Housing Discrimination

Adriana Alvarado, 2023, alvadri@bu.edu

Tianyi Bao, 2023, tianyib@bu.edu

Rafael Perron, 2023, perronr@bu.edu

David Sullo, 2023, dsullo@bu.edu

Song Xu, Masters, songxu@bu.edu

**Introduction:**

*# Rafael contributed work*

In this project, we have been looking at the effect of discrimination on housing loans across Boston. We look at loan rejection rates and reasons and look at what demographics are more disproportionately affected by overall loan rejection or certain types of reasons for rejection. We use our trends and findings to try to predict which areas and demographics are more likely to participate in first time home ownership programs after having a loan rejected. We also extend our data to look at Boston’s history with redlining to guage if historical practices still have a lasting effect on discrimination today.

Our goal with this project is to provide evidence for or against the presence of discrimination in housing loans and more importantly, find potential causes and areas that city council workers can focus on when making policies. This work is super important, and breaking down the data is super helpful to finding a solution to discrimination in this city.

We want to be open minded in this project, and explore different trends we found and intuitions we had. This is a super difficult problem to break down and solve, so we experimented with different theories and hypotheses that could explain the findings in our data.

**Base Analysis:**

*# Tianyi contributed work*

We have been able to answer the following questions through our analysis with the dataset:

1. By county how many loans were taken.

By doing some primary research on the HDMA dataset, we were able to determine the number of loans taken by each county. Below is the histogram we have plotted.

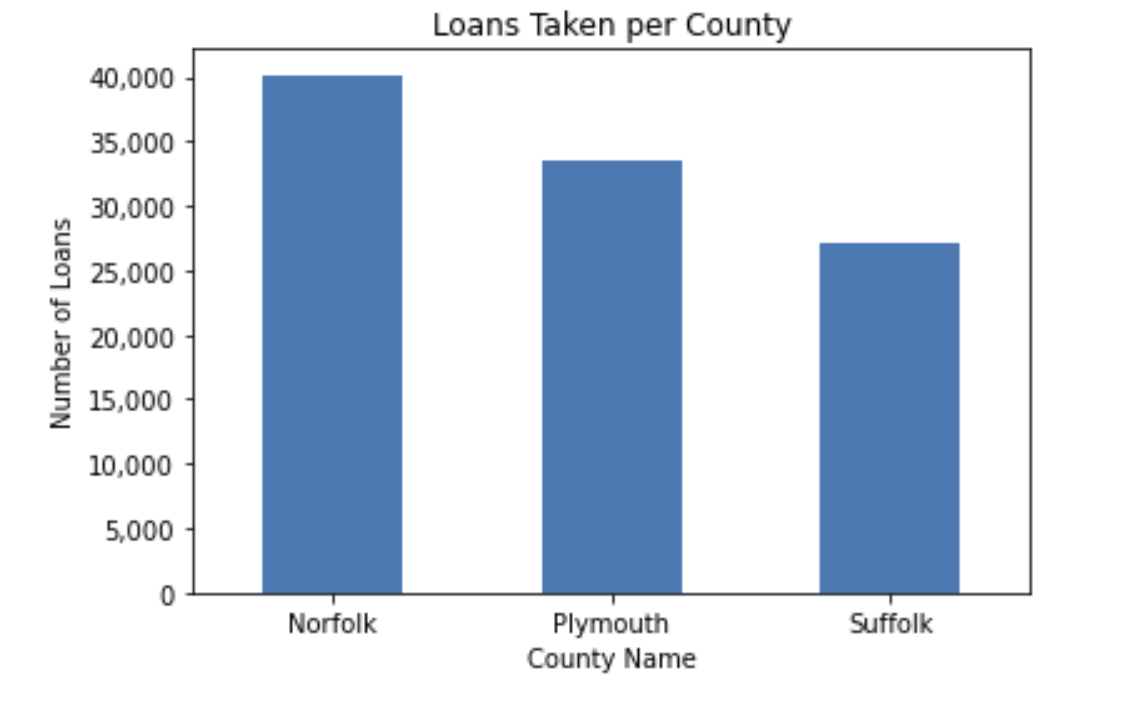


Figure 1: Loans taken per county. The y-axis represents the number of loans taken and the x-axis represents the county code. 25021 is the FIPS code for Norfolk, 25023 is the FIPS code for Plymouth, and 25025 is the FIPS code for Suffolk.

Based on the analysis of the provided data, the following observations were made:

1. Norfolk residents took the highest number of loans among the three counties, with 40,175 loans, representing 39.86% of the total loans taken.
2. Plymouth residents took the second-highest number of loans, with 33,453 loans, representing 33.20% of the total loans taken.
3. Suffolk residents took the lowest number of loans among the three counties, with 27,140 loans, representing 26.94% of the total loans taken.

Furthermore, we have also investigated the mean income in those counties. Below are the results:

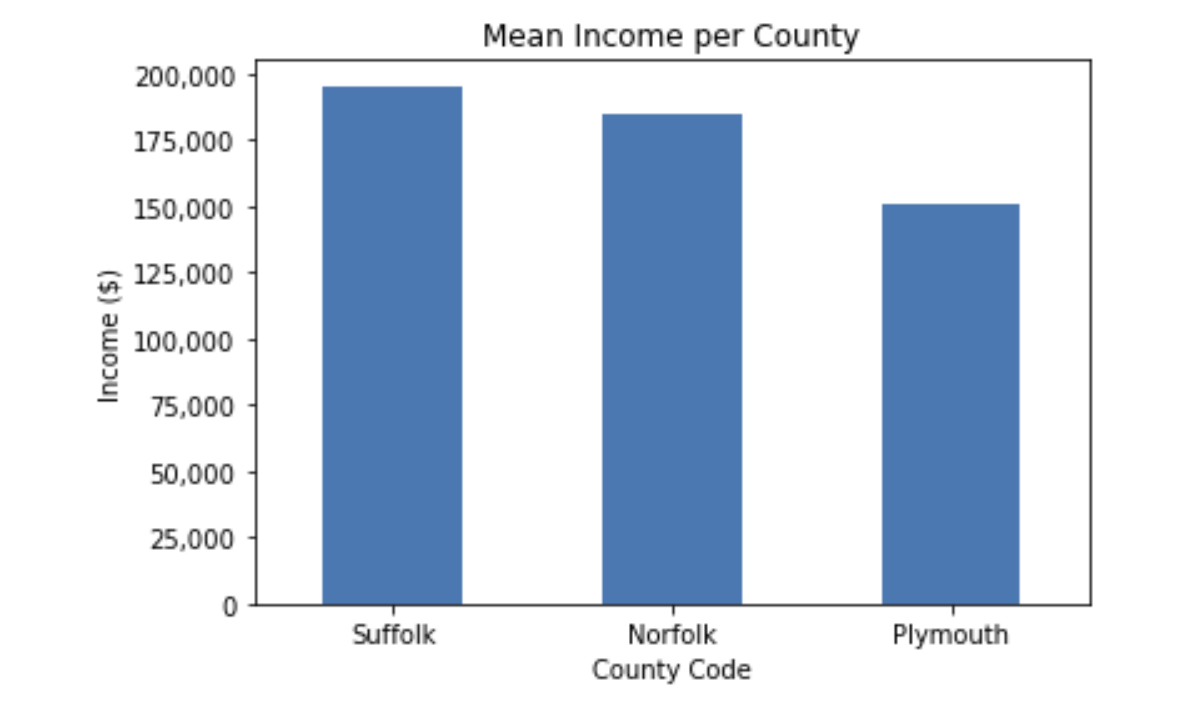


Figure 2: Mean income per county. The x-axis represents county code and the y-axis represents mean income, in thousands of dollars.

