Uncovering Traffic Stop and Arrest Patterns with Data-Driven Insights and Predictive Modeling

Problem Statement

Terry Stops are a critical law enforcement practice, but only a small portion of stops result in arrests.

This raises concerns about efficiency, fairness, and the potential for bias in stop-and-arrest outcomes.

Our goal is to analyze stop-level data to identify the key factors that influence whether an arrest is made, and to build predictive models that can support more data-driven, ethical decision-making.

Objectives

- To identify patterns in traffic stops and arrest outcomes.
- To understand key factors that drive arrests.
- To build predictive models for arrest outcomes.
- To provide actionable recommendations.

Business Understanding

Traffic stops are a routine part of law enforcement, but patterns in who is stopped and arrested can reveal potential disparities. Understanding these patterns is critical for ensuring fair and equitable policing.

This project aims to analyze the Terry Traffic Stops dataset to identify trends and factors that influence arrest outcomes, such as race, gender, age, and stop context, and to provide actionable insights that help guide policy, training, and resource allocation decisions.

Data Understanding

The Terry Traffic Stops dataset has 64737 records and 23 features, with information on:

1. **Demographic features:** race, gender, age group.
2. **Stop features:** reason for stop, location, if the subject was frisked, weapon presence.
3. **Time features:** date and time of stop, reported year, and time of day.
4. **Outcome variable:** arrest_flag (whether an arrest was made).

Most of the features in the dataset were categorical, with only a small number of numerical features.

Some of the columns had inconsistent values, while there were some missing values in some columns as well.

Columns such as ‘reported_time’ and ‘reported_date’ were in incorrect data types.

These issues needed to be corrected to ensure the dataset was accurate and ready for analysis.

Data Preparation

The following steps were carried out to ensure that the dataset is clean:

- Removed unnecessary or redundant columns to focus on relevant information.
- Filled missing values to ensure completeness and reliability.
- Standardized column names for consistency across datasets.
- Converted data types where needed to make analysis and modeling possible.

This helped to ensure the dataset was clean, consistent, and ready for exploratory data analysis and predictive modeling.

Feature engineering was performed to create new variables and transform existing ones that helped to improve insights and supported more accurate analysis and modeling.

Data Analysis

Explored the dataset to identify patterns and trends in traffic stops and arrests.

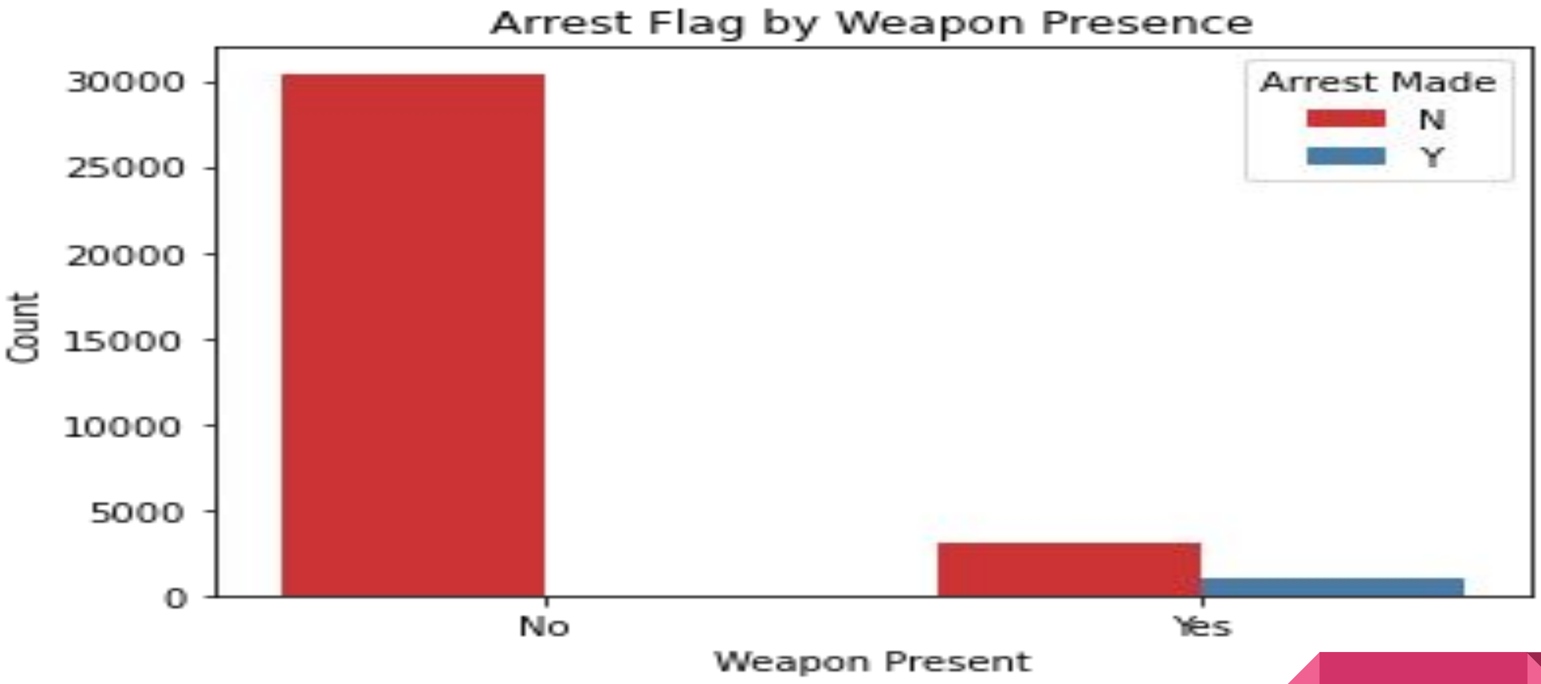
Examined the influence of demographics, stop context, location, and time on arrest outcomes.

