

Data Dictionary Guide | CRA Analytics Data Tables

Board of Governors of the Federal Reserve System

September 2023

Table of Contents

Overview	2
Retail Loan Table.....	2
Retail Loan Table Naming Convention.....	3
Retail Loan Lending Test Table	5
Performance Evaluation (PE) Table	6
Merged Data Table	7
AA Definition Table	8

Overview

This data dictionary guide provides an overview of the variables that make up the data in the CRA Analytics Data Tables. This document is accompanied by the [User Guide](#), which provides more information on the collection of the data and the creation of the data tables, and the [Data Dictionary](#), which provides a complete list of tables and variables with individual descriptions of each variable across all tables in the dataset.

Retail Loan Table

The Retail Loan Table contains HMDA LAR (Loan Application Register), CRA small business and small farm loan data for banks and select demographic data for the counties in which they lend, aggregated at the year, bank, and county level from 2005 through 2021. The table contains loan counts and amounts for originations and purchases, specified as inside or outside assessment area values, and grouped by borrower and tract income categories. For HMDA, breakouts are provided for the various product types. For convenience, a selection of columns from eight of the retail loan tables described below were merged to create the “retail_loan_lending_test” table which contains most of the variables needed to replicate the Retail Lending Test as described in the Notice of Proposed Rulemaking (NPR) is also provided. A “retail_loan_county_agg_all” table, described below, was added for the final release.

Retail Loan Table Descriptions

1. “retail_loan_hmda_bank_total”: This table has bank HMDA lending data at the year, bank, county level with the total loan counts and amounts for originations and purchases for Multi-Family loans and Single Family Loans. Single-Family loans are further broken out into Open-End and Closed-End. Single-Family Closed-End mortgage loans are further broken out by loan purpose (home purchase, home refinance, home improvement, other). Please note that Open-End and Closed-End breakouts are only provided for years on or after 2018 and will have a value of zero for years prior to 2018. All loan types listed above are then broken out by the income level of the tract in which the loan was made (low, moderate, middle, upper, and unknown) and the income level of the borrower (low, moderate, middle, upper, and unknown).
2. “retail_loan_hmda_bank_inside”: This table has the same loan data as the “retail_loan_hmda_bank_total” table above but limited to include only counties, or partial county areas, that are part of the bank’s assessment area.
3. “retail_loan_hmda_agg_total”: This table has the total aggregate HMDA lending at the year, bank, and county level by all HMDA lenders in the county in which the bank reported loans. This table has the same level of detail for loans as the “retail_loan_hmda_bank_total” table except that it includes all HMDA loans by all HMDA lenders.
4. “retail_loan_hmda_agg_inside”: This table has the same loan data as the “retail_loan_hmda_agg_total” table above but limited to include only counties, or partial county areas, that are part of the bank’s assessment area.
5. “retail_loan_cra_all”: This table has the bank-reported CRA small business and small farm lending aggregated at the year, bank, and county level for counts and amounts of originations and purchases with breakouts for the census tract income (low, moderate, middle, upper, unknown) and borrower (gross annual revenue less than \$1 million) income categories. Separate breakouts are also provided for the bank’s total lending in the county as well as the portion inside the assessment area only, where applicable.

6. “retail_loan_demo_all”: This table has the select demographic data at the year, bank, and county level for each county in which a bank reported either HMDA or CRA loans. Separate breakouts are also provided for the entire county as well as the portion inside the assessment area only, where applicable.
7. “retail_loan_bank_attributes”: This table has select bank level attributes for each year a bank submitted either HMDA or CRA loan data.
8. “retail_loan_county_attributes”: This table has select bank and/or county level attributes for each year a bank submitted either HMDA or CRA loan data. This table contains information such as the number of branches and amounts of deposits the bank had in the county, CRA Assessment Area Number, and the MSA/MD code of the county.
9. “retail_loan_lending_test”: This table has a subset of variables from each of the above eight tables and can be used “out-of-the-box” to replicate a bank’s performance on the retail lending test as proposed in the NPR.
10. “retail_loan_county_agg_all”: This table has county-level aggregate loans and select demographic data at the year/county level for each county where at least one bank reported one loan of any type.

