**Single Family Loan Level dataset**

For each calendar quarter, there is one file containing loan origination data and one file containing monthly performance data for each loan in the origination datafile. There might be cases when the data is in the origination file but not in the performance file (this happens if the loan is paid in the month of origination, before first cycle begins)

**Sample dataset:**

Random sample of 50,000 loans selected in a year.

What is a vintage year?

Each vintage year has an origination datafile and one monthly performance data file.

**Column description:**

**ORIGINATION FILE:**

1. Credit Score: Summarizes the borrower’s credit worthiness, which may be indicative of the likelihood that the borrower will timely repay future obligations. This is used while originating the mortgage.

Values: 301 - 850

Space (3) = Unknown,

if Credit Score is < 301 or > 850.

Things to be done:

Handle missing values: Replace it with average credit score (use Property State Column 17 to determine the average Credit Score)

1. First Payment Date: - The date of the first scheduled mortgage payment due under the terms of the mortgage note.

Value: YYYYMM

Things to be done:

Convert it into date format

Handle missing values: Count the maximum number of occurrence and replace the value with that

1. First time home buyer flag: Indicates if the borrower will be the primary resident of the mortgage. Has no ownership interest in a residential property during three-year period other than ownership interest in the marital residence with a spouse.

Value: Y = Yes  N = No  Space (1) = Unknown or Not Applicable

Things to be done:

Replace space with “Not Applicable”

Replace missing value with “Not available”

1. Maturity Date: The month in which the final monthly payment on the mortgage is scheduled to be made

Things to be done:

Replace missing values with : check original loan term/12 to get the number of year and add it to the First payment due date

1. METROPOLITAN STATISTICAL AREA (MSA) OR METROPOLITAN DIVISION:

If an MSA applies to a mortgaged property, the applicable five-digit value is disclosed; however, if the mortgaged property also falls within a Metropolitan Division classification, the applicable five-digit value for the Metropolitan Division takes precedence and is disclosed instead.

Value:

Metropolitan Division or MSA Code.

Space (5) = Indicates that the area in which the mortgaged property is located is a) neither an MSA nor a Metropolitan Division, or b) unknown.

Things to be done:

Replace missing values: “Neither MSA nor MD/Unknown”

1. Mortgage Insurance Percentage: The percentage of loss coverage on the loan, at the time of Freddie Mac’s purchase of the mortgage loan that a mortgage insurer is providing to cover losses incurred because of a default on the loan.

Value:

1% - 55%

000 = No MI

Space (3) = Unknown

Things to be done:

Replace space/Missing values by “Unknown”

1. Number of Units: Denotes whether the mortgage is a one-, two-, three-, or four-unit property.

Values: 1 = one-unit

2 = two-unit

3 = three-unit

4 = four-unit

Space(1) = Unknown

Things to be done:

Replace space/Missing Value by “Unknown”

1. Occupancy Status: Denotes whether the mortgage type is owner occupied, second home, or investment property.

Value: O = Owner Occupied  I = Investment Property  S = Second Home  Space (1) = Unknown

Things to be done:

Replace Space/Missing Value: “Unknown”

Occupancy Status and First Time home buyer flag related?

1. Original combined loan-to-value: It is a proportion of loans (secured by a property) in relation to its value.

Loan to Value which simply indicates the ratio between one primary loan and the property value

This field is used to verify borrower’s secondary mortgage loan amount

Value:

If CLTV is < 0 or > 200, and the LTV is > 80 and <= 200 set the CLTV to the LTV.

If the LTV is < 80 or > 200 or unknown, set the CLTV to ‘unknown.’

If the CLTV is < LTV, set the CLTV to ‘unknown.’ Unknown is indicated by a null value.

Things to be done:

Replace null value with “Unknown”

Look into this (imp feature)

1. Original debt to income ratio: the sum of the borrower's monthly debt payments, including monthly housing expenses that incorporate the mortgage payment the borrower is making at the time of the delivery of the mortgage loan to Freddie Mac, divided by (2) the total monthly income used to underwrite the borrower as of the date of the origination of the mortgage loan.

This is used to verify borrower’s assets and liabilities

Values:

0%<DTI<=65%

Space(3) >65%

Null = Unknown

Things to be done:

Look into this (imp feature)

Replace spaces with >65%

Replace null/missing values with “Unknown”

1. Original UPB: Unpaid principal balance: It is the portion of a loan (e.g. a mortgage loan) at a certain point in time that has not yet been remitted to the lender.

