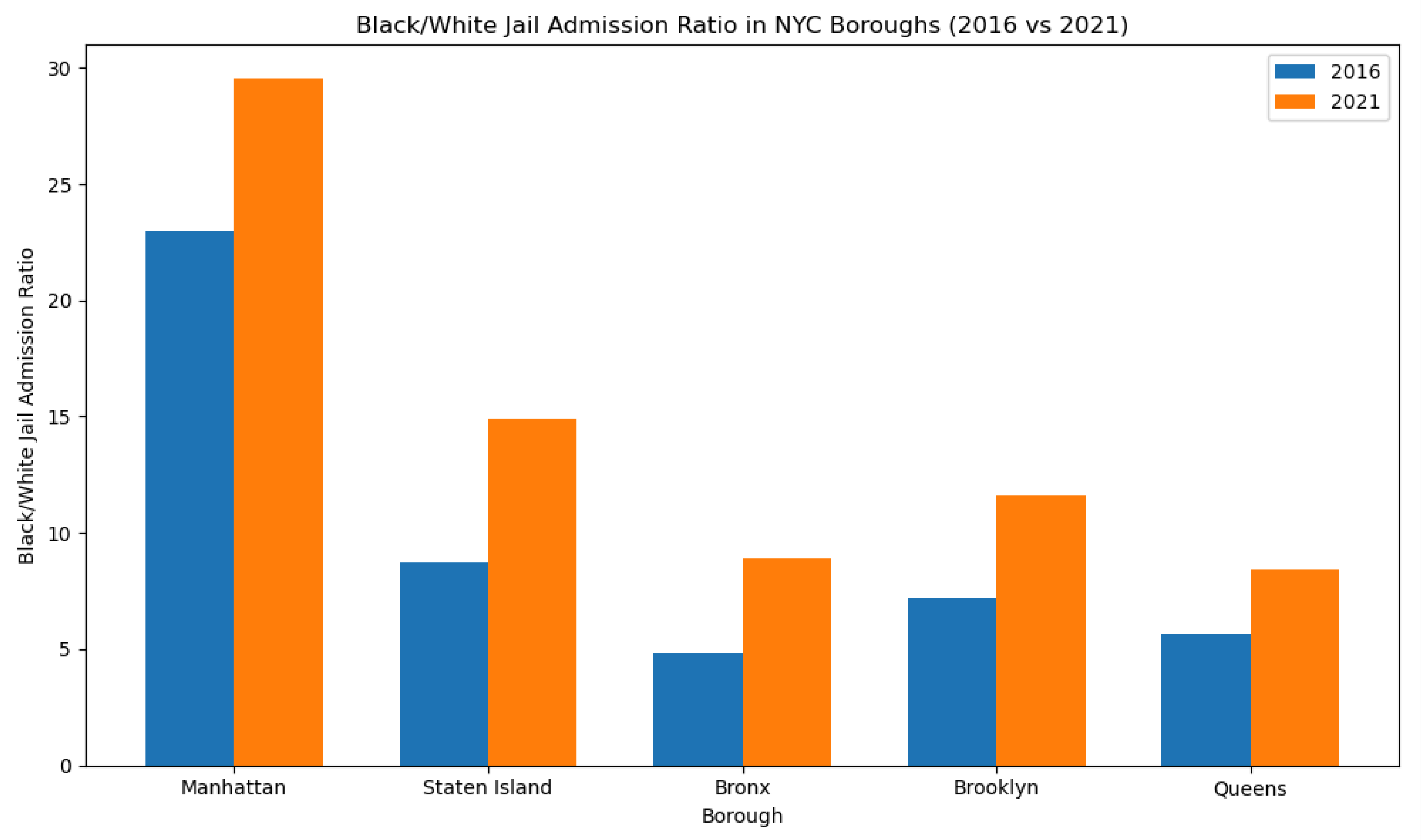
INTRODUCTION

Current Analysis - Racial Disparities in Arrests and Prison Sentences



Black individuals face the highest arrest and prison sentence rates, followed by Hispanics with notable disparities, while White individuals, despite being the majority, have significantly lower rates.

Current Analysis - Geographic Disparities in Incarceration Rates

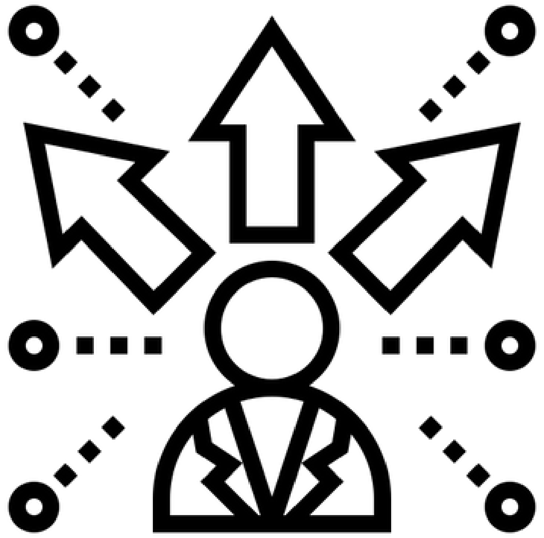


Black individuals in New York City face disproportionately high incarceration rates across all boroughs, particularly in Brooklyn and Staten Island, compounded by neighborhood-level disparities in high-poverty areas and pretrial incarceration due to inability to afford bail, highlighting systemic racial and geographic inequalities.

THE CONTEXT

Objective:

Address systemic bias in the N.Y. criminal justice system using advanced data analytics and machine learning.



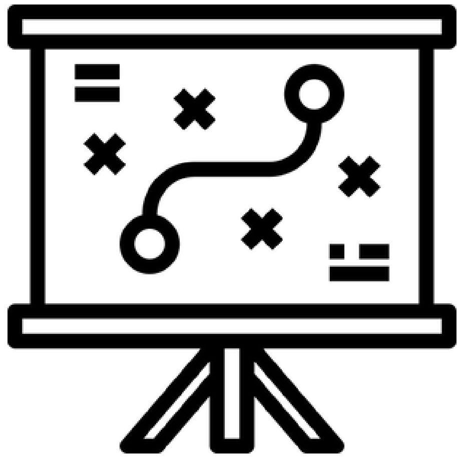
Opportunity:

Analyze vast datasets (arrest records, demographics) to uncover trends and systemic disparities based on race, age, gender, and geography.

THE CONTEXT

Goal:

Provide data-driven insights to reduce racial, geographic, and socio-economic disparities in arrests and guide reforms in law enforcement practices.



Approach:

Responding to growing public scrutiny by promoting fairness, equity, and transparency in the judicial system.

Broader Context:

Leverage predictive models to identify biases and recommend informed policy changes.

PRIMARY TARGETS FOR CHANGE

Policy-makers and Government Officials

Responsible for enacting laws and making decisions on resource allocation in the criminal justice system. They could use the findings of this project to drive legislative changes that promote fairer treatment across different demographics and regions.

Law Enforcement Agencies

Police departments, sheriffs, and other local law enforcement agencies could use the predictive models to re-evaluate arrest procedures and potentially adjust their training or resource allocation to minimize biased practices.

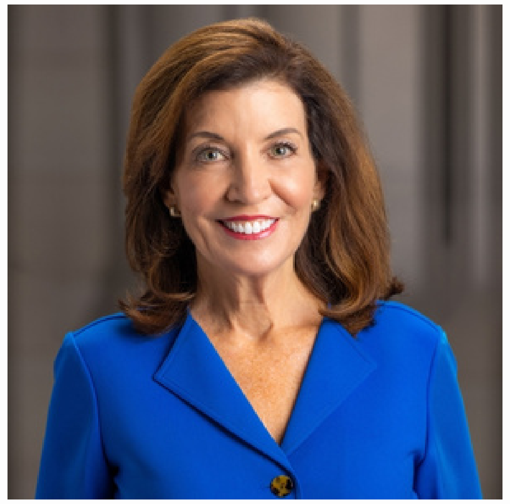
Criminal Justice Reform Advocates

Non-profit organizations and advocacy groups such as the ACLU or The Sentencing Project that work on criminal justice reform would be key stakeholders. They could use this data to highlight systemic problems and push for change within the judicial system.

KEY STAKEHOLDERS IN THE PROCESS

Federal, State, and Local Government Bodies

Agencies that manage law enforcement and policy creation are the key stakeholders, as they can directly implement changes based on the project's findings.



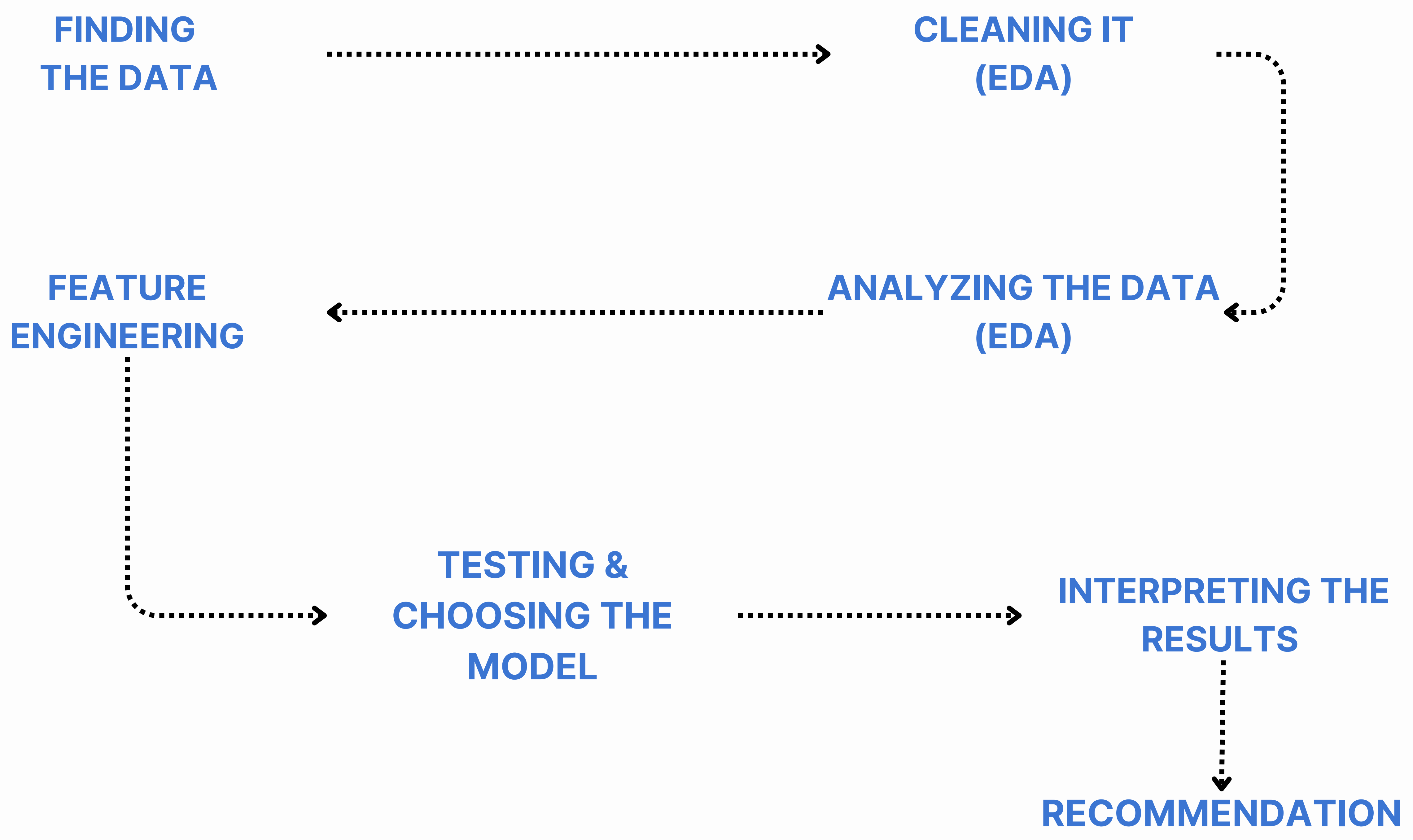
Community Leaders

Those representing populations disproportionately affected by biased arrests, such as minority and marginalized communities, would be essential stakeholders in ensuring that the findings lead to actionable reforms.

Technology and Data Vendors

If the model proves successful, vendors offering tools to manage and implement predictive analytics in law enforcement agencies may also be stakeholders.

THE PROCESS


FINDING THE DATA
CLEANING IT (EDA)
ANALYZING THE DATA (EDA)
FEATURE ENGINEERING
TESTING & CHOOSING THE MODEL
INTERPRETING THE RESULTS
RECOMMENDATION

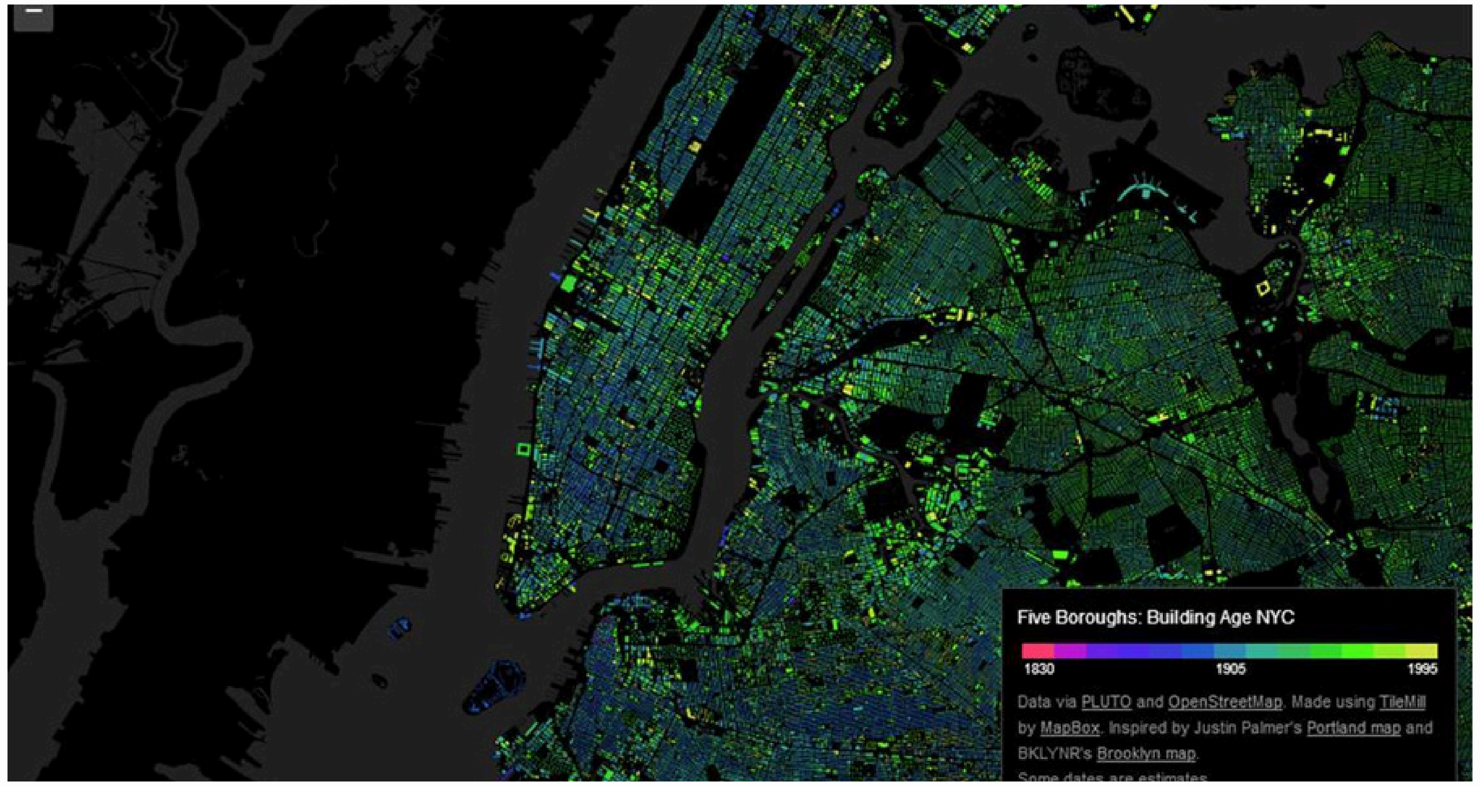
DESCRIPTION OF THE DATASET

01 **Stop, Question and Frisk Data.**
Last Updated: year 2022
Data records from the NYPD Stop.

https://www.nyc.gov/site/nypd/stats/reports-analysis/stopfrisk.page

02 **Neighborhood Financial Health Digital Mapping and Data.**
Last Updated: June 27, 2022
Data Provided By the Department of Consumer and Worker Protection (DCWP).

https://data.cityofnewyork.us/Business/Neighborhood-Financial-Health-Digital-Mapping-and-/r3dx-pew9/about_data



1- STOP, QUESTION AND FRISK DATA.

