Capstone TCD project note II

Suprasanna Pradhan

26 September 2019

```
library(caret)
```

```
## Loading required package: lattice
```

```
## Loading required package: ggplot2
```

```
library(class)
library(devtools)
```

```
## Loading required package: usethis
```

```
library(e1071)
library(ggplot2)
library(Hmisc)
```

```
## Loading required package: survival
```

```
##
## Attaching package: 'survival'
```

```
## The following object is masked from 'package:caret':
##
##     cluster
```

```
## Loading required package: Formula
```

```
##
## Attaching package: 'Hmisc'
```

```
## The following object is masked from 'package:e1071':
##
##     impute
```

```
## The following objects are masked from 'package:base':
##
##     format.pval, units
```

```
library(klaR)
```

```
## Loading required package: MASS
```

```
library(MASS)
library(nnet)
library(plyr)
```

```
##
## Attaching package: 'plyr'
```

```
## The following objects are masked from 'package:Hmisc':
##
##     is.discrete, summarize
```

```
library(pROC)
```

```
## Type 'citation("pROC")' for a citation.
```

```
##
## Attaching package: 'pROC'
```

```
## The following objects are masked from 'package:stats':
##
##     cov, smooth, var
```

```
library(psych)
```

```
##
## Attaching package: 'psych'
```

```
## The following object is masked from 'package:Hmisc':
##
##     describe
```

```
## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```

```
library(scatterplot3d)
library(SDMTools)
```

```
## Registered S3 method overwritten by 'R.oo':
##   method        from
##   throw.default R.methodsS3
```

```
##
## Attaching package: 'SDMTools'
```

```
## The following object is masked from 'package:pROC':
##
##     auc
```

```
## The following objects are masked from 'package:caret':
##
##     sensitivity, specificity
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:plyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize
```

```
## The following object is masked from 'package:MASS':
##
##     select
```

```
## The following objects are masked from 'package:Hmisc':
##
##     src, summarize
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ElemStatLearn)
```

```
##
## Attaching package: 'ElemStatLearn'
```

```
## The following object is masked from 'package:plyr':
##
##     ozone
```

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

```
## randomForest 4.6-14
```

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

```
##
## Attaching package: 'randomForest'
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

```
## The following object is masked from 'package:psych':
##
##     outlier
```

```
## The following object is masked from 'package:ggplot2':
##
##     margin
```

```
library(neuralnet)
```

```
##
## Attaching package: 'neuralnet'
```

```
## The following object is masked from 'package:dplyr':
##
##     compute
```

Recap of Project note 1

. In the data set earlier we found it consists 30000 observations with 25 variables. . The categorical data value like Sex , martial status and education has changed to numeric value . . We have realized that 22.1 % percent defaulter and 77.9% are not default cases . Default category whereas male customer are 9.6% and female category shows 12.5% . . University level - graduate or PG is more into default side . Married customers somehow leaning to tend defaulter . Average age of 25 to 30 is the highest risk . . We have also checked the multicollinearity problems is existed In the data set . . pay status categorical variables are dependent on each other and impact of REPAY_ SEP to REPAY_APR variables to default. Payment DEFAULT is high. . We have also created some dummy variables like ratio of the payment for each month SEP to APR and balance amount month wise from SEP to APR .

Supplimantry Over view of EDA

We have been checking here the relationship of paid amount SEP to APR and balance amount((billed outstanding amount - paid amount) we assumed that bank is keep on adding the preceding month balance to current month)

importing the data from Note 1

Ratio is alrady added in the data set .

```
library(readxl)
#Importing Data set
setwd("C:/Users/SuprasannaPradhan/Documents/My Files/Great Lakes Projects/Capstone Pro
ject TCD")
bank_data_2nd=read.csv("bank_data_new.csv")
bank_data_2=read.csv("taiwan_bank3.csv")
names(bank_data_2)
```

```
##  [1] "X"            "BILL_AMT_SEP" "BILL_AMT_AUG" "BILL_AMT_JUL"
##  [5] "BILL_AMT_JUN" "BILL_AMT_MAY" "BILL_AMT_APR" "PAID_AMT_SEP"
##  [9] "PAID_AMT_AUG" "PAID_AMT_JUL" "PAID_AMT_JUN" "PAID_AMT_MAY"
## [13] "PAID_AMT_APR" "DEFAULT"      "RATIO_SEP"    "RATIO_AUG"
## [17] "RATIO_JUL"    "RATIO_JUN"    "RATIO_MAY"    "RATIO_APR"
## [21] "balance_SEP"  "balance_AUG"  "balance_JUL"  "balance_JUN"
## [25] "balance_MAY"  "balance_APR"
```

Finding outliers and NA values

```
bank_data_n1 <- subset(bank_data_2nd, select = -c(1:2))
bank_data1 <- cbind(bank_data_n1,bank_data_2[21:26])
str(bank_data1)
```

```
## 'data.frame':    30000 obs. of  36 variables:
##  $ LIMIT_BAL   : num  20000 120000 90000 50000 50000 50000 500000 100000 140000 200
00 ...
##  $ SEX         : int  2 2 2 2 1 1 1 2 2 1 ...
##  $ EDUCATION   : int  2 2 2 2 2 1 1 2 3 3 ...
##  $ MARRIAGE    : int  1 2 2 1 1 2 2 2 1 2 ...
##  $ AGE         : int  24 26 34 37 57 37 29 23 28 35 ...
##  $ REPAY_SEP   : int  2 -1 0 0 -1 0 0 0 0 -2 ...
##  $ REPAY_AUG   : int  2 2 0 0 0 0 0 -1 0 -2 ...
##  $ REPAY_JUL   : int  -1 0 0 0 -1 0 0 -1 2 -2 ...
##  $ REPAY_JUN   : int  -1 0 0 0 0 0 0 0 0 -2 ...
##  $ REPAY_MAY   : int  -2 0 0 0 0 0 0 0 0 -1 ...
##  $ REPAY_APR   : int  -2 2 0 0 0 0 0 -1 0 -1 ...
##  $ BILL_AMT_SEP: num  3913 2682 29239 46990 8617 ...
##  $ BILL_AMT_AUG: num  3102 1725 14027 48233 5670 ...
##  $ BILL_AMT_JUL: num  689 2682 13559 49291 35835 ...
##  $ BILL_AMT_JUN: num  0 3272 14331 28314 20940 ...
##  $ BILL_AMT_MAY: num  0 3455 14948 28959 19146 ...
##  $ BILL_AMT_APR: num  0 3261 15549 29547 19131 ...
##  $ PAID_AMT_SEP: num  0 0 1518 2000 2000 ...
##  $ PAID_AMT_AUG: num  689 1000 1500 2019 36681 ...
##  $ PAID_AMT_JUL: num  0 1000 1000 1200 10000 657 38000 0 432 0 ...
##  $ PAID_AMT_JUN: num  0 1000 1000 1100 9000 ...
##  $ PAID_AMT_MAY: num  0 0 1000 1069 689 ...
##  $ PAID_AMT_APR: num  0 2000 5000 1000 679 ...
##  $ DEFAULT     : int  1 1 0 0 0 0 0 0 0 0 ...
##  $ RATIO_SEP   : num  0 0 5 4 23 4 15 3 29 NA ...
##  $ RATIO_AUG   : num  22 58 11 4 647 3 10 158 0 NA ...
##  $ RATIO_JUL   : num  0 37 7 2 28 1 9 0 4 NA ...
##  $ RATIO_JUN   : num  NA 31 7 4 43 ...
##  $ RATIO_MAY   : num  NA 29 7 4 47 5 4 -365 8 100 ...
##  $ RATIO_APR   : num  NA 31 6 4 47 5 4 102 27 93 ...
##  $ balance_SEP : num  3913 2682 27721 44990 6617 ...
##  $ balance_AUG : num  2413 725 12527 46214 -31011 ...
##  $ balance_JUL : num  689 1682 12559 48091 25835 ...
##  $ balance_JUN : num  0 2272 13331 27214 11940 ...
##  $ balance_MAY : num  0 3455 13948 27890 18457 ...
##  $ balance_APR : num  0 1261 10549 28547 18452 ...
```

```
summary(bank_data1)
```

```
##    LIMIT_BAL            SEX          EDUCATION        MARRIAGE
##  Min.   :  10000   Min.   :1.000   Min.   :0.000   Min.   :0.000
##  1st Qu.:  50000   1st Qu.:1.000   1st Qu.:1.000   1st Qu.:1.000
##  Median : 140000   Median :2.000   Median :2.000   Median :2.000
##  Mean   : 167484   Mean   :1.604   Mean   :1.853   Mean   :1.552
##  3rd Qu.: 240000   3rd Qu.:2.000   3rd Qu.:2.000   3rd Qu.:2.000
##  Max.   :1000000   Max.   :2.000   Max.   :6.000   Max.   :3.000
##
##       AGE          REPAY_SEP         REPAY_AUG         REPAY_JUL
##  Min.   :21.00   Min.   :-2.0000   Min.   :-2.0000   Min.   :-2.0000
##  1st Qu.:28.00   1st Qu.:-1.0000   1st Qu.:-1.0000   1st Qu.:-1.0000
##  Median :34.00   Median : 0.0000   Median : 0.0000   Median : 0.0000
##  Mean   :35.49   Mean   :-0.0167   Mean   :-0.1338   Mean   :-0.1662
##  3rd Qu.:41.00   3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.: 0.0000
##  Max.   :79.00   Max.   : 8.0000   Max.   : 8.0000   Max.   : 8.0000
##
##    REPAY_JUN         REPAY_MAY         REPAY_APR        BILL_AMT_SEP
##  Min.   :-2.0000   Min.   :-2.0000   Min.   :-2.0000   Min.   :-165580
##  1st Qu.:-1.0000   1st Qu.:-1.0000   1st Qu.:-1.0000   1st Qu.:   3559
##  Median : 0.0000   Median : 0.0000   Median : 0.0000   Median :  22382
##  Mean   :-0.2207   Mean   :-0.2662   Mean   :-0.2911   Mean   :  51223
##  3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.:  67091
##  Max.   : 8.0000   Max.   : 8.0000   Max.   : 8.0000   Max.   : 964511
##
##   BILL_AMT_AUG     BILL_AMT_JUL      BILL_AMT_JUN      BILL_AMT_MAY
##  Min.   :-69777   Min.   :-157264   Min.   :-170000   Min.   :-81334
##  1st Qu.:  2985   1st Qu.:   2666   1st Qu.:   2327   1st Qu.:  1763
##  Median : 21200   Median :  20089   Median :  19052   Median : 18105
##  Mean   : 49179   Mean   :  47013   Mean   :  43263   Mean   : 40311
##  3rd Qu.: 64006   3rd Qu.:  60165   3rd Qu.:  54506   3rd Qu.: 50191
##  Max.   :983931   Max.   :1664089   Max.   : 891586   Max.   :927171
##
##   BILL_AMT_APR      PAID_AMT_SEP      PAID_AMT_AUG      PAID_AMT_JUL
##  Min.   :-339603   Min.   :     0   Min.   :      0   Min.   :     0
##  1st Qu.:   1256   1st Qu.:  1000   1st Qu.:    833   1st Qu.:   390
##  Median :  17071   Median :  2100   Median :   2009   Median :  1800
##  Mean   :  38872   Mean   :  5664   Mean   :   5921   Mean   :  5226
##  3rd Qu.:  49198   3rd Qu.:  5006   3rd Qu.:   5000   3rd Qu.:  4505
##  Max.   : 961664   Max.   :873552   Max.   :1684259   Max.   :896040
##
##   PAID_AMT_JUN      PAID_AMT_MAY       PAID_AMT_APR         DEFAULT
##  Min.   :     0   Min.   :     0.0   Min.   :     0.0   Min.   :0.0000
##  1st Qu.:   296   1st Qu.:   252.5   1st Qu.:   117.8   1st Qu.:0.0000
##  Median :  1500   Median :  1500.0   Median :  1500.0   Median :0.0000
##  Mean   :  4826   Mean   :  4799.4   Mean   :  5215.5   Mean   :0.2212
##  3rd Qu.:  4013   3rd Qu.:  4031.5   3rd Qu.:  4000.0   3rd Qu.:0.0000
##  Max.   :621000   Max.   :426529.0   Max.   :528666.0   Max.   :1.0000
##
```

```
##    RATIO_SEP          RATIO_AUG          RATIO_JUL
##  Min.   :-3543600   Min.   :-1025950   Min.   :-8215000
##  1st Qu.:       4   1st Qu.:       4   1st Qu.:       3
##  Median :       6   Median :       6   Median :       5
##  Mean   :     Inf   Mean   :     Inf   Mean   :     Inf
##  3rd Qu.:      30   3rd Qu.:      30   3rd Qu.:      25
##  Max.   :     Inf   Max.   :     Inf   Max.   :     Inf
##  NA's   :1468       NA's   :1867       NA's   :2246
##    RATIO_JUN          RATIO_MAY          RATIO_APR          balance_SEP
##  Min.   :-1726667   Min.   :-430667   Min.   :-1942500   Min.   :-733744
##  1st Qu.:       3   1st Qu.:      4   1st Qu.:       4   1st Qu.:    745
##  Median :       4   Median :      5   Median :       5   Median :  18551
##  Mean   :     Inf   Mean   :    Inf   Mean   :     Inf   Mean   :  45560
##  3rd Qu.:      22   3rd Qu.:     70   3rd Qu.:      38   3rd Qu.:  62242
##  Max.   :     Inf   Max.   :    Inf   Max.   :     Inf   Max.   : 913727
##  NA's   :2470       NA's   :3449      NA's   :2952
##   balance_AUG          balance_JUL         balance_JUN
##  Min.   :-1702347.0   Min.   :-854641.0   Min.   :-667000
##  1st Qu.:     329.5   1st Qu.:    262.8   1st Qu.:    230
##  Median :   18102.5   Median :  17769.0   Median :  16970
##  Mean   :   43257.9   Mean   :  41787.5   Mean   :  38437
##  3rd Qu.:   59077.8   3rd Qu.:  56294.2   3rd Qu.:  50260
##  Max.   :  933208.0   Max.   :1542258.0   Max.   : 841586
##
##   balance_MAY       balance_APR
##  Min.   :-414380   Min.   :-684896
##  1st Qu.:      0   1st Qu.:      0
##  Median :  15538   Median :  13926
##  Mean   :  35512   Mean   :  33656
##  3rd Qu.:  46962   3rd Qu.:  46067
##  Max.   : 877171   Max.   : 911408
##
```

we got here few variables are having outliers ,found almost all continours frequancy based variable are having outliers but variables are having extem outlier values these are limit balance, BILL_AMT_JUL ,PAID_AMT_AUG,balance_JUL Dummy ration created from SEP to APR shows lot of NA and inf values.