To analyze the relationship between loan distribution and mean income, the number of loans taken per $1,000 of mean income is calculated for each county:

1. Norfolk: 40,175 loans / ($184,513.91 / $1,000) = 217.60 loans per $1,000 of mean income
2. Plymouth: 33,453 loans / ($151,138.41 / $1,000) = 221.29 loans per $1,000 of mean income
3. Suffolk: 27,140 loans / ($195,312.54 / $1,000) = 138.91 loans per $1,000 of mean income

The analysis of loan distribution in relation to mean income reveals the following insights:

1. Plymouth residents have the lowest mean income among the three counties but take a higher number of loans per $1,000 of mean income (221.29) than Norfolk residents (217.60).
2. Norfolk residents have a higher mean income than Plymouth residents, but their loan distribution per $1,000 of mean income is only slightly lower (217.60).
3. Suffolk residents have the highest mean income among the three counties but take significantly fewer loans per $1,000 of mean income (138.91) compared to Norfolk and Plymouth residents.

The additional analysis indicates that there is a relationship between loan distribution and mean income among the three counties. While Suffolk residents have the highest mean income, they take substantially fewer loans per $1,000 of mean income compared to residents of Norfolk and Plymouth. This may suggest that higher-income individuals in Suffolk have less reliance on loans or may have alternative means of financing.

Further research could explore factors such as the cost of living, credit scores, and borrowing habits in each county, to better understand the reasons behind these differences in loan distribution and mean income.

1. Who is participating in first time home ownership programs?

As there are hundreds of data fields available in the dataset, we have to pick the essential parameters before trying to answer this question. As a result, we have picked the following data fields:

1. 'applicant\_age': The age of the applicant.
2. 'derived\_dwelling\_category': Derived dwelling type from Construction Method and Total Units fields for easier querying of specific records.
3. 'derived\_race': Single aggregated race categorization derived from applicant/borrower and co-applicant/co-borrower race fields.
4. 'income': The gross annual income, in thousands of dollars, relied on in making the credit decision, or if a credit decision was not made, the gross annual income relied on in processing the application.
5. ‘denial\_reason-1’: The principal reason, or reasons, for denial.
6. 'derived\_sex': Single aggregated sex categorization derived from applicant/borrower and co-applicant/co-borrower sex fields.
7. 'action\_taken': The action taken on the covered loan or application.
8. 'loan\_type': The type of covered loan or application.

As there is no clear definition of how first time home ownership programs should be counted, we hypothesized that members of these programs would be previously denied from loans before entering the programs. Here are the results we have found:

1. Distribution of different races:

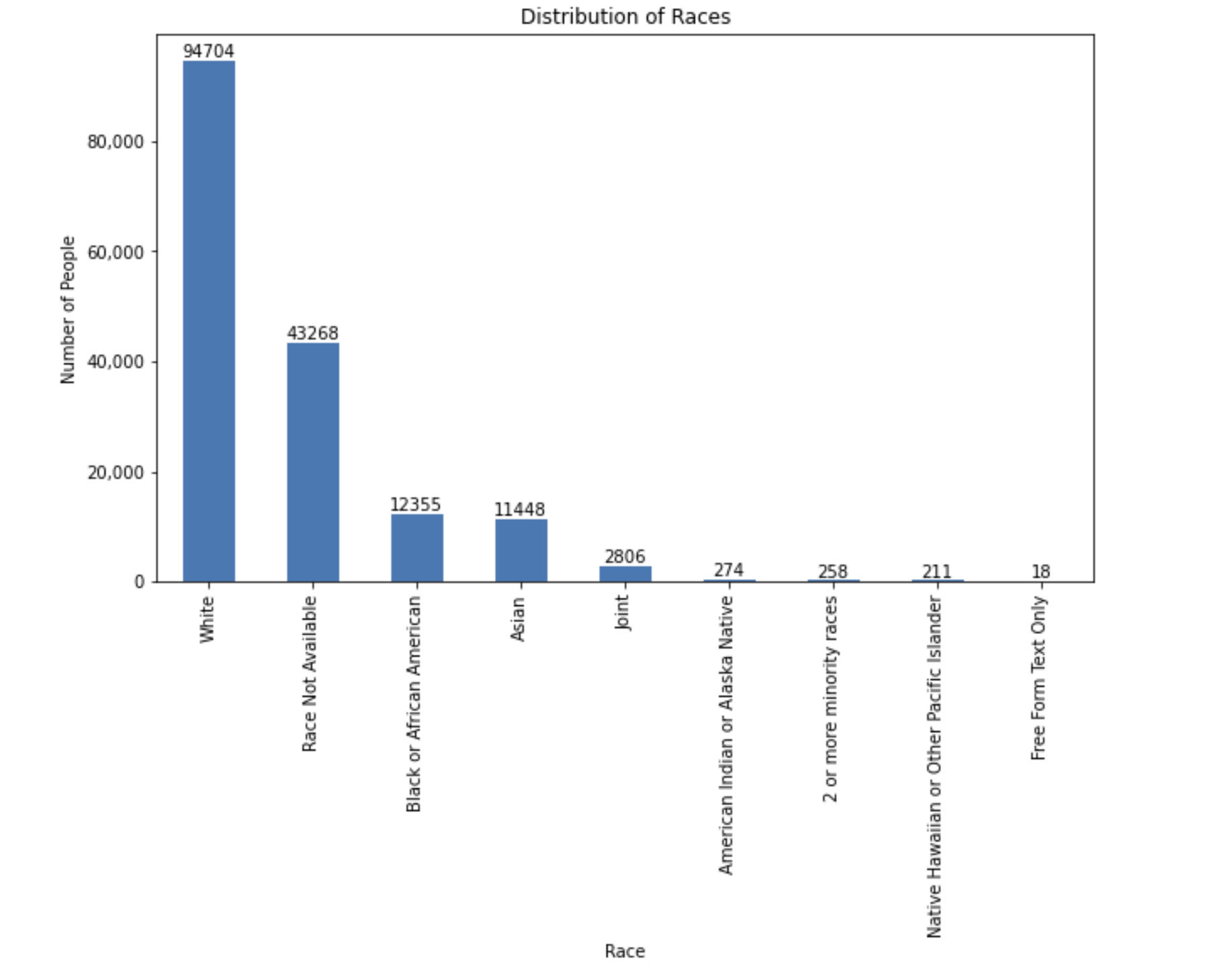


Figure 3: Distribution of different races participating in first time home ownership programs. The x-axis represents races, the y-axis represents the corresponding number of people

* + White: 16,332
  + Race Not Available: 4,340
  + Asian: 3,023
  + Black or African American: 1,654
  + Joint: 681
  + American Indian or Alaska Native: 37
  + 2 or more minority races: 39
  + Native Hawaiian or Other Pacific Islander: 26

1. The distribution of income for different derived races shows considerable variation, with American Indian or Alaska Native applicants having the highest standard deviation, and White applicants having the highest maximum income.

