MIS 510 Portfolio Project Option 2

Deepak Gupta

5/30/2020

## R Markdown

# load library arules to convert the binary incidence matrix into a transactions database  
  
setwd("D:/CSUGLOBAL/MIS510 Data Mining and Visualization/Module 8")  
  
#load the packages for association rules and R sample datasets  
  
library(arules) #arules: Visualizing Association Rules

## Loading required package: Matrix

##   
## Attaching package: 'arules'

## The following objects are masked from 'package:base':  
##   
## abbreviate, write

library(arulesViz) #arulesViz: Visualizing Association Rules

## Loading required package: grid

## Registered S3 method overwritten by 'seriation':  
## method from   
## reorder.hclust gclus

crossSell.df <- read.csv("CatalogCrossSell.csv", nrows = 4998)  
  
crossSell.df <- crossSell.df[,2:10] # Removing customer id from dataset, otherwise will be taken as an "item"  
  
head(crossSell.df)

## Clothing.Division Housewares.Division Health.Products.Division  
## 1 0 1 1  
## 2 0 1 1  
## 3 0 1 1  
## 4 0 0 1  
## 5 0 0 1  
## 6 0 1 1  
## Automotive.Division Personal.Electronics.Division Computers.Division  
## 1 1 1 0  
## 2 1 1 0  
## 3 1 1 0  
## 4 1 1 0  
## 5 0 1 0  
## 6 1 1 0  
## Garden.Division Novelty.Gift.Division Jewelry.Division  
## 1 0 1 0  
## 2 1 1 1  
## 3 1 1 1  
## 4 1 1 0  
## 5 1 1 0  
## 6 1 1 1

tail(crossSell.df)

## Clothing.Division Housewares.Division Health.Products.Division  
## 4993 0 0 1  
## 4994 0 0 1  
## 4995 0 1 1  
## 4996 0 0 1  
## 4997 0 0 1  
## 4998 0 0 1  
## Automotive.Division Personal.Electronics.Division Computers.Division  
## 4993 0 0 0  
## 4994 0 0 0  
## 4995 0 0 0  
## 4996 0 1 0  
## 4997 0 0 0  
## 4998 0 0 0  
## Garden.Division Novelty.Gift.Division Jewelry.Division  
## 4993 0 0 0  
## 4994 0 0 0  
## 4995 1 0 1  
## 4996 0 0 1  
## 4997 0 0 0  
## 4998 0 0 0

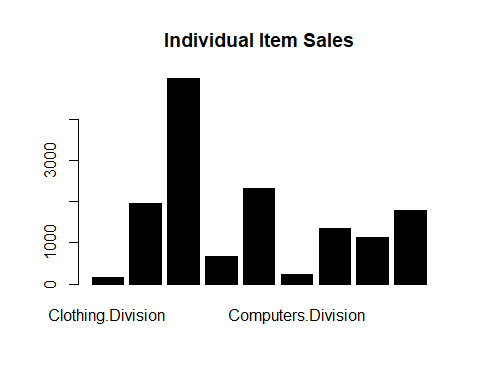
dim(crossSell.df)

## [1] 4998 9

#View individual item names  
t(t(names(crossSell.df)))

## [,1]   
## [1,] "Clothing.Division"   
## [2,] "Housewares.Division"   
## [3,] "Health.Products.Division"   
## [4,] "Automotive.Division"   
## [5,] "Personal.Electronics.Division"  
## [6,] "Computers.Division"   
## [7,] "Garden.Division"   
## [8,] "Novelty.Gift.Division"   
## [9,] "Jewelry.Division"

#Convert to matrix  
crossSelldata.df <- as.matrix(crossSell.df)  
  
#Analyzing Dataset with bar plot to view distribution of sales  
barplot(crossSelldata.df, main = "Individual Item Sales")



#convert the binary incidence matrix into a transactions database  
crossSelltrans.df <- as(crossSelldata.df, "transactions")  
  
inspect(head(crossSelltrans.df))

## items   
## [1] {Housewares.Division,   
## Health.Products.Division,   
## Automotive.Division,   
## Personal.Electronics.Division,  
## Novelty.Gift.Division}   
## [2] {Housewares.Division,   
## Health.Products.Division,   
## Automotive.Division,   
## Personal.Electronics.Division,  
## Garden.Division,   
## Novelty.Gift.Division,   
## Jewelry.Division}   
## [3] {Housewares.Division,   
## Health.Products.Division,   
## Automotive.Division,   
## Personal.Electronics.Division,  
## Garden.Division,   
## Novelty.Gift.Division,   
## Jewelry.Division}   
## [4] {Health.Products.Division,   
## Automotive.Division,   
## Personal.Electronics.Division,  
## Garden.Division,   
## Novelty.Gift.Division}   
## [5] {Health.Products.Division,   
## Personal.Electronics.Division,  
## Garden.Division,   
## Novelty.Gift.Division}   
## [6] {Housewares.Division,   
## Health.Products.Division,   
## Automotive.Division,   
## Personal.Electronics.Division,  
## Garden.Division,   
## Novelty.Gift.Division,   
## Jewelry.Division}

