User Guide | CRA Analytics Data Tables

Board of Governors of the Federal Reserve System

March 6, 2020

Table of Contents

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 "aa_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 it to the assigned ratings, Home Mortgage Disclosure Act (HMDA) data and CRA data were combined, and AA data were extracted from the PEs for banks that did not report CRA data. HMDA and CRA small business and small farm loan data were collected for the years of 2005 to 2017. The data were aggregated by tract income status and/or borrower income level, depending on the variables. The coverage for the PE data is also 2005 to 2017.

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 due to resubmissions.

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 were used to aggregate the data to the county level with breakouts for total, inside AA and outside AA lending. The data tables use the post-publication version of the CRA data, which includes 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.

Data Tables

There are four main tables: Retail Loan Table, Performance Evaluation (PE) Table, Merged Data Table, and AA Definitions Table. The fifth table, PE Accuracy Table Check, contains quality control data for the PE Table. Descriptions of variables across the four main tables are included in the table-specific data dictionaries and the Data Dictionary Guide.

1. Retail Loan Table

The Retail Loan Table contains HMDA LAR and CRA small business and small farm data for depository institutions, aggregated at the year, bank, and county level from 2005 through 2017. 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, where applicable.

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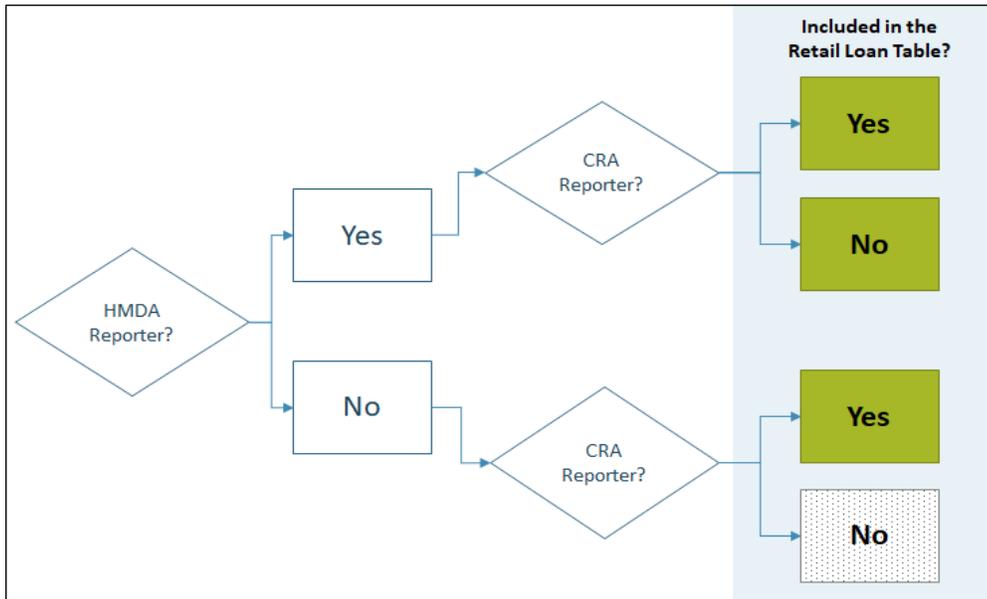

Figure 1. Reporters included in the Retail Loan Table - *Green, solid boxes indicate inclusion in the Retail Loan Table, i.e., the bank must be either a HMDA reporter or a CRA reporter to be included in the Retail Loan Table. If neither HMDA nor CRA reporter (patterned white box), then the bank's data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

2. Performance Evaluation (PE) Table

The PE data were hand collected from a sample of over 6,000 CRA PEs for a diverse sample of bank sizes and business models. The PEs were selected using a semi-random sampling strategy (see Sampling Procedure).

