# 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, affordable housing has decreased as rent has steadily increased while incomes have stagnated [3]. 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. Additionally, Black Americans face discriminatory lending practices in the mortgage application process [5].

Traditionally, decisions about mortgage loan approval were made by human underwriters who reviewed a combination of financial and personal information provided by applicants. When purchasing a house with a mortgage, one must first be pre-approved for a loan. 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 choosing a property, applicants must submit more comprehensive information on employment, income, assets, debt, property, and credit history [1]. Using this data, lenders generate a loan estimate and decide whether to approve or deny the application.

However, lenders today increasingly rely on algorithms with the aim of being more fair, accurate, and efficient than human underwriters. Mortgage lenders hope algorithms will streamline the decision-making process and eliminate human bias and inconsistency. Now, hidden algorithms are used throughout the process to evaluate home prices and mortgage applications and to determine final interest rates [4]. During the application process, these algorithms assess risk and predict loan repayment. Applicants who are misclassified in terms of risk or repayment capability may receive higher interest rates, smaller loans, or even outright rejection. This increases financial burden and restricts financial mobility. Although these algorithms promise greater efficiency and objectivity, growing evidence suggests that they may perpetuate the very biases they were designed 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. Their findings revealed that even after accounting for factors such as debt-to-income ratio, loan-to-value ratio, and credit score—elements lenders often cite 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 their white counterparts with similar financial profiles. The investigation highlighted that high-income Black applicants with less debt were denied more often than high-income white applicants with more debt, suggesting that algorithmic underwriting systems may perpetuate existing biases rather than eliminate them. The process of getting approved for a loan can be found in Appendix, Figure 2.

## RESEARCH FOCUS

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

We address four central questions:

(1) What are the potential impacts of the loan-approval system algorithms on the approval rates of applicants of diverse identities and geographical backgrounds?
(2) Do applicants of different race and gender backgrounds experience significantly different loan approval rates?
(3) What disparities, if any, exist in the loan interest rates awarded to applicants of different race and gender backgrounds?
(4) How can we create and assess a more fair credit-scoring model while still prioritizing accuracy?

The demographic distribution in the New York State HMDA dataset is as follows:

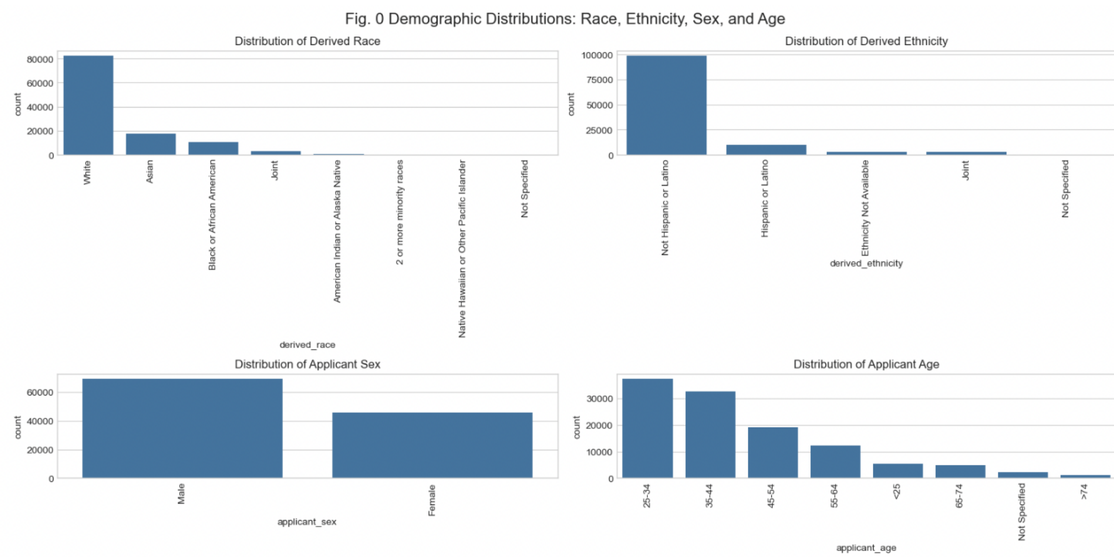


Fig. 1. Demographic Distribution Based on Four Variables.

## 2 DATA

The dataset 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—such as `debt_to_income_ratio`, `property_value`, and `interest_rate`—by coercing them to numeric types and dropping null values where appropriate.