‘The New York police department release it’s stop-and-frisk incident recorded by the NYPD. The columns provide detailed information about various aspects of these incidents.”

Incident Details

Date, time, location (precinct, sector, address, etc.), duration, reason for the stop (radio run, self-initiated, etc.)

Officer and Supervisor Information

Rank, command code, and whether the supervising officer reviewed the activity log entry.

Suspect Demographics

Age, sex, race, height, weight, build, eye color, hair color, and other identifying descriptions.

Circumstances of the Stop

Suspected crime, presence of weapons, actions of the suspect (casing, concealed possession, etc.), background circumstances (violent crime, suspect known to carry a weapon, etc.).

Justification and Basis for the Stop

Suspected Crime, Whether the officer explained the stop, the suspect's demeanor, the basis for any searches conducted (admission, consent, hard object, etc.).

Outcomes of the Stop

Whether the suspect was arrested, frisked, searched, if a summons was issued, and if any contraband or weapons were found.

Use of Force

Whether any physical force was used during the stop, including CEW, firearm, handcuffs, OC spray, etc.

2- NEIGHBORHOOD FINANCIAL HEALTH DIGITAL MAPPING AND DATA.

“ Provide an overview of financial health within neighborhoods across New York City. Each row in the dataset represents a specific Public Use Microdata Area (PUMA) in NYC, which is a statistical geographic unit with at least 100,000 people. “

Geographic and Demographic Information

Borough and Neighborhood details.
Community Districts within each PUMA.
Poverty rate, median income, and racial/ethnic composition of the population.

Specific indicators for each Indicators

- Homeownership rates.
- Access to banks and credit unions.
- Job training and placement support.
- Housing affordability.
- Health insurance coverage.
- Retirement security.
- Community resources.

Five key financial health Indicators

- Access to affordable and high-quality financial services.
- Access to affordable and essential goods and services.
- Access to quality jobs and income supports.
- Stability of housing and ability to manage financial shocks.
- Opportunities to build assets and plan for the future.

Performance and Ranking

- Overall index score for each financial health goal.
- Ranking of each PUMA compared to others based on the index score.
- Outcome or raw score for each indicator.
- Ranking of each PUMA based on its performance on individual indicators.

QUANTIFYING POLICE BIAS: KEY INDICATORS

The Challenge of Quantifying Bias:

- Identifying and measuring bias is crucial, but **bias itself is intangible and difficult to quantify.**
- **Traditional metrics**, such as arrest rates, **often fall short** of capturing the nuances of police suspicion and decision-making, and whether or not these arrests were justified.
- We need **robust indicators** that allow us to measure unjustified actions as a reflection of police bias.

To uncover bias, we focus on actions that do not result in arrests—revealing unwarranted suspicion and disproportionate scrutiny, without true cause to arrest:

- **Unjustified Stop Rates**: Highlight situations where individuals were stopped due to being suspected of a crime, but not arrested, suggesting police suspicion without justifiable cause.
- **Unjustified Frisks and Searches**: Reflect excessive perceptions of threat when no arrest follows.

Unjustified indicators allow us to uncover patterns of police behavior that disproportionately affect certain groups. They enable fair comparisons across racial and socioeconomic groups, even when crime rates vary.

DATA PRE-PROCESSING WORKFLOW

Data Cleaning: dropping Irrelevant Columns:

- **Neighborhood Data:** Removed attributes that do not contribute to identifying the neighborhood or its financial health. For example, descriptive information or redundant identifiers that were not necessary for modeling were excluded to reduce noise.
- **Stop & Frisk Data:** Eliminated attributes that were irrelevant to assessing unjustifiable police actions. For instance, columns indicating outcomes like whether a weapon was found during a stop were removed, as they do not align with the focus of evaluating unjustified stops, frisks, or searches.

Dealing with NULL variables:

- For binary variables, often times NULL meant False. These values were replaced accordingly.
- When NULL due to lack of information, rows were dropped.

Merging Datasets:

- An intermediate table was manually created to link Stop and Frisk Precinct location to its Neighborhood PUMA (Public Use Microdata Area)
- Neighborhood Financial Health and Stop and Frisk datasets were joined with the help of this intermediate dataset

FEATURE ENGINEERING FOR BIAS DETECTION

Target Engineering:

Created target variables for the model:

- **Unjustified Stop:** Defined as a stop without an arrest (SUSPECT_ARRESTED_FLAG = 0).
- **Unjustified Frisk:** Defined as a frisk where the individual was not arrested.
- **Unjustified Search**: Defined as a search where the individual was not arrested.

Cyclical Encoding of Time Data:

- Captures the periodic nature of time-related variables (e.g., hour, weekday, month) to reflect their cyclical relationships and ensures the model understands temporal proximity (e.g., hour 23 and hour 0 are close).
- How: Applied sine (sin) and cosine (cos) transformations to variables like Hour, Day of the Week, and Month.
- Created pairs of features (e.g., Hour_sin and Hour_cos) to encode cyclicality.

Dummification of Categorical Variables:
Transformed categorical columns like DEMEANOR_OF_PERSON_STOPPED and STOP_WAS_INITIATED (by radio or self) into dummy variables using one-hot encoding.

MULTICOLLINEARITY ANALYSIS:

Correlation Analysis:

- Generated a correlation matrix to identify highly correlated variable pairs.
- Focused on pairs with absolute correlation ≥ 0.8.
- Removed highly correlated variables to prevent redundancy and multicollinearity.

Variance Inflation Factor (VIF):

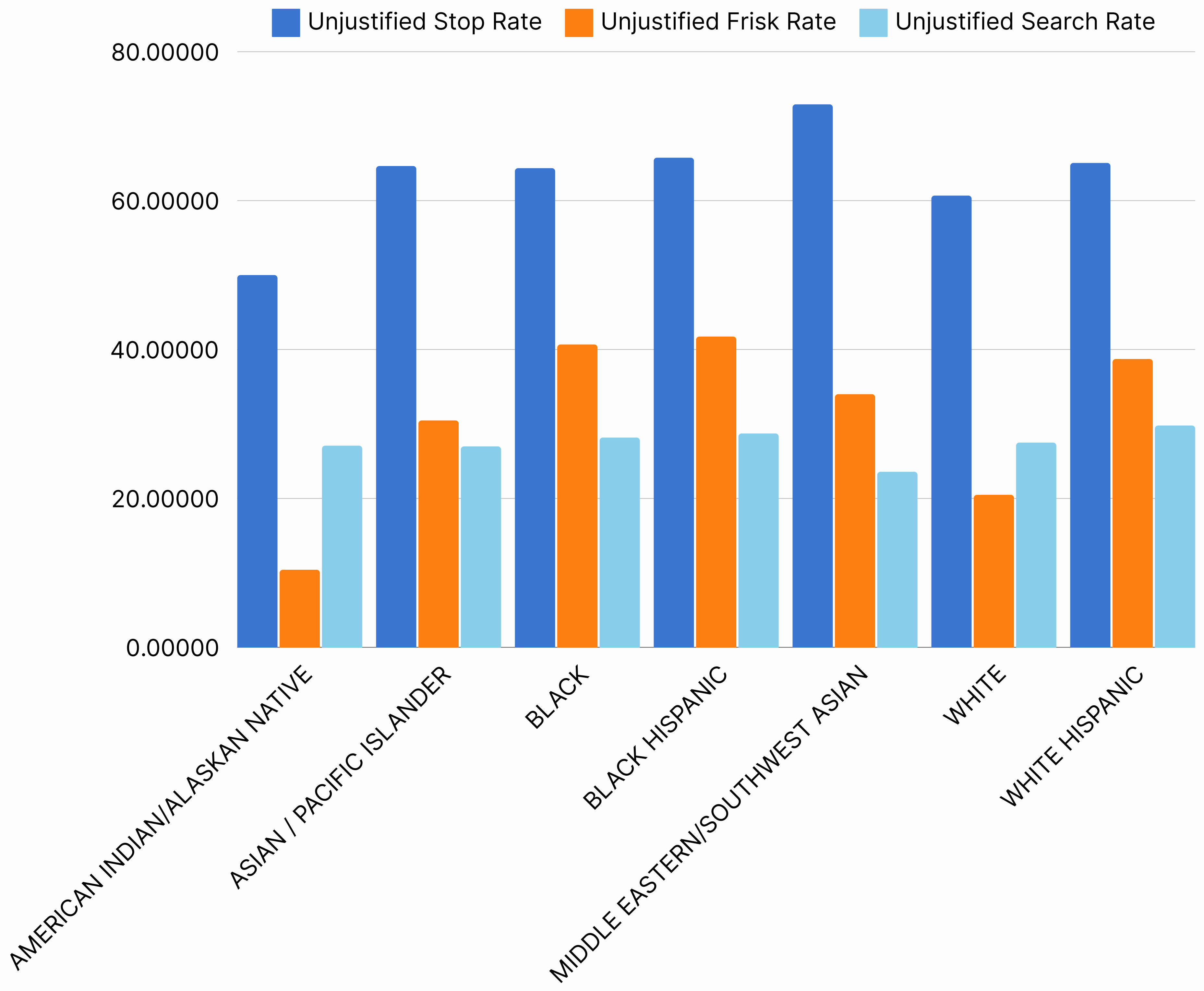
- Converted boolean columns (True/False) into integers (1/0).
- Calculated VIF for each feature to quantify multicollinearity.
- Removed variables with VIF > 10.

Examples of multicollinearity in our dataset:

- Area median income and other financial health indicators.
- Area median income and % of Hispanic population.

EXPLORATORY DATA ANALYSIS - EDA

RACIAL DISPARITIES IN UNJUSTIFIED POLICE ACTIONS



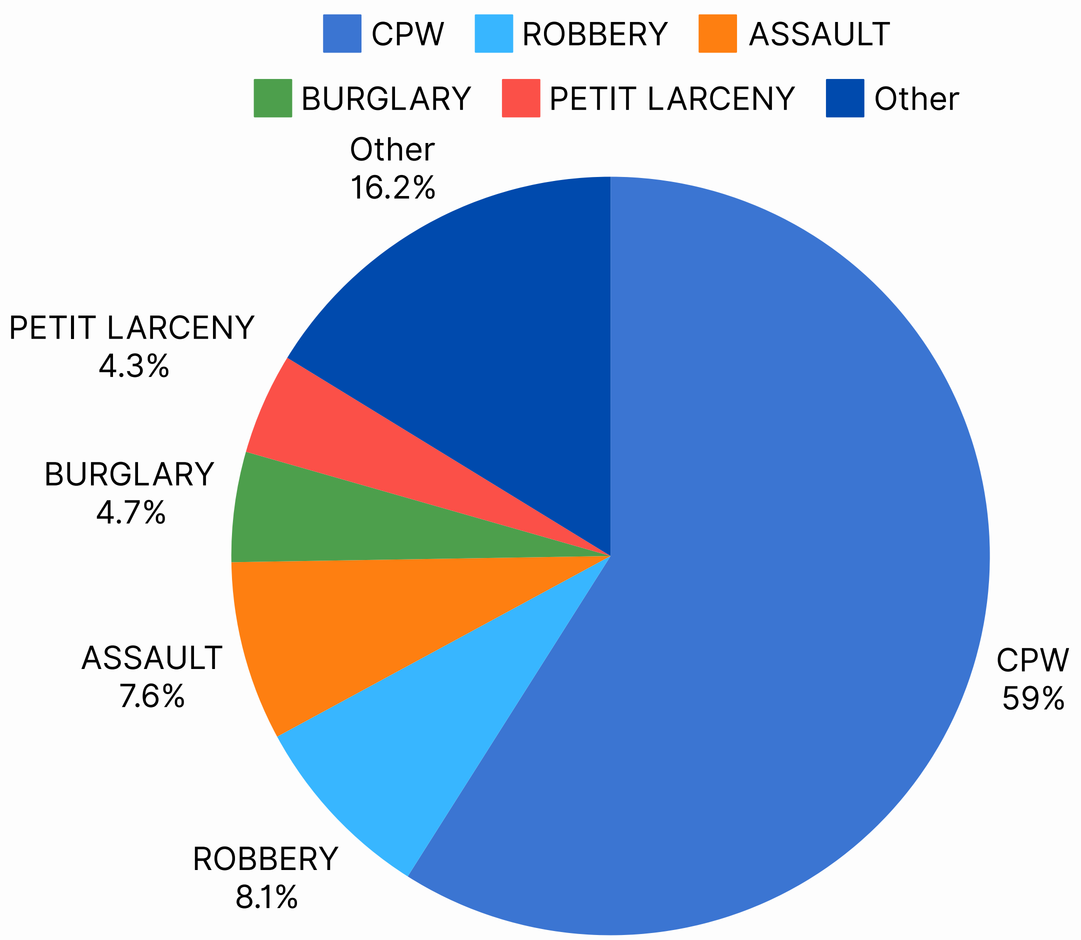
This chart shows **racial disparities in stop outcomes**, especially after the stop occurs.

Small differences in unjustified stop rates, with Black and Hispanic individuals slightly more affected.

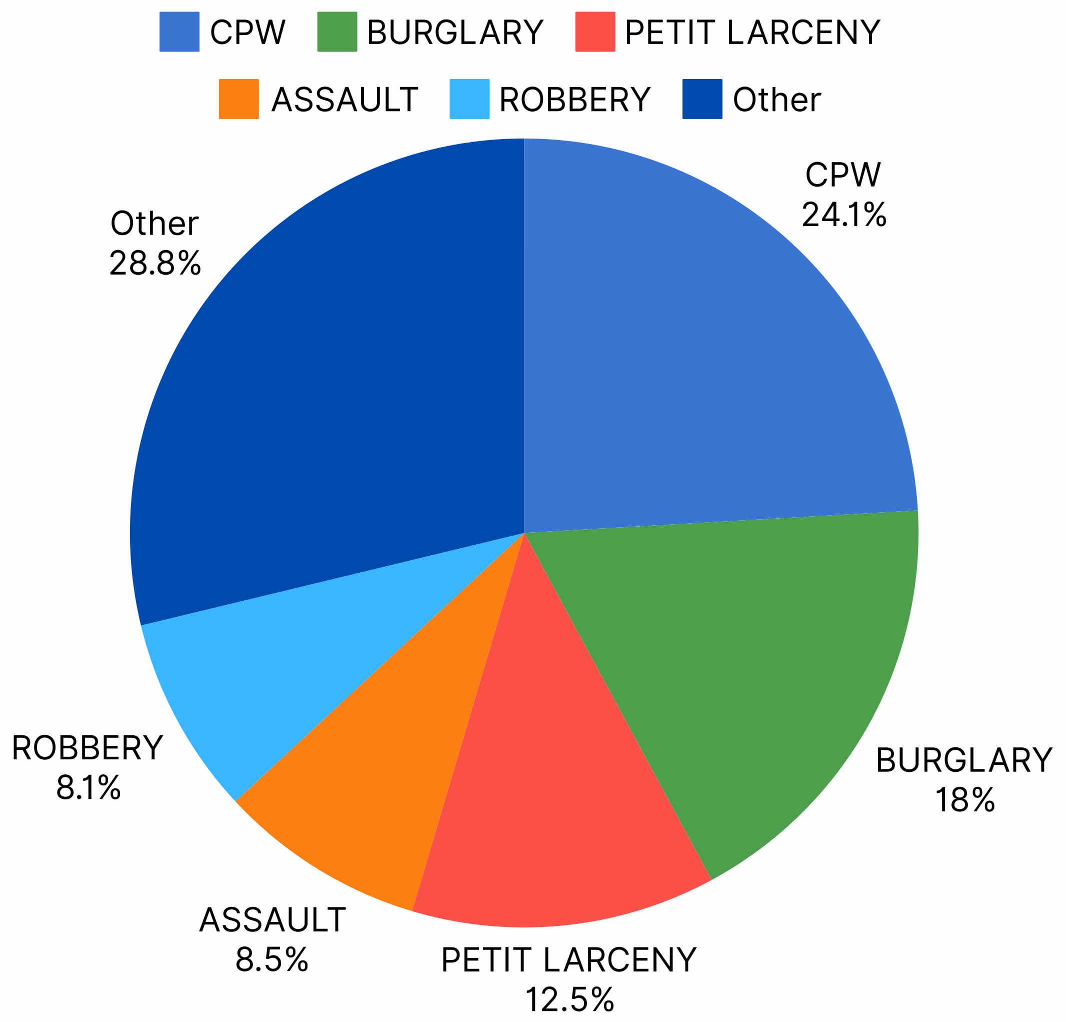
The **disparities grow significantly after the stop**, as Black and Hispanic individuals face **much higher unjustified frisk and search rates** compared to White and Asian individuals.

