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

```
## randomForest 4.7-1.1

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

```
library(gbm)
```

```
## Loaded gbm 2.1.9

## This version of gbm is no longer under development. Consider transitioning to gbm3, https://github.c
```

```
library(MASS)
library(Rtsne)
library(ggplot2)
```

```
##
## Attaching package: 'ggplot2'

## The following object is masked from 'package:randomForest':
##
##     margin
```

```
library(flexclust)
```

```
## Loading required package: grid

## Loading required package: lattice

## Loading required package: modeltools

## Loading required package: stats4
```

```
data <- read.csv("C:/Users/hp/Downloads/HMEQ_Scrubbed (3)/HMEQ_Scrubbed.csv")
str(data)
```

Step-1

```
## 'data.frame':    5960 obs. of  29 variables:
##  $ TARGET_BAD_FLAG    : int  1 1 1 1 0 1 1 1 1 1 ...
##  $ TARGET_LOSS_AMT    : int  641 1109 767 1425 0 335 1841 373 1217 1523 ...
##  $ LOAN               : int  1100 1300 1500 1500 1700 1700 1800 1800 2000 2000 ...
##  $ IMP_MORTDUE        : num  25860 70053 13500 65000 97800 ...
```

```
##  $ M_MORTDUE          : int  0 0 0 1 0 0 0 0 0 1 ...
##  $ IMP_VALUE          : num  39025 68400 16700 89000 112000 ...
##  $ M_VALUE            : int  0 0 0 1 0 0 0 0 0 0 ...
##  $ IMP_YOJ            : num  10.5 7 4 7 3 9 5 11 3 16 ...
##  $ M_YOJ              : int  0 0 0 1 0 0 0 0 0 0 ...
##  $ IMP_DEROG          : int  0 0 0 1 0 0 3 0 0 0 ...
##  $ M_DEROG            : int  0 0 0 1 0 0 0 0 0 0 ...
##  $ IMP_DELINQ         : int  0 2 0 1 0 0 2 0 2 0 ...
##  $ M_DELINQ           : int  0 0 0 1 0 0 0 0 0 0 ...
##  $ IMP_CLAGE          : num  94.4 121.8 149.5 174 93.3 ...
##  $ M_CLAGE            : int  0 0 0 1 0 0 0 0 0 0 ...
##  $ IMP_NINQ           : int  1 0 1 1 0 1 1 0 1 0 ...
##  $ M_NINQ             : int  0 0 0 1 0 0 0 0 0 0 ...
##  $ IMP_CLNO           : int  9 14 10 20 14 8 17 8 12 13 ...
##  $ M_CLNO             : int  0 0 0 1 0 0 0 0 0 0 ...
##  $ IMP_DEBTINC        : num  35 35 35 35 35 ...
##  $ M_DEBTINC          : int  1 1 1 1 1 0 1 0 1 1 ...
##  $ FLAG.Job.Mgr       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ FLAG.Job.Office    : int  0 0 0 0 1 0 0 0 0 0 ...
##  $ FLAG.Job.Other     : int  1 1 1 0 0 1 1 1 1 0 ...
##  $ FLAG.Job.ProfExe   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ FLAG.Job.Sales     : int  0 0 0 0 0 0 0 0 0 1 ...
##  $ FLAG.Job.Self      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ FLAG.Reason.DebtCon: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ FLAG.Reason.HomeImp: int  1 1 1 0 1 1 1 1 1 1 ...
```

```
summary(data)
```

```
##  TARGET_BAD_FLAG  TARGET_LOSS_AMT      LOAN        IMP_MORTDUE
##  Min.   :0.0000   Min.   :    0   Min.   : 1100   Min.   :  2063
##  1st Qu.:0.0000   1st Qu.:    0   1st Qu.:11100   1st Qu.: 48139
##  Median :0.0000   Median :    0   Median :16300   Median : 65000
##  Mean   :0.1995   Mean   : 2676   Mean   :18608   Mean   : 72999
##  3rd Qu.:0.0000   3rd Qu.:    0   3rd Qu.:23300   3rd Qu.: 88200
##  Max.   :1.0000   Max.   :78987   Max.   :89900   Max.   :399550
##    M_MORTDUE         IMP_VALUE         M_VALUE           IMP_YOJ
##  Min.   :0.00000   Min.   :  8000   Min.   :0.00000   Min.   : 0.000
##  1st Qu.:0.00000   1st Qu.: 66490   1st Qu.:0.00000   1st Qu.: 3.000
##  Median :0.00000   Median : 89000   Median :0.00000   Median : 7.000
##  Mean   :0.08691   Mean   :101536   Mean   :0.01879   Mean   : 8.756
##  3rd Qu.:0.00000   3rd Qu.:119005   3rd Qu.:0.00000   3rd Qu.:12.000
##  Max.   :1.00000   Max.   :855909   Max.   :1.00000   Max.   :41.000
##      M_YOJ           IMP_DEROG          M_DEROG         IMP_DELINQ
##  Min.   :0.00000   Min.   : 0.0000   Min.   :0.0000   Min.   : 0.000
##  1st Qu.:0.00000   1st Qu.: 0.0000   1st Qu.:0.0000   1st Qu.: 0.000
##  Median :0.00000   Median : 0.0000   Median :0.0000   Median : 0.000
##  Mean   :0.08641   Mean   : 0.3431   Mean   :0.1188   Mean   : 0.503
##  3rd Qu.:0.00000   3rd Qu.: 0.0000   3rd Qu.:0.0000   3rd Qu.: 1.000
##  Max.   :1.00000   Max.   :10.0000   Max.   :1.0000   Max.   :15.000
##     M_DELINQ         IMP_CLAGE         M_CLAGE           IMP_NINQ
##  Min.   :0.00000   Min.   :   0.0   Min.   :0.00000   Min.   : 0.00
##  1st Qu.:0.00000   1st Qu.: 117.4   1st Qu.:0.00000   1st Qu.: 0.00
##  Median :0.00000   Median : 174.0   Median :0.00000   Median : 1.00
##  Mean   :0.09732   Mean   : 179.5   Mean   :0.05168   Mean   : 1.17
```

```
##  3rd Qu.:0.00000   3rd Qu.: 227.1   3rd Qu.:0.00000   3rd Qu.: 2.00
##  Max.   :1.00000   Max.   :1168.2   Max.   :1.00000   Max.   :17.00
##      M_NINQ           IMP_CLNO         M_CLNO          IMP_DEBTINC
##  Min.   :0.00000   Min.   : 0.00   Min.   :0.00000   Min.   :  0.5245
##  1st Qu.:0.00000   1st Qu.:15.00   1st Qu.:0.00000   1st Qu.: 30.7632
##  Median :0.00000   Median :20.00   Median :0.00000   Median : 35.0000
##  Mean   :0.08557   Mean   :21.25   Mean   :0.03725   Mean   : 34.0393
##  3rd Qu.:0.00000   3rd Qu.:26.00   3rd Qu.:0.00000   3rd Qu.: 37.9499
##  Max.   :1.00000   Max.   :71.00   Max.   :1.00000   Max.   :203.3122
##    M_DEBTINC       FLAG.Job.Mgr    FLAG.Job.Office  FLAG.Job.Other
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :0.0000   Median :0.0000
##  Mean   :0.2126   Mean   :0.1287   Mean   :0.1591   Mean   :0.4007
##  3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##  FLAG.Job.ProfExe FLAG.Job.Sales    FLAG.Job.Self     FLAG.Reason.DebtCon
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :0.0000   Median :0.00000   Median :0.00000   Median :1.0000
##  Mean   :0.2141   Mean   :0.01829   Mean   :0.03238   Mean   :0.6591
##  3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.00000   Max.   :1.00000   Max.   :1.0000
##  FLAG.Reason.HomeImp
##  Min.   :0.0000
##  1st Qu.:0.0000
##  Median :0.0000
##  Mean   :0.2987
##  3rd Qu.:1.0000
##  Max.   :1.0000
```

```
head(data,6)
```

```
##   TARGET_BAD_FLAG TARGET_LOSS_AMT LOAN IMP_MORTDUE M_MORTDUE IMP_VALUE M_VALUE
## 1               1             641 1100       25860         0     39025       0
## 2               1            1109 1300       70053         0     68400       0
## 3               1             767 1500       13500         0     16700       0
## 4               1            1425 1500       65000         1     89000       1
## 5               0               0 1700       97800         0    112000       0
## 6               1             335 1700       30548         0     40320       0
##   IMP_YOJ M_YOJ IMP_DEROG M_DEROG IMP_DELINQ M_DELINQ IMP_CLAGE M_CLAGE
## 1    10.5     0         0       0          0        0  94.36667       0
## 2     7.0     0         0       0          2        0 121.83333       0
## 3     4.0     0         0       0          0        0 149.46667       0
## 4     7.0     1         1       1          1        1 174.00000       1
## 5     3.0     0         0       0          0        0  93.33333       0
## 6     9.0     0         0       0          0        0 101.46600       0
##   IMP_NINQ M_NINQ IMP_CLNO M_CLNO IMP_DEBTINC M_DEBTINC FLAG.Job.Mgr
## 1        1      0        9      0    35.00000         1            0
## 2        0      0       14      0    35.00000         1            0
## 3        1      0       10      0    35.00000         1            0
## 4        1      1       20      1    35.00000         1            0
## 5        0      0       14      0    35.00000         1            0
## 6        1      0        8      0    37.11361         0            0
```

```
##   FLAG.Job.Office FLAG.Job.Other FLAG.Job.ProfExe FLAG.Job.Sales FLAG.Job.Self
## 1               0              1                0              0             0
## 2               0              1                0              0             0
## 3               0              1                0              0             0
## 4               0              0                0              0             0
## 5               1              0                0              0             0
## 6               0              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
```

STEP-2

```
SEED=1
set.seed(SEED)
```

```
data_flag= data
data_flag$TARGET_LOSS_AMT= NULL
head(data_flag)
```

```
##   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_DELINQ M_DELINQ 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
```

```
FLAG= sample(c(TRUE,FALSE), nrow(data_flag), replace=TRUE, prob=c(0.7,0.3))
data_train= data_flag[FLAG, ]
data_test= data_flag[! FLAG, ]
```

```
dim(data_flag)
```

```
## [1] 5960   28
```

```
dim(data_train)
```

```
## [1] 4142   28
```

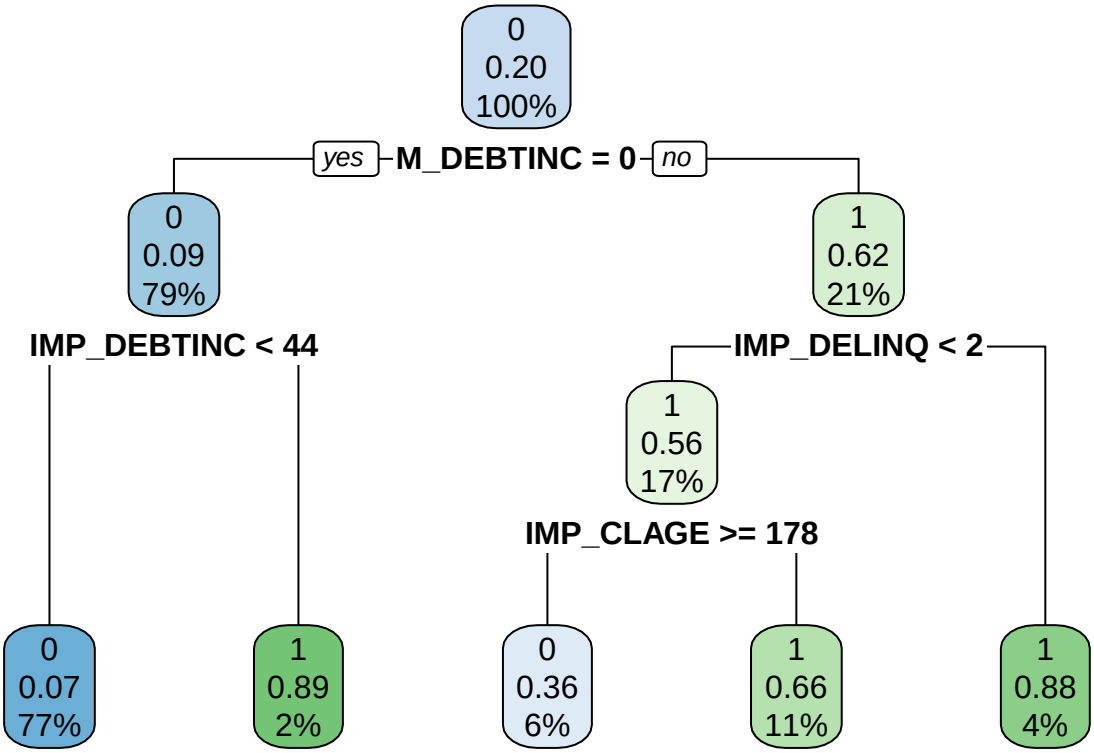
```
dim(data_test)
```

```
## [1] 1818   28
```

Decision Tree

```
tr_set=rpart.control(maxdepth=10)
t1E = rpart(data=data_train, TARGET_BAD_FLAG ~ ., control = tr_set, method = "class", parms = list(spli
```

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




```
t1E$variable.importance
```

```
##   M_DEBTINC IMP_DEBTINC  IMP_DELINQ   IMP_CLAGE        LOAN     M_VALUE
##  533.397481  134.588883   46.494397   30.749923   24.521888   22.199895
##   IMP_VALUE IMP_MORTDUE    IMP_CLNO     IMP_YOJ
##    7.967967    5.783975    2.459994    2.090995
```

