Logistic Regression Business Case (Proactive Attrition Management)

## Inpoting the Data set

Dataset<-read.csv("Proactive Attrition Management-Logistic Regression Case Study.csv")  
View(Dataset)

## To get the all column names

colnames(Dataset)

## [1] "REVENUE" "MOU" "RECCHRGE" "DIRECTAS" "OVERAGE" "ROAM"   
## [7] "CHANGEM" "CHANGER" "DROPVCE" "BLCKVCE" "UNANSVCE" "CUSTCARE"  
## [13] "THREEWAY" "MOUREC" "OUTCALLS" "INCALLS" "PEAKVCE" "OPEAKVCE"  
## [19] "DROPBLK" "CALLFWDV" "CALLWAIT" "CHURN" "MONTHS" "UNIQSUBS"  
## [25] "ACTVSUBS" "CSA" "PHONES" "MODELS" "EQPDAYS" "CUSTOMER"  
## [31] "AGE1" "AGE2" "CHILDREN" "CREDITA" "CREDITAA" "CREDITB"   
## [37] "CREDITC" "CREDITDE" "CREDITGY" "CREDITZ" "PRIZMRUR" "PRIZMUB"   
## [43] "PRIZMTWN" "REFURB" "WEBCAP" "TRUCK" "RV" "OCCPROF"   
## [49] "OCCCLER" "OCCCRFT" "OCCSTUD" "OCCHMKR" "OCCRET" "OCCSELF"   
## [55] "OWNRENT" "MARRYUN" "MARRYYES" "MARRYNO" "MAILORD" "MAILRES"   
## [61] "MAILFLAG" "TRAVEL" "PCOWN" "CREDITCD" "RETCALLS" "RETACCPT"  
## [67] "NEWCELLY" "NEWCELLN" "REFER" "INCMISS" "INCOME" "MCYCLE"   
## [73] "CREDITAD" "SETPRCM" "SETPRC" "RETCALL" "CALIBRAT" "CHURNDEP"

#Spliting the data into two datasets Numerical dataset variables and categorical dataset variables and removing unnessary variables from the Dataset.

## Selecting only categorical variables or features and create a new dataset named ‘Dataset\_cat’

#Categorical Variables  
library(dplyr)

## Warning: package 'dplyr' was built under R version 4.0.5

##   
## Attaching package: 'dplyr'

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

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

Dataset\_cat=data.frame(Dataset%>%select\_("CHURN","UNIQSUBS","ACTVSUBS","PHONES","CHILDREN","CREDITA","CREDITAA","CREDITB","CREDITC","CREDITDE","CREDITGY","CREDITZ","PRIZMRUR","PRIZMUB","PRIZMTWN","REFURB","WEBCAP","TRUCK","RV","OCCPROF","OCCCLER","OCCCRFT","OCCSTUD", "OCCHMKR","OCCRET","OCCSELF","OWNRENT","MARRYUN","MARRYYES","MARRYNO","MAILORD","MAILRES", "MAILFLAG","TRAVEL","PCOWN","CREDITCD","RETCALLS","RETACCPT","NEWCELLY","NEWCELLN","REFER","INCMISS","INCOME","MCYCLE","CREDITAD","SETPRCM","RETCALL","CALIBRAT"))

## Warning: `select\_()` was deprecated in dplyr 0.7.0.  
## Please use `select()` instead.

head(Dataset\_cat,5)

## CHURN UNIQSUBS ACTVSUBS PHONES CHILDREN CREDITA CREDITAA CREDITB CREDITC  
## 1 0 1 1 7 0 0 0 0 0  
## 2 0 2 2 9 0 0 0 0 1  
## 3 0 2 2 2 0 1 0 0 0  
## 4 0 2 2 3 1 1 0 0 0  
## 5 0 2 2 2 0 1 0 0 0  
## CREDITDE CREDITGY CREDITZ PRIZMRUR PRIZMUB PRIZMTWN REFURB WEBCAP TRUCK RV  
## 1 1 0 0 0 0 0 0 1 1 1  
## 2 0 0 0 0 0 0 0 1 0 0  
## 3 0 0 0 0 0 0 0 1 0 0  
## 4 0 0 0 0 0 0 0 1 0 0  
## 5 0 0 0 0 0 0 0 1 1 0  
## OCCPROF OCCCLER OCCCRFT OCCSTUD OCCHMKR OCCRET OCCSELF OWNRENT MARRYUN  
## 1 0 0 0 0 0 0 0 1 0  
## 2 0 0 0 0 0 0 0 0 0  
## 3 0 0 0 0 0 0 1 0 0  
## 4 1 0 0 0 0 0 0 0 0  
## 5 0 0 0 0 0 0 0 1 1  
## MARRYYES MARRYNO MAILORD MAILRES MAILFLAG TRAVEL PCOWN CREDITCD RETCALLS  
## 1 0 1 1 1 0 0 0 1 0  
## 2 0 1 1 1 0 0 0 1 0  
## 3 1 0 1 1 0 1 0 1 0  
## 4 0 1 1 1 0 0 0 1 0  
## 5 0 0 0 0 0 0 0 0 0  
## RETACCPT NEWCELLY NEWCELLN REFER INCMISS INCOME MCYCLE CREDITAD SETPRCM  
## 1 0 0 1 0 0 5 0 1 0  
## 2 0 1 0 0 0 6 0 0 0  
## 3 0 0 1 0 0 9 0 1 0  
## 4 0 1 0 0 0 6 0 0 0  
## 5 0 1 0 0 0 7 0 0 0  
## RETCALL CALIBRAT  
## 1 0 0  
## 2 0 0  
## 3 0 0  
## 4 0 0  
## 5 0 0

##Creating numerical variables dataset ##

#Numerical Variables  
Dataset\_num=Dataset%>%select\_("REVENUE","MOU","RECCHRGE","DIRECTAS","OVERAGE","ROAM","CHANGEM","CHANGER","DROPVCE","BLCKVCE","UNANSVCE","CUSTCARE","THREEWAY","MOUREC","OUTCALLS","INCALLS","PEAKVCE","OPEAKVCE","DROPBLK","CALLFWDV","CALLWAIT","AGE1","AGE2")  
head(Dataset\_num,5)

## REVENUE MOU RECCHRGE DIRECTAS OVERAGE ROAM CHANGEM CHANGER DROPVCE  
## 1 57.49 482.75 37.43 0.25 22.75 0 532.25 50.99 8.33  
## 2 82.28 1312.25 75.00 1.24 0.00 0 156.75 8.14 52.00  
## 3 31.66 25.50 29.99 0.25 0.00 0 59.50 4.03 0.00  
## 4 62.13 97.50 65.99 2.48 0.00 0 23.50 6.82 0.00  
## 5 25.23 2.50 25.00 0.00 0.00 0 -2.50 -0.23 0.00  
## BLCKVCE UNANSVCE CUSTCARE THREEWAY MOUREC OUTCALLS INCALLS PEAKVCE OPEAKVCE  
## 1 1.00 61.33 1.67 0.33 55.28 46.33 6.33 83.67 157.00  
## 2 7.67 76.00 4.33 1.33 200.32 370.33 147.00 555.67 303.67  
## 3 1.00 2.33 0.00 0.00 0.00 0.00 0.00 1.67 1.67  
## 4 0.33 4.00 4.00 0.00 0.00 3.67 0.00 7.67 7.33  
## 5 0.00 0.33 0.00 0.00 1.13 0.33 0.00 0.67 0.67  
## DROPBLK CALLFWDV CALLWAIT AGE1 AGE2  
## 1 9.33 0 5.67 30 0  
## 2 59.67 0 22.67 30 0  
## 3 1.00 0 0.00 52 58  
## 4 0.33 0 0.00 46 46  
## 5 0.00 0 0.00 0 0

## Converting categorical datset into factor and identify number of missing values in each feature by using describe function under “Hmisc” package

Dataset\_cat\_factor=data.frame(lapply(Dataset\_cat, as.factor))  
str(Dataset\_cat\_factor)

## 'data.frame': 71047 obs. of 48 variables:  
## $ CHURN : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ UNIQSUBS: Factor w/ 15 levels "1","2","3","4",..: 1 2 2 2 2 5 2 3 1 1 ...  
## $ ACTVSUBS: Factor w/ 12 levels "0","1","2","3",..: 2 3 3 3 3 2 3 3 2 2 ...  
## $ PHONES : Factor w/ 24 levels "1","2","3","4",..: 7 9 2 3 2 10 5 6 4 4 ...  
## $ CHILDREN: Factor w/ 2 levels "0","1": 1 1 1 2 1 2 2 2 1 2 ...  
## $ CREDITA : Factor w/ 2 levels "0","1": 1 1 2 2 2 1 1 2 2 1 ...  
## $ CREDITAA: Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ CREDITB : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 2 1 1 2 ...  
## $ CREDITC : Factor w/ 2 levels "0","1": 1 2 1 1 1 2 1 1 1 1 ...  
## $ CREDITDE: Factor w/ 2 levels "0","1": 2 1 1 1 1 1 1 1 1 1 ...  
## $ CREDITGY: Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ CREDITZ : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ PRIZMRUR: Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ PRIZMUB : Factor w/ 2 levels "0","1": 1 1 1 1 1 2 2 1 2 1 ...  
## $ PRIZMTWN: Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ REFURB : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ WEBCAP : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2 ...  
## $ TRUCK : Factor w/ 2 levels "0","1": 2 1 1 1 2 1 1 2 1 1 ...  
## $ RV : Factor w/ 2 levels "0","1": 2 1 1 1 1 1 1 2 1 1 ...  
## $ OCCPROF : Factor w/ 2 levels "0","1": 1 1 1 2 1 1 1 1 1 1 ...  
## $ OCCCLER : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ OCCCRFT : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ OCCSTUD : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ OCCHMKR : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ OCCRET : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ OCCSELF : Factor w/ 2 levels "0","1": 1 1 2 1 1 1 1 1 1 1 ...  
## $ OWNRENT : Factor w/ 2 levels "0","1": 2 1 1 1 2 1 1 1 1 1 ...  
## $ MARRYUN : Factor w/ 2 levels "0","1": 1 1 1 1 2 1 2 1 1 1 ...  
## $ MARRYYES: Factor w/ 2 levels "0","1": 1 1 2 1 1 1 1 1 1 1 ...  
## $ MARRYNO : Factor w/ 2 levels "0","1": 2 2 1 2 1 2 1 2 2 2 ...  
## $ MAILORD : Factor w/ 2 levels "0","1": 2 2 2 2 1 1 1 2 2 2 ...  
## $ MAILRES : Factor w/ 2 levels "0","1": 2 2 2 2 1 2 1 2 2 2 ...  
## $ MAILFLAG: Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 2 1 1 ...  
## $ TRAVEL : Factor w/ 2 levels "0","1": 1 1 2 1 1 1 1 1 1 1 ...  
## $ PCOWN : Factor w/ 2 levels "0","1": 1 1 1 1 1 2 1 1 1 2 ...  
## $ CREDITCD: Factor w/ 2 levels "0","1": 2 2 2 2 1 2 2 2 2 1 ...  
## $ RETCALLS: Factor w/ 5 levels "0","1","2","3",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ RETACCPT: Factor w/ 5 levels "0","1","2","3",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ NEWCELLY: Factor w/ 2 levels "0","1": 1 2 1 2 2 2 1 1 1 2 ...  
## $ NEWCELLN: Factor w/ 2 levels "0","1": 2 1 2 1 1 1 2 2 2 1 ...  
## $ REFER : Factor w/ 13 levels "0","1","2","3",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ INCMISS : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ INCOME : Factor w/ 10 levels "0","1","2","3",..: 6 7 10 7 8 4 2 5 4 2 ...  
## $ MCYCLE : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ CREDITAD: Factor w/ 16 levels "0","1","2","3",..: 2 1 2 1 1 2 2 2 1 1 ...  
## $ SETPRCM : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ RETCALL : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ CALIBRAT: Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...

#Missing values  
colSums(is.na(Dataset\_cat\_factor))

## CHURN UNIQSUBS ACTVSUBS PHONES CHILDREN CREDITA CREDITAA CREDITB   
## 0 0 0 1 0 0 0 0   
## CREDITC CREDITDE CREDITGY CREDITZ PRIZMRUR PRIZMUB PRIZMTWN REFURB   
## 0 0 0 0 0 0 0 0   
## WEBCAP TRUCK RV OCCPROF OCCCLER OCCCRFT OCCSTUD OCCHMKR   
## 0 0 0 0 0 0 0 0   
## OCCRET OCCSELF OWNRENT MARRYUN MARRYYES MARRYNO MAILORD MAILRES   
## 0 0 0 0 0 0 0 0   
## MAILFLAG TRAVEL PCOWN CREDITCD RETCALLS RETACCPT NEWCELLY NEWCELLN   
## 0 0 0 0 0 0 0 0   
## REFER INCMISS INCOME MCYCLE CREDITAD SETPRCM RETCALL CALIBRAT   
## 0 0 0 0 0 0 0 0

#There is only one column or calegorical variable, i.e., PHONES where we are getting missing values.