This suggests that police bias increases after an officer observes the individual's race.

RACIAL DISPARITIES IN UNJUSTIFIED POLICE ACTIONS



Distribution of Suspected Crimes (No Arrest) for Black Individuals

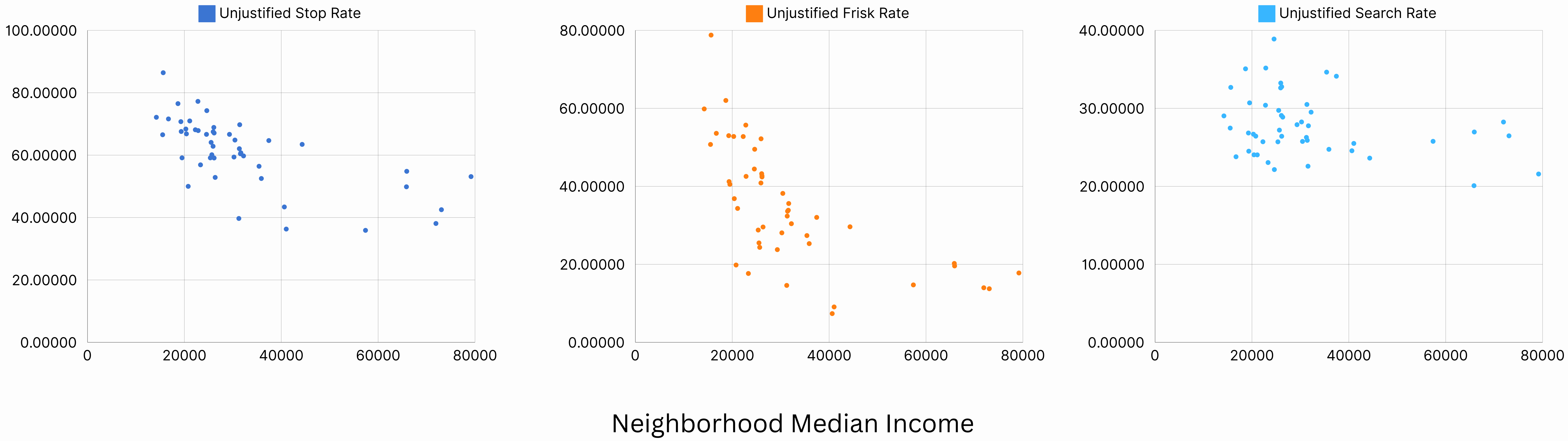


Distribution of Suspected Crimes (No Arrest) for White Individuals

- **Black** individuals are overwhelmingly **suspected of CPW** compared to White individuals, who face a broader spectrum of suspected crimes.

This reflect potential racial profiling, where systemic biases in law enforcement disproportionately target Black individuals for weapon-related suspicions.

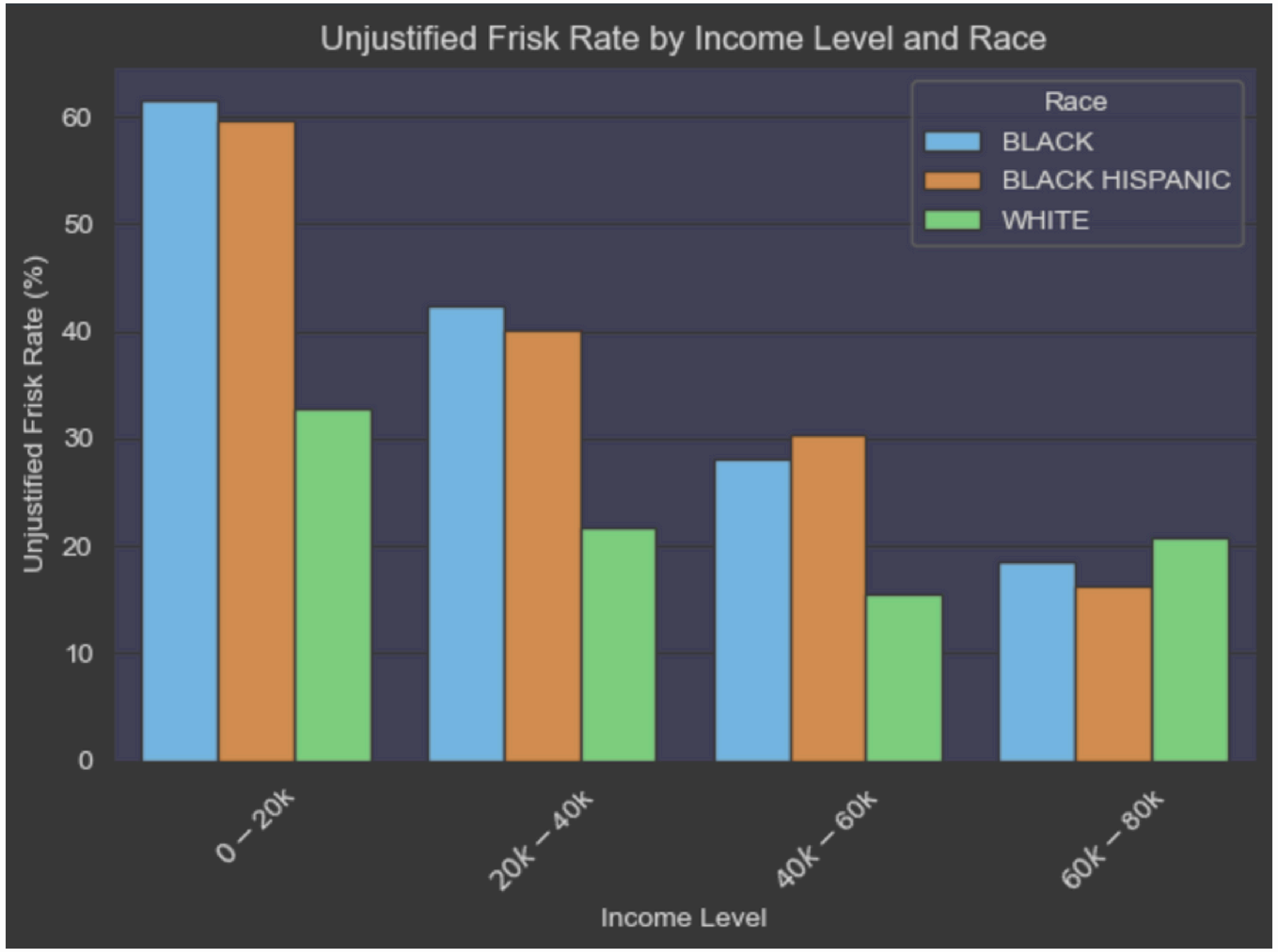
ECONOMIC DISPARITIES IN UNJUSTIFIED POLICE ACTIONS



Lower-income neighborhoods show disproportionately **higher unjustified stop rates**.

This highlights how economic vulnerability correlates with increased policing actions and suspicion, regardless of arrest outcomes.

ECONOMIC DISPARITIES IN UNJUSTIFIED POLICE ACTIONS

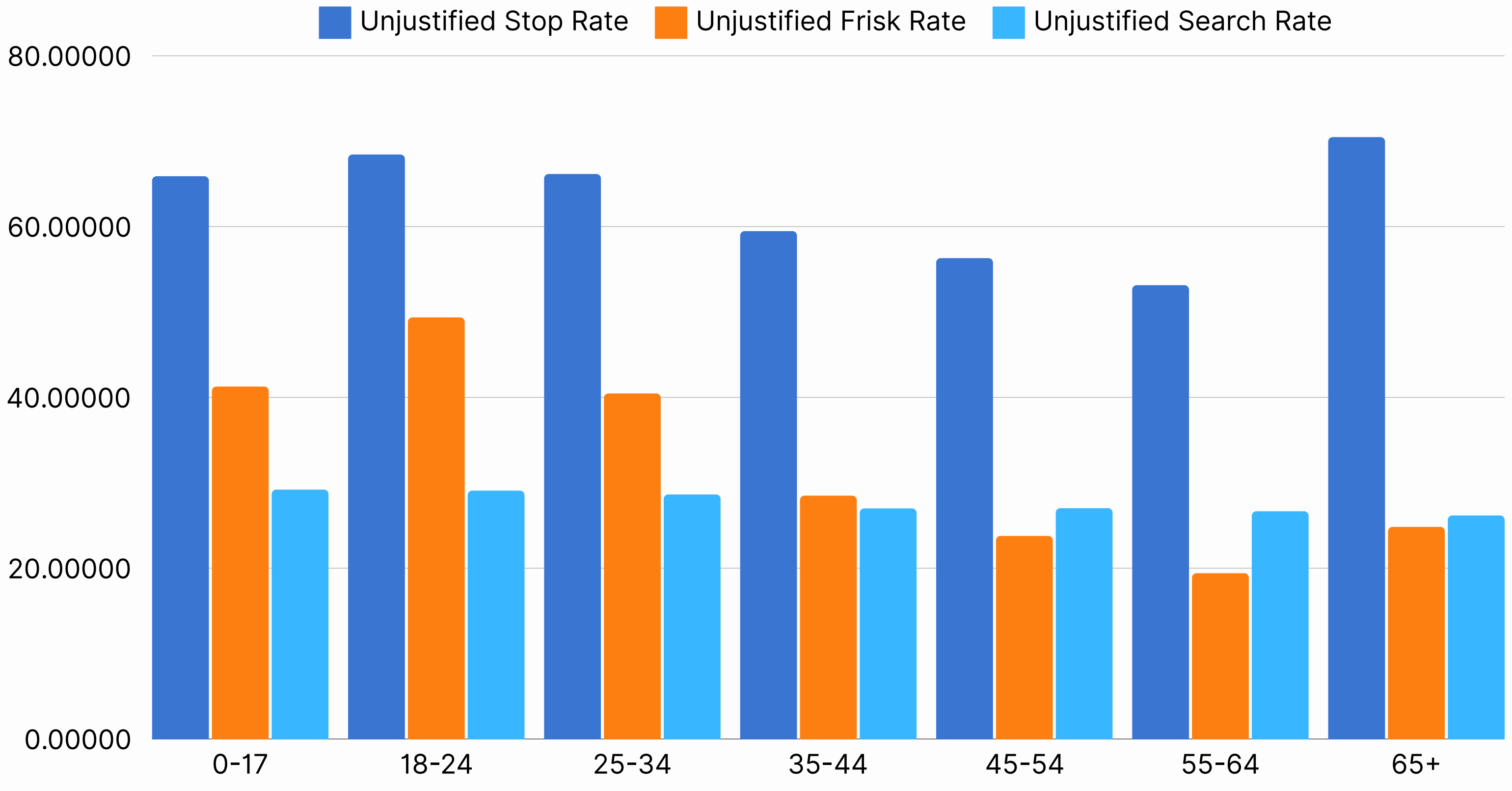


These findings reveal an **intersection between racial and economic disparities**: racial minorities in lower-income neighborhoods face the highest levels of unjustified police actions.

in low income areas, minorities are a lot more threatening than white individuals.

As the neighborhood income level increases, racial disparities diminish.

BEHAVIOR AND DEMOGRAPHICS IN UNJUSTIFIED POLICE ACTIONS



Younger individuals (0-17 and 18-24) experience the **highest** unjustified stop rates (approximately 65-70%), indicating a disproportionate targeting of younger populations in police actions.

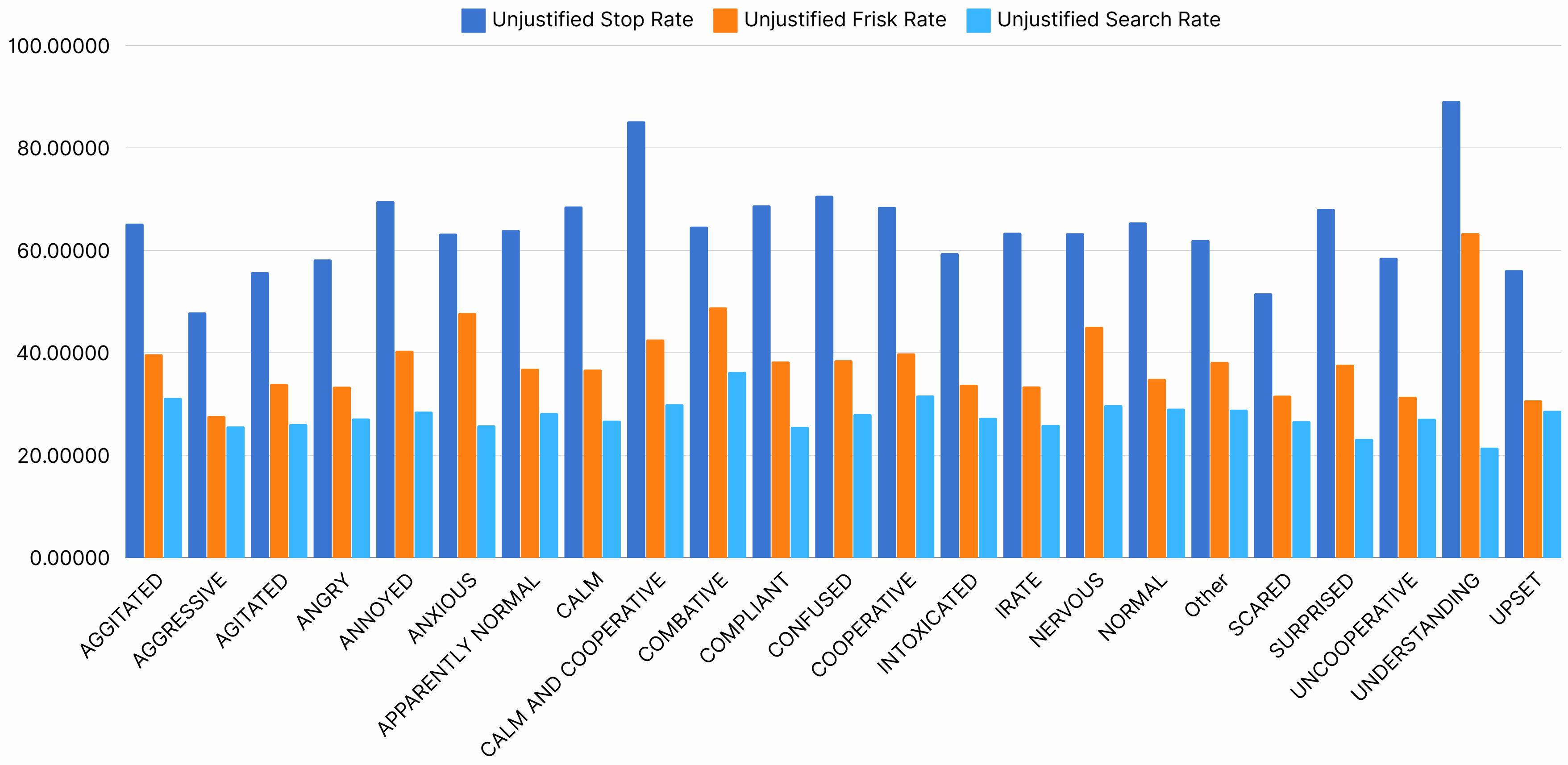
As **age increases**, unjustified frisk and search **rates decrease slightly**, but older individuals (65+) still face substantial unjustified stop rates, comparable to the youngest groups.

Younger individuals may face unjustified stops due to stereotypes or perceptions of delinquency, raising questions about fairness in age-based profiling.

