8\_1.R

dell

Fri Sep 09 00:04:32 2016

rm(list=ls(all=T))  
  
set.seed(999)  
  
setwd("C:/Users/dell/Dropbox/compitition/nasscom/BFSI")  
library(data.table)

## Warning: package 'data.table' was built under R version 3.2.5

library(mice)

## Warning: package 'mice' was built under R version 3.2.5

## Loading required package: Rcpp

## Warning: package 'Rcpp' was built under R version 3.2.5

## mice 2.25 2015-11-09

library(dummies)

## Warning: package 'dummies' was built under R version 3.2.3

## dummies-1.5.6 provided by Decision Patterns

library(xgboost)

## Warning: package 'xgboost' was built under R version 3.2.5

library(Matrix)  
library(corrplot)

## Warning: package 'corrplot' was built under R version 3.2.5

library(car)

## Warning: package 'car' was built under R version 3.2.5

library(caret)

## Warning: package 'caret' was built under R version 3.2.5

## Loading required package: lattice

## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 3.2.5

library(leaps)

## Warning: package 'leaps' was built under R version 3.2.5

library(xlsx)

## Warning: package 'xlsx' was built under R version 3.2.3

## Loading required package: rJava

## Warning: package 'rJava' was built under R version 3.2.3

## Loading required package: xlsxjars

## Warning: package 'xlsxjars' was built under R version 3.2.3

library(randomForest)

## Warning: package 'randomForest' was built under R version 3.2.3

## randomForest 4.6-12

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

##   
## Attaching package: 'randomForest'

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

require('vcd')

## Loading required package: vcd

## Warning: package 'vcd' was built under R version 3.2.5

## Loading required package: grid

source("funplot.R")  
#1.Data Preparation   
  
tr<-fread("BFSI Stage 1 Train data.csv",stringsAsFactors=F)  
te<-fread("BFSI Stage 1 Test data.csv",stringsAsFactors=F)  
  
sol<-fread("BFSI - Solution submission template.csv",stringsAsFactors=T)  
te<-te[,c("Behavioural\_Score","Application\_Score"):= NA]  
  
c<-list(tr, te)  
a<- rbindlist(c)  
  
a$Total\_Bank\_Products<-ordered(a$Total\_Bank\_Products,levels=c(">=6","5","2","1","3","0","4"),labels=c(6,5,2,1,3,0,4))  
a$Active\_Bank\_Products<-ordered(a$Active\_Bank\_Products,levels=c(">=5","2","1","3","0","4"),labels=c(5,2,1,3,0,4))  
a$Tenure\_with\_Bank\_Group<-ordered(a$Tenure\_with\_Bank\_Group,levels=c("<=5 YRS","<=10 YRS",">10 YRS","<=2 YRS", "<=1 YRS"),labels=c(2,4,5,1,0))  
a$Education<-ordered(a$Education,levels=c("University","Graduate and Higher","Intermediate","High School","Unknown","No Education","Primary School" ),labels=c(5,4,3,2,NA,0,1))  
  
colSums(is.na(a))

## Identifier Age   
## 0 0   
## Education Indutry\_Groups   
## 0 0   
## Industry\_Domain Marital\_Status   
## 0 0   
## Occupation Customer\_Segment   
## 0 0   
## Gender Metropolitan\_City   
## 0 0   
## Residence Insurance\_Product\_type   
## 0 0   
## Insurance\_Acquisition\_Channel Active\_Bank\_Products   
## 0 0   
## Total\_Bank\_Products Commercial\_Loan   
## 0 0   
## Consumer\_Auto\_Loan Mortgage\_Loan   
## 0 0   
## Personal\_Loan Credit\_Card   
## 0 0   
## Deposit Pre-Approved Auto Limit   
## 0 0   
## Pre-Approved Mortgage Limit Pre-Approved Personal Limit   
## 0 0   
## Tenure\_with\_Bank\_Group Tenure\_of\_Insurance   
## 0 0   
## Salary\_Amount Credit\_Limit   
## 976 1681   
## Total\_Asset\_Under\_Mngmnt Avg\_Monthly\_Balance   
## 0 0   
## Application\_Score Behavioural\_Score   
## 2307 2307

a[,.(.N),.(`Pre-Approved Auto Limit`)]

## Pre-Approved Auto Limit N  
## 1: MISSING 6084  
## 2: >480,000 1944  
## 3: <360,000 328  
## 4: <480,000 354  
## 5: <240,000 344  
## 6: <120,000 151  
## 7: ZERO 26

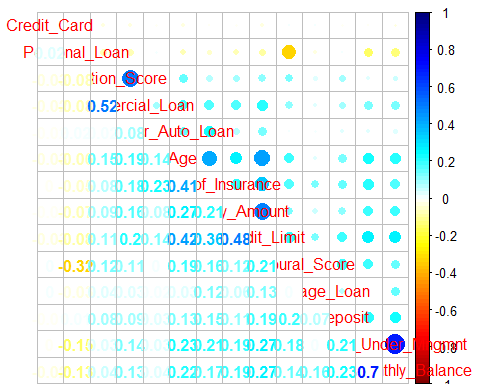
a[,.(.N),.(`Pre-Approved Mortgage Limit`)]

## Pre-Approved Mortgage Limit N  
## 1: MISSING 6084  
## 2: >=5,000,000 1434  
## 3: <2,500,000 612  
## 4: <5,000,000 505  
## 5: <1,000,000 335  
## 6: ZERO 261

a[,.(.N),.(`Pre-Approved Personal Limit`)]

## Pre-Approved Personal Limit N  
## 1: MISSING 6084  
## 2: ZERO 2068  
## 3: <1,500,000 782  
## 4: <1,000,000 250  
## 5: <500,000 47

a\_train<-a[,":="(`Pre-Approved Mortgage Limit`=NULL,  
 `Pre-Approved Auto Limit`=NULL,  
 `Pre-Approved Personal Limit`=NULL)]  
  
a0<-rbindlist(c)  
cp<-cor(a0[,c(2,16,17,18,19,20,21,26:32),with=F],use="complete.obs")  
col1 <- colorRampPalette(c("#7F0000","red","#FF7F00","yellow","white",   
 "cyan", "#007FFF", "blue","#00007F"))  
corrplot.mixed(cp,order="AOE",col=col1(100))



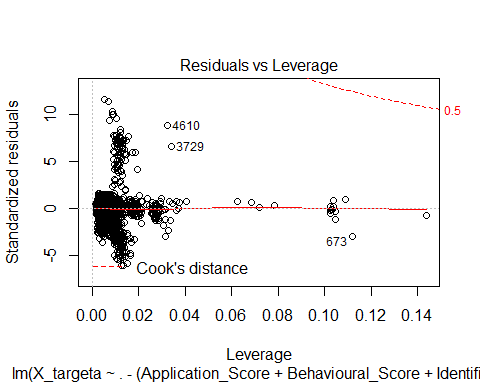
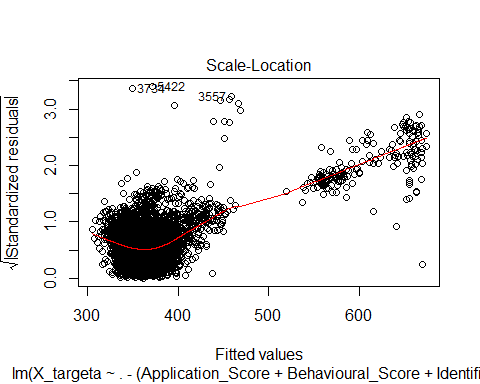
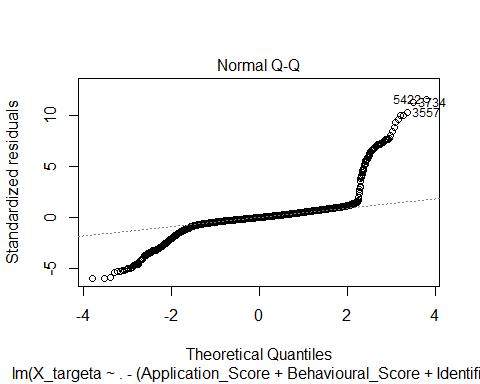
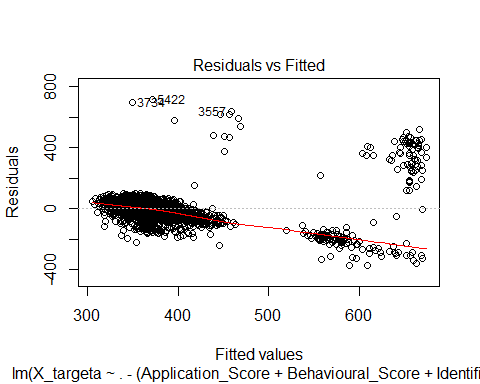
a\_train <- a[,":="(Identifier =as.numeric(Identifier),  
 Age=as.numeric(Age),  
 Education=(Education),  
 Indutry\_Groups=as.numeric(as.factor(Indutry\_Groups)),  
 Industry\_Domain=as.numeric(as.factor(Industry\_Domain)),  
 Marital\_Status=as.numeric(as.factor(Marital\_Status)),  
 Occupation=as.numeric(as.factor(Occupation)),  
 Customer\_Segment=as.numeric(as.factor(Customer\_Segment)),  
 Gender=ifelse(Gender=="Male",1,0),  
 Metropolitan\_City=as.numeric(as.factor(Metropolitan\_City)),  
 Residence=as.numeric(as.factor(Residence)),  
 Insurance\_Product\_type=as.numeric(as.factor(Insurance\_Product\_type)),  
 Insurance\_Acquisition\_Channel=as.numeric(as.factor(Insurance\_Acquisition\_Channel)),  
 Active\_Bank\_Products=(Active\_Bank\_Products),  
 Total\_Bank\_Products=as.numeric(Total\_Bank\_Products),  
 Commercial\_Loan=as.numeric(as.factor(Commercial\_Loan)),  
 Consumer\_Auto\_Loan=as.numeric(as.factor(Consumer\_Auto\_Loan)),  
 Mortgage\_Loan=as.numeric(as.factor(Mortgage\_Loan)),  
 Personal\_Loan=as.numeric(as.factor(Personal\_Loan)),  
 Credit\_Card=as.numeric(as.factor(Credit\_Card)),  
 Deposit=as.numeric(as.factor(Deposit)),  
 Tenure\_with\_Bank\_Group=(Tenure\_with\_Bank\_Group),  
 Tenure\_of\_Insurance=as.numeric(Tenure\_of\_Insurance),  
 Salary\_Amount=as.numeric(Salary\_Amount),  
 Credit\_Limit=as.numeric(Credit\_Limit),  
 Total\_Asset\_Under\_Mngmnt=as.numeric(as.factor(Total\_Asset\_Under\_Mngmnt)),  
 Avg\_Monthly\_Balance=as.numeric(Avg\_Monthly\_Balance),  
 Application\_Score=as.numeric(Application\_Score),  
 Behavioural\_Score=as.numeric(Behavioural\_Score)  
)]  
  
  
Dat1 <- subset(a\_train, select=c(Salary\_Amount,Credit\_Limit,Avg\_Monthly\_Balance,Tenure\_of\_Insurance,  
 Active\_Bank\_Products,Total\_Bank\_Products,Age,Total\_Asset\_Under\_Mngmnt))   
imp <- mice(Dat1, m=5, maxit=10,method="pmm", seed=999)

##   
## iter imp variable  
## 1 1 Salary\_Amount Credit\_Limit  
## 1 2 Salary\_Amount Credit\_Limit  
## 1 3 Salary\_Amount Credit\_Limit  
## 1 4 Salary\_Amount Credit\_Limit  
## 1 5 Salary\_Amount Credit\_Limit  
## 2 1 Salary\_Amount Credit\_Limit  
## 2 2 Salary\_Amount Credit\_Limit  
## 2 3 Salary\_Amount Credit\_Limit  
## 2 4 Salary\_Amount Credit\_Limit  
## 2 5 Salary\_Amount Credit\_Limit  
## 3 1 Salary\_Amount Credit\_Limit  
## 3 2 Salary\_Amount Credit\_Limit  
## 3 3 Salary\_Amount Credit\_Limit  
## 3 4 Salary\_Amount Credit\_Limit  
## 3 5 Salary\_Amount Credit\_Limit  
## 4 1 Salary\_Amount Credit\_Limit  
## 4 2 Salary\_Amount Credit\_Limit  
## 4 3 Salary\_Amount Credit\_Limit  
## 4 4 Salary\_Amount Credit\_Limit  
## 4 5 Salary\_Amount Credit\_Limit  
## 5 1 Salary\_Amount Credit\_Limit  
## 5 2 Salary\_Amount Credit\_Limit  
## 5 3 Salary\_Amount Credit\_Limit  
## 5 4 Salary\_Amount Credit\_Limit  
## 5 5 Salary\_Amount Credit\_Limit  
## 6 1 Salary\_Amount Credit\_Limit  
## 6 2 Salary\_Amount Credit\_Limit  
## 6 3 Salary\_Amount Credit\_Limit  
## 6 4 Salary\_Amount Credit\_Limit  
## 6 5 Salary\_Amount Credit\_Limit  
## 7 1 Salary\_Amount Credit\_Limit  
## 7 2 Salary\_Amount Credit\_Limit  
## 7 3 Salary\_Amount Credit\_Limit  
## 7 4 Salary\_Amount Credit\_Limit  
## 7 5 Salary\_Amount Credit\_Limit  
## 8 1 Salary\_Amount Credit\_Limit  
## 8 2 Salary\_Amount Credit\_Limit  
## 8 3 Salary\_Amount Credit\_Limit  
## 8 4 Salary\_Amount Credit\_Limit  
## 8 5 Salary\_Amount Credit\_Limit  
## 9 1 Salary\_Amount Credit\_Limit  
## 9 2 Salary\_Amount Credit\_Limit  
## 9 3 Salary\_Amount Credit\_Limit  
## 9 4 Salary\_Amount Credit\_Limit  
## 9 5 Salary\_Amount Credit\_Limit  
## 10 1 Salary\_Amount Credit\_Limit  
## 10 2 Salary\_Amount Credit\_Limit  
## 10 3 Salary\_Amount Credit\_Limit  
## 10 4 Salary\_Amount Credit\_Limit  
## 10 5 Salary\_Amount Credit\_Limit