Performed summary statistics, visualizations, and correlations to understand the data.

Insights from this analysis will inform predictive modeling and recommendations.

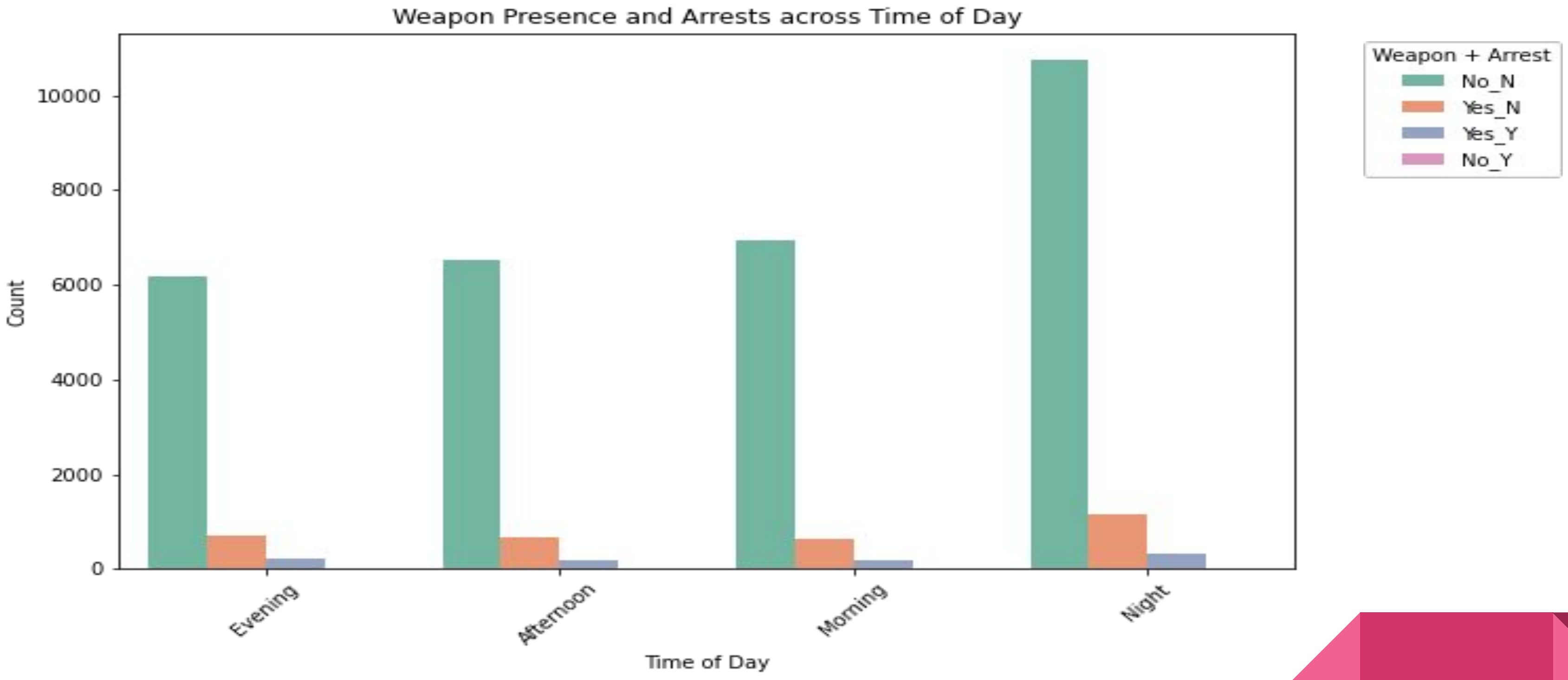
Key visualizations that supported our analysis are shown in the next slides.