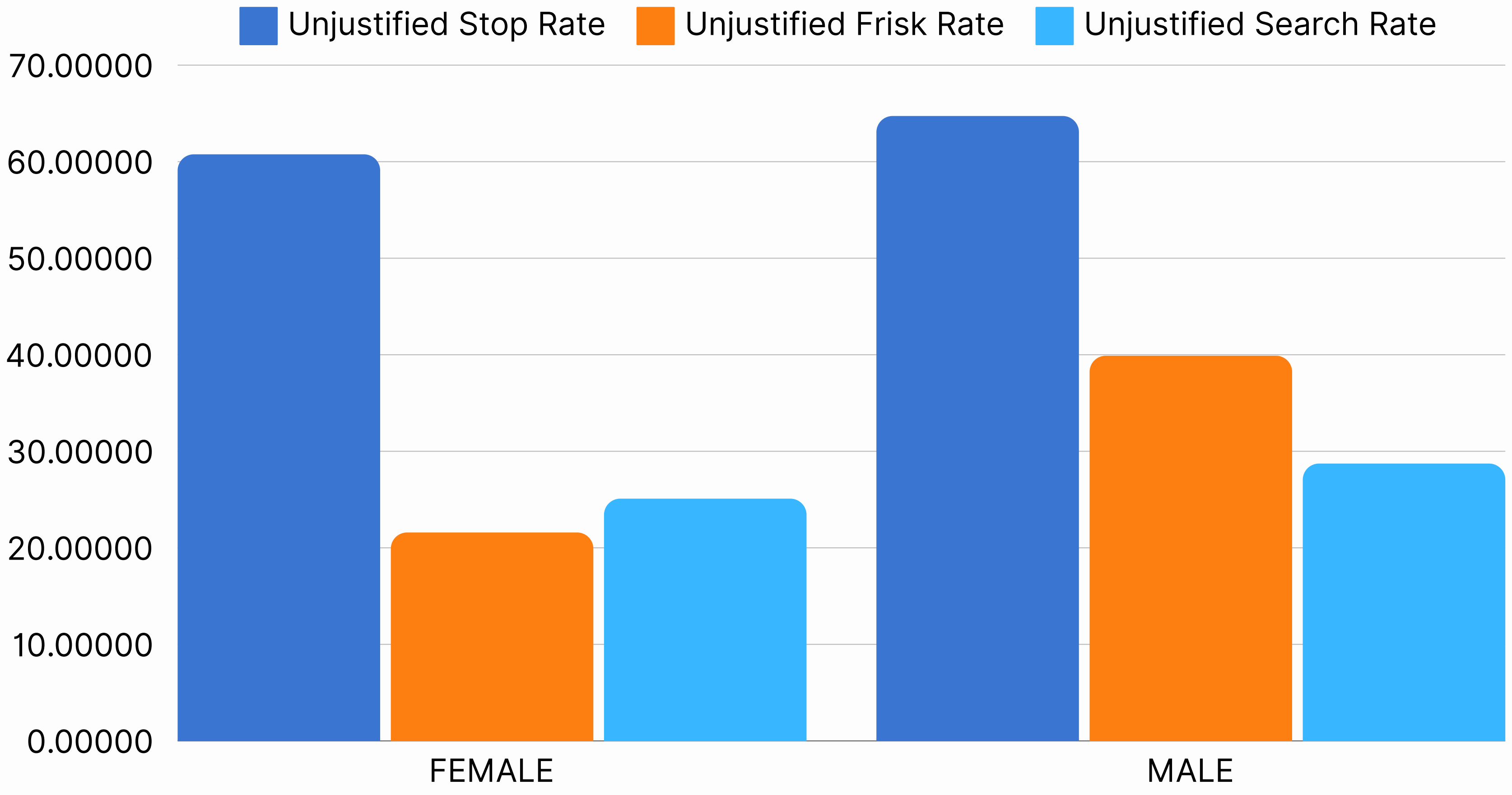
Individuals exhibiting **hostile, anxious, or aggressive behaviors** are more likely to experience stops, frisks, and searches that align with arrest outcomes, resulting in **lower unjustified rates**.

Individuals displaying **calm or cooperative behaviors** are less likely result in arrests.

This pattern indicates that perceived demeanor is often a good indicator of whether an arrest is warranted.



PHYSIOLOGY IN UNJUSTIFIED POLICE ACTIONS

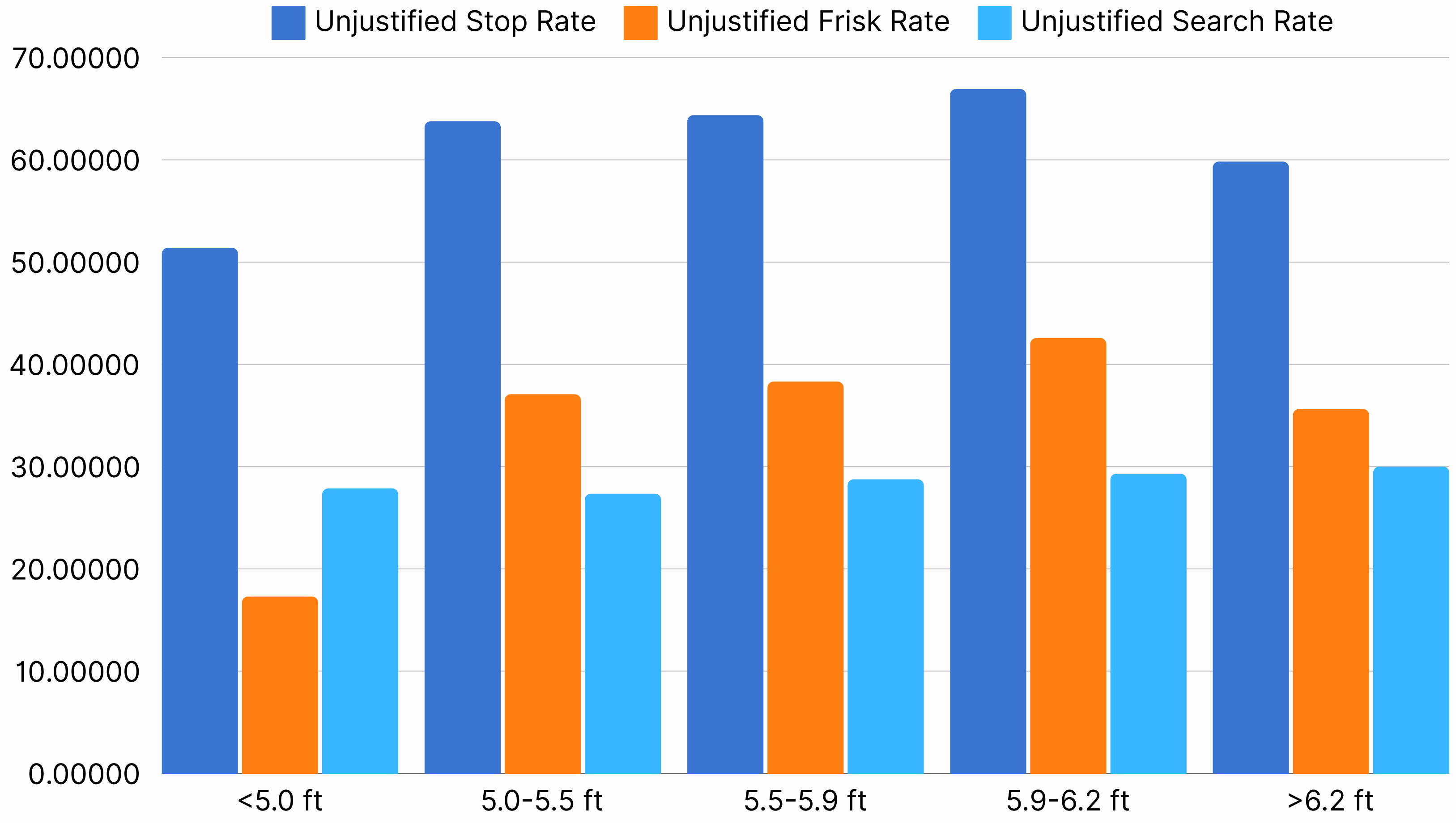


The data suggests that police officers are more inclined to subject men to additional scrutiny and suspicion, often engaging in more invasive actions that ultimately do not yield arrests.

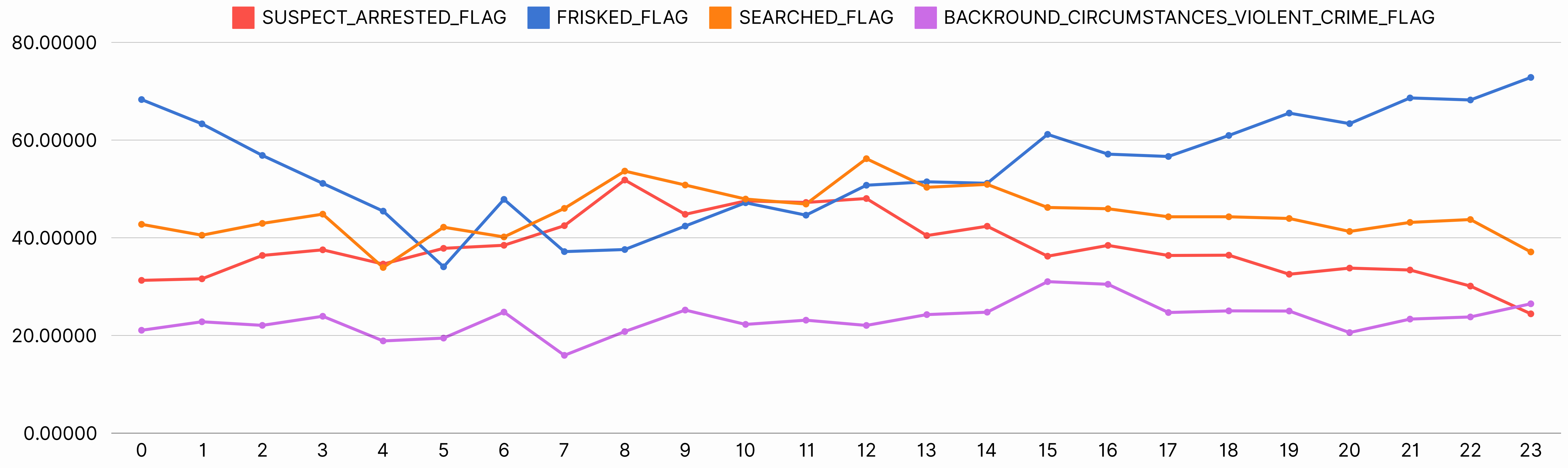
This disparity highlights a **potential bias in policing practices**, where **men, compared to women**, are treated with heightened suspicion during stops.

Unjustified actions are **relatively consistent for all heights above 5.0 ft.**

Shorter individuals (<5.0 ft) experience relatively **lower frisk and search rates** which may indicate lower perceived threat by the police officer.



TEMPORAL PATTERNS OF UNJUSTIFIED POLICE ACTIONS

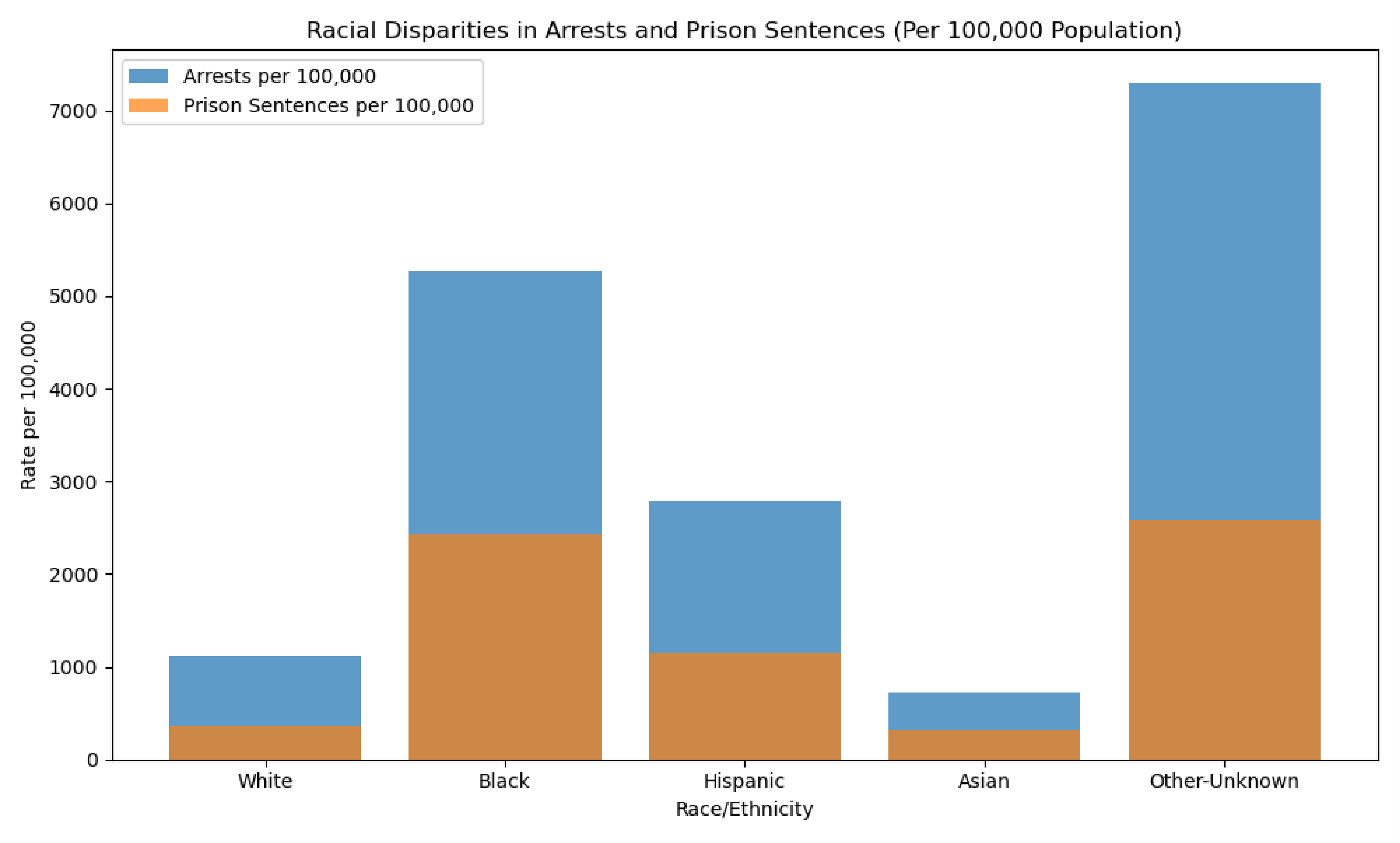


Hourly trends show that frisk and search rates **peak during the late afternoon and evening**, reflecting increased enforcement during busier times of the day.

PREDICTIVE MODEL DEVELOPMENT

INTRODUCTION

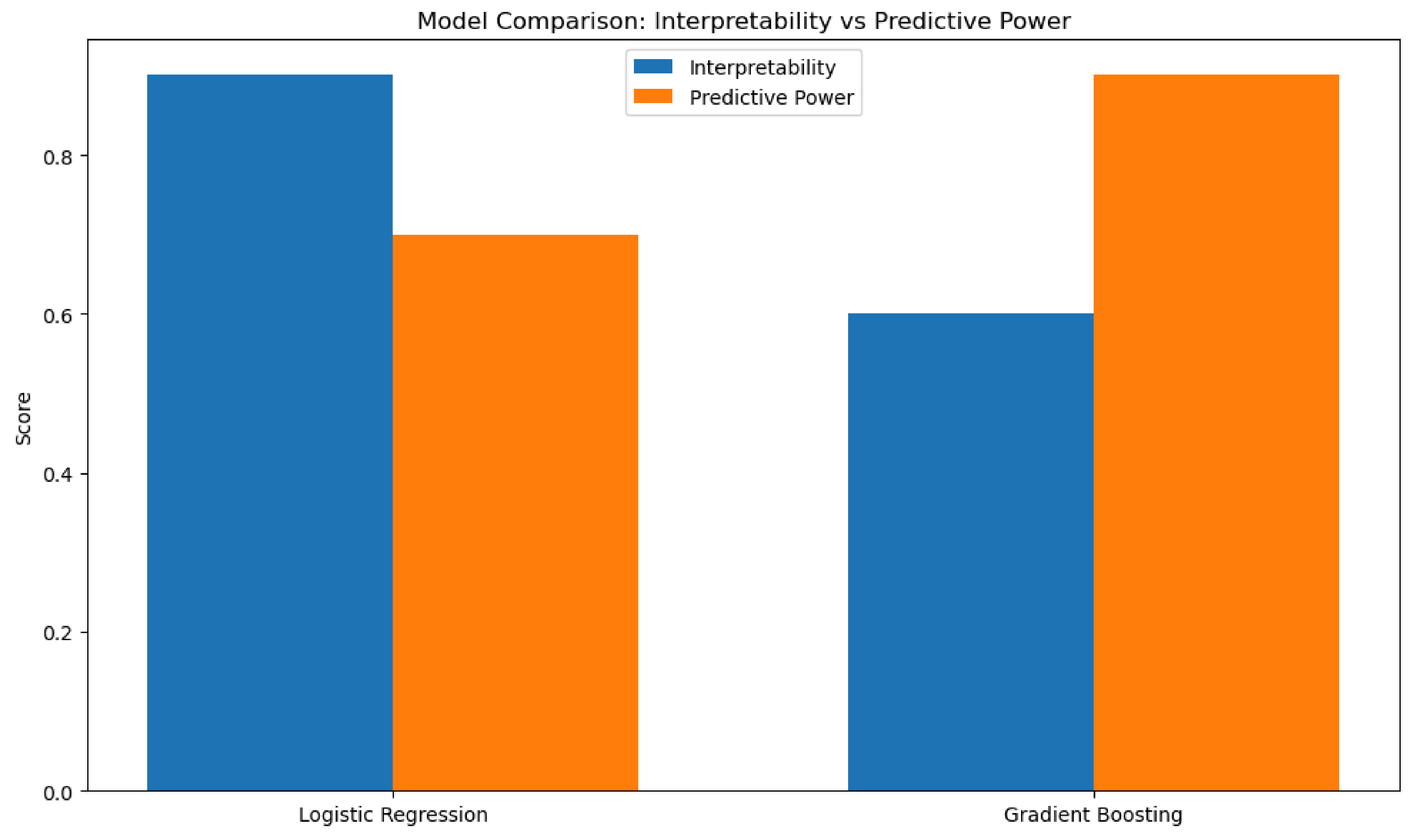
Current Analysis - Racial Disparities in Arrests and Prison Sentences:



- **Black** individuals have the **highest rate** of both **arrests** and **prison** sentences **per 100,000 people**, with prison sentences significantly higher compared to other groups.
- **Hispanic** individuals also experience a notable disparity, with **prison** sentence rates **higher** than the White and Asian populations.
- **White** individuals make up the majority of the population but have **significantly lower arrest** and **prison** sentence rates.