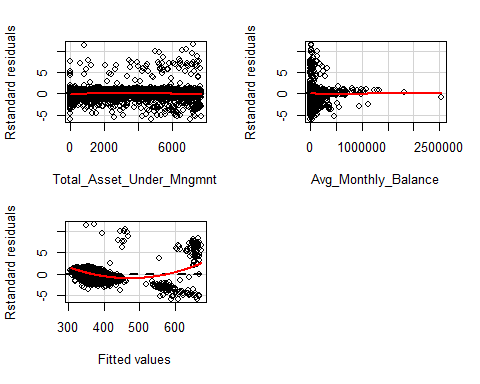
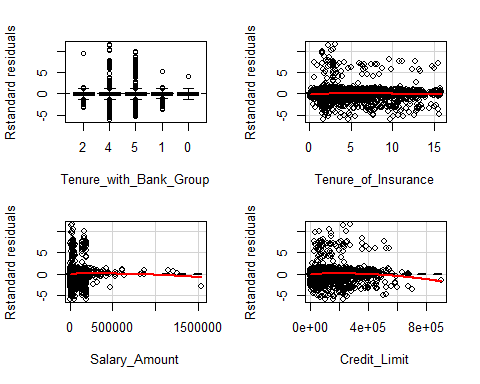
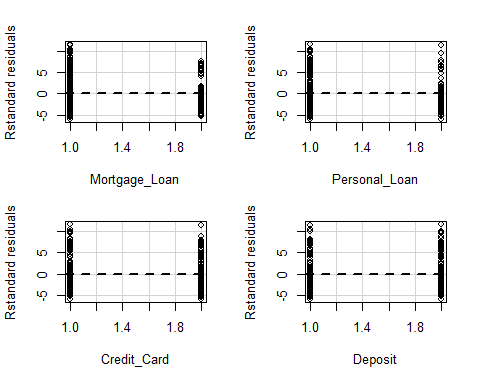
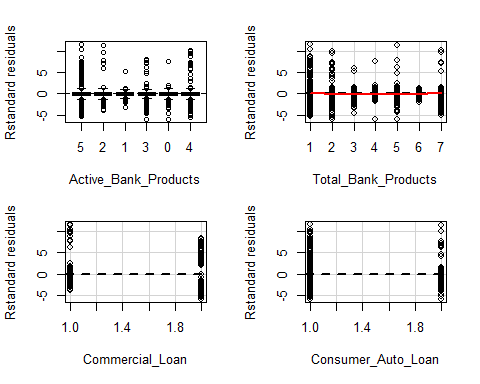
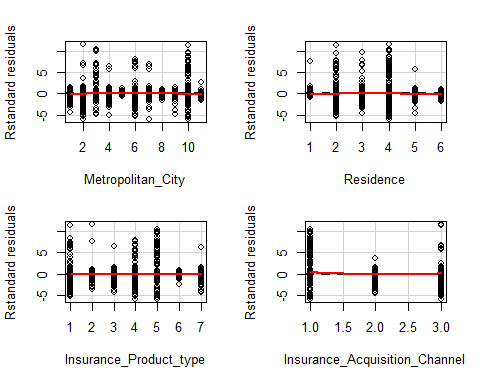
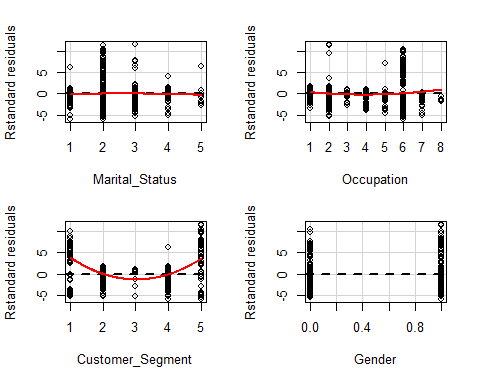
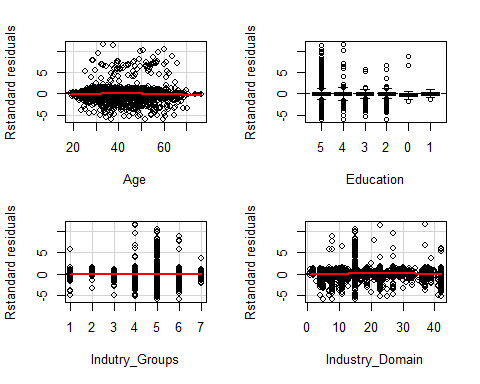
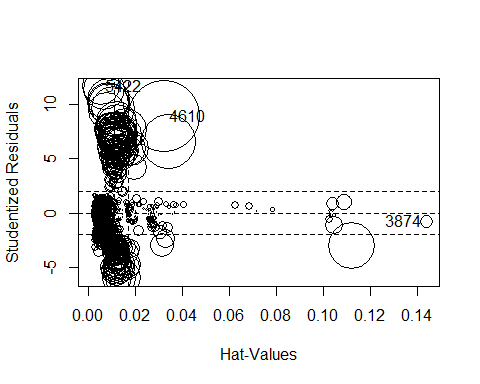
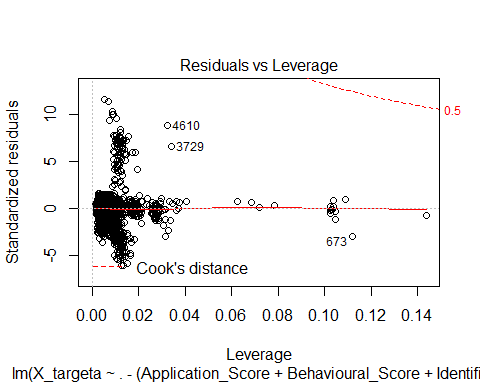
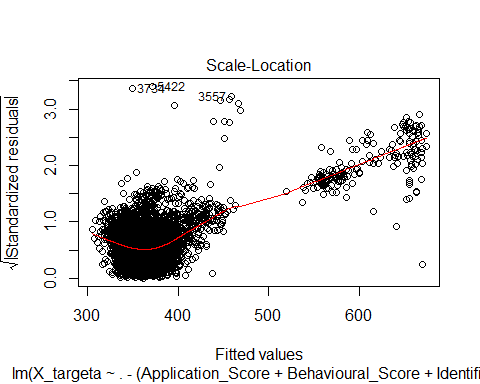
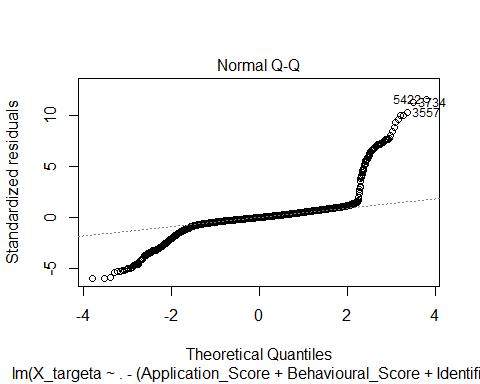
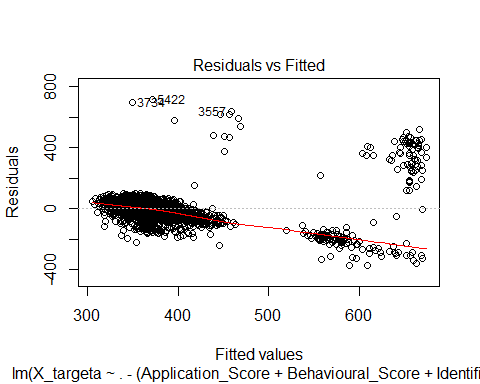
Data.imp<- complete(imp,4)  
  
X\_train <- a\_train[,":="(Salary\_Amount=Data.imp$Salary\_Amount,  
 Credit\_Limit=Data.imp$Credit\_Limit)][1:6924,]  
X\_train<-X\_train[-c(5255,which(X\_train$Application\_Score==0)),]  
  
X\_test <- a\_train[,":="(Salary\_Amount=Data.imp$Salary\_Amount,  
 Credit\_Limit=Data.imp$Credit\_Limit)][6925:9231,]  
X\_targeta <- as.numeric(X\_train$Application\_Score)  
X\_targetb <- as.numeric(X\_train$Behavioural\_Score)  
  
ma<-lm(X\_targeta~.-(Application\_Score+Behavioural\_Score+Identifier),data = X\_train)  
summary(ma)

##   
## Call:  
## lm(formula = X\_targeta ~ . - (Application\_Score + Behavioural\_Score +   
## Identifier), data = X\_train)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -369.44 -18.65 -0.56 19.75 716.08   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.823e+02 1.231e+01 14.806 < 2e-16 \*\*\*  
## Age 4.739e-02 1.210e-01 0.392 0.695197   
## Education.L -7.552e+00 6.867e+00 -1.100 0.271461   
## Education.Q 3.772e+00 7.772e+00 0.485 0.627442   
## Education.C -6.612e+00 9.656e+00 -0.685 0.493524   
## Education^4 -1.659e+01 6.065e+00 -2.735 0.006260 \*\*   
## Education^5 6.599e+00 1.169e+01 0.564 0.572489   
## Education^6 5.448e+00 1.010e+01 0.540 0.589555   
## Indutry\_Groups 6.717e-01 7.283e-01 0.922 0.356396   
## Industry\_Domain -2.033e-01 6.477e-02 -3.139 0.001700 \*\*   
## Marital\_Status -2.600e+00 1.351e+00 -1.925 0.054331 .   
## Occupation 1.257e+01 8.730e-01 14.402 < 2e-16 \*\*\*  
## Customer\_Segment -3.351e-01 1.156e+00 -0.290 0.771980   
## Gender -3.078e+00 1.715e+00 -1.795 0.072722 .   
## Metropolitan\_City 1.490e+00 2.346e-01 6.350 2.29e-10 \*\*\*  
## Residence -2.442e-02 7.281e-01 -0.034 0.973245   
## Insurance\_Product\_type 5.081e-02 4.269e-01 0.119 0.905264   
## Insurance\_Acquisition\_Channel -1.490e+01 1.284e+00 -11.605 < 2e-16 \*\*\*  
## Active\_Bank\_Products.L -3.353e-01 2.172e+00 -0.154 0.877285   
## Active\_Bank\_Products.Q 9.004e+00 2.341e+00 3.846 0.000121 \*\*\*  
## Active\_Bank\_Products.C -4.615e+00 2.560e+00 -1.803 0.071479 .   
## Active\_Bank\_Products^4 1.246e-01 2.776e+00 0.045 0.964208   
## Active\_Bank\_Products^5 2.467e+00 2.720e+00 0.907 0.364491   
## Total\_Bank\_Products -3.845e-01 4.852e-01 -0.793 0.428068   
## Commercial\_Loan 2.013e+02 5.189e+00 38.795 < 2e-16 \*\*\*  
## Consumer\_Auto\_Loan -5.914e+00 2.486e+00 -2.379 0.017389 \*   
## Mortgage\_Loan 1.436e+00 1.942e+00 0.739 0.459764   
## Personal\_Loan -5.698e+00 1.788e+00 -3.187 0.001444 \*\*   
## Credit\_Card -1.209e+00 1.575e+00 -0.768 0.442583   
## Deposit 2.311e+00 1.765e+00 1.310 0.190387   
## Tenure\_with\_Bank\_Group.L -1.601e+01 3.433e+00 -4.663 3.18e-06 \*\*\*  
## Tenure\_with\_Bank\_Group.Q -1.550e+01 3.092e+00 -5.014 5.45e-07 \*\*\*  
## Tenure\_with\_Bank\_Group.C 7.346e+00 2.288e+00 3.211 0.001330 \*\*   
## Tenure\_with\_Bank\_Group^4 8.664e+00 1.864e+00 4.649 3.40e-06 \*\*\*  
## Tenure\_of\_Insurance -8.970e-01 3.090e-01 -2.903 0.003707 \*\*   
## Salary\_Amount -4.740e-06 1.440e-05 -0.329 0.742010   
## Credit\_Limit -2.088e-05 1.189e-05 -1.756 0.079116 .   
## Total\_Asset\_Under\_Mngmnt 1.321e-03 4.080e-04 3.238 0.001211 \*\*   
## Avg\_Monthly\_Balance -4.638e-05 9.548e-06 -4.858 1.21e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 62 on 6864 degrees of freedom  
## Multiple R-squared: 0.3373, Adjusted R-squared: 0.3336   
## F-statistic: 91.94 on 38 and 6864 DF, p-value: < 2.2e-16

plot(ma)



dia(ma)

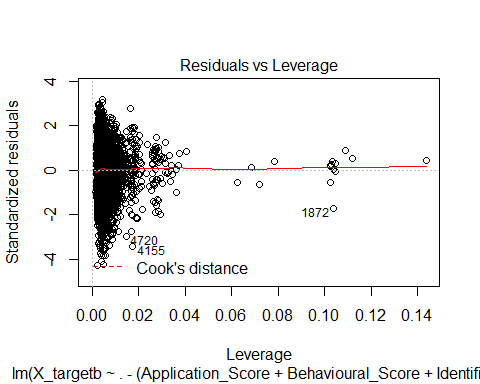
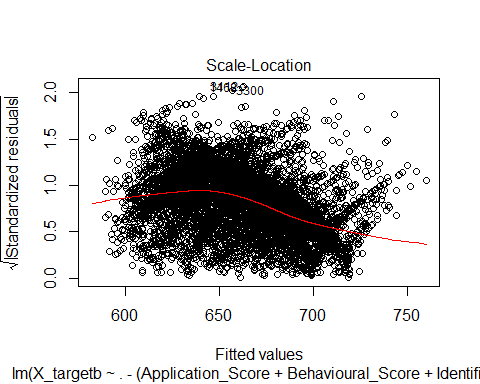
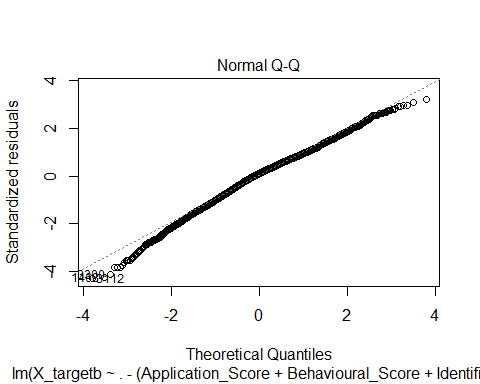
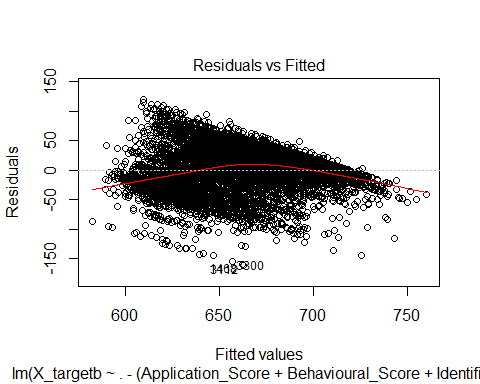


## Test stat Pr(>|t|)  
## Age -3.050 0.002  
## Education NA NA  
## Indutry\_Groups -0.493 0.622  
## Industry\_Domain -3.836 0.000  
## Marital\_Status -2.398 0.016  
## Occupation 7.570 0.000  
## Customer\_Segment 71.560 0.000  
## Gender -0.159 0.874  
## Metropolitan\_City -0.624 0.533  
## Residence -2.344 0.019  
## Insurance\_Product\_type -0.256 0.798  
## Insurance\_Acquisition\_Channel 11.180 0.000  
## Active\_Bank\_Products NA NA  
## Total\_Bank\_Products 2.417 0.016  
## Commercial\_Loan -0.230 0.818  
## Consumer\_Auto\_Loan -0.237 0.813  
## Mortgage\_Loan -0.254 0.800  
## Personal\_Loan -0.245 0.806  
## Credit\_Card -0.234 0.815  
## Deposit -0.255 0.799  
## Tenure\_with\_Bank\_Group NA NA  
## Tenure\_of\_Insurance -1.682 0.093  
## Salary\_Amount -1.470 0.142  
## Credit\_Limit -5.061 0.000  
## Total\_Asset\_Under\_Mngmnt -2.852 0.004  
## Avg\_Monthly\_Balance 0.068 0.946  
## Tukey test 56.941 0.000

