Freddie Mac Final Report

Thomas Billman

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.

3. Linear Regression

3.1. Simple Linear

We begin our analysis with a simple linear regression using all the variables our dataset provides to us. Two variables have been removed due to a lack of contrast. These columns only contained one value so they are not useful for prediction. Additional categorical variables such as zip code and MSA code were removed due to a high number of insignificant dummy variables being produced. The resulting summary of our first linear regression can be found in Figure 31. This shows that some of our variables have very little impact on predicting NPV. Particularly the number of borrowers associated with the loan. The introduction of a second borrower was only shown to lower the value of the NPV by around 11 cents which is very insignificant. However, were many variables that showed correlations that were very statistically significant. However, our R_a^2 value is very low at only .01026. Additionally the residual plots found in Figure 32 show strong heteroscedasticity in our dataset particularly around the mean. In an attempt to boost predictive power we look to remove some variables.

3.2. Variable Reduction

We begin a simple variable reduction by removing variables that are not statistically significant in our first regression. A summary of this regression can be shown in Figure 33 and the residual plots can be found in Figure 34. This summary shows that our R_a^2 value actually decreased and our residual plots look largely the same. However, this should be useful for the next step where we check for interactions between these variables.

3.3. Interaction Terms

Due to screen size limitations Figure 35 does not show all variables included in the regression, however the R_a^2 value jumped to .01152. Other interesting finds were that variables that were previously significant such as Credit Score now become insignificant as most of the predictive power was wrapped up in interactions with other variables in the dataset. Again, the residuals shows in Figure 36 show similar trends to the previous residual plots. These models show that there is a lot of discretion available for variable selection when creating a linear model for this dataset. Although this is not a comprehensive review, it provides a good baseline to build upon with more advanced statistical techniques.

4. Logistic Regression

We begin our further analysis by predicting whether the NPV of an observation will be positive or negative. The covariates considered include Credit Score, Mortgage Insurance, Debt to Income ratio, among others. This model predicted that every mortgage would have a positive NPV, so it's accuracy measures were as follows:

Sensitivity	1
Specificity	0
Accuracy	0.98125

Additionally the ROC Curve can be found in Figure 37. The area under the curve was .6417, which is not very significant.

5. Linear Discriminant Analysis

We continue by applying Linear Discriminant Analysis. These results were the same as logistic regression, by predicting all rows as positive NPV.

Sensitivity	1
Specificity	0
Accuracy	0.98125

6. Quadratic Discriminant Analysis

Quadratic Discriminant Analysis begins predicting NPVs that are not all positive. The table of results is as follows:

	True -	True +
Predicted -	246	1587
Predicted +	3676	203619

This yields the following sensitivity, specificity, and accuracy:

Sensitivity	0.9922663
Specificity	0.0627231
Accuracy	0.9748336

It is worth noting that although a specificity of .06 seems poor, it is nearly 3 times more accurate than random guessing.

7. KNN Analysis

We continue through to KNN analysis, which produced even better specificity than QDA. This was the resulting confusion matrix:

	True -	True +
Predicted -	601	209
Predicted +	3321	204997

This yields the following sensitivity, specificity, and accuracy:

Sensitivity	0.1532381
Specificity	0.9989815
Accuracy	0.9831204

Of the four methods tested this had the best accuracy and impressive specificity for this dataset.

8. Cross Validated Results

We repeat the above methods using 5-fold cross validation.

8.1. Logistic Regression

Overall average accuracy was 0.981719 and took 15 seconds to run. All sensitivities were 1 and all specificities were 0.

8.2. LDA Analysis

Overall average accuracy was 0.9812459 and took 5 seconds to run. All sensitivities were 1 and all specificities were 0.

8.3. QDA Analysis

Overall average accuracy was 0.9747953 and took 4 seconds to run. All sensitivities were .992 and specificities were between .049 and .075.

8.4. KNN Analysis

Overall average accuracy was 0.9806004 and took 9 minutes to run. All sensitivities were .998 and specificities were between .077 and .087.

9. Recap

These models show that QDA and KNN were better classifiers for our data than logistic or LDA. Additionally, cross validated results show lower accuracies than testing on your training data. However, this is to be expected.

