Predicting defaults on small loans from LendingClub

Predicting loan defaults is a classic application of logistic regression. This report attempts to develop a predictive model using the LendingClub's 2015 dataset without access to FICO scores. Major banks are heavily reliant on FICO scores, despite flaws with the measure, and end up denying credit or charging higher rates to good borrowers when their scores miss a certain standard. Furthermore, credit agencies are notoriously opaque with how they determine FICO scores, which creates a frustrating experience for many borrowers. Therefore, developing a good predictive model without FICO scores could be an attractive selling point for lenders and may alleviate some of the distress that borrowers face.

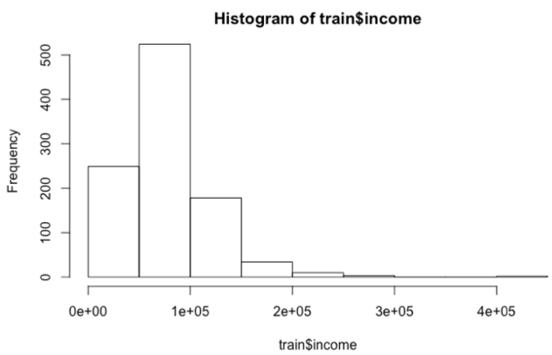
The data contains 421,095 loans that LendingClub originated in 2015. There are three main outcomes for each loan— fully paid, current, charge-off and default. The current loans were ignored. The charge-off loans were grouped with defaulted loans since the borrower is already 4 months late on their payment and the likelihood of collection is slim. A big caveat for this dataset is that some of the metrics, like income, are self-reported and are not always validated. Finally, since the dataset is large, smaller samples were created by random sampling. While the base-rate for defaults in the full dataset is 20%, the smaller sample was set to 50% default with 1,000 observations. Hence, this is a retrospective analysis and β_0 needs to be adjusted for predictions.

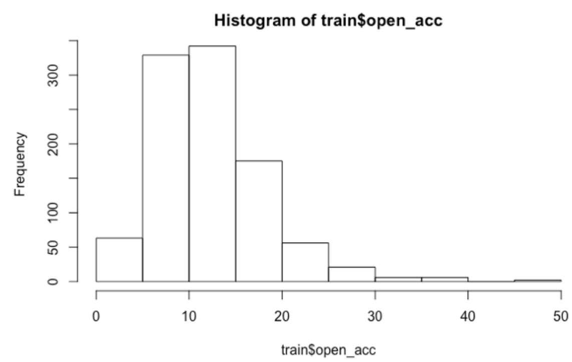
1. Definition of Variables:

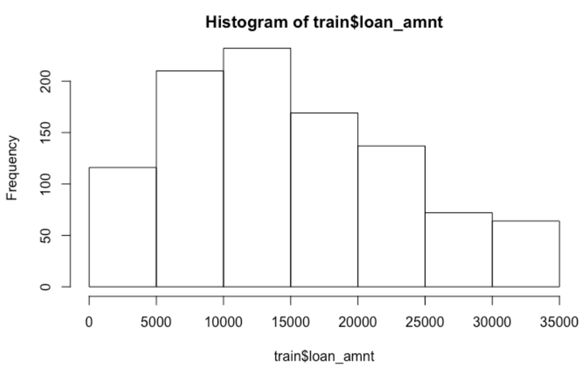
1. *Response Variable Default:* A binary response variable, 1 = Default and 0 = Fully Paid.
2. *ln(Income):* Natural log of the annual income in dollars
3. *ln(Open Accounts)*: Natural log of the number of open credit lines the borrower has.
4. *ln(Loan Amount)*: Natural log of the size of the loan in dollars
5. *Debt-to-Income* Ratio (DTI): The borrower's total debt obligations excluding this loan divided by their income.
6. *Homeownership:* A binary variable, 1 = Homeowner and 0 = Renter. Includes mortgages.
7. *Employment:* A binary variable, 1 = More than 10 years of employment history and 0 = Less than 10 years of employment history.
8. *Bankruptcy:* A binary variable, 1 = Had at least one bankruptcy in the past and 0 = Never filed for bankruptcy.

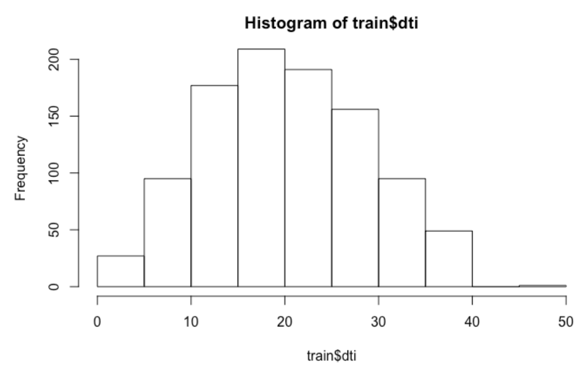
9. *Derogatory:* A binary variable, 1 = Had at least one serious delinquency or late payment in the past and 0 = Has no record of delinquencies or late payments.
10. *Delinquencies (2 Years):* A binary variable, 1 = Had at least one delinquency in the past two years and 0 = Had no delinquencies in the past two years.