Arrest Flag by Weapon Presence



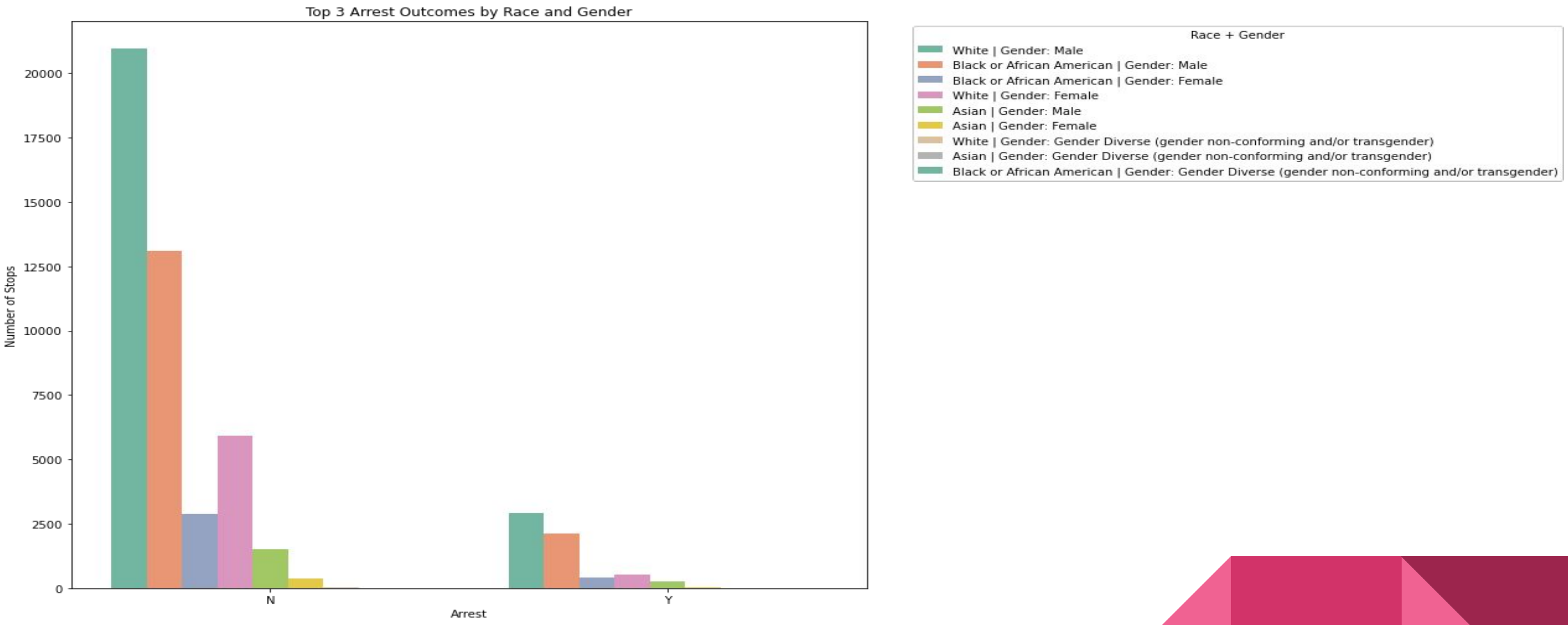
When no weapon was involved, most stops ended without an arrest, and the few arrests made were due to the presence of a weapon.

