Freddie Mac Initial Data Analysis

Anonymous CVPR submission

Paper ID ****

1. Introduction

Freddie Mac holds a large portion of the United States of America's home mortgages. As such looking for trends in the values and performance of these loans is crucial. Our dataset consists of data collected at the time loans were issued and a calculated Net Present Value (NPV) that acts as a proxy for the total value of each loan. This paper summarizes a subset of the different variables found in the dataset. The dataset can be found on my GitHub at the following address: https://tinyurl.com/ycpr8lrj

2. Variables

2.1. Credit Score

Figure 1 shows us the summary statistics of Credit Score, and Figure 2 shows us the distribution. This distribution is skewed left with many individuals in the 650 - 775 range.

2.2. First Payment

Figure 3 shows us the frequency table of the year and month the first payment was made for each loan.Figure 4 shows us the distribution as a bar chart. This distribution shows us that most borrowers made their first payment in the first half of 1999 with a few outliers all the way out into 2004.

2.3. First Time Home Buyer Flag

Figure 5 shows us the frequency table of the flag marking if the borrowers for each loan were first time home buyers. This shows that very few (10%) of the loans were from these borrowers. Another notable quality is the amount of unreported data (33%). Figure 6 shows us the distribution as a bar chart. This shows the same data as the summary.

2.4. Loan Maturity Date

Figure 7 shows us the frequency table of the date the loans will mature or end. This shows a large spike in early 2029, which would be 30 years from most of the loans' origination date. This makes sense, and like the previous variable we also have some outliers sprinkled around the

surrounding years. Figure 6 shows us the distribution as a bar chart. This shows the same data as the summary.

2.5. Metropolitan Statistical Area (MSA) Codes

Figure 9 shows us the frequency table of the MSA codes of these homes. The actual codes do not give a lot of pertinent information, but we can not the large count of those marked Other and NA. "Other" likely refers to rural homes not included in an MSA. Figure 10 shows us the distribution as a bar chart. This shows the same data as the summary.

2.6. Mortgage Insurance Percentage

Figure 11 shows us the summary statistics of the mortgage insurance percentages of the homes. It is interesting to note that the majority of homes do not require mortgage insurance and the large amount of underreporting. Figure 12 shows us the distribution as a histogram. This also shows us that the majority of mortgage insurance policies are in round percentages(10%, 25%, etc.).

2.7. Number of Units

Figure 11 shows us the summary statistics of the number of units each loan covers. It is interesting to note that the vast majority of loans cover just one loan, however there are some that go as high as covering four units. It is also worth noting that in this case there are very few NAs. Figure 12 shows us the distribution as a bar chart.

2.8. Occupancy Status

Figure 15 shows us the frequency table of the occupancy status of each home. This shows that the vast majority of these homes are owner occupied, while there is a subset that are either investments or secondary homes. Figure 16 shows us the distribution as a bar chart.

2.9. Combined Loan to Value

Figure 17 shows us the Combined Loan to Value Ratio of each home. This is the initial loan amount plus any lent to refinance divided by the cost of the home. This shows that most loans have a CLTV of around 80%. Figure 18 shows us the distribution as a histogram. This shows us

that many people try to get as close to 80% as they can without going over. Usually going over requires mortgage insurance, which many borrowers prefer not to pay.

2.10. Debt to Income Ratio

Figure 19 shows us the summary statistics of the Debt to Income Ratio of each loan. This is the percent of the borrower's income each month that is paid to the mortgage. Most of the data is in the 20-40% range. Figure 20 shows us the distribution as a histogram. This distribution is surprisingly very normal.

2.11. Unpaid Balance

Figure 21 shows us the summary statistics of the initial unpaid balances of each loan. This is the same thing as the amount loaned. This shows that most loans at this time were around $100,000. Figure 22 shows us the distribution as a histogram. It is interesting that there is a sharp decrease around $250,000. This is because Freddie Mac generally does not accept loans with balances over $250,000.

2.12. Interest Rate

Figure 23 shows us the summary statistics of the interest rate of each loan. At this time most interest rates were around 6-7%. Figure 24 shows us the distribution as a histogram.