10. KNN Selection of K

We begin this project by testing different values of K for accuracy. For this I decided to use 5-fold cross validation in order to evaluate choices of K by testing accuracy rather than training accuracy. I tested values of 1,3,5,7,9, and 11.

K	Accuracy
1	0.9806052
3	0.9806052
5	0.9806195
7	0.98061
9	0.98061
11	0.98061

This code took 37 minutes to run. This shows that all choices of K amongst this set have essentially identical accuracies. We will continue this analysis using K = 3.

11. 5-Fold Cross Validation

In this section we apply 5-fold cross validation to the Logistic, LDA, QDA, and KNN models previously selected.

Model	Acc	Sens	Spec	Run Time
Logistic	0.9812	1	0	10 s
LDA	0.9812	1	0	4.7 s
QDA	0.9747	0.9922	0.06339	4.2 s
KNN	0.9806	0.9978	0.08076	6.5 m

This shows that Logistic and LDA predicted all observations as positive NPV to achieve the greatest overall accuracy. However, I would pick KNN as a model given that it can predict negative NPVs at a value over four times their relative occurrence in the population. Additionally, the accuracies for these models can be found in Figure 38

12. Leave-One-Out Cross Validation (LOOCV)

We continue this cross validation by utilizing LOOCV. This is the least biased as it does not involve any randomization. However, it is also the most computationally intensive. Due to our large dataset it was necessary to utilize parallel computation techniques and High Performance Computing resources. The results of our LOOCV were as follows:

Model	Acc	Sens	Spec	Time to Run
Logistic	0.9812	1	0	1.1 days
LDA	0.9812	1	0	7.5 hours
QDA	0.9748	0.9922	0.07503	6.6 hours
KNN	0.9805	0.9972	0.09778	26 minutes

It is also worth noting that any boxplots for these accuracies would be degenerate as accuracies in LOOCV can only take values of 1 or 0. Due to this lack of interpretation, they are not included in this paper.These models show that KNN was our best classifier for our data, particularly when utilizing LOOCV.

13. Random Forest

Near the end of this course, we learned about Random Forest classification and regression. This method involves randomly selecting subsets of your covariates and training decision trees. After training 500 trees, a majority vote or average vote is taken.

13.1. Classification

By applying this technique to our dataset and using 5-fold cross validation, we derived the following results:

Fold	Acc	Sens	Spec
1	0.9801	0.9988	0.1700
2	0.9837	0.9987	0.1551
3	0.9822	0.9988	0.1330
4	0.9840	0.9987	0.1489
5	0.9845	0.9992	0.1436

This gives us an average accuracy similar to our other classifiers, however it is very notable that Random Forest classification can pick out loans with negative NPVs at a rate nearly ten times their rate of occurrence in the data set. This makes it the best classifier out of all that we have considered. We can compare Random Forest accuracy to our other classification methods in Figure 39

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

13.2. Regression

By applying this method to estimate NPV as a continuous variable we get the following results:

Fold	Testing R^2
1	.019
2	.0208
3	.0201
4	.0173
5	.0205

This gives us a mean R^2 of .0195. This explains nearly twice the variance of the dataset compared to our reduced variable linear model, however still provides very little insight into our dataset.

14. Conclusion

Although it would be nice to be able to consistently predict what the value of a loan would be at the time of origination, there seems to be too much that goes into whether or not a borrower defaults. An individual's credit score and income can not predict whether the borrower will lose their job some time in the next 30 years. By adding micro and macroeconomic data and collating additional quarters of loan data, we may improve predictive power in the future.

15. References

A. Liaw and M. Wiener (2002). Classication and Regression by randomForest. R News 2(3), 1822.

Breiman, Leo. ”Random forests.” Machine learning 45.1 (2001): 5-32.

Deng, Grace. ”Analyzing the Risk of Mortgage Default” (2016)

Freddie Mac. Single Family Loan-Level Dataset General User Guide: http://www.freddiemac.com/research/pdf/user_guide.pdf

Kutner, Michael H. Applied linear regression models. 4th ed. Michael H. Kutner, Christopher J. Nachtsheim, John Neter.