## Replacement of NA from PHONES variable

#Getting number of counts of number of handsets.  
table(Dataset\_cat\_factor$PHONES)

##   
## 1 2 3 4 5 6 7 8 9 10 11 12 13   
## 40062 17741 7098 3105 1437 658 409 219 117 71 45 26 19   
## 14 15 16 17 18 19 20 21 22 24 28   
## 8 6 8 3 3 4 1 1 2 2 1

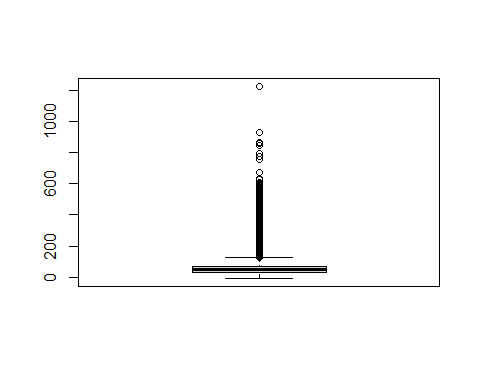
#Replacing NA with 1.   
Dataset\_cat\_factor$PHONES=is.na(Dataset\_cat\_factor$PHONES)<-1  
  
#Again checking for NA  
colSums(is.na(Dataset\_cat\_factor))

## CHURN UNIQSUBS ACTVSUBS PHONES CHILDREN CREDITA CREDITAA CREDITB   
## 0 0 0 0 0 0 0 0   
## CREDITC CREDITDE CREDITGY CREDITZ PRIZMRUR PRIZMUB PRIZMTWN REFURB   
## 0 0 0 0 0 0 0 0   
## WEBCAP TRUCK RV OCCPROF OCCCLER OCCCRFT OCCSTUD OCCHMKR   
## 0 0 0 0 0 0 0 0   
## OCCRET OCCSELF OWNRENT MARRYUN MARRYYES MARRYNO MAILORD MAILRES   
## 0 0 0 0 0 0 0 0   
## MAILFLAG TRAVEL PCOWN CREDITCD RETCALLS RETACCPT NEWCELLY NEWCELLN   
## 0 0 0 0 0 0 0 0   
## REFER INCMISS INCOME MCYCLE CREDITAD SETPRCM RETCALL CALIBRAT   
## 0 0 0 0 0 0 0 0

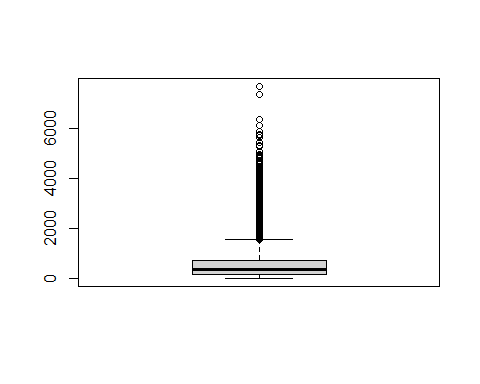
#Now, Categorical dataset is NA free.

## Checking outliers for numeric dataset variables using Boxplot

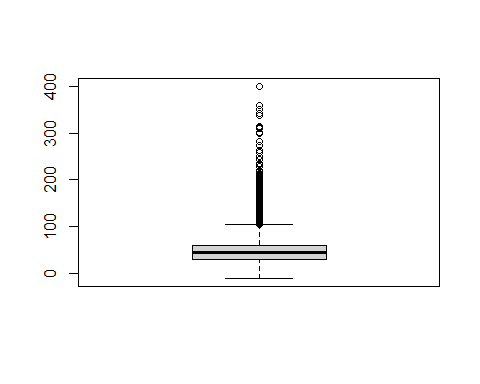
boxplot(Dataset\_num$REVENUE)



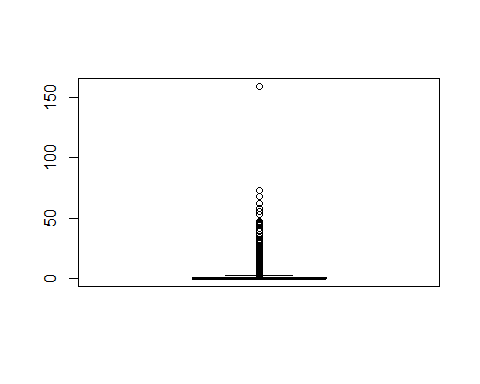
boxplot(Dataset\_num$MOU)



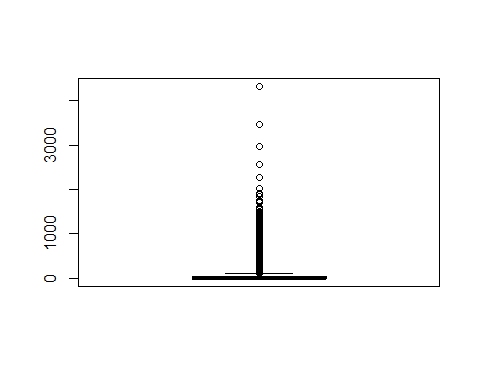
boxplot(Dataset\_num$RECCHRGE)



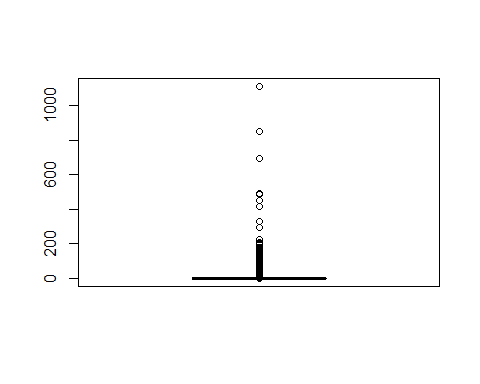
boxplot(Dataset\_num$DIRECTAS)



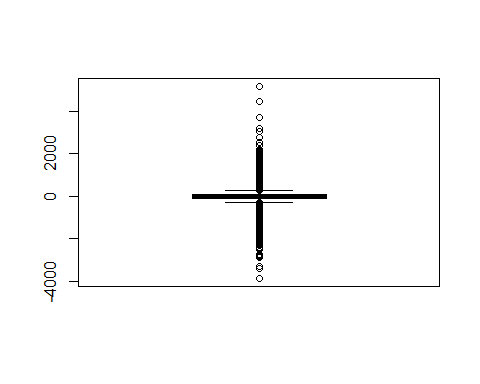
boxplot(Dataset\_num$OVERAGE)



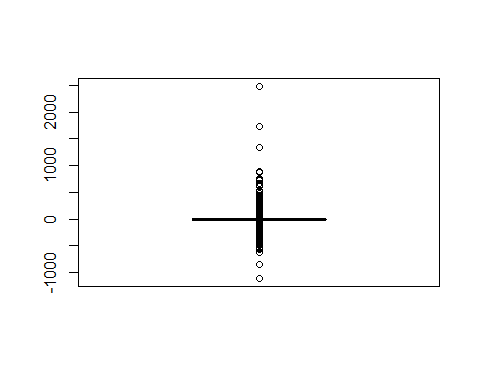
boxplot(Dataset\_num$ROAM)



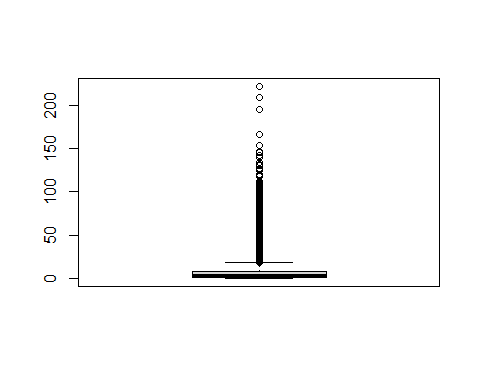
boxplot(Dataset\_num$CHANGEM)



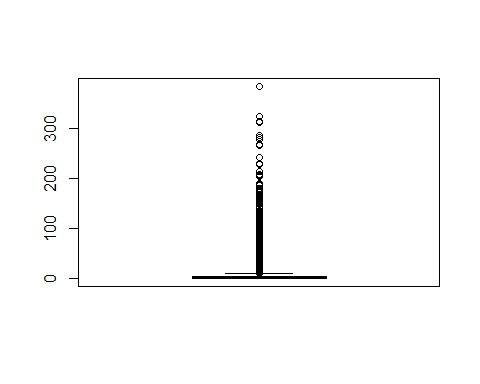
boxplot(Dataset\_num$CHANGER)



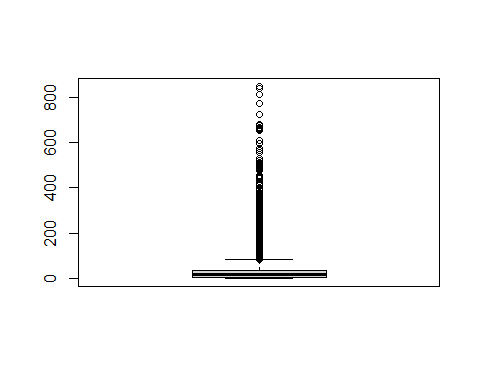
boxplot(Dataset\_num$DROPVCE)



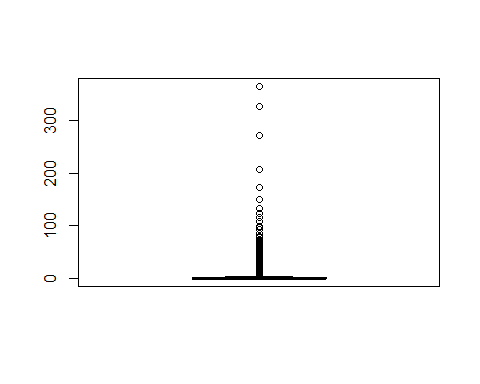
boxplot(Dataset\_num$BLCKVCE)



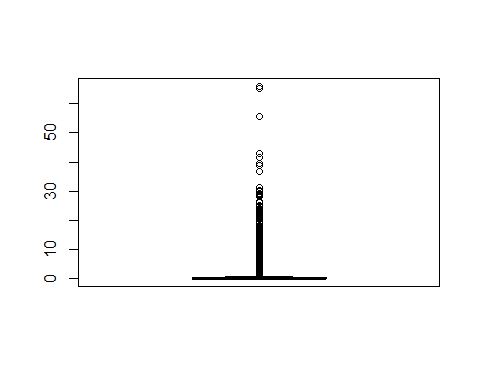
boxplot(Dataset\_num$UNANSVCE)



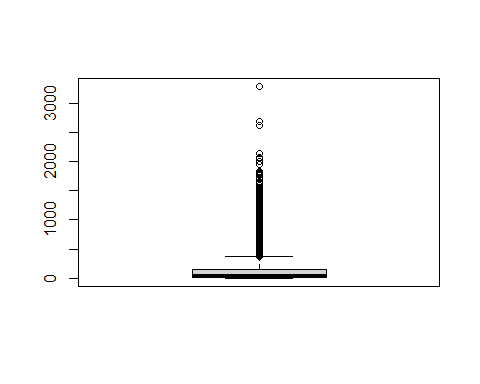
boxplot(Dataset\_num$CUSTCARE)



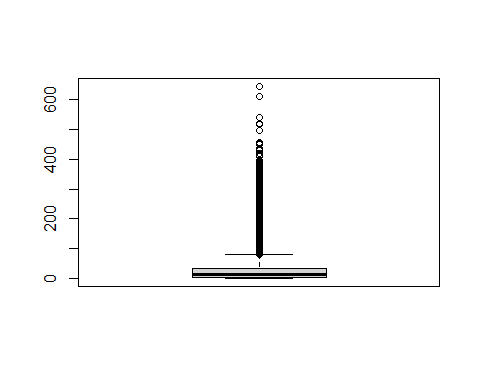
boxplot(Dataset\_num$THREEWAY)



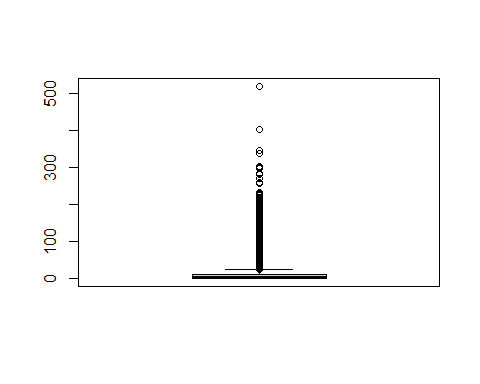
boxplot(Dataset\_num$MOUREC)



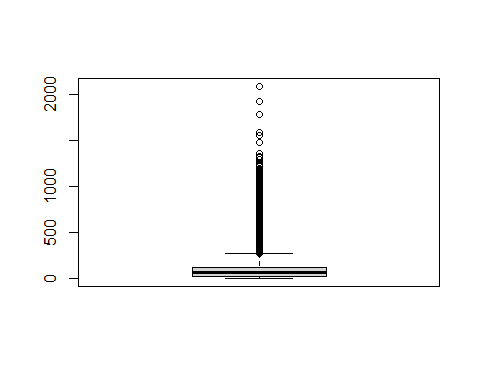
boxplot(Dataset\_num$OUTCALLS)