1. New York State Division of Criminal Justice Services. (2023). 2022 Population, arrests, prison by race. NYS DCJS-OJRP.
2. Monaghan, S., Rempel, M., & Lin, T. (2023, February). Racial disparities in the use of jail across New York City, 2016-2021. Data Collaborative for Justice.
3. McCormack, S., & Barber, J. (n.d.). A racial disparity across New York that is truly jarring.

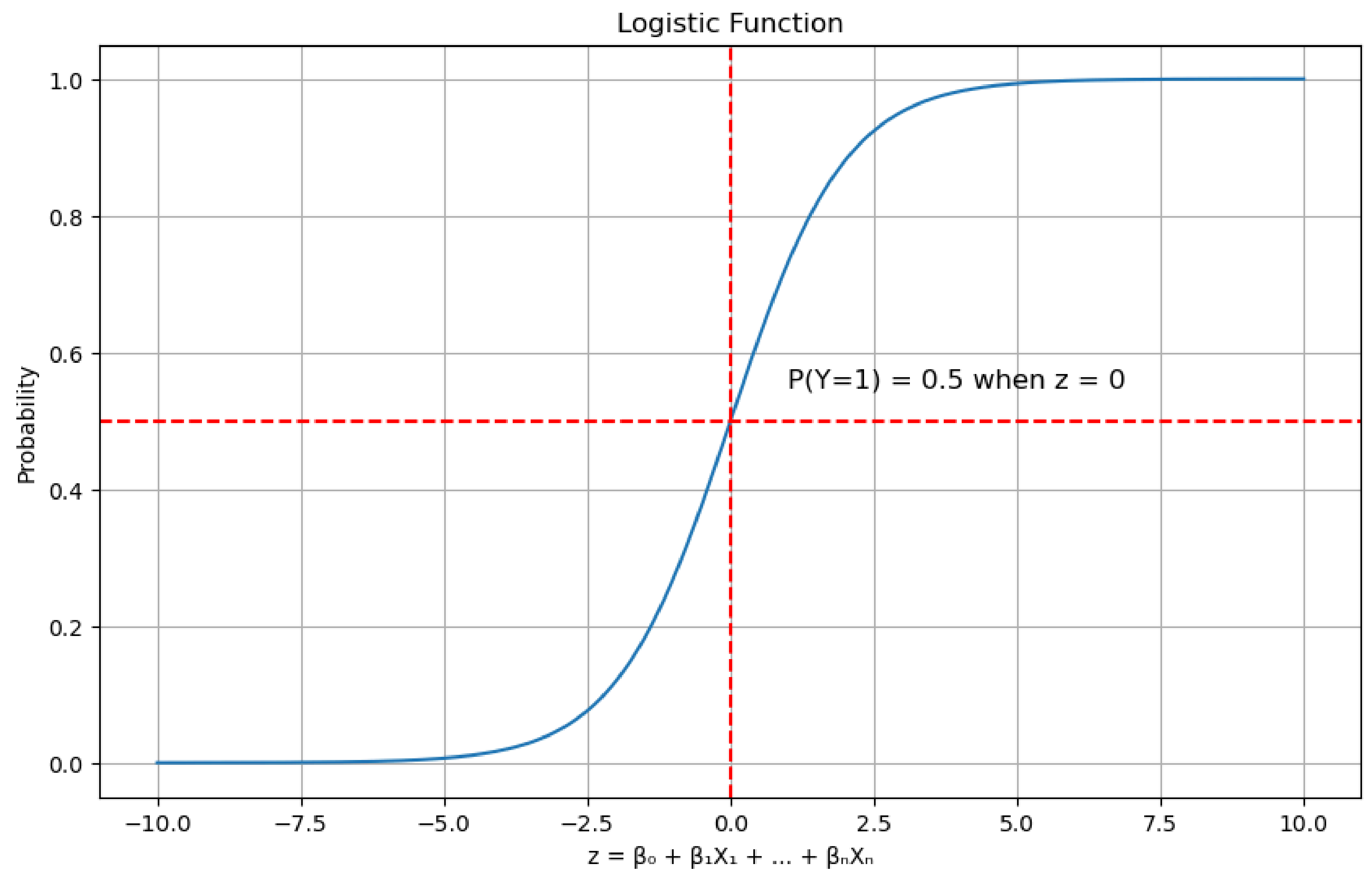
PREDICTIVE MODEL SELECTION PROCESS



- We chose **Logistic Regression** and **Gradient Boosting** for our analysis.
- Logistic Regression is ideal for **binary classification problems**.
- Gradient Boosting is a powerful ensemble method that **combines weak learners**.
- These models allow us to **predict unjustified stops, frisks, and searches.**
- Logistic Regression offers **interpretability**, while Gradient Boosting provides **high predictive power**.

Our goal is to compare their performance and gain insights into factors influencing police actions.

MODEL SELECTION: LOGISTIC REGRESSION



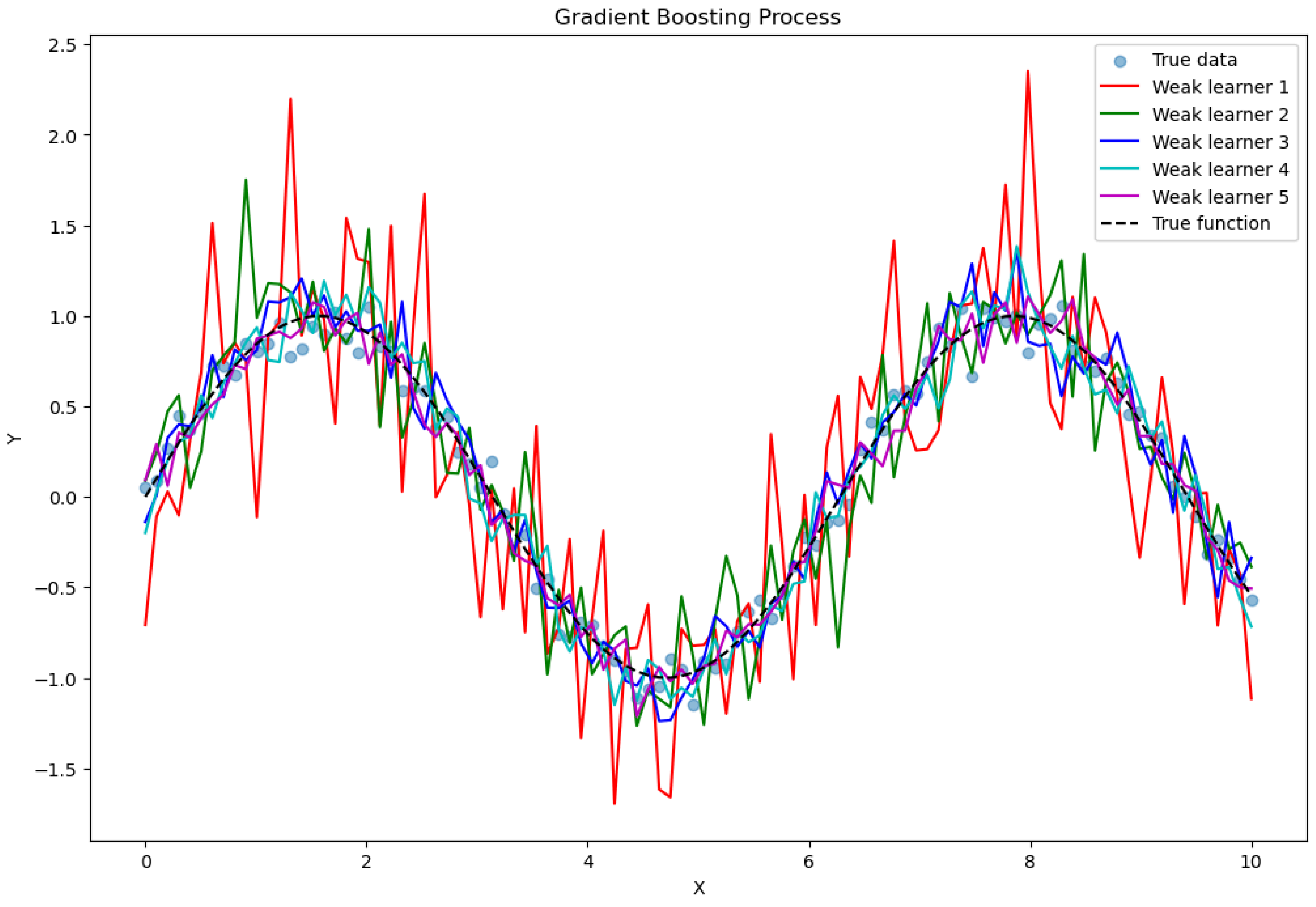
Logistic Regression estimates the probability of a binary outcome.

- It uses a logistic function to model the relationship between features and target.
- **Advantages**: interpretable, efficient, and less prone to overfitting.
- **Limitations**: assumes linear relationship between features and log-odds.
- We applied **Lasso** regularization to **reduce overfitting** and **perform feature selection**.
- The logistic function is defined as:

P(Y=1) = 1 / (1 + e^(-z)),
where $z = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n$

- Coefficients (β) represent the change in log-odds for a unit change in the predictor.

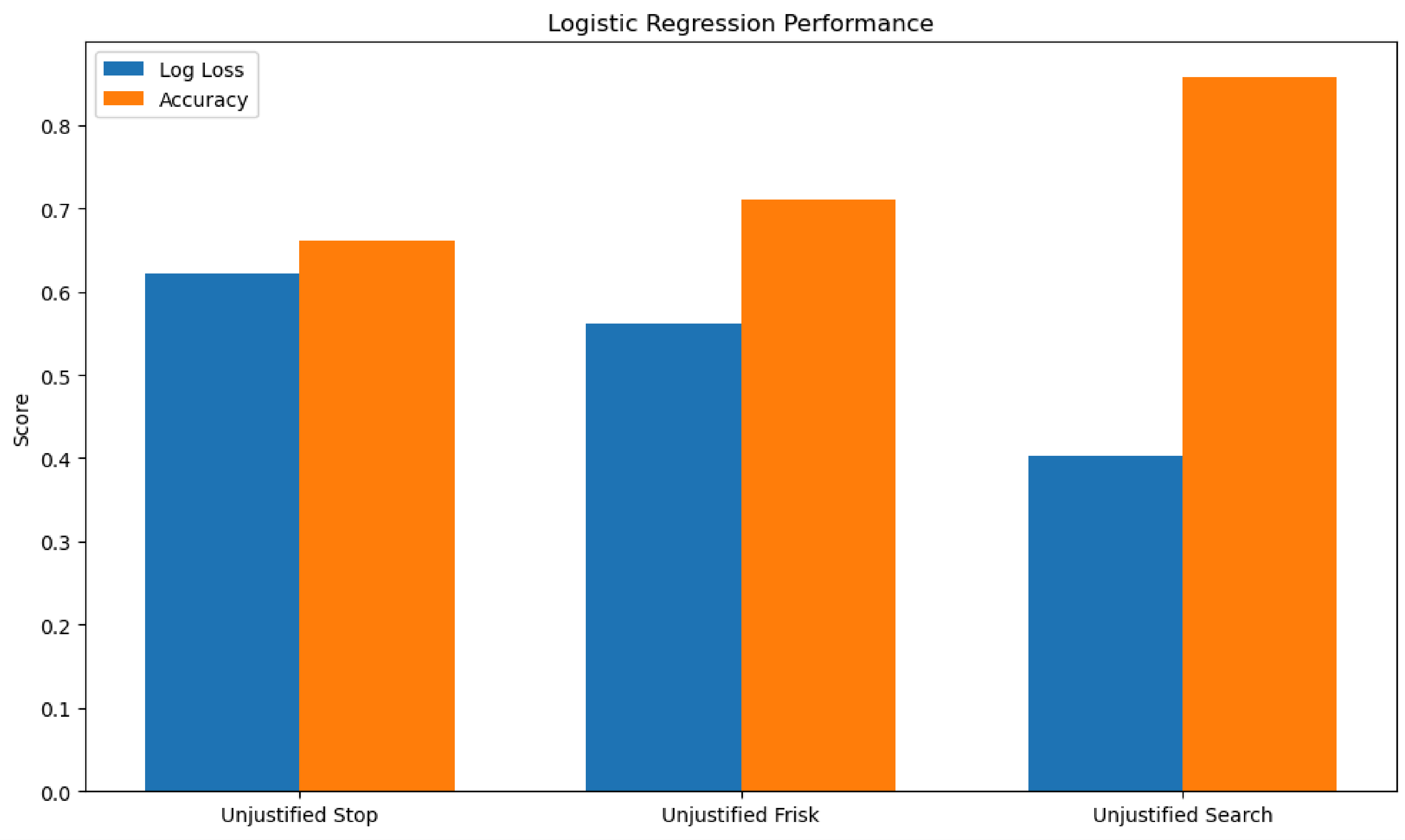
MODEL SELECTION: GRADIENT BOOSTING



Gradient Boosting builds an ensemble of weak learners sequentially.

- Each new model corrects errors made by the previous models.
- **Advantages**: high predictive power, handles non-linear relationships.
- **Limitations**: can be prone to overfitting, less interpretable than linear models.
- We used **hyperparameter tuning** to optimize model performance.
- Key hyperparameters: number of estimators, learning rate, max depth, and subsample.
- The algorithm minimizes a loss function by adding weak learners iteratively.
- Feature importance is calculated based on how often a feature is used to split the data.