Handling of Ratio paybill amount

Ratio Pay Amount Bill Amount Negative/NaN number 1. If Bill Amount =0 : Then convert Ratio to positive 2. Impute NaN to 1 (Higher the ratio, lesser the chance of Default)

```
bank_data1[is.na(bank_data1)] <- 0
sum(is.na(bank_data1))
```

```
## [1] 0
```

```
bank_data1$RATIO_SEP=ifelse(bank_data1$RATIO_SEP<=100,0,1)
bank_data1$RATIO_AUG=ifelse(bank_data1$RATIO_AUG<=100,0,1)
bank_data1$RATIO_JUL=ifelse(bank_data1$RATIO_JUL<=100,0,1)
bank_data1$RATIO_JUN=ifelse(bank_data1$RATIO_JUN<=100,0,1)
bank_data1$RATIO_MAY=ifelse(bank_data1$RATIO_MAY<=100,0,1)
bank_data1$RATIO_APR=ifelse(bank_data1$RATIO_APR<=100,0,1)
str(bank_data1[25:30])
```

```
## 'data.frame':    30000 obs. of  6 variables:
##  $ RATIO_SEP: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ RATIO_AUG: num  0 0 0 0 1 0 0 1 0 0 ...
##  $ RATIO_JUL: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ RATIO_JUN: num  0 0 0 0 0 0 0 1 0 1 ...
##  $ RATIO_MAY: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ RATIO_APR: num  0 0 0 0 0 0 0 1 0 0 ...
```

```
View(bank_data1)
```

```
#write.csv(bank_data1, file = 'bank_data1.csv')
table(bank_data1$RATIO_SEP)
```

```
##
##     0     1
## 26302  3698
```

```
table(bank_data1$RATIO_AUG)
```

```
##
##     0     1
## 26174  3826
```

```
table(bank_data1$RATIO_JUL)
```

```
##
##     0     1
## 26423  3577
```

```
table(bank_data1$RATIO_JUN)
```

```
##
##     0     1
## 26613  3387
```

```
table(bank_data1$RATIO_MAY)
```

```
##
##     0     1
## 28992  1008
```

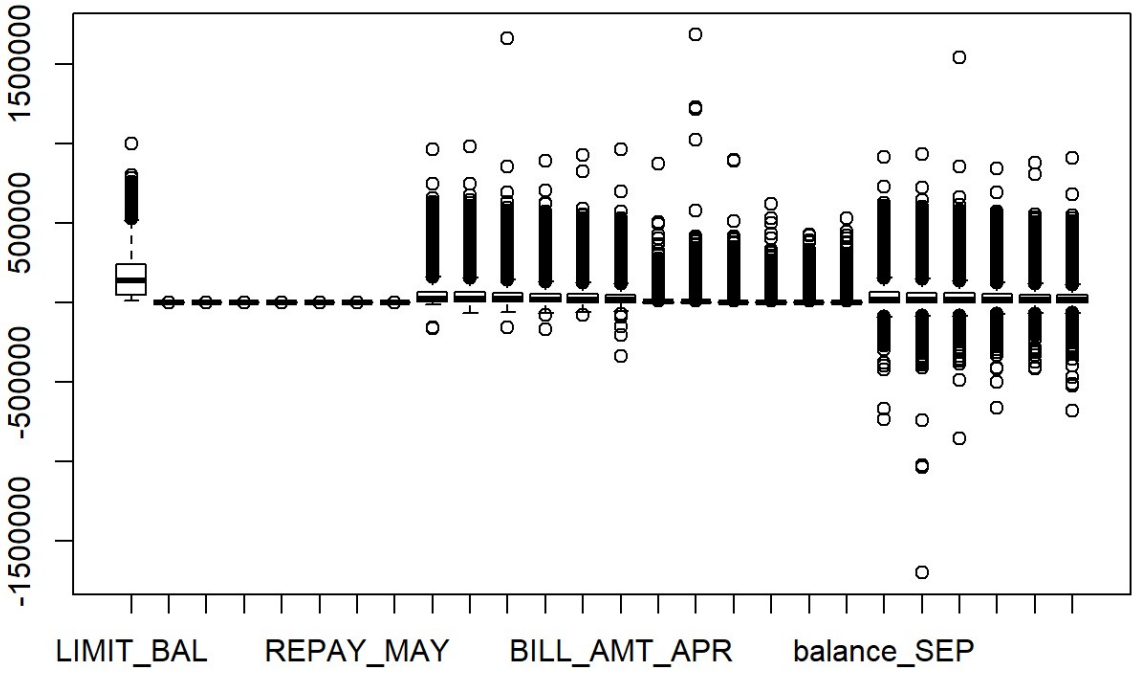
```
table(bank_data1$RATIO_APR)
```

```
##
##     0     1
## 26043  3957
```

```
bank_data_bp <- subset(bank_data1,select = -c(2,3,4,24:30))
names(bank_data_bp)
```

```
##  [1] "LIMIT_BAL"    "AGE"          "REPAY_SEP"    "REPAY_AUG"
##  [5] "REPAY_JUL"    "REPAY_JUN"    "REPAY_MAY"    "REPAY_APR"
##  [9] "BILL_AMT_SEP" "BILL_AMT_AUG" "BILL_AMT_JUL" "BILL_AMT_JUN"
## [13] "BILL_AMT_MAY" "BILL_AMT_APR" "PAID_AMT_SEP" "PAID_AMT_AUG"
## [17] "PAID_AMT_JUL" "PAID_AMT_JUN" "PAID_AMT_MAY" "PAID_AMT_APR"
## [21] "balance_SEP"  "balance_AUG"  "balance_JUL"  "balance_JUN"
## [25] "balance_MAY"  "balance_APR"
```

```
boxplot(bank_data_bp)
```

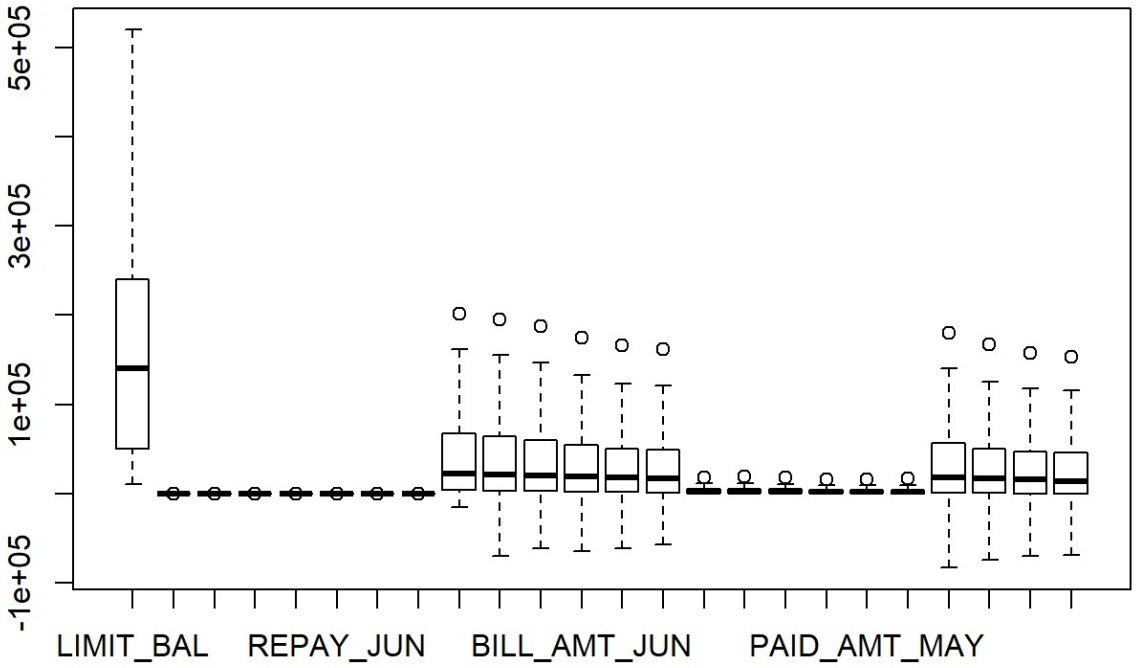


Ratio culmun we are not checking the outliers becuse it values are very low .it is show us the size of the amount ahs been paid for each month ,hence we not incluing these variabels for outlier checking

```
#Capping vairbale outlieres
capOutlier <- function(x){
   qnt <- quantile(x, probs=c(.25, .75), na.rm = T)
   caps <- quantile(x, probs=c(.05, .95), na.rm = T)
   H <- 1.5 * IQR(x, na.rm = T)
   x[x < (qnt[1] - H)] <- caps[1]
   x[x > (qnt[2] + H)] <- caps[2]
   return(x)
}
bank_data1$LIMIT_BAL=capOutlier(bank_data1$LIMIT_BAL)
bank_data1$BILL_AMT_SEP=capOutlier(bank_data1$BILL_AMT_SEP)
bank_data1$BILL_AMT_AUG=capOutlier(bank_data1$BILL_AMT_AUG)
bank_data1$BILL_AMT_JUL=capOutlier(bank_data1$BILL_AMT_JUL)
bank_data1$BILL_AMT_JUN=capOutlier(bank_data1$BILL_AMT_JUN)
bank_data1$BILL_AMT_MAY=capOutlier(bank_data1$BILL_AMT_MAY)
bank_data1$BILL_AMT_APR=capOutlier(bank_data1$BILL_AMT_APR)
bank_data1$PAID_AMT_SEP=capOutlier(bank_data1$PAID_AMT_SEP)
bank_data1$PAID_AMT_AUG=capOutlier(bank_data1$PAID_AMT_AUG)
bank_data1$PAID_AMT_JUL=capOutlier(bank_data1$PAID_AMT_JUL)
bank_data1$PAID_AMT_JUN=capOutlier(bank_data1$PAID_AMT_JUN)
bank_data1$PAID_AMT_MAY=capOutlier(bank_data1$PAID_AMT_MAY)
bank_data1$PAID_AMT_APR=capOutlier(bank_data1$PAID_AMT_APR)
bank_data1$balance_SEP=capOutlier(bank_data1$balance_SEP)
bank_data1$balance_AUG=capOutlier(bank_data1$balance_AUG)
bank_data1$balance_JUL=capOutlier(bank_data1$balance_JUL)
bank_data1$balance_JUN=capOutlier(bank_data1$balance_JUN)
bank_data1$balance_MAY=capOutlier(bank_data1$balance_MAY)
bank_data1$balance_APR=capOutlier(bank_data1$balance_APR)
```

Cappped outliers for these variables

```
bank_data_bp1 <- subset(bank_data1,select = -c(2,3,4,24:32))
boxplot(bank_data_bp1)
```




```
##Checking Data ##
colSums(is.na(bank_data1))
```

```
##    LIMIT_BAL          SEX    EDUCATION     MARRIAGE          AGE
##            0            0            0            0            0
##    REPAY_SEP    REPAY_AUG    REPAY_JUL    REPAY_JUN    REPAY_MAY
##            0            0            0            0            0
##    REPAY_APR BILL_AMT_SEP BILL_AMT_AUG BILL_AMT_JUL BILL_AMT_JUN
##            0            0            0            0            0
## BILL_AMT_MAY BILL_AMT_APR PAID_AMT_SEP PAID_AMT_AUG PAID_AMT_JUL
##            0            0            0            0            0
## PAID_AMT_JUN PAID_AMT_MAY PAID_AMT_APR      DEFAULT    RATIO_SEP
##            0            0            0            0            0
##    RATIO_AUG    RATIO_JUL    RATIO_JUN    RATIO_MAY    RATIO_APR
##            0            0            0            0            0
##  balance_SEP  balance_AUG  balance_JUL  balance_JUN  balance_MAY
##            0            0            0            0            0
##  balance_APR
##            0
```

All NA are shown in the dummy variable of ratio, we understand this values are occured due to the paid amount by the cutomser are not paid as the outstanding bill amount shown each month or very less againg of outstandin billed amount .So we would be keeping in as it is , if needed in future we will make them zero .

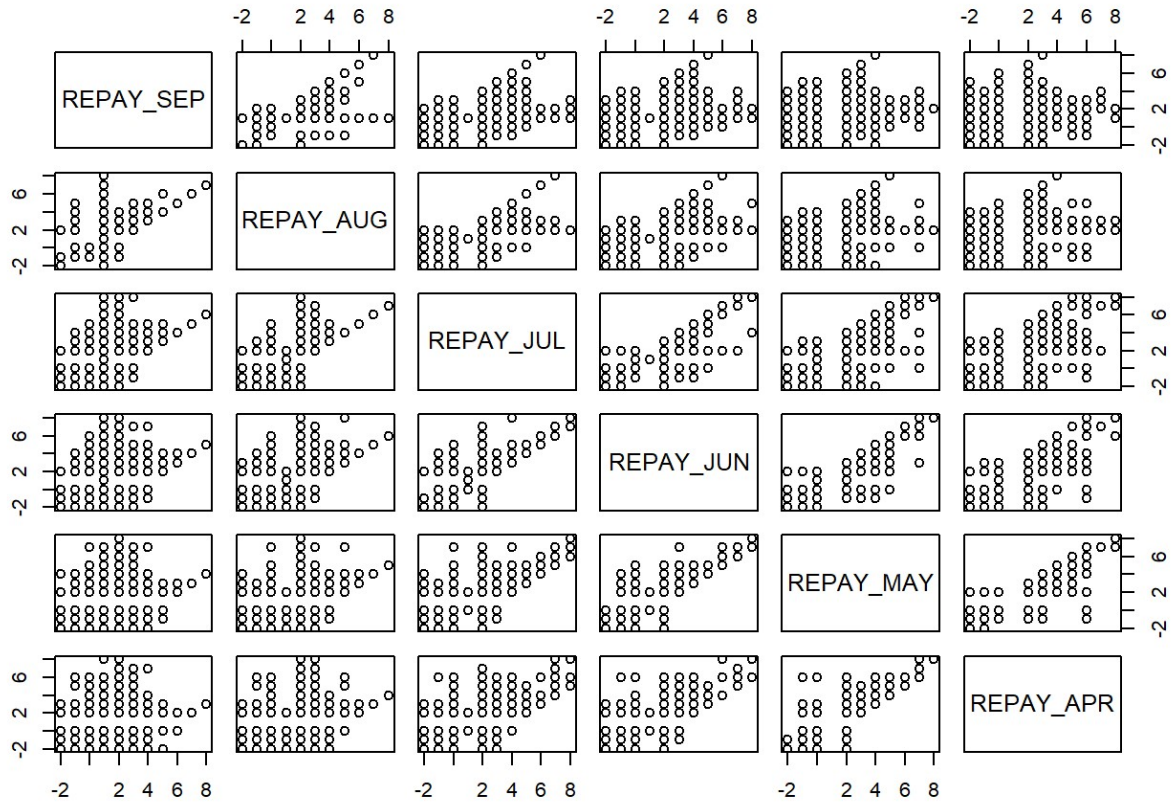
let us chekc the REPAY variables and clsifiy them , timely paid and delayed in payeemnt

According to the description, this REPAY_SEP TO REPAy_APR is a set of categorical variables with the levels:-2 is No consumption; -1 is Paid in full; 0 is The use of revolving credit; 1 is payment delay for one month; 2 is payment delay for two months; . . .; . 8=payment delay for eight months, We have already checked almost all months tend is very high with value of “0”, so its indicate moreover people are paid very minimal amount with revolving credit

```
names(bank_data1[6:11])
```

```
## [1] "REPAY_SEP" "REPAY_AUG" "REPAY_JUL" "REPAY_JUN" "REPAY_MAY" "REPAY_APR"
```

```
plot(bank_data1[6:11])
```




```
table(bank_data1$REPAY_SEP)
```

```
##
##    -2    -1     0     1     2     3     4     5     6     7     8
##  2759  5686 14737  3688  2667   322    76    26    11     9    19
```

```
table(bank_data1$REPAY_AUG)
```

```
##
##    -2    -1     0     1     2     3     4     5     6     7     8
##  3782  6050 15730    28  3927   326    99    25    12    20     1
```

```
table(bank_data1$REPAY_JUL)
```

```
##
##    -2    -1     0     1     2     3     4     5     6     7     8
##  4085  5938 15764     4  3819   240    76    21    23    27     3
```

```
table(bank_data1$REPAY_JUN)
```

```
##
##    -2    -1     0     1     2     3     4     5     6     7     8
##  4348  5687 16455     2  3159   180    69    35     5    58     2
```

