Freddie Mac Projects

Table of Contents

[Dataset Overview 3](#_Toc59198619)

[Origination Data 3](#_Toc59198620)

[Monthly Performance Data 5](#_Toc59198621)

[Project 1 – Changes in Mortgage Portfolio 8](#_Toc59198622)

[1.1 Background 8](#_Toc59198623)

[1.2 Sample 8](#_Toc59198624)

[1.3 Hypotheses - origination 8](#_Toc59198625)

[1.4 Definition of Delinquency - Performance 8](#_Toc59198626)

[Project 2 – Risk / Delinquency Model 10](#_Toc59198627)

[2.1 Business Purpose 10](#_Toc59198628)

[2.2 Sample 10](#_Toc59198629)

[2.3 Data Cleaning 10](#_Toc59198630)

[2.4 Modeling Method 11](#_Toc59198631)

[2.5 Out-of-time Validation 11](#_Toc59198632)

[2.6 A discussion on the model score from logistic regression 12](#_Toc59198633)

[Project 3 – Prepayment Model 13](#_Toc59198634)

[3.1 Definition of prepayment 13](#_Toc59198635)

[3.2 Criteria for defining prepayment in our data 13](#_Toc59198636)

[3.3 A note of a weak model 13](#_Toc59198637)

[Project 4 – Transition of Loan Status 14](#_Toc59198638)

[4.1 Business Purpose 14](#_Toc59198639)

[4.2 Common account statuses in consumer lending 14](#_Toc59198640)

[4.4 Final output of a transition matrix 14](#_Toc59198641)

[Project 5 – Survival Analysis 16](#_Toc59198642)

[5.1 Business Purpose 16](#_Toc59198643)

[5.2 Final Output of Survival Analysis 16](#_Toc59198644)

[Summary of Project Experience 17](#_Toc59198645)

# Dataset Overview

Freddie Mac Single Family Loan-Level Dataset

<http://www.freddiemac.com/research/datasets/sf_loanlevel_dataset.html>

1. Download the quarterly datasets of 2008 and 2009. Each zipped file contains two parts:
   1. Origination data
      1. Containing loan-level origination information for all the loans originated during the quarter.
      2. Loan characteristics such as FICO, LTV, CLTV, property type, occupancy status, etc.
   2. Performance data by month
      1. All of the respective loans originated during the quarter.
2. Status of every loan in each month since it enters the Freddie Mac book. - The documents serve as a data dictionary. Carefully review them to understand the meaning of each variable.

