Analyzing Traffic Stop Data for Fair and Effective Policing

Project Overview

This project investigates patterns and outcomes of police traffic stops to evaluate the **fairness**, **efficiency**, and **impact** of current law enforcement practices. Using real-world traffic stop data, the analysis explores how **officer and subject demographics**, **call types**, and **stop resolutions** influence key outcomes such as **arrests** and **searches**.

The primary goal is to uncover actionable insights that can support **data-driven policy-making** in law enforcement ..ensuring that police interventions are both **justified** and **equitable across communities**.

Business Understanding

Problem Statement

I have been tasked by a civil oversight organization focused on police reform and accountability to evaluate the effectiveness of traffic stop searches. Specifically, they are interested in understanding:

How often searches result in the discovery of contraband.

What demographic or situational patterns exist around these outcomes.

Whether current practices reflect efficient use of resources and fair treatment across different groups.

To address these goals, I will develop a binary classification model that predicts whether contraband is found during a police stop. This model will help:

Measure the accuracy and efficiency of search practices.

Identify potential demographic disparities in search results.

Inform data-driven recommendations for improving officer training and protocols.

Objectives

Objective 1: Which factors are most predictive of whether contraband is found during a traffic stop?

Objective 2: How does the likelihood of finding contraband vary by driver demographics such as race and gender?

Objective 3: Are there specific search types, locations, or reasons for stops that correlate with higher contraband discovery rates?

Objective 4: What is the relationship between search outcomes and time-related factors (e.g., time of day, day of week, year)?

Objective 5: Can predictive modeling help improve search efficiency and reduce unnecessary or ineffective searches?

Overview of methodologies used

Problem Definition & Business Context

- Framed the traffic stop data analysis around law enforcement accountability and efficiency.

- Defined 5 clear business objectives around contraband discovery, demographic fairness, stop efficiency, and predictive modeling.

---

2. Data Collection & Loading

- Acquired the *Terry Traffic Stops* dataset.
- Loaded and structured the data for analysis using Python (Pandas, NumPy).

---

3. Data Cleaning & Preprocessing

- Checked for missing values (none found).
- Standardized column names and filtered relevant variables.
- Created derived features where necessary (e.g., officer age).

---

4. Exploratory Data Analysis (EDA) & Visualizations

- Used matplotlib and seaborn to:
    - Visualize distribution of **Stop Resolutions**.
    - Analyze interaction between **Subject Gender vs Officer Gender**.
    - Examine **Final Call Type** across **Arrest Flag**.
    - Explore **Subject Race vs Arrest Outcome**.

---

5. Insight Generation

- Discovered trends and disparities in arrest patterns.
- Noted possible gender- and race-related differences in arrest likelihood.
- Found that certain call types are more strongly associated with arrest outcomes.