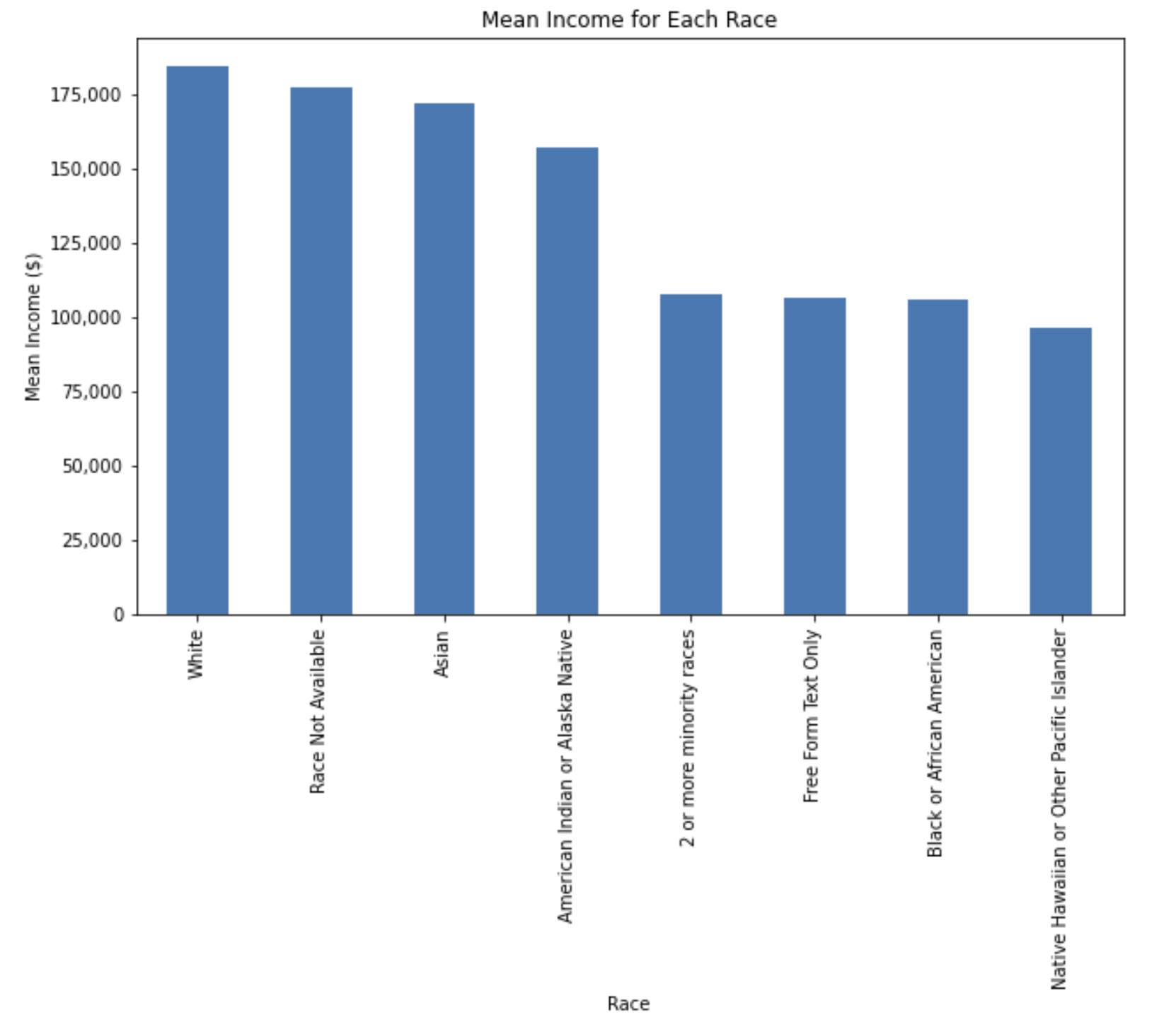


Figure 4: Distribution of income among races. The x-axis represents race, the y-axis represents the corresponding income, in thousands of dollars.

1. Distribution for applicant's age and income:
   * 25-34 years old: Mean income of 149.01
   * Below 25 years old: Mean income of 140.29

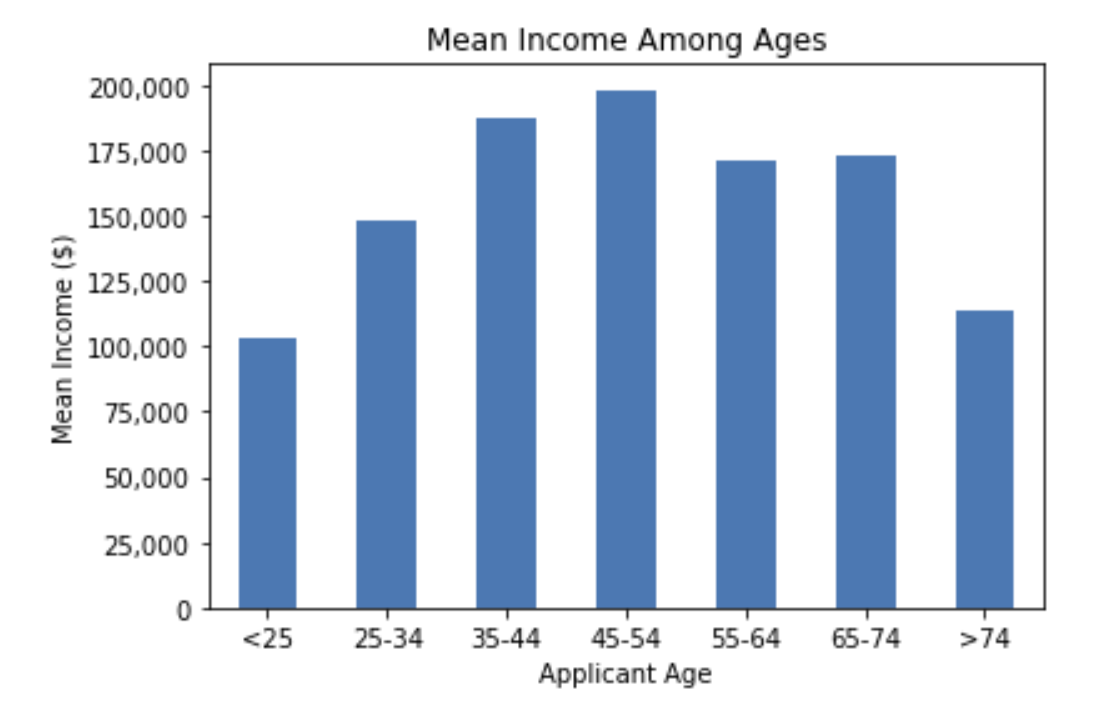


Figure 5: Distribution of income among ages. The x-axis represents applicants’ age, the y-axis represents their corresponding income.

