

# Rui\_Peng\_Week\_7.R

raypeng

2025-04-26

*#Step 1: Use the Decision Tree / Random Forest / Decision Tree / Regression code from Week 6 as a Start*

```
library( rpart )
library( rpart.plot )
library( ROCR )
library( MASS )
library( randomForest )
```

```
## randomForest 4.7-1.2
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
library( Rtsne )
```

```
SEED = 1
```

```
set.seed( SEED )
```

```
TARGET = "TARGET_BAD_FLAG"
```

```
PATH = "/Users/raypeng/Documents/IS 5213 Data science and big data/HMEQ_Scrubbed"
```

```
FILE_NAME = "HMEQ_Scrubbed.csv"
```

```
INFILE = paste(PATH, FILE_NAME, sep = "/")
```

```
setwd(PATH)
```

```
df = read.csv(FILE_NAME)
```

*#Step 2: PCA Analysis*

*#Use only the input variables. Do not use either of the target variables.*

```
df_pca = df
```

```
df_pca$TARGET_BAD_FLAG = NULL
```

```
df_pca$TARGET_LOSS_AMT = NULL
```

*#Use only the continuous variables. Do not use any of the flag variables.*

*#Do a Principal Component Analysis (PCA) on the continuous variables.*

```
pca2 = prcomp(df_pca[,c(1,2,4,6,8,10,12,14,16,18)] ,center=TRUE, scale=TRUE)
```

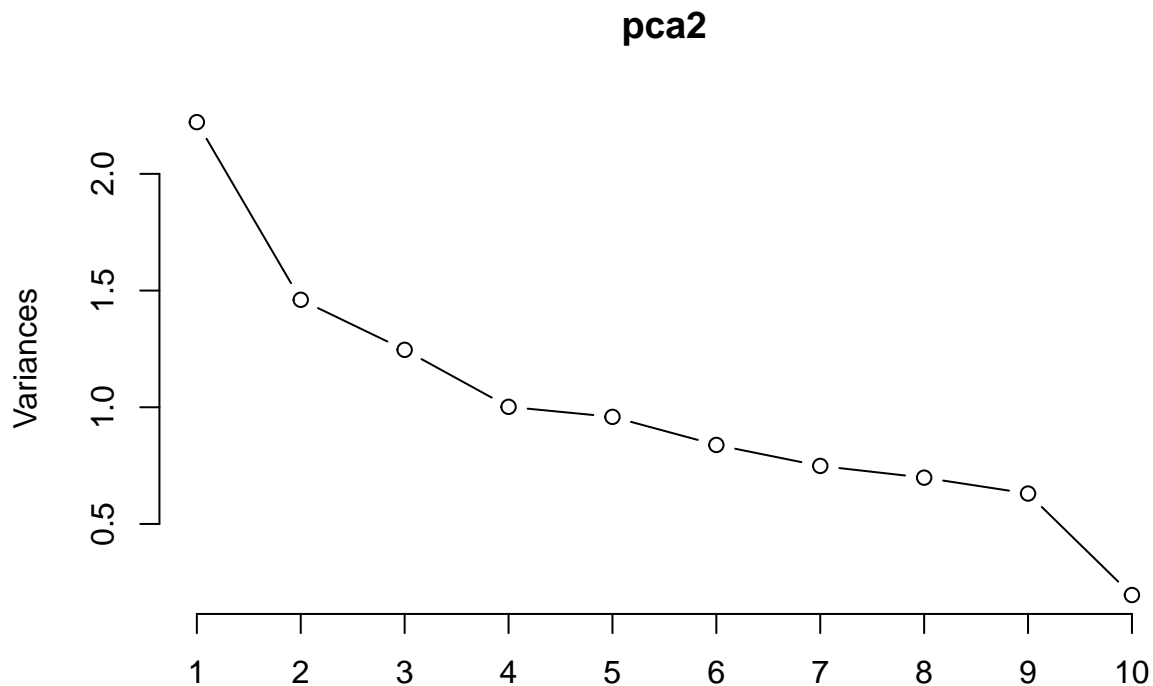
```
summary(pca2)
```

```
## Importance of components:
```

```
##              PC1    PC2    PC3    PC4    PC5    PC6    PC7
## Standard deviation  1.4905 1.2085 1.1163 1.0009 0.97918 0.91572 0.86520
## Proportion of Variance 0.2222 0.1461 0.1246 0.1002 0.09588 0.08385 0.07486
## Cumulative Proportion 0.2222 0.3682 0.4928 0.5930 0.68889 0.77274 0.84760
##              PC8    PC9    PC10
## Standard deviation  0.83568 0.79387 0.44203
## Proportion of Variance 0.06984 0.06302 0.01954
## Cumulative Proportion 0.91744 0.98046 1.00000
```

```
#Display the Scree Plot of the PCA analysis.
```

```
plot(pca2, type = "l")
```



```
#Using the Scree Plot, determine how many Principal Components you wish to use.
```

```
#I decide to use 4 PCs with PC4 has a standard deviation above 1.
```

```
#This means the first 4 PCs contain most of the information of the imputed dataset.
```

```
#Print the weights of the Principal Components. Use the weights to tell a story on what the Principal C
```

```
print(pca2)
```

```
## Standard deviations (1, ..., p=10):
```

```
## [1] 1.4905331 1.2085158 1.1162779 1.0009159 0.9791769 0.9157223 0.8652018
```

```
## [8] 0.8356779 0.7938727 0.4420313
```

```
##
```

```
## Rotation (n x k) = (10 x 10):
```

```
##              PC1          PC2          PC3          PC4          PC5
```

```
## LOAN      0.31425517 -0.104598465  0.05295727 -0.53771580  0.419827766
## IMP_MORTDUE 0.57476524  0.001640244  0.19466925  0.22040956  0.098110092
## IMP_VALUE  0.58633796 -0.078601929  0.15458274  0.10039762  0.186199448
## IMP_YOJ    0.03435411 -0.260508848 -0.53332969 -0.51783332  0.018198368
## IMP_DEROG -0.03192356  0.555370079 -0.18904164  0.05750818  0.383539825
## IMP_DELIHQ 0.02493396  0.459862520 -0.43274648  0.17040014  0.147886344
## IMP_CLAGE  0.23297961 -0.242635491 -0.53339115  0.09355940 -0.172544558
## IMP_NINQ   0.04386120  0.461243097  0.14867761 -0.46490354  0.006637185
## IMP_CLNO   0.36929326  0.217794624 -0.29861623  0.14890732 -0.349807190
## IMP_DEBTINC 0.17802968  0.275775654  0.17937104 -0.32348860 -0.676730173
##          PC6          PC7          PC8          PC9          PC10
## LOAN      -0.213495057 -0.44043005 -0.35336106  0.21745400 -0.101959313
## IMP_MORTDUE 0.006342209  0.23976652  0.13640577 -0.13289245 -0.692636273
## IMP_VALUE  -0.050919084  0.18362254  0.13777813 -0.15610907  0.708322628
## IMP_YOJ    -0.106699879  0.47595145  0.36416419  0.06189337 -0.060587595
## IMP_DEROG  -0.040965995 -0.34472757  0.61903640  0.02468630 -0.008212326
## IMP_DELIHQ -0.397005325  0.27884949 -0.51269531 -0.22668592  0.010743180
## IMP_CLAGE  0.314965078 -0.46591623 -0.05424561 -0.49387254 -0.026149737
## IMP_NINQ   0.655283652  0.20913489 -0.15817225 -0.22004875  0.021740265
## IMP_CLNO   0.225736964 -0.03028645 -0.09454951  0.72011437  0.055669736
## IMP_DEBTINC -0.448947365 -0.17308851  0.14821557 -0.20966727  0.005391156
```

*#PC1 is more about MORTDUE, VALUE and CLNO. I call this "Financial Capacity".*

*#PC2 is more about DEROG, NINQ and DELINQ. I call this "Credit Risk".*

*#PC3 is more about YOJ, CLAGE and DELINQ. I call this "Financial Responsibility".*

*#PC4 is more about LOAN, YOJ and NINQ. I call this "Borrowing Intensity".*

*#Perform a scatter plot using the first two Principal Components. Color the scatter plot dots using the*  
`df_new = predict( pca2, df_pca )`

```
df_flags = df
df_flags$PC1 = df_new[, "PC1"]
df_flags$PC2 = df_new[, "PC2"]
```

*#If you believe the graph is too cluttered, you are free to do a random sample of the data to make it m*

```
df_flags$RAND1 = sample(100, size = nrow(df_flags), replace = TRUE)
df_flags$RAND2 = sample(100, size = nrow(df_flags), replace = TRUE)
```

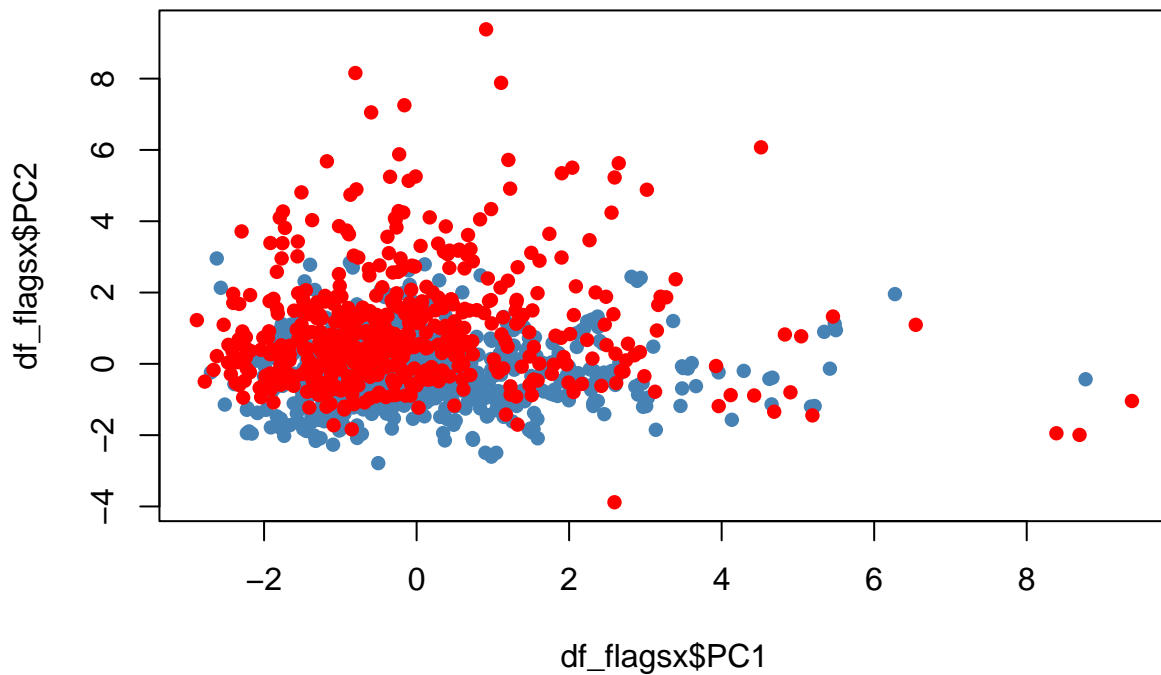
```
df_flags0 = df_flags[ which(df_flags$TARGET_BAD_FLAG == 0), ]
df_flags1 = df_flags[ which(df_flags$TARGET_BAD_FLAG == 1), ]
```

```
df_flags0 = df_flags0[ df_flags0$RAND1 < 30, ]
df_flags1 = df_flags1[ df_flags1$RAND1 < 90, ]
```

```
df_flagsx = rbind( df_flags0, df_flags1 )
df_flagsx = df_flagsx[ df_flagsx$RAND2 < 50, ]
```

*#One color will represent "defaults" and the other color will represent "non defaults".*

```
colors <- c("steelblue", "red")
colors <- colors[df_flagsx$TARGET_BAD_FLAG + 1]
plot( df_flagsx$PC1, df_flagsx$PC2, col = colors, pch = 16 )
```



*#Comment on whether you consider the first two Principal Components to be predictive.  
 #I think they are predictive as we can see from the plot:  
 #Defaulted red flags are generally above the blue safe ones.  
 #Higher PC2 means higher credit risk and easier to default.*

*#However, I feel PC1 is not doing so good on distinguish the two sides.  
 #One possible reason is that people with different financial capacities  
 #may default on loans. Rich people may also overborrow and then default.*

### *#Step 3: tSNE Analysis*

*#Use only the input variables. Do not use either of the target variables.*  
`dfu = df`  
`dfu$TARGET_LOSS_AMT = NULL`  
`dfu = unique(dfu)`  
`head( dfu )`

##	TARGET_BAD_FLAG	LOAN	IMP_MORTDUE	M_MORTDUE	IMP_VALUE	M_VALUE	IMP_YOJ	M_YOJ
## 1	1	1100	25860	0	39025	0	10.5	0
## 2	1	1300	70053	0	68400	0	7.0	0
## 3	1	1500	13500	0	16700	0	4.0	0
## 4	1	1500	65000	1	89000	1	7.0	1
## 5	0	1700	97800	0	112000	0	3.0	0
## 6	1	1700	30548	0	40320	0	9.0	0

```

##      IMP_DEROG M_DEROG IMP_DELIHQ M_DELIHQ IMP_CLAGE M_CLAGE IMP_NINQ M_NINQ
## 1          0      0          0          0 94.36667      0          1      0
## 2          0      0          2          0 121.83333      0          0      0
## 3          0      0          0          0 149.46667      0          1      0
## 4          1      1          1          1 174.00000      1          1      1
## 5          0      0          0          0 93.33333      0          0      0
## 6          0      0          0          0 101.46600      0          1      0
##      IMP_CLNO M_CLNO IMP_DEBTINC M_DEBTINC FLAG.Job.Mgr FLAG.Job.Office
## 1          9      0    35.00000      1          0          0
## 2         14      0    35.00000      1          0          0
## 3         10      0    35.00000      1          0          0
## 4         20      1    35.00000      1          0          0
## 5         14      0    35.00000      1          0          1
## 6          8      0    37.11361      0          0          0
##      FLAG.Job.Other FLAG.Job.ProfExe FLAG.Job.Sales FLAG.Job.Self
## 1          1          0          0          0
## 2          1          0          0          0
## 3          1          0          0          0
## 4          0          0          0          0
## 5          0          0          0          0
## 6          1          0          0          0
##      FLAG.Reason.DebtCon FLAG.Reason.HomeImp
## 1          0          1
## 2          0          1
## 3          0          1
## 4          0          0
## 5          0          1
## 6          0          1

```

```

#Use only the continuous variables. Do not use any of the flag variables.
#Do a tSNE analysis on the data. Set the dimensions to 2.
#Run two tSNE analysis for Perplexity=30.
theTSNE = Rtsne( dfu[,c(2,3,5,7,9,11,13,15,17,19)], dims = 2,
                 perplexity = 30, verbose = TRUE, max_iter = 500)

```

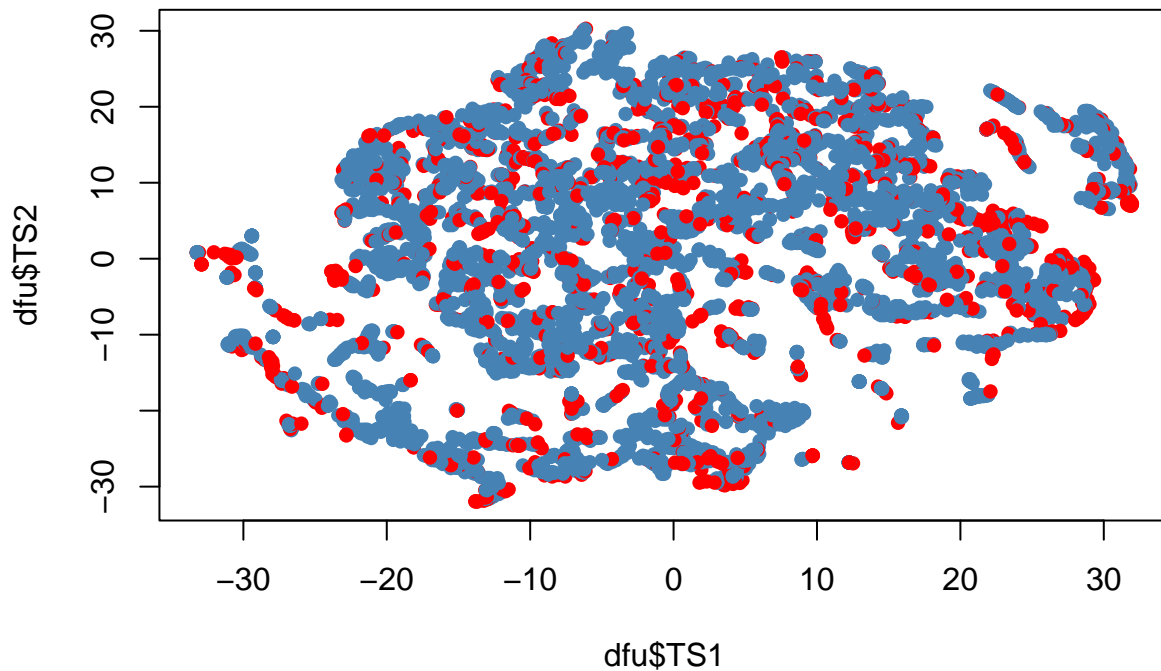
```

## Performing PCA
## Read the 5960 x 10 data matrix successfully!
## Using no_dims = 2, perplexity = 30.000000, and theta = 0.500000
## Computing input similarities...
## Building tree...
## Done in 0.41 seconds (sparsity = 0.019121)!
## Learning embedding...
## Iteration 50: error is 92.480472 (50 iterations in 0.91 seconds)
## Iteration 100: error is 74.158377 (50 iterations in 0.86 seconds)
## Iteration 150: error is 70.540698 (50 iterations in 0.77 seconds)
## Iteration 200: error is 69.257199 (50 iterations in 0.76 seconds)
## Iteration 250: error is 68.610904 (50 iterations in 0.77 seconds)
## Iteration 300: error is 2.090753 (50 iterations in 0.71 seconds)
## Iteration 350: error is 1.659516 (50 iterations in 0.67 seconds)
## Iteration 400: error is 1.423800 (50 iterations in 0.68 seconds)
## Iteration 450: error is 1.277760 (50 iterations in 0.68 seconds)
## Iteration 500: error is 1.179853 (50 iterations in 0.69 seconds)
## Fitting performed in 7.50 seconds.