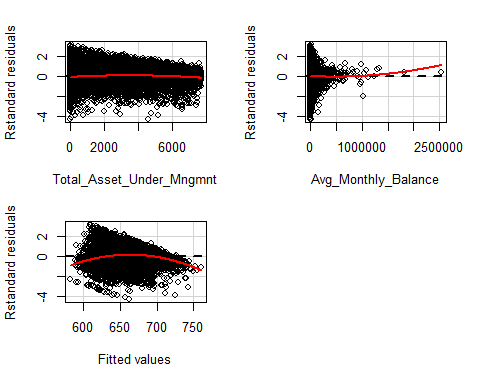
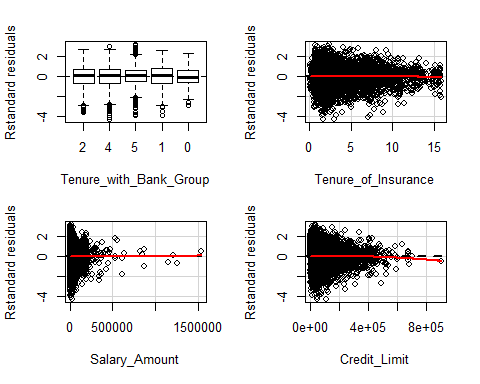
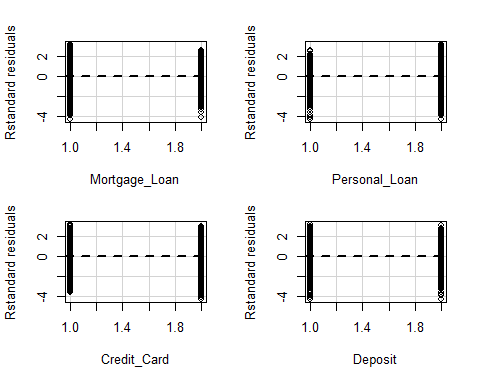
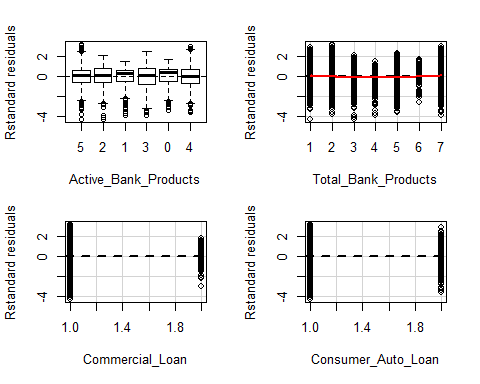
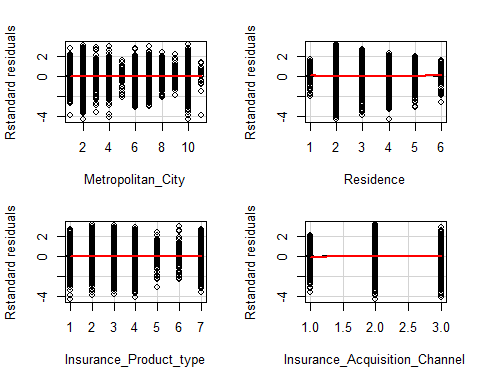
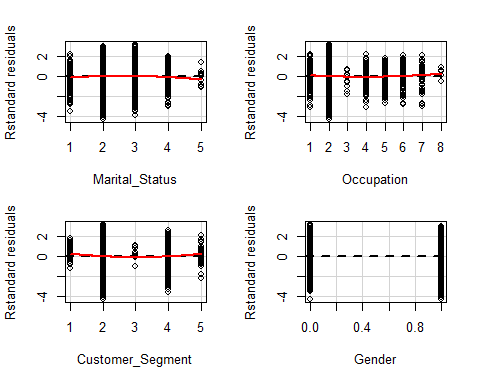
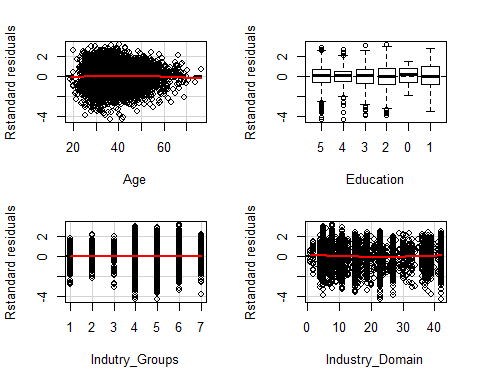
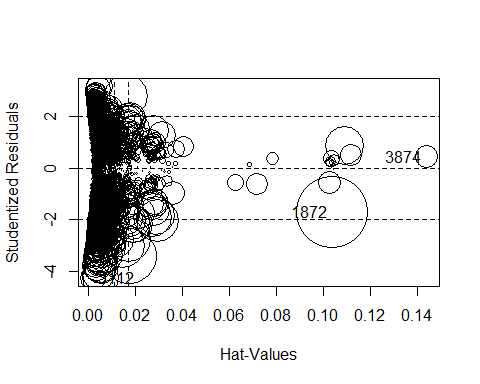
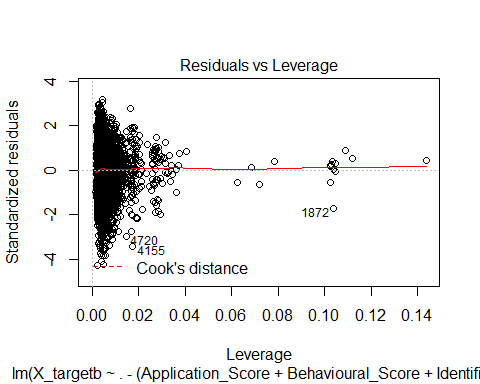
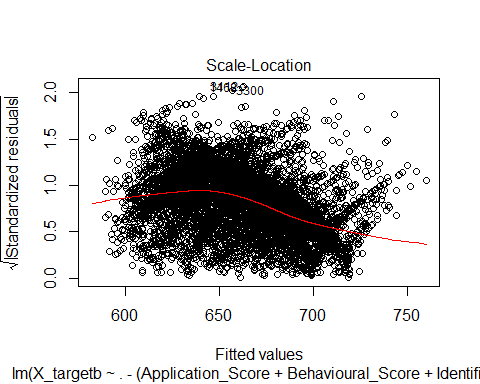
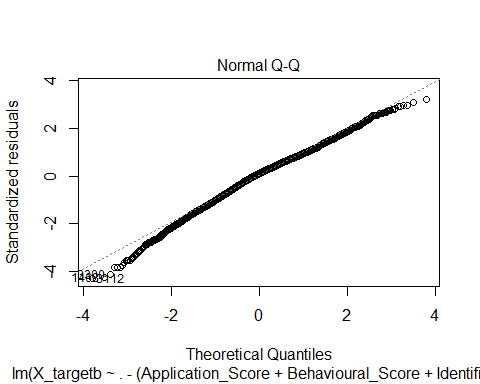
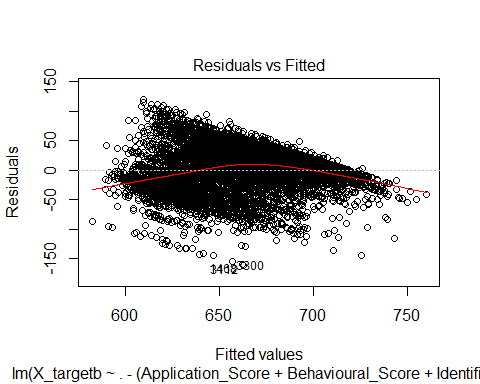
mb<-lm(X\_targetb~.-(Application\_Score+Behavioural\_Score+Identifier),data = X\_train)  
summary(mb)

##   
## Call:  
## lm(formula = X\_targetb ~ . - (Application\_Score + Behavioural\_Score +   
## Identifier), data = X\_train)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -160.608 -24.262 3.931 25.216 119.501   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6.542e+02 7.469e+00 87.582 < 2e-16 \*\*\*  
## Age 1.156e-01 7.336e-02 1.576 0.115068   
## Education.L -3.905e+00 4.165e+00 -0.938 0.348433   
## Education.Q -9.449e+00 4.714e+00 -2.004 0.045062 \*   
## Education.C -6.941e+00 5.857e+00 -1.185 0.235983   
## Education^4 6.258e-01 3.679e+00 0.170 0.864919   
## Education^5 1.296e+01 7.091e+00 1.828 0.067557 .   
## Education^6 6.892e+00 6.125e+00 1.125 0.260540   
## Indutry\_Groups -2.312e-01 4.417e-01 -0.523 0.600658   
## Industry\_Domain -1.652e-01 3.928e-02 -4.204 2.65e-05 \*\*\*  
## Marital\_Status -1.593e+00 8.193e-01 -1.944 0.051936 .   
## Occupation -3.348e-01 5.295e-01 -0.632 0.527209   
## Customer\_Segment 6.007e-01 7.012e-01 0.857 0.391698   
## Gender 1.006e+00 1.040e+00 0.967 0.333400   
## Metropolitan\_City 2.332e-01 1.423e-01 1.639 0.101244   
## Residence 2.065e+00 4.416e-01 4.677 2.97e-06 \*\*\*  
## Insurance\_Product\_type 3.596e-02 2.589e-01 0.139 0.889560   
## Insurance\_Acquisition\_Channel -2.970e+00 7.787e-01 -3.814 0.000138 \*\*\*  
## Active\_Bank\_Products.L 1.191e+01 1.317e+00 9.041 < 2e-16 \*\*\*  
## Active\_Bank\_Products.Q -4.150e+01 1.420e+00 -29.230 < 2e-16 \*\*\*  
## Active\_Bank\_Products.C 2.731e+00 1.553e+00 1.758 0.078731 .   
## Active\_Bank\_Products^4 -2.486e+01 1.684e+00 -14.766 < 2e-16 \*\*\*  
## Active\_Bank\_Products^5 -3.462e+01 1.650e+00 -20.984 < 2e-16 \*\*\*  
## Total\_Bank\_Products -1.155e+00 2.943e-01 -3.926 8.73e-05 \*\*\*  
## Commercial\_Loan 6.108e+00 3.147e+00 1.940 0.052367 .   
## Consumer\_Auto\_Loan -3.772e+00 1.508e+00 -2.502 0.012390 \*   
## Mortgage\_Loan -1.597e+00 1.178e+00 -1.355 0.175344   
## Personal\_Loan -1.445e+01 1.084e+00 -13.322 < 2e-16 \*\*\*  
## Credit\_Card 3.322e+00 9.553e-01 3.478 0.000509 \*\*\*  
## Deposit 5.923e+00 1.070e+00 5.534 3.25e-08 \*\*\*  
## Tenure\_with\_Bank\_Group.L -1.385e+01 2.082e+00 -6.652 3.11e-11 \*\*\*  
## Tenure\_with\_Bank\_Group.Q -1.324e+01 1.875e+00 -7.059 1.85e-12 \*\*\*  
## Tenure\_with\_Bank\_Group.C 2.303e+00 1.388e+00 1.660 0.096991 .   
## Tenure\_with\_Bank\_Group^4 4.322e+00 1.130e+00 3.824 0.000133 \*\*\*  
## Tenure\_of\_Insurance 2.704e-01 1.874e-01 1.443 0.149143   
## Salary\_Amount -1.364e-05 8.732e-06 -1.562 0.118313   
## Credit\_Limit 1.258e-05 7.210e-06 1.745 0.081069 .   
## Total\_Asset\_Under\_Mngmnt 1.025e-02 2.475e-04 41.414 < 2e-16 \*\*\*  
## Avg\_Monthly\_Balance -1.186e-05 5.791e-06 -2.048 0.040563 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 37.6 on 6864 degrees of freedom  
## Multiple R-squared: 0.3922, Adjusted R-squared: 0.3889   
## F-statistic: 116.6 on 38 and 6864 DF, p-value: < 2.2e-16

plot(mb)



dia(mb)



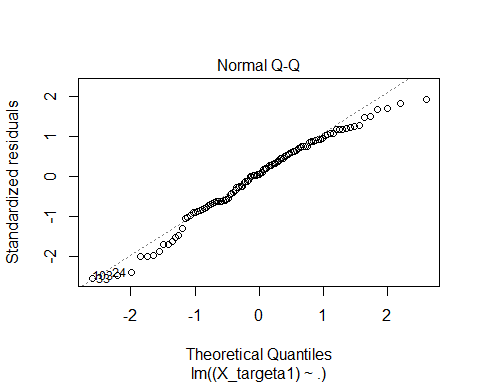
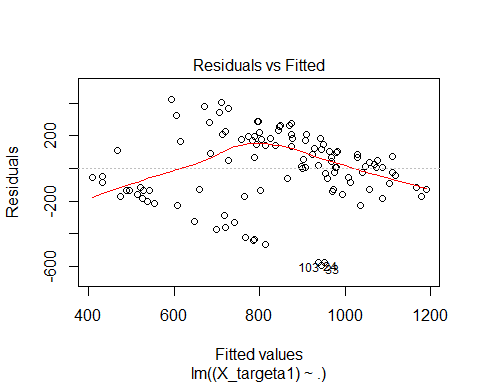
## Test stat Pr(>|t|)  
## Age -1.764 0.078  
## Education NA NA  
## Indutry\_Groups -0.602 0.547  
## Industry\_Domain 4.689 0.000  
## Marital\_Status -2.835 0.005  
## Occupation 1.898 0.058  
## Customer\_Segment 3.544 0.000  
## Gender 1.440 0.150  
## Metropolitan\_City 0.019 0.985  
## Residence 1.861 0.063  
## Insurance\_Product\_type 0.331 0.741  
## Insurance\_Acquisition\_Channel -2.889 0.004  
## Active\_Bank\_Products NA NA  
## Total\_Bank\_Products 8.208 0.000  
## Commercial\_Loan -1.119 0.263  
## Consumer\_Auto\_Loan -1.115 0.265  
## Mortgage\_Loan -1.105 0.269  
## Personal\_Loan -1.102 0.270  
## Credit\_Card -1.118 0.263  
## Deposit -1.111 0.267  
## Tenure\_with\_Bank\_Group NA NA  
## Tenure\_of\_Insurance -1.228 0.219  
## Salary\_Amount 0.116 0.907  
## Credit\_Limit -1.310 0.190  
## Total\_Asset\_Under\_Mngmnt -6.353 0.000  
## Avg\_Monthly\_Balance 1.560 0.119  
## Tukey test -14.749 0.000

X\_train1<-X\_train[X\_train$Customer\_Segment %in% c(1,5)]  
X\_train2<-X\_train[X\_train$Customer\_Segment %in% c(2,3,4)]  
X\_targeta1 <- as.numeric(X\_train1$Application\_Score)  
X\_targetb1 <- as.numeric(X\_train1$Behavioural\_Score)  
X\_targeta2 <- as.numeric(X\_train2$Application\_Score)  
X\_targetb2 <- as.numeric(X\_train2$Behavioural\_Score)  
  
I<-X\_train$Identifier  
I1<-X\_train1$Identifier  
I2<-X\_train2$Identifier  
  
X\_train<-X\_train[,":="(Application\_Score=NULL,Behavioural\_Score=NULL,Identifier=NULL)]  
X\_train1<-X\_train1[,":="(Application\_Score=NULL,Behavioural\_Score=NULL,Identifier=NULL)]  
X\_train2<-X\_train2[,":="(Application\_Score=NULL,Behavioural\_Score=NULL,Identifier=NULL)]  
  
X\_test1<-X\_test[X\_test$Customer\_Segment %in% c(1,5)]  
X\_test2<-X\_test[X\_test$Customer\_Segment %in% c(2,3,4)]  
  
X\_test1 <- X\_test1[, .SDcols=names(X\_train)]  
X\_test2 <- X\_test2[, .SDcols=names(X\_train)]  
  
ma1<-lm((X\_targeta1)~.,data = X\_train1)  
summary(ma1)

