

# Housing Discrimination and Algorithmic Bias

*Group 1: William Holt, Vaishnavi Paniki, Sebastian Segura*

## Introduction and Historical Context

Housing discrimination in the US has and still has a major impact on urban development and property valuation. Racial covenants, legal clauses in property deeds that prohibited sale or occupancy of homes by non-white peoples, were implemented between 1910 and 1960. These covenants reinforced racial segregation and limited opportunities for countless minority communities. In 1968, the Fair Housing Act outlawed these discriminatory practices, yet their effects remain apparent in modern housing patterns, property values, and wealth in areas.

The University of Minnesota published the Mapping Prejudice Project by documenting 30,000 racial covenant cases in Hennepin County, showing that the historical practices still shape modern residential patterns. Areas that historically partook in the practice of such racial covenants conveniently find their homes with higher average property values than other areas, while those that had no such covenants find themselves more despondent and contribute to driving deeper the wedge between minorities and generational wealth and opportunities today.

Modern mortgage lenders rely heavily on algorithmic decision-making tools. These tools often simply spit out a yes or no answer without the user understanding the reasoning behind the decisions. These systems do not explicitly consider race but instead may use proxy variables such as property values, income levels, and loan size to determine the race of the individual based on past discrimination. Our project examines the correlation between historical discrimination and modern lending patterns through the investigation of both data bias and algorithmic biases.

## Methodology

To conduct our analysis, we employed a mixed-method approach. The analysis was conducted using a preprocessed dataset that combined historical covenant geographical data with modern mortgage lending data and census demographics information in Hennepin County. This allowed us to assess both historical data bias and contemporary algorithmic bias in lending decisions. The following are the steps we utilized for this analysis.

### *Data Cleaning and Preparation*

We removed rows with missing values in key columns such as 'loan\_amount' or 'income' and created a binary approved column to classify loan outcomes as 'yes' or 'no' data. Categorical variables such as race and loan purpose were encoded into. Numerical representation for this analysis.

### *Exploratory Analysis*

We analyzed loan approval rates by race and minority population categories to find correlations between a myriad of factors and loan approval rates. Then, we visualized trends with different visuals. Looking at the differences between loan amounts and interest rates across racial and socioeconomic groups. We computed correlation coefficients to detect patterns that might indicate bias when it comes to decision-making.

### *Bias Investigation*

We identified proxy variables for historical discrimination, such as neighborhood population demographics and income to compare lending patterns across different communities.

## Algorithmic Bias

Finally, we investigated how modern lending systems might perpetuate poor data through biased approval rates and risk assessments. Assessing whether underrepresented applicants were assigned higher interest rates.

## Findings

### *Data Bias Analysis*

#### *Historical Discrimination in Today's Data*

The dataset showed major discrepancies in property value and lending outcomes between covenant and non-covenant areas. Areas containing higher covenant density also showed higher property values and much more favorable lending terms, and those areas in which covenants were not used showed much lower average property values and way higher loan denial rates.

