Bias and Feedback Loops in Predictive Policing:

A Data-Driven and Ethical Analysis Using the Stanford Open Policing Project

Regis University

MSDS640 Ethics, Privacy, and Social Justice in Data Science

Professor: Dr. Ghulam Mujtaba

Yassine Alami

May 5, 2025

Abstract

I analyse **227 912 085** traffic-stop records (**41 GB**) from **88** U.S. agencies released by the Stanford Open Policing Project (OPP). After a robust extraction pipeline, I quantify *search rates*, *hit rates*, false-positive rates, and implied suspicion thresholds disaggregated by race. Beyond outcome disparities, I expose large, uneven pockets of *missing or ambiguous race data*, a structural weakness that jeopardises any downstream predictive-policing model. Results confirm systematic over-policing of Black drivers and lower evidence thresholds for searches. I argue that unless historical enforcement data are audited and repaired, predictive systems will encode and amplify these inequities.

1 Introduction

Predictive policing is often presented as a technological solution to crime prevention. However, when historical data are racially biased, algorithmic tools replicate, and can amplify, systemic injustice. This paper presents a large-scale, multi-agency analysis of racial disparities in U.S. traffic-stop data and demonstrates how these disparities translate into risks for AI-driven policing.

Research question: *How do patterns of missing data, label inconsistency, and differential search outcomes affect fairness metrics, and what feedback loops arise when such data train predictive-policing models?*

Using the Stanford OPP corpus, I dissect enforcement disparities through key metrics and situate the quantitative findings within ethical and policy frameworks. I ultimately argue that *data fairness is a prerequisite for algorithmic fairness.*

2 Data and Methodology

OPP aggregates traffic-stop data from 200+ agencies; my cleaned corpus spans **227 M rows** across **88** jurisdictions. For in-depth fairness metrics I selected the ten most complete agencies:[1]

Agency (City/State)	Years	Rows (M)

[1]Selected by $< 5\%$ race missingness and full search/outcome fields.

Mesa, AZ	2014-19	1.6
San Francisco, CA	2010-20	0.9
Oakland, CA	2016-20	0.7
Los Angeles, CA	2016-20	7.8
Hartford, CT	2013-20	0.4
Statewide CT	2013-20	3.0
Charlotte, NC	2012-20	1.2
Durham, NC	2010-20	0.5
Seattle, WA	2011-20	1.1
Madison, WI	2016-22	0.3

Pipeline highlights

- **Multi-encoding chunk reader** (UTF-8, Latin-1, UTF-16, CP-1252, ISO-8859-1) to stream 200 k-row blocks without crashes.
- Harmonised race labels (*Black*, *White*, *Hispanic*, *Asian/PI*, *Native*, *Other/Unknown*; 43 raw variants detected).
- Computed, per agency and race:
  - *Search Rate*
  - *Hit Rate* (contraband found)
  - *False-Positive Rate* (1 − Hit)
  - *Risk Ratio* (Black / White search likelihood)

– *Proxy Threshold* (hit-rate-as-evidence)

3 Data Acquisition and Computational Process

The raw OPP archive (41 GB of ZIP-wrapped CSVs) was downloaded, decompressed, and organised via an automated Python script (Requests, ZipFile, Pandas). Chunked loading kept RAM usage < 8 GB per pass. The full, reproducible notebook and helper scripts are available at:

`https://github.com/Regis-University-Data-Science/msds640_caseStudy`

4 Generated Data and Public Availability

The repository above includes:

- Cleaned CSVs of fairness metrics (`fairness_rates_top10.csv`, `risk_ratios.csv`, etc.).
- `missingness_by_agency.csv` (race-field completeness for all 88 agencies; worst: NE Statewide and OR Statewide at 100% missing).
- All code and high-resolution figures (`.png`).

5 Results and Analysis

Search Rate Disparities

Black drivers are searched at 2–5× the White rate (*see Figure 1*).