```
pt = predict(t1E, data_test, type= "prob")
head(pt)
```

```
##            0         1
## 4  0.3354839 0.66451613
## 6  0.9315112 0.06848885
## 7  0.1206897 0.87931034
## 15 0.3354839 0.66451613
## 17 0.1206897 0.87931034
## 18 0.9315112 0.06848885
```

```
pt2 = prediction(pt[,2], data_test$TARGET_BAD_FLAG)
pt3 = performance(pt2, "tpr", "fpr")
```

```
pt = predict(t1E, data_test)
head(pt)
```

```
##            0         1
## 4  0.3354839 0.66451613
## 6  0.9315112 0.06848885
## 7  0.1206897 0.87931034
## 15 0.3354839 0.66451613
## 17 0.1206897 0.87931034
## 18 0.9315112 0.06848885
```

```
RMSEt = sqrt(mean((data_test$TARGET_BAD_FLAG - pt)^2))
```

```
rf_model = randomForest(data = data_train, TARGET_BAD_FLAG ~., ntree=500, importance= TRUE)
```

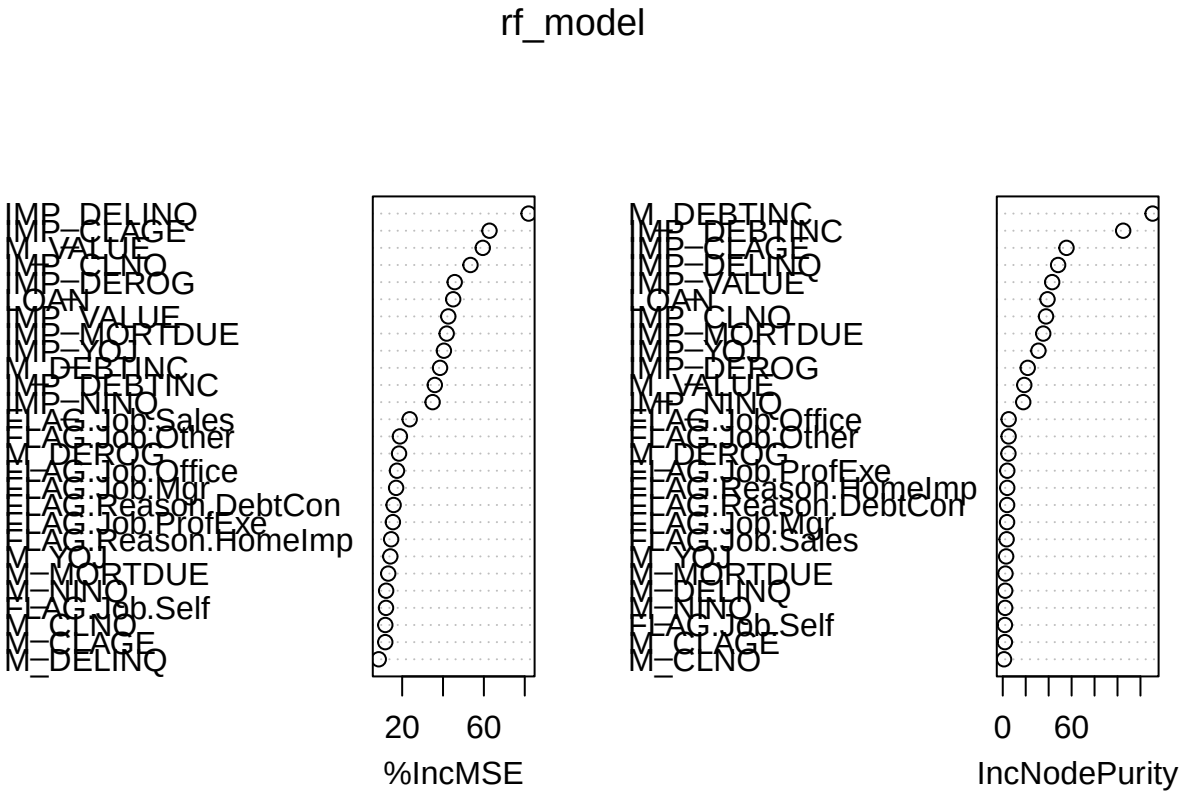
RF

```
## Warning in randomForest.default(m, y, ...): The response has five or fewer
## unique values.  Are you sure you want to do regression?
```

```
importance(rf_model)
```

```
##                       %IncMSE IncNodePurity
## LOAN                44.967305     38.993106
## IMP_MORTDUE         41.789344     35.266432
## M_MORTDUE           13.172790      2.349943
## IMP_VALUE           42.571780     43.056864
## M_VALUE             59.462869     18.739138
## IMP_YOJ             40.432069     31.086101
## M_YOJ               14.199750      2.944031
## IMP_DEROG           45.717324     21.754538
## M_DEROG             18.473874      4.951353
## IMP_DELINQ          81.884902     48.238429
## M_DELINQ             8.547209      2.107158
## IMP_CLAGE           62.739262     55.559885
## M_CLAGE             11.726470      1.840701
## IMP_NINQ            34.906103     17.852157
## M_NINQ              12.162285      2.082765
## IMP_CLNO            53.442201     37.694496
## M_CLNO              11.739013      1.103966
## IMP_DEBTINC         35.982961    104.953372
## M_DEBTINC           38.536277    130.591066
## FLAG.Job.Mgr        17.040006      3.765830
## FLAG.Job.Office     17.467948      5.148490
## FLAG.Job.Other      18.895132      5.102794
## FLAG.Job.ProfExe    15.455256      3.935220
## FLAG.Job.Sales      23.675630      3.364372
## FLAG.Job.Self       12.098948      1.883075
## FLAG.Reason.DebtCon 15.811019      3.814255
## FLAG.Reason.HomeImp 14.617020      3.908644
```

```
varImpPlot(rf_model)
```




```
pr= predict(rf_model, data_test)
head(pr)
```

```
##         4         6         7        15        17        18
## 0.7966524 0.8313333 0.9241667 0.7746333 0.9212667 0.3462000
```

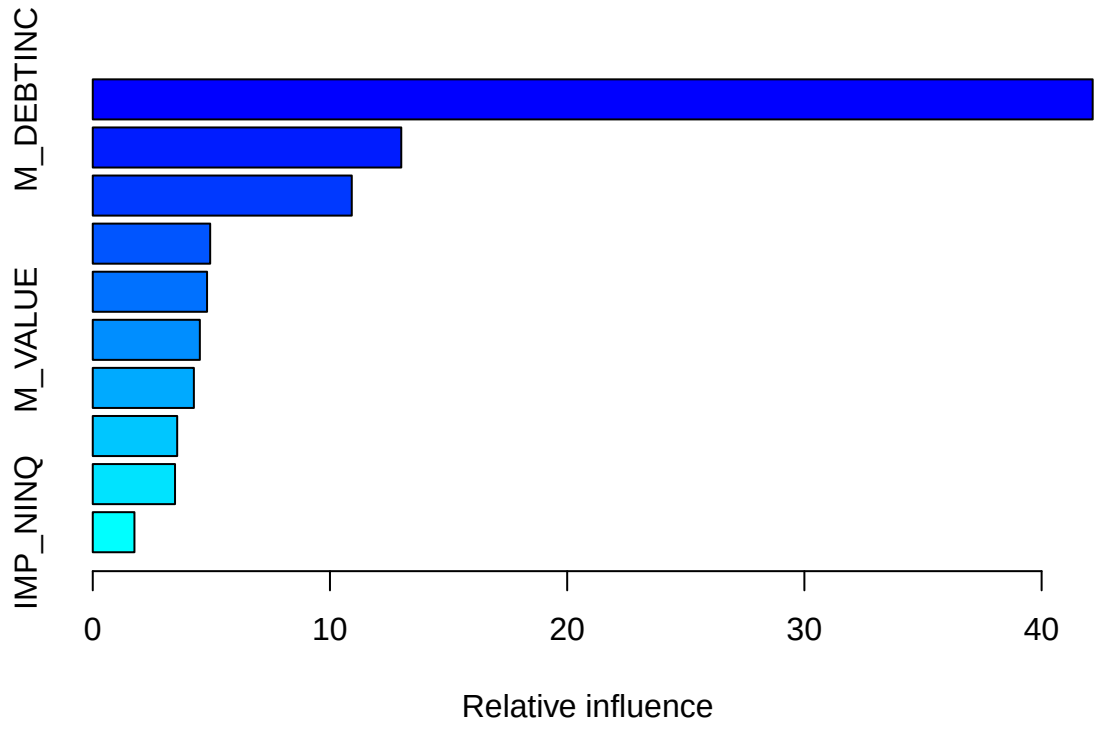
```
pr2= prediction(pr, data_test$TARGET_BAD_FLAG)
pr3= performance(pr2, "tpr", "fpr")
```

```
pr = predict(rf_model, data_test)
head(pr)
```

```
##         4         6         7        15        17        18
## 0.7966524 0.8313333 0.9241667 0.7746333 0.9212667 0.3462000
```

```
RMSEr = sqrt(mean((data_test$TARGET_BAD_FLAG - pr)^2))
```

```
gb_model = gbm(data=data_train, TARGET_BAD_FLAG ~., n.tree=500, distribution = "bernoulli")
summary.gbm(gb_model, cBars = 10)
```



GB

```
##                                  var     rel.inf
## M_DEBTINC                  M_DEBTINC 42.15276212
## IMP_DEBTINC              IMP_DEBTINC 13.00578764
## IMP_DELINQ                IMP_DELINQ 10.91954001
## IMP_CLAGE                  IMP_CLAGE  4.94970498
## IMP_VALUE                  IMP_VALUE  4.81763715
## M_VALUE                      M_VALUE  4.51409149
## IMP_DEROG                  IMP_DEROG  4.26284516
## LOAN                            LOAN  3.55916423
## IMP_CLNO                    IMP_CLNO  3.46797820
## IMP_NINQ                    IMP_NINQ  1.75727930
## IMP_YOJ                      IMP_YOJ  1.65837025
## IMP_MORTDUE              IMP_MORTDUE  1.64426153
## M_DEROG                      M_DEROG  1.27305534
## FLAG.Job.Sales        FLAG.Job.Sales  0.73536853
## M_CLNO                        M_CLNO  0.54901563
## M_YOJ                          M_YOJ  0.16877535
## FLAG.Job.Other        FLAG.Job.Other  0.13372343
## FLAG.Job.Office      FLAG.Job.Office  0.12346471
## FLAG.Job.Mgr            FLAG.Job.Mgr  0.06870764
## FLAG.Reason.DebtCon FLAG.Reason.DebtCon  0.06701971
## M_DELINQ                    M_DELINQ  0.06598496
## M_MORTDUE                  M_MORTDUE  0.05293782
## FLAG.Job.Self          FLAG.Job.Self  0.05252482
## M_CLAGE                      M_CLAGE  0.00000000
## M_NINQ                        M_NINQ  0.00000000
```

```
## FLAG.Job.ProfExe       FLAG.Job.ProfExe  0.00000000
## FLAG.Reason.HomeImp FLAG.Reason.HomeImp  0.00000000
```

```
pg = predict(gb_model, data_test)
```

```
## Using 500 trees...
```

```
head(pg)
```

```
## [1]  3.367087 -0.272228  5.113228  1.428145  6.221386  3.551554
```

```
RMSEg = sqrt(mean((data_test$TARGET_BAD_FLAG - pg)^2))
```

```
pg= predict(gb_model, data_test, type= "response")
```

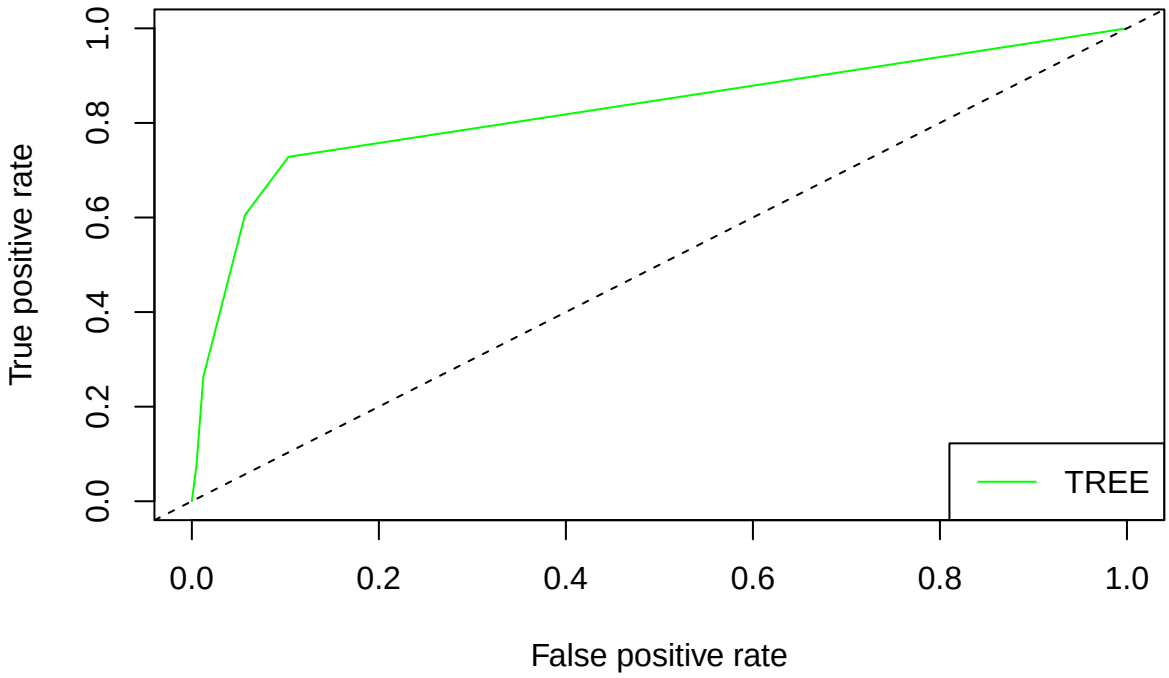
```
## Using 500 trees...
```

