

HMDA Mortgage Data

Finn, Owen, Ben, Joe





What the data shows



The unfiltered HMDA Mortgage data set consist of 143349 (when filtered 2251) rows and 99 columns of data from loans given by The Bank of America in Oregon, 2024, consisting of variables such as race, ethnicity, sex, interest rate, property value, etc, which determine the loans given to applicants.

→ Problem Statement



How do variables/lack of variables such as race and gender result in inequality for factors like loan statements, interest rates, and property value?

Variable definitions

Applicant_race

Race of the applicant or borrower.

Each number represents a certain race

Values are represented by numbers 1-50, that each represent a certain race, ethnicity, age, and gender

co-applicant/applicant sex

Sex of the first co-applicant or co-borrower/ applicant

Property_value

The value of the property securing the covered loan or, in the case of an application, proposed to secure the covered loan, relied on in making the credit decision

Co-applicant_ethnicity_observed

Whether the ethnicity of the first co-applicant or co-borrower was collected on the basis of visual observation or surname

Co-applicant_race_observed

Whether the race of the first co-applicant or co-borrower was collected on the basis of visual observation or surname

total_loan_cost

The amount, in dollars, of total loan costs

What races are included in this dataset?

derived_ethnicity

- Description: Single aggregated ethnicity categorization derived from applicant/borrower and co-applicant/co-borrower ethnicity fields
- Values:
 - Hispanic or Latino
 - Not Hispanic or Latino
 - Joint
 - Ethnicity Not Available
 - Free Form Text Only

derived_race

- Description: Single aggregated race categorization derived from applicant/borrower and co-applicant/co-borrower race fields
- Values:
 - American Indian or Alaska Native
 - Asian
 - Black or African American
 - Native Hawaiian or Other Pacific Islander
 - White
 - 2 or more minority races
 - Joint
 - Free Form Text Only
 - Race Not Available

Why is grouping races and ethnicity harmful

Specific races/ethnicities that may be grouped into a broader category could be discriminated against more than groups who share that same category. This dilutes the discrimination they face, creates blind spots in data, negates ability to identify unethical practices which was why this data was created.

- When multiple racial or ethnic groups are placed in one category, it becomes harder to see which specific groups face discrimination in mortgage lending.
- For example, no middle eastern ethnicities are mentioned in this data set
- This makes it more difficult for researchers and regulators to identify unfair lending patterns.
- As a result, efforts to prevent discrimination may be less effective because the data is less clear.

Methods for Research

Poiriers Methods

Connotative and Deconstructive reading were useful for interpreting both the meaning of the data in relation to its context as well as looking deeper at what data is left out and the implications of that. In 2015 the CFPB rule expanded the data required be reported, even still we can see there are not enough for genuine fair distribution of loans

Koopmans Format Anatomies Method

Using the levels of granularity such as Macro, we can identify the conceptual assumptions made within the data set such as 'observed ethnicity' that group certain races together



Thank You for listening!