Police Stops in San Diego 2014-2017 EDA

Source: https://openpolicing.stanford.edu/data/

David Aminifard

A15451805

Aaron Mcmillan Fraenkel

Introduction & Context:

America's history of disproportionate treatment of people based on their backgrounds has long been subject to scrutinty, and yet has often been perpetuated by bad policy and bad policing. The article, "Targeting young men of color for search and arrest during traffic stops: evidence from North Carolina, 2002–2013", written by, Frank R. Baumgartner, Derek A. Epp, Kelsey Shoub & Bayard Love, showed evidence of strong racial bias in the policing. This bias is often directed towards young men of color and isolates minority communities, and much of this has been worsening in the wake of the war on drugs in North Carolina. Their analysis of North Carolina police stop data revealed that "North Carolina's black men, who are searched at much higher rates than their white counterparts, but are less likely to be found with contraband in discretionary police searches"(18, Frank R. Baumgartner, Derek A. Epp, Kelsey Shoub & Bayard Love). This indicates that there is racial bias in how North Carolina's motorways are policed.

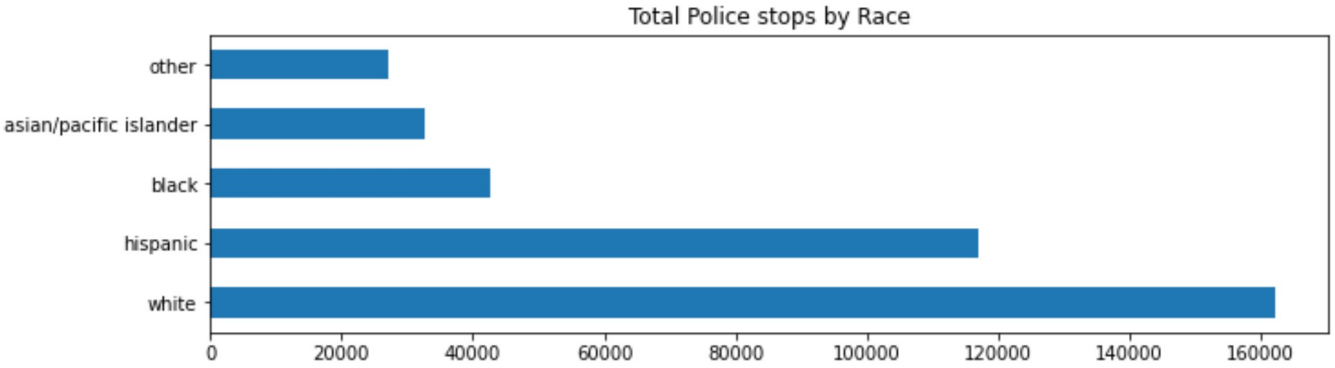
Seeing the situation in North Carolina, this begs the question: Does racial bias in San Diego's motorways unfairly affect the outcomes of police stops?