Weapon Presence and Arrests across Time of Day



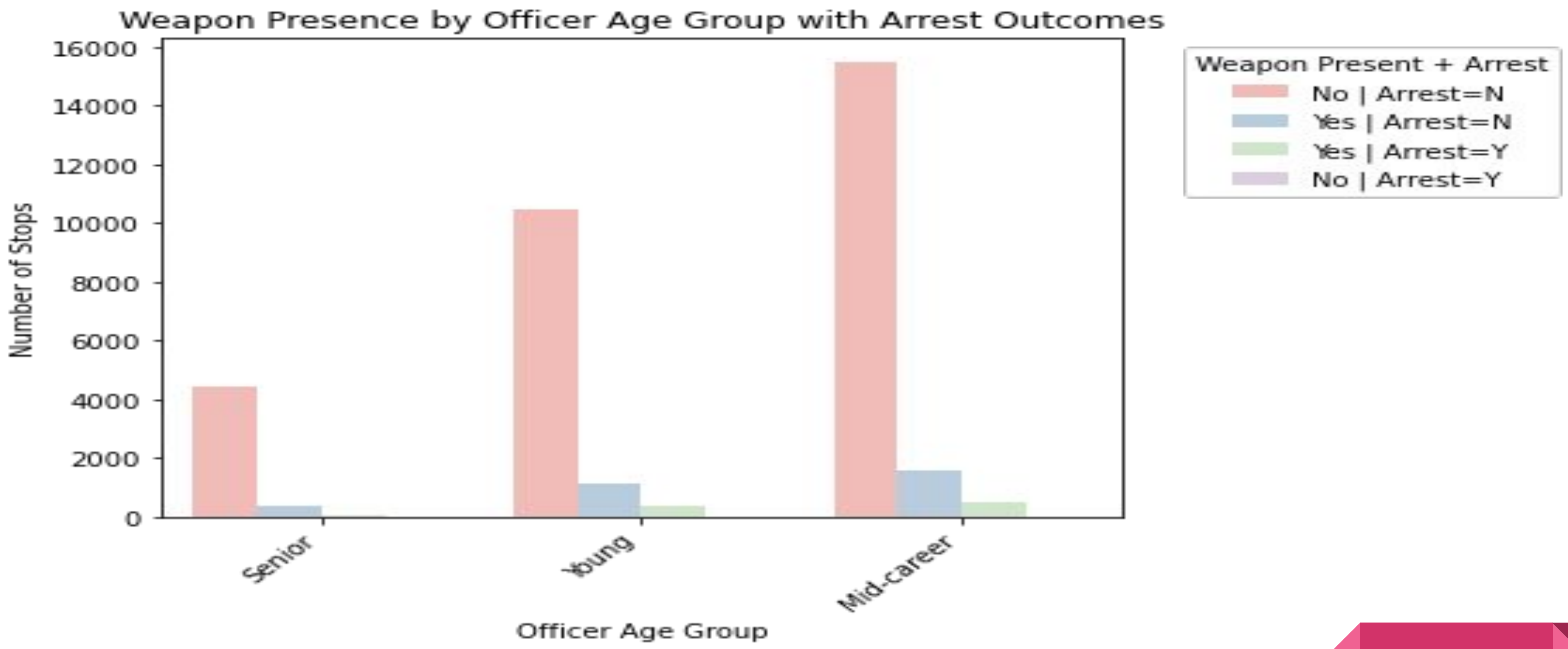
Most arrests happened at night when a weapon was present.