For each PE reviewed, the following data were extracted where available:

1. Lending, Community Development, Investment and overall test ratings for institution, state, multi-state metropolitan statistical area (MSA), MSAs, and non-MSA AAs
2. Evaluation periods for each product the examiner included as a part of the PE
3. Community Development loan counts and amounts as well as investment amounts for each AA

For banks that were not required to report CRA data at the time of the exam, the following were collected:

1. Geographical boundaries of their AAs
2. Number and amount of retail lending and community development lending and investments conducted inside and outside of the banks' AAs
3. Information on Loan Production Offices (LPOs), where available

The integration of these Retail Loan and PE data constitutes the Merged Data Table.

Sampling Procedure

Data were recorded from a stratified random sample of approximately 6,300 PEs. A number of considerations drove the sampling procedure design for selecting PEs to include in the assembled database. The sample was drawn from evaluations conducted from 2005 to 2017. Evaluations cover bank performance over a period of years

preceding the evaluation start date, so this sample 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 (roughly less than $300 million in assets), Intermediate Small (roughly between $300 million and $1.2 billion), and Large (roughly greater than $1.2 billion).

Historically, ratings below a "Satisfactory" have been relatively uncommon. Staff, therefore, oversampled banks with "Needs to Improve" or "Substantial Noncompliance" 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. In all, approximately 22,000 PEs with an evaluation start date from 2005 through 2017 are on the websites of the Federal Reserve Board (the Board), the Federal Deposit Insurance Corporation (FDIC), and the Office of the Comptroller of the Currency (OCC). Of these, staff collected data from a sample of approximately 6,300 PEs.

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 multi-state 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:

    i)   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 multi-state MSA
    ii) FDIC bank evaluations where the bank level lending test rating (or overall rating, for Small Banks) was "Needs to Improve" or worse
    iii) OCC bank evaluations where the bank level overall rating was "Needs to Improve" or worse

*   *Examiners assign banks ratings on the various CRA tests at the state, multi-state 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 multi-state MSA level. Therefore, for OCC banks, ratings on each test for an AA are available only when that AA was the only one the bank had in a particular state or multi-state MSA.

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

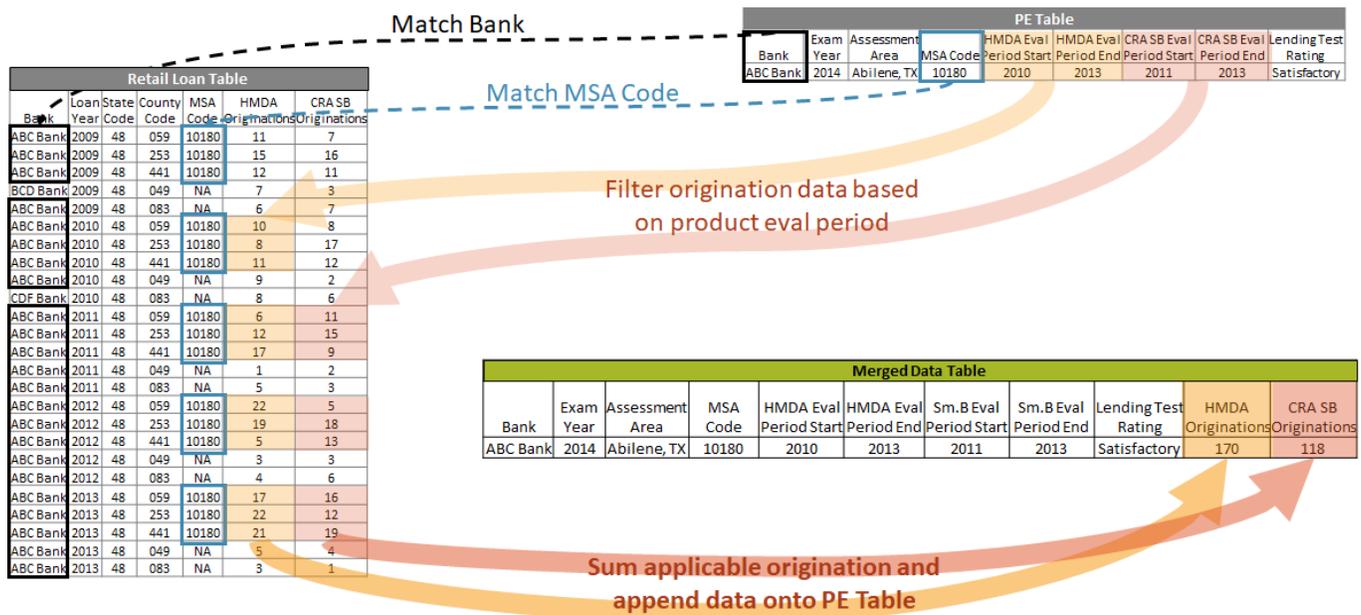# Merging Retail Loan and PE Tables to Produce Merged Data Table

