DATA 6550 – Algorithmic Bias

Group 3 – Clay Skiles, Seth Phillips, Yousaf Khaliq, John Allard

INTRODUCTION/HISTORICAL CONTEXT

As the fields of data science, analytics, and machine learning continue to emerge and impact industries from medical imaging to behavioral science to retail, one of the most prevalent issues, and most difficult to eradicate, has been bias. Bias in data can exist in the form of human bias of those involved in the analysis, bias prevalent in the data collection, to bias in the implementation of the analysis results, and these biases can render the project ineffective if not addressed properly from the onset.

Within the lending industry, banks have attempted to create algorithmic models to help predict whether or not potential customers will default on loans. Mortgage lending is a critical step on a person's path to buying a home, and the majority of Americans are not able to afford a home upfront, making securing a lending agency a necessary step towards home-ownership. Ideally, these algorithmic models accurately advise on lending terms while improving company efficiency, increasing profits, and providing more accurate lending terms than a human alone could reliably provide. However, creating and implementing these models has been a process rife with historical bias. Redlining, racial covenants, and generational wealth disparities have all been present in the datasets used to train today's predictive models. This bias, in turn, is passed on to today's potential home-buyers, perpetuating historical bias today.

Racial covenants were legal clauses included in property deeds prohibiting the sale of the property to individuals of minority races - typically Black Americans, but also Asian Americans and other minorities. They were most commonly implemented in developing suburbs and urban neighborhoods between the 1910's and 1960's due to predominantly white communities' fears that a racially diversifying neighborhood would result in lower property values and decreased safety. These clauses were upheld by the federal government and actively encouraged for decades, until the Fair Housing Act deemed it illegal to discriminate based on race, religion, sex, and national origin in 1968.

Redlining, similar to racial covenants, was a zoning regulation method implemented by the Home Owners' Loan Corporation in the 1930's meant to categorize neighborhoods by mortgage approval eligibility. Redlining criteria explicitly included racial bias, referring to immigrants and minority communities as "detrimental influences" on a neighborhood's redlining categorization. Regardless of blatant racial bias, all citizens in the lowest ranking neighborhoods were either denied loans, or offered loans with exploitative terms, effectively limiting their opportunities to build wealth, creating financial segregation and generational wealth disparities.

Even though these practices are no longer upheld today, their presence can be seen in the racial and economic make up of neighborhoods, upholding unseen lines between communities.

And when current algorithmic models are trained on data gathered during this era, they can reassert the same disparities in our communities today.

METHODOLOGYDATA BIAS

Prior to performing the analysis, the dataset was cleaned and prepared to ensure accuracy. We noticed missing values in the dataset and handled those for key variables (e.g., income, interest rates) which were critical for evaluating loan approvals and financial disparities. To maintain the integrity of the analysis, rows with missing values were removed prior to statistical modeling. Additionally, we standardized data types to ensure that numerical variables were formatted correctly.

To find patterns of data bias, we performed an initial EDA to uncover patterns using various visualizations to examine potential disparities in loan approvals, interest rates, and property values across racial and different economic groups. Loan approval rates were compared between different demographic groups to determine whether certain groups faced higher rejection rates. Similarly, interest rate distributions were analyzed to detect patterns of discriminatory lending practices.

In addition to EDA, we used statistical modeling to quantify the impact of race, income, property value, and neighborhood demographics on loan outcomes. We performed a regression analysis to evaluate whether racial disparities continued and utilized a heatmap to identify key relationships between financial metrics and demographic characteristics.