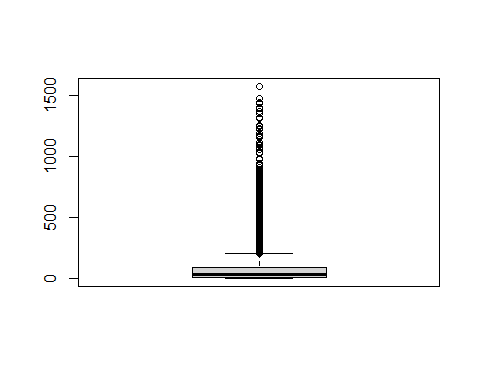
boxplot(Dataset\_num$INCALLS)



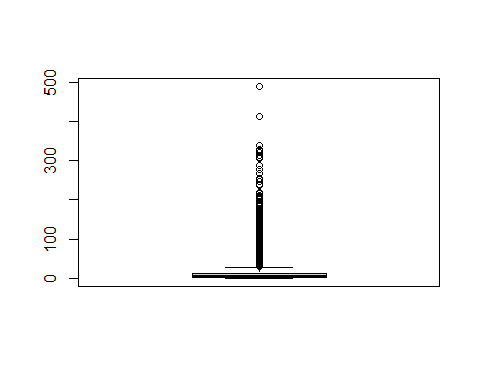
boxplot(Dataset\_num$PEAKVCE)



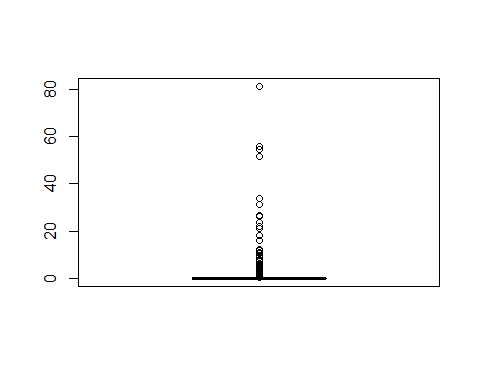
boxplot(Dataset\_num$OPEAKVCE)



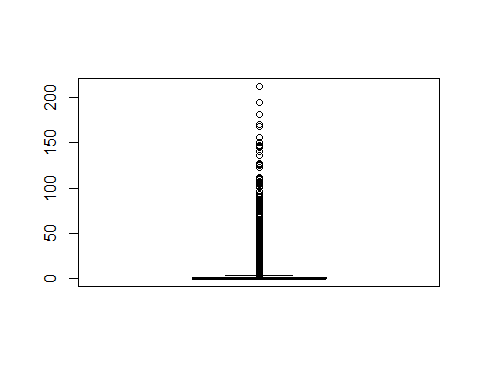
boxplot(Dataset\_num$DROPBLK)



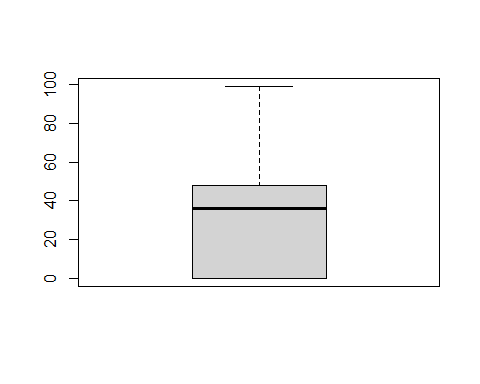
boxplot(Dataset\_num$CALLFWDV)



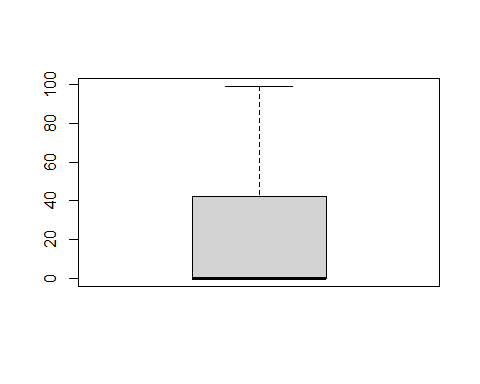
boxplot(Dataset\_num$CALLWAIT)



boxplot(Dataset\_num$AGE1)



boxplot(Dataset\_num$AGE2)

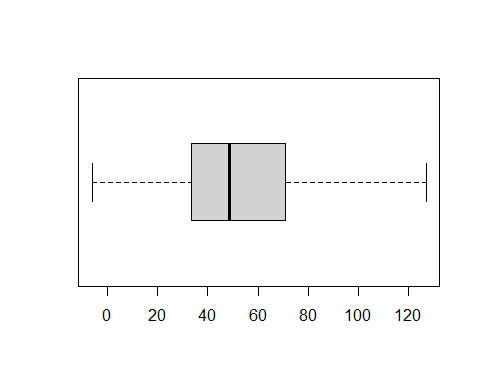


## Replacing outliers from the first and third quartiles for each variable of numerical dataset.

#For REVENUE  
upper\_whisker\_revenue <- quantile(Dataset\_num$REVENUE,0.75,na.rm=T)+1.5\*(quantile(Dataset\_num$REVENUE,0.75,na.rm=T)-quantile(Dataset\_num$REVENUE,0.25,na.rm=T))  
upper\_whisker\_revenue

## 75%   
## 127.115

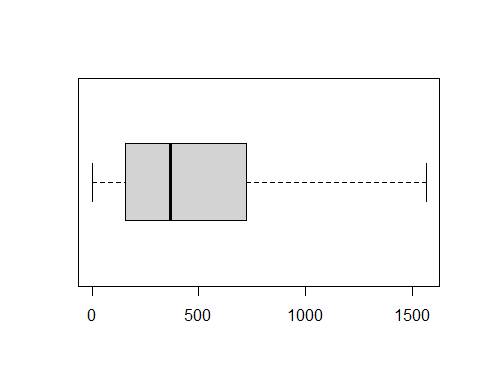
Dataset\_num$REVENUE[Dataset\_num$REVENUE>upper\_whisker\_revenue]<-quantile(Dataset\_num$REVENUE,0.75,na.rm=T)  
boxplot(Dataset\_num$REVENUE, horizontal=T)



#For MOU  
upper\_whisker\_MOU <- quantile(Dataset\_num$MOU,0.75,na.rm=T)+1.5\*(quantile(Dataset\_num$MOU,0.75,na.rm=T)-quantile(Dataset\_num$MOU,0.25,na.rm=T))  
upper\_whisker\_MOU

## 75%   
## 1567

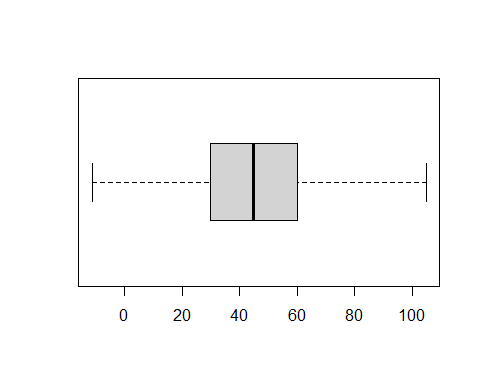
Dataset\_num$MOU[Dataset\_num$MOU>upper\_whisker\_MOU]<-quantile(Dataset\_num$MOU,0.75,na.rm=T)  
boxplot(Dataset\_num$MOU, horizontal=T)



#For RECCHRGE  
upper\_whisker\_RECCHRGE <- quantile(Dataset\_num$RECCHRGE,0.75,na.rm=T)+1.5\*(quantile(Dataset\_num$RECCHRGE,0.75,na.rm=T)-quantile(Dataset\_num$RECCHRGE,0.25,na.rm=T))  
upper\_whisker\_RECCHRGE

## 75%   
## 104.975

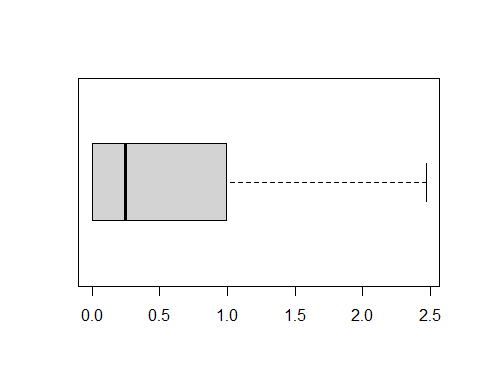
Dataset\_num$RECCHRGE[Dataset\_num$RECCHRGE>upper\_whisker\_RECCHRGE]<-quantile(Dataset\_num$RECCHRGE,0.75,na.rm=T)  
boxplot(Dataset\_num$RECCHRGE, horizontal=T)



#For DIRECTAS  
upper\_whisker\_DIRECTAS <- quantile(Dataset\_num$DIRECTAS,0.75,na.rm=T)+1.5\*(quantile(Dataset\_num$DIRECTAS,0.75,na.rm=T)-quantile(Dataset\_num$DIRECTAS,0.25,na.rm=T))  
upper\_whisker\_DIRECTAS

## 75%   
## 2.475

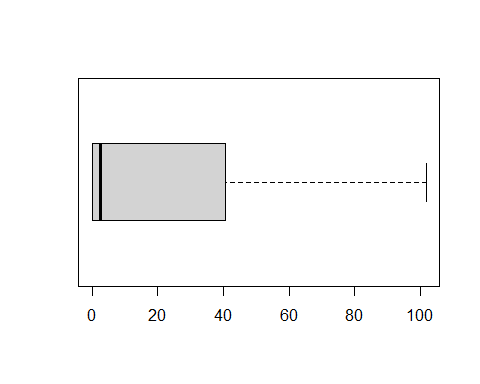
Dataset\_num$DIRECTAS[Dataset\_num$DIRECTAS>upper\_whisker\_DIRECTAS]<-quantile(Dataset\_num$DIRECTAS,0.75,na.rm=T)  
boxplot(Dataset\_num$DIRECTAS, horizontal=T)



#For OVERAGE  
upper\_whisker\_OVERAGE <- quantile(Dataset\_num$OVERAGE,0.75,na.rm=T)+1.5\*(quantile(Dataset\_num$OVERAGE,0.75,na.rm=T)-quantile(Dataset\_num$OVERAGE,0.25,na.rm=T))  
upper\_whisker\_OVERAGE

## 75%   
## 101.875

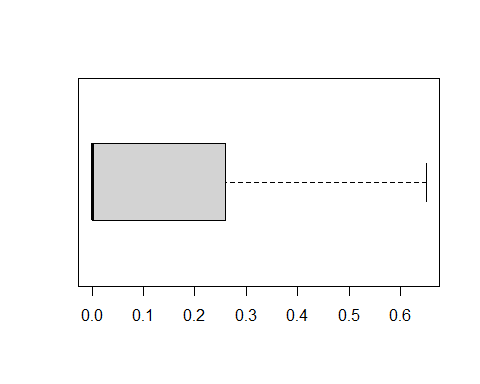
Dataset\_num$OVERAGE[Dataset\_num$OVERAGE>upper\_whisker\_OVERAGE]<-quantile(Dataset\_num$OVERAGE,0.75,na.rm=T)  
boxplot(Dataset\_num$OVERAGE, horizontal=T)



#For ROAM  
upper\_whisker\_ROAM <- quantile(Dataset\_num$ROAM,0.75,na.rm=T)+1.5\*(quantile(Dataset\_num$ROAM,0.75,na.rm=T)-quantile(Dataset\_num$ROAM,0.25,na.rm=T))  
upper\_whisker\_ROAM

## 75%   
## 0.65

Dataset\_num$ROAM[Dataset\_num$ROAM>upper\_whisker\_ROAM]<-quantile(Dataset\_num$ROAM,0.75,na.rm=T)  
boxplot(Dataset\_num$ROAM, horizontal=T)



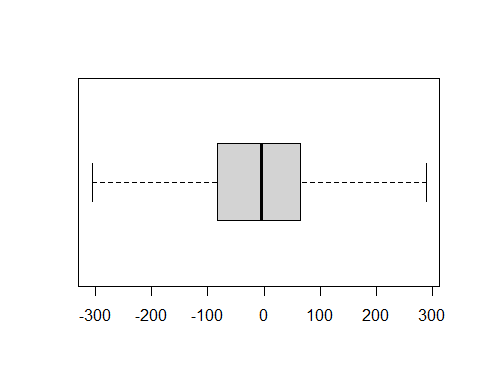
#For CHANGEM  
upper\_whisker\_CHANGEM <- quantile(Dataset\_num$CHANGEM,0.75,na.rm=T)+1.5\*(quantile(Dataset\_num$CHANGEM,0.75, na.rm=T)-quantile(Dataset\_num$CHANGEM,0.25, na.rm=T))  
upper\_whisker\_CHANGEM

