

User Guide | CRA Analytics Data Tables

Board of Governors of the Federal Reserve System

May 2022

Table of Contents

Purpose	2
Main Data Sources	2
CRA Small Business and Small Farm Data	2
HMDA Data	2
Performance Evaluation Data	3
Data Tables	3
1. Retail Loan Table	3
2. Performance Evaluation (PE) Table	5
3. Merged Data Table	8
4. Assessment Area Definition Table	8
Quality Control	9
Key Concepts and Variables of Note	10
Low-and-Moderate-Income (LMI) Tracts and Borrowers	10
Summary of Deposits Proxy (SoD_Proxy_AA_Flag)	10
County Assessment Area Flag (County_AA_Flag)	10
Partial Indicator (Partial_Ind)	10
Additional Data Sources	11
National Information Center (NIC) Data	11
Summary of Deposits (SoD)	11
Census	11
Vendor Data	11
Use Cases & Analytic Examples	11
NPR Analytics Table	11
Retail Loan Table Examples	12
PE Table Example	15
Merged Data Table Example	16

Purpose

The primary purpose of the data tables is to provide Community Reinvestment Act (CRA) stakeholders a better understanding of the historical relationship between bank lending activity and the conclusions and ratings that regulators assigned on CRA performance evaluations (PEs).

For the purposes of this document, (1) “ratings” refer to assessment area (AA) performance conclusions and ratings; (2) AAs refer to areas that received a conclusion or rating as a part of the exam and are captured through different AA types (see “assessment_area_type” in [Data Dictionary](#)); and (3) “CRA data” refers to data reported under the CRA.

Main Data Sources

In order to gain a comprehensive view of bank lending activities and connect them to the assigned ratings, Home Mortgage Disclosure Act (HMDA) and CRA loan data were combined with data collected from banks’ historical CRA PEs.

The first release of the CRA Analytics Data Tables was in March 2020 and contained data from 2005–17. The data have since been updated with a second release including data from 2018 and 2019. This guide will discuss the original data release (vintage 1) and the subsequent vintages (vintage 2) of the data tables.

Please note that there may be discrepancies between the CRA Analytics Data Tables and publicly available data on the Federal Financial Institutions Examination Council (FFIEC) website because of resubmissions. The bank-reported HMDA and CRA data included in these tables were frozen on 12/31/2020.

CRA Small Business and Small Farm Data

The census tract level CRA small business and small farm data are aggregated to the year, bank, and county level. Bank-provided AA delineations are used to aggregate the data to the county level. The tables include the total loan counts and dollar amounts for originations and purchases of small business and small farm lending by the bank in the county with breakouts for tract income levels (low, moderate, middle, upper, and unknown) and borrower income (gross annual revenue less than \$1 million).

In some cases, banks claim a portion of a county, rather than an entire county, as being in an AA. In those cases, the same data are provided as above for the portion of the county that was inside the bank’s AA. In addition to the bank’s lending for a given year, bank, and county combination, the same loan data breakouts are provided as above for all lending done by all banks in the county for the same year. Separate aggregates are provided for the total county and the inside AA portion only where applicable.

The data tables use the post-publication version of the CRA data, which include additional updates to the yearly publications currently available on the FFIEC website.

HMDA Data

The HMDA loan application register (LAR) data are aggregated to the year, bank, and county level. Where applicable, these lending categories are further grouped into inside or outside of the banks’ AAs. This grouping uses the AA delineation files submitted with the CRA data.

The data tables use the post-publication version of the HMDA data, which includes additional updates to the yearly publications currently available on the FFIEC website. Row level (year, bank, county) loan data only include depository institutions.

The CRA Analytics Data Tables utilize a subset of loans from the HMDA LARs. The way in which HMDA is reported has changed over the years, and the filters used to aggregate the HMDA data in the Retail Loan Table over time can be summarized as follows:

- For HMDA LARs from 2005 through 2017, the following filters were used:
 - Action Types of either 1 or 6
 - Loan Purposes of 1, 2, or 3
 - Lien Status not equal to 3
- For HMDA LARs from 2018 through 2019, the following filters were used:
 - Action Types of either 1 or 6
 - Loan Purposes not equal to 4

The ability to differentiate between open- and closed-end mortgage products was introduced to HMDA in 2018. At this time, open-end and closed-end loans are combined in the loan counts and amounts in the Retail Loan Table.

Performance Evaluation Data

To help fill in the data gaps identified by CRA reform efforts, staff collected a wide range of data elements from a sample of CRA PEs. For each PE, staff collected information related to the product evaluation periods and the ratings and conclusions the bank received in its assessment areas. For small and intermediate small banks, additional information related to assessment area lending and the geographic boundaries of those assessment areas was collected. These data were extracted from the PEs manually by staff over time to supplement the data needs of CRA reform efforts.

Data Tables

The first vintage contained four primary tables: Retail Loan Table, Performance Evaluation (PE) Table, Merged Data Table, and AA Definition Table. In the second vintage, the Retail Loan Table was broken out into eight separate tables, primarily because of the addition of the breakouts for the individual census tract and borrower income categories and mortgage loan products.

Definitions of variables across the primary tables are included in the table-specific [Data Dictionary](#).

1. Retail Loan Table

The Retail Loan Table contains HMDA LAR, 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 2019. 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, we also provide breakouts for the various product types. Because of the expansive number of columns, the Retail Loan

Table is broken out into eight separate tables. For ease of use, a ninth table containing only the variables needed to replicate the Retail Lending Test as described in the Notice of Proposed Rulemaking (NPR) is also provided.

