Terry Stop Legal Analysis and Prediction

Project Overview

The goal of this project is to build a machine learning model that predicts whether an arrest was made following a Terry Stop.

This Photo by Unknown Author is licensed under CC BY-SA-NC

Business Problem

- Analysis of the landmark 1968 U.S. Supreme Court case Terry v. Ohio, which established the legal standard of "reasonable suspicion" and introduced the concept of Terry Stops.

- While originally intended as a crime prevention tool, Terry Stops have increasingly been criticized for enabling **racial profiling**, **biased policing**, and in some cases, contributing to **police brutality.**

- Using data from police department records, this project analyzes the circumstances surrounding Terry Stops such as the individual's **gender, and age,** to build a predictive model that determines whether an **arrest occurred** following a stop.

- This is framed as a **binary classification problem**, focusing on two outcomes: *arrested* or *not arrested*.

Data Understanding

This project aims to analyze and evaluate the use of Terry Stops by addressing the following goals:

- Examine whether factors like gender, age, and reason for the stop influence the likelihood of an arrest.

- -Build a machine learning model that balances accuracy with fairness and justice.

- - Identify potential patterns or biases in stop-and-arrest decisions.

- - Investigate the presence of disparities in arrests based on demographic variables.

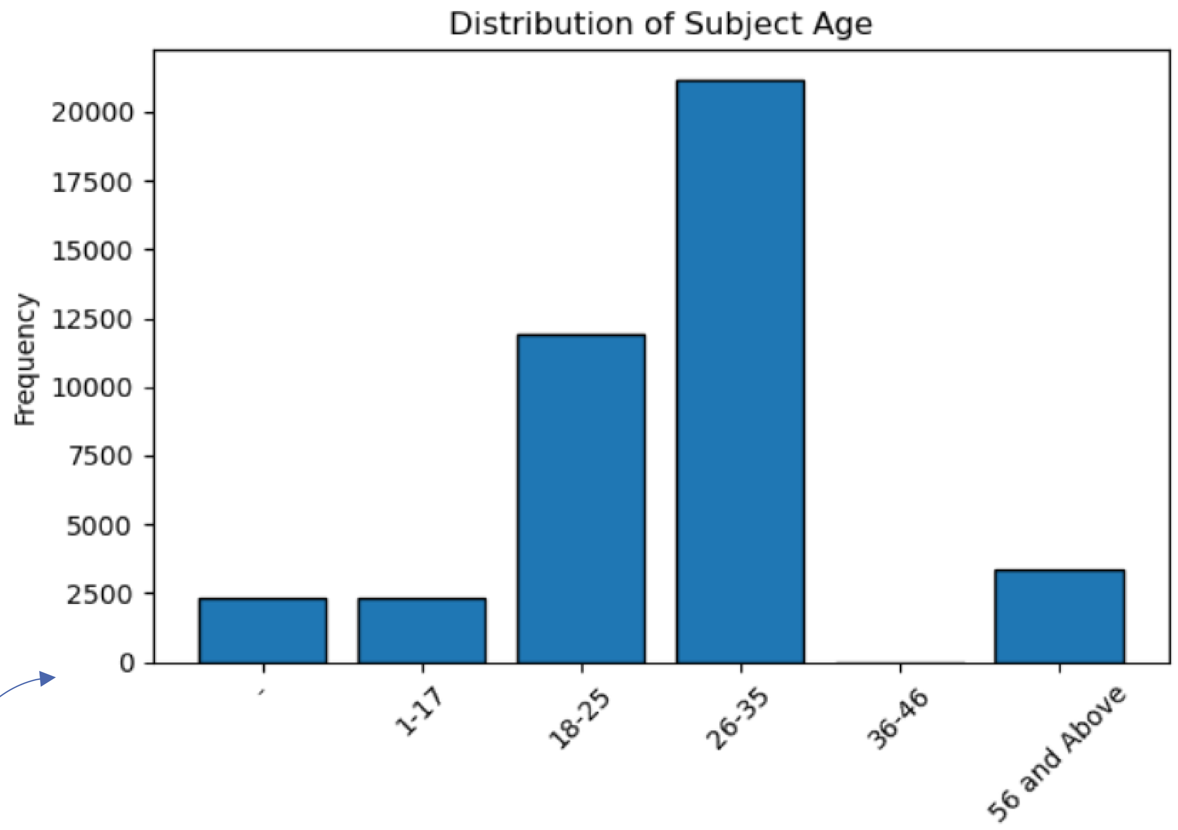
The dataset contains 63461 records with 8 categorical features, such as 'Arrest Made' (Yes/No). Most data types are objects, with only one date at Officer YOB and three other IDs.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 63462 entries, 0 to 63461