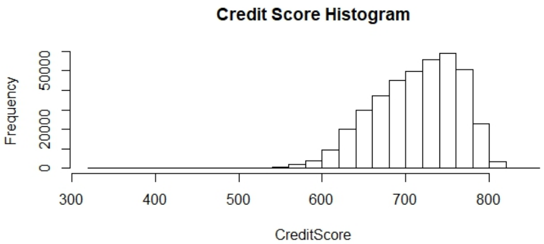


Figure 2. Histogram of Credit Score

199901	199902	199903	199904	199905	199906	199907	199908	199909	199910	199911
10	3015	124593	123525	134865	2789	103	170	384	460	412
199912	200001	200002	200003	200004	200005	200006	200007	200008	200009	200010
131	106	86	63	48	41	21	16	8	10	9
200011	200012	200101	200102	200103	200104	200105	200106	200107	200108	200109
11	8	14	33	35	21	24	25	10	7	7
200110	200111	200112	200201	200202	200203	200204	200205	200206	200207	200208
16	32	30	20	26	9	5	4	16	1	5
200209	200210	200211	200212	200301	200302	200303	200304	200305	200306	200307
6	11	11	7	11	11	22	13	24	11	11
200308	200309	200310	200311	200312	200401	200403	200404	200405	201111	201303
15	23	8	1	3	1	1	1	2	1	1

Figure 3. Frequency Table of First Payment

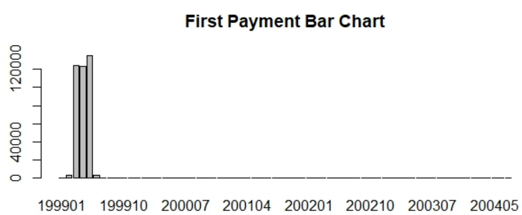


Figure 4. Histogram of Credit Score

N	Y	NA's
243025	34721	113673

Figure 5. Frequency Table of First Time Home Buyers

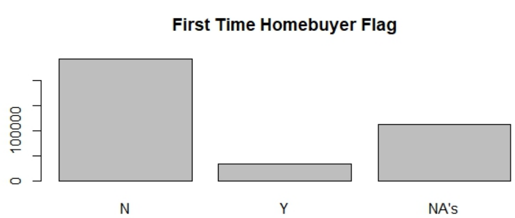


Figure 6. Bar Chart of First Time Home Buyer Flag

202402	202403	202404	202405	202406	202407	202408	202409	202410	202411	202412
1	2	5	7	7	7	13	17	19	13	1
202501	202502	202503	202504	202505	202506	202507	202508	202509	202510	202511
9	39	28	30	9	7	8	10	12	13	13
202512	202601	202602	202603	202604	202605	202606	202607	202608	202609	202610
9	13	52	41	55	14	20	28	37	26	23
202611	202612	202701	202702	202703	202704	202705	202706	202707	202708	202709
23	30	37	103	87	81	20	29	20	25	14
202710	202711	202712	202801	202802	202803	202804	202805	202806	202807	202808
18	18	10	16	65	44	40	21	43	57	103
202809	202810	202811	202812	202901	202902	202903	202904	202905	202906	202907
50	33	29	30	2998	124296	123236	134918	2768	77	124
202908	202909	202910	202911	202912	203001	203002	203003	203004	203005	203006
273	326	267	110	89	81	57	53	38	17	8
203007	203008	203009	203010	203011	203012	203101	203104	203105	203106	203111
6	5	5	7	1	4	3	3	1	3	2
203202	203210	203211	203212	203304	203308	203312	204111	204302		
1	1	1	1	1	1	1	1	1		