<u>2. Initial Plots and Unusual Observations:</u>

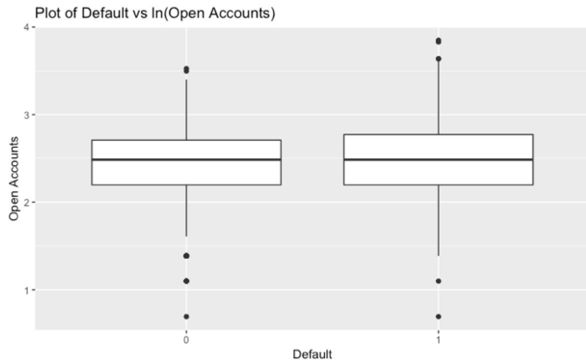


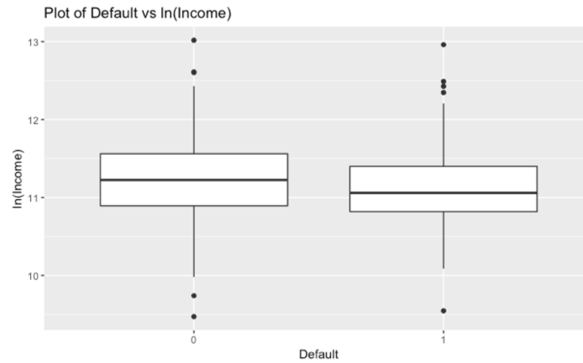


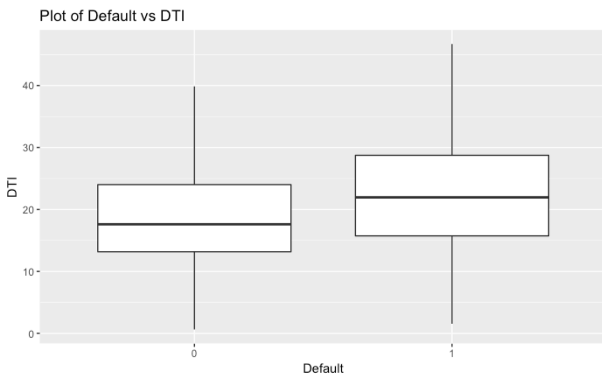


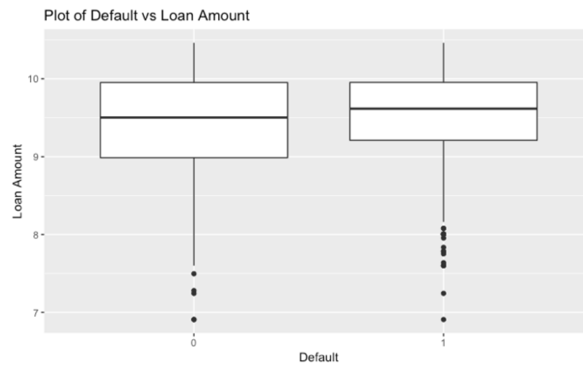


The predictors Income and Open Accounts have severe long right tails, while the predictor Loan Amount has a slightly long right tail. Taking the natural logs of these variables may improve the model. The multiplicative relationship also makes sense. For example, a 1% increase in income may be associated with a β% decrease in the odds; that implies income at lower levels has a greater impact on the odds of default than higher levels. DTI looks fairly normally distributed and the right tail values are probably unusual observations, therefore, it does not need to be transformed. The boxplots and tables (for binary predictors) are displayed below:









	Mortgage	Rent
Paid	327	173
Default	246	254

	No Bankruptcy	Bankruptcy
Paid	436	64
Default	427	73

	Not Derogatory	Derogatory
Paid	407	93
Default	396	104

	No Delinquency	Delinquency
Paid	397	103
Default	394	106

	≤10 years employment	> 10 years employment
Paid	191	309
Default	147	353

For the numerical predictors, it seems like most of them do not provide much information on whether a loan will default. The debt-to-income ratio, however, clearly shows that defaults tend to have higher debt-to-income ratio. The binary variables also seem to have very little relationship with defaults except Homeownership and Employment which have some separation. The output for an initial regression with all predictors is displayed below:

```
Likelihood Ratio Tests (Single term deletions):
             Df  Deviance  AIC     LRT   Pr(>Chi)
<none>             1298.6 1318.6
ln_income     1    1307.9 1325.9  9.2538   0.00235 **
ln_open_acc   1    1299.1 1317.1  0.4851   0.48612
ln_loan_amnt  1    1314.0 1332.0 15.3598 8.886e-05 ***
dti           1    1317.1 1335.1 18.4656 1.730e-05 ***
homeowner     1    1318.4 1336.4 19.7437 8.855e-06 ***
employment    1    1301.6 1319.6  2.9553   0.08560 .
pub_bankrupt  1    1298.6 1316.6  0.0021   0.96310
pub_derog     1    1298.7 1316.7  0.0852   0.77037
delinq_2yrs   1    1300.7 1318.7  2.0279   0.15443

Null Deviance - Deviance
gstat
87.6771 4.773959e-15

Coefficients:        Estimate Std. Error  z value Pr(>|z|)
(Intercept)          0.158611   1.775841   0.089 0.928831
ln_income           -0.518384   0.171811  -3.017 0.002551 **
ln_open_acc          0.113300   0.162826   0.696 0.486534
ln_loan_amnt         0.432334   0.112048   3.858 0.000114 ***
dti                  0.037802   0.008908   4.244 2.20e-05 ***
homeownerRENT        0.625617   0.141594   4.418 9.94e-06 ***
employmentbelow_10   0.244589   0.142336   1.718 0.085726 .
pub_bankrupt        -0.015005   0.324337  -0.046 0.963101
pub_derog            0.081598   0.279641   0.292 0.770442
delinq_2yrs          0.235194   0.165429   1.422 0.155106
```

Odds Ratios:

ln_income	ln_open_acc	ln_loan_amnt	dti	homeownerRENT
0.5954821	1.1199677	1.5408492	1.0385254	1.8693989

employmentbelow_10	pub_bankrupt	pub_derog	delinq_2yrs
1.2770958	0.9851073	1.0850197	1.2651545

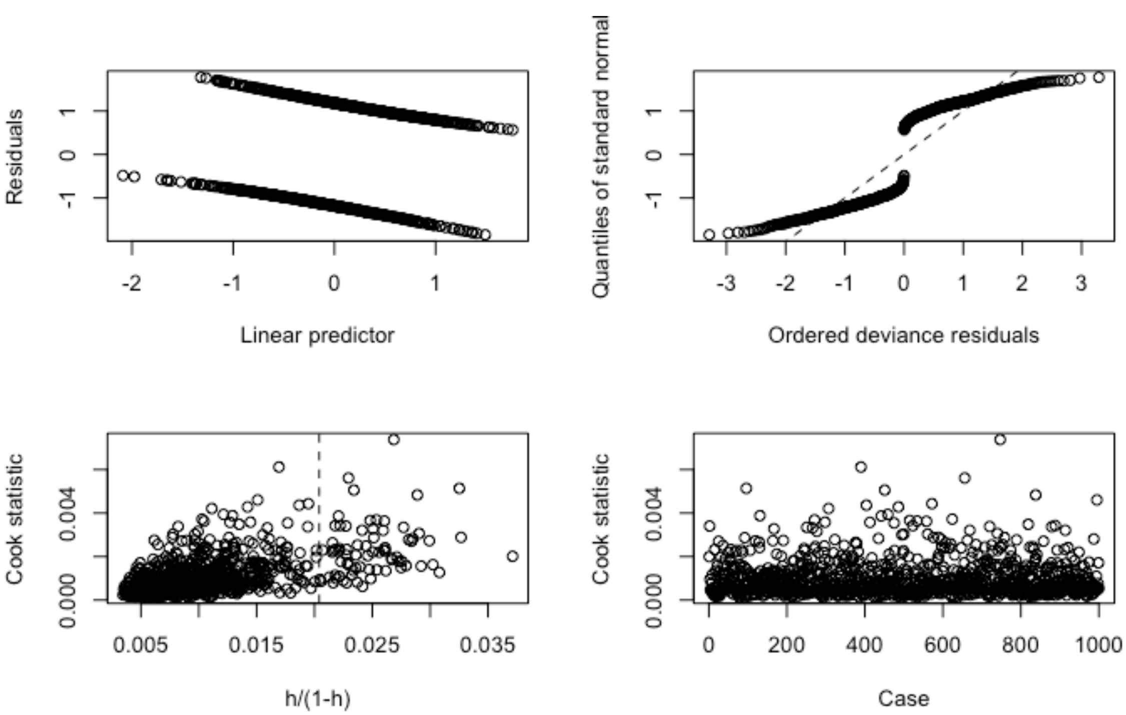
Hosmer and Lemeshow goodness of fit (GOF) test

X-squared = 4.5309, df = 8, p-value = 0.8063

Measure of Association

Somers D = 0.3355440

The initial model looks good in terms of goodness of fit and the statistical significance of the predictors. The Hosmer-Lemeshow p-value is extremely high, which suggests that logistic regression is a good fit for the data. The p-value for overall significance is very low and the predictors ln(Income), ln(Loan Amount), DTI and Homeownership are very significant. Employment is also near significance. Somers D, however, is quite low and suggests the strength of the fit is weak. The diagnostic plots are shown below. The glm.diag.plots() function only displays deviance residuals, but standardized pearson residuals were used for identifying outliers.