## 75%   
## 288.875

Dataset\_num$CHANGEM[Dataset\_num$CHANGEM>upper\_whisker\_CHANGEM]<-quantile(Dataset\_num$CHANGEM,0.75, na.rm = T)  
  
lower\_whisker\_CHANGEM <- quantile(Dataset\_num$CHANGEM,0.25,na.rm=T)-1.5\*(quantile(Dataset\_num$CHANGEM,0.75,na.rm=T)-quantile(Dataset\_num$CHANGEM,0.25, na.rm=T))  
lower\_whisker\_CHANGEM

## 25%   
## -306.125

Dataset\_num$CHANGEM[Dataset\_num$CHANGEM<lower\_whisker\_CHANGEM]<-quantile(Dataset\_num$CHANGEM,0.25, na.rm=T)  
  
boxplot(Dataset\_num$CHANGEM, horizontal=T)



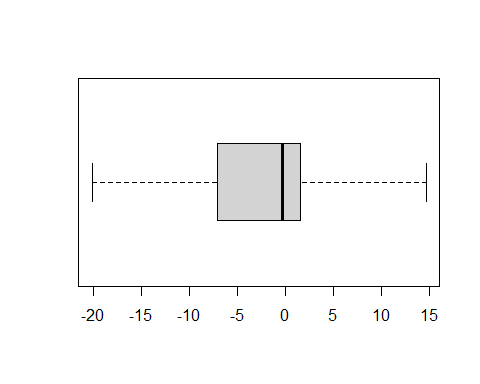
#For CHANGER  
upper\_whisker\_CHANGER <- quantile(Dataset\_num$CHANGER,0.75,na.rm=T)+1.5\*(quantile(Dataset\_num$CHANGER,0.75, na.rm=T)-quantile(Dataset\_num$CHANGER,0.25, na.rm=T))  
upper\_whisker\_CHANGER

## 75%   
## 14.665

Dataset\_num$CHANGER[Dataset\_num$CHANGER>upper\_whisker\_CHANGER]<-quantile(Dataset\_num$CHANGER,0.75, na.rm = T)  
  
lower\_whisker\_CHANGER <- quantile(Dataset\_num$CHANGER,0.25,na.rm=T)-1.5\*(quantile(Dataset\_num$CHANGER,0.75,na.rm=T)-quantile(Dataset\_num$CHANGER,0.25, na.rm=T))  
lower\_whisker\_CHANGER

## 25%   
## -20.175

Dataset\_num$CHANGER[Dataset\_num$CHANGER<lower\_whisker\_CHANGER]<-quantile(Dataset\_num$CHANGER,0.25, na.rm=T)  
  
boxplot(Dataset\_num$CHANGER, horizontal=T)



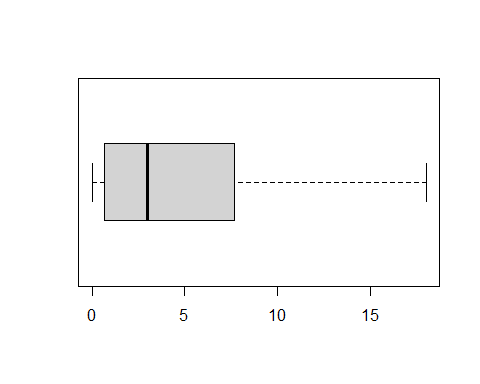
#For DROPVCE  
upper\_whisker\_DROPVCE <- quantile(Dataset\_num$DROPVCE,0.75,na.rm=T)+1.5\*(quantile(Dataset\_num$DROPVCE,0.75, na.rm=T)-quantile(Dataset\_num$DROPVCE,0.25, na.rm=T))  
upper\_whisker\_DROPVCE

## 75%   
## 18.17

Dataset\_num$DROPVCE[Dataset\_num$DROPVCE>upper\_whisker\_DROPVCE]<-quantile(Dataset\_num$DROPVCE,0.75, na.rm = T)  
  
lower\_whisker\_DROPVCE <- quantile(Dataset\_num$DROPVCE,0.25,na.rm=T)-1.5\*(quantile(Dataset\_num$DROPVCE,0.75,na.rm=T)-quantile(Dataset\_num$DROPVCE,0.25, na.rm=T))  
lower\_whisker\_DROPVCE

## 25%   
## -9.83

Dataset\_num$DROPVCE[Dataset\_num$DROPVCE<lower\_whisker\_DROPVCE]<-quantile(Dataset\_num$DROPVCE,0.25, na.rm=T)  
  
boxplot(Dataset\_num$DROPVCE, horizontal=T)



#For BLCKVCE  
upper\_whisker\_BLCKVCE <- quantile(Dataset\_num$BLCKVCE,0.75,na.rm=T)+1.5\*(quantile(Dataset\_num$BLCKVCE,0.75, na.rm=T)-quantile(Dataset\_num$BLCKVCE,0.25, na.rm=T))  
upper\_whisker\_BLCKVCE

## 75%   
## 9.175

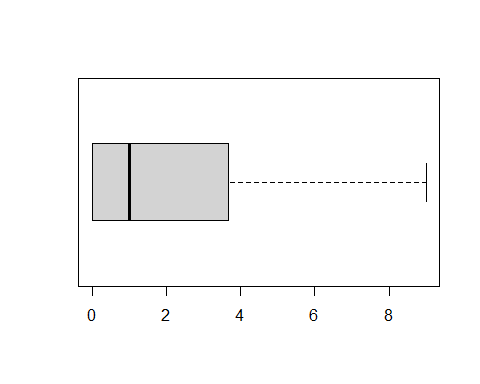
Dataset\_num$BLCKVCE[Dataset\_num$BLCKVCE>upper\_whisker\_BLCKVCE]<-quantile(Dataset\_num$BLCKVCE,0.75, na.rm = T)  
  
lower\_whisker\_BLCKVCE <- quantile(Dataset\_num$BLCKVCE,0.25,na.rm=T)-1.5\*(quantile(Dataset\_num$BLCKVCE,0.75,na.rm=T)-quantile(Dataset\_num$BLCKVCE,0.25, na.rm=T))  
lower\_whisker\_BLCKVCE

## 25%   
## -5.505

Dataset\_num$BLCKVCE[Dataset\_num$BLCKVCE<lower\_whisker\_BLCKVCE]<-quantile(Dataset\_num$BLCKVCE,0.25, na.rm=T)  
  
boxplot(Dataset\_num$BLCKVCE, horizontal=T)  
  
#For UNANSVCE  
upper\_whisker\_UNANSVCE <- quantile(Dataset\_num$UNANSVCE,0.75,na.rm=T)+1.5\*(quantile(Dataset\_num$UNANSVCE,0.75, na.rm=T)-quantile(Dataset\_num$UNANSVCE,0.25, na.rm=T))  
upper\_whisker\_UNANSVCE

## 75%   
## 83.68

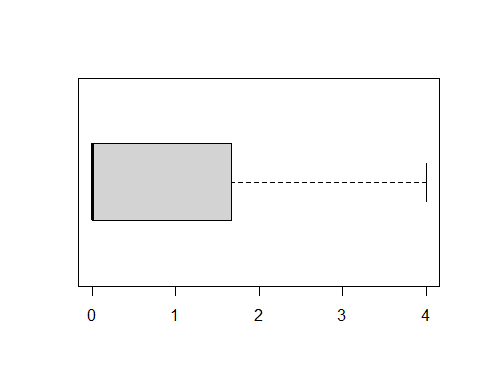
Dataset\_num$UNANSVCE[Dataset\_num$UNANSVCE>upper\_whisker\_UNANSVCE]<-quantile(Dataset\_num$UNANSVCE,0.75, na.rm = T)  
  
boxplot(Dataset\_num$BLCKVCE, horizontal=T)



#For CUSTCARE  
upper\_whisker\_CUSTCARE <- quantile(Dataset\_num$CUSTCARE,0.75,na.rm=T)+1.5\*(quantile(Dataset\_num$CUSTCARE,0.75, na.rm=T)-quantile(Dataset\_num$CUSTCARE,0.25, na.rm=T))  
upper\_whisker\_CUSTCARE

## 75%   
## 4.175

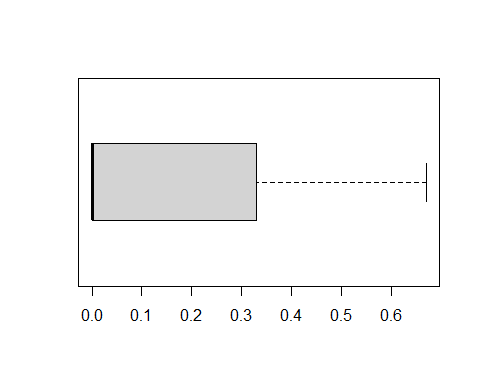
Dataset\_num$CUSTCARE[Dataset\_num$CUSTCARE>upper\_whisker\_CUSTCARE]<-quantile(Dataset\_num$CUSTCARE,0.75, na.rm = T)  
  
boxplot(Dataset\_num$CUSTCARE, horizontal=T)



#For THREEWAY  
upper\_whisker\_THREEWAY <- quantile(Dataset\_num$THREEWAY,0.75,na.rm=T)+1.5\*(quantile(Dataset\_num$THREEWAY,0.75, na.rm=T)-quantile(Dataset\_num$THREEWAY,0.25, na.rm=T))  
upper\_whisker\_THREEWAY

## 75%   
## 0.825

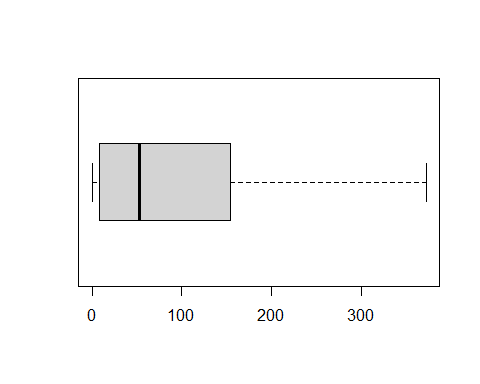
Dataset\_num$THREEWAY[Dataset\_num$THREEWAY>upper\_whisker\_THREEWAY]<-quantile(Dataset\_num$THREEWAY,0.75, na.rm = T)  
  
boxplot(Dataset\_num$THREEWAY, horizontal=T)



#For MOUREC  
upper\_whisker\_MOUREC <- quantile(Dataset\_num$MOUREC,0.75,na.rm=T)+1.5\*(quantile(Dataset\_num$MOUREC,0.75, na.rm=T)-quantile(Dataset\_num$MOUREC,0.25, na.rm=T))  
upper\_whisker\_MOUREC

## 75%   
## 372.6925

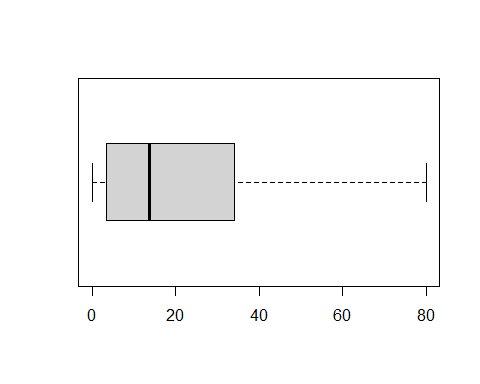
Dataset\_num$MOUREC[Dataset\_num$MOUREC>upper\_whisker\_MOUREC]<-quantile(Dataset\_num$MOUREC,0.75, na.rm = T)  
  
boxplot(Dataset\_num$MOUREC, horizontal=T)



#For OUTCALLS  
upper\_whisker\_OUTCALLS <- quantile(Dataset\_num$OUTCALLS,0.75,na.rm=T)+1.5\*(quantile(Dataset\_num$OUTCALLS,0.75, na.rm=T)-quantile(Dataset\_num$OUTCALLS,0.25, na.rm=T))  
upper\_whisker\_OUTCALLS

## 75%   
## 80.005

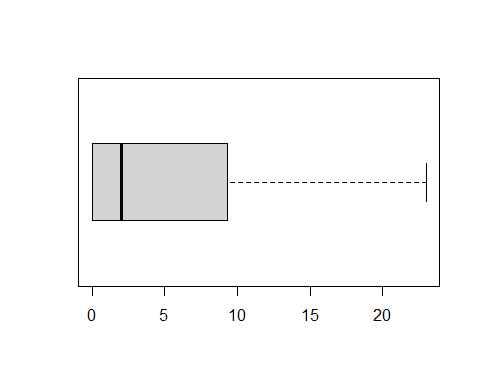
Dataset\_num$OUTCALLS[Dataset\_num$OUTCALLS>upper\_whisker\_OUTCALLS]<-quantile(Dataset\_num$OUTCALLS,0.75, na.rm = T)  
  
boxplot(Dataset\_num$OUTCALLS, horizontal=T)



#For INCALLS  
upper\_whisker\_INCALLS <- quantile(Dataset\_num$INCALLS,0.75,na.rm=T)+1.5\*(quantile(Dataset\_num$INCALLS,0.75, na.rm=T)-quantile(Dataset\_num$INCALLS,0.25, na.rm=T))  
upper\_whisker\_INCALLS