```

```
dfu$TS1 = theTSNE$Y[,1]
dfu$TS2 = theTSNE$Y[,2]

#Color the scatter plot dots using the Target Flag.
#One color will represent "defaults" and the other color will represent "non defaults".
colors <- c("steelblue", "red")
colors <- colors[dfu$TARGET_BAD_FLAG + 1]
plot( dfu$TS1, dfu$TS2, col = colors, pch = 16 )
```



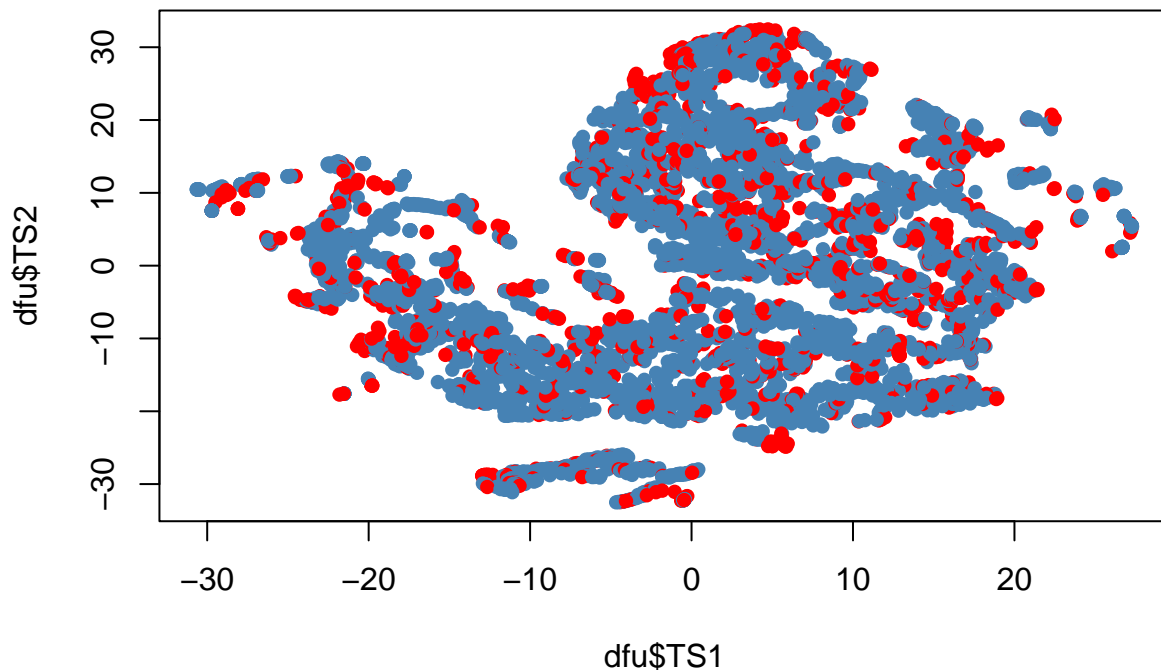
```
#Comment on whether you consider the tSNE values to be predictive.
#This TSNE ananlysis with perplexity of 30 is not very efficient.
#We can see some segments but there still exist many overlaps.
#So maybe larger perplexity would be better and more cluttered.

#Repeat the previous step with a Perplexity greater than 30 (try to get a value much higher than 30).
theTSNE = Rtsne( dfu[,c(2,3,5,7,9,11,13,15,17,19)], dims = 2,
                 perplexity = 50, verbose = TRUE, max_iter = 500)

## Performing PCA
## Read the 5960 x 10 data matrix successfully!
## Using no_dims = 2, perplexity = 50.000000, and theta = 0.500000
## Computing input similarities...
## Building tree...
## Done in 0.73 seconds (sparsity = 0.032172)!
## Learning embedding...
```

```
## Iteration 50: error is 86.335781 (50 iterations in 1.02 seconds)
## Iteration 100: error is 70.471389 (50 iterations in 0.97 seconds)
## Iteration 150: error is 68.493062 (50 iterations in 0.88 seconds)
## Iteration 200: error is 67.183005 (50 iterations in 0.90 seconds)
## Iteration 250: error is 66.388505 (50 iterations in 0.89 seconds)
## Iteration 300: error is 1.787510 (50 iterations in 0.83 seconds)
## Iteration 350: error is 1.400510 (50 iterations in 0.82 seconds)
## Iteration 400: error is 1.207694 (50 iterations in 0.82 seconds)
## Iteration 450: error is 1.092022 (50 iterations in 0.83 seconds)
## Iteration 500: error is 1.017409 (50 iterations in 0.83 seconds)
## Fitting performed in 8.79 seconds.
```

```
dfu$TS1 = theTSNE$Y[,1]
dfu$TS2 = theTSNE$Y[,2]
colors <- c("steelblue", "red")
colors <- colors[dfu$TARGET_BAD_FLAG + 1]
plot( dfu$TS1, dfu$TS2, col = colors, pch = 16 )
```

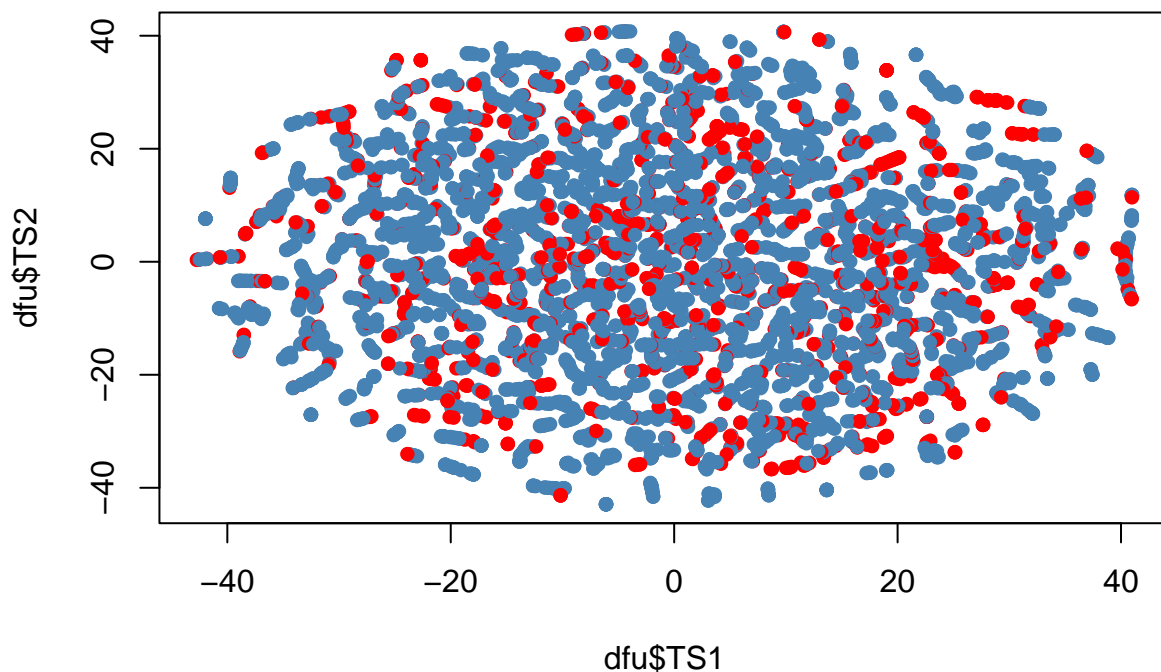


```
##Repeat the previous step with a Perplexity less than 30 (try to get a value much lower than 30).
theTSNE = Rtsne( dfu[,c(2,3,5,7,9,11,13,15,17,19)], dims = 2,
                 perplexity = 5, verbose = TRUE, max_iter = 500)
```

```
## Performing PCA
## Read the 5960 x 10 data matrix successfully!
## Using no_dims = 2, perplexity = 5.000000, and theta = 0.500000
```

```
## Computing input similarities...
## Building tree...
## Done in 0.09 seconds (sparsity = 0.003081)!
## Learning embedding...
## Iteration 50: error is 113.233374 (50 iterations in 0.63 seconds)
## Iteration 100: error is 93.433723 (50 iterations in 0.57 seconds)
## Iteration 150: error is 87.102198 (50 iterations in 0.55 seconds)
## Iteration 200: error is 83.704114 (50 iterations in 0.56 seconds)
## Iteration 250: error is 81.490318 (50 iterations in 0.57 seconds)
## Iteration 300: error is 3.438796 (50 iterations in 0.56 seconds)
## Iteration 350: error is 2.806279 (50 iterations in 0.56 seconds)
## Iteration 400: error is 2.388733 (50 iterations in 0.58 seconds)
## Iteration 450: error is 2.096944 (50 iterations in 0.58 seconds)
## Iteration 500: error is 1.882032 (50 iterations in 0.59 seconds)
## Fitting performed in 5.76 seconds.
```

```
dfu$TS1 = theTSNE$Y[,1]
dfu$TS2 = theTSNE$Y[,2]
colors <- c("steelblue", "red")
colors <- colors[dfu$TARGET_BAD_FLAG + 1]
plot( dfu$TS1, dfu$TS2, col = colors, pch = 16 )
```



```
##Decide on which value of Perplexity best predicts the Target Flag.
# 50 works best. I also tried 100 but it seemed too smooth and crowded.
# 5 is a small perplexity but it is too spread out and hard to find a pattern.
```



```

#Train two Random Forest Models to predict each of the tSNE values.
P = paste(colnames(dfu)[c(2,3,5,7,9,11,13,15,17,19)], collapse = "+")
F1 = as.formula( paste("TS1 ~", P ) )
F2 = as.formula( paste("TS2 ~", P ) )

print( F1 )

## TS1 ~ LOAN + IMP_MORTDUE + IMP_VALUE + IMP_YOJ + IMP_DEROG +
##      IMP_DELIHQ + IMP_CLAGE + IMP_NINQ + IMP_CLNO + IMP_DEBTINC

print( F2 )

## TS2 ~ LOAN + IMP_MORTDUE + IMP_VALUE + IMP_YOJ + IMP_DEROG +
##      IMP_DELIHQ + IMP_CLAGE + IMP_NINQ + IMP_CLNO + IMP_DEBTINC

ts1_model = lm( F1, data = dfu )
ts2_model = lm( F2, data = dfu )

ts1_model_rf = randomForest( data = dfu, F1, ntree = 200, importance = TRUE )
ts2_model_rf = randomForest( data = dfu, F2, ntree = 200, importance = TRUE )

df_tsne = df

df_tsne$TS1M = predict( ts1_model, df_tsne )
df_tsne$TS2M = predict( ts2_model, df_tsne )

df_tsne$TS1M_RF = predict( ts1_model_rf, df_tsne )
df_tsne$TS2M_RF = predict( ts2_model_rf, df_tsne )

df_tsne$RAND1 = sample( 100, size = nrow(df_tsne), replace = TRUE )
df_tsne$RAND2 = sample( 100, size = nrow(df_tsne), replace = TRUE )

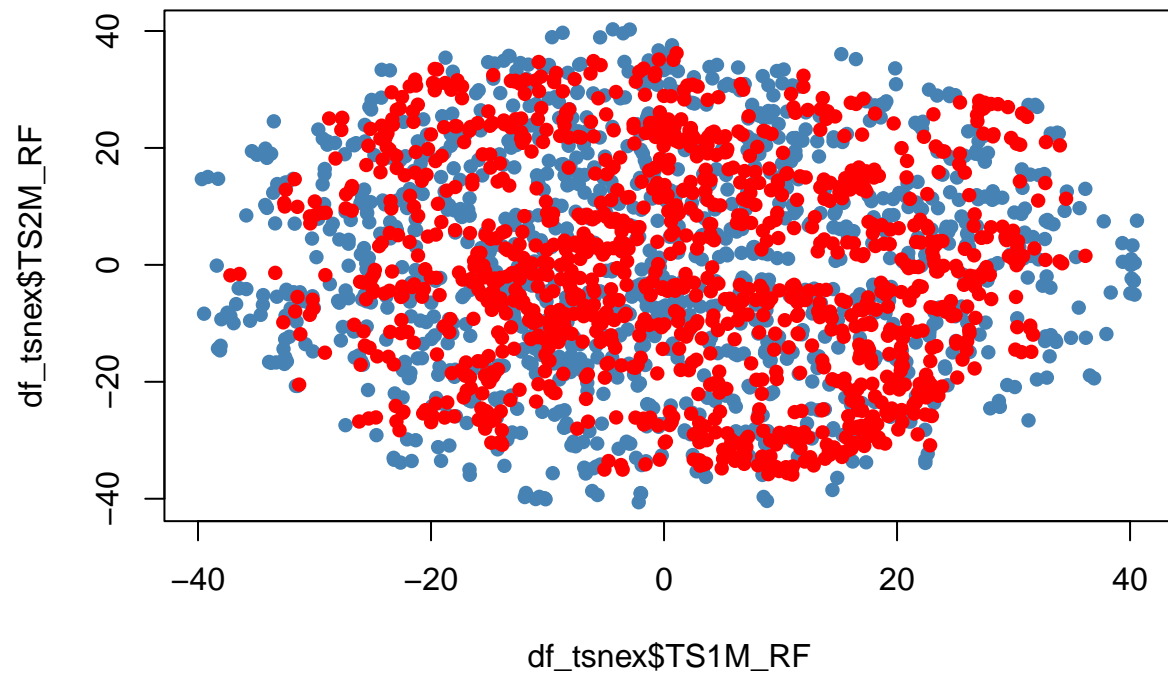
df_tsne0 = df_tsne[ which(df_tsne$TARGET_BAD_FLAG == 0), ]
df_tsne1 = df_tsne[ which(df_tsne$TARGET_BAD_FLAG == 1), ]

df_tsne$RAND1 = sample( 100, size = nrow(df_tsne), replace = TRUE )
df_tsne$RAND2 = sample( 100, size = nrow(df_tsne), replace = TRUE )

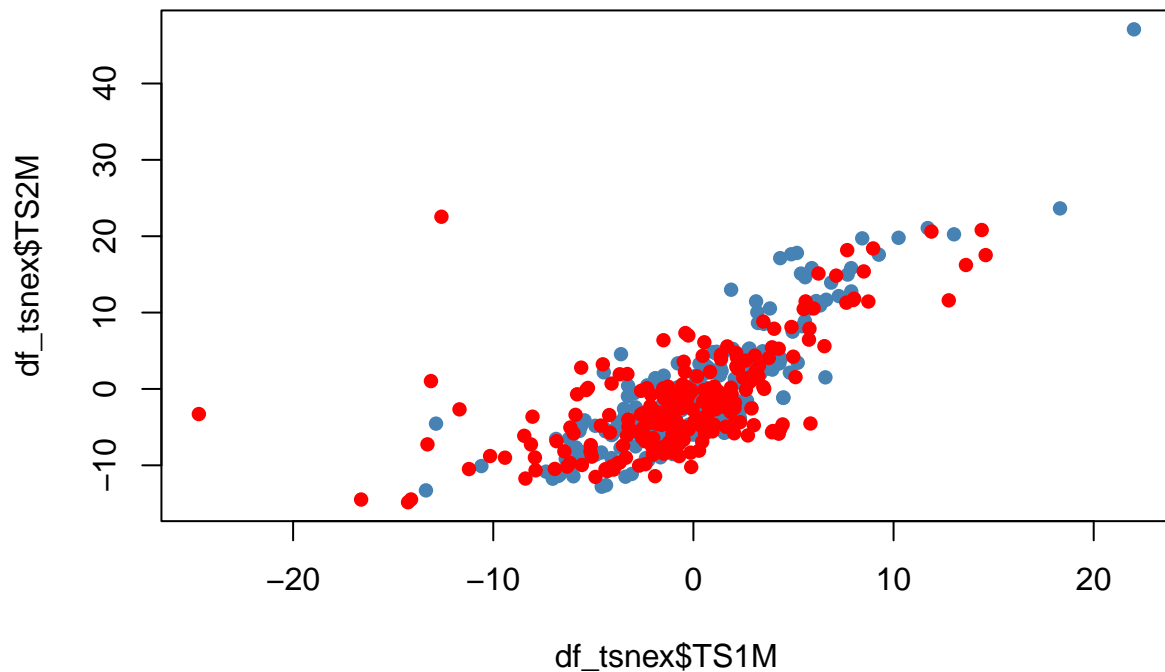
df_tsne0 = df_tsne0[df_tsne0$RAND1 < 25, ]

df_tsnex = rbind( df_tsne0, df_tsne1 )
colors <- c("steelblue", "red")
colors <- colors[df_tsnex$TARGET_BAD_FLAG + 1]
plot( df_tsnex$TS1M_RF, df_tsnex$TS2M_RF, col = colors, pch = 16 )

