Racial and Ethnic Disparities in Home Lending Approval in Mecklenburg County, NC

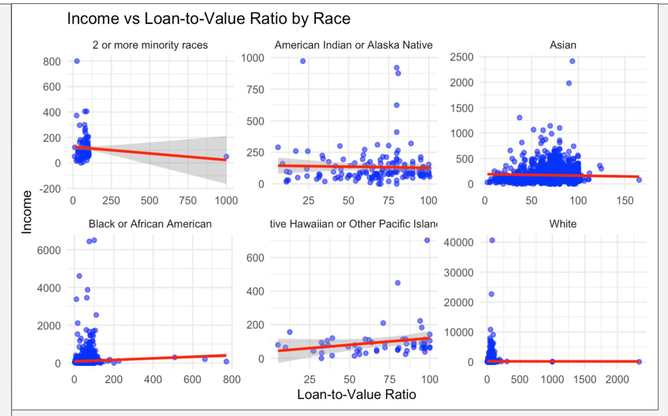
AUTHORS

Alessandra Bielli
Aidan Hennessy

01. Introduction

By analyzing the impact of factors such as income, race, and credit score on loan-to-value ratios and loan approval rates, we aim to provide actionable insights into these disparities and contribute to a deeper understanding of equity in housing finance.

Data Source: HMDA Data Browser
Stakeholders: Mecklenburg County Commissioners



02. Research Questions

- How do an applicant's income, race, and credit score (the race of the primary applicant) predict their loan-to-value (LTV) ratio?
- Can we predict whether a loan application will be approved or denied (action_taken) based on factors like income, race, credit score, and loan-to-value ratio?

Anova Classification Test

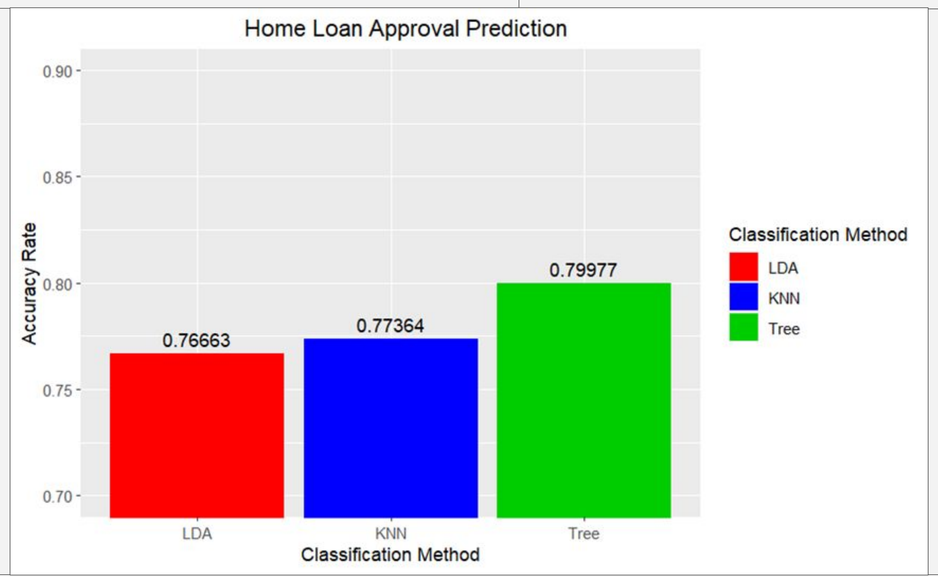
Model 1 includes race and ethnicity as predictors for classification outcome
Model 2 excludes race and ethnicity as predictors for classification outcome

Model	Residual DF	Residual Deviance (Dev)	DF	Deviance	P-Value
1	26219	22710			
2	26210	22570	9	140.46	< 0.001 *

*Not including race and ethnicity in predicting the classification outcome makes it harder to predict the classification outcome.

03. Methodology

- Linear Regression
- Ridge Regression
- PLS Regression
- KNN Classification
- LDA Classification
- Tree Classification



04. Results/Findings

- Regression Testing
  - **Income** showed a small positive impact on LTV (coefficient of 0.02156)
  - **White** applicants showed a slight positive association with LTV (β=0.0031)
  - **Asian** applicants demonstrated a stronger positive relationship (β=1.99)
  - **Black or African American** (β=−0.63) and **American Indian or Alaska Native** (β=−7.32) had lower LTV ratios
  - **Higher credit scores** positively impacted LTV ratios
  - **Lower credit scores** (types 7 or 8) negatively influenced LTV
  - **Ridge regression** outperformed other models, with the lowest RMSE (**27.08**) and the best R2(**0.050**)
- 2. Classification Testing
  - KNN accuracy rate is 76.66 percent.
  - LDA accuracy rate is 77.81 percent
  - Tree accuracy rate is 79.98 percent

05. Conclusion

- Regression: All 3 methods show associations between applicant race, income, and credit score and LTV ratios. We'd recommend the ridge regression model as it has the lowest RMSE and the best R2.
- Classification: Each of the three methods of classification predict the action taken on home loan requests accurately more than 75 percent of the time. Tree has the highest accuracy prediction rate, so we'd recommend using the Tree method to predict action taken. The Anova classification test shows race and ethnicity are strong predictors in predicting action taken on home loan requests.
- Ethical Implications: Racial and ethnic disparities still exist in home loan approvals in Mecklenburg County, NC.