## 75%   
## 23.325

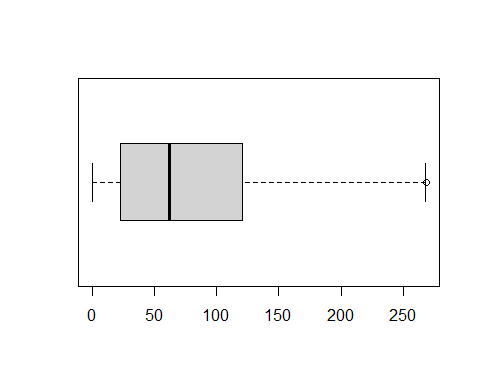
Dataset\_num$INCALLS[Dataset\_num$INCALLS>upper\_whisker\_INCALLS]<-quantile(Dataset\_num$INCALLS,0.75, na.rm = T)  
  
boxplot(Dataset\_num$INCALLS, horizontal=T)



#For PEAKVCE  
upper\_whisker\_PEAKVCE <- quantile(Dataset\_num$PEAKVCE,0.75,na.rm=T)+1.5\*(quantile(Dataset\_num$PEAKVCE,0.75, na.rm=T)-quantile(Dataset\_num$PEAKVCE,0.25, na.rm=T))  
upper\_whisker\_PEAKVCE

## 75%   
## 268.4125

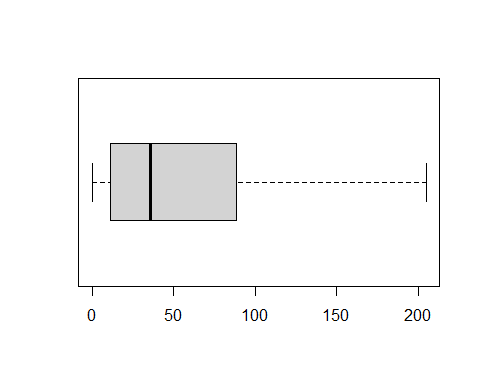
Dataset\_num$PEAKVCE[Dataset\_num$PEAKVCE>upper\_whisker\_PEAKVCE]<-quantile(Dataset\_num$PEAKVCE,0.75, na.rm = T)  
  
boxplot(Dataset\_num$PEAKVCE, horizontal=T)



#For OPEAKVCE  
upper\_whisker\_OPEAKVCE <- quantile(Dataset\_num$OPEAKVCE,0.75,na.rm=T)+1.5\*(quantile(Dataset\_num$OPEAKVCE,0.75, na.rm=T)-quantile(Dataset\_num$OPEAKVCE,0.25, na.rm=T))  
upper\_whisker\_OPEAKVCE

## 75%   
## 205.175

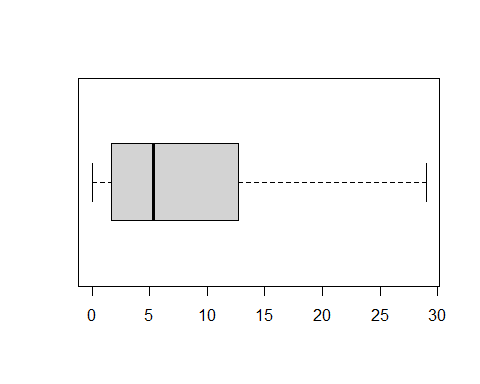
Dataset\_num$OPEAKVCE[Dataset\_num$OPEAKVCE>upper\_whisker\_OPEAKVCE]<-quantile(Dataset\_num$OPEAKVCE,0.75, na.rm = T)  
  
boxplot(Dataset\_num$OPEAKVCE, horizontal=T)



#For DROPBLK  
upper\_whisker\_DROPBLK <- quantile(Dataset\_num$DROPBLK,0.75,na.rm=T)+1.5\*(quantile(Dataset\_num$DROPBLK,0.75, na.rm=T)-quantile(Dataset\_num$DROPBLK,0.25, na.rm=T))  
upper\_whisker\_DROPBLK

## 75%   
## 29.17

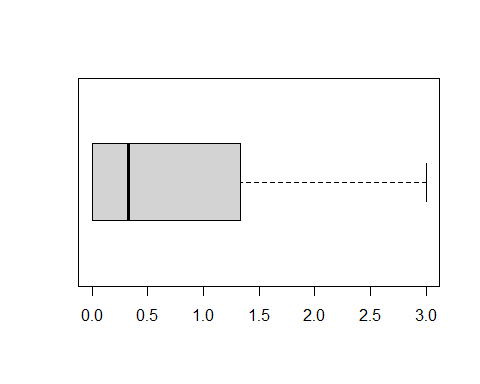
Dataset\_num$DROPBLK[Dataset\_num$DROPBLK>upper\_whisker\_DROPBLK]<-quantile(Dataset\_num$DROPBLK,0.75, na.rm = T)  
  
boxplot(Dataset\_num$DROPBLK, horizontal=T)



#For CALLWAIT  
upper\_whisker\_CALLWAIT <- quantile(Dataset\_num$CALLWAIT,0.75,na.rm=T)+1.5\*(quantile(Dataset\_num$CALLWAIT,0.75, na.rm=T)-quantile(Dataset\_num$CALLWAIT,0.25, na.rm=T))  
upper\_whisker\_CALLWAIT

## 75%   
## 3.325

Dataset\_num$CALLWAIT[Dataset\_num$CALLWAIT>upper\_whisker\_CALLWAIT]<-quantile(Dataset\_num$CALLWAIT,0.75, na.rm = T)  
  
boxplot(Dataset\_num$CALLWAIT, horizontal=T)



## Checking NA’s for numberical dataset

colSums(is.na(Dataset\_num))

## REVENUE MOU RECCHRGE DIRECTAS OVERAGE ROAM CHANGEM CHANGER   
## 216 216 216 216 216 216 502 502   
## DROPVCE BLCKVCE UNANSVCE CUSTCARE THREEWAY MOUREC OUTCALLS INCALLS   
## 0 0 0 0 0 0 0 0   
## PEAKVCE OPEAKVCE DROPBLK CALLFWDV CALLWAIT AGE1 AGE2   
## 0 0 0 0 0 1244 1244

#There are few variables such as “REVENUE”, “MOU”, “RECCHRGE”,“DIRECTAS”,“OVERAGE”,“ROAM” ,“CHANGEM”,“CHANGER”, “AGE1”, and “AGE2” where we are getting NA’s.

## Missing Value treatment

#Treat the missing values for REVENUE  
mean(Dataset\_num$REVENUE, na.rm=T)

## [1] 52.03004

Dataset\_num$REVENUE= ifelse(is.na(Dataset\_num$REVENUE),  
 mean(Dataset\_num$REVENUE, na.rm = T),Dataset\_num$REVENUE)  
any(is.na(Dataset\_num$REVENUE))

## [1] FALSE

#Treat the missing values for "MOU"  
mean(Dataset\_num$MOU, na.rm=T)

## [1] 454.5023

Dataset\_num$MOU= ifelse(is.na(Dataset\_num$MOU),  
 mean(Dataset\_num$MOU, na.rm = T),Dataset\_num$MOU)  
any(is.na(Dataset\_num$MOU))

## [1] FALSE

#Treat the missing values for "RECCHRGE"  
mean(Dataset\_num$RECCHRGE, na.rm=T)

## [1] 45.41282

Dataset\_num$RECCHRGE= ifelse(is.na(Dataset\_num$RECCHRGE),  
 mean(Dataset\_num$RECCHRGE, na.rm = T),Dataset\_num$RECCHRGE)  
any(is.na(Dataset\_num$RECCHRGE))

## [1] FALSE

#Treat the missing values for "DIRECTAS"  
mean(Dataset\_num$DIRECTAS, na.rm=T)

## [1] 0.4317024

Dataset\_num$DIRECTAS= ifelse(is.na(Dataset\_num$DIRECTAS),  
 mean(Dataset\_num$DIRECTAS, na.rm = T),Dataset\_num$DIRECTAS)  
any(is.na(Dataset\_num$DIRECTAS))

## [1] FALSE

#Treat the missing values for "OVERAGE"  
mean(Dataset\_num$OVERAGE, na.rm=T)

## [1] 18.08484

Dataset\_num$OVERAGE= ifelse(is.na(Dataset\_num$OVERAGE),  
 mean(Dataset\_num$OVERAGE, na.rm = T),Dataset\_num$OVERAGE)  
any(is.na(Dataset\_num$OVERAGE))

## [1] FALSE

#Treat the missing values for "ROAM"  
mean(Dataset\_num$ROAM, na.rm=T)

## [1] 0.08590363

Dataset\_num$ROAM= ifelse(is.na(Dataset\_num$ROAM),  
 mean(Dataset\_num$ROAM, na.rm = T),Dataset\_num$ROAM)  
any(is.na(Dataset\_num$ROAM))

## [1] FALSE

#Treat the missing values for "CHANGEM"  
mean(Dataset\_num$CHANGEM, na.rm=T)

## [1] -7.875646

Dataset\_num$CHANGEM= ifelse(is.na(Dataset\_num$CHANGEM),  
 mean(Dataset\_num$CHANGEM, na.rm = T),Dataset\_num$CHANGEM)  
any(is.na(Dataset\_num$CHANGEM))

## [1] FALSE

#Treat the missing values for "CHANGER"  
mean(Dataset\_num$CHANGER, na.rm=T)

## [1] -1.980514

Dataset\_num$CHANGER= ifelse(is.na(Dataset\_num$CHANGER),  
 mean(Dataset\_num$CHANGER, na.rm = T),Dataset\_num$CHANGER)  
any(is.na(Dataset\_num$CHANGER))

## [1] FALSE

#Treat the missing values for "AGE1"  
mean(Dataset\_num$AGE1, na.rm=T)

## [1] 31.37511

Dataset\_num$AGE1= ifelse(is.na(Dataset\_num$AGE1),  
 mean(Dataset\_num$AGE1, na.rm = T),Dataset\_num$AGE1)  
any(is.na(Dataset\_num$AGE1))

## [1] FALSE

#Treat the missing values for "AGE2"  
mean(Dataset\_num$AGE2, na.rm=T)

## [1] 21.15772

Dataset\_num$AGE2= ifelse(is.na(Dataset\_num$AGE2),  
 mean(Dataset\_num$AGE2, na.rm = T),Dataset\_num$AGE2)  
any(is.na(Dataset\_num$AGE2))

## [1] FALSE

## Again checking numerical dataset for NA’s

colSums(is.na(Dataset\_num))

## REVENUE MOU RECCHRGE DIRECTAS OVERAGE ROAM CHANGEM CHANGER   
## 0 0 0 0 0 0 0 0   
## DROPVCE BLCKVCE UNANSVCE CUSTCARE THREEWAY MOUREC OUTCALLS INCALLS   
## 0 0 0 0 0 0 0 0   
## PEAKVCE OPEAKVCE DROPBLK CALLFWDV CALLWAIT AGE1 AGE2   
## 0 0 0 0 0 0 0

#Now, numerical dataset is free from NA’s.

## Combining numerical and categorical data sets

dataset\_combined = cbind(Dataset\_num,Dataset\_cat\_factor)  
View(dataset\_combined)

## Fitting Logistic regression model

Model\_logistic\_1 = glm(CHURN~.,data=dataset\_combined,family = binomial(logit))  
summary(Model\_logistic\_1)