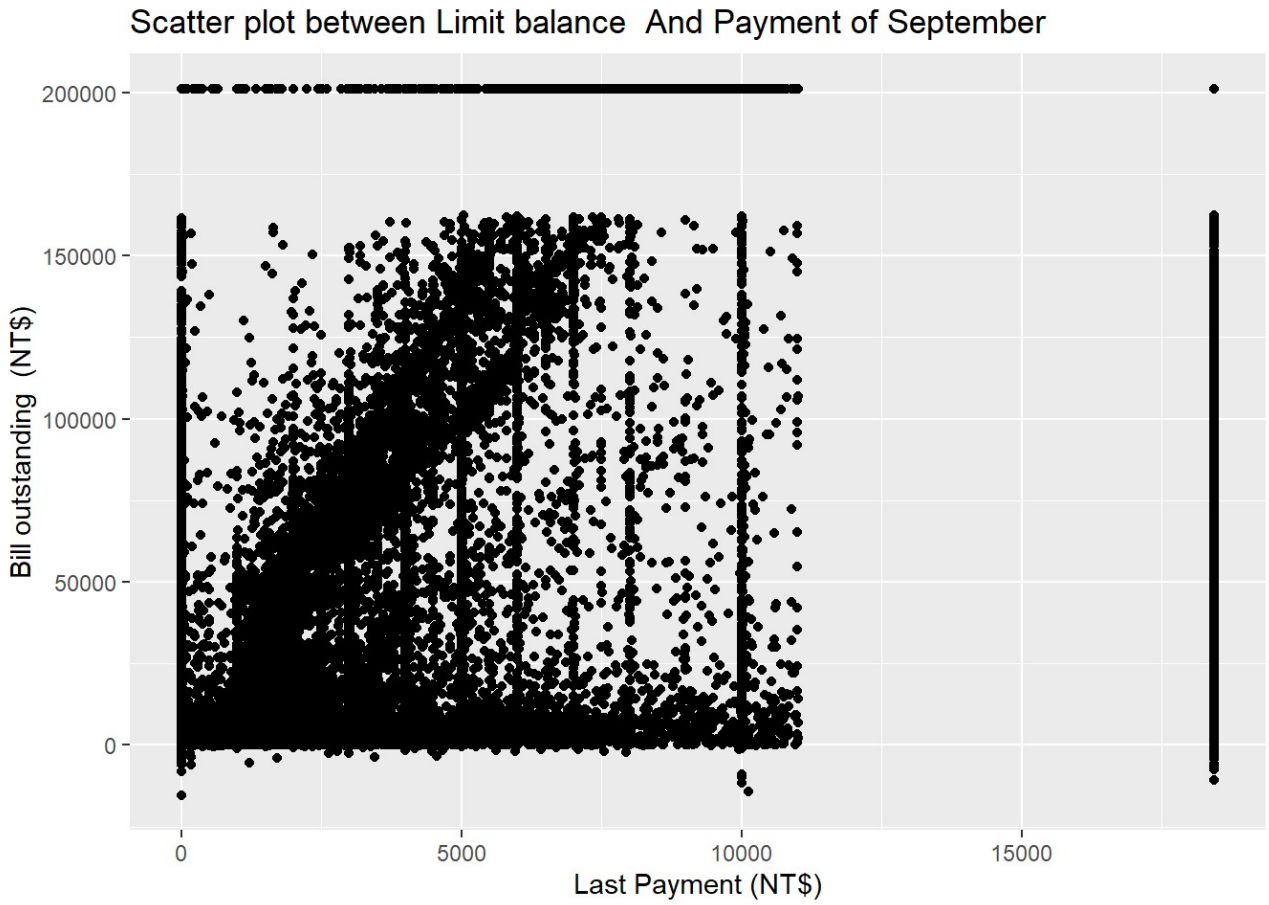
```
table(bank_data1$REPAY_MAY)
```

```
##
##    -2    -1     0     2     3     4     5     6     7     8
##  4546  5539 16947  2626   178    84    17     4    58     1
```

```
table(bank_data1$REPAY_APR)
```

```
##
##    -2    -1     0     2     3     4     5     6     7     8
##  4895  5740 16286  2766   184    49    13    19    46     2
```

```
library(ggplot2)
# plotting scatter plot
sepp <- ggplot(bank_data1, aes(x = PAID_AMT_SEP, y =BILL_AMT_SEP)) +
  ylab("Bill outstanding  (NT$)") + geom_point()
sepp + labs(x = "Last Payment (NT$)") +
  labs(title="Scatter plot between Limit balance  And Payment of September")
```



#Important variables

Creat dummy variables from REPAY_SEP to REPAY_APR

-2 and -1 we have consiered here that customerare paid on time other than that all 0 to 8, we ahvecatogarized as late payer

```
bank_data1$TIMELY_PD_SEP=ifelse(bank_data1$REPAY_SEP>0,0,1)
bank_data1$TIMELY_PD_AUG=ifelse(bank_data1$REPAY_AUG>0,0,1)
bank_data1$TIMELY_PD_JUL=ifelse(bank_data1$REPAY_JUL>0,0,1)
bank_data1$TIMELY_PD_JUN=ifelse(bank_data1$REPAY_JUN>0,0,1)
bank_data1$TIMELY_PD_MAY=ifelse(bank_data1$REPAY_MAY>0,0,1)
bank_data1$TIMELY_PD_APR=ifelse(bank_data1$REPAY_APR>0,0,1)
str(bank_data1)
```

```
## 'data.frame':    30000 obs. of  42 variables:
##  $ LIMIT_BAL    : num  20000 120000 90000 50000 50000 50000 500000 100000 140000 20
000 ...
##  $ SEX          : int  2 2 2 2 1 1 1 2 2 1 ...
##  $ EDUCATION    : int  2 2 2 2 2 1 1 2 3 3 ...
##  $ MARRIAGE     : int  1 2 2 1 1 2 2 2 1 2 ...
##  $ AGE          : int  24 26 34 37 57 37 29 23 28 35 ...
##  $ REPAY_SEP    : int  2 -1 0 0 -1 0 0 0 0 -2 ...
##  $ REPAY_AUG    : int  2 2 0 0 0 0 0 -1 0 -2 ...
##  $ REPAY_JUL    : int  -1 0 0 0 -1 0 0 -1 2 -2 ...
##  $ REPAY_JUN    : int  -1 0 0 0 0 0 0 0 0 -2 ...
##  $ REPAY_MAY    : int  -2 0 0 0 0 0 0 0 0 -1 ...
##  $ REPAY_APR    : int  -2 2 0 0 0 0 0 -1 0 -1 ...
##  $ BILL_AMT_SEP : num  3913 2682 29239 46990 8617 ...
##  $ BILL_AMT_AUG : num  3102 1725 14027 48233 5670 ...
##  $ BILL_AMT_JUL : num  689 2682 13559 49291 35835 ...
##  $ BILL_AMT_JUN : num  0 3272 14331 28314 20940 ...
##  $ BILL_AMT_MAY : num  0 3455 14948 28959 19146 ...
##  $ BILL_AMT_APR : num  0 3261 15549 29547 19131 ...
##  $ PAID_AMT_SEP : num  0 0 1518 2000 2000 ...
##  $ PAID_AMT_AUG : num  689 1000 1500 2019 19004 ...
##  $ PAID_AMT_JUL : num  0 1000 1000 1200 10000 ...
##  $ PAID_AMT_JUN : num  0 1000 1000 1100 9000 ...
##  $ PAID_AMT_MAY : num  0 0 1000 1069 689 ...
##  $ PAID_AMT_APR : num  0 2000 5000 1000 679 ...
##  $ DEFAULT      : int  1 1 0 0 0 0 0 0 0 0 ...
##  $ RATIO_SEP    : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ RATIO_AUG    : num  0 0 0 0 1 0 0 1 0 0 ...
##  $ RATIO_JUL    : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ RATIO_JUN    : num  0 0 0 0 0 0 0 1 0 1 ...
##  $ RATIO_MAY    : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ RATIO_APR    : num  0 0 0 0 0 0 0 1 0 0 ...
##  $ balance_SEP  : num  3913 2682 27721 44990 6617 ...
##  $ balance_AUG  : num  2413 725 12527 46214 -31011 ...
##  $ balance_JUL  : num  689 1682 12559 48091 25835 ...
##  $ balance_JUN  : num  0 2272 13331 27214 11940 ...
##  $ balance_MAY  : num  0 3455 13948 27890 18457 ...
##  $ balance_APR  : num  0 1261 10549 28547 18452 ...
##  $ TIMELY_PD_SEP: num  0 1 1 1 1 1 1 1 1 1 ...
##  $ TIMELY_PD_AUG: num  0 0 1 1 1 1 1 1 1 1 ...
##  $ TIMELY_PD_JUL: num  1 1 1 1 1 1 1 1 0 1 ...
##  $ TIMELY_PD_JUN: num  1 1 1 1 1 1 1 1 1 1 ...
##  $ TIMELY_PD_MAY: num  1 1 1 1 1 1 1 1 1 1 ...
##  $ TIMELY_PD_APR: num  1 0 1 1 1 1 1 1 1 1 ...
```

Checking the percentage of customers paid timely theri dues

```
table(bank_data1$TIMELY_PD_APR)
```

```
##
##     0     1
##  3079 26921
```

```
table(bank_data1$TIMELY_PD_MAY)
```

```
##
##     0     1
##  2968 27032
```

```
table(bank_data1$TIMELY_PD_JUN)
```

```
##
##     0     1
##  3510 26490
```

```
table(bank_data1$TIMELY_PD_JUL)
```

```
##
##     0     1
##  4213 25787
```

```
table(bank_data1$TIMELY_PD_AUG)
```

```
##
##     0     1
##  4438 25562
```

```
table(bank_data1$TIMELY_PD_SEP)
```

```
##
##     0     1
##  6818 23182
```

```
Per_Sep <- nrow(subset(bank_data1, TIMELY_PD_SEP == "0"))/nrow(bank_data1)
Per_Aug <- nrow(subset(bank_data1, TIMELY_PD_AUG == "0"))/nrow(bank_data1)
Per_Jul <- nrow(subset(bank_data1, TIMELY_PD_JUL == "0"))/nrow(bank_data1)
Per_Jun <- nrow(subset(bank_data1, TIMELY_PD_JUN == "0"))/nrow(bank_data1)
Per_May <- nrow(subset(bank_data1, TIMELY_PD_MAY == "0"))/nrow(bank_data1)
Per_Apr <- nrow(subset(bank_data1, TIMELY_PD_APR == "0"))/nrow(bank_data1)
Per_Sep
```

```
## [1] 0.2272667
```

```
Per_Aug
```

```
## [1] 0.1479333
```

```
Per_Jul
```

```
## [1] 0.1404333
```

```
Per_Jun
```

```
## [1] 0.117
```

```
Per_May
```

```
## [1] 0.09893333
```

```
Per_Apr
```

```
## [1] 0.1026333
```

we found in September22%,Auguest 14%, July 14%, June 11%, May 09 and April it is 10% customer are paid on time

