Analysis of Arrests

By Dylan Garsee

The Terry Stop

In the 1968 Supreme Court case "Terry v. Ohio", the court found that a police officer was not in violation of the "unreasonable search and seizure" clause of the Fourth Amendment after he stopped and frisked suspects only because their behavior was suspicious. Thus the phrase "Terry Stops" are in reference to stops made of suspicious drivers.

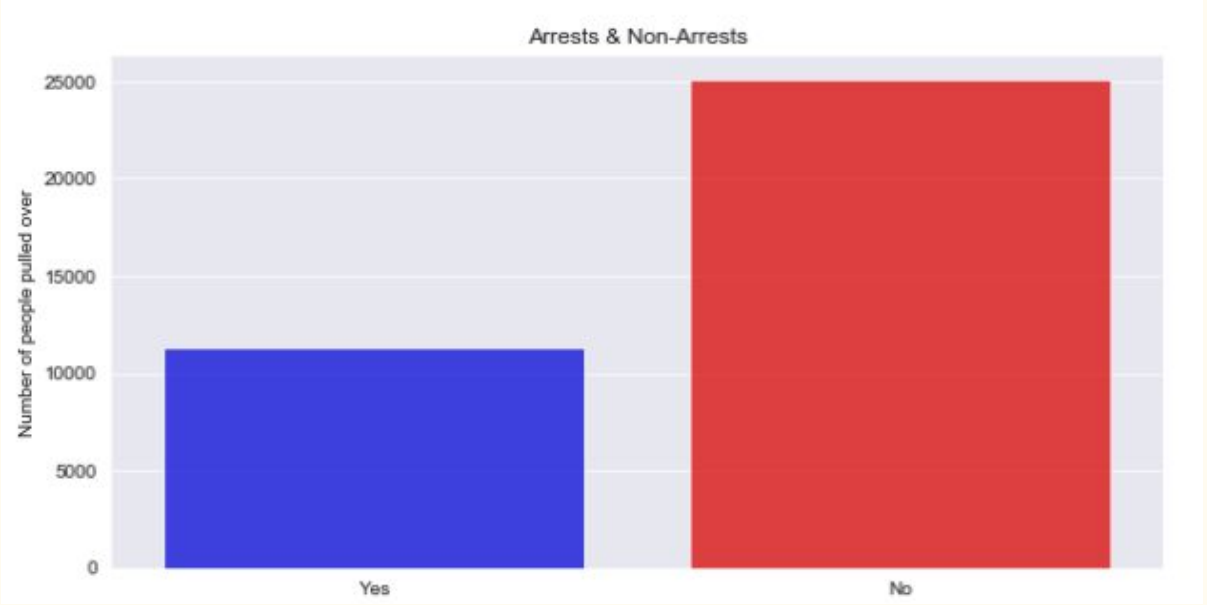
Goals of Analysis

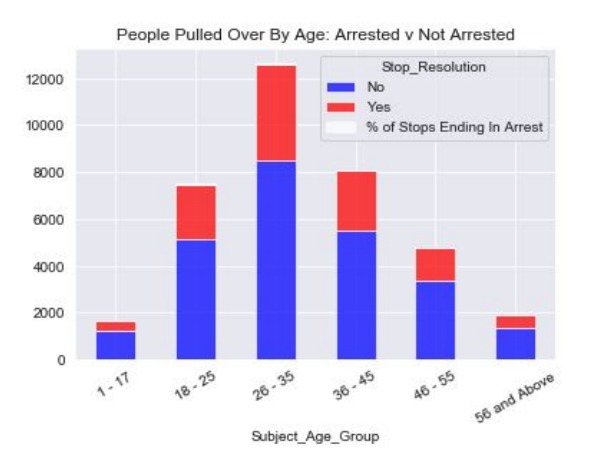
This is an analysis of over 36,000 Terry Stops, with a goal of predicting if an arrest will be made based off time of day, whether a suspect was frisked, and racial & gender demographics of both the suspects and officers.