1. Average Income by Sex:

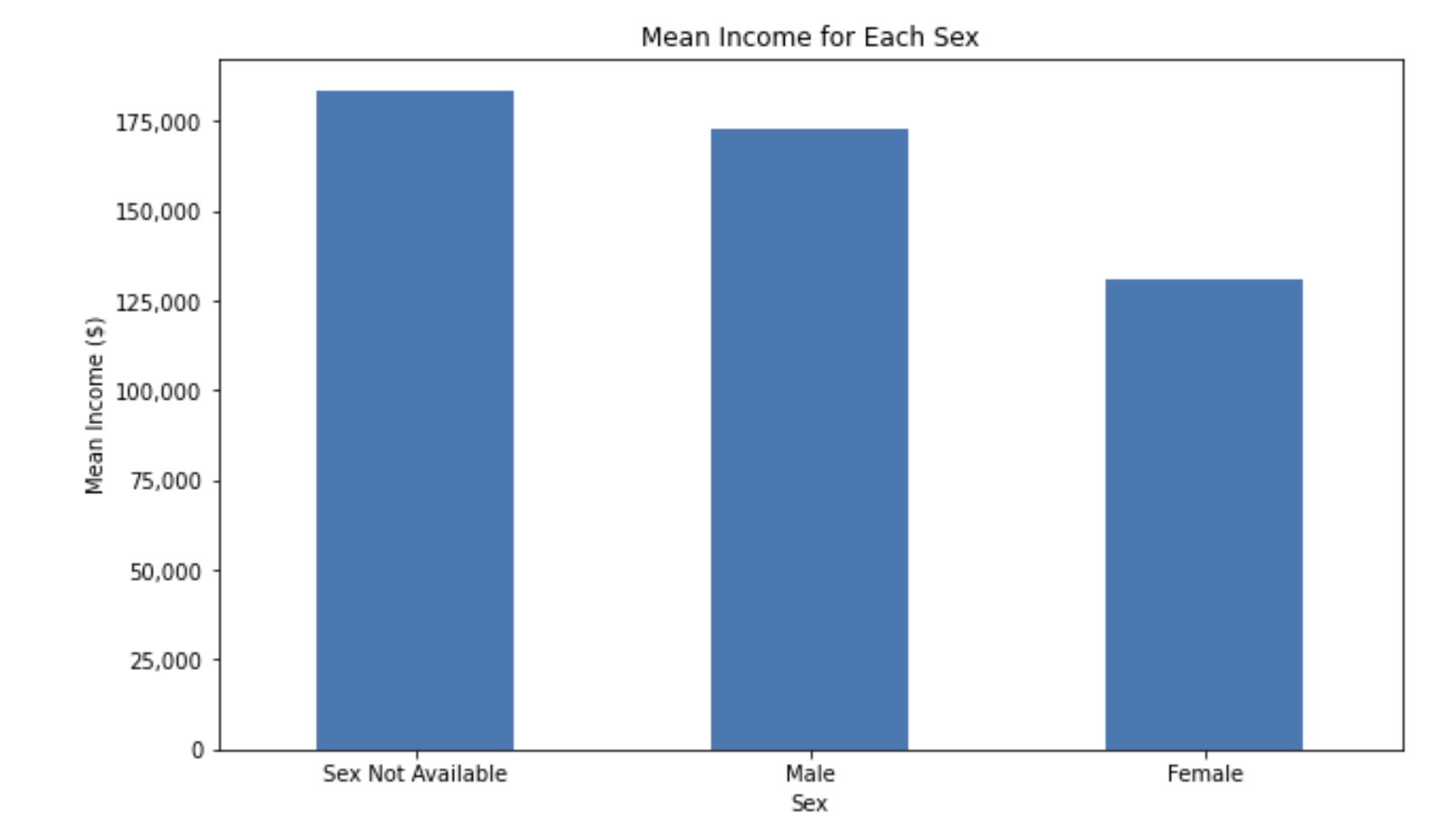


Figure 8: Average income by sex. The x-axis represents sex categories, the y-axis represents their corresponding average income.