Top 3 Arrest Outcomes by Race and Gender



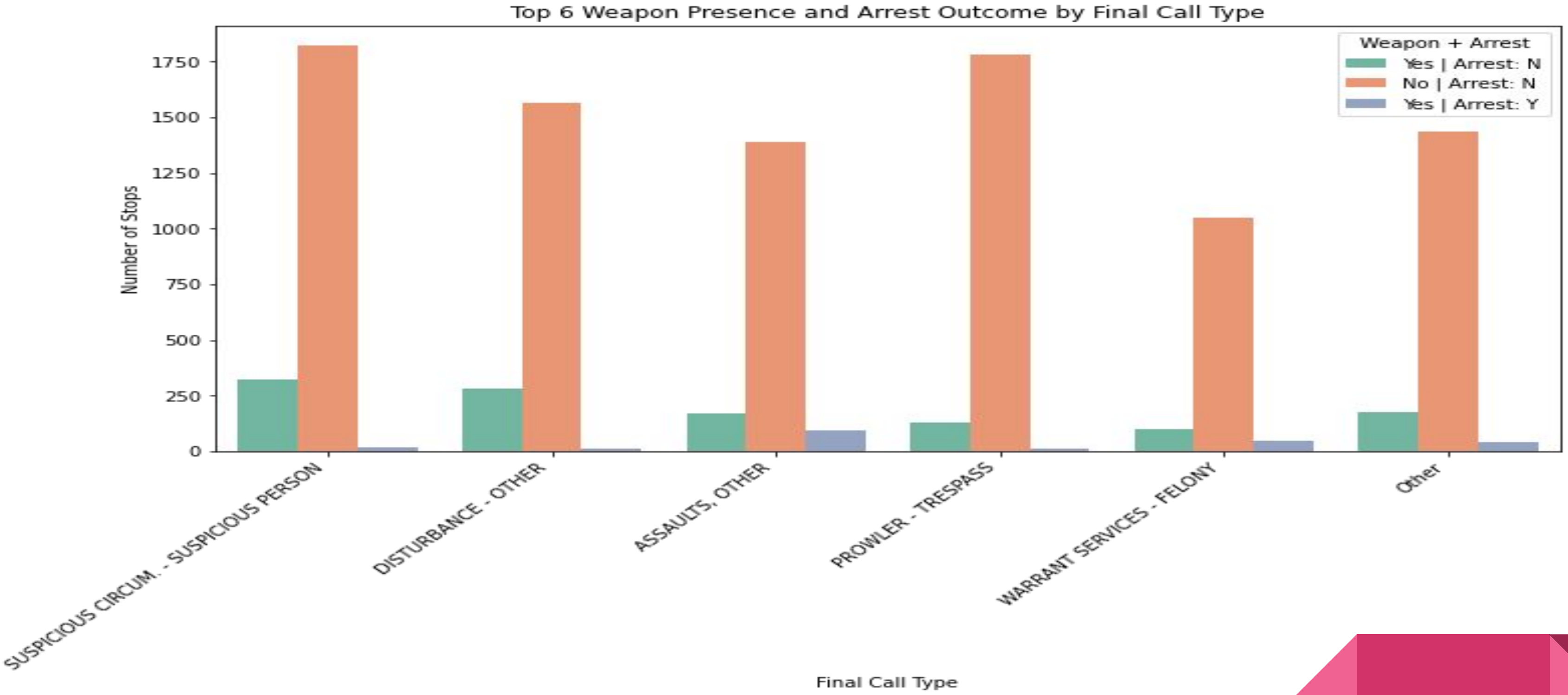
Most arrests were made up of white males, followed closely by Black or African American males

Weapon by Officer Age Group with Arrest Outcomes



Most arrests were made by mid-career officers in the presence of a weapon.

Top 6 Weapon Presence and Arrests by Final Call Type



Most arrests happened when a weapon was present and when the final call type was classified as Assaults.