Retail Loan Table Descriptions

1. “retail_loan_hmda_bank_total”: This table has the bank HMDA lending at the year, bank, county level with the total loan counts and amounts for originations and purchases as well as breakouts for the income level of the census tract in which the loan was made (low, moderate, middle, upper, and unknown), income level of the borrower (low, moderate, middle, upper, and unknown), loan purpose (home purchase, home refinance, home improvement), and each of the possible combinations of the census tract income level, borrower income, and product type.
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.

A bank must be a HMDA reporter, a CRA reporter, or both to be included in the Retail Loan Table (figure 1).

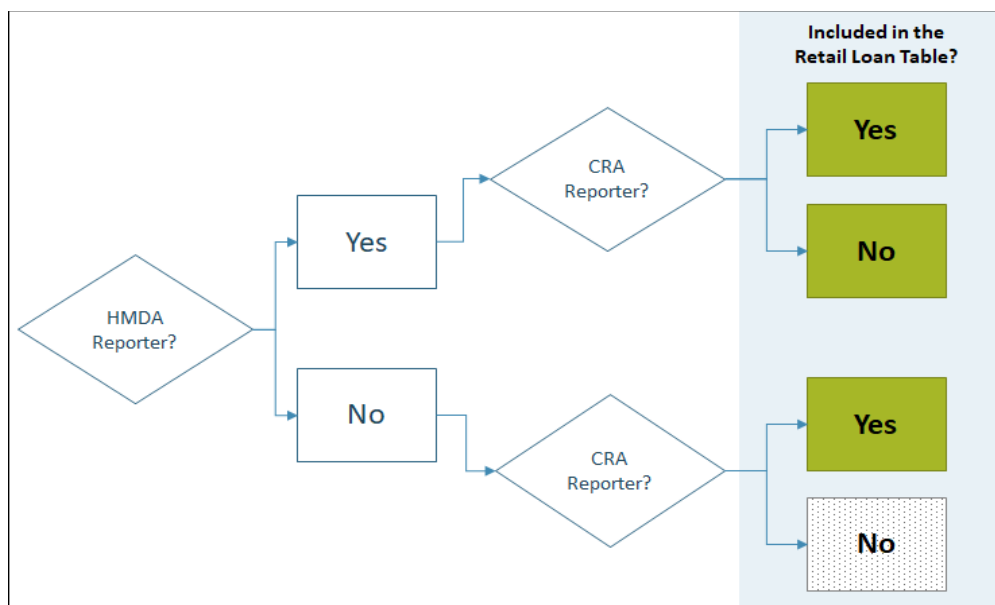


Figure 1. Reporters included in the Retail Loan Table - Green, solid boxes indicate inclusion in the Retail Loan Table, i.e., the bank must be either a HMDA reporter or a CRA reporter to be included in the Retail Loan Table. If neither HMDA nor CRA reporter (patterned white box), then the bank’s data are not included in the Retail Loan Table.

2. Performance Evaluation (PE) Table

The PE data has been hand-collected over time in separate vintages. The first vintage of PE data was collected from a sample of over 6,000 CRA PEs over the 2005–17 period and consisted of a diverse sample of bank sizes and business models. The second vintage consisted of an additional 600 PEs sampled from the 2018–19 period. The PEs were selected using a semi-random sampling strategy (see [Sampling Procedure](#)).

PE Vintage 1

For each PE reviewed in the first vintage (2005–17), the following data were extracted, where available:

1. Evaluation periods for each product the examiner included as a part of the PE and a flag to denote when a statistical sample of loans was used by the examiner.
2. Lending, community development, investment, and overall test ratings/conclusions for institution, state, multistate metropolitan statistical area (MSA), MSAs, and non-MSA AAs.
 - For each rating/conclusion recorded, a set of “seed” geography codes were collected to help map the rating/conclusion to the appropriate data in the retail loan tables.
3. Community development loan counts and amounts as well as investment amounts made in each assessment area.
4. Information on loan production offices (LPOs), where available.

For banks that were not required to report CRA and/or HMDA data at the time of the exam, the following data fields were collected:

1. County-level geographical boundaries of their AAs with a flag to denote when the county was a partial county.

2. Where applicable, the counts of mortgage, small business, and small farm loans made in each assessment area were collected.
 - Loan count data were collected from the Geographic and/or Borrower Distribution tables. In the first vintage, low- and moderate-income loan counts were combined into one “LMI” value.

PE Vintage 2

There were several changes made to the data collection efforts for the second vintage of PE data (2018–19). In addition to the data collected in the first vintage, the following data were also extracted:

1. Total inside and outside assessment area counts and amounts for each loan product reviewed as a part of the performance evaluation.
2. Assessment area loan count data were collected from the Geographic and/or Borrower Distribution tables using all of the available income breakouts (low, moderate, middle, upper, unknown, <=\$1M GAR, >\$1M GAR).
3. Community development loan and investments made outside of an assessment area were recorded separately from loans assigned to an individual assessment area. The integration of these Retail Loan and PE data constitutes the Merged Data Table.

Sampling Procedure

For each vintage, the PEs were selected using a stratified random sample method. The first vintage of PEs was drawn from evaluations conducted from 2005 to 2017, and the second vintage was from 2018 to 2019. Evaluations cover bank performance over a period of years preceding the evaluation start date. The sample therefore, covers the market during a boom period, a severe recession, and then an extended recovery.

