Homework 1 | Group 5 | SCMT650

Group 5

2/7/19

QUESTION 1: Print top 10 customers by total sales.

```r
rm(list=ls())
load("transaction.rdata")

# Aggreate the sales amount by custcode without deducting refunds
temp1=aggregate(trans$saleamount,by=list(trans$custcode),sum)

# Sort by sales amount from highest to lowest
temp2=temp1[order(-temp1$x),]
names(temp2)[1]="Customer_Code"
names(temp2)[2]="Total_$_Amt"

# Take the top ten values
Q1=head(temp2,10)


#Question 1 Answer
Q1

##      Customer_Code Total_$_Amt
## 518         930092   130835.12
## 212         919592    76182.72
## 582         932514    43319.59
## 68          914338    39367.12
## 936         944992    36239.69
## 1427        961842    35634.29
## 793         939474    33569.69
## 1138        952006    31005.90
## 915         944079    29302.69
## 1312        958127    25913.51
```

QUESTION 2: Customers with hightest return rates.

```r
rm(list=ls())
load("transaction.rdata")

# Extract return purchases
returns=subset(trans,trans$numpurchases<0)

# Aggregrate returns by custcode, summing number of returns and total
returend sales value
```

```r
returnsAgg=aggregate(returns[,c(7:8)],list(returns$custcode),sum,na.rm=TRUE)
colnames(returnsAgg)= c("custcode", "returned_numpurchases",
"returned_sale_amount")

# Extract purchases
purchases=subset(trans,trans$numpurchases>0)

# Aggregate sales by custcode, summing number of purchases and total sales
value
purchasesAgg=aggregate(purchases[,c(7:8)],list(purchases$custcode),sum,na.rm=
TRUE)
colnames(purchasesAgg)= c("custcode", "nonreturned_numpurchases",
"nonreturned_sale_amount")

# Merge refund and purchases data frames
mergedTrans=merge(returnsAgg,purchasesAgg)

# Extract customers with 5 or more returns
fiveReturns=subset(mergedTrans,mergedTrans$returned_numpurchases<= -5)

# Calcualte return rate
returnRate = transform(fiveReturns, returnRate =
fiveReturns$returned_sale_amount/fiveReturns$nonreturned_sale_amount)
returnRate$returnRate = (-1*returnRate$returnRate)

# Order return rate from highest to lowest
highestReturners = returnRate[order(-returnRate$returnRate),]


# Extract top 10 customers and their return rates
temp=head(highestReturners,10)
Q2=temp[c(1,6)]

#Question 2 Answer
Q2

##      custcode returnRate
## 753   957743  0.8964878
## 43    914514  0.8277993
## 354   933179  0.8215661
## 31    913938  0.8058871
## 212   925214  0.8045915
## 281   928792  0.7997604
## 37    914242  0.7771765
## 546   945243  0.7296235
## 335   932343  0.7055097
## 155   921714  0.6932584
```

QUESTION 3: Customer length of relationship compared to high vs low returners

```r
rm(list=ls())
load("customer.rdata")
load("transaction.rdata")

# Extract subset of returned transactions
returns=subset(trans,trans$numpurchases<0)

# Aggregate returns by custcode, summing number of returns and total returend
sales value
returnsAgg=aggregate(returns[,c(7:8)],list(returns$custcode),sum,na.rm=TRUE)
colnames(returnsAgg)= c("custcode", "returned_numpurchases",
"returned_sale_amount")

# Calcualte average number of returns
meanRenturns=mean(returnsAgg$returned_numpurchases)

# Catgorize custcode as high or low returner
returnsAgg$hilo = ifelse(returnsAgg$returned_numpurchases<meanRenturns,
"high", "low")

# Find length of relationahip for each customer
trans$length=as.numeric(trans$saledate)
rtemp1=aggregate(trans$length, by =list(trans$custcode) ,max)
rtemp2=aggregate(trans$length, by =list(trans$custcode) ,min)
customer$Relationship_Length=rtemp1$x-rtemp2$x

# Merge returnsAgg and customer data.frames
customer$custcode <- as.numeric(paste(customer$custcode))
Q3=merge(returnsAgg, customer)
Q3=Q3[,c(1:4,10)]

# Subset of full data, for example
head(Q3,10)
```

```
##     custcode returned_numpurchases returned_sale_amount hilo
## 1    912074                   -11              -537.75  low
## 2    912079                    -2               -43.80  low
## 3    912138                    -5              -141.00  low
## 4    912214                    -2              -650.00  low
## 5    912238                    -5              -258.97  low
## 6    912279                    -2               -49.95  low
## 7    912295                    -5              -294.60  low
## 8    912442                    -2               -64.90  low
## 9    912479                    -7              -481.90  low
## 10   912514                    -7              -854.90  low
##     Relationship_Length
## 1                   1739
## 2                   1761
```

```
## 3                      1789
## 4                       139
## 5                      2164
## 6                      2250
## 7                       897
## 8                      1925
## 9                      2617
## 10                     2182
```

```r
# Find average relatinoship length for high and low categories
temp=aggregate(Q3$Relationship_Length, by =list(Q3$hilo) ,mean)
colnames(temp)=c("return_cat", "avg_length")
temp
```

```
##    return_cat avg_length
## 1        high   2321.196
## 2         low   1732.092
```

Answer to question 3a: Yes, customers with a high number of returns have a longer customer relationship length than customers with a low number of returns.