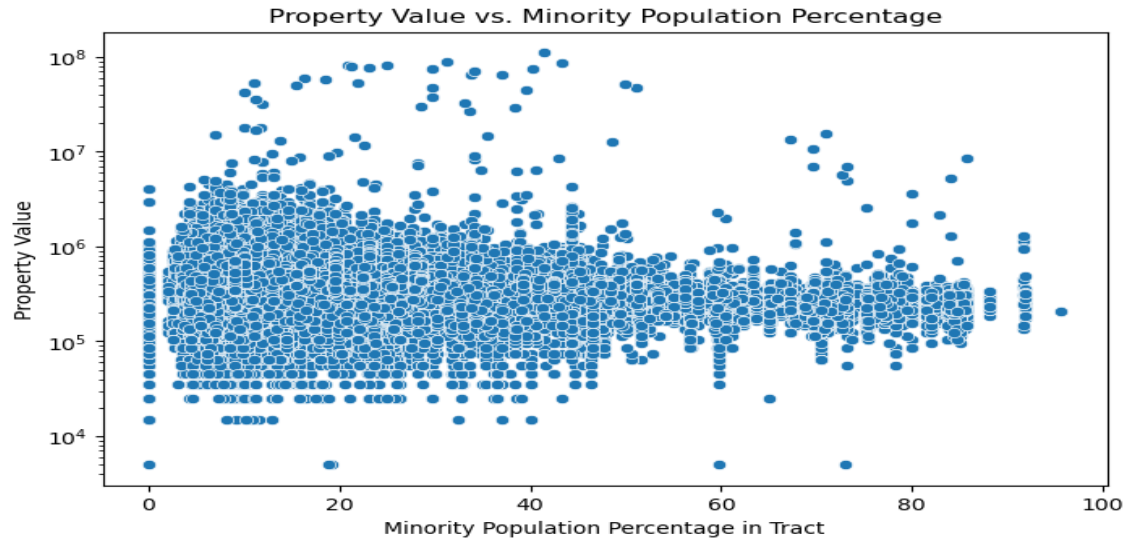


Figure 1. Scatterplot comparing property values and tract minority population percentage.

#### Property Values and Neighborhood Characteristics

Visuals such as scatter plots and box plots highlight the lasting impact of historical discrimination. For example, areas in which more minorities lived than average had lower property values and higher loan denial rates, indicating that historical biases are ever presently embedded in data we use today.

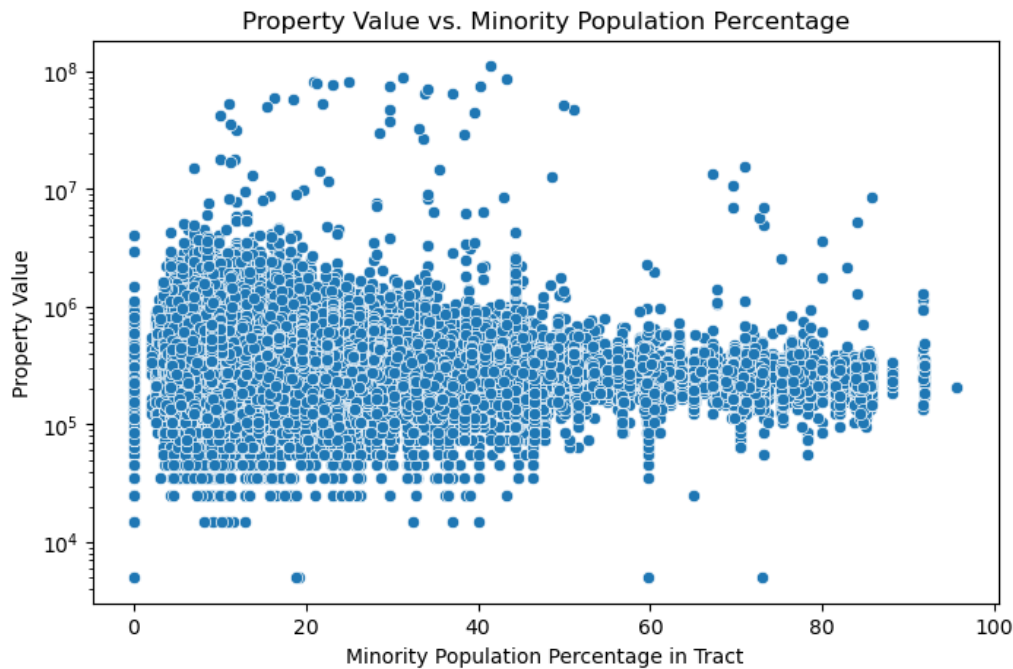


Figure 2. The scatter plot shows the changes in property value as the percentage of minority in the area changes.

## Generational Wealth and Investment

This analysis explained how generational wealth accumulation is highly affected by historically discriminatory policies. The areas which benefited from the past's discriminatory policies continue to reap the benefits of wealth accumulation, while historically excluded areas still lack those same advantages.

## Algorithmic Bias Analysis

### Loan Approval Rates by Race

This section's analysis focuses on loan approval rates by racial groups. Notable conclusions include those white applicants had the highest approval rates (74%) followed by Asian individuals (69%), while black, native Hawaiian, and Pacific Islander individuals all had approval rates around 55%. These findings show that racial disparities persist in lending outcomes, despite excluding race from the algorithmic analysis altogether.

derived_race	
2 or more minority races	0.622222
American Indian or Alaska Native	0.597163
Asian	0.687393
Black or African American	0.585913
Free Form Text Only	0.800000
Joint	0.729461
Native Hawaiian or Other Pacific Islander	0.553571
Race Not Available	0.611235
White	0.739222