```
head(pg)
```

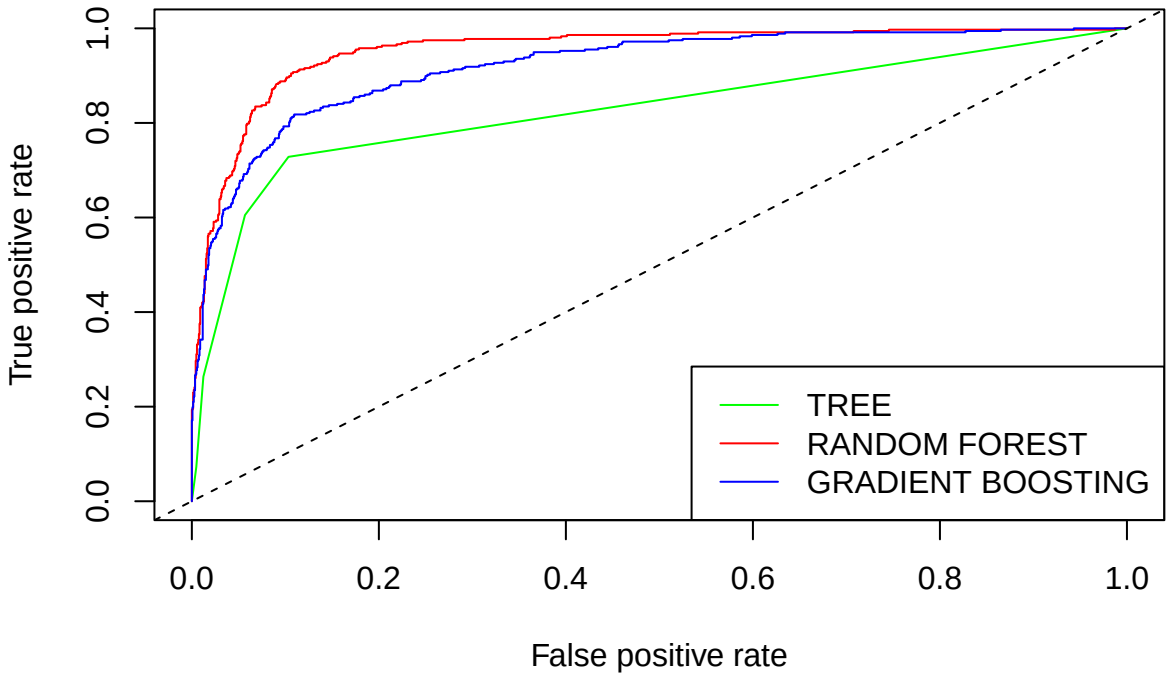
```
## [1] 0.9666599 0.4323602 0.9940194 0.8066121 0.9980174 0.9721196
```

```
pg2=prediction(pg, data_test$TARGET_BAD_FLAG)
pg3 = performance(pg2, "tpr", "fpr")
```

```
plot(pt3, col= "green")
abline(0,1,lty=2)
legend("bottomright", c("TREE"), col=c("green"), bty="y", lty=1)
```




```
plot(pt3, col="green")
plot(pr3, col="red", add=TRUE)
plot(pg3, col="blue", add=TRUE)
abline(0,1,lty=2)
legend("bottomright", c("TREE", "RANDOM FOREST", "GRADIENT BOOSTING"), col=c("green", "red", "blue"), b
```




```
aucT = performance(pt2, "auc")@y.values
aucR = performance(pr2, "auc")@y.values
aucG = performance(pg2, "auc")@y.values
```

```
print(paste("TREE AUC=",aucT))
```

```
## [1] "TREE AUC= 0.826618121581281"
```

```
print(paste("RF AUC",aucR))
```

```
## [1] "RF AUC 0.953436405362943"
```

```
print(paste("GB AUC",aucG))
```

```
## [1] "GB AUC 0.920521802150007"
```

```
aucT = performance(pt2, "auc")@y.values
aucR = performance(pr2, "auc")@y.values
aucG = performance(pg2, "auc")@y.values
```

```
print(paste("TREE AUC=",aucT))
```

```
## [1] "TREE AUC= 0.826618121581281"
```

```
print(paste("RF AUC",aucR))
```

```
## [1] "RF AUC 0.953436405362943"
```

```
print(paste("GB AUC",aucG))
```

```
## [1] "GB AUC 0.920521802150007"
```

The ROC curves for all trees are optimal According to the results of my code Random Forest is slightly better than gradient boosting

```
SEED=1
set.seed(SEED)
```

```
data_flag1= data
data_flag1$TARGET_BAD_FLAG= NULL
head(data_flag1)
```

LINEAR REGRESSION

```
##   TARGET_LOSS_AMT LOAN IMP_MORTDUE M_MORTDUE IMP_VALUE M_VALUE IMP_YOJ M_YOJ
## 1             641 1100       25860         0     39025       0    10.5     0
## 2            1109 1300       70053         0     68400       0     7.0     0
## 3             767 1500       13500         0     16700       0     4.0     0
## 4            1425 1500       65000         1     89000       1     7.0     1
## 5               0 1700       97800         0    112000       0     3.0     0
## 6             335 1700       30548         0     40320       0     9.0     0
##   IMP_DEROG M_DEROG IMP_DELINQ M_DELINQ 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
```

```
FLAG= sample(c(TRUE,FALSE), nrow(data_flag), replace=TRUE, prob=c(0.7,0.3))
data_train1= data_flag1[FLAG, ]
data_test1= data_flag1[! FLAG, ]
```

```
dim(data_flag1)
```

```
## [1] 5960   28
```

```
dim(data_train1)
```

```
## [1] 4142   28
```

```
dim(data_test1)
```

```
## [1] 1818   28
```

```
theUpper_LR1 = lm(TARGET_LOSS_AMT ~ ., data=data_train1)
theLower_LR1 = lm(TARGET_LOSS_AMT ~ 1, data=data_train1)
summary(theUpper_LR1)
```

```
##
## Call:
## lm(formula = TARGET_LOSS_AMT ~ ., data = data_train1)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -32389  -2537   -335   1536  58383
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -6.167e+03  7.462e+02  -8.264  < 2e-16 ***
## LOAN                 1.407e-01  8.541e-03  16.476  < 2e-16 ***
## IMP_MORTDUE         -7.051e-03  3.449e-03  -2.044 0.040992 *
## M_MORTDUE            9.756e+02  3.588e+02   2.719 0.006582 **
## IMP_VALUE            1.189e-02  2.632e-03   4.516 6.49e-06 ***
## M_VALUE              5.695e+03  6.379e+02   8.927  < 2e-16 ***
## IMP_YOJ             -4.460e+01  1.252e+01  -3.563 0.000371 ***
## M_YOJ               -9.450e+02  3.466e+02  -2.727 0.006423 **
## IMP_DEROG            9.082e+02  1.145e+02   7.929 2.82e-15 ***
## M_DEROG             -2.791e+03  4.350e+02  -6.416 1.56e-10 ***
## IMP_DELINQ           1.794e+03  8.430e+01  21.276  < 2e-16 ***
## M_DELINQ            -1.426e+03  5.550e+02  -2.570 0.010214 *
```

```
## IMP_CLAGE           -8.837e+00  1.102e+00  -8.018 1.39e-15 ***
## M_CLAGE              5.643e+02  8.113e+02   0.696 0.486760
## IMP_NINQ             2.198e+02  5.574e+01   3.944 8.15e-05 ***
## M_NINQ              -5.651e+01  5.264e+02  -0.107 0.914523
## IMP_CLNO             4.846e+01  1.005e+01   4.820 1.49e-06 ***
## M_CLNO               2.246e+03  1.041e+03   2.157 0.031026 *
## IMP_DEBTINC          1.104e+02  1.144e+01   9.649  < 2e-16 ***
## M_DEBTINC            6.305e+03  2.246e+02  28.069  < 2e-16 ***
## FLAG.Job.Mgr         6.379e+02  5.473e+02   1.166 0.243855
## FLAG.Job.Office      3.706e+02  5.403e+02   0.686 0.492845
## FLAG.Job.Other       9.849e+02  5.089e+02   1.935 0.053027 .
## FLAG.Job.ProfExe     6.239e+02  5.342e+02   1.168 0.242856
## FLAG.Job.Sales       3.390e+03  8.134e+02   4.168 3.14e-05 ***
## FLAG.Job.Self        2.594e+03  6.998e+02   3.707 0.000212 ***
## FLAG.Reason.DebtCon -5.226e+02  4.837e+02  -1.080 0.279986
## FLAG.Reason.HomeImp -1.028e+03  4.914e+02  -2.093 0.036435 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5531 on 4114 degrees of freedom
## Multiple R-squared:  0.4354, Adjusted R-squared:  0.4317
## F-statistic: 117.5 on 27 and 4114 DF,  p-value: < 2.2e-16
```

```
summary(theLower_LR1)
```

```
##
## Call:
## lm(formula = TARGET_LOSS_AMT ~ 1, data = data_train1)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -2726  -2726  -2726  -2726  76261
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)     2726        114   23.92   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7337 on 4141 degrees of freedom
```

```
lr_model1 = stepAIC(theUpper_LR1, direction = "backward", scope = list(lower=theLower_LR1, upper = theUp
```

```
## Start:  AIC=71420.12
## TARGET_LOSS_AMT ~ LOAN + IMP_MORTDUE + M_MORTDUE + IMP_VALUE +
##     M_VALUE + IMP_YOJ + M_YOJ + IMP_DEROG + M_DEROG + IMP_DELINQ +
##     M_DELINQ + IMP_CLAGE + M_CLAGE + IMP_NINQ + M_NINQ + IMP_CLNO +
##     M_CLNO + IMP_DEBTINC + M_DEBTINC + FLAG.Job.Mgr + FLAG.Job.Office +
##     FLAG.Job.Other + FLAG.Job.ProfExe + FLAG.Job.Sales + FLAG.Job.Self +
##     FLAG.Reason.DebtCon + FLAG.Reason.HomeImp
##
##                       Df  Sum of Sq        RSS   AIC
## - M_NINQ               1 3.5246e+05 1.2585e+11 71418
```

```
## - FLAG.Job.Office      1 1.4389e+07 1.2586e+11 71419
## - M_CLAGE              1 1.4798e+07 1.2586e+11 71419
## - FLAG.Reason.DebtCon  1 3.5712e+07 1.2588e+11 71419
## - FLAG.Job.Mgr         1 4.1558e+07 1.2589e+11 71419
## - FLAG.Job.ProfExe     1 4.1734e+07 1.2589e+11 71419
## <none>                              1.2585e+11 71420
## - FLAG.Job.Other       1 1.1457e+08 1.2596e+11 71422
## - IMP_MORTDUE          1 1.2783e+08 1.2597e+11 71422
## - FLAG.Reason.HomeImp  1 1.3397e+08 1.2598e+11 71423
## - M_CLNO               1 1.4239e+08 1.2599e+11 71423
## - M_DELINQ             1 2.0199e+08 1.2605e+11 71425
## - M_MORTDUE            1 2.2609e+08 1.2607e+11 71426
## - M_YOJ                1 2.2744e+08 1.2607e+11 71426
## - IMP_YOJ              1 3.8825e+08 1.2623e+11 71431
## - FLAG.Job.Self        1 4.2044e+08 1.2627e+11 71432
## - IMP_NINQ             1 4.7583e+08 1.2632e+11 71434
## - FLAG.Job.Sales       1 5.3135e+08 1.2638e+11 71436
## - IMP_VALUE            1 6.2376e+08 1.2647e+11 71439
## - IMP_CLNO             1 7.1072e+08 1.2656e+11 71441
## - M_DEROG              1 1.2591e+09 1.2710e+11 71459
## - IMP_DEROG            1 1.9232e+09 1.2777e+11 71481
## - IMP_CLAGE            1 1.9665e+09 1.2781e+11 71482
## - M_VALUE              1 2.4375e+09 1.2828e+11 71498
## - IMP_DEBTINC          1 2.8480e+09 1.2869e+11 71511
## - LOAN                 1 8.3042e+09 1.3415e+11 71683
## - IMP_DELINQ           1 1.3847e+10 1.3969e+11 71850
## - M_DEBTINC            1 2.4101e+10 1.4995e+11 72144
##
## Step:  AIC=71418.13
## TARGET_LOSS_AMT ~ LOAN + IMP_MORTDUE + M_MORTDUE + IMP_VALUE +
##     M_VALUE + IMP_YOJ + M_YOJ + IMP_DEROG + M_DEROG + IMP_DELINQ +
##     M_DELINQ + IMP_CLAGE + M_CLAGE + IMP_NINQ + IMP_CLNO + M_CLNO +
##     IMP_DEBTINC + M_DEBTINC + FLAG.Job.Mgr + FLAG.Job.Office +
##     FLAG.Job.Other + FLAG.Job.ProfExe + FLAG.Job.Sales + FLAG.Job.Self +
##     FLAG.Reason.DebtCon + FLAG.Reason.HomeImp
##
##                        Df  Sum of Sq        RSS   AIC
## - FLAG.Job.Office      1 1.4504e+07 1.2586e+11 71417
## - M_CLAGE              1 1.4938e+07 1.2586e+11 71417
## - FLAG.Reason.DebtCon  1 3.5706e+07 1.2588e+11 71417
## - FLAG.Job.Mgr         1 4.1764e+07 1.2589e+11 71418
## - FLAG.Job.ProfExe     1 4.1931e+07 1.2589e+11 71418
## <none>                              1.2585e+11 71418
## - FLAG.Job.Other       1 1.1491e+08 1.2596e+11 71420
## - IMP_MORTDUE          1 1.2780e+08 1.2597e+11 71420
## - FLAG.Reason.HomeImp  1 1.3398e+08 1.2598e+11 71421
## - M_CLNO               1 1.4626e+08 1.2599e+11 71421
## - M_MORTDUE            1 2.2584e+08 1.2607e+11 71424
## - M_YOJ                1 2.3039e+08 1.2608e+11 71424
## - M_DELINQ             1 2.6415e+08 1.2611e+11 71425
## - IMP_YOJ              1 3.8824e+08 1.2623e+11 71429
## - FLAG.Job.Self        1 4.2047e+08 1.2627e+11 71430
## - IMP_NINQ             1 4.7986e+08 1.2633e+11 71432
## - FLAG.Job.Sales       1 5.3338e+08 1.2638e+11 71434
```