Correlation between Variable

```
#Checking umeric values

names(bank_data1)
```

```
##  [1] "LIMIT_BAL"     "SEX"           "EDUCATION"     "MARRIAGE"
##  [5] "AGE"           "REPAY_SEP"     "REPAY_AUG"     "REPAY_JUL"
##  [9] "REPAY_JUN"     "REPAY_MAY"     "REPAY_APR"     "BILL_AMT_SEP"
## [13] "BILL_AMT_AUG"  "BILL_AMT_JUL"  "BILL_AMT_JUN"  "BILL_AMT_MAY"
## [17] "BILL_AMT_APR"  "PAID_AMT_SEP"  "PAID_AMT_AUG"  "PAID_AMT_JUL"
## [21] "PAID_AMT_JUN"  "PAID_AMT_MAY"  "PAID_AMT_APR"  "DEFAULT"
## [25] "RATIO_SEP"     "RATIO_AUG"     "RATIO_JUL"     "RATIO_JUN"
## [29] "RATIO_MAY"     "RATIO_APR"     "balance_SEP"   "balance_AUG"
## [33] "balance_JUL"   "balance_JUN"   "balance_MAY"   "balance_APR"
## [37] "TIMELY_PD_SEP" "TIMELY_PD_AUG" "TIMELY_PD_JUL" "TIMELY_PD_JUN"
## [41] "TIMELY_PD_MAY" "TIMELY_PD_APR"
```

```
bank_data_num <- subset(bank_data1,select= -c(2,3,4))
numeric.list <- sapply(bank_data_num,is.numeric)
numeric.list
```

```
##     LIMIT_BAL           AGE     REPAY_SEP     REPAY_AUG     REPAY_JUL
##          TRUE          TRUE          TRUE          TRUE          TRUE
##     REPAY_JUN     REPAY_MAY     REPAY_APR  BILL_AMT_SEP  BILL_AMT_AUG
##          TRUE          TRUE          TRUE          TRUE          TRUE
##  BILL_AMT_JUL  BILL_AMT_JUN  BILL_AMT_MAY  BILL_AMT_APR  PAID_AMT_SEP
##          TRUE          TRUE          TRUE          TRUE          TRUE
##  PAID_AMT_AUG  PAID_AMT_JUL  PAID_AMT_JUN  PAID_AMT_MAY  PAID_AMT_APR
##          TRUE          TRUE          TRUE          TRUE          TRUE
##       DEFAULT     RATIO_SEP     RATIO_AUG     RATIO_JUL     RATIO_JUN
##          TRUE          TRUE          TRUE          TRUE          TRUE
##     RATIO_MAY     RATIO_APR   balance_SEP   balance_AUG   balance_JUL
##          TRUE          TRUE          TRUE          TRUE          TRUE
##   balance_JUN   balance_MAY   balance_APR TIMELY_PD_SEP TIMELY_PD_AUG
##          TRUE          TRUE          TRUE          TRUE          TRUE
## TIMELY_PD_JUL TIMELY_PD_JUN TIMELY_PD_MAY TIMELY_PD_APR
##          TRUE          TRUE          TRUE          TRUE
```

```
sum(numeric.list)
```

```
## [1] 39
```

```
numeric.df1 <- bank_data_num[, numeric.list]
```

```
cor.mat <- cor(numeric.df1)
cor.mat
```

```
##                 LIMIT_BAL          AGE   REPAY_SEP   REPAY_AUG   REPAY_JUL
## LIMIT_BAL      1.00000000  0.144411208 -0.27539728 -0.30132429 -0.29085210
## AGE            0.14441121  1.000000000 -0.03944738 -0.05014778 -0.05304844
## REPAY_SEP     -0.27539728 -0.039447376  1.00000000  0.67216438  0.57424509
## REPAY_AUG     -0.30132429 -0.050147776  0.67216438  1.00000000  0.76655168
## REPAY_JUL     -0.29085210 -0.053048437  0.57424509  0.76655168  1.00000000
## REPAY_JUN     -0.27188031 -0.049721674  0.53884063  0.66206713  0.77735887
## REPAY_MAY     -0.25386748 -0.053825976  0.50942606  0.62278025  0.68677451
## REPAY_APR     -0.23914673 -0.048773426  0.47455309  0.57550086  0.63268359
## BILL_AMT_SEP   0.22574850  0.044197956  0.21169675  0.27123718  0.24327442
## BILL_AMT_AUG   0.21928462  0.042399024  0.21503351  0.27104464  0.27565894
## BILL_AMT_JUL   0.22750648  0.041973628  0.20737569  0.26239738  0.26793965
## BILL_AMT_JUN   0.23861623  0.037627195  0.20266794  0.25622637  0.26389145
## BILL_AMT_MAY   0.24164395  0.035040604  0.19997954  0.25153785  0.25757598
## BILL_AMT_APR   0.23546407  0.032943061  0.19718917  0.24936048  0.25353730
## PAID_AMT_SEP   0.34472260  0.042630410 -0.12484300 -0.10516756  0.03961255
## PAID_AMT_AUG   0.35155111  0.050324437 -0.12468940 -0.07193099 -0.10179544
## PAID_AMT_JUL   0.35154679  0.043465898 -0.11133798 -0.06310105 -0.05852685
## PAID_AMT_JUN   0.34010202  0.044050199 -0.08842719 -0.04586198 -0.03315491
## PAID_AMT_MAY   0.34316453  0.039833639 -0.08117923 -0.03531956 -0.02375237
## PAID_AMT_APR   0.35701221  0.035382567 -0.09170658 -0.04581591 -0.03972182
## DEFAULT       -0.15487161  0.013889834  0.32479373  0.26355120  0.23525251
## RATIO_SEP      0.17013688  0.033436180 -0.27133659 -0.30231144 -0.18319677
## RATIO_AUG      0.17174885  0.027530355 -0.27549031 -0.27338650 -0.29374069
## RATIO_JUL      0.16138355  0.015291893 -0.24294682 -0.23830100 -0.24051176
## RATIO_JUN      0.13896766  0.029853508 -0.22779673 -0.23147323 -0.23498005
## RATIO_MAY      0.07041878  0.011651845 -0.10768036 -0.10648743 -0.10779970
## RATIO_APR      0.10713783  0.025828607 -0.21659451 -0.21259679 -0.21114396
## balance_SEP    0.18863296  0.039584764  0.23322320  0.29261651  0.24747342
## balance_AUG    0.17931334  0.038081037  0.23749793  0.28911232  0.29593154
## balance_JUL    0.18651671  0.038051625  0.22706449  0.27786665  0.28247089
## balance_JUN    0.20133969  0.032883393  0.22024351  0.26959462  0.27695959
## balance_MAY    0.20146957  0.030734691  0.21873873  0.26544400  0.27012025
## balance_APR    0.18973546  0.029652051  0.21400990  0.26206908  0.26574909
## TIMELY_PD_SEP  0.15921009  0.003073199 -0.76562288 -0.42343527 -0.29675183
## TIMELY_PD_AUG  0.20228766  0.009968067 -0.54364452 -0.79951598 -0.53347863
## TIMELY_PD_JUL  0.19645811  0.015262820 -0.41315133 -0.54291333 -0.78736082
## TIMELY_PD_JUN  0.18695974  0.007899190 -0.38081781 -0.44045458 -0.53977177
## TIMELY_PD_MAY  0.17589366  0.016556216 -0.36097705 -0.41144520 -0.44338770
## TIMELY_PD_APR  0.17327390  0.020612667 -0.32951861 -0.37321749 -0.40496551
##                  REPAY_JUN   REPAY_MAY    REPAY_APR  BILL_AMT_SEP
## LIMIT_BAL     -0.271880313 -0.25386748 -0.239146726  2.257485e-01
## AGE           -0.049721674 -0.05382598 -0.048773426  4.419796e-02
## REPAY_SEP      0.538840627  0.50942606  0.474553086  2.116968e-01
## REPAY_AUG      0.662067131  0.62278025  0.575500862  2.712372e-01
## REPAY_JUL      0.777358873  0.68677451  0.632683593  2.432744e-01
## REPAY_JUN      1.000000000  0.81983531  0.716449482  2.382699e-01
## REPAY_MAY      0.819835311  1.00000000  0.816900160  2.400896e-01
```

```
## REPAY_APR      0.716449482  0.81690016  1.000000000  2.407940e-01
## BILL_AMT_SEP   0.238269941  0.24008957  0.240793973  1.000000e+00
## BILL_AMT_AUG   0.264887662  0.26413104  0.264006062  9.464939e-01
## BILL_AMT_JUL   0.289982280  0.28763824  0.283762534  8.988444e-01
## BILL_AMT_JUN   0.287197846  0.31721377  0.309800209  8.517570e-01
## BILL_AMT_MAY   0.280099508  0.31022107  0.334516643  8.190675e-01
## BILL_AMT_APR   0.276027978  0.30222883  0.327066503  7.930090e-01
## PAID_AMT_SEP   0.013794760  0.01322631  0.020143779  3.834999e-01
## PAID_AMT_AUG   0.043295545  0.02931824  0.018074934  3.413154e-01
## PAID_AMT_JUL  -0.078448789  0.05826027  0.042543946  3.198864e-01
## PAID_AMT_JUN  -0.025814044 -0.04034934  0.080924968  3.355537e-01
## PAID_AMT_MAY  -0.007393487  0.00146192 -0.020659098  3.305207e-01
## PAID_AMT_APR  -0.021266061 -0.00471918  0.007348771  3.121608e-01
## DEFAULT        0.216613637  0.20414891  0.186866362 -2.814973e-02
## RATIO_SEP     -0.212483389 -0.21113715 -0.210955276 -2.651461e-01
## RATIO_AUG     -0.185284356 -0.20154998 -0.206663748 -2.418946e-01
## RATIO_JUL     -0.268930470 -0.16836533 -0.185488444 -2.138381e-01
## RATIO_JUN     -0.239609132 -0.25741487 -0.159466469 -1.978175e-01
## RATIO_MAY     -0.107523848 -0.14849786 -0.089690212 -7.965437e-02
## RATIO_APR     -0.213916829 -0.24676769 -0.165679338 -1.912840e-01
## balance_SEP    0.244368739  0.24585107  0.246105673  9.829857e-01
## balance_AUG    0.267579018  0.26743677  0.268899915  9.316748e-01
## balance_JUL    0.308161389  0.28923989  0.286952430  8.827723e-01
## balance_JUN    0.300018629  0.33293165  0.310949495  8.349008e-01
## balance_MAY    0.291352835  0.32093794  0.350326140  8.013733e-01
## balance_APR    0.286372817  0.31107739  0.334783721  7.696848e-01
## TIMELY_PD_SEP -0.259522794 -0.24244755 -0.221528224  6.326166e-02
## TIMELY_PD_AUG -0.429263308 -0.40063275 -0.365837679 -1.209009e-02
## TIMELY_PD_JUL -0.519510913 -0.42371136 -0.384725641  1.709903e-02
## TIMELY_PD_JUN -0.757320032 -0.53157511 -0.434655357  1.235628e-02
## TIMELY_PD_MAY -0.552610424 -0.73252120 -0.530848179  5.796195e-04
## TIMELY_PD_APR -0.447518324 -0.53617698 -0.734811807 -7.790246e-05
##               BILL_AMT_AUG BILL_AMT_JUL BILL_AMT_JUN BILL_AMT_MAY
## LIMIT_BAL      0.219284616   0.22750648   0.23861623   0.24164395
## AGE            0.042399024   0.04197363   0.03762719   0.03504060
## REPAY_SEP      0.215033512   0.20737569   0.20266794   0.19997954
## REPAY_AUG      0.271044643   0.26239738   0.25622637   0.25153785
## REPAY_JUL      0.275658942   0.26793965   0.26389145   0.25757598
## REPAY_JUN      0.264887662   0.28998228   0.28719785   0.28009951
## REPAY_MAY      0.264131042   0.28763824   0.31721377   0.31022107
## REPAY_APR      0.264006062   0.28376253   0.30980021   0.33451664
## BILL_AMT_SEP   0.946493906   0.89884436   0.85175700   0.81906751
## BILL_AMT_AUG   1.000000000   0.93867414   0.88704848   0.84930320
## BILL_AMT_JUL   0.938674143   1.00000000   0.93047351   0.88564143
## BILL_AMT_JUN   0.887048480   0.93047351   1.00000000   0.93603470
## BILL_AMT_MAY   0.849303197   0.88564143   0.93603470   1.00000000
## BILL_AMT_APR   0.820514292   0.85169140   0.89641600   0.94392171
## PAID_AMT_SEP   0.468478029   0.43426181   0.41388446   0.39492124
## PAID_AMT_AUG   0.351474735   0.44508575   0.40911219   0.38322288
```

```
## PAID_AMT_JUL   0.327217570   0.33625027   0.43489337   0.39365285
## PAID_AMT_JUN   0.339482536   0.35041040   0.36097186   0.44884284
## PAID_AMT_MAY   0.340047122   0.35350936   0.36421171   0.36807416
## PAID_AMT_APR   0.316717024   0.33128864   0.33916249   0.34296527
## DEFAULT       -0.021723287  -0.02032750  -0.01636235  -0.01388112
## RATIO_SEP     -0.182940008  -0.18606840  -0.17594266  -0.17464959
## RATIO_AUG     -0.267373326  -0.17500257  -0.17928025  -0.17638374
## RATIO_JUL     -0.225673075  -0.25214778  -0.15575877  -0.16466294
## RATIO_JUN     -0.210049070  -0.21774596  -0.24343448  -0.14927365
## RATIO_MAY     -0.086221937  -0.08883390  -0.10337786  -0.07914399
## RATIO_APR     -0.199394553  -0.20128598  -0.21133533  -0.17267815
## balance_SEP    0.910737084   0.86959437   0.82344593   0.79382779
## balance_AUG    0.982600033   0.90326440   0.85756999   0.82311488
## balance_JUL    0.922485864   0.98167162   0.89312953   0.85632752
## balance_JUN    0.871773905   0.91427476   0.98187199   0.90253251
## balance_MAY    0.831076074   0.86538377   0.91496456   0.97903387
## balance_APR    0.798042792   0.82733590   0.87109056   0.91915086
## TIMELY_PD_SEP  0.049986593   0.04105088   0.03156677   0.02353475
## TIMELY_PD_AUG -0.016100950  -0.02609613  -0.03478733  -0.04198107
## TIMELY_PD_JUL -0.005309878  -0.00919752  -0.02525567  -0.03411177
## TIMELY_PD_JUN -0.003318187  -0.02144914  -0.03458244  -0.04836318
## TIMELY_PD_MAY -0.012438353  -0.02736164  -0.05241539  -0.06393935
## TIMELY_PD_APR -0.012980333  -0.02633602  -0.04799944  -0.07125941
##               BILL_AMT_APR PAID_AMT_SEP PAID_AMT_AUG PAID_AMT_JUL
## LIMIT_BAL       0.23546407   0.34472260   0.35155111   0.35154679
## AGE             0.03294306   0.04263041   0.05032444   0.04346590
## REPAY_SEP       0.19718917  -0.12484300  -0.12468940  -0.11133798
## REPAY_AUG       0.24936048  -0.10516756  -0.07193099  -0.06310105
## REPAY_JUL       0.25353730   0.03961255  -0.10179544  -0.05852685
## REPAY_JUN       0.27602798   0.01379476   0.04329554  -0.07844879
## REPAY_MAY       0.30222883   0.01322631   0.02931824   0.05826027
## REPAY_APR       0.32706650   0.02014378   0.01807493   0.04254395
## BILL_AMT_SEP    0.79300899   0.38349991   0.34131544   0.31988643
## BILL_AMT_AUG    0.82051429   0.46847803   0.35147474   0.32721757
## BILL_AMT_JUL    0.85169140   0.43426181   0.44508575   0.33625027
## BILL_AMT_JUN    0.89641600   0.41388446   0.40911219   0.43489337
## BILL_AMT_MAY    0.94392171   0.39492124   0.38322288   0.39365285
## BILL_AMT_APR    1.00000000   0.37405952   0.36165151   0.37493209
## PAID_AMT_SEP    0.37405952   1.00000000   0.49183698   0.45954130
## PAID_AMT_AUG    0.36165151   0.49183698   1.00000000   0.46780328
## PAID_AMT_JUL    0.37493209   0.45954130   0.46780328   1.00000000
## PAID_AMT_JUN    0.41044422   0.43651346   0.44472487   0.46276419
## PAID_AMT_MAY    0.45242774   0.43100306   0.43118614   0.45087100
## PAID_AMT_APR    0.33548428   0.41310351   0.42140998   0.43565445
## DEFAULT        -0.01040406  -0.12745405  -0.12966646  -0.11305607
## RATIO_SEP      -0.17491029   0.32347950   0.10842229   0.10237912
## RATIO_AUG      -0.17848914  -0.04576346   0.34781141   0.10361116
## RATIO_JUL      -0.16375576   0.02914445  -0.04090644   0.37682163
## RATIO_JUN      -0.16314901   0.02292415   0.01921562  -0.04807343
```

```
## RATIO_MAY      -0.07273748   0.03928512   0.03895988   0.04167422
## RATIO_APR      -0.24001114   0.04529049   0.05321766   0.05318352
## balance_SEP     0.77112130   0.26548174   0.28824504   0.27244967
## balance_AUG     0.79762194   0.41329519   0.23088484   0.27858519
## balance_JUL     0.82714221   0.38802551   0.39351355   0.21415716
## balance_JUN     0.87055604   0.37127391   0.36345339   0.38708605
## balance_MAY     0.90619776   0.34978538   0.33679100   0.34710418
## balance_APR     0.97700844   0.32474088   0.31168270   0.32654135
## TIMELY_PD_SEP   0.02043208   0.20228245   0.15917116   0.13824251
## TIMELY_PD_AUG  -0.04533688   0.19666442   0.12073611   0.10116083
## TIMELY_PD_JUL  -0.03736249   0.04891961   0.19153266   0.11014202
## TIMELY_PD_JUN  -0.05633509   0.08136572   0.05606436   0.15725552
## TIMELY_PD_MAY  -0.07233785   0.08778706   0.07641425   0.04343556
## TIMELY_PD_APR  -0.07519381   0.07360718   0.07879006   0.05991920
##               PAID_AMT_JUN PAID_AMT_MAY PAID_AMT_APR      DEFAULT
## LIMIT_BAL       0.34010202  0.343164533  0.357012214 -0.154871612
## AGE             0.04405020  0.039833639  0.035382567  0.013889834
## REPAY_SEP      -0.08842719 -0.081179226 -0.091706581  0.324793728
## REPAY_AUG      -0.04586198 -0.035319556 -0.045815908  0.263551202
## REPAY_JUL      -0.03315491 -0.023752366 -0.039721816  0.235252514
## REPAY_JUN      -0.02581404 -0.007393487 -0.021266061  0.216613637
## REPAY_MAY      -0.04034934  0.001461920 -0.004719180  0.204148914
## REPAY_APR       0.08092497 -0.020659098  0.007348771  0.186866362
## BILL_AMT_SEP    0.33555369  0.330520667  0.312160768 -0.028149729
## BILL_AMT_AUG    0.33948254  0.340047122  0.316717024 -0.021723287
## BILL_AMT_JUL    0.35041040  0.353509362  0.331288637 -0.020327498
## BILL_AMT_JUN    0.36097186  0.364211711  0.339162495 -0.016362354
## BILL_AMT_MAY    0.44884284  0.368074164  0.342965266 -0.013881118
## BILL_AMT_APR    0.41044422  0.452427743  0.335484278 -0.010404064
## PAID_AMT_SEP    0.43651346  0.431003065  0.413103509 -0.127454051
## PAID_AMT_AUG    0.44472487  0.431186142  0.421409979 -0.129666463
## PAID_AMT_JUL    0.46276419  0.450870997  0.435654446 -0.113056071
## PAID_AMT_JUN    1.00000000  0.474034682  0.449923096 -0.102892735
## PAID_AMT_MAY    0.47403468  1.000000000  0.477744166 -0.102013983
## PAID_AMT_APR    0.44992310  0.477744166  1.000000000 -0.105654753
## DEFAULT        -0.10289274 -0.102013983 -0.105654753  1.000000000
## RATIO_SEP       0.06446055  0.067084558  0.070884202 -0.068646633
## RATIO_AUG       0.07731871  0.066132090  0.079786768 -0.074470210
## RATIO_JUL       0.08560472  0.078354500  0.072281342 -0.079113555
## RATIO_JUN       0.33503403  0.068170763  0.072204428 -0.063494323
## RATIO_MAY       0.17953527  0.070945958  0.057443651 -0.041880561
## RATIO_APR       0.28291331 -0.083522152  0.047428424 -0.071744303
## balance_SEP     0.29228390  0.287637378  0.269239018 -0.012262235
## balance_AUG     0.29429610  0.296699738  0.271733317 -0.005010962
## balance_JUL     0.30253850  0.307315217  0.283736658 -0.005809474
## balance_JUN     0.24241953  0.314799952  0.292058510 -0.004326666
## balance_MAY     0.40015487  0.244349847  0.290201237 -0.000316773
## balance_APR     0.36181742  0.397909698  0.202999860  0.002000906
## TIMELY_PD_SEP   0.12799792  0.122184613  0.127541281 -0.368116996
```

```
## TIMELY_PD_AUG   0.09269627  0.084497695  0.089446466 -0.338022325
## TIMELY_PD_JUL   0.08834239  0.084769838  0.094339069 -0.293583004
## TIMELY_PD_JUN   0.09344797  0.080055375  0.088959256 -0.275494664
## TIMELY_PD_MAY   0.12530591  0.077100169  0.081013660 -0.266958955
## TIMELY_PD_APR   0.01556949  0.124737381  0.074600218 -0.246088912
##                 RATIO_SEP   RATIO_AUG   RATIO_JUL   RATIO_JUN   RATIO_MAY
## LIMIT_BAL      0.17013688  0.17174885  0.16138355  0.13896766  0.07041878
## AGE            0.03343618  0.02753035  0.01529189  0.02985351  0.01165185
## REPAY_SEP     -0.27133659 -0.27549031 -0.24294682 -0.22779673 -0.10768036
## REPAY_AUG     -0.30231144 -0.27338650 -0.23830100 -0.23147323 -0.10648743
## REPAY_JUL     -0.18319677 -0.29374069 -0.24051176 -0.23498005 -0.10779970
## REPAY_JUN     -0.21248339 -0.18528436 -0.26893047 -0.23960913 -0.10752385
## REPAY_MAY     -0.21113715 -0.20154998 -0.16836533 -0.25741487 -0.14849786
## REPAY_APR     -0.21095528 -0.20666375 -0.18548844 -0.15946647 -0.08969021
## BILL_AMT_SEP  -0.26514609 -0.24189455 -0.21383812 -0.19781747 -0.07965437
## BILL_AMT_AUG  -0.18294001 -0.26737333 -0.22567308 -0.21004907 -0.08622194
## BILL_AMT_JUL  -0.18606840 -0.17500257 -0.25214778 -0.21774596 -0.08883390
## BILL_AMT_JUN  -0.17594266 -0.17928025 -0.15575877 -0.24343448 -0.10337786
## BILL_AMT_MAY  -0.17464959 -0.17638374 -0.16466294 -0.14927365 -0.07914399
## BILL_AMT_APR  -0.17491029 -0.17848914 -0.16375576 -0.16314901 -0.07273748
## PAID_AMT_SEP   0.32347950 -0.04576346  0.02914445  0.02292415  0.03928512
## PAID_AMT_AUG   0.10842229  0.34781141 -0.04090644  0.01921562  0.03895988
## PAID_AMT_JUL   0.10237912  0.10361116  0.37682163 -0.04807343  0.04167422
## PAID_AMT_JUN   0.06446055  0.07731871  0.08560472  0.33503403  0.17953527
## PAID_AMT_MAY   0.06708456  0.06613209  0.07835450  0.06817076  0.07094596
## PAID_AMT_APR   0.07088420  0.07978677  0.07228134  0.07220443  0.05744365
## DEFAULT       -0.06864663 -0.07447021 -0.07911355 -0.06349432 -0.04188056
## RATIO_SEP      1.00000000  0.16426179  0.23187564  0.19560123  0.09157725
## RATIO_AUG      0.16426179  1.00000000  0.15258243  0.21378463  0.09396729
## RATIO_JUL      0.23187564  0.15258243  1.00000000  0.12778559  0.09008316
## RATIO_JUN      0.19560123  0.21378463  0.12778559  1.00000000  0.19417478
## RATIO_MAY      0.09157725  0.09396729  0.09008316  0.19417478  1.00000000
## RATIO_APR      0.19484283  0.21036767  0.18487752  0.51709987  0.27390498
## balance_SEP   -0.32851497 -0.24087456 -0.22286600 -0.20680206 -0.08945092
## balance_AUG   -0.20753379 -0.33240787 -0.22503166 -0.21702722 -0.09386626
## balance_JUL   -0.20497116 -0.19886523 -0.32286731 -0.21401356 -0.09703492
## balance_JUN   -0.19124513 -0.19777880 -0.17650497 -0.30802110 -0.13837634
## balance_MAY   -0.19143096 -0.19354789 -0.18268254 -0.16459314 -0.09172357
## balance_APR   -0.19121364 -0.19539494 -0.17958327 -0.17855459 -0.08296204
## TIMELY_PD_SEP  0.07414362  0.09145351  0.07951001  0.07735065  0.04462044
## TIMELY_PD_AUG  0.14909760  0.09625866  0.09072825  0.09198303  0.04695335
## TIMELY_PD_JUL  0.05846244  0.14677139  0.07470394  0.08629354  0.04501988
## TIMELY_PD_JUN  0.09106354  0.08445320  0.12753058  0.08070484  0.03506964
## TIMELY_PD_MAY  0.09809638  0.09689908  0.08539828  0.11009603  0.05868736
## TIMELY_PD_APR  0.09306310  0.09373738  0.08952012  0.06789340  0.03197300
##                 RATIO_APR  balance_SEP  balance_AUG  balance_JUL
## LIMIT_BAL      0.10713783  0.188632956  0.179313344  0.186516714
## AGE            0.02582861  0.039584764  0.038081037  0.038051625
## REPAY_SEP     -0.21659451  0.233223202  0.237497925  0.227064488
```

```
## REPAY_AUG     -0.21259679  0.292616508  0.289112324  0.277866646
## REPAY_JUL     -0.21114396  0.247473418  0.295931537  0.282470887
## REPAY_JUN     -0.21391683  0.244368739  0.267579018  0.308161389
## REPAY_MAY     -0.24676769  0.245851073  0.267436766  0.289239887
## REPAY_APR     -0.16567934  0.246105673  0.268899915  0.286952430
## BILL_AMT_SEP  -0.19128404  0.982985681  0.931674805  0.882772346
## BILL_AMT_AUG  -0.19939455  0.910737084  0.982600033  0.922485864
## BILL_AMT_JUL  -0.20128598  0.869594373  0.903264398  0.981671619
## BILL_AMT_JUN  -0.21133533  0.823445930  0.857569987  0.893129526
## BILL_AMT_MAY  -0.17267815  0.793827793  0.823114884  0.856327521
## BILL_AMT_APR  -0.24001114  0.771121295  0.797621936  0.827142214
## PAID_AMT_SEP   0.04529049  0.265481743  0.413295191  0.388025513
## PAID_AMT_AUG   0.05321766  0.288245042  0.230884841  0.393513548
## PAID_AMT_JUL   0.05318352  0.272449674  0.278585187  0.214157165
## PAID_AMT_JUN   0.28291331  0.292283904  0.294296105  0.302538495
## PAID_AMT_MAY  -0.08352215  0.287637378  0.296699738  0.307315217
## PAID_AMT_APR   0.04742842  0.269239018  0.271733317  0.283736658
## DEFAULT       -0.07174430 -0.012262235 -0.005010962 -0.005809474
## RATIO_SEP      0.19484283 -0.328514970 -0.207533786 -0.204971163
## RATIO_AUG      0.21036767 -0.240874559 -0.332407866 -0.198865235
## RATIO_JUL      0.18487752 -0.222866005 -0.225031662 -0.322867312
## RATIO_JUN      0.51709987 -0.206802059 -0.217027223 -0.214013558
## RATIO_MAY      0.27390498 -0.089450921 -0.093866264 -0.097034923
## RATIO_APR      1.00000000 -0.203458612 -0.211308750 -0.213464434
## balance_SEP   -0.20345861  1.000000000  0.905501078  0.860485908
## balance_AUG   -0.21130875  0.905501078  1.000000000  0.895522728
## balance_JUL   -0.21346443  0.860485908  0.895522728  1.000000000
## balance_JUN   -0.26274193  0.813521394  0.849299964  0.885126588
## balance_MAY   -0.16603559  0.783262007  0.811843057  0.843325604
## balance_APR   -0.25367998  0.754704888  0.782494739  0.810332388
## TIMELY_PD_SEP  0.08539791  0.042350580  0.031100511  0.023867637
## TIMELY_PD_AUG  0.09416249 -0.034901234 -0.032439019 -0.040489083
## TIMELY_PD_JUL  0.09064253  0.008255684 -0.028078790 -0.023218988
## TIMELY_PD_JUN  0.09561139  0.001047892 -0.012850703 -0.040607646
## TIMELY_PD_MAY  0.12553147 -0.010795433 -0.023549771 -0.036193790
## TIMELY_PD_APR  0.06723022 -0.010149039 -0.024113268 -0.036032296
##                balance_JUN  balance_MAY  balance_APR TIMELY_PD_SEP
## LIMIT_BAL      0.201339693  0.201469566  0.189735458   0.159210087
## AGE            0.032883393  0.030734691  0.029652051   0.003073199
## REPAY_SEP      0.220243515  0.218738727  0.214009900  -0.765622881
## REPAY_AUG      0.269594616  0.265443998  0.262069079  -0.423435268
## REPAY_JUL      0.276959593  0.270120248  0.265749088  -0.296751825
## REPAY_JUN      0.300018629  0.291352835  0.286372817  -0.259522794
## REPAY_MAY      0.332931646  0.320937944  0.311077386  -0.242447549
## REPAY_APR      0.310949495  0.350326140  0.334783721  -0.221528224
## BILL_AMT_SEP   0.834900833  0.801373312  0.769684811   0.063261655
## BILL_AMT_AUG   0.871773905  0.831076074  0.798042792   0.049986593
## BILL_AMT_JUL   0.914274761  0.865383767  0.827335904   0.041050885
## BILL_AMT_JUN   0.981871989  0.914964557  0.871090561   0.031566769
```

```
## BILL_AMT_MAY   0.902532508  0.979033867  0.919150859   0.023534752
## BILL_AMT_APR   0.870556039  0.906197757  0.977008443   0.020432076
## PAID_AMT_SEP   0.371273911  0.349785380  0.324740876   0.202282446
## PAID_AMT_AUG   0.363453394  0.336791004  0.311682698   0.159171164
## PAID_AMT_JUL   0.387086052  0.347104185  0.326541348   0.138242507
## PAID_AMT_JUN   0.242419533  0.400154865  0.361817420   0.127997924
## PAID_AMT_MAY   0.314799952  0.244349847  0.397909698   0.122184613
## PAID_AMT_APR   0.292058510  0.290201237  0.202999860   0.127541281
## DEFAULT       -0.004326666 -0.000316773  0.002000906  -0.368116996
## RATIO_SEP     -0.191245132 -0.191430961 -0.191213644   0.074143617
## RATIO_AUG     -0.197778802 -0.193547895 -0.195394939   0.091453506
## RATIO_JUL     -0.176504971 -0.182682538 -0.179583266   0.079510006
## RATIO_JUN     -0.308021098 -0.164593139 -0.178554594   0.077350648
## RATIO_MAY     -0.138376341 -0.091723574 -0.082962039   0.044620436
## RATIO_APR     -0.262741930 -0.166035590 -0.253679984   0.085397909
## balance_SEP    0.813521394  0.783262007  0.754704888   0.042350580
## balance_AUG    0.849299964  0.811843057  0.782494739   0.031100511
## balance_JUL    0.885126588  0.843325604  0.810332388   0.023867637
## balance_JUN    1.000000000  0.890808375  0.853357164   0.016879958
## balance_MAY    0.890808375  1.000000000  0.892324948   0.007576236
## balance_APR    0.853357164  0.892324948  1.000000000   0.005369907
## TIMELY_PD_SEP  0.016879958  0.007576236  0.005369907   1.000000000
## TIMELY_PD_AUG -0.046665721 -0.054256351 -0.056938304   0.668175817
## TIMELY_PD_JUL -0.037384255 -0.046173663 -0.049948700   0.430527007
## TIMELY_PD_JUN -0.047592928 -0.061058867 -0.068687831   0.371030465
## TIMELY_PD_MAY -0.068534628 -0.076165754 -0.084544434   0.348586753
## TIMELY_PD_APR -0.054088112 -0.089079844 -0.086055344   0.309866549
##               TIMELY_PD_AUG TIMELY_PD_JUL TIMELY_PD_JUN TIMELY_PD_MAY
## LIMIT_BAL      0.202287660   0.196458106   0.186959736  0.1758936622
## AGE            0.009968067   0.015262820   0.007899190  0.0165562164
## REPAY_SEP     -0.543644516  -0.413151326  -0.380817805 -0.3609770469
## REPAY_AUG     -0.799515983  -0.542913332  -0.440454576 -0.4114452007
## REPAY_JUL     -0.533478628  -0.787360818  -0.539771775 -0.4433877046
## REPAY_JUN     -0.429263308  -0.519510913  -0.757320032 -0.5526104236
## REPAY_MAY     -0.400632748  -0.423711357  -0.531575109 -0.7325212043
## REPAY_APR     -0.365837679  -0.384725641  -0.434655357 -0.5308481793
## BILL_AMT_SEP  -0.012090086   0.017099033   0.012356283  0.0005796195
## BILL_AMT_AUG  -0.016100950  -0.005309878  -0.003318187 -0.0124383530
## BILL_AMT_JUL  -0.026096132  -0.009197520  -0.021449145 -0.0273616413
## BILL_AMT_JUN  -0.034787329  -0.025255667  -0.034582445 -0.0524153950
## BILL_AMT_MAY  -0.041981071  -0.034111766  -0.048363175 -0.0639393540
## BILL_AMT_APR  -0.045336876  -0.037362487  -0.056335088 -0.0723378503
## PAID_AMT_SEP   0.196664423   0.048919610   0.081365716  0.0877870606
## PAID_AMT_AUG   0.120736112   0.191532657   0.056064365  0.0764142506
## PAID_AMT_JUL   0.101160830   0.110142024   0.157255518  0.0434355606
## PAID_AMT_JUN   0.092696271   0.088342394   0.093447967  0.1253059102
## PAID_AMT_MAY   0.084497695   0.084769838   0.080055375  0.0771001690
## PAID_AMT_APR   0.089446466   0.094339069   0.088959256  0.0810136598
## DEFAULT       -0.338022325  -0.293583004  -0.275494664 -0.2669589548
```