Property values in areas with/without historical covenants

To assess property values in areas with/without historical covenants, the main priority was to identify correlations between race with property values, income, approval standing, and loan amount. These correlations can indicate possible selection and sampling bias in certain races when it comes to their financial and housing situation as the main concern is the grandfathering of covenant laws into modern day housing prices which affect minority groups. Many of these observations were explored in EDA:

  - Property Values vs. Covenant Density Compared for Each Race (The Main Observation)
  - Income distribution by Race (Assesses the economic equality of each race when looking at property values districted by covenants the county)
- Observing counts of covenants for whites vs. minorities (Harbor an idea of the covenanted districts for minority groups)Neighborhood demographic and economic characteristics

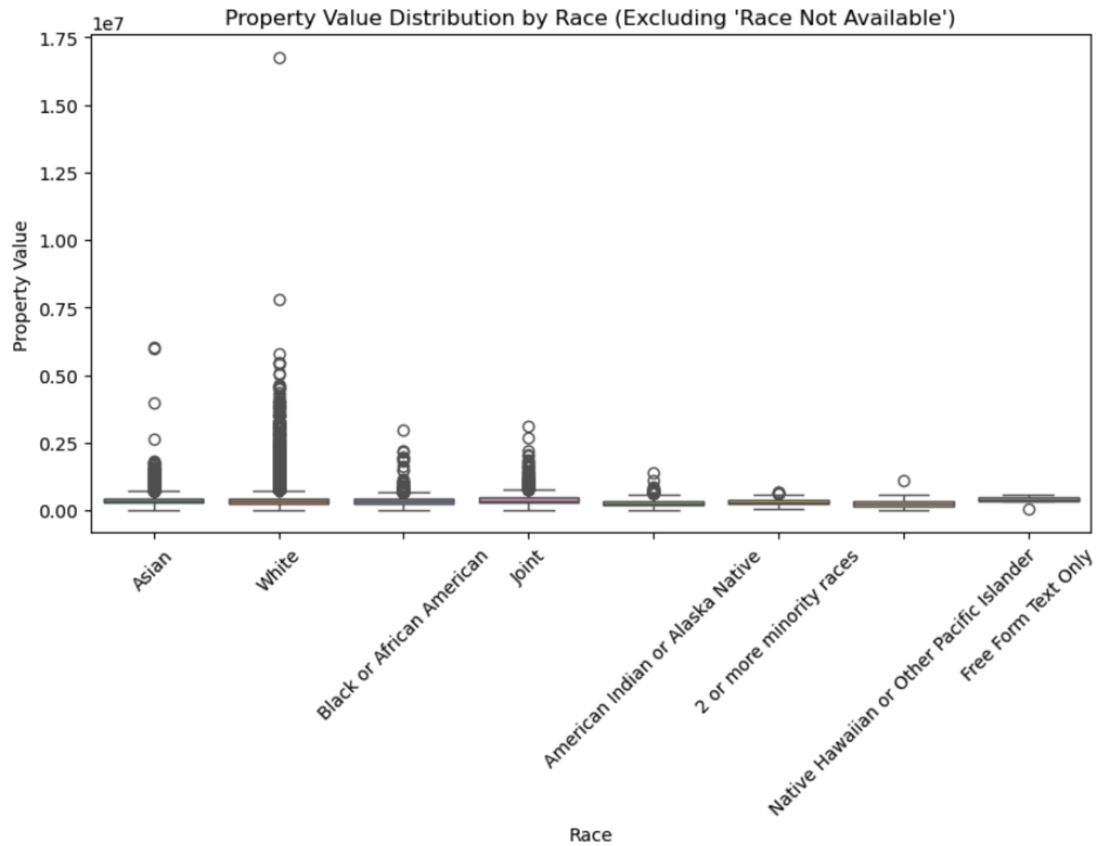
    - To find trends in demographic and economic characteristics in the data set we evaluated various fields such as: loan approvals, interest rates, and property valuations. These analyses are very influenced by the socioeconomic composition of neighborhoods and demographic data.
    - Some key observations:
        - Higher minority population percentages correlated with lower property values and higher interest rates.
        - Racial covenants continue to impact financial opportunities today as properties in those racially covenant areas are subjected to less favorable mortgage terms today.
- Patterns of generational wealth and investment
    - To find and assess patterns of generational wealth, we performed several analyses that looked into loan approvals disparities by race, interest rates for minority borrowers, and loan amounts vs the property values.
    - Some key observations:
        - White borrowers receive higher loan amounts for similar property values.
        - Minority borrowers face higher borrowing costs due to increased interest rates and loan denials, making it harder to build home equity.
        - Homeownership remains a primary driver of generational wealth, but the data shows that historical racial covenants (along with other historical barriers) continue to be a barrier for growth for non-white and marginalized groups.