```



```
df_tsnext = df_tsnext[ df_tsnext$RAND2 < 20, ]  
colors <- c("steelblue", "red")  
colors <- colors[df_tsnext$TARGET_BAD_FLAG + 1]  
plot( df_tsnext$TS1M, df_tsnext$TS2M, col = colors, pch = 16 )
```



*#Step 4: Tree and Regression Analysis on the Original Data*  
*#Create a Decision Tree to predict Loan Default (Target Flag=1).*

```
df_model = df
df_model$TARGET_LOSS_AMT = NULL
```

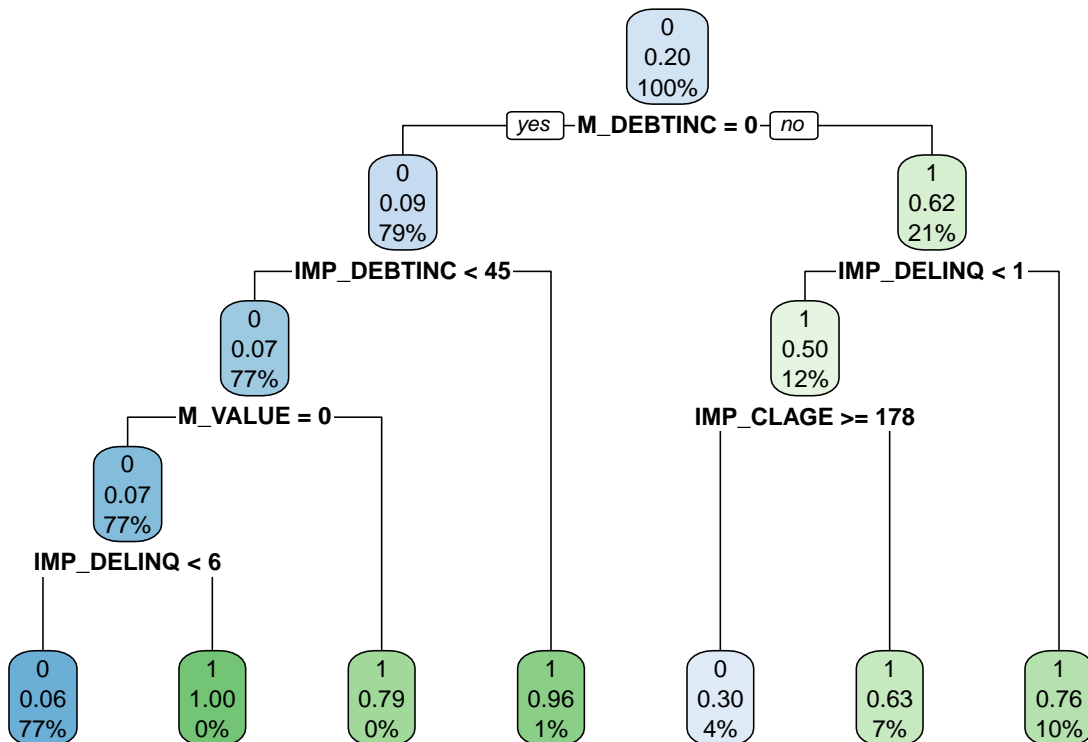
```
head( df_model )
```

```
##   TARGET_BAD_FLAG LOAN IMP_MORTDUE M_MORTDUE IMP_VALUE M_VALUE IMP_YOJ M_YOJ
## 1              1 1100      25860         0    39025      0    10.5    0
## 2              1 1300      70053         0    68400      0     7.0    0
## 3              1 1500     13500         0    16700      0     4.0    0
## 4              1 1500     65000         1    89000      1     7.0    1
## 5              0 1700     97800         0   112000      0     3.0    0
## 6              1 1700     30548         0    40320      0     9.0    0
##   IMP_DEROG M_DEROG IMP_DELIQ M_DELIQ IMP_CLAGE M_CLAGE IMP_NINQ M_NINQ
## 1          0      0         0         0  94.36667      0         1      0
## 2          0      0         2         0 121.83333      0         0      0
## 3          0      0         0         0 149.46667      0         1      0
## 4          1      1         1         1 174.00000      1         1      1
## 5          0      0         0         0  93.33333      0         0      0
## 6          0      0         0         0 101.46600      0         1      0
##   IMP_CLNO M_CLNO IMP_DEBTINC M_DEBTINC FLAG.Job.Mgr FLAG.Job.Office
## 1         9      0   35.00000         1           0           0
## 2        14      0   35.00000         1           0           0
## 3        10      0   35.00000         1           0           0
## 4        20      1   35.00000         1           0           0
```

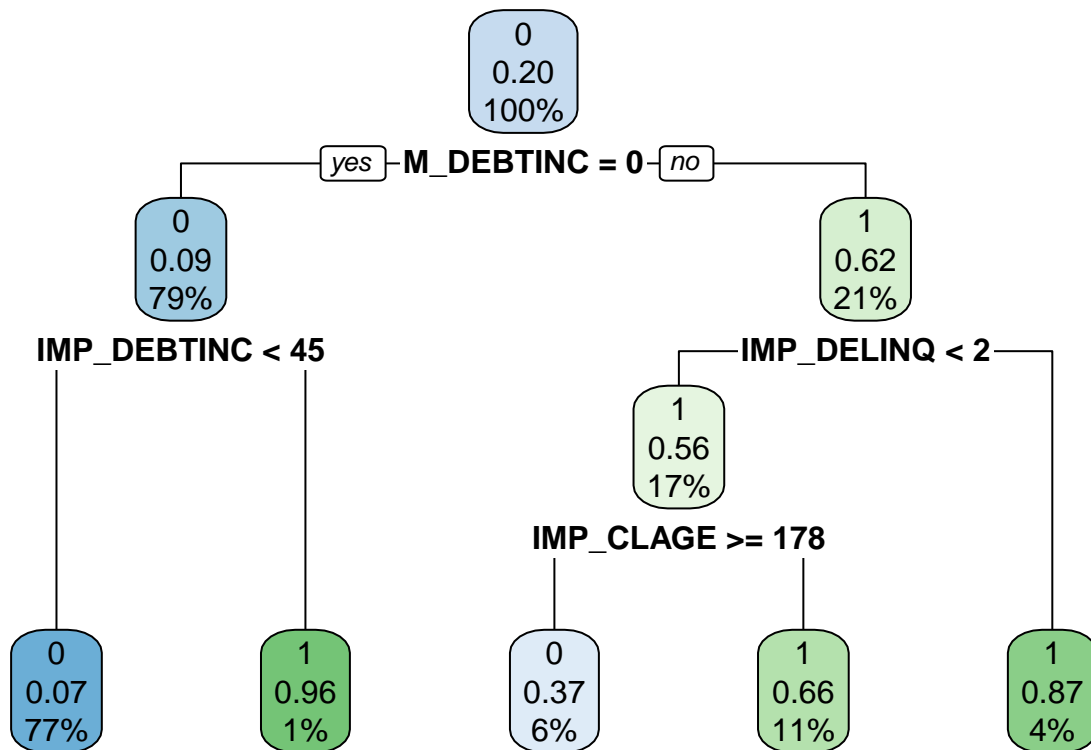
	14	0	35.00000	1	0	1
## 5	14	0	35.00000	1	0	1
## 6	8	0	37.11361	0	0	0
##	FLAG.Job.Other	FLAG.Job.ProfExe	FLAG.Job.Sales	FLAG.Job.Self		
## 1	1	0	0	0		
## 2	1	0	0	0		
## 3	1	0	0	0		
## 4	0	0	0	0		
## 5	0	0	0	0		
## 6	1	0	0	0		
##	FLAG.Reason.DebtCon	FLAG.Reason.HomeImp				
## 1	0	1				
## 2	0	1				
## 3	0	1				
## 4	0	0				
## 5	0	1				
## 6	0	1				

```
tr_set = rpart.control( maxdepth = 10 )
t1G = rpart( data = df_model, TARGET_BAD_FLAG ~ .,
             control = tr_set, method = "class", parms = list(split='gini') )
t1E = rpart( data = df_model, TARGET_BAD_FLAG ~ .,
             control = tr_set, method = "class", parms = list(split='information') )

rpart.plot( t1G )
```



```
rpart.plot( t1E )
```



```
t1G$variable.importance
```

```
##  M_DEBTINC IMP_DEBTINC IMP_DELTINQ  M_VALUE  IMP_CLAGE      LOAN
##  570.021010 128.539072  77.371518  51.334486  36.076295  25.645675
##  IMP_DEROG      M_DEROG  IMP_VALUE  M_DELTINQ      M_NINQ  IMP_YOJ
##  22.501563   9.540586   8.551021   7.632469   6.311465   4.323751
##      M_CLNO      IMP_CLNO IMP_MORTDUE
##      4.256569   2.837461   1.621407
```

```
t1E$variable.importance
```

```
##  M_DEBTINC IMP_DEBTINC IMP_DELTINQ  IMP_CLAGE      LOAN      M_VALUE
##  762.591210 188.922871  68.152477  40.125205  34.053718  30.094365
##  IMP_DEROG  IMP_VALUE      IMP_YOJ  IMP_CLNO IMP_MORTDUE
##  12.037746  10.263083   3.436136   3.075170   1.219274
```

```

#Comment on the variables that were included in the model.
#For both the gini and information method,
#they think debt income ratio is the most significant variable.
#And then both of them think DELINQ and CLAGE are important.
#Actually I remember on step 2 the PC2 is about those variables and we can conclude them as

```

```

#Credit Risk. Those variables are all about the risk of defaulting.

#Create a Logistic Regression model to predict Loan Default (Target Flag=1).
#Use either Forward, Backward, or Stepwise variable selection.
theUpper_LR = glm( TARGET_BAD_FLAG ~ ., family = "binomial", data = df_model )
theLower_LR = glm( TARGET_BAD_FLAG ~ 1, family = "binomial", data = df_model )

summary( theUpper_LR )

```

```

##
## Call:
## glm(formula = TARGET_BAD_FLAG ~ ., family = "binomial", data = df_model)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -7.217e+00  5.622e-01 -12.837  < 2e-16 ***
## LOAN          -7.945e-06  4.833e-06  -1.644  0.100181
## IMP_MORTDUE    -3.604e-06  1.732e-06  -2.081  0.037446 *
## M_MORTDUE       5.284e-01  2.031e-01   2.602  0.009270 **
## IMP_VALUE       3.972e-06  1.248e-06   3.182  0.001464 **
## M_VALUE        5.159e+00  5.358e-01   9.628  < 2e-16 ***
## IMP_YOJ        -1.629e-02  6.898e-03  -2.361  0.018222 *
## M_YOJ          -6.176e-01  1.972e-01  -3.132  0.001739 **
## IMP_DEROG       5.219e-01  6.258e-02   8.339  < 2e-16 ***
## M_DEROG        -2.548e+00  2.983e-01  -8.540  < 2e-16 ***
## IMP_DELINQ      8.002e-01  5.263e-02  15.204  < 2e-16 ***
## M_DELINQ       -1.603e+00  4.198e-01  -3.818  0.000135 ***
## IMP_CLAGE       -5.976e-03  6.806e-04  -8.780  < 2e-16 ***
## M_CLAGE        1.109e+00  3.433e-01   3.230  0.001237 **
## IMP_NINQ        1.453e-01  2.611e-02   5.565  2.61e-08 ***
## M_NINQ         -1.492e-01  3.816e-01  -0.391  0.695827
## IMP_CLNO        -1.306e-02  5.329e-03  -2.451  0.014266 *
## M_CLNO         3.242e+00  6.324e-01   5.127  2.95e-07 ***
## IMP_DEBTINC     9.416e-02  8.783e-03  10.721  < 2e-16 ***
## M_DEBTINC       2.668e+00  9.545e-02  27.951  < 2e-16 ***
## FLAG.Job.Mgr    2.243e+00  4.312e-01   5.201  1.98e-07 ***
## FLAG.Job.Office 1.553e+00  4.333e-01   3.585  0.000337 ***
## FLAG.Job.Other  2.339e+00  4.179e-01   5.597  2.19e-08 ***
## FLAG.Job.ProfExe 2.104e+00  4.285e-01   4.909  9.14e-07 ***
## FLAG.Job.Sales   3.421e+00  5.031e-01   6.801  1.04e-11 ***
## FLAG.Job.Self    2.649e+00  4.827e-01   5.488  4.07e-08 ***
## FLAG.Reason.DebtCon 4.464e-02  3.138e-01   0.142  0.886878
## FLAG.Reason.HomeImp 1.655e-01  3.185e-01   0.520  0.603245
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 5956.5  on 5959  degrees of freedom
## Residual deviance: 3233.3  on 5932  degrees of freedom
## AIC: 3289.3
##
## Number of Fisher Scoring iterations: 6

