Data feminism to mitigate algorithmic harm in tenant screening and promote housing justice

Author: Sonia Torres Rodríguez

Collaborators: Katie Fallon, Judah Axelrod, Brendan Chen

Contact Information: storresrod@urban.org

Introduction

The home is a critical space for the maintenance of patriarchal and capitalist systems (Austerberry and Watson 1981). This is exemplified by the not-too-distant past of credit, labor, and property ownership discrimination that systematically excluded women and racial/ethnic minorities from economic and political power (Kazis 2021).

These pasts and current realities are now further complicated by the introduction of AI in housing. Because they reproduce existing inequities, new AI technologies have been shown to develop and produce gendered and racialized algorithmic bias (Smith and Rustagi, 2021), and to pose significant anti-democracy implications via population surveillance.

This poster focuses on a small piece of this complex feminist housing and AI intersection: algorithms utilized in renter or tenant screening. In a small pilot exploratory data analysis, we employ an intersectional data feminist framework (D'Ignazio and Klein 2020) to document the introduction of possible sources of inequities in the aggregation of the data utilized in tenant screening reports. Specifically, we utilize criminal and eviction court record data from the state of Pennsylvania to replicate the cleaning and name matching that tenant screening companies undergo to prior to generating risk scores. We apply four of the seven principles in the data feminist framework to trace how inequitable distribution of power and assertions of the neutrality of algorithms can generate inequities. Our feminist exploratory data analysis reveals three potential ways that reliance on name-only matching, incorrect counting and cleaning of record data, and over-representation of records relative to population for marginalized populations, produces opportunities for algorithms to perpetuate existing patriarchal and racism in the housing space.

What is Tenant Screening?

Today, almost 90% of landlords in the United States use tenant screening reports (Choi et al. 2022). These third-party private corporations typically compile a combination of applicant credit, housing and employment, criminal, and eviction history data to provide to landlords as they decide whether to accept a renter's application or not. Companies are increasingly using this data to algorithmically assign scores or provide recommendations on whether to rent to a given prospective tenant. For example, Consumer Reports found that all eight prominent tenant screening companies in the US included an algorithmically generated score or a recommendation to accept or reject an applicant (Waddell 2021). These scores can be binary or a continuous scale.

An example of a tenant screening report is shown in Figure 1.

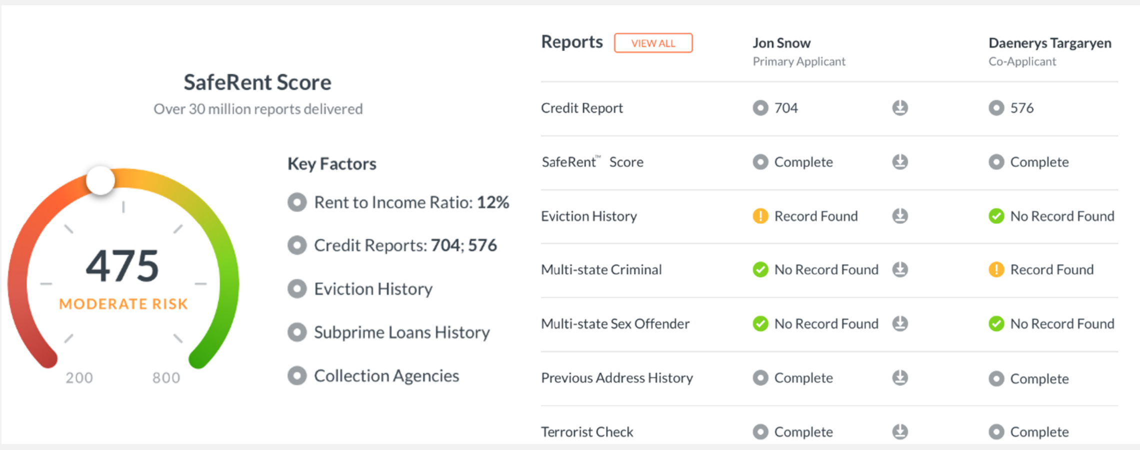

Figure 1: Tenant screening report sample from the CoreLogic company

Policy Background

There has been significant recent interest shown by the US federal government to regulate bias in tenant screening. For example:

- **US White House**: Published its 2023 Executive Order on Artificial Intelligence and its 2023 Statement on Actions to Protect Renters to state its interest in protecting renters from algorithmic discrimination or bias.
- **US Department of Housing and Urban Development (HUD):** Developed guidance for tenant screening companies and landlords to be in compliance with the Fair Chance at Housing Act. For example, in June 2022, HUD issued a memorandum that reviewed fair housing principles related to the use of criminal records.
- **US Consumer Financial Protection Bureau (CFPB):** Issued a report on the tenant background check industry in November 2022, which included an analysis of more than 24,000 complaints and describes how the industry's failures to remove wrong, old, or misleading information may contribute to higher costs and barriers to quality rental housing.
- **US Federal Trade Commission (FTC): C**ompiled a Know-Your-Rights resource for tenants undergoing a tenant screening check, including steps a prospective tenant can take if they believe their fair housing rights were violated in March 2024.

In a set of **virtual stakeholder interviews** with Community Legal Service attorneys and housing justice advocates, we surfaced two primary policy gaps:

- **Fair Housing Protections:** Although US fair housing law prohibits discrimination on protected classes, it is unclear if a landlord or software company can be held liable under fair housing laws for discrimination from a "race-neutral" algorithm.
- **Enforcement**: Even in US states like New Jersey, which have "Ban the Box" policies for housing that restrict the type and age of a criminal record that can be reviewed by a landlord, low enforcement of restrictions over data that can be used means landlords and tenant screening companies may still continue to utilize inadmissible data in an algorithmic risk scoring model.

Intersectional Data Feminist Framework

Catherine D'Ignazio and Dr. Lauren Klein outline present an intersectional feminist framework to unpack the power dynamics in data science and algorithms in their book titled Data Feminism (D'Ignazio and Klein, 2020). They outline seven principles of their framework, which include:

1. **Examine power:** Analyzing how power operates in the world.
2. **Challenge power:** Committing to challenging unequal power structures and working toward justice.
3. **Elevate emotion and embodiment:** Valuing multiple forms of knowledge, including the knowledge that comes from people as living, feeling bodies in the world.
4. **Rethink binaries and hierarchies:** Challenging the gender binary, along with other systems of counting and classification that perpetuate oppression.
5. **Embrace pluralism:** Insisting that the most complete knowledge comes from synthesizing multiple perspectives, with priority given to local, Indigenous, and experiential ways of knowing.
6. **Consider context:** Asserting that data are not neutral or objective. They are the products of unequal social relations, and this context is essential for conducting accurate, ethical analysis.
7. **Make labor visible:** Acknowledging the work of many hands. Data feminism makes this labor visible so that it can be recognized and valued.

Figure 2: Intersectional Data Feminist Framework by Catherine D'Ignazio and Dr. Lauren Klein

This framework spotlights the power dynamics of automated systems in a privatized and racialized housing market. For example:

Principle 1, Examine Power: Spotlighting the prevalence of error-prone data matching

As our societies become increasingly digitized, our names and the digital records attached to them grow in power and influence. This is apparent in tenant screening, where companies rely heavily on name-only matching, or even sometimes matching on the first few letters of names in "wild-card matching," which frequently leads to false-positives (Kirchner and Goldstein 2020). Asian and Latinx applicants are also in particular danger of being linked to data that is not theirs, because of lower name diversity (Census Bureau 2017).

Examining the power of names and the disparate impact of data-matching errors based on name-only matching is difficult because the data included in tenant screening reports have missing or unreliable race, ethnicity, and gender variables. In credit data, demographic variables are not included by design due to fair lending protections, which protect against discrimination on protected clauses; in criminal record data, race and ethnicity are typically based on observation by police or court staff, rather than self-identification, where observed race is found to be more likely to differ from self-identification for non-white individuals (Boehmer et al. 2002).

Principle 2, Challenge Power: Individualizing responsibility and burden for identifying errors

The current regulatory landscape diffuses and obscures responsibility for discriminatory impact away from companies and landlords when using algorithms to screen tenants. This obfuscation of power also correlates with placing the onus of identifying and correcting errors in reports on prospective tenants. One common source of errors is tenant screening reports includes incorrectly characterizing the presence of any record under someone's name as a negative factor. Even if a prospective tenant is not found guilty of a crime or formally evicted, for example if they were only arrested or served an eviction but not found guilty, tenant screening companies still sometimes disqualify those prospective tenants due to the mere presence of a record (Housing Justice Center 2021). This penalizes candidates for membership in overpoliced and over-filed communities. Challenging the power of screening companies will require an update of fair housing protections and enforcement.