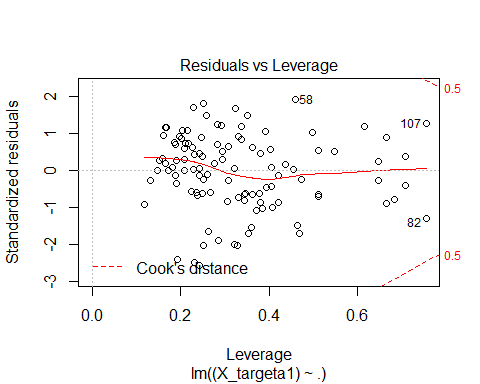
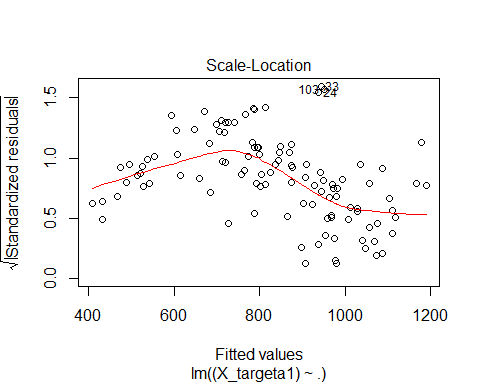
##   
## Call:  
## lm(formula = (X\_targeta1) ~ ., data = X\_train1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -593.4 -134.2 14.2 168.9 422.3   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.971e+03 6.812e+02 2.894 0.00502 \*\*  
## Age 3.175e+00 3.198e+00 0.993 0.32399   
## Education.L 1.670e+02 1.543e+02 1.082 0.28263   
## Education.Q -1.149e+02 1.850e+02 -0.621 0.53669   
## Education.C -3.750e+01 1.509e+02 -0.248 0.80445   
## Education^4 1.704e+02 2.039e+02 0.836 0.40588   
## Indutry\_Groups -8.022e+01 6.307e+01 -1.272 0.20744   
## Industry\_Domain 4.166e+00 4.521e+00 0.922 0.35979   
## Marital\_Status 6.195e+01 6.409e+01 0.967 0.33696   
## Occupation -1.862e+01 3.544e+01 -0.526 0.60078   
## Customer\_Segment -1.493e+01 2.006e+01 -0.745 0.45889   
## Gender -4.672e+01 6.658e+01 -0.702 0.48508   
## Metropolitan\_City -8.169e+00 9.503e+00 -0.860 0.39281   
## Residence -3.030e+01 4.021e+01 -0.754 0.45356   
## Insurance\_Product\_type -2.495e+01 2.030e+01 -1.229 0.22300   
## Insurance\_Acquisition\_Channel -2.348e+02 7.279e+01 -3.226 0.00188 \*\*  
## Active\_Bank\_Products.L -7.005e+01 9.287e+01 -0.754 0.45308   
## Active\_Bank\_Products.Q 2.852e+02 1.275e+02 2.237 0.02833 \*   
## Active\_Bank\_Products.C 2.213e+00 1.297e+02 0.017 0.98644   
## Active\_Bank\_Products^4 -1.019e+02 1.446e+02 -0.705 0.48314   
## Active\_Bank\_Products^5 4.494e+02 1.816e+02 2.475 0.01563 \*   
## Total\_Bank\_Products -7.092e-01 2.398e+01 -0.030 0.97649   
## Commercial\_Loan -9.372e+01 1.075e+02 -0.872 0.38611   
## Consumer\_Auto\_Loan -6.401e+01 8.867e+01 -0.722 0.47266   
## Mortgage\_Loan -8.738e+01 8.820e+01 -0.991 0.32512   
## Personal\_Loan -2.308e+01 9.328e+01 -0.247 0.80524   
## Credit\_Card -6.187e+01 6.148e+01 -1.006 0.31755   
## Deposit 7.998e+01 7.600e+01 1.052 0.29605   
## Tenure\_with\_Bank\_Group.L 2.778e+02 2.962e+02 0.938 0.35151   
## Tenure\_with\_Bank\_Group.Q -2.726e+02 2.299e+02 -1.186 0.23953   
## Tenure\_with\_Bank\_Group.C -5.682e+01 1.935e+02 -0.294 0.76989   
## Tenure\_with\_Bank\_Group^4 -8.911e+01 1.349e+02 -0.660 0.51108   
## Tenure\_of\_Insurance 7.797e-01 9.737e+00 0.080 0.93640   
## Salary\_Amount -7.006e-04 6.279e-04 -1.116 0.26816   
## Credit\_Limit 4.039e-04 3.837e-04 1.053 0.29598   
## Total\_Asset\_Under\_Mngmnt 2.204e-03 1.609e-02 0.137 0.89143   
## Avg\_Monthly\_Balance -9.323e-04 4.571e-04 -2.040 0.04501 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 267.8 on 73 degrees of freedom  
## Multiple R-squared: 0.4258, Adjusted R-squared: 0.1426   
## F-statistic: 1.503 on 36 and 73 DF, p-value: 0.07074

plot(ma1)

## Warning: not plotting observations with leverage one:  
## 59

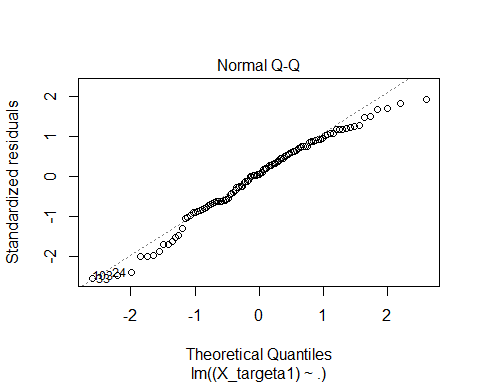
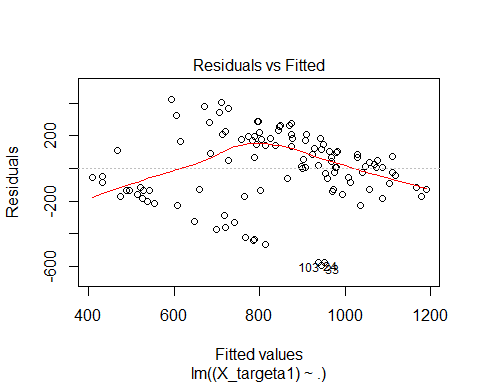


## Warning: not plotting observations with leverage one:  
## 59

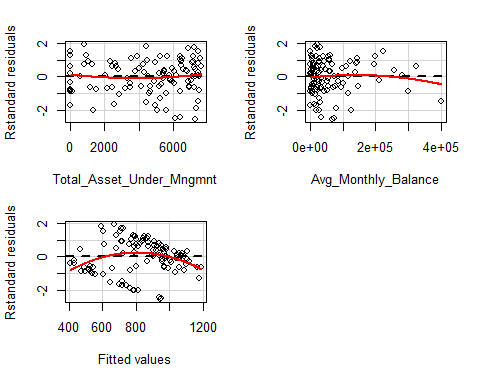
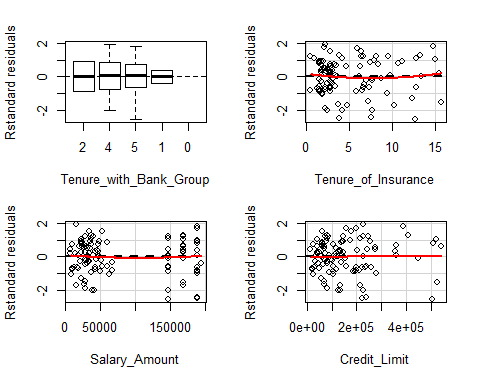
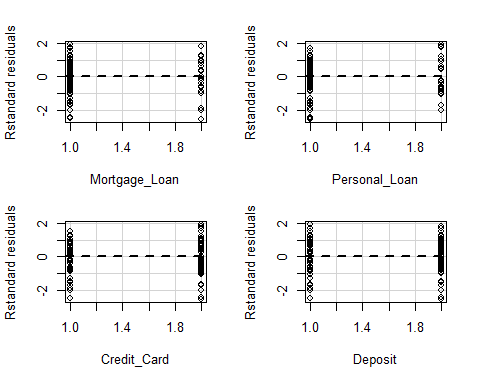
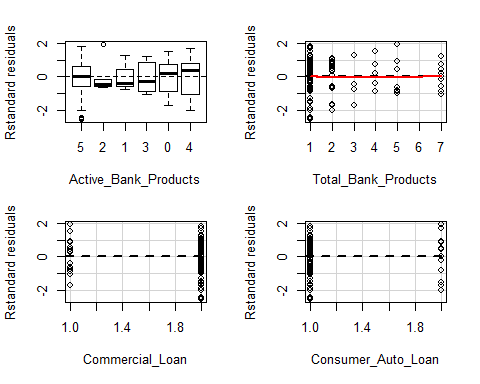
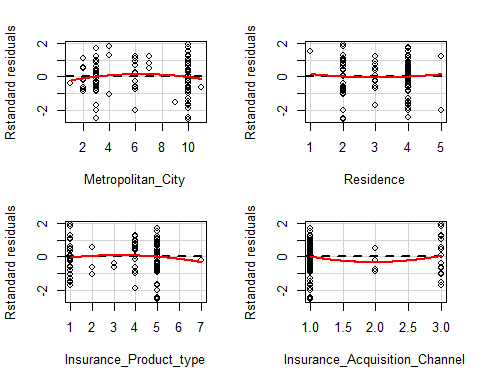
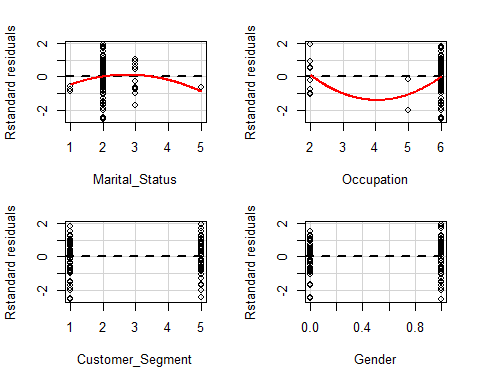
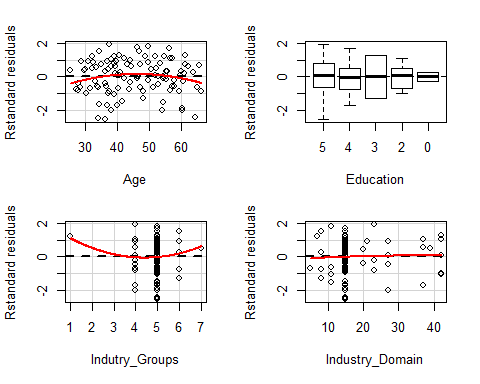
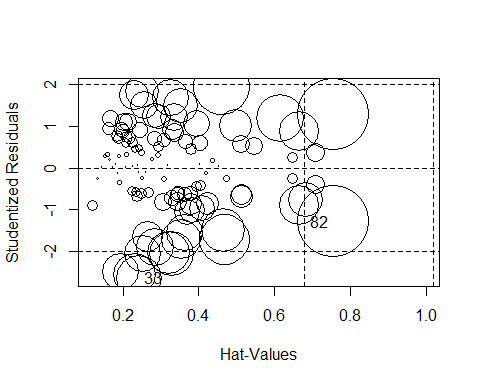
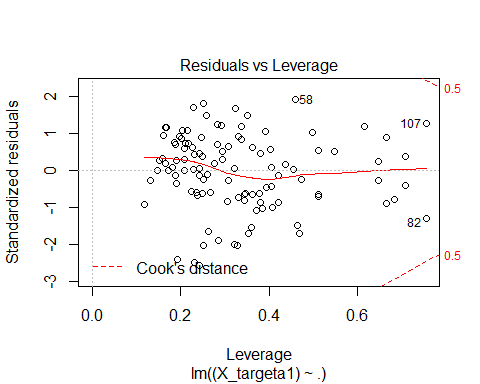
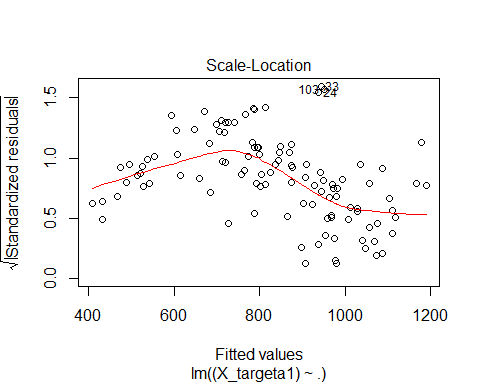


dia(ma1)

## Warning: not plotting observations with leverage one:  
## 59



## Warning: not plotting observations with leverage one:  
## 59

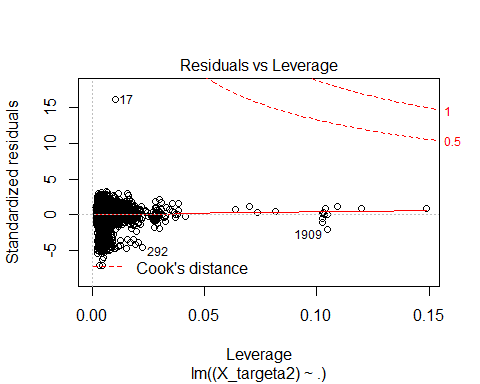
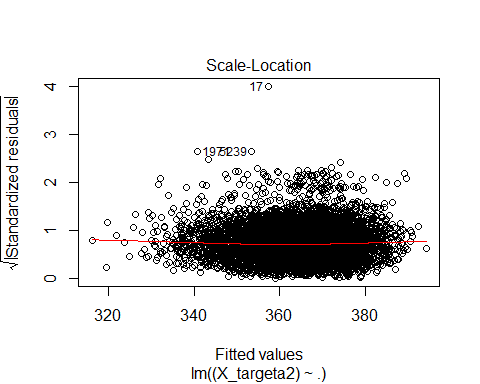
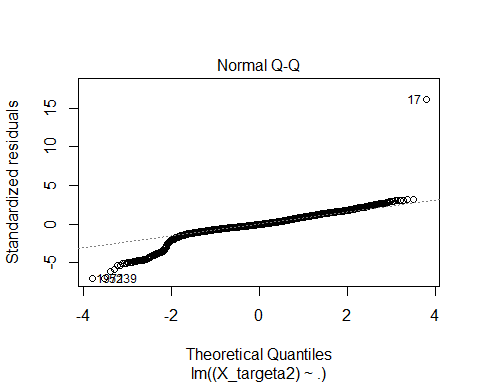
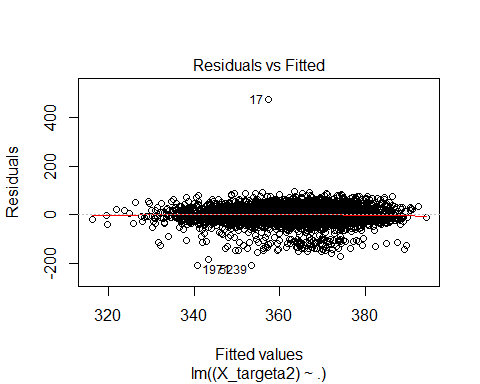


## Test stat Pr(>|t|)  
## Age -1.825 0.072  
## Education NA NA  
## Indutry\_Groups 1.146 0.256  
## Industry\_Domain -0.179 0.858  
## Marital\_Status -1.028 0.308  
## Occupation 1.555 0.124  
## Customer\_Segment -0.036 0.971  
## Gender 1.588 0.117  
## Metropolitan\_City -0.813 0.419  
## Residence 0.137 0.891  
## Insurance\_Product\_type -0.390 0.698  
## Insurance\_Acquisition\_Channel 0.760 0.450  
## Active\_Bank\_Products NA NA  
## Total\_Bank\_Products 0.482 0.631  
## Commercial\_Loan 1.544 0.127  
## Consumer\_Auto\_Loan 0.121 0.904  
## Mortgage\_Loan -0.009 0.993  
## Personal\_Loan -0.091 0.928  
## Credit\_Card -0.240 0.811  
## Deposit 0.175 0.861  
## Tenure\_with\_Bank\_Group NA NA  
## Tenure\_of\_Insurance 0.746 0.458  
## Salary\_Amount 0.438 0.663  
## Credit\_Limit -0.054 0.957  
## Total\_Asset\_Under\_Mngmnt 0.815 0.418  
## Avg\_Monthly\_Balance -0.664 0.509  
## Tukey test -2.740 0.006

