df1 <-read.csv('data.csv',header = TRUE)

df <- read.csv('data.csv',header = TRUE)

unique(df$Address)

str(df$Address)

unique(df$C.O)

# according to our pre knowledge (0 is 'Markazi')&(13 is 'Hamedan')

# &(25 is 'Qom')&(26 is 'Qazvin')&(30 is 'Alborz')

for (i in 1:nrow(df)) {

if(is.na(df$Daramad\_Mozd\_Month[i])){

df$Daramad\_month\_menhaie\_year[i] = NA

}

else df$Daramad\_month\_menhaie\_year[i]= 12\*(df$Daramad\_Mozd\_Month[i])- df$Daramad\_Mozd\_Year[i]

}

ggplot(df, aes(x= Daramad\_kol\_Mounth))+

xlim(0, 200000000)+

geom\_histogram()

nrow(df[is.na(df$Daramad\_Azad),])

#1960 Na value in Daramad\_Azad which we will consider them as 1

for (i in 1:nrow(df)) {

if (is.na(df$Daramad\_Azad[i])) {

df$Daramad\_Azad[i]= 1

}

}

nrow(df[which(df$Daramad\_Azad==1),])

str(df$Daramad\_Azad)

#In Following Lines we will do the same thing to other

for (i in 1:nrow(df)) {

if (is.na(df$Daramad\_Yarane[i])) {

df$Daramad\_Yarane[i]= 1

}

}

for (i in 1:nrow(df)) {

if (is.na(df$Daramad\_Motefaraghe[i])) {

df$Daramad\_Motefaraghe[i]= 1

}

}

for (i in 1:nrow(df)) {

if (is.na(df$Daramad\_Mozd\_Month[i])) {

df$Daramad\_Mozd\_Month[i]= 1

}

}

for (i in 1:nrow(df)) {

if (is.na(df$Daramad\_Mozd\_Year[i])) {

df$Daramad\_Mozd\_Year[i]= 1

}

}

Daramad\_month\_menhaie\_year<- df$Daramad\_month\_menhaie\_year

df$Daramad\_month\_menhaie\_year <- NULL

#Makinig y

df$Daramad\_kol\_Mounth <-( df$Daramad\_Azad+ df$Daramad\_Motefaraghe+ df$Daramad\_Mozd\_Year+ df$Daramad\_Yarane)/12

quantile(df$Daramad\_kol\_Mounth,probs = 0.7)

unique(df$C.O)

str(df$C.O)

table(df$C.O)

for (i in 1:nrow(df)) {

if (df$C.O[i]== 0) {

df$C.O[i] = 'Markazi'

}else if (df$C.O[i]==13) {

df$C.O[i]= 'Hamedan'

}else if (df$C.O[i]==25){

df$C.O[i]= 'Qom'

}else if (df$C.O[i]==26) {

df$C.O[i]= 'Qazvin'

}else if (df$C.O[i]==30) {

df$C.O[i]= 'Alborz'

}

}

quantile(df$Daramad\_kol\_Mounth,probs = 0.7)

# = 24464000

for (i in 1:nrow(df)) {

if (df$Daramad\_kol\_Mounth[i] > 24464000) {

df$Cat.Daramad\_kol\_Mounth[i]= 1

}else df$Cat.Daramad\_kol\_Mounth[i]= 0

}

#visualization

options(scipen = 999)

library(ggplot2)

ggplot(df, aes(x= C.O))+

geom\_bar(width = .6)+

ggtitle('نمودار 1')

C.O1<- aggregate(df[,73], by= list(df$C.O), FUN= mean)

ggplot(C.O1, aes(x= Group.1, y= x))+

geom\_bar(stat = 'identity')+

xlab('Ostan ha')+

ylab('miangine daramad kol dar mah')+

ggtitle('نمودار 2')

C.O2<- aggregate(df[,74],by= list(df$C.O), FUN= mean)

df$Cat.Daramad\_kol\_Mounth <- as.factor(df$Cat.Daramad\_kol\_Mounth)

ggplot(C.O2, aes(x= Group.1, y= x))+

geom\_bar(stat = 'identity')+

xlab('Ostan ha')+

ylab('Darsad taloq be 3 dahak bartar')+

ggtitle('نمودار 3')

str(df$Jens)

for (i in 1:nrow(df)) {

if (df$Jens[i]==2) {

df$Jens[i]=0

}

}

table(df$Jens)

#0 female and 1 is male

df$Jens <- as.factor(df$Jens)

ggplot(df, aes(x= Jens))+

geom\_bar(width = 0.6)+

ggtitle('نمودار 4')

jens1<- aggregate(df[,73],by= list(df$Jens),FUN= mean)

ggplot(jens1,aes(x= Group.1, y=x))+

geom\_bar(width = 0.6, stat = 'identity')+

xlab('Jens')+

ylab('Miangine daramad kol mahiane')+

ggtitle('نمودار 5')

table(df$Sen)

ggplot(df, aes(x= Sen))+

geom\_histogram(binwidth =5)+

ggtitle('نمودار 6 ')

ggplot(df, aes(x= Cat.Daramad\_kol\_Mounth, y= Sen))+

geom\_boxplot()+

ggtitle('6.1')

ggplot(df, aes(x= Sen, y= Daramad\_kol\_Mounth))+

geom\_point(color= 'navy', alpha= 0.2)+

ylim(0,100000000)+

ggtitle('نمودار 7')

quantile(df$Sen, probs = c(0.2, 0.4, 0.6, 0.8))

Quantile\_Sen= array()

for (i in 1:nrow(df)) {

if (df$Sen[i]<= 35) {

df$Quantile\_Sen[i] = 'less than 35'

}else if (df$Sen[i]>35 &df$Sen[i]<=43) {

df$Quantile\_Sen[i]='betwen 35 and 43'

}else if (df$Sen[i]>43 & df$Sen[i]<=53) {

df$Quantile\_Sen[i]= 'between 43 and 53'

}else if (df$Sen[i]>53& df$Sen[i]<=64) {

df$Quantile\_Sen[i]= 'between 53 and 64'

}else if (df$Sen[i]>64) {

df$Quantile\_Sen[i]= 'more than 64'

}

}

Quantile\_Sen

unique(df$Quantile\_Sen)

quantil1<- aggregate(df[,73],by= list(df$Quantile\_Sen),FUN= mean)

q1 <- quantil1[4,]

q2 <- quantil1[3,]

q3 <- quantil1[1,]

q4 <- quantil1[2,]

q5 <- quantil1[5,]

quantil11<- rbind(q1,q2,q3,q4,q5)

ggplot(quantil11,aes(x=Group.1, y= x))+

geom\_bar(stat = 'identity')+

xlab('Goroohe senni')+

ylab('miangine daramad kol mahane')+

ggtitle('نمودار 8')

ggplot(df, aes(x= Quantile\_Sen, fill= factor(Cat.Daramad\_kol\_Mounth)))+

geom\_bar(width = .5)+

labs(fill='Cat.Daramad\_kol\_Mounth' )+

ggtitle('نمودار 8.1')

for (i in 1:nrow(df)) {

if (df$Quantile\_Sen[i]=='less than 35') {

df$Quantile\_Sen[i] = 28

}else if (df$Quantile\_Sen[i]=='betwen 35 and 43') {

df$Quantile\_Sen[i]= 40

}else if (df$Quantile\_Sen[i]== 'between 43 and 53') {

df$Quantile\_Sen[i]= 48

}else if (df$Quantile\_Sen[i]== 'between 53 and 64') {

df$Quantile\_Sen[i]= 58

}else if (df$Quantile\_Sen[i]== 'more than 64') {

df$Quantile\_Sen[i] = 78

}

}

unique(df$Quantile\_Sen)

str(df$Quantile\_Sen)

df$Quantile\_Sen <- as.integer(df$Quantile\_Sen)

str(df$Quantile\_Sen)

#Savad 1= ba savad va 2 = bi savad

for (i in 1:nrow(df)) {

if (df$Savad[i]==2) {

df$Savad[i]= 0

}

}

#Savad 1 mean darad , savad 0 means nadarad

df$Savad <- as.factor(df$Savad)

ggplot(df, aes(x= Savad))+

geom\_bar(width = 0.5)+

ggtitle('نمودار 9 ')

Savad1 <- aggregate(df[,73],by= list(df$Savad),FUN= mean)

ggplot(Savad1, aes(x= Group.1, y= x))+

geom\_bar(width = .5, stat = 'identity')+

xlab('Savad')+

ylab('Miangin darmad kolle mahane')+

ggtitle('نمودار 10')

#tahsil mikonad:

nulltahsilat <- df[which(is.na(df$Tahsil.Mikonad)),]

table(df$Tahsil.Mikonad)

for (i in 1:nrow(df)) {

if (is.na(df$Tahsil.Mikonad[i])) {

if(df$Savad[i]==0){

df$Tahsil.Mikonad[i]= 0

}

}else if (df$Tahsil.Mikonad[i] ==2) {

df$Tahsil.Mikonad[i]= 0

}

}

table(df$Tahsil.Mikonad)

df$Tahsil.Mikonad <- as.factor(df$Tahsil.Mikonad)

ggplot(df, aes(x= Tahsil.Mikonad))+

geom\_bar(width = .5)+

ggtitle('نمودار 11')

Tahsil.1<- aggregate(df[,73],by= list(df$Tahsil.Mikonad), FUN= mean)

ggplot(Tahsil.1, aes(x= Group.1,y =x))+

geom\_bar( width = .5, stat = 'identity')+

xlab('Tahsi.Mikonad')+

ylab('Miangin daramad mahane kol')+

ggtitle('نمودار 12')

#Madrak variable:

NA.Madrak<- df[which(is.na(df$Madrak)),c(7,9)]

for (i in 1:nrow(df)) {

if (is.na(df$Madrak[i])) {

df$Madrak[i]= 0

}else if (df$Madrak[i]==9|df$Madrak[i]==1) {

df$Madrak[i]= '1& 9'

}else if (df$Madrak[i]==8|df$Madrak[i]==7) {

df$Madrak[i]= '7& 8'

}else if (df$Madrak[i]==3|df$Madrak[i]==4) {

df$Madrak[i]= '3& 4'

}

}

unique(df$Madrak)

df$Madrak <- as.factor(df$Madrak)

ggplot(df, aes(x= Madrak))+

geom\_bar()+

ggtitle('نمودار 13')

Madrak.1 <- aggregate(df[,73],by= list(df$Madrak), FUN= mean)

ggplot(Madrak.1, aes(x= Group.1, y= x))+

geom\_bar(stat = 'identity')+

xlab('Madrak')+

ylab('Miangin daramad kol mahiane')+

ggtitle('نمودار 14')

# faaliat variable

faaaliat.Nul <- df[which(is.na(df$Faaliat)),c(4,5,6,7,8,9)]