Retail Loan Table Naming Convention

Retail Loan Table is organized at the *year*, *county*, and *bank* level; each row is uniquely identified by a combination of year (*ActivityYear*), county (*State_Code + County_Code*), and bank (*id_rssd*) fields. Unique identifiers such as the LEI, Respondent ID, and Agency Codes from the source datasets, HMDA and CRA, are also provided for each lender.

Aggregate Variables: Aggregate variables begin with the prefix “Agg” and include loans by all lenders at the county level. Except for the table “retail_loan_county_agg_all,” all other variables not beginning with “Agg” are at the year, county, and bank level.

Inside Variables: If a variable is designated as “Inside,” then it refers to the portion of the county in which the bank has claimed as being in their assessment area. Otherwise, it refers to data for the entire county.

Tract and Borrower Income Levels: Income levels for borrowers and tracts are classified as compared to the median family income of the county, defined as follows:

- Low: <50% of median
- Moderate: 50 – <80% of median
- Middle: 80 – 120% of median
- Upper: ≥ 120% of median

Loan Count and Loan Amount Variables: Apart from identification variables (*ActivityYear*, *hmda_lender_id*, *id_rssd*, *State_Code*, *County_Code*, and *hmda_lender_id*), the tables—aside from the demographic table “retail_loan_demo_all”—provide the count of loans, and the loan amount, filtered in the following ways, as described in the variable names. Variables are labeled using a variety of standardized “stubs,” each separated by an underscore “_”, which can be used to describe the variable.

A breakdown of these naming conventions is below:

Amt	Loan amount (in \$1000s)
Loan	Count of loans
Orig	Refers to loans originated by the evaluated bank
Purch	Refers to loans originated by another lender and purchased by the bank
MFam	Refers to mortgage loans that were for structures with 5 or more units
SFam	Refers to mortgage loans that were for structures with 1-4 units
Closed	Refers to a Closed-End mortgage loans
Open	Refers to a Open-End mortgage loans
BILow/BIMod/BIMid/BIUpp/BIUnk	Specifies loan data for borrowers designated as low-, moderate-, middle-, upper-, or unknown-income levels
TILow/TIMod/TIMid/TIUpp/TIUnk	Specifies loan data in census tracts designated as low-, moderate-, middle-, upper-, or unknown-income levels
GAR_1_to_250k	Specific to small business and small farm loans; specifies loans where the loan recipient has a gross annual revenue between 1 and \$250K
GAR_250k_to_1M	Specific to small business and small farm loans; specifies loans where the loan recipient has a gross annual revenue between \$250K and \$1M
GAR_less_1m	Specific to small business and small farm loans; specifies loans where the loan recipient has a gross annual revenue less than \$1M
HI	Mortgage loan that was designated as a home improvement loan
HP	Mortgage loan that was designated as a home purchase loan
HR	Mortgage loan that was designated as a home refinance loan
Inside	Data inside the specific assessment area
Rpt_by_aff	Data reported by an affiliate or subsidiary of the evaluated bank
SB	Small business loan data
SF	Small farm loan data
Establishments	Count of establishments designated by Dun & Bradstreet as “Small Business” or “Small Farm” in the relevant area

Demographic Variables: The demographic table “retail_loan_demo_all” contains the following information about the county and assessment areas of the evaluated banks that do not appear in the other tables, as well as some of the above data filters. Except for “Establishments” data, all demographic data are sourced from census data products available on the FFIEC website.

A breakdown of these naming conventions is below:

Family_Count	Total family count in the relevant area
Low/Moderate/Middle/ Upper_Income_Family_Count	Count of families of low-, moderate-, middle-, or upper-income designations within the relevant area. Family counts by income group are estimated using the counts of families per income buckets provided in census data
HouseHold_Count	Count of households in the relevant area
Population	Total population in the relevant area
Minority_Population	Minority population in the relevant area
Total_Housing_Units	Count of all housing units in the relevant area regardless of occupancy status

Occupied_Housing_Units	Count of housing units in the relevant area that are occupied
Vacant_Units	Count of housing units in the relevant area that are not occupied
Owner_Occupied_Units	Count of housing units in the relevant area occupied by the unit's owner

Bank and County Attributes Variables: Variables in the attributes tables “retail_loan_bank_attributes” and “retail_loan_county_attributes,” including possible values for the variables, are described below.

The Bank Attributes table is organized at the *year* and *bank* level; each row is uniquely identified by a combination of year (*ActivityYear*) and bank (*id_rssd*) fields. The County Attributes table is organized at the *year*, *bank*, and *county* (*State_Code* + *County_Code*) level, with each row identifying a specific bank as it exists in a specific county.

