rm(list=ls(all=T))

setwd('D:/INSOFE/Insofe\_ project/Target\_Marketing\_and\_cross\_selling\_\_Data')

library(XLConnect)

library(dplyr)

library(compare)

df = readWorksheet(loadWorkbook("Target Marketing and cross selling - Data.xls"), sheet = "Sheet1", header = TRUE)

str(df)

colnames(df) <- tolower(gsub("\\.","\_",colnames(df)))

View(df)

cutomertype <- df %>% group\_by(customer\_type) %>% arrange(desc(customer\_type))

customerID <- data.frame(cutomertype %>% group\_by(customer\_id) %>% arrange(desc(customer\_id)))

View(cutomertype)

View(customerID)

class(customerID)

str(customerID)

customerID %>% group\_by(customer\_id,customer\_type,call\_date) %>% summarise(n=n())

days = as.numeric(as.Date(unique(customerID)$complete\_date[2:5284],format = "%y-%m-%d") - as.Date(customerID$call\_date[1:5283],format = "%y-%m-%d"))

i <- 1

days[1:(i+4)==(i+4)] = 0

days = data.frame(days)

nrow(days)

days = rbind(c(0),days)

class(days)

customerID = cbind(customerID,days)

customerID$days = as.numeric(lapply(customerID$days, abs))

customerID %>% group\_by(customer\_id) %>% summarise(n= n(),mean(days))

customerID %>% group\_by(customer\_type) %>% summarise(n= n(),mean(days))

raw\_data = customerID %>% filter(days == 0)

train\_data = anti\_join(customerID,raw\_data)

View(train\_data)

View(test\_data)

train\_data <- train\_data %>% group\_by(customer\_type) %>% arrange(desc(customer\_type))

train\_data <- data.frame(train\_data %>% group\_by(customer\_id) %>% arrange(desc(customer\_id)))

test\_data = data.frame()

j= 4

while (j< 4145){

raw = train\_data[j,]

test\_data = rbind(test\_data,raw)

j = j+4

}

View(test\_data)

test\_data = test\_data[,-32]

rm(raw\_data)

rm(raw)

train\_data %>% group\_by(customer\_id) %>% summarise(n= n(),mean(days))

train\_data %>% group\_by(customer\_type) %>% summarise(n= n(),mean(days))

churn = c(rep(1,4145))

train\_data = cbind(train\_data,churn)

cust1 = train\_data %>% filter(customer\_type=="CustType01")

cust2 = train\_data %>% filter(customer\_type=="CustType02")

cust3 = train\_data %>% filter(customer\_type=="CustType03")

Days1 = cust1$days

cust1$churn = lapply(Days1,function(x){

if(x>c(180)) x = 0

else x= 1

})

Days2 = cust2$days

cust2$churn = lapply(Days2,function(x){

if(x>c(180)) x = 0

else x= 1

})

Days3 = cust3$days

cust3$churn = lapply(Days3,function(x){

if(x>c(180)) x = 0

else x= 1

})

train\_data = rbind(cust1,cust2,cust3)

train\_data$churn = as.factor(as.character(train\_data$churn))

str(train\_data)

library(glm2)

model = glm(churn ~ .,data = train\_data,family = binomial)