Sampling was stratified by the existing three bank asset threshold categories: small, intermediate small, and large; the agency which performs the performance evaluation (Office of the Comptroller of the Currency (OCC), Federal Deposit Insurance Corporation (FDIC), or the Federal Reserve); and whether or not the bank received a rating of “Needs to Improve” or “Substantial Non-Compliance.” Historically, ratings below a “Satisfactory” have been relatively uncommon. Staff, therefore, oversampled banks with “Needs to Improve” or “Substantial Non-Compliance” ratings to obtain a sufficient sample size of these performances. PEs were grouped by year, regulator, bank size, and the performance category described above. PEs were randomly selected from each group for data collection.

The first vintage of PEs contained approximately 22,000 PEs with an evaluation start date from 2005 through 2017 available on the websites of Federal Reserve Board (the Board), the FDIC, and the OCC. Of these, staff selected approximately 6,300 PEs from which to collect data. In the second vintage of PEs, staff collected data from 600 of the approximately 2,900 PEs available from the 2018–19 timeframe.

Additional Notes on Performance Evaluations

- *The scope of information on performance ratings varies across agencies.*

Board staff had access to ratings for each test at the state and multistate MSA level for Federal Reserve System (FRS) banks, to ratings for each test at the individual institution level for FDIC banks, and to overall institution level ratings for OCC banks. To address this variation in availability of digitized information, the following evaluations were oversampled:

- FRS state member bank evaluations that received a “Needs to Improve” or worse on a lending test rating (or overall rating, for small banks) for any state or multistate MSA;

- FDIC bank evaluations where the bank level lending test rating (or overall rating, for small banks) was "Needs to Improve" or worse; and
- OCC bank evaluations where the bank level overall rating was "Needs to Improve" or worse
- *Examiners assign bank ratings on the various CRA tests at the state, multistate MSA, and institution level, but only draw conclusions at the AA level.*
 - Currently, evaluation conclusions do not use the ratings terminology. As a result, staff often had to judge what language in a conclusion corresponded to a particular rating. For example, if an AA was within a state that received an "Outstanding" rating, and the conclusion language for the AA stated that the bank's performance was consistent with its state-level rating, the bank would be recorded in the database as having received an "Outstanding" on the test in question for that AA. Alternatively, if the state-level rating was "Satisfactory," but the AA conclusion was "poor" or "below state-level performance," then the bank would be recorded as "Needs to Improve" for that AA, i.e., below "Satisfactory."
- *Evaluation structure differs by regulatory agency.*
 - FRS and FDIC evaluations generally stated conclusions about banks' retail and community development performance, as applicable, at the AA levels. OCC evaluations, instead, recorded conclusions for components of those tests (i.e., geographic and borrower income distributions of various retail loan products) at the AA levels, with ratings on the full test recorded at the state or multistate MSA level. Therefore, for OCC banks, ratings on each test for an AA are often only available when that AA was the only one the bank had in a particular state or multistate MSA. However, staff did make efforts to map state ratings to multiple assessment area conclusions when the conclusions were clearly described as consistent with state ratings.

Retail Loan and PE Tables are used to produce the Merged Data Table by matching the bank, geography, and product evaluation period (figure 2).