#So no Null value is in faaliat variable

for (i in 1:nrow(df)) {

if (df$Faaliat[i]== 1) {

df$Faaliat[i]= 'Shaghel'

}else if (df$Faaliat[i]== 3) {

df$Faaliat[i]= 'Daraie daramad bedune kar'

}else if (df$Faaliat[i]== 6| df$Faaliat[i]== 5|df$Faaliat[i]== 2|df$Faaliat[i]== 4) {

df$Faaliat[i]= 'Sayer'

}

}

unique(df$Faaliat)

table(df$Faaliat)

#since we just hav one person who is Mohassel we put it in Sayer

ggplot(df, aes(x= Faaliat))+

geom\_bar()+

ggtitle('نمودار 15')

Faaliat.1<- aggregate(df$Daramad\_kol\_Mounth,by= list(df$Faaliat),FUN= mean)

ggplot(Faaliat.1, aes(x= Group.1, y=x))+

geom\_bar(stat = 'identity')+

xlab('Faaliat')+

ylab('Miangine darmada kolle mahiane')+

ggtitle('نمودار 16')

#Zanashooi variable

table(df$Zanashoi)

for (i in 1:nrow(df)) {

if (df$Zanashoi[i]== 1) {

df$Zanashoi[i]= 'daraye Hamsar'

}else if (df$Zanashoi[i]== 2) {

df$Zanashoi[i]= 'Fote hamsar'

}else if (df$Zanashoi[i]== 3) {

df$Zanashoi[i]= 'talaghe hamsar'

}else if (df$Zanashoi[i]== 4) {

df$Zanashoi[i]= 'ezdevaj nakarde'

}

}

unique(df$Zanashoi)

table(df$Zanashoi)

ggplot(df, aes(x= Zanashoi))+

geom\_bar()+

ggtitle('نمودار 17')

zanashoi.1<- aggregate(df$Daramad\_kol\_Mounth,by=list(df$Zanashoi,df$Jens),FUN= mean)

zanashoi.2<- aggregate(df$Daramad\_kol\_Mounth,by=list(df$Zanashoi,df$Jens),FUN= mean, drop= FALSE)

ggplot(zanashoi.1, aes(x= Group.1, y= x))+

geom\_bar(stat = 'identity')+

xlab('Zanashoi')+

ylab('Miangine darmada koll')+

ggtitle('نمودار 18 ')

ggplot(zanashoi.1, aes(x= Group.1, y= x))+

geom\_bar(stat = 'identity')+

facet\_wrap(~Group.2)+

xlab('Zanashoi')+

ylab('Miangine darmada koll')+

ggtitle('نمودار 19 Jens')

#tedad.a variable

alaki <- df[which(df$tedad.a==10),]

str(df$tedad.a)

sort(unique(df$tedad.a))

table(df$tedad.a)

for (i in 1:nrow(df)) {

if (df$tedad.a[i]>=6) {

df$tedad.a[i]= 6

}

}

unique(df$tedad.a)

table(df$tedad.a)

for (i in 1:nrow(df)) {

if (df$tedad.a[i]==6) {

df$tedad.a[i]= 'greater than 6'

}

}

ggplot(df, aes(x= tedad.a))+

geom\_bar()+

ggtitle('نمو دار 20')

tedad.1<- aggregate(df$Daramad\_kol\_Mounth,by= list(df$tedad.a),FUN= mean)

ggplot(tedad.1, aes(x= Group.1,y= x))+

geom\_bar(stat = 'identity')+

xlab('tedad.a')+

ylab('Miangine kolle daramad mahiane')+

ggtitle('نمودار 21')

for (i in 1:nrow(df)) {

if (df$tedad.a[i]== 3| df$tedad.a[i]== 4) {

df$tedad.a[i]= '3& 4'

} else if (df$tedad.a[i]== 5| df$tedad.a[i]== 'greater than 6') {

df$tedad.a[i]= 'greater than 5'

}

}

unique(df$tedad.a)

#Now n.t.m variable

#1 mean Melki arse o ayan

str(df$n.t.m)

table(df$n.t.m)

for (i in 1:nrow(df)) {

if (df$n.t.m[i]== 1) {

df$n.t.m[i]= 'Melki-Arse-Ayan'

}else if (df$n.t.m[i]== 2) {

df$n.t.m[i]= 'Melki-Ayan'

}else if (df$n.t.m[i]== 3) {

df$n.t.m[i]= 'Ejari'

}else if (df$n.t.m[i]== 4) {

df$n.t.m[i]= 'Rahn'

}else if (df$n.t.m[i]== 5) {

df$n.t.m[i]= 'Khedmat'

}else if (df$n.t.m[i]== 6) {

df$n.t.m[i]= 'Raygan'

}else df$n.t.m[i]= 'Sayer'

}

ggplot(df, aes(x= n.t.m))+

geom\_bar()+

ggtitle('نمودار22')

table(df$n.t.m)

n.tm.1 <- aggregate(df$Daramad\_kol\_Mounth,by= list(df$n.t.m),FUN= mean)

ggplot(n.tm.1, aes(x= Group.1, y= x))+

geom\_bar(stat = 'identity')+

xlab('n.t.m')+

ylab('Miangin Daramad Kol mahane')+

ggtitle('نمودار 23')

for (i in 1:nrow(df)) {

if (df$n.t.m[i]=='Sayer'|df$n.t.m[i]== 'Raygan') {

df$n.t.m[i]= 'Sayer&Raygan'

}else if (df$n.t.m[i]== 'Khedmat'| df$n.t.m[i]== 'Melki-Ayan') {

df$n.t.m[i]= 'Khedmat&Melki-Ayan'

}

}

table(df$n.t.m)

#t.o variable

unique(df$t.o)

ggplot(df, aes(x= factor(df$t.o)))+

geom\_bar()+

ggtitle('نمودار 24')

table(df$t.o)

t.o.1 <- aggregate(df$Daramad\_kol\_Mounth,by= list(df$t.o),FUN= mean)

ggplot(t.o.1,aes(x=factor(Group.1), y= x))+

geom\_bar(stat = 'identity')+

xlab('t.O')+

ylab('Miangine kole daramad mahiane')+

ggtitle('نمودار 25')

t.o.2<- aggregate(df$Daramad\_kol\_Mounth,by=list(df$t.o, df$n.t.m),drop= FALSE,FUN= mean)

ggplot(t.o.2, aes(x= factor(Group.1), y= x))+

geom\_bar(stat = 'identity')+

facet\_wrap(~factor(Group.2))+

xlab('T.O')+

ylab('Miangine daramad kolle mahane')+

ggtitle('نحوه تصرف منزل نمودار26')

unique(df$t.o)

for (i in 1:nrow(df)) {

if (df$t.o[i]== 1| df$t.o[i]== 2) {

df$t.o[i]= '1 or 2'

}else if (df$t.o[i]>= 7) {

df$t.o[i] = '7 or greater'

}

}

#S.Z sathe zir bana variable

ggplot(df, aes(x= s.z, y= Daramad\_kol\_Mounth))+

geom\_point(alpha= 0.6)+

ylim(0, 10000000)+

ggtitle('نمودار 27')

#n.e noe eskelete bana

unique(df$n.e)

table(df$n.e)

for (i in 1:nrow(df)) {

if (df$n.e[i]== 1) {

df$n.e[i]= 'Felezi'

}else if (df$n.e[i]== 2) {

df$n.e[i]= 'Beton-Arme'

}else if (df$n.e[i]== 3) {

df$n.e[i]= 'Sayer'

}

}

ggplot(df, aes(x= n.e))+

geom\_bar(width = .5)+

ggtitle('نمودار28')

n.e.1 <- aggregate(df$Daramad\_kol\_Mounth,by= list(df$n.e),FUN= mean)

ggplot(n.e.1, aes(x= Group.1, y= x))+

geom\_bar(width = 0.5, stat = 'identity')+

xlab('n.e')+

ylab('miangine daramad kol mahane')+

ggtitle('نمودار 29')

#m.o.b: masale omde bana variable

for (i in 1:nrow(df)) {

if (is.na(df$m.o.b[i])) {

df$m.o.b[i]='NULL'

}

}

unique(df$m.o.b)

table(df$m.o.b)

ggplot(df,aes(x= m.o.b))+

geom\_bar(width = .5)+

ggtitle('نمودار 30')

# otoo variable

for (i in 1:nrow(df)) {

if (is.na(df$oto[i])) {

df$oto[i]= 0

}

}

tabel(df$oto)

df$oto <- as.factor(df$oto)

ggplot(df, aes(x= oto))+

geom\_bar(width = 0.5)+

ggtitle('نمودار 31')

oto.1<- aggregate(df$Daramad\_kol\_Mounth,by=list(df$oto), mean)

ggplot(oto.1, aes(x= Group.1, y= x))+

geom\_bar(width = 0.5, stat = 'identity')+

xlab('oto')+

ylab('miangine darmad kolli mahane')+

ggtitle('نمودار 32')

ggplot(df, aes(x= oto, y= Daramad\_kol\_Mounth))+

geom\_boxplot()+

ylim(0,100000000)+

ggtitle('نمودار33')

box.oto<- ggplot(df, aes(x= oto, y= Daramad\_kol\_Mounth))+

geom\_boxplot()+

ylim(0,100000000)

#mo: motor variable

for (i in 1:nrow(df)) {

if (is.na(df$mo[i])) {

df$mo[i]= 0

}

}

df$mo <- as.factor(df$mo)

ggplot(df, aes(x= mo))+

geom\_bar(width = 0.5)+

ggtitle('نمودار 34')

box.mo<- ggplot(df, aes(x= mo, y= Daramad\_kol\_Mounth))+

geom\_boxplot()+

ylim(0,100000000)

#do: docharkhe variable

for (i in 1:nrow(df)) {

if (is.na(df$do[i])) {

df$do[i]= 0

}

}

df$do <- as.factor(df$do)

ggplot(df, aes(x= do))+

geom\_bar(width = 0.5)+

ggtitle('نمودار 35')

box.do<- ggplot(df, aes(x= do, y= Daramad\_kol\_Mounth))+

geom\_boxplot()+

ylim(0,100000000)

#radio variable

for (i in 1:nrow(df)) {

if (is.na(df$radio[i])) {

df$radio[i]= 0

}

}

df$radio <- as.factor(df$radio)

ggplot(df, aes(x= radio))+

geom\_bar(width = 0.5)+

ggtitle('نمودار 36')

box.radio<- ggplot(df, aes(x= radio, y= Daramad\_kol\_Mounth))+

geom\_boxplot()+

ylim(0,100000000)