```

```
summary( theLower_LR )
```

```
##
## Call:
## glm(formula = TARGET_BAD_FLAG ~ 1, family = "binomial", data = df_model)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.38944    0.03241  -42.87  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 5956.5  on 5959  degrees of freedom
## Residual deviance: 5956.5  on 5959  degrees of freedom
## AIC: 5958.5
##
## Number of Fisher Scoring iterations: 4
```

```
lr_model = stepAIC( theLower_LR, direction="forward",
                    scope =list(lower=theLower_LR, upper=theUpper_LR))
```

```
## Start:  AIC=5958.47
## TARGET_BAD_FLAG ~ 1
##
##              Df Deviance    AIC
## + M_DEBTINC      1  4431.3 4435.3
## + IMP_DELINQ      1  5411.8 5415.8
## + M_VALUE         1  5659.8 5663.8
## + IMP_DEROG       1  5689.5 5693.5
## + IMP_CLAGE       1  5771.5 5775.5
## + IMP_DEBTINC     1  5798.0 5802.0
## + IMP_NINQ        1  5807.6 5811.6
## + LOAN            1  5919.8 5923.8
## + FLAG.Job.Office 1  5921.4 5925.4
## + M_DEROG         1  5923.8 5927.8
## + FLAG.Job.Other  1  5930.4 5934.4
## + M_DELINQ        1  5931.2 5935.2
## + M_YOJ           1  5935.5 5939.5
## + IMP_YOJ         1  5938.5 5942.5
## + FLAG.Job.Sales  1  5943.0 5947.0
## + IMP_MORTDUE     1  5943.1 5947.1
## + FLAG.Job.ProfExe 1  5944.8 5948.8
## + FLAG.Job.Self   1  5945.0 5949.0
## + M_NINQ          1  5946.2 5950.2
## + IMP_VALUE       1  5948.0 5952.0
## + FLAG.Reason.HomeImp 1  5948.2 5952.2
## + FLAG.Reason.DebtCon 1  5949.6 5953.6
## + FLAG.Job.Mgr    1  5950.4 5954.4
## + M_CLAGE         1  5950.9 5954.9
## + M_CLNO          1  5954.3 5958.3
```

```

## <none>                5956.5 5958.5
## + IMP_CLNO            1   5956.3 5960.3
## + M_MORTDUE           1   5956.4 5960.4
##
## Step:  AIC=4435.29
## TARGET_BAD_FLAG ~ M_DEBTINC
##
##              Df Deviance    AIC
## + IMP_DELINQ      1   4134.4 4140.4
## + IMP_DEBTINC      1   4263.3 4269.3
## + M_VALUE          1   4265.9 4271.9
## + IMP_DEROG        1   4289.6 4295.6
## + IMP_CLAGE         1   4300.3 4306.3
## + IMP_NINQ          1   4371.4 4377.4
## + M_DEROG           1   4401.3 4407.3
## + FLAG.Job.Office   1   4409.9 4415.9
## + M_DELINQ          1   4409.9 4415.9
## + IMP_YOJ           1   4412.0 4418.0
## + FLAG.Job.Other    1   4412.3 4418.3
## + M_YOJ             1   4417.6 4423.6
## + FLAG.Job.Sales    1   4419.9 4425.9
## + FLAG.Job.ProfExe  1   4421.6 4427.6
## + FLAG.Job.Self     1   4422.8 4428.8
## + M_NINQ            1   4423.8 4429.8
## + FLAG.Reason.HomeImp 1   4427.4 4433.4
## + FLAG.Job.Mgr      1   4427.8 4433.8
## + LOAN              1   4428.4 4434.4
## + FLAG.Reason.DebtCon 1   4428.7 4434.7
## <none>              4431.3 4435.3
## + M_CLAGE           1   4429.4 4435.4
## + IMP_MORTDUE        1   4430.0 4436.0
## + IMP_CLNO           1   4430.8 4436.8
## + M_CLNO             1   4431.0 4437.0
## + IMP_VALUE          1   4431.0 4437.0
## + M_MORTDUE          1   4431.3 4437.3
##
## Step:  AIC=4140.38
## TARGET_BAD_FLAG ~ M_DEBTINC + IMP_DELINQ
##
##              Df Deviance    AIC
## + IMP_CLAGE         1   3970.7 3978.7
## + IMP_DEBTINC        1   3972.8 3980.8
## + M_VALUE            1   4010.8 4018.8
## + M_DEROG            1   4040.9 4048.9
## + IMP_DEROG          1   4061.8 4069.8
## + IMP_NINQ           1   4080.3 4088.3
## + M_DELINQ           1   4092.3 4100.3
## + IMP_YOJ            1   4105.9 4113.9
## + FLAG.Job.Other     1   4107.8 4115.8
## + FLAG.Job.Office    1   4109.3 4117.3
## + M_NINQ             1   4110.3 4118.3
## + FLAG.Job.Sales     1   4117.7 4125.7
## + IMP_CLNO           1   4118.9 4126.9
## + M_YOJ              1   4120.1 4128.1

```



```

## + FLAG.Job.Self      1  4126.0 4134.0
## + FLAG.Job.ProfExe   1  4126.2 4134.2
## + FLAG.Reason.HomeImp 1  4129.5 4137.5
## + LOAN               1  4132.4 4140.4
## <none>              4134.4 4140.4
## + IMP_MORTDUE        1  4132.4 4140.4
## + FLAG.Reason.DebtCon 1  4132.8 4140.8
## + M_CLAGE            1  4133.3 4141.3
## + FLAG.Job.Mgr       1  4133.8 4141.8
## + M_CLNO             1  4134.0 4142.0
## + IMP_VALUE          1  4134.3 4142.3
## + M_MORTDUE          1  4134.3 4142.3
##
## Step:  AIC=3978.67
## TARGET_BAD_FLAG ~ M_DEBTINC + IMP_DELINQ + IMP_CLAGE
##
##              Df Deviance    AIC
## + IMP_DEBTINC      1   3818.2 3828.2
## + M_VALUE          1   3849.9 3859.9
## + M_DEROG          1   3884.9 3894.9
## + IMP_DEROG        1   3911.2 3921.2
## + M_DELINQ         1   3923.8 3933.8
## + IMP_NINQ         1   3933.8 3943.8
## + FLAG.Job.Office   1   3946.5 3956.5
## + M_NINQ           1   3946.9 3956.9
## + FLAG.Job.Sales    1   3951.0 3961.0
## + FLAG.Job.Other    1   3954.3 3964.3
## + M_YOJ            1   3960.3 3970.3
## + IMP_YOJ          1   3961.5 3971.5
## + FLAG.Job.Self     1   3963.3 3973.3
## + FLAG.Reason.HomeImp 1   3965.2 3975.2
## + IMP_VALUE         1   3965.4 3975.4
## + FLAG.Reason.DebtCon 1   3967.8 3977.8
## <none>             3970.7 3978.7
## + M_CLAGE          1   3968.9 3978.9
## + FLAG.Job.ProfExe   1   3969.3 3979.3
## + IMP_MORTDUE       1   3970.2 3980.2
## + FLAG.Job.Mgr      1   3970.5 3980.5
## + M_CLNO           1   3970.5 3980.5
## + M_MORTDUE         1   3970.5 3980.5
## + IMP_CLNO          1   3970.5 3980.5
## + LOAN             1   3970.7 3980.7
##
## Step:  AIC=3828.24
## TARGET_BAD_FLAG ~ M_DEBTINC + IMP_DELINQ + IMP_CLAGE + IMP_DEBTINC
##
##              Df Deviance    AIC
## + M_VALUE          1   3676.5 3688.5
## + M_DEROG          1   3731.6 3743.6
## + IMP_DEROG        1   3759.6 3771.6
## + M_DELINQ         1   3780.3 3792.3
## + FLAG.Job.Office   1   3789.3 3801.3
## + IMP_NINQ         1   3791.3 3803.3
## + FLAG.Job.Other    1   3802.8 3814.8

```

```

## + FLAG.Job.Sales      1  3803.6 3815.6
## + M_NINQ              1  3805.6 3817.6
## + M_YOJ               1  3809.7 3821.7
## + FLAG.Reason.HomeImp 1  3812.1 3824.1
## + M_CLAGE             1  3812.2 3824.2
## + IMP_YOJ             1  3812.5 3824.5
## + FLAG.Job.Self       1  3812.8 3824.8
## + FLAG.Reason.DebtCon 1  3813.7 3825.7
## + IMP_CLNO            1  3814.2 3826.2
## <none>                3818.2 3828.2
## + M_MORTDUE           1  3816.5 3828.5
## + IMP_VALUE           1  3816.7 3828.7
## + LOAN                1  3817.4 3829.4
## + IMP_MORTDUE         1  3818.0 3830.0
## + FLAG.Job.ProfExe    1  3818.0 3830.0
## + M_CLNO              1  3818.1 3830.1
## + FLAG.Job.Mgr        1  3818.2 3830.2
##
## Step: AIC=3688.49
## TARGET_BAD_FLAG ~ M_DEBTINC + IMP_DELINQ + IMP_CLAGE + IMP_DEBTINC +
## M_VALUE
##
##              Df Deviance    AIC
## + M_DEROG      1  3569.9 3583.9
## + M_DELINQ     1  3624.3 3638.3
## + IMP_DEROG    1  3625.7 3639.7
## + IMP_NINQ     1  3647.4 3661.4
## + FLAG.Job.Office 1  3647.8 3661.8
## + M_NINQ       1  3658.2 3672.2
## + FLAG.Job.Other 1  3658.3 3672.3
## + FLAG.Job.Sales 1  3660.8 3674.8
## + M_YOJ        1  3664.9 3678.9
## + FLAG.Reason.HomeImp 1  3669.4 3683.4
## + IMP_YOJ      1  3669.7 3683.7
## + FLAG.Job.Self 1  3671.6 3685.6
## + IMP_CLNO     1  3672.8 3686.8
## + FLAG.Reason.DebtCon 1  3673.6 3687.6
## + IMP_VALUE    1  3674.3 3688.3
## <none>         3676.5 3688.5
## + M_CLAGE      1  3674.5 3688.5
## + LOAN         1  3674.6 3688.6
## + M_CLNO       1  3675.7 3689.7
## + IMP_MORTDUE  1  3676.0 3690.0
## + FLAG.Job.ProfExe 1  3676.4 3690.4
## + FLAG.Job.Mgr  1  3676.5 3690.5
## + M_MORTDUE    1  3676.5 3690.5
##
## Step: AIC=3583.94
## TARGET_BAD_FLAG ~ M_DEBTINC + IMP_DELINQ + IMP_CLAGE + IMP_DEBTINC +
## M_VALUE + M_DEROG
##
##              Df Deviance    AIC
## + IMP_DEROG    1  3482.7 3498.7
## + M_CLAGE      1  3511.6 3527.6

```

```

## + M_CLNO          1    3524.8 3540.8
## + IMP_NINQ         1    3534.5 3550.5
## + FLAG.Job.Office  1    3537.5 3553.5
## + FLAG.Job.Sales   1    3553.8 3569.8
## + FLAG.Job.Other   1    3554.8 3570.8
## + IMP_YOJ          1    3560.8 3576.8
## + IMP_CLNO         1    3564.2 3580.2
## + M_NINQ           1    3564.3 3580.3
## + FLAG.Reason.HomeImp 1    3564.8 3580.8
## + M_MORTDUE        1    3566.0 3582.0
## + LOAN             1    3566.0 3582.0
## + FLAG.Reason.DebtCon 1    3566.5 3582.5
## + FLAG.Job.Self    1    3566.6 3582.6
## + M_YOJ            1    3567.5 3583.5
## <none>              3569.9 3583.9
## + IMP_VALUE        1    3568.3 3584.3
## + FLAG.Job.ProfExe  1    3569.0 3585.0
## + IMP_MORTDUE      1    3569.6 3585.6
## + FLAG.Job.Mgr     1    3569.7 3585.7
## + M_DELINQ         1    3569.8 3585.8
##
## Step:  AIC=3498.73
## TARGET_BAD_FLAG ~ M_DEBTINC + IMP_DELINQ + IMP_CLAGE + IMP_DEBTINC +
##      M_VALUE + M_DEROG + IMP_DEROG
##
##              Df Deviance    AIC
## + M_CLAGE      1    3428.4 3446.4
## + M_CLNO       1    3439.7 3457.7
## + FLAG.Job.Office 1    3454.1 3472.1
## + IMP_NINQ      1    3459.5 3477.5
## + FLAG.Job.Other 1    3469.8 3487.8
## + FLAG.Job.Sales 1    3470.1 3488.1
## + IMP_CLNO      1    3473.2 3491.2
## + FLAG.Reason.HomeImp 1    3476.0 3494.0
## + M_YOJ         1    3476.5 3494.5
## + IMP_YOJ       1    3476.6 3494.6
## + LOAN          1    3477.2 3495.2
## + FLAG.Reason.DebtCon 1    3478.3 3496.3
## + FLAG.Job.Self  1    3478.3 3496.3
## + IMP_VALUE     1    3480.2 3498.2
## <none>          3482.7 3498.7
## + M_MORTDUE     1    3481.0 3499.0
## + M_DELINQ      1    3481.3 3499.3
## + M_NINQ        1    3481.9 3499.9
## + FLAG.Job.ProfExe 1    3482.1 3500.1
## + FLAG.Job.Mgr   1    3482.6 3500.6
## + IMP_MORTDUE   1    3482.6 3500.6
##
## Step:  AIC=3446.38
## TARGET_BAD_FLAG ~ M_DEBTINC + IMP_DELINQ + IMP_CLAGE + IMP_DEBTINC +
##      M_VALUE + M_DEROG + IMP_DEROG + M_CLAGE
##
##              Df Deviance    AIC
## + IMP_NINQ      1    3402.2 3422.2

```

```

## + FLAG.Job.Office      1    3403.0 3423.0
## + FLAG.Job.Sales       1    3413.3 3433.3
## + M_YOJ                1    3414.4 3434.4
## + M_DELINQ             1    3416.0 3436.0
## + FLAG.Job.Other       1    3416.6 3436.6
## + FLAG.Reason.HomeImp  1    3422.1 3442.1
## + LOAN                 1    3423.7 3443.7
## + IMP_YOJ              1    3424.0 3444.0
## + FLAG.Job.Self        1    3424.1 3444.1
## + M_NINQ               1    3424.8 3444.8
## + IMP_VALUE            1    3424.9 3444.9
## + M_CLNO               1    3425.1 3445.1
## + IMP_CLNO             1    3425.1 3445.1
## + FLAG.Reason.DebtCon  1    3426.0 3446.0
## <none>                  3428.4 3446.4
## + FLAG.Job.Mgr         1    3427.8 3447.8
## + M_MORTDUE            1    3428.1 3448.1
## + FLAG.Job.ProfExe     1    3428.2 3448.2
## + IMP_MORTDUE          1    3428.4 3448.4
##
## Step: AIC=3422.25
## TARGET_BAD_FLAG ~ M_DEBTINC + IMP_DELINQ + IMP_CLAGE + IMP_DEBTINC +
##      M_VALUE + M_DEROG + IMP_DEROG + M_CLAGE + IMP_NINQ
##
##
##      Df Deviance    AIC
## + FLAG.Job.Office      1    3379.3 3401.3
## + FLAG.Job.Sales       1    3385.1 3407.1
## + M_YOJ                1    3388.7 3410.7
## + M_DELINQ             1    3391.9 3413.9
## + FLAG.Reason.HomeImp  1    3392.0 3414.0
## + FLAG.Job.Other       1    3392.1 3414.1
## + LOAN                 1    3395.0 3417.0
## + M_CLNO               1    3396.9 3418.9
## + FLAG.Reason.DebtCon  1    3397.2 3419.2
## + IMP_CLNO             1    3397.4 3419.4
## + FLAG.Job.Self        1    3398.2 3420.2
## + IMP_YOJ              1    3398.5 3420.5
## + IMP_VALUE            1    3398.7 3420.7
## <none>                  3402.2 3422.2
## + M_NINQ               1    3400.5 3422.5
## + M_MORTDUE            1    3401.5 3423.5
## + FLAG.Job.Mgr         1    3402.0 3424.0
## + IMP_MORTDUE          1    3402.1 3424.1
## + FLAG.Job.ProfExe     1    3402.2 3424.2
##
## Step: AIC=3401.27
## TARGET_BAD_FLAG ~ M_DEBTINC + IMP_DELINQ + IMP_CLAGE + IMP_DEBTINC +
##      M_VALUE + M_DEROG + IMP_DEROG + M_CLAGE + IMP_NINQ + FLAG.Job.Office
##
##
##      Df Deviance    AIC
## + M_YOJ                1    3362.7 3386.7
## + FLAG.Job.Sales       1    3364.5 3388.5
## + M_DELINQ             1    3368.4 3392.4
## + FLAG.Reason.HomeImp  1    3369.6 3393.6