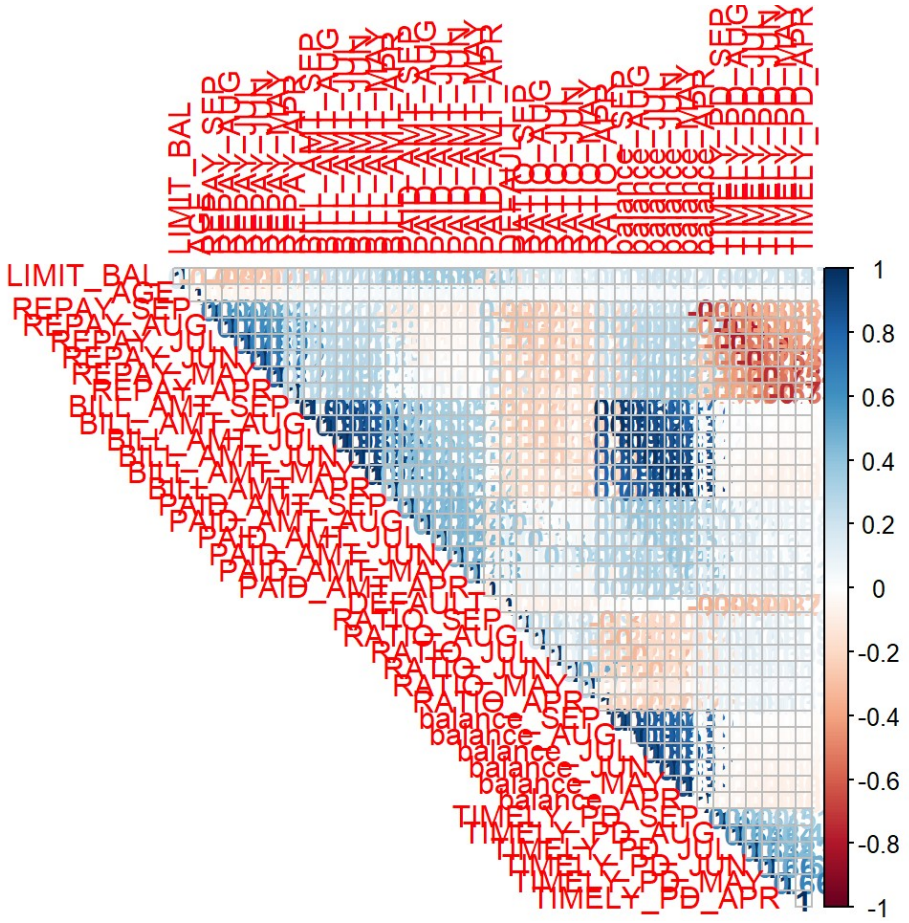
```
## RATIO_SEP       0.149097603   0.058462440   0.091063536  0.0980963770
## RATIO_AUG       0.096258655   0.146771385   0.084453199  0.0968990803
## RATIO_JUL       0.090728247   0.074703943   0.127530575  0.0853982782
## RATIO_JUN       0.091983026   0.086293539   0.080704840  0.1100960299
## RATIO_MAY       0.046953352   0.045019877   0.035069640  0.0586873582
## RATIO_APR       0.094162488   0.090642530   0.095611389  0.1255314676
## balance_SEP    -0.034901234   0.008255684   0.001047892 -0.0107954332
## balance_AUG    -0.032439019  -0.028078790  -0.012850703 -0.0235497711
## balance_JUL    -0.040489083  -0.023218988  -0.040607646 -0.0361937896
## balance_JUN    -0.046665721  -0.037384255  -0.047592928 -0.0685346276
## balance_MAY    -0.054256351  -0.046173663  -0.061058867 -0.0761657544
## balance_APR    -0.056938304  -0.049948700  -0.068687831 -0.0845444337
## TIMELY_PD_SEP   0.668175817   0.430527007   0.371030465  0.3485867526
## TIMELY_PD_AUG   1.000000000   0.625247223   0.478976661  0.4430469603
## TIMELY_PD_JUL   0.625247223   1.000000000   0.625659761  0.4823817389
## TIMELY_PD_JUN   0.478976661   0.625659761   1.000000000  0.6626371432
## TIMELY_PD_MAY   0.443046960   0.482381739   0.662637143  1.0000000000
## TIMELY_PD_APR   0.397391292   0.434246955   0.496442845  0.6626834794
##               TIMELY_PD_APR
## LIMIT_BAL      1.732739e-01
## AGE            2.061267e-02
## REPAY_SEP     -3.295186e-01
## REPAY_AUG     -3.732175e-01
## REPAY_JUL     -4.049655e-01
## REPAY_JUN     -4.475183e-01
## REPAY_MAY     -5.361770e-01
## REPAY_APR     -7.348118e-01
## BILL_AMT_SEP  -7.790246e-05
## BILL_AMT_AUG  -1.298033e-02
## BILL_AMT_JUL  -2.633602e-02
## BILL_AMT_JUN  -4.799944e-02
## BILL_AMT_MAY  -7.125941e-02
## BILL_AMT_APR  -7.519381e-02
## PAID_AMT_SEP   7.360718e-02
## PAID_AMT_AUG   7.879006e-02
## PAID_AMT_JUL   5.991920e-02
## PAID_AMT_JUN   1.556949e-02
## PAID_AMT_MAY   1.247374e-01
## PAID_AMT_APR   7.460022e-02
## DEFAULT       -2.460889e-01
## RATIO_SEP      9.306310e-02
## RATIO_AUG      9.373738e-02
## RATIO_JUL      8.952012e-02
## RATIO_JUN      6.789340e-02
## RATIO_MAY      3.197300e-02
## RATIO_APR      6.723022e-02
## balance_SEP   -1.014904e-02
## balance_AUG   -2.411327e-02
## balance_JUL   -3.603230e-02
```

```
## balance_JUN   -5.408811e-02
## balance_MAY   -8.907984e-02
## balance_APR   -8.605534e-02
## TIMELY_PD_SEP  3.098665e-01
## TIMELY_PD_AUG  3.973913e-01
## TIMELY_PD_JUL  4.342470e-01
## TIMELY_PD_JUN  4.964428e-01
## TIMELY_PD_MAY  6.626835e-01
## TIMELY_PD_APR  1.000000e+00
```

```
library(corrplot)
```

```
## corrplot 0.84 loaded
```

```
taiwan_bank1_matrix = cor(bank_data_num)
corrplot(taiwan_bank1_matrix, type="upper", method="number")
```