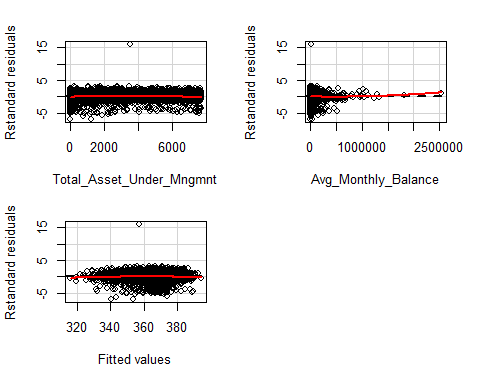
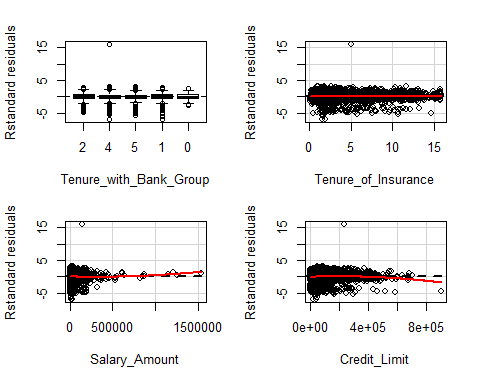
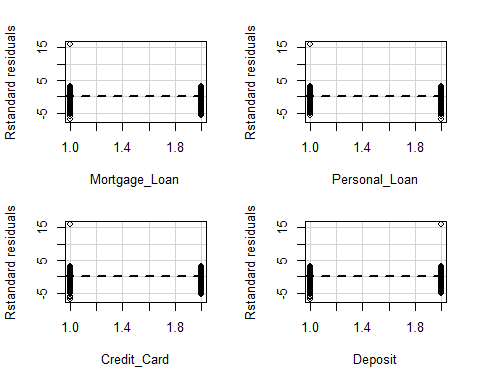
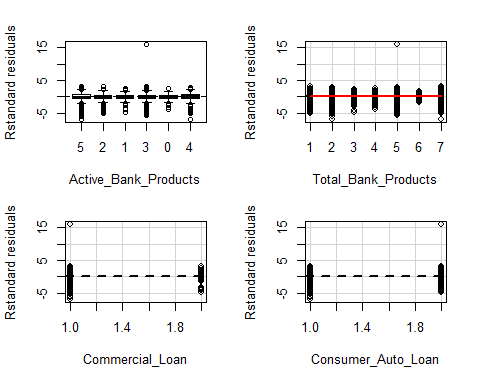
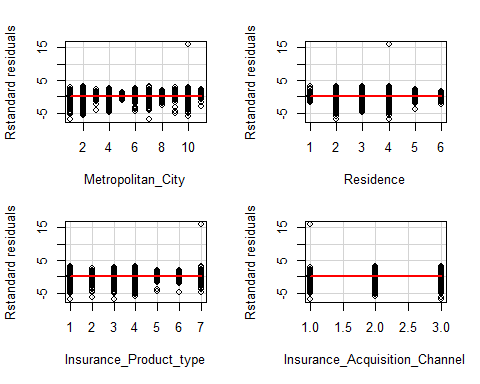
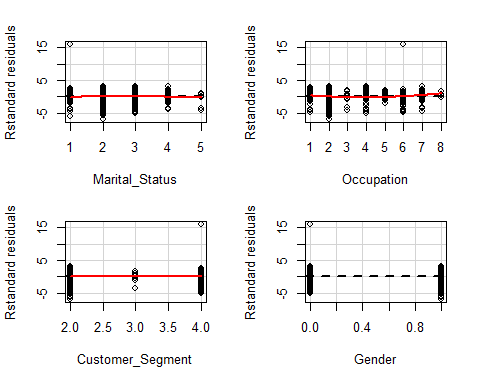
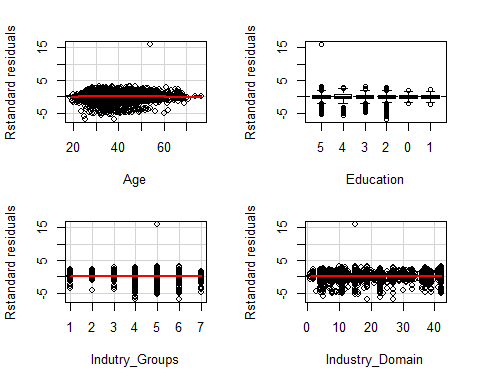
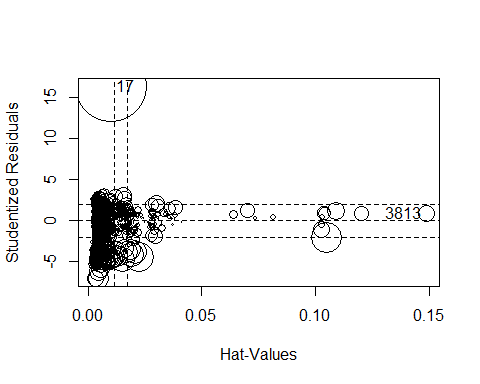
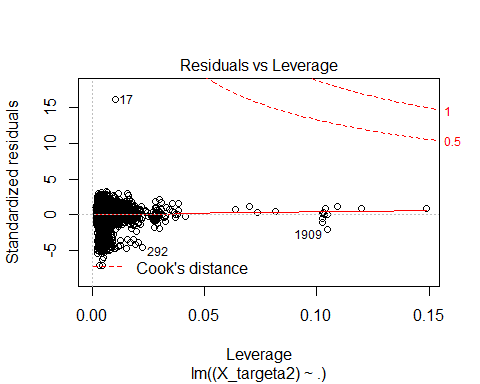
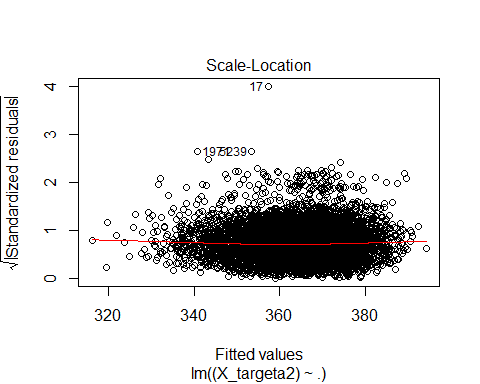
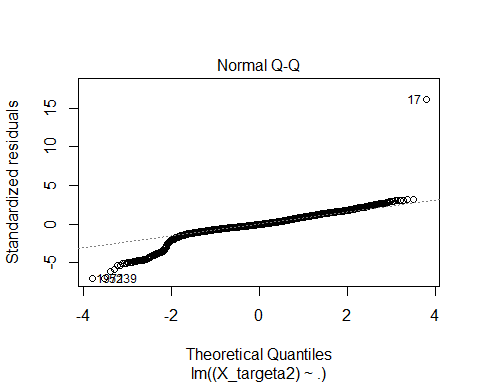
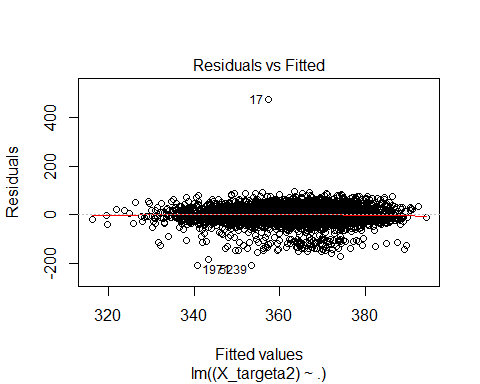
ma2<-lm((X\_targeta2)~.,data = X\_train2)  
summary(ma2)

##   
## Call:  
## lm(formula = (X\_targeta2) ~ ., data = X\_train2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -205.81 -13.97 -1.18 16.40 471.72   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.396e+02 6.150e+00 55.224 < 2e-16 \*\*\*  
## Age 2.986e-01 5.891e-02 5.068 4.12e-07 \*\*\*  
## Education.L -8.387e+00 3.293e+00 -2.547 0.010891 \*   
## Education.Q 4.432e+00 3.700e+00 1.198 0.231068   
## Education.C 4.140e+00 4.617e+00 0.897 0.369904   
## Education^4 -6.350e+00 2.940e+00 -2.160 0.030835 \*   
## Education^5 5.757e+00 5.581e+00 1.032 0.302315   
## Education^6 3.444e+00 4.811e+00 0.716 0.474127   
## Indutry\_Groups -1.019e+00 3.480e-01 -2.929 0.003409 \*\*   
## Industry\_Domain -1.149e-02 3.100e-02 -0.371 0.710897   
## Marital\_Status -1.245e+00 6.470e-01 -1.925 0.054272 .   
## Occupation -3.846e+00 4.431e-01 -8.680 < 2e-16 \*\*\*  
## Customer\_Segment -8.382e-02 6.179e-01 -0.136 0.892095   
## Gender 4.446e-03 8.237e-01 0.005 0.995693   
## Metropolitan\_City 2.467e-01 1.132e-01 2.178 0.029420 \*   
## Residence -1.289e-01 3.487e-01 -0.370 0.711710   
## Insurance\_Product\_type -6.559e-02 2.046e-01 -0.321 0.748504   
## Insurance\_Acquisition\_Channel 8.194e-02 6.263e-01 0.131 0.895920   
## Active\_Bank\_Products.L 2.418e+00 1.043e+00 2.319 0.020442 \*   
## Active\_Bank\_Products.Q -1.096e+00 1.124e+00 -0.976 0.329223   
## Active\_Bank\_Products.C -4.751e+00 1.227e+00 -3.873 0.000108 \*\*\*  
## Active\_Bank\_Products^4 -4.915e+00 1.330e+00 -3.696 0.000221 \*\*\*  
## Active\_Bank\_Products^5 -4.497e+00 1.303e+00 -3.451 0.000563 \*\*\*  
## Total\_Bank\_Products 2.534e-01 2.327e-01 1.089 0.276174   
## Commercial\_Loan -3.304e+00 3.210e+00 -1.029 0.303445   
## Consumer\_Auto\_Loan 2.745e+00 1.197e+00 2.294 0.021849 \*   
## Mortgage\_Loan 6.438e+00 9.339e-01 6.894 5.92e-12 \*\*\*  
## Personal\_Loan 4.931e+00 8.607e-01 5.730 1.05e-08 \*\*\*  
## Credit\_Card 1.381e+00 7.568e-01 1.825 0.068069 .   
## Deposit 2.438e+00 8.471e-01 2.878 0.004012 \*\*   
## Tenure\_with\_Bank\_Group.L -1.313e+01 1.639e+00 -8.013 1.31e-15 \*\*\*  
## Tenure\_with\_Bank\_Group.Q -1.072e+01 1.478e+00 -7.254 4.50e-13 \*\*\*  
## Tenure\_with\_Bank\_Group.C 2.753e+00 1.092e+00 2.521 0.011740 \*   
## Tenure\_with\_Bank\_Group^4 3.674e+00 8.943e-01 4.109 4.03e-05 \*\*\*  
## Tenure\_of\_Insurance -1.148e-01 1.492e-01 -0.769 0.441645   
## Salary\_Amount -6.533e-06 6.930e-06 -0.943 0.345799   
## Credit\_Limit 1.183e-06 5.885e-06 0.201 0.840654   
## Total\_Asset\_Under\_Mngmnt 5.987e-04 1.960e-04 3.054 0.002264 \*\*   
## Avg\_Monthly\_Balance 9.407e-07 4.613e-06 0.204 0.838404   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 29.51 on 6754 degrees of freedom  
## Multiple R-squared: 0.1102, Adjusted R-squared: 0.1052   
## F-statistic: 22 on 38 and 6754 DF, p-value: < 2.2e-16

plot(ma2)



dia(ma2)



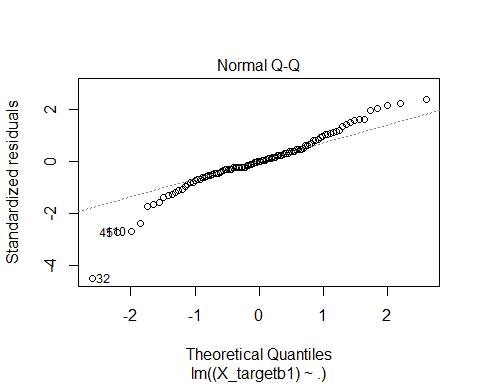
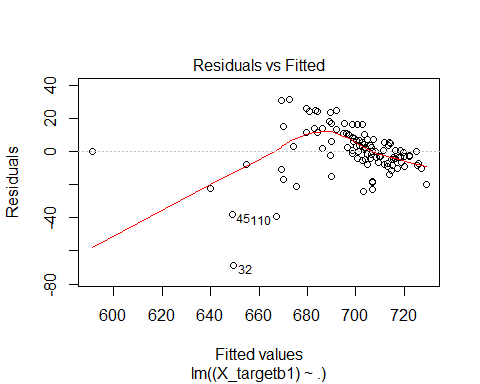
## Test stat Pr(>|t|)  
## Age -1.193 0.233  
## Education NA NA  
## Indutry\_Groups 0.631 0.528  
## Industry\_Domain 0.010 0.992  
## Marital\_Status -1.523 0.128  
## Occupation 7.368 0.000  
## Customer\_Segment -0.217 0.828  
## Gender -0.302 0.763  
## Metropolitan\_City -0.716 0.474  
## Residence -0.350 0.727  
## Insurance\_Product\_type 1.481 0.139  
## Insurance\_Acquisition\_Channel 1.217 0.224  
## Active\_Bank\_Products NA NA  
## Total\_Bank\_Products -0.097 0.923  
## Commercial\_Loan 0.124 0.901  
## Consumer\_Auto\_Loan 0.157 0.875  
## Mortgage\_Loan 0.134 0.894  
## Personal\_Loan 0.117 0.907  
## Credit\_Card 0.131 0.895  
## Deposit 0.132 0.895  
## Tenure\_with\_Bank\_Group NA NA  
## Tenure\_of\_Insurance 0.486 0.627  
## Salary\_Amount 2.631 0.009  
## Credit\_Limit -5.068 0.000  
## Total\_Asset\_Under\_Mngmnt -4.837 0.000  
## Avg\_Monthly\_Balance 1.759 0.079  
## Tukey test -2.970 0.003

mb1<-lm((X\_targetb1)~.,data = X\_train1)  
summary(mb1)

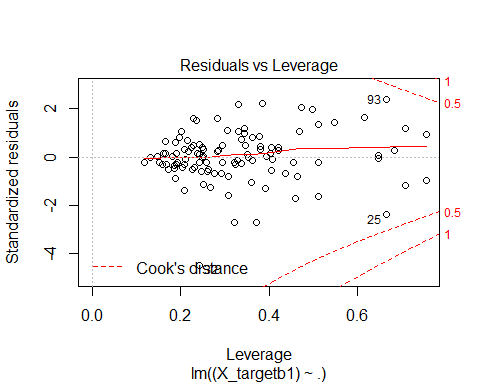
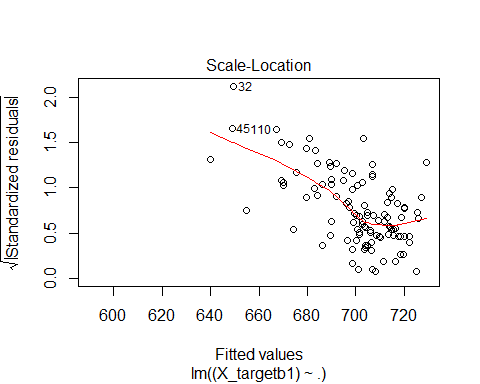
##   
## Call:  
## lm(formula = (X\_targetb1) ~ ., data = X\_train1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -68.539 -6.915 0.102 7.433 31.278   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6.893e+02 4.470e+01 15.421 < 2e-16 \*\*\*  
## Age 4.185e-03 2.098e-01 0.020 0.984141   
## Education.L 2.944e+00 1.012e+01 0.291 0.771992   
## Education.Q -1.309e+01 1.214e+01 -1.079 0.284348   
## Education.C -2.829e+00 9.902e+00 -0.286 0.775926   
## Education^4 -8.426e+00 1.338e+01 -0.630 0.530772   
## Indutry\_Groups -1.111e+00 4.138e+00 -0.268 0.789084   
## Industry\_Domain 4.420e-02 2.966e-01 0.149 0.881955   
## Marital\_Status 4.484e+00 4.205e+00 1.066 0.289865   
## Occupation -1.775e-01 2.325e+00 -0.076 0.939372   
## Customer\_Segment -3.755e-01 1.316e+00 -0.285 0.776187   
## Gender -8.026e+00 4.369e+00 -1.837 0.070273 .   
## Metropolitan\_City 2.815e-01 6.236e-01 0.452 0.652960   
## Residence 1.887e+00 2.639e+00 0.715 0.476744   
## Insurance\_Product\_type 7.919e-01 1.332e+00 0.594 0.554061   
## Insurance\_Acquisition\_Channel -8.399e+00 4.776e+00 -1.759 0.082817 .   
## Active\_Bank\_Products.L 1.310e+01 6.093e+00 2.149 0.034923 \*   
## Active\_Bank\_Products.Q -3.250e+00 8.364e+00 -0.389 0.698694   
## Active\_Bank\_Products.C -1.195e+01 8.510e+00 -1.404 0.164595   
## Active\_Bank\_Products^4 -1.165e+01 9.490e+00 -1.228 0.223489   
## Active\_Bank\_Products^5 -2.609e+00 1.191e+01 -0.219 0.827235   
## Total\_Bank\_Products -1.697e+00 1.574e+00 -1.079 0.284357   
## Commercial\_Loan 7.968e-01 7.053e+00 0.113 0.910364   
## Consumer\_Auto\_Loan -3.476e+00 5.818e+00 -0.597 0.552057   
## Mortgage\_Loan 2.820e+00 5.787e+00 0.487 0.627574   
## Personal\_Loan -1.641e+01 6.121e+00 -2.680 0.009084 \*\*   
## Credit\_Card 2.307e+00 4.034e+00 0.572 0.569236   
## Deposit -4.300e-01 4.986e+00 -0.086 0.931513   
## Tenure\_with\_Bank\_Group.L -9.838e+01 1.944e+01 -5.062 3.01e-06 \*\*\*  
## Tenure\_with\_Bank\_Group.Q -5.760e+01 1.508e+01 -3.819 0.000279 \*\*\*  
## Tenure\_with\_Bank\_Group.C -3.322e+01 1.270e+01 -2.616 0.010793 \*   
## Tenure\_with\_Bank\_Group^4 2.037e+00 8.854e+00 0.230 0.818643   
## Tenure\_of\_Insurance 6.013e-01 6.389e-01 0.941 0.349744   
## Salary\_Amount -1.174e-05 4.120e-05 -0.285 0.776551   
## Credit\_Limit -4.042e-06 2.518e-05 -0.161 0.872896   
## Total\_Asset\_Under\_Mngmnt 2.450e-03 1.056e-03 2.320 0.023136 \*   
## Avg\_Monthly\_Balance 2.892e-05 2.999e-05 0.964 0.338073   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 17.57 on 73 degrees of freedom  
## Multiple R-squared: 0.6674, Adjusted R-squared: 0.5034   
## F-statistic: 4.069 on 36 and 73 DF, p-value: 1.838e-07

plot(mb1)

