Predicting Racial Bias Tendencies in Police Stops

Team(0) - Mahir Bathija, Taylor Merry, Hyewon Kim, Christopher Painton

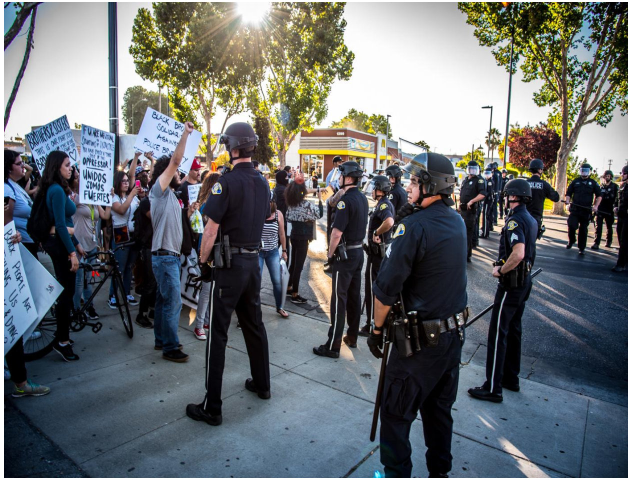
Motivation

We wanted to research the belief that police forces within the United States have adopted policies of discriminatory behavior

These racial injustice claims are highly relevant in our current society

Investigate police practices within our own home state of Washington

Our hope is that we can provide some clarity into an ongoing and evolving issues



Previous Work

Stanford Open Policing Project (Our Primary Data Source)

After stops, officers are more like to ticket, search, and arrest black and Hispanics over white drivers

Are Traffic Stops Prone to Racial Bias

Explains Stanford Open Policing Project data collection strategies
Washington State Patrol collect and interprets their data internally twice a year to fight racial discrepancy

Fast Threshold Tests for Detecting Discrimination

Stop thresholds for whites are lower than blacks and hispanics

Precinct or Prejudice? Understanding Racial Disparities in New York City's Stop and Frisk Policy

Lower thresholds to stop minorities compared to whites, despite stimulating situations of whites within these high-crime rate areas

Research Question

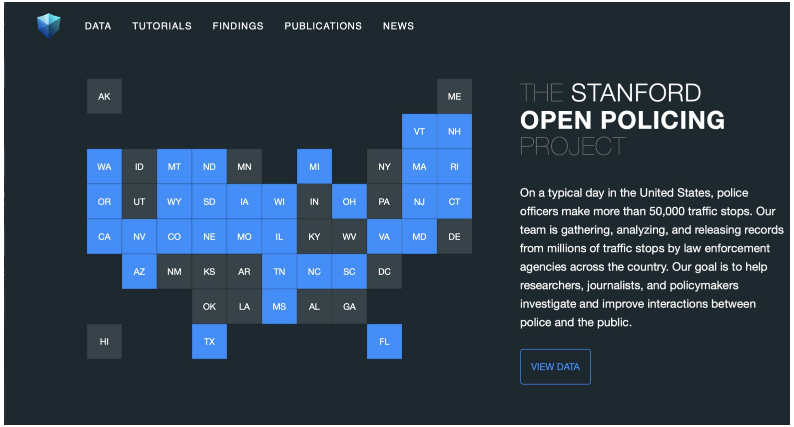
Can we predict the race of an offender from a traffic/pedestrian stop given the stop reasoning along with county information, demographics, and voting history?

Data Sources

Stanford Open Policing Project

Provides traffic stop data for 31 states containing over 130 million traffic stop records.

Dedicated to the goal “to help researchers, journalists, and policymakers investigate and improve interactions between police and the public”



Data Sources

Presidential Election Voting History

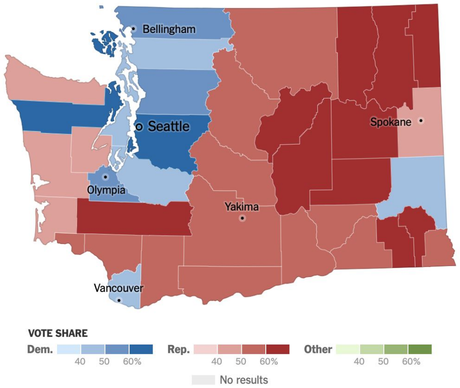
From the Secretary of State for Washington website