ALGORITHMIC BIAS

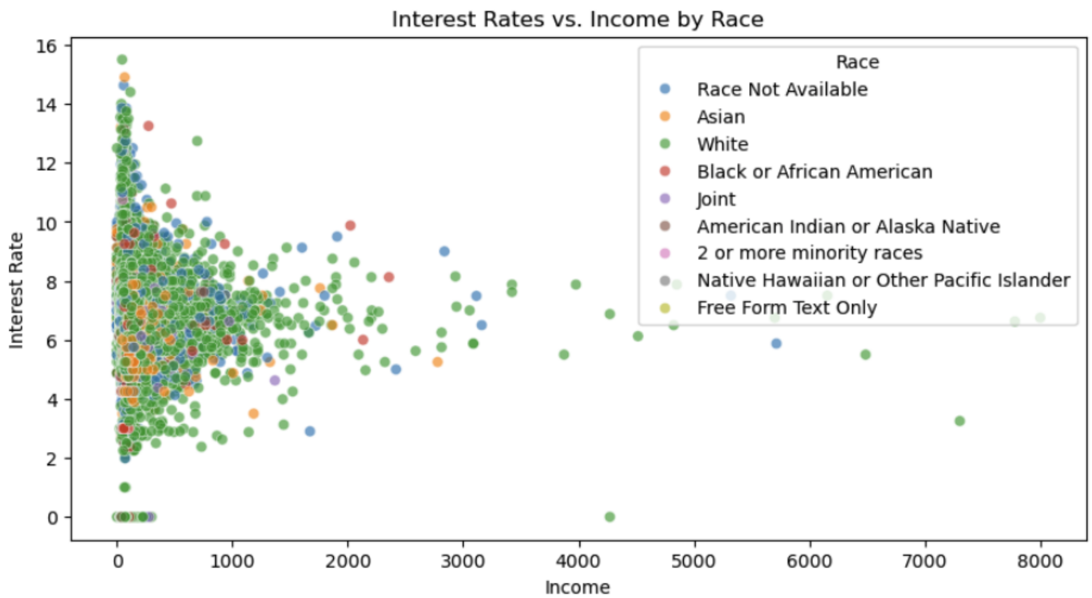
- Mortgage approval rates and terms
    - Kolmogorov-Smirnov Test (Interest Rate Disparities)
    - Demographic Parity (Approval Rates by Race)
- Risk assessment criteria
    - Logistic Regression (Effect of Race on Loan Approval)
- Box plot and Bar plot to visualize interest rates and feature importance

RESULTS

- DATA BIASProperty value distribution by race



    - White applicants tend to have higher property values compared to other racial groups.
    - Asian applicants also show higher property values, though slightly lower than the white population
    - Black, indigenous, and other minority applicants generally have lower property values, with fewer high-value outliers.
- Interest Rates vs Income by Race

