DETECTING IMPLICIT BIAS IN POLICE TRAFFIC STOPS

Unit 3 Supervised Learning
Capstone Presentation
By Mark Ferguson
July 2018

DATASET DETAILS

Source:

- Connecticut Racial Profiling Prohibition Project
- Ctrp3.ctdata.org

Details:

- Oct 1st, 2013 to Sep 30, 2015
- 850,000 rows
- Subject details: Age, race, sex, residency
- LE details: Officer ID, police dept.
- Stop details: Time, reason, vehicle searched, contraband, outcome of stop

HYPOTHESES

- Two hypotheses arise from an initial look at the data
- This study will test these hypotheses using various statistical techniques

HYPOTHESIS 1

ımplicit racial bias plays a role in at least some aspects of some interactions on the road between police and motorists.

HYPOTHESIS 2

The outcome of interactions between police and motorists can be statistically modeled and predicted as a function of demographic factors, specifically race.

DATA CLEANING

Main Issues:

- Missingness, e.g. most location fields null
- Questionable Data – subject ages ranged from 0 to 250+
- Large number of object fields with hundreds of possible values

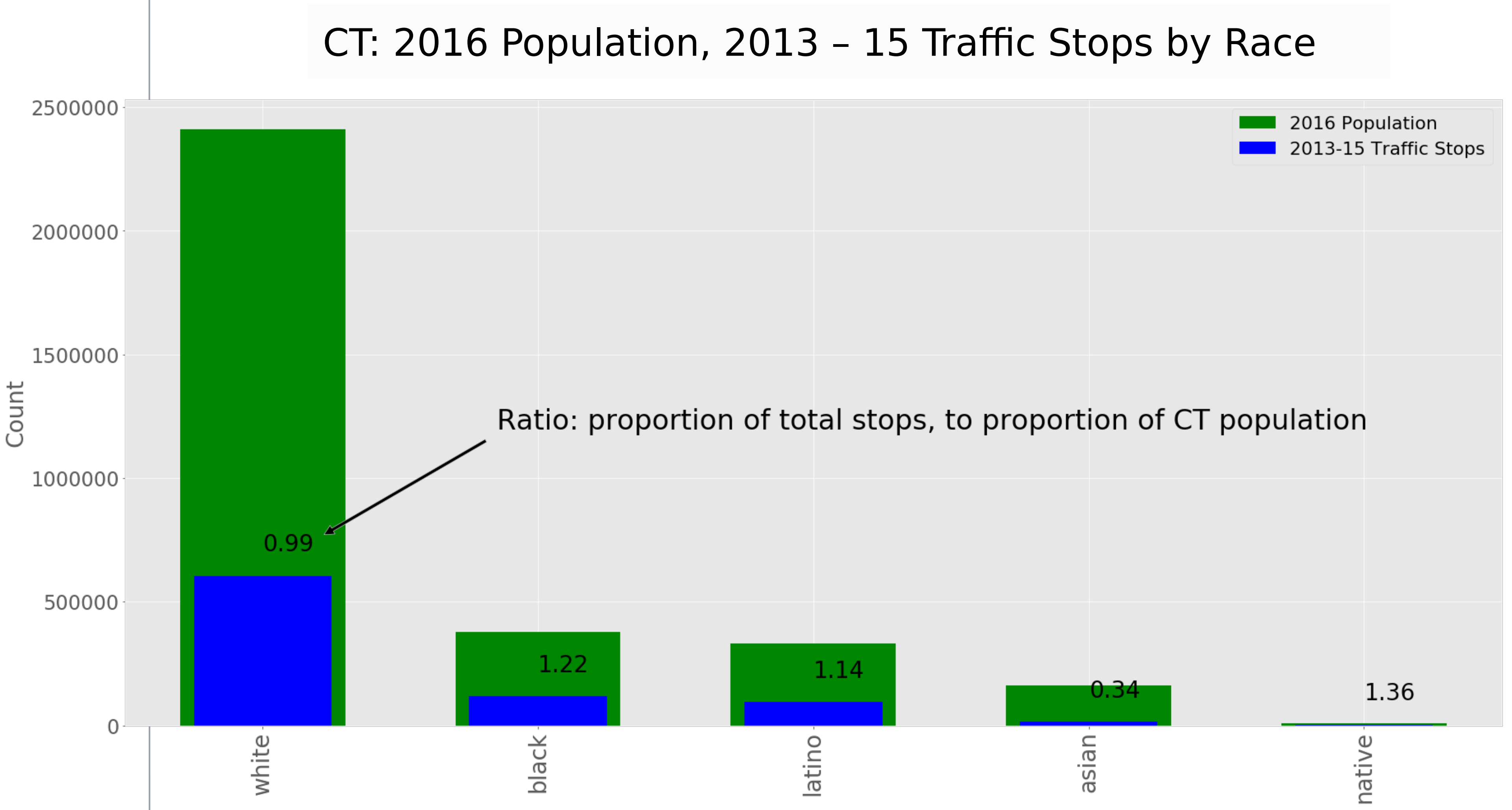
Corrective Measures:

- Remove columns that were mostly null
- Set outlying ages to the mean value
- Set object fields to ‘category’ datatype

Original Dataset Shape: (857895, 42)

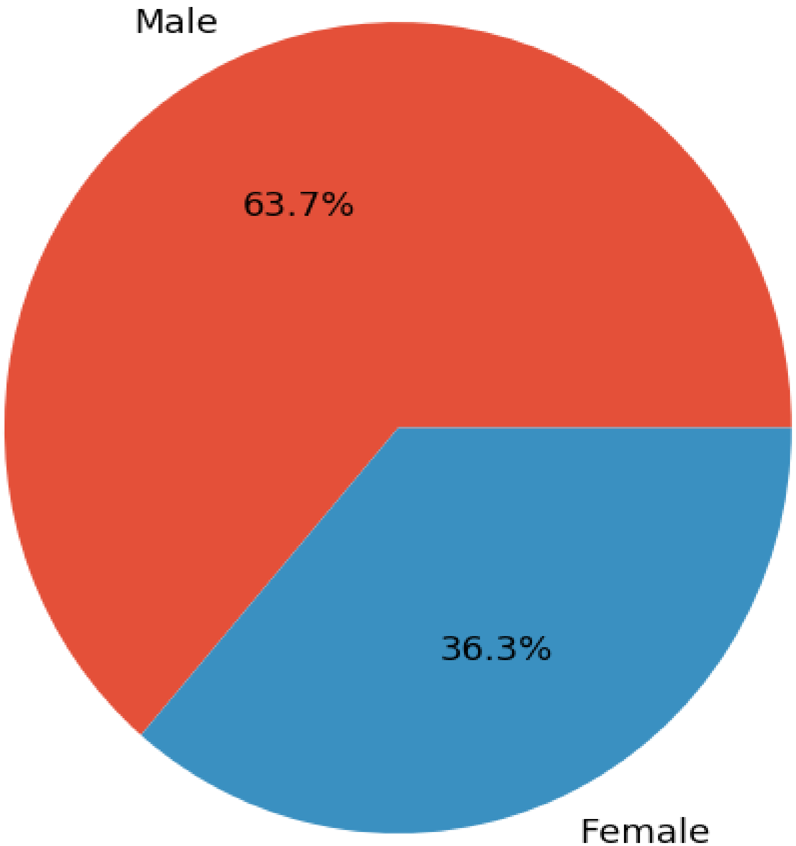
Final Dataset Shape: (817091, 23)

