Documentation for the Urban Institute's Census Tract–Level Summary Files of the Home Mortgage Disclosure Act (HMDA)

The Home Mortgage Disclosure Act (HMDA) is a federal act that requires mortgage lenders to keep records of loans and lending practices and submit data to regulatory authorities. HMDA reporting allows government regulators to analyze information on mortgage loans and mortgage lending trends. HMDA data are used to understand credit accessibility, fair lending, and the mortgage market. For example, Urban researchers have used HMDA data to analyze the state of the mortgage market and understand how well GSEs serve minority borrowers. HMDA data are also an important research tool understanding housing market dynamics and neighborhood change. The data are of special interest to the National Neighborhood Indicators Partnership (NNIP) because the data includes information about the location (census tract) of the property for each individual loan application. Individual loan application data can be aggregated to the census-tract level to provide insights on the demographic and economic characteristics of people purchasing homes.

These data files are very large and cumbersome. The 2018 file has over 15 million observations and is 5.7 gigabytes. To make these data accessible to more analysts, the Urban Institute has published a public use census-tract level files with selected indicators that focus on monitoring neighborhood change. This is a brief description of the source and data provided by Urban. If we see field demand for this data, we will continue to update as additional years of HMDA are released.

Source data

HMDA's Snapshot National Loan Level Dataset contains the applications for a calendar year as of a fixed date for all HMDA reporters, as modified by the Consumer Financial Protection Bureau to protect applicant and borrower privacy. For example, the 2018 Snapshot National Loan Level Dataset reflects all the loan applications reported for the calendar year 2018 as of August 7th, 2019. For more information on the source data, see the full documentation from the Consumer Financial Protection Bureau.

Notes on definitions and creation of variables

The indicators included in this dataset are restricted to loans that meet the following criteria.

1. Loan completed the origination process (action_taken=1 - Loan originated)
2. Loan is for a home purchase (loan_purpose=1 – Home Purchase)
3. Loan is first lien (lien_status=1 – Secured by a first lien)
4. Home is single family or 1-4 unit building (derived_dwelling_category=Single Family (1-4 Units):Site-Built or Manufactured)
5. Home is owner occupied (occupancy_type=1 – Principal Residence)

Within this universe, the Urban Institute uses a three-step process to create HMDA neighborhood variables. First, we use information from individual loan applications to flag characteristics of that application based on several key measures described below. Second, we combine those characteristic flags to indicate whether a loan application meets all the criteria for that indicator. And third, loan applications that meet the criteria are counted for each census tract.

Counts are then created for the number of loans:

1. By race/ethnicity
2. By income level (relative to surrounding geographic area)
3. By race/ethnicity AND relative income

The median loan amount and median borrower income are also calculated using the same universe as described above.

Geography

This data set is published at the census-tract level. Census tract codes follow the format: 2-digit state FIPS code, 3-digit county 6-digit FIPS code, census tract FIPS code (ie: ssccctttttt).

However, approximately one percent of records on the Snapshot National Loan Level Dataset are missing census-tract information. For that reason, attempting to aggregate all of the census tracts in a county or state will not add up to the total number of records in that geography.

Rather that drop these records, we created special census tract codes for them to facilitate accurate summaries of larger geographies. For partial geographic matches we use the following format:

- If a record has a known county it is assigned: ssccc000000.
- If a record only has a known state, it is assigned: ss000000000.
- If a record has no geography details at all, it is assigned 99999999999.

Race/Ethnicity

To create the indicators of race ethnicity, we evaluate the race and ethnicity variables for both the applicant and co-applicant. A record may have up to five race/ethnicities each for the applicant and co-applicant. In the source data, the field series we use are applicant_race_1 - applicant_race_5; applicant_ethnicity_1 - applicant_ethnicity_5; co_applicant_race_1 – co_applicant_race_5; and co_applicant_ethnicity_1 – co_applicant_ethnicity_5. Note, the source data file includes two other variable series about race on the data file that we do not use: derived race (Derived_race) and race based on observation (applicant_race_observed).

The following steps are used to create the "Household race/ethnicity" indicator:

First, we determine a race/ethnicity category for the applicant:

- IF applicant_ethnicity is Hispanic then they are coded: Hispanic
- ELSE IF their race is missing or is marked as "Not applicable" or "Not provided" then they are coded as: Race not available
- ELSE IF the applicant indicated identifying with multiple races (using fields applicant_race_2 - applicant_race_5) then they are coded as: Multiple races

- ELSE they are coded as the race indicated in applicant_race_1

Second, the same logic is followed to determine a race/ethnicity category for the co-applicant (if there is a co-applicant).

Third, the values determined in steps one and two are used to determine the household race/ethnicity category for each record:

- IF there is a co-applicant AND both the applicant and co-applicant race are available AND they are not the same, then the record is coded as: Multiple races
- ELSE IF the applicant and co-applicant's race/ethnicity are the same OR there is no co-applicant, then the record is coded as the applicant's race/ethnicity

Relative Income

We provide indicators based relative income categories compared to the median area income for metropolitan areas (for counties in those areas) or counties (for non-metropolitan areas). Using the relative income level is helpful in comparing across areas of the country with different housing costs and other costs of living. Income limit data are published by the Department of Housing and Urban Development (HUD) and can be acquired from the HUD USER website. We merged the dataset with the Snapshot National Loan Level Dataset using county identifier.

We compare the borrower's income to the Area Median Income to determine whether the applicant falls into the category of:

- Very low income (0-50 percent of Area Median Income)
- Low income (50.1-80 percent of Area Median Income)
- Moderate income (80.1-120 percent of Area Median Income)
- High income (120.1 percent+ of Area Median Income)

Citation and license

Urban Institute. 2020. Census Tract–Level Home Mortgage Disclosure Act Indicators for Neighborhood Change. Accessible from https://datacatalog.urban.org/dataset/home-mortgage-disclosure-act-hmda-neighborhood-summary-files-census-tract-level