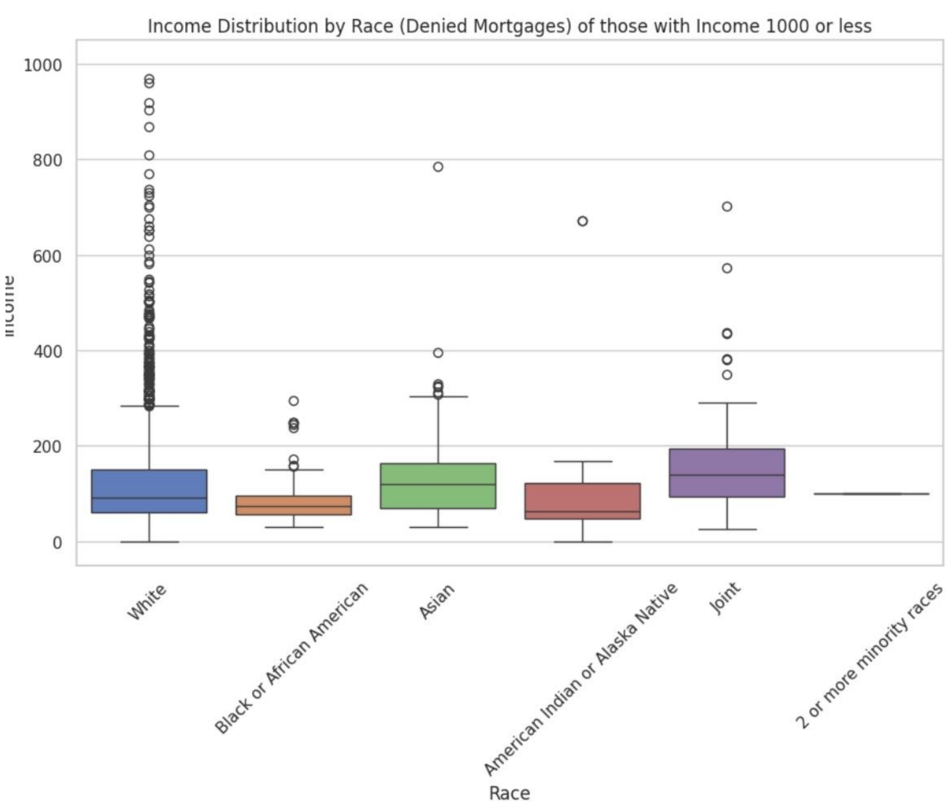


-
- The left side of the scatterplot is showing high concentration of points, showing higher interest rates for lower-income individuals.
- This suggests that lower-income borrowers face worse loan terms, regardless of race.
- There are race differences in certain racial groups seem to have more points in the higher interest rate range compared to white applicants.
- There are extreme outliers in the higher income levels that show high interest rates. Higher income should correspond with lower interest rates, but this doesn't seem consistent across all races.

- Property values in areas with/without historical covenants
  - There is a surplus of white people in the data set compared to the actual proportion of white people in the Hennepin County, MN demographic census. (84% of the dataset is white while the census of Hennepin County is 65% white)

derived_race	proportion
White	84.553570
Asian	6.571857
Black or African American	5.041665
Joint	2.918574
American Indian or Alaska Native	0.667180
2 or more minority races	0.166120
Native Hawaiian or Other Pacific Islander	0.081034

    - Suggests sampling bias from the oversampling of white citizens in the county, which could skew the outlook of property housing for minorities.)
- There is an abundance of outliers between the wealth in property values of white people in Hennepin County, MN compared to other demographics.
    - Although the IQR for each race is similar, the presence of outliers gives the impression of a sampling of minorities in underprivileged backgrounds, which may not apply to certain communities.



The counts of covenants in minority groups is larger with minority groups underrepresented in the dataset than white people.

	count	mean	std	min	25%	50%	75%	max
race_binary								
0	4269.0	5.959241	42.108751	0.0	0.0	0.0	0.0	561.0
1	9692.0	4.948617	41.569669	0.0	0.0	0.0	0.0	965.0