```

```

## + LOAN                1    3372.2 3396.2
## + IMP_CLNO             1    3373.7 3397.7
## + M_CLNO              1    3374.2 3398.2
## + FLAG.Reason.DebtCon  1    3374.7 3398.7
## + IMP_YOJ              1    3375.0 3399.0
## + IMP_VALUE            1    3376.6 3400.6
## + FLAG.Job.Self        1    3376.6 3400.6
## + M_NINQ              1    3376.8 3400.8
## + FLAG.Job.Other       1    3376.9 3400.9
## <none>                 3379.3 3401.3
## + FLAG.Job.ProfExe     1    3377.7 3401.7
## + M_MORTDUE            1    3378.3 3402.3
## + IMP_MORTDUE          1    3378.8 3402.8
## + FLAG.Job.Mgr         1    3379.1 3403.1
##
## Step:  AIC=3386.68
## TARGET_BAD_FLAG ~ M_DEBTINC + IMP_DELINQ + IMP_CLAGE + IMP_DEBTINC +
##      M_VALUE + M_DEROG + IMP_DEROG + M_CLAGE + IMP_NINQ + FLAG.Job.Office +
##      M_YOJ
##
##              Df Deviance    AIC
## + FLAG.Job.Sales      1    3347.2 3373.2
## + FLAG.Reason.HomeImp  1    3354.6 3380.6
## + IMP_CLNO             1    3355.2 3381.2
## + M_DELINQ             1    3355.2 3381.2
## + M_CLNO              1    3356.3 3382.3
## + LOAN                1    3357.3 3383.3
## + IMP_YOJ              1    3357.7 3383.7
## + FLAG.Reason.DebtCon  1    3358.2 3384.2
## + FLAG.Job.Other       1    3358.4 3384.4
## + M_MORTDUE            1    3358.8 3384.8
## + FLAG.Job.ProfExe     1    3359.3 3385.3
## + FLAG.Job.Self        1    3360.1 3386.1
## <none>                 3362.7 3386.7
## + IMP_VALUE            1    3360.9 3386.9
## + IMP_MORTDUE          1    3361.6 3387.6
## + M_NINQ              1    3362.1 3388.1
## + FLAG.Job.Mgr         1    3362.2 3388.2
##
## Step:  AIC=3373.21
## TARGET_BAD_FLAG ~ M_DEBTINC + IMP_DELINQ + IMP_CLAGE + IMP_DEBTINC +
##      M_VALUE + M_DEROG + IMP_DEROG + M_CLAGE + IMP_NINQ + FLAG.Job.Office +
##      M_YOJ + FLAG.Job.Sales
##
##              Df Deviance    AIC
## + FLAG.Reason.HomeImp  1    3338.2 3366.2
## + IMP_CLNO             1    3338.4 3366.4
## + FLAG.Job.Other       1    3339.6 3367.6
## + M_DELINQ             1    3340.1 3368.1
## + M_CLNO              1    3340.5 3368.5
## + FLAG.Reason.DebtCon  1    3341.9 3369.9
## + IMP_YOJ              1    3342.9 3370.9
## + M_MORTDUE            1    3343.0 3371.0
## + LOAN                1    3343.1 3371.1

```

```

## + FLAG.Job.Self          1    3344.2 3372.2
## + FLAG.Job.ProfExe       1    3345.1 3373.1
## <none>                   1    3347.2 3373.2
## + IMP_VALUE              1    3345.5 3373.5
## + IMP_MORTDUE            1    3345.9 3373.9
## + M_NINQ                 1    3346.9 3374.9
## + FLAG.Job.Mgr           1    3347.0 3375.0
##
## Step:  AIC=3366.16
## TARGET_BAD_FLAG ~ M_DEBTINC + IMP_DELINQ + IMP_CLAGE + IMP_DEBTINC +
##      M_VALUE + M_DEROG + IMP_DEROG + M_CLAGE + IMP_NINQ + FLAG.Job.Office +
##      M_YOJ + FLAG.Job.Sales + FLAG.Reason.HomeImp
##
##              Df Deviance    AIC
## + M_DELINQ          1    3330.3 3360.3
## + FLAG.Job.Other    1    3330.8 3360.8
## + IMP_CLNO          1    3331.7 3361.7
## + M_CLNO            1    3332.4 3362.4
## + IMP_YOJ           1    3333.8 3363.8
## + M_MORTDUE         1    3335.4 3365.4
## + LOAN              1    3335.7 3365.7
## + FLAG.Reason.DebtCon 1    3336.0 3366.0
## + FLAG.Job.ProfExe   1    3336.1 3366.1
## + FLAG.Job.Self      1    3336.1 3366.1
## <none>              1    3338.2 3366.2
## + IMP_VALUE          1    3336.5 3366.5
## + IMP_MORTDUE        1    3337.1 3367.1
## + M_NINQ             1    3337.6 3367.6
## + FLAG.Job.Mgr       1    3338.1 3368.1
##
## Step:  AIC=3360.32
## TARGET_BAD_FLAG ~ M_DEBTINC + IMP_DELINQ + IMP_CLAGE + IMP_DEBTINC +
##      M_VALUE + M_DEROG + IMP_DEROG + M_CLAGE + IMP_NINQ + FLAG.Job.Office +
##      M_YOJ + FLAG.Job.Sales + FLAG.Reason.HomeImp + M_DELINQ
##
##              Df Deviance    AIC
## + M_CLNO          1    3314.5 3346.5
## + FLAG.Job.Other    1    3322.6 3354.6
## + IMP_CLNO          1    3324.7 3356.7
## + IMP_YOJ           1    3325.5 3357.5
## + M_MORTDUE         1    3325.9 3357.9
## + FLAG.Job.Self      1    3327.9 3359.9
## + FLAG.Job.ProfExe   1    3328.1 3360.1
## + LOAN              1    3328.3 3360.3
## <none>              1    3330.3 3360.3
## + FLAG.Reason.DebtCon 1    3328.3 3360.3
## + IMP_VALUE          1    3328.8 3360.8
## + M_NINQ             1    3329.1 3361.1
## + IMP_MORTDUE        1    3329.3 3361.3
## + FLAG.Job.Mgr       1    3330.2 3362.2
##
## Step:  AIC=3346.53
## TARGET_BAD_FLAG ~ M_DEBTINC + IMP_DELINQ + IMP_CLAGE + IMP_DEBTINC +
##      M_VALUE + M_DEROG + IMP_DEROG + M_CLAGE + IMP_NINQ + FLAG.Job.Office +

```

```

##      M_YOJ + FLAG.Job.Sales + FLAG.Reason.HomeImp + M_DELINQ +
##      M_CLNO
##
##              Df Deviance    AIC
## + IMP_CLNO          1   3304.2 3338.2
## + FLAG.Job.Other     1   3304.5 3338.5
## + IMP_YOJ            1   3309.3 3343.3
## + M_MORTDUE          1   3310.1 3344.1
## + FLAG.Reason.DebtCon 1   3311.4 3345.4
## + FLAG.Job.ProfExe    1   3311.8 3345.8
## + FLAG.Job.Self       1   3312.0 3346.0
## <none>                3314.5 3346.5
## + IMP_VALUE          1   3312.9 3346.9
## + LOAN                1   3313.4 3347.4
## + IMP_MORTDUE         1   3313.7 3347.7
## + FLAG.Job.Mgr        1   3314.4 3348.4
## + M_NINQ              1   3314.5 3348.5
##
## Step:  AIC=3338.24
## TARGET_BAD_FLAG ~ M_DEBTINC + IMP_DELINQ + IMP_CLAGE + IMP_DEBTINC +
##      M_VALUE + M_DEROG + IMP_DEROG + M_CLAGE + IMP_NINQ + FLAG.Job.Office +
##      M_YOJ + FLAG.Job.Sales + FLAG.Reason.HomeImp + M_DELINQ +
##      M_CLNO + IMP_CLNO
##
##              Df Deviance    AIC
## + FLAG.Job.Other     1   3297.2 3333.2
## + IMP_YOJ            1   3298.8 3334.8
## + IMP_VALUE          1   3300.7 3336.7
## + FLAG.Reason.DebtCon 1   3301.1 3337.1
## + FLAG.Job.Self       1   3301.2 3337.2
## + M_MORTDUE          1   3301.4 3337.4
## <none>                3304.2 3338.2
## + FLAG.Job.ProfExe    1   3302.9 3338.9
## + LOAN                1   3303.6 3339.6
## + M_NINQ              1   3304.2 3340.2
## + FLAG.Job.Mgr        1   3304.2 3340.2
## + IMP_MORTDUE         1   3304.2 3340.2
##
## Step:  AIC=3333.16
## TARGET_BAD_FLAG ~ M_DEBTINC + IMP_DELINQ + IMP_CLAGE + IMP_DEBTINC +
##      M_VALUE + M_DEROG + IMP_DEROG + M_CLAGE + IMP_NINQ + FLAG.Job.Office +
##      M_YOJ + FLAG.Job.Sales + FLAG.Reason.HomeImp + M_DELINQ +
##      M_CLNO + IMP_CLNO + FLAG.Job.Other
##
##              Df Deviance    AIC
## + IMP_VALUE          1   3291.2 3329.2
## + IMP_YOJ            1   3291.4 3329.4
## + FLAG.Job.Self       1   3291.4 3329.4
## + M_MORTDUE          1   3294.7 3332.7
## + FLAG.Reason.DebtCon 1   3294.9 3332.9
## <none>                3297.2 3333.2
## + FLAG.Job.Mgr        1   3295.6 3333.6
## + IMP_MORTDUE         1   3296.8 3334.8
## + LOAN                1   3296.8 3334.8

```

```

## + FLAG.Job.ProfExe      1    3297.0 3335.0
## + M_NINQ                1    3297.1 3335.1
##
## Step:  AIC=3329.21
## TARGET_BAD_FLAG ~ M_DEBTINC + IMP_DELINQ + IMP_CLAGE + IMP_DEBTINC +
##      M_VALUE + M_DEROG + IMP_DEROG + M_CLAGE + IMP_NINQ + FLAG.Job.Office +
##      M_YOJ + FLAG.Job.Sales + FLAG.Reason.HomeImp + M_DELINQ +
##      M_CLNO + IMP_CLNO + FLAG.Job.Other + IMP_VALUE
##
##
##      Df Deviance    AIC
## + IMP_YOJ          1    3285.1 3325.1
## + FLAG.Job.Self    1    3286.3 3326.3
## + M_MORTDUE        1    3288.5 3328.5
## + IMP_MORTDUE      1    3288.9 3328.9
## + LOAN             1    3288.9 3328.9
## + FLAG.Job.Mgr     1    3289.0 3329.0
## + FLAG.Reason.DebtCon 1    3289.1 3329.1
## <none>              3291.2 3329.2
## + M_NINQ           1    3291.2 3331.2
## + FLAG.Job.ProfExe  1    3291.2 3331.2
##
## Step:  AIC=3325.11
## TARGET_BAD_FLAG ~ M_DEBTINC + IMP_DELINQ + IMP_CLAGE + IMP_DEBTINC +
##      M_VALUE + M_DEROG + IMP_DEROG + M_CLAGE + IMP_NINQ + FLAG.Job.Office +
##      M_YOJ + FLAG.Job.Sales + FLAG.Reason.HomeImp + M_DELINQ +
##      M_CLNO + IMP_CLNO + FLAG.Job.Other + IMP_VALUE + IMP_YOJ
##
##
##      Df Deviance    AIC
## + FLAG.Job.Self    1    3280.6 3322.6
## + IMP_MORTDUE      1    3281.7 3323.7
## + M_MORTDUE        1    3282.4 3324.4
## + FLAG.Job.Mgr     1    3282.5 3324.5
## <none>              3285.1 3325.1
## + FLAG.Reason.DebtCon 1    3283.1 3325.1
## + LOAN             1    3283.6 3325.6
## + M_NINQ           1    3285.1 3327.1
## + FLAG.Job.ProfExe  1    3285.1 3327.1
##
## Step:  AIC=3322.64
## TARGET_BAD_FLAG ~ M_DEBTINC + IMP_DELINQ + IMP_CLAGE + IMP_DEBTINC +
##      M_VALUE + M_DEROG + IMP_DEROG + M_CLAGE + IMP_NINQ + FLAG.Job.Office +
##      M_YOJ + FLAG.Job.Sales + FLAG.Reason.HomeImp + M_DELINQ +
##      M_CLNO + IMP_CLNO + FLAG.Job.Other + IMP_VALUE + IMP_YOJ +
##      FLAG.Job.Self
##
##
##      Df Deviance    AIC
## + FLAG.Job.Mgr     1    3276.0 3320.0
## + IMP_MORTDUE      1    3277.4 3321.4
## + M_MORTDUE        1    3277.9 3321.9
## + LOAN             1    3278.3 3322.3
## <none>              3280.6 3322.6
## + FLAG.Reason.DebtCon 1    3278.7 3322.7
## + FLAG.Job.ProfExe  1    3280.0 3324.0
## + M_NINQ           1    3280.5 3324.5

```



```

##
## Step: AIC=3319.97
## TARGET_BAD_FLAG ~ M_DEBTINC + IMP_DELINQ + IMP_CLAGE + IMP_DEBTINC +
## M_VALUE + M_DEROG + IMP_DEROG + M_CLAGE + IMP_NINQ + FLAG.Job.Office +
## M_YOJ + FLAG.Job.Sales + FLAG.Reason.HomeImp + M_DELINQ +
## M_CLNO + IMP_CLNO + FLAG.Job.Other + IMP_VALUE + IMP_YOJ +
## FLAG.Job.Self + FLAG.Job.Mgr
##
##
## Df Deviance AIC
## + FLAG.Job.ProfExe 1 3245.9 3291.9
## + IMP_MORTDUE 1 3272.8 3318.8
## + M_MORTDUE 1 3273.2 3319.2
## + LOAN 1 3273.5 3319.5
## <none> 3276.0 3320.0
## + FLAG.Reason.DebtCon 1 3274.3 3320.3
## + M_NINQ 1 3275.9 3321.9
##
## Step: AIC=3291.89
## TARGET_BAD_FLAG ~ M_DEBTINC + IMP_DELINQ + IMP_CLAGE + IMP_DEBTINC +
## M_VALUE + M_DEROG + IMP_DEROG + M_CLAGE + IMP_NINQ + FLAG.Job.Office +
## M_YOJ + FLAG.Job.Sales + FLAG.Reason.HomeImp + M_DELINQ +
## M_CLNO + IMP_CLNO + FLAG.Job.Other + IMP_VALUE + IMP_YOJ +
## FLAG.Job.Self + FLAG.Job.Mgr + FLAG.Job.ProfExe
##
##
## Df Deviance AIC
## + M_MORTDUE 1 3240.7 3288.7
## + IMP_MORTDUE 1 3242.2 3290.2
## + LOAN 1 3243.9 3291.9
## <none> 3245.9 3291.9
## + M_NINQ 1 3245.8 3293.8
## + FLAG.Reason.DebtCon 1 3245.9 3293.9
##
## Step: AIC=3288.73
## TARGET_BAD_FLAG ~ M_DEBTINC + IMP_DELINQ + IMP_CLAGE + IMP_DEBTINC +
## M_VALUE + M_DEROG + IMP_DEROG + M_CLAGE + IMP_NINQ + FLAG.Job.Office +
## M_YOJ + FLAG.Job.Sales + FLAG.Reason.HomeImp + M_DELINQ +
## M_CLNO + IMP_CLNO + FLAG.Job.Other + IMP_VALUE + IMP_YOJ +
## FLAG.Job.Self + FLAG.Job.Mgr + FLAG.Job.ProfExe + M_MORTDUE
##
##
## Df Deviance AIC
## + IMP_MORTDUE 1 3236.3 3286.3
## + LOAN 1 3238.1 3288.1
## <none> 3240.7 3288.7
## + M_NINQ 1 3240.6 3290.6
## + FLAG.Reason.DebtCon 1 3240.7 3290.7
##
## Step: AIC=3286.25
## TARGET_BAD_FLAG ~ M_DEBTINC + IMP_DELINQ + IMP_CLAGE + IMP_DEBTINC +
## M_VALUE + M_DEROG + IMP_DEROG + M_CLAGE + IMP_NINQ + FLAG.Job.Office +
## M_YOJ + FLAG.Job.Sales + FLAG.Reason.HomeImp + M_DELINQ +
## M_CLNO + IMP_CLNO + FLAG.Job.Other + IMP_VALUE + IMP_YOJ +
## FLAG.Job.Self + FLAG.Job.Mgr + FLAG.Job.ProfExe + M_MORTDUE +
## IMP_MORTDUE
##