## Origination Data

|  |  |
| --- | --- |
| **Column** | **Note** |
| 1. Credit Score 🡪 fico | FICO - Summarizing the borrower’s creditworthiness  Subprime mortgage 🡪 fico <= 600  (Higher credit risk comes with a higher interest rate and closing costs) |
| 1. First Payment Date 🡪 dt\_first\_pi | The date of the first scheduled mortgage payment due under the terms of the mortgage note |
| 1. First Time Homebuyer Flag 🡪 flag\_fthb   Y = Yes  N = No  9 = Not Available | First Time Homebuyer Flag - Indicates whether the Borrower, or one of a group of Borrowers, is an individual who:   1. Purchasing the mortgaged property 2. Reside in the mortgaged property as a primary residence 3. Had no ownership interest (sole or joint) in a residential property during the three-year period preceding the date of the purchase of the mortgaged property.   Not Available: Investment Properties, Second Homes and Refinance transactions |
| 1. Maturity Date 🡪 dt\_matr | The final monthly payment on the mortgage is scheduled to be made |
| 1. Metropolitan Statistical Area (MSA) or Metropolitan Division 🡪 cd\_msa | A geographical region with a relatively high population density at its core and close economic ties throughout the area |
| 1. Mortgage Insurance Percentage (MI%) 🡪 mi\_pct | The percentage of loss coverage on the loan  1% - 55%  000 = No MI  999 = Not Available |
| 1. Number of Units 🡪 cnt\_units | Denotes whether the mortgage is a one-, two-, or four-unit property  99 = Not Available |
| 1. Occupancy Status 🡪 occpy\_sts | P = Primary Residence  I = Investment Property  S = Second Home  9 = Not Available |
| 1. Original Combined Loan-To-Value. (CLTV) 🡪 cltv   6% - 200%  HARP: 1% - 998%  999 = Not Available (when CLTV < LTV) | The combined loan-to-value (CLTV) ratio - The ratio of all [secured loans](https://www.investopedia.com/ask/answers/110614/what-difference-between-secured-and-unsecured-debts.asp) on a property to the value of a property. (second mortgages, home equity loans or home equity lines of credit)  Home equity loan - A[home equity loan](https://www.discover.com/home-equity-loans/) is essentially a one-time consumer loan using your home as collateral. If your home is worth more than you owe on it, you have equity, and may be able to use this equity to borrow money.  Home equity lines of credit (HELOCs) - Works like a credit card, that you can access as you choose  Home Affordable Refinance Program (HARP) - Designed for homeowners who are current on their mortgage payments, but who haven't been able to refinance because they have limited equity, no equity or negative equity in their homes. |
| 1. Original Debt-To-Income (DTI) Ratio 🡪 dti   0% < DTI < 65%  999 = Not Available | Debt-to-income ratio (DTI) - A personal finance measure that compares the amount of debt you have to your overall income. Lenders, including issuers of mortgages, use it as a way to measure your ability to manage the payments you make each month and repay the money you have borrowed. |
| 1. Original UPB 🡪 orig\_upb | UPB (Unpaid principal balance) - The portion of a mortgage loan at a certain point in time that has not yet been remitted to the lender |
| 1. Original Loan-To-Value (LTV) 🡪 ltv   6% - 105%  HARP: 1% - 998%  999 = Not Available | Loan-to-value - The ratio of your first mortgage balance to your home value (sales price or appraised value of your home)  CLTV ratio always >= LTV ratio |
| 1. Original Interest Rate 🡪 int\_rt | |
| 1. Channel 🡪 channel | R = Retail  B = Broker  C = Correspondent  T = TPO Not Specified (TPO 🡪 third-party originator)  9 = Not Available |
| 1. Prepayment Penalty Mortgage (PPM) Flag 🡪 ppmt\_pnlty   Y = PPM  N = Not PPM | Prepayment Penalty Mortgage (PPM) – Also called an early payoff penalty, is the fee that’s charged if you pay off your principal balance early |
| 1. Product Type 🡪 prod\_type | FRM – Fixed Rate Mortgage |
| 1. Property State 🡪 st | |
| 1. Property Type 🡪 prop\_type | CO = Condo  PU = PUD 🡪 Planned Unit Development  MH = Manufactured Housing  SF = Single Family  CP = Co-op 🡪 Cooperative Share  99 = Not Available |
| 1. Postal Code 🡪 zipcode | |
| 1. Loan Sequence Number 🡪 id\_loan | Unique identifier assigned to each loan  F1YYQnXXXXXX   * F1 = product (Fixed Rate Mortgage) * YYQn = origination year and quarter * XXXXXX = randomly assigned digits |
| 1. Loan Purpose 🡪 loan\_purpose   P = Purchase  C = Refinance – Cash Out  N = Refinance – No Cash Out  R = Refinance 🡪 Not Specified  9 = Not Available | A cash-out refinance is a mortgage refinancing option in which an old mortgage is replaced for a new one with a larger amount than owed on the previously existing loan, helping borrowers use their home mortgage to get some cash  A no cash-out refinance refers to the refinancing of an existing [mortgage](https://www.investopedia.com/terms/m/mortgage.asp) for an amount equal to or less than the existing outstanding loan balance plus any additional loan settlement costs. It is done primarily to lower the interest rate charge on the loan and/or to change some of the terms of the mortgage |
| 1. Original Loan Term 🡪 orig\_loan\_term | A calculation of the number of scheduled monthly payments of the mortgage based on the First Payment Date and Maturity Date  Calculation: (Loan Maturity Date (MM/YY) – Loan First Payment Date (MM/YY) + 1) |
| 1. Number of Borrowers 🡪 cnt\_borr | The number of borrower (s) who are obligated to repay the mortgage note secured by the mortgaged property |
| 1. Seller Name 🡪 seller\_name | |
| 1. Servicer Name 🡪 servicer\_name | |
| 1. Super Conforming Flag 🡪 flag\_sc   Y = Yes  N = Not Super Conforming | Freddie Mac's super conforming mortgages are mortgages originated using higher maximum loan limits that are permitted in designated high-cost areas |
| 1. Pre-HARP Loan Sequence Number 🡪 flag\_sc | |
| 1. Program Indicator 🡪 program\_ind | The indicator that identifies if a loan participates in the following Freddie Mac program |