```
## - IMP_VALUE            1 6.2345e+08 1.2647e+11 71437
## - IMP_CLNO             1 7.1050e+08 1.2656e+11 71439
## - M_DEROG              1 1.2679e+09 1.2711e+11 71458
## - IMP_DEROG            1 1.9257e+09 1.2777e+11 71479
## - IMP_CLAGE            1 1.9665e+09 1.2781e+11 71480
## - M_VALUE              1 2.4434e+09 1.2829e+11 71496
## - IMP_DEBTINC          1 2.8779e+09 1.2872e+11 71510
## - LOAN                 1 8.3252e+09 1.3417e+11 71681
## - IMP_DELINQ           1 1.3858e+10 1.3970e+11 71849
## - M_DEBTINC            1 2.4108e+10 1.4995e+11 72142
##
## Step:  AIC=71416.61
## TARGET_LOSS_AMT ~ LOAN + IMP_MORTDUE + M_MORTDUE + IMP_VALUE +
##     M_VALUE + IMP_YOJ + M_YOJ + IMP_DEROG + M_DEROG + IMP_DELINQ +
##     M_DELINQ + IMP_CLAGE + M_CLAGE + IMP_NINQ + IMP_CLNO + M_CLNO +
##     IMP_DEBTINC + M_DEBTINC + FLAG.Job.Mgr + FLAG.Job.Other +
##     FLAG.Job.ProfExe + FLAG.Job.Sales + FLAG.Job.Self + FLAG.Reason.DebtCon +
##     FLAG.Reason.HomeImp
##
##                        Df  Sum of Sq        RSS   AIC
## - M_CLAGE               1 1.1611e+07 1.2587e+11 71415
## - FLAG.Reason.DebtCon   1 2.8931e+07 1.2589e+11 71416
## - FLAG.Job.Mgr          1 3.3928e+07 1.2589e+11 71416
## - FLAG.Job.ProfExe      1 3.7375e+07 1.2590e+11 71416
## <none>                               1.2586e+11 71417
## - FLAG.Reason.HomeImp   1 1.2210e+08 1.2598e+11 71419
## - IMP_MORTDUE           1 1.3002e+08 1.2599e+11 71419
## - M_CLNO                1 1.3662e+08 1.2600e+11 71419
## - M_MORTDUE             1 2.1870e+08 1.2608e+11 71422
## - M_YOJ                 1 2.3425e+08 1.2609e+11 71422
## - FLAG.Job.Other        1 2.3818e+08 1.2610e+11 71422
## - M_DELINQ              1 2.5831e+08 1.2612e+11 71423
## - IMP_YOJ               1 3.9080e+08 1.2625e+11 71427
## - IMP_NINQ              1 4.7697e+08 1.2634e+11 71430
## - FLAG.Job.Self         1 5.6184e+08 1.2642e+11 71433
## - IMP_VALUE             1 6.2667e+08 1.2649e+11 71435
## - FLAG.Job.Sales        1 6.4442e+08 1.2650e+11 71436
## - IMP_CLNO              1 7.1373e+08 1.2657e+11 71438
## - M_DEROG               1 1.2731e+09 1.2713e+11 71456
## - IMP_DEROG             1 1.9173e+09 1.2778e+11 71477
## - IMP_CLAGE             1 1.9533e+09 1.2781e+11 71478
## - M_VALUE               1 2.4410e+09 1.2830e+11 71494
## - IMP_DEBTINC           1 2.9273e+09 1.2879e+11 71510
## - LOAN                  1 8.3163e+09 1.3418e+11 71680
## - IMP_DELINQ            1 1.3900e+10 1.3976e+11 71849
## - M_DEBTINC             1 2.4143e+10 1.5000e+11 72141
##
## Step:  AIC=71414.99
## TARGET_LOSS_AMT ~ LOAN + IMP_MORTDUE + M_MORTDUE + IMP_VALUE +
##     M_VALUE + IMP_YOJ + M_YOJ + IMP_DEROG + M_DEROG + IMP_DELINQ +
##     M_DELINQ + IMP_CLAGE + IMP_NINQ + IMP_CLNO + M_CLNO + IMP_DEBTINC +
##     M_DEBTINC + FLAG.Job.Mgr + FLAG.Job.Other + FLAG.Job.ProfExe +
##     FLAG.Job.Sales + FLAG.Job.Self + FLAG.Reason.DebtCon + FLAG.Reason.HomeImp
##
```

```
##                         Df  Sum of Sq        RSS   AIC
## - FLAG.Reason.DebtCon    1 3.2415e+07 1.2590e+11 71414
## - FLAG.Job.Mgr           1 3.3084e+07 1.2591e+11 71414
## - FLAG.Job.ProfExe       1 3.6162e+07 1.2591e+11 71414
## <none>                                1.2587e+11 71415
## - IMP_MORTDUE            1 1.2503e+08 1.2600e+11 71417
## - FLAG.Reason.HomeImp    1 1.3075e+08 1.2600e+11 71417
## - M_MORTDUE              1 2.2243e+08 1.2609e+11 71420
## - M_YOJ                  1 2.3442e+08 1.2611e+11 71421
## - FLAG.Job.Other         1 2.3678e+08 1.2611e+11 71421
## - M_DELINQ               1 2.6349e+08 1.2614e+11 71422
## - IMP_YOJ                1 3.9885e+08 1.2627e+11 71426
## - IMP_NINQ               1 4.8244e+08 1.2635e+11 71429
## - M_CLNO                 1 5.2769e+08 1.2640e+11 71430
## - FLAG.Job.Self          1 5.5803e+08 1.2643e+11 71431
## - IMP_VALUE              1 6.2261e+08 1.2649e+11 71433
## - FLAG.Job.Sales         1 6.4188e+08 1.2651e+11 71434
## - IMP_CLNO               1 7.0567e+08 1.2658e+11 71436
## - M_DEROG                1 1.2720e+09 1.2714e+11 71455
## - IMP_DEROG              1 1.9307e+09 1.2780e+11 71476
## - IMP_CLAGE              1 1.9435e+09 1.2782e+11 71476
## - M_VALUE                1 2.4363e+09 1.2831e+11 71492
## - IMP_DEBTINC            1 2.9166e+09 1.2879e+11 71508
## - LOAN                   1 8.3244e+09 1.3420e+11 71678
## - IMP_DELINQ             1 1.3889e+10 1.3976e+11 71847
## - M_DEBTINC              1 2.4178e+10 1.5005e+11 72141
##
## Step:  AIC=71414.06
## TARGET_LOSS_AMT ~ LOAN + IMP_MORTDUE + M_MORTDUE + IMP_VALUE +
##     M_VALUE + IMP_YOJ + M_YOJ + IMP_DEROG + M_DEROG + IMP_DELINQ +
##     M_DELINQ + IMP_CLAGE + IMP_NINQ + IMP_CLNO + M_CLNO + IMP_DEBTINC +
##     M_DEBTINC + FLAG.Job.Mgr + FLAG.Job.Other + FLAG.Job.ProfExe +
##     FLAG.Job.Sales + FLAG.Job.Self + FLAG.Reason.HomeImp
##
##                         Df  Sum of Sq        RSS   AIC
## - FLAG.Job.Mgr           1 2.8951e+07 1.2593e+11 71413
## - FLAG.Job.ProfExe       1 3.0177e+07 1.2593e+11 71413
## <none>                                1.2590e+11 71414
## - IMP_MORTDUE            1 1.2798e+08 1.2603e+11 71416
## - M_YOJ                  1 2.1556e+08 1.2612e+11 71419
## - FLAG.Job.Other         1 2.2152e+08 1.2613e+11 71419
## - FLAG.Reason.HomeImp    1 2.2951e+08 1.2613e+11 71420
## - M_MORTDUE              1 2.4234e+08 1.2615e+11 71420
## - M_DELINQ               1 2.6331e+08 1.2617e+11 71421
## - IMP_YOJ                1 3.9098e+08 1.2630e+11 71425
## - IMP_NINQ               1 4.8232e+08 1.2639e+11 71428
## - FLAG.Job.Self          1 5.5359e+08 1.2646e+11 71430
## - M_CLNO                 1 6.1654e+08 1.2652e+11 71432
## - FLAG.Job.Sales         1 6.2636e+08 1.2653e+11 71433
## - IMP_VALUE              1 6.4069e+08 1.2655e+11 71433
## - IMP_CLNO               1 7.0088e+08 1.2661e+11 71435
## - M_DEROG                1 1.2917e+09 1.2720e+11 71454
## - IMP_CLAGE              1 1.9245e+09 1.2783e+11 71475
## - IMP_DEROG              1 1.9314e+09 1.2784e+11 71475
```

```
## - M_VALUE              1 2.4817e+09 1.2839e+11 71493
## - IMP_DEBTINC          1 2.8949e+09 1.2880e+11 71506
## - LOAN                 1 8.2938e+09 1.3420e+11 71676
## - IMP_DELINQ           1 1.4036e+10 1.3994e+11 71850
## - M_DEBTINC            1 2.4162e+10 1.5007e+11 72139
##
## Step:  AIC=71413.01
## TARGET_LOSS_AMT ~ LOAN + IMP_MORTDUE + M_MORTDUE + IMP_VALUE +
##     M_VALUE + IMP_YOJ + M_YOJ + IMP_DEROG + M_DEROG + IMP_DELINQ +
##     M_DELINQ + IMP_CLAGE + IMP_NINQ + IMP_CLNO + M_CLNO + IMP_DEBTINC +
##     M_DEBTINC + FLAG.Job.Other + FLAG.Job.ProfExe + FLAG.Job.Sales +
##     FLAG.Job.Self + FLAG.Reason.HomeImp
##
##                       Df  Sum of Sq        RSS   AIC
## - FLAG.Job.ProfExe     1 1.1519e+07 1.2594e+11 71411
## <none>                              1.2593e+11 71413
## - IMP_MORTDUE          1 1.1924e+08 1.2605e+11 71415
## - FLAG.Job.Other       1 1.9904e+08 1.2613e+11 71418
## - M_YOJ                1 2.1459e+08 1.2615e+11 71418
## - M_MORTDUE            1 2.3395e+08 1.2617e+11 71419
## - FLAG.Reason.HomeImp  1 2.3452e+08 1.2617e+11 71419
## - M_DELINQ             1 2.6039e+08 1.2619e+11 71420
## - IMP_YOJ              1 3.7964e+08 1.2631e+11 71423
## - IMP_NINQ             1 4.9548e+08 1.2643e+11 71427
## - FLAG.Job.Self        1 5.2479e+08 1.2646e+11 71428
## - M_CLNO               1 5.9562e+08 1.2653e+11 71431
## - FLAG.Job.Sales       1 5.9773e+08 1.2653e+11 71431
## - IMP_VALUE            1 6.3879e+08 1.2657e+11 71432
## - IMP_CLNO             1 7.0308e+08 1.2664e+11 71434
## - M_DEROG              1 1.2953e+09 1.2723e+11 71453
## - IMP_CLAGE            1 1.9280e+09 1.2786e+11 71474
## - IMP_DEROG            1 1.9464e+09 1.2788e+11 71475
## - M_VALUE              1 2.4793e+09 1.2841e+11 71492
## - IMP_DEBTINC          1 2.9129e+09 1.2885e+11 71506
## - LOAN                 1 8.2784e+09 1.3421e+11 71675
## - IMP_DELINQ           1 1.4095e+10 1.4003e+11 71850
## - M_DEBTINC            1 2.4218e+10 1.5015e+11 72140
##
## Step:  AIC=71411.39
## TARGET_LOSS_AMT ~ LOAN + IMP_MORTDUE + M_MORTDUE + IMP_VALUE +
##     M_VALUE + IMP_YOJ + M_YOJ + IMP_DEROG + M_DEROG + IMP_DELINQ +
##     M_DELINQ + IMP_CLAGE + IMP_NINQ + IMP_CLNO + M_CLNO + IMP_DEBTINC +
##     M_DEBTINC + FLAG.Job.Other + FLAG.Job.Sales + FLAG.Job.Self +
##     FLAG.Reason.HomeImp
##
##                       Df  Sum of Sq        RSS   AIC
## <none>                              1.2594e+11 71411
## - IMP_MORTDUE          1 1.1543e+08 1.2606e+11 71413
## - FLAG.Job.Other       1 1.9410e+08 1.2614e+11 71416
## - M_YOJ                1 2.1752e+08 1.2616e+11 71417
## - M_MORTDUE            1 2.2920e+08 1.2617e+11 71417
## - FLAG.Reason.HomeImp  1 2.3046e+08 1.2618e+11 71417
## - M_DELINQ             1 2.5770e+08 1.2620e+11 71418
## - IMP_YOJ              1 3.8259e+08 1.2633e+11 71422
```