```

```
##              Df Deviance    AIC
## + LOAN              1  3233.5 3285.5
## <none>              3236.3 3286.3
## + M_NINQ            1  3236.1 3288.1
## + FLAG.Reason.DebtCon 1  3236.2 3288.2
##
## Step:  AIC=3285.51
## TARGET_BAD_FLAG ~ M_DEBTINC + IMP_DELINQ + IMP_CLAGE + IMP_DEBTINC +
##     M_VALUE + M_DEROG + IMP_DEROG + M_CLAGE + IMP_NINQ + FLAG.Job.Office +
##     M_YOJ + FLAG.Job.Sales + FLAG.Reason.HomeImp + M_DELINQ +
##     M_CLNO + IMP_CLNO + FLAG.Job.Other + IMP_VALUE + IMP_YOJ +
##     FLAG.Job.Self + FLAG.Job.Mgr + FLAG.Job.ProfExe + M_MORTDUE +
##     IMP_MORTDUE + LOAN
##
##              Df Deviance    AIC
## <none>              3233.5 3285.5
## + M_NINQ            1  3233.4 3287.4
## + FLAG.Reason.DebtCon 1  3233.5 3287.5
```

```
summary( lr_model )
```

```
##
## Call:
## glm(formula = TARGET_BAD_FLAG ~ M_DEBTINC + IMP_DELINQ + IMP_CLAGE +
##     IMP_DEBTINC + M_VALUE + M_DEROG + IMP_DEROG + M_CLAGE + IMP_NINQ +
##     FLAG.Job.Office + M_YOJ + FLAG.Job.Sales + FLAG.Reason.HomeImp +
##     M_DELINQ + M_CLNO + IMP_CLNO + FLAG.Job.Other + IMP_VALUE +
##     IMP_YOJ + FLAG.Job.Self + FLAG.Job.Mgr + FLAG.Job.ProfExe +
##     M_MORTDUE + IMP_MORTDUE + LOAN, family = "binomial", data = df_model)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -7.198e+00  5.281e-01 -13.630 < 2e-16 ***
## M_DEBTINC      2.669e+00  9.539e-02  27.977 < 2e-16 ***
## IMP_DELINQ      7.989e-01  5.252e-02  15.210 < 2e-16 ***
## IMP_CLAGE     -5.975e-03  6.789e-04  -8.801 < 2e-16 ***
## IMP_DEBTINC      9.445e-02  8.761e-03  10.780 < 2e-16 ***
## M_VALUE        5.164e+00  5.305e-01   9.735 < 2e-16 ***
## M_DEROG       -2.546e+00  2.980e-01  -8.545 < 2e-16 ***
## IMP_DEROG       5.195e-01  6.212e-02   8.362 < 2e-16 ***
## M_CLAGE        1.111e+00  3.429e-01   3.241 0.001192 **
## IMP_NINQ       1.459e-01  2.609e-02   5.592 2.25e-08 ***
## FLAG.Job.Office 1.569e+00  4.188e-01   3.747 0.000179 ***
## M_YOJ         -6.297e-01  1.952e-01  -3.226 0.001256 **
## FLAG.Job.Sales  3.443e+00  4.886e-01   7.046 1.84e-12 ***
## FLAG.Reason.HomeImp 1.215e-01  1.050e-01   1.157 0.247415
## M_DELINQ     -1.675e+00  3.763e-01  -4.452 8.51e-06 ***
## M_CLNO        3.166e+00  6.001e-01   5.276 1.32e-07 ***
## IMP_CLNO     -1.304e-02  5.329e-03  -2.447 0.014408 *
## FLAG.Job.Other  2.354e+00  4.020e-01   5.855 4.76e-09 ***
## IMP_VALUE      3.951e-06  1.243e-06   3.178 0.001485 **
## IMP_YOJ     -1.636e-02  6.894e-03  -2.372 0.017672 *
## FLAG.Job.Self  2.653e+00  4.702e-01   5.642 1.68e-08 ***
## FLAG.Job.Mgr   2.256e+00  4.173e-01   5.407 6.41e-08 ***
```

```
## FLAG.Job.ProfExe      2.117e+00  4.145e-01  5.108 3.26e-07 ***
## M_MORTDUE             5.247e-01  2.027e-01  2.589 0.009616 **
## IMP_MORTDUE           -3.596e-06  1.728e-06  -2.080 0.037498 *
## LOAN                  -7.897e-06  4.827e-06  -1.636 0.101864
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 5956.5 on 5959 degrees of freedom
## Residual deviance: 3233.5 on 5934 degrees of freedom
## AIC: 3285.5
##
## Number of Fisher Scoring iterations: 6
```

```
#Comment on the variables that were included in the model.
#For this logistic regression model, also IMP_DEBTINC has highest weight.
#Then it is IMP_DELTINQ, LOAN, IMP_CLAGE. These are the same as the previous models.
```

```
#Create a ROC curve showing the accuracy of the model.
```

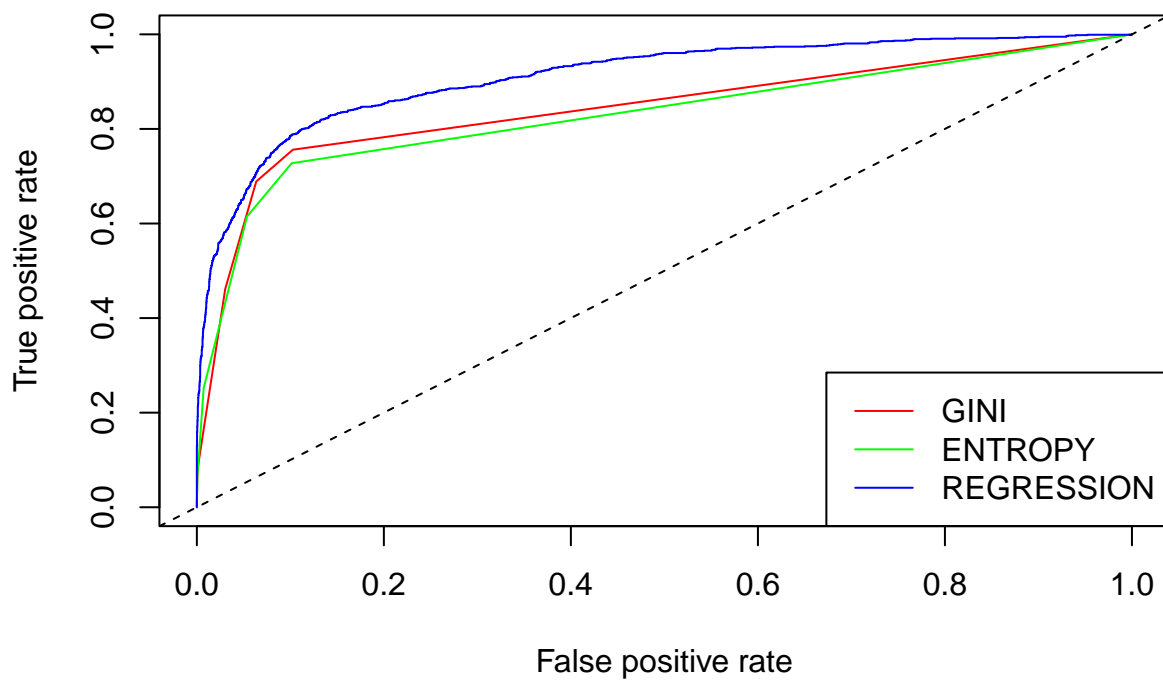
```
pG = predict( t1G, df_model )
pG2 = prediction( pG[,2], df_model$TARGET_BAD_FLAG )
pG3 = performance( pG2, "tpr", "fpr" )
```

```
pE = predict( t1E, df_model )
pE2 = prediction( pE[,2], df_model$TARGET_BAD_FLAG )
pE3 = performance( pE2, "tpr", "fpr" )
```

```
plr = predict( lr_model, df_model, type="response" )
plr2 = prediction( plr, df_model$TARGET_BAD_FLAG )
plr3 = performance( plr2, "tpr", "fpr" )
```

```
#Calculate and display the Area Under the ROC Curve (AUC).
```

```
plot( pG3, col="red" )
plot( pE3, col="green", add=TRUE )
plot( plr3, col="blue", add=TRUE )
abline(0,1,lty=2)
legend("bottomright",c("GINI","ENTROPY","REGRESSION"),
      col = c("red","green","blue"), bty="y", lty=1 )
```



```
aucG = performance( pG2, "auc" )@y.values
aucE = performance( pE2, "auc" )@y.values
aucR = performance( plr2, "auc" )@y.values
```

```
print( aucG )
```

```
## [[1]]
## [1] 0.8433084
```

```
print( aucE )
```

```
## [[1]]
## [1] 0.8293732
```

```
print( aucR )
```

```
## [[1]]
## [1] 0.9105166
```

*#Step 5: Tree and Regression Analysis on the PCA/tSNE Data*

*#Append the Principal Component values from Step 2 to your data set.*

*#Using the Random Forest models from Step 3, append the two tSNE values to the data set.*

```

df_model = df
df_model$TARGET_LOSS_AMT = NULL

df_model$PC1 = df_new[, "PC1"]
df_model$PC2 = df_new[, "PC2"]
df_model$PC3 = df_new[, "PC3"]
df_model$PC4 = df_new[, "PC4"]

df_model$TS1M_RF = predict( ts1_model_rf, df_model )
df_model$TS2M_RF = predict( ts2_model_rf, df_model )

#Remove all of the continuous variables from the data set (set them to NULL).
#Keep the flag variables in the data set.
df_model$LOAN = NULL
df_model$IMP_MORTDUE = NULL
df_model$IMP_VALUE = NULL
df_model$IMP_YOJ = NULL
df_model$IMP_DEROG = NULL
df_model$IMP_DELINQ = NULL
df_model$IMP_CLAGE = NULL
df_model$IMP_NINQ = NULL
df_model$IMP_CLNO = NULL
df_model$IMP_DEBTINC = NULL

head( df_model )

##      TARGET_BAD_FLAG M_MORTDUE M_VALUE M_YOJ M_DEROG M_DELINQ M_CLAGE M_NINQ
## 1             1         0         0         0         0         0         0
## 2             1         0         0         0         0         0         0
## 3             1         0         0         0         0         0         0
## 4             1         1         1         1         1         1         1
## 5             0         0         0         0         0         0         0
## 6             1         0         0         0         0         0         0
##      M_CLNO M_DEBTINC FLAG.Job.Mgr FLAG.Job.Office FLAG.Job.Other FLAG.Job.ProfExe
## 1         0         1         0         0         1         0
## 2         0         1         0         0         1         0
## 3         0         1         0         0         1         0
## 4         1         1         0         0         0         0
## 5         0         1         0         1         0         0
## 6         0         0         0         0         1         0
##      FLAG.Job.Sales FLAG.Job.Self FLAG.Reason.DebtCon FLAG.Reason.HomeImp
## 1             0         0         0         1
## 2             0         0         0         1
## 3             0         0         0         1
## 4             0         0         0         0
## 5             0         0         0         1
## 6             0         0         0         1
##      PC1      PC2      PC3      PC4      TS1M_RF      TS2M_RF
## 1 -2.4361630 -0.2914953 0.60058199 -0.013149908 17.184316 -30.976658
## 2 -1.2657133 0.3930930 -0.07345171 1.208924161 -9.059882 -13.214553
## 3 -2.6621119 -0.1696773 0.58110691 0.404535849 9.599395 -34.560871
## 4 -0.7828377 0.8659403 -0.29103382 0.988600888 -5.239384 1.771037

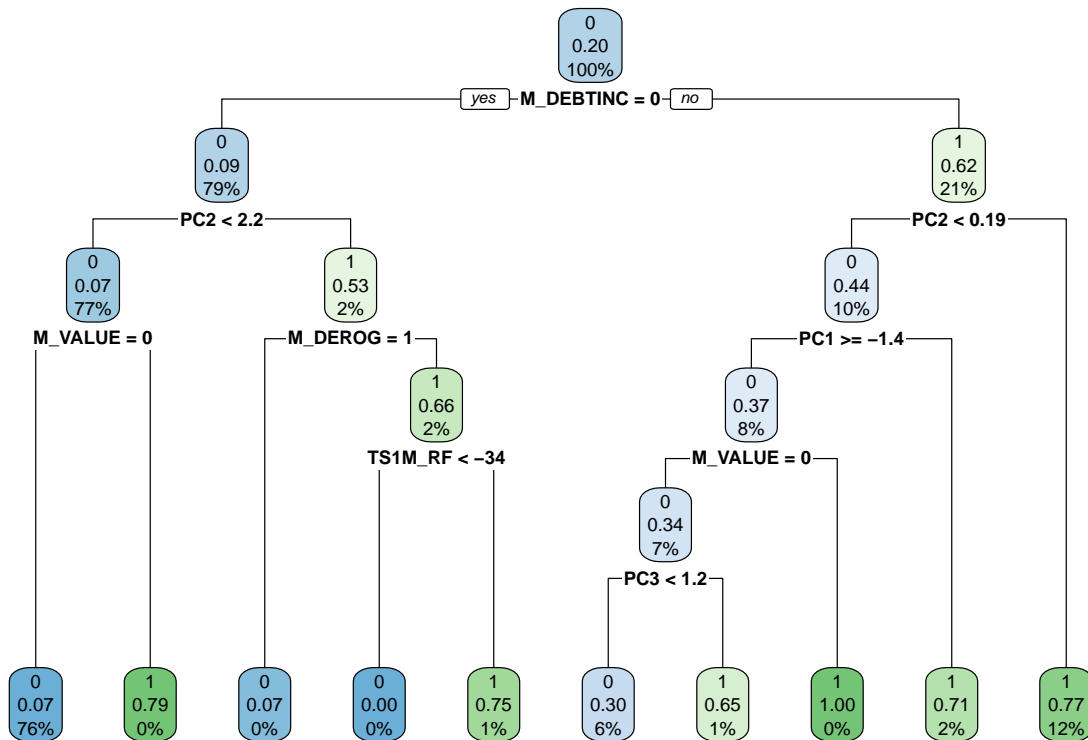
```

```
## 5 -0.5746093 -0.2924981 1.44854530 1.349271483 -16.430504 10.430375
## 6 -2.3178901 -0.2111695 0.77284627 -0.004773515 17.474265 -29.442652
```

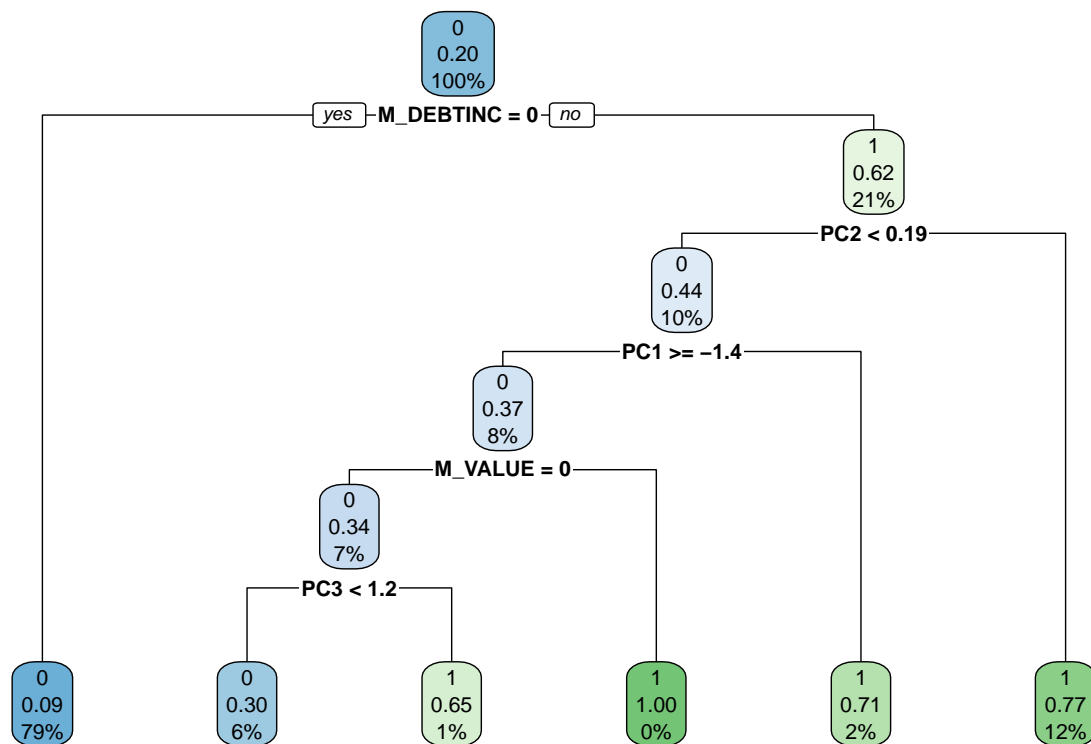
*#Create a Decision Tree to predict Loan Default (Target Flag=1).*

```
tr_set = rpart.control( maxdepth = 10 )
t1G = rpart( data=df_model, TARGET_BAD_FLAG ~ .,
             control=tr_set, method="class", parms=list(split='gini') )
t1E = rpart( data=df_model, TARGET_BAD_FLAG ~ .,
             control=tr_set, method="class", parms=list(split='information') )

rpart.plot( t1G )
```



```
rpart.plot( t1E )
```



```
t1G$variable.importance
```