MODEL PERFORMANCE: LOGISTIC REGRESSION



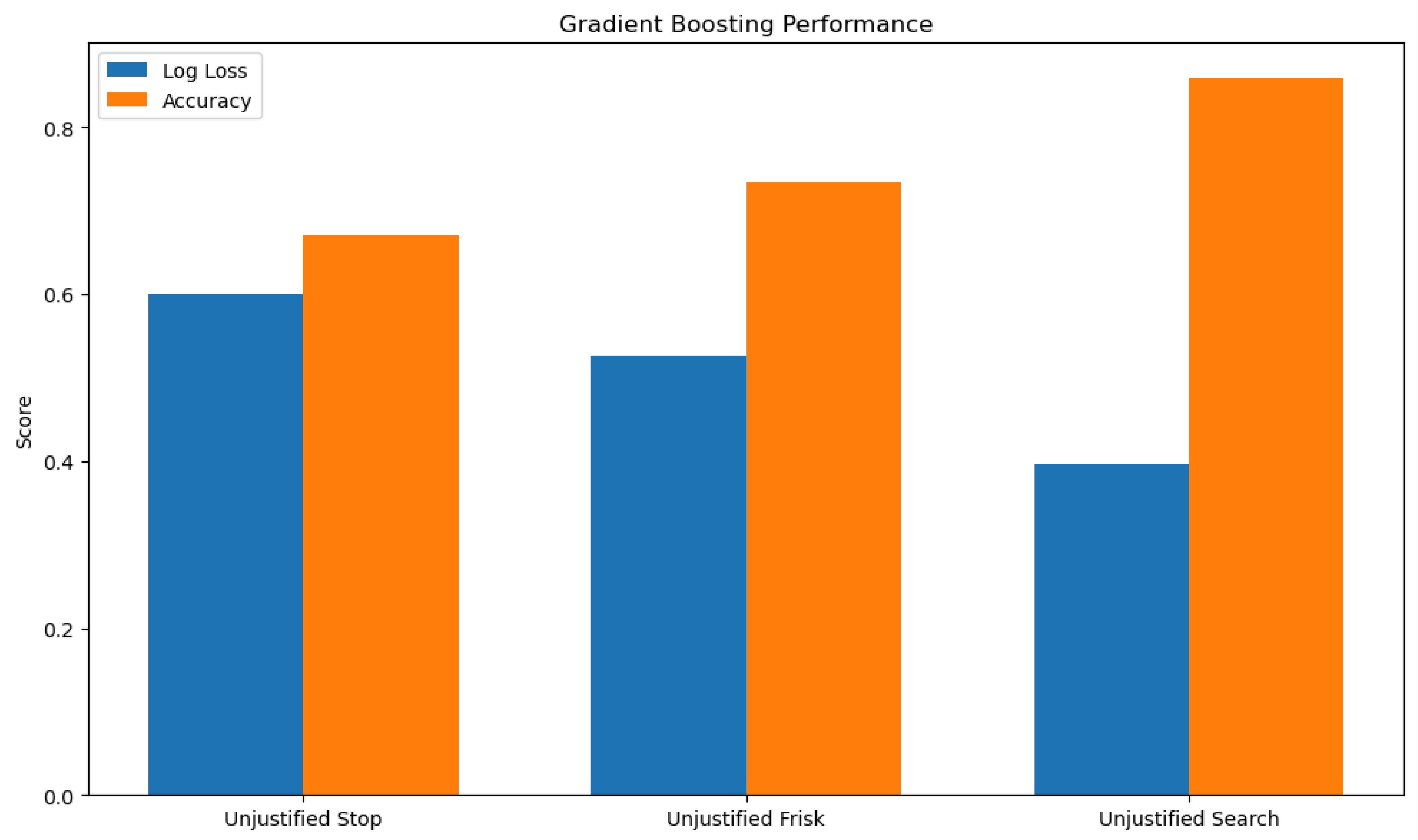
Unjustified Stop: Log Loss: 0.622, Accuracy: 0.661

Unjustified Frisk: Log Loss: 0.562, Accuracy: 0.710

Unjustified Search: Log Loss: 0.402, Accuracy: 0.857

- Log Loss measures the accuracy of predicted probabilities.
- Accuracy represents the proportion of correct predictions.
- Lower Log Loss indicates better model performance.
- Unjustified Search prediction shows the highest accuracy.
- Results suggest the model performs best for predicting unjustified searches.

MODEL PERFORMANCE: GRADIENT BOOSTING



Unjustified Stop: Log Loss: 0.600, Accuracy: 0.670

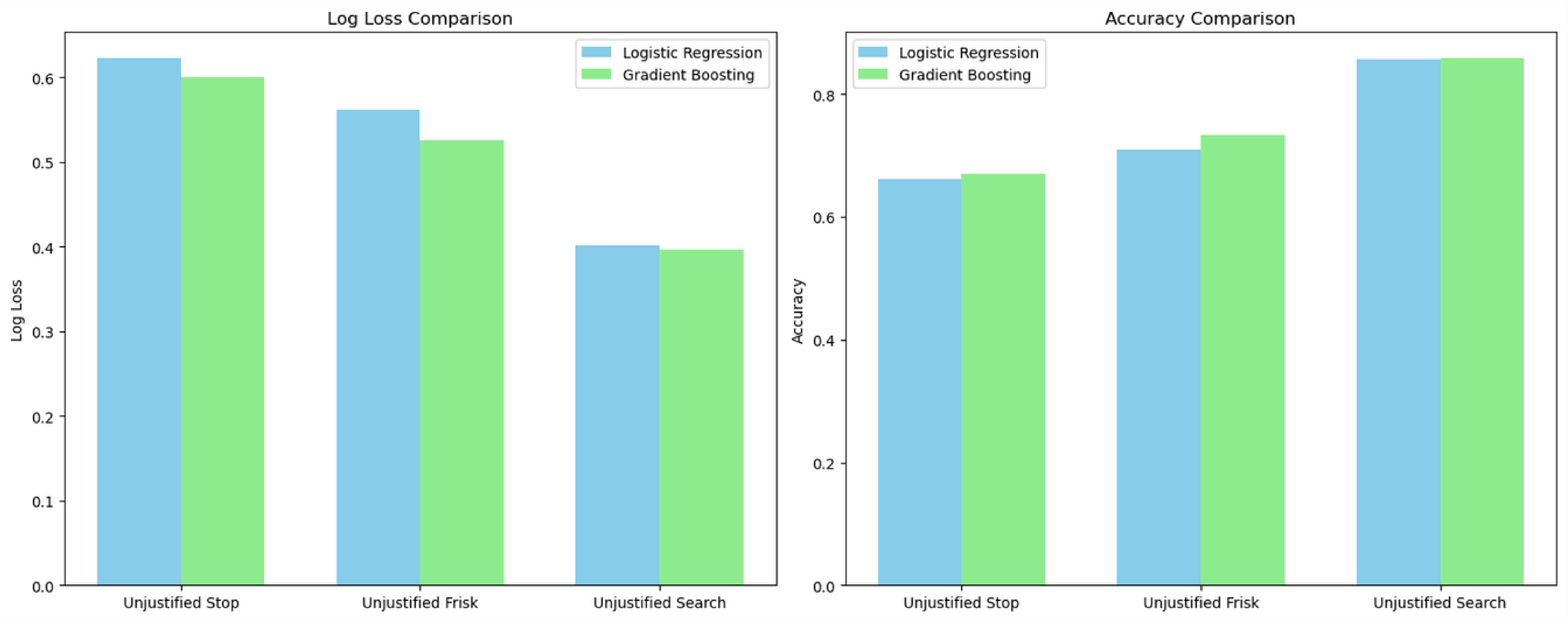
Unjustified Frisk: Log Loss: 0.526, Accuracy: 0.733

Unjustified Search: Log Loss: 0.396, Accuracy: 0.858

- Gradient Boosting slightly outperforms Logistic Regression.
- Improvement is most notable for Unjustified Frisk predictions.
- The model shows consistent performance across all three prediction tasks.
- Unjustified Search prediction remains the most accurate.
- The smaller gap between Log Loss and Accuracy suggests better calibrated probabilities.

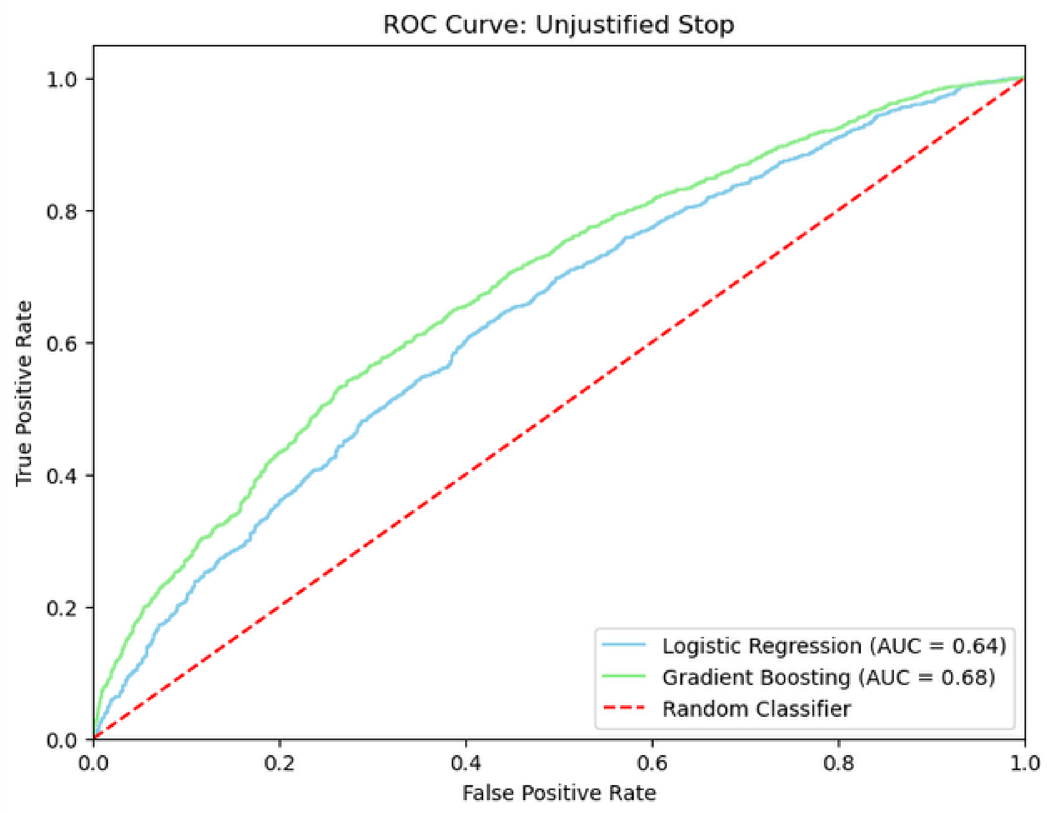
MODEL PERFORMANCE: COMPARISON

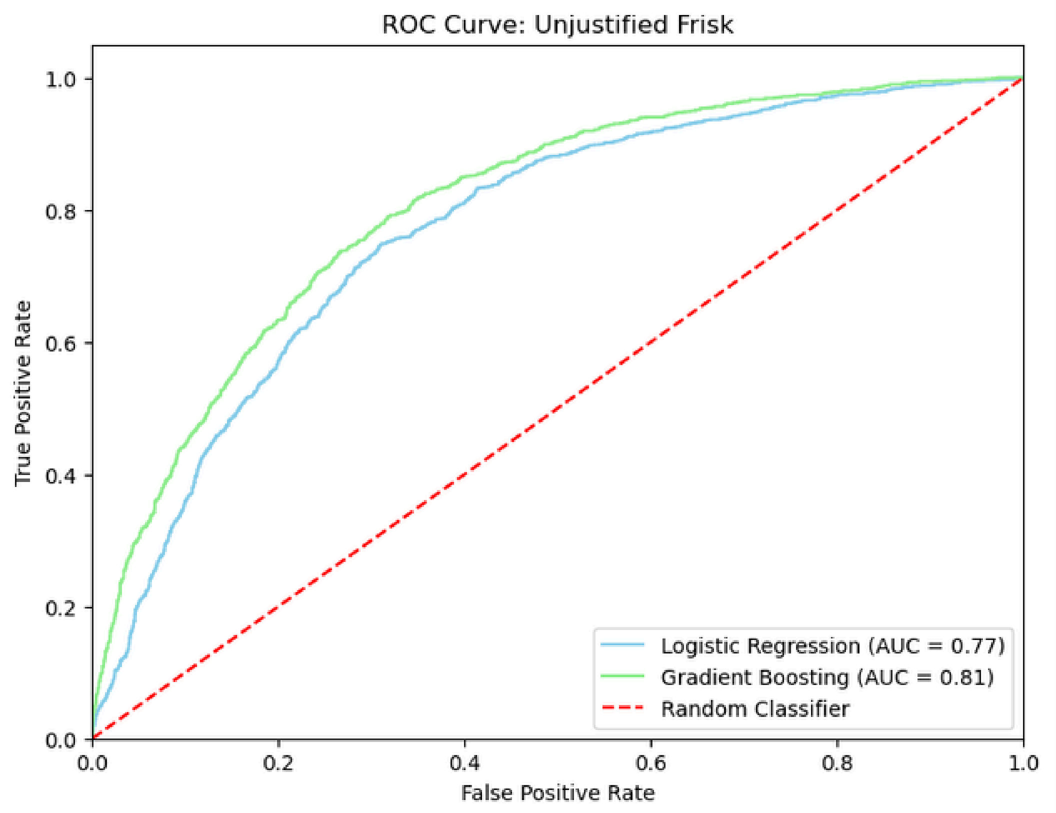
- Performance metrics used: **Log Loss** (measure of model uncertainty) and **Accuracy** (measure of correct predictions).
- **Lower Log Loss** indicates **better** model calibration and confidence in predictions.
- **Higher Accuracy** indicates a **greater** proportion of correct classifications.
- **Gradient Boosting** slightly **outperforms** Logistic Regression across all tasks.
- Both models perform best on the Unjustified Search task, with lowest Log Loss and highest Accuracy.
- Unjustified Stop task proves most challenging, with highest Log Loss and lowest Accuracy for both models.