## Warning: not plotting observations with leverage one:  
## 59

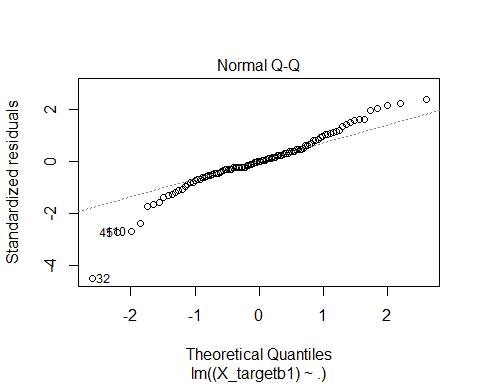
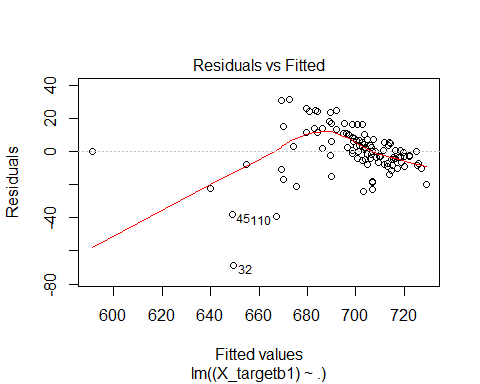


## Warning: not plotting observations with leverage one:  
## 59

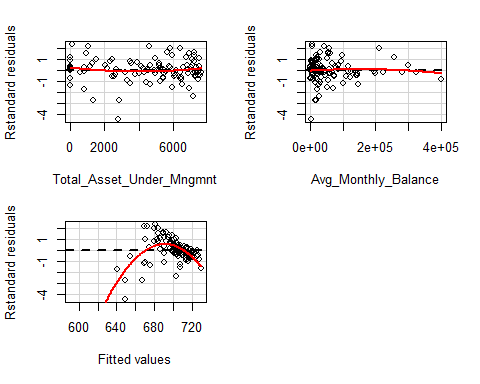
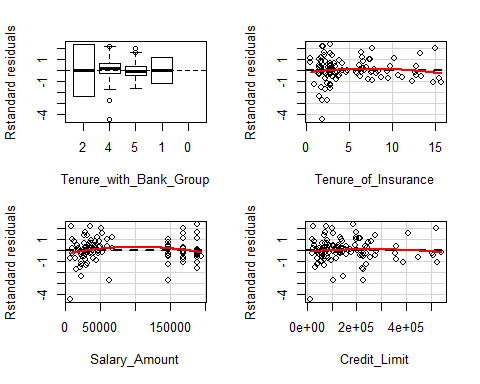
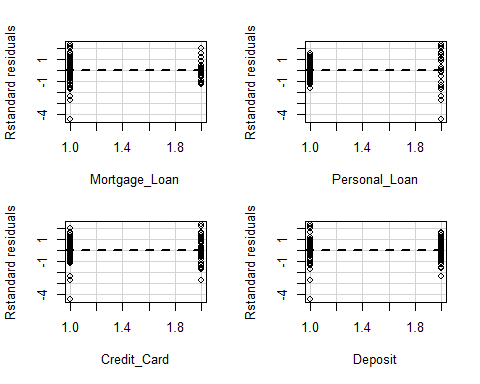
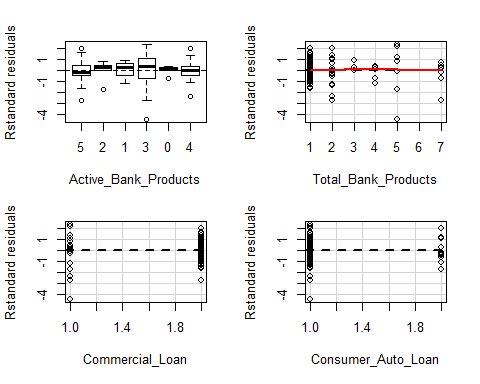
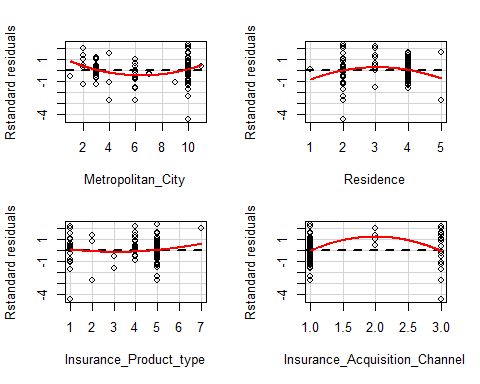
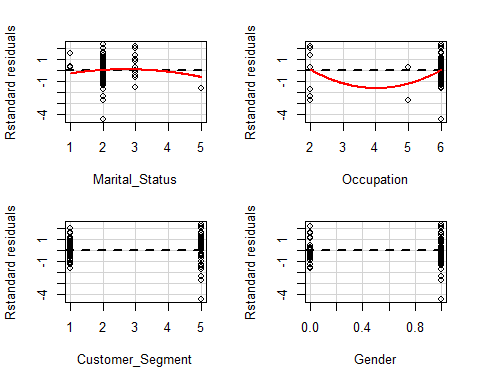
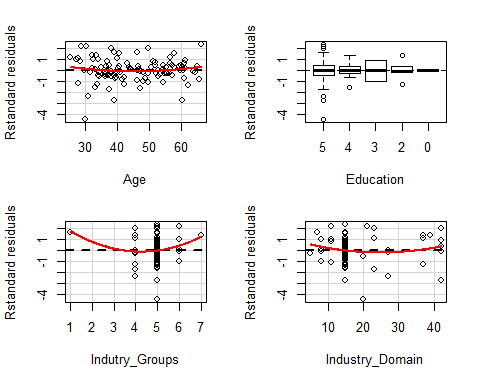
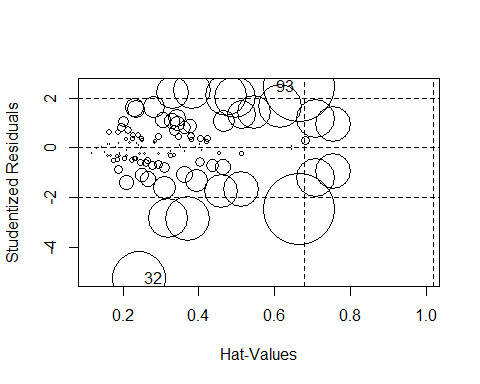
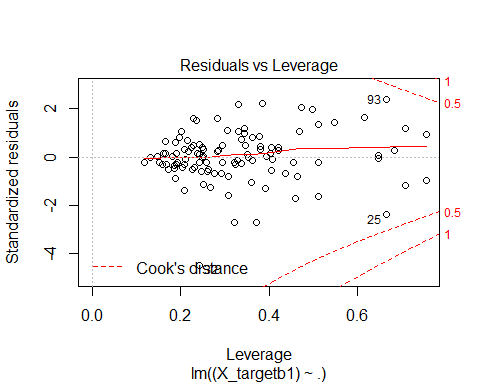
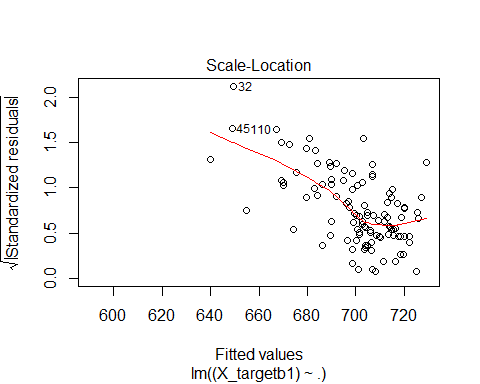


dia(mb1)

## Warning: not plotting observations with leverage one:  
## 59



## Warning: not plotting observations with leverage one:  
## 59

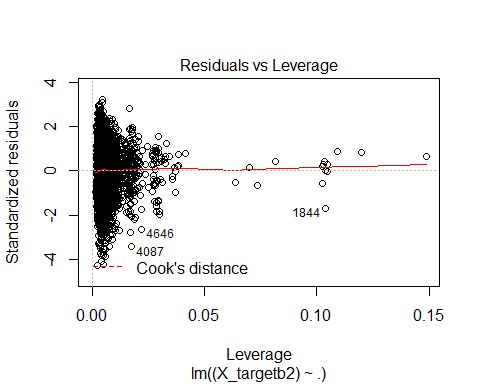
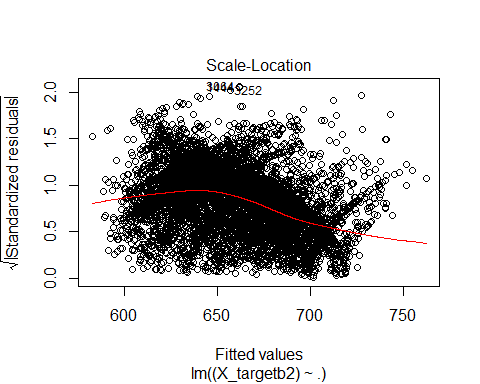
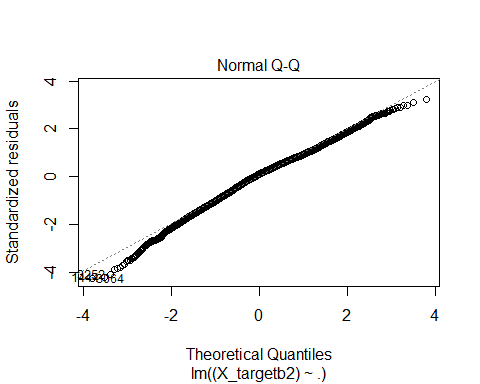
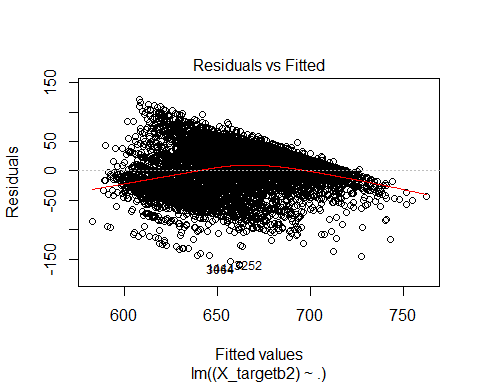


## Test stat Pr(>|t|)  
## Age 0.877 0.383  
## Education NA NA  
## Indutry\_Groups 2.048 0.044  
## Industry\_Domain 1.639 0.106  
## Marital\_Status -0.581 0.563  
## Occupation 1.824 0.072  
## Customer\_Segment 0.513 0.609  
## Gender 0.896 0.373  
## Metropolitan\_City 2.325 0.023  
## Residence -1.774 0.080  
## Insurance\_Product\_type 0.949 0.346  
## Insurance\_Acquisition\_Channel -2.629 0.010  
## Active\_Bank\_Products NA NA  
## Total\_Bank\_Products -0.187 0.852  
## Commercial\_Loan -0.523 0.602  
## Consumer\_Auto\_Loan -0.579 0.564  
## Mortgage\_Loan -0.392 0.696  
## Personal\_Loan -0.878 0.383  
## Credit\_Card -0.652 0.516  
## Deposit -0.377 0.708  
## Tenure\_with\_Bank\_Group NA NA  
## Tenure\_of\_Insurance -0.955 0.343  
## Salary\_Amount -1.495 0.139  
## Credit\_Limit -0.810 0.421  
## Total\_Asset\_Under\_Mngmnt 1.313 0.193  
## Avg\_Monthly\_Balance -0.579 0.564  
## Tukey test -8.318 0.000