QUESTION 4: Total Sales by income and age.

```r
rm(list=ls())
load("customer.rdata")
load("transaction.rdata")

# Extract purchases
purchases=subset(trans,trans$numpurchases>0)

# Aggregate sales by custcode, summing number of purchases and total sales
value
purchasesAgg=aggregate(purchases[,c(7:8)],list(purchases$custcode),sum,na.rm=
TRUE)
colnames(purchasesAgg)= c("custcode", "nonreturned_numpurchases",
"nonreturned_sale_amount")

# Create subset with specific columns, then merge
sub1=purchasesAgg[c(1,3)]
sub2=customer[c(1:3)]

Q4merged=merge(sub1,sub2)

# Aggregate non_returned_Sale_amount, by both inccome and agecode
Q4agg=aggregate(Q4merged$nonreturned_sale_amount , by=list(Q4merged$inccode,
Q4merged$agecode),sum,na.rm=TRUE)
colnames(Q4agg)= c("income_code", "age_code", "nonreturned_totalsale_amount")
```

```r
# Create 5x6 matrix populated with 0s
Q4matrix=matrix(0,nrow=5,ncol=6,byrow=F)

# Input nonreturned_totalsale_amount into correct matrix cell, with income
code as row and age code as column
for (i in 1:nrow(Q4agg)) {
  income_code=Q4agg[i,1]
  age_code=Q4agg[i,2]
  Q4matrix[income_code,age_code] <- Q4agg$nonreturned_totalsale_amount[i]
}

colnames(Q4matrix) <- c(1:6)
rownames(Q4matrix) <- c(1:5)

#Answer to question 4
Q4matrix
```

```
##      1         2         3        4        5         6
## 1    0 265331.77 264135.32 316635.0 271551.6 414826.68
## 2  157  55342.40 112126.65 279605.7 311952.9 148480.23
## 3    0  10244.41  94509.86 180575.2 206615.8 113203.61
## 4  506   8219.10 101308.49 107086.1 124171.7  94428.48
## 5    0  33974.41 105762.76 258766.5 236885.6 129779.57
```

QUESTION 5: return rate by product category

```r
rm(list=ls())
load("customer.rdata")
load("transaction.rdata")
category=read.csv("category.csv")

# Correct error in catname data
category$catname[category$catname == "Apparel Chilldren"] <- "Apparel
Children"

# Merge
cattrans=merge(trans, category)

# Subset returns and nonreturned purchases
cattransRet=subset(cattrans,cattrans$numpurchases<0 )
cattransPur=subset(cattrans,cattrans$numpurchases>0 )

# Aggregate saleamount by catname
RetAgg=aggregate(cattransRet$saleamount, list(cattransRet$catname), sum)
SaleAgg=aggregate(cattransPur$saleamount, list(cattransPur$catname), sum)
```

```
colnames(RetAgg)= c("Catname", "total_return_amount")
colnames(SaleAgg)= c("Catname", "total_sale_amount")

# Merge returns and sales and calculate the return rate
Q5merge=merge(RetAgg, SaleAgg)
Q5merge$total_return_amount = (-1*Q5merge$total_return_amount)
returnRate = transform(Q5merge, returnRate =
Q5merge$total_return_amount/Q5merge$total_sale_amount)
returnRate = returnRate[order(returnRate$returnRate),]

# Sort return rate from highest to lowest
Q5 = returnRate[order(-returnRate$returnRate),]

#Answer to question 5
Q5
```

```
##                 Catname total_return_amount total_sale_amount returnRate
## 1           Accessories            47650.70         215448.20 0.22117010
## 4               Apparel            24029.24         114484.02 0.20989165
## 10              Jewelry            27958.65         139498.29 0.20042289
## 7         Apparel Women           424298.63        2351566.35 0.18043234
## 14          Shoes Women           130756.12         725135.51 0.18031957
## 5      Apparel Children            21482.47         144893.76 0.14826360
## 13            Shoes Men            16953.74         127589.95 0.13287677
## 12       Shoes Children             5195.84          43112.90 0.12051706
## 6           Apparel Men            31578.42         311256.83 0.10145454
## 2       Accessories Men             7947.57          86982.92 0.09136932
## 3     Accessories Women             1415.96          20008.68 0.07076729
## 11              Service            21648.24         357838.36 0.06049726
## 8             Cosmetics             6774.37         117107.25 0.05784757
## 9            Home Goods             1105.07          22109.51 0.04998166
```