To clean the data, we converted all object-type variables into either categorical, binary, or numerical values to facilitate model training and analysis. For predictors that used numeric codes to represent categories, we filtered to include only relevant codes. For example, for the variable `action_taken`, we retained only values 1, 2, and 7, ignoring values 3 through 6. Additionally, we removed all rows with missing values or unusable entries such as `NaN`.

## 3 METHODS

To explore the feasibility of building a fairer loan-approval model without sacrificing accuracy, we trained three new models using the 2023 HMDA dataset. Our aim was to evaluate whether the original HMDA model may have generated biased outputs that could pose ethical and practical concerns if used in future credit evaluations.

We began by refining the dataset for our analysis. Applicant race was recoded as a binary variable to focus specifically on disparities between white and Black applicants. Loan outcomes were reclassified into binary categories: *"Loan Originated"* and *"Application Approved but Not Accepted"* were grouped as successful applications, while *"Loan Denied"* and *"Pre-Approval Request Denied"* were categorized as rejections.

Our primary research goal was to determine whether race and gender have a statistically significant impact on loan-related outcomes. Figure 3 in Appendix shows the spread of different loan outcomes in the dataset. To test this, we constructed three Ordinary Least Squares (OLS) regression models using the `statsmodels` library. All three models incorporated a consistent set of independent variables covering demographic characteristics (e.g., race, sex, age), financial indicators (e.g., income, debt-to-income ratio, credit score), and loan features (e.g., loan type, loan purpose, property value).

- **Model 1** predicted the approved loan amount to investigate whether borrowers of different demographic profiles receive unequal loan values.
- **Model 2** estimated interest rates to examine potential disparities in the cost of borrowing.
- **Model 3** assessed the rate spread, which measures how much more a borrower pays relative to the most favorable market rate.

Categorical variables were one-hot encoded using `pandas.get_dummies()`, and all predictors were standardized for consistency. Each model included a constant term and was evaluated using R-squared values and p-values to assess model fit and the statistical significance of key variables.

In parallel, we implemented three classification algorithms—K-Nearest Neighbors, Decision Tree Classifier, and Logistic Regression—chosen for their compatibility with binary outcome prediction. These models were trained on a subset of financial variables, such as applicant income and debt-to-income ratio, to predict loan approval or denial. As shown in Figure 3, we first examined the success rates by race and then tested fairness more formally. We retrained the models with race included as an input and analyzed fairness metrics such as selection rate and false positive rate across demographic groups.

To evaluate counterfactual fairness, we introduced synthetic applicant profiles where a white and a Black applicant shared identical financial characteristics and loan requests. If the predicted outcomes differed solely due to race, the model exhibited explicit bias beyond reliance on indirect proxies like income or location. These insights offer critical guidance for stakeholders considering the HMDA dataset for future credit modeling applications.

## 4 RESULTS

Our first objective was to train models on various loan terms—such as loan amount, interest rate, and interest spread—to gain a better understanding of how applicant demographics and financial status influence the terms accepted applicants ultimately receive.

### Model 1: Predicting Loan Amount (Multiple Linear Regression)

We employed a multiple linear regression model using selected variables to predict the loan amount approved for applicants. The model demonstrated a strong fit, with an $R^2$ value of 0.790, indicating that it explains approximately 79% of the variance in the data.

Age exhibited a predictable pattern: as applicants aged, the loan amount they received tended to increase. All age groups except for the 64+ demographic showed statistical significance ($p < 0.001$). Notably, applicants aged 35–44 were predicted to receive the highest loan amounts, with a coefficient of $22,100.

Gender and race also displayed significant effects. When controlling for all other variables, female applicants were predicted to receive $15,570 less than male applicants ($p < 0.001$), while White applicants were predicted to receive $16,760 more than their counterparts ($p < 0.001$).

Among the financial variables, income and property value were the most significant predictors of loan amount. Income had a positive coefficient of 46.446 ($p < 0.001$), debt-to-income ratio had a positive coefficient of 360.525 ($p = 0.04$), and property value had a positive coefficient of 0.4472 ($p < 0.001$).

### Model 2: Predicting Interest Rate (Multiple Linear Regression)

Our second model predicted the interest rate offered to applicants. This model had a weaker fit, with an $R^2$ of 0.171. While gender was not statistically significant, White applicants were predicted to receive slightly higher rates than Black applicants, with a coefficient of 0.0542 ($p = 0.007$).

Financial variables again played a central role. Income was negatively associated with interest rates (coefficient: $-4.96 \times 10^{-5}$, $p = 0.001$), and debt-to-income ratio had a positive coefficient of 0.0190 ($p < 0.001$), indicating that higher income reduces interest rates while higher debt ratios increase them.

