

HMDA Data

Metadata File

Loans Data

Field	Type	Description	Data Validation
As_of_Year	integer	Year in which loan is originated and for which the data is reported. Keep in mind these are real people reporting data, and often variations can crop up in how they handle each year	<ul style="list-style-type: none"> This should have only integers in the format YYYY
Agency_Code	alphanumeric	The agency to which the Respondent_ID "responded" with its information	<ul style="list-style-type: none"> This should be the type Factor and have values of 1,2,3,5,7,9 only The mappings to descriptions should match
Agency_Code_Description	string	Expanded form of the agency code for readability.	<ul style="list-style-type: none"> This should be of type Factor and all the values with match with appropriate agency codes
Respondent_ID	string	When paired with the Agency_Code, provides a unique identifier for the institution that year.	<ul style="list-style-type: none"> This should be of type char and should have 10 characters If it has less than 10 characters, append 0's before it
Sequence_Number	integer	Provides a unique identifier corresponding to an application or loan for a given responding institution. I.e., each application at a responding institution must be assigned its own unique sequence number in a given year.	<ul style="list-style-type: none"> This should be of type integer There should be no duplicates for any selection of responding institution and year combination
Loan_Amount_000	integer	Amount of the loan, in \$000's	<ul style="list-style-type: none"> This should be of type integer
Applicant_Income_000	integer	The total income used for decisioning the loan	<ul style="list-style-type: none"> This should be of the type integer
Loan_Purpose_Description	string	The purpose of the loan is whether it is used for purchasing a home, refinancing an already-existing loan, or used for "home improvement" purposes. ("Home Improvement" loans have been removed from this data file.)	<ul style="list-style-type: none"> This should be of the type Character This should have values Refinance and Purchase (At least for this analysis as the remodeling rows are removed from data)
Loan_Type_Description	string	Specifies whether the loan is "conventional," i.e., not backed by a government-sponsored program or government-insured or government-guaranteed	<ul style="list-style-type: none"> This should be of the type Character This should have one of these only: Conventional, FHA Insured, FSA/RHS insured, VA guaranteed
Lien_Status_Description	string	Lien position of the loan. If first lien position, then this loan carries the least risk if the borrower defaults because the lender has first claims to proceeds from the sale of a home. Subordinate liens carry more risk because they are paid after the lender for the first lien is paid.	<ul style="list-style-type: none"> This should be either First Lien or Subordinate Lien Should be of type Factor
State_Code	alphanumeric	Two-digit FIPS code for the state in which the home is located.	<ul style="list-style-type: none"> This should be of the type integer State code should be valid (This can be eliminated for mediated Schema)
State	string	Standard 2-letter postal code for the state based on the State_Code given	<ul style="list-style-type: none"> This should be of the type character with length 2 This should have a valid state value
County_Code	alphanumeric	Numeric code designating a county within a state in which the home is located	<ul style="list-style-type: none"> This should be of the type integer State code should be valid (This can be eliminated for mediated Schema)

MSA_MD	alphanumeric	code designating a Metropolitan Statistical Area (MSA) / Metropolitan Division (MD) in which the home is located. These codes are periodically updated, and are last updated for HMDA between the 2013 and 2014 reporting years.	<ul style="list-style-type: none"> This should be integer with 5 digits ID should match the descriptions
MSA_MD_Description	string	a human-readable format for describing MSA's. These map 1:1 with MSA_MD codes	<ul style="list-style-type: none"> This should be of type character State Names after comma can be removed
Census_Tract_Number	alphanumeric	Identifier for the census tract in which the home is located	<ul style="list-style-type: none"> This should be of the type float
FFIEC_Median_Family_Income	alphanumeric	Median family income for the MSA/MD in which the census tract is located	<ul style="list-style-type: none"> This should be of the type integer To be divided by 1000 for data uniformity
Tract_to_MSA_MD_Income_Pct	alphanumeric	Percentage of the MSA/MD median Family income for the census tract; 2 decimal places of precision	<ul style="list-style-type: none"> This should be of the type number
Number_of_Owner_Occupied_Units	alphanumeric	Number of units, in the census tract and including condominiums, in which the owner lives	<ul style="list-style-type: none"> This should be of the type integer Needs to be check with a functional person about this column
County_Name	alphanumeric	Name of the county as pulled in from the FHFA Conforming Loan Limits as calculated under HERA. Please note that county names are only unique within a state	<ul style="list-style-type: none"> This should be of the type character This should have a valid county value
Conforming_Limit_000	numeric	Single Unit Conforming Loan limit for the given county.	<ul style="list-style-type: none"> This should be of the type integer
Conventional_Status	string	Signifies whether a loan type is '1' or not.	<ul style="list-style-type: none"> This should be a factor
Conforming_Status	string	'Conforming' for loans with amounts less than their county's respective conforming loan limit. "Jumbo" if the loan amount exceeds the county limit	<ul style="list-style-type: none"> This should be of the type factor
Conventional_Conforming_Flag	string	'Y' if Conventional_Status is "Conventional" and "Conforming_Status" is "Conforming"	<ul style="list-style-type: none"> This should a factor

Institutions Data

Field	Type	Description	Data Validation
As_of_Year	integer	Year in which loan was originated and for which the data was reported. Keep in mind these are real people reporting data, and often variations can crop up in how they handle each year	<ul style="list-style-type: none"> This should have only integers in the format YYYY
Agency_Code	alphanumeric	The agency to which the Respondent_ID "responded" with its information	<ul style="list-style-type: none"> This should be of the type integer and have values of 1,2,3,5,7,9 only
Respondent_ID	string	When paired with the Agency_Code, provides a unique identifier for the institution that year	<ul style="list-style-type: none"> This should be of type char and should have 10 characters If it has less than 10 characters, append 0's before it
Respondent_Name_TS	string	name of the company providing the HMDA data (and who originated the loan)	<ul style="list-style-type: none"> This should be of type CHAR
Respondent_City_TS	string	City where the originating institution is located.	<ul style="list-style-type: none"> This should be of type CHAR

Respondent_State_TS	string	State where the originating institution is located	<ul style="list-style-type: none"> • This should be of type CHAR • State ID must be valid
Respondent_ZIP_Code	string	State where the originating institution is located	<ul style="list-style-type: none"> • This should be of type CHAR • ZIP code should be of 5 digits only (Take first 5 digits)
Parent_Name_TS	string	The reported Parent to the responding insitutuion. Parent_Name_TS may refer to another entry within the institution file. Parent_Name_TS may be blank if the responding instution has no parent to report (and hence could be regarded as its own parent in certain circumstances). Finally, a Parent institution may have multiple children institutions within the file. Parent institutions may change with each reporting year.	<ul style="list-style-type: none"> • This should be of type CHAR
Parent_City_TS	string	City where the parent institution is located	<ul style="list-style-type: none"> • This should be of type CHAR
Parent_State_TS	string	State where the parent institution is located	<ul style="list-style-type: none"> • This should be of type CHAR • State should be a valid state
Parent_ZIP_Code	string	ZIP where the parent institution is located	<ul style="list-style-type: none"> • This should be of type CHAR • ZIP code should be of 5 digits only (Take first 5 digits)
Assets_000_Panel	int64	Total assets reported to the "Agency" for the top parent by the top level parent corporation across the nation.	<ul style="list-style-type: none"> • This should be of the type Integer

Area Data

Field	Type	Description	Data Validation
State	factor	These ate the states which are present in the Loans and institutions data	NA
areas	numeric	The area of these state, collected from google. (5 values)	NA