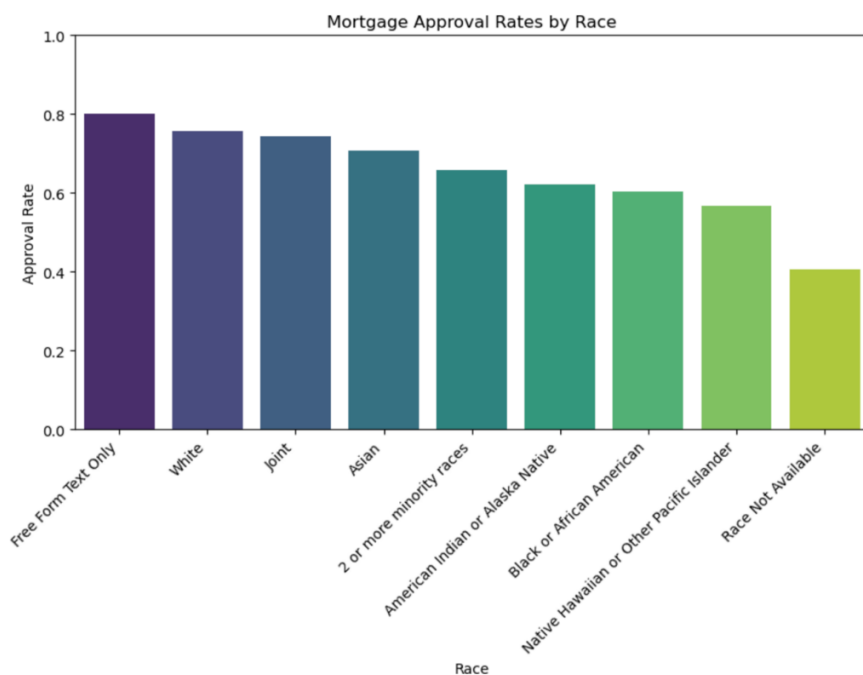
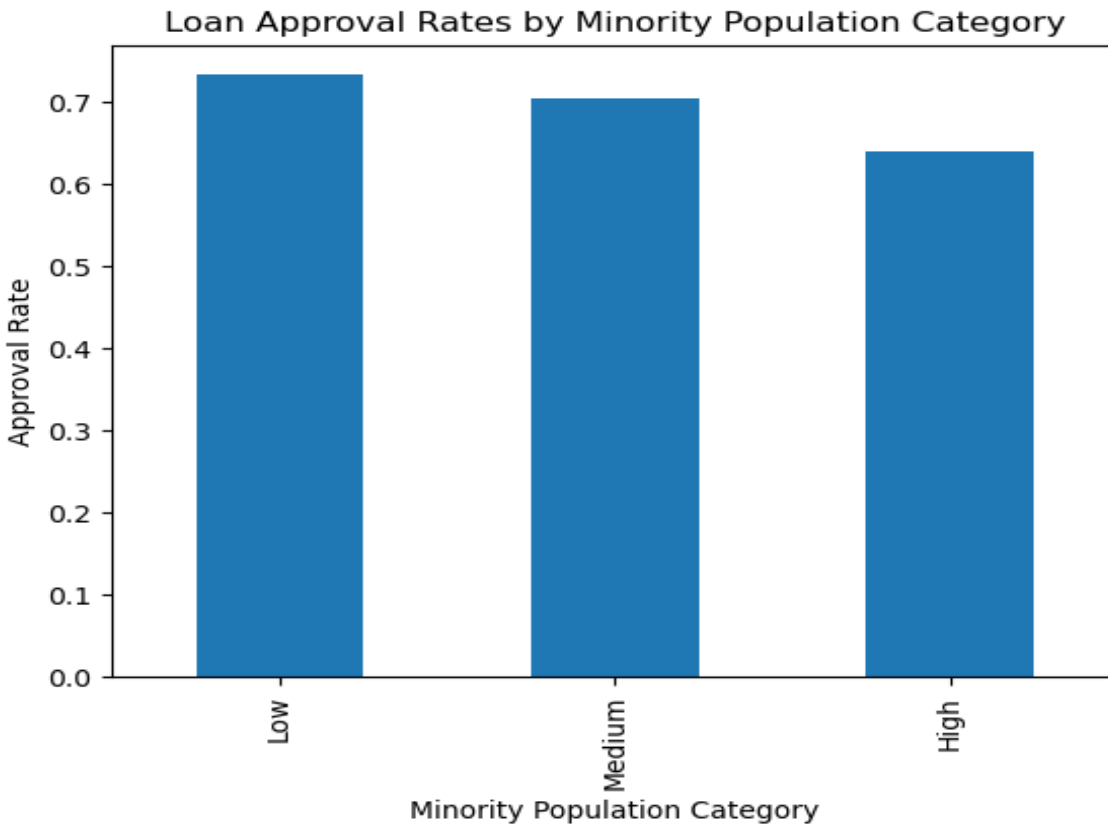


Figure 4. The chart supports the assertion that loan approval rates are still affected to this day based on historically racist and discriminatory practices.