### Model 3: Predicting Interest Spread (Multiple Linear Regression)

The third model estimated the interest spread, which standardizes interest rates across different economic periods. Neither gender nor race were significant predictors in this model. However, financial variables were once again strongly correlated.

Income had a negative coefficient of $-5.897 \times 10^{-5}$ ($p < 0.001$), and debt-to-income ratio had a positive coefficient of 0.0121 ($p < 0.001$), suggesting that higher income and lower debt ratios are associated with more favorable (i.e., lower) interest spreads.

## LOAN APPROVAL MODELS AND FAIRNESS EVALUATION

Our second objective was to analyze the fairness of loan approval decisions. We trained three models—Logistic Regression, Decision Tree, and K-Nearest Neighbors (KNN)—to predict application outcomes (accepted vs. denied).

- **Logistic Regression:** Accuracy = 86.49%
- **Decision Tree:** Accuracy = 78.51%
- **K-Nearest Neighbors:** Accuracy = 83.59%

To assess potential racial bias, we compared accuracy across White and Black applicants. Each model showed a discrepancy of over 5 percentage points:

- **Logistic Regression:** 9.13% difference
- **KNN:** 7.95% difference
- **Decision Tree:** 5.02% difference

We then retrained each model to explicitly include race as an input feature. The performance metrics are summarized below:

- **Logistic Regression:**
  - Accuracy: 86.9% (White), 77.6% (Black)
  - False Positive Rate (FPR): 99% (both groups)
  - False Negative Rate (FNR): <1% (both groups)
- **K-Nearest Neighbors:**
  - Accuracy: 83.78% (White), 75.89% (Black)
  - Selection Rate: 94.04% (White), 92.86% (Black)
  - FPR: 89.16% (White), 88.13% (Black)
  - FNR: ~5% (both groups)
- **Decision Tree:**
  - Accuracy: 79.65% (White), 67.35% (Black)
  - Selection Rate: 74.39% (White), 86.2% (Black)
  - FPR: 74.98% (White), 65.81% (Black)
  - FNR: 12.11% (White), 23.15% (Black)

These disparities suggest that while all three models perform reasonably well in aggregate, racial disparities in predictive accuracy and error rates persist—raising potential fairness concerns that warrant further investigation.

## 5 DISCUSSION

In their 2020 study, Lee et al. examined mortgage lending as a case study to reconceptualize fairness not as a rigid mathematical constraint, but as a negotiation between competing objectives [2]. This framework may be valuable for synthesizing the multiple fairness metrics we used to assess racial disparities in our models. Rather than attempting to optimize for a single definition of fairness, this approach allows for the identification of the most harmful inequities and the articulation of trade-offs—such as potential reductions in financial returns—in service of achieving more equitable outcomes. Prior research has also explored techniques like data sampling to balance fairness and performance in machine learning classifiers, as well as fairness metrics designed for datasets with multiple sensitive features [6]. Although we did not incorporate these specific methods into our models, they represent promising avenues for future work and could enhance the robustness and equity of mortgage approval algorithms.

The findings from our three regression models provide evidence of systemic disparities in mortgage lending outcomes based on race and gender, even after accounting for relevant financial and loan-specific variables.

The first model revealed that differences in approved loan amounts could not be fully explained by financial characteristics or property values. These persistent gaps suggest the presence of racial and gender-based bias in lending decisions. Our results align with the hypothesis that Black female applicants, in particular, may receive less favorable loan terms than both Black male and White female applicants. Research from the Brookings Institution supports this finding, showing that homes in majority-Black neighborhoods are undervalued by an average of 48,000 dollars compared to those in white neighborhoods, even when controlling for key features [5]. This structural devaluation results in significantly lower borrowing capacity and contributes to 156 billion dollars in lost equity for Black homeowners. Such patterns reinforce the idea that systemic bias, rather than individual applicant qualifications alone, continues to affect loan outcomes.

Interestingly, in the second model, White applicants were found to be charged interest rates that were on average 0.054 percentage points higher than those charged to Black applicants, holding all other financial indicators constant. Though numerically small, this difference becomes meaningful over the long term of a mortgage. This outcome contrasts with our initial hypothesis that Black applicants would face higher interest rates, raising questions about the role of algorithmic decision-making in risk-based pricing and the possibility of lender overcorrection or other hidden biases.