Merging Retail Loan and PE Tables to Produce Merged Data Table

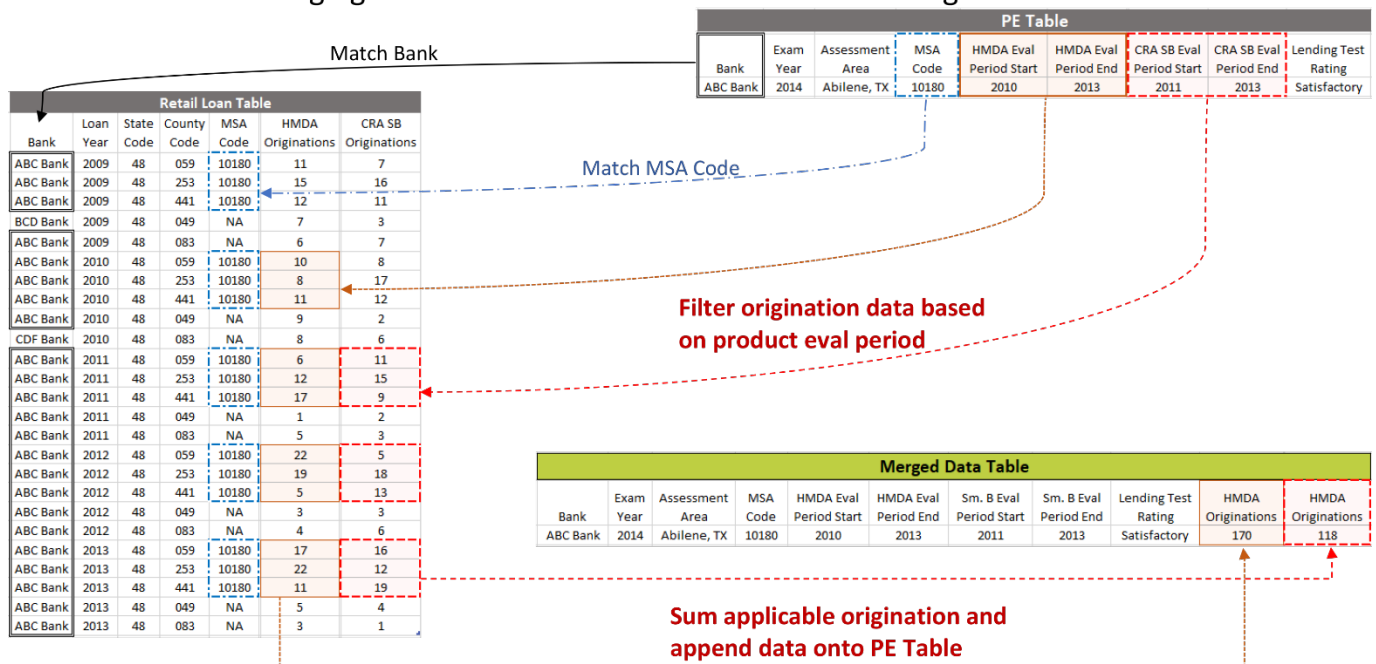


Figure 2. Merging the Retail Loan Table and PE Table to produce the Merged Data Table - *Applicable originations from the Retail Loan Table are summed and matched with the correct bank, MSA code, and evaluation period from the PE Table. The aggregated originations are then merged onto the original PE Table data to create the Merged Data Table.*

3. Merged Data Table

One of the primary purposes of collecting the loan and PE data was to link each bank's lending activity to individual AAs and their corresponding ratings for a given exam period. The product evaluation period information was used to independently aggregate the HMDA and CRA lending data within each of the banks' AAs for each of the filing years within the product-specific evaluation periods. For banks that did not report HMDA and/or CRA data during an evaluation period, the lending activity AA information from the PEs was used to fill in the AAs and lending activity data fields, where applicable. This process led to the creation of the Merged Data Table, which has one row per bank, exam, and AA, and summarizes all of the bank's lending activity in addition to the deposits, demographics, and vendor data (see [Additional Data Sources](#)).

For each of the individual AAs in the PE table, where all of the necessary data were available, bank lending activity, deposits, aggregate lending, and select demographic data from the AA were aggregated over the product-specific evaluation periods. Generally, only MSA, multistate MSA, and non-MSA AAs were mapped over from the PE Table to the Merged Data Table. However, if the bank only had one assessment area in a given state or only a single overall assessment area in the PE Table, that overall or state rating would be mapped to the assessment area. The Merged Data Table contains AAs from exams with evaluation periods starting no earlier than 2005 and evaluation periods ending no later than 2019. The updated Merged Data Table also includes the more granular low- and moderate-geographic and borrower income breakouts described in the Retail Loan and PE table sections.

There are a substantial number of AAs in the PE table where a conclusion was not extracted from the PE. These AAs were previously dropped in the first vintage of the Merged Data Table. However, if all of the other information was available to accurately map the AA and aggregate the lending activity, then these AAs were included in this second vintage of the Merged Data Table even if they did not have a lending test rating or conclusion.

4. Assessment Area Definition Table

This table combines AA geography data collected from PEs with AA information from annually reported CRA data. Joining these AA data together results in (approximate) county level geographic definitions for each AA included in the Merged Data Table.

- Banks that report their CRA lending data annually provide a list of all census tracts where they originated or purchased loans and flag them if they were in their AA for the filing year. However, information to accurately assign these tracts to specific AAs and their respective ratings for a given exam were unavailable. To solve this problem, MSA, state, and county geography codes for each AA in each exam were collected, where available, and used to associate the reported lending with a specific AA. It should be noted that neither small banks nor ISBs are required to report their CRA data, but both small banks and ISBs have the option to report CRA data.
- "MSA AAs" refer to metropolitan statistical areas (MSA), metropolitan division (MD), or combined statistical area (CSA). Whenever these geography codes were used to define the geography of an AA in the exam, they were collected and mapped to all counties in the annually reported CRA data within that same MSA code.

- For “non-MSA AAs,” only a single “seed” county was recorded during the PE data collection for each exam. These seed county codes were a starting point to build the whole, approximate non-MSA AA. First, the seed county was selected and matched to the same county inside the AA of the corresponding bank in the Retail Loan Table. Then, the list of non-MSA counties inside the AA was examined to see if any of the remaining counties were contiguous to the seed county, using the census contiguous counties files. If a candidate county was (1) contiguous to the seed county, (2) inside the bank’s AA, (3) in the same state as the seed county, and (4) not a seed county itself, then the county was assigned to the seed county’s AA. Iterations of this process continued until no new counties were added to the AA.
- Small banks do not report CRA data; therefore, data defining their AA geographies were only available through PEs. For these banks, the entire set of counties that were at least partially included in each AA was collected. Census-tract-level information for partial counties was not collected; instead, a flag was created to note when only a portion of a county was included in the AA.

Quality Control

Various quality control (QC) strategies were applied to assess the data tables. The sections below discuss the approaches taken for each of the primary tables.

Retail Loan Table

As the content for the Retail Loan Table comes from validated sources like CRA and HMDA, the QC process focused on validating the application of the business requirements and programming procedures used to create the Retail Loan Table. After the knowledge transfer of the materials, an internal IT team independently and successfully replicated the Retail Loan Table components and the final table itself.

PE Table

Two approaches were used to provide quality control and validation of the PE data. First, the data collection included continuous monitoring. The data were processed and integrated into the internal database, which included procedures to detect and correct errors in the data. Second, a separate QC process for each vintage reevaluated and re-extracted data from a sample of PEs including 200 from the first vintage and 50 from the second vintage. The re-extracted data were compared to the previously collected data for each major type of data point collected. The review found that data from the first vintage were misreported at a rate of 1.0 percent and missing in error 3.0 percent of the time. Results from the second vintage showed an overall rate of misreporting of 1.4 percent and missing in error 0.1 percent for the data fields reviewed.