Things to be done:

Look into this (imp feature)

Replace missing value with “Unknown”

1. Original Loan-to-value:

Ratios below 6% or greater than 105% will be disclosed as “Unknown,” indicated by three (3) blank spaces

Values:

6% - 105%

Space(3) = Unknown

Things to be done:

Look into this (imp prop)

Replace space/missing value with “Unknown”

1. Original interest rate: Interest rate

Things to be done:

Replace space/missing value with Mean value

1. Channel: Who was involved in the origination of the mortgage loan

Values:

 R = Retail

 B = Broker

 C = Correspondent

 T = TPO Not Specified

 Space = Unknown

Things to be done:

Replace missing/space with “unknown”

1. Pre-payment penalty mortgage flag: A PPM is a mortgage with respect to which the borrower is, or at any time has been, obligated to pay a penalty in the event of certain repayments of principals

Values:

Y = PPM

N = Not PPM

Space (1) = Unknown

Things to be done:

Replace missing/space with “Unknown”

1. Product type: Denotes if product is fixed rate mortgage

Value: FRM – Fixed Rate Mortgage

Things to be done:

Replace missing values with “FRM”

1. Property State: A two-letter abbreviation indicating the state or territory within which the property securing the mortgage is located

Things to be done:

Replace missing value with “Unknown”

1. Property Type: Type of the property

 CO = Condo

 LH = Leasehold

 PU = PUD

 MH = Manufactured Housing

 SF = 1-4 Fee Simple

 CP = Co-op

 Space (2) = Unknown

Things to be done:

Replace missing value with “Unknown”

1. Postal Code: First three digits of the zipcode followed by zeros

Things to be done:

Fetch the City from first three digits and append a column

Replace spaces/missing values with “Unknown”

1. Loan Sequence Number: Unique id for each loan

F1YYQnXXXXXX

 F1 = product (Fixed Rate Mortgage);

 YYQn = origination year and quarter;

 XXXXXX = randomly assigned digits

Things to be done:

If missing, discard the record

1. Loan Purpose: Indicates whether the mortgage loan is a Cashout Refinance mortgage, No Cash-out Refinance mortgage, or a Purchase mortgage.

Value:

P = Purchase

 C = Cash-out Refinance

 N = No Cash-out Refinance

 Space = Unknown

Things to be done:

Replace Space/Missing value with “Unknown”

Look into in (imp feature)

1. Original Loan Term: A calculation of the number of scheduled monthly payments of the mortgage based on the First Payment Date and Maturity Date.

Things to be done:

Apply the formula to get the value

Calculation: (Loan Maturity Date (MM/YY) – Loan First Payment Date (MM/YY) + 1)

1. Number of borrowers: The number of Borrower(s) who are obligated to repay the mortgage note secured by the mortgaged property

Values:

 01 = 1 borrower

 02 = > 1 borrowers

 Space (2) = Unknown

Things to be done:

Replace missing/space with “Unknown”

1. Seller name: The entity acting in its capacity as a seller of mortgages to Freddie Mac at the time of acquisition

Values:

Name of the seller, or “Other Sellers”

Seller Name will be disclosed for sellers with a total Original UPB representing 1% or more of the total Original UPB of all loans in the Dataset for a given calendar quarter

How to use this information?

Things to be done:

Replace missing/space with “Unknown”

1. Servicer name: The entity acting in its capacity as the servicer of mortgages to Freddie Mac as of the last period for which loan activity is reported in the Dataset.

Values: Name of the servicer, or “Other Servicers”

Servicer Name will be disclosed for servicers with a total Original UPB representing 1% or more of the total Original UPB of all loans in the Dataset for a given calendar quarter

How to use this information?

Things to be done:

Replace missing/space with “Unknown”

1. Super confirming flag: For mortgages that exceed conforming loan limits with origination dates on or after 10/1/2008 and settlements on or after 1/1/2009

Values: Y = Yes  Space (1) = Not Super Conforming

Things to be done:

Replace space with “No”

What is NATL?

**PERFORMANCE DATA FILE:**

1. Loan Sequence number: Unique identifier for the loan amount paid. Map this in the originations file

Things to be done:

If missing, discard the result

1. Monthly Reporting Period: The as-of month for loan information contained in the loan record.

What about the rest of the months for which the user was supposed to pay??

1. Current Actual UPB: The Current Actual UPB reflects the mortgage ending balance as reported by the servicer for the corresponding monthly reporting period

??? Did not understand this term

1. Current loan delinquency status: A value corresponding to the number of days the borrower is delinquent, based on the due date of last paid installment (“DDLPI”) reported by servicers to Freddie Mac, and is calculated under the Mortgage Bankers Association (MBA) method.

Values:

XX = Unknown

 0 = Current, or less than 30 days past due

 1 = 30-59 days delinquent

 2 = 60 – 89 days delinquent

 3 = 90 – 119 days delinquent

 And so on…

 R = REO Acquisition

 Space (3) = Unavailable

If a loan has been acquired by REO, then the Current Loan Delinquency Status will reflect the value corresponding to that status (instead of the value corresponding to the number of days the borrower is delinquent).