##   
## Call:  
## glm(formula = CHURN ~ ., family = binomial(logit), data = dataset\_combined)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.0480 -0.9658 -0.1905 0.9852 3.2360   
##   
## Coefficients: (6 not defined because of singularities)  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -3.998e+00 3.802e-01 -10.518 < 2e-16 \*\*\*  
## REVENUE 3.674e-03 7.384e-04 4.976 6.49e-07 \*\*\*  
## MOU -2.432e-04 5.042e-05 -4.822 1.42e-06 \*\*\*  
## RECCHRGE -5.902e-03 7.869e-04 -7.501 6.36e-14 \*\*\*  
## DIRECTAS -1.334e-02 1.972e-02 -0.677 0.498633   
## OVERAGE 4.248e-03 5.402e-04 7.864 3.71e-15 \*\*\*  
## ROAM -1.879e-02 7.242e-02 -0.259 0.795277   
## CHANGEM -1.007e-03 1.014e-04 -9.934 < 2e-16 \*\*\*  
## CHANGER 4.371e-03 1.982e-03 2.206 0.027395 \*   
## DROPVCE 1.302e-02 4.038e-03 3.224 0.001266 \*\*   
## BLCKVCE -3.218e-03 6.298e-03 -0.511 0.609395   
## UNANSVCE 8.979e-05 8.017e-04 0.112 0.910830   
## CUSTCARE -6.776e-02 1.183e-02 -5.727 1.02e-08 \*\*\*  
## THREEWAY -1.175e-01 5.460e-02 -2.152 0.031430 \*   
## MOUREC 4.587e-04 2.022e-04 2.268 0.023313 \*   
## OUTCALLS 2.097e-03 9.016e-04 2.326 0.020043 \*   
## INCALLS -1.677e-02 2.595e-03 -6.462 1.04e-10 \*\*\*  
## PEAKVCE -9.675e-04 3.014e-04 -3.210 0.001329 \*\*   
## OPEAKVCE -1.059e-03 3.859e-04 -2.744 0.006073 \*\*   
## DROPBLK 7.503e-03 2.891e-03 2.595 0.009466 \*\*   
## CALLFWDV -3.114e-04 2.049e-02 -0.015 0.987878   
## CALLWAIT -3.188e-03 1.750e-02 -0.182 0.855432   
## AGE1 -3.416e-03 8.184e-04 -4.174 2.99e-05 \*\*\*  
## AGE2 -1.127e-03 6.549e-04 -1.720 0.085412 .   
## UNIQSUBS2 2.557e-01 3.713e-02 6.886 5.73e-12 \*\*\*  
## UNIQSUBS3 3.991e-01 5.731e-02 6.964 3.30e-12 \*\*\*  
## UNIQSUBS4 4.610e-01 8.003e-02 5.760 8.41e-09 \*\*\*  
## UNIQSUBS5 6.571e-01 1.357e-01 4.843 1.28e-06 \*\*\*  
## UNIQSUBS6 6.857e-01 1.901e-01 3.607 0.000310 \*\*\*  
## UNIQSUBS7 1.168e+00 3.363e-01 3.472 0.000516 \*\*\*  
## UNIQSUBS8 1.120e+00 5.850e-01 1.915 0.055536 .   
## UNIQSUBS9 1.884e+00 7.378e-01 2.554 0.010649 \*   
## UNIQSUBS10 1.960e+00 1.099e+00 1.784 0.074411 .   
## UNIQSUBS11 1.380e+01 1.827e+02 0.076 0.939802   
## UNIQSUBS12 -8.052e+00 1.843e+02 -0.044 0.965154   
## UNIQSUBS13 1.377e+01 2.176e+02 0.063 0.949555   
## UNIQSUBS18 1.187e+01 3.247e+02 0.037 0.970845   
## UNIQSUBS196 1.303e+01 3.247e+02 0.040 0.968000   
## ACTVSUBS1 3.857e-01 3.667e-01 1.052 0.292883   
## ACTVSUBS2 2.061e-01 3.682e-01 0.560 0.575717   
## ACTVSUBS3 -5.608e-03 3.728e-01 -0.015 0.987998   
## ACTVSUBS4 -2.125e-01 3.881e-01 -0.548 0.583957   
## ACTVSUBS5 -1.322e+00 4.436e-01 -2.979 0.002891 \*\*   
## ACTVSUBS6 -1.191e-01 7.166e-01 -0.166 0.867939   
## ACTVSUBS7 -1.347e+01 1.872e+02 -0.072 0.942614   
## ACTVSUBS8 1.229e+00 1.752e+00 0.701 0.483004   
## ACTVSUBS9 1.096e+01 3.247e+02 0.034 0.973087   
## ACTVSUBS11 -2.182e+01 3.726e+02 -0.059 0.953297   
## ACTVSUBS53 NA NA NA NA   
## PHONES NA NA NA NA   
## CHILDREN1 9.344e-02 2.718e-02 3.438 0.000586 \*\*\*  
## CREDITA1 6.079e-02 5.742e-02 1.059 0.289726   
## CREDITAA1 7.682e-02 5.303e-02 1.449 0.147424   
## CREDITB1 9.864e-02 5.556e-02 1.775 0.075834 .   
## CREDITC1 -9.398e-02 5.819e-02 -1.615 0.106318   
## CREDITDE1 -2.733e-01 5.707e-02 -4.789 1.67e-06 \*\*\*  
## CREDITGY1 -3.912e-02 8.341e-02 -0.469 0.639026   
## CREDITZ1 NA NA NA NA   
## PRIZMRUR1 8.180e-02 4.768e-02 1.716 0.086225 .   
## PRIZMUB1 -2.954e-02 2.336e-02 -1.264 0.206136   
## PRIZMTWN1 5.296e-02 3.028e-02 1.749 0.080281 .   
## REFURB1 2.394e-01 3.016e-02 7.937 2.07e-15 \*\*\*  
## WEBCAP1 -3.384e-01 3.409e-02 -9.927 < 2e-16 \*\*\*  
## TRUCK1 4.556e-02 3.479e-02 1.310 0.190336   
## RV1 1.523e-04 4.633e-02 0.003 0.997378   
## OCCPROF1 -2.019e-02 3.140e-02 -0.643 0.520221   
## OCCCLER1 9.961e-02 7.213e-02 1.381 0.167326   
## OCCCRFT1 -1.246e-02 6.081e-02 -0.205 0.837685   
## OCCSTUD1 1.668e-01 1.172e-01 1.423 0.154602   
## OCCHMKR1 2.739e-01 1.806e-01 1.517 0.129265   
## OCCRET1 -3.604e-02 8.734e-02 -0.413 0.679875   
## OCCSELF1 -3.790e-02 7.767e-02 -0.488 0.625552   
## OWNRENT1 2.123e-02 4.119e-02 0.515 0.606280   
## MARRYUN1 8.825e-02 3.280e-02 2.691 0.007128 \*\*   
## MARRYYES1 5.926e-02 3.137e-02 1.889 0.058928 .   
## MARRYNO1 NA NA NA NA   
## MAILORD1 -3.144e-02 8.242e-02 -0.381 0.702850   
## MAILRES1 -1.086e-01 8.277e-02 -1.312 0.189636   
## MAILFLAG1 -5.495e-02 8.183e-02 -0.671 0.501914   
## TRAVEL1 1.538e-02 4.562e-02 0.337 0.736076   
## PCOWN1 1.608e-02 2.985e-02 0.539 0.590161   
## CREDITCD1 6.711e-02 4.209e-02 1.595 0.110786   
## RETCALLS1 9.247e-01 7.483e-02 12.358 < 2e-16 \*\*\*  
## RETCALLS2 1.083e+00 2.414e-01 4.486 7.27e-06 \*\*\*  
## RETCALLS3 8.094e-01 7.360e-01 1.100 0.271458   
## RETCALLS4 1.206e+01 2.241e+02 0.054 0.957080   
## RETACCPT1 -3.654e-01 1.060e-01 -3.448 0.000564 \*\*\*  
## RETACCPT2 -3.855e-02 4.595e-01 -0.084 0.933147   
## RETACCPT3 -2.745e-01 1.145e+00 -0.240 0.810527   
## RETACCPT4 -5.599e-02 3.946e+02 0.000 0.999887   
## NEWCELLY1 -4.428e-02 2.593e-02 -1.707 0.087745 .   
## NEWCELLN1 8.330e-04 2.988e-02 0.028 0.977760   
## REFER1 1.067e-02 5.008e-02 0.213 0.831217   
## REFER2 -5.290e-01 2.595e-01 -2.039 0.041491 \*   
## REFER3 7.093e-01 4.804e-01 1.477 0.139786   
## REFER4 1.242e+00 8.777e-01 1.415 0.157142   
## REFER5 -9.499e-01 1.251e+00 -0.759 0.447733   
## REFER6 -1.187e+01 1.480e+02 -0.080 0.936085   
## REFER7 -1.180e+01 1.826e+02 -0.065 0.948474   
## REFER8 -8.479e+00 3.247e+02 -0.026 0.979170   
## REFER9 1.263e+01 3.247e+02 0.039 0.968977   
## REFER11 -1.230e+01 3.247e+02 -0.038 0.969798   
## REFER16 -8.274e+00 3.247e+02 -0.025 0.979673   
## REFER35 -8.154e+00 3.247e+02 -0.025 0.979969   
## INCMISS1 2.274e-02 6.128e-02 0.371 0.710604   
## INCOME1 9.281e-02 6.142e-02 1.511 0.130734   
## INCOME2 9.279e-02 7.456e-02 1.245 0.213275   
## INCOME3 4.943e-02 5.369e-02 0.921 0.357232   
## INCOME4 3.712e-02 4.832e-02 0.768 0.442462   
## INCOME5 9.692e-03 4.696e-02 0.206 0.836500   
## INCOME6 -4.627e-03 3.861e-02 -0.120 0.904623   
## INCOME7 7.897e-02 4.209e-02 1.876 0.060631 .   
## INCOME8 -5.798e-02 5.365e-02 -1.081 0.279865   
## INCOME9 NA NA NA NA   
## MCYCLE1 7.622e-02 8.557e-02 0.891 0.373108   
## CREDITAD1 -1.219e-01 6.378e-02 -1.911 0.056069 .   
## CREDITAD2 -3.483e-01 1.478e-01 -2.358 0.018397 \*   
## CREDITAD3 -3.011e-01 2.323e-01 -1.296 0.195030   
## CREDITAD4 -1.776e-01 3.482e-01 -0.510 0.610057   
## CREDITAD5 6.249e-02 4.259e-01 0.147 0.883361   
## CREDITAD6 -1.168e+00 8.458e-01 -1.381 0.167405   
## CREDITAD7 -2.222e-01 7.588e-01 -0.293 0.769625   
## CREDITAD8 -2.188e-02 1.350e+00 -0.016 0.987073   
## CREDITAD9 1.738e+00 1.426e+00 1.219 0.223032   
## CREDITAD10 1.931e-01 1.387e+00 0.139 0.889308   
## CREDITAD11 -1.057e+01 1.352e+02 -0.078 0.937699   
## CREDITAD12 -8.168e+00 3.247e+02 -0.025 0.979933   
## CREDITAD14 -8.712e+00 3.247e+02 -0.027 0.978599   
## CREDITAD21 -8.669e+00 3.247e+02 -0.027 0.978704   
## CREDITAD25 -8.736e+00 3.247e+02 -0.027 0.978537   
## SETPRCM1 1.476e-01 2.295e-02 6.431 1.27e-10 \*\*\*  
## RETCALL1 NA NA NA NA   
## CALIBRAT1 3.915e+00 4.232e-02 92.512 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 85572 on 71046 degrees of freedom  
## Residual deviance: 59847 on 70920 degrees of freedom  
## AIC: 60101  
##   
## Number of Fisher Scoring iterations: 11

## Fitting model only with significant features

Model\_logistic\_2 = glm(CHURN~REVENUE+MOU+RECCHRGE+OVERAGE+ROAM+CHANGEM+CHANGER+DROPVCE+UNANSVCE+CUSTCARE+THREEWAY+INCALLS+PEAKVCE  
+AGE1+AGE2+CHILDREN+CREDITA+CREDITAA+CREDITB+CREDITDE+REFURB+WEBCAP+MARRYUN+MARRYYES+RETCALL+NEWCELLY+SETPRCM+CALIBRAT  
,data=dataset\_combined,family = binomial(logit))  
summary(Model\_logistic\_2)

##   
## Call:  
## glm(formula = CHURN ~ REVENUE + MOU + RECCHRGE + OVERAGE + ROAM +   
## CHANGEM + CHANGER + DROPVCE + UNANSVCE + CUSTCARE + THREEWAY +   
## INCALLS + PEAKVCE + AGE1 + AGE2 + CHILDREN + CREDITA + CREDITAA +   
## CREDITB + CREDITDE + REFURB + WEBCAP + MARRYUN + MARRYYES +   
## RETCALL + NEWCELLY + SETPRCM + CALIBRAT, family = binomial(logit),   
## data = dataset\_combined)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.9480 -0.9862 -0.1921 1.0041 3.1983   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -3.574e+00 7.018e-02 -50.922 < 2e-16 \*\*\*  
## REVENUE 3.763e-03 7.304e-04 5.151 2.59e-07 \*\*\*  
## MOU -2.162e-04 4.511e-05 -4.793 1.65e-06 \*\*\*  
## RECCHRGE -5.941e-03 7.801e-04 -7.616 2.62e-14 \*\*\*  
## OVERAGE 4.268e-03 5.372e-04 7.944 1.96e-15 \*\*\*  
## ROAM -1.037e-02 7.168e-02 -0.145 0.88491   
## CHANGEM -1.021e-03 1.010e-04 -10.107 < 2e-16 \*\*\*  
## CHANGER 4.108e-03 1.975e-03 2.081 0.03748 \*   
## DROPVCE 1.822e-02 3.172e-03 5.745 9.18e-09 \*\*\*  
## UNANSVCE -4.111e-05 7.554e-04 -0.054 0.95660   
## CUSTCARE -6.603e-02 1.166e-02 -5.661 1.51e-08 \*\*\*  
## THREEWAY -1.164e-01 5.381e-02 -2.163 0.03054 \*   
## INCALLS -1.280e-02 2.222e-03 -5.761 8.35e-09 \*\*\*  
## PEAKVCE -6.832e-04 2.757e-04 -2.478 0.01321 \*   
## AGE1 -3.811e-03 6.959e-04 -5.477 4.33e-08 \*\*\*  
## AGE2 -1.563e-03 6.291e-04 -2.484 0.01299 \*   
## CHILDREN1 8.250e-02 2.586e-02 3.190 0.00142 \*\*   
## CREDITA1 1.438e-01 3.566e-02 4.034 5.49e-05 \*\*\*  
## CREDITAA1 1.460e-01 3.013e-02 4.846 1.26e-06 \*\*\*  
## CREDITB1 1.890e-01 3.438e-02 5.499 3.83e-08 \*\*\*  
## CREDITDE1 -2.154e-01 3.800e-02 -5.668 1.44e-08 \*\*\*  
## REFURB1 2.377e-01 2.997e-02 7.930 2.19e-15 \*\*\*  
## WEBCAP1 -3.368e-01 3.384e-02 -9.951 < 2e-16 \*\*\*  
## MARRYUN1 8.344e-02 2.826e-02 2.953 0.00315 \*\*   
## MARRYYES1 4.122e-02 2.975e-02 1.385 0.16593   
## RETCALL1 7.644e-01 5.296e-02 14.435 < 2e-16 \*\*\*  
## NEWCELLY1 -5.867e-02 2.513e-02 -2.334 0.01957 \*   
## SETPRCM1 1.450e-01 2.262e-02 6.407 1.48e-10 \*\*\*  
## CALIBRAT1 3.912e+00 4.226e-02 92.568 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 85572 on 71046 degrees of freedom  
## Residual deviance: 60133 on 71018 degrees of freedom  
## AIC: 60191  
##   
## Number of Fisher Scoring iterations: 6