-
-

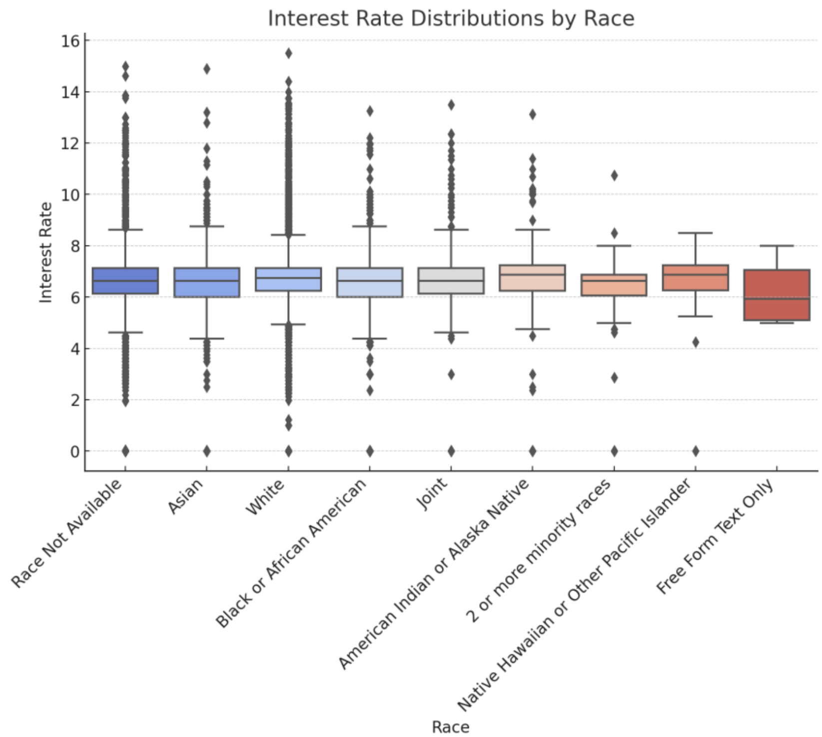
ALGORITHMIC BIAS

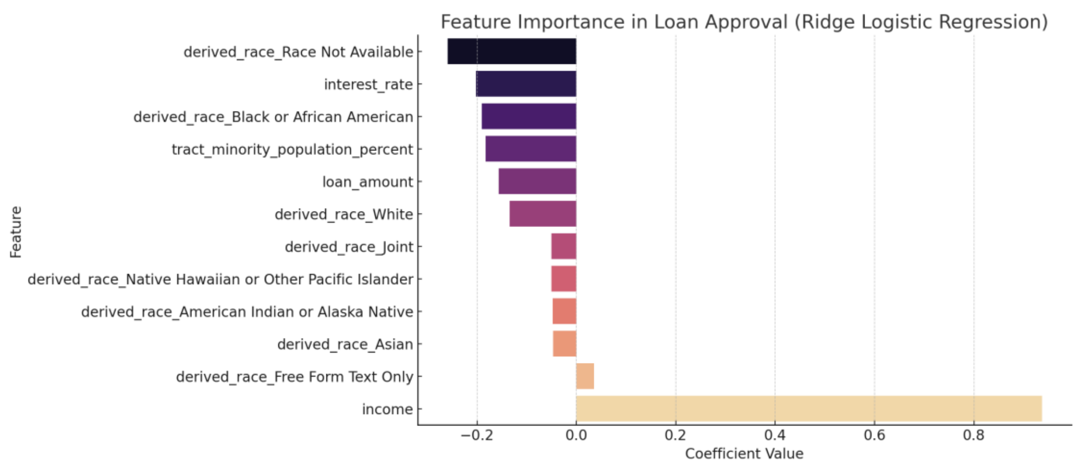
Logistic Regression:

- **Model Accuracy:**
  - **Training:** 90.3%
  - **Testing:** 90.6% (indicating a well-performing model)
- **Key Findings from Model Coefficients:**
  - **Income (+0.94)** → Higher income **increases** loan approval likelihood.
  - **Minority Population Percentage (-0.18)** → Higher minority population in a census tract **lowers** loan approval odds.
  - **Loan Amount (-0.16) & Interest Rate (-0.20)** → Higher values reduce approval chances.
  - **Race-Based Effects:**
    - **Black or African American (-0.19), Race Not Available (-0.26), White (-0.13)** → Negative impact on approval.
    - **Free Form Text Only (+0.035)** → Positive association.

Kolmogorov-Smirnov Test:

- Interest rate distributions **significantly differ** across racial groups.
- The **largest disparities** are observed in:
  - **Black or African American applicants ($p < 10^{-13}$)**
  - **American Indian or Alaska Native ($p < 10^{-5}$)**
  - **2 or more minority races ($p \approx 0.028$)**
- Suggests **systematic differences in interest rate assignments** based on racial identity.





<u>COMPARATIVE ANALYSIS</u>

DATA BIAS

Our analysis identified some biases in mortgage lending, demonstrating that historical housing discrimination continues to affect marginalized groups today.