#tv.s : siah sefid variable

for (i in 1:nrow(df)) {

if (is.na(df$tv.s[i])) {

df$tv.s[i]= 0

}

}

df$tv.s <- as.factor(df$tv.s)

ggplot(df, aes(x= tv.s))+

geom\_bar(width = 0.5)+

ggtitle('نمودار 37')

box.tv.s<- ggplot(df, aes(x= tv.s, y= Daramad\_kol\_Mounth))+

geom\_boxplot()+

ylim(0,100000000)

#tv.r : rangi variable

for (i in 1:nrow(df)) {

if (is.na(df$tv.r[i])) {

df$tv.r[i]= 0

}

}

df$tv.r <- as.factor(df$tv.r)

ggplot(df, aes(x= tv.r))+

geom\_bar(width = 0.5)+

ggtitle('نمودار 38')

box.tv.r<- ggplot(df, aes(x= tv.r, y= Daramad\_kol\_Mounth))+

geom\_boxplot()+

ylim(0,100000000)

# video : variable

for (i in 1:nrow(df)) {

if (is.na(df$video[i])) {

df$video[i]= 0

}

}

df$video <- as.factor(df$video)

ggplot(df, aes(x= video))+

geom\_bar(width = 0.5)+

ggtitle('نمودار 39')

box.video<- ggplot(df, aes(x= video, y= Daramad\_kol\_Mounth))+

geom\_boxplot()+

ylim(0,100000000)

#pc variable

for (i in 1:nrow(df)) {

if (is.na(df$pc[i])) {

df$pc[i]= 0

}

}

df$pc <- as.factor(df$pc)

ggplot(df, aes(x= pc))+

geom\_bar(width = 0.5)+

ggtitle('نمودار 40')

box.pc<- ggplot(df, aes(x= pc, y= Daramad\_kol\_Mounth))+

geom\_boxplot()+

ylim(0,100000000)

#mobile variable

for (i in 1:nrow(df)) {

if (is.na(df$mobile[i])) {

df$mobile[i]= 0

}

}

df$mobile <- as.factor(df$mobile)

ggplot(df, aes(x= mobile))+

geom\_bar(width = 0.5)+

ggtitle('نمودار 41')

box.mobile<- ggplot(df, aes(x= mobile, y= Daramad\_kol\_Mounth))+

geom\_boxplot()+

ylim(0,100000000)

#freeizer: freezer

for (i in 1:nrow(df)) {

if (is.na(df$freeizer[i])) {

df$freeizer[i]= 0

}

}

df$freeizer <- as.factor(df$freeizer)

ggplot(df, aes(x= freeizer))+

geom\_bar(width = 0.5)+

ggtitle('نمودار 42')

box.freeizer<- ggplot(df, aes(x= freeizer, y= Daramad\_kol\_Mounth))+

geom\_boxplot()+

ylim(0,100000000)

#yakhchal variable

for (i in 1:nrow(df)) {

if (is.na(df$yakhchal[i])) {

df$yakhchal[i]= 0

}

}

df$yakhchal <- as.factor(df$yakhchal)

ggplot(df, aes(x= yakhchal))+

geom\_bar(width = 0.5)+

ggtitle('نمودار 43')

box.yakhchal<- ggplot(df, aes(x= yakhchal, y= Daramad\_kol\_Mounth))+

geom\_boxplot()+

ylim(0,100000000)

#yakhchal.f

for (i in 1:nrow(df)) {

if (is.na(df$yakhchal.f[i])) {

df$yakhchal.f[i]= 0

}

}

df$yakhchal.f <- as.factor(df$yakhchal.f)

ggplot(df, aes(x= yakhchal.f))+

geom\_bar(width = 0.5)+

ggtitle('نمودار 44')

box.yakhchal.f<- ggplot(df, aes(x= yakhchal.f, y= Daramad\_kol\_Mounth))+

geom\_boxplot()+

ylim(0,100000000)

#gaz variable

for (i in 1:nrow(df)) {

if (is.na(df$gaz[i])) {

df$gaz[i]= 0

}

}

df$gaz <- as.factor(df$gaz)

ggplot(df, aes(x= gaz))+

geom\_bar(width = 0.5)+

ggtitle('نمودار 45')

box.gaz<- ggplot(df, aes(x= gaz, y= Daramad\_kol\_Mounth))+

geom\_boxplot()+

ylim(0,100000000)

#jaro.b variable

for (i in 1:nrow(df)) {

if (is.na(df$jaro.b[i])) {

df$jaro.b[i]= 0

}

}

df$jaro.b <- as.factor(df$jaro.b)

ggplot(df, aes(x= jaro.b))+

geom\_bar(width = 0.5)+

ggtitle('نمودار 46')

box.jaro.b<- ggplot(df, aes(x= jaro.b, y= Daramad\_kol\_Mounth))+

geom\_boxplot()+

ylim(0,100000000)

#m.lebas : mashine lebas shooi

for (i in 1:nrow(df)) {

if (is.na(df$m.lebas[i])) {

df$m.lebas[i]= 0

}

}

df$m.lebas <- as.factor(df$m.lebas)

ggplot(df, aes(x= m.lebas))+

geom\_bar(width = 0.5)+

ggtitle('نمودار 47')

box.m.lebas<- ggplot(df, aes(x= m.lebas, y= Daramad\_kol\_Mounth))+

geom\_boxplot()+

ylim(0,100000000)

# charkh.kh : khayyati variable

for (i in 1:nrow(df)) {

if (is.na(df$charkh.kh[i])) {

df$charkh.kh[i]= 0

}

}

df$charkh.kh <- as.factor(df$charkh.kh)

ggplot(df, aes(x= charkh.kh))+

geom\_bar(width = 0.5)+

ggtitle('نمودار 48')

box.charkh.kh<- ggplot(df, aes(x= charkh.kh, y= Daramad\_kol\_Mounth))+

geom\_boxplot()+

ylim(0,100000000)

#panke variable

for (i in 1:nrow(df)) {

if (is.na(df$panke[i])) {

df$panke[i]= 0

}

}

df$panke <- as.factor(df$panke)

ggplot(df, aes(x= panke))+

geom\_bar(width = 0.5)+

ggtitle('نمودار 49')

box.panke<- ggplot(df, aes(x= panke, y= Daramad\_kol\_Mounth))+

geom\_boxplot()+

ylim(0,100000000)

#cooler.a: cooler abi

for (i in 1:nrow(df)) {

if (is.na(df$cooler.a[i])) {

df$cooler.a[i]= 0

}

}

df$cooler.a <- as.factor(df$cooler.a)

ggplot(df, aes(x= cooler.a))+

geom\_bar(width = 0.5)+

ggtitle('نمودار 50')

box.cooler.a<- ggplot(df, aes(x= cooler.a, y= Daramad\_kol\_Mounth))+

geom\_boxplot()+

ylim(0,100000000)

#cooler.g variable

for (i in 1:nrow(df)) {

if (is.na(df$cooler.g[i])) {

df$cooler.g[i]= 0

}

}

df$cooler.g <- as.factor(df$cooler.g)

ggplot(df, aes(x= cooler.g))+

geom\_bar(width = 0.5)+

ggtitle('نمودار 51')

box.cooler.g<- ggplot(df, aes(x= cooler.g, y= Daramad\_kol\_Mounth))+

geom\_boxplot()+

ylim(0,100000000)

#m.zarf: zarfshoi variable

for (i in 1:nrow(df)) {

if (is.na(df$m.zarf[i])) {

df$m.zarf[i]= 0

}

}

df$m.zarf <- as.factor(df$m.zarf)

ggplot(df, aes(x= m.zarf))+

geom\_bar(width = 0.5)+

ggtitle('نمودار 52')

box.m.zarf<- ggplot(df, aes(x= m.zarf, y= Daramad\_kol\_Mounth))+

geom\_boxplot()+

ylim(0,100000000)

#microfer variable

for (i in 1:nrow(df)) {

if (is.na(df$microfer[i])) {

df$microfer[i]= 0

}

}

df$microfer <- as.factor(df$microfer)

ggplot(df, aes(x= microfer))+

geom\_bar(width = 0.5)+

ggtitle('نمودار 53')

box.microfer<- ggplot(df, aes(x= microfer, y= Daramad\_kol\_Mounth))+

geom\_boxplot()+

ylim(0,100000000)

#ab.l: ab loolekeshi

for (i in 1:nrow(df)) {

if (is.na(df$ab.l[i])) {

df$ab.l[i]= 0

}

}

df$ab.l <- as.factor(df$ab.l)

ggplot(df, aes(x= ab.l))+

geom\_bar(width = 0.1)+

ggtitle('نمودار 54')

box.ab.l<- ggplot(df, aes(x= ab.l, y= Daramad\_kol\_Mounth))+

geom\_boxplot()+

ylim(0,100000000)

# bargh is like ab.l

#gaz.l is like the jjus a few 0 so we shuld delete it soon

for (i in 1:nrow(df)) {

if (is.na(df$gaz.l[i])) {

df$gaz.l[i]= 0

}

}

#tel variable

for (i in 1:nrow(df)) {

if (is.na(df$tel[i])) {

df$tel[i]= 0

}

}

df$tel <- as.factor(df$tel)

ggplot(df, aes(x= tel))+

geom\_bar(width = 0.5)+

ggtitle('نمودار 54')

box.tel<- ggplot(df, aes(x= tel, y= Daramad\_kol\_Mounth))+

geom\_boxplot()+

ylim(0,100000000)

#internet

for (i in 1:nrow(df)) {

if (is.na(df$internet[i])) {

df$internet[i]= 0

}

}

df$internet <- as.factor(df$internet)

ggplot(df, aes(x= internet))+

geom\_bar(width = 0.5)+

ggtitle('نمودار 55')

box.internet<- ggplot(df, aes(x= internet, y= Daramad\_kol\_Mounth))+

geom\_boxplot()+

ylim(0,100000000)

# hamam variable

for (i in 1:nrow(df)) {

if (is.na(df$hamam[i])) {

df$hamam[i]= 0

}

}

df$hamam <- as.factor(df$hamam)

ggplot(df, aes(x= hamam))+

geom\_bar(width = 0.5)+

ggtitle('نمودار 56')

box.hamam<- ggplot(df, aes(x= hamam, y= Daramad\_kol\_Mounth))+

geom\_boxplot()+

ylim(0,100000000)

#cooler.a.s

for (i in 1:nrow(df)) {

if (is.na(df$cooler.a.s[i])) {

df$cooler.a.s[i]= 0

}

}

df$cooler.a.s <- as.factor(df$cooler.a.s)

ggplot(df, aes(x= cooler.a.s))+

geom\_bar(width = 0.5)+

ggtitle('نمودار 57')

box.cooler.a.s<- ggplot(df, aes(x= cooler.a.s, y= Daramad\_kol\_Mounth))+

geom\_boxplot()+

ylim(0,100000000)