Figure 7. Frequency Table of loan maturity date

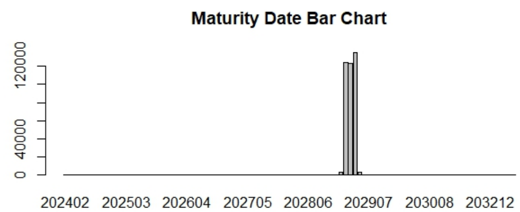


Figure 8. Bar Chart of loan maturity date

31084	16974	12060	38060	47644	42044	19740	42644	41740
13239	12046	8816	8233	7765	7340	7241	6716	6625
47894	33460	36084	38900	35644	40140	17140	40900	26420
6609	6359	6340	6106	5536	4897	4673	4287	4280
19124	41940	41180	45300	37964	29820	12580	35004	20764
4213	4026	3969	3496	3381	3216	3134	3085	3015
19804	28140	14484	16740	22744	18140	41620	15764	36740
2988	2859	2841	2826	2791	2751	2683	2664	2651
41884	26900	35084	31140	13644	37100	48424	39580	12420
2525	2342	2245	2224	2169	2150	2102	1947	1898
34980	33124	17460	46060	23104	40060	33340	38300	29404
1882	1845	1844	1777	1757	1700	1650	1628	1609
42220	13820	47260	39300	45104	27260	10740	25540	17820
1606	1583	1578	1537	1513	1501	1428	1427	1370
19380	14260	24340	37764	14860	36260	15804	36540	14500
1354	1305	1296	1290	1249	1242	1167	1124	1108
35840	39340	49340	41700	44700	35380	15980	35300	32820
1062	1061	1028	991	989	976	974	957	951
36420	17900	24660	44060	21660	23844	39900	41420	24860
942	937	937	901	898	865	864	856	847
37340	48864	22660	45780	30460	42100	40484	42020	(Other)
829	822	815	815	810	809	800	792	71373
NA's								
52849								