A full breakdown of the percentages of misreported and underreported data for each of the PE Table data fields can be found in the “pe_accuracy_check_npr” table included in this release. While the data collection process included robust QC checks, the data may nevertheless be subject to error. To submit a question or feedback, please fill out the [feedback form](#). To ensure that your question is properly routed, please select the Community Reinvestment Act as the “staff group” and select no other options above the field labeled “Type your message.”

Merged Data Table

The Merged Data Table was validated using a combination of code review, user testing, and spot checks of random samples. A code review to map the ratings, evaluation periods, and the AA code data from the PE tables to the matching year, bank, and county records in the Retail Loan Table was conducted to ensure that the code aggregated the data in accordance with the business requirements of the project. In addition to the code review, analysts randomly sampled rows in the Merged Data Table (exam AAs) and manually validated the aggregated

loan data and compared it to the sum of the matching loan data from the Retail Loan Table (for the set of counties which constituted the AA over the specific evaluation period from the PE data).

Key Concepts and Variables of Note

This section provides additional information on key concepts across the data tables as well as notable variables within the data tables.

Low-and-Moderate-Income (LMI) Tracts and Borrowers

Low-and-Moderate-Income (LMI) borrowers and geographies are key concepts in the CRA Analytics Data Tables. Definitions of LMI *geographies* are consistent across HMDA and CRA small business and small farm lending, where census tracts with a median family income (MFI) of less than 50 percent of the area MFI are considered “low” and those with MFI of less than 80 percent are considered “moderate.” However, each primary dataset provides different definitions for loans to LMI borrowers. For HMDA purposes, “low” income *borrowers* refer to borrowers with incomes less than 50 percent of the area MFI and “moderate” income borrowers are those with incomes less than 80 percent of the area MFI. For CRA small business and small farm lending, an LMI borrower refers to a small business with gross annual revenue of less than \$1 million.

Summary of Deposits Proxy (SoD_Proxy_AA_Flag)

The Summary of Deposits (SoD) Proxy variable is a proxy for defining AAs in the loan data for banks that did not report CRA lending data for a given year. If a bank operated a branch within a county for a given year, then that county was assumed to be included in the bank’s AA for that year. The FDIC’s SoD data were used to create this flag. They were aggregated into a table with one row per year, bank, and county. Each row contained the number of branches within the county and a binary flag, indicating that the bank had at least one branch in this county during the time of reporting. These data were merged to the loan data by year, bank, and geography to create the SoD table. For the variable “SoD_Proxy_AA_Flag” in the “retail_loan_county_attributes,” records from the loan table that matched a row in the SoD table have a value of 1 and otherwise have a value of 0. For cases where the lender is not in the annual SoD data, the value is NA.

County Assessment Area Flag (County_AA_Flag)

County AA Flag in the “retail_loan_county_attributes” is used to indicate that at least one census tract within the county is inside the bank’s AA.

- If this flag is positive (value = 1) and the Partial Indicator variable is negative (value = “N”), then the entire county is within the bank’s AA.
- If the flag is positive (value = 1), and the Partial Indicator variable is positive (value = “Y”), then only a portion of the county is inside the bank’s AA.
- If the flag is negative (value = 0), then the county is completely outside the bank’s AA.
- If the value is missing or “NA” then the bank does not report annual CRA data, and its assessment area status is unknown.

Partial Indicator (Partial_Ind)

The Partial Indicator variable in the “retail_loan_county_attributes” table is positive (value = “Y”) when a bank claims part of the county in its AA. Banks that report CRA lending data also provide a list of all tracts included

within their AAs. However, banks are not required to add entire counties to their AAs. This variable comes from annually reported bank CRA data.

- If the flag is positive (value = “Y”), the bank includes only a portion of the county in its AA.
- If the flag is negative (value = “N”), the bank includes the whole county in its AA.

Additional Data Sources

National Information Center (NIC) Data

[The National Information Center \(NIC\)](#) provides comprehensive information on banks and other institutions for which the FRS has a supervisory, regulatory, or research interest, including both domestic and foreign banking organizations operating in the United States.

Summary of Deposits (SoD)

The [FDIC SoD](#) is an annually reported dataset, which contains the location of each branch office for all FDIC-insured banks for the reporting year. The SoD was used to aggregate the total sum of deposits within a given county for each bank and HMDA/CRA filing year and to count the number of branches (if any) that a bank has in a county.

Census

The FFIEC census data, available on the [FFIEC Online Census Data System](#), identify LMI census tracts. They also provide data for several county-level demographic comparators, such as the count of LMI families and owner-occupied housing units within a given county.

Vendor Data

Dun & Bradstreet (D&B) data identify the number of small businesses within a county for a given year. The D&B 2010 census file data comprise all D&B records in the United States, Puerto Rico, and the Virgin Islands. These data include fields for two-digit North American Industry Classification System (NAICS) codes; annual sales volume; year started; control year ranges; and census state, county, and tract codes. The 2018 data included updates to correct for changes in census and American Community Survey (ACS) data.