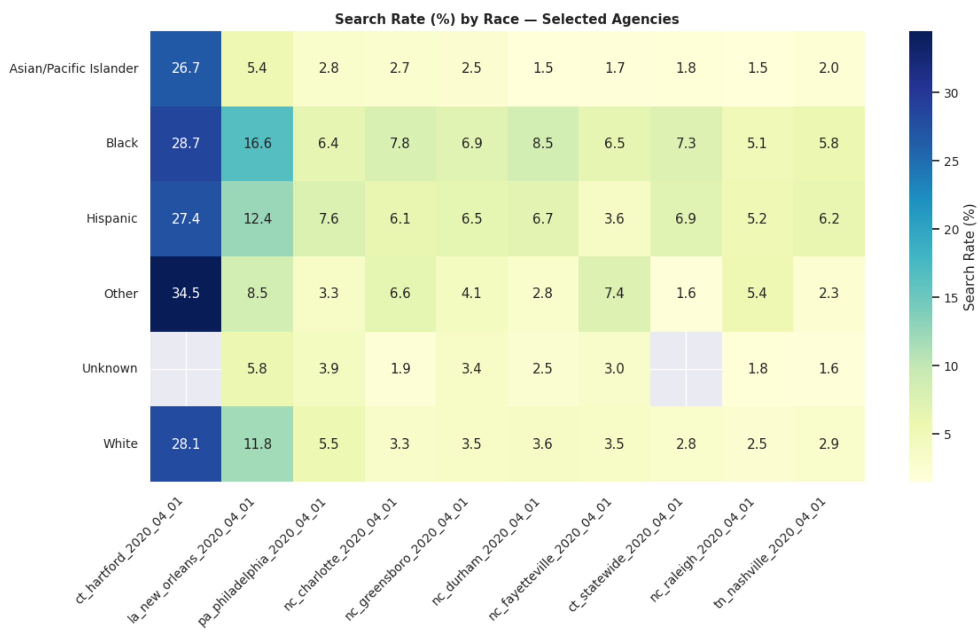

Figure 1: Search Rate (%) by Race across Selected Agencies.

Hit Rates

Black and Hispanic hit rates trail White drivers in every agency (*Figure 2*).

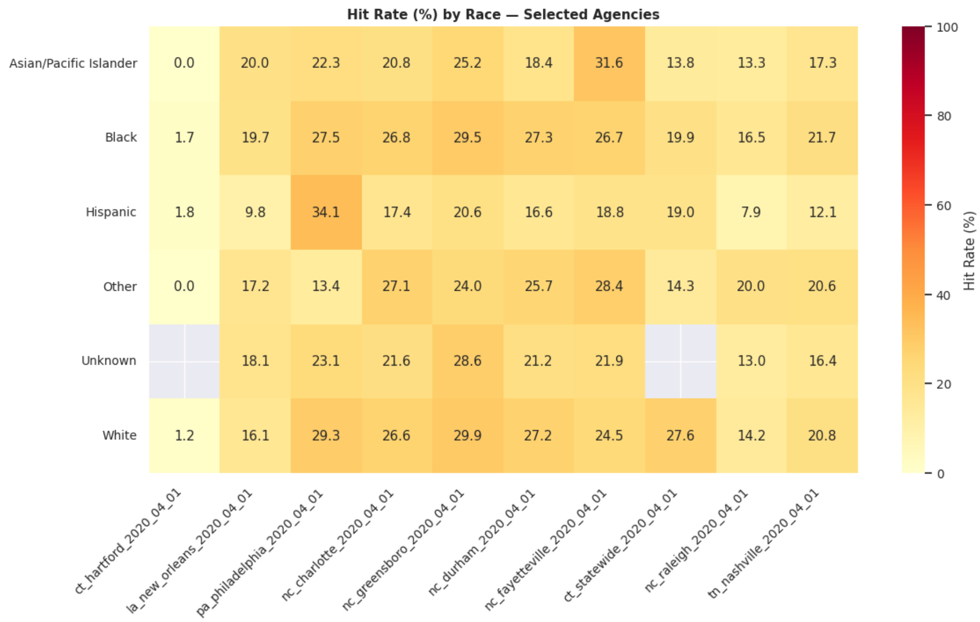

Figure 2: Hit Rate (%) by Race.

False Positives

False-positive searches disproportionately affect Black drivers (*Figure 3*).

