PHASE 3 PROJECT: Terry Stops Analysis

JESSE GITOBU

Business Understanding

- The goal of this project is to predict whether an arrest was made following a Terry Stop. This is a binary classification problem, where the target variable is `Arrest Flag`. By building this model, we aim to identify the factors that contribute to arrests during such stops and to explore potential disparities related to race, gender, or other variables.
- Key Considerations:
- Ethical Implications: This project uses sensitive information such as race and gender. Transparency in analysis and handling these features responsibly are critical.
- Bias Awareness: Models and analyses must be evaluated for potential biases that could perpetuate systemic inequities.

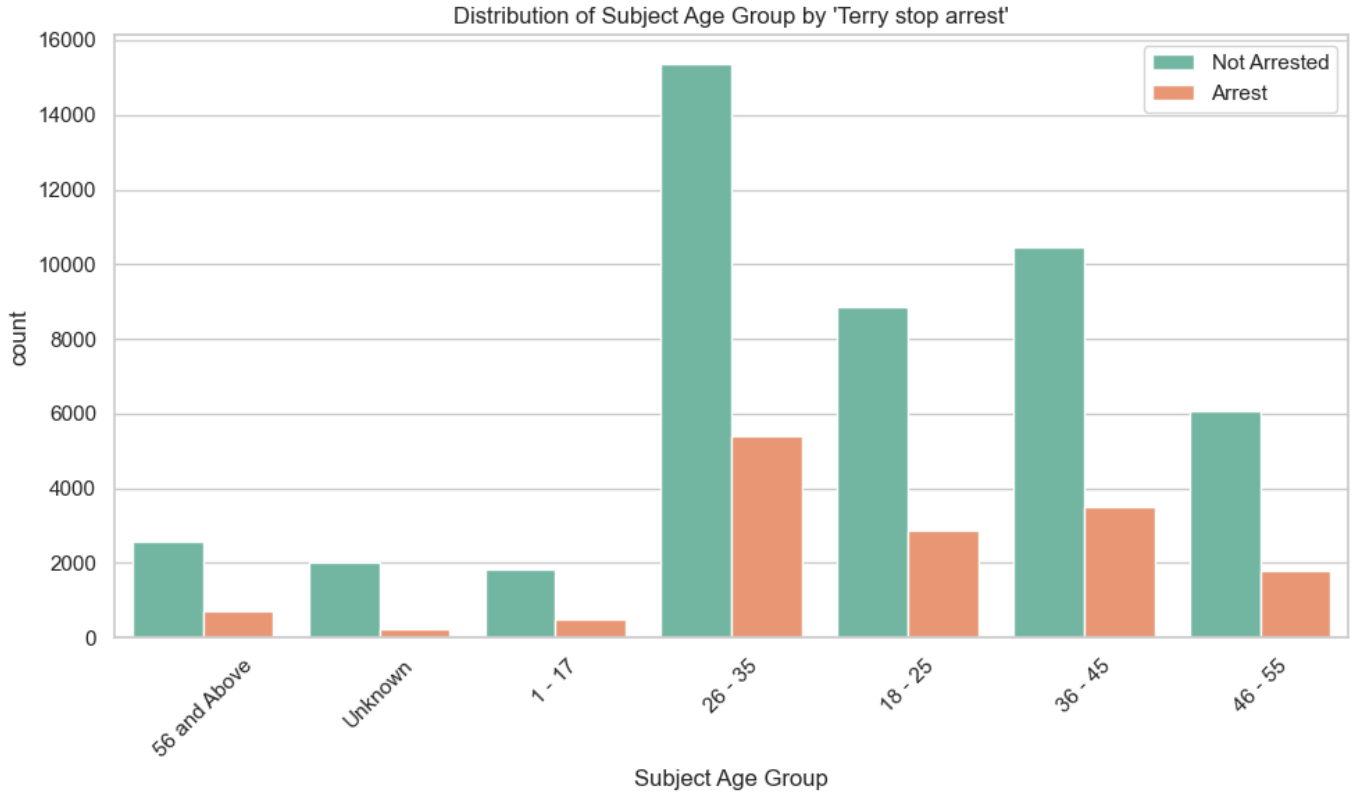
Key Questions:

- 1. What features (e.g., time of day, weapon presence, officer demographics) are most predictive of arrests?