Use Cases & Analytic Examples

NPR Analytics Table

To assess the performance of large banks under the proposed Retail Lending Test, a Retail Loan NPR Analytics Table was created from the base Retail Loan Table. This analytics table grouped each Large Bank’s lending into approximate assessment areas over a hypothetical evaluation period of 2017-2019. These assessment areas are referred to as approximate because they rely on the Assessment Area Number (AAN) field reported by banks in their annual CRA reporting. These AANs may differ from actual assessment areas defined during a bank’s regular CRA PE. To replicate this table used in the NPR analyses, the data user should do the following:

1. **Utilize the retail loan table data from 2017 to 2019.**
2. **Remove wholesale/limited purpose/strategic plan banks from the sample.** These banks were identified using the CRA exam results data (from the FFIEC website). Any bank examined during the evaluation

period as wholesale/limited purpose/strategic plan is removed from the sample (regardless of exam year), as are any banks whose first exam after the evaluation period is wholesale/limited purpose/strategic plan.

3. **Remove bank-county observations that are part of non-state territories (Puerto Rico, Guam, etc.).** These observations are often missing the data needed to construct the various community benchmarks, so they should be removed.
4. **Create a consistent assessment area (AA) identification (ID) variable that is constant across time.** The AA ID variable used in the NPR analyses focused on large bank performance on the retail lending test is a concatenation of the Assessment_Area_Number (AAN) and the MSA_Code (or state code in the case of non-MSA areas), both of which can be found in the “retail_loan_county_attributes” table. However, the AA numbers reported by banks in the CRA data and included in the retail loan tables are not consistent from year to year. For example, a bank’s AAN “1” in 2017 may not necessarily correspond to the same geography as the same bank’s AAN “1” in 2019. To obtain a consistent AA identifier across all years of the sample, the data user may perform the following steps:
 - a. Create a crosswalk that links a bank’s 2017 and 2018 AA IDs to the correct 2019 AA ID.
 - b. Reshape the data so that each bank-county has an AA ID 2017, AA ID 2018, etc. in one row of data. This means that counties that exit the data before 2019 would have a missing value for AA ID 2019.
 - c. Group by the AA ID 2017 and take the mode of AA ID 2019 within that group.
 - d. Backfill the missing AA ID 2019 values with the mode value of the group.
5. **Restrict the sample to bank AAs that exist in 2019.** This means that any bank-AA that dropped out of the sample between 2017 and 2019 should be removed. Despite this, *individual counties* that drop out of an individual AA are still included as long as the MSA in which that AA is located still has data in 2019.
6. **Aggregate (sum) the county-level data within each assessment area and then across years.** In the final dataset, each bank AA should only have one row that represents data from all three years.

The final sample includes approximately 6,880 unique bank AAs and 606 banks. The data user’s sample may differ somewhat, depending on the exclusion rules applied to wholesale/limited purpose banks.

Retail Loan Table Examples

These scripts and tables show the geographic and borrower income distributions for HMDA and CRA loans at the county and MSA level for 2019.

HMDA Lending in R Code

Use the “retail_loan_lending_test” table to calculate the borrower income distribution ratios for all counties in the country for 2019.

```
#R Code, Loans_2019 is the retail_loan_lending_test table filtered for year 2019
hmda_county_2019 <- retail_loan_lending_test %>%
  filter(ActivityYear == 2019) %>%
  filter(Lender_in_HMDA == "Y") %>%
  filter(Lender_in_CRA == "Y") %>%
  group_by(ActivityYear, State_Code, County_Code) %>%
```

```

summarise(across(c(Loan_Orig, Loan_Orig_BILow, Loan_Orig_BIMod, Loan_Orig_TILow, Loan_Orig_TIMod), sum, na.rm=T)) %>%
mutate(hmda_low_borrower_ratio = Loan_Orig_BILow/Loan_Orig,
       hmda_mod_borrower_ratio = Loan_Orig_BIMod/Loan_Orig,
       hmda_low_tract_ratio = Loan_Orig_TILow/Loan_Orig,
       hmda_mod_tract_ratio = Loan_Orig_TIMod/Loan_Orig)

#Look at the distribution of the HMDA Low/moderate income borrower ratios for all counties in US for 2019
quants = c(0.1,0.25,0.5,0.75,0.9)

hmda_low_borr_quants = quantile(hmda_county_2019$hmda_low_borrower_ratio, probs=quants, na.rm=T)
hmda_mod_borr_quants = quantile(hmda_county_2019$hmda_mod_borrower_ratio, probs=quants, na.rm=T)

#HMDA ratios for banks within Morgan County, Ohio
hmda_county_borr_median <- tibble(`Percentiles`=names(hmda_low_borr_quants),
                                   `Low Income Borrower`=hmda_low_borr_quants,
                                   `Moderate Income Borrower`=hmda_mod_borr_quants) %>%
mutate(across(c(`Low Income Borrower`:`Moderate Income Borrower`), round, 2))

```

Table 1. HMDA bank county borrower income distribution ratios, 2019

Percentiles	Low-income borrower	Moderate-income borrower
10	0.00	0.06
25	0.03	0.11
50	0.05	0.16
75	0.09	0.21
90	0.13	0.26