Principle 4, Rethink Binaries and Hierarchies: Binary Classification of Good/Bad Renters

Tenant screening companies do not publish the methods or algorithms they use to construct their scores due to proprietary protections. Lack of transparency around data sources, data fields, model selection, weighing of certain characteristics, creates a binary and static perspective on what constitutes a "good" or "bad" renter.

Principle 6, Consider Context: Questioning the objectivity of behavior predicting algorithms and the data that feed them

People of color are overrepresented in arrest and criminal record data due to racist policing and incarceration practices (Pinard 2013). For example, in a 2020 study, Black individuals made up 19.9% of all adult renters in sampled counties, but 32.7% of all eviction filing defendants (Hepburn, Louis, and Desmond 2020). Tenant screening scores create a guise of objectivity around untested predictors and a sense of neutrality that papers over the ways marginalized communities are overpoliced and over-evicted into data.

Research Questions

Our pilot project aims to study the data acquisition and aggregation process used by companies to compile tenant screening reports, with the goal of highlighting new intersectional feminist opportunities for regulation and oversight. These include:

- **Q1:** How can a data feminist framework highlight the effects of variability in the quality of data used by tenant screening companies to assess tenants, specifically linked eviction and criminal record data?
- **Q2:** What are the feminist implications of data aggregation and linking across eviction and criminal records, including in name matching?

Methods

We constructed a pilot database that aggregates information on eviction filing and criminal data and probabilistic (or fuzzy) matching techniques for the US state of Pennsylvania. All data is from sourced from the Administrative Office of Pennsylvania Courts and includes criminal and landlord/tenant record data from 2014-2024.

Following an intersectional data feminism framework (D'Ignazio and Klein, 2020), we employ a feminist exploratory data analysis (EDA) to study the inherent biases, power dynamics, and inequities associated with each of the data sources, and summarize how various linking processes may disparately impact marginalized communities.

Preliminary Findings

Finding 1: Error-prone data matching

Given the common practice of name-only matching by tenant screening companies, we attempt to quantify the potential drawbacks of that strategy. After finding the total number of unique combinations of name (first and last), date of birth, and zip code, we find that 43% of those unique combinations also are unique on just the name and date of birth field. However, we find that 7% of those unique combinations have the same name but a different date of birth. This shows that name-only matching may generate false positives that could be avoided with the utilization of other identifying information. This could be especially important for marginalized communities with less diversity of names.

Finding 2: Identifying potential sources of error

When analyzing the details included in the criminal and landlord/tenant records, we find a complicated picture. For the criminal data, we find 182 case dispositions, which describe outcomes, and over 5000 offense descriptions, which describe criminal charges and severity. Figure 3 shows that about 9% and 4% of cases for landlord/tenant and criminal records respectively are ruled for the defendant, settled, or withdrawn, and that about 45% of records include procedural information (about an incomplete trial). Without care, it would be very simple to accidentally include these records as an incorrectly presumed guilty charge or eviction.

Type of Record	Procedural Disposition	Judgement for Plaintiff or Guilty	Judgement for Defendant, Settled, Withdrawn	Missing
Landlord/Tenant	N/A	90.69%	8.79%	N/A
Criminal	44.92%	31.9%	4.13%	8.82%

Figure 3: Distribution of case dispositions in criminal and landlord/tenant court record data

Finding 3: Structural racism in algorithm input data

To assess the extent to which marginalized communities are over-represented in criminal and tenant-landlord data, we conduct a simple difference between demographic population and demographic presence in court data. Over-representation, which may not necessarily mean guilt or eviction, may indicate greater likelihood to have error-prone data included in a tenant screening report.

We find that landlord and tenant has an exceedingly high missingness for race (79.7%) and gender (75.7%), making it not usable for this exercise. This high level of missingness was corroborated by the Eviction Lab's recent efforts to link ACS microdata with eviction data (Graetz et al. 2023). Still, it is useful to keep in mind that there has been a documented pattern of landlords on occasion serially filing evictions, and that measures to prevent landlords from this pattern of behavior lead to the greatest decrease in evictions in Black-majority neighborhoods (Gomory et al. 2023).

There was almost no missingness of race in the criminal record data, though note that the data is observed (not self-identification). Figure 4 below confirms a pattern of over-representation of court record data compared to the underlying population in a given zip code. The context of over-representation for Black individuals is observed in larger cities in Pennsylvania, particularly in Pittsburgh.