Figure 9. Frequency Table of Metropolitan Statistical Area locations

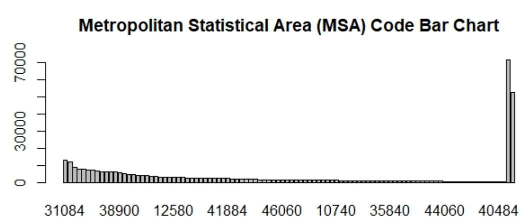


Figure 10. Bar Chart of Metropolitan Statistical Area locations

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

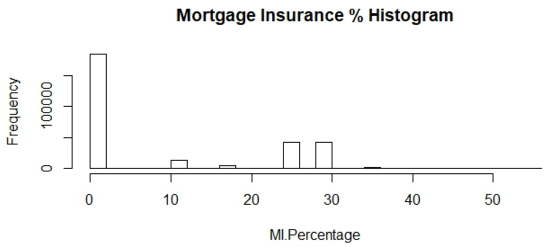


Figure 12. Bar Chart of Mortgage Insurance Percentage

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

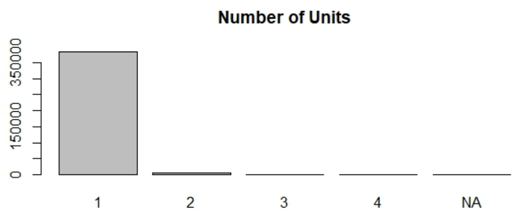


Figure 14. Bar Chart of Number of Units

I	O	S
13071	369213	9135

Figure 15. Frequency Table of Occupancy Status

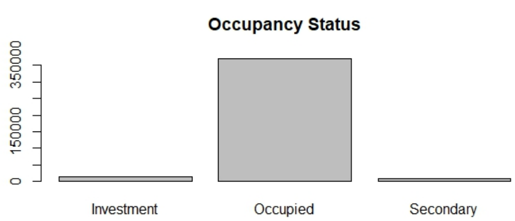


Figure 16. Bar Chart of Occupancy Status

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

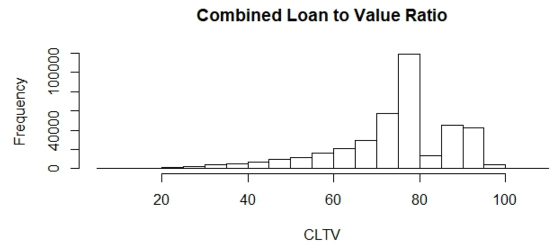


Figure 18. Histogram of Combined Loan to Value Ratio

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

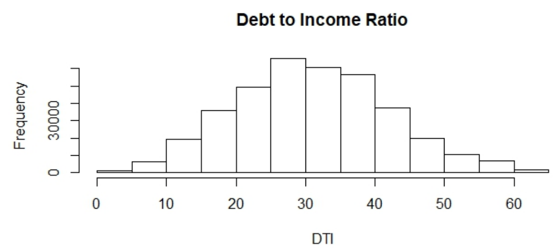


Figure 20. Histogram of Debt to Income Ratio

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

Figure 21. Summary Statistics of Unpaid Balance

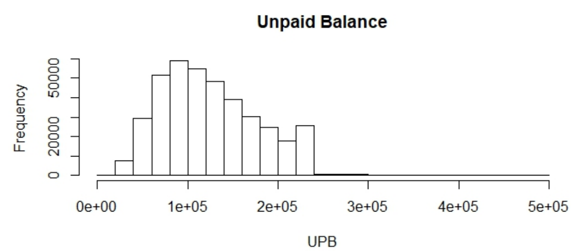


Figure 22. Histogram of Unpaid Balance

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

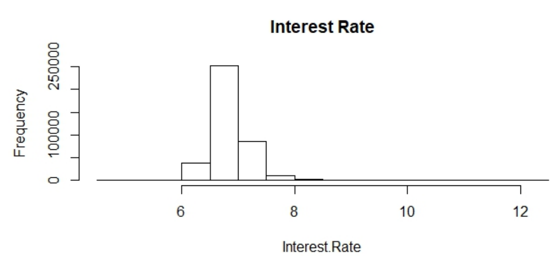


Figure 24. Histogram of Interest Rate

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

Figure 25. Summary Statistics of Property Type

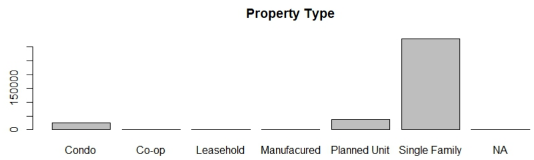


Figure 26. Bar Chart of Property Type

```
 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

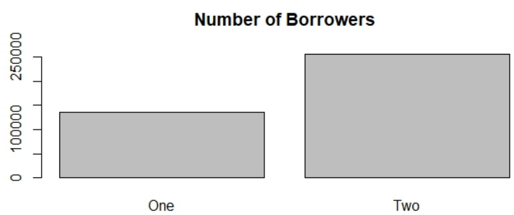


Figure 28. Bar Chart of Number of Borrowers

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

Figure 29. Summary Statistics of Net Present Value

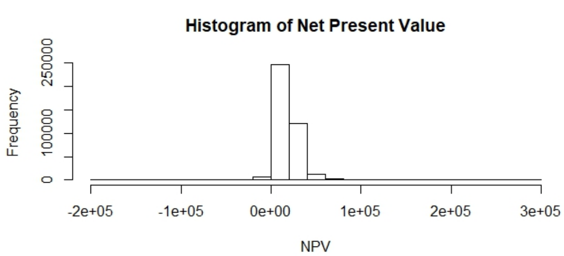


Figure 30. Histogram of Net Present Value