#broodat.m is almost always 0 so we shuld delete it

for (i in 1:nrow(df)) {

if (is.na(df$broodat.m[i])) {

df$broodat.m[i]= 0

}

}

#hararat.m variable

for (i in 1:nrow(df)) {

if (is.na(df$hararat.m[i])) {

df$hararat.m[i]= 0

}

}

df$hararat.m <- as.factor(df$hararat.m)

ggplot(df, aes(x= hararat.m))+

geom\_bar(width = 0.5)+

ggtitle('نمودار 58')

box.hararat.m<- ggplot(df, aes(x= hararat.m, y= Daramad\_kol\_Mounth))+

geom\_boxplot()+

ylim(0,100000000)

#package variable

for (i in 1:nrow(df)) {

if (is.na(df$package[i])) {

df$package[i]= 0

}

}

df$package <- as.factor(df$package)

ggplot(df, aes(x= package))+

geom\_bar(width = 0.5)+

ggtitle('نمودار 59')

box.package<- ggplot(df, aes(x= package, y= Daramad\_kol\_Mounth))+

geom\_boxplot()+

ylim(0,100000000)

#cooler.g.s

for (i in 1:nrow(df)) {

if (is.na(df$cooler.g.s[i])) {

df$cooler.g.s[i]= 0

}

}

df$cooler.g.s <- as.factor(df$cooler.g.s)

ggplot(df, aes(x= cooler.g.s))+

geom\_bar(width = 0.5)+

ggtitle('نمودار 60')

box.cooler.g.s<- ggplot(df, aes(x= cooler.g.s, y= Daramad\_kol\_Mounth))+

geom\_boxplot()+

ylim(0,100000000)

#fazelab variable

for (i in 1:nrow(df)) {

if (is.na(df$fazelab[i])) {

df$fazelab[i]= 0

}

}

df$fazelab <- as.factor(df$fazelab)

ggplot(df, aes(x= fazelab))+

geom\_bar(width = 0.5)+

ggtitle('نمودار 61')

box.fazelab<- ggplot(df, aes(x= fazelab, y= Daramad\_kol\_Mounth))+

geom\_boxplot()+

ylim(0,100000000)

#zabt variable

for (i in 1:nrow(df)) {

if (is.na(df$zabt[i])) {

df$zabt[i]= 0

}

}

df$zabt <- as.factor(df$zabt)

ggplot(df, aes(x= zabt))+

geom\_bar(width = 0.5)+

ggtitle('نمودار 62')

box.zabt<- ggplot(df, aes(x= zabt, y= Daramad\_kol\_Mounth))+

geom\_boxplot()+

ylim(0,100000000)

# box plots

library(gridExtra)

grid.arrange(box.oto, box.mo, box.do, box.radio, box.zabt, box.tv.s, ncol= 3)

grid.arrange(box.tv.r,box.video, box.pc, box.mobile, box.freeizer, box.yakhchal, ncol= 3)

grid.arrange(box.yakhchal.f, box.gaz, box.jaro.b, box.m.lebas, box.charkh.kh, box.panke, ncol= 3)

grid.arrange(box.cooler.a, box.cooler.g,box.m.zarf, box.microfer, box.tel, box.internet, ncol= 3)

grid.arrange(box.hamam, box.cooler.a.s, box.hararat.m, box.package, box.cooler.g.s,

box.fazelab, ncol= 3)

grid.arrange(box.m.lebas, box.charkh.kh, box.panke(box.cooler.a, box.cooler.g,

box.m.zarf, box.microfer, box.tel, box.internet,

box.hamam, box.cooler.a.s, box.hararat.m, box.package, box.cooler.g.s,

box.fazelab, ncol= 5, top= 'نمودار 64')

#sookht.p variable

table(df$sookht.p)

for (i in 1:nrow(df)) {

if (df$sookht.p[i]== 3) {

df$sookht.p[i]= 'gaz maye'

}else if (df$sookht.p[i]== 4) {

df$sookht.p[i]= 'gaz tabiE'

}else if (df$sookht.p[i]== 5) {

df$sookht.p[i]= 'bargh'

}

}

ggplot(df, aes(x= sookht.p))+

geom\_bar(width = .5)+

ggtitle('نمودار 65')

#sookht.g variable

table(df$sookht.g)

for (i in 1:nrow(df)) {

if (df$sookht.g[i]== 11) {

df$sookht.g[i]= 'nafte-sefid'

}else if (df$sookht.g[i]== 13) {

df$sookht.g[i]= 'gaz-maye'

}else if (df$sookht.g[i]== 14) {

df$sookht.g[i]= 'gaze tabie'

}else if (df$sookht.g[i]== 15) {

df$sookht.g[i]= 'bargh'

}

}

ggplot(df, aes(x= sookht.g))+

geom\_bar(width = 0.5)+

ggtitle('نمودار 66')

#sookht.ab

table(df$sookht.ab)

#this is like sookht.g and and sookht.p

#Hazine ha

#hazine behdasht

for (i in 1:nrow(df)) {

if (is.na(df$Hazine\_Behdashti[i])) {

df$Hazine\_Behdashti[i]= 1

}

}

ggplot(df, aes(x= Hazine\_Behdashti))+

geom\_histogram()+

xlim(0, 20000000)+

ggtitle('نمودار 66')

#It is not normal distribution:

ggplot(df, aes(x= log10(Hazine\_Behdashti)))+

geom\_histogram()+

# xlim(0, 20000000)+

ggtitle('نمودار 66.1')

ggplot(df, aes(x= Hazine\_Behdashti, y= Daramad\_kol\_Mounth))+

geom\_point(alpha= 0.3,colour= 'navy')+

xlim(0, 25000000)+

ylim(0,200000000)+

ggtitle('نمودار 67')

ggplot(df, aes(x= Cat.Daramad\_kol\_Mounth, y= Hazine\_Behdashti))+

geom\_boxplot()+

ylim(0, 2000000)+

ggtitle('نمودار 67')

#Hazine\_ertebatat

for (i in 1:nrow(df)) {

if (is.na(df$Hazine\_Ertebatat[i])) {

df$Hazine\_Ertebatat[i]= 1

}

}

ggplot(df, aes(x= Hazine\_Ertebatat))+

geom\_histogram()+

xlim(0,3000000)+

ggtitle('نمودار 68')

ggplot(df, aes(x= log10(Hazine\_Ertebatat)))+

geom\_histogram()+

# xlim(0,3000000)+

ggtitle('نمودار 68.1')

ggplot(df, aes(x= Cat.Daramad\_kol\_Mounth, y= Hazine\_Ertebatat))+

geom\_boxplot()+

ylim(0,3000000)+

ggtitle('نمودار 69')

# Hazine\_Ghazayeamade varibale:

for (i in 1:nrow(df)) {

if (is.na(df$Hazine\_Ghazayeamade[i])) {

df$Hazine\_Ghazayeamade[i]= 1

}

}

mean(df$Hazine\_Ghazayeamade)

range(df$Hazine\_Ghazayeamade)

ggplot(df, aes(x= Hazine\_Ghazayeamade))+

geom\_histogram()+

xlim(0,10000000)+

ylim(0,550)+

ggtitle('نمودار 70')

ggplot(df, aes(x= log10(Hazine\_Ghazayeamade)))+

geom\_histogram()+

# xlim(0,10000000)+

# ylim(0,550)+

ggtitle('نمودار 70.1')

ggplot(df, aes(x= Cat.Daramad\_kol\_Mounth, y= Hazine\_Ghazayeamade))+

geom\_boxplot()+

ylim(0,1000000)+

ggtitle('نمودار 71')

#Hazine\_Hamlonaghl variable

for (i in 1:nrow(df)) {

if (is.na(df$Hazine\_Hamlonaghl[i])) {

df$Hazine\_Hamlonaghl[i]= 1

}

}

mean(df$Hazine\_Hamlonaghl)

range(df$Hazine\_Hamlonaghl)

ggplot(df, aes(x= Hazine\_Hamlonaghl))+

geom\_histogram()+

# xlim(0,10000000)+

# ylim(0,550)+

ggtitle('نمودار 72')

ggplot(df, aes(x= log10(Hazine\_Hamlonaghl)))+

geom\_histogram()+

# xlim(0,10000000)+

# ylim(0,550)+

ggtitle('نمودار 72.1')

ggplot(df, aes(x= Cat.Daramad\_kol\_Mounth, y= Hazine\_Hamlonaghl))+

geom\_boxplot()+

ylim(0,2700000)+

ggtitle('نمودار 73')

#Hazine\_kalavakhedmat variable

for (i in 1:nrow(df)) {

if (is.na(df$Hazine\_kalavakhadamat[i])) {

df$Hazine\_kalavakhadamat[i]= 1

}

}

mean(df$Hazine\_kalavakhadamat)

range(df$Hazine\_kalavakhadamat)

ggplot(df, aes(x= Hazine\_kalavakhadamat))+

geom\_histogram()+

xlim(0,7500000)+

# ylim(0,550)+

ggtitle('نمودار 74')

ggplot(df, aes(x= log10(Hazine\_kalavakhadamat)))+

geom\_histogram()+

# xlim(0,7500000)+

# ylim(0,550)+

ggtitle('نمودار 74.1')

ggplot(df, aes(x= Cat.Daramad\_kol\_Mounth, y= Hazine\_kalavakhadamat))+

geom\_boxplot()+

ylim(0,2500000)+

ggtitle('نمودار 75')

#Hazine\_khorakivadokhani

for (i in 1:nrow(df)) {

if (is.na(df$Hazine\_Khorakivadokhani[i])) {

df$Hazine\_Khorakivadokhani[i]= 1

}

}

mean(df$Hazine\_Khorakivadokhani)

range(df$Hazine\_Khorakivadokhani)

ggplot(df, aes(x= Hazine\_Khorakivadokhani))+

geom\_histogram()+

xlim(0,40000000)+

# ylim(0,550)+

ggtitle('نمودار 76')

ggplot(df, aes(x= log10(Hazine\_Khorakivadokhani)))+

geom\_histogram()+

# xlim(0,40000000)+

# ylim(0,550)+

ggtitle('نمودار 76.1')

ggplot(df, aes(x= Cat.Daramad\_kol\_Mounth, y= Hazine\_Khorakivadokhani))+

geom\_boxplot()+

ylim(1500000,3500000)+

ggtitle('نمودار 77')

#Hazine\_lavamekhanegi variable

for (i in 1:nrow(df)) {

if (is.na(df$Hazine\_lavazemkhanegi[i])) {

df$Hazine\_lavazemkhanegi[i]= 1

}

}

mean(df$Hazine\_lavazemkhanegi)

range(df$Hazine\_lavazemkhanegi)

ggplot(df, aes(x= Hazine\_lavazemkhanegi))+

geom\_histogram()+

xlim(0,2500000)+

# ylim(0,550)+

ggtitle('نمودار 78')

ggplot(df, aes(x= log10(Hazine\_lavazemkhanegi)))+

geom\_histogram()+

# xlim(0,2500000)+

# ylim(0,550)+

ggtitle('نمودار 78.1')

ggplot(df, aes(x= Cat.Daramad\_kol\_Mounth, y= Hazine\_lavazemkhanegi))+

geom\_boxplot()+

ylim(0,1000000)+

ggtitle('نمودار 79')