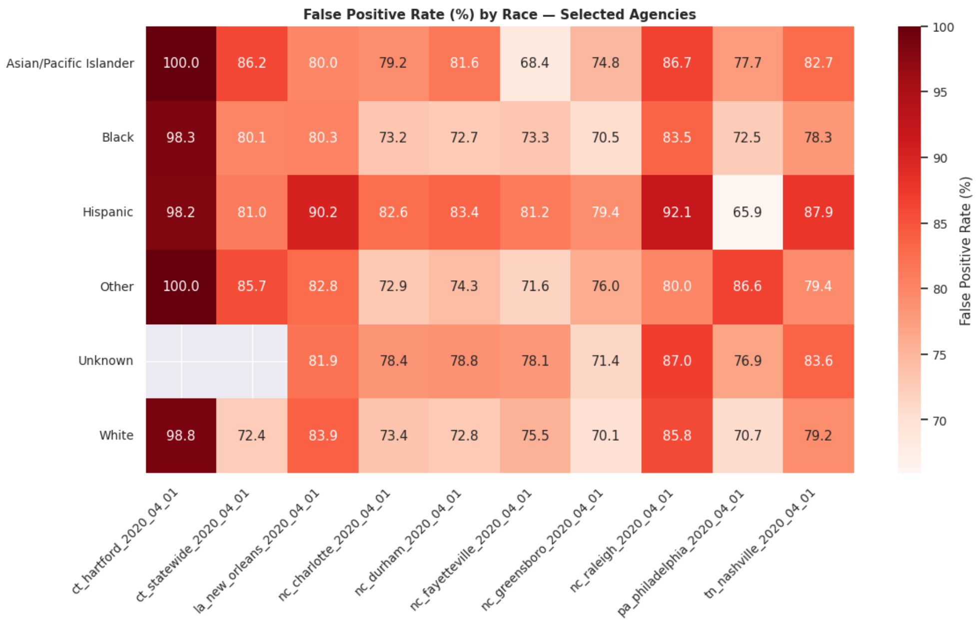

Figure 3: False-Positive Rate by Race.

Risk Ratios

Some agencies exceed a 5:1 Black-to-White search likelihood (*Figure 4*).

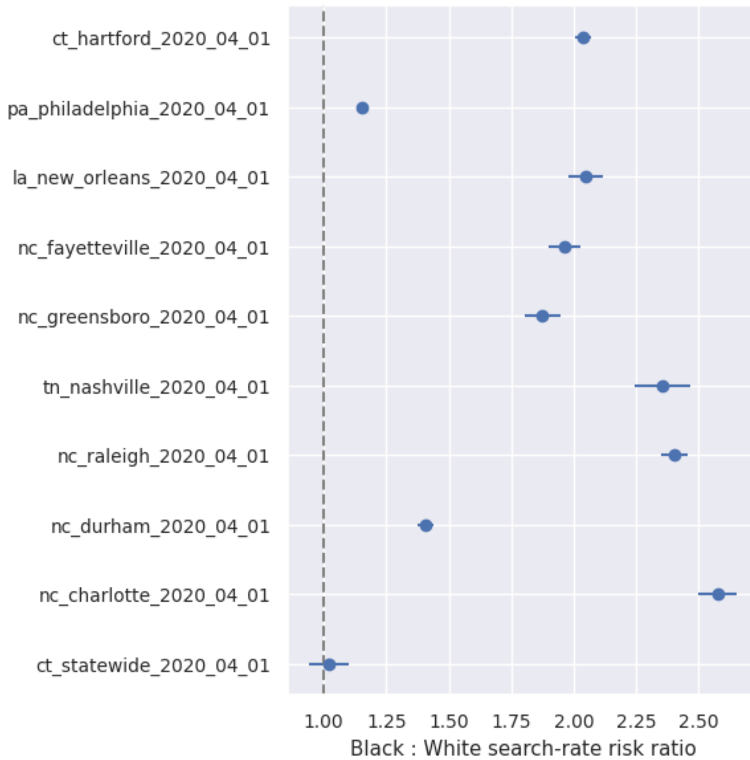

Figure 4: Black:White Search Risk Ratios.

Search Thresholds

Lower proxy thresholds for Black drivers suggest weaker evidentiary standards (*Figure 5*).