1. Loan Approval Disparities

- Minority applicants face higher denial rates.
- Lending models seem to still have discriminative historical biases which results in marginalized groups receiving unequal access to homeownership

2. Higher Interest Rates for Minority Borrowers
    - Even with comparable financial profiles, minority borrowers are more likely to receive higher interest rates than more privileged demographic groups.
3. The Generational Wealth of Minorities
    - With an evident presence of generational wealth in the county, there is apparent bias to approve mortgages towards those who exemplify a more privileged background. In the case of the dataset studied, most of this wealth is evident in white and Asian citizens of the county who live in districts with fewer covenants.
    - Covenants are usually found in regions of lower income. As minority groups such as African Americans show high minority percentage statistics in these regions, the grandfathering of covenants into these neighborhoods leaves people in these communities barricaded from the opportunity to afford housing for the more privileged.

ALGORITHMIC BIAS

- Loan approvals are influenced by neighborhood demographics, reinforcing potential redlining effects.
- Clear racial disparities exist in approval rates and interest rate assignments.
- Black, Pacific Islander, and Native American applicants face systemic disadvantages.
- Income plays a significant role, but racial factors still have independent effects.
- The bar plot shows the importance of each feature regarding loan approval. Income has the strongest positive effect on approval, confirming that financial status is a major determinant. Loan amount and interest rate have negative coefficients, meaning larger loans and higher interest rates reduce approval chances. Race-based coefficients (e.g., Black or African American: -0.19) suggest racial disparities remain even after accounting for financial factors.
- The box plot shows the interest rate distributions by race. Minority groups tend to receive higher interest rates than White applicants. Black and Native Hawaiian applicants show a wider spread of higher interest rates compared to other groups.

Kolmogorov-Smirnov test confirms statistically significant differences in interest rate distributions, meaning minority groups face systematically different (worse) lending terms.

-

RECOMMENDATIONS

DATA BIAS MITIGATION

a) Balance the Training Data

- Ensure representation from underrepresented communities through targeted data collection.
- Apply techniques like SMOTE or reduce overrepresented groups to balance datasets.
- Regularly check for class imbalances and apply corrective measures.

b) Human-Centered Approaches

- Routinely audit the training data using metrics like disparate impact and demographic parity.
- Use techniques like reweighting, adversarial debiasing, or post-processing adjustments.

c) Avoid Proxy Variables

- Detect and remove variables that correlate with sensitive attributes (e.g., ZIP codes, occupations).
- Use tools like SHAP or LIME to evaluate feature importance and monitor for bias.

d) Transparent Data Collection

- Maintain detailed records of data origins, collection methods, and potential biases.
- Follow ethical guidelines and obtain informed consent, especially from vulnerable groups.

e) Diverse Stakeholder Involvement

- Collaborate with diverse stakeholders to identify biases in data collection.
- Form advisory groups to monitor and evaluate data practices for fairness.

BIAS MITIGATION FOR ALGORITHMIC BIAS FOR MORTGAGE LENDING ALGORITHMS

a) Fairness Constraints in Model Training

- Use fairness-aware algorithms such as:
  - **Demographic Parity Loss**: Ensures similar approval rates across groups.
  - **Equalized Odds Constraints**: Ensures similar recall and false-positive rates.

b) Algorithmic Adjustments

- Implement **"Reject Option Classification"**, where **uncertain decisions are manually reviewed** to prevent biased denials.
- Use **Fairness-Aware Boosting Models** that **penalize unfair decisions** during training.

c) Regularization for Fairness

- Apply **Fairness Constraints in Ridge Regression** (modified L2 regularization) to ensure race-related variables do not excessively influence predictions.

d) Regular Fairness Testing & Reporting

- Conduct periodic audits using fairness metrics (demographic parity, equalized odds, disparate impact).
- Compare loan decisions against historical biases to measure progress.

e) Comply with Fair Lending Laws

- Ensure compliance with **Equal Credit Opportunity Act (ECOA)** and **Fair Housing Act**.
- If using AI for lending, document bias mitigation efforts for regulatory compliance.