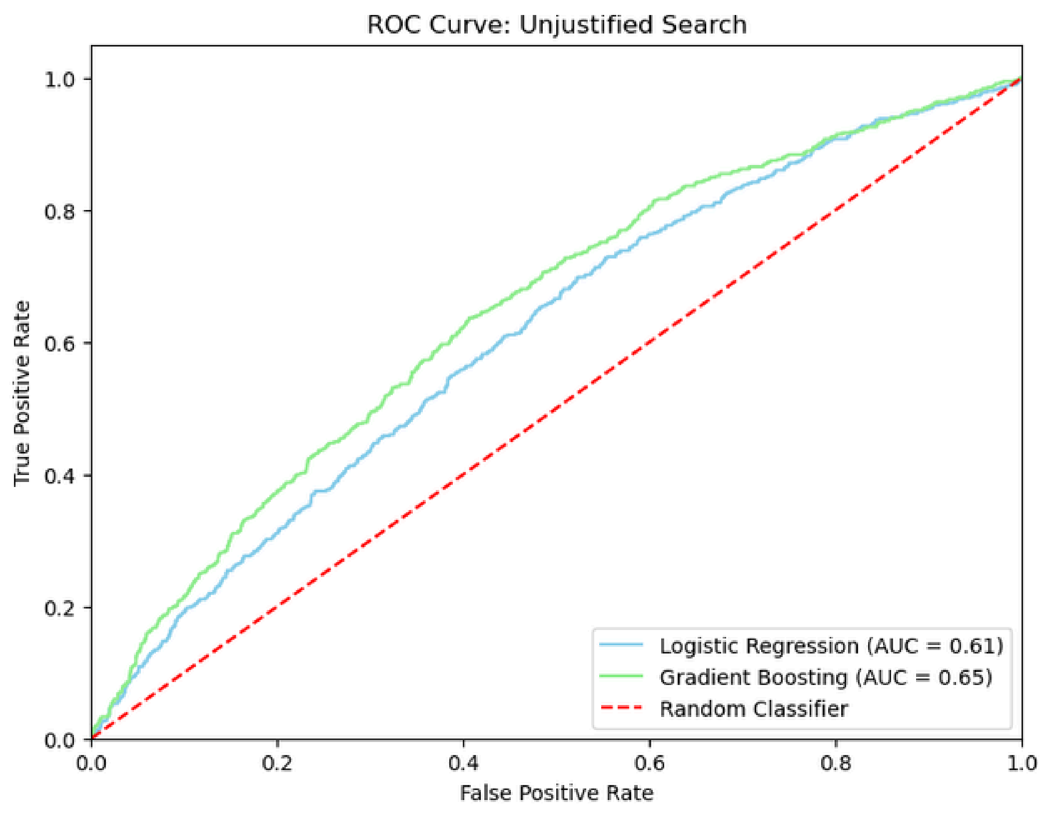


MODEL PERFORMANCE: ROC CURVE ANALYSIS

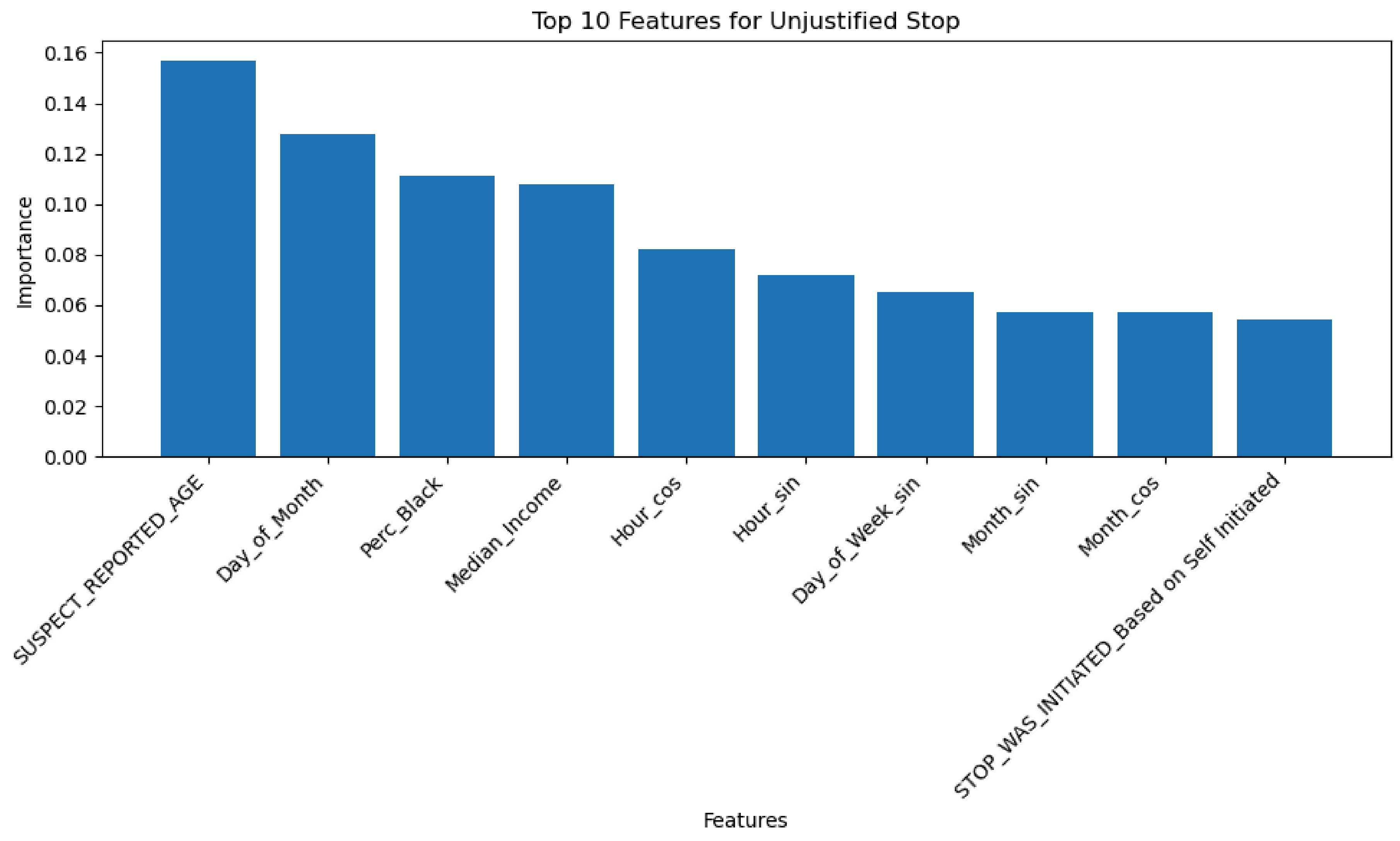
- **Gradient Boosting** consistently **outperforms** Logistic Regression across all tasks.
- **Unjustified Frisk** shows the **strongest predictive performance** (AUC: 0.81 GB, 0.77 LR).
- Both models perform significantly better than random classification (AUC = 0.5).
- Trade-off between sensitivity and specificity can be adjusted based on threshold selection.
- Frisk predictions are most reliable for operational use.
- Gradient Boosting should be preferred for all three classification tasks.





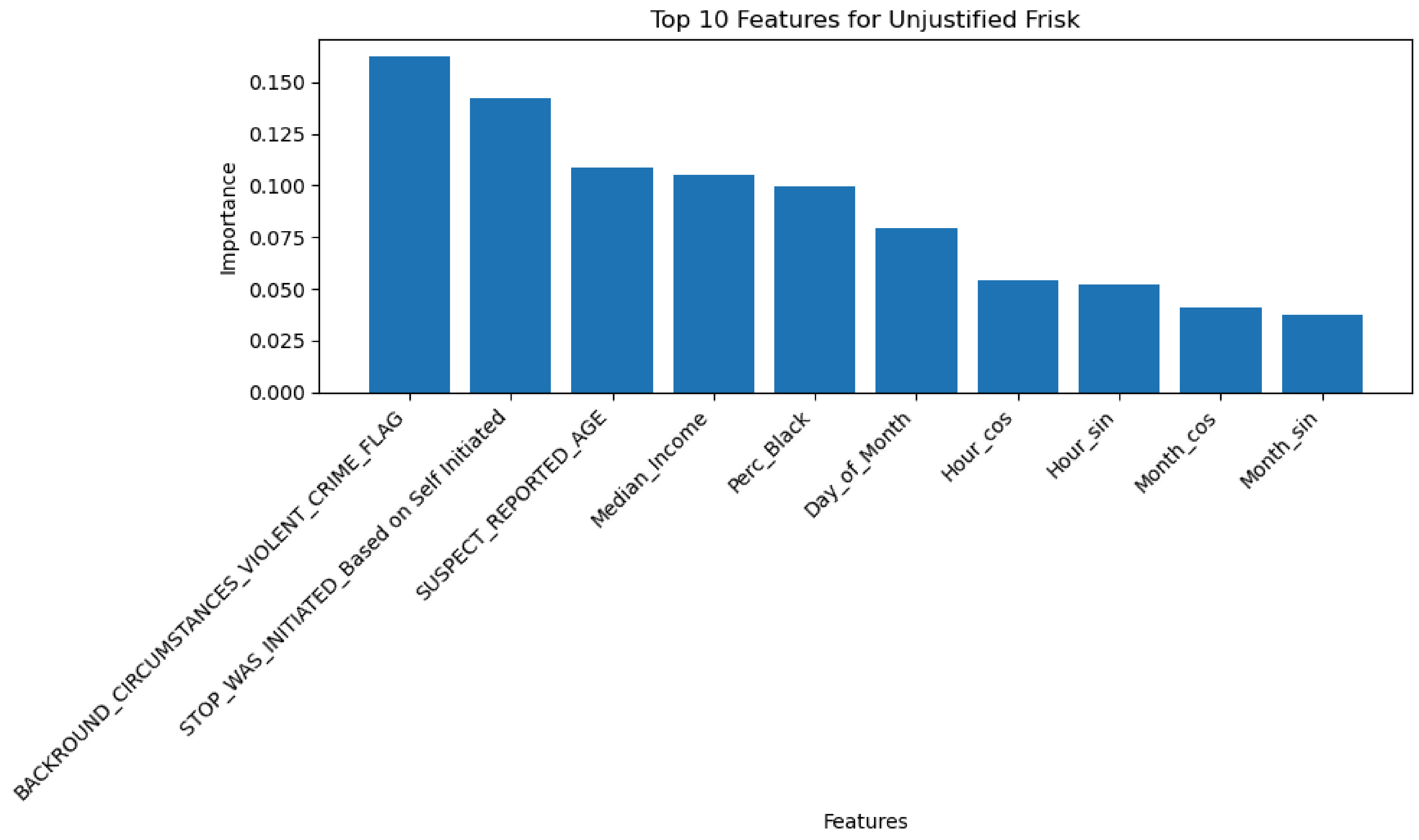


FEATURE IMPORTANCE: UNJUSTIFIED STOP



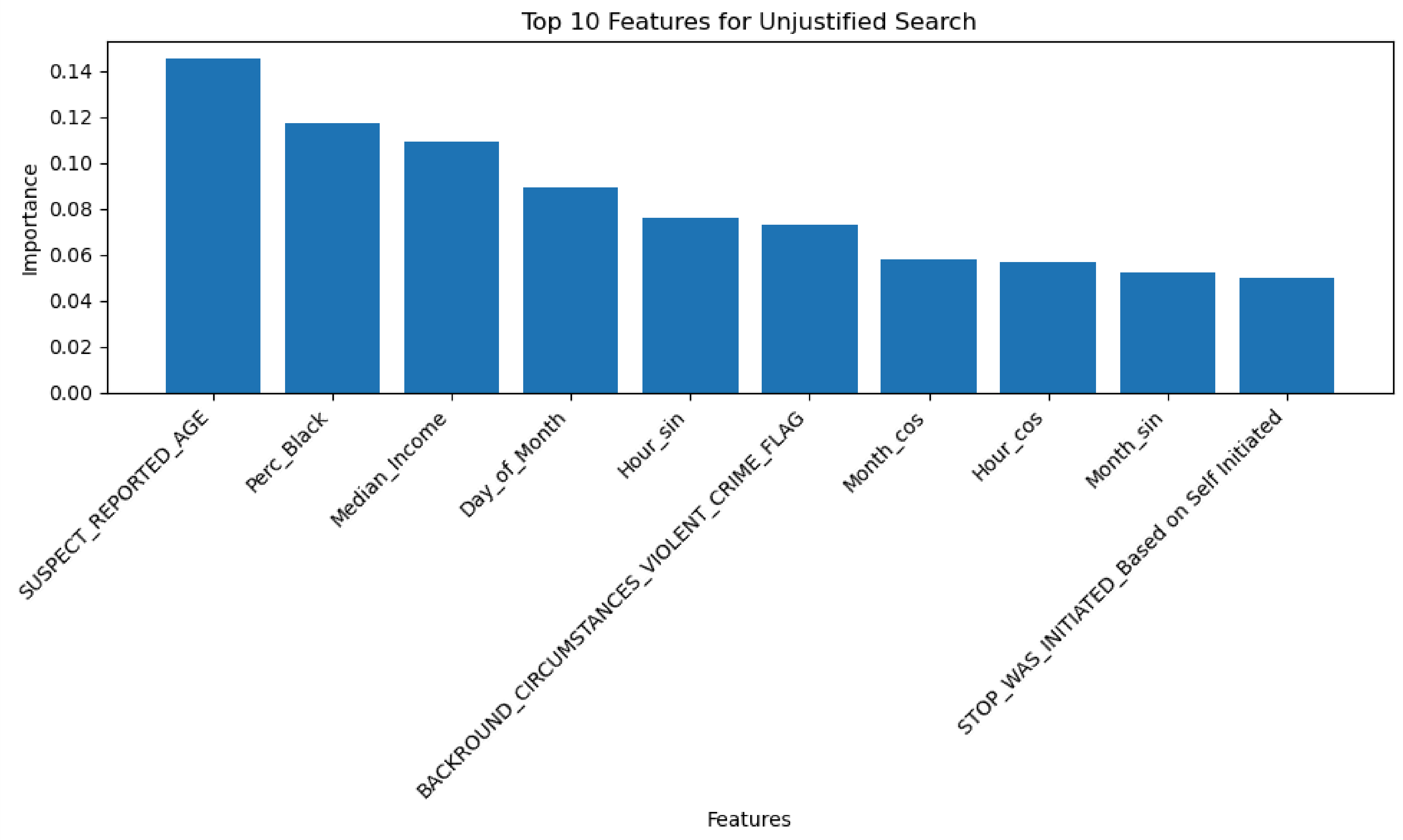
- **Top features:** SUSPECT_REPORTED_AGE, Median_Income, Perc_Black.
- **Age** is the most significant predictor of unjustified stops.
- **Median income** of the area also plays a crucial role.
- **Percentage of Black population** in the area is another important factor.
- **Time-related features** (Hour_sin, Hour_cos) are also influential.
- These findings suggest **potential age and racial biases** in stop decisions.
- **Economic factors** of the area significantly impact stop justification.
- Time of day affects the likelihood of an unjustified stop.

FEATURE IMPORTANCE: UNJUSTIFIED FRISK



- **Top features**: SUSPECT_REPORTED_AGE, Median_Income, STOP_WAS_INITIATED_Based on Self Initiated.
- **Age** remains the most significant predictor for unjustified frisks.
- **Median income** continues to be an important factor.
- **Self-initiated stops** are more likely to **result in unjustified frisks**.
- The importance of self-initiated stops suggests **potential officer bias.**
- Age and **economic factors** consistently influence police actions.
- **Day** of the week and **time** of day affect frisk justification.

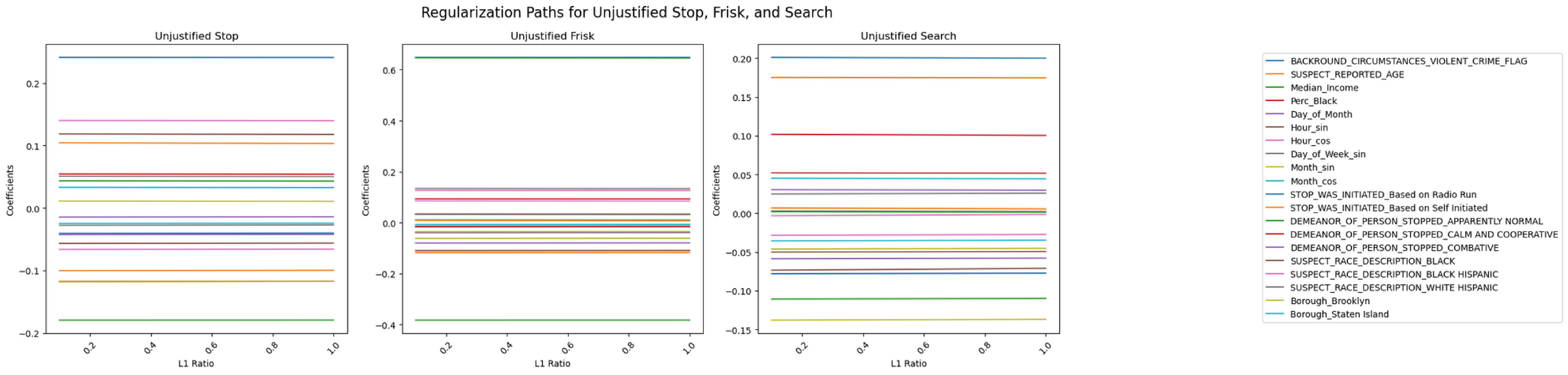
FEATURE IMPORTANCE: UNJUSTIFIED SEARCH



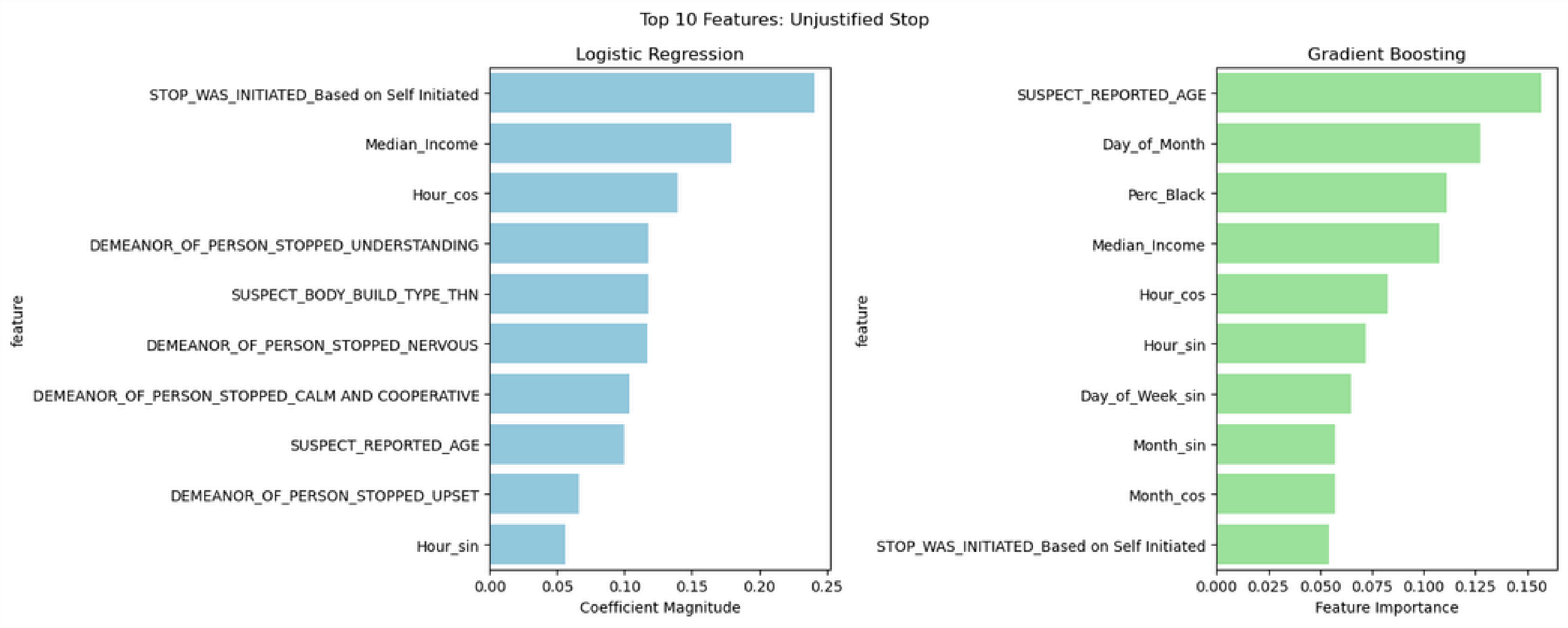
- **Top features:** SUSPECT_REPORTED_AGE, Median_Income, STOP_WAS_INITIATED_Based on Self Initiated.
- **Age** is consistently the top predictor across all three outcomes.
- **Median income** remains a significant factor in predicting unjustified searches.
- **Self-initiated stops** are more likely to lead to unjustified searches.
- Suspect's **race** (BLACK, BLACK HISPANIC) appears as important features.
- The consistency of age and income factors across all actions is noteworthy.
- **Racial factors** play a more prominent role in search decisions.
- Officer discretion (self-initiated stops) significantly impacts search justification.