* + Sex Not Available: 183.45
  + Male: 172.86
  + Female: 130.71

1. Loan Rejection Rates by Race:

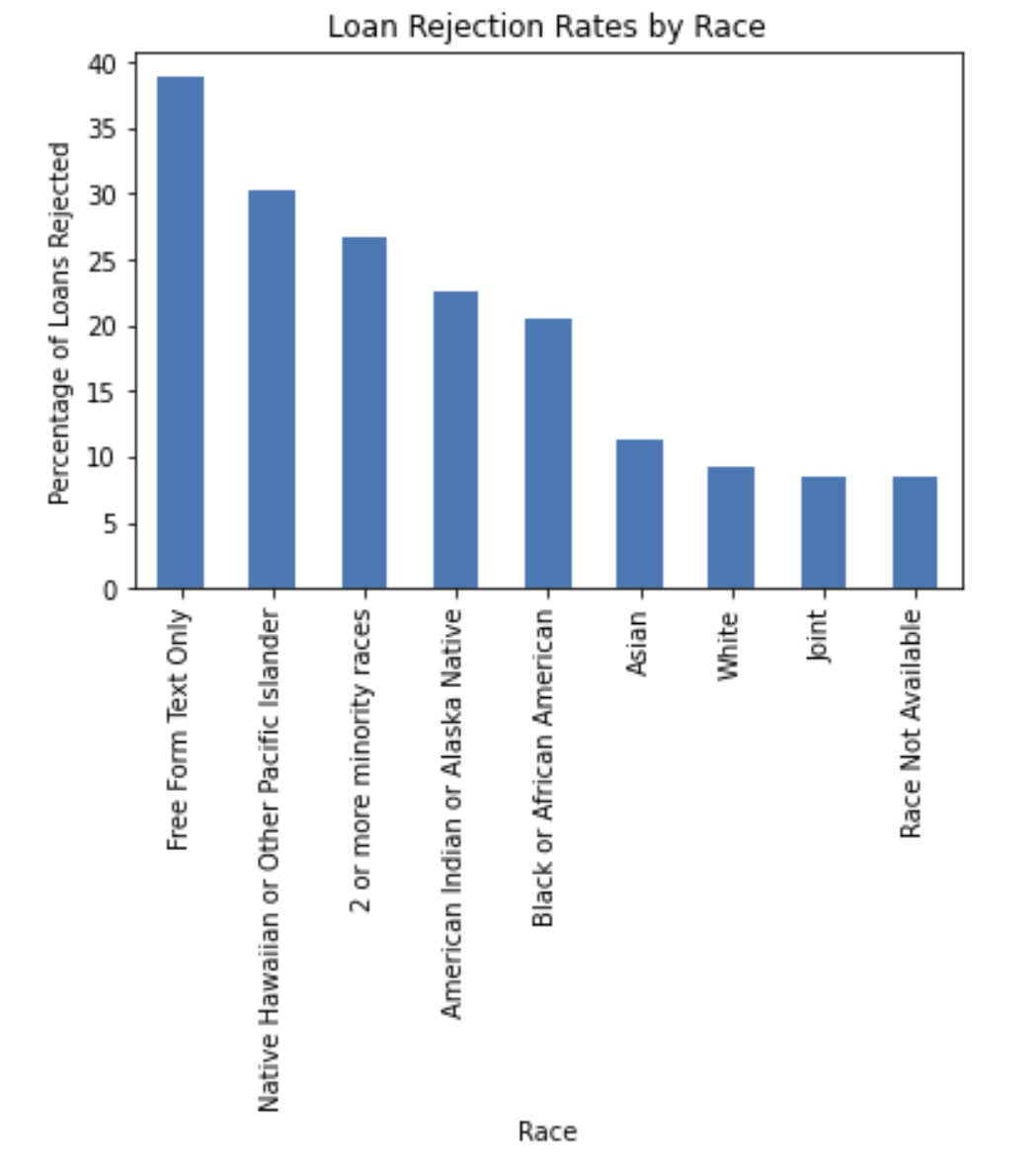


Figure 9: Loan rejection rates by race. The x-axis represents race groups, the y-axis represents percentage of loans rejected.

* + Highest: Free Form Text Only (38.89%)
  + Lowest: Race not Available (8.52%)

1. Loan Rejection Rates by Sex:

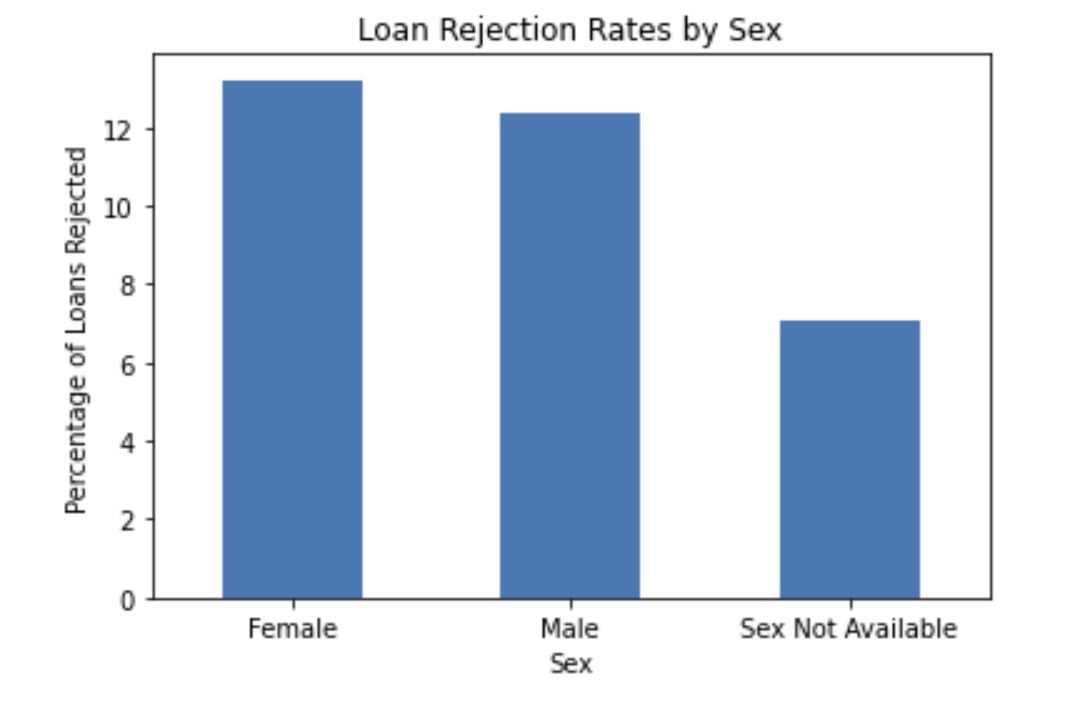


Figure 10: Loan rejection rates by sex. The x-axis represents sex categories, the y- axis represents percentage of loans rejected.