Data columns (total 23 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   Subject Age Group    63462 non-null  object
 1   Subject ID           63462 non-null  object
 2   GO / SC Num          63462 non-null  object
 3   Terry Stop ID        63462 non-null  object
 4   Stop Resolution      63462 non-null  object
 5   Weapon Type          63462 non-null  object
 6   Officer ID           63462 non-null  object
 7   Officer YOB          63462 non-null  datetime64[ns]
 8   Officer Gender       63462 non-null  object
```

Data Preparation and Preprocessing

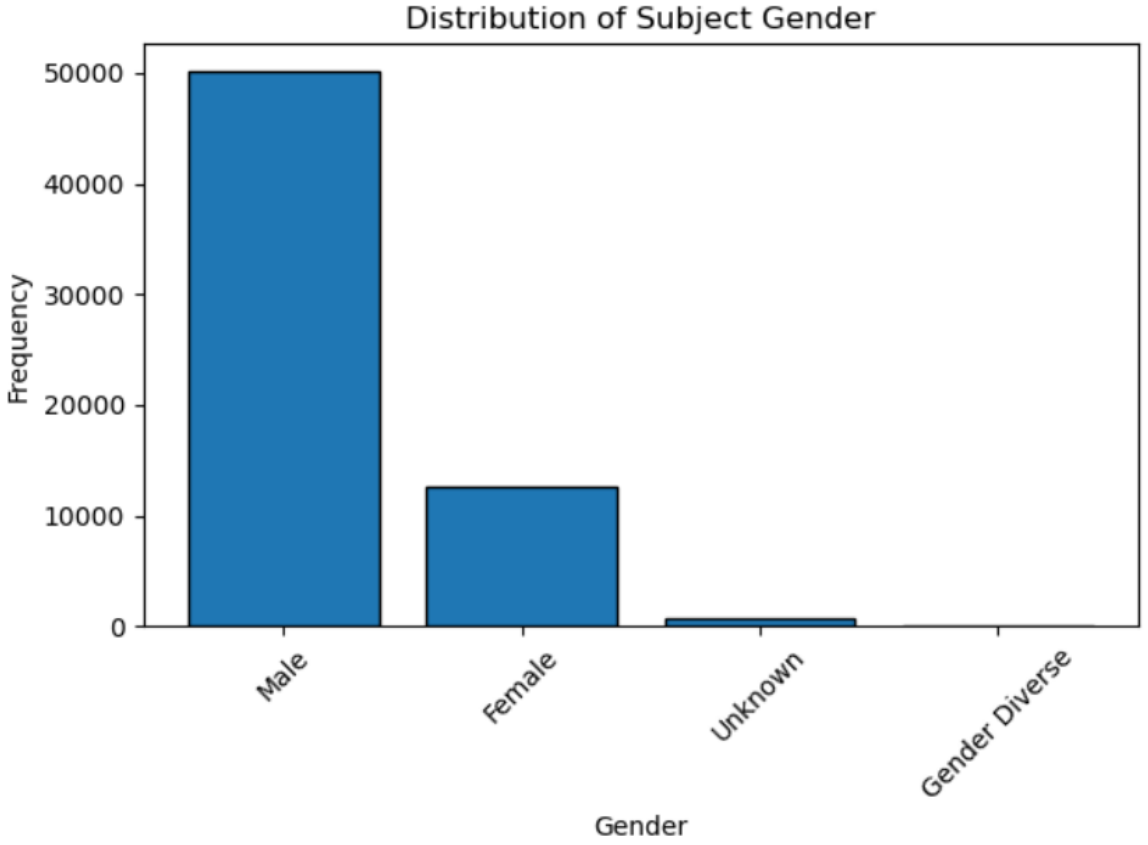
- Checked for null values, duplicate rows, and corrected data types for consistency.- cleaning of categorical variables that would then be used as features e g Subject age

- Explored feature distributions to gain better understanding:- The most common age group with the highest number of stops is 26–35 years old.

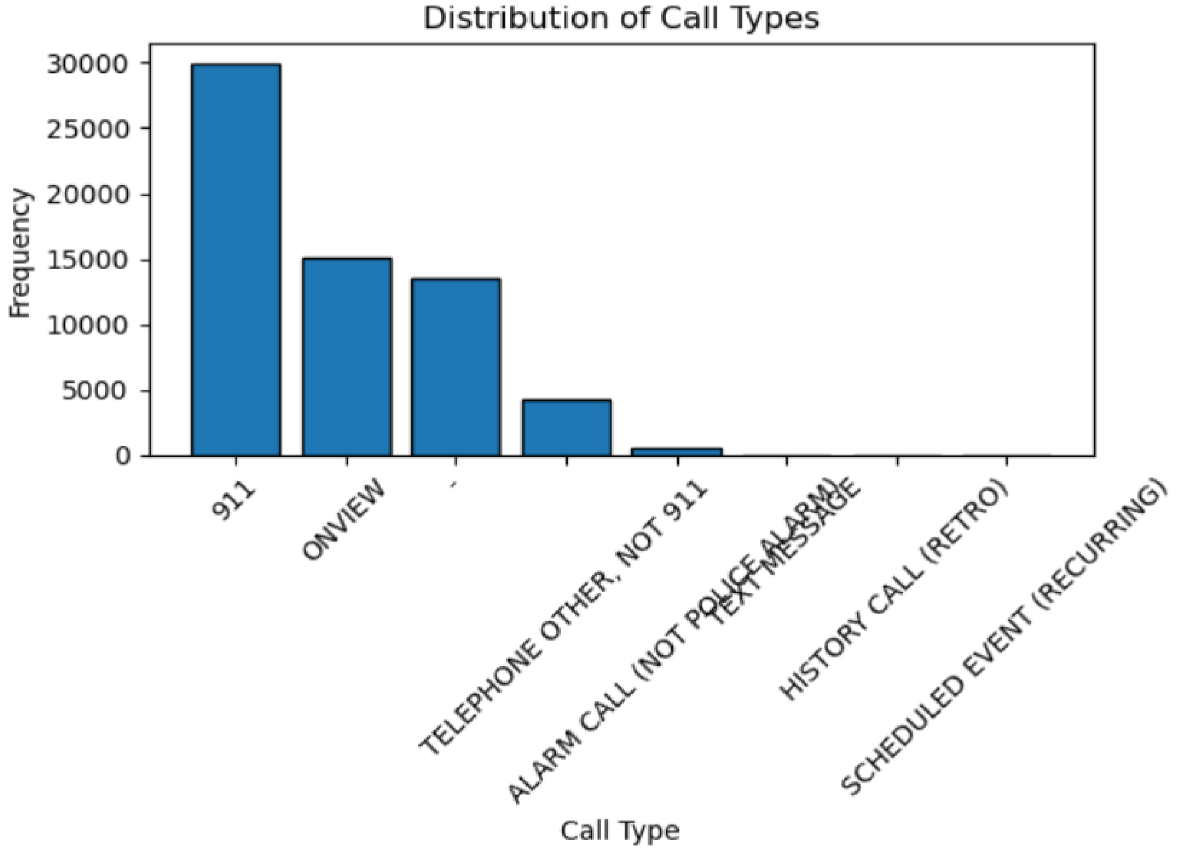


Data Preparation and Preprocessing cont.

The majority of stops involved males.



The most frequent origin of stops was 911 calls, followed by on-view observations by officers.



Modeling : Logistic Regression

- I used this model because it's well-suited for binary classification problems where there are only two possible outcomes, like whether an arrest was made (1) or not (0).
- The model looks at features like age, gender, and location, and then estimates the probability of an arrest happening.
- Based on that probability, it predicts either a 1 or a 0. This helps us uncover patterns in the Terry Stop data and understand what factors might influence an arrest.

Logistic Regression Evaluation

The results show:

The precision for Not arrested is at 0.95 but a low for Arrested at 0.15. This means the model identifies those who are not Arrested accurately but fails to identify those that have been Arrested.

The recall is at a low of 0.43 for Not Arrested while Arrested is at 0.83. Meaning that The model identifies the majority of actual arrests well but predicts that too many will be arrested.

The overall accuracy is 0.48, which is very low. But this is because of class imbalance in the dataset as logically many stops do not end in Arrests.

However, we focus on Recall in the Arrested calls as the community and police force should be making sure few real arrests are not being missed

	precision	recall	f1-score	support
Not Arrested	0.95	0.43	0.59	11286
Arrested	0.15	0.83	0.26	1407
accuracy			0.48	12693
macro avg	0.55	0.63	0.43	12693
weighted avg	0.86	0.48	0.56	12693

Modeling : Decision Tree Classifier

- I used this model because it's well-suited for binary classification problems where there are only two possible outcomes, like whether an arrest was made (1) or not (0).
- The model looks at features like age, gender, and location, and then estimates the probability of an arrest happening.
- Based on that probability, it predicts either a 1 or a 0. This helps us uncover patterns in the Terry Stop data and understand what factors might influence an arrest.

Decision Tree Classifier Evaluation

The results show:

A precision for Not Arrested at 0.95 and low Arrested at 0.15. Meaning, with a very low precision, the model wrongly predicts arrests for people who were not arrested, which is dangerous.

A recall for Not Arrested at 0.43 and Arrested at 0.83. This means the model catches actual arrests very well.

The accuracy score of 0.48 is also low, and the model is failing to identify those who are arrested overall.