Now use the “retail_loan_lending_test” table to calculate the borrower and tract income distribution ratios for all MSAs in the country for 2019.

```

#Do the same but for only metropolitan statistical areas
hmda_msa_2019 <- retail_loan_lending_test %>%
  filter(ActivityYear == 2019) %>%
  filter(Lender_in_HMDA == "Y") %>%
  filter(Lender_in_CRA == "Y") %>%
  group_by(ActivityYear, MSA_Code) %>%
  summarise(across(c(Loan_Orig, Loan_Orig_BILow, Loan_Orig_BIMod, Loan_Orig_TILow, Loan_Orig_TIMod), sum, na.rm=T)) %>%
  mutate(hmda_low_borrower_ratio = Loan_Orig_BILow/Loan_Orig,
         hmda_mod_borrower_ratio = Loan_Orig_BIMod/Loan_Orig,
         hmda_low_tract_ratio = Loan_Orig_TILow/Loan_Orig,
         hmda_mod_tract_ratio = Loan_Orig_TIMod/Loan_Orig)

```

```

hmda_msa_low_borr_quants = quantile(hmda_msa_2019$hmda_low_borrower_ratio,probs=quants,na.rm=T)
hmda_msa_mod_borr_quants = quantile(hmda_msa_2019$hmda_mod_borrower_ratio,probs=quants,na.rm=T)
hmda_msa_low_tract_quants = quantile(hmda_msa_2019$hmda_low_tract_ratio,probs=quants,na.rm=T)
hmda_msa_mod_tract_quants = quantile(hmda_msa_2019$hmda_mod_tract_ratio,probs=quants,na.rm=T)

#HMDA ratios for banks within Morgan County, Ohio
hmda_msa_borr_tract_quantiles <- tibble(`Percentiles`=names(hmda_msa_low_borr_quants),
                                       `Low Income Borrower`=hmda_msa_low_borr_quants,
                                       `Moderate Income Borrower`=hmda_msa_mod_borr_quants,
                                       `Low Income Tract`=hmda_msa_low_tract_quants,
                                       `Moderate Income Tract`=hmda_msa_mod_tract_quants) %>%
  mutate(across(c(`Low Income Borrower`:`Moderate Income Tract`),round,2))

```

Table 2. HMDA bank MSA borrower and tract income distribution ratios, 2019

Percentiles	Low-income borrower	Moderate-income borrower	Low-income tract	Moderate-income tract
10%	0.02	0.09	0.00	0.06
25%	0.04	0.13	0.00	0.08
50%	0.05	0.16	0.01	0.12
75%	0.08	0.19	0.02	0.14
90%	0.10	0.22	0.03	0.16

CRA Lending in R Code

Use the “retail_loan_lending_test” table to calculate the inside assessment area small business loan borrower and tract income distribution ratios for MSAs in 2019.

```

cra_borr_tract_msa_2019 <- retail_loan_lending_test %>%
  filter(ActivityYear == 2019 & Lender_in_CRA == "Y" & County_AA_Flag == 1) %>%
  group_by(ActivityYear,id_rssd,MSA_Code) %>%
  summarise(across(c(SB_Loan_Orig_Inside,SB_Loan_Orig_TILow_Inside,SB_Loan_Orig_TIMod_Inside,
                    SB_Loan_Orig_GAR_less_1m_Inside),sum,na.rm=T)) %>%
  filter(SB_Loan_Orig_Inside > 0) %>%
  mutate(`SB Borrower Ratio` = SB_Loan_Orig_GAR_less_1m_Inside/SB_Loan_Orig_Inside,
         `SB Tract Low Ratio` = SB_Loan_Orig_TILow_Inside/SB_Loan_Orig_Inside,
         `SB Tract Moderate Ratio` = SB_Loan_Orig_TIMod_Inside/SB_Loan_Orig_Inside)

```

#distributions

```

sb_borr_quants = quantile(cra_borr_tract_msa_2019$`SB Borrower Ratio`,probs=quants,na.rm=T)
sb_tract_low_quants = quantile(cra_borr_tract_msa_2019$`SB Tract Low Ratio`,probs=quants,na.rm=T)
sb_tract_mod_quants = quantile(cra_borr_tract_msa_2019$`SB Tract Moderate Ratio`,probs=quants,na.rm=T)

sb_borr_tract_percentiles <- tibble(Percentiles = names(sb_borr_quants),
  `SB Borrower Ratio` = sb_borr_quants,
  `SB Tract Low Ratio` = sb_tract_low_quants,
  `SB Tract Moderate Ratio` = sb_tract_mod_quants)
%>%
  mutate(across(c(`SB Borrower Ratio`:`SB Tract Moderate Ratio`),round,2))

```

Table 3: CRA small business loan borrower and tract income ratio distributions in MSA assessment areas, 2019

Percentiles	SB borrower ratio	SB tract low ratio	SB tract moderate ratio
10	0.29	0.00	0.01
25	0.44	0.00	0.11
50	0.58	0.03	0.17
75	0.70	0.08	0.25
90	0.84	0.14	0.36

PE Table Example

The code and chart below show a distribution of PEs by bank agency and state lending test rating specifically for the PE Table.