There are no outliers based on standardized pearson residuals; the highest standardized pearson residual was only 2.18. The cook's distance on that observation was unusually high, however, so it was removed. There are many clear leverage points. While these are estimates, several of the observations had unusually high hat values and were removed. Unfortunately, the loans are anonymous and there is no practical way to investigate why they are leverage points. This process was repeated two more times and no other outliers were identified, only leverage points. In total, 65 observations (6.50% of the sample) were removed. The full list of removed observations is in Appendix 1. The final regression outputs, tests and diagnostic plots are shown below:

Likelihood Ratio Tests (Single term deletions):

```
            Df Deviance    AIC     LRT  Pr(>Chi)
<none>            1218.8 1236.8
ln_income     1   1227.5 1243.5  8.6560 0.0032599 **
ln_open_acc   1   1219.7 1235.7  0.8527 0.3557969
ln_loan_amnt  1   1230.7 1246.7 11.8994 0.0005615 ***
dti           1   1233.2 1249.2 14.3730 0.0001499 ***
homeowner     1   1236.3 1252.3 17.4591 2.936e-05 ***
employment    1   1221.7 1237.7  2.8992 0.0886248 .
pub_bankrupt  1   1218.9 1234.9  0.0675 0.7950727
delinq_2yrs   1   1220.1 1236.1  1.2571 0.2621974
```

Null Deviance - Deviance

```
gstat
77.36079 1.658673e-13
```

Coefficients:

```
                   Estimate  Std. Error z value Pr(>|z|)
(Intercept)        0.436860   1.857180   0.235 0.814032
ln_income         -0.521968   0.178803  -2.919 0.003509 **
ln_open_acc        0.156640   0.169888   0.922 0.356518
ln_loan_amnt       0.403950   0.118749   3.402 0.000670 ***
dti                0.034514   0.009193   3.754 0.000174 ***
homeownerRENT      0.609004   0.146467   4.158 3.21e-05 ***
```

```
employmentbelow_10   0.251657   0.147839   1.702 0.088712 .
pub_bankrupt         0.050926   0.196105   0.260 0.795104
delinq_2yrs          0.191754   0.171203   1.120 0.262697
```

Odds Ratios:

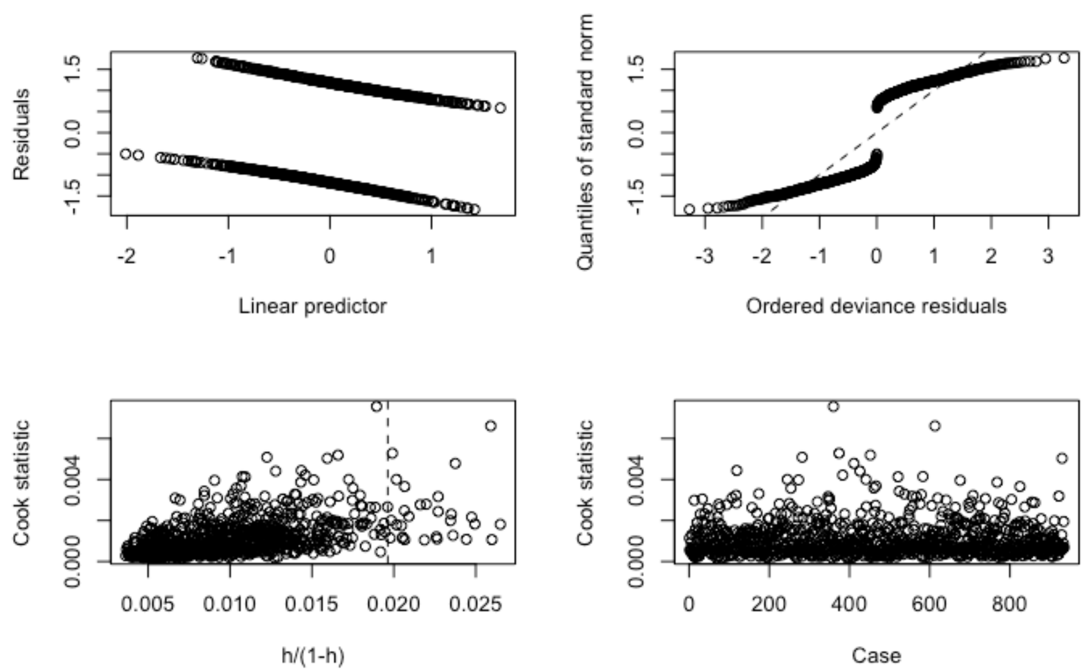
```
ln_income    ln_open_acc    ln_loan_amnt    dti          homeownerRENT
0.5933516    1.1695748      1.4977288       1.0351169    1.8385991
employmentbelow_10   pub_bankrupt      delinq_2yrs
1.2861548            1.0522451          1.2113730
```

Hosmer and Lemeshow goodness of fit (GOF) test

```
X-squared = 4.4835, df = 8, p-value = 0.8111
```