```
Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)        -1.033e+05  7.152e+03 -14.447  < 2e-16 ***
CreditScore        -5.628e+00  7.624e-01  -7.382 1.57e-13 ***
FirstTimeHomebuyerY -7.164e+02  1.152e+02  -6.217 5.09e-10 ***
MI.Percentage       3.518e+01  4.130e+00   8.519  < 2e-16 ***
Number.of.Units    -4.557e+02  2.259e+02  -2.017 0.043692 *
Occupancy.StatusO   2.709e+02  2.699e+02   1.004 0.315556
Occupancy.StatusS   1.485e+02  4.160e+02   0.357 0.721125
CLTV                6.637e+01  6.767e+01   0.981 0.326661
DTI                -2.008e+01  3.452e+00  -5.816 6.04e-09 ***
UPB                 1.557e-02  7.387e-04  21.077  < 2e-16 ***
LTV                -5.854e+01  6.784e+01  -0.863 0.388199
Interest.Rate      -9.437e+02  1.134e+02  -8.318  < 2e-16 ***
ChannelC            2.123e+03  2.005e+03   1.059 0.289607
ChannelR            3.720e+02  1.521e+03   0.245 0.806816
ChannelT            1.512e+03  1.521e+03   0.994 0.320088
PPMY                9.370e+02  2.612e+02   3.587 0.000334 ***
Property.TypeCP    -1.537e+03  2.095e+03  -0.734 0.463091
Property.TypeLH    -6.241e+03  2.346e+03  -2.660 0.007804 **
Property.TypeMH     1.053e+03  9.341e+02   1.127 0.259666
Property.TypePU     1.926e+03  1.812e+02  10.631  < 2e-16 ***
Property.TypeSF     1.001e+03  1.524e+02   6.567 5.14e-11 ***
Loan.PurposeN       4.882e+01  1.162e+02   0.420 0.674292
Loan.PurposeP       1.374e+02  1.216e+02   1.130 0.258366
Original.Term       3.561e+02  1.909e+01  18.651  < 2e-16 ***
Borrower.Num       -1.102e+01  7.962e+01  -0.138 0.889927
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15350 on 178033 degrees of freedom
