Disparities in Mortgage Lending: A Statistical Analysis of Race and Gender Bias in New York State

VICTOR ZHANG, ELLIE KOGAN, ERIN MUTCHEK, ZHEKA CHYZHYKOVA AND MICHAEL LEE, CMPU 250, Vassar College

1 INTRODUCTION

Over the past several decades, rising rents have outpaced stagnant wages, leading to a sharp decline in accessible, budget-friendly housing [7]. According to the U.S. Department of Housing and Urban Development (HUD), the national homeownership rate declined from 69% in 2004 to around 65% in 2020, reflecting growing instability in housing accessibility [9]. In New York State, homeownership rates have consistently remained below the national average, hovering around 54% as of 2022, according to the U.S. Census Bureau [8]. As a result, home ownership has become increasingly more important as a means to build wealth, especially among Black Americans.

Historically, Black homeowners have long faced systemic barriers to owning property due to the legacies of slavery, redlining, and ongoing discriminatory lending practices. Homes in predominantly Black neighborhoods are significantly devalued compared to similar homes in White neighborhoods, and Black Americans face discriminatory lending practices in the mortgage application process [6]. Traditionally, decisions about mortgage loan approval were made by human underwriters who reviewed a combination of financial and personal information provided by applicants. During the pre-approval process, applicants submit basic budget information and lenders use credit reports to determine the maximum amount they are willing to lend. After selecting a property, applicants submit comprehensive employment, income, asset, debt, property, and credit history information [2]. Using this information, lenders choose to approve or deny applications, generating a loan estimate in the process.

However, today lenders increasingly rely on algorithms with the goal of being fairer, more accurate, and more efficient than their human counterparts. Mortgage lenders hope these algorithms will streamline decision-making and eliminate human bias and inconsistency. Hidden algorithms are now used to evaluate home prices and mortgage applications, determining final interest rates [1]. During the application process, these algorithms assess risk and predict loan repayment; applicants who are misclassified may receive higher interest rates, smaller loan amounts, or outright rejections, thereby increasing financial burden and reducing mobility. While these systems promise objectivity, growing evidence suggests they may perpetuate the very biases they are intended to eliminate.

In their August 2021 investigation, "The Secret Bias Hidden in Mortgage-Approval Algorithms," The Markup's Emmanuel Martinez and Lauren Kirchner analyzed over 2 million conventional mortgage applications from 2019. Even after accounting for debt-to-income ratio, loan-to-value ratio, and credit score—factors often cited to explain disparities—people of color were denied mortgages at significantly higher rates than White applicants. Nationally, Black applicants were 80% more likely to be denied than White applicants with similar financial profiles [4]. High-income Black applicants with less debt were denied more often than high-income White applicants with more debt, suggesting that algorithmic underwriting may reinforce rather than eliminate existing biases.

This paper examines whether and how bias is embedded within algorithmic decision making in the mortgage application process, with a specific focus on New York's housing market. Although algorithms are used at various stages: prequalification, application review, and fraud detection, our analysis centers on the underwriting stage, where risk assessments and loan recommendations directly affect approval outcomes and interest rates.

We investigate three research questions:

(1) Do applicants of different race and gender backgrounds experience significantly different loan amounts?
(2) What disparities, if any, exist in the loan interest rates awarded to applicants of different race and gender backgrounds?
(3) How can we create and assess a more fair loan-approval model while still prioritizing accuracy?

To address these questions, we analyze a detailed HMDA Loan Application Register (LAR) dataset for 2023. We begin with exploratory data analysis to uncover general trends and establish baseline comparisons for predictive variables. We then construct linear regression models to assess the influence of race and gender on loan amounts and interest rates, comparing their impact against other financial indicators. Finally, we build and evaluate machine-learning classifiers to measure accuracy and fairness across demographic groups. Our findings demonstrate that racial and gender-based biases persist in loan approval and pricing, and that models trained on this data risk perpetuating systemic inequities.