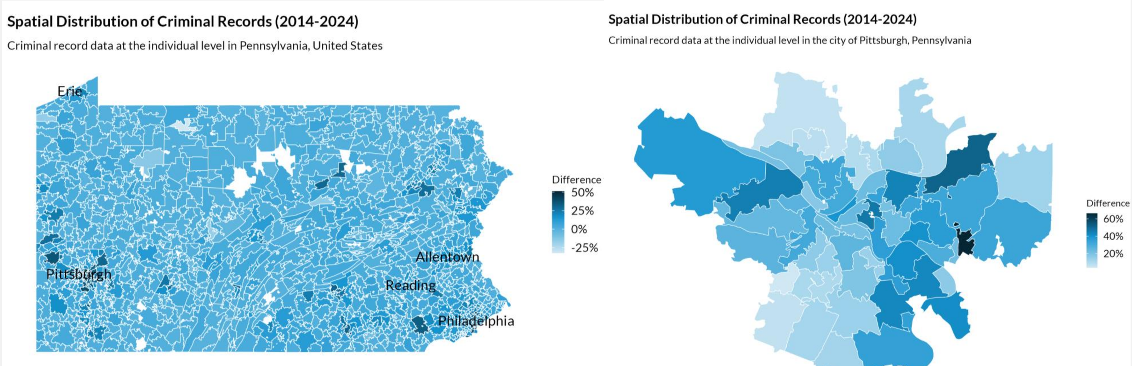

Figure 4: Population comparison between Black population and Black individuals with a criminal court record at the zip code level

Limitations and Next Steps

- **Limitations:** This feminist EDA is preliminary and does not include name matching across landlord/tenant and criminal data. Additionally, due to data access limitations, we do not include credit data in our analysis, which limits our ability to replicate scoring.
- **Next steps:** Explore possibilities for extending EDA and scoring replication, as well as potential regulatory procedures for the data aggregation processes of tenant screening companies.

References

Austerberry, H., & Watson, S. 1981. A Woman's Place: A Feminist Approach to Housing in Britain. Feminist Review, 8(1), 49-62. https://doi.org/10.1057/fr.1981.11

Choi, J.H., Goodman, L., & Pang, D. 2022. The Real Rental Housing Crisis is on the Horizon. Urban Institute. https://www.urban.org/urban-wire/real-rental-housing-crisis-horizon

D'Ignazio C., & Klein, L. 2020. Data Feminism. The MIT Press.

Kazis, Noah. 2021. Fair Housing for a Non-Sexist City. 134 Harv. L. Rev. 1683 (2021). http://dx.doi.org/10.2139/ssrn.3552190

Graetz, N., Gershenson, C., Hepburn, P., & Desmond, M. 2023. Who is Evicted in America? Eviction Lab. https://evictionlab.org/who-is-evicted-in-america/#methods

Gomory, H., Massey, D.S., Hendrickson, J.R., & Desmond, M. 2023. The Racially Disparate Influence of Filing Fees on Eviction Rates. Housing Policy Debate, 33(6), 1463-1483. https://doi.org/10.1080/10511482.2023.2212662

Hepburn, Peter; Renee Louis; and Matt Desmond. 2020. "Racial and Gender Disparities Among Evicted Americans". The Eviction Lab. https://evictionlab.org/demographics-of-eviction/

Housing Justice Center. 2021. "Opening the Door: Tenant Screening and Selection – How it works in the Twin Cities Metro Area and Opportunities for Improvement". Housing Justice Center and Family Housing Fund. https://www.hjcmn.org/wp-content/uploads/2021/04/Tenant-Screening-Report.pdf

Kirchner, Lauren, and Matthew Goldstein. 2020. "Access Denied: Faulty Automated Background Checks Freeze Out Renters." The Markup. https://themarkup.org/locked-out/2020/05/28/access-denied-faulty-automated-background-checks-freeze-out-renters

Pinard, Michael. 2013. Criminal Records, Race and Redemption, 16 N.Y.U. J. Legis. & PUB. POL'y 963 (2013).Purser, G. (2016). The circle of dispossession: Evicting the urban poor in Baltimore. Critical Sociology, 42(3), 393–415.

Smith, G., & Rustagi, I. 2021. When Good Algorithms Go Sexist: Why and How to Advance AI Gender Equity. Stanford Social Innovation Review. https://doi.org/10.48558/A179-B138

U.S. Census Bureau. 2017. "What's in a Name? Hispanic Surnames Rise in Popularity." https://www.census.gov/library/stories/2017/08/what-is-in-a-name.html