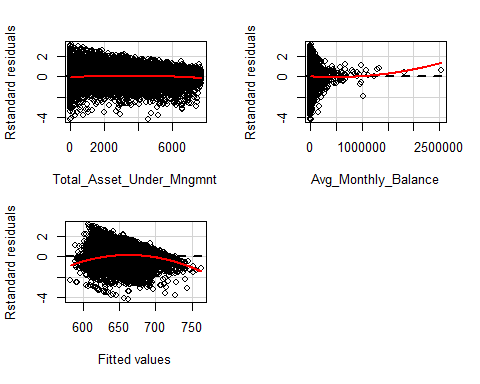
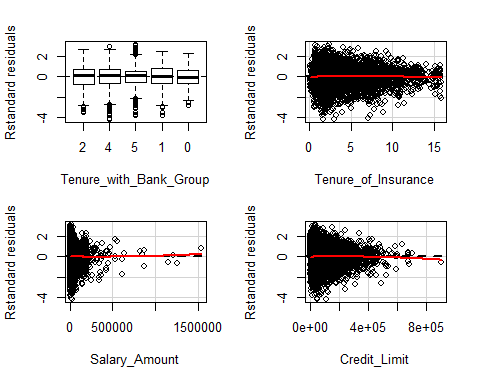
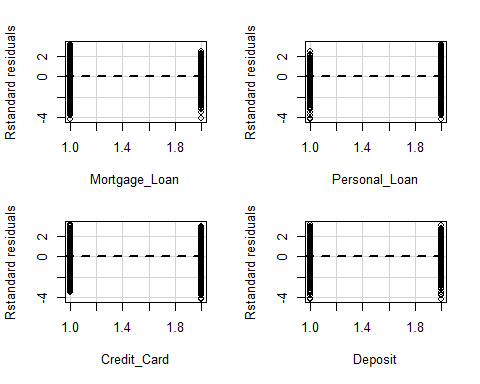
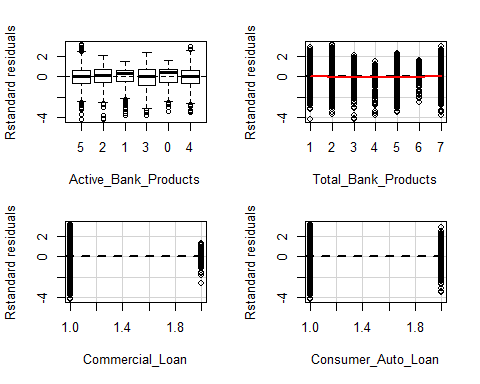
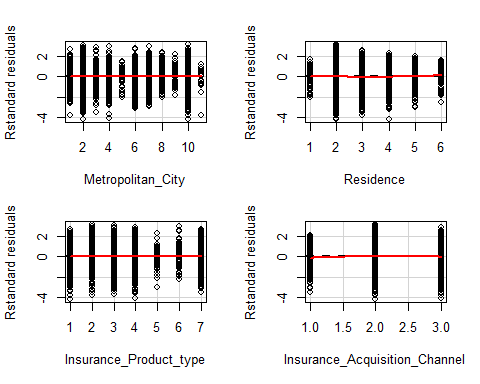
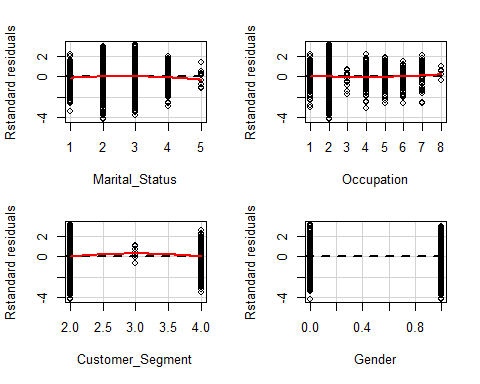
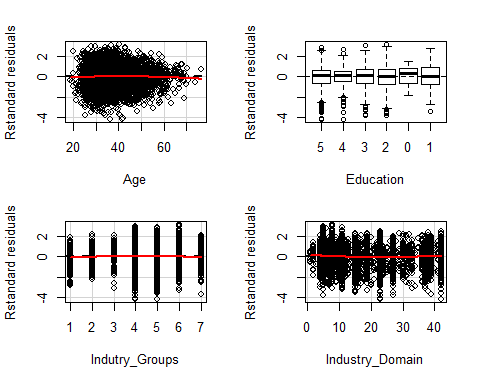
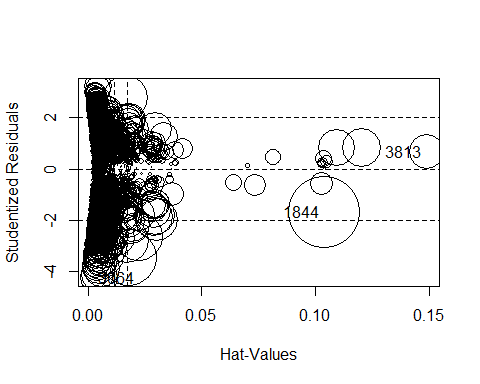
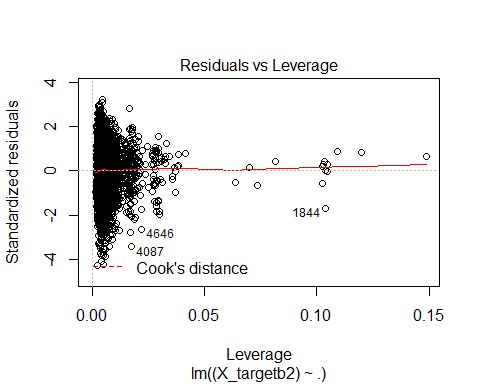
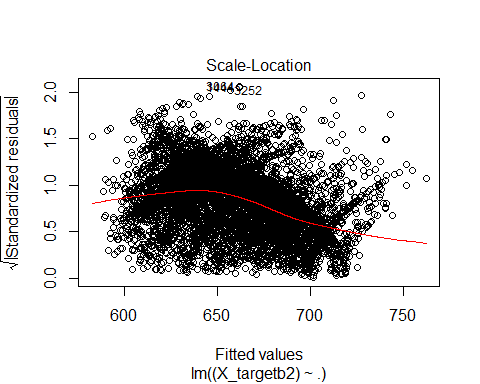
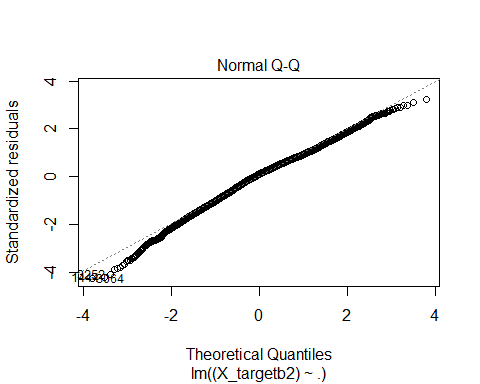
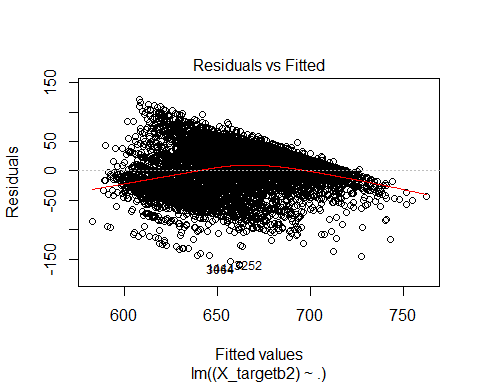
mb2<-lm((X\_targetb2)~.,data = X\_train2)  
summary(mb2)

##   
## Call:  
## lm(formula = (X\_targetb2) ~ ., data = X\_train2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -160.014 -24.599 3.888 25.525 121.057   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6.615e+02 7.862e+00 84.146 < 2e-16 \*\*\*  
## Age 1.308e-01 7.531e-02 1.737 0.082375 .   
## Education.L -3.535e+00 4.209e+00 -0.840 0.401117   
## Education.Q -9.477e+00 4.730e+00 -2.004 0.045156 \*   
## Education.C -6.806e+00 5.902e+00 -1.153 0.248832   
## Education^4 5.517e-01 3.759e+00 0.147 0.883314   
## Education^5 1.253e+01 7.134e+00 1.756 0.079159 .   
## Education^6 6.656e+00 6.150e+00 1.082 0.279195   
## Indutry\_Groups -3.120e-01 4.448e-01 -0.701 0.483119   
## Industry\_Domain -1.557e-01 3.962e-02 -3.930 8.57e-05 \*\*\*  
## Marital\_Status -1.492e+00 8.271e-01 -1.803 0.071369 .   
## Occupation -1.083e+00 5.664e-01 -1.912 0.055959 .   
## Customer\_Segment 9.939e-01 7.899e-01 1.258 0.208330   
## Gender 1.282e+00 1.053e+00 1.217 0.223470   
## Metropolitan\_City 1.835e-01 1.447e-01 1.267 0.205051   
## Residence 2.090e+00 4.458e-01 4.688 2.82e-06 \*\*\*  
## Insurance\_Product\_type 1.429e-02 2.615e-01 0.055 0.956419   
## Insurance\_Acquisition\_Channel -2.425e+00 8.007e-01 -3.029 0.002460 \*\*   
## Active\_Bank\_Products.L 1.188e+01 1.333e+00 8.912 < 2e-16 \*\*\*  
## Active\_Bank\_Products.Q -4.211e+01 1.436e+00 -29.312 < 2e-16 \*\*\*  
## Active\_Bank\_Products.C 2.939e+00 1.568e+00 1.874 0.060914 .   
## Active\_Bank\_Products^4 -2.527e+01 1.700e+00 -14.864 < 2e-16 \*\*\*  
## Active\_Bank\_Products^5 -3.518e+01 1.666e+00 -21.114 < 2e-16 \*\*\*  
## Total\_Bank\_Products -1.075e+00 2.975e-01 -3.613 0.000305 \*\*\*  
## Commercial\_Loan -4.664e+00 4.104e+00 -1.136 0.255814   
## Consumer\_Auto\_Loan -3.321e+00 1.530e+00 -2.170 0.030021 \*   
## Mortgage\_Loan -1.489e+00 1.194e+00 -1.247 0.212473   
## Personal\_Loan -1.386e+01 1.100e+00 -12.597 < 2e-16 \*\*\*  
## Credit\_Card 3.428e+00 9.674e-01 3.544 0.000397 \*\*\*  
## Deposit 6.097e+00 1.083e+00 5.630 1.87e-08 \*\*\*  
## Tenure\_with\_Bank\_Group.L -1.331e+01 2.095e+00 -6.352 2.27e-10 \*\*\*  
## Tenure\_with\_Bank\_Group.Q -1.270e+01 1.890e+00 -6.722 1.93e-11 \*\*\*  
## Tenure\_with\_Bank\_Group.C 2.241e+00 1.396e+00 1.605 0.108563   
## Tenure\_with\_Bank\_Group^4 4.032e+00 1.143e+00 3.526 0.000424 \*\*\*  
## Tenure\_of\_Insurance 2.972e-01 1.908e-01 1.558 0.119330   
## Salary\_Amount -1.483e-05 8.858e-06 -1.675 0.094070 .   
## Credit\_Limit 1.290e-05 7.523e-06 1.714 0.086523 .   
## Total\_Asset\_Under\_Mngmnt 1.034e-02 2.506e-04 41.254 < 2e-16 \*\*\*  
## Avg\_Monthly\_Balance -1.123e-05 5.897e-06 -1.905 0.056799 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 37.73 on 6754 degrees of freedom  
## Multiple R-squared: 0.3897, Adjusted R-squared: 0.3862   
## F-statistic: 113.5 on 38 and 6754 DF, p-value: < 2.2e-16

plot(mb2)



dia(mb2)



## Test stat Pr(>|t|)  
## Age -1.622 0.105  
## Education NA NA  
## Indutry\_Groups -0.591 0.555  
## Industry\_Domain 4.898 0.000  
## Marital\_Status -2.618 0.009  
## Occupation 1.638 0.101  
## Customer\_Segment -0.984 0.325  
## Gender -0.073 0.942  
## Metropolitan\_City -0.071 0.943  
## Residence 1.978 0.048  
## Insurance\_Product\_type 0.338 0.735  
## Insurance\_Acquisition\_Channel -3.500 0.000  
## Active\_Bank\_Products NA NA  
## Total\_Bank\_Products 8.047 0.000  
## Commercial\_Loan -1.091 0.275  
## Consumer\_Auto\_Loan -1.110 0.267  
## Mortgage\_Loan -1.075 0.282  
## Personal\_Loan -1.086 0.278  
## Credit\_Card -1.105 0.269  
## Deposit -1.113 0.266  
## Tenure\_with\_Bank\_Group NA NA  
## Tenure\_of\_Insurance -0.945 0.345  
## Salary\_Amount 0.435 0.664  
## Credit\_Limit -0.872 0.383  
## Total\_Asset\_Under\_Mngmnt -6.265 0.000  
## Avg\_Monthly\_Balance 1.901 0.057  
## Tukey test -14.575 0.000

xgb\_cva1 <- xgb.cv(data=data.matrix(X\_train1), nfold=10,  
 label=as.matrix(X\_targeta1), booster = "gblinear",   
 objective = "reg:linear",   
 max.depth = 20,  
 nrounds = 35,  
 lambda = 0,   
 lambda\_bias = 0,   
 alpha = 0)

## [0] train-rmse:268.542203+5.993551 test-rmse:281.681717+47.959464  
## [1] train-rmse:258.882213+5.988069 test-rmse:284.306982+49.300313  
## [2] train-rmse:253.233989+6.169361 test-rmse:285.888432+54.049501  
## [3] train-rmse:249.202208+6.543612 test-rmse:285.489118+58.841164  
## [4] train-rmse:246.236188+6.933319 test-rmse:287.926198+63.217849  
## [5] train-rmse:243.990390+7.197850 test-rmse:289.086063+65.197260  
## [6] train-rmse:242.306621+7.331760 test-rmse:289.760640+65.242979  
## [7] train-rmse:240.902525+7.411744 test-rmse:291.295843+65.268577  
## [8] train-rmse:239.718109+7.470191 test-rmse:292.574722+65.057470  
## [9] train-rmse:238.729671+7.565799 test-rmse:293.873543+65.897310  
## [10] train-rmse:237.834548+7.692964 test-rmse:295.243838+65.514681  
## [11] train-rmse:237.060295+7.684435 test-rmse:295.376152+66.609819  
## [12] train-rmse:236.414850+7.752705 test-rmse:296.655328+66.272843  
## [13] train-rmse:235.768393+7.741185 test-rmse:296.778111+66.115360  
## [14] train-rmse:235.162965+7.787639 test-rmse:297.128700+67.366057  
## [15] train-rmse:234.728564+7.787644 test-rmse:298.610458+67.568482  
## [16] train-rmse:234.312497+7.850409 test-rmse:299.488507+68.413896  
## [17] train-rmse:233.878073+7.889415 test-rmse:300.066466+68.842378  
## [18] train-rmse:233.541479+7.903786 test-rmse:301.171072+69.172026  
## [19] train-rmse:233.225000+7.956764 test-rmse:301.546252+69.467942  
## [20] train-rmse:232.929906+8.004204 test-rmse:302.341519+69.989872  
## [21] train-rmse:232.628401+8.064950 test-rmse:302.512444+70.683278  
## [22] train-rmse:232.350128+8.125691 test-rmse:302.776733+71.145533  
## [23] train-rmse:232.108650+8.160940 test-rmse:303.880568+71.319327  
## [24] train-rmse:231.897705+8.168223 test-rmse:304.287642+71.294330  
## [25] train-rmse:231.681818+8.187622 test-rmse:304.601904+72.052719  
## [26] train-rmse:231.477801+8.216954 test-rmse:304.832280+72.149471  
## [27] train-rmse:231.306175+8.268526 test-rmse:305.401885+72.147622  
## [28] train-rmse:231.162518+8.286570 test-rmse:305.907022+72.881256  
## [29] train-rmse:231.001508+8.290823 test-rmse:306.200265+72.867142  
## [30] train-rmse:230.874495+8.313947 test-rmse:306.691798+73.091478  
## [31] train-rmse:230.742200+8.340851 test-rmse:306.940305+72.929384  
## [32] train-rmse:230.627069+8.349438 test-rmse:307.223167+73.270517  
## [33] train-rmse:230.522604+8.346824 test-rmse:307.507863+73.424109  
## [34] train-rmse:230.404288+8.359510 test-rmse:307.567020+73.844758

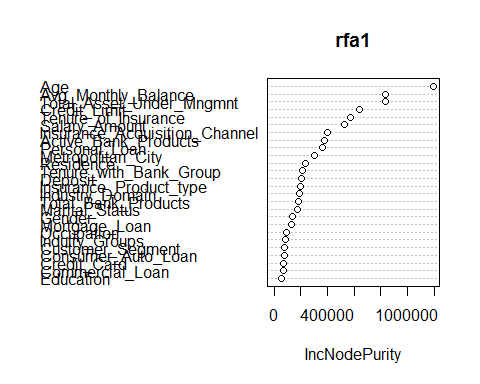
#one-hot-encoding categorical features  
ohe\_feats = c('Insurance\_Acquisition\_Channel','Insurance\_Product\_type','Residence',  
 'Metropolitan\_City','Customer\_Segment')  
  
dummies <- dummyVars(~ (Insurance\_Acquisition\_Channel+Insurance\_Product\_type  
 +Residence+Metropolitan\_City+Customer\_Segment+Identifier),sep=':',  
 fullRank=T,sparse=T,data = a0)  
a\_ohe <- as.data.table(predict(dummies, newdata =a0))  
#a\_cbd <- merge(a[,-c(which(colnames(a) %in% ohe\_feats))],a\_ohe,  
# by='Identifier',all=TRUE)  
a\_cbd <- cbind(a[,-c(which(colnames(a0) %in% ohe\_feats)),with=F],a\_ohe)  
a\_cbd <-(a\_cbd[,":="(Application\_Score=NULL,  
 Behavioural\_Score=NULL)])  
X\_train0 = a\_cbd[a\_cbd$Identifier %in% I,2:49,with=FALSE]  
X\_train01 = a\_cbd[a\_cbd$Identifier %in% I1,2:49,with=FALSE]  
X\_train02 = a\_cbd[a\_cbd$Identifier %in% I2,2:49,with=FALSE]  
ma01<-lm((X\_targeta1)~.,data = X\_train01)  
summary(ma01)