DATA OVERVIEW



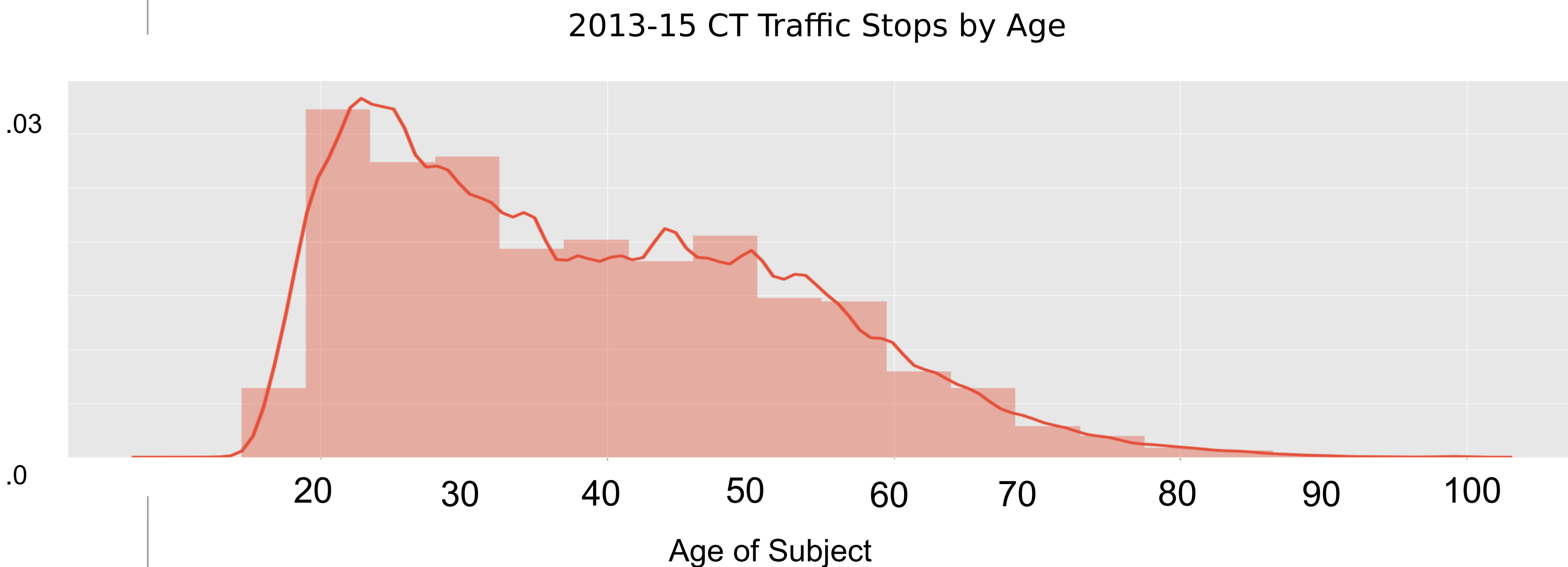
DATA OVERVIEW

1b. Race and Sex Distribution



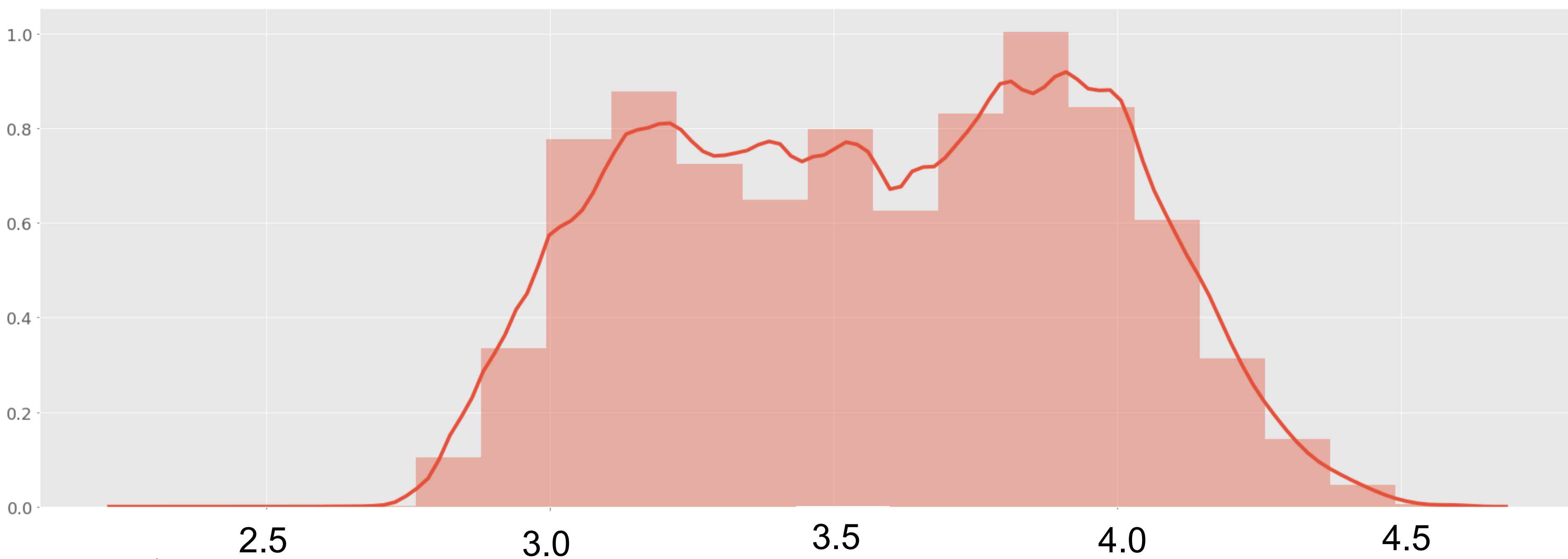
DATA OVERVIEW

2. Age Distribution



DATA OVERVIEW

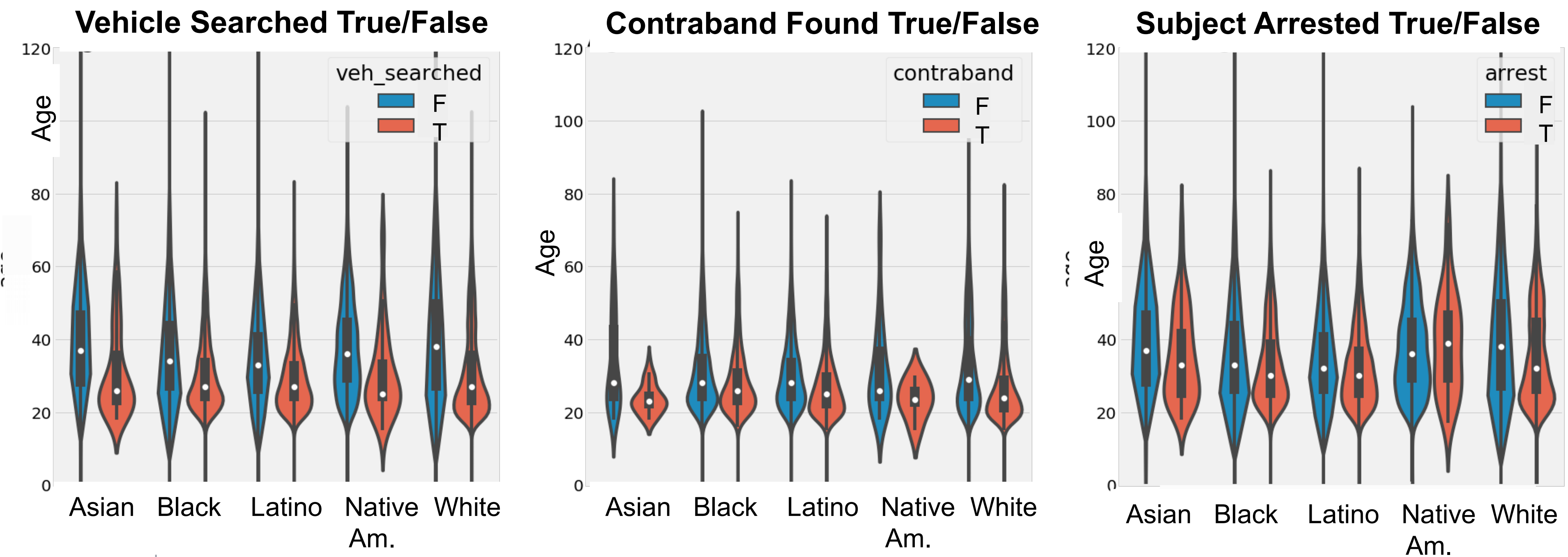
2013-15 CT Traffic Stops by (log) Age



Age of Subject (log scale)

DATA OVERVIEW

2a. Age Distribution



t-Tests

- Outcomes were coded as either 0 or 1.

- The expected values of these outcomes were then computed for each race group.

- These expected values were then compared vs. the overall population mean.

- t-tests were conducted to determine whether these differences vs. the population mean were statistically significant or not.