KMO Test

```
describe(bank_data1)
```

```
##                vars     n      mean        sd   median   trimmed       mad
## LIMIT_BAL         1 30000 166438.66 126722.92 140000.0 151607.40 133434.00
## SEX               2 30000      1.60      0.49      2.0      1.63      0.00
## EDUCATION         3 30000      1.85      0.79      2.0      1.78      1.48
## MARRIAGE          4 30000      1.55      0.52      2.0      1.55      0.00
## AGE               5 30000     35.49      9.22     34.0     34.69      8.90
## REPAY_SEP         6 30000     -0.02      1.12      0.0     -0.06      1.48
## REPAY_AUG         7 30000     -0.13      1.20      0.0     -0.20      0.00
## REPAY_JUL         8 30000     -0.17      1.20      0.0     -0.24      0.00
## REPAY_JUN         9 30000     -0.22      1.17      0.0     -0.31      0.00
## REPAY_MAY        10 30000     -0.27      1.13      0.0     -0.36      0.00
## REPAY_APR        11 30000     -0.29      1.15      0.0     -0.39      0.00
## BILL_AMT_SEP     12 30000  47402.10  58937.52  22381.5  35359.66  32321.42
## BILL_AMT_AUG     13 30000  45528.33  57005.11  21200.0  33836.10  30852.91
## BILL_AMT_JUL     14 30000  43534.93  55081.75  20088.5  32064.43  29219.82
## BILL_AMT_JUN     15 30000  40185.58  51604.88  19052.0  29212.37  27659.39
## BILL_AMT_MAY     16 30000  37657.51  49383.59  18104.5  26920.95  26224.97
## BILL_AMT_APR     17 30000  36262.58  48281.79  17071.0  25726.08  24840.96
## PAID_AMT_SEP     18 30000   4175.04   5193.02   2100.0   2997.21   2864.38
## PAID_AMT_AUG     19 30000   4123.77   5330.17   2009.0   2876.43   2951.86
## PAID_AMT_JUL     20 30000   3634.39   4918.11   1800.0   2468.91   2661.27
## PAID_AMT_JUN     21 30000   3359.59   4686.97   1500.0   2199.23   2223.90
## PAID_AMT_MAY     22 30000   3350.00   4659.06   1500.0   2202.19   2223.90
## PAID_AMT_APR     23 30000   3455.95   5037.28   1500.0   2165.33   2223.90
## DEFAULT          24 30000      0.22      0.42      0.0      0.15      0.00
## RATIO_SEP        25 30000      0.12      0.33      0.0      0.03      0.00
## RATIO_AUG        26 30000      0.13      0.33      0.0      0.03      0.00
## RATIO_JUL        27 30000      0.12      0.32      0.0      0.02      0.00
## RATIO_JUN        28 30000      0.11      0.32      0.0      0.02      0.00
## RATIO_MAY        29 30000      0.03      0.18      0.0      0.00      0.00
## RATIO_APR        30 30000      0.13      0.34      0.0      0.04      0.00
## balance_SEP      31 30000  42407.17  57907.66  18550.5  31661.62  28042.64
## balance_AUG      32 30000  40712.70  55965.98  18102.5  30247.54  27328.02
## balance_JUL      33 30000  39106.07  54197.24  17769.0  28875.89  26348.77
## balance_JUN      34 30000  36240.28  50806.44  16970.0  26314.16  25159.72
## balance_MAY      35 30000  33546.52  48076.22  15538.0  23994.38  23036.64
## balance_APR      36 30000  32103.70  47142.10  13926.5  22794.62  20647.43
## TIMELY_PD_SEP    37 30000      0.77      0.42      1.0      0.84      0.00
## TIMELY_PD_AUG    38 30000      0.85      0.36      1.0      0.94      0.00
## TIMELY_PD_JUL    39 30000      0.86      0.35      1.0      0.95      0.00
## TIMELY_PD_JUN    40 30000      0.88      0.32      1.0      0.98      0.00
## TIMELY_PD_MAY    41 30000      0.90      0.30      1.0      1.00      0.00
## TIMELY_PD_APR    42 30000      0.90      0.30      1.0      1.00      0.00
##                   min       max     range  skew kurtosis     se
## LIMIT_BAL       10000 520000.00 510000.00  0.87    -0.03 731.64
## SEX                 1      2.00      1.00 -0.42    -1.82   0.00
## EDUCATION           0      6.00      6.00  0.97     2.08   0.00
## MARRIAGE            0      3.00      3.00 -0.02    -1.36   0.00
```

```
## AGE               21     79.00     58.00  0.73     0.04   0.05
## REPAY_SEP         -2      8.00     10.00  0.73     2.72   0.01
## REPAY_AUG         -2      8.00     10.00  0.79     1.57   0.01
## REPAY_JUL         -2      8.00     10.00  0.84     2.08   0.01
## REPAY_JUN         -2      8.00     10.00  1.00     3.50   0.01
## REPAY_MAY         -2      8.00     10.00  1.01     3.99   0.01
## REPAY_APR         -2      8.00     10.00  0.95     3.43   0.01
## BILL_AMT_SEP  -15308 201203.05 216511.05  1.49     1.23 340.28
## BILL_AMT_AUG  -69777 194792.20 264569.20  1.50     1.27 329.12
## BILL_AMT_JUL  -61506 187821.05 249327.05  1.53     1.34 318.01
## BILL_AMT_JUN  -65167 174333.35 239500.35  1.56     1.38 297.94
## BILL_AMT_MAY  -61372 165794.30 227166.30  1.59     1.46 285.12
## BILL_AMT_APR  -57060 161912.00 218972.00  1.59     1.47 278.76
## PAID_AMT_SEP       0  18428.20  18428.20  1.82     2.39  29.98
## PAID_AMT_AUG       0  19004.35  19004.35  1.90     2.67  30.77
## PAID_AMT_JUL       0  17589.40  17589.40  1.93     2.81  28.39
## PAID_AMT_JUN       0  16014.95  16014.95  1.88     2.44  27.06
## PAID_AMT_MAY       0  16000.00  16000.00  1.89     2.48  26.90
## PAID_AMT_APR       0  17343.80  17343.80  1.98     2.81  29.08
## DEFAULT            0      1.00      1.00  1.34    -0.20   0.00
## RATIO_SEP          0      1.00      1.00  2.29     3.25   0.00
## RATIO_AUG          0      1.00      1.00  2.23     2.99   0.00
## RATIO_JUL          0      1.00      1.00  2.35     3.52   0.00
## RATIO_JUN          0      1.00      1.00  2.45     3.98   0.00
## RATIO_MAY          0      1.00      1.00  5.18    24.79   0.00
## RATIO_APR          0      1.00      1.00  2.18     2.73   0.00
## balance_SEP   -90862 191911.60 282773.60  1.40     1.13 334.33
## balance_AUG   -87288 185663.85 272951.85  1.42     1.17 323.12
## balance_JUL   -83226 179756.00 262982.00  1.44     1.24 312.91
## balance_JUN   -74731 166742.10 241473.10  1.48     1.30 293.33
## balance_MAY   -70034 157664.75 227698.75  1.51     1.38 277.57
## balance_APR   -69089 153520.35 222609.35  1.50     1.37 272.18
## TIMELY_PD_SEP      0      1.00      1.00 -1.30    -0.31   0.00
## TIMELY_PD_AUG      0      1.00      1.00 -1.98     1.93   0.00
## TIMELY_PD_JUL      0      1.00      1.00 -2.07     2.28   0.00
## TIMELY_PD_JUN      0      1.00      1.00 -2.38     3.68   0.00
## TIMELY_PD_MAY      0      1.00      1.00 -2.69     5.22   0.00
## TIMELY_PD_APR      0      1.00      1.00 -2.62     4.86   0.00
```

```
bank_data_cor <- subset(bank_data1, select = -c(1,2,3,4,5,24))
#is.na(bank_data1)
cormatrix_bank = cor(bank_data_cor)
library(psych)
cortest.bartlett(cormatrix_bank,100)
```

```
## $chisq
## [1] 4723.631
##
## $p.value
## [1] 0
##
## $df
## [1] 630
```

```
KMO(cormatrix_bank)
```

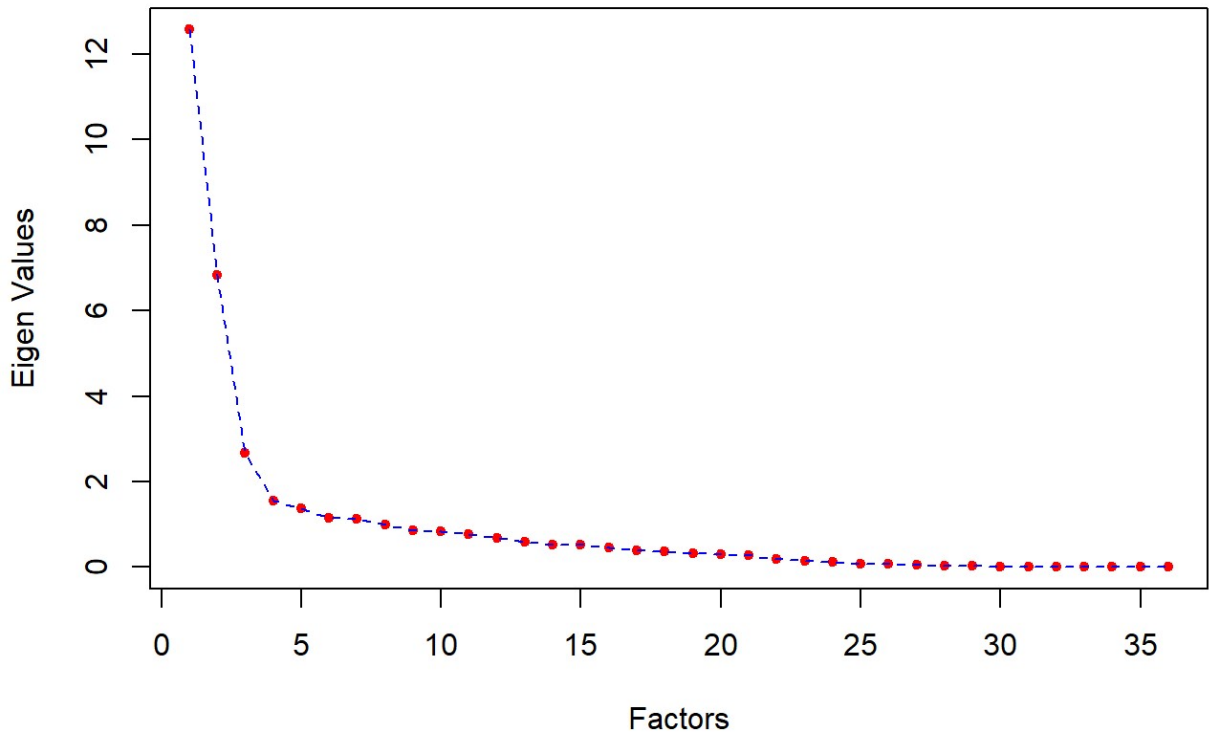
```
## Kaiser-Meyer-Olkin factor adequacy
## Call: KMO(r = cormatrix_bank)
## Overall MSA =  0.84
## MSA for each item =
##     REPAY_SEP     REPAY_AUG     REPAY_JUL     REPAY_JUN     REPAY_MAY
##          0.72          0.72          0.78          0.78          0.76
##     REPAY_APR  BILL_AMT_SEP  BILL_AMT_AUG  BILL_AMT_JUL  BILL_AMT_JUN
##          0.79          0.89          0.91          0.91          0.91
##  BILL_AMT_MAY  BILL_AMT_APR  PAID_AMT_SEP  PAID_AMT_AUG  PAID_AMT_JUL
##          0.91          0.89          0.84          0.83          0.82
##  PAID_AMT_JUN  PAID_AMT_MAY  PAID_AMT_APR     RATIO_SEP     RATIO_AUG
##          0.84          0.83          0.77          0.90          0.88
##     RATIO_JUL     RATIO_JUN     RATIO_MAY     RATIO_APR   balance_SEP
##          0.86          0.85          0.87          0.86          0.88
##   balance_AUG   balance_JUL   balance_JUN   balance_MAY   balance_APR
##          0.90          0.90          0.90          0.90          0.89
## TIMELY_PD_SEP TIMELY_PD_AUG TIMELY_PD_JUL TIMELY_PD_JUN TIMELY_PD_MAY
##          0.56          0.62          0.69          0.69          0.66
## TIMELY_PD_APR
##          0.70
```