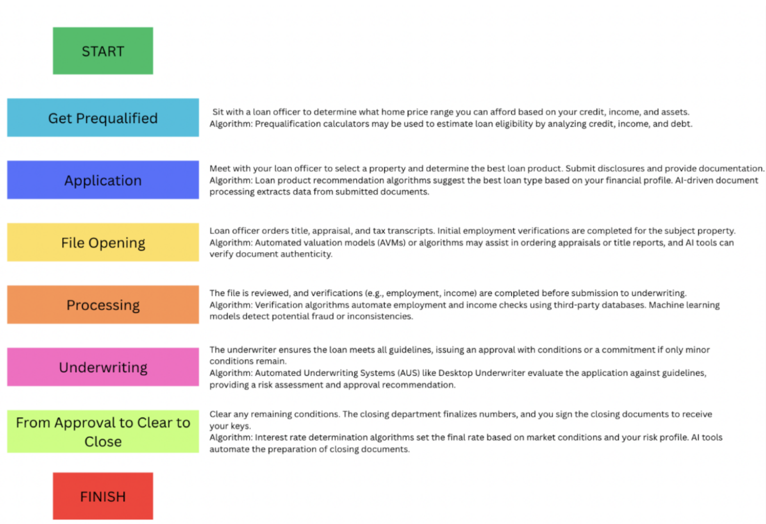

Fig. 1. Loan Application Process

2 DATA

The data set comes from the 2023 Home Mortgage Disclosure Act's (HMDA) Loan Application Register (LAR) dataset. HMDA is a federal law that requires financial institutions to report data on mortgage applications, originations, and purchases. The data is collected and maintained by the Consumer Financial Protection Bureau (CFPB) and made available through the Federal Financial Institutions Examination Council (FFIEC). Financial institutions that meet certain criteria (based on asset size, location, and lending activity) are required to report HMDA data. They collect this data as part of their mortgage application and origination processes throughout the year. The data is submitted electronically to the CFPB in a standardized format and the CFPB then processes, validates, and releases the data to the public through the FFIEC website.

We chose the following variables because they were either related to the financial or demographic composition of the applicant. Specifically, we focused on 20 variables including `derived_race`, `derived_ethnicity`, `income`, `loan_amount`, `interest_rate`, and `rate_spread`. Records with missing or unspecified race information were removed to ensure clarity in group-level comparisons. We also cleaned numerical fields—like `debt_to_income_ratio`, `property_value`, and `interest_rate`—by coercing to numeric types and dropping null values where appropriate.

To clean the data, we changed all the type objects into either categorical, binary, or numerical values. This is to help make them more usable in the models and data analysis. For some predictors that had numerical values to represent categorical variables, we removed some of the numbers and only used some. For example, for `action_taken` we only used 1, 2, and 7 because these outcomes were most relevant to the specific analysis we intended to perform.

We excluded categories like 3 (Application denied), 4 (Withdrawn), 5 (File closed), and 6 (Loan purchased) because they either represented intermediate steps, incomplete processes, or outcomes not directly related to our focus on application success and preapproval outcomes. We also removed all the rows that either had a missing value or an unusable value such as NaN. Ultimately, our clean dataset was left with 233,145 applications. The complete codebook can be found in Appendix, under Table 1.

3 METHODOLOGY

In order to investigate the possibility of creating and assessing a more fair loan model while still prioritizing accuracy, we chose to train a set of three statistical models to assess bias and then create three predictive ML models as proof of concept. By analyzing the behavior of the first three models, we aimed to investigate how the HDMA model may have produced a biased dataset that could be harmful to use in the future. In order to hone in on White and Black applicants specifically, we converted applicant race into a binary variable where '1' denotes a White applicant and '0' a Black applicant. In addition, we combined the outcomes listed as "Loan Originated" and "Application approved but not accepted" to indicate success, and "Loan Denied" or "Pre Approval request denied" to indicate rejection. For gender, we encoded it as a categorical variable where "1" denoted male, "2" denoted female, "3" denoted not received, and "6" denoted that the applicant selected both male, and female.