#rahn variable:

for (i in 1:nrow(df)) {

if (is.na(df$Rahn[i])|df$Rahn[i]==0) {

df$Rahn[i]= 1

}

}

mean(df$Rahn)

range(df$Rahn)

ggplot(df, aes(x= Rahn))+

geom\_histogram()+

xlim(0,500000000)+

ylim(0,350)+

ggtitle('نمودار 80')

ggplot(df, aes(x= log10(Rahn)))+

geom\_histogram()+

# xlim(0,500000000)+

# ylim(0,350)+

ggtitle('نمودار 80.1')

ggplot(df, aes(x= Cat.Daramad\_kol\_Mounth, y= log10(Rahn)))+

geom\_boxplot()+

ylim(0,10)+

ggtitle('نمودار 80.2 ')

quantile(df$Rahn,probs = .3)

Rahn1 = df[which(df$Rahn>=96000.04),]

cat.1.Rahn2<- df[which(df$Cat.Daramad\_kol\_Mounth==1 & df$Rahn>2),]

aggregate(df$Rahn, by= list(df$Cat.Daramad\_kol\_Mounth), FUN = mean)

#hazine maskan variable:

ggplot(df, aes(x= Hazine\_Maskan))+

geom\_histogram()+

# xlim(0,500000000)+

# ylim(0,350)+

ggtitle('نمودار 81')

ggplot(df, aes(x= log10(Hazine\_Maskan)))+

geom\_histogram()+

# xlim(0,500000000)+

# ylim(0,350)+

ggtitle('نمودار 81.1')

ggplot(df, aes(x= Cat.Daramad\_kol\_Mounth, y= Hazine\_Maskan))+

geom\_boxplot()+

ylim(0,20000000)+

ggtitle(' نمودار 82')

#hazine noshidani

for (i in 1:nrow(df)) {

if (is.na(df$Hazine\_Noshidani[i])) {

df$Hazine\_Noshidani[i]= 1

}

}

mean(df$Hazine\_Noshidani)

range(df$Hazine\_Noshidani)

ggplot(df, aes(x= Hazine\_Noshidani))+

geom\_histogram()+

xlim(0,4000000)+

ylim(0,150)+

ggtitle('نمودار83')

ggplot(df, aes(x= log10(Hazine\_Noshidani)))+

geom\_histogram()+

# xlim(0,4000000)+

# ylim(0,150)+

ggtitle('نمودار 83.1')

ggplot(df, aes(x= Cat.Daramad\_kol\_Mounth, y= Hazine\_Noshidani))+

geom\_boxplot()+

# ylim(0,1000000)+

ggtitle('')

#hazinepushak variable

for (i in 1:nrow(df)) {

if (is.na(df$Hazine\_Pushak[i])) {

df$Hazine\_Pushak[i]= 1

}

}

ggplot(df, aes(x= Hazine\_Pushak))+

geom\_histogram()+

xlim(0, 20000000)+

ggtitle('نمودار 84')

ggplot(df, aes(x= log10(Hazine\_Pushak)))+

geom\_histogram()+

# xlim(0, 20000000)+

ggtitle('نمودار 84.1')

ggplot(df, aes(x= Cat.Daramad\_kol\_Mounth, y= Hazine\_Pushak))+

geom\_boxplot()+

ylim(0, 2000000)+

ggtitle('نمودار 85')

#Hazine\_tafrihat variable

for (i in 1:nrow(df)) {

if (is.na(df$Hazine\_Tafrihat[i])) {

df$Hazine\_Tafrihat[i]= 1

}

}

ggplot(df, aes(x= Hazine\_Tafrihat))+

geom\_histogram()+

ylim(0,200)+

xlim(0, 8000000)+

ggtitle('نمودار 86')

ggplot(df, aes(x= log10(Hazine\_Tafrihat)))+

geom\_histogram()+

# ylim(0,200)+

# xlim(0, 8000000)+

ggtitle('نمودار 86.1')

ggplot(df, aes(x= Cat.Daramad\_kol\_Mounth, y= Hazine\_Tafrihat))+

geom\_boxplot()+

ylim(0, 500000)+

ggtitle('نمودار 87')

#Heat-Map

#making correlation matrix:

names(df)

#sen, tedad.a, t.o, s.z, hazine\_(behdashti, ertebati, ghazaieamade, hamlonaghl,

#kalava khedamat, khorakivadokhani, lavazemkhanegi)

for (i in 1:nrow(df1)) {

if (is.na(df1$Hazine\_Behdashti[i])) {

df1$Hazine\_Behdashti[i]= 0

}

}

for (i in 1:nrow(df1)) {

if (is.na(df1$Hazine\_Ertebatat[i])) {

df1$Hazine\_Ertebatat[i]= 0

}

}

for (i in 1:nrow(df1)) {

if (is.na(df1$Hazine\_Ghazayeamade[i])) {

df1$Hazine\_Ghazayeamade[i]= 0

}

}

for (i in 1:nrow(df1)) {

if (is.na(df1$Hazine\_Hamlonaghl[i])) {

df1$Hazine\_Hamlonaghl[i]= 0

}

}

for (i in 1:nrow(df1)) {

if (is.na(df1$Hazine\_kalavakhadamat[i])) {

df1$Hazine\_kalavakhadamat[i]= 0

}

}

for (i in 1:nrow(df1)) {

if (is.na(df1$Hazine\_Khorakivadokhani[i])) {

df1$Hazine\_Khorakivadokhani[i]= 0

}

}

for (i in 1:nrow(df1)) {

if (is.na(df1$Hazine\_lavazemkhanegi[i])) {

df1$Hazine\_lavazemkhanegi[i]= 0

}

}

for (i in 1:nrow(df1)) {

if (is.na(df1$Rahn[i])) {

df1$Rahn[i]= 0

}

}

for (i in 1:nrow(df1)) {

if (is.na(df1$Hazine\_Noshidani[i])) {

df1$Hazine\_Noshidani[i]= 0

}

}

for (i in 1:nrow(df1)) {

if (is.na(df1$Hazine\_Pushak[i])) {

df1$Hazine\_Pushak[i]= 0

}

}

for (i in 1:nrow(df1)) {

if (is.na(df1$Hazine\_Tafrihat[i])) {

df1$Hazine\_Tafrihat[i]= 0

}

}

df1$Hazine\_koll <- df1$Hazine\_Behdashti+ df1$Hazine\_Ertebatat+ df1$Hazine\_Ghazayeamade+ df1$Hazine\_Hamlonaghl+ df1$Hazine\_kalavakhadamat+ df1$Hazine\_Khorakivadokhani+ df1$Hazine\_lavazemkhanegi+ df1$Hazine\_Maskan+ df1$Hazine\_Noshidani+ df1$Hazine\_Pushak+ df1$Hazine\_Tafrihat+ df1$Rahn

df.numeric<- df1[,c('Sen','tedad.a','t.o', 's.z', 'Hazine\_Behdashti', 'Hazine\_Ertebatat',

'Hazine\_Ghazayeamade', 'Hazine\_Hamlonaghl', 'Hazine\_kalavakhadamat',

'Hazine\_Khorakivadokhani', 'Hazine\_lavazemkhanegi', 'Rahn', 'Hazine\_Maskan',

'Hazine\_Noshidani', 'Hazine\_Pushak', 'Hazine\_Tafrihat','Hazine\_koll')]

df1$Hazine\_koll <- df1$Hazine\_Behdashti+ df1$Hazine\_Ertebatat+ df1$Hazine\_Ghazayeamade+ df1$Hazine\_Hamlonaghl+ df1$Hazine\_kalavakhadamat+ df1$Hazine\_Khorakivadokhani+ df1$Hazine\_lavazemkhanegi+ df1$Hazine\_Maskan+ df1$Hazine\_Noshidani+ df1$Hazine\_Pushak+ df1$Hazine\_Tafrihat+ df1$Rahn

df$Hazine\_Kol

str(df.numeric$Rahn)

for (i in 1:16) {

print (typeof(df.numeric[i,i]))

}

str(df.numeric$tedad.a)

cor.matrix <- cor(df.numeric, method = 'pearson')

library(reshape)

cor.matrix <- round(cor.matrix, 2)

melted.cor.mat <- melt(cor.matrix)

ggplot(melted.cor.mat, aes(x= X1, y= X2, fill= value))+

geom\_tile()+

geom\_text(aes(x= X1, y= X2, label= value))+

ggtitle('نمودار 88')

df$Hazine\_Kol <- df1$Hazine\_koll

for (i in 1:nrow(df)) {

if (df$Hazine\_Kol[i]== 0) {

df$Hazine\_Kol[i]= 1

}

}

ggplot(df, aes(x= log10(Hazine\_Kol)))+

geom\_histogram()+

ggtitle('نمودار فراوانی برای لگاریتم هزینه کل')

ggplot(df, aes(x= Cat.Daramad\_kol\_Mounth, y= Hazine\_Kol))+

geom\_boxplot()+

ylim(0, 100000000)+

ggtitle('نمودار جعبه ای هزینه کل')

#PCA

pca <- prcomp(data.frame(df.numeric$t.o, df.numeric$s.z), scale= TRUE)

pca

summary(pca)

options(scipen = 999)

pca$sdev

pca$rotation[2,1]

t.o.scale <- scale(df.numeric$t.o)

s.z.scale <- scale(df.numeric$s.z)

ss<- pca$rotation[1,1]\* t.o.scale+ pca$rotation[2,1]\* s.z.scale

df$c\_t.oands.z<- ss[,1]

class(df$c\_t.oands.z)

str(df$c\_t.oands.z)