```
## - IMP_NINQ             1 4.9230e+08 1.2644e+11 71426
## - FLAG.Job.Self        1 5.1615e+08 1.2646e+11 71426
## - FLAG.Job.Sales       1 5.8630e+08 1.2653e+11 71429
## - M_CLNO               1 5.9038e+08 1.2654e+11 71429
## - IMP_VALUE            1 6.5743e+08 1.2660e+11 71431
## - IMP_CLNO             1 7.2048e+08 1.2667e+11 71433
## - M_DEROG              1 1.3193e+09 1.2726e+11 71453
## - IMP_CLAGE            1 1.9181e+09 1.2786e+11 71472
## - IMP_DEROG            1 1.9389e+09 1.2788e+11 71473
## - M_VALUE              1 2.4799e+09 1.2842e+11 71490
## - IMP_DEBTINC          1 2.9025e+09 1.2885e+11 71504
## - LOAN                 1 8.2679e+09 1.3421e+11 71673
## - IMP_DELINQ           1 1.4083e+10 1.4003e+11 71848
## - M_DEBTINC            1 2.4293e+10 1.5024e+11 72140
```

```
summary(lr_model1)
```

```
##
## Call:
## lm(formula = TARGET_LOSS_AMT ~ LOAN + IMP_MORTDUE + M_MORTDUE +
##     IMP_VALUE + M_VALUE + IMP_YOJ + M_YOJ + IMP_DEROG + M_DEROG +
##     IMP_DELINQ + M_DELINQ + IMP_CLAGE + IMP_NINQ + IMP_CLNO +
##     M_CLNO + IMP_DEBTINC + M_DEBTINC + FLAG.Job.Other + FLAG.Job.Sales +
##     FLAG.Job.Self + FLAG.Reason.HomeImp, data = data_train1)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -32477  -2520   -340   1525  58418
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -6.183e+03  4.874e+02 -12.686  < 2e-16 ***
## LOAN                 1.398e-01  8.498e-03  16.446  < 2e-16 ***
## IMP_MORTDUE         -6.647e-03  3.420e-03  -1.943 0.052055 .
## M_MORTDUE            9.695e+02  3.541e+02   2.738 0.006204 **
## IMP_VALUE            1.214e-02  2.617e-03   4.637 3.64e-06 ***
## M_VALUE              5.727e+03  6.359e+02   9.007  < 2e-16 ***
## IMP_YOJ             -4.409e+01  1.246e+01  -3.538 0.000408 ***
## M_YOJ               -9.121e+02  3.419e+02  -2.668 0.007671 **
## IMP_DEROG            9.088e+02  1.141e+02   7.964 2.14e-15 ***
## M_DEROG             -2.840e+03  4.323e+02  -6.570 5.67e-11 ***
## IMP_DELINQ           1.801e+03  8.391e+01  21.464  < 2e-16 ***
## M_DELINQ            -1.431e+03  4.929e+02  -2.903 0.003710 **
## IMP_CLAGE           -8.680e+00  1.096e+00  -7.921 3.00e-15 ***
## IMP_NINQ             2.226e+02  5.546e+01   4.013 6.10e-05 ***
## IMP_CLNO             4.769e+01  9.823e+00   4.855 1.25e-06 ***
## M_CLNO               2.680e+03  6.097e+02   4.395 1.14e-05 ***
## IMP_DEBTINC          1.098e+02  1.127e+01   9.744  < 2e-16 ***
## M_DEBTINC            6.320e+03  2.242e+02  28.190  < 2e-16 ***
## FLAG.Job.Other       4.791e+02  1.901e+02   2.520 0.011779 *
## FLAG.Job.Sales       2.842e+03  6.489e+02   4.379 1.22e-05 ***
## FLAG.Job.Self        2.064e+03  5.022e+02   4.109 4.05e-05 ***
## FLAG.Reason.HomeImp -5.427e+02  1.976e+02  -2.746 0.006064 **
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5529 on 4120 degrees of freedom
## Multiple R-squared:  0.435,  Adjusted R-squared:  0.4321
## F-statistic:   151 on 21 and 4120 DF,  p-value: < 2.2e-16
```

```
plr = predict(lr_model1, data_test1)
head(plr)
```

```
##         4         6         7        15        17        18
## 11097.448 -2321.894 11008.296  5898.499 16905.210  1133.279
```

```
RMSElr = sqrt(mean((data_test1$TARGET_LOSS_AMT - plr)^2))
```

```
treeVars1 = t1E$variable.importance
treeVars1 = names(treeVars1)
treeVarsPlus1 = paste(treeVars1, collapse ="+")
F1 = as.formula(paste("TARGET_LOSS_AMT ~", treeVarsPlus1))
```

```
tree_LR1 = glm(F1, data=data_train1)
theLower_LR1 = lm(TARGET_LOSS_AMT ~ 1, data=data_train1)
```

```
summary(tree_LR1)
```

```
##
## Call:
## glm(formula = F1, data = data_train1)
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.873e+03  4.696e+02 -12.504  < 2e-16 ***
## M_DEBTINC    6.964e+03  2.248e+02  30.977  < 2e-16 ***
## IMP_DEBTINC  1.102e+02  1.130e+01   9.748  < 2e-16 ***
## IMP_DELINQ   1.723e+03  8.277e+01  20.813  < 2e-16 ***
## IMP_CLAGE   -1.043e+01  1.100e+00  -9.489  < 2e-16 ***
## LOAN         1.506e-01  8.321e-03  18.093  < 2e-16 ***
## M_VALUE      5.887e+03  6.472e+02   9.097  < 2e-16 ***
## IMP_VALUE    1.231e-02  2.590e-03   4.753 2.07e-06 ***
## IMP_MORTDUE -7.699e-03  3.418e-03  -2.252 0.024346 *
## IMP_CLNO     5.404e+01  9.685e+00   5.580 2.56e-08 ***
## IMP_YOJ     -4.872e+01  1.259e+01  -3.870 0.000111 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 32202311)
##
##     Null deviance: 2.2290e+11  on 4141  degrees of freedom
## Residual deviance: 1.3303e+11  on 4131  degrees of freedom
## AIC: 83372
##
## Number of Fisher Scoring iterations: 2
```

```r
summary(theLower_LR1)
```

```
##
## Call:
## lm(formula = TARGET_LOSS_AMT ~ 1, data = data_train1)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -2726  -2726  -2726  -2726  76261
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)     2726        114   23.92   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7337 on 4141 degrees of freedom
```

```r
lrt_model1 = stepAIC(theLower_LR1, direction = "both", scope = list(lower=theLower_LR1, upper=tree_LR1))
```

```
## Start:  AIC=73734.01
## TARGET_LOSS_AMT ~ 1
##
##               Df  Sum of Sq        RSS   AIC
## + M_DEBTINC    1 4.7153e+10 1.7575e+11 72752
## + IMP_DELINQ   1 2.9356e+10 1.9355e+11 73151
## + M_VALUE      1 1.0850e+10 2.1205e+11 73529
## + LOAN         1 9.4873e+09 2.1341e+11 73556
## + IMP_DEBTINC  1 7.7321e+09 2.1517e+11 73590
## + IMP_CLNO     1 3.7741e+09 2.1913e+11 73665
## + IMP_CLAGE    1 3.3814e+09 2.1952e+11 73673
## + IMP_VALUE    1 3.0724e+09 2.1983e+11 73679
## + IMP_MORTDUE  1 1.5436e+09 2.2136e+11 73707
## + IMP_YOJ      1 2.7913e+08 2.2262e+11 73731
## <none>                      2.2290e+11 73734
##
## Step:  AIC=72751.57
## TARGET_LOSS_AMT ~ M_DEBTINC
##
##               Df  Sum of Sq        RSS   AIC
## + IMP_DELINQ   1 1.6625e+10 1.5912e+11 72342
## + LOAN         1 1.4130e+10 1.6162e+11 72406
## + IMP_DEBTINC  1 5.6846e+09 1.7006e+11 72617
## + M_VALUE      1 4.4592e+09 1.7129e+11 72647
## + IMP_VALUE    1 4.4448e+09 1.7130e+11 72647
## + IMP_CLNO     1 3.5915e+09 1.7216e+11 72668
## + IMP_MORTDUE  1 2.7257e+09 1.7302e+11 72689
## + IMP_CLAGE    1 1.2992e+09 1.7445e+11 72723
## + IMP_YOJ      1 2.2179e+08 1.7553e+11 72748
## <none>                      1.7575e+11 72752
## - M_DEBTINC    1 4.7153e+10 2.2290e+11 73734
##
```

```
## Step:  AIC=72341.97
## TARGET_LOSS_AMT ~ M_DEBTINC + IMP_DELINQ
##
##               Df  Sum of Sq        RSS   AIC
## + LOAN         1 1.4375e+10 1.4475e+11 71952
## + IMP_DEBTINC  1 5.2306e+09 1.5389e+11 72206
## + IMP_VALUE    1 4.7653e+09 1.5436e+11 72218
## + IMP_MORTDUE  1 2.6109e+09 1.5651e+11 72275
## + M_VALUE      1 2.5070e+09 1.5662e+11 72278
## + IMP_CLAGE    1 1.7654e+09 1.5736e+11 72298
## + IMP_CLNO     1 1.6652e+09 1.5746e+11 72300
## + IMP_YOJ      1 4.7816e+08 1.5865e+11 72332
## <none>                      1.5912e+11 72342
## - IMP_DELINQ   1 1.6625e+10 1.7575e+11 72752
## - M_DEBTINC    1 3.4422e+10 1.9355e+11 73151
##
## Step:  AIC=71951.81
## TARGET_LOSS_AMT ~ M_DEBTINC + IMP_DELINQ + LOAN
##
##               Df  Sum of Sq        RSS   AIC
## + IMP_DEBTINC  1 4.1659e+09 1.4058e+11 71833
## + IMP_CLAGE    1 2.7166e+09 1.4203e+11 71875
## + M_VALUE      1 2.3297e+09 1.4242e+11 71887
## + IMP_YOJ      1 1.1225e+09 1.4363e+11 71922
## + IMP_CLNO     1 1.0000e+09 1.4375e+11 71925
## + IMP_VALUE    1 9.9762e+08 1.4375e+11 71925
## + IMP_MORTDUE  1 6.9311e+08 1.4406e+11 71934
## <none>                      1.4475e+11 71952
## - LOAN         1 1.4375e+10 1.5912e+11 72342
## - IMP_DELINQ   1 1.6870e+10 1.6162e+11 72406
## - M_DEBTINC    1 3.8351e+10 1.8310e+11 72923
##
## Step:  AIC=71832.86
## TARGET_LOSS_AMT ~ M_DEBTINC + IMP_DELINQ + LOAN + IMP_DEBTINC
##
##               Df  Sum of Sq        RSS   AIC
## + M_VALUE      1 2.5528e+09 1.3803e+11 71759
## + IMP_CLAGE    1 2.4359e+09 1.3815e+11 71762
## + IMP_YOJ      1 8.9256e+08 1.3969e+11 71808
## + IMP_VALUE    1 6.7675e+08 1.3991e+11 71815
## + IMP_CLNO     1 5.1313e+08 1.4007e+11 71820
## + IMP_MORTDUE  1 3.6240e+08 1.4022e+11 71824
## <none>                      1.4058e+11 71833
## - IMP_DEBTINC  1 4.1659e+09 1.4475e+11 71952
## - LOAN         1 1.3310e+10 1.5389e+11 72206
## - IMP_DELINQ   1 1.6450e+10 1.5703e+11 72289
## - M_DEBTINC    1 3.6770e+10 1.7735e+11 72793
##
## Step:  AIC=71758.95
## TARGET_LOSS_AMT ~ M_DEBTINC + IMP_DELINQ + LOAN + IMP_DEBTINC +
##     M_VALUE
##
##               Df  Sum of Sq        RSS   AIC
## + IMP_CLAGE    1 2.3352e+09 1.3570e+11 71690
```