Our first objective was to assess whether race and gender have a statistically significant effect on loan outcomes. To do so, we constructed three Ordinary Least Squares (OLS) regression models using the statsmodels library. These models employed a consistent set of predictors representing the applicant's demographic profile (e.g., race, sex, age), financial condition (e.g., income, debt-to-income ratio, credit score), and loan characteristics (e.g., loan type, purpose, property value). In Model 1 we wanted to predict the loan amount, aiming to identify disparities in how much is lent to similarly qualified applicants based on demographic attributes. Model 2 predicted interest rate, evaluating whether borrower characteristics influence the cost of borrowing. And Model 3 predicted rate spread, measuring how much more applicants pay relative to the best-available market rates. Categorical variables were one-hot encoded using `pandas.get_dummies()`, and we ensured all features were numeric and standardized in format. Each regression model included a constant term and was evaluated using R-squared values and significance levels to interpret explanatory power and identify potential systemic biases.

The second part of our methodology addressed our third research question. We created three predictive ML models to demonstrate how any future model trained on this data may exhibit bias. The algorithms we selected were K-Nearest Neighbors, a Decision Tree Classifier, and Logistic Regression, because they are appropriate for categorical outcomes such as approval or denial. Each took in a small set of key financial factors as input variables and predicted either an approval or denial (regardless of when it occurred in the application process). These variables included debt-to-income ratio, a categorical variable with different percentage ranges; and then a list of continuous numeric values: income, loan to value ratio, loan amount, and property value.

Figure 2 below shows the overall odds of success per racial group in the ground truth outcomes for reference. To check for bias in our own predictive models, we compared the accuracy rates between White and Black applicants. Then, we retrained the models to include race as an explicit input and compared fairness metrics of selection rate, false positive rate, and false negative rate for Black versus White applicants. By demonstrating racial bias in the application decisions our models would have advised, we hope to illustrate the potential dangers of using the outcomes of one predictive model as inputs to another.

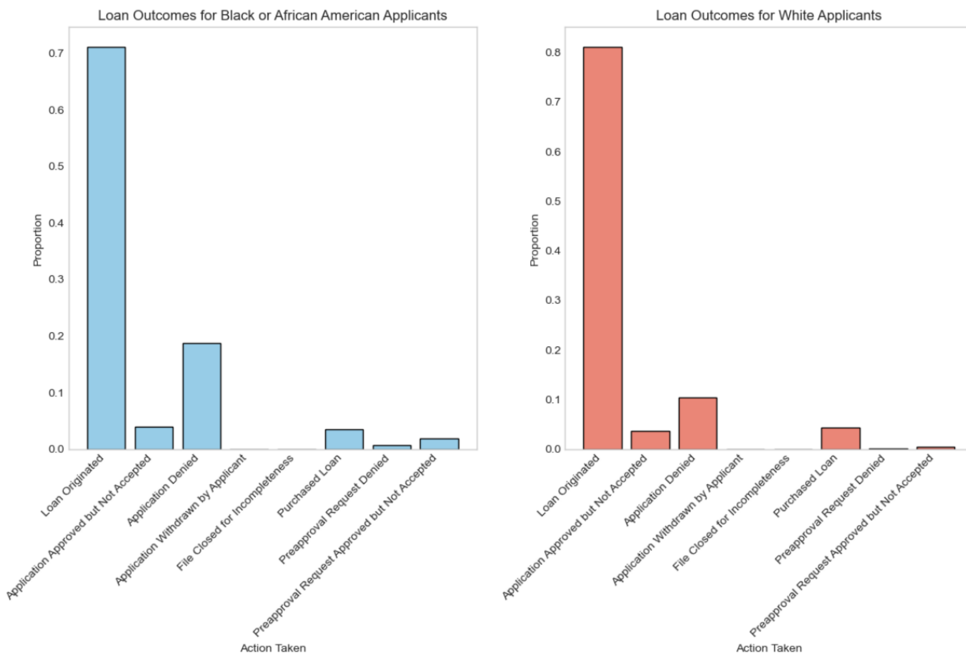

Fig. 2. Loan Outcomes per Racial Group

4 RESULTS

Our first goal was to train a model on the different terms of the loan such as loan amount, interest rate, and interest spread to get a better understanding of how applicant demographics and financial status influence the loan terms that the accepted applicants ultimately receive.

