Terry Stops and the Legitimacy of "Reasonable Suspicion"

Kenny Oh and Moses Lin

Overview

This data represents records of police reported stops under Terry v. Ohio, 392 U.S. 1 (1968). Each row represents a unique stop.

- Each record contains perceived demographics of the subject, as reported by the officer making the stop and officer demographics as reported to the Seattle Police Department, for employment purposes.
- Where available, data elements from the associated Computer Aided Dispatch (CAD) event (e.g. Call Type, Initial Call Type, Final Call Type) are included.

Deployment

Evaluation

Modeling

Data Preparation

Exploratory Data Analysis

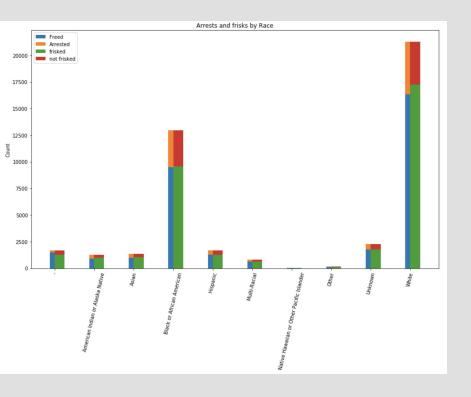
Business Understanding

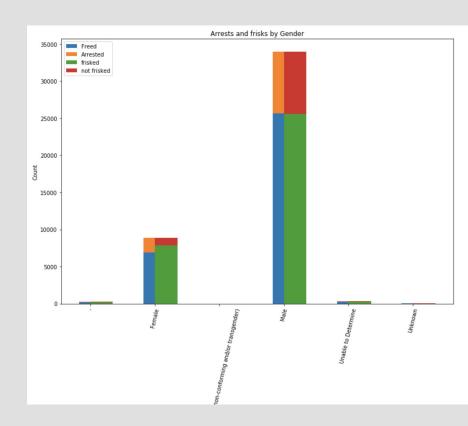
Business Understanding

We aim to answer the following questions:

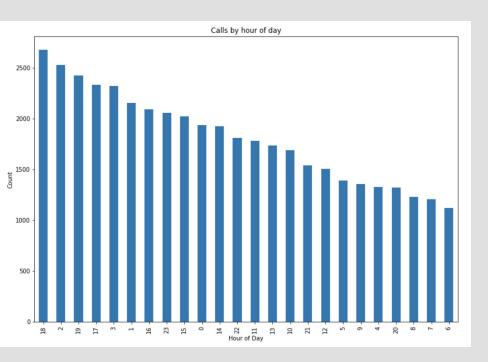
- Are Terry Stops motivated by perceptions of race and/or gender?
- Does the type of weapon a subject carries determine his likelihood of being arrested?
- Which features best predict an arrest?

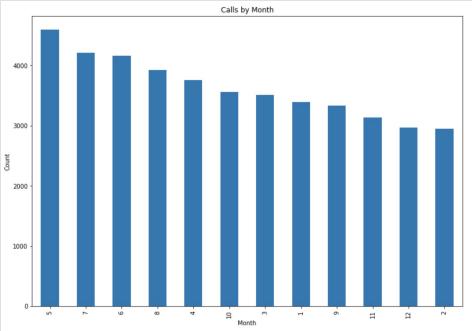
Looking at Race and Gender



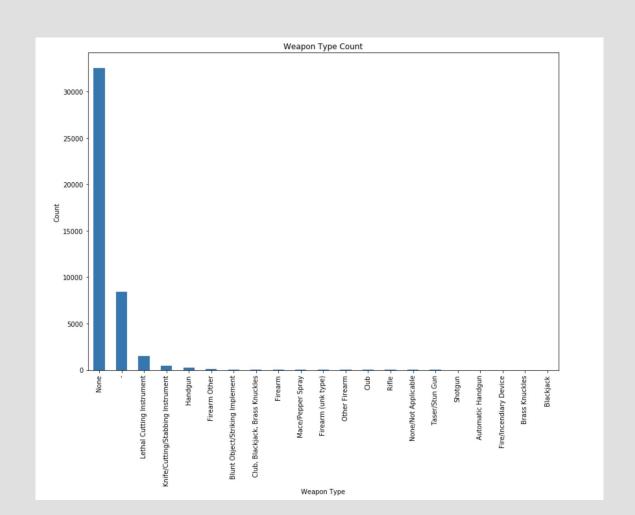


Time of calls





Weapon Type



Hypothesis Testing

arrest_flag	0	1
subject_perceived_race		
American Indian or Alaska Native	1180	61
Asian	1235	96
Black or African American	11769	767
Hispanic	1643	0
Multi-Racial	788	0
Native Hawaiian or Other Pacific Islander	25	9
Other	147	0
Unknown	1845	160
White	19624	1092

arrest_flag	0	1
subject_perceived_gender		
Female	7932	409
Gender Diverse (gender non-conforming and/or transgender)	2	1
Male	30202	1775
Unable to Determine	120	0

P-value: 3.086e-44

Hypothesis 1: Is there an association between Subject Race and Arrest Flags?

- . HO: Subject Race is not associated with Arrest.
- . H1: Subject Race is associated with Arrest.

P-value: 0.00075

Hypothesis 4: Is there an association between Subject Gender and Arrest Flags?

- . HO: Subject Gender is not associated with Arrest.
- · H1: Subject Gender is associated with Arrest.

Classification Models

Logistic Regression	Decision Tree	Random Forest	KNN	XGBoost	Voting Classifier
<u>F1:</u> 0.394	<u>F1:</u> 0.370	<u>F1:</u> 0.357	<u>F1:</u> 0.383	<u>F1:</u> 0.357	<u>F1:</u> 0.384
<u>Acc:</u> 0.867	<u>Acc:</u> 0.837	Acc: 0.909	<u>Acc:</u> 0.855	<u>Acc:</u> 0.909	<u>Acc:</u> 0.853

Using Grid Search-Validation did not increase these scores by much!

Conclusion

- Logistic Regression is the most successful model
- R1 score of 39.4%

Feature Importance

- Weapon was not checked 0.491
- Officer Age 0.492
- Subject Female 0.415
- Subject White 0.399
- Afternoon 0.374
- No Frisk 0.394

Future Work

- We could apply our model to data from other states to see if it holds up.
- We could look modeling using data of police reported stops in other regions, such as in other states or countries.
- We could classify some of the features differently, for example, not categorizing certain hours into morning, afternoon, evening.
- We could focus more on specifics of the Seattle Police Department. We did not fully utilize all the data in the Terry stops Dataset.
- We could gather data on arrests in general, not just as Terry stops.

Thank You!

Are there any questions?