  (213361 observations deleted due to missingness)
Multiple R-squared:  0.01039,	Adjusted R-squared:  0.01026
F-statistic: 77.92 on 24 and 178033 DF,  p-value: < 2.2e-16
```

Figure 31. Summary from simple linear regression

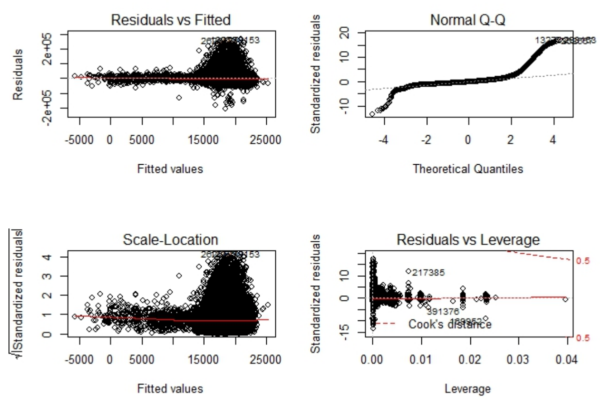


Figure 32. Residual plots of simple linear regression

```
Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)        -1.001e+05  5.977e+03 -16.753  < 2e-16 ***
CreditScore        -5.967e+00  6.852e-01  -8.708  < 2e-16 ***
FirstTimeHomebuyerY -7.824e+02  1.015e+02  -7.707 1.29e-14 ***
MI.Percentage       3.985e+01  2.720e+00  14.654  < 2e-16 ***
Number.of.Units    -5.377e+02  1.828e+02  -2.941  0.00327 **
DTI                -1.847e+01  3.136e+00  -5.889 3.88e-09 ***
UPB                 1.774e-02  6.479e-04  27.379  < 2e-16 ***
Interest.Rate      -1.060e+03  1.014e+02 -10.446  < 2e-16 ***
Property.TypeCP    -1.063e+03  2.075e+03  -0.512  0.60870
Property.TypeLH    -5.435e+03  2.021e+03  -2.690  0.00715 **
Property.TypeMH     1.535e+03  7.189e+02   2.135  0.03273 *
Property.TypePU     1.812e+03  1.702e+02  10.647  < 2e-16 ***
Property.TypeSF     9.330e+02  1.399e+02   6.669 2.57e-11 ***
Original.Term       3.549e+02  1.638e+01  21.662  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15360 on 209114 degrees of freedom
  (182291 observations deleted due to missingness)
Multiple R-squared:  0.009676,	Adjusted R-squared:  0.009615
F-statistic: 157.2 on 13 and 209114 DF,  p-value: < 2.2e-16
```

Figure 33. Summary from regression after variable reduction

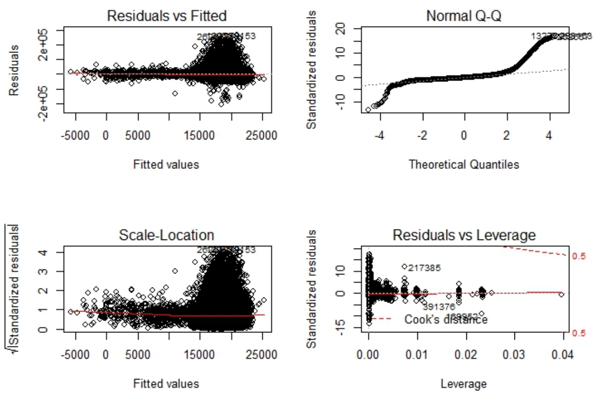
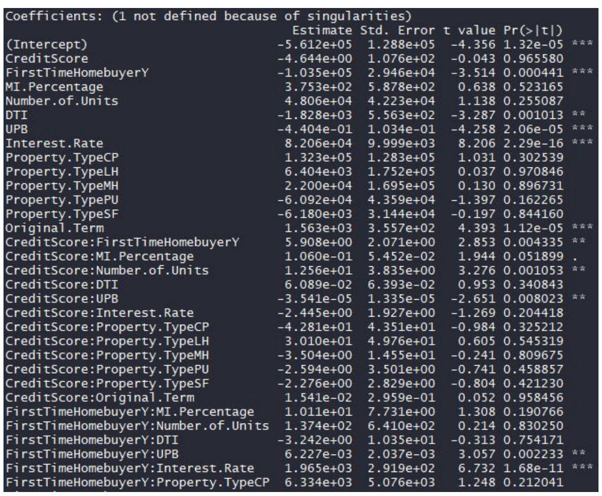


Figure 34. Residual plots after variable reduction