Our first multiple linear regression, used our selected variables to predict the loan amount given to applicants. Our model overall had a very strong fit and had a r-squared value of 0.790, indicating it does a very good job of predicting the variance in the data. Age followed a predictable pattern as applicants got older their loan amounts would increase. All p-values for age outside of the 64+ demographic were statistically significant with p-values less than 0.001. One interesting thing to note was that applicants between the ages 35–44 were predicted to have the highest loan amounts with a coefficient of $22,100. This model supported the conclusion when controlling for all other factors, women were given \$15,570 less than men (p-value < 0.001) and that White applicants were given \$16,760 more than Black applicants (p-value < 0.001). In terms of the financial variables, income and property value were the most significant in predicting the loan outcome as expected due to their statistical significance and coefficient values. Income had a positive coefficient of 46.446 (p-value < 0.001), debt-to-income ratio had a positive coefficient of 360.5253 (p-value = 0.04), and property value had a positive coefficient of 0.4472 (p-value < 0.001).

Our second regression model used our selected variables to predict the interest rates given to applicants. Our model overall had a weaker fit with a r-squared of 0.171. While sex was not statistically significant, White applicants were predicted to have slightly higher rates than Black applicants with an increase of 0.0542 (p-value = 0.007). For predicting interest rates, the financial predictors were the most strongly correlated. Income had a coefficient of -0.496×10^{-5} (p-value = 0.001) and debt-to-income ratio had a coefficient of 0.0190 (p-value < 0.001) indicating that as income increases and debt-to-income ratio decreases, interest rates decrease as expected.

Our third regression model used the selected variables to predict the interest spread given to applicants. Interest spread refers to the interest rates adjusted for the market value, standardizing interest rates for different economic periods. For this model, neither sex nor race were statistically significant. Once again for predicting interest spread, the financial predictors were the most strongly correlated: income had a coefficient of -5.897×10^{-5} (p-value < 0.001) and debt-to-income ratio had a coefficient of 0.0121 (p-value < 0.001), indicating that as income increases and debt-to-income ratio decreases, interest spread decreases.

The second part of our methodology was to train predictive models on loan application outcomes (denial vs. acceptance). This idea was inspired by previous literature focusing on bias in the loan application process. First, we made a Logistic Regression model, which yielded an accuracy of 86.49%. Our Decision Tree model yielded an accuracy of 78.51%. Our K-Nearest Neighbors model yielded an accuracy of 83.59%. We then assessed bias in our models by comparing accuracy rates across races, namely between White and Black individuals. All three models showed a difference of over 5% in accuracy rates between Black and White individuals. Next, we retrained the three predictive models to also take in race as an input so that we could calculate fairness metrics. This time, our models' accuracy rates at predicting the real-world, ground-truth outcomes was at least 8% lower for Black applicants across the board. The decision-tree model exhibited notable bias in other metrics as well: the selection rate was higher for White individuals (85.4%) than for Black individuals (76.0%). FPR was higher for White individuals (75.6%) than Black individuals (63.7%), meaning that when the model incorrectly predicted an acceptance it was more likely to be for a White applicant. FNR was higher for Black individuals (20.6%) than White individuals (13.1%), meaning that when the model incorrectly predicted a denial it was more likely to be for a Black applicant. In conclusion, the algorithm used to train the model influenced how many signs of racial bias were found, but all of the predictive models had some racial discrepancies.