FEATURE IMPORTANCE: COMPARISON

- **Coefficient stability varies significantly** across the three tasks, with Unjustified Search showing the most stable patterns.
- Most features maintain **consistent directional influence** (positive/negative) across different L1 ratios.
- Demographic and circumstantial variables show stronger regularization effects than temporal features.
- The stability of paths suggests good model robustness across different regularization strengths.
- Feature selection can be confidently performed based on coefficient magnitudes.



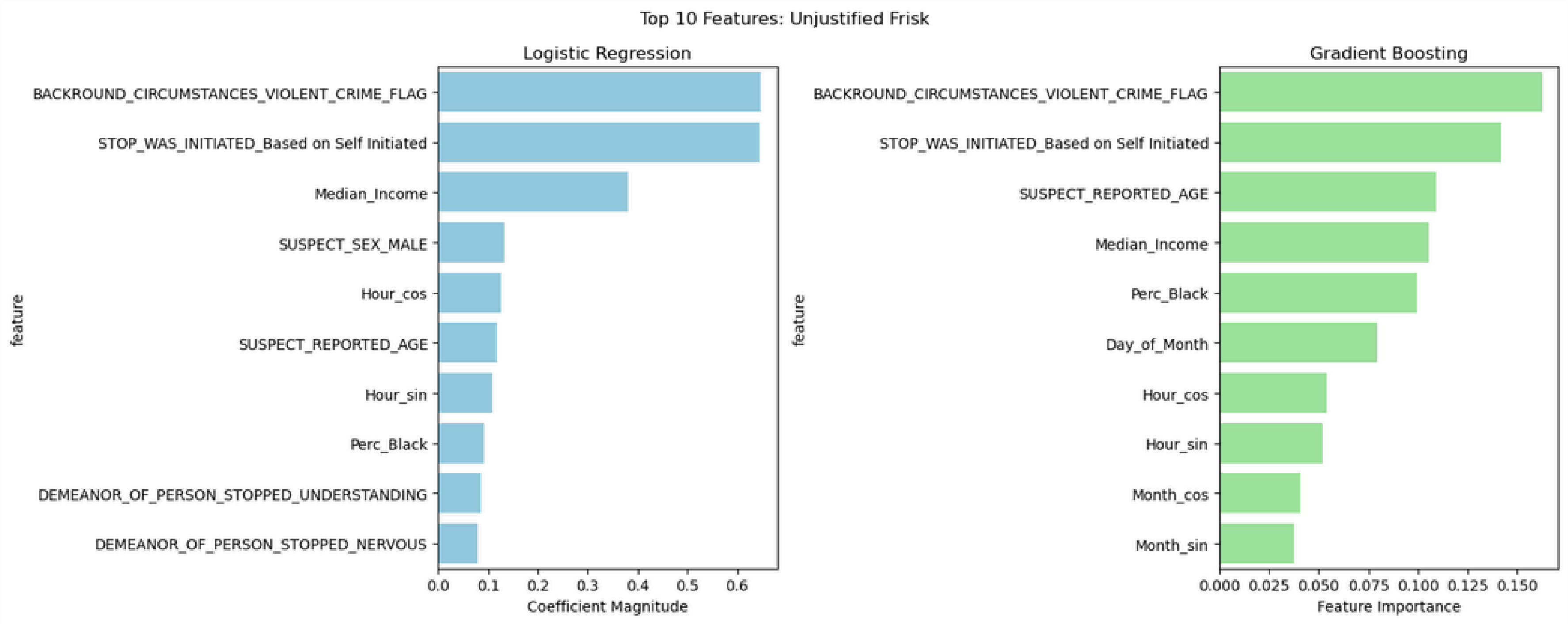
FEATURE IMPORTANCE: COMPARISON

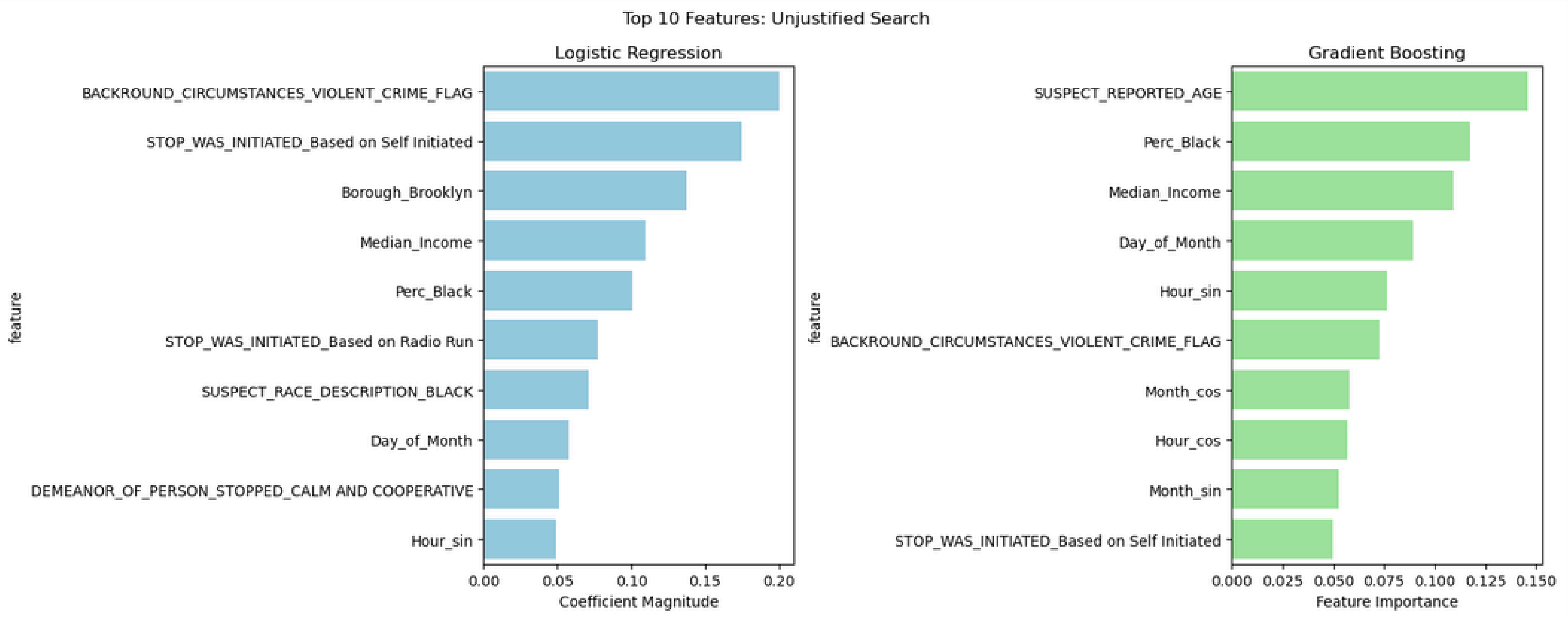


- **Unjustified Stop** models diverge in primary predictors: Logistic Regression emphasizes self-initiated stops (0.25 coefficient), while Gradient Boosting prioritizes suspect age and temporal patterns (0.15 importance score).

- **Unjustified Frisk** demonstrates violent crime flags and self-initiated stops as dominant features (0.6 coefficient) in both models, with demographic factors showing secondary influence.

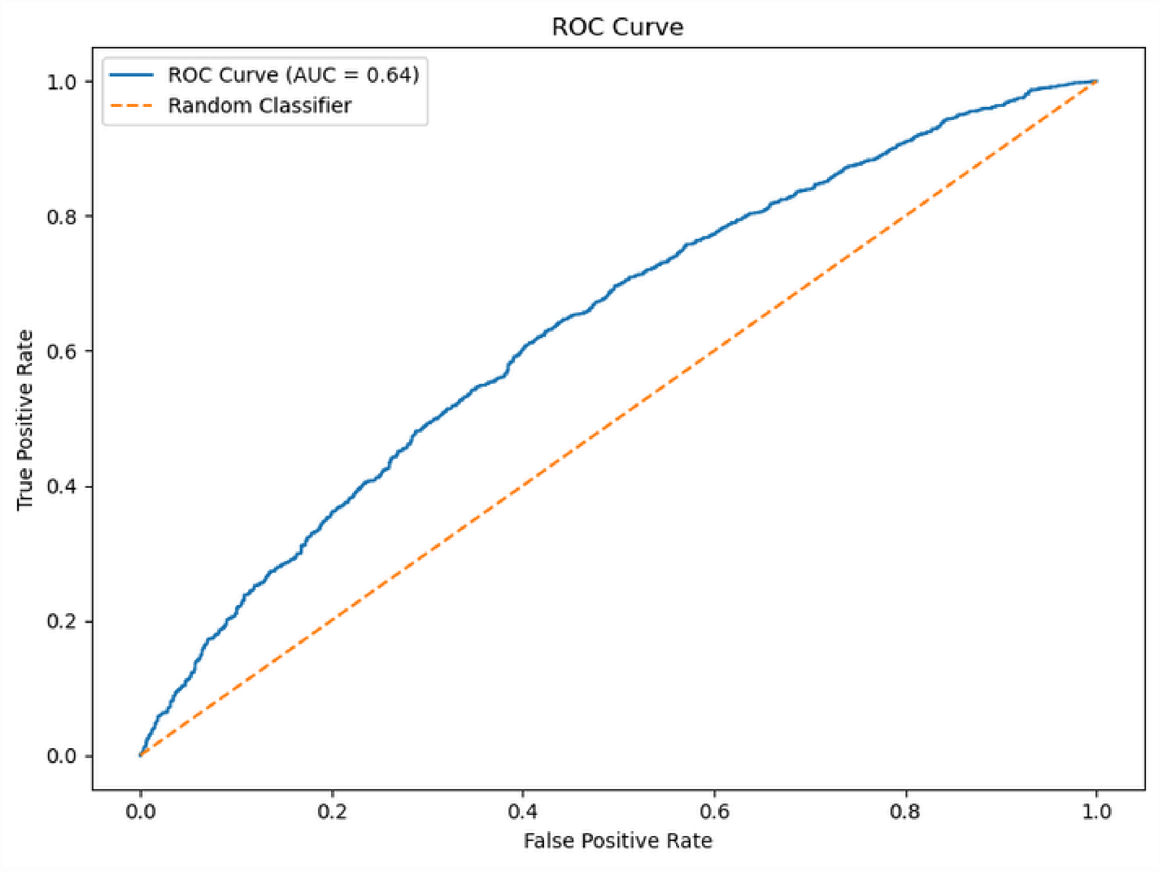
- **Unjustified Search** reveals model differences: Logistic Regression weights violent crime and self-initiated stops heavily, while Gradient Boosting emphasizes demographic factors (age, race) and socioeconomic indicators.

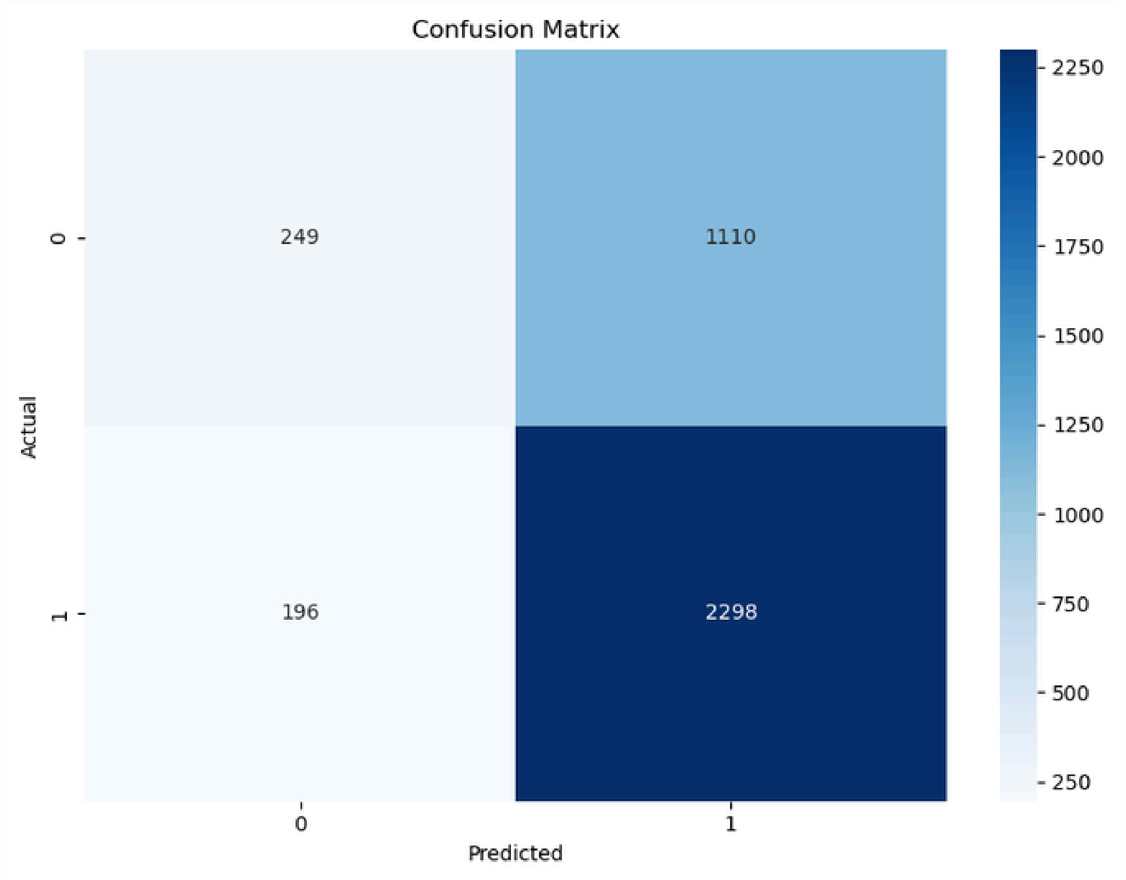




THE RESULTS

MODEL RESULTS:





- **Age** is the most crucial factor in predicting unjustified police actions.
- **Younger** individuals are more likely to experience **unjustified stops, frisks, and searches**.
- **Economic factors** (median income) significantly influence police behavior.
- **Lower-income** areas are more prone to **unjustified** police **actions**.
- **Self-initiated** stops have a higher likelihood of resulting in **unjustified actions**.
- **Race plays a role**, with Black and Hispanic individuals more likely to experience unjustified actions.
- Time and location factors consistently influence police decision-making.

The models reveal potential systemic biases in law enforcement practices.

BUSINESS IMPLICATION

IMPLICATIONS FOR LAW ENFORCEMENT AGENCIES

- Need for **age-sensitive training** to reduce bias against younger individuals.
- Implement **community-based policing strategies** in lower-income areas.
- Review and **revise protocols for self-initiated stops** to reduce unjustified actions.
- Enhance **diversity and inclusion training** to address racial biases.
- Develop data-driven performance metrics that **discourage unjustified actions**.
- Implement body-worn cameras and regular audits to **increase accountability**.
- Create specialized units for youth engagement and community relations.
- Establish an early warning system to **identify officers with patterns** of unjustified actions.

RECOMMENDATIONS FOR FUTURE RESEARCH

- Investigate the **long-term effects of unjustified police actions** on communities.
- Analyze the **effectiveness** of various **police training programs** on reducing biases.
- Study the impact of body-worn cameras on the frequency of unjustified actions.
- Examine the relationship between community policing initiatives and unjustified stops.
- Investigate the role of **implicit bias in self-initiated stops**.
- Conduct longitudinal studies to track **changes in police behavior over time**.
- Explore the intersection of age, race, and socioeconomic factors in police interactions.
- Assess the effectiveness of civilian oversight boards in reducing unjustified actions.

Thank you!