Postscript: Reporting Racial Equity in Lending

My investigation began with a question: in the wake of George Floyd's murder and the flood of corporate pledges to address systemic racism, had the largest U.S. mortgage lenders actually narrowed racial approval gaps? To answer it, I built a multi-year lending dataset, standardized lender identities, analyzed public filings, and spoke to experts in fair housing and mortgage markets.

Discovering the Disparity

The project started when I noticed a gap between public commitments and lending results. JPMorgan Chase, Bank of America, and Wells Fargo all announced billion-dollar racial equity initiatives in 2020, yet watchdog groups continued to report wide denial gaps for Black and Latino borrowers. This difference between stated intentions and measurable outcomes motivated me to compare what the banks promised with what they delivered.

Building the Dataset

I chose the public Home Mortgage Disclosure Act (HMDA) dataset because it records every mortgage application and outcome in the United States, including borrower race, property type, and lender identity. I downloaded annual HMDA files for 2018 through 2024, totaling roughly 124 million records, and processed them in Python.

The raw data required extensive cleaning. Race was coded numerically, with sub-codes for Asian subgroups but no specific "Latino" race code, as Latino identity is stored in a separate ethnicity field. Financial fields such as income, loan amount, property value, and debt-to-income ratio contained non-numeric characters, placeholders such as "NA," and inconsistent formats. The 2024 release was in text format, so I converted it to CSV and validated it for parsing errors.

I created a race-mapping dictionary that grouped sub-codes into broad categories of Asian, Black, and White. Parsing functions converted all financial fields to numeric values. A unique row ID was created by combining the lender's Legal Entity Identifier (LEI) with the record index to ensure each record could be tracked through the process.

To track individual records through the pipeline, I generated a unique row ID by concatenating each lender's Legal Entity Identifier (LEI) with the row index.

Applying Strict Filter

To make results comparable across lenders and years, I applied a strict filter that matched standards used in regulatory and academic studies. The filter retained only:

Conventional (loan_type = 1) first-lien mortgages

Owner-occupied, site-built properties with four or fewer units

Applications with complete income, loan amount, property value, and debt-to-income ratio data

Applications coded as originated, approved but not accepted, or denied

Each condition was applied in sequence, with the number of dropped rows logged at every step. I saved the filtered dataset for each year along with a drop log. The strict filter reduced the dataset to roughly 10 million comparable records.

Calculating Approval Rate Gaps

From the filtered data, I grouped records by lender, state, year, and race, counting approvals, denials, and total applications using pandas (Python). Approval rates were calculated as approvals divided by total applications. The approval gap was defined as the White approval rate minus the Black approval rate and, where ethnicity data was available, the White approval rate minus the Latino approval rate. I also calculated median loan amounts by race and year.

Analyzing Corporate Language

To see whether rhetoric and results moved together, I collected annual reports for the largest banks and major non-bank lenders (if they exist since private companies are not obliged to produce annual reports). Using a curated set of diversity, equity, and inclusion (DEI) terms, I counted case-insensitive matches in each report. I converted the reports to plain text using PyMuPDF to ensure consistent parsing.

The counts showed a rise in explicit racial equity terms in 2020 and 2021, followed by a sharp drop in 2022. By 2023 and 2024, banks had replaced explicit terms like “racial equity” and “systemic racism” with broader language such as “economic mobility” and “financial inclusion.”

Resolving Data Gaps

A small but important detail: the HMDA dataset lists lenders by their LEI, and my name-matching file had missing entries for some institutions. To ensure accuracy, I wrote an asynchronous scraper using the Playwright library. For each missing name, the script opened the company’s Bloomberg profile, waited for the element containing the legal name to load, extracted the text and filled it into the dataset. The process took several hours but prevented erroneous attribution in the final statistics. The rest of the missing files, I manually went through and filled them in and cross-checked that the names filed were correct.

Cross-checking with Experts and Enforcement Actions

To understand why gaps persisted, I interviewed experts on fair lending.

I also examined Department of Justice redlining cases under the Combatting Redlining Initiative. Gaps narrowed after some settlements but often widened again. In April 2025, a federal executive order ended DEI programs and halted disparate-impact enforcement, making it harder to prove discrimination. Fair housing attorneys warned that disparities would likely persist.

Visualization and Interpretation

The processed data supported several visualizations:

- Choropleth maps showing state-level approval gaps
- Time series plots of approval rates by bank and race
- Line charts showing changes in DEI language usage in annual reports

These visualizations made it possible to see how approval gaps persisted while explicit equity language declined.

Limitations

The public HMDA data does not include credit scores or some underwriting variables, so the analysis cannot determine whether gaps are the result of unlawful discrimination. Latino borrower analysis depends on the ethnicity field, which is incomplete in some cases.

Public HMDA data does not include key underwriting factors such as credit scores, so it cannot show whether lenders treated comparable applicants differently. Fairway Mortgage said that without this information, no conclusions about its fair-lending practices can be drawn. Other data sources show that U.S. credit scores have risen by only about 10 points in the past seven years, remaining in the same bracket generally considered favorable for mortgage lending.

The DEI vocabularies cannot capture every euphemism companies might use for racial equity. And non-bank lenders rarely publish detailed reports, so my analysis of their commitments relies on third-party research.

Reflection

Through a transparent filtering pipeline, reproducible statistical measures, corporate language tracking, and expert context, the investigation documented that post-2020 equity pledges did not produce sustained reductions in racial mortgage approval gaps. The disparities remain, the language has softened, and policy changes may weaken oversight.

List of DEI words used to filter annual reports:

 # Racial justice and equity terms
"racial justice", "racial equity", "racial equality",
"systemic racism", "institutional racism", "structural racism",
"racial bias", "racial discrimination", "racial disparities",
"antiracism", "anti-racism", "racial healing",
"white privilege", "white supremacy", "racial privilege",

 # Gender equity and inclusion
"gender equity", "gender equality", "gender justice",
"gender bias", "gender discrimination", "gender disparities",
"gender inclusivity", "gender-neutral", "gender identity",
"sexual orientation", "lgbtq+", "transgender rights",

 # General DEI terminology
"diversity and inclusion", "equity and inclusion",
"belonging and inclusion", "inclusive excellence",
"cultural competency", "cultural humility", "unconscious bias",
"implicit bias", "microaggressions", "allyship",
"intersectionality", "social justice", "inclusive leadership",
"inclusive practices", "equity lens", "inclusive design",
"accessibility", "neurodiversity", "ableism",

 # Workplace DEI terms
"inclusive hiring", "diverse workforce", "pay equity",
"equal opportunity", "affirmative action", "minority representation",
"underrepresented groups", "marginalized communities",
"employee resource groups", "inclusion training",
"diversity metrics", "equity audit"