Provides csv files for 2008, 2012, and 2016 presidential elections by county and candidate

Used 2008 election data for stops within 2009-2010

Used 2012 election data for stops within 2011-2014

Used 2016 election data for stops within 2015-2016



Vote by county	Clinton	Trump
King	718,322	216,339
Pierce	172,538	146,824
Snohomish	185,227	128,255
Spokane	93,767	113,435
Clark	92,757	92,441
Thurston	68,798	48,624
Kitsap	63,156	49,018
Whatcom	60,340	40,599

Data Sources

U.S Census Quick Facts (County Demographics)

Provides information on a national, statewide, and county-wide level.

Scraped data from each URL into csv file

This file contains the county name, population of county, race, and the percentage breakdown of the races made up from that county.

All Topics	Snohomish County, Washington
Population estimates, July 1, 2017, (V2017)	801,633
PEOPLE	
Population	
Population estimates, July 1, 2017, (V2017)	801,633
Population estimates base, April 1, 2010, (V2017)	713,308
Population, percent change - April 1, 2010 (estimates base) to July 1, 2017, (V2017)	12.4%
Population, Census, April 1, 2010	713,335
Age and Sex	
Persons under 5 years, percent	6.4%
Persons under 18 years, percent	22.7%
Persons 65 years and over, percent	13.0%
Female persons, percent	49.8%
Race and Hispanic Origin	
White alone, percent (a)	78.4%
Black or African American alone, percent (a)	3.5%
American Indian and Alaska Native alone, percent (a)	1.6%
Asian alone, percent (a)	11.1%
Native Hawaiian and Other Pacific Islander alone, percent (a)	0.7%
Two or More Races, percent	4.7%
Hispanic or Latino, percent (b)	10.2%

Data Analysis Process

We tested each of our predictor variables to see their effects on driver race:

For numerical variables, we used analysis of variance (**ANOVA**) to determine if the means for each race were statistically different, and then we used **Tukey tests** to determine which races had different means and by how much.

For categorical variables, we used **contingency tables** and **chi square independence tests** to test whether or not each categorical predictor was independent of driver race.

Lastly, we fit an **XGBoost model** to predict if the driver was a minority based on all of the predictor variables.

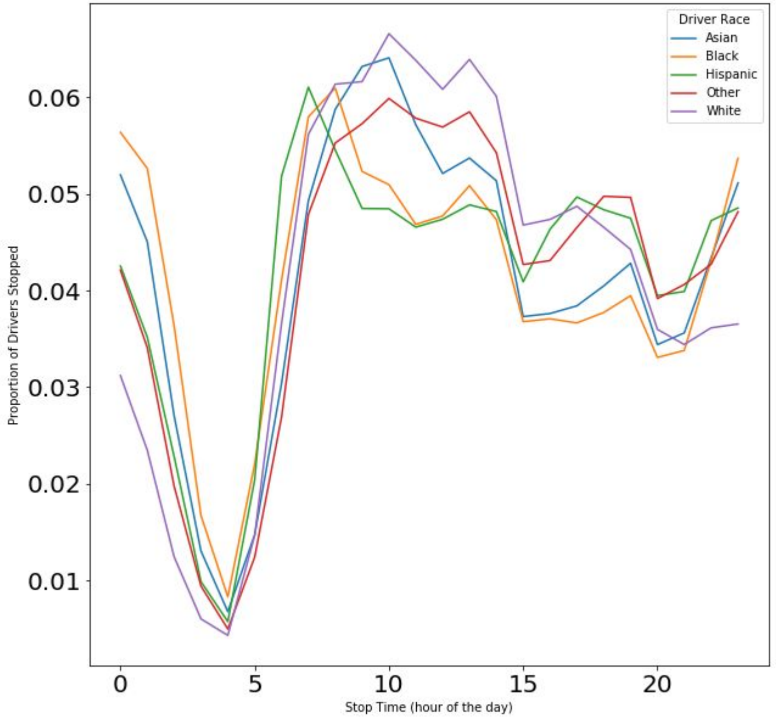
Results (EDA)

10AM to 8PM - White drivers stopped more than all other races.

8PM to 4AM - All other races are stopped significantly more than white drivers.

Black drivers are stopped almost twice as often during this period.