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). This process is carried out within an AA for a given exam period and links these lending activity data to a lending test rating. There are fewer observations in the Merged Data Table compared to the PE Table, which results from losses in the merge process. These losses are explained in greater detail in figure 3.

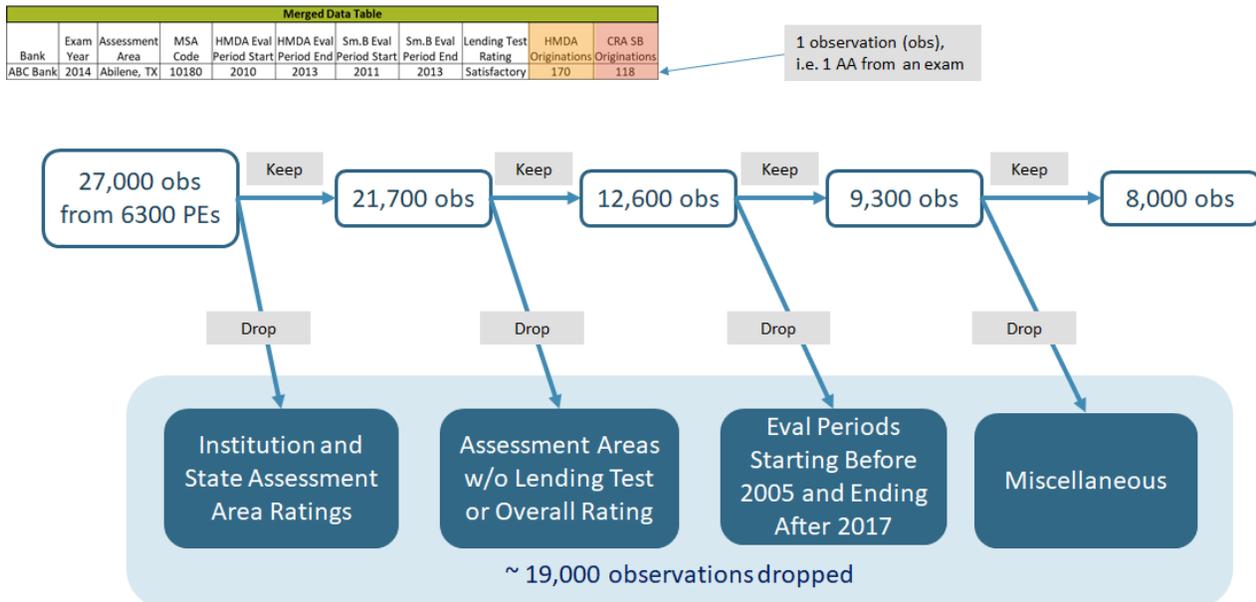## Loss of Observations from PE Table to Merged Data Table