```
## + IMP_YOJ      1 8.8687e+08 1.3714e+11 71734
## + IMP_VALUE    1 7.4466e+08 1.3729e+11 71739
## + IMP_CLNO     1 5.8622e+08 1.3745e+11 71743
## + IMP_MORTDUE  1 3.3016e+08 1.3770e+11 71751
## <none>                      1.3803e+11 71759
## - M_VALUE      1 2.5528e+09 1.4058e+11 71833
## - IMP_DEBTINC  1 4.3890e+09 1.4242e+11 71887
## - LOAN         1 1.3105e+10 1.5114e+11 72133
## - IMP_DELINQ   1 1.4479e+10 1.5251e+11 72170
## - M_DEBTINC    1 3.3115e+10 1.7115e+11 72648
##
## Step:  AIC=71690.28
## TARGET_LOSS_AMT ~ M_DEBTINC + IMP_DELINQ + LOAN + IMP_DEBTINC +
##     M_VALUE + IMP_CLAGE
##
##               Df  Sum of Sq        RSS   AIC
## + IMP_CLNO     1 1.3616e+09 1.3433e+11 71651
## + IMP_VALUE    1 1.2123e+09 1.3448e+11 71655
## + IMP_MORTDUE  1 5.5962e+08 1.3514e+11 71675
## + IMP_YOJ      1 4.7246e+08 1.3522e+11 71678
## <none>                      1.3570e+11 71690
## - IMP_CLAGE    1 2.3352e+09 1.3803e+11 71759
## - M_VALUE      1 2.4521e+09 1.3815e+11 71762
## - IMP_DEBTINC  1 4.1018e+09 1.3980e+11 71812
## - LOAN         1 1.3972e+10 1.4967e+11 72094
## - IMP_DELINQ   1 1.5032e+10 1.5073e+11 72123
## - M_DEBTINC    1 3.1160e+10 1.6686e+11 72544
##
## Step:  AIC=71650.51
## TARGET_LOSS_AMT ~ M_DEBTINC + IMP_DELINQ + LOAN + IMP_DEBTINC +
##     M_VALUE + IMP_CLAGE + IMP_CLNO
##
##               Df  Sum of Sq        RSS   AIC
## + IMP_VALUE    1 7.2837e+08 1.3361e+11 71630
## + IMP_YOJ      1 4.6595e+08 1.3387e+11 71638
## + IMP_MORTDUE  1 1.7734e+08 1.3416e+11 71647
## <none>                      1.3433e+11 71651
## - IMP_CLNO     1 1.3616e+09 1.3570e+11 71690
## - M_VALUE      1 2.5480e+09 1.3688e+11 71726
## - IMP_CLAGE    1 3.1106e+09 1.3745e+11 71743
## - IMP_DEBTINC  1 3.3059e+09 1.3764e+11 71749
## - IMP_DELINQ   1 1.3432e+10 1.4777e+11 72043
## - LOAN         1 1.3507e+10 1.4784e+11 72045
## - M_DEBTINC    1 3.1140e+10 1.6547e+11 72512
##
## Step:  AIC=71629.99
## TARGET_LOSS_AMT ~ M_DEBTINC + IMP_DELINQ + LOAN + IMP_DEBTINC +
##     M_VALUE + IMP_CLAGE + IMP_CLNO + IMP_VALUE
##
##               Df  Sum of Sq        RSS   AIC
## + IMP_YOJ      1 4.1518e+08 1.3319e+11 71619
## + IMP_MORTDUE  1 9.6294e+07 1.3351e+11 71629
## <none>                      1.3361e+11 71630
## - IMP_VALUE    1 7.2837e+08 1.3433e+11 71651
```

```
## - IMP_CLNO     1 8.7768e+08 1.3448e+11 71655
## - M_VALUE      1 2.5925e+09 1.3620e+11 71708
## - IMP_DEBTINC  1 3.1145e+09 1.3672e+11 71723
## - IMP_CLAGE    1 3.3650e+09 1.3697e+11 71731
## - LOAN         1 1.0411e+10 1.4402e+11 71939
## - IMP_DELINQ   1 1.3752e+10 1.4736e+11 72034
## - M_DEBTINC    1 3.1181e+10 1.6479e+11 72497
##
## Step:  AIC=71619.1
## TARGET_LOSS_AMT ~ M_DEBTINC + IMP_DELINQ + LOAN + IMP_DEBTINC +
##     M_VALUE + IMP_CLAGE + IMP_CLNO + IMP_VALUE + IMP_YOJ
##
##               Df  Sum of Sq        RSS   AIC
## + IMP_MORTDUE  1 1.6338e+08 1.3303e+11 71616
## <none>                      1.3319e+11 71619
## - IMP_YOJ      1 4.1518e+08 1.3361e+11 71630
## - IMP_VALUE    1 6.7759e+08 1.3387e+11 71638
## - IMP_CLNO     1 8.8567e+08 1.3408e+11 71645
## - M_VALUE      1 2.5943e+09 1.3579e+11 71697
## - IMP_CLAGE    1 2.8747e+09 1.3607e+11 71706
## - IMP_DEBTINC  1 3.0040e+09 1.3620e+11 71709
## - LOAN         1 1.0716e+10 1.4391e+11 71938
## - IMP_DELINQ   1 1.3941e+10 1.4713e+11 72029
## - M_DEBTINC    1 3.1248e+10 1.6444e+11 72490
##
## Step:  AIC=71616.01
## TARGET_LOSS_AMT ~ M_DEBTINC + IMP_DELINQ + LOAN + IMP_DEBTINC +
##     M_VALUE + IMP_CLAGE + IMP_CLNO + IMP_VALUE + IMP_YOJ + IMP_MORTDUE
##
##               Df  Sum of Sq        RSS   AIC
## <none>                      1.3303e+11 71616
## - IMP_MORTDUE  1 1.6338e+08 1.3319e+11 71619
## - IMP_YOJ      1 4.8226e+08 1.3351e+11 71629
## - IMP_VALUE    1 7.2760e+08 1.3376e+11 71637
## - IMP_CLNO     1 1.0027e+09 1.3403e+11 71645
## - M_VALUE      1 2.6649e+09 1.3569e+11 71696
## - IMP_CLAGE    1 2.8996e+09 1.3593e+11 71703
## - IMP_DEBTINC  1 3.0601e+09 1.3609e+11 71708
## - LOAN         1 1.0542e+10 1.4357e+11 71930
## - IMP_DELINQ   1 1.3949e+10 1.4698e+11 72027
## - M_DEBTINC    1 3.0900e+10 1.6393e+11 72479
```

```
summary(lrt_model1)
```

```
##
## Call:
## lm(formula = TARGET_LOSS_AMT ~ M_DEBTINC + IMP_DELINQ + LOAN +
##     IMP_DEBTINC + M_VALUE + IMP_CLAGE + IMP_CLNO + IMP_VALUE +
##     IMP_YOJ + IMP_MORTDUE, data = data_train1)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -28161  -2513   -439   1586  57349
##
```

```
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.873e+03  4.696e+02 -12.504  < 2e-16 ***
## M_DEBTINC    6.964e+03  2.248e+02  30.977  < 2e-16 ***
## IMP_DELINQ   1.723e+03  8.277e+01  20.813  < 2e-16 ***
## LOAN         1.506e-01  8.321e-03  18.093  < 2e-16 ***
## IMP_DEBTINC  1.102e+02  1.130e+01   9.748  < 2e-16 ***
## M_VALUE      5.887e+03  6.472e+02   9.097  < 2e-16 ***
## IMP_CLAGE   -1.043e+01  1.100e+00  -9.489  < 2e-16 ***
## IMP_CLNO     5.404e+01  9.685e+00   5.580 2.56e-08 ***
## IMP_VALUE    1.231e-02  2.590e-03   4.753 2.07e-06 ***
## IMP_YOJ     -4.872e+01  1.259e+01  -3.870 0.000111 ***
## IMP_MORTDUE -7.699e-03  3.418e-03  -2.252 0.024346 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5675 on 4131 degrees of freedom
## Multiple R-squared:  0.4032, Adjusted R-squared:  0.4018
## F-statistic: 279.1 on 10 and 4131 DF,  p-value: < 2.2e-16
```

```
plr_tree1 = predict(tree_LR1, data_test1)
head(plr_tree1)
```

```
##         4         6         7        15        17        18
## 12303.995 -2330.417  8863.354  6664.713 15744.060  1735.312
```

```
RMSElr_tree1 = sqrt(mean(data_test1$TARGET_LOSS_AMT - plr_tree1)^2)
```

```
plr_tree1_step = predict(lrt_model1, data_test1)
head(plr_tree1_step)
```

```
##         4         6         7        15        17        18
## 12303.995 -2330.417  8863.354  6664.713 15744.060  1735.312
```

```
RMSElr_tree1_step = sqrt(mean(data_test1$TARGET_LOSS_AMT - plr_tree1_step)^2)
```

```
print(paste("TREE RMSE=", RMSEt))
```

```
## [1] "TREE RMSE= 0.638005314291569"
```

```
print(paste("RF RMSE=", RMSEr))
```

```
## [1] "RF RMSE= 0.254043534143728"
```

```
print(paste("GB RMSE=", RMSEg))
```

```
## [1] "GB RMSE= 3.12180712334603"
```

```
print(paste("LR BACK RMSE=", RMSElr))
```

```
## [1] "LR BACK RMSE= 5340.94504235833"
```

```
print(paste("LR TREE RMSE=", RMSElr_tree1))
```

```
## [1] "LR TREE RMSE= 4.85688854590777"
```

```
print(paste("LR TREE STEP RMSE=", RMSElr_tree1_step))
```

```
## [1] "LR TREE STEP RMSE= 4.85688854590556"
```

The trees all are optimal According to the results of my code, random forest trees are better than remaining models, and random forest has the lowest value also and I recommend random forest.

Week-8

```
print(head(data))
```

PCA Analysis

```
##   TARGET_BAD_FLAG TARGET_LOSS_AMT LOAN IMP_MORTDUE M_MORTDUE IMP_VALUE M_VALUE
## 1               1             641 1100       25860         0     39025       0
## 2               1            1109 1300       70053         0     68400       0
## 3               1             767 1500       13500         0     16700       0
## 4               1            1425 1500       65000         1     89000       1
## 5               0               0 1700       97800         0    112000       0
## 6               1             335 1700       30548         0     40320       0
##   IMP_YOJ M_YOJ IMP_DEROG M_DEROG IMP_DELINQ M_DELINQ IMP_CLAGE M_CLAGE
## 1    10.5     0         0       0          0        0  94.36667       0
## 2     7.0     0         0       0          2        0 121.83333       0
## 3     4.0     0         0       0          0        0 149.46667       0
## 4     7.0     1         1       1          1        1 174.00000       1
## 5     3.0     0         0       0          0        0  93.33333       0
## 6     9.0     0         0       0          0        0 101.46600       0
##   IMP_NINQ M_NINQ IMP_CLNO M_CLNO IMP_DEBTINC M_DEBTINC FLAG.Job.Mgr
## 1        1      0        9      0    35.00000         1            0
## 2        0      0       14      0    35.00000         1            0
## 3        1      0       10      0    35.00000         1            0
## 4        1      1       20      1    35.00000         1            0
## 5        0      0       14      0    35.00000         1            0
## 6        1      0        8      0    37.11361         0            0
##   FLAG.Job.Office FLAG.Job.Other FLAG.Job.ProfExe FLAG.Job.Sales FLAG.Job.Self
## 1               0              1                0              0             0
## 2               0              1                0              0             0
## 3               0              1                0              0             0
## 4               0              0                0              0             0
## 5               1              0                0              0             0
```

```
## 6               0             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
```

```
df_pca = data
df_pca$TARGET_BAD_FLAG = NULL
df_pca$TARGET_LOSS_AMT = NULL
head(df_pca)
```

```
##   LOAN IMP_MORTDUE M_MORTDUE IMP_VALUE M_VALUE IMP_YOJ M_YOJ IMP_DEROG M_DEROG
## 1 1100       25860         0     39025       0    10.5     0         0       0
## 2 1300       70053         0     68400       0     7.0     0         0       0
## 3 1500       13500         0     16700       0     4.0     0         0       0
## 4 1500       65000         1     89000       1     7.0     1         1       1
## 5 1700       97800         0    112000       0     3.0     0         0       0
## 6 1700       30548         0     40320       0     9.0     0         0       0
##   IMP_DELINQ M_DELINQ IMP_CLAGE M_CLAGE IMP_NINQ M_NINQ IMP_CLNO M_CLNO
## 1          0        0  94.36667       0        1      0        9      0
## 2          2        0 121.83333       0        0      0       14      0
## 3          0        0 149.46667       0        1      0       10      0
## 4          1        1 174.00000       1        1      1       20      1
## 5          0        0  93.33333       0        0      0       14      0
## 6          0        0 101.46600       0        1      0        8      0
##   IMP_DEBTINC M_DEBTINC FLAG.Job.Mgr FLAG.Job.Office FLAG.Job.Other
## 1    35.00000         1            0               0              1
## 2    35.00000         1            0               0              1
## 3    35.00000         1            0               0              1
## 4    35.00000         1            0               0              0
## 5    35.00000         1            0               1              0
## 6    37.11361         0            0               0              1
##   FLAG.Job.ProfExe FLAG.Job.Sales FLAG.Job.Self FLAG.Reason.DebtCon
## 1                0              0             0                   0
## 2                0              0             0                   0
## 3                0              0             0                   0
## 4                0              0             0                   0
## 5                0              0             0                   0
## 6                0              0             0                   0
##   FLAG.Reason.HomeImp
## 1                   1
## 2                   1
## 3                   1
## 4                   0
## 5                   1
## 6                   1
```