- P-value threshold: 0.05.

t-TEST RESULTS

- Considering Stops for (Initially) Non-investigative Reasons:

Q. In what percentage of these stops do searches occur?

Racial Groups:

Mean of Overall Population: 3.0%

Group	Mean: p-value < 0.05
Black Male	0.069208
Latino Male	0.061417
	0.030127
White Male	0.026879
Latina Female	0.023390
Black female	0.019927
Native Male	0.014036
Asian Male	0.012159
White Female	0.011926
Native Female	0.010169
Asian Female	0.006298

t-TEST RESULTS
- Searches from Non-Investigative Stops

Q. In what percentage of these searches is contraband found?

Racial Groups:

Mean of Overall Population: 31.3%

Group	Mean: p-value < 0.05
White Male	0.368033
	0.312563
Black Male	0.270601
Latino Male	0.243375
Latina Female	0.239003
Asian Male	0.226087
Black Female	0.209476
Native Male	0.135135

t-TEST RESULTS

- In stops for some less-serious reason, e.g. defective lights:

Q. What is the split between punitive and non-punitive outcomes?

Racial Groups:

Mean of Overall Population: 13.6%

Group	Mean: p-Value < 0.05
White Female	0.065579
Asian Female	0.078431
Black Female	0.084977
Native Male	0.092
White Male	0.093603
Asian Male	0.094318
	0.135881
Latino Male	0.197874

Supervised Learning

Various supervised learning techniques were used
to model and better understand these police interactions,
and the factors that influence
their outcomes.

- Target variable: punitive/non-punitive outcome, defined
as 1 (ticket, arrest, summons) or 0 (warning, no warning).
- Techniques Used:

Logistic Regression (Standard, L2, L1)
K-Nearest Neighbors Classifier
Random Forest Classifier
PCA with Random Forest
Gradient Boosting

Support Vector Machines
- abandoned modeling with this technique
due to chronic slowness

TESTING ROUND 1

- 1 % sample of dataset (approx. 8,300 records)
- All features used as well as dummies
- Where applicable, input parameters were varied
- Models evaluated on accuracy, consistency, and efficiency
- The best predictors/ parameters are shown for each.

<table>
<tr><th rowspan="2">Model</th><th rowspan="2">Parameters</th><th colspan="2">Results</th></tr>
<tr><th>R²</th><th>Run Time</th></tr>
<tr><td>KNN Classifier</td><td>N = 10</td><td>0.6461</td><td>15.2 minutes</td></tr>
<tr><td>Logistic Regression</td><td>C = 1E9</td><td>0.70094</td><td rowspan="3">2.8 minutes</td></tr>
<tr><td>Ridge (L2 Reg)</td><td>C = 0.1</td><td>0.74676</td></tr>
<tr><td>LASSO (L1 Reg)</td><td>C = 1</td><td>0.75536</td></tr>
<tr><td rowspan="2">Random Forest Classifier</td><td>n = 10</td><td rowspan="2">0.71594</td><td rowspan="2">0.56 minutes</td></tr>
<tr><td>Max depth: None</td></tr>
<tr><td>PCA with Random Forest</td><td>20 Components</td><td>0.64916</td><td>0.07 minutes</td></tr>
<tr><td rowspan="3">Gradient Boosting</td><td>50 Estimators</td><td rowspan="3">0.75327</td><td rowspan="3">15.4 minutes</td></tr>
<tr><td>Max Depth 10</td></tr>
<tr><td>Loss: Deviance</td></tr>
</table>

TESTING ROUND 2

- 10 % of dataset (83,000 records)
- Successively fewer features used - optimal results at 100 (Top 50 +ve and top 50 -ve correlators vs. target variable).
- Models evaluated on precision, recall, consistency, efficiency

Model	Parameters	Results				
		R^2	SD	Precision	Recall	Run Time
Ridge (L2 Reg)	C = 0.1	0.73643	0.0044	0.74	0.79	55.3 min
LASSO (L1 Reg)	C = 1	0.69441	0.0015	0.74	0.65	
RFC	n = 10 Max depth: None	0.6916	0.0015	0.75	0.64	1.0 min
PCA w/ RFC	5 Components	0.6915	0.0015	0.75	0.64	1.3 min
KNN Class	n = 250	0.74529	0.0061	0.75	0.76	25.2 min
Gradient Boosting	50 Estimators Max Depth 10 Loss: Deviance	0.69614	0.0015	0.75	0.64	25.0 min
Gradient Boosting	50 Estimators Max Depth 10 Loss: Deviance	0.69615	0.0015	0.75	0.64	19.8 min

OBSERVATIONS

- Preferred model: K-Nearest Neighbors Classifier - offered best combination of precision, recall and efficiency.