5 DISCUSSION

Our analysis demonstrates that despite the promise of fairness and efficiency, algorithmic underwriting systems can reproduce and in some cases amplify the racial and gender disparities that have long existed in mortgage lending. By applying both linear regression and machine-learning classifiers to the 2023 HMDA LAR dataset for New York, we directly connect our empirical findings to each of our three research questions and highlight the urgent need for fairness interventions.

First, when we asked whether loan amount allocations differ by race and gender (RQ1), our regression results were unequivocal. Controlling for income, debt-to-income ratio, property value, age, and loan purpose, Model 1 predicts that White applicants receive \$16,760 more financing than Black applicants, and male applicants receive \$15,570 more than female applicants (both $p < 0.001$). These gaps are economically substantial—on the order of a typical down payment—and they persist amid New York's challenging housing landscape, where homeownership remains one of the principal avenues for building intergenerational wealth. The size of these coefficients underscores how automated underwriting could codify historical patterns of devaluation in predominantly Black neighborhoods and long-standing gender credit gaps [6].

Our second and third models examined interest pricing (RQ2). Model 2 reveals that White applicants are charged base interest rates 0.054 percentage points higher than Black applicants ($p = 0.007$), a result that runs counter to our hypothesis but nonetheless signals demographic signal leakage in pricing algorithms. Although numerically small, this rate differential accrues meaningful cost over a 30-year mortgage term—particularly in markets where nominal rates have hovered near historic lows [1]. When we shift to Model 3's "rate spread" (interest relative to benchmark market rates), race and gender lose statistical significance, indicating that while headline rate-setting may embed demographic proxies, market-driven competitive pressures largely standardize spreads.

Because the racial and gender disparities are strongest for the loan amount, whereas interest rate and rate spread are best explained by financial terms, this indicates that there may be more bias in the pre-approval process rather than the application process. During the pre-approval process, in addition to an approval applicants also receive an estimate for how much money they can receive. Influencing the loan amount that applicants ultimately ask for on their applications, if Black and female applicants are receiving lower estimates affecting their loan amounts, but ultimately similar interest rates (determined in the underwriting process), this would indicate bias in the pre-approval process and relative fairness in the underwriting process.

To explore how a fairer approval system might balance equity with predictive performance (RQ3), we constructed three classifiers—logistic regression, decision tree, and K-nearest neighbors—to predict approval outcomes. Even without race as an input, each model predicts approvals for Black applicants with at least a five-point lower accuracy than for White applicants, demonstrating that historical HMDA decisions encode bias. Introducing race as a feature further widens this gap by eight points or more, and fairness metrics diverge sharply—for example, the decision tree's false negative rate is 20.6% for Black applicants versus 13.1% for White applicants. These results mirror large-scale audits of automated underwriting [4], showing that naïve machine-learning pipelines can not only inherit but also magnify systemic inequities.

It is more difficult to draw concrete conclusions from the predictive portion of this project to answer the third research question, given that these recreations were trained on a more limited dataset than the ones used by financial institutions. However, our work shows a strong chance that any future model trained on the New York state outcomes from 2023 may inadvertently reinforce the same racial biases. The financial variables employed in the models seem objective on the surface, but we demonstrated that it is more difficult for a model to accurately predict the outcomes of applications from Black individuals, likely because the real–world decisions don't always make objective choices.

Taken together, our findings illuminate precisely where algorithmic bias enters the mortgage processes—loan sizing, rate-setting, and approval decisions—and quantify its magnitude. They also suggest practical pathways forward. Drawing on Lee et al.'s reconceptualization of fairness as a trade-off among objectives [3], lenders can employ fairness-aware training methods—such as demographic parity or equalized odds constraints—and data-sampling strategies to mitigate disparate impacts [5]. Routine subgroup audits and human-in-the-loop reviews for borderline cases can further ensure that automated decisions do not become unchallengeable "black-box" edicts.