##	M_DEBTINC	PC2	M_VALUE	PC1	PC3
##	570.0210103	134.1439768	65.9302691	26.4676580	21.8390681
##	TS1M_RF	M_DEROG	M_DELINQ	M_YOJ	TS2M_RF
##	15.5301423	15.0433225	10.9292905	10.9292905	10.5012784
##	M_NINQ	M_MORTDUE	PC4	FLAG.Job.Office	M_CLAGE
##	9.9357186	8.9421468	0.8033029	0.6005814	0.2111148

```
t1E$variable.importance
```

##	M_DEBTINC	PC2	M_VALUE	PC1	PC3
##	762.5912102	92.9118996	48.6278204	26.6244030	23.1220866
##	TS2M_RF	TS1M_RF	FLAG.Job.Office	M_CLAGE	
##	7.0199247	2.5842634	0.6460658	0.2209512	

*#Comment on the variables that were included in the model.*

*#Did any of the Principal Components or tSNE values make it into the model? Discuss why or why not.*

*#The trees are amazing with those man-made PC and TS variables included!*

*#For both trees, debt-income ratio flag is the most important one.*

*#However, our PC2 plays a very big role. And then PC1 and PC3.*

*#This means our PC analysis provides extra information for our trees.*

```
#Create a Logistic Regression model to predict Loan Default (Target Flag=1).
#Use either Forward, Backward, or Stepwise variable selection.
theUpper_LR = glm( TARGET_BAD_FLAG ~ ., family = "binomial", data = df_model )
theLower_LR = glm( TARGET_BAD_FLAG ~ 1, family = "binomial", data = df_model )
```

```
summary( theUpper_LR )
```

```
##
## Call:
## glm(formula = TARGET_BAD_FLAG ~ ., family = "binomial", data = df_model)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -4.8086924   0.4409957  -10.904 < 2e-16 ***
## M_MORTDUE         0.5000348   0.1916988    2.608 0.009096 **
## M_VALUE          4.8361373   0.5007873    9.657 < 2e-16 ***
## M_YOJ           -0.5714085   0.1926845   -2.966 0.003022 **
## M_DEROG         -1.9950009   0.2699225   -7.391 1.46e-13 ***
## M_DELINQ        -1.9031009   0.3868964   -4.919 8.70e-07 ***
## M_CLAGE          1.1716156   0.3193534    3.669 0.000244 ***
## M_NINQ          -0.0039769   0.3508355   -0.011 0.990956
## M_CLNO           2.8977404   0.5719932    5.066 4.06e-07 ***
## M_DEBTINC        2.6057042   0.0912360   28.560 < 2e-16 ***
## FLAG.Job.Mgr      2.4314441   0.4285850    5.673 1.40e-08 ***
## FLAG.Job.Office   1.9128561   0.4296918    4.452 8.52e-06 ***
## FLAG.Job.Other    2.5564828   0.4159263    6.146 7.92e-10 ***
## FLAG.Job.ProfExe  2.2010474   0.4255837    5.172 2.32e-07 ***
## FLAG.Job.Sales    3.4342196   0.4985177    6.889 5.62e-12 ***
## FLAG.Job.Self     2.8587349   0.4741742    6.029 1.65e-09 ***
## FLAG.Reason.DebtCon -0.0605061   0.2984749   -0.203 0.839356
## FLAG.Reason.HomeImp 0.2496983   0.3032695    0.823 0.410306
## PC1              -0.0331033   0.0346914   -0.954 0.339972
## PC2               0.9375971   0.0448293   20.915 < 2e-16 ***
## PC3               0.0209800   0.0416561    0.504 0.614508
## PC4              -0.0258861   0.0456354   -0.567 0.570552
## TS1M_RF           0.0079293   0.0027006    2.936 0.003323 **
## TS2M_RF          -0.0003729   0.0026198   -0.142 0.886800
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 5956.5  on 5959  degrees of freedom
## Residual deviance: 3451.9  on 5936  degrees of freedom
## AIC: 3499.9
##
## Number of Fisher Scoring iterations: 6
```

```
summary( theLower_LR )
```

```
##
## Call:
```



```
## glm(formula = TARGET_BAD_FLAG ~ 1, family = "binomial", data = df_model)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.38944    0.03241  -42.87  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 5956.5  on 5959  degrees of freedom
## Residual deviance: 5956.5  on 5959  degrees of freedom
## AIC: 5958.5
##
## Number of Fisher Scoring iterations: 4
```

```
lr_model = stepAIC(theLower_LR, direction="forward",
                  scope=list(lower=theLower_LR, upper=theUpper_LR))
```

```
## Start:  AIC=5958.47
## TARGET_BAD_FLAG ~ 1
##
##              Df Deviance    AIC
## + M_DEBTINC    1  4431.3 4435.3
## + PC2          1  5125.1 5129.1
## + M_VALUE      1  5659.8 5663.8
## + FLAG.Job.Office 1  5921.4 5925.4
## + M_DEROG      1  5923.8 5927.8
## + FLAG.Job.Other 1  5930.4 5934.4
## + M_DELINQ     1  5931.2 5935.2
## + M_YOJ        1  5935.5 5939.5
## + PC1          1  5939.9 5943.9
## + FLAG.Job.Sales 1  5943.0 5947.0
## + FLAG.Job.ProfExe 1  5944.8 5948.8
## + FLAG.Job.Self 1  5945.0 5949.0
## + M_NINQ       1  5946.2 5950.2
## + TS1M_RF      1  5946.3 5950.3
## + TS2M_RF      1  5947.6 5951.6
## + FLAG.Reason.HomeImp 1  5948.2 5952.2
## + FLAG.Reason.DebtCon 1  5949.6 5953.6
## + FLAG.Job.Mgr  1  5950.4 5954.4
## + M_CLAGE      1  5950.9 5954.9
## + PC3          1  5951.6 5955.6
## + PC4          1  5953.3 5957.3
## + M_CLNO       1  5954.3 5958.3
## <none>         5956.5 5958.5
## + M_MORTDUE    1  5956.4 5960.4
##
## Step:  AIC=4435.29
## TARGET_BAD_FLAG ~ M_DEBTINC
##
##              Df Deviance    AIC
## + PC2          1  3956.0 3962.0
## + M_VALUE      1  4265.9 4271.9
```

```

## + M_DEROG          1  4401.3 4407.3
## + FLAG.Job.Office   1  4409.9 4415.9
## + M_DELINQ          1  4409.9 4415.9
## + FLAG.Job.Other    1  4412.3 4418.3
## + M_YOJ             1  4417.6 4423.6
## + FLAG.Job.Sales    1  4419.9 4425.9
## + FLAG.Job.ProfExe  1  4421.6 4427.6
## + FLAG.Job.Self     1  4422.8 4428.8
## + M_NINQ            1  4423.8 4429.8
## + FLAG.Reason.HomeImp 1  4427.4 4433.4
## + PC4               1  4427.8 4433.8
## + FLAG.Job.Mgr      1  4427.8 4433.8
## + FLAG.Reason.DebtCon 1  4428.7 4434.7
## + TS1M_RF           1  4428.8 4434.8
## <none>              4431.3 4435.3
## + M_CLAGE           1  4429.4 4435.4
## + PC1               1  4429.5 4435.5
## + TS2M_RF           1  4430.6 4436.6
## + M_CLNO            1  4431.0 4437.0
## + PC3               1  4431.1 4437.1
## + M_MORTDUE         1  4431.3 4437.3
##
## Step:  AIC=3962.01
## TARGET_BAD_FLAG ~ M_DEBTINC + PC2
##
##              Df Deviance    AIC
## + M_VALUE          1  3817.4 3825.4
## + M_DEROG          1  3835.3 3843.3
## + M_DELINQ         1  3875.1 3883.1
## + M_NINQ           1  3923.3 3931.3
## + FLAG.Job.Other   1  3939.0 3947.0
## + FLAG.Job.Office  1  3939.1 3947.1
## + FLAG.Reason.HomeImp 1  3939.4 3947.4
## + M_YOJ            1  3939.6 3947.6
## + FLAG.Reason.DebtCon 1  3944.0 3952.0
## + FLAG.Job.Sales   1  3945.8 3953.8
## + FLAG.Job.Self    1  3947.2 3955.2
## + PC3              1  3950.3 3958.3
## + TS1M_RF          1  3951.3 3959.3
## + PC1              1  3952.3 3960.3
## + TS2M_RF          1  3953.2 3961.2
## + FLAG.Job.ProfExe 1  3953.8 3961.8
## <none>             3956.0 3962.0
## + PC4              1  3954.9 3962.9
## + M_CLNO           1  3955.1 3963.1
## + M_CLAGE          1  3955.1 3963.1
## + M_MORTDUE        1  3955.9 3963.9
## + FLAG.Job.Mgr     1  3956.0 3964.0
##
## Step:  AIC=3825.36
## TARGET_BAD_FLAG ~ M_DEBTINC + PC2 + M_VALUE
##
##              Df Deviance    AIC
## + M_DEROG          1  3668.1 3678.1

```

```

## + M_DELINQ          1    3712.6 3722.6
## + M_NINQ             1    3773.4 3783.4
## + M_YOJ              1    3796.7 3806.7
## + FLAG.Job.Other     1    3797.2 3807.2
## + FLAG.Reason.HomeImp 1    3799.6 3809.6
## + FLAG.Job.Office    1    3800.2 3810.2
## + FLAG.Job.Sales     1    3805.9 3815.9
## + PC3                1    3806.2 3816.2
## + FLAG.Job.Self      1    3808.9 3818.9
## + TS1M_RF            1    3809.4 3819.4
## + FLAG.Reason.DebtCon 1    3809.4 3819.4
## + M_CLNO             1    3811.5 3821.5
## + TS2M_RF            1    3812.6 3822.6
## + PC1                1    3814.0 3824.0
## + PC4                1    3815.2 3825.2
## <none>               3817.4 3825.4
## + FLAG.Job.ProfExe    1    3815.9 3825.9
## + M_MORTDUE           1    3816.4 3826.4
## + M_CLAGE             1    3817.3 3827.3
## + FLAG.Job.Mgr        1    3817.3 3827.3
##
## Step:  AIC=3678.08
## TARGET_BAD_FLAG ~ M_DEBTINC + PC2 + M_VALUE + M_DEROG
##
##              Df Deviance    AIC
## + M_CLAGE          1    3626.8 3638.8
## + M_CLNO           1    3632.1 3644.1
## + FLAG.Job.Office   1    3650.2 3662.2
## + FLAG.Reason.HomeImp 1    3650.8 3662.8
## + FLAG.Job.Other     1    3653.4 3665.4
## + FLAG.Reason.DebtCon 1    3655.7 3667.7
## + FLAG.Job.Sales     1    3657.6 3669.6
## + TS1M_RF           1    3659.4 3671.4
## + M_YOJ             1    3661.7 3673.7
## + FLAG.Job.Self      1    3662.2 3674.2
## + M_DELINQ          1    3662.5 3674.5
## + FLAG.Job.ProfExe    1    3664.7 3676.7
## + PC1               1    3665.1 3677.1
## + TS2M_RF           1    3665.5 3677.5
## + PC3               1    3665.8 3677.8
## <none>              3668.1 3678.1
## + M_MORTDUE          1    3666.7 3678.7
## + M_NINQ            1    3667.8 3679.8
## + FLAG.Job.Mgr       1    3668.0 3680.0
## + PC4               1    3668.1 3680.1
##
## Step:  AIC=3638.82
## TARGET_BAD_FLAG ~ M_DEBTINC + PC2 + M_VALUE + M_DEROG + M_CLAGE
##
##              Df Deviance    AIC
## + M_DELINQ          1    3607.5 3621.5
## + FLAG.Reason.HomeImp 1    3608.6 3622.6
## + M_YOJ             1    3611.2 3625.2
## + FLAG.Job.Office    1    3612.2 3626.2

```

```

## + FLAG.Job.Other      1  3613.3 3627.3
## + FLAG.Job.Sales      1  3614.9 3628.9
## + TS1M_RF            1  3615.0 3629.0
## + FLAG.Reason.DebtCon 1  3617.1 3631.1
## + FLAG.Job.Self       1  3620.7 3634.7
## + M_CLNO             1  3622.4 3636.4
## + M_NINQ             1  3623.6 3637.6
## + FLAG.Job.ProfExe    1  3624.4 3638.4
## <none>               3626.8 3638.8
## + TS2M_RF            1  3625.4 3639.4
## + PC1                1  3625.8 3639.8
## + PC3                1  3626.0 3640.0
## + FLAG.Job.Mgr       1  3626.4 3640.4
## + M_MORTDUE          1  3626.7 3640.7
## + PC4                1  3626.8 3640.8
##
## Step:  AIC=3621.51
## TARGET_BAD_FLAG ~ M_DEBTINC + PC2 + M_VALUE + M_DEROG + M_CLAGE +
##      M_DELINQ
##
##              Df Deviance    AIC
## + FLAG.Reason.HomeImp 1  3588.1 3604.1
## + M_CLNO              1  3589.8 3605.8
## + FLAG.Job.Office     1  3591.9 3607.9
## + FLAG.Job.Other      1  3592.6 3608.6
## + FLAG.Job.Sales      1  3595.4 3611.4
## + FLAG.Reason.DebtCon 1  3596.2 3612.2
## + M_YOJ               1  3597.0 3613.0
## + TS1M_RF            1  3597.1 3613.1
## + FLAG.Job.Self       1  3600.6 3616.6
## + FLAG.Job.ProfExe    1  3604.7 3620.7
## <none>               3607.5 3621.5
## + M_MORTDUE          1  3606.0 3622.0
## + PC1                1  3606.1 3622.1
## + TS2M_RF            1  3606.2 3622.2
## + PC3                1  3606.5 3622.5
## + M_NINQ             1  3606.6 3622.6
## + FLAG.Job.Mgr       1  3607.3 3623.3
## + PC4                1  3607.5 3623.5
##
## Step:  AIC=3604.13
## TARGET_BAD_FLAG ~ M_DEBTINC + PC2 + M_VALUE + M_DEROG + M_CLAGE +
##      M_DELINQ + FLAG.Reason.HomeImp
##
##              Df Deviance    AIC
## + M_CLNO              1  3571.9 3589.9
## + FLAG.Job.Office     1  3573.1 3591.1
## + FLAG.Job.Other      1  3574.1 3592.1
## + FLAG.Job.Sales      1  3574.7 3592.7
## + TS1M_RF            1  3577.2 3595.2
## + M_YOJ               1  3578.8 3596.8
## + FLAG.Job.Self       1  3583.5 3601.5
## + FLAG.Reason.DebtCon 1  3584.1 3602.1
## + FLAG.Job.ProfExe    1  3585.1 3603.1

```

```

## <none>                3588.1 3604.1
## + PC3                  1   3587.0 3605.0
## + PC1                  1   3587.5 3605.5
## + TS2M_RF              1   3587.5 3605.5
## + M_NINQ               1   3587.5 3605.5
## + FLAG.Job.Mgr         1   3587.6 3605.6
## + M_MORTDUE            1   3587.7 3605.7
## + PC4                  1   3587.7 3605.7
##
## Step:  AIC=3589.9
## TARGET_BAD_FLAG ~ M_DEBTINC + PC2 + M_VALUE + M_DEROG + M_CLAGE +
##      M_DELINQ + FLAG.Reason.HomeImp + M_CLNO
##
##              Df Deviance    AIC
## + FLAG.Job.Other      1   3554.3 3574.3
## + FLAG.Job.Office     1   3556.8 3576.8
## + FLAG.Job.Sales      1   3558.0 3578.0
## + TS1M_RF             1   3561.7 3581.7
## + M_YOJ               1   3562.7 3582.7
## + FLAG.Reason.DebtCon  1   3566.1 3586.1
## + FLAG.Job.Self       1   3567.3 3587.3
## + FLAG.Job.ProfExe    1   3568.3 3588.3
## <none>                3571.9 3589.9
## + PC3                  1   3570.3 3590.3
## + PC4                  1   3570.9 3590.9
## + PC1                  1   3571.1 3591.1
## + TS2M_RF             1   3571.2 3591.2
## + M_MORTDUE           1   3571.2 3591.2
## + FLAG.Job.Mgr        1   3571.4 3591.4
## + M_NINQ              1   3571.7 3591.7
##
## Step:  AIC=3574.34
## TARGET_BAD_FLAG ~ M_DEBTINC + PC2 + M_VALUE + M_DEROG + M_CLAGE +
##      M_DELINQ + FLAG.Reason.HomeImp + M_CLNO + FLAG.Job.Other
##
##              Df Deviance    AIC
## + FLAG.Job.Sales      1   3535.6 3557.6
## + M_YOJ               1   3541.1 3563.1
## + TS1M_RF             1   3543.8 3565.8
## + FLAG.Job.Self       1   3545.8 3567.8
## + FLAG.Job.Office     1   3547.7 3569.7
## + FLAG.Job.Mgr        1   3549.1 3571.1
## + FLAG.Reason.DebtCon  1   3549.9 3571.9
## <none>                3554.3 3574.3
## + PC3                  1   3553.0 3575.0
## + PC4                  1   3554.1 3576.1
## + M_MORTDUE           1   3554.1 3576.1
## + M_NINQ              1   3554.1 3576.1
## + PC1                  1   3554.3 3576.3
## + TS2M_RF             1   3554.3 3576.3
## + FLAG.Job.ProfExe    1   3554.3 3576.3
##
## Step:  AIC=3557.61
## TARGET_BAD_FLAG ~ M_DEBTINC + PC2 + M_VALUE + M_DEROG + M_CLAGE +