Figure 3. Loss of observations from PE Table to Merged Data Table - *This figure depicts the observations retained and dropped in the merge process between the Retail Loan Table and the PE table. There are five light-colored boxes in a horizontal line, each representing the number of observations in the table kept at a particular step in the process. There are arrow pointing downward from these boxes to a set of dark-colored boxes, annotated with the word, "Drop"; these represent the number of observations that were dropped at the corresponding step of the process. Starting with the first box on the left, which represents the initial 27,000 observations from 6,300 PEs, the "Drop" arrow indicates that 5,300 "Institution and State Assessment Area Ratings" were dropped. The second box shows the remaining 21,700 observations with a drop of 9,100 observations due to "Assessment Areas without Lending Test or overall Ratings". The third box shows the remaining 12,600 observations with a drop of 3,300 observations due to "Evaluation Periods Starting Before 2005 or Ending After 2017". The fourth box shows the remaining 9,300 resulting observations with a drop of 1,300 observations due to miscellaneous reasons. The final box in the sequence represents the 8,000 observations in the Merged Data table.*

4. Assessment Area Definitions Table

This table combines AA geography data collected from PEs for Large Banks with AA information from CRA data. Joining these AA data together results in (approximate) county level geographic definitions for each AA included in the Merged Data Table. AA geography definitions are also available for all of the Small Bank and Intermediate Small Bank (ISB) AAs collected from the PEs.

• 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 core based statistical areas (CBSA), metropolitan division (MD), or combined statistical area (CSA). These geography codes used in the exam for the AA were collected and mapped to all counties in the CRA data within that same MSA code. These counties were also flagged by the lenders in the CRA data as being in the banks' AA during the CRA evaluation period.

- 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 AA were 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 core tables.

Retail Loan Table

As the content for the Retail Loan Table come 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 reevaluated and re-extracted data from a sample of 200 PEs. The re-extracted data were compared to the previously collected data for each major type of data point collected. The review found that the data were misreported one percent and missing in error three percent of the time.

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" table included in this release. While the data collection process included robust quality control checks, the data may nevertheless be subject to error. To submit a question or feedback, please fill out our 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 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

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 of less than 80 percent of the area median family income are considered LMI. However, each primary dataset provides different definitions for loans to LMI borrowers. For HMDA purposes, LMI *borrowers* refer to borrowers with incomes less than 80 percent of the area median family income. For CRA small business and small farm lending, an LMI borrower refers to a small business with gross annual revenue of less than one million dollars.

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. Records from the loan table which matched a row in the SoD table were flagged as "Inside_SoD_Proxy_AA" and otherwise flagged as "Outside_SoD_Proxy_AA". For cases where the lender is not in the annual SoD data, the value "Lender_Not_in_SoD" was assigned.

County Assessment Area Flag (County_AA_Flag)

County AA Flag 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.

Partial Indicator (partial_aa)

The Partial Indicator 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.

- 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 U.S.

Summary of Deposits (SoD)

The FDIC SoD data is an annually reported dataset, which contains the location of each branch office for all FDIC-insured banks for the reporting year. 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 U.S., Puerto Rico, and the Virgin Islands. This data includes 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.

Analytic Examples

Below are analytic examples from the data tables using R code and the tidyverse package.

Retail Loan Table

These scripts and tables show the ratio of banks' LMI lending to overall lending for HMDA and CRA loans by borrower and tract income.

HMDA lending in R Code

```
#R Code, loans_2017 is the Retail Loan Table filtered for year 2017
hmda_lmi <- loans_2017%>%
  filter(Lender_in_HMDA == "Lender_in_HMDA") %>%
  group_by(state_code_fips, county_code_fips) %>%
 mutate(hmda_lmi_ratio =  H_L_1/ H_1,
        hmda_lmi_borrower_ratio = H_bL_1/ H_1,
        hmda_lmi_tract_ratio = H_tL_1/ H_1,
        hmda_inside_AA_ratio = H_in_1/ H_1,
        hmda_outside_AA_ratio = H_out_1/ H_1) %>%
  select(state_code_fips, county_code_fips, id_rssd, hmda_lmi_ratio, hmda_lmi_borrow
er_ratio, hmda_lmi_tract_ratio)
```

```
#HMDA ratios for banks within Morgan County, Ohio
hmda_lmi_Morgan_OH <- hmda_lmi %>%
  filter(state_code_fips == "39" & county_code_fips == "115",
         !is.na(hmda_lmi_ratio)) %>%
  ungroup() %>%
  select(id_rssd, hmda_lmi_ratio, hmda_lmi_borrower_ratio, hmda_lmi_tract_ratio) %>%
  head()
```

