Ramya Prasanna

August 14, 2025

Stanford Open Policing Data Analysis Report

The main focus of my study was on the disparities in searches conducted, as well as the reasoning for searches, subject age, race, and sex data found in various counties. I decided to study this topic not only because there was so much geographic and demographic data available, but also because such data could give a lot of insight into any bias or patterns concerning the subjects of the report. The question I have chosen is important to analyze as it comes to the root of why policing occurs and whether it is being done fairly.

I started my study by looking at whether some counties conducted more searches than others. Here, I found a stark difference between the number of searches conducted, ranging from 200,000 searches to only one search. My work further studied whether there was some factor driving this stark difference among the top four counties with the most searches. The counties with the most searches I found were Mecklenburg County, Los Angeles, Wake County, and Guilford County. Through my analysis, I first discovered that there is a pattern of slightly more person searches than vehicle searches. Next, looking at the search basis, all counties presented a similar pattern where consent held the majority as the basis for searching, followed by other bases that were not specifically listed. However, Los Angeles skewed from these results, with approximately 90% of the search bases not specifically listed. Regarding reasons for frisk, which further dwell on the slightly higher person searches, it was found that what is reported as generally erratic or suspicious behavior alternated with other possible official information as the majority reasoning behind a pat-down. Finally, I studied whether race, sex, or age could be used to understand the high number of searches. For the three counties other than Los Angeles, the

Black population was the majority searched, with the white population as the runner-up. Los Angeles, in contrast, had a majority Hispanic population being searched, followed by white. In addition, all four counties showed a varying majority of male subjects being searched over female, and searches were primarily conducted on the population within their mid to late 20s.

During this analysis, I came across an issue that caught my attention: Los Angeles was missing much of the data I analyzed. I decided to specifically analyze L.A.'s data and first discovered that more than half of the columns in L.A.-specific data contained no value other than null. I decided to focus on the district, race, and outcomes data. Looking at the distribution of policing reports among the various districts, it can be noted that most have a somewhat equal share, aside from some outliers such as Castaic CVEF, Los Angeles Communications Center, and Los Angeles District. The majority of the outcomes were summons, followed by warnings, and, as seen for summons, arrests, and warnings, the Hispanic population held a large majority of the distribution.

The main conclusion I drew from my analysis was that there is a concerning lack of detail in policing reports in counties with the most policing activity. Various factors presented this to me, such as the lack of specificity in the reasons for frisk—the majority reason being generally erratic or suspicious behavior with no further explanation—consent and "other" being the two main search bases, and the fact that much of L.A.'s data is not reported clearly. This shows how a higher quantity of policing may lead to lower quality reports, or that there needs to be a higher standard of data reporting within these specific police departments. In addition, some other conclusions I was able to derive were that there is some level of racial bias and sexism influencing how many searches are conducted in a region. For example, the Black and Hispanic populations represented a majority of the searches among the other outcomes, and both are

known to be systematically discriminated communities. There is also a stark contrast in the number of men compared to women being searched, and both of these statistical points highlight a bias toward who is seen as more threatening and in need of being searched. Finally, seeing how person searches are slightly more frequent compared to vehicle searches, there could be a chance that personal information is being given greater weight than vehicular information.