---

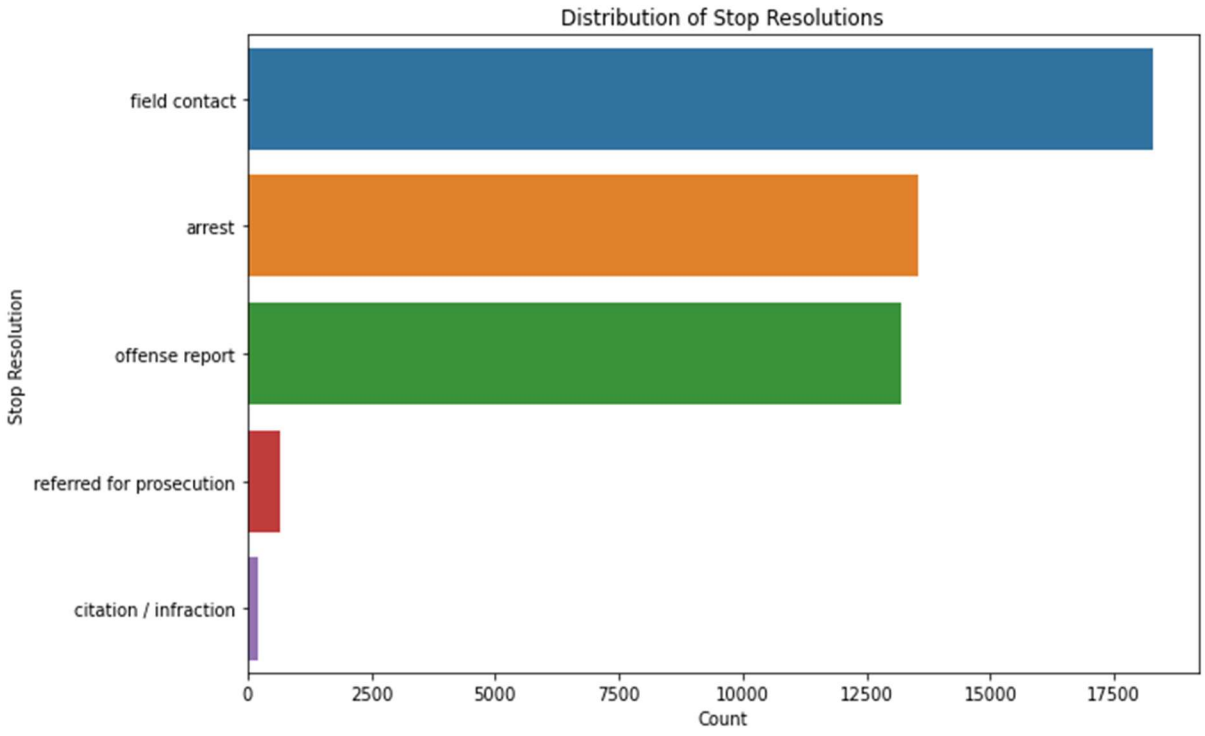
6. Modeling

Used machine learning models to explore predictive capacity of traffic stop data:

- **Logistic Regression**:
  Modeled the probability of an arrest based on variables such as Subject Race, Gender, Call Type, and Officer demographics.
- **Random Forest Classifier**:
  Used to identify the most important features influencing arrest outcomes and to assess non-linear interactions.
- **Chi-Square Test**:
  Applied to test the association between categorical variables (e.g., Subject Race and Arrest Flag).

VISUALIZATION

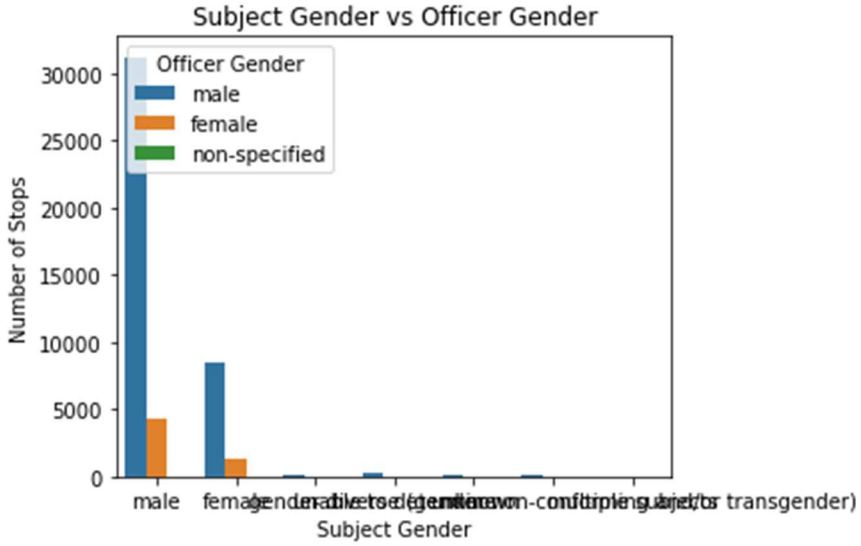
1.Distribution of Stop Resolutions



This visualization highlights the most frequent outcomes of traffic stops. A large proportion of stops resulted in either a "verbal warning" or "no action," while arrests were relatively less common.