Proportions of Drivers Stopped During Stop Times



Results (EDA)

Searches are conducted on **white offenders with a rate of ~2.4%**

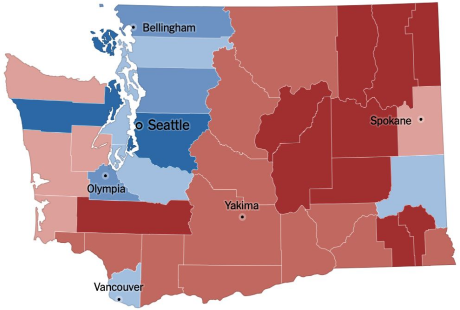
Searches conducted on **black offenders at a rate of ~5.2%** and **Hispanics at a rate of ~4.1%.**

Race	Search Conducted	Search Not Conducted	Percentage of Stops with Search
Asian	6,838	333,620	2.0084 %
Black	12,840	235,789	**5.1643 %**
Hispanic	18,970	446,694	**4.0738 %**
White	97,471	4,007,539	**2.3744 %**
Other	4,346	72,799	5.6335 %

Race Breakdown of Searches Conducted

Results (ANOVA)

Hispanics, on average, were being pulled over in more Republican counties than any other race (republican to democrat ratio of over .15 higher than any other race).



Race 1 vs Race 2	Difference of Means
Hispanic vs Asian	0.375
Hispanic vs Black	0.362
Hispanic vs Other	0.227
Hispanic vs White	0.151

Tukey Test Table for Republican to Democrat Voting Ratio

Results (ANOVA)

Hispanic drivers were **almost 2 years younger than every other race** on average.

White drivers were **2 years older** than drivers of any other race.

Race 1 vs Race 2	Difference of Means
Asian vs Hispanic	3.901
White vs Asian	2.272
Black vs Hispanic	1.989
White vs Black	4.185
Other vs Hispanic	2.992
White vs Hispanic	6.173
White vs Other	3.181

Tukey Test Table for Driver Age

Results (Chi2 Independence Test)

All chi squared independence tests came up with low p-values, indicating categorical variables are not independent of one another

Comparison between expected and observed counts for arrests or citations

Whites only race that had their counts decrease from expected to observed.

	Arrest/Citation Expected	Arrest/Citation Observed	Percentage Change
Asian	176,229.54	198,652	**+ 12.72 %**
Black	128,696.56	139,903	**+ 8.70 %**
Hispanic	241,039.29	249,637	**+ 3.57 %**
Other	39,932.17	46,152	**+ 15.58 %**
White	2,124,855.43	2,076,409	**- 2.33 %**

Expected vs Observed Counts for Arrests/Citations

Results (Chi2 Independence Test)

Displays the violations the had greater than a 25 percent increase of expected violations count to observed violations count

Whites were the only racial group to not have a large increase from expected to observed counts for violations

For transparency, it is worth noting that there was a higher total count for whites

Race	Violation (Percentage Increase from Expected to Observed)
Asian	Moving (45.63%)
Black	DUI (63.43%), **License (179.87%)**, Lights (35.7%), **Moving (116.84%)**
Hispanic	DUI (51.61%), License (59.26%), Paperwork (34.63%), Stopping (26.12%), Truck (27.90%)
Other	**DUI (81.65%)**, **License (137.63%)**
White	

Violations with Percentage Increase from Expected to Observed

XGBoost Model

XGBoost is a gradient boosting framework. Gradient boosting is an ensemble method. Ensemble methods combine the predictions of several models in order to improve performance.

Boosting is when the model starts with weak learners (simple decision trees) and then iteratively combines them while reducing bias/error to create a strong learner.

Used **Binary Classification** to predict the Minority Race variable

We chose to prioritize accuracy over implementation time and explainability because we don't plan on implementing our model to make real time predictions and we aren't trying to explain our algorithm to non-technical shareholders. Therefore, we chose a model praised for its accuracy and spent a lot of time optimizing the hyperparameters.