## Monthly Performance Data

|  |  |
| --- | --- |
| **Column** | **Note** |
| 1. Loan Sequence Number 🡪 id\_loan | |
| 1. Monthly Reporting Period 🡪 Period | The as-of month for loan information contained in the loan record  🡪 YYYYMM |
| 1. Current Actual UPB 🡪 Act\_endg\_upb | Current Actual UPB = (interest bearing UPB) + (non-interest bearing UPB) |
| 1. Current Loan Delinquency Status 🡪 delq\_sts   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 | Delinquent (late) - when a scheduled payment is not made on or before the due date  REO (Real Estate Owned) Acquisition – The bank owns the property as the result of a foreclosure  Foreclosure - a legal process that allows a lender, or the subsequent loan owner, to sell your property to satisfy the debt you owe. (when you fall far enough behind in your mortgage payments) |
| 1. Loan Age 🡪 loan\_age | The number of months since the note origination month of the mortgage  = Monthly Reporting Period – Loan Origination Date – 1 month |
| 1. Remaining Months to Legal Maturity 🡪 mths\_remng | The remaining number of months to the mortgage maturity date  = Maturity Date – Monthly Reporting Period |
| 1. Repurchase Flag 🡪 repch\_flag | Indicates loans that have been repurchased or made whole |
| 1. Modification Flag 🡪 flag\_mod | A mortgage loan modification is a change in your loan terms. The modification is a type of loss mitigation.  The modification can reduce your monthly payment to an amount you can afford.  Modifications may involve extending the number of years you have to repay the loan, reducing your interest rate, and/or forbearing or reducing your principal balance. |
| 1. Zero Balance Code 🡪 CD\_Zero\_BAL   01 = Prepaid or Matured (Voluntary Payoff)  02 = Third Party Sale  03 = Short Sale / Charge Off  06 = Repurchase prior to Property Disposition  09 = REO Disposition  15 = Note sale / Reperforming sale | Short Sale – the home is being sold for less than the balance remaining on the mortgage  Charge Off - A charge-off occurs when a lender writes off unpaid debt for tax purposes. A charge-off is merely an accounting term. You're still on the hook for the debt you owe the bank  Repurchase (putback/buyback) - A financial vehicle by which a previously approved loan is taken back by the originator of the loan.  Originator - A mortgage originator is an institution or individual that works with a borrower to complete a home loan transaction.  Disposition - When a bank reviews its loans and decides to sell the collateral that has been taken in a foreclosure  Mortgage note - A financial document that details a loan agreement used to purchase property.  Reperforming loan - A mortgage that became delinquent because the borrower was behind on payments by at least 90 days, but it is "performing" again because the borrower has resumed making payments. |
| 1. Zero Balance Effective Date 🡪 Dt\_zero\_BAL | |
| 1. Current Interest Rate 🡪 New\_Int\_rt | |
| 1. Current Deferred UPB 🡪 Amt\_Non\_Int\_Brng\_Upb | The current non-interest bearing UPB of the modified mortgage. |
| 1. Due Date of Last Paid Installment (DDLPI) 🡪 Dt\_Lst\_Pi | |
| 1. Mortgage Insurance Recoveries 🡪 MI\_Recoveries | Mortgage insurance lowers the risk to the lender of making a loan to you, so you can qualify for a loan that you might not otherwise be able to get.  Mortgage Insurance Recoveries – The proceeds received by Freddie Mac in the event of credit losses. These proceeds are based on claims under a mortgage insurance policy. |
| 1. Net Sales Proceeds 🡪 Net\_Sale\_Proceed | Gross Sale Proceeds – Allowable Selling Expenses |
| 1. Non Mortgage Insuarance Recoveries 🡪 Non\_MI\_Recoveries | Non-MI recoveries include non-sale income such as tax and insurance refunds, hazard insurance proceeds, rental receipts, positive escrow, and other miscellaneous credits. |
| 1. Expenses 🡪 Expenses | Allowable Expenses (process of acquiring, maintaining, disposing a property) |
| 1. Legal Costs 🡪 legal\_costs | |
| 1. Maintenance and Preservation Costs 🡪 maint\_pres\_costs | |
| 1. Taxes and Insurance 🡪 taxes\_ins\_costs | |
| 1. Miscellaneous Expenses 🡪 misc\_costs | |
| 1. Actual Loss Calculation 🡪 actual\_loss | Actual Loss = (Default UPB – Net Sale\_Proceeds) + Delinquent Accrued Interest - Expenses – MI Recoveries – Non MI  Delinquent Accrued Interest = (Default\_UPB – Non Interest bearing UPB)\* (Current Interest rate – 0.35) \* (Months between Last Principal & Interest paid to date and zero balance date) \* 30/360/100 |
| 1. Modification Cost 🡪 modcost | |
| 1. Step Modification Flag 🡪 stepmod\_ind | To denote if the terms of modification agreement call for note rate to increase over time |
| 1. Deferred Payment Plan 🡪 dpm\_ind | |
| 1. Estimated Loan to Value (ELTV) 🡪 eltv | |
| 1. Zero Balance Removal UPB 🡪 zb\_removal\_upb | |
| 1. Delinquent Accrued Interest 🡪 dlq\_acrd\_int | |
| 1. Delinquency Due to Disaster 🡪 disaster\_hardship\_ind | |
| 1. Borrower Assistance Status Code 🡪 borrower\_assist\_flag | F = Forbearance  R = Repayment  T = Trial Period |