We have checked Kaiser-Meyer-Olkin (KMO) to find the Test for Sampling Adequacy whereas the values in this case is greater than .5 , hence the data set is occurred with enough samples .

```
#Check eigen values
evector = eigen(cormatrix_bank)
eigen_value = evector$values
eigen_value
```

```
##  [1] 12.564536917  6.837047228  2.673468931  1.545928586  1.384570984
##  [6]  1.147488741  1.117272212  0.984109325  0.860073851  0.845526832
## [11]  0.777669220  0.675126384  0.595873143  0.529439215  0.520280068
## [16]  0.454700120  0.395735592  0.372796398  0.324856581  0.292909561
## [21]  0.268496618  0.191038994  0.152532812  0.111790321  0.082502461
## [26]  0.076962738  0.065080659  0.043633264  0.028124844  0.020848150
## [31]  0.017292527  0.011301741  0.009076509  0.008154693  0.007504493
## [36]  0.006249286
```

```
plot(eigen_value, xlab = "Factors", ylab = "Eigen Values", col="red", pch=20)
lines(eigen_value, col="blue", lty = 2)
```



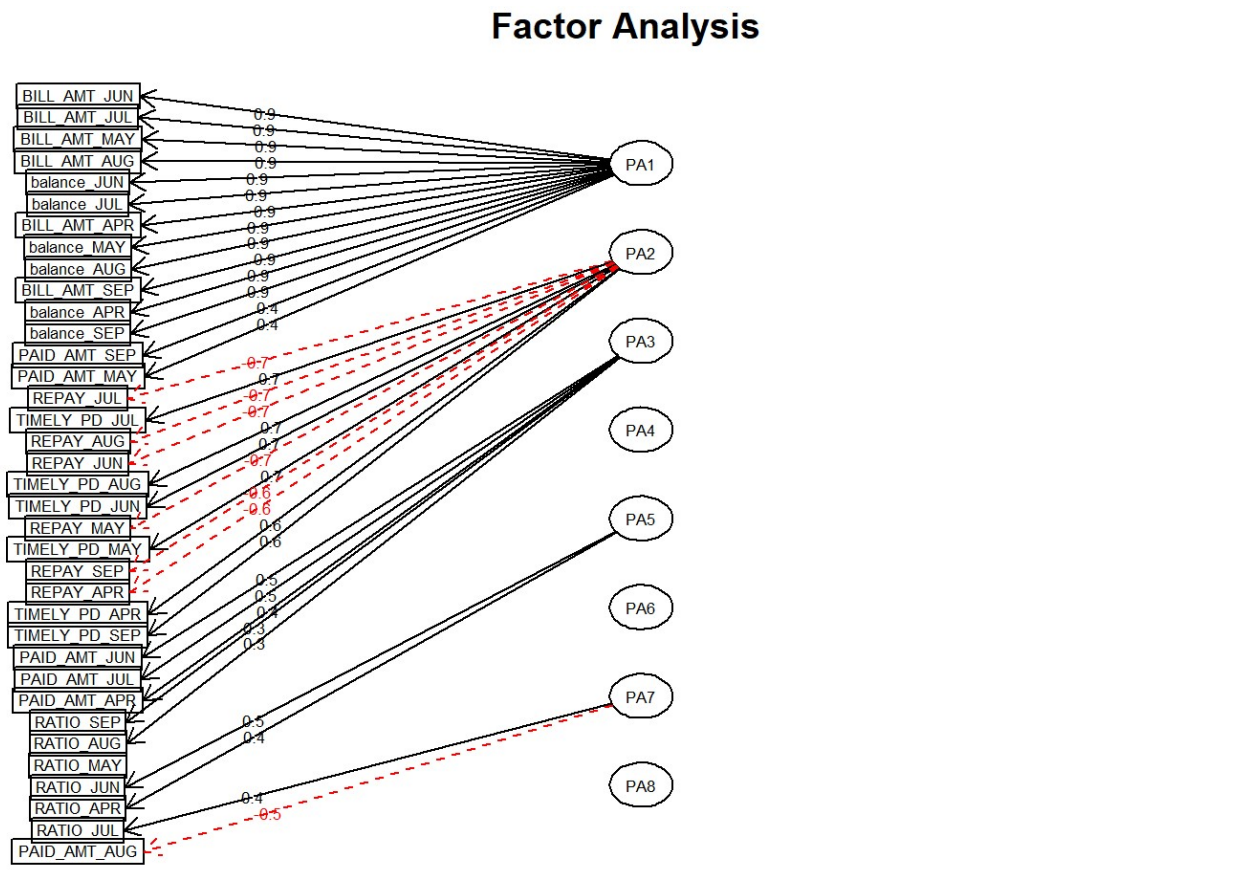

```
fa1 = fa(r= bank_data_cor, nfactors =8, rotate ="none", fm ="pa")
print(fa1)
```

```
## Factor Analysis using method =  pa
## Call: fa(r = bank_data_cor, nfactors = 8, rotate = "none", fm = "pa")
## Standardized loadings (pattern matrix) based upon correlation matrix
##                 PA1   PA2   PA3   PA4   PA5   PA6   PA7   PA8   h2    u2
## REPAY_SEP      0.35 -0.64 -0.10  0.35 -0.11 -0.12 -0.09  0.16 0.72 0.280
## REPAY_AUG      0.43 -0.70 -0.01  0.23 -0.05  0.12  0.00  0.24 0.80 0.200
## REPAY_JUL      0.45 -0.71  0.09  0.00  0.01  0.39  0.13  0.04 0.89 0.111
## REPAY_JUN      0.46 -0.69  0.15 -0.24  0.05  0.17 -0.15  0.02 0.83 0.171
## REPAY_MAY      0.47 -0.67  0.19 -0.36 -0.04 -0.08  0.00  0.16 0.86 0.137
## REPAY_APR      0.47 -0.61  0.22 -0.31  0.14 -0.19  0.06  0.25 0.85 0.151
## BILL_AMT_SEP   0.90  0.21 -0.15  0.06  0.08  0.07 -0.01  0.11 0.90 0.101
## BILL_AMT_AUG   0.93  0.20 -0.12  0.05  0.06  0.12  0.01  0.04 0.93 0.066
## BILL_AMT_JUL   0.94  0.20 -0.08  0.02  0.03  0.04 -0.12 -0.04 0.95 0.051
## BILL_AMT_JUN   0.94  0.19 -0.02 -0.01 -0.06 -0.08  0.03 -0.09 0.94 0.057
## BILL_AMT_MAY   0.93  0.19  0.03  0.02  0.06 -0.15  0.07 -0.12 0.94 0.057
## BILL_AMT_APR   0.91  0.17  0.02  0.01  0.00 -0.14  0.07 -0.13 0.89 0.105
## PAID_AMT_SEP   0.40  0.36  0.37 -0.05 -0.10  0.27  0.03 -0.05 0.52 0.481
## PAID_AMT_AUG   0.37  0.38  0.45 -0.06 -0.19  0.02 -0.48  0.03 0.76 0.238
## PAID_AMT_JUL   0.36  0.37  0.48  0.00 -0.37 -0.09  0.25  0.14 0.72 0.277
## PAID_AMT_JUN   0.36  0.37  0.54  0.18  0.23 -0.01  0.04  0.15 0.66 0.340
## PAID_AMT_MAY   0.36  0.33  0.36  0.06 -0.18  0.09 -0.02  0.05 0.41 0.588
## PAID_AMT_APR   0.32  0.32  0.37  0.05 -0.12  0.09 -0.04  0.12 0.38 0.616
## RATIO_SEP     -0.24  0.21  0.31 -0.01 -0.07  0.12  0.06 -0.21 0.26 0.736
## RATIO_AUG     -0.25  0.21  0.30  0.00 -0.05 -0.12 -0.30 -0.09 0.31 0.687
## RATIO_JUL     -0.24  0.19  0.29  0.04 -0.20 -0.12  0.36 -0.01 0.36 0.637
## RATIO_JUN     -0.26  0.20  0.37  0.22  0.49 -0.04  0.00 -0.01 0.54 0.464
## RATIO_MAY     -0.11  0.11  0.21  0.11  0.18  0.01  0.02  0.02 0.11 0.887
## RATIO_APR     -0.26  0.19  0.36  0.20  0.42 -0.02  0.03  0.03 0.45 0.549
## balance_SEP    0.87  0.16 -0.21  0.07  0.10  0.03 -0.01  0.13 0.86 0.138
## balance_AUG    0.90  0.16 -0.19  0.06  0.09  0.12  0.09  0.04 0.91 0.088
## balance_JUL    0.92  0.16 -0.15  0.02  0.10  0.06 -0.17 -0.06 0.94 0.062
## balance_JUN    0.92  0.15 -0.11 -0.05 -0.11 -0.07  0.03 -0.12 0.92 0.079
## balance_MAY    0.91  0.14 -0.03  0.01  0.10 -0.16  0.07 -0.12 0.90 0.099
## balance_APR    0.88  0.13 -0.04  0.00  0.03 -0.16  0.08 -0.15 0.86 0.143
## TIMELY_PD_SEP -0.06  0.58 -0.02 -0.57  0.18  0.27  0.11  0.08 0.80 0.203
## TIMELY_PD_AUG -0.17  0.69 -0.12 -0.43  0.10  0.02  0.02  0.01 0.71 0.289
## TIMELY_PD_JUL -0.16  0.70 -0.21 -0.17  0.00 -0.29 -0.18  0.27 0.78 0.225
## TIMELY_PD_JUN -0.18  0.68 -0.25  0.09 -0.07 -0.08  0.11  0.29 0.68 0.323
## TIMELY_PD_MAY -0.19  0.66 -0.26  0.22 -0.01  0.18  0.01  0.14 0.64 0.363
## TIMELY_PD_APR -0.17  0.59 -0.26  0.20 -0.13  0.25 -0.05  0.01 0.57 0.430
##                com
## REPAY_SEP      2.7
## REPAY_AUG      2.3
## REPAY_JUL      2.5
## REPAY_JUN      2.4
## REPAY_MAY      2.8
## REPAY_APR      3.6
## BILL_AMT_SEP   1.2
```

```
## BILL_AMT_AUG  1.2
## BILL_AMT_JUL  1.1
## BILL_AMT_JUN  1.1
## BILL_AMT_MAY  1.2
## BILL_AMT_APR  1.2
## PAID_AMT_SEP  4.0
## PAID_AMT_AUG  4.2
## PAID_AMT_JUL  4.7
## PAID_AMT_JUN  3.6
## PAID_AMT_MAY  3.7
## PAID_AMT_APR  3.6
## RATIO_SEP     4.2
## RATIO_AUG     4.3
## RATIO_JUL     4.4
## RATIO_JUN     3.4
## RATIO_MAY     3.9
## RATIO_APR     3.7
## balance_SEP   1.3
## balance_AUG   1.2
## balance_JUL   1.2
## balance_JUN   1.2
## balance_MAY   1.2
## balance_APR   1.2
## TIMELY_PD_SEP 2.8
## TIMELY_PD_AUG 2.0
## TIMELY_PD_JUL 2.4
## TIMELY_PD_JUN 2.0
## TIMELY_PD_MAY 2.0
## TIMELY_PD_APR 2.4
##
##                         PA1  PA2  PA3  PA4  PA5  PA6  PA7  PA8
## SS loadings           12.43 6.57 2.25 1.26 0.95 0.80 0.71 0.60
## Proportion Var         0.35 0.18 0.06 0.03 0.03 0.02 0.02 0.02
## Cumulative Var         0.35 0.53 0.59 0.63 0.65 0.67 0.69 0.71
## Proportion Explained   0.49 0.26 0.09 0.05 0.04 0.03 0.03 0.02
## Cumulative Proportion  0.49 0.74 0.83 0.88 0.92 0.95 0.98 1.00
##
## Mean item complexity =  2.6
## Test of the hypothesis that 8 factors are sufficient.
##
## The degrees of freedom for the null model are  630  and the objective function was
54.82 with Chi Square of  1643833
## The degrees of freedom for the model are 370  and the objective function was  16.01
##
## The root mean square of the residuals (RMSR) is  0.03
## The df corrected root mean square of the residuals is  0.04
##
## The harmonic number of observations is  30000 with the empirical chi square  31288.
72  with prob <  0
```

```
## The total number of observations was  30000  with Likelihood Chi Square =  479933.5
with prob <  0
##
## Tucker Lewis Index of factoring reliability =  0.503
## RMSEA index =  0.208  and the 90 % confidence intervals are  0.207 0.208
## BIC =  476119.2
## Fit based upon off diagonal values = 0.99
## Measures of factor score adequacy
##                                                  PA1  PA2  PA3  PA4  PA5
## Correlation of (regression) scores with factors   1.00 0.98 0.94 0.93 0.87
## Multiple R square of scores with factors          0.99 0.97 0.88 0.87 0.76
## Minimum correlation of possible factor scores     0.99 0.94 0.75 0.73 0.52
##                                                  PA6  PA7  PA8
## Correlation of (regression) scores with factors   0.93 0.88 0.87
## Multiple R square of scores with factors          0.86 0.78 0.75
## Minimum correlation of possible factor scores     0.72 0.56 0.51
```

```
fa.diagram(fa1)
```



