Predicting Arrests During Terry Stops: A Data-Driven Approach

Understanding Terry Stops

- Terry Stops are critical law enforcement tools utilized by the Seattle Police Department to maintain public safety.
- These stops involve questioning and potentially frisking individuals based on reasonable suspicion of criminal activity.
- However, their effectiveness and fairness continue to face scrutiny, especially concerning bias and arrest rates.
- This analysis seeks to shed light on these concerns by developing a predictive model for arrest outcomes. By exploring historical data, we aim to identify key factors influencing arrest likelihood during these stops.

Major Objective: Important Features in Predicting Arrests

By identifying key features, we can inform training and policy adjustments within SPD.

Minor Objectives:

- **Building a Predictive Model**
- **EDA & Feature Engineering**
- **Tackle Class Imbalance**

3

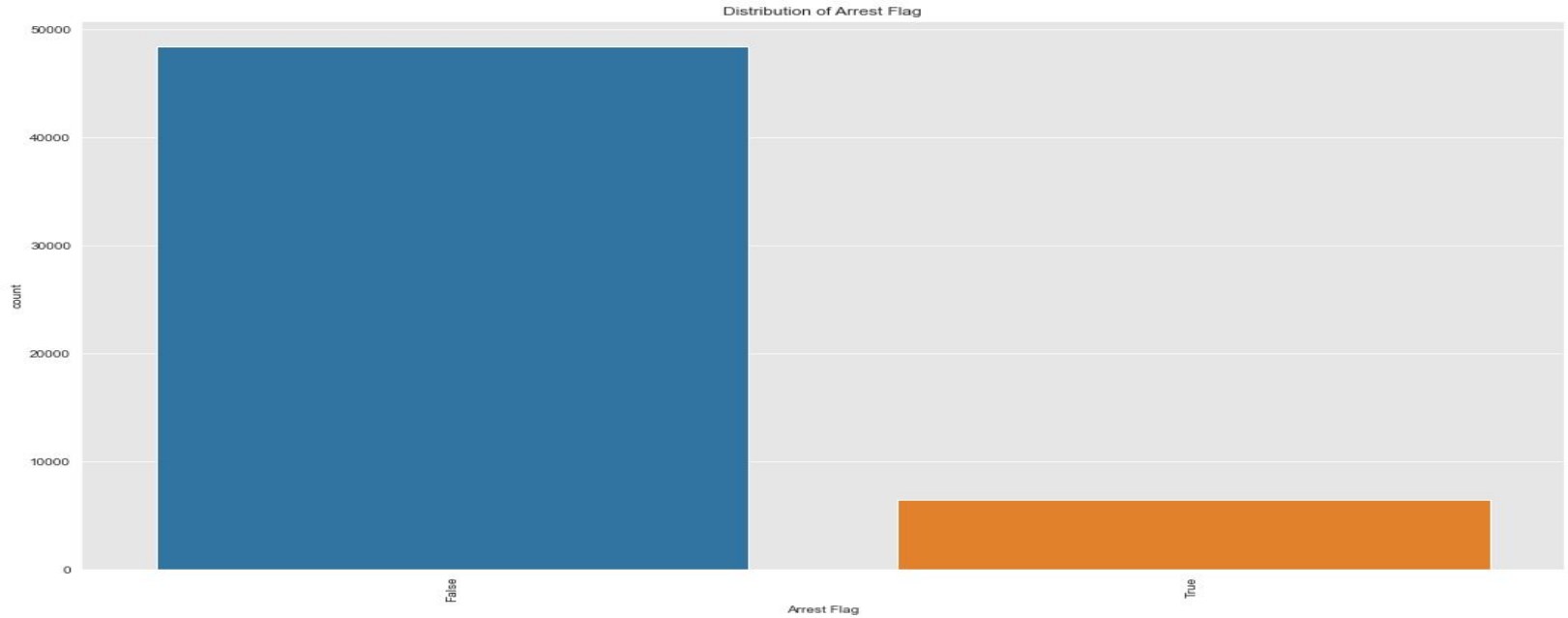
Exploratory Data Analysis & Feature Engineering

- Some of our features have high cardinality. We used target encoding which replaces each category with the mean of the target variable for that category, to reduce the dimensionality of our data

- We also used one hot encoding for our low cardinality features to prepare our data for logistic regression.