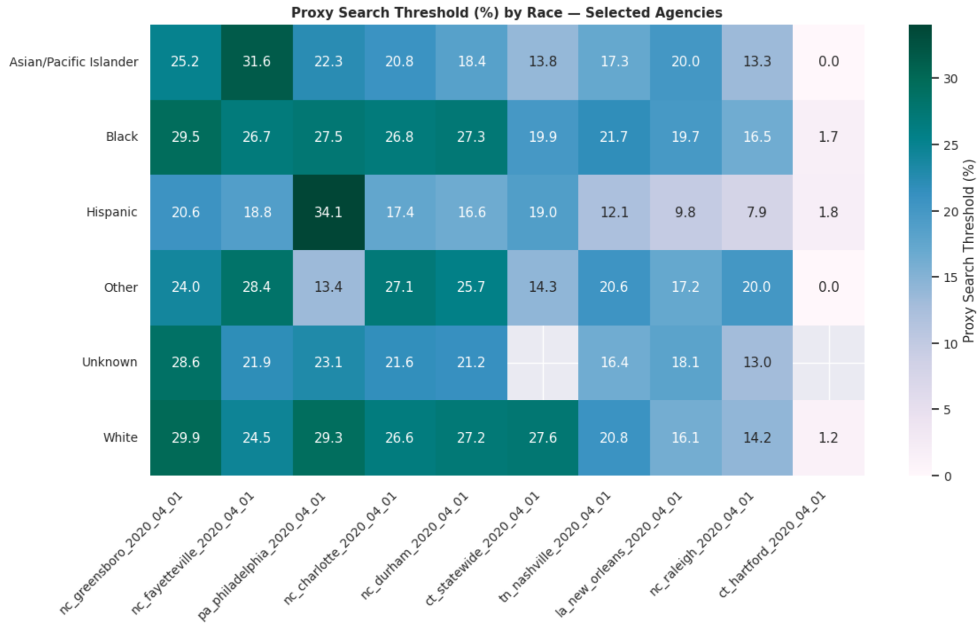

Figure 5: Proxy Thresholds by Race.

Search Volume

Aggregate volumes mirror rate disparities (*Figure 6*).

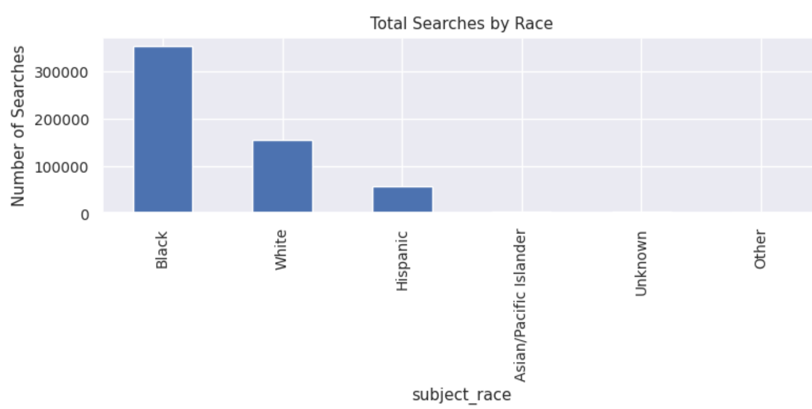

Figure 6: Total Searches by Race.

Demographic Context (U.S. Census Reference)

To provide critical context for the racial disparities observed in my analysis, I compare my findings to national-level population statistics. The following figures are drawn from the 2020 United States Census:

- White alone (non-Hispanic): 57.8%
- Black or African American alone (non-Hispanic): 12.1%
- Hispanic or Latino (of any race): 18.7%
- Asian alone: 6.0%
- Two or more races: 10.2%
- American Indian and Alaska Native alone: 1.1%
- Native Hawaiian and Other Pacific Islander alone: 0.2%

These proportions serve as a population baseline against which to evaluate enforcement practices. For example, while Black individuals make up roughly 12% of the national population, they often account for more than 30% of traffic searches in multiple jurisdictions, as shown earlier. This discrepancy supports the conclusion that stop-and-search practices are not race-neutral but exhibit patterns of structural bias.