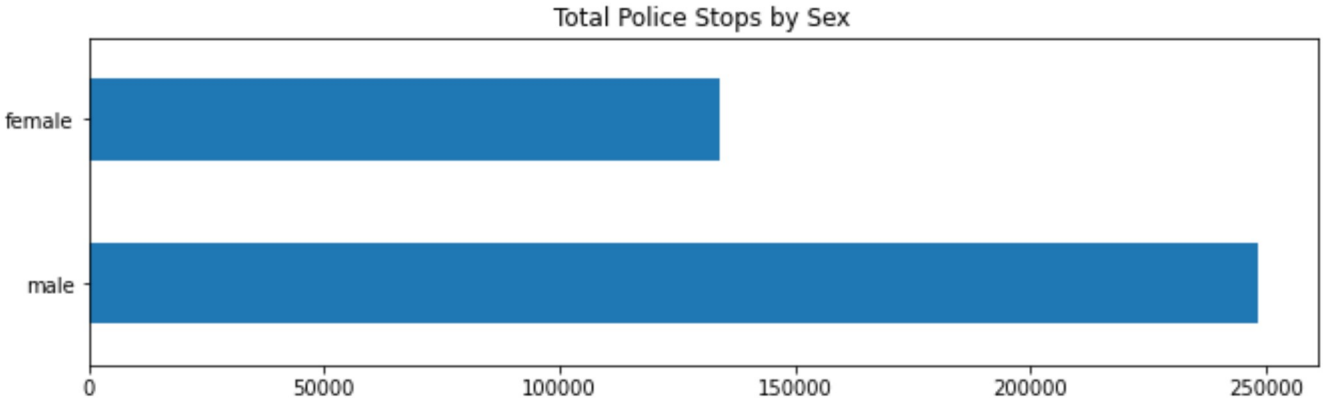
Given this context, this report will analyse police stop data made available by The Stanford Open Policing Project. In particular, we will look at San Diego's police stop data, and identify and analyze potential inequity in light of the evidence there is for North Carolina's bad policing using parity measures and a hypothetical classification model.

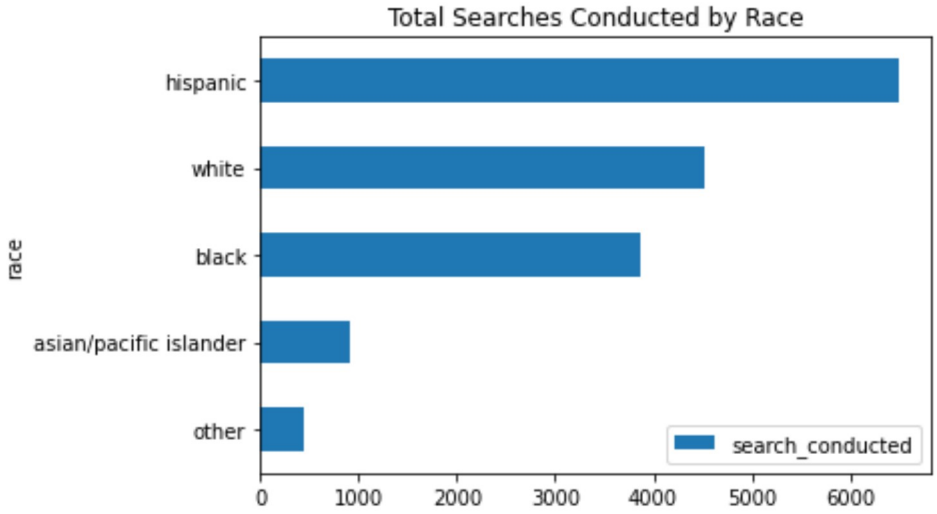
This report will consider fairness in terms of luck egalitarianism because it asserts that inequalities should follow from an individuals choices rather than circumstances, and our analysis of San Diego police stop data may provide evidence of disparate treatement based on a person's background, i.e. circumstances, which would violate this notion of fairness.

EDA & Data Description

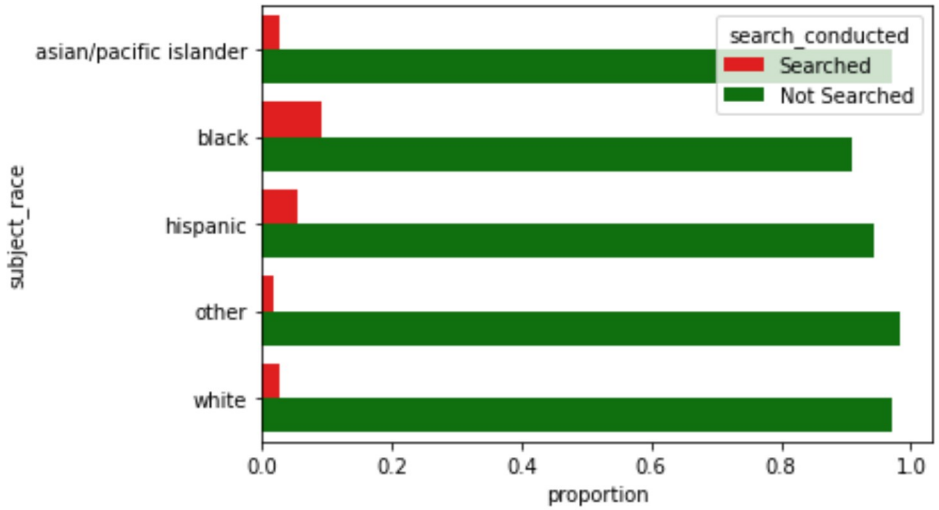
The visualizations below show that white people and males are stopped by police the most, hispanics are subject to the most total searches. However, they do not take into account the amount of people per each group there are in the population of San Diego.