```
Coefficients: (1 not defined because of singularities)
                                      Estimate Std. Error t value Pr(>|t|)
(Intercept)                          -5.612e+05  1.288e+05  -4.356 1.32e-05 ***
CreditScore                          -4.644e+00  1.076e+02  -0.043 0.965580
FirstTimeHomebuyerY                  -1.035e+05  2.946e+04  -3.514 0.000441 ***
MI.Percentage                         3.753e+02  5.878e+02   0.638 0.523165
Number.of.Units                       4.806e+04  4.223e+04   1.138 0.255087
DTI                                  -1.828e+03  5.563e+02  -3.287 0.001013 **
UPB                                  -4.404e-01  1.034e-01  -4.258 2.06e-05 ***
Interest.Rate                         8.206e+04  9.999e+03   8.206 2.29e-16 ***
Property.TypeCP                       1.323e+05  1.283e+05   1.031 0.302539
Property.TypeLH                       6.404e+03  1.752e+05   0.037 0.970846
Property.TypeMH                       2.200e+04  1.695e+05   0.130 0.896731
Property.TypePU                      -6.092e+04  4.359e+04  -1.397 0.162265
Property.TypeSF                      -6.180e+03  3.144e+04  -0.197 0.844160
Original.Term                         1.563e+03  3.557e+02   4.393 1.12e-05 ***
CreditScore:FirstTimeHomebuyerY       5.908e+00  2.071e+00   2.853 0.004335 **
CreditScore:MI.Percentage             1.060e-01  5.452e-02   1.944 0.051899 .
CreditScore:Number.of.Units           1.256e+01  3.835e+00   3.276 0.001053 **
CreditScore:DTI                       6.089e-02  6.393e-02   0.953 0.340843
CreditScore:UPB                      -3.541e-05  1.335e-05  -2.651 0.008023 **
CreditScore:Interest.Rate            -2.445e+00  1.927e+00  -1.269 0.204418
CreditScore:Property.TypeCP          -4.281e+01  4.351e+01  -0.984 0.325212
CreditScore:Property.TypeLH           3.010e+01  4.976e+01   0.605 0.545319
CreditScore:Property.TypeMH          -3.504e+00  1.455e+01  -0.241 0.809675
CreditScore:Property.TypePU          -2.594e+00  3.501e+00  -0.741 0.458857
CreditScore:Property.TypeSF          -2.276e+00  2.829e+00  -0.804 0.421230
CreditScore:Original.Term             1.541e-02  2.959e-01   0.052 0.958456
FirstTimeHomebuyerY:MI.Percentage     1.011e+01  7.731e+00   1.308 0.190766
FirstTimeHomebuyerY:Number.of.Units   1.374e+02  6.410e+02   0.214 0.830250
FirstTimeHomebuyerY:DTI              -3.242e+00  1.035e+01  -0.313 0.754171
FirstTimeHomebuyerY:UPB               6.227e-03  2.037e-03   3.057 0.002233 **
FirstTimeHomebuyerY:Interest.Rate     1.965e+03  2.919e+02   6.732 1.68e-11 ***
FirstTimeHomebuyerY:Property.TypeCP   6.334e+03  5.076e+03   1.248 0.212041
```

Figure 35. Head of the summary of regression with interactions

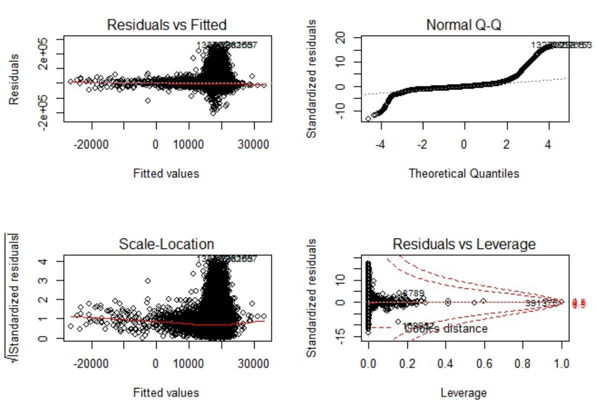


Figure 36. Residual plots with interaction terms

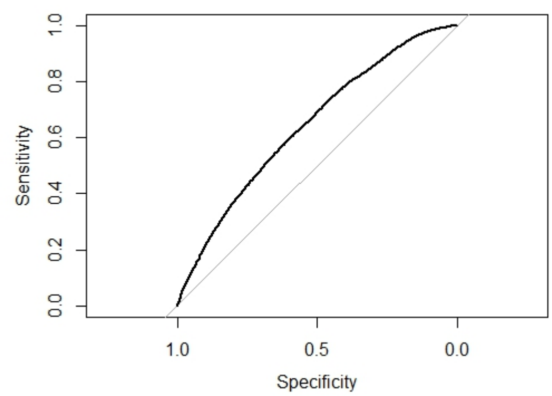


Figure 37. ROC Curve for Logistic Regression

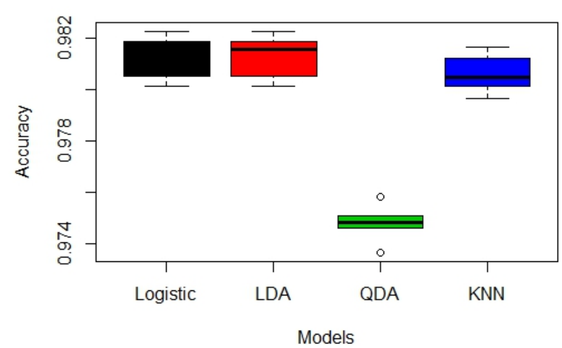


Figure 38. Accuracies for 5-Fold CV

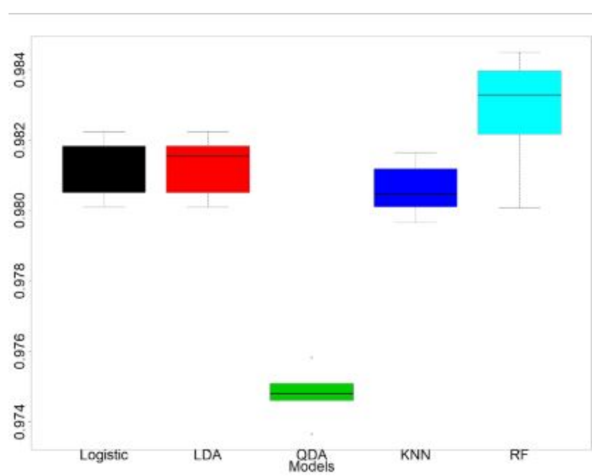


Figure 39. Accuracies for all 5-Fold CV