Modeling

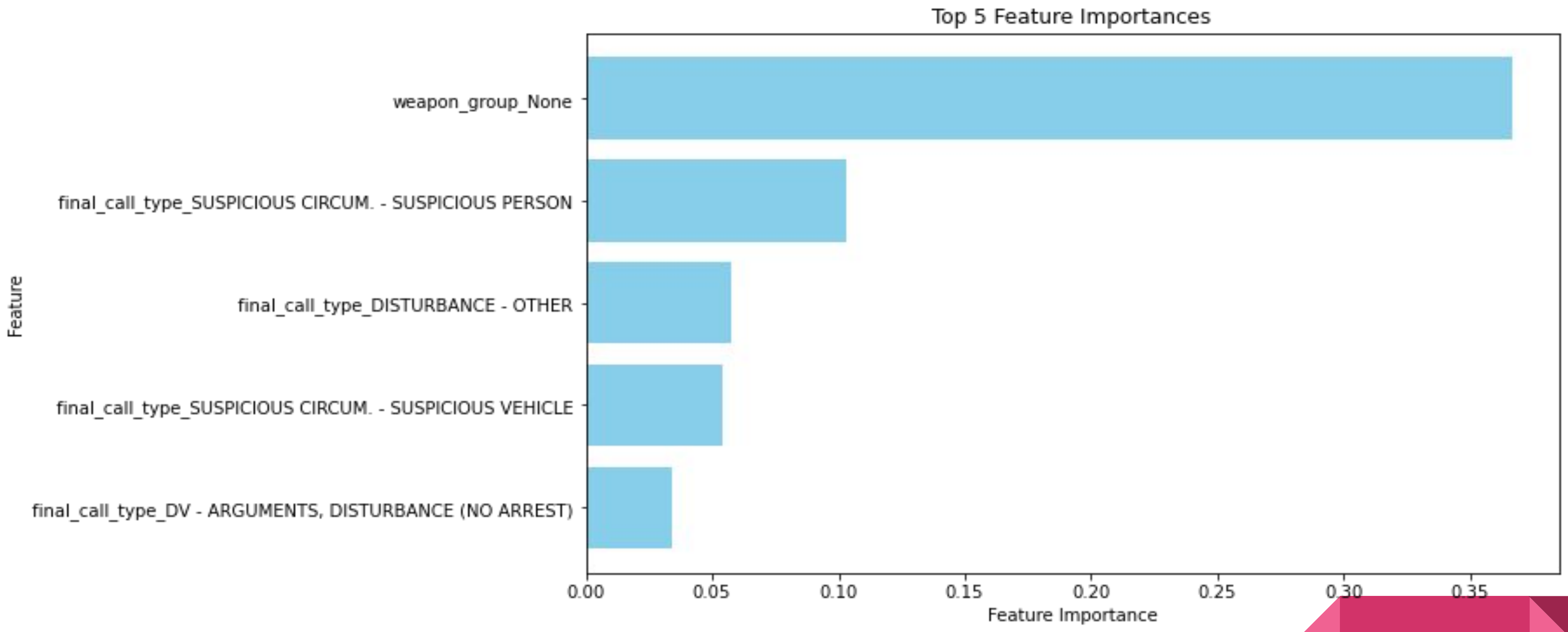
Sensitive variables such as race and gender, were tested but excluded from the final model to avoid bias. Race and gender exclusion showed minimal impact on performance, confirming fairness in predictions.

Weapon presence was one of the strongest predictors of arrest, and this is because stops involving weapons are higher-risk, making arrests more likely.

Built and compared four models, and out of those four the final model chosen was XGBoost, because it had strong performance across all metrics.

Final model provides reliable insights into arrest likelihood without relying on sensitive demographics.

Feature Importance



'weapon_group_None' has the highest importance hence it heavily influences the model's decisions.

Conclusion

1. The majority of Terry Stops involved male subjects, and most arrests were concentrated among White and Black/African American males.
2. Most Terry Stops did not result in an arrest, which suggests that the majority of stops end without escalation.
3. Subject perceived race and gender showed disparities in stop frequency, even though they were excluded from the final model to avoid bias.
4. Stop context like suspicion of weapons or specific behaviors, proved to be more informative for predicting arrests than demographic characteristics.
5. Frisk requests were common but did not always lead to arrests, showing that frisks are not a perfect predictor of criminal activity.

Recommendations

1. Given that most stops don't end in arrests, policy should aim to minimize unnecessary stops that may erode community trust.
2. Focus training on handling high-risk stop situations like the weapon-related ones, since they strongly drive arrest outcomes.
3. Officers and policymakers should focus more on situational factors like weapons and behaviors, rather than personal demographics in decision-making.
4. Even though race and gender weren't used in the final model, disparities in stop rates suggest a need for regular fairness audits.
5. Since many frisks do not result in arrests, refine frisk criteria to improve efficiency and reduce unnecessary searches.

Data-driven analysis can support fairer, evidence-based policing strategies.

THANK YOU!

Questions?