The Bank Attributes table contains information about the bank as reported through HMDA and CRA data panels, such as the ID and name of the bank and any parent banks in those contexts, the overseeing agency, whether the bank operates as a credit card bank, the type of bank, and the assets both for the year and for the prior year. Much of these bank attributes data come from the respective panel datasets available on the FFIEC HMDA and/or CRA websites.

The County Attributes table contains information about the counties in which banks have been assessed, including the metropolitan statistical area (MSA) or metropolitan division (MD), state, and county codes for identification of the county. This table also includes bank-specific data for the county, such as whether or not the county is in the bank’s assessment area, if it is a partial county, the presence of a physical bank branch in that county, as well as the number of branches and amounts of deposits at branches within that county.

The Bank and County Attributes tables contain additional information on the banks and counties that can be utilized and cross-referenced with the other retail loan data tables, all of which contain the *ActivityYear*, *id_rssd*, and the *State_Code* and *County_Code*.

Individual variables are described in the [Data Dictionary](#).

Retail Loan Lending Test Table

For ease of use, the table “retail_loan_lending_test” has been created to supplement the Retail Loan Table. This table has most of the variables from the other Retail Loan sub tables and can be used “out-of-the-box” as a base table for reproducing a bank’s performance on the retail lending test as proposed in the NPR.

Variables in the Retail Loan Lending Test table represent a subset of data from the Retail Loan tables, as well as both Bank and County Attributes tables, and provide the information necessary to start simulating bank’s performance based on the Retail Lending Test. It is designed for ease of use for most applications that do not require the full breakdown of data provided in the Retail Loan sub tables.

The table “retail_loan_lending_test” is organized at the *year*, *county*, and *bank* level and contains both aggregate data and information about the counties where banks have a presence, as well as the loan counts and amounts of loans originated by the bank. Additionally, a sample of fields from the demographic table are included regarding the families, establishments, and housing units within the county.

Individual variables are described in the [Data Dictionary](#) and their components in the [Retail Loan Table](#) section of this guide.

Performance Evaluation (PE) Table

The PE data encompass the data obtained by examiners from each of the supervising agencies— Office of the Comptroller of the Currency (OCC), Federal Deposit Insurance Corporation (FDIC), and the Federal Reserve Board—for the purposes of evaluating a bank’s performance under the Community Reinvestment Act (CRA), and include data for mortgage, consumer, small farm, small business, and community development loans, as applicable for each bank, product-specific evaluation periods, and the ratings/conclusions received, combined into a single dataset. Descriptions of variables in the “performance_evaluation” table, including possible values for the variables, can be found below.

The PE table is organized at the *exam*, *bank*, and *rating/conclusion* level; each row is uniquely identified by a combination of exam date (*exam_start_date*), bank (*id_rssd*), and assessment area fields. The *bank_identifier* also provides an identifier for the respondent from the supervising agency (if the bank is an OCC bank it is the charter number, and if it is an FDIC bank then it is the certification number).

Generally, for each exam there will be an overall, state, and multistate MSA rating where applicable, as well as conclusions for full and limited scope, where applicable, and assessment areas. For each loan product examined, the evaluation period is included as well as the size of the sample of loans used to evaluate a bank’s performance, where applicable.

Within this framework, if a bank was not a CRA or HMDA reporter for the year of the exam, the dataset collected the applicable count of loan originations of each loan product (mortgage, small business/farm, consumer) reviewed for the bank within each of the assessment areas and reviewed as a part of the examination. If the bank was already a HMDA and/or CRA reporter, the dataset did not record the loans for the associated products. For those loan products where data were collected, the dataset pulled loan data from the Geographic and Borrower Income Distribution tables. For data collected from the first vintage of exams (2005–17), counts of loans to low- and moderate-income tracts and borrowers were combined into one LMI (Low-and-Moderate-Income) field. For the second and third vintage of exams (2018–20) all income breakouts were collected from the exams, where applicable.