id_rssd	hmda_lmi_ratio	hmda_lmi_borrower_ratio	hmda_lmi_tract_ratio
44929	0.6271186	0.2203390	0.5084746
783910	0.5500000	0.2500000	0.4500000
451965	0.4000000	0.2000000	0.2000000
3282852	1.0000000	0.0000000	1.0000000
584724	0.3333333	0.3333333	0.0000000
619877	0.0000000	0.0000000	0.0000000

CRA lending in R Code
```
cra_lmi <- loans_2017 %>%
  filter(Lender_in_CRA == "Lender_in_CRA") %>%
  mutate(cra_lmi_borrower_ratio = C_bL_1/ C_1,
         sb_lmi_borrower_ratio = sb_bL_1/ sb_1,
         sf_lmi_borrower_ratio = sf_bL_1/ sf_1)

#CRA ratios for banks within Morgan County, Ohio
cra_lmi_Morgan_OH <- cra_lmi %>%
  filter(state_code_fips == "39" & county_code_fips == "115",
         !is.na(cra_lmi_borrower_ratio)) %>%
  ungroup() %>%
  select(id_rssd, cra_lmi_borrower_ratio, sb_lmi_borrower_ratio, sf_lmi_borrower_rat
io) %>%
  head()
```

id_rssd	cra_lmi_borrower_ratio	sb_lmi_borrower_ratio	sf_lmi_borrower_ratio
451965	1.0000000	1.0000000	NaN
476810	1.0000000	1.0000000	NaN
504713	1.0000000	1.0000000	1
12311	0.5000000	0.5000000	NaN
577128	0.7777778	0.7142857	1
817824	0.4285714	0.4285714	NaN

PE Table

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

```
PE_2017_lending_test <- PE_2017 %>%
  filter(aa_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", "Substan
tial Non-Compliance")))
```
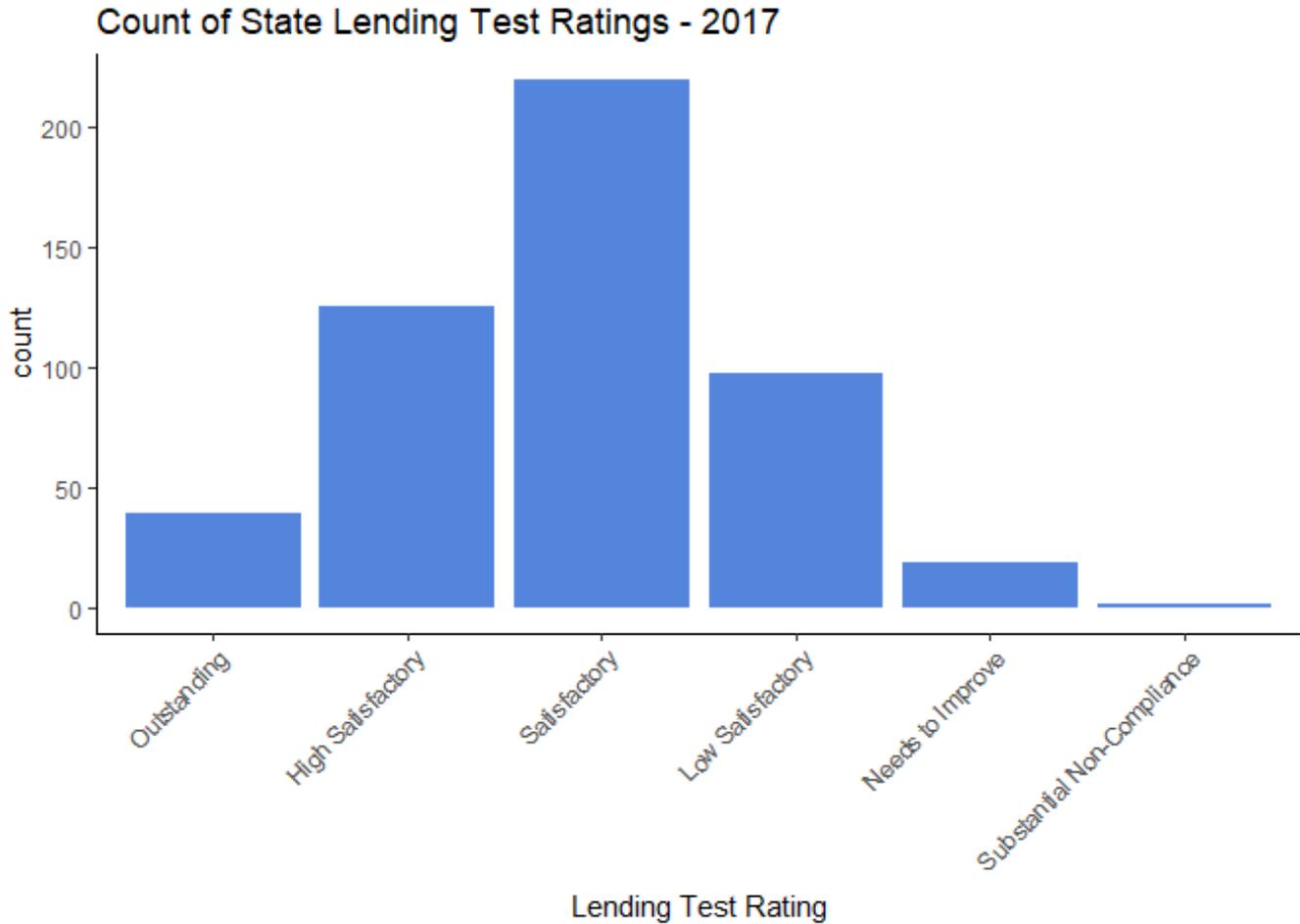