Measure of Association

```
Somers D = 0.3280883
```

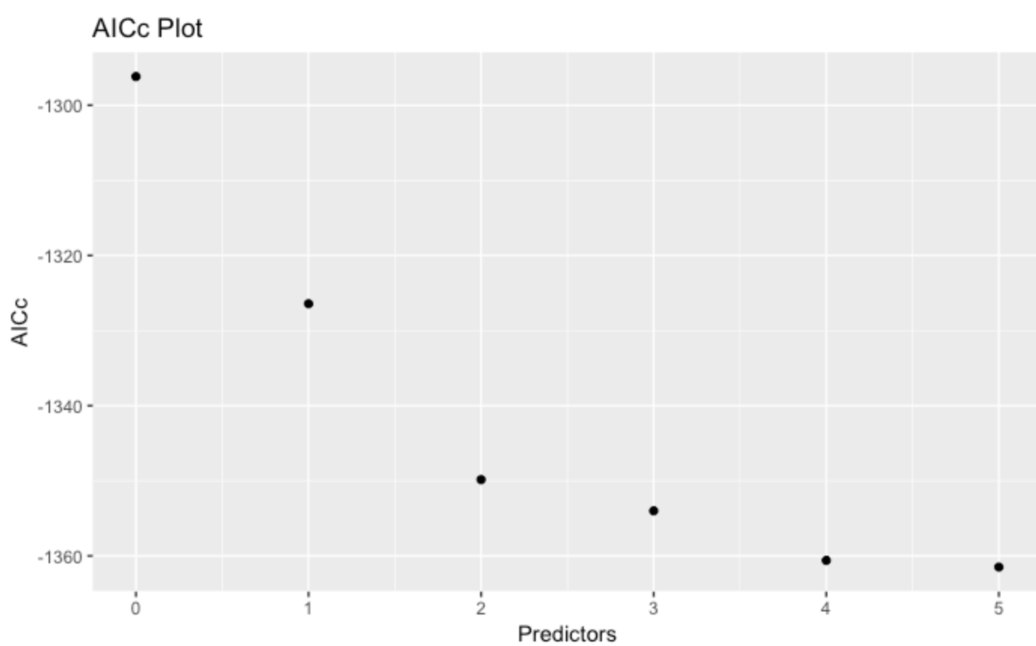


Removing the unusual observations has not changed the interpretation of the model by much. The odd ratios are roughly the same and the predictors ln(Open Accounts), Bankruptcy and Delinquencies (2 Years) are still not significant. The Hosmer-Lemeshow test continues to show a good fit while Somers D continues to show poor separation and a generally weak model.

3. Model Selection:

The predictors ln(Income), ln(Loan Amount), DTI, Homeownership and Employment were included for model selection. The bestglm() output is shown below, as well as a plot of the number of predictors versus AICc:

Predictors	ln_income	ln_loan_amnt	dti	homeowner	employment	AIC
0	FALSE	FALSE	FALSE	FALSE	FALSE	-1296.195
1	FALSE	FALSE	TRUE	FALSE	FALSE	-1326.452
2	FALSE	FALSE	TRUE	TRUE	FALSE	-1349.891
3	FALSE	TRUE	TRUE	TRUE	FALSE	-1354.071
4	TRUE	TRUE	TRUE	TRUE	FALSE	-1360.691
5	TRUE	TRUE	TRUE	TRUE	TRUE	-1361.608



The AICc plot suggests that the 3, 4 and 5-predictor models may be good candidates for the best model. There regression output and tests are displayed below:

Three Predictor Model:

Likelihood Ratio Tests (Single term deletions):

```
             Df Deviance    AIC    LRT  Pr(>Chi)
<none>            1232.6 1240.6
ln_loan_amnt  1   1238.7 1244.7  6.135   0.01325 *
dti           1   1262.8 1268.8 30.211 3.875e-08 ***
homeowner     1   1261.2 1267.2 28.642 8.707e-08 ***
```

Null Deviance - Deviance

```
gstat
63.56823 1.014744e-13
```

```
Coefficients:  Estimate  Std. Error z value Pr(>|z|)
(Intercept)   -3.662318   1.021165  -3.586 0.000335 ***
ln_loan_amnt   0.259332   0.105317   2.462 0.013802 *
dti            0.043879   0.008146   5.387 7.18e-08 ***
homeownerRENT  0.738747   0.139538   5.294 1.20e-07 ***
```

Odds Ratios:

```
ln_loan_amnt       dti          homeownerRENT
1.296064           1.044856      2.093312
```

VIF:

```
ln_loan_amnt       dti          homeownerRENT
1.033583           1.002987      1.033600
```

Hosmer and Lemeshow goodness of fit (GOF) test

```
X-squared = 11.612, df = 8, p-value = 0.1694
```

Measure of Association

```
Somers D = 0.2868307
```

Four Predictor Model:

Likelihood Ratio Tests (Single term deletions):

```
             Df Deviance    AIC     LRT  Pr(>Chi)
<none>            1223.9 1233.9
ln_income     1   1232.6 1240.6  8.6751 0.0032259 **
ln_loan_amnt  1   1235.8 1243.8 11.8157 0.0005873 ***
dti           1   1245.0 1253.0 21.0198 4.546e-06 ***
homeowner     1   1244.0 1252.0 20.0554 7.523e-06 ***
```

Null Deviance - Deviance

```
gstat
72.24334 7.660539e-15
```

Coefficients:

```
               Estimate   Std. Error z value Pr(>|z|)
(Intercept)     0.767624   1.821681   0.421  0.67348
ln_income      -0.497215   0.170173  -2.922  0.00348 **
ln_loan_amnt    0.394359   0.116206   3.394  0.00069 ***
dti             0.037951   0.008392   4.522 6.11e-06 ***
homeownerRENT   0.639669   0.143720   4.451 8.55e-06 ***
```

Odds Ratios:

```
ln_income    ln_loan_amnt   dti           homeownerRENT
0.608222      1.483433      1.038680      1.895853
```

VIF:

```
Ln_income    ln_loan_amnt    dti          homeownerRENT
1.332372     1.238067        1.055322     1.086458
```

Hosmer and Lemeshow goodness of fit (GOF) test

```
X-squared = 4.9161, df = 8, p-value = 0.7665
```

Measure of Association

```
Somers D = 0.3139956
```

Five Predictor Model:

Likelihood Ratio Tests (Single term deletions):

	Df	Deviance	AIC	LRT	Pr(>Chi)	
<none>		1221.1	1233.1			
ln_income	1	1228.4	1238.4	7.3352	0.0067617	**
ln_loan_amnt	1	1232.2	1242.2	11.0966	0.0008649	***
dti	1	1242.0	1252.0	20.8754	4.901e-06	***
homeowner	1	1238.4	1248.4	17.3249	3.150e-05	***
employment	1	1223.9	1233.9	2.8250	0.0928081	.

Null Deviance - Deviance

gstat

75.06832 8.992806e-15

Coefficients:	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	0.329773	1.842576	0.179	0.857958	
ln_income	-0.461861	0.171706	-2.690	0.007149	**
ln_loan_amnt	0.383409	0.116502	3.291	0.000998	***
dti	0.037874	0.008404	4.507	6.58e-06	***
homeownerRENT	0.602823	0.145502	4.143	3.43e-05	***
employmentbelow_10	0.247797	0.147461	1.680	0.092876	.