### *Loan Approval Rates by Minority Population*

The loan approval rates were also analyzed in relation to predominantly minority neighborhoods. The low minority population tracts had 73% approval rates, and the high minority population tracts showed significantly lower approval rates, 64%. The logical conclusion here is that neighborhoods with high percentages of minorities face discriminatory and biased lending decisions, probably due to the use of proxy variables that will reflect historical discrimination.



*Figure 5. Bar chart explaining loan approval rates by minority population category (low, med, high).*

### *Proxy Variables*

Weak correlations were found between income and loan approval (0.0038) and between loan amount and loan approval (0.0488). This suggests that both loan amount and income are poor predictors of approval, and some other factors may be responsible for being racial proxies.

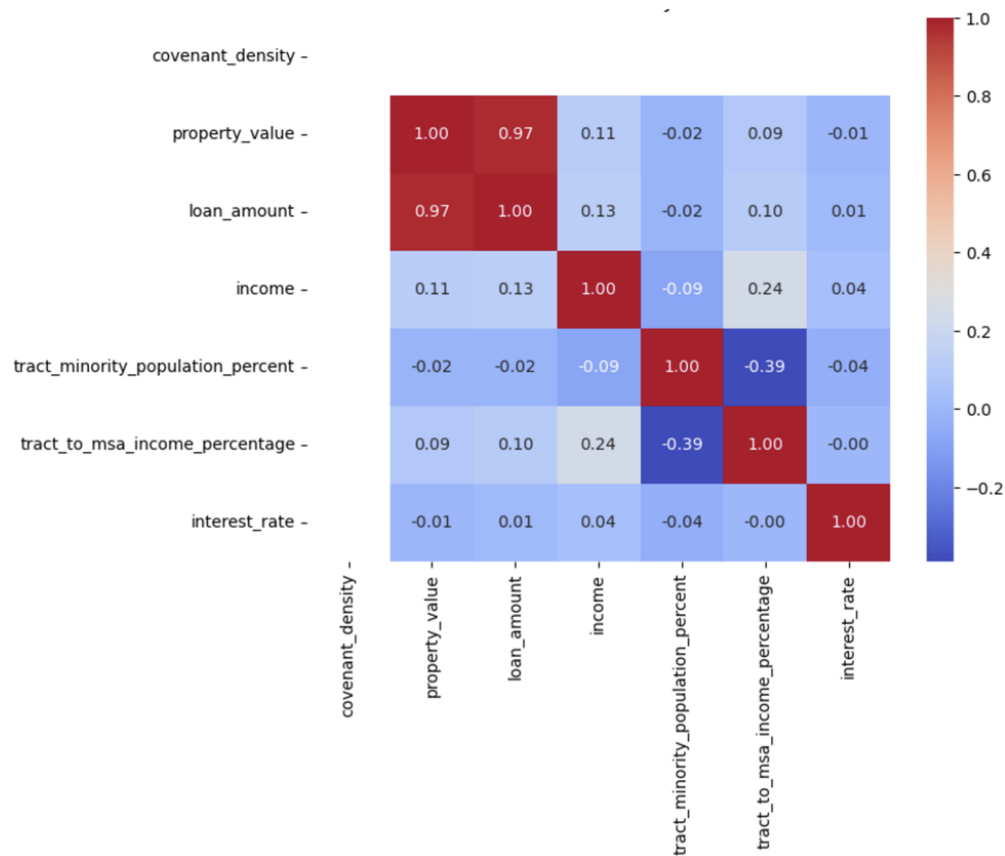


Figure 6. Plot variables and loan approval. This figure explains the weak correlation between the two factors, but stronger correlations for others.