# seeing predective accuracy measures and charts od predictio err distribution FASLE 5:

# set.seed(2564)

# library(caTools)

# split <- sample.split(df$Daramad\_kol\_Mounth, SplitRatio = 0.7)

# split

#

# train <- subset(df, split== TRUE)

# test <- subset(df, split== FALSE)

# train$m.o.b <- NULL

# train$ab.l <- NULL

# train$bargh<- NULL

# train$Tahsil.Mikonad <- NULL

# train$Savad <- NULL

# train$tv.s <- NULL

# train$tv.r <- NULL

# train$gaz <- NULL

# train$cooler.a <- NULL

# train$cooler.g <- NULL

# train$gaz.l <- NULL

# train$hamam <- NULL

# train$ashpazkhane <- NULL

# train$broodat.m <- NULL

# train$hararat.m <- NULL

# train$sookht.p <- NULL

# train$sookht.g <- NULL

# train$sookht.ab <- NULL

# train$Sen <- NULL

# train$Rahn <- NULL

# train$Address <- NULL

# train$MahMorajeh <- NULL

# train$Fasl <- NULL

# library(forecast)

# Model <- lm(formula = Daramad\_kol\_Mounth ~ ., data = train[,-c(45:49,51)])

# Model

# #a<- lapply(df.notnumeric, unique)

# summary(Model)

# pred.train <- predict(Model, train)

# pred.train <- as.numeric(pred.train)

# rmse.train <- sqrt(mean(pred.train- train$Daramad\_kol\_Mounth)^2)

# rmse

# MAPE.train<- 100\*mean(abs((pred.train - train$Daramad\_kol\_Mounth)/train$Daramad\_kol\_Mounth))

# MAPE.train

# accuracy(pred.train, train$Daramad\_kol\_Mounth)

#

#

#

# pred.test <- predict(Model, test[,which(names(test)%in% names(train))])

# pred.test <- as.numeric(pred.test)

# rmse.test <- sqrt(mean(pred.test- test$Daramad\_kol\_Mounth)^2)

# rmse.test

#

# MAPE.test<- 100\*mean(abs((pred.test- test$Daramad\_kol\_Mounth)/test$Daramad\_kol\_Mounth))

# MAPE.test

# accuracy(pred.test, test$Daramad\_kol\_Mounth)

# train$Daramad\_kol\_Mounth[978]

# train$error <- pred.train- train$Daramad\_kol\_Mounth

# pred.train <- as.numeric(pred.train)

# test

# ggplot(train, aes(x= error))+

# geom\_histogram(binwidth = 300000)+

# scale\_x\_continuous(breaks = round(seq(-100000000,100000000, by= 5)))

#

#LOgestic reg

#View(df)

set.seed(2564)

library(caret)

library(caTools)

split <- sample.split(df$Daramad\_kol\_Mounth, SplitRatio = 0.7)

split

train <- subset(df, split== TRUE)

train.copy<- df[split== TRUE, which(names(df)%in%names(train))]

test <- subset(df, split== FALSE)

train$m.o.b <- NULL

train$ab.l <- NULL

train$bargh<- NULL

train$Tahsil.Mikonad <- NULL

train$Savad <- NULL

train$tv.s <- NULL

train$tv.r <- NULL

train$gaz <- NULL

train$cooler.a <- NULL

train$cooler.g <- NULL

train$gaz.l <- NULL

train$hamam <- NULL

train$ashpazkhane <- NULL

train$broodat.m <- NULL

train$hararat.m <- NULL

train$sookht.p <- NULL

train$sookht.g <- NULL

train$sookht.ab <- NULL

train$Sen <- NULL

train$Rahn <- NULL

train$Address <- NULL

train$MahMorajeh <- NULL

train$Fasl <- NULL

#View(train)

nrow(train)+ nrow(test)== nrow(df)

library(forecast)

str(train$Cat.Daramad\_kol\_Mounth)

#logestic <- glm(Cat.Daramad\_kol\_Mounth~., data = train[,c(1, 2, 5, 7, 52, 3, 6, 8, 11, 17, 22, 23, 27, 29, 34, 35, 37, 41, 44, 53, 51)], family = "binomial")

logestic <- glm(Cat.Daramad\_kol\_Mounth~., data = train[,c(1, 2, 5, 7, 52, 3, 6, 8, 11, 17, 22, 23, 27, 29, 34, 35, 37, 41, 44, 53, 51)], family = "binomial")

logestic

summary(logestic)

#train set

glm.probs.train <- predict(logestic, newdata = train, type= 'response')

glm.probs.train

glm.pred.train <- ifelse(glm.probs.train > 0.5, 1, 0)

glm.pred.train

table(glm.pred.train, train$Cat.Daramad\_kol\_Mounth)

aa <- table(glm.pred.train, train$Cat.Daramad\_kol\_Mounth)

(aa[1,1]+ aa[2,2])/(aa[1,1]+ aa[2,2]+ aa[1,2]+ aa[2,1])

#Test set

glm.probs.test = predict(logestic, newdata = test, type = "response")

glm.probs.test

glm.pred.test <- ifelse(glm.probs.test > 0.5, 1, 0)

glm.pred.test

table(glm.pred.test, test$Cat.Daramad\_kol\_Mounth)

bb <- table(glm.pred.test, test$Cat.Daramad\_kol\_Mounth)

(bb[1,1]+ bb[2,2])/(bb[1,1]+ bb[2,2]+ bb[1,2]+ bb[2,1])

# acc <- 0.7927273

logestic.back <- glm(Cat.Daramad\_kol\_Mounth~., data = train[,-c(45:50)], family = "binomial")

logestic.back

step(logestic.back, direction = 'backward')

logestic.back <- glm(formula = Cat.Daramad\_kol\_Mounth ~ C.O + Jens + Madrak +

tedad.a + n.t.m + oto + zabt + video + pc + yakhchal + m.lebas +

charkh.kh + m.zarf + tel + internet + cooler.g.s + Hazine\_Ertebatat +

Hazine\_Ghazayeamade + Hazine\_Khorakivadokhani + Hazine\_lavazemkhanegi +

Hazine\_Maskan + Hazine\_Pushak + Quantile\_Sen, family = "binomial",

data = train[, -c(45:50)])

glm.probs.train.back <- predict(logestic.back, newdata = train, type= 'response')

glm.probs.train.back

glm.pred.train.back <- ifelse(glm.probs.train.back > 0.5, 1, 0)

glm.pred.train.back

table(glm.pred.train.back, train$Cat.Daramad\_kol\_Mounth)

cc <- table(glm.pred.train.back, train$Cat.Daramad\_kol\_Mounth)

(cc[1,1]+ cc[2,2])/(cc[1,1]+ cc[2,2]+ cc[1,2]+ cc[2,1])

glm.probs.test.back <- predict(logestic.back, newdata = test, type= 'response')

glm.probs.test.back

glm.pred.test.back <- ifelse(glm.probs.test.back > 0.5, 1, 0)

glm.pred.test.back

table(glm.pred.test.back, test$Cat.Daramad\_kol\_Mounth)

dd <- table(glm.pred.test.back, test$Cat.Daramad\_kol\_Mounth)

(dd[1,1]+ dd[2,2])/(dd[1,1]+ dd[2,2]+ dd[1,2]+ dd[2,1])

#0.7830303

#Add log to see diffrences

train.log<- train

train.log$logHazine\_Ertebatat <- log10(train$Hazine\_Ertebatat)

train.log$logHazine\_Ghazayeamade <- log10(train$Hazine\_Ghazayeamade)

train.log$logHazine\_Khorakivadokhani <- log10(train$Hazine\_Khorakivadokhani)

train.log$logHazine\_lavazemkhanegi <- log10(train$Hazine\_lavazemkhanegi)

train.log$logHazine\_Maskan <- log10(train$Hazine\_Maskan)

train.log$logHazine\_Pushak <- log10(train$Hazine\_Pushak)

test.log<- test

test.log$logHazine\_Ertebatat <- log10(test$Hazine\_Ertebatat)

test.log$logHazine\_Ghazayeamade <- log10(test$Hazine\_Ghazayeamade)

test.log$logHazine\_Khorakivadokhani <- log10(test$Hazine\_Khorakivadokhani)

test.log$logHazine\_lavazemkhanegi <- log10(test$Hazine\_lavazemkhanegi)

test.log$logHazine\_Maskan <- log10(test$Hazine\_Maskan)

test.log$logHazine\_Pushak <- log10(test$Hazine\_Pushak)

logestic.back.log <- glm(formula = Cat.Daramad\_kol\_Mounth ~ C.O + Jens + Madrak +

tedad.a + n.t.m + oto + zabt + video + pc + yakhchal + m.lebas +

charkh.kh + m.zarf + tel + internet + cooler.g.s + logHazine\_Ertebatat +

logHazine\_Ghazayeamade + logHazine\_Khorakivadokhani + logHazine\_lavazemkhanegi +

logHazine\_Maskan + logHazine\_Pushak + Quantile\_Sen, family = "binomial",

data = train.log[, -c(45:50)])

glm.probs.train.back.log <- predict(logestic.back.log, newdata = train.log, type= 'response')

glm.probs.train.back.log

glm.pred.train.back.log <- ifelse(glm.probs.train.back.log > 0.5, 1, 0)

glm.pred.train.back.log

table(glm.pred.train.back.log, train$Cat.Daramad\_kol\_Mounth)

ee <- table(glm.pred.train.back.log, train$Cat.Daramad\_kol\_Mounth)

(ee[1,1]+ ee[2,2])/(ee[1,1]+ ee[2,2]+ ee[1,2]+ ee[2,1])

confusionMatrix(as.factor(glm.pred.train.back.log), train$Cat.Daramad\_kol\_Mounth)

glm.probs.test.back.log <- predict(logestic.back.log, newdata = test.log, type= 'response')

glm.probs.test.back.log

glm.pred.test.back.log <- ifelse(glm.probs.test.back.log > 0.5, 1, 0)

glm.pred.test.back.log

table(glm.pred.test.back.log, test$Cat.Daramad\_kol\_Mounth)

ff <- table(glm.pred.test.back.log, test$Cat.Daramad\_kol\_Mounth)

(ff[1,1]+ ff[2,2])/(ff[1,1]+ ff[2,2]+ ff[1,2]+ ff[2,1])

confusionMatrix(as.factor(glm.pred.test.back.log), test$Cat.Daramad\_kol\_Mounth)