# Project 1 – Changes in Mortgage Portfolio

Examine changes in mortgage portfolio after the financial meltdown in 2008. This is a pure exploratory data analysis.

## 1.1 Background

Before the financial meltdown in the third quarter of 2008, many people purchased houses for investment purpose, i.e., they expected to the property value to go up quickly so that they could profit from reselling. Lenders also applied very lax policies in underwriting, i.e., no need for income verification or for down payment when reviewing loan applications.

Since the financial meltdown, the US government imposed more strict policies for underwriting. For example, requirement for FICO score has become higher, and income verification is needed.

* Underwriting 🡪 Your lender verifies your income, assets, debt and property details in order to issue final approval for your loan

## 1.2 Sample

Loans from 2008 Q1 to 2009 Q4 that meet the following criteria:

* Single family home
* 30-year fixed rate
* FICO between 550 and 850

(A random sample of 1% is good enough. If your computer constantly runs out of memory, use data from 2008 Q3, 2008 Q4, 2009 Q1 and 2009Q2 will be sufficient)

## 1.3 Hypotheses - origination

1. After the financial meltdown, loans in general have high credit scores (FICO).
2. After the financial meltdown, the portion of subprime loans (FICO < 600) decreased

substantially in the mortgage portfolio.

1. Before the financial meltdown, many borrowers thought they could profit from the

increased value of the houses in a few years, so portion of purchases (i.e., buying a house rather than refinancing an existing mortgage) was higher. After the meltdown, the portion of purchases went down. Lower interest rate in 2009 also boosted the market for refinancing

4) The quality of loans originated after meltdown is considerably higher, i.e., their delinquency rate is much lower compared to those loans originated before the meltdown.

* See the plot in 1.4

## 1.4 Definition of Delinquency - Performance

1. A loan has reached 90+ days in delinquency
2. Balance becomes zero, with delinquency flags 90+ days
3. A loan has shown recovery amount as recovery usually comes after foreclosure and repossession by the bank (REO)
4. A loan has been modified, i.e., interest rate reduction after negotiating with the bank.

Performance data need to be aggregated to derive a delinquency flag for each loan.

As older loans have longer performance histories and hence are more likely to encounter delinquency, we select performance data of 5 years for each loan.

Delinquency will be used in the next project of building a regression/classification model for risk prediction.

# Project 2 – Risk / Delinquency Model

A regression/classification model for predicting delinquency. Either logistic regression or decision tree will be used for building the risk model.

The methods and practices introduced here are also used in modeling risk for credit card, auto loans, personal loans, etc. Many companies also use this method for building marketing models.

## 2.1 Business Purpose

Predicting delinquency rate (usually called bad rate in industry) will help a bank or an investor to do the following:

-  For underwriting, i.e., approving or declining a loan application. Applicants with very high probability estimates for delinquency are declined.

-  A better pricing strategy. Loans with higher probabilities of delinquency should be charged a higher interest rate. (Interest rate is considered the price of a loan.)

-  A better valuation of loan quality. When a bank or an investor buys a mortgage portfolio from another company, quality of loans needs to be evaluated. A portfolio of higher delinquency rate will be evaluated for a lower price for purchase.

## 2.2 Sample

1)  6-%-10% of all eligible loans generated in 2009. Sample size depends on the memory size of your computer. Use the same criteria as in Project 1 to select loans eligible or modeling (see 1.3).

2)  Use the same definition of delinquency as in Project 1 (see 1.4).

3)  Split the sample by 60/40 into a modeling and validation sample.