##   
## Call:  
## lm(formula = (X\_targeta1) ~ ., data = X\_train01)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -570.25 -94.71 3.18 125.95 428.21   
##   
## Coefficients: (7 not defined because of singularities)  
## Estimate  
## (Intercept) 2.754e+03  
## Age 3.781e+00  
## Education.L 2.838e+02  
## Education.Q -1.077e+01  
## Education.C -3.018e+01  
## Education^4 1.564e+02  
## Indutry\_Groups -1.636e+02  
## Industry\_Domain 1.844e+00  
## Marital\_Status 1.713e+02  
## Occupation -3.739e+01  
## Gender 1.347e+01  
## Active\_Bank\_Products.L -1.384e+02  
## Active\_Bank\_Products.Q 3.159e+02  
## Active\_Bank\_Products.C 6.710e+01  
## Active\_Bank\_Products^4 -7.482e+01  
## Active\_Bank\_Products^5 4.932e+02  
## Total\_Bank\_Products 1.053e+01  
## Commercial\_Loan -3.346e+01  
## Consumer\_Auto\_Loan -6.511e+01  
## Mortgage\_Loan -1.735e+02  
## Personal\_Loan 3.474e+01  
## Credit\_Card -1.373e+02  
## Deposit 8.823e+01  
## Tenure\_with\_Bank\_Group.L 2.182e+02  
## Tenure\_with\_Bank\_Group.Q -3.934e+02  
## Tenure\_with\_Bank\_Group.C -1.467e+02  
## Tenure\_with\_Bank\_Group^4 -1.774e+02  
## Tenure\_of\_Insurance -9.881e-01  
## Salary\_Amount -5.886e-04  
## Credit\_Limit 5.281e-04  
## Total\_Asset\_Under\_Mngmnt -2.010e-03  
## Avg\_Monthly\_Balance -9.080e-04  
## Insurance\_Acquisition\_ChannelBank -4.176e+02  
## Insurance\_Acquisition\_ChannelDirect -5.064e+02  
## `Insurance\_Product\_typeCredit Card Insurance` -3.309e+02  
## `Insurance\_Product\_typeHealth Insurance` -4.892e+02  
## `Insurance\_Product\_typeHome Insurance` 1.137e+02  
## `Insurance\_Product\_typeSmall and Medium Enterprises Insurance` -9.675e+01  
## `Insurance\_Product\_typeTheft Insurance` NA  
## `Insurance\_Product\_typeTheft Insurance for Credit Card` -2.326e+02  
## `ResidenceLiving in a house owned by a family member` -6.216e+02  
## `ResidenceLiving with Family` -6.340e+02  
## `ResidenceOwn House` -6.197e+02  
## `ResidenceRented House` -1.164e+03  
## ResidenceUnknown NA  
## Metropolitan\_CityBengaluru -2.466e+02  
## Metropolitan\_CityChandigarh -3.043e+02  
## Metropolitan\_CityChennai 2.732e+01  
## Metropolitan\_CityDelhi NA  
## Metropolitan\_CityHyderabad -2.961e+02  
## Metropolitan\_CityIndore -1.511e+02  
## Metropolitan\_CityLucknow NA  
## Metropolitan\_CityMumbai -7.465e+02  
## Metropolitan\_CityOthers -4.120e+02  
## Metropolitan\_CityPune -8.017e+02  
## Customer\_SegmentIndividual NA  
## `Customer\_SegmentPrivate Banking` NA  
## `Customer\_SegmentPrivilege Banking` NA  
## Customer\_SegmentSME -5.059e+01  
## Std. Error  
## (Intercept) 1.035e+03  
## Age 3.616e+00  
## Education.L 1.671e+02  
## Education.Q 2.003e+02  
## Education.C 1.646e+02  
## Education^4 2.311e+02  
## Indutry\_Groups 9.007e+01  
## Industry\_Domain 5.268e+00  
## Marital\_Status 1.028e+02  
## Occupation 4.270e+01  
## Gender 7.415e+01  
## Active\_Bank\_Products.L 1.070e+02  
## Active\_Bank\_Products.Q 1.466e+02  
## Active\_Bank\_Products.C 1.489e+02  
## Active\_Bank\_Products^4 1.656e+02  
## Active\_Bank\_Products^5 2.162e+02  
## Total\_Bank\_Products 2.657e+01  
## Commercial\_Loan 1.311e+02  
## Consumer\_Auto\_Loan 1.024e+02  
## Mortgage\_Loan 1.030e+02  
## Personal\_Loan 1.119e+02  
## Credit\_Card 6.731e+01  
## Deposit 9.271e+01  
## Tenure\_with\_Bank\_Group.L 3.240e+02  
## Tenure\_with\_Bank\_Group.Q 2.936e+02  
## Tenure\_with\_Bank\_Group.C 2.110e+02  
## Tenure\_with\_Bank\_Group^4 1.626e+02  
## Tenure\_of\_Insurance 1.085e+01  
## Salary\_Amount 6.975e-04  
## Credit\_Limit 4.315e-04  
## Total\_Asset\_Under\_Mngmnt 1.817e-02  
## Avg\_Monthly\_Balance 5.036e-04  
## Insurance\_Acquisition\_ChannelBank 2.325e+02  
## Insurance\_Acquisition\_ChannelDirect 2.042e+02  
## `Insurance\_Product\_typeCredit Card Insurance` 2.578e+02  
## `Insurance\_Product\_typeHealth Insurance` 4.120e+02  
## `Insurance\_Product\_typeHome Insurance` 1.166e+02  
## `Insurance\_Product\_typeSmall and Medium Enterprises Insurance` 8.938e+01  
## `Insurance\_Product\_typeTheft Insurance` NA  
## `Insurance\_Product\_typeTheft Insurance for Credit Card` 4.762e+02  
## `ResidenceLiving in a house owned by a family member` 3.393e+02  
## `ResidenceLiving with Family` 3.431e+02  
## `ResidenceOwn House` 3.428e+02  
## `ResidenceRented House` 4.796e+02  
## ResidenceUnknown NA  
## Metropolitan\_CityBengaluru 5.665e+02  
## Metropolitan\_CityChandigarh 5.640e+02  
## Metropolitan\_CityChennai 5.448e+02  
## Metropolitan\_CityDelhi NA  
## Metropolitan\_CityHyderabad 5.625e+02  
## Metropolitan\_CityIndore 5.877e+02  
## Metropolitan\_CityLucknow NA  
## Metropolitan\_CityMumbai 6.155e+02  
## Metropolitan\_CityOthers 5.578e+02  
## Metropolitan\_CityPune 6.619e+02  
## Customer\_SegmentIndividual NA  
## `Customer\_SegmentPrivate Banking` NA  
## `Customer\_SegmentPrivilege Banking` NA  
## Customer\_SegmentSME 8.270e+01  
## t value  
## (Intercept) 2.661  
## Age 1.046  
## Education.L 1.698  
## Education.Q -0.054  
## Education.C -0.183  
## Education^4 0.677  
## Indutry\_Groups -1.816  
## Industry\_Domain 0.350  
## Marital\_Status 1.667  
## Occupation -0.876  
## Gender 0.182  
## Active\_Bank\_Products.L -1.293  
## Active\_Bank\_Products.Q 2.154  
## Active\_Bank\_Products.C 0.451  
## Active\_Bank\_Products^4 -0.452  
## Active\_Bank\_Products^5 2.282  
## Total\_Bank\_Products 0.396  
## Commercial\_Loan -0.255  
## Consumer\_Auto\_Loan -0.636  
## Mortgage\_Loan -1.685  
## Personal\_Loan 0.310  
## Credit\_Card -2.040  
## Deposit 0.952  
## Tenure\_with\_Bank\_Group.L 0.673  
## Tenure\_with\_Bank\_Group.Q -1.340  
## Tenure\_with\_Bank\_Group.C -0.695  
## Tenure\_with\_Bank\_Group^4 -1.091  
## Tenure\_of\_Insurance -0.091  
## Salary\_Amount -0.844  
## Credit\_Limit 1.224  
## Total\_Asset\_Under\_Mngmnt -0.111  
## Avg\_Monthly\_Balance -1.803  
## Insurance\_Acquisition\_ChannelBank -1.796  
## Insurance\_Acquisition\_ChannelDirect -2.480  
## `Insurance\_Product\_typeCredit Card Insurance` -1.283  
## `Insurance\_Product\_typeHealth Insurance` -1.188  
## `Insurance\_Product\_typeHome Insurance` 0.975  
## `Insurance\_Product\_typeSmall and Medium Enterprises Insurance` -1.083  
## `Insurance\_Product\_typeTheft Insurance` NA  
## `Insurance\_Product\_typeTheft Insurance for Credit Card` -0.488  
## `ResidenceLiving in a house owned by a family member` -1.832  
## `ResidenceLiving with Family` -1.848  
## `ResidenceOwn House` -1.808  
## `ResidenceRented House` -2.428  
## ResidenceUnknown NA  
## Metropolitan\_CityBengaluru -0.435  
## Metropolitan\_CityChandigarh -0.540  
## Metropolitan\_CityChennai 0.050  
## Metropolitan\_CityDelhi NA  
## Metropolitan\_CityHyderabad -0.526  
## Metropolitan\_CityIndore -0.257  
## Metropolitan\_CityLucknow NA  
## Metropolitan\_CityMumbai -1.213  
## Metropolitan\_CityOthers -0.739  
## Metropolitan\_CityPune -1.211  
## Customer\_SegmentIndividual NA  
## `Customer\_SegmentPrivate Banking` NA  
## `Customer\_SegmentPrivilege Banking` NA  
## Customer\_SegmentSME -0.612  
## Pr(>|t|)   
## (Intercept) 0.0101 \*  
## Age 0.3000   
## Education.L 0.0948 .  
## Education.Q 0.9573   
## Education.C 0.8552   
## Education^4 0.5013   
## Indutry\_Groups 0.0745 .  
## Industry\_Domain 0.7275   
## Marital\_Status 0.1009   
## Occupation 0.3848   
## Gender 0.8565   
## Active\_Bank\_Products.L 0.2011   
## Active\_Bank\_Products.Q 0.0354 \*  
## Active\_Bank\_Products.C 0.6538   
## Active\_Bank\_Products^4 0.6530   
## Active\_Bank\_Products^5 0.0262 \*  
## Total\_Bank\_Products 0.6932   
## Commercial\_Loan 0.7994   
## Consumer\_Auto\_Loan 0.5272   
## Mortgage\_Loan 0.0973 .  
## Personal\_Loan 0.7574   
## Credit\_Card 0.0459 \*  
## Deposit 0.3452   
## Tenure\_with\_Bank\_Group.L 0.5033   
## Tenure\_with\_Bank\_Group.Q 0.1855   
## Tenure\_with\_Bank\_Group.C 0.4897   
## Tenure\_with\_Bank\_Group^4 0.2798   
## Tenure\_of\_Insurance 0.9277   
## Salary\_Amount 0.4022   
## Credit\_Limit 0.2260   
## Total\_Asset\_Under\_Mngmnt 0.9123   
## Avg\_Monthly\_Balance 0.0766 .  
## Insurance\_Acquisition\_ChannelBank 0.0777 .  
## Insurance\_Acquisition\_ChannelDirect 0.0161 \*  
## `Insurance\_Product\_typeCredit Card Insurance` 0.2045   
## `Insurance\_Product\_typeHealth Insurance` 0.2399   
## `Insurance\_Product\_typeHome Insurance` 0.3334   
## `Insurance\_Product\_typeSmall and Medium Enterprises Insurance` 0.2835   
## `Insurance\_Product\_typeTheft Insurance` NA   
## `Insurance\_Product\_typeTheft Insurance for Credit Card` 0.6271   
## `ResidenceLiving in a house owned by a family member` 0.0721 .  
## `ResidenceLiving with Family` 0.0697 .  
## `ResidenceOwn House` 0.0758 .  
## `ResidenceRented House` 0.0183 \*  
## ResidenceUnknown NA   
## Metropolitan\_CityBengaluru 0.6650   
## Metropolitan\_CityChandigarh 0.5915   
## Metropolitan\_CityChennai 0.9602   
## Metropolitan\_CityDelhi NA   
## Metropolitan\_CityHyderabad 0.6006   
## Metropolitan\_CityIndore 0.7979   
## Metropolitan\_CityLucknow NA   
## Metropolitan\_CityMumbai 0.2301   
## Metropolitan\_CityOthers 0.4632   
## Metropolitan\_CityPune 0.2308   
## Customer\_SegmentIndividual NA   
## `Customer\_SegmentPrivate Banking` NA   
## `Customer\_SegmentPrivilege Banking` NA   
## Customer\_SegmentSME 0.5431   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 265.9 on 58 degrees of freedom  
## Multiple R-squared: 0.5504, Adjusted R-squared: 0.155   
## F-statistic: 1.392 on 51 and 58 DF, p-value: 0.1112

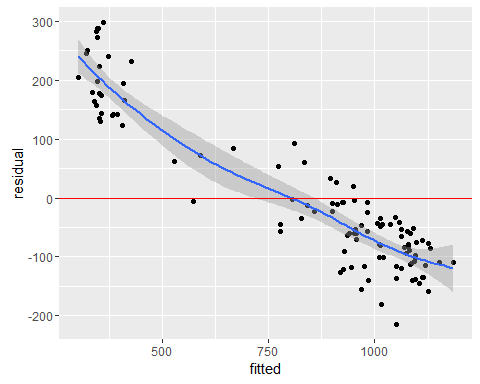
rfa1<-randomForest((X\_targeta1)~.,data = X\_train1)  
p\_rf1<-predict(rfa1,X\_train1)  
RMS(p\_rf1,X\_targeta1,36)[1]

## [[1]]  
## [1] 155.688

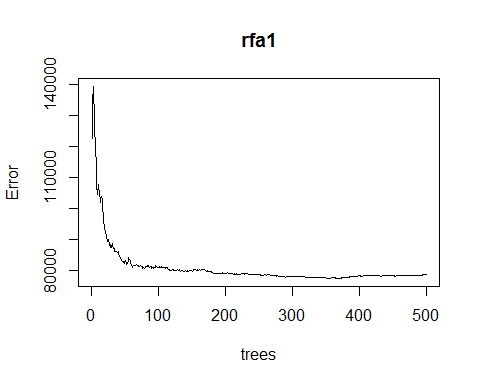
varImpPlot(rfa1)



res\_plot2(X\_targeta1,p\_rf1)



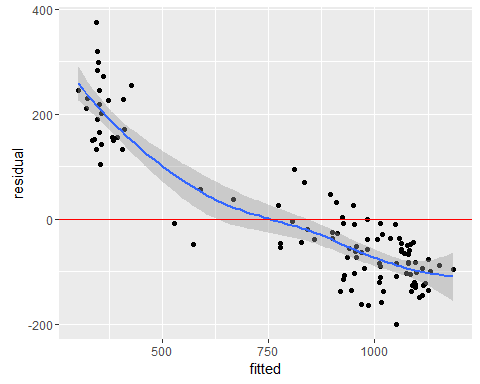
plot(rfa1)



rfa1<-randomForest((X\_targeta1)~.-(Gender+Occupation+Indutry\_Groups+  
 Consumer\_Auto\_Loan+Commercial\_Loan+Credit\_Card+  
 Customer\_Segment+Education+Mortgage\_Loan),  
 importance=F,ntree =100,data = X\_train1)  
p\_rf1<-predict(rfa1,X\_train1)  
RMS(p\_rf1,X\_targeta1,36)[1]

## [[1]]  
## [1] 160.6635

res\_plot2(X\_targeta1,p\_rf1)



pa1<-predict(rfa1,X\_test1)  
pa2<-predict(ma2,X\_test2)  
  
da1<-as.data.table(cbind(X\_test1$Identifier,pa1),check.names=T)[,.(P\_Application\_Score=pa1,Identifier=V1)]  
sol<-merge(x = sol, y = da1, by = "Identifier", all.x = TRUE)  
da2<-as.data.table(cbind(X\_test2$Identifier,pa2),check.names=T)[,.(P\_Application\_Score=pa2,Identifier=V1)]  
sol<-merge(x = sol, y = da2, by = "Identifier", all.x = TRUE)  
sol$P\_Application\_Score<- rowSums(sol[,c("P\_Application\_Score.y","P\_Application\_Score"),with=F], na.rm=T)  
  
  
p1<-sol$P\_Application\_Score  
p2<-predict(mb,X\_test)  
  
sol<-fread("BFSI - Solution submission template.csv",stringsAsFactors=T)  
sol[,c("P\_Application\_Score","P\_Behavioural\_Score"):=list(p1,p2)]  
write.xlsx(sol,"BFSI - Solution submission template.xlsx",row.names=FALSE)