#KNN

library(caret)

str(df)

df.KNN <- df1

#KNN works only numeric so we shuld do some pre processing

df.KNN$m.o.b <- NULL

df.KNN$ab.l <- NULL

df.KNN$bargh<- NULL

df.KNN$Tahsil.Mikonad <- NULL

df.KNN$Savad <- NULL

df.KNN$tv.s <- NULL

df.KNN$tv.r <- NULL

df.KNN$gaz <- NULL

df.KNN$cooler.a <- NULL

df.KNN$cooler.g <- NULL

df.KNN$gaz.l <- NULL

df.KNN$hamam <- NULL

df.KNN$ashpazkhane <- NULL

df.KNN$broodat.m <- NULL

df.KNN$hararat.m <- NULL

df.KNN$sookht.p <- NULL

df.KNN$sookht.g <- NULL

df.KNN$sookht.ab <- NULL

df.KNN$Sen <- NULL

df.KNN$Rahn <- NULL

df.KNN$Address <- NULL

df.KNN$MahMorajeh <- NULL

df.KNN$Fasl <- NULL

df.KNN$C.O <- df$C.O

for (i in 1:nrow(df.KNN)) {

if (df.KNN$C.O[i]== 'Alborz') {

df.KNN$C.O[i]= 1

}else df.KNN$C.O[i]= 0

}

df.KNN$C.O<- as.numeric(df.KNN$C.O)

str(df.KNN$C.O)

for (i in 1:nrow(df.KNN)) {

if (is.na(df.KNN$Madrak[i])) {

df.KNN$Madrak[i]= 0

}

}

for (i in 1:nrow(df.KNN)) {

if (is.na(df.KNN$oto[i])) {

df.KNN$oto[i]= 0

}

}

for (i in 1:nrow(df.KNN)) {

if (is.na(df.KNN$mo[i])) {

df.KNN$mo[i]= 0

}

}

for (i in 1:nrow(df.KNN)) {

if (is.na(df.KNN$do[i])) {

df.KNN$do[i]= 0

}

}

for (i in 1:nrow(df.KNN)) {

if (is.na(df.KNN$radio[i])) {

df.KNN$radio[i]= 0

}

}

for (i in 1:nrow(df.KNN)) {

if (is.na(df.KNN$zabt[i])) {

df.KNN$zabt[i]= 0

}

}

for (i in 1:nrow(df.KNN)) {

if (is.na(df.KNN$video[i])) {

df.KNN$video[i]= 0

}

}

for (i in 1:nrow(df.KNN)) {

if (is.na(df.KNN$pc[i])) {

df.KNN$pc[i]= 0

}

}

for (i in 1:nrow(df.KNN)) {

if (is.na(df.KNN$mobile[i])) {

df.KNN$mobile[i]= 0

}

}

for (i in 1:nrow(df.KNN)) {

if (is.na(df.KNN$freeizer[i])) {

df.KNN$freeizer[i]= 0

}

}

for (i in 1:nrow(df.KNN)) {

if (is.na(df.KNN$yakhchal[i])) {

df.KNN$yakhchal[i]= 0

}

}

for (i in 1:nrow(df.KNN)) {

if (is.na(df.KNN$yakhchal.f[i])) {

df.KNN$yakhchal.f[i]= 0

}

}

for (i in 1:nrow(df.KNN)) {

if (is.na(df.KNN$jaro.b[i])) {

df.KNN$jaro.b[i]= 0

}

}

for (i in 1:nrow(df.KNN)) {

if (is.na(df.KNN$m.lebas[i])) {

df.KNN$m.lebas[i]= 0

}

}

for (i in 1:nrow(df.KNN)) {

if (is.na(df.KNN$charkh.kh[i])) {

df.KNN$charkh.kh[i]= 0

}

}

for (i in 1:nrow(df.KNN)) {

if (is.na(df.KNN$panke[i])) {

df.KNN$panke[i]= 0

}

}

for (i in 1:nrow(df.KNN)) {

if (is.na(df.KNN$m.zarf[i])) {

df.KNN$m.zarf[i]= 0

}

}

for (i in 1:nrow(df.KNN)) {

if (is.na(df.KNN$microfer[i])) {

df.KNN$microfer[i]= 0

}

}

for (i in 1:nrow(df.KNN)) {

if (is.na(df.KNN$tel[i])) {

df.KNN$tel[i]= 0

}

}

for (i in 1:nrow(df.KNN)) {

if (is.na(df.KNN$internet[i])) {

df.KNN$internet[i]= 0

}

}

for (i in 1:nrow(df.KNN)) {

if (is.na(df.KNN$cooler.a.s[i])) {

df.KNN$cooler.a.s[i]= 0

}

}

for (i in 1:nrow(df.KNN)) {

if (is.na(df.KNN$package[i])) {

df.KNN$package[i]= 0

}

}

for (i in 1:nrow(df.KNN)) {

if (is.na(df.KNN$cooler.g.s[i])) {

df.KNN$cooler.g.s[i]= 0

}

}

for (i in 1:nrow(df.KNN)) {

if (is.na(df.KNN$fazelab[i])) {

df.KNN$fazelab[i]= 0

}

}

str(df.KNN$oto)

df.KNN$Cat.Daramad\_kol\_Mounth <- df$Cat.Daramad\_kol\_Mounth

df.KNN$Hazine\_koll<- df$Hazine\_Kol

set.seed(2564)

library(caTools)

split.knn <- sample.split(df.KNN$Cat.Daramad\_kol\_Mounth, SplitRatio = 0.7)

split.knn

train.knn <- subset(df.KNN, split== TRUE)

#train.copy<- df[split== TRUE, which(names(df)%in%names(train))]

test.knn <- subset(df.KNN, split== FALSE)

train.knn$Daramad\_Motefaraghe<-NULL

train.knn$Daramad\_Mozd\_Month<-NULL

train.knn$Daramad\_Mozd\_Year<-NULL

train.knn$Daramad\_Yarane<- NULL

train.knn$Daramad\_Azad<- NULL

test.knn$Daramad\_Motefaraghe<-NULL

test.knn$Daramad\_Mozd\_Month<-NULL

test.knn$Daramad\_Mozd\_Year<-NULL

test.knn$Daramad\_Yarane<- NULL

test.knn$Daramad\_Azad<- NULL

library(class)

for (i in 1:20) {

knn.pred.test <- knn(train = train.knn[,!(names(train.knn)%in%drop)],

test = test.knn[,!(names(test.knn)%in%drop)],

cl = train.knn[,names(train.knn)%in%drop],

k= i)

knn.pred.train <- knn(train = train.knn[,!(names(train.knn)%in%drop)],

test = train.knn[,!(names(test.knn)%in%drop)],

cl = train.knn[,names(train.knn)%in%drop],

k= i)

knn.pred

kh<- confusionMatrix(knn.pred.test, test.normalized.knn$Cat.Daramad\_kol\_Mounth)

kh1<- confusionMatrix(knn.pred.train, train.normalized.knn$Cat.Daramad\_kol\_Mounth)

print(i)

print(kh1[["overall"]][["Accuracy"]])

print(kh[["overall"]][["Accuracy"]])

}

knn.pred.train.15 <- knn(train = train.knn[,!(names(train.knn)%in%drop)],

test = train.knn[,!(names(test.knn)%in%drop)],

cl = train.knn[,names(train.knn)%in%drop],

k= 15)

confusionMatrix(knn.pred.train.15, train.normalized.knn$Cat.Daramad\_kol\_Mounth)

knn.pred.test.15 <- knn(train = train.knn[,!(names(train.knn)%in%drop)],

test = test.knn[,!(names(test.knn)%in%drop)],

cl = train.knn[,names(train.knn)%in%drop],

k= 15)

confusionMatrix(knn.pred.test.15, test.knn$Cat.Daramad\_kol\_Mounth)

#######Neural

library(neuralnet)

library(caret)

library(nnet)

library(caTools)

df.neural <- df

df.neural$ab.l<- NULL

df.neural$tedad.a<- df1$tedad.a

df.neural$t.o<- df1$t.o

str(df.neural)

df.neural$Hazine\_Behdashti<- log10(df.neural$Hazine\_Behdashti)

df.neural$Hazine\_Ertebatat<- log10(df.neural$Hazine\_Ertebatat)

df.neural$Hazine\_Ghazayeamade<- log10(df.neural$Hazine\_Ghazayeamade)

df.neural$Hazine\_Hamlonaghl<- log10(df.neural$Hazine\_Hamlonaghl)

df.neural$Hazine\_kalavakhadamat<- log10(df.neural$Hazine\_kalavakhadamat)

df.neural$Hazine\_lavazemkhanegi<- log10(df.neural$Hazine\_lavazemkhanegi)

df.neural$Rahn<- log10(df.neural$Rahn)

df.neural$Hazine\_Maskan<- log10(df.neural$Hazine\_Maskan)

df.neural$Hazine\_Noshidani<- log10(df.neural$Hazine\_Noshidani)

df.neural$Hazine\_Pushak<- log10(df.neural$Hazine\_Pushak)

df.neural$Hazine\_Tafrihat<- log10(df.neural$Hazine\_Tafrihat)

df.neural$Hazine\_Kol<- log10(df.neural$Hazine\_Kol)

df.neural <- df.neural[, c(2, 5, 9, 12, 13, 18, 22, 25, 26, 29, 33, 38, 42, 43, 50, 74, 56, 57, 60, 61, 63, 65,

73)]

for (i in 16:22) {

df.neural[,i]<- (df.neural[,i]- min(df.neural[,i]))/(max(df.neural[,i])- min(df.neural[,i]))

}

z1<-data.frame(class.ind(df.neural$C.O))

z2<-data.frame(class.ind(df.neural$Madrak))

z3<- data.frame(class.ind(df.neural$n.t.m))

df.neural <- cbind(df.neural, z1, z2, z3)

df.neural <- df.neural[, -c(1, 3, 5)]

names(df.neural)[c(26, 27, 28, 29, 30, 31, 32)]=c("Madrak.0","Madrak.1.9","Madrak.2","Madrak.3.4",

"Madrak.5","Madrak.6","Madrak.7.8")

df.neural$Jens <- as.numeric(as.character(df.neural$Jens))

df.neural$oto <- as.numeric(as.character(df.neural$oto))

df.neural$zabt <- as.numeric(as.character(df.neural$zabt))

df.neural$video <- as.numeric(as.character(df.neural$video))

df.neural$pc <- as.numeric(as.character(df.neural$pc))

df.neural$yakhchal <- as.numeric(as.character(df.neural$yakhchal))

df.neural$m.lebas <- as.numeric(as.character(df.neural$m.lebas))

df.neural$m.zarf <- as.numeric(as.character(df.neural$m.zarf))

df.neural$tel <- as.numeric(as.character(df.neural$tel))

df.neural$internet <- as.numeric(as.character(df.neural$internet))

df.neural$cooler.g.s <- as.numeric(as.character(df.neural$cooler.g.s))

df.neural$tedad.a <- as.numeric(as.character(df.neural$tedad.a))

set.seed(2564)

split.neural <- sample.split(df.neural, SplitRatio = 0.7)

train.neural<- subset(df.neural, split.neural==TRUE)

test.neural<- subset(df.neural, split.neural== FALSE)

str(train.neural)

library(doSNOW)

c1<- makeCluster(7, type = 'SOCK')

registerDoSNOW(c1)

nn<- neuralnet(as.factor(Cat.Daramad\_kol\_Mounth)~Jens+ tedad.a+ oto+ zabt+ video+ pc+ yakhchal+

m.lebas+ m.zarf+ tel+ internet+ cooler.g.s+

Quantile\_Sen+ Hazine\_Ertebatat+ Hazine\_Ghazayeamade+ Hazine\_Khorakivadokhani+

Hazine\_lavazemkhanegi+ Hazine\_Maskan+ Hazine\_Pushak+ Alborz+ Hamedan+ Markazi+

Qazvin+

Qom+ Madrak.0+ Madrak.1.9+ Madrak.2+ Madrak.3.4+ Madrak.5+ Madrak.6+ Madrak.7.8+

Ejari+

Khedmat.Melki.Ayan+ Melki.Arse.Ayan+ Rahn+ Sayer.Raygan, data= train.neural, linear.output =

