Deriverable2

Code:

R coding for data exploration

data2 <- read.csv("C:/wellsfargo2/data2.csv", header=TRUE)

install.packages("sqldf")

library(sqldf)

sqldf("SELECT count(\*) from data2")

count(\*)

1 120000

New Rows added

products\_computed <- (data2$C\_product\_presence + data2$D\_product\_presence + data2$E\_product\_presence + data2$F\_product\_presence + data2$G\_product\_presence)

outreaches\_8\_computed <- (data3$Type\_A\_customer\_outreaches + data3$Type\_B\_customer\_outreaches + data3$Type\_C\_customer\_outreaches + data3$Type\_D\_customer\_outreaches + data3$Type\_E\_customer\_outreaches + data3$Type\_F\_customer\_outreaches + data3$Type\_G\_customer\_outreaches + data3$Typ\_H\_customer\_outreaches )

channel\_outreaches\_computed <- (data3$Channel\_1\_outreach + data3$service\_contact\_using\_channel\_2 + data3$X3\_market\_outreach\_using\_channe + data3$X4\_market\_outreach\_using\_channel )

data3 <- cbind(data2, products\_computed)

data3 <- cbind(data3, channel\_outreaches\_computed)

data3 <- cbind(data3, outreaches\_8\_computed)

**Untargeted customers with no products, no outreach and no outreach through channels**

**Unexplored customers:**

sqldf("SELECT count(\*) FROM data3 WHERE products\_computed ='0'AND channel\_outreaches\_computed = '0' and outreaches\_8\_computed ='0' ")

count(\*)

1 16730

**Untargeted customers with products**

sqldf("SELECT count(\*) FROM data3 WHERE products\_computed >'0'AND channel\_outreaches\_computed = '0' and outreaches\_8\_computed ='0' ")

count(\*)

1 2676

**Total 19406 untargeted**

Data4<-data3

**customer with increased b type account**

**subset(data4, data4$cust\_num==436)[which(diff(subset(data4[,c("B\_type\_account\_balance")], data4$cust\_num==436)) >0),]**

**Each customers demographic data**

**for (i in 1:10000) print(unique(subset(data4[,c("cust\_num","cust\_demographics\_ai","cust\_demographics\_aii")], cust\_num==i)))**

> b\_bal\_inc<-(final2[,], final2$b\_bal.B\_bal\_change >0 )

a\_bal\_inc<-(final2[,], final2$a\_bal.a\_bal\_change >0 )

final2<- cbind(data4,b\_bal$B\_bal\_change,a\_bal$A\_bal\_change)

**Similarly, we have created other datasets**

nrow((subset(final2[,],final2$Prod\_inc.Prod\_changed ==0 & final2$a\_bal.A\_bal\_change >0 | final2$b\_bal.B\_bal\_change >0)))

[1] 17176

> nrow((subset(final2[,],final2$Prod\_inc.Prod\_changed ==0 & final2$a\_bal.A\_bal\_change >0 & final2$b\_bal.B\_bal\_change >0)))

[1] 700

> nrow((subset(final2[,], final2$a\_bal.A\_bal\_change >0 )))

[1] 13374

> nrow((subset(final2[,], final2$b\_bal.B\_bal\_change >0 )))

[1] 4753

For Visualization

> count(subset(prod\_purchased[,c(27)], prod\_purchased$Prod\_inc.Prod\_changed>0))

x freq

1 0 340

2 1 231

> count(subset(prod\_purchased[,c(28)], prod\_purchased$Prod\_inc.Prod\_changed>0))

x freq

1 0 143

2 1 428

> count(subset(prod\_purchased[,c(29)], prod\_purchased$Prod\_inc.Prod\_changed>0))

x freq

1 0 571

> count(subset(prod\_purchased[,c(30)], prod\_purchased$Prod\_inc.Prod\_changed>0))

x freq

1 0 461

2 1 110

**Customers in demography 0,1**

**data4[ which( data4$cust\_demographics\_ai ==0 & data4$cust\_demographics\_aii ==1 ) , c(1,3,4) ]**

for (i in 0:5) for(j in 1:5) print(count(data4[ which( data4$cust\_demographics\_ai ==i & data4$cust\_demographics\_aii ==j ) , c(3,4) ] ))

|  |
| --- |
|  |

**Note: for creating models , SAS Enterprise miner is used and requires no code**