As policymakers consider algorithmic accountability frameworks, our work underscores the necessity of embedding fairness checks into every stage of mortgage underwriting. Only by proactively auditing and constraining these systems can we harness the efficiencies of automation without perpetuating the very disparities that homeownership was meant to overcome.

6 LIMITATIONS

While our models control for many observable factors, several important limitations remain. First, the models assume linear relationships and independence among predictors, yet interactions (e.g., between income and race) may be complex and nonlinear. Second, multicollinearity may distort coefficient interpretations, especially with high-cardinality categorical variables. The large condition numbers in our model outputs suggest this could be a concern. Additionally, all models rely on OLS, which assumes homoscedasticity and normality of residuals—assumptions we did not explicitly test. Using more flexible models (e.g., random forests or gradient boosting) could improve predictive performance, though at the cost of interpretability.

Another key limitation lies in the imbalanced distribution of outcomes and applicant demographics within the dataset. The vast majority of applications were approved, leaving a relatively small sample of rejected cases, which limits our ability to robustly analyze denial patterns. Furthermore, White applicants comprised a disproportionate majority, which may reduce the statistical power to detect disparities affecting underrepresented racial groups and could bias model training toward majority group patterns.

Moreover, we lack visibility into the internal decision-making models or criteria used by lenders—such as proprietary risk scoring algorithms, underwriting thresholds, or qualitative assessments. Because we do not know the exact structure or logic of the models lenders use to assess loan applications, it's difficult to determine whether observed disparities are the result of bias, justified financial risk assessments, or omitted variables. This absence of ground truth limits our ability to draw definitive conclusions about the fairness or discriminatory nature of the lending process.

7 ETHICAL CONSIDERATIONS

This project raises several ethical questions. By analyzing race and gender in lending decisions, we are directly engaging with issues of fairness, systemic bias, and discrimination. Our findings support the concern that, even when controlling for financial indicators, minority and female applicants may receive less favorable loan terms. This reinforces historical patterns of redlining and credit exclusion. However, ethical care must also be taken in interpreting results: labeling disparities as discrimination without understanding all underlying variables risks misattribution. Still, transparency in these processes is crucial, and projects like this can pressure institutions toward more equitable practices.

8 IMPROVEMENTS

To improve our study on algorithmic bias in mortgage lending, future research could add several improvements. First, we would prioritize analyzing geographical trends, such as redlining and zoning patterns, to assess their influence on loan approvals and disparities, particularly in New York's diverse neighborhoods. Second, selecting a state with higher denial rates and greater demographic diversity would address the limitations of our current dataset, which was skewed toward approvals, complicating the modeling of rejection patterns. Third, incorporating additional predictors or refining existing ones to help refine our models or uncover hidden biases. Finally, a deeper investigation into specific algorithmic underwriting systems, comparing their outcomes to identify which algorithms exhibit greater bias, would provide good insights for developing fairer lending practices.

REFERENCES

[1] Mortgage Bankers Association. 2024. *Algorithmic Property Data Collection.* Technical Report. Mortgage Bankers Association. https://www.mba.org/docs/default-source/policy/mba_worksession_propertydatacollection_2.27.24.pdf?sfvrsn=f2caae76_1 Accessed: 2025-02-25.

[2] Investopedia. 2023. Mortgage Process Explained. *Investopedia* (2023). https://www.investopedia.com/mortgage-process-explained-5213694 Accessed: 2025-02-25.

[3] M. S. A. Lee and L. Floridi. 2021. Algorithmic Fairness in Mortgage Lending: from Absolute Conditions to Relational Trade-offs. *Minds & Machines* 31 (2021), 165–191. https://doi.org/10.1007/s11023-020-09529-4 Accessed: 2025-05-13.

[4] Emmanuel Martinez and Lauren Kirchner. 2021. The Secret Bias Hidden in Mortgage-Approval Algorithms. *The Markup* (2021). https://themarkup.org Accessed: 2025-02-25.