In the third model, race did not emerge as a statistically significant predictor of rate spread. Instead, financial variables such as the debt-to-income ratio, loan-to-value ratio, and automated underwriting system (AUS) decision type were more influential. This may indicate that while demographic variables affect base interest rates, adjustments made relative to prevailing market rates are more standardized and less susceptible to individual-level bias.

Overall, our results suggest that racial and gender disparities continue to shape mortgage lending outcomes, with the clearest evidence found in the loan amount model. While disparities in interest rates and rate spreads are less pronounced, they may still reflect subtle or indirect forms of bias. The strongest disparities appear at the pre-approval stage—before rates are finalized—suggesting that the initial decision to approve and the loan terms offered may be more prone to demographic bias than the later stages of the loan process.

## 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.

## 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.

## REFERENCES

[1] Investopedia Staff. 2023. The Mortgage Process, Explained. https://www.investopedia.com/mortgage-process-explained-5213694 Accessed: April 11, 2025.

[2] 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

[3] Local Housing Solutions. 2024. Why Housing Matters. https://localhousingsolutions.org/bridge/why-housing-matters/ Accessed: April 11, 2025.

[4] Mortgage Bankers Association. 2024. *Property Data Collection and Algorithmic Technology in Appraisals*. Technical Report. Mortgage Bankers Association. https://www.mba.org/docs/default-source/policy/mba_worksession_propertydatacollection_2.27.24.pdf Accessed: April 11, 2025.

[5] Andre M. Perry, Jonathan Rothwell, and David Harshbarger. 2020. Homeownership, Racial Segregation, and Policy Solutions to Racial Wealth Equity. https://www.brookings.edu/articles/homeownership-racial-segregation-and-policies-for-racial-wealth-equity/ Accessed: April 11, 2025.

[6] Pedro Saleiro, Ben Kuester, Ben Heller, Jesse London, and Rayid Ghani. 2020. *Aequitas: A Bias and Fairness Audit Toolkit*. Technical Report. MIT CSAIL. https://cap.csail.mit.edu/sites/default/files/research-pdfs/make-04-00011-v2.pdf Accessed: April 11, 2025.

## 6 APPENDIX

The process of getting approved for a loan is as follows:

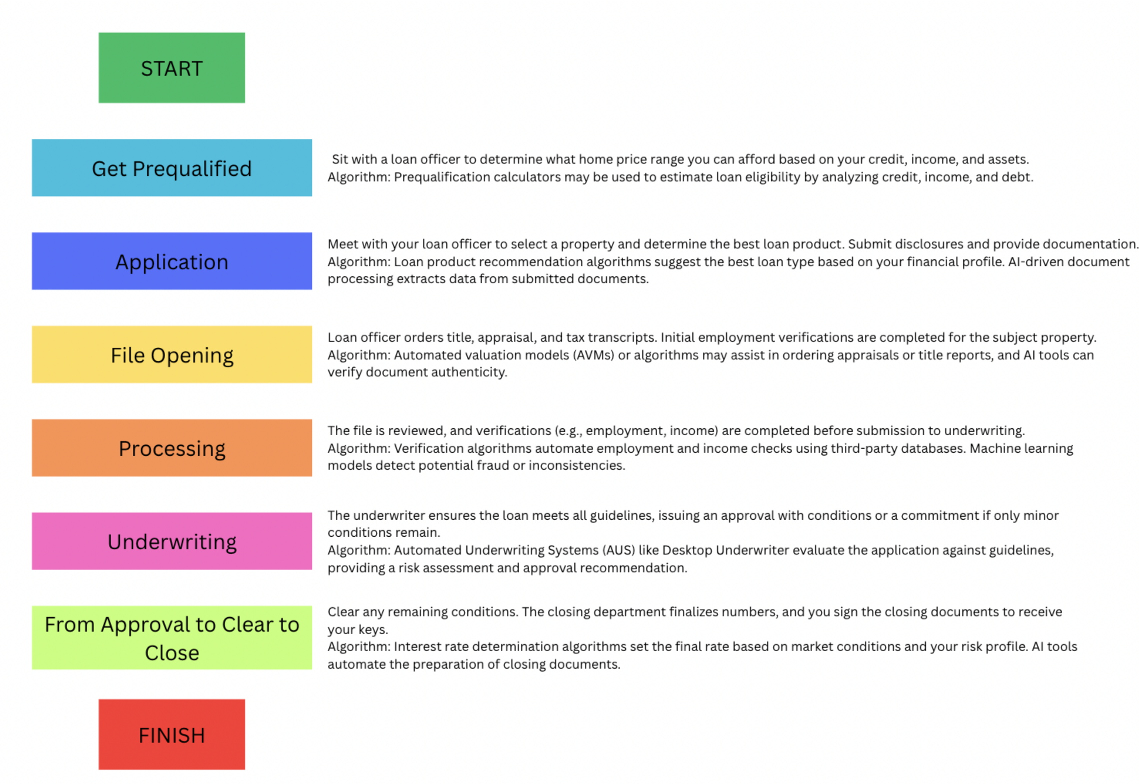


Fig. 2. Diagram of Loan Application and Approval Process.