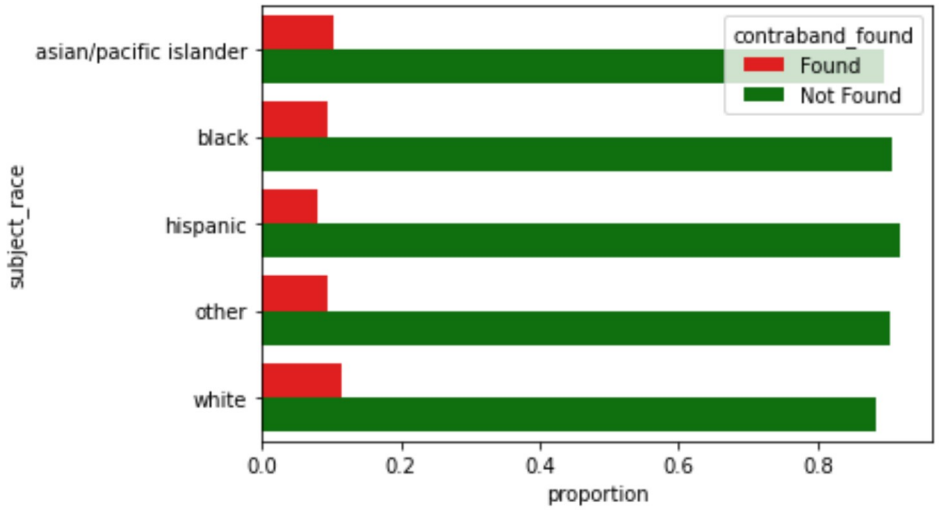


Since we do not know the proportion of people in each group with respect to the population of San Diego, we will investigate further in the rates of searches with respect to the amount of people in each group in the dataset. The visualizations below reveal that blacks and hispanics are searched at a much higher rate than whites and asians/pacific islanders.



	proportion	
search_conducted	**Not Searched**	**Searched**
subject_race		
asian/pacific islander	0.972035	0.027965
black	0.909308	0.090692
hispanic	0.944475	0.055525
other	0.983442	0.016558
white	0.972199	0.027801

Seeing that blacks and hispanics have a higher rate of being searched, we'd assume that there would be a higher rate of contraband being found. However, the visualizations reveal below that blacks and hispanics have a lower rate of contraband being found than whites and asians/pacific islanders.



	proportion	
contraband_found	**Found**	**Not Found**

subject_race		
asian/pacific islander	0.104396	0.895604
black	0.093209	0.906791
hispanic	0.081218	0.918782
other	0.095344	0.904656
white	0.114856	0.885144

Since P(C=search_conducted | A=black) is greater than P(C=search_conducted | A=white) we would assume that P(C=contraband_found | A=black) is proportionally greater than P(C=contraband_found | A=white).

However P(C=contraband_found | A=black) is less than P(C=contraband_found | A=white), (0.093209 < 0.114856), which indicates that there is a severe inequity that may be based on race. There dataset does not satisfy accuracy parity.