XGBoost Design and Hyperparameters

To adjust for the unbalanced dataset (~80/20 split of whites to minorities), we **undersampled the white population** for the training set (within KFold loop)

K Fold Cross Validation (k = 5): for scale_pos_weight, learning rate, and n_estimators

For Other Hyperparameter Tuning:

Grid Search (with the Grid Search CV function) to iterate through different parameter combinations and return the best combination (after 3 cv folds) based on the highest mean auroc (receiver operating characteristic)

max depth (maximum depth of a tree)
min_child_weight (min sum of weights of all observations)
subsample (fraction of observations to be randomly sampled for each tree)
colsampe_bytree(fraction of observations to be randomly sampled for each tree)
gamma and alpha

XGBoost Results

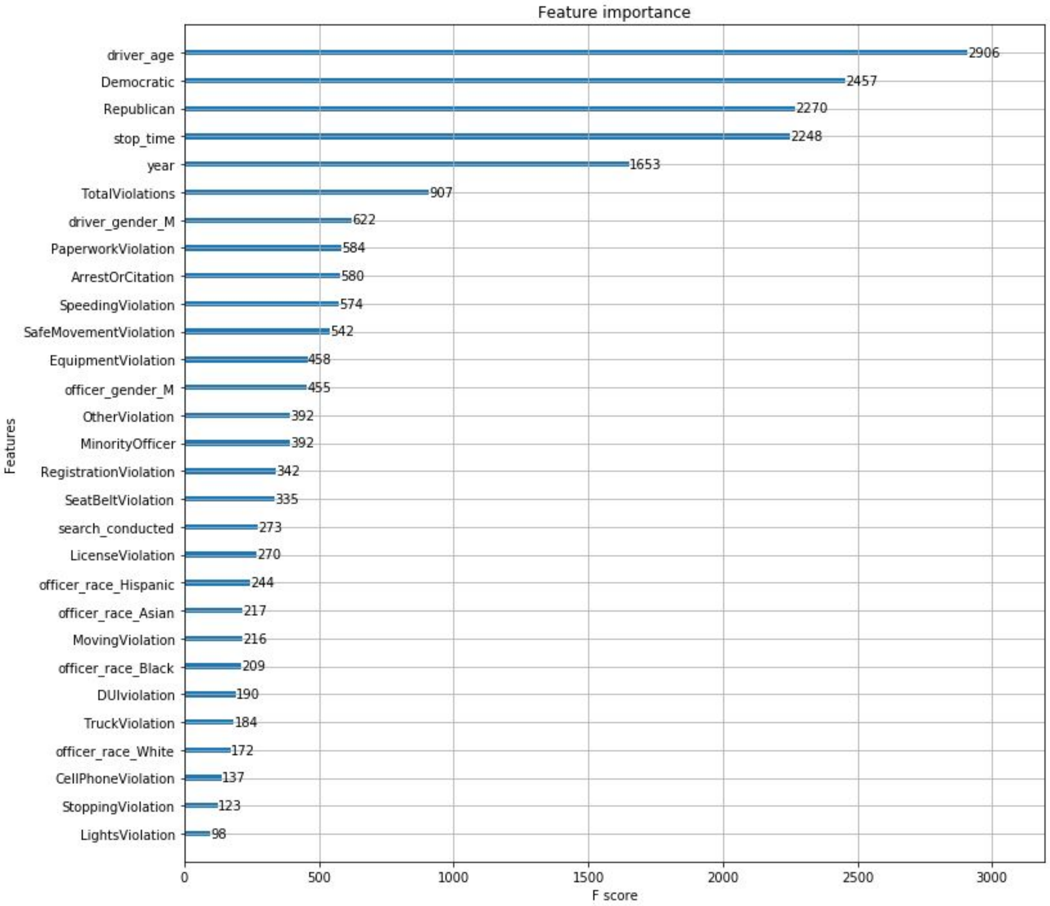
Accuracy: 62.89%

F1 Score: 0.442

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

$$F1 = \frac{2 \times precision \times recall}{precision + recall}$$



Limitations & Future Directions

Shortcomings:

Undersampling issue

Lack of time and resources

Only able to use k=3 k-fold due to the large number of rows

Therefore, even if predicted correct, we still can't attribute it completely to racial bias.

Engineer more features/variables starting with search conducted.

We only scoped one of the 50 states.

Closing

Although model was unsuccessful, perhaps it is good that we couldn't predict the driver race

Massive issue that continues to grow and so will the data

We hope that our research, although on a smaller scale, will motivate others and add valuable information to this ongoing topic

Questions?