## Correlation Matrix

corr\_mat=cor(dataset\_combined[,c("REVENUE","MOU","RECCHRGE","OVERAGE","ROAM","CHANGEM","CHANGER","DROPVCE","UNANSVCE","CUSTCARE","THREEWAY","INCALLS","PEAKVCE","AGE1","AGE2")])  
corr\_mat

## REVENUE MOU RECCHRGE OVERAGE ROAM  
## REVENUE 1.00000000 0.62612010 0.702852725 0.51585282 0.1524590478  
## MOU 0.62612010 1.00000000 0.511476029 0.44859517 0.0992566382  
## RECCHRGE 0.70285273 0.51147603 1.000000000 0.12485395 0.0966471634  
## OVERAGE 0.51585282 0.44859517 0.124853950 1.00000000 0.0809461699  
## ROAM 0.15245905 0.09925664 0.096647163 0.08094617 1.0000000000  
## CHANGEM -0.01669546 -0.01388070 -0.006466461 -0.01557694 0.0009958237  
## CHANGER -0.17035666 -0.13799798 -0.052079347 -0.23353148 -0.0385791897  
## DROPVCE 0.43511704 0.56707705 0.347993417 0.30879567 0.0778383253  
## UNANSVCE 0.44451698 0.61378360 0.353063528 0.33403910 0.0695128565  
## CUSTCARE 0.26201208 0.39287191 0.219481542 0.20209525 0.0139911142  
## THREEWAY 0.19201233 0.24517662 0.154997911 0.15146436 0.0377315419  
## INCALLS 0.33611052 0.48381581 0.275208037 0.26350340 0.0370276390  
## PEAKVCE 0.57842395 0.69247397 0.446027785 0.41723797 0.0959147331  
## AGE1 -0.13803622 -0.18059402 -0.116679476 -0.10013921 -0.0163473023  
## AGE2 -0.12793210 -0.15298520 -0.108800767 -0.09112577 -0.0109175336  
## CHANGEM CHANGER DROPVCE UNANSVCE CUSTCARE  
## REVENUE -0.0166954591 -0.17035666 0.43511704 0.44451698 0.26201208  
## MOU -0.0138806983 -0.13799798 0.56707705 0.61378360 0.39287191  
## RECCHRGE -0.0064664614 -0.05207935 0.34799342 0.35306353 0.21948154  
## OVERAGE -0.0155769412 -0.23353148 0.30879567 0.33403910 0.20209525  
## ROAM 0.0009958237 -0.03857919 0.07783833 0.06951286 0.01399111  
## CHANGEM 1.0000000000 0.37053804 -0.04696893 -0.04000725 -0.02992923  
## CHANGER 0.3705380394 1.00000000 -0.11598041 -0.12649280 -0.09193441  
## DROPVCE -0.0469689339 -0.11598041 1.00000000 0.56957753 0.33442685  
## UNANSVCE -0.0400072536 -0.12649280 0.56957753 1.00000000 0.40414894  
## CUSTCARE -0.0299292292 -0.09193441 0.33442685 0.40414894 1.00000000  
## THREEWAY -0.0119685163 -0.04589086 0.27942671 0.26538500 0.18940902  
## INCALLS -0.0213406767 -0.08794082 0.41034632 0.44063393 0.25987197  
## PEAKVCE -0.0434135321 -0.14725932 0.61445539 0.68655770 0.36882261  
## AGE1 0.0171810382 0.04819447 -0.13882735 -0.13456041 -0.12890103  
## AGE2 0.0093554813 0.04268038 -0.12380727 -0.11447516 -0.11123095  
## THREEWAY INCALLS PEAKVCE AGE1 AGE2  
## REVENUE 0.19201233 0.33611052 0.57842395 -0.13803622 -0.127932105  
## MOU 0.24517662 0.48381581 0.69247397 -0.18059402 -0.152985199  
## RECCHRGE 0.15499791 0.27520804 0.44602779 -0.11667948 -0.108800767  
## OVERAGE 0.15146436 0.26350340 0.41723797 -0.10013921 -0.091125771  
## ROAM 0.03773154 0.03702764 0.09591473 -0.01634730 -0.010917534  
## CHANGEM -0.01196852 -0.02134068 -0.04341353 0.01718104 0.009355481  
## CHANGER -0.04589086 -0.08794082 -0.14725932 0.04819447 0.042680376  
## DROPVCE 0.27942671 0.41034632 0.61445539 -0.13882735 -0.123807274  
## UNANSVCE 0.26538500 0.44063393 0.68655770 -0.13456041 -0.114475158  
## CUSTCARE 0.18940902 0.25987197 0.36882261 -0.12890103 -0.111230950  
## THREEWAY 1.00000000 0.17666259 0.27720027 -0.05604032 -0.054227147  
## INCALLS 0.17666259 1.00000000 0.53749442 -0.13387837 -0.104607797  
## PEAKVCE 0.27720027 0.53749442 1.00000000 -0.13434431 -0.118106727  
## AGE1 -0.05604032 -0.13387837 -0.13434431 1.00000000 0.675380534  
## AGE2 -0.05422715 -0.10460780 -0.11810673 0.67538053 1.000000000

#Detection and Removal of Multicollinearity using Variance Inflation Factors. #Detection: As a rule of thumb if VIF\_j > 5 or 10 then x\_j can be taken to have strong linear relationship with the other regressiors.

## Removal: Deleting the corresponding x\_j’s will solve the problem

library(car)

## Loading required package: carData

##   
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':  
##   
## recode

vif(Model\_logistic\_2)

## REVENUE MOU RECCHRGE OVERAGE ROAM CHANGEM CHANGER DROPVCE   
## 3.402955 2.703641 2.441004 1.797506 1.030268 1.183356 1.251640 1.846550   
## UNANSVCE CUSTCARE THREEWAY INCALLS PEAKVCE AGE1 AGE2 CHILDREN   
## 2.208925 1.307071 1.120812 1.494529 2.987241 2.382773 2.268057 1.271971   
## CREDITA CREDITAA CREDITB CREDITDE REFURB WEBCAP MARRYUN MARRYYES   
## 1.861068 2.201323 1.703940 1.509311 1.138995 1.093760 1.952525 2.114195   
## RETCALL NEWCELLY SETPRCM CALIBRAT   
## 1.014619 1.010147 1.284325 1.002058

#Dropping the “REVENUE” explanatory variable from the dataset\_cobined because VIF > 10.(Assuming Threshold VIF = 10).

## Fitting model without “REVENUE” explanatory variable

Model\_logistic\_3 = glm(CHURN~MOU+RECCHRGE+OVERAGE+ROAM+CHANGEM+CHANGER+DROPVCE+UNANSVCE+CUSTCARE+THREEWAY+INCALLS+PEAKVCE  
+AGE1+AGE2+CHILDREN+CREDITA+CREDITAA+CREDITB+CREDITDE+REFURB+WEBCAP+MARRYUN+MARRYYES+RETCALL+NEWCELLY+SETPRCM+CALIBRAT  
,data=dataset\_combined,family = binomial(logit))  
summary(Model\_logistic\_3)

##   
## Call:  
## glm(formula = CHURN ~ MOU + RECCHRGE + OVERAGE + ROAM + CHANGEM +   
## CHANGER + DROPVCE + UNANSVCE + CUSTCARE + THREEWAY + INCALLS +   
## PEAKVCE + AGE1 + AGE2 + CHILDREN + CREDITA + CREDITAA + CREDITB +   
## CREDITDE + REFURB + WEBCAP + MARRYUN + MARRYYES + RETCALL +   
## NEWCELLY + SETPRCM + CALIBRAT, family = binomial(logit),   
## data = dataset\_combined)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.9469 -0.9877 -0.1923 1.0064 3.2005   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -3.537e+00 6.978e-02 -50.685 < 2e-16 \*\*\*  
## MOU -1.840e-04 4.463e-05 -4.122 3.76e-05 \*\*\*  
## RECCHRGE -3.359e-03 5.971e-04 -5.626 1.84e-08 \*\*\*  
## OVERAGE 5.573e-03 4.739e-04 11.760 < 2e-16 \*\*\*  
## ROAM 2.351e-02 7.134e-02 0.330 0.74177   
## CHANGEM -1.011e-03 1.009e-04 -10.014 < 2e-16 \*\*\*  
## CHANGER 3.564e-03 1.971e-03 1.808 0.07066 .   
## DROPVCE 1.853e-02 3.170e-03 5.847 5.01e-09 \*\*\*  
## UNANSVCE -1.319e-04 7.549e-04 -0.175 0.86131   
## CUSTCARE -6.813e-02 1.165e-02 -5.846 5.04e-09 \*\*\*  
## THREEWAY -1.179e-01 5.379e-02 -2.192 0.02839 \*   
## INCALLS -1.335e-02 2.219e-03 -6.018 1.77e-09 \*\*\*  
## PEAKVCE -5.092e-04 2.734e-04 -1.862 0.06256 .   
## AGE1 -3.790e-03 6.957e-04 -5.448 5.10e-08 \*\*\*  
## AGE2 -1.567e-03 6.289e-04 -2.492 0.01272 \*   
## CHILDREN1 8.187e-02 2.585e-02 3.167 0.00154 \*\*   
## CREDITA1 1.417e-01 3.564e-02 3.975 7.05e-05 \*\*\*  
## CREDITAA1 1.404e-01 3.010e-02 4.666 3.07e-06 \*\*\*  
## CREDITB1 1.871e-01 3.437e-02 5.444 5.20e-08 \*\*\*  
## CREDITDE1 -2.154e-01 3.798e-02 -5.669 1.43e-08 \*\*\*  
## REFURB1 2.376e-01 2.996e-02 7.929 2.20e-15 \*\*\*  
## WEBCAP1 -3.401e-01 3.383e-02 -10.054 < 2e-16 \*\*\*  
## MARRYUN1 8.405e-02 2.825e-02 2.976 0.00292 \*\*   
## MARRYYES1 4.047e-02 2.975e-02 1.360 0.17368   
## RETCALL1 7.658e-01 5.293e-02 14.468 < 2e-16 \*\*\*  
## NEWCELLY1 -5.714e-02 2.512e-02 -2.274 0.02294 \*   
## SETPRCM1 1.385e-01 2.258e-02 6.134 8.59e-10 \*\*\*  
## CALIBRAT1 3.913e+00 4.226e-02 92.589 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 85572 on 71046 degrees of freedom  
## Residual deviance: 60159 on 71019 degrees of freedom  
## AIC: 60215  
##   
## Number of Fisher Scoring iterations: 6

## New dataset without multicollinearity

library(dplyr)  
Dataset\_new =dataset\_combined %>%select\_("MOU","RECCHRGE","OVERAGE","ROAM","CHANGEM","CHANGER","DROPVCE","UNANSVCE","CUSTCARE","THREEWAY","INCALLS","PEAKVCE","AGE1","AGE2","CHILDREN","CREDITA","CREDITAA","CREDITB","CREDITDE","REFURB","WEBCAP","MARRYUN","MARRYYES","RETCALL","NEWCELLY","SETPRCM","CALIBRAT", "CHURN")  
head(Dataset\_new,5)