All predictors (or independent variables) should be from the origination file. No information from the performance file can be used as predictors because any of the activities have not occurred at the time of origination.

## 2.3 Data Cleaning

1. Remove invalid values
   1. Such as 999 represent for “not available”
2. Make derived variables
   1. Combine multiple variables
   2. Transform a raw variable
   3. Create dummy variable (for both numeric and categoric variables)
   4. Create log of dollar amount
3. Remove outlines
   1. Suppress the outliers for continuous values 🡪 create 20 bins and use 98 percentiles for an upper cap
   2. If a variable behaves the same after a certain threshold, use this threshold to cap the variable (FICO)
4. Fix with missing value
   1. All missing values need to be imputed in order not to maintain an adequate sample size.
      1. (If records with missing values are few, they can be deleted)
5. Remove uncorrelated variables
6. VIF analysis to detect multicollinearity

## 2.4 Modeling Method

Only ‘white-box’ methods can be used due to the following reasons:

-  This model can be used to approve or decline a loan application. As underwriting is subject to regulation from the government, behaviors of all variables should be explainable to meet the requirement for compliance. Also, if a loan application is declined, we need to tell the consumer the reasons for decline, e.g., his/her FICO score is too low and his/her DTI is too high.

-  After the model has been built and approved by the legal/compliance department, the algorithm will be handed to the IT (or operation) team for implementation (also called deployment).

Therefore, logistic regression or decision tree are preferred here. Black-box algorithm cannot be used. (Black-box example: Deep Learning)

## 2.5 Out-of-time Validation

After a model has been deployed, we need to monitor its performance regularly by using new data. For this exercise, we can use loans originated in 2010 or later to check the external validity of the model. When the model has shown significant decay, it is time to refresh or to rebuild the model.

1. Refreshing a model

-  KS is the most commonly used statistics to determine whether a model is strong or

not. However, a very high KS can arouse suspicion.

-  Always test the model on the validation sample. Performance in the validation

sample should be quite comparable to similar to that in the modeling sample.

-  keep the same variable treatment and variable selection.

-  Only re-run the regression with new data. Only the coefficients of regression

equation will change.

-  This is a very quick fix, but the improvement for business is limited.

2. Rebuilding a model

-  Start from scratch and go through all steps from variable screening, variable treatment and regression.

-  The results are usually better than just refreshing a model, but it takes much more time. Also, the model needs to be reviewed by compliance department again.

## 2.6 A discussion on the model score from logistic regression

Although logistic regression is not preferred by many machine learning specialists, it exhibits great strength in the following aspects:

-  The regression equation is very easily explainable.

-  A probability estimate provides a great flexibility to meet changing business needs (see the following three cases).

Case 1: If the company wants to increase the loan volume originated, it can approve more applicants by raising the risk threshold.

Case 2: The company can also assign different interest rates according to applicants’ estimated probability of delinquency. Loans with higher probabilities of delinquency will be charged high interest rates. (The same practice also applies to personal loans, auto loans and credit cards.)

Case 3: When a company sells a portfolio of mortgage loans to investors, segments of loans (usually called “tranches”) with lower risk can ask for a higher price for selling.

# Project 3 – Prepayment Model

A regression/classification model for predicting prepayment. Either logistic regression or decision tree will be used.

## 3.1 Definition of prepayment

When the balance of a mortgage loans becomes zero before it reaches expected maturity, a prepayment occurs. Broadly speaking, there are two types of prepayment:

-  Involuntary prepayment, usually related to foreclosure or bankruptcy. The balance of such a loan is written off and the bank takes a loss.

-  Voluntary prepayment. Common cases are sale of the property and refinance of a mortgage loan.

We will focus on voluntary prepayment in this exercise. (Involuntary prepayment has been covered in Project 2.)

## 3.2 Criteria for defining prepayment in our data

- The balance of a mortgage loans becomes zero in some month, i.e., *CD\_Zero\_BAL*=1.0. - The loan has never been ‘bad’. Otherwise it could be an involuntary prepayment. (Use the logic and code for defining ‘bad’ in Project 2.)

## 3.3 A note of a weak model

The KS statistics for a delinquency model score is around 0.45. For a prepayment model, the KS statistics drops to 0.24. The prepayment model is much weaker than a default model.

FICO is still a good predictor for repayment, especially for rate/term refinance and cash-out refinance. But there are other important causes for early payoffs: marriage, birth of new children, divorce, etc. These will prompt borrowers to sell their current houses. However, our data such as FICO or property type are not suitable predictors for these events.