* + Female: 13.22%
  + Male: 12.39%
  + Sex Not Available: 7.05%

1. Loan Rejection Rates by Age:

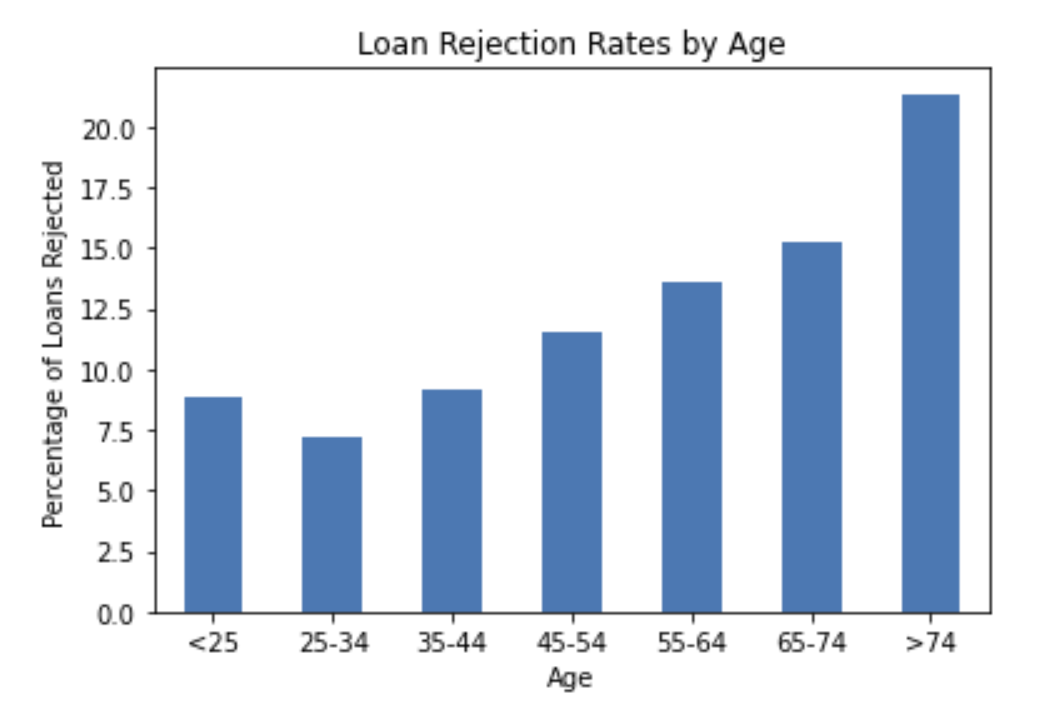


Figure 10: Loan rejection rates by age. The x-axis represents age ranges, the y- axis represents percentage of loans rejected.

1. Loan Rejection Rates by Loan Type:

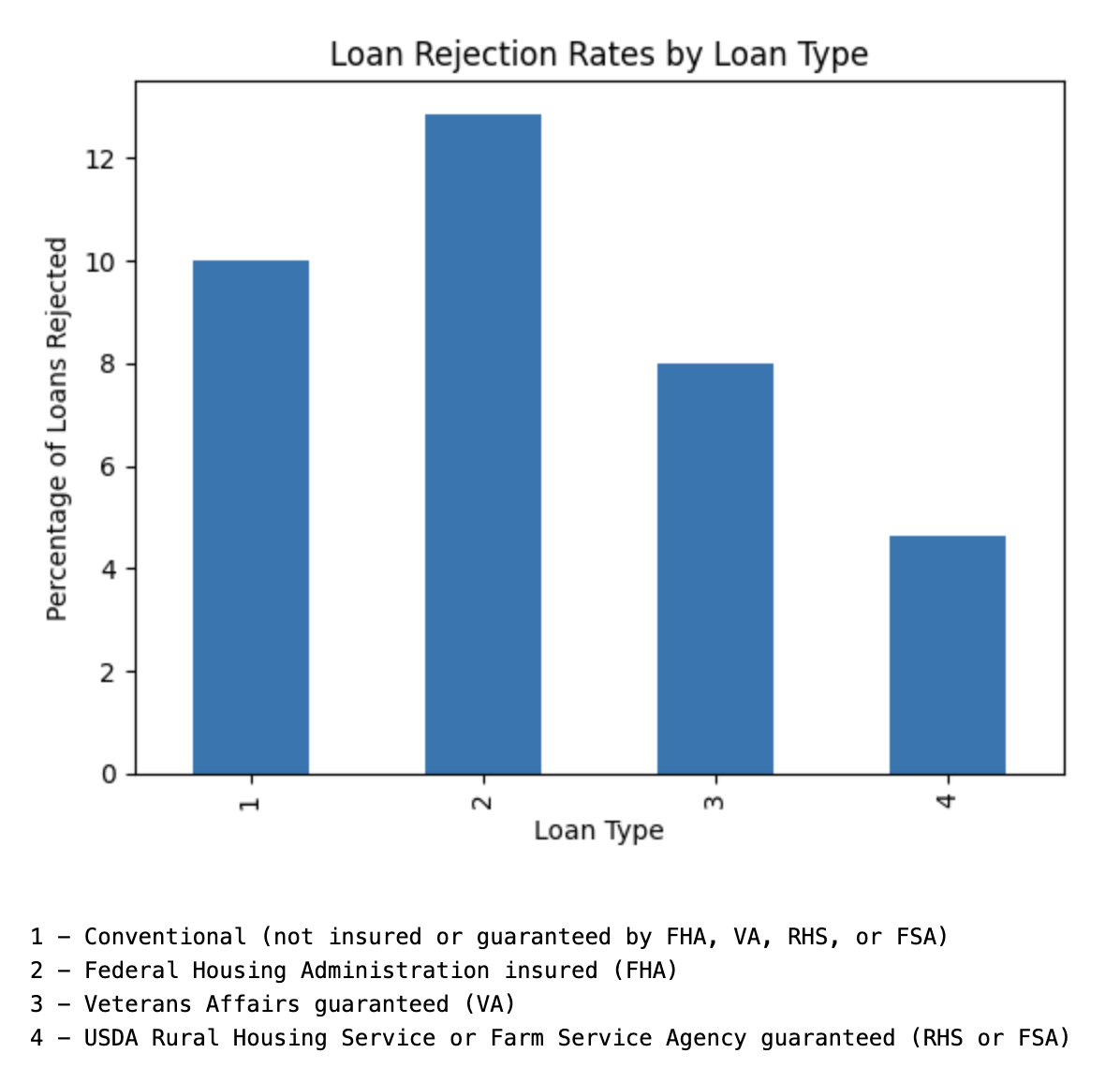


Figure 11: Loan rejection rates by loan type. The x-axis’ numbers represent loan type, as described above. The y-axis represents the percentage of loans rejected.

* + Highest: Federal Housing Administration insured (FHA) (12.86%)
  + Lowest: USDA Rural Housing Service or Farm Service Agency guaranteed (RHS or FSA) (4.63%)