Odds Ratios:

ln_income	ln_loan_amnt	dti	homeownerRENT	employmentbelow_10
0.630110	1.467279	1.038601	1.827270	1.281199

VIF:

ln_income	ln_loan_amnt	dti	homeownerRENT	employmentbelow_10
1.355021	1.242475	1.056657	1.110108	1.046183

Hosmer and Lemeshow goodness of fit (GOF) test

X-squared = 4.9161, df = 8, p-value = 0.7665

Measure of Association

Somers D = 0.3139956

The classification tables with threshold = 0.5 are displayed below. It includes 64 leverage points that were re-classified using the new models as well as the one outlier.

Model 3	Predicted 0	Predicted 1
Actual 0	318	182
Actual 1	221	279

Model 4	Predicted 0	Predicted 1
Actual 0	326	174
Actual 1	204	296

Model 5	Predicted 0	Predicted 1
Actual 0	318	182
Actual 1	205	295

% Correctly classified by model 3 = 59.70%

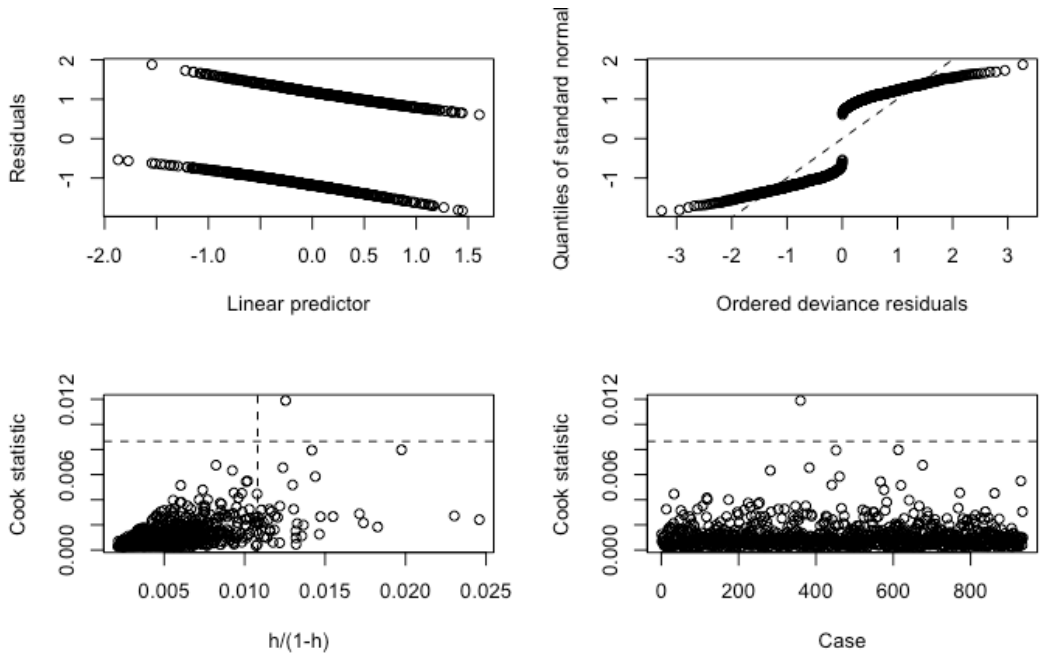
% Correctly classified by model 4 = 62.20%

% Correctly classified by model 5 = 61.30%

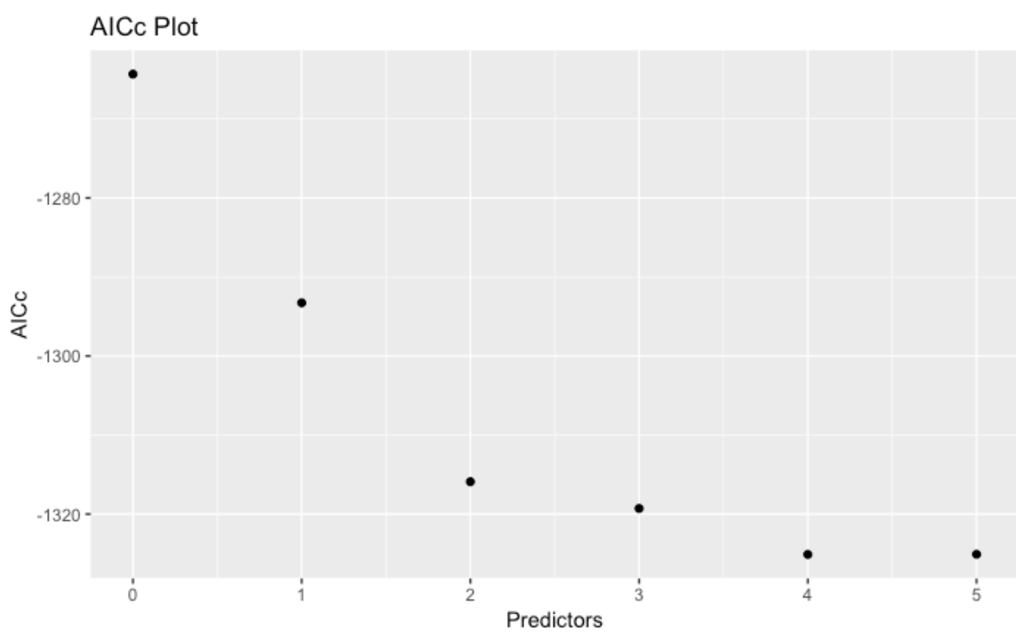
C_{max} = 50.00%

C_{pro} = 62.50% (after x1.25 adjustment for using the same data)

Unfortunately, for all three models, the % correctly classified is below C_{pro}. This was expected since Somers D was low for all the models. However, the models do perform better than C_{max}, so it has some predictive power. Among the three models, model 4 is likely the best. Firstly, the model 3 Hosmer-Lemeshow test is much closer to significance which implies a significantly worse goodness-of-fit. Secondly, model 5 performs too similarly to model 4— a chi-squared test of the deviance(model 4) – deviance(model 5) results in a p-value of 0.09, which suggests model 5 is not significantly different from Model 4. Moreover, model 5 actually predicted fewer correctly than model 4. All three models have very low VIFs, which suggests that they are stable and generalizes well. The diagnostic plots are displayed below. 22 leverage points were removed and the bestglm algorithm was ran again:



Predictors	ln_income	ln_loan_amnt	dti	homeowner	employment	AIC
0	FALSE	FALSE	FALSE	FALSE	FALSE	-1264.340
1	FALSE	FALSE	TRUE	FALSE	FALSE	-1293.295
2	FALSE	FALSE	TRUE	TRUE	FALSE	-1315.952
3	FALSE	TRUE	TRUE	TRUE	FALSE	-1319.358
4	TRUE	TRUE	TRUE	TRUE	FALSE	-1325.200
5	TRUE	TRUE	TRUE	TRUE	TRUE	-1325.209



The leverage points were not influential; the bestglm output is the same and Model 4 continues to be the best model for the reasons stated above. The final regression outputs, tests and diagnostic plots are shown below:

Likelihood Ratio Tests (Single term deletions):

```
             Df Deviance    AIC     LRT  Pr(>Chi)
<none>             1195.6 1205.6
ln_income     1   1203.5 1211.5  7.9066  0.004925 **
ln_loan_amnt  1   1206.3 1214.3 10.6354  0.001109 **
dti           1   1215.0 1223.0 19.3083 1.112e-05 ***
homeowner     1   1214.1 1222.1 18.4295 1.763e-05 ***
```

Null Deviance - Deviance

```
gstat
68.61868 4.440892e-14
```

Coefficients:

```
              Estimate  Std. Error z value Pr(>|z|)
(Intercept)    0.733435   1.902562   0.385  0.69987
ln_income     -0.500193   0.179109  -2.793  0.00523 **
ln_loan_amnt   0.401916   0.124611   3.225  0.00126 **
dti            0.037304   0.008602   4.337 1.45e-05 ***
homeownerRENT  0.626118   0.146692   4.268 1.97e-05 ***
```

Odds Ratios:

```
ln_income     ln_loan_amnt   dti          homeownerRENT
0.6064133     1.4946859      1.0380084     1.8703353
```

VIF:

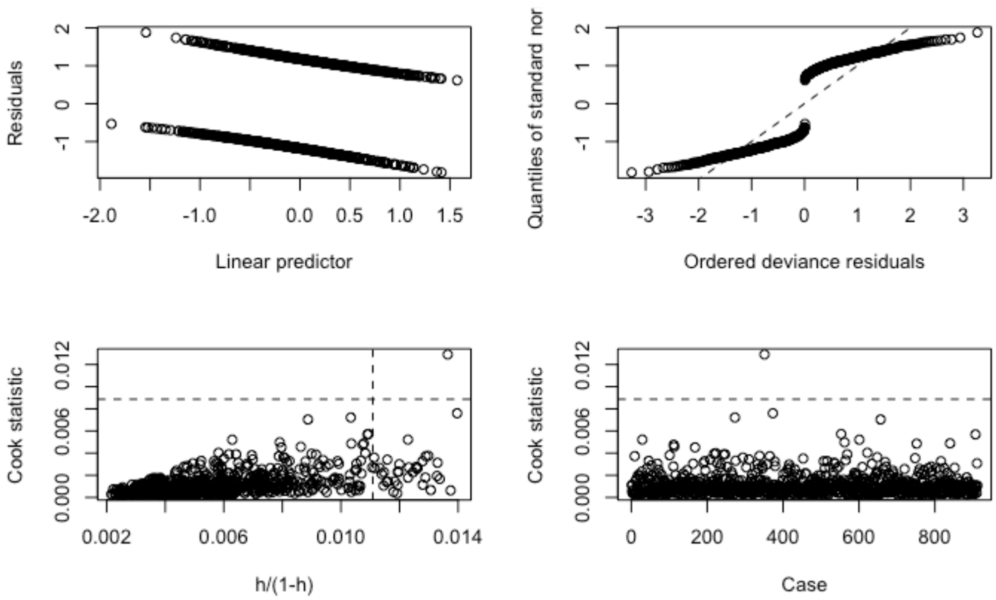
```
ln_income     ln_loan_amnt   dti          homeownerRENT
1.368157      1.257562      1.054530     1.103346
```

Hosmer and Lemeshow goodness of fit (GOF) test

```
X-squared = 4.2848, df = 8, p-value = 0.8208
```

Measure of Association

```
Somers D = 0.3090123
```



4. Prediction and Performance on Test Set

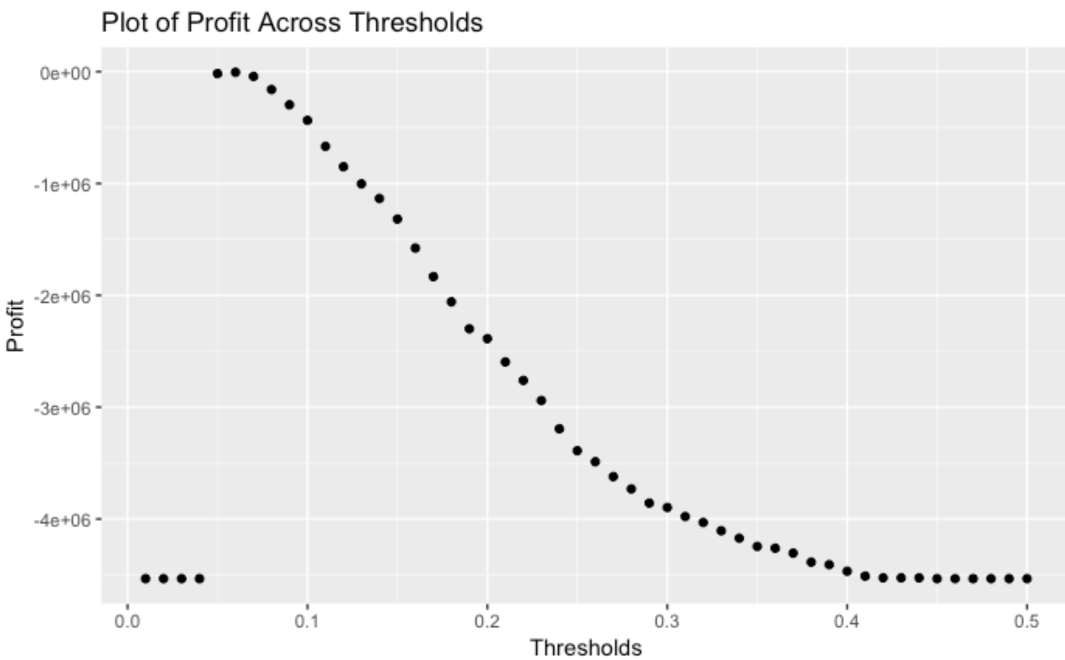
A new dataset with 1,000 observations was constructed to test the model. It excludes any observations selected for modelling and was randomly sampled from the original dataset such that it has 20% defaults— matching the base rate observed in the full 2015 dataset. Since this is a retrospective analysis, β_0 is adjusted assuming a 20% prior probability of default:

Unadjusted p > 0.5	Predicted 0	Predicted 1
Actual 0	113	87
Actual 1	315	485

Adjusted p > 0.2	Predicted 0	Predicted 1
Actual 0	149	53
Actual 1	459	331

$$\tilde{\beta}_0 = \hat{\beta}_0 + \ln\left(\frac{0.2*437}{0.8*563}\right)$$

Against the new dataset, a threshold of p > 0.5 is too high and predicts every observation as a good loan with default = 0. With p > 0.2, the adjusted model correctly classifies only 48% of the data. A better way to decide the threshold is to calculate the estimated profits for different thresholds. In the full dataset, the average loan amount is $14,657; the average annual interest is 12.39% and the average maturity is 3.66 years. Ignoring the time value of money, a rough estimate of the revenue from originating a loan is $14,657 * 0.1239 * 3.66 = $6,646.57. The cost of a default is the principal $14,657. For this cost-benefit analysis, it is assumed that on average LendingClub recovers half the principal when a borrower defaults. Therefore, the estimated profit = (True Positives * $6,646.57) – (False Positives * $14,657 * 0.5). A plot of this estimated profit is displayed below:



The model makes a loss at all thresholds. However, around $p > 0.06$ the loss is minimized. This makes sense as the lending business is really about limiting losses; an established lender like LendingClub will rarely struggle to attract a surplus of applicants, which means their profits are driven by their ability to identify and reject bad loans. Hence, a low threshold that predicts the majority of loans as defaults is well-suited.