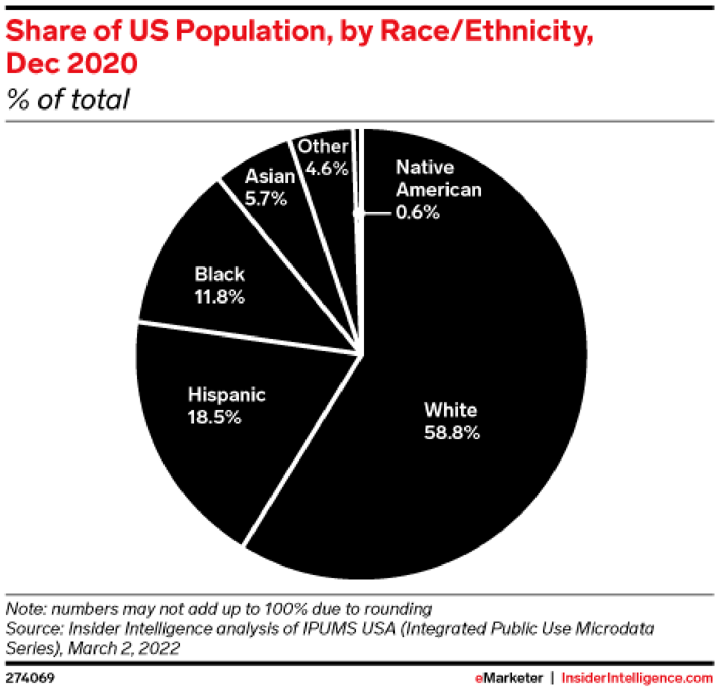

Figure 7: U.S. Census Racial Composition (2020). Baseline demographic data used for evaluating enforcement disparities.

6 Data Quality and Bias in Reporting

Across 88 agencies, 6.8% of rows lacked race data; "Unknown/Other" comprised 3.4 M additional rows. Labels were fragmented into **43** unique strings (top five: *White*, *Black*, *Hispanic*, *Asian*, *Other*). Such inconsistency obscures true disparities and can mislead models.

7 Ethical Implications and Feedback Loops

Under ACM 1.2 (avoid harm) and 1.4 (transparency), deploying models trained on biased, incomplete data is unethical. Feedback loops, where over-policing of Black communities generates more training data and thus more patrol recommendations, mirror the dynamics described by Lum and Isaac [2] and Eubanks [4].

8 Policy Recommendations

1. Mandate pre-deployment demographic audits of predictive-policing systems.
2. Prohibit use of datasets with unresolved bias or >5% demographic missingness.
3. Adopt a national schema for race/ethnicity (drop-down taxonomy, audit trail).
4. Require public release of model inputs, weights, and performance by subgroup.
5. Establish community oversight boards with veto power over algorithmic tools.

9 Conclusion

Traffic-stop data reveal entrenched racial bias: Black drivers face higher search likelihoods and lower evidence thresholds. If these patterns feed predictive models, policing enters a self-reinforcing cycle of injustice. *Fairness must start at the data layer.* Without rigorous audits and standardisation, no algorithmic veneer can guarantee justice.

References

[1] Pierson, E., et al. (2020). A large-scale analysis of racial disparities in police stops across the United States. *Nature Human Behaviour*, **4**(7), 736–745.

[2] Lum, K., & Isaac, W. (2016). To predict and serve? *Significance*, **13**(5), 14–19.

[3] Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, **104**(3), 671–732.

[4] Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press.

[5] U.S. Census Bureau. (2021). *2020 Census Redistricting (Public Law 94-171) Summary File*. Retrieved from `https://www.census.gov`.

[6] Stanford Open Policing Project. (2025). *Traffic-Stop Data Repository*. Retrieved April 23, 2025, from `https://openpolicing.stanford.edu/data/`

[7] Stanford Open Policing Project. (2025). *Open Policing Standardised CSV Archive [GitHub repository]*. Github. Retrieved April 23, 2025, from `https://github.com/5harad/openpolicing`

[8] Association for Computing Machinery. (2018). *ACM Code of Ethics and*

Professional Conduct. Retrieved from `https://www.acm.org/code-of-ethics`

[9] U.S. Census Bureau. (2021). *QuickFacts: United States (2020 Census release)*. Retrieved from `https://www.census.gov/quickfacts/fact/table/US`