## MOU RECCHRGE OVERAGE ROAM CHANGEM CHANGER DROPVCE UNANSVCE CUSTCARE  
## 1 482.75 37.43 22.75 0 65.75 1.60 8.33 61.33 1.67  
## 2 1312.25 75.00 0.00 0 156.75 8.14 7.67 76.00 1.67  
## 3 25.50 29.99 0.00 0 59.50 4.03 0.00 2.33 0.00  
## 4 97.50 65.99 0.00 0 23.50 6.82 0.00 4.00 4.00  
## 5 2.50 25.00 0.00 0 -2.50 -0.23 0.00 0.33 0.00  
## THREEWAY INCALLS PEAKVCE AGE1 AGE2 CHILDREN CREDITA CREDITAA CREDITB CREDITDE  
## 1 0.33 6.33 83.670 30 0 0 0 0 0 1  
## 2 0.33 9.33 121.165 30 0 0 0 0 0 0  
## 3 0.00 0.00 1.670 52 58 0 1 0 0 0  
## 4 0.00 0.00 7.670 46 46 1 1 0 0 0  
## 5 0.00 0.00 0.670 0 0 0 1 0 0 0  
## REFURB WEBCAP MARRYUN MARRYYES RETCALL NEWCELLY SETPRCM CALIBRAT CHURN  
## 1 0 1 0 0 0 0 0 0 0  
## 2 0 1 0 0 0 1 0 0 0  
## 3 0 1 0 1 0 0 0 0 0  
## 4 0 1 0 0 0 1 0 0 0  
## 5 0 1 1 0 0 1 0 0 0

##Splitting data into Training (Development) and Testing (Validation) Dataset #install.packages(“caTools”) #Caret package that we can use to split the data #This is a package we use to break the data into training and test.

library(caTools)  
set.seed(123)  
  
split= sample.split(Dataset\_new$CHURN,SplitRatio = 2/3)  
  
train\_dataset=subset(Dataset\_new, split==T)  
  
test\_dataset=subset(Dataset\_new, split==F)

## IMPLEMENTING LOGISTIC REGRESSION

Final\_model = glm(CHURN~.,data = train\_dataset,family = binomial(logit))  
summary(Final\_model)

##   
## Call:  
## glm(formula = CHURN ~ ., family = binomial(logit), data = train\_dataset)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.9643 -0.9853 -0.1947 1.0078 3.1840   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -3.565e+00 8.513e-02 -41.877 < 2e-16 \*\*\*  
## MOU -1.734e-04 5.451e-05 -3.182 0.001464 \*\*   
## RECCHRGE -3.632e-03 7.306e-04 -4.972 6.62e-07 \*\*\*  
## OVERAGE 5.147e-03 5.794e-04 8.882 < 2e-16 \*\*\*  
## ROAM 4.751e-02 8.692e-02 0.547 0.584598   
## CHANGEM -9.632e-04 1.234e-04 -7.804 6.02e-15 \*\*\*  
## CHANGER 2.395e-03 2.418e-03 0.990 0.322011   
## DROPVCE 2.021e-02 3.897e-03 5.186 2.15e-07 \*\*\*  
## UNANSVCE 8.372e-04 9.272e-04 0.903 0.366563   
## CUSTCARE -8.095e-02 1.438e-02 -5.630 1.80e-08 \*\*\*  
## THREEWAY -8.984e-02 6.571e-02 -1.367 0.171584   
## INCALLS -1.423e-02 2.709e-03 -5.255 1.48e-07 \*\*\*  
## PEAKVCE -5.021e-04 3.353e-04 -1.497 0.134274   
## AGE1 -3.925e-03 8.510e-04 -4.612 3.98e-06 \*\*\*  
## AGE2 -1.386e-03 7.657e-04 -1.810 0.070227 .   
## CHILDREN1 5.525e-02 3.161e-02 1.748 0.080528 .   
## CREDITA1 1.616e-01 4.363e-02 3.703 0.000213 \*\*\*  
## CREDITAA1 1.659e-01 3.687e-02 4.500 6.78e-06 \*\*\*  
## CREDITB1 2.130e-01 4.201e-02 5.070 3.98e-07 \*\*\*  
## CREDITDE1 -2.113e-01 4.667e-02 -4.527 5.99e-06 \*\*\*  
## REFURB1 2.355e-01 3.667e-02 6.423 1.34e-10 \*\*\*  
## WEBCAP1 -3.230e-01 4.127e-02 -7.826 5.05e-15 \*\*\*  
## MARRYUN1 1.001e-01 3.462e-02 2.891 0.003835 \*\*   
## MARRYYES1 4.226e-02 3.639e-02 1.161 0.245561   
## RETCALL1 7.999e-01 6.462e-02 12.378 < 2e-16 \*\*\*  
## NEWCELLY1 -6.608e-02 3.077e-02 -2.148 0.031740 \*   
## SETPRCM1 1.611e-01 2.762e-02 5.832 5.47e-09 \*\*\*  
## CALIBRAT1 3.881e+00 5.125e-02 75.734 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 57047 on 47363 degrees of freedom  
## Residual deviance: 40236 on 47336 degrees of freedom  
## AIC: 40292  
##   
## Number of Fisher Scoring iterations: 6

## Coming Up with the Predicted Probabilities for test\_dataset

CHURNDEP=predict(Final\_model,newdata = test\_dataset, type="response")  
head(CHURNDEP,10)

## 2 4 5 8 11 13   
## 0.007819759 0.010263175 0.022271729 0.018853311 0.018974827 0.021998229   
## 16 20 21 22   
## 0.016948612 0.027586474 0.027053480 0.019499541

test\_dataset$CHURNDEP = CHURNDEP  
View(test\_dataset)

## CONCORDANCE

library(Metrics)

## Warning: package 'Metrics' was built under R version 4.0.5

#install.packages("InformationValue")  
library(InformationValue)

## Warning: package 'InformationValue' was built under R version 4.0.5

##   
## Attaching package: 'InformationValue'

## The following object is masked from 'package:Metrics':  
##   
## precision

# Concordance  
Concordance(test\_dataset$CHURN,test\_dataset$CHURNDEP)

## $Concordance  
## [1] 0.8268479  
##   
## $Discordance  
## [1] 0.1731521  
##   
## $Tied  
## [1] 0  
##   
## $Pairs  
## [1] 115505310

#The concordance value is 83% (approx.). The concordance value indicates that model fit is good.

## AUC

#install.packages("pROC")  
library(pROC)

## Warning: package 'pROC' was built under R version 4.0.5

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

##   
## Attaching package: 'pROC'

## The following object is masked from 'package:Metrics':  
##   
## auc

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

roc\_obj <- roc(test\_dataset$CHURN, test\_dataset$CHURNDEP)

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

roc\_obj

##   
## Call:  
## roc.default(response = test\_dataset$CHURN, predictor = test\_dataset$CHURNDEP)  
##   
## Data: test\_dataset$CHURNDEP in 16813 controls (test\_dataset$CHURN 0) < 6870 cases (test\_dataset$CHURN 1).  
## Area under the curve: 0.8268

auc(roc\_obj)

## Area under the curve: 0.8268

#The Area Under the Curve (AUC) is the measure of the ability of a classifier to distinguish between classes and is used as a summary of the ROC curve. The higher the AUC, the better the performance of the model at distinguishing between the positive and negative classes. #Model performance is good with 83% (approx.) accuracy.

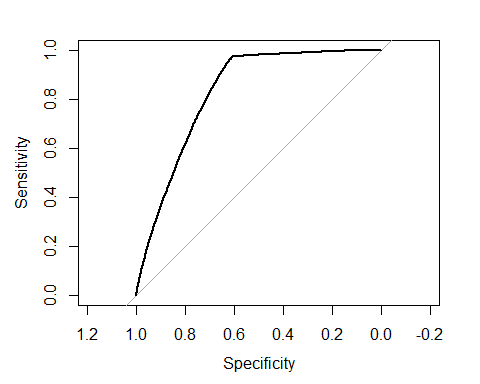
## FINDING THE BEST VALUE OF THRESHOLD

# Method: pROC  
roc\_obj <- roc(test\_dataset$CHURN, test\_dataset$CHURNDEP)

## Setting levels: control = 0, case = 1

## Setting direction: controls < cases

plot(roc\_obj)



x <- coords(roc\_obj, "best", "threshold", transpose = TRUE)  
x

## threshold specificity sensitivity   
## 0.2659997 0.6097663 0.9708879

#We consider the threshold value 0.1283726.

## Labelling

test\_dataset$CHURNDEP <- ifelse(test\_dataset$CHURNDEP > 0.1283726,1,0)  
head(test\_dataset,5)

## MOU RECCHRGE OVERAGE ROAM CHANGEM CHANGER DROPVCE UNANSVCE CUSTCARE  
## 2 1312.25 75.00 0.00 0 156.75 8.14 7.67 76.00 1.67  
## 4 97.50 65.99 0.00 0 23.50 6.82 0.00 4.00 4.00  
## 5 2.50 25.00 0.00 0 -2.50 -0.23 0.00 0.33 0.00  
## 8 153.00 29.99 16.00 0 30.00 7.35 2.00 17.33 0.00  
## 11 299.50 24.99 0.75 0 -47.50 -0.50 2.00 14.33 1.00  
## THREEWAY INCALLS PEAKVCE AGE1 AGE2 CHILDREN CREDITA CREDITAA CREDITB  
## 2 0.33 9.33 121.165 30 0 0 0 0 0  
## 4 0.00 0.00 7.670 46 46 1 1 0 0  
## 5 0.00 0.00 0.670 0 0 0 1 0 0  
## 8 0.00 4.67 48.330 46 0 1 1 0 0  
## 11 0.33 5.00 45.670 0 0 0 0 0 1  
## CREDITDE REFURB WEBCAP MARRYUN MARRYYES RETCALL NEWCELLY SETPRCM CALIBRAT  
## 2 0 0 1 0 0 0 1 0 0  
## 4 0 0 1 0 0 0 1 0 0  
## 5 0 0 1 1 0 0 1 0 0  
## 8 0 0 1 0 0 0 0 0 0  
## 11 0 0 1 0 1 0 1 0 0  
## CHURN CHURNDEP  
## 2 0 0  
## 4 0 0  
## 5 0 0  
## 8 0 0  
## 11 0 0

## Confusion Matrix

#install.packages("caret")  
library(caret)

## Warning: package 'caret' was built under R version 4.0.5

## Loading required package: lattice

## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 4.0.4

##   
## Attaching package: 'caret'

## The following objects are masked from 'package:InformationValue':  
##   
## confusionMatrix, precision, sensitivity, specificity

## The following objects are masked from 'package:Metrics':  
##   
## precision, recall

confusionMatrix(as.factor(test\_dataset$CHURNDEP), as.factor(test\_dataset$CHURN))

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction 0 1  
## 0 10222 194  
## 1 6591 6676  
##   
## Accuracy : 0.7135   
## 95% CI : (0.7077, 0.7193)  
## No Information Rate : 0.7099   
## P-Value [Acc > NIR] : 0.1131   
##   
## Kappa : 0.4546   
##   
## Mcnemar's Test P-Value : <2e-16   
##   
## Sensitivity : 0.6080   
## Specificity : 0.9718   
## Pos Pred Value : 0.9814   
## Neg Pred Value : 0.5032   
## Prevalence : 0.7099   
## Detection Rate : 0.4316   
## Detection Prevalence : 0.4398   
## Balanced Accuracy : 0.7899   
##   
## 'Positive' Class : 0   
##

## Computation of sensitivity and specificity

# Senstivity  
Senstivity = sensitivity(as.factor(test\_dataset$CHURNDEP),as.factor(test\_dataset$CHURN))  
Senstivity

## [1] 0.6079819

# Specificity  
specificity = specificity(as.factor(test\_dataset$CHURNDEP),as.factor(test\_dataset$CHURN))  
specificity

## [1] 0.9717613

#Sensitivity and specificity should be as large as possible.

## F1 Score

#install.packages("MLmetrices")  
library(MLmetrics)

## Warning: package 'MLmetrics' was built under R version 4.0.5

##   
## Attaching package: 'MLmetrics'

## The following objects are masked from 'package:caret':  
##   
## MAE, RMSE

## The following object is masked from 'package:base':  
##   
## Recall

F1\_Score(test\_dataset$CHURN, test\_dataset$CHURNDEP)

## [1] 0.7508171

#Decision Rule:

#F1 score = 0; Model is bad.

#F1 score !=0, > 0; Model is good.

#F1 score = 1; Perfect Model.

#Therefore, The model fitting is good with 75% F1 score.

## CONCLUSION

#The key factors are “MOU”, “RECCHRGE”, “OVERAGE”, “ROAM”, “CHANGEM”, “CHANGER”, “DROPVCE”, “UNANSVCE”, “CUSTCARE”, “THREEWAY”, “INCALLS”, “PEAKVCE”, “AGE1”, “AGE2” , “CHILDREN”, “CREDITA”, “CREDITAA”, “CREDITB”, “CREDITDE”, “REFURB”, “WEBCAP”, “MARRYUN”, “MARRYYES”, “RETCALL”, “NEWCELLY”, “SETPRCM”, “CALIBRAT”, “CHURN” ,“CHURNDEP”. These are imporatant factor or significant variables for the prediction of customer churn or customer attrition. These factors can lead to customer attrition. Customer attrition happens when a business losses a customer for whatever reason. It is a normal part of the customer life cycle. By actively tracking customers on the basis of the determined factors, we can take a proactive approach to identify the reason why a customer is leaving and win them back. This insight is also very useful to understand the demand trends in the market and ensure that customers are happy and less likely to churn.