Figure 4. Count of State Lending Test Ratings - 2017 - *This is a bar chart titled "Count of State Lending Test Ratings by Agency - 2017." The chart was created using the PE table. The x-axis range is organized left to right by state lending test "ratings," starting with "Outstanding" (the highest rating), "High Satisfactory," "Satisfactory," "Low Satisfactory," "Needs to Improve," and lastly, "Substantial Noncompliance" (the lowest rating). The y-axis is a count of 2017 ratings and ranges from 0 to 250. "Outstanding" ratings are approximately 40 in count. "High Satisfactory" ratings are approximately 130 in count. "Satisfactory" ratings have the highest overall count of approximately 220. "Low Satisfactory" ratings are about 100 in count. There are approximately 20 "Needs to Improve" ratings and one "Substantial Noncompliance" state lending rating.*

Merged Data Table

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.

11

```
#R Code, merged_2017 is the Merged Data Table filtered for year 2017
merged_data_2017_lending_test <- merged_data_2017 %>%
  filter(aa_type %in% c("msa", "non_msa")) %>%
  group_by(lending_test_rating)%>%
  summarise(AA_count = n())
```
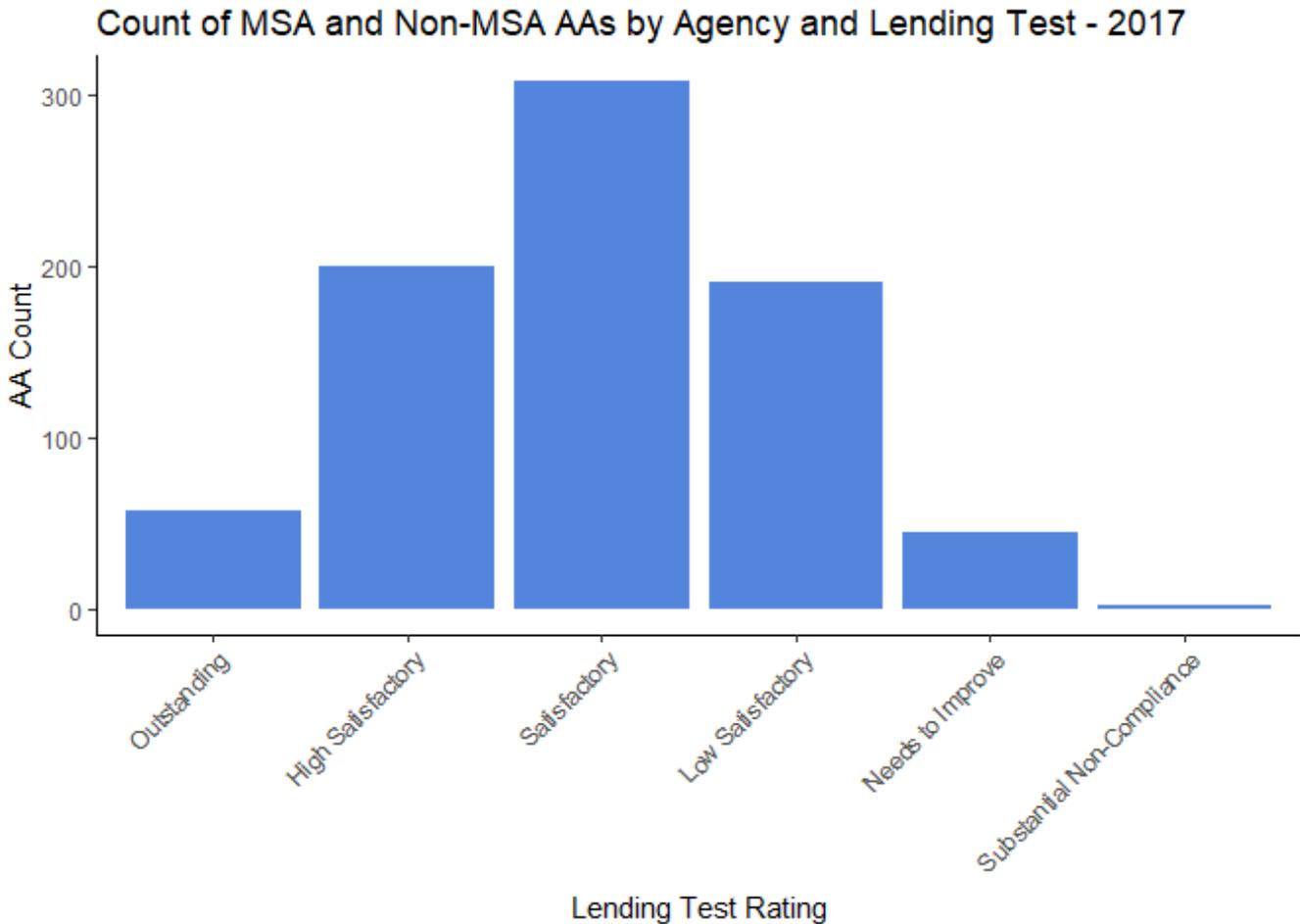

Figure 5. Count of MSA and Non-MSA AAs by lending test ratings - 2017 - *This is a bar chart titled "Count of MSA and Non-MSA AAs by Agency and Lending Test Ratings - 2017." The chart was created using the Merged Data Table. The x-axis range is organized left to right by lending test "ratings," starting with "Outstanding" (the highest rating),"High Satisfactory," "Satisfactory," "Low Satisfactory," "Needs to Improve," and lastly, "Substantial Noncompliance" (the lowest rating). The y-axis is a count of 2017 ratings and ranges from 0 to 300. Approximately 70 AAs received "Outstanding" ratings. Approximately 210 AAs received "High Satisfactory" ratings. Approximately 330 AAs received "Satisfactory" ratings. Approximately 200 AAs received "Low Satisfactory" ratings. Approximately 50 AAs received "Needs to Improve" ratings and two AAs received "Substantial Noncompliance" ratings*