Project Proposal

Erik Stenberg 3035428720
Huang Zanming 3035182081
Wong Kwan Yi 3035503192
Wong Yick Lun 3035281407

Abstract

An abstract is a self-contained, short, and powerful statement that describes a larger work. Components vary according to discipline. An abstract of a social science or scientific work may contain the scope, purpose, results, and contents of the work. An abstract of a humanities work may contain the thesis, background, and conclusion of the larger work. An abstract is not a review, nor does it evaluate the work being abstracted. s

what we need to do:

Description of data and preliminary data preparation: Source of data: Freddie Mac, The Federal Home Loan Mortage Corp. (FHLMC). A government sponsored corporation that buys mortages and packages them into mortage-backed securities. (How they collect the data?) (What is the data about?)

Description of major attributes (variables):

Monthly Performance Dataset

	Variable Name	Description	Type
1	Loan Sequence Number	Unique identifier assigned to each loan.	Categorical
2	Monthly reporting period	The as-of month for loan information contained in the loan record.	Date
3	Current Actual UPB	The Current Actual UPB reflects the mortgage ending balance as reported by the servicer for the corresponding monthly reporting period.	Numeric
4	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.	Categorical
5	Loan Age	The number of months since the note origination month of the mortgage.	Numeric
6	Remaining months to legal maturity	Unique identifier assigned to each loan.	categorical
7	Repurchase flag	Indicates loans that have been repurchased or made whole (not inclusive of pool-level repurchase settlements).	categorical

	Variable Name	Description	Type
8	Modification Flag	For mortgages with loan modifications, indicates that the loan has been modified.	Categorical
9	Zero Balance Code	A code indicating the reason the loan's balance was reduced to zero.	Categorical
10	Zero Balance Effective Date	The date on which the event triggering the Zero Balance Code took place.	Date
11	Current Interest Rate	Reflects the current interest rate on the mortgage note, taking into account any loan modifications.	Numeric
12	Current Deferred UPB	The current non-interest bearing UPB of the modified mortgage.	Numeric
13	Due Date of Last Paid Installment (DDLPI)	The due date that the loan's scheduled principal and interest is paid through, regardless of when the installment payment was actually made.	Date
14	MI Recoveries	Mortgage Insurance Recoveries are proceeds received by Freddie Mac in the event of credit losses. These proceeds are based on claims under a mortgage insurance policy.	Numeric
15	Net Sales Proceeds	The amount remitted to Freddie Mac resulting from a property disposition once allowable selling expenses have been deducted from the gross sales proceeds of the property.	Categorical,Numeric
16	Non-MI Recoveries	Non-MI Recoveries are proceeds received by Freddie Mac based on repurchase/make whole proceeds, non-sale income such as refunds (tax or insurance), hazard insurance proceeds, rental receipts, positive escrow and/or other miscellaneous credits.	Numeric
17	Expenses	Expenses will include allowable expenses that Freddie Mac bears in the process of acquiring, maintaining and/ or disposing a property (excluding selling expenses, which are subtracted from gross sales proceeds to derive net sales proceeds). This is an aggregation of Legal Costs, Maintenance and Preservation Costs, Taxes and Insurance, and Miscellaneous Expenses	Numeric

	Variable Name	Description	Type
18	Legal Costs	The amount of legal costs associated with the sale of a property (but not included in Net Sale Proceeds). Only applicable when the loan is ended with ends with foreclosure or REO Disposition.	Numeric
19	Maintenance and preservation costs	The amount of maintenance, preservation, and repair costs, including but not limited to property inspection, homeownerâ€™s association, utilities, and REO management, that is associated with the sale of a property (but not included in Net Sale Proceeds).Only applicable when the loan is ended with ends with foreclosure or REO Disposition.	Numeric
20	Taxes and Insurance	The amount of taxes and insurance owed that are associated with the sale of a property (but not included in Net Sale Proceeds)Only applicable when the loan is ended with ends with foreclosure or REO Disposition.	Numeric
21	Miscellanesous Expenses	Miscellaneous expenses associated with the sale of a property (but not included in net sales proceeds)	Numeric
22	Actual Loss Calculation	Actual Loss of the loan	Numeric
23	Modification Cost	The cumulative modification cost amount calculated when Freddie Mac determines such mortgage loan has experienced a rate modification event. Modification Cost is applicable for loans with rate changes only. This amount will be calculated on a monthly basis beginning with the first reporting period a modification event is reported, and disclosed in the last performance record.	Numeric

Total of 26 variables in the original data set. And a total of 24 variables in the monthly data set. (What are the variables exactly, categorical or numerical?) (What are the variables about?)

Quality of the data: - Many empty or N/A values in the monthly performance data set. The Dataset is sparse in nature. 15 out of 23 total columns are mostly empty or not avalible. Some columns such as Repurchase Flag, Modification Flag, Zero Balance Code, and Zero Balance Effective Data are left empty on purpose, as These data are only avalible in the loan termination months. Also columns like Legal Costs, Tax and Insurance, Miscellaneous Expenses are left empty as the data are only applicable when loan ends with foreclosure or REO Disposition. However other columns such as Actual Loss Calculation are left empty without expanation. Due to the lack of data, there will be certain limitations in future analysis.

- Some of the values are coded in a way that we will not know about the exact value of the variable. (Eg: Credit score–below 301:space, above 850:space), poses problems. Some of the variables are unknown

too.

Objectives

Data

Description

Quality of the data

Many of the variables in the data are coded in ways that may not be optimal for analysis. We therefore plan to recode many of the variable into forms that are easier to handle. Below we propose a scheme for such data manipulations

Preparation

Preparation. What changes? Recoding i.e. mutate, filter, remove NA etc.

Study Plan

Referenecs