6. Interpreting the best model:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$$

$$\ln\left(\frac{p}{1-p}\right) = 0.73 - 0.50\, x_1 + 0.40 x_2 + 0.04 x_3 + 0.63 x_4$$

β_1: ln(Income)

β_2: ln(Loan Amount)

β_3: Debt-to-Income Ratio

β_4: Homeownership binary variable (1= Homeowner, 0 = Renter)

The intercept has no practical meaning. The two logged variables can be interpreted as a multiplicative relationship. For example, a 1% increase in income is associated with a -0.50% decrease in the odds of a default, holding all else constant. The DTI coefficient can be interpreted as an additive-multiplicative relationship: a one unit increase in DTI is associated with multiplying $e^{0.04}$ to the odds, holding all else constant. Finally, the binary variable Homeownership can be interpreted as multiplying by a constant: when a borrower owns a home, that is associated with multiplying $e^{0.63}$ to the odds, holding all else constant. The direction of the coefficients generally make sense. When income is higher, the odds of default is lower; when the debt-to-income ratio is higher, the odds of default is higher. For Loan Amount, larger loans are harder to pay off when a borrower faces financial issues, especially since these loans are not collateralized, which might explain why the coefficient is positive. Finally, for Homeownership, the most likely explanation is that many borrowers that own a home are actively paying off their mortgage, which adds additional pressure to their finances and increases the odds of default.

6. Conclusion:

The best model is a great fit for the data but an extremely poor classifier of defaults. It predicts fewer loans accurately than C_{pro} expects, and in the test set, it operates at a loss across all thresholds. The predictors, however, are highly significant and should be included into models that aim for higher predictive power. Overall, it seems like a deeper credit history is needed to accurately predict defaults, which grading metrics like FICO scores can represent. Until large lenders like banks can collect and aggregate such information at a low cost, it is probably still most effective to use models that rely on FICO scores or other equivalent metrics produced by credit agencies.

Data:

1) “LendingClub Statistics.” *LendingClub Corporation*, https://www.lendingclub.com/info/download-data.action.

Appendix 1:

default	ln_income	ln_open_acc	ln_loan_amnt	dti	homeowner	employment	pub_derog	pub_bankrupt	delinq_2yrs	Std. Pearson Residuals	Hatvalues	Cook's Distance
1	10.30895	2.484907	8.517193	9.16	RENT	below_10	1	0	0	-	0.031639368	-
1	11.11245	2.484907	10.085809	29.07	MORTGAGE	10_or_above	1	0	1	-	0.031502924	-
1	11.19821	2.079442	10.085809	19.76	RENT	10_or_above	1	0	1	-	0.029884756	-
1	11.60824	2.639057	9.952278	22.5	MORTGAGE	below_10	1	0	1	-	0.029352958	-
1	11.86312	3.135494	8.517193	25.58	MORTGAGE	below_10	1	0	1	-	0.029061358	-
1	10.4631	2.833213	9.769956	35.12	MORTGAGE	10_or_above	1	0	0	-	0.028393852	-
1	10.61644	1.94591	8.006368	28.53	RENT	10_or_above	1	0	0	-	0.028233175	-
1	11.77529	2.197225	10.085809	16.22	MORTGAGE	below_10	1	0	1	-	0.028084225	-
1	11.89136	1.94591	10.373491	10.26	RENT	below_10	1	0	0	-	0.027709326	-
1	10.4631	1.609438	8.517193	39.37	MORTGAGE	below_10	1	0	1	-	0.027602682	-
1	11.08214	2.890372	10.085809	27.58	MORTGAGE	10_or_above	1	0	1	-	0.027369813	-
1	11.40756	2.397895	9.242323	31.85	MORTGAGE	10_or_above	1	0	1	-	0.027159195	-
1	11.73607	3.637586	10.463103	18.15	RENT	below_10	1	0	0	-	0.026953922	-
1	10.49127	2.890372	8.853665	21.4	RENT	10_or_above	1	0	0	-	0.026802572	-
0	10.91509	3.178054	8.881836	17.35	RENT	below_10	1	0	0	-	0.026751304	-
0	10.859	2.772589	7.783224	28.38	MORTGAGE	below_10	1	0	1	-	0.026542835	-
0	11.0021	2.564949	9.259131	15.39	MORTGAGE	below_10	1	0	1	-	0.026399753	-
0	10.81978	2.302585	9.729134	5.76	RENT	below_10	1	0	0	-	0.026155895	-
0	10.81978	1.386294	9.862666	19.24	MORTGAGE	below_10	1	0	0	-	0.026015013	-
0	12.0137	3.044522	10.463103	18.89	MORTGAGE	below_10	1	0	0	-	0.025500243	-
0	11.09741	3.258097	10.085809	23.98	RENT	10_or_above	1	0	1	-	0.025459429	-
0	10.92777	2.302585	10.234588	26.18	MORTGAGE	10_or_above	1	0	1	-	0.025438757	-
0	12.07254	2.944439	9.546813	25.43	MORTGAGE	below_10	1	0	0	-	0.025379374	-
0	13.017	1.609438	10.085809	0.7	RENT	below_10	0	0	1	-	0.025342391	-
0	11.84223	2.639057	9.903488	9.24	RENT	10_or_above	1	0	0	-	0.025282726	-
0	10.77896	1.791759	8.517193	17.48	RENT	10_or_above	1	0	0	-	0.024964193	-
0	10.50567	3.135494	6.907755	35.72	MORTGAGE	10_or_above	0	0	0	-	0.024766882	-
0	10.4631	1.94591	7.600902	4.01	RENT	below_10	1	0	0	-	0.024731009	-
1	11.91839	2.890372	9.798127	9.39	RENT	below_10	1	1	0	2.18526	-	0.007380653