Data understanding

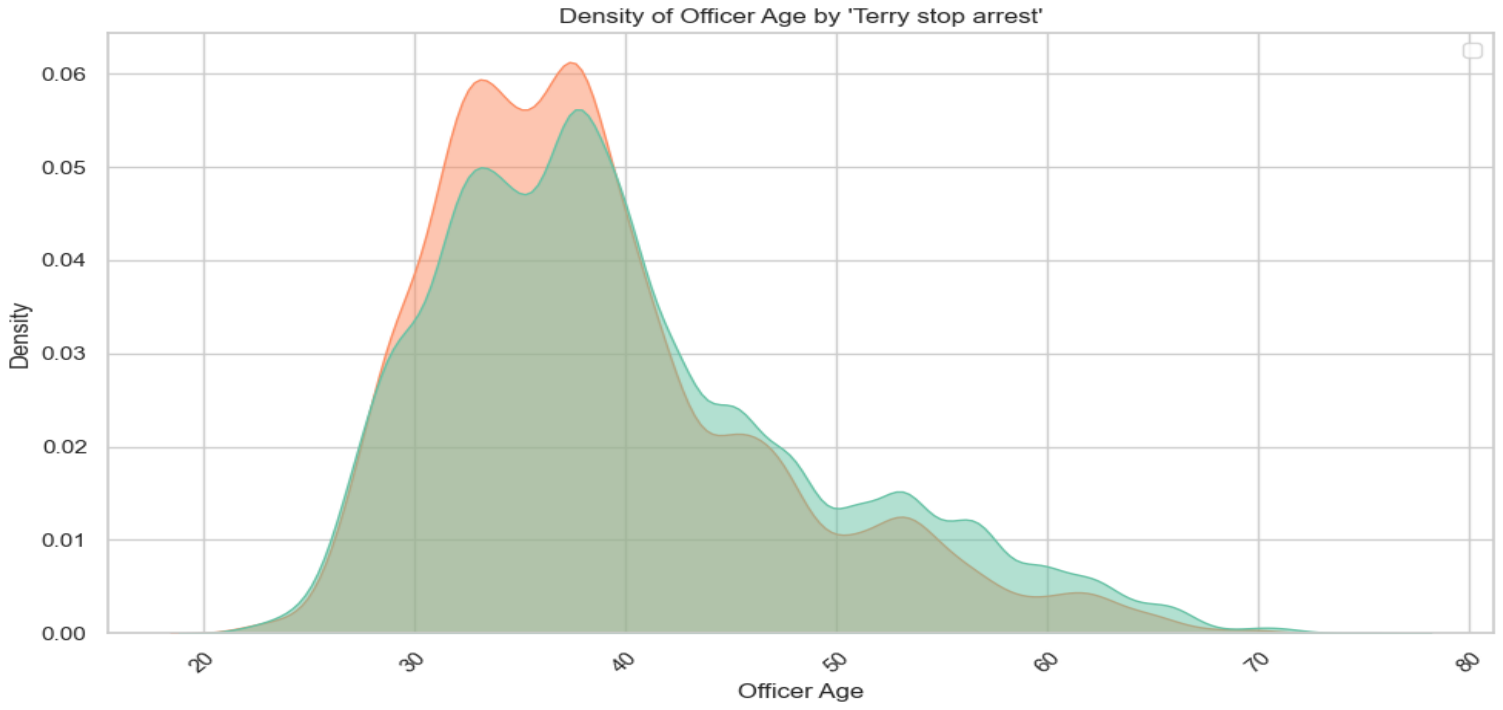
- File Name: Terry_Stops_20241220.csv
  - Rows: 62,191
  - Columns: 23
- Key Features:
  - Demographics (age group, gender, perceived race)
  - Officer attributes (gender, race, year of birth)
  - Stop context (call type, weapon type, arrest and frisk flags)
- Target Variable:
  - Arrest Flag (Binary): 'Y' (Yes) or 'N' (No)
- Class Distribution:
  - No Arrest (N): 89.01%
  - Arrest (Y): 10.99%

Exploratory Data analysis



We can see majority of thearreststst we made on subjects who were aged 26-35

Distribution of Officer's Age by the arrest



Majority of the arrest were made by officer whose age was about 30 to 40 years.

Modelling

1. **Logistic Regression**
    - **Confusion Matrix:**
    - Predicted well for non-arrest cases (True Negatives: 14,043) but struggled significantly with arrest cases (True Positives: 195).
    - High false negative rate indicates difficulty in identifying arrests.
    - **Metrics:**
    - Precision (Arrest): 0.53
    - Recall (Arrest): 0.04
    - F1-Score (Arrest): 0.08
    - Accuracy: 76.3%
    - Macro Avg F1-Score: 0.47
    - **Insights:**
    - The model is biased towards predicting non-arrest outcomes. Despite high overall accuracy (driven by non-arrest predictions), its ability to detect arrests is inadequate.

Modelling

2. **Decision Tree**
   - **Confusion Matrix:**
   - Almost perfect classification for non-arrest cases (True Negatives: 14,206), but nearly no capability to classify arrests correctly (True Positives: 5).
   - False negatives dominate arrest cases.
   - **Metrics:**
   - Precision (Arrest): 0.42
   - Recall (Arrest): 0.00
   - F1-Score (Arrest): 0.00
   - Accuracy: 76.2%
   - Macro Avg F1-Score: 0.43
   - **Insights:**
   - The Decision Tree model has very low utility for predicting arrests. Its focus on maximizing accuracy for non-arrest cases leads to a failure to identify arrests.

Modelling

3. **Random Forest**
   - **Confusion Matrix:**
   - Improved balance between non-arrest (True Negatives: 13,790) and arrest predictions (True Positives: 644). Reduction in false negatives for arrest cases compared to Logistic Regression and Decision Tree.
   - **Metrics:**
   - Precision (Arrest): 0.60
   - Recall (Arrest): 0.14
   - F1-Score (Arrest): 0.23
   - Accuracy: 77.4%
   - Macro Avg F1-Score: 0.55

Recommendations for Stakeholders

- **Focus Areas:**
  - Use results to address fairness in Terry Stops, particularly addressing race and gender biases.
- **Deployment Considerations:**
  - Test model robustness with unseen data before implementation.
- **Policy Implications:**
  - Develop interventions based on identified predictors (e.g., targeted training for officers).

Next Steps

- Evaluate additional algorithms (e.g., XGBoost, Neural Networks) for performance improvements.
- Perform interpretability analysis to make the model outputs more transparent.
- Expand EDA with time-series analysis to capture trends over time.