F,

hidden = 1, threshold = 0.1)

stopCluster(c1)

nn$weights

prediction(nn)

plot(nn, rep="best")

train.p=compute(nn,train.neural)

train.c=apply(train.p$net.result,1,which.max)-1

confusionMatrix(as.factor(train.c),as.factor(train.neural$Cat.Daramad\_kol\_Mounth))

test.p=compute(nn,test.neural)

test.c=apply(test.p$net.result,1,which.max)-1

confusionMatrix(as.factor(test.c),as.factor(test.neural$Cat.Daramad\_kol\_Mounth))

#TREE

library(caTools)

library(caret)

df.tree<- df

df.tree$m.o.b <- NULL

df.tree$ab.l <- NULL

df.tree$bargh<- NULL

df.tree$Tahsil.Mikonad <- NULL

df.tree$Savad <- NULL

df.tree$tv.s <- NULL

df.tree$tv.r <- NULL

df.tree$gaz <- NULL

df.tree$cooler.a <- NULL

df.tree$cooler.g <- NULL

df.tree$gaz.l <- NULL

df.tree$hamam <- NULL

df.tree$ashpazkhane <- NULL

df.tree$broodat.m <- NULL

df.tree$hararat.m <- NULL

df.tree$sookht.p <- NULL

df.tree$sookht.g <- NULL

df.tree$sookht.ab <- NULL

df.tree$Sen <- NULL

df.tree$Rahn <- NULL

df.tree$Address <- NULL

df.tree$MahMorajeh <- NULL

df.tree$Fasl <- NULL

df.tree$Daramad\_Motefaraghe<-NULL

df.tree$Daramad\_Mozd\_Month<-NULL

df.tree$Daramad\_Mozd\_Year<-NULL

df.tree$Daramad\_Yarane<- NULL

df.tree$Daramad\_Azad<- NULL

df.tree$Daramad\_Motefaraghe<-NULL

df.tree$Daramad\_Mozd\_Month<-NULL

df.tree$Daramad\_Mozd\_Year<-NULL

df.tree$Daramad\_Yarane<- NULL

df.tree$Daramad\_Azad<- NULL

df.tree$Daramad\_kol\_Mounth<- NULL

library(rpart)

library(rpart.plot)

set.seed(2564) # partition

split.tree <- sample.split(df.tree$Cat.Daramad\_kol\_Mounth, SplitRatio = 0.7)

split.tree

train.tree <- subset(df.tree, split== TRUE)

#train.copy<- df[split== TRUE, which(names(df)%in%names(train))]

test.tree <- subset(df.tree, split== FALSE)

#plotting a tree with just one variable for clustering

class.tree <- rpart(Cat.Daramad\_kol\_Mounth ~ ., data = train.tree, control = rpart.control(maxdepth = 2),

method = "class")

prp(class.tree, type = 1, extra = 1, split.font = 1, varlen = -10)

#compue accuracy for one.variable tree

one.variable.pred.train <- predict(class.tree,train.tree,type = "class")

# generate confusion matrix for training data

confusionMatrix(one.variable.pred.train, train.tree$Cat.Daramad\_kol\_Mounth)

one.variable.pred.test <- predict(class.tree,test.tree,type = "class")

# generate confusion matrix for test data

confusionMatrix(one.variable.pred.test, test.tree$Cat.Daramad\_kol\_Mounth)

#derakht pish farz

# plot tree

default.ct <- rpart(Cat.Daramad\_kol\_Mounth ~ ., data = train.tree, method = "class")

prp(default.ct, type = 1, extra = 1, under = TRUE, split.font = 1, varlen = -10)

#derakht amigh

deeper.ct <- rpart(Cat.Daramad\_kol\_Mounth ~ ., data = train.tree, method = "class", cp = 0, minsplit = 1)

# count number of leaves

length(deeper.ct$frame$var[deeper.ct$frame$var == "<leaf>"])

# plot tree

prp(deeper.ct, type = 1, extra = 1, under = TRUE, split.font = 1, varlen = -10,

box.col=ifelse(deeper.ct$frame$var == "<leaf>", 'gray', 'white'))

#compue accuracy for default tree

default.ct.point.pred.train <- predict(default.ct,train.tree,type = "class")

# generate confusion matrix for training data

confusionMatrix(default.ct.point.pred.train, train.tree$Cat.Daramad\_kol\_Mounth)

# repeat the code for the validation set

default.ct.point.pred.test <- predict(default.ct,test.tree,type = "class")

# generate confusion matrix for test data

confusionMatrix(default.ct.point.pred.test, test.tree$Cat.Daramad\_kol\_Mounth)

#compue accuracy for deep tree

deeper.ct.point.pred.train <- predict(deeper.ct,train.tree,type = "class")

# generate confusion matrix for training data

confusionMatrix(deeper.ct.point.pred.train, train.tree$Cat.Daramad\_kol\_Mounth)

# repeat the code for the validation set

deeper.ct.point.pred.test <- predict(deeper.ct,test.tree,type = "class")

# generate confusion matrix for test data

confusionMatrix(deeper.ct.point.pred.test, test.tree$Cat.Daramad\_kol\_Mounth)

##Etebar sanji moteqate va prune kardan

# argument xval refers to the number of folds to use in rpart's built-in

# cross-validation procedure

# argument cp sets the smallest value for the complexity parameter.

cv.ct <- rpart(Cat.Daramad\_kol\_Mounth ~ ., data = train.tree, cp = 0.00001, minsplit = 7, xval = 3)

# use printcp() to print the table.

printcp(cv.ct)

#prune by lower cp

pruned.ct <- prune(cv.ct, cp = cv.ct$cptable[which.min(cv.ct$cptable[,"xerror"]),"CP"])

length(pruned.ct$frame$var[pruned.ct$frame$var == "<leaf>"])

prp(pruned.ct, type = 1, extra = 1, split.font = 1, varlen = -10, digits=-3)

pruned.ct.point.pred.train <- predict(pruned.ct,train.tree,type = "class")

# generate confusion matrix for training data

confusionMatrix(pruned.ct.point.pred.train, train.tree$Cat.Daramad\_kol\_Mounth)

# repeat the code for the validation set

pruned.ct.point.pred.test <- predict(pruned.ct,test.tree,type = "class")

# generate confusion matrix for test data

confusionMatrix(pruned.ct.point.pred.test, test.tree$Cat.Daramad\_kol\_Mounth)

##Implimention

df.piadesazi <- read.csv('implimantion - Copy.csv', header = TRUE)

df.piadesazi<- df.piadesazi[1:14, 1:9]

t.o.scale.piadesazi <- scale(df.piadesazi$t.o)

s.z.scale.piadesazi <- scale(df.piadesazi$s.z)

ss<- pca$rotation[1,1]\* t.o.scale.piadesazi+ pca$rotation[2,1]\* s.z.scale.piadesazi

df.piadesazi$c\_t.oands.z<- ss[,1]

nrow(df.piadesazi)

for (i in 1:nrow(df.piadesazi)) {

if (df.piadesazi[i,'Hazine\_Khorakivadokhani']<= 8399945& df.piadesazi[i,'Hazine\_Ertebatat']<= 545000)

{

df.piadesazi[i,'pred.Cat.Daramad\_kol\_Mounth'] = 0

}else if (df.piadesazi[i,'Hazine\_Khorakivadokhani']<= 5545500& df.piadesazi[i,'Hazine\_Ertebatat']>

545000& df.piadesazi[i,'m.zarf']== 0) {

df.piadesazi[i,'pred.Cat.Daramad\_kol\_Mounth'] = 0

}else if (df.piadesazi[i,'Hazine\_Khorakivadokhani']<= 5545500& df.piadesazi[i,'Hazine\_Ertebatat']>

545000& df.piadesazi[i,'m.zarf']== 1) {

df.piadesazi[i,'pred.Cat.Daramad\_kol\_Mounth'] = 1

}else if (df.piadesazi[i,'Hazine\_Khorakivadokhani']<= 8399945& df.piadesazi[i,'Hazine\_Ertebatat']>

545000& df.piadesazi[i,'Hazine\_Khorakivadokhani']> 5545500& df.piadesazi[i,'c\_t.oands.z']>=.587) {

df.piadesazi[i,'pred.Cat.Daramad\_kol\_Mounth'] = 0

}else if (df.piadesazi[i,'Hazine\_Khorakivadokhani']<= 8399945& df.piadesazi[i,'Hazine\_Ertebatat']>

545000& df.piadesazi[i,'Hazine\_Khorakivadokhani']> 5545500& df.piadesazi[i,'c\_t.oands.z']<.587) {

df.piadesazi[i,'pred.Cat.Daramad\_kol\_Mounth'] = 1

}else if (df.piadesazi[i,'Hazine\_Khorakivadokhani']> 8399945& df.piadesazi[i,

'Hazine\_Maskan']>9805000) {

df.piadesazi[i,'pred.Cat.Daramad\_kol\_Mounth'] = 1

}else if (df.piadesazi[i,'Hazine\_Khorakivadokhani']> 8399945& df.piadesazi[i,

'Hazine\_Maskan']<=9805000& df.piadesazi[i, 'Hazine\_Ertebatat']>= 402500& df.piadesazi[i,

'Quantile\_Sen']> 44) {

df.piadesazi[i,'pred.Cat.Daramad\_kol\_Mounth'] = 1

}else if (df.piadesazi[i,'Hazine\_Khorakivadokhani']> 8399945& df.piadesazi[i,

'Hazine\_Maskan']<=9805000& df.piadesazi[i, 'Hazine\_Ertebatat']>= 402500& df.piadesazi[i,

'Quantile\_Sen']<= 44) {

df.piadesazi[i,'pred.Cat.Daramad\_kol\_Mounth'] = 0

}else if (df.piadesazi[i,'Hazine\_Khorakivadokhani']> 8399945& df.piadesazi[i,

'Hazine\_Maskan']<=9805000& df.piadesazi[i, 'Hazine\_Ertebatat']< 402500& df.piadesazi[i, 'Madrak']==

6) {

df.piadesazi[i,'pred.Cat.Daramad\_kol\_Mounth'] = 1

}else if (df.piadesazi[i,'Hazine\_Khorakivadokhani']> 8399945& df.piadesazi[i,

'Hazine\_Maskan']<=9805000& df.piadesazi[i, 'Hazine\_Ertebatat']< 402500& df.piadesazi[i, 'Madrak']!=

6) {

df.piadesazi[i,'pred.Cat.Daramad\_kol\_Mounth'] = 0

}

}

confusionMatrix(as.factor(df.piadesazi$Cat.Daramad\_kol\_Mounth),as.factor(df.piadesazi$pred.Cat.Dara

mad\_kol\_Mounth))