Target Variable Exhibits Class Imbalance

Class imbalance would result in a model that more commonly predicts No Arrests, which is the most common class in our dataset. We address this using SMOTE which creates



Baseline Model

```
Baseline Model Accuracy: 0.1184

Baseline Model Classification Report:
              precision    recall  f1-score   support

       False       0.00      0.00      0.00     24185
        True       0.12      1.00      0.21      3249
```

- Our baseline model which always predicts the positive class (Arrests) shows low accuracy in correctly predicting arrests.
- We use logistic regression and class imbalance handling techniques to improve on it

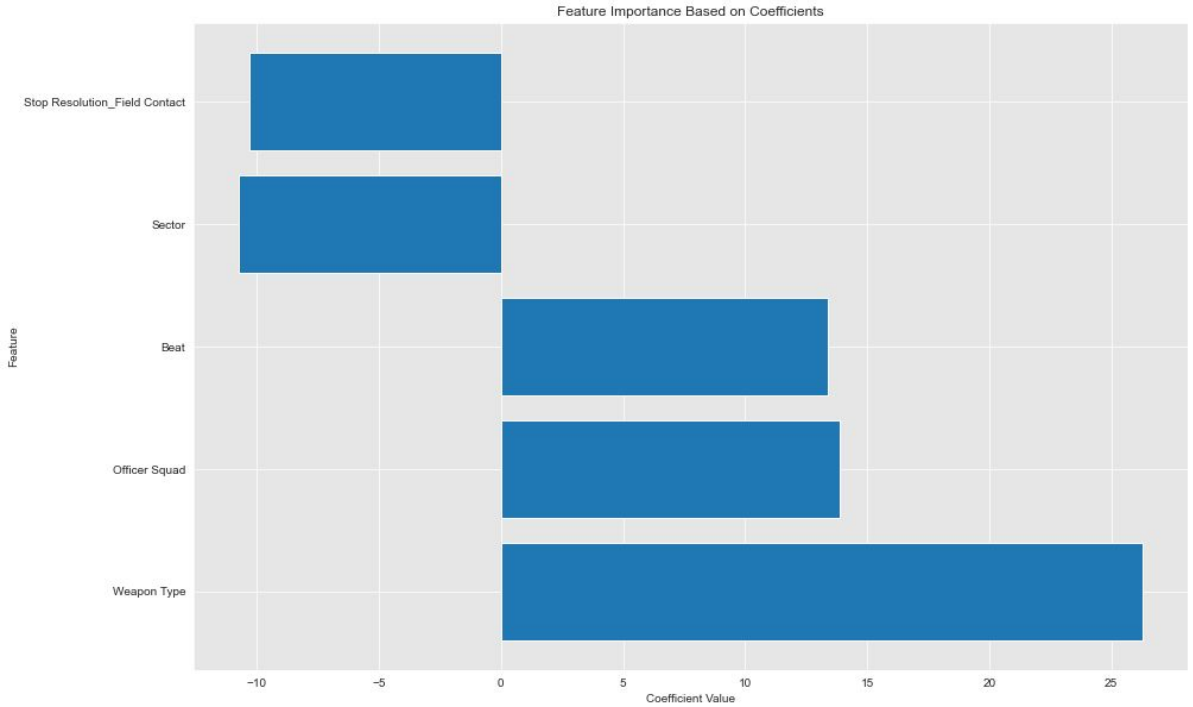
Final Model

```
Model accuracy score: 0.9166
              precision    recall  f1-score   support

       False       0.97      0.93      0.95     24185
        True       0.61      0.81      0.70      3249
```

- Our model accuracy improves significantly from 11.8% to 91.7% showing that we can rely on it to accurately predict Arrests vs our baseline model.
- We have to make a trade off between precision (correctly identifying arrests) and recall (actual arrests vs false positives) in order to improve our overall model accuracy.

Most Important Aspects in Predicting Arrests



Recommendations

- Use Model to Improve Law Enforcement Practices
- Enhance Training and Protocols for Weapon-Related Stops.
- Evaluate and Optimize Squad Practices.
- Adjust Resource Allocation Based on Beats and Sectors.
- Review and Standardize Field Contact Procedures.