```
df_pca = df_pca[c(1,2,4,6,8,10,12,14,16,18)]
```

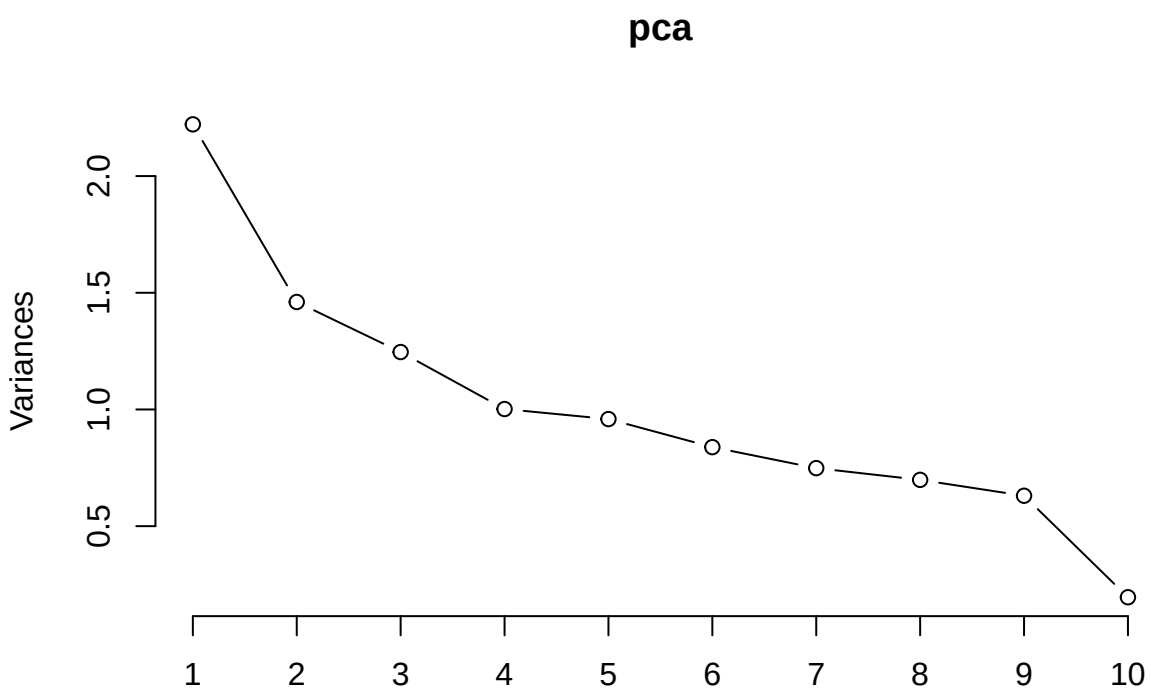
```r
head(df_pca)
```

```
##   LOAN IMP_MORTDUE IMP_VALUE IMP_YOJ IMP_DEROG IMP_DELINQ IMP_CLAGE IMP_NINQ
## 1 1100       25860     39025    10.5         0          0  94.36667        1
## 2 1300       70053     68400     7.0         0          2 121.83333        0
## 3 1500       13500     16700     4.0         0          0 149.46667        1
## 4 1500       65000     89000     7.0         1          1 174.00000        1
## 5 1700       97800    112000     3.0         0          0  93.33333        0
## 6 1700       30548     40320     9.0         0          0 101.46600        1
##   IMP_CLNO IMP_DEBTINC
## 1        9    35.00000
## 2       14    35.00000
## 3       10    35.00000
## 4       20    35.00000
## 5       14    35.00000
## 6        8    37.11361
```

```r
pca = prcomp(df_pca, center=TRUE, scale=TRUE)
summary(pca)
```

```
## 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
```

```r
plot(pca, type="l")
```




```
df_new = data.frame(predict(pca, df_pca))
```

```
head(df_new)
```

```
##          PC1        PC2         PC3          PC4        PC5         PC6
## 1 -2.4361630 -0.2914953  0.60058199 -0.013149908 -0.6714839 -0.16425779
## 2 -1.2657133  0.3930930 -0.07345171  1.208924161 -0.4381651 -1.04852828
## 3 -2.6621119 -0.1696773  0.58110691  0.404535849 -0.9232975  0.17214915
## 4 -0.7828377  0.8659403 -0.29103382  0.988600888 -0.3638980 -0.02551371
## 5 -0.5746093 -0.2924981  1.44854530  1.349271483 -0.4406566 -0.40675473
## 6 -2.3178901 -0.2111695  0.77284627 -0.004773515 -0.8042225 -0.27408475
##           PC7         PC8        PC9        PC10
## 1  0.81640905  0.52557847 -0.2980650  0.08839448
## 2  1.14291972 -0.35965773 -0.6286026 -0.21161255
## 3 -0.07742133  0.04792639 -0.4990894  0.05048375
## 4  0.31774708  0.70029614 -0.4336507  0.13596624
## 5  0.80616649  0.58663236 -0.2743750 -0.10138159
## 6  0.64071766  0.49542829 -0.4896611  0.02895187
```

```
df_flags = data
df_flags$PC1 = df_new["PC1"]
df_flags$PC2 = df_new["PC2"]
```

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

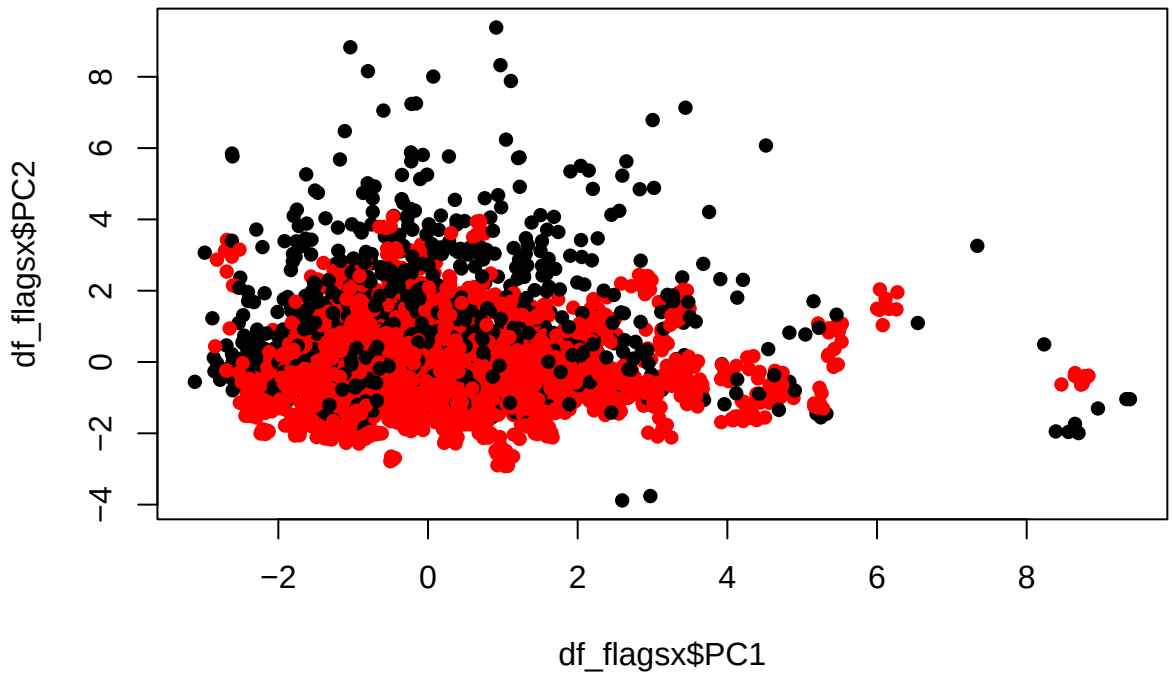
```
df_flags0 =df_flags[which(df_flags$TARGET_BAD_FLAG ==0), ]
df_flags1 =df_flags[which(df_flags$TARGET_BAD_FLAG ==1), ]
```

```
df_flags0 = df_flags0[df_flags$RAND1 < 25, ]
df_flags1 = df_flags1[df_flags$RAND1 < 75, ]
```

```
df_flagsx = rbind(df_flags0, df_flags1)
df_flagsx = df_flagsx[df_flagsx$RAND2 < 15, ]
```

```
df_flagsx = df_flags
```

```
colors = c("#00AFBB", "#E7B800")
colors = c("red","black")
colors=colors[df_flagsx$TARGET_BAD_FLAG + 1]
plot(df_flagsx$PC1, df_flagsx$PC2, col=colors, pch=16)
```

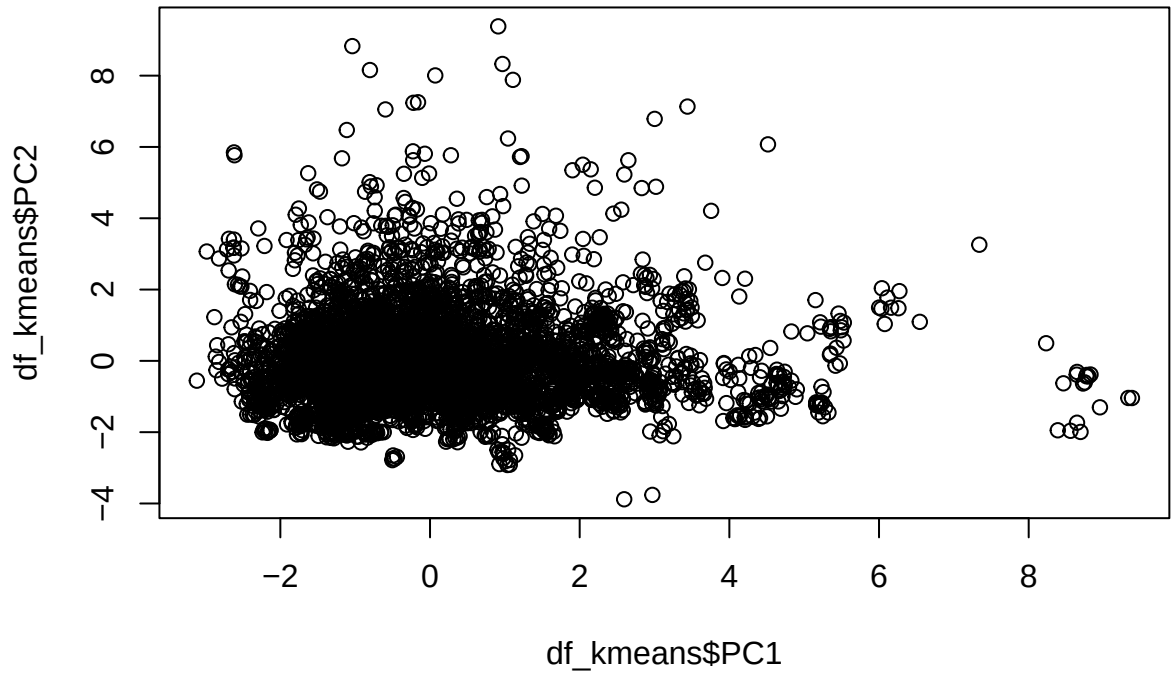



```
df_kmeans = df_new[1:2]
print(head(df_kmeans))
```

Cluster Analysis - Find the Number of Clusters

```
##          PC1        PC2
## 1 -2.4361630 -0.2914953
## 2 -1.2657133  0.3930930
## 3 -2.6621119 -0.1696773
## 4 -0.7828377  0.8659403
## 5 -0.5746093 -0.2924981
## 6 -2.3178901 -0.2111695
```

```
plot(df_kmeans$PC1, df_kmeans$PC2)
```




```
MAX_N = 10

WSS = numeric( MAX_N )

for ( N in 1:MAX_N )
    {
    km = kmeans( df_kmeans, centers=N, nstart=20  )
    WSS[N] = km$tot.withinss
    }
```

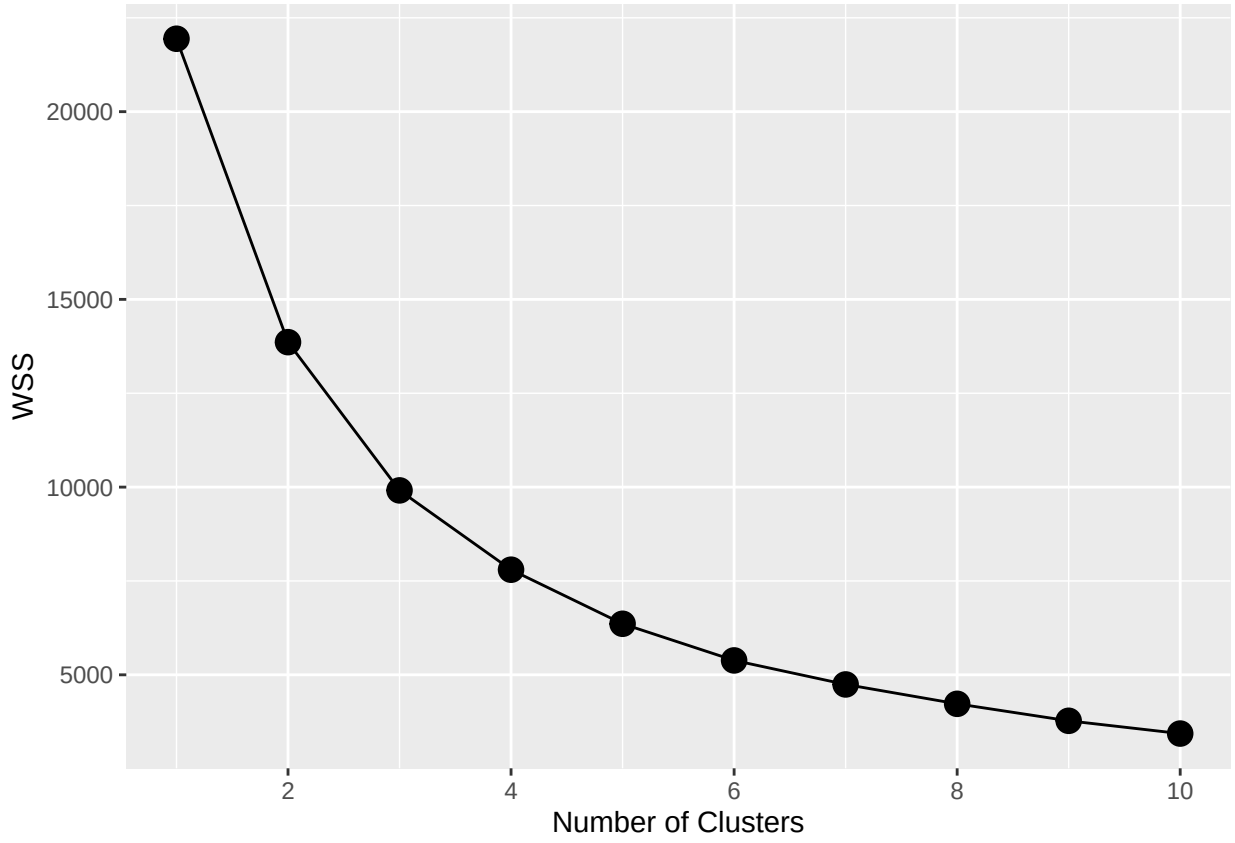
```
## Warning: did not converge in 10 iterations
```