```

```

##      M_DELINQ + FLAG.Reason.HomeImp + M_CLNO + FLAG.Job.Other +
##      FLAG.Job.Sales
##
##              Df Deviance      AIC
## + M_YOJ          1    3521.8 3545.8
## + TS1M_RF        1    3525.7 3549.7
## + FLAG.Job.Self   1    3525.8 3549.8
## + FLAG.Job.Mgr    1    3527.7 3551.7
## + FLAG.Job.Office 1    3531.3 3555.3
## + FLAG.Reason.DebtCon 1    3531.7 3555.7
## <none>              3535.6 3557.6
## + PC3            1    3534.4 3558.4
## + PC4            1    3535.0 3559.0
## + FLAG.Job.ProfExe 1    3535.3 3559.3
## + M_MORTDUE       1    3535.4 3559.4
## + M_NINQ          1    3535.5 3559.5
## + PC1            1    3535.5 3559.5
## + TS2M_RF        1    3535.6 3559.6
##
## Step:  AIC=3545.79
## TARGET_BAD_FLAG ~ M_DEBTINC + PC2 + M_VALUE + M_DEROG + M_CLAGE +
##      M_DELINQ + FLAG.Reason.HomeImp + M_CLNO + FLAG.Job.Other +
##      FLAG.Job.Sales + M_YOJ
##
##              Df Deviance      AIC
## + FLAG.Job.Self   1    3511.7 3537.7
## + TS1M_RF        1    3513.0 3539.0
## + FLAG.Job.Mgr    1    3513.9 3539.9
## + FLAG.Job.Office 1    3517.2 3543.2
## + FLAG.Reason.DebtCon 1    3519.8 3545.8
## <none>              3521.8 3545.8
## + M_MORTDUE       1    3520.3 3546.3
## + PC3            1    3520.7 3546.7
## + PC4            1    3520.9 3546.9
## + TS2M_RF        1    3521.6 3547.6
## + FLAG.Job.ProfExe 1    3521.7 3547.7
## + M_NINQ          1    3521.7 3547.7
## + PC1            1    3521.7 3547.7
##
## Step:  AIC=3537.68
## TARGET_BAD_FLAG ~ M_DEBTINC + PC2 + M_VALUE + M_DEROG + M_CLAGE +
##      M_DELINQ + FLAG.Reason.HomeImp + M_CLNO + FLAG.Job.Other +
##      FLAG.Job.Sales + M_YOJ + FLAG.Job.Self
##
##              Df Deviance      AIC
## + FLAG.Job.Mgr    1    3500.6 3528.6
## + TS1M_RF        1    3503.8 3531.8
## + FLAG.Job.Office 1    3508.9 3536.9
## + FLAG.Reason.DebtCon 1    3509.7 3537.7
## <none>              3511.7 3537.7
## + M_MORTDUE       1    3510.0 3538.0
## + FLAG.Job.ProfExe 1    3510.6 3538.6
## + PC3            1    3511.0 3539.0
## + PC4            1    3511.1 3539.1

```

```

## + PC1                1    3511.6 3539.6
## + TS2M_RF            1    3511.6 3539.6
## + M_NINQ             1    3511.7 3539.7
##
## Step:  AIC=3528.63
## TARGET_BAD_FLAG ~ M_DEBTINC + PC2 + M_VALUE + M_DEROG + M_CLAGE +
##      M_DELINQ + FLAG.Reason.HomeImp + M_CLNO + FLAG.Job.Other +
##      FLAG.Job.Sales + M_YOJ + FLAG.Job.Self + FLAG.Job.Mgr
##
##              Df Deviance    AIC
## + FLAG.Job.ProfExe    1    3490.8 3520.8
## + TS1M_RF            1    3492.7 3522.7
## <none>                3500.6 3528.6
## + M_MORTDUE          1    3498.7 3528.7
## + FLAG.Reason.DebtCon 1    3499.0 3529.0
## + PC3                1    3500.0 3530.0
## + FLAG.Job.Office     1    3500.4 3530.4
## + PC4                1    3500.5 3530.5
## + PC1                1    3500.6 3530.6
## + TS2M_RF            1    3500.6 3530.6
## + M_NINQ             1    3500.6 3530.6
##
## Step:  AIC=3520.78
## TARGET_BAD_FLAG ~ M_DEBTINC + PC2 + M_VALUE + M_DEROG + M_CLAGE +
##      M_DELINQ + FLAG.Reason.HomeImp + M_CLNO + FLAG.Job.Other +
##      FLAG.Job.Sales + M_YOJ + FLAG.Job.Self + FLAG.Job.Mgr + FLAG.Job.ProfExe
##
##              Df Deviance    AIC
## + FLAG.Job.Office     1    3467.7 3499.7
## + TS1M_RF            1    3483.8 3515.8
## + M_MORTDUE          1    3487.8 3519.8
## <none>                3490.8 3520.8
## + FLAG.Reason.DebtCon 1    3489.7 3521.7
## + PC1                1    3490.0 3522.0
## + PC3                1    3490.2 3522.2
## + TS2M_RF            1    3490.5 3522.5
## + PC4                1    3490.5 3522.5
## + M_NINQ             1    3490.8 3522.8
##
## Step:  AIC=3499.66
## TARGET_BAD_FLAG ~ M_DEBTINC + PC2 + M_VALUE + M_DEROG + M_CLAGE +
##      M_DELINQ + FLAG.Reason.HomeImp + M_CLNO + FLAG.Job.Other +
##      FLAG.Job.Sales + M_YOJ + FLAG.Job.Self + FLAG.Job.Mgr + FLAG.Job.ProfExe +
##      FLAG.Job.Office
##
##              Df Deviance    AIC
## + TS1M_RF            1    3460.8 3494.8
## + M_MORTDUE          1    3462.6 3496.6
## <none>                3467.7 3499.7
## + PC1                1    3466.7 3500.7
## + PC3                1    3466.9 3500.9
## + TS2M_RF            1    3467.3 3501.3
## + PC4                1    3467.5 3501.5
## + FLAG.Reason.DebtCon 1    3467.6 3501.6

```

```
## + M_NINQ          1    3467.7 3501.7
##
## Step:  AIC=3494.81
## TARGET_BAD_FLAG ~ M_DEBTINC + PC2 + M_VALUE + M_DEROG + M_CLAGE +
##      M_DELINQ + FLAG.Reason.HomeImp + M_CLNO + FLAG.Job.Other +
##      FLAG.Job.Sales + M_YOJ + FLAG.Job.Self + FLAG.Job.Mgr + FLAG.Job.ProfExe +
##      FLAG.Job.Office + TS1M_RF
##
##
##              Df Deviance    AIC
## + M_MORTDUE      1    3453.8 3489.8
## <none>              3460.8 3494.8
## + PC1            1    3459.4 3495.4
## + PC4            1    3460.4 3496.4
## + PC3            1    3460.5 3496.5
## + TS2M_RF        1    3460.7 3496.7
## + FLAG.Reason.DebtCon 1    3460.7 3496.7
## + M_NINQ          1    3460.8 3496.8
##
## Step:  AIC=3489.78
## TARGET_BAD_FLAG ~ M_DEBTINC + PC2 + M_VALUE + M_DEROG + M_CLAGE +
##      M_DELINQ + FLAG.Reason.HomeImp + M_CLNO + FLAG.Job.Other +
##      FLAG.Job.Sales + M_YOJ + FLAG.Job.Self + FLAG.Job.Mgr + FLAG.Job.ProfExe +
##      FLAG.Job.Office + TS1M_RF + M_MORTDUE
##
##
##              Df Deviance    AIC
## <none>              3453.8 3489.8
## + PC1            1    3452.6 3490.6
## + PC4            1    3453.4 3491.4
## + PC3            1    3453.5 3491.5
## + TS2M_RF        1    3453.5 3491.5
## + FLAG.Reason.DebtCon 1    3453.8 3491.8
## + M_NINQ          1    3453.8 3491.8
```

```
summary( lr_model )
```

```
##
## Call:
## glm(formula = TARGET_BAD_FLAG ~ M_DEBTINC + PC2 + M_VALUE + M_DEROG +
##      M_CLAGE + M_DELINQ + FLAG.Reason.HomeImp + M_CLNO + FLAG.Job.Other +
##      FLAG.Job.Sales + M_YOJ + FLAG.Job.Self + FLAG.Job.Mgr + FLAG.Job.ProfExe +
##      FLAG.Job.Office + TS1M_RF + M_MORTDUE, family = "binomial",
##      data = df_model)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -4.839458   0.405074 -11.947  < 2e-16 ***
## M_DEBTINC      2.612057   0.090460  28.875  < 2e-16 ***
## PC2           0.938396   0.044692  20.997  < 2e-16 ***
## M_VALUE        4.844441   0.497376   9.740  < 2e-16 ***
## M_DEROG       -2.034149   0.266839  -7.623 2.48e-14 ***
## M_CLAGE        1.218129   0.317273   3.839 0.000123 ***
## M_DELINQ       -1.895128   0.346811  -5.464 4.64e-08 ***
## FLAG.Reason.HomeImp 0.310595   0.097906   3.172 0.001512 **
## M_CLNO         2.867242   0.545798   5.253 1.49e-07 ***
```



```
## FLAG.Job.Other      2.550325    0.399034    6.391 1.65e-10 ***
## FLAG.Job.Sales      3.385296    0.482748    7.013 2.34e-12 ***
## M_YOJ              -0.565182    0.190062   -2.974 0.002943 **
## FLAG.Job.Self       2.799692    0.457214    6.123 9.16e-10 ***
## FLAG.Job.Mgr        2.401406    0.412815    5.817 5.99e-09 ***
## FLAG.Job.ProfExe    2.142394    0.409591    5.231 1.69e-07 ***
## FLAG.Job.Office     1.887622    0.414580    4.553 5.29e-06 ***
## TS1M_RF             0.007777    0.002624    2.963 0.003045 **
## M_MORTDUE           0.512219    0.190443    2.690 0.007153 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 5956.5  on 5959  degrees of freedom
## Residual deviance: 3453.8  on 5942  degrees of freedom
## AIC: 3489.8
##
## Number of Fisher Scoring iterations: 6
```

```
#Comment on the variables that were included in the model.
#Did any of the Principal Components or tSNE values make it into the model? Discuss why or why not.

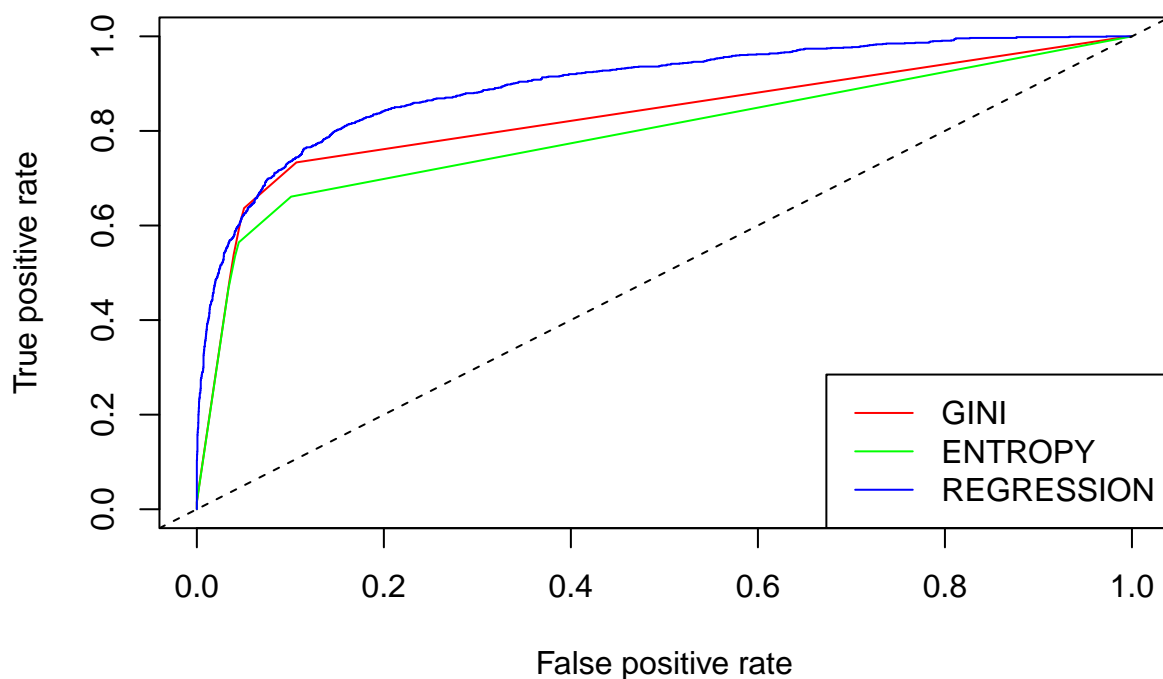
#Fortunately we have our PC2 and TS1M_RF included in the model.
#As we have expected, PC2 has a small Std. Error and a high z value,
#which means PC2 is very precise and accurate when it comes to predicting.
#TS1M_RF is not so good in this model.

#Create a ROC curve showing the accuracy of the model.
#Calculate and display the Area Under the ROC Curve (AUC).
pG = predict( t1G, df_model )
pG2 = prediction( pG[,2], df_model$TARGET_BAD_FLAG )
pG3 = performance( pG2, "tpr", "fpr" )

pE = predict( t1E, df_model )
pE2 = prediction( pE[,2], df_model$TARGET_BAD_FLAG )
pE3 = performance( pE2, "tpr", "fpr" )

plr = predict( lr_model, df_model, type="response" )
plr2 = prediction( plr, df_model$TARGET_BAD_FLAG )
plr3 = performance( plr2, "tpr", "fpr" )

plot( pG3, col="red" )
plot( pE3, col="green", add=TRUE )
plot( plr3, col="blue", add=TRUE )
abline(0,1,lty=2)
legend("bottomright",c("GINI","ENTROPY","REGRESSION"),
      col=c("red","green","blue"), bty="n", lty=1 )
```



```
aucG = performance( pG2, "auc" )@y.values
aucE = performance( pE2, "auc" )@y.values
aucR = performance( plr2, "auc" )@y.values
```

```
print( aucG )
```

```
## [[1]]
## [1] 0.830688
```

```
print( aucE )
```

```
## [[1]]
## [1] 0.7949586
```

```
print( aucR )
```

```
## [[1]]
## [1] 0.8974398
```

```
#Step 6: Comment
#Discuss how the PCA / tSNE values performed when compared to the original data set.
#I think PCA/tSNE is a really good way to make our prediction precise.
```

*#In the original data set, there are so many variables and when we create models,  
#Some important information may be ignored.  
#However, with the engineered lower-dimension values added to the original data set,  
#we can create a model more robust and include more details.  
#It is also a good way to find some potential risks that are not so easy to be  
#identified in the original data set.  
#So if we have a data set with many variables and rows,  
#it is wise to use PCA or tSNE to get some neutral and extra information.  
#This is more persuasive when we deliver our analysis to the manager or clients.*