Here we focus on Precision for Arrested, as we do not want to wrongly label a person as Arrested.

	precision	recall	f1-score	support
Not Arrested	0.95	0.43	0.59	11286
Arrested	0.15	0.83	0.26	1407
accuracy			0.48	12693
macro avg	0.55	0.63	0.43	12693
weighted avg	0.86	0.48	0.56	12693

Recommendations

The model would be useful in:

- Reviewing and updating Policy: This could help police departments understand which types of stops most often lead to arrests and investigate patterns.
- To detect bias: To look at areas where Arrests are made and flag harmful patterns that can be later on discussed.
- In Training and Awareness: The data and model can be used to show officers or the community where certain factors may unintentionally influence bias; this will help in promoting more informed decision making.

Conclusions

- The dataset was highly imbalanced, with far more "Not Arrested" cases, which is expected given the nature of police work in Chicago.
- In predictive policing, high recall ensures potential arrests are identified, while high precision prevents false labels, helping to reduce bias.
- The final model chosen is the Logistic Regression model as it was better able to better handle the imbalance of the Terry stop data, decision trees have clear splits based on feature values, which is not as flexible in adjusting for imbalance.
- The model performed well on "Not Arrested" cases but struggled with "Arrested" cases, leading to many false positives and negatives. This indicates both Type I and Type II errors, with Type II being more critical.
- These results reflect the challenge of predicting rare events and suggest a need for better features, model tuning, or fairness-focused methods.

Thank you

Any Questions?

Michelle Chekwoti