1. Loan age: The number of months since the note origination month of the mortgage.

Calculation: ((Monthly Reporting Period) – Loan Origination Date (MM/YY)) – 1 month

Things to do:

If missing compute using this formula

1. Remaining months to legal maturity: The remaining number of months to the mortgage maturity date.

Calculation: (Maturity Date (MM/YY) – Monthly Reporting Period (MM/YY)

Things to do:

If missing compute using this formula

For mortgages with loan modifications, as indicated by “Y” in the Modification Flag field, the calculation uses the modified maturity date.

1. Repurchase flag: Indicates loans that have been repurchased or made whole (not inclusive of pool-level repurchase settlements).

This field is only populated only at loan termination month.

Values:  N = Not Repurchased  Y = Repurchased  Space (1) = Not Applicable

1. Modification flag: For mortgages with loan modifications, indicates that the loan has been modified.

Values: Y = Yes  Space (1) = Not Modified

1. Zero balance code: A code indicating the reason the loan's balance was reduced to zero

 01 = Prepaid or Matured (Voluntary Payoff)

 03 = Foreclosure Alternative Group (Short Sale, Third Party Sale, Charge Off or Note Sale)

 06 = Repurchase prior to Property Disposition

 09 = REO Disposition  Space(2) = Not Applicable

1. Zero balance effective date: The date on which the event triggering the Zero Balance Code took place.

Value:

YYYYMM

 Space(6) = Not Applicable

1. Current interest rates: Reflects the current interest rate on the mortgage note, taking into account any loan modifications.
2. Current deferred UPB: The current non-interest bearing UPB of the modified mortgage.
3. Due date of last paid installment: The due date that the loan’s scheduled principal and interest is paid

During early and mid 2000’s people started taking second loan

Case 1: Risk of extending loan to a borrower:

Lower LTV means that it is less risky to lend them money

Case 2: Calculate the prepaid ratio for every quarter using zero balance code. Pranjal

Case 3: Qualify for home loan, total monthly gross income, monthly expenses (property tax, PMI, association dues, credit card payments)

Calculation for highest payment to be qualified:

Categorize credit into buckets Medium and Good credit

Medium: 28% of gross income

Or

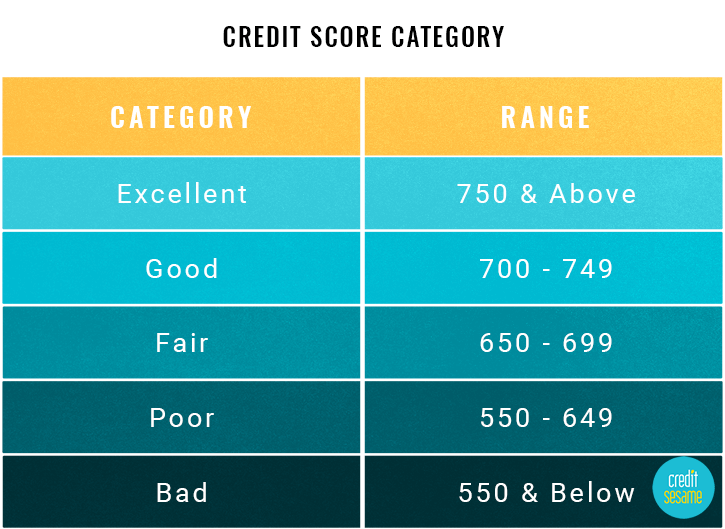
36% of gross income less fixed monthly expenses

Good credit: 36% of gross income

Or

42% of gross income less fixed monthly expenses

Credit buckets:



Medium 🡪 650 and below

Good 🡪 650 and above

**Steps**:

1. Create a column “Credit Bucket”, where if >650 user is in Good Credit whereas if < = 650 user is in Medium Credit
2. Create a column “Gross Income”, fetch original debt to income(DI) and unpaid principal balance (UPB) and apply formula:

Assumption: that the user has only one property in which he has done investment

UPB \* 100/DI

1. Create a column “Gross income less fixed monthly expenses” = Gross Income-UPB –additional expenses

Note: As per Bureau of Statistics the average expenditure of a married couple with children is 13% of their gross income

1. Create a column “Mortgage Qualification Amount” =take the minimum of (gross income \* percent\_as per\_credit\_score , Gross income less fixed monthly expenses \* percent\_as per\_credit\_score)

Ref: <https://www.bls.gov/news.release/cesan.nr0.htm>

<http://tcalc.timevalue.com/all-financial-calculators/mortgage-calculators/mortgage-qualification-calculator.aspx>