[5] MIT CSAIL. 2020. Fairness Metrics for Machine Learning: a Retrospective. https://cap.csail.mit.edu/sites/default/files/research-pdfs/make-04-00011-v2.pdf. Accessed: 2025-05-13.

[6] Andre M. Perry, Jonathan Rothwell, and David Harshbarger. 2023. Homeownership, racial segregation, and policies for racial wealth equity. *Brookings Institution* (2023). https://www.brookings.edu/articles/homeownership-racial-segregation-and-policies-for-racial-wealth-equity/ Accessed: 2025-02-25.

[7] Local Housing Solutions. 2023. Why Housing Matters. *Local Housing Solutions* (2023). https://localhousingsolutions.org/bridge/why-housing-matters/ Accessed: 2025-02-25.

[8] U.S. Census Bureau. 2023. Homeownership Rates by State. https://www.census.gov/data/tables/time-series/demo/homeownership/rate-children.html Accessed: 2025-05-13.

[9] U.S. Department of Housing and Urban Development. 2021. *U.S. Housing Market Conditions: National Data Summary.* Technical Report. Accessed: 2025-05-13.

Table 1. Codebook for Loan Application Data

Field Name	**Description**	**Data Type / Values**
action_taken	Final decision made on the loan application.	`int`: 1 (Loan originated), 2 (Application approved but not accepted), 3 (Application denied), 4 (Application withdrawn by applicant), 5 (File closed for incompleteness), 6 (Purchased loan).
derived_race	Race of the applicant as determined by the data system.	`object`
derived_ethnicity	Ethnicity of the applicant as derived from application data.	`object`
applicant_sex	Sex of the primary loan applicant.	`int`
applicant_age	Age of the primary loan applicant.	`object`: < 25, 25–34, 35–44, 45–54, 55–64, > 65.
income	Applicant's annual income (in thousands of dollars).	`float`
debt_to_income_ratio	Ratio of applicant's monthly debt payments to income.	`float`: < 20%, 20%– < 30%, 30%– < 36%, 36%– < 40%, 40%– < 45%, 45%– < 50%, 50%–60%, > 60%.
applicant_credit_score_type	Type of credit score used for the applicant.	`int`
loan_amount	Total amount of the loan applied for.	`float`
loan_to_value_ratio	Ratio of the loan amount to the appraised value of the property.	`float`
interest_rate	Interest rate charged on the loan.	`float`
rate_spread	Difference between the loan's interest rate and the average prime offer rate.	`object`
loan_type	Type of loan.	`int`: 1 (Conventional), 2 (FHA), 3 (VA), 4 (RHS or FSA).
loan_purpose	Purpose of the loan.	`int`: 1 (Home purchase), 2 (Home improvement), 31 (Refinancing), 32 (Cash-out refinancing), 4 (Other), 5 (Not applicable).
lien_status	Indicates whether the loan is a first or subordinate lien.	`int`: 1 (First lien), 2 (Subordinate lien).
property_value	Appraised value of the property backing the loan.	`float`
occupancy_type	Indicates whether the property is owner-occupied, rental, etc.	`int`: 1 (Principal residence), 2 (Second residence), 3 (Investment property).
tract_minority_population_percent	Percentage of minority population in the census tract.	`float`
aus-1	Automated Underwriting System used for the loan decision.	`int`: 1 (Desktop Underwriter), 2 (Loan Prospector), 3 (Technology Open to Approved Lenders), 4 (Guaranteed Underwriting System), 5 (Other), 6 (Not applicable).
denial_reason-1	Primary reason for denial if the loan was denied.	`int`: 1 (Debt-to-income ratio), 2 (Employment history), 3 (Credit history), 4 (Collateral), 5 (Insufficient cash), 6 (Unverifiable information), 7 (Credit application incomplete), 8 (Mortgage insurance denied), 9 (Other), 10 (Not applicable).