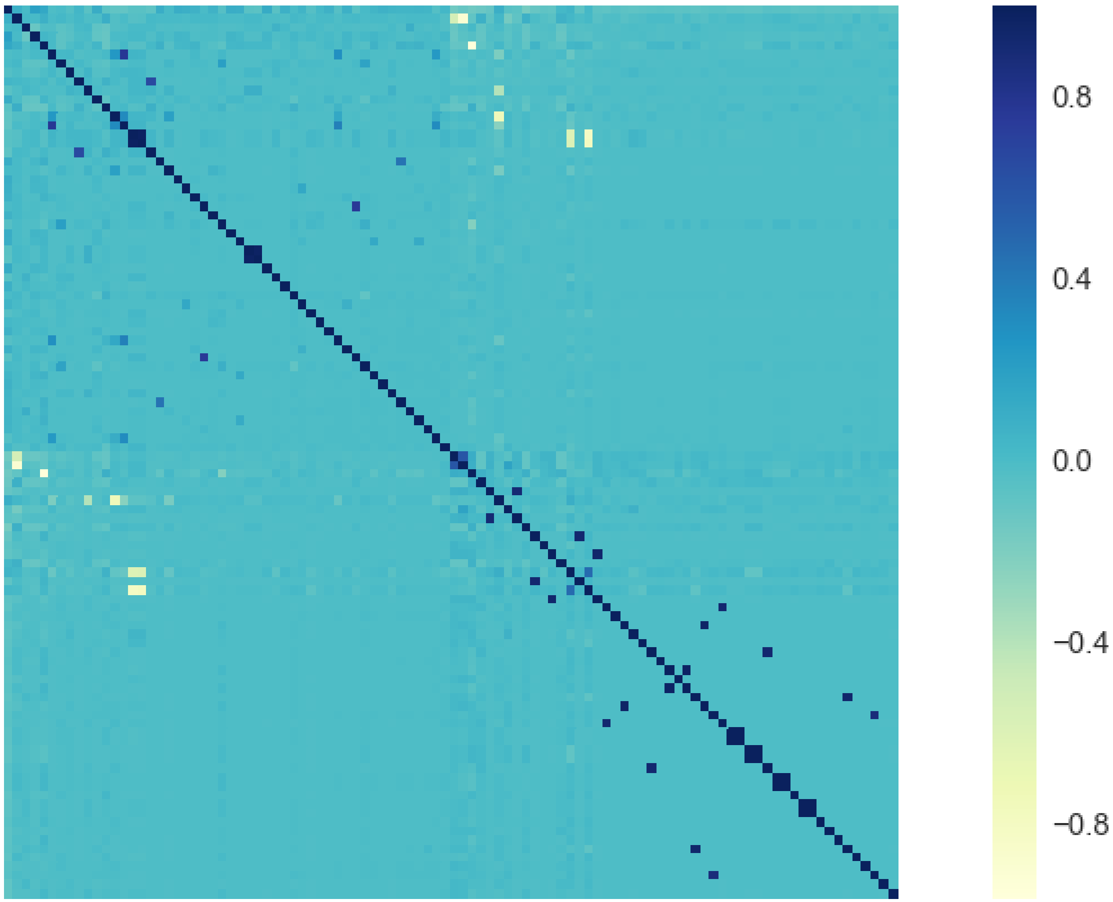
- Accuracy seems to plateau around 75%.

- ~ 69% accuracy can be achieved with as few as 10 highly - correlated variables.

- Race is a factor in traffic stop outcomes but not a top one.

HEAT MAP
– TOP 100 CORRELATORS (50+ve, 50 -ve) vs. OUTCOME

Mostly Low Co-Linearity Among Variables



SELECTED CORRELATORS vs. TRAFFIC STOP OUTCOME

Positive:

1. State Police
2. Reason Vehicle

26. Month September
36. Month May
38. Ethnicity Hispanic
43. Race Latino
50. Race Black

Negative:

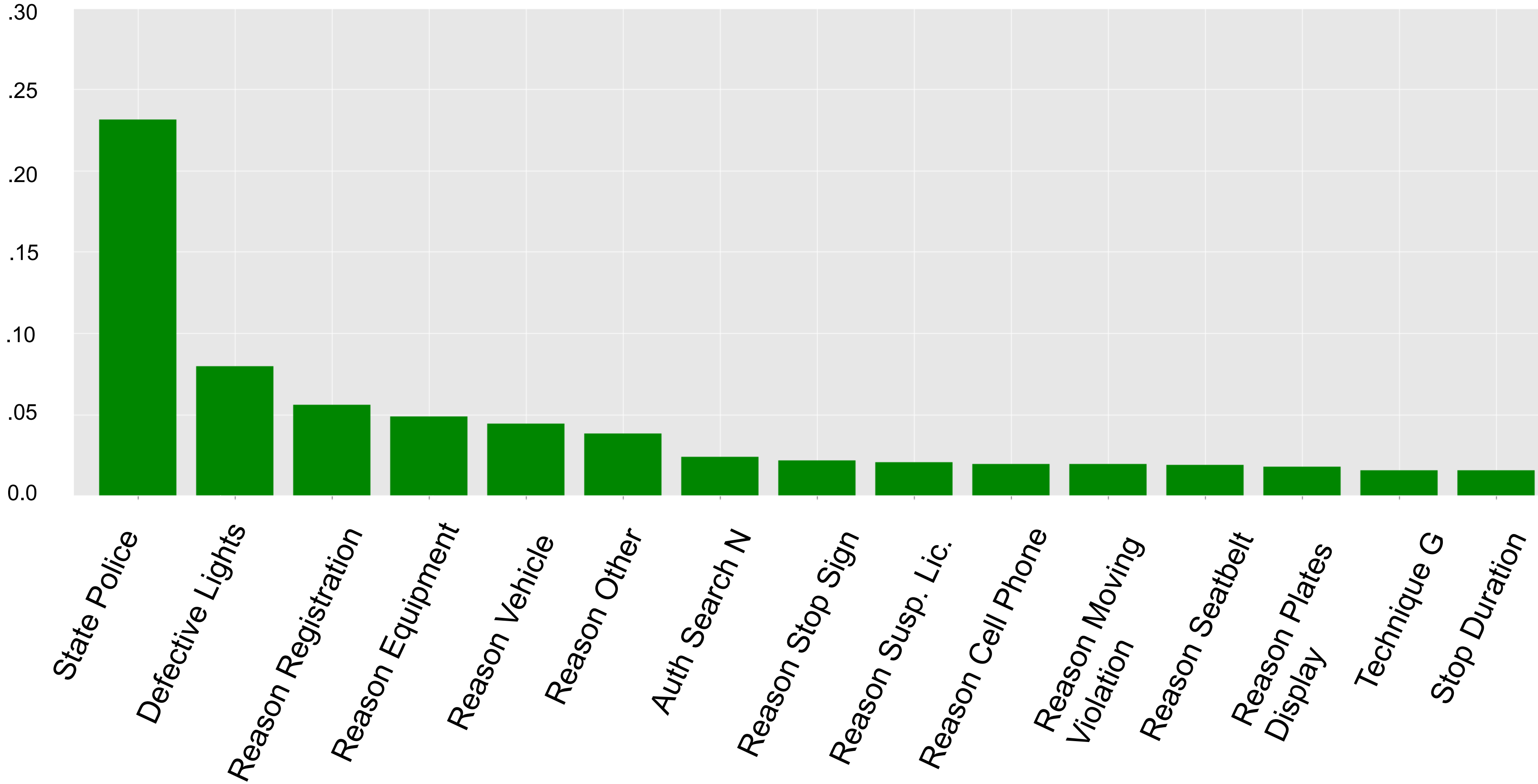
1. Reason Defective Lights
2. Reason Equipment

4. Reason Stop Sign

24. Race White

RANDOM FOREST FEATURE IMPORTANCES

Feature Importances taken from top 50 and bottom 50 correlators – Top 15 Shown



Conclusions

- Some implicit bias does seem to be at play.

- Black drivers are stopped at a higher rate than their proportion of the CT population (14.1% vs. 11.5%)
- Black and Latino males are more likely to be searched even when the stop is for a non-investigative reason
- Searches of white males most likely to yield contraband
- Females (most of all white) more likely to be let off for minor infractions. Least likely by far: Latino males

- Race ‘black’ or ‘latino’ correlates positively with punitive outcomes.

- Race ‘white’ correlates negatively with punitive outcomes.

- Some other features/ factors more influential than race.

Suggestions for Further Study

- Run models on entire dataset.
- Fully investigate the role of geographical (lat, long) location.
- Look at consistency among depts, officers.
- Compare CT with other states, parts of the country.

Thank You

- Questions?