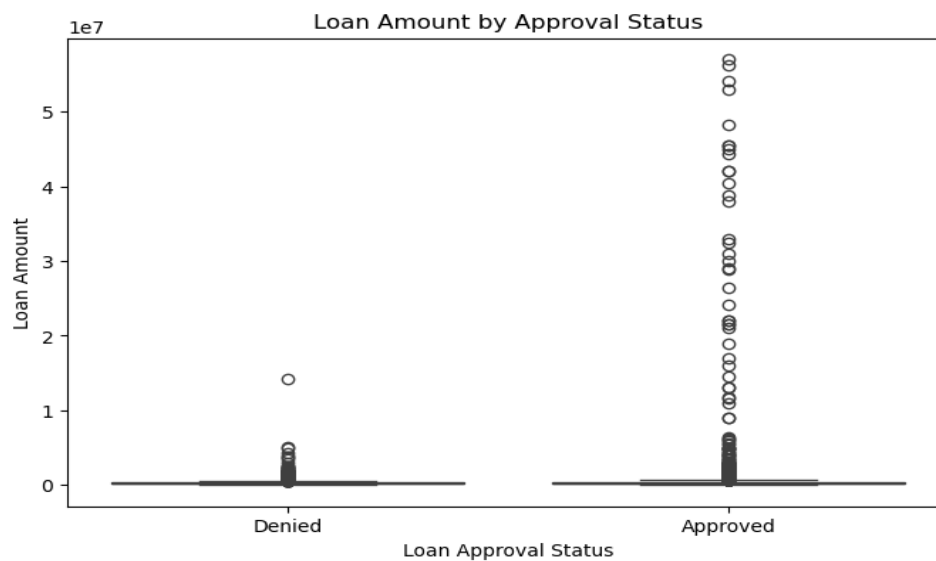


Figure 7. Plot of the loan amount and loan approval, showing that higher loans are more often approved.

## Discussion

### *Implications for Modern Lending Practices*

Our findings display the potential for modern lending systems to extend the life of historical discrimination. Although the systems are designed with race-neutrality in mind, they seldom are completely free of bias and often unknowingly rely on proxy variables (such as neighborhood characteristics and property values) which reflect historic biases and ultimately lead to unequal access to lines of credit and reinforce long-standing wealth disparities.

Moreover, the increasing use of machine learning algorithms in mortgage approvals and risk assessments could potentially have some risks. Without oversight, these models could make the disparities bigger by prioritizing “safe” lending decisions that disproportionately favor privileged groups.

### *Potential Mitigation Strategies*

To address these deeply embedded issues, the following strategies are recommended:

1. **Integration of Historical Context:** Lending models should explicitly account for historical discrimination by incorporating variables which would adjust for past injustice.
2. **Use of Alternative Data:** By including alternative data sources, such as rental history, utility payments, etc. may provide a more comprehensive understanding of each applicant as an individual with an ability to repay a loan, as opposed to simply a minority from a poor area.
3. **Improving Data Collection:** Collaborating between institutions and government agencies will help see that data being used is more representative and inclusive towards excluded communities who would otherwise be limited in their loan approval luck.

## Limitations and Future Work

### Limitations

Multiple potential issues arise when discussing the future of historically racist loan data. First, the availability of quality data is limited, and this specific dataset lacked comprehensive information related to historical covenant density, slowing the ability to fully analyze its potential impact as it relates to modern lending algorithms in depth. Next, the reliance on proxy variables may introduce measurement bias and further obscure the true effects of historical discrimination unknowingly. Finally, these findings in these two

counties may not generalize to other regions, so any conclusions or workarounds uncovered may be needlessly costly and waste time in communities where they simply are not needed. Moreover, policy changes and economic shifts could alter the landscape of mortgage approval in ways that were accounted for in this study.

## Future Work

Moving forward, expanding the scope of our analysis to a nationwide scale would provide a much broader understanding of racial lending disparities through a historical and regional lens. Advanced techniques would further help undo some of the effects of historical discrimination from other urban factors. Algorithmic fairness may only be achieved when every individual working with the data or its models is aware of their conscious or unconscious biases, and all future research should be conducted with fairness at the forefront of every analyst's or scientist's mind. Future Research should place fairness, transparency, and accountability at the forefront by continuously refining our methodologies and expanding the scope of our analysis.

## Conclusion

This report demonstrated the lasting impacts of historical housing discrimination on modern lending practices and explains the need for many more proactive measures to address algorithmic biases. By integrating historical context and improving data representation to include everyone, we will be able to move forward and provide equitable housing markets across the country despite the past. Lastly, addressing these biases will require Government intervention and community-driven initiatives to help promote housing equity. We can move towards a more fair and equitable future in housing through continuous monitoring and policy adaptation.