Home Mortgage Data

Discovering threats to equality in Oregon

Home Mortgage Data

The Home Mortgage Data Set

The dataset is the result of the Home Mortgage Disclosure Act, a 1975 US law which requires banks to publish relevant information on their mortgage lending activities, in an effort to identify systematic discrimination within the home mortgage system.

Our data has been filtered to include the lending activities of The Bank of America within Oregon during the fiscal year of 2022. It is arranged in 75 columns, each representing a specific type of data, and has 3533 rows, each representing an individual mortgage.

Both are examples of homes which can be mortgaged.

Problem Statement

How do the variables of the data represent the equality of the loans given?

What variables are not tracked which would indicate discriminatory practices?

We have identified that ethnicity and race related fields lack proper specificity which could lead to blind spots in the data. This, in tandem with the formatting of the dataset could cause a misrepresentation of the equality of the mortgage systems.

Important Data Points - HMDA definitions

derived_ethnicty - Single aggregated ethnicity categorization derived from applicant/borrower and co-applicant/co-borrower ethnicity fields

derived_race - Single aggregated race categorization derived from applicant/borrower and co-applicant/co-borrower race fields

applicant_ethnicity-1 - Ethnicity of the applicant or borrower

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

applicant_race-1 - Race of the applicant or borrower

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

Definitions

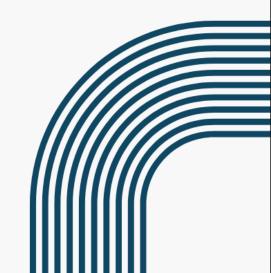|	
applicant_race	Self reported values for race
applicant_ethnicity	Self reported values for ethnicity
derived_race	A column which further groups races into broad categories for easier data analysis
derived_ethnicity	Same as above, but for ethnicity
applicant_race_observed	If the race data was derived from how they look or from common surnames from certain races
applicant_ethnicity_observed	Same as above, but for ethnicity

applicant_race allowed values

1 – American Indian or Alaska Native

2 – Asian

3 – Black or African American

4 – Native Hawaiian or Other Pacific Islander

5 – White

6 – Information not provided

7 – Not applicable

21 – Asian Indian

22 – Chinese

23 – Filipino

24 – Japanese

25 – Korean

26 – Vietnamese

27 – Other Asian

41 – Native Hawaiian

42 – Guamanian or Chamorro

43 – Samoan

44 – Other Pacific Islander

applicant_ethnicity allowed values

1 – Hispanic or Latino

11 – Mexican

12 – Puerto Rican

13 – Cuban

14 – Other Hispanic or Latino

2 – Not Hispanic or Latino

3 – Information not provided

4 – Not applicable

Derived columns lose more granularity

- derived_race and derived_ethnicity are columns created for easier analysis of data
- Groups race and ethnicity data into even more broad categories

derived_race allowed values

American Indian or Alaska Native

Asian

Black or African American

Native Hawaiian or Other Pacific Islander

White

2 or more minority races

Joint

Race Not Available

derived_ethnicity allowed values

Hispanic or Latino

Not Hispanic or Latino

Joint

Ethnicity Not Available