Income levels for borrowers and tracts are classified as compared to the median family income of either the MSA or the non-MSA portion of the state of the assessment area, defined as follows:

- Low: <50% of median
- Moderate: 50 – <80% of median
- Middle: 80 – 120% of median
- Upper: ≥ 120% of median
- LMI: < 80% of median

Additionally, the borrower income for small business and small farm loans is recorded as either having “less than” or “greater than” gross annual revenue of \$1 million. Community Development (CD) loans and investments are included in the table as well but track the count and amount of the loans and the dollar amount of investments in an assessment area. If the row pertains to an overall or state rating, there is information regarding the total amount of CD loans and investments made outside of an assessment area but still included in consideration of the overall/state rating.

A breakdown of loan counts and percentages naming conventions is below:

Pct_PE	Percentage of loans of this type that apply to this income level
PE	Count of loans of this type for this income level
TI_LMI	Specifies loan counts for loans made in low- and moderate-income census tracts
BI_LMI	Specifies loan counts made to low- and moderate-income borrowers
BI _{Low} /BI _{Mod} /BI _{Mid} /BI _{Upp} /BI _{Unk}	Specifies loan data for borrowers designated as low-, moderate-, middle-, upper-, or unknown-income levels
TI _{Low} /TI _{Mod} /TI _{Mid} /TI _{Upp} /TI _{Unk}	Specifies loan data in census tracts designated as low-, moderate-, middle-, upper-, or unknown-income levels
GAR_les/greater/unknown_1m	Specific to small business and small farm loans; specifies loans where the recipient of the loan had a gross annual revenue of less than or greater than \$1M, or the revenue of the business was unknown
ML	Mortgage loan data
CL	Consumer loan data
SB	Small business loan data
SF	Small farm loan data

Individual variables are described in the [Data Dictionary](#).

Merged Data Table

The Merged Data Table, “merged_data,” is a combination of the Retail Loan and Performance Evaluation tables and is organized at the *bank*, *exam*, and *Assessment Area* level. Each row represents one Performance Evaluation (PE) for a given bank in a given assessment area, and it contains both information about the banks activities as well as demographic information about the assessment area.

Not all data are available for all banks or assessment areas, and these omissions are represented in the table. Generally, if the bank reported HMDA or CRA data at any point during the evaluation period, then that data will be in the Merged Data Table. If the bank did not report HMDA or CRA data during the exam year (*Lender_in_CRA* = “N” or *Lender_in_HMDA* = “N”), then the corresponding loan data were collected from the PE and can be found in the Merged Data Table from the fields with a “PE” stub.

Unlike the Retail Loan Table, there are no “Inside” stubs in the Merged Data Table. Wherever possible, the loan values in the Merged Data Table use “Inside” assessment area data. For instance, if a bank is a CRA reporter (*Lender_in_CRA* = “Y”), then all of the loan values from the Retail Loan Table that will be in the Merged Data Table will be from the “Inside” Retail Loan Table. If the bank is not a CRA reporter (*Lender_in_CRA* = “N”) but is a HMDA reporter (*Lender_in_HMDA* = “Y”), then the Retail Loan Table values in the Merged Data Table will be from the “Total” Retail Loan Tables.

The *Lender_in_HMDA/CRA* variables are determined based on the exam year. It is possible that a bank was not a reporter during the exam year but did report loan data at some point during the exam’s evaluation period. If the bank did report some loan data at some point during the evaluation period, then the corresponding Retail Loan Table values will be populated with whatever data were reported during the evaluation period. If the bank did not report any HMDA or CRA loan data at all during the evaluation period, then the corresponding columns of Retail Loan Table variables will be “NA.”

Additionally, if the PE loan data came from a statistical sample of loans in the Performance Evaluation table (*statistical_sample*), the “PE” counts/amounts will be based on the sample data.

Individual variables are described in the [Data Dictionary](#) and their components in the [Performance Evaluation Table](#) and [Retail Loan Table](#) sections of this guide.

AA Definition Table

The Assessment Area Definition Table, “assessment_area_definition,” provides general geographic information on the assessment areas for each bank for each year of data. The table provides state and county codes for each county included in an assessment area. This information is drawn from the CRA examinations for banks which were examined, and from performance evaluation (PE) data for all banks. The information in this table is at a county level, so assessment areas containing only certain tracts from a county are approximated.

For each *ActivityYear* and bank *id_rssd*, the *State_Code* and *County_Code* for each assessed county is provided in a unique row, along with the *assessment_area* identifier to which they belong. For organizations submitting CRA examination data, the *exam_start_date* and *exam_year* provide the date and year of the CRA examination.

Individual variables are described in the [Data Dictionary](#).