```
## Warning: Quick-TRANSfer stage steps exceeded maximum (= 298000)
```

```
df_wss = as.data.frame( WSS )
df_wss$clusters = 1:MAX_N

scree_plot = ggplot( df_wss, aes( x=clusters, y=WSS, group=1 )) +
        geom_point( size=4 ) +
        geom_line() +
        scale_x_continuous( breaks=c(2,4,6,8,10)) +
        xlab("Number of Clusters")

scree_plot
```




```
BEST_N = 4
km = kmeans( df_kmeans, centers=BEST_N, nstart=20  )

print( km$size )
```

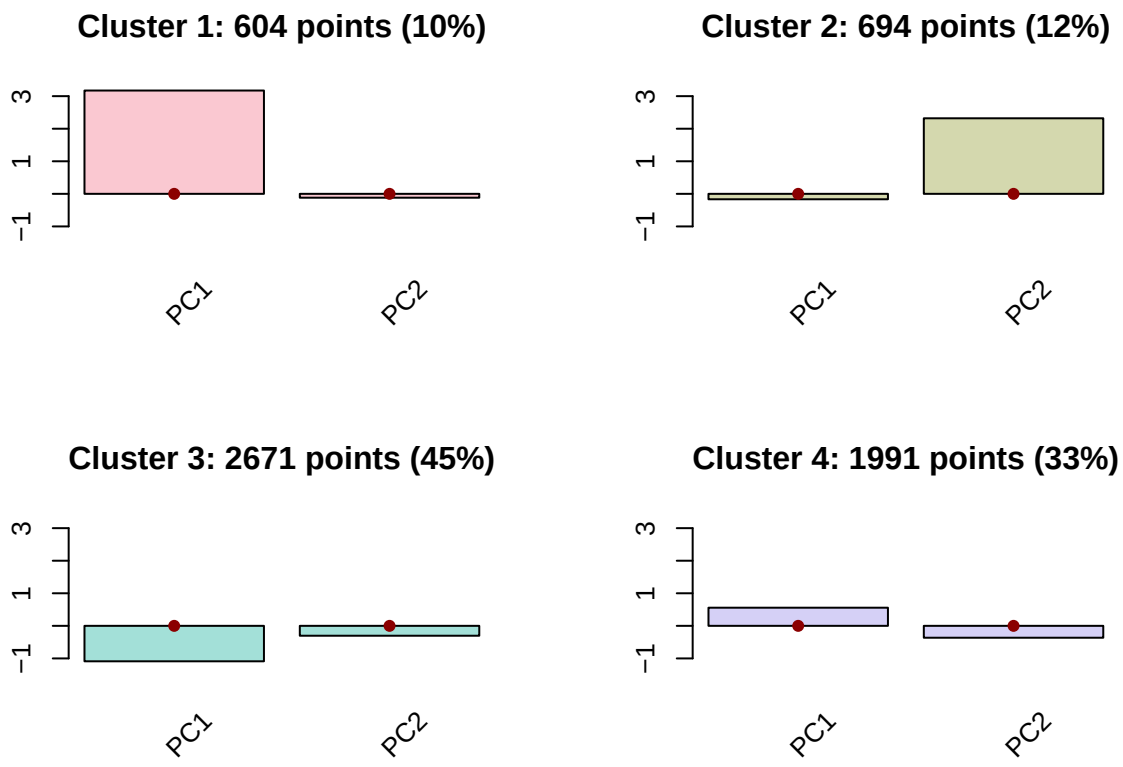
```
## [1]  604  694 2671 1991
```

```
print( km$centers )
```

```
##          PC1        PC2
## 1  3.1726984 -0.1185246
## 2 -0.1673616  2.3202858
## 3 -1.0891828 -0.3035157
## 4  0.5570299 -0.3656450
```

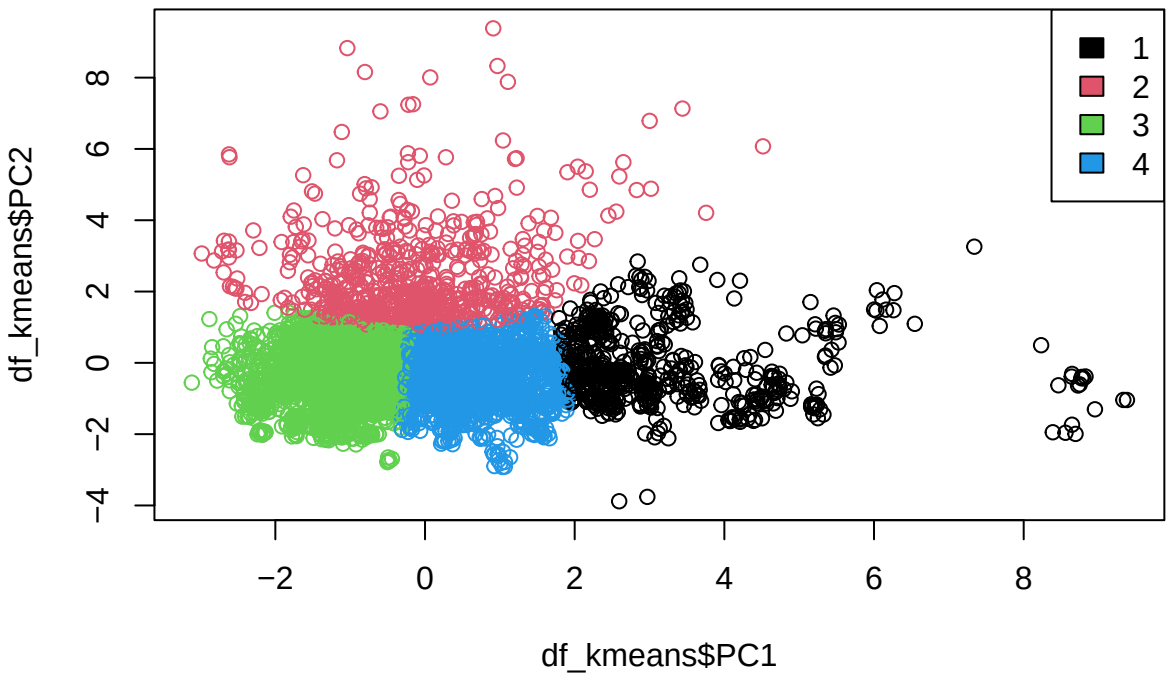
```
kf = as.kcca( object=km, data=df_kmeans, save.data=TRUE )
kfi = kcca2df( kf )
agg = aggregate( kfi$value, list( kfi$variable, kfi$group ), FUN=mean )
```

```
barplot(kf)
```



In cluster 1, the loan amount was lower and mortgage was also not much In cluster 2, the loan amount was higher and mortgage was lower In cluster 3, the loan amount was lower and mortgage was higher In cluster 4, the loan amount was not much higher and mortgage was lower

```
clus = predict( kf, df_kmeans )
plot( df_kmeans$PC1, df_kmeans$PC2, col=clus )
legend( x="topright", legend=c(1:BEST_N), fill=c(1:BEST_N) )
```




```
data$CLUSTER = clus
```

```
head(data)
```

```
##   TARGET_BAD_FLAG TARGET_LOSS_AMT LOAN IMP_MORTDUE M_MORTDUE IMP_VALUE M_VALUE
## 1               1             641 1100       25860         0     39025       0
## 2               1            1109 1300       70053         0     68400       0
## 3               1             767 1500       13500         0     16700       0
## 4               1            1425 1500       65000         1     89000       1
## 5               0               0 1700       97800         0    112000       0
## 6               1             335 1700       30548         0     40320       0
##   IMP_YOJ M_YOJ IMP_DEROG M_DEROG IMP_DELINQ M_DELINQ IMP_CLAGE M_CLAGE
## 1    10.5     0          0       0          0        0  94.36667       0
## 2     7.0     0          0       0          2        0 121.83333       0
## 3     4.0     0          0       0          0        0 149.46667       0
## 4     7.0     1          1       1          1        1 174.00000       1
## 5     3.0     0          0       0          0        0  93.33333       0
## 6     9.0     0          0       0          0        0 101.46600       0
##   IMP_NINQ M_NINQ IMP_CLNO M_CLNO IMP_DEBTINC M_DEBTINC FLAG.Job.Mgr
## 1        1      0        9      0    35.00000         1            0
## 2        0      0       14      0    35.00000         1            0
## 3        1      0       10      0    35.00000         1            0
## 4        1      1       20      1    35.00000         1            0
## 5        0      0       14      0    35.00000         1            0
## 6        1      0        8      0    37.11361         0            0
```

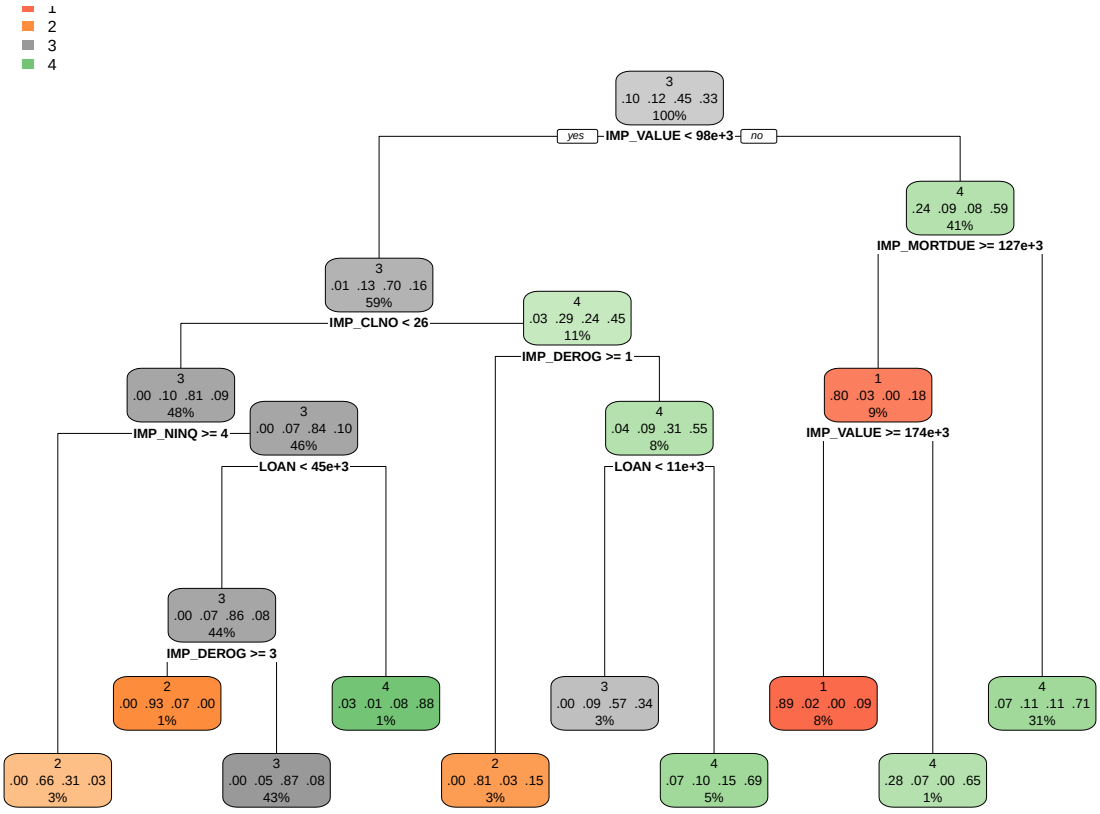
```
##   FLAG.Job.Office FLAG.Job.Other FLAG.Job.ProfExe FLAG.Job.Sales FLAG.Job.Self
## 1               0              1                0              0             0
## 2               0              1                0              0             0
## 3               0              1                0              0             0
## 4               0              0                0              0             0
## 5               1              0                0              0             0
## 6               0              1                0              0             0
##   FLAG.Reason.DebtCon FLAG.Reason.HomeImp CLUSTER
## 1                   0                   1       3
## 2                   0                   1       3
## 3                   0                   1       3
## 4                   0                   0       3
## 5                   0                   1       3
## 6                   0                   1       3
```

```
agg = aggregate( data$TARGET_BAD_FLAG, list( data$CLUSTER ), FUN=mean )
agg
```

```
##   Group.1         x
## 1       1 0.1705298
## 2       2 0.5547550
## 3       3 0.1860726
## 4       4 0.1024611
```

```
df_tree = df_pca
df_tree$CLUSTER = as.factor(clus)
dt = rpart( CLUSTER ~ . , data=df_tree )
```

```
rpart.plot( dt )
```



The decision tree model analyzes loan approval based on various features related to an individual's financial situation. It starts by evaluating the house value (IMP_VALUE), with a threshold set at 58,823 dollars. If the house value falls below this threshold, the model proceeds to assess other factors. Otherwise, it moves on to the next step, considering the amount owed on the mortgage (IMP_MORTGAGE). If the mortgage debt is less than or equal to $127,623, the model predicts loan approval; otherwise, it predicts denial. Additionally, the model incorporates additional features such as the number of lines of credit (IMP_CLNO), inquiries (IMP_NINQ), home equity loan amount (LOAN), and derogatory loans (IMP_DEBROG) to refine its prediction. These conditions collectively determine whether a loan is approved or denied.