1. Number of Loans Rejected for Each Denial Reason:

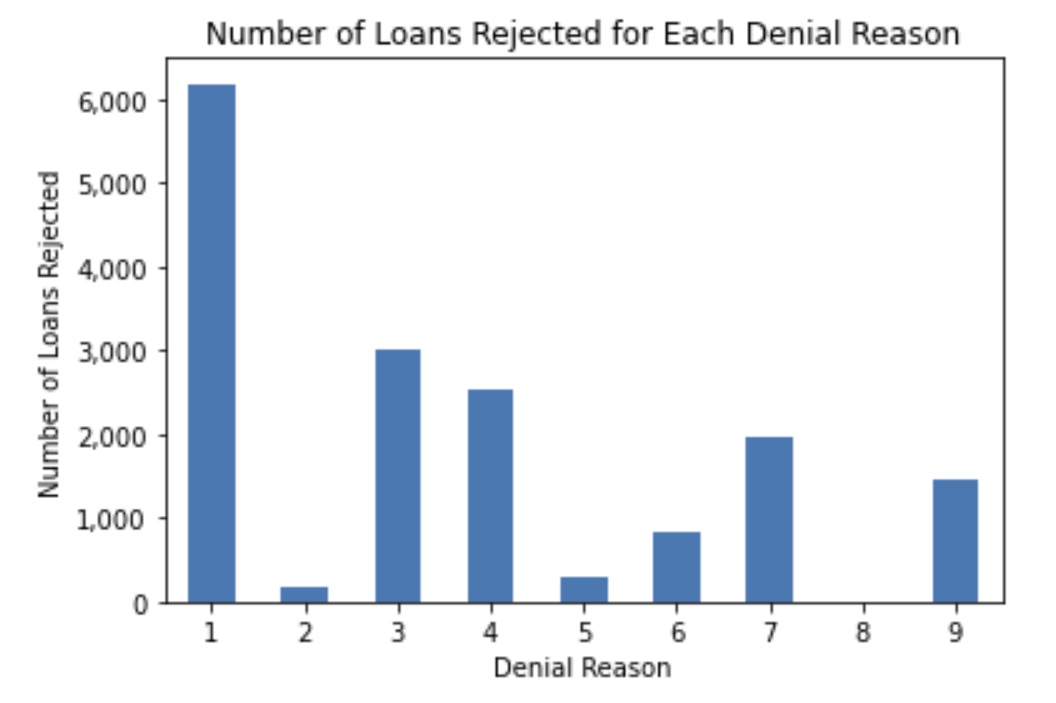


Figure 12: Number of loans rejected for each denial reason. The x-axis’ numbers represent denial reasons, as described below. The y-axis represents the number of loans rejected. It is important to note that the majority of reasons (1, 3, and 4) are monetary.

* + Debt-to-income ratio: 6,178
  + Employment history: 181
  + Credit history: 3,014
  + Collateral: 2,525
  + Insufficient cash (downpayment, closing costs): 278
  + Unverifiable information: 827
  + Credit application incomplete: 1,954
  + Mortgage insurance denied: 3
  + Other: 1,449

The analysis of first-time homeownership program participation reveals disparities in income, loan rejection rates, and other factors based on race, sex, dwelling type, and loan type. Some key findings include higher loan rejection rates among certain races, differences in average income based on sex, and variations in mean income for different dwelling types among accepted and denied applicants. The most common reasons for loan rejection include debt-to-income ratio, credit history, collateral, and incomplete credit applications. These findings can help stakeholders develop targeted strategies to address these challenges and improve access to first-time homeownership programs. Potential solutions may include financial education programs, credit counseling services, and alternative financing options for applicants with less than ideal financial profiles.

**Extension Analysis:**

*# David contributed work*

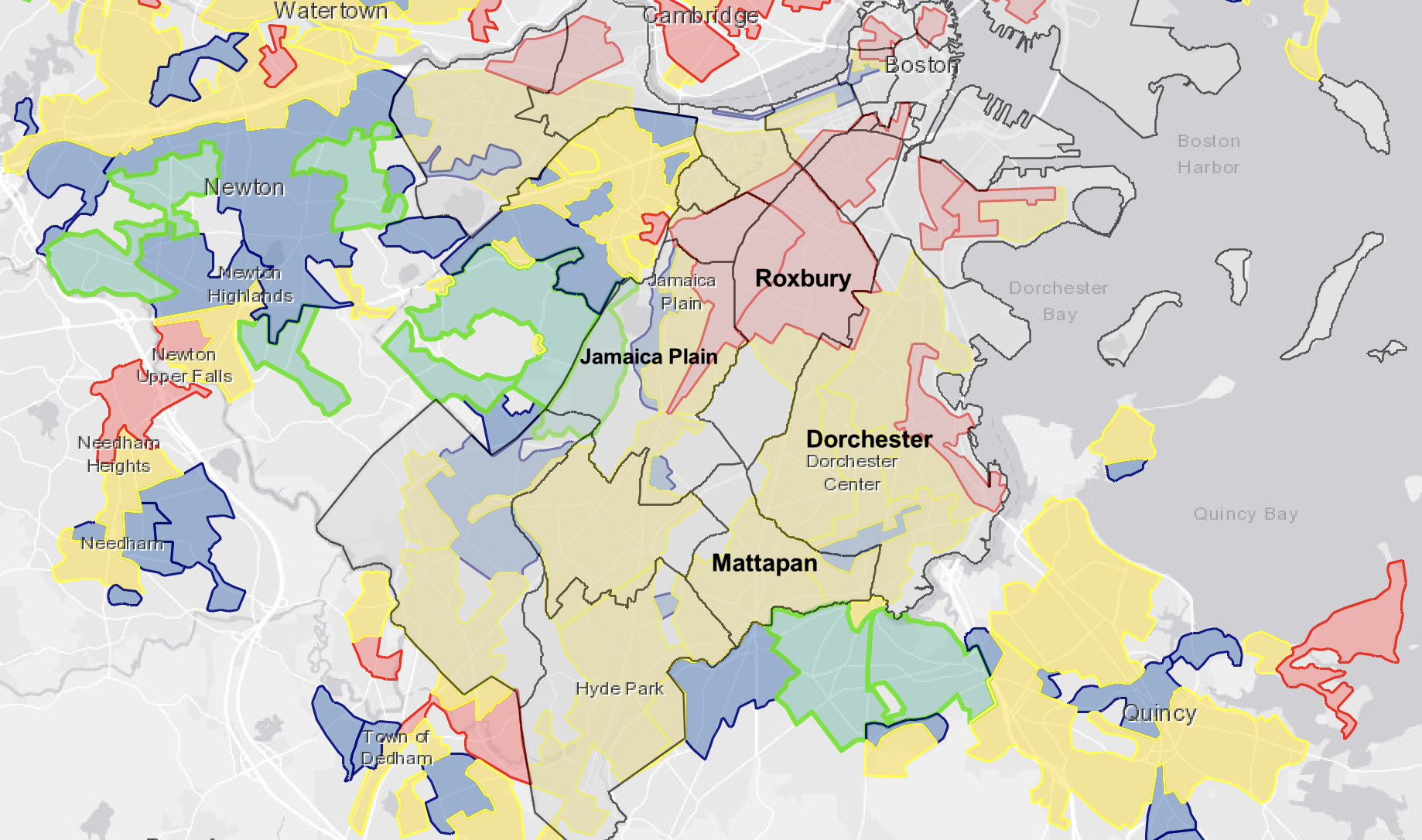
After our analysis of the initially provided HMDA data set, our team was able to identify potential points of discrimination in loan approval rates by race. While we have been able to do productive analysis on this data, we believe that looking at data for Redlining and Neighborhood Composition in Boston will allow us to gain a deeper understanding of first-time home buyers discrimination. We chose these points of data as Redlining highlights the perception of neighborhoods by loaning corporations and we can then make connections based on the ethnic compositions of the neighborhoods. Additionally, we believe that by making these additional discoveries, we will be able to help guide Councilor Louijeune on how to approach this discrimination problem.

Figure 13: This map shows the history of redlining in Boston. Yellow and red indicate “Declining” and “Hazardous” neighborhoods. Blue and Green indicate “Still Desirable” and “Best” neighborhoods.