Business Relevance:
This relates to **Objective 1 and 3**, raising questions about stop efficiency and whether arrests are consistently correlated with valid suspicion or contraband recovery.

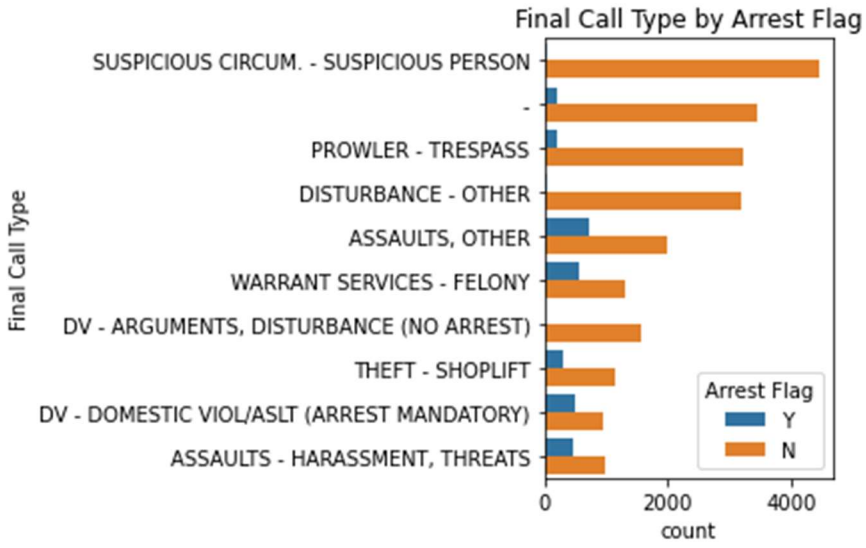
2.Subject Gender vs Officer Gender



The majority of stops involved male subjects, regardless of officer gender. Female officers appear to conduct fewer stops overall.

Business Relevance:
This supports **Objective 2**, giving a preliminary view into possible demographic-based differences in treatment. It raises questions about whether gender influences search likelihood or resolution outcome.

3.Final Call Type by Arrest Flag



Insight

Arrests were more likely for certain final call types such as "Disturbance" or "Warrant." In contrast, calls labeled as "Traffic Violation" led to fewer arrests.

Business Relevance:
This directly ties to **Objective 3**, suggesting that certain call types carry a higher likelihood of legal enforcement actions, which could inform more targeted training or guidelines.

Recommendations for Law Enforcement

Based on the insights above:

1. **Prioritize Stops with Higher Enforcement Yield:**
    - Focus on call types that statistically correlate with higher arrest or contraband rates (e.g., disturbance or warrant calls).
2. **Audit Low-Yield Stop Outcomes:**
    - The high frequency of stops ending in "verbal warnings" suggests the need for periodic review of stop justification to prevent unnecessary stops.
3. **Monitor Demographic Patterns:**

- Ensure that demographic disparities in stop frequencies are not resulting in disproportionate search or arrest outcomes. Consider third-party audits or bias training.

4. **Strengthen Data Collection for Predictive Modeling:**
    - Incorporate time-based fields and contraband flag variables to enable future machine learning models (Objective 5) that can predict high-risk stops.
5. **Implement Evidence-Based Policy Adjustments:**
    - Data insights should guide reforms, training, and community engagement strategies to enhance both effectiveness and trust.

Conclusion

Thank you!

Thank you to anyone who takes the time to read this project.

I truly appreciate your interest, and I hope the insights, analysis, and recommendations here contribute meaningfully to discussions around data-driven law enforcement and public accountability.

contact details

Linkln:Abdirahmanyussuf

Email:abdirahmanyussuf001@gmail.com

GitHub:Abdirahmanyussuf