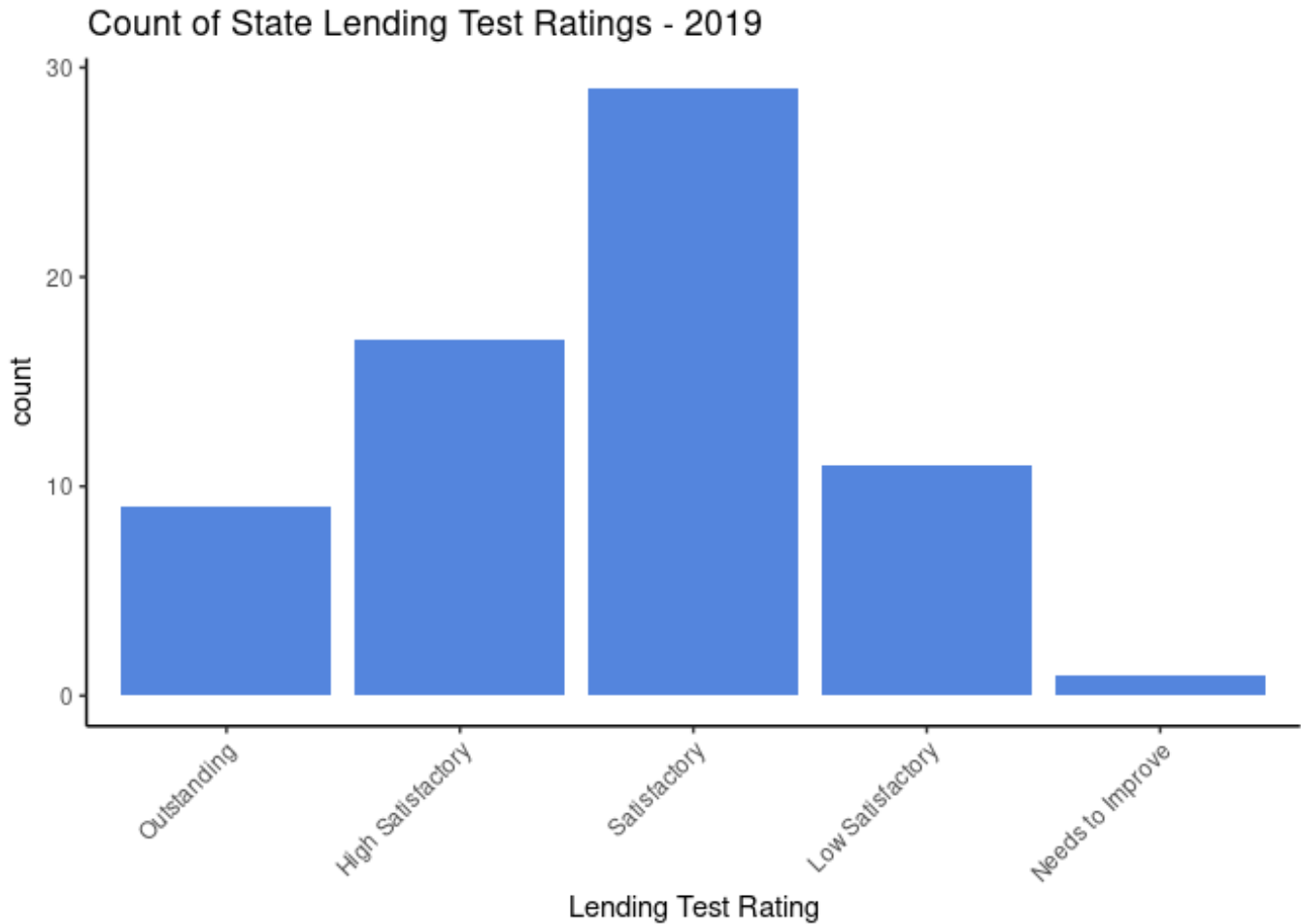
```

PE_2019_lending_test <- PE_2019 %>%
  filter(assessment_area_type == "state")%>%
  group_by(lending_test_rating) %>%
  summarise(count = n()) %>%
  mutate(lending_test_rating = factor(lending_test_rating, levels = c("Outstanding",
"High Satisfactory", "Satisfactory", "Low Satisfactory","Needs to Improve", "Substantial Non-Compliance")))

```

Figure 3. Count of State Lending Test Ratings, 2019

The chart was created using the PE table. The x-axis range is organized left to right by state lending test ratings, starting with “Outstanding” (the highest rating), “High Satisfactory,” “Satisfactory,” “Low Satisfactory,” “Needs to Improve.” The y-axis is a count of 2019 ratings and ranges from 0 to 30. “Outstanding” ratings are approximately 10 in count. “High Satisfactory” ratings are approximately 20 in count. “Satisfactory” ratings have the highest overall count of approximately 30. “Low Satisfactory” ratings are about 10 in count. There is approximately 1 “Needs to Improve” rating.



Merged Data Table Example

The code and chart below show count of MSA and non-MSA AAs in the Merged Data Table by lending test rating and regulatory agency.

```
#R Code, merged_2017 is the Merged Data Table filtered for year 2017  
merged_data_2019_lending_test <- merged_data_2019 %>%  
  filter(assessment_area_type %in% c("msa", "non_msa")) %>%  
  group_by(lending_test_rating)%>%  
  summarise(AA_count = n())
```


Figure 4. Count of MSA and non-MSA AAs by Lending Test Conclusion, 2019

The chart was created using the Merged Data Table. The x-axis range is organized left to right by lending test ratings, starting with "Outstanding" (the highest rating), "High Satisfactory," "Satisfactory," "Low Satisfactory," "Needs to Improve," and lastly, "Substantial Non-Compliance" (the lowest rating). The y-axis is a count of 2019 assessment area conclusions and ranges from 0 to 300. Approximately 50 AAs received "Outstanding" ratings. Approximately 60 AAs received "High Satisfactory" ratings. Approximately 300 AAs received "Satisfactory" ratings. Approximately 40 AAs received "Low Satisfactory" ratings. Approximately 30 AAs received "Needs to Improve" ratings, and one AA received a "Substantial Non-Compliance" rating.