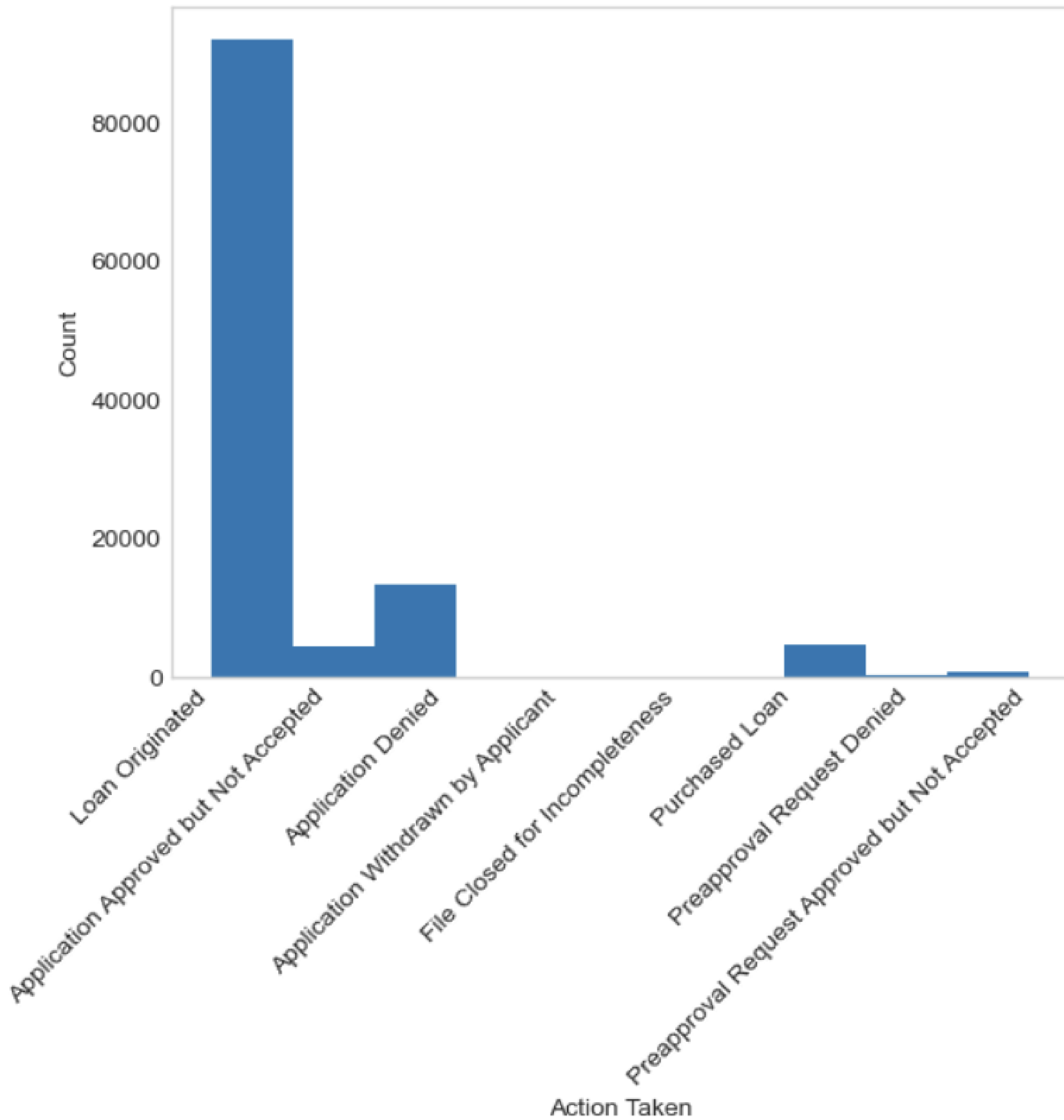


Fig. 3. Distribution of Loan Application Outcomes.