2.13. Property Type

Figure 25 shows us the frequency table of the property type of each home. This shows that the majority of the loans are for single family homes, planned units, and condominiums. Figure 26 shows us the distribution as a bar chart.

2.14. Number of Borrowers

Figure 27 shows us the summary statistics of the number of borrowers on each loan. This shows that the majority of loans actually have two borrowers. Figure 28 shows us the distribution as a bar chart.

2.15. Net Present Value (NPV)

Figure 29 shows us the summary statistics of the NPV of each loan. The NPV is our proxy of how profitable a loan ended up being for the lender. This shows that most loans end up netting the lender around $10,000 - $20,000 over what a 30 year government bond would yield. Figure 30 shows us the distribution as a histogram. This shows the high volatility associated with the NPVs of these loans. Additionally how far out the outlying data points are.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
335.0	678.0	719.0	713.1	754.0	849.0	1536

Figure 1. Summary Statistics of Credit Score
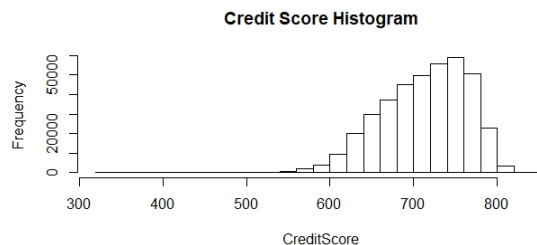
Figure 2. Histogram of Credit Score

Figure 3. Frequency Table of First Payment
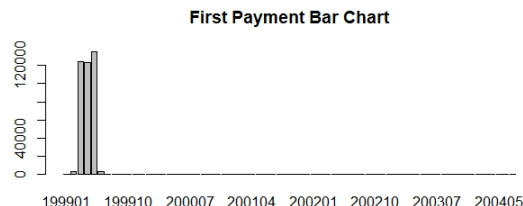
Figure 4. Histogram of Credit Score

N	Y	NA's
243025	34721	113673

Figure 5. Frequency Table of First Time Home Buyers
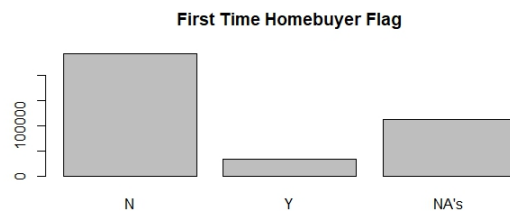
Figure 6. Bar Chart of First Time Home Buyer Flag

Figure 7. Frequency Table of loan maturity date

Figure 8. Bar Chart of loan maturity date

Figure 9. Frequency Table of Metropolitan Statistical Area locations
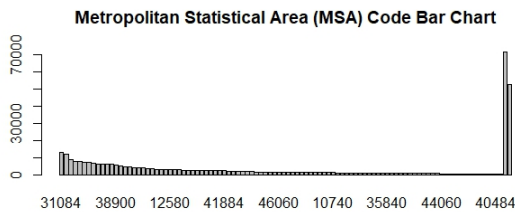
Figure 10. Bar Chart of Metropolitan Statistical Area locations

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
0.00	0.00	0.00	9.22	25.00	55.00	100981

Figure 11. Summary Statistics of Mortgage Insurance Percentage

Figure 12. Bar Chart of Mortgage Insurance Percentage

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
1.000	1.000	1.000	1.029	1.000	4.000	2

Figure 13. Frequency Table of Number of Units

Figure 14. Bar Chart of Number of Units

Figure 15. Frequency Table of Occupancy Status
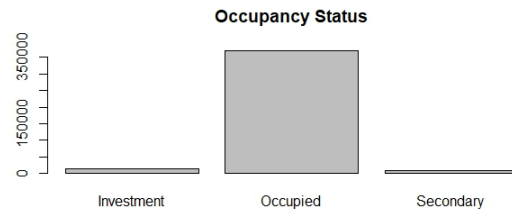
Figure 16. Bar Chart of Occupancy Status

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
6.00	69.00	79.00	75.16	85.00	110.00	5