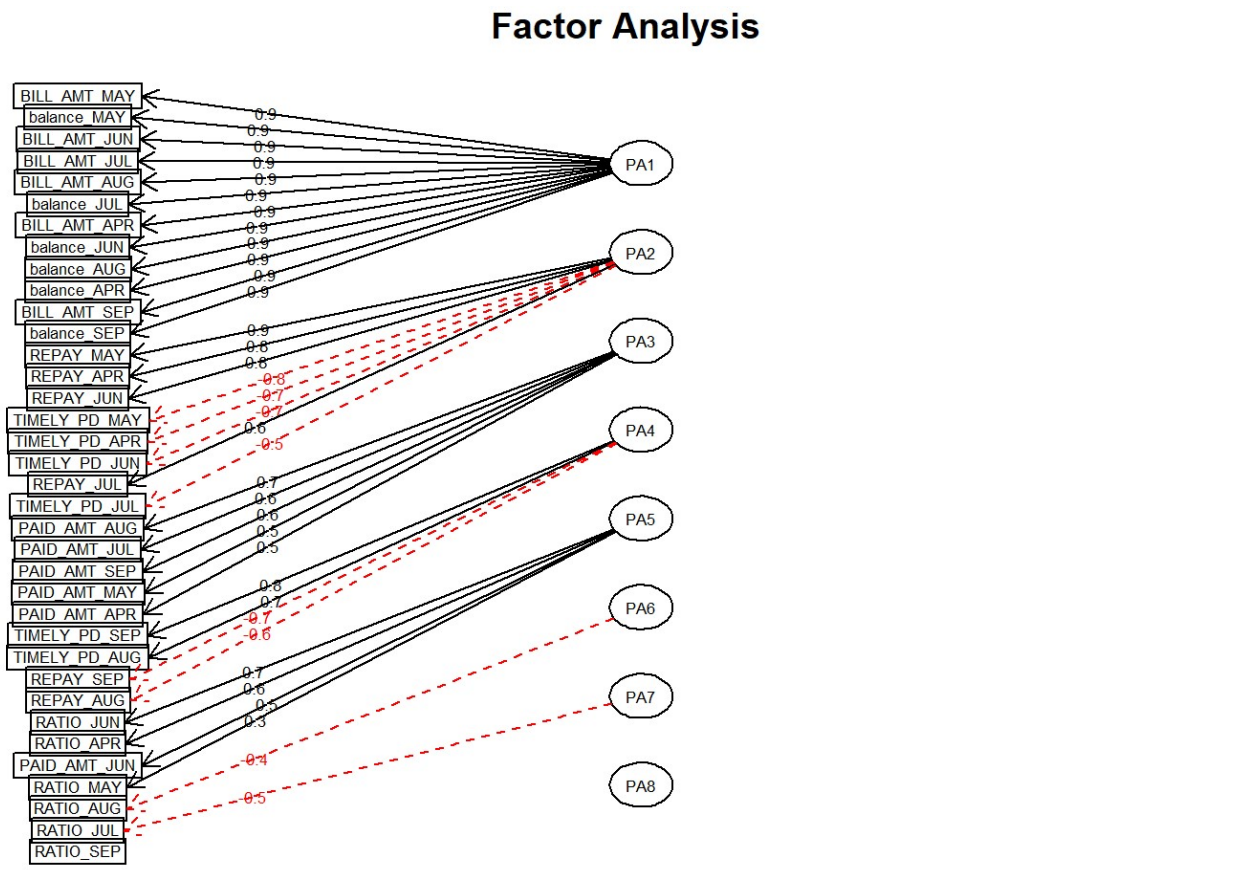
```
fa2 = fa(r= bank_data_cor, nfactors = 8
         , rotate ="varimax", fm ="pa")
print(fa2)
```

```
## Factor Analysis using method =  pa
## Call: fa(r = bank_data_cor, nfactors = 8, rotate = "varimax", fm = "pa")
## Standardized loadings (pattern matrix) based upon correlation matrix
##                 PA1   PA2   PA3   PA4   PA5   PA6   PA7   PA8   h2    u2
## REPAY_SEP      0.18  0.36 -0.11 -0.68 -0.17  0.12  0.11  0.15 0.72 0.280
## REPAY_AUG      0.19  0.51 -0.01 -0.56 -0.14  0.36  0.17  0.11 0.80 0.200
## REPAY_JUL      0.18  0.65  0.02 -0.30 -0.14  0.51  0.17 -0.19 0.89 0.111
## REPAY_JUN      0.19  0.79  0.05 -0.17 -0.17  0.16  0.29 -0.06 0.83 0.171
## REPAY_MAY      0.20  0.85  0.05 -0.10 -0.23  0.09  0.03  0.15 0.86 0.137
## REPAY_APR      0.24  0.83 -0.02 -0.08 -0.04  0.10 -0.02  0.28 0.85 0.151
## BILL_AMT_SEP   0.89  0.00  0.17 -0.03 -0.07  0.18  0.14  0.13 0.90 0.101
## BILL_AMT_AUG   0.92  0.02  0.19 -0.01 -0.08  0.18  0.13  0.04 0.93 0.066
## BILL_AMT_JUL   0.92  0.06  0.23 -0.02 -0.10  0.02  0.17  0.01 0.95 0.051
## BILL_AMT_JUN   0.93  0.10  0.22 -0.02 -0.14 -0.01 -0.05  0.00 0.94 0.057
## BILL_AMT_MAY   0.94  0.14  0.16 -0.02  0.00 -0.05 -0.08  0.00 0.94 0.057
## BILL_AMT_APR   0.91  0.13  0.17 -0.03 -0.06 -0.05 -0.10 -0.02 0.89 0.105
## PAID_AMT_SEP   0.35 -0.03  0.55  0.18  0.09  0.08 -0.04 -0.19 0.52 0.481
## PAID_AMT_AUG   0.28 -0.01  0.71  0.07  0.05 -0.36  0.20  0.02 0.76 0.238
## PAID_AMT_JUL   0.28 -0.02  0.64  0.05  0.00  0.00 -0.48  0.08 0.72 0.277
## PAID_AMT_JUN   0.36  0.00  0.48  0.03  0.53  0.01 -0.11  0.08 0.66 0.340
## PAID_AMT_MAY   0.31 -0.07  0.55  0.02  0.07 -0.01 -0.10 -0.03 0.41 0.588
## PAID_AMT_APR   0.26 -0.05  0.54  0.04  0.13  0.00 -0.07  0.03 0.38 0.616
## RATIO_SEP     -0.21 -0.11  0.23  0.14  0.16 -0.08 -0.15 -0.29 0.26 0.736
## RATIO_AUG     -0.24 -0.10  0.25  0.06  0.16 -0.39  0.04 -0.04 0.31 0.687
## RATIO_JUL     -0.22 -0.12  0.18  0.06  0.10  0.00 -0.50 -0.05 0.36 0.637
## RATIO_JUN     -0.15 -0.09  0.02  0.05  0.70 -0.08 -0.01 -0.03 0.54 0.464
## RATIO_MAY     -0.08 -0.05  0.07  0.02  0.31 -0.01 -0.03 -0.01 0.11 0.887
## RATIO_APR     -0.17 -0.09  0.05  0.06  0.64 -0.03 -0.03 -0.01 0.45 0.549
## balance_SEP    0.88  0.01  0.09 -0.05 -0.08  0.16  0.15  0.17 0.86 0.138
## balance_AUG    0.91  0.03  0.09 -0.02 -0.09  0.25  0.09  0.04 0.91 0.088
## balance_JUL    0.91  0.07  0.14 -0.02 -0.10  0.02  0.26  0.00 0.94 0.062
## balance_JUN    0.91  0.10  0.16 -0.02 -0.24 -0.02 -0.03 -0.02 0.92 0.079
## balance_MAY    0.93  0.15  0.08 -0.03 -0.01 -0.04 -0.07  0.01 0.90 0.099
## balance_APR    0.90  0.14  0.09 -0.03 -0.08 -0.05 -0.09 -0.02 0.86 0.143
## TIMELY_PD_SEP  0.04 -0.21  0.14  0.84  0.04  0.13  0.04  0.04 0.80 0.203
## TIMELY_PD_AUG  0.02 -0.40  0.05  0.72  0.02 -0.12 -0.03  0.09 0.71 0.289
## TIMELY_PD_JUL  0.03 -0.54  0.06  0.41  0.00 -0.31 -0.01  0.46 0.78 0.225
## TIMELY_PD_JUN  0.02 -0.69  0.06  0.24  0.03  0.03 -0.15  0.33 0.68 0.323
## TIMELY_PD_MAY  0.01 -0.76  0.08  0.17  0.08  0.11  0.06  0.09 0.64 0.363
## TIMELY_PD_APR -0.01 -0.72  0.12  0.13 -0.05  0.08  0.10 -0.06 0.57 0.430
##                com
## REPAY_SEP      2.2
## REPAY_AUG      3.4
## REPAY_JUL      3.1
## REPAY_JUN      1.7
## REPAY_MAY      1.4
## REPAY_APR      1.5
## BILL_AMT_SEP   1.3
```

```
## BILL_AMT_AUG  1.2
## BILL_AMT_JUL  1.2
## BILL_AMT_JUN  1.2
## BILL_AMT_MAY  1.1
## BILL_AMT_APR  1.2
## PAID_AMT_SEP  2.4
## PAID_AMT_AUG  2.1
## PAID_AMT_JUL  2.3
## PAID_AMT_JUN  2.9
## PAID_AMT_MAY  1.7
## PAID_AMT_APR  1.7
## RATIO_SEP     5.2
## RATIO_AUG     3.2
## RATIO_JUL     1.9
## RATIO_JUN     1.2
## RATIO_MAY     1.3
## RATIO_APR     1.2
## balance_SEP   1.3
## balance_AUG   1.2
## balance_JUL   1.3
## balance_JUN   1.2
## balance_MAY   1.1
## balance_APR   1.1
## TIMELY_PD_SEP 1.2
## TIMELY_PD_AUG 1.7
## TIMELY_PD_JUL 3.5
## TIMELY_PD_JUN 1.8
## TIMELY_PD_MAY 1.2
## TIMELY_PD_APR 1.2
##
##                         PA1  PA2  PA3  PA4  PA5  PA6  PA7  PA8
## SS loadings           11.01 5.10 2.56 2.51 1.67 1.05 0.97 0.71
## Proportion Var         0.31 0.14 0.07 0.07 0.05 0.03 0.03 0.02
## Cumulative Var         0.31 0.45 0.52 0.59 0.63 0.66 0.69 0.71
## Proportion Explained   0.43 0.20 0.10 0.10 0.07 0.04 0.04 0.03
## Cumulative Proportion  0.43 0.63 0.73 0.83 0.89 0.93 0.97 1.00
##
## Mean item complexity =  1.8
## Test of the hypothesis that 8 factors are sufficient.
##
## The degrees of freedom for the null model are  630  and the objective function was
54.82 with Chi Square of  1643833
## The degrees of freedom for the model are 370  and the objective function was  16.01
##
## The root mean square of the residuals (RMSR) is  0.03
## The df corrected root mean square of the residuals is  0.04
##
## The harmonic number of observations is  30000 with the empirical chi square  31288.
72  with prob <  0
```

```
## The total number of observations was  30000  with Likelihood Chi Square =  479933.5
with prob <  0
##
## Tucker Lewis Index of factoring reliability =  0.503
## RMSEA index =  0.208  and the 90 % confidence intervals are  0.207 0.208
## BIC =  476119.2
## Fit based upon off diagonal values = 0.99
## Measures of factor score adequacy
##                                                  PA1  PA2  PA3  PA4  PA5
## Correlation of (regression) scores with factors   0.99 0.97 0.92 0.95 0.88
## Multiple R square of scores with factors          0.98 0.94 0.85 0.89 0.78
## Minimum correlation of possible factor scores     0.96 0.88 0.70 0.79 0.55
##                                                  PA6  PA7  PA8
## Correlation of (regression) scores with factors   0.91 0.90 0.88
## Multiple R square of scores with factors          0.84 0.81 0.77
## Minimum correlation of possible factor scores     0.67 0.63 0.54
```

```
fa.diagram(fa2)
```



#Renaming vairbles

```
bank_final <- subset(bank_data1,select = c(1,2,3,4,5,24))
bank_final_data = cbind(bank_final,fa2$scores)
head(bank_final_data)
```

```
##   LIMIT_BAL SEX EDUCATION MARRIAGE AGE DEFAULT         PA1          PA2
## 1     20000   2         2        1  24       1 -0.65505048 -1.566069537
## 2    120000   2         2        2  26       1 -0.85452952  1.053568856
## 3     90000   2         2        2  34       0 -0.52269375  0.096085124
## 4     50000   2         2        1  37       0 -0.04871894 -0.002544749
## 5     50000   1         2        1  57       0 -0.85332389  0.266568820
## 6     50000   1         1        2  37       0 -0.08186224 -0.065224365
##          PA3        PA4         PA5        PA6        PA7       PA8
## 1 -0.5853008 -2.7396537 -0.28868694 -0.3336919  0.4391205 0.1522075
## 2 -0.4161835  0.1636740 -0.03377417  0.4646900 -0.1320396 1.4610357
## 3 -0.3000248  0.4873752 -0.43826364  0.4397578  0.2643110 0.7402958
## 4 -0.4353623  0.5266916 -0.39476438  0.4034434  0.5480870 0.5670916
## 5  1.9285488 -0.1913143 -0.19966378 -1.9722762  0.8661627 1.0896000
## 6 -0.3937278  0.5364438 -0.32483433  0.6391223  0.9395406 0.7666797
```

```
colnames(bank_final_data) = c("LIMIT_BAL","SEX","EDUCATION","MARRIAGE","AGE","DEFAUL
T","BILLED _AMT",   "REPAY_STATUS ",    "PAID_AMT", "TIMELY_PAID_AMT",  "RATIO_PADI_AM
T1",  "RATIO_PADI_AMT2",  "RATIO_PADI_AMT3","RATIO_PADI_AMT4")
str(bank_final_data)
```

```
## 'data.frame':    30000 obs. of  14 variables:
##  $ LIMIT_BAL      : num  20000 120000 90000 50000 50000 50000 500000 100000 140000
20000 ...
##  $ SEX            : int  2 2 2 2 1 1 1 2 2 1 ...
##  $ EDUCATION      : int  2 2 2 2 2 1 1 2 3 3 ...
##  $ MARRIAGE       : int  1 2 2 1 1 2 2 2 1 2 ...
##  $ AGE            : int  24 26 34 37 57 37 29 23 28 35 ...
##  $ DEFAULT        : int  1 1 0 0 0 0 0 0 0 0 ...
##  $ BILLED _AMT    : num  -0.6551 -0.8545 -0.5227 -0.0487 -0.8533 ...
##  $ REPAY_STATUS   : num  -1.56607 1.05357 0.09609 -0.00254 0.26657 ...
##  $ PAID_AMT       : num  -0.585 -0.416 -0.3 -0.435 1.929 ...
##  $ TIMELY_PAID_AMT: num  -2.74 0.164 0.487 0.527 -0.191 ...
##  $ RATIO_PADI_AMT1: num  -0.2887 -0.0338 -0.4383 -0.3948 -0.1997 ...
##  $ RATIO_PADI_AMT2: num  -0.334 0.465 0.44 0.403 -1.972 ...
##  $ RATIO_PADI_AMT3: num  0.439 -0.132 0.264 0.548 0.866 ...
##  $ RATIO_PADI_AMT4: num  0.152 1.461 0.74 0.567 1.09 ...
```

The variable we would be considering

Check the proportion of data

```
### Check the proportion of data #
#Default payment in June, 2005 (1=yes, 0=no)
table(bank_final_data$DEFAULT)
```

```
##
##     0     1
## 23364  6636
```

```
nrow(subset(bank_final_data, DEFAULT == 1))/nrow(bank_final_data)
```

```
## [1] 0.2212
```

Take the sample subset from the major class (here negative)

```
set.seed(123)
train_idx <- sample(c(6:nrow(bank_final_data)), round(nrow(bank_final_data) * 0.7,0),
replace = FALSE)
train_data <- bank_final_data[train_idx,]
test_data <- bank_final_data[-train_idx,]


table(train_data$DEFAULT)
```

```
##
##     0     1
## 16360  4640
```

```
table(test_data$DEFAULT)
```

```
##
##    0    1
## 7004 1996
```

Checking the samples counts

```
train.pos <- subset(train_data, DEFAULT == 1)
train.neg <- subset(train_data, DEFAULT == 0)

dim(train.pos)
```

```
## [1] 4640   14
```

```
dim(train.neg)
```

```
## [1] 16360    14
```

Take the sample subset from the major class (here negative)

```
## Set the seed
set.seed(108)

train.neg.sub_idx <- sample(c(24:nrow(train.neg)), nrow(train.pos), replace = FALSE)
train_new <- train.neg[train.neg.sub_idx,]
dim(train_new)
```

```
## [1] 4640   14
```

Merge the negative and positive cases

```
train_new <- rbind(train_new, train.pos)
table(train_new$DEFAULT)
```

```
##
##    0    1
## 4640 4640
```

Rendomizing the data

```
train_new <- train_new[sample(6:nrow(train_new)),]
### Now check the proportion of target in the sample
## in train_data
nrow(subset(train_data, DEFAULT == 1))/nrow(train_data)
```

```
## [1] 0.2209524
```

```
## in train.new
nrow(subset(train_new, DEFAULT == 1))/nrow(train_new)
```

```
## [1] 0.5002695
```

```
table(train_new$DEFAULT)
```

```
##
##    0    1
## 4635 4640
```

```
str(train_new)
```

```
## 'data.frame':    9275 obs. of  14 variables:
##  $ LIMIT_BAL      : num  80000 30000 150000 40000 500000 180000 20000 310000 10000
30000 ...
##  $ SEX            : int  1 1 2 2 2 1 2 2 1 1 ...
##  $ EDUCATION      : int  3 1 1 3 1 1 2 3 2 1 ...
##  $ MARRIAGE       : int  2 1 2 1 2 2 1 2 2 1 ...
##  $ AGE            : int  30 37 52 41 34 31 25 38 43 32 ...
##  $ DEFAULT        : int  1 1 0 0 0 0 1 0 1 1 ...
##  $ BILLED _AMT    : num  -0.57 -0.212 -0.585 -0.585 0.724 ...
##  $ REPAY_STATUS   : num  0.0364 0.08 0.0362 1.7462 -0.0812 ...
##  $ PAID_AMT       : num  -0.5159 -0.3857 2.9588 -0.0822 -0.4153 ...
##  $ TIMELY_PAID_AMT: num  0.4542 0.5058 0.0266 -1.7071 0.481 ...
##  $ RATIO_PADI_AMT1: num  -0.4017 -0.4105 0.0517 -0.4185 -0.3699 ...
##  $ RATIO_PADI_AMT2: num  0.672 0.303 0.966 0.126 0.206 ...
##  $ RATIO_PADI_AMT3: num  0.567 0.152 0.235 0.181 0.233 ...
##  $ RATIO_PADI_AMT4: num  1.292 0.519 1.933 -1.44 0.348 ...
```