default	ln_income	ln_open_acc	ln_loan_amnt	dti	homeowner	employment	pub_derog	pub_bankrupt	delinq_2yrs	Std. Pearson Residuals	Hatvalues	Cook's Distance
1	11.08214	1.94591	9.680344	18.96	MORTGAGE	10_or_above	1	0	0	-	0.039819296	-
1	10.94376	2.079442	9.863967	18.21	MORTGAGE	below_10	1	0	0	-	0.039451677	-
1	10.59663	2.944439	8.987197	25.15	RENT	below_10	1	0	0	-	0.039248465	-
1	11.22524	3.218876	9.952278	21.56	MORTGAGE	below_10	1	0	0	-	0.039014753	-
1	10.29297	2.197225	9.076809	46.71	RENT	below_10	1	0	0	-	0.038694173	-
1	10.30895	2.564949	6.907755	31.08	RENT	10_or_above	0	0	0	-	0.038602512	-
1	11.26766	2.944439	9.769956	24.17	RENT	10_or_above	1	0	0	-	0.038398527	-
1	10.49127	1.386294	7.600902	29.46	MORTGAGE	below_10	0	0	1	-	0.038311172	-
1	10.30895	1.791759	9.217812	18.88	RENT	below_10	1	0	0	-	0.038260804	-
1	10.71442	2.70805	9.169518	37.81	RENT	below_10	1	0	0	-	0.03787546	-
1	11.35041	2.484907	10.385914	25.92	RENT	10_or_above	1	0	0	-	0.037856934	-
1	11.08214	2.079442	8.881836	14.03	RENT	below_10	1	0	0	-	0.036811371	-
1	10.91509	2.079442	8.987197	10.16	RENT	below_10	1	0	0	-	0.036719745	-
1	11.46163	1.791759	9.615805	3.21	RENT	below_10	1	0	0	-	0.036424642	-
1	11.14186	1.94591	10.002201	26.83	RENT	below_10	1	0	0	-	0.036398463	-
1	11.81303	2.890372	10.463103	21.05	MORTGAGE	below_10	1	0	0	-	0.036267863	-
1	10.54534	2.564949	9.680344	18.95	RENT	10_or_above	1	0	0	-	0.036135234	-
1	10.91509	2.302585	9.615805	32.05	MORTGAGE	10_or_above	1	0	0	-	0.036123681	-
1	12.07254	2.079442	10.085809	11.38	MORTGAGE	below_10	1	0	0	-	0.035999018	-
0	10.77896	2.397895	9.680344	33.18	MORTGAGE	10_or_above	1	0	0	-	0.035864715	-
0	11.49272	2.70805	9.392662	24.04	RENT	below_10	1	0	0	-	0.035477926	-
0	10.40426	2.484907	9.21034	21.75	RENT	below_10	1	0	0	-	0.035388227	-
0	11.56172	2.197225	8.699515	14.45	MORTGAGE	10_or_above	1	0	0	-	0.035210327	-
0	11.15625	2.564949	10.065819	13.33	MORTGAGE	below_10	1	0	0	-	0.034319778	-
0	11.15625	2.197225	8.678461	9.53	RENT	below_10	1	0	0	-	0.034138152	-
0	11.22524	2.639057	9.079662	20.37	RENT	10_or_above	1	0	0	-	0.033857598	-
0	11.15625	2.197225	8.720134	17.36	RENT	below_10	1	0	0	-	0.033110846	-
0	11.0021	2.564949	7.718685	16.88	MORTGAGE	below_10	1	0	0	-	0.032610097	-
0	11.38509	1.791759	8.455318	14.21	MORTGAGE	below_10	1	0	0	-	0.031916888	-
0	11.4721	2.564949	10.085809	22.7	MORTGAGE	below_10	1	0	0	-	0.031853207	-
0	10.82775	1.609438	7.824046	12.48	MORTGAGE	below_10	1	0	0	-	0.02860644	-
0	10.859	2.302585	9.959395	26.17	MORTGAGE	below_10	1	0	0	-	0.028216548	-
0	11.08214	2.079442	9.903488	16.49	RENT	10_or_above	1	0	0	-	0.028112942	-
0	12.38839	2.302585	8.794825	7.04	MORTGAGE	below_10	1	0	1	-	0.026396602	-
0	10.79958	2.302585	8.070906	25.91	MORTGAGE	10_or_above	1	0	0	-	0.026189277	-
0	11.91839	1.791759	9.169518	5.54	MORTGAGE	below_10	1	0	0	-	0.025905776	-

default	ln_income	ln_open_acc	ln_loan_amnt	dti	homeowner	employment	pub_derog	pub_bankrupt	delinq_2yrs	Std. Pearson Residuals	Hatvalues	Cook's Distance
1	11.759786	2.890372	9.21034	36.23	RENT	below_10	0	0	1	-	0.023991691	-
1	12.345835	2.564949	10.085809	14.44	RENT	below_10	1	1	0	-	0.022512344	-
1	11.155136	3.044522	8.006368	38.86	RENT	below_10	0	0	1	-	0.019364909	-
1	10.858999	2.302585	7.600902	29.26	RENT	10_or_above	0	0	0	-	0.017937211	-
1	10.085809	2.564949	9.21034	3.75	RENT	below_10	0	0	0	-	0.017087886	-
1	10.266393	2.302585	7.833996	37.91	MORTGAGE	below_10	0	0	0	-	0.016843464	-
1	10.634532	2.772589	7.637716	5.03	RENT	below_10	0	0	0	-	0.015257794	-
1	10.858999	2.197225	7.244228	20.01	MORTGAGE	below_10	0	0	0	-	0.014502293	-
1	12.425208	2.833213	9.798127	5.6	RENT	below_10	0	0	0	-	0.014433007	-
0	11.066638	2.833213	7.244228	34.95	MORTGAGE	below_10	0	0	0	-	0.014187246	-

default	ln_income	ln_open_acc	ln_loan_amnt	dti	homeowner	employment	pub_derog	pub_bankrupt	delinq_2yrs	Std. Pearson Residuals	Hatvalues	Cook's Distance
0	1	12.601487	2.484907	12.69	RENT	10_or_above	0	0	1	-	0.01397934	-
0	0	12.100712	2.302585	12.66	MORTGAGE	10_or_above	0	0	0	-	0.013471157	-
0	0	12.301383	2.079442	4.76	RENT	below_10	1	1	0	-	0.013370181	-
0	1	9.472705	1.791759	34.07	MORTGAGE	below_10	0	0	1	-	0.013094167	-
0	0	10.819778	2.197225	20.38	RENT	10_or_above	0	0	0	-	0.013023199	-
0	0	12.230765	2.302585	5.71	RENT	10_or_above	0	0	0	-	0.013020358	-
0	0	11.362103	2.302585	20.64	RENT	10_or_above	0	0	0	-	0.013010575	-
0	0	10.463103	1.791759	18.59	MORTGAGE	below_10	0	0	0	-	0.012879664	-
0	0	11.647456	1.94591	6.34	RENT	below_10	0	0	0	-	0.012816212	-