Figure 17. Summary Statistics of Combined Loan to Value Ratio

3

Combined Loan to Value Ratio
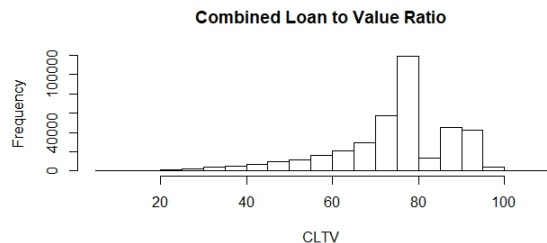
Figure 18. Histogram of Combined Loan to Value Ratio

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
1.00	24.00	31.00	31.53	39.00	65.00	18532

Figure 19. Summary Statistics of Debt to Income Ratio

Debt to Income Ratio
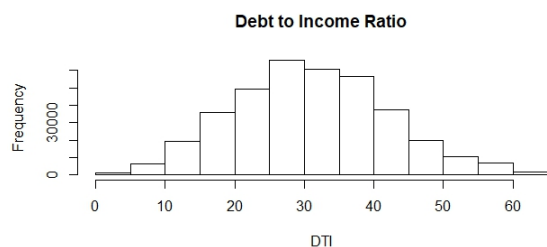
Figure 20. Histogram of Debt to Income Ratio

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
8000	84000	119000	126661	163000	497000

Figure 21. Summary Statistics of Unpaid Balance
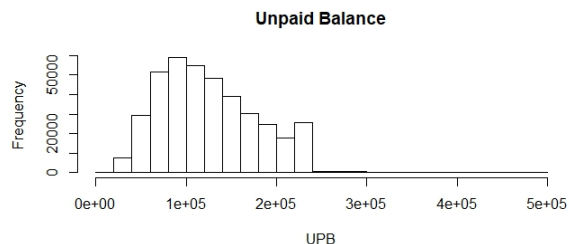
Unpaid Balance

Figure 22. Histogram of Unpaid Balance

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
4.750	6.750	6.875	6.925	7.125	12.350

Figure 23. Summary Statistics of Interest Rate
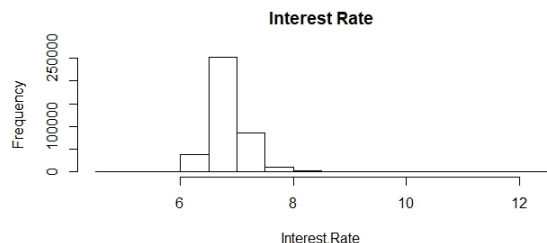
Interest Rate

Figure 24. Histogram of Interest Rate

CO	CP	LH	MH	PU	SF	NA's
23539	204	145	837	37594	329078	22

Figure 25. Summary Statistics of Property Type
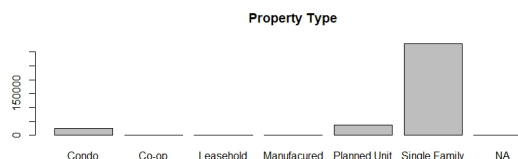
Property Type

Figure 26. Bar Chart of Property Type

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
1.000	1.000	2.000	1.653	2.000	2.000	344

Figure 27. Summary Statistics of Number of Borrowers

Number of Borrowers
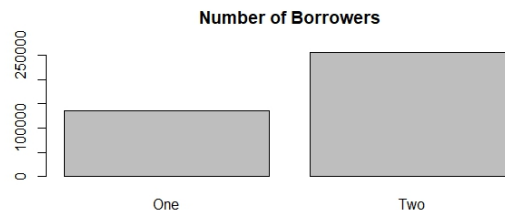
Figure 28. Bar Chart of Number of Borrowers

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-198406	9776	16001	18290	23697	286472

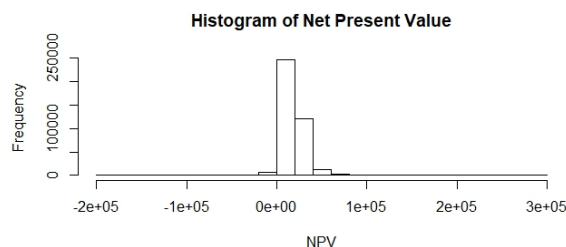Figure 29. Summary Statistics of Net Present Value

Histogram of Net Present Value

Figure 30. Histogram of Net Present Value

4