**Assignment – Market Basket Analysis**

Initial Data file Need to be analyzed



**r. Script file**

#read csv into R dataframe

library(plyr)

retail <- read.csv("file:///D:/DataScientist/kmohans/DataScience/Assignment/data.csv",header = T)

#Consolidate the item describtion based upon the Number and date

transactionData <- ddply(retail,c("InvoiceNo","InvoiceDate"),

function(df1)paste(df1$Description,

collapse = ","))

#Null both the Invoice Number and Date columns

transactionData$InvoiceNo <- NULL

transactionData$InvoiceDate <- NULL

# Change column Title alone

colnames(transactionData) <- c("items")

#Print the data in console

head(transactionData,10)

#Write the DataItems in the temp file

write.csv(transactionData,"file:///D:/DataScientist/kmohans/DataScience/Assignment/dataItems.csv", quote = FALSE, row.names = TRUE)

#Read it into Variable

tr <- read.transactions('file:///D:/DataScientist/kmohans/DataScience/Assignment/dataItems.csv', format = 'basket', sep=',')

#Plot the top N items

itemFrequencyPlot(tr,topN=20,type="absolute")

#Create the rules based upon the Algorithm

rules <- apriori(tr,

parameter = list(supp = 0.001, conf = 0.8))

# Show the top 5 rules, but only 2 digits

summary(rules)

options(digits=2)

inspect(rules[1:5])

# Sorting the rules

rules<-sort(rules, by="confidence", decreasing=TRUE)

inspect(rules[1:5])

### Write the rules into a data.frame for easier viewing

rule.df <- as(rules,"data.frame")

View(rule.df)

### Set rule with maximum length to subset limited items

rules <- apriori(tr,

parameter = list(supp = 0.001,

conf = 0.8,maxlen=3))

# Cleaning the rules

subset.matrix <- is.subset(rules, rules)

subset.matrix[lower.tri(subset.matrix, diag=T)] <- NA

redundant <- colSums(subset.matrix, na.rm=T) >= 1

redundant

rules.pruned <- rules[!redundant]

rules<-rules.pruned

## Mining rules based on a items with specific product

rules<-apriori(data=tr, parameter=list(supp=0.001,conf = 0.08),

appearance = list(default="rhs",lhs="BILLBOARD FONTS DESIGN"),

control = list(verbose=F))

## Mining rules based on specific product with other items

rules<-sort(rules, decreasing=TRUE,by="confidence")

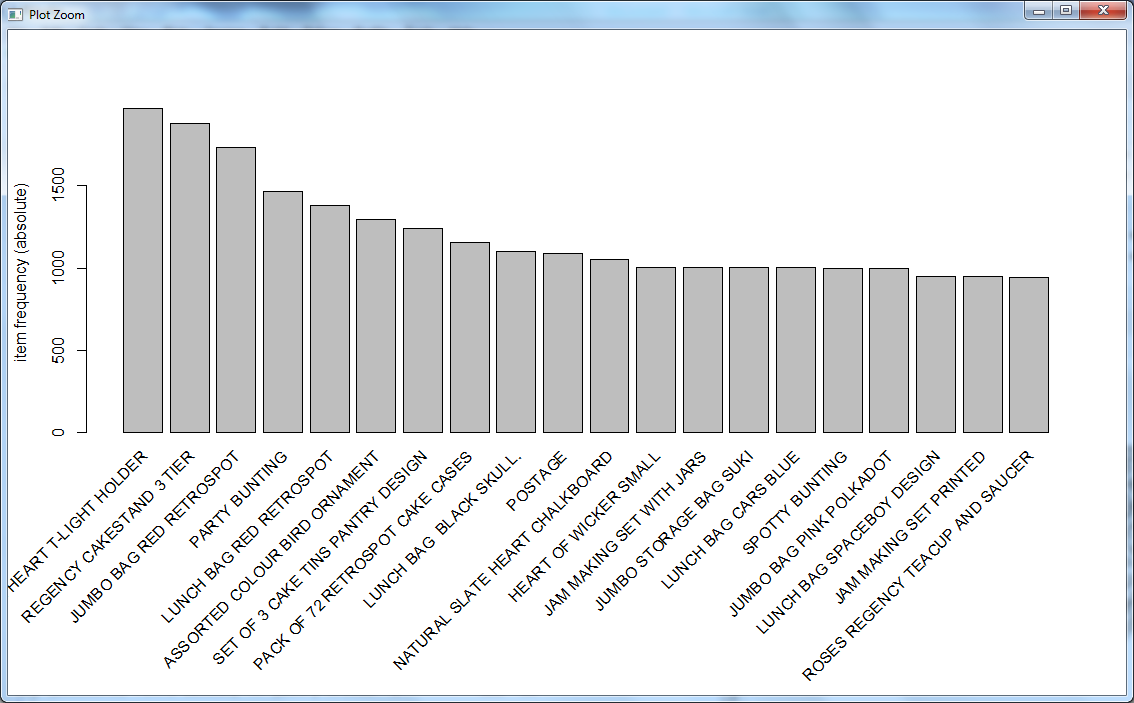
inspect(rules[1])

**Results**

Transaction file created for processing:



Top 20 Transaction Items



Possibility of selecting the following items after selecting “**BILLBOARD FONTS DESIGN**”

> rules<-apriori(data=tr, parameter=list(supp=0.001,conf = 0.08),

+ appearance = list(default="rhs",lhs="BILLBOARD FONTS DESIGN"),

+ control = list(verbose=F))

lhs rhs support confidence lift count

[1] {BILLBOARD FONTS DESIGN} => {WRAP} 0.0012 1 811 30