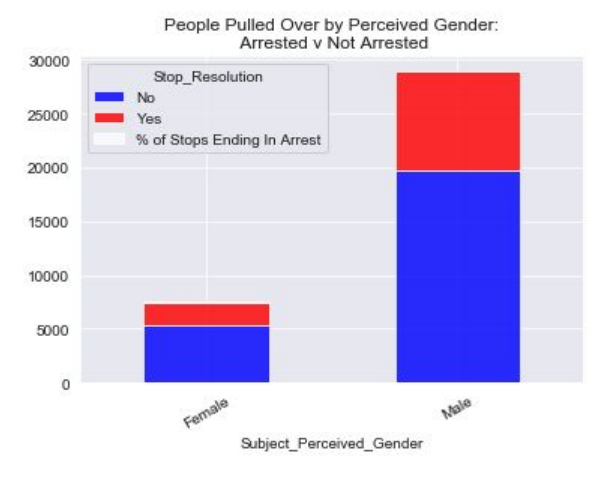
The overall goal of the analysis is to have the highest possible accuracy, to be able to predict an arrest after a Terry Stop.

Initial Analysis

The dataset used in this analysis had over 48,000 entries of Terry Stop data. After removing duplicate and unnecessary values, 36363 Terry Stops were used in analysis. Of the 36363 entries, 11,278 of the stops ended in with an arrest of the subject, or 31% of total stops.







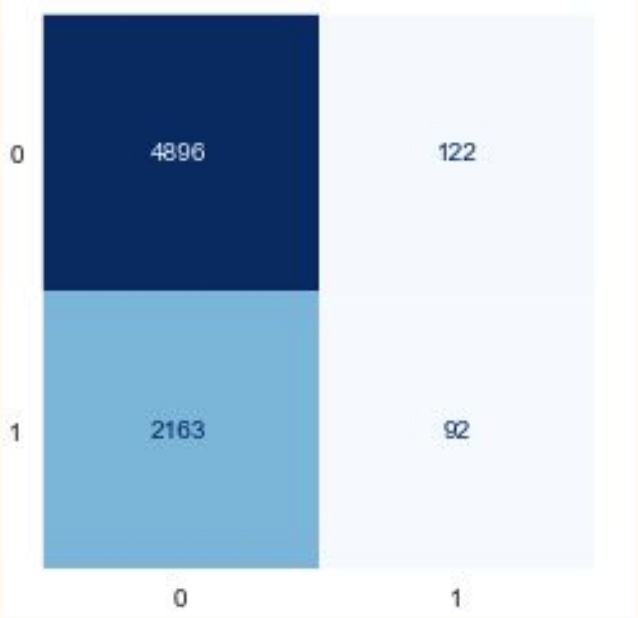
Initial Analysis

Of the age groups of subjects pulled over, 26-35 were the most pulled over. Nearly one-third of stops of subject in this age range resulted in an arrest, as opposed to the 25%-30% of the other age groups recorded.

In terms of gender divide, those perceived to be male were stopped nearly 4 times as often as those presenting female. Those presenting male were arrested 31% of the time while female presenting were arrested at a rate of 28%.

Modeling

Four different predictive models were used in this analysis, starting with Logistic Regression. This basic regression model returned an accuracy of nearly 54%. Upon further analysis, the biggest challenge this model had was labeling subjects who were arrested as not arrested.



Final modeling

While subsequent modeling used more advanced modeling techniques, the highest accuracy rate achieved was 69%. While 2/3 is good gambling odds, this is data that is directly tied to real people, so any amount of human error is a slippery slope to travel down.

Recommandations

1. After many attempts of modeling, we were still only able to achieve a 69% in accuracy. Be aware of using this model or any model if there are plans to retroactively adjust records of people who have been arrested, because human error is not worth accidentally ruining anyone's life by marking they were arrested when they weren't.
2. Hire more women and non-white officers as there is a discrepancy in officer demographics.

Future Work

Apply more advanced techniques in order to get the data to work more efficiently.


THANK
YOU