I began by looking through the Redlining data and selected several neighborhoods in various areas of Boston with different Redlining Grades to look for any correlations. From here, I identified the following neighborhoods as “Positive Rated” based on having a grade of either “Best” or “Still Desirable”: Fenway, Jamaica Plain, Back Bay, West Roxbury, Newton, and Brookline. Next, I grouped together some neighborhoods with “Declining” or “Hazardous” ratings: North End, West End, Roxbury, South Boston, Dorchester, Longwood, Roslindale, and Mattapan.

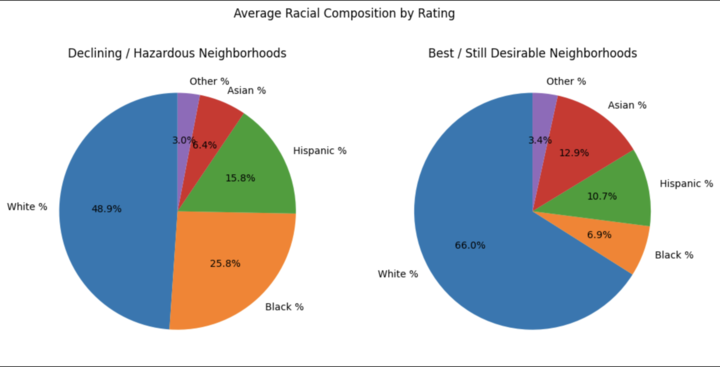
The first thing I noticed was that many of the “Declining” neighborhoods were located next to each other, with typically one or two of these boarding a “Hazardous” neighborhood. A similar pattern was found with the “Still Desirable” neighborhoods being adjacent to a “Best”. Additionally, I found that within the city limits of Boston, there are not any neighborhoods where a “Best” neighborhood can be located next to a “Hazardous” one, indicating that “Best” neighborhoods are well isolated from poorer graded neighborhoods.

Figure 14: These pie charts indicate the percentage of each race make up neighborhoods originally classified as “hazardous”, and the neighborhoods originally classified as “best.”

Next, I looked at the ethnic compositions of these areas to look for any other noteworthy trends. I used US Census Data for Newton and Brookline, and then “Boston Neighborhood Demographics, 2015-2019” for specific neighborhoods to characterize their population breakdown. For ease of interpretation, I am again using the simplified Positive and Negative ratings from the Redlining Grading. Amongst neighborhoods with a negative rating, on average, there are nearly 19% more Black and African American people and just over 5% more Hispanic people. In positively-rated neighborhoods, there are roughly 17% more White People and 6.4% more Asian People. Moreover, for both types of neighborhoods, this data backs up our previous research which showed that White and Asian People see much higher loan approval rates compared to African Americans, Hispanics, and other ethnic minorities.

**Overarching Project Questions:**

*#Adriana contributed work*  
1. Which (census tract) areas receive what type of loans? How many loans does each area receive? Are there any discrepancies between certain areas or the rate at which certain types of loans are given?

We can use the Home Mortgage Disclosure Act (HMDA) database to determine which census tracts receive what type of loans and how many loans each area receives. By analyzing loan data at the census tract level, we can create maps and charts that display loan information, including the number of loans, loan amounts, and loan types. HMDA allows for a myriad of options for filtering across different map levels, year, action taken, loan type, loan purpose, ethnicity, race, age, sex, lien status, construction method, total units, loan product and dwelling category. This will allow as much flexibility in terms of finding the evidence that will support the above questions, in addition to answering any lingering questions that come to mind during the analysis as well as analysis of the extension proposal. In addition, there is a supplemental dataset that would be useful in answering the above questions, which is the datasets and visualizations from the Massachusetts Community and Banking Council (MCBC), which includes key mortgage lending and small business lending data. For example, to target these questions, we observed a relationship between a higher mean income leading to a lower loan distribution, while a lower mean income lead to a higher loan distribution - portrayed in Figure 2.

*#Adriana contributed work*  
2. Establish home environment: view of housing by neighborhood/more granular level: who is receiving loans and who are purchasing homes.

To establish a view of housing by neighborhood or at a more granular level, we can examine loan and purchase data at the census tract level. By analyzing this data, we can identify who is receiving loans and purchasing homes in a particular neighborhood or area. Using the MCBC, we can look into visualizations such as home loans by income and race, home loans by municipality by income and race, home loans by municipality by lender and type and many more. Using the Home Mortgage Disclosure Act (HMDA) database, we can examine loan and purchase data by demographic characteristics, such as race, ethnicity, gender, and income. This can help us identify any patterns or trends in the data that suggest unequal access to credit or discrimination, such as lower loan approval rates for borrowers in certain demographic groups, or lower rates of home ownership among certain groups. Within our findings, it was concluded that Norfolk county (county with highest mean income), took the highest percentage of loans among the other two counties, Suffolk and Plymouth - portrayed in Figure 1.

*#Adriana contributed work*

3. Establish who is participating in first time home ownership programs

To establish who is participating in first-time home ownership programs, we can gather data from program administrators or agencies that offer such programs. This may include government agencies, non-profit organizations, and community development corporations.

Once we have identified the relevant programs, we can examine the demographic characteristics of participants, such as race, ethnicity, gender, age, and income. This can help us identify any patterns or trends in participation that suggest unequal access to home ownership programs or discrimination. Further analysis into factors that first time home ownership programs look for like credit score/type and income would be beneficial as it proves that lower income people will be more likely to join the program, and this confirms that those applying for single family are the lowest income (done through correlating average income and housing type for those who were denied vs with those who weren’t denied). To have a better understanding as to why these discrepancies occur, looking into datasets from census block groups in Boston, Boston’s social vulnerability, and the United States census can be helpful in doing so.

**Project Takeaways:**

*# Rafael contributed work*

1. While minimal concrete answers were found in this project, our data gave us intriguing insights on which further groups can extrapolate. People of color do face higher rates of loan rejection, but they also have lower average income. As we saw, most denial reasons given for loans are monetary, so this begs the question: are people of color being more frequently denied, or strictly because they more often, make a lower wage? The answer to that question will determine what actions councilors will need to take; although, both answers point to deep rooted issues of either personal, or systemic discrimination. Accounting for this is the next step of the project that we were not able to reach. By modeling income and rejection rates and removing bias as best we can, we can predict an expected rate of rejection for every demographic and compare the actual rates to see how far off they really are. That would give greater insights into what types of issues people are facing in the loaning process.
2. Simply outlawing redlining has not erased the damage it has done and continues to do. Neighborhoods deemed unsafe often have lower values and access to resources like transportation and quality education. Moving away from these neighborhoods is often tough, and it can cause cycles of families to stay in these neighborhoods. These neighborhoods being predominantly composed of people of color shifts people’s view of the demographics, and, in our opinion, can contribute heavily to discrimination in the loan process.
3. The job is far from done, and more work such as this needs to be done to uncover the root of this problem in Boston.