This is a good example of data limitations. Data are not omnipotent in prediction. If the connection between the data and the question is weak, we can at most build a weak model no matter how high the programming skills we might have.

# Project 4 – Transition of Loan Status

In this exercise, we will build a transition matrix of loan statuses. (Markov Chain model)

## 4.1 Business Purpose

The status of a loan can change. Some examples of status transitions:

-  If a borrower misses a monthly payment, the loan becomes delinquent.

-  A delinquent loan can become good (called “current”) again once the borrower has

repaid all past due payments.

-  A delinquent loan might also deteriorate to a worse status, such as transitioning

from 30-day delinquency to 60-day delinquency.

-  If a loan is refinanced or if a borrower sells his/her property, it results in a prepayment, i.e., current to prepayment.

A transition matrix is a cross-tabulation table that presents the probability of transition from one status to another. It is the foundation for building a Markov Chain model for forecasting risk and for cash flow analysis.

## 4.2 Common account statuses in consumer lending

Use *delq\_sts* to define delinquency status as follows:

0.0: Current or account in good standing, with no past due amount

1.0: Bucket 1 or 30 days delinquent  
2.0: Bucket 2 or 60 days delinquent  
3.0: Bucket 3 or 90 days delinquent

4.0: Bucket 4 or 120 days delinquent

5.0: Bucket 5 or 150 days delinquent

6.0: Bucket 6 or 180 days delinquent

7.0: Charge off or 180+ days delinquent

180+ days delinquent or prepayment usually means the termination of a loan performance. Any performance data after termination should be removed. For example, if a loan has been prepayment in 2011/06, any performance information for 2011/07 and on should be excluded.

## 4.4 Final output of a transition matrix

Table

Description automatically generated

* Linking statuses from two adjacent periods
  + For example, if we use the status of 2012/01 as the begin\_status, the status of 2012/02 is the end\_status. Moving from begin\_status to end\_status is considered a transition.
* If a loan is current (status = 0), its probability of becoming 30-day delinquent (status = 1) is 0.58%. Its probability of getting prepaid is 1.74%.
* If a loan is 30-day delinquent, its probability of going back to a good standing (current) is 34.93%.
* The higher the bucket (or delinquency status), the more likely the loan will deteriorate in status.
* Once a loan become charged-off/prepaid, the bank sometimes is willing to negotiate with the borrower for some arrangement, such as extend the term to reduce the monthly payment. Under these arrangements, the borrower usually gains some incentive to pay off past due amount. Therefore, some loans become current again.

# Project 5 – Survival Analysis

Use survival analysis to examine how characteristics of loans impact loan performance in terms of delinquency and prepayment.

## 5.1 Business Purpose

Survival analysis sheds insights on probability of default (or prepayment) as well as when. We will focus on Kaplan-Meir curves, which is a pool-level analysis. (It does not drill down to loan-level probability.)

## 5.2 Final Output of Survival Analysis

The following is a cross-group comparison of prepaid curves. Loans with higher FICO scores are more likely to get prepaid (mostly because of being refinanced).

Chart, scatter chart

Description automatically generated

Banks use above results to project or forecast loan performance.

Survival analysis can differentiate loan performance in terms of prepayment. This insight has impact on loan valuation. Loans estimated to survive longer are considered to be more valuable.

# Summary of Project Experience

Compared the portfolio mix of mortgage loans originated before and after the financial meltdown in 2008 Q3. Exploratory data analysis showed that new regulatory requirements by the government (for debt-to-income ratio and documentation, etc.) had impact on the improvement of loan quality, changes in loan purposes (purchase/property resale vs. refinance) and other aspects.

* Built a delinquency model by using logistic regression to predict the probability of mortgage loans going ‘bad’ (90+ days in past due). The same modeling method can be used for building underwriting models for credit card, auto loans, personal loans, etc.
* Built a prepayment model to predict the likelihood of loans getting prepaid (because of refinance or property resale).
* Constructed a transition matrix illustrating changes in loan statuses between two adjacent months, such as current to 30-day delinquent, deterioration in delinquency, current to prepaid, etc. The transition matrix is the foundation for building Markov Chain model
* Used Survival Analysis to produce Kaplan-Meier curves to predict prepayment trend by loan age. Differentiation in performance between different segments of loans (by FICO band, property type, etc.) can lend insights on prediction or forecasting of loan performance in the future by loan age.