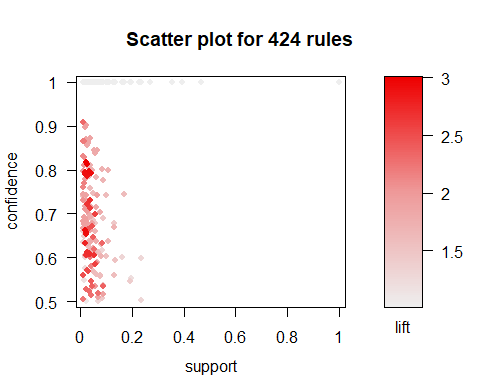
#Create Association Rules by running apriori(), with the minimum support as 0.01, minimum confidence as 0.5, and target as rules in arguments.  
rules <- apriori(crossSelltrans.df, parameter = list(supp = 0.01, conf = 0.5, target = "rules"))

## Apriori  
##   
## Parameter specification:  
## confidence minval smax arem aval originalSupport maxtime support minlen  
## 0.5 0.1 1 none FALSE TRUE 5 0.01 1  
## maxlen target ext  
## 10 rules FALSE  
##   
## Algorithmic control:  
## filter tree heap memopt load sort verbose  
## 0.1 TRUE TRUE FALSE TRUE 2 TRUE  
##   
## Absolute minimum support count: 49   
##   
## set item appearances ...[0 item(s)] done [0.00s].  
## set transactions ...[9 item(s), 4998 transaction(s)] done [0.00s].  
## sorting and recoding items ... [9 item(s)] done [0.00s].  
## creating transaction tree ... done [0.00s].  
## checking subsets of size 1 2 3 4 5 6 7 done [0.00s].  
## writing ... [424 rule(s)] done [0.00s].  
## creating S4 object ... done [0.00s].

# inspect the first 20 rules, sorted by their lift  
inspect(head(sort(rules, by = "lift"), n = 40))

## lhs rhs support confidence lift count  
## [1] {Automotive.Division,   
## Personal.Electronics.Division,   
## Novelty.Gift.Division,   
## Jewelry.Division} => {Garden.Division} 0.02400960 0.8163265 3.000000 120  
## [2] {Health.Products.Division,   
## Automotive.Division,   
## Personal.Electronics.Division,   
## Novelty.Gift.Division,   
## Jewelry.Division} => {Garden.Division} 0.02400960 0.8163265 3.000000 120  
## [3] {Automotive.Division,   
## Novelty.Gift.Division,   
## Jewelry.Division} => {Garden.Division} 0.02781112 0.8128655 2.987281 139  
## [4] {Health.Products.Division,   
## Automotive.Division,   
## Novelty.Gift.Division,   
## Jewelry.Division} => {Garden.Division} 0.02781112 0.8128655 2.987281 139  
## [5] {Automotive.Division,   
## Personal.Electronics.Division,   
## Novelty.Gift.Division} => {Garden.Division} 0.03541417 0.7972973 2.930068 177  
## [6] {Health.Products.Division,   
## Automotive.Division,   
## Personal.Electronics.Division,   
## Novelty.Gift.Division} => {Garden.Division} 0.03541417 0.7972973 2.930068 177  
## [7] {Housewares.Division,   
## Automotive.Division,   
## Novelty.Gift.Division,   
## Jewelry.Division} => {Garden.Division} 0.01920768 0.7933884 2.915702 96  
## [8] {Housewares.Division,   
## Health.Products.Division,   
## Automotive.Division,   
## Novelty.Gift.Division,   
## Jewelry.Division} => {Garden.Division} 0.01920768 0.7933884 2.915702 96  
## [9] {Automotive.Division,   
## Novelty.Gift.Division} => {Garden.Division} 0.04401761 0.7913669 2.908273 220  
## [10] {Health.Products.Division,   
## Automotive.Division,   
## Novelty.Gift.Division} => {Garden.Division} 0.04401761 0.7913669 2.908273 220  
## [11] {Housewares.Division,   
## Automotive.Division,   
## Personal.Electronics.Division,   
## Garden.Division,   
## Jewelry.Division} => {Novelty.Gift.Division} 0.01720688 0.6615385 2.907976 86  
## [12] {Housewares.Division,   
## Health.Products.Division,   
## Automotive.Division,   
## Personal.Electronics.Division,   
## Garden.Division,   
## Jewelry.Division} => {Novelty.Gift.Division} 0.01720688 0.6615385 2.907976 86  
## [13] {Housewares.Division,   
## Automotive.Division,   
## Personal.Electronics.Division,   
## Novelty.Gift.Division} => {Garden.Division} 0.02420968 0.7908497 2.906373 121  
## [14] {Housewares.Division,   
## Health.Products.Division,   
## Automotive.Division,   
## Personal.Electronics.Division,   
## Novelty.Gift.Division} => {Garden.Division} 0.02420968 0.7908497 2.906373 121  
## [15] {Housewares.Division,   
## Automotive.Division,   
## Personal.Electronics.Division,   
## Novelty.Gift.Division,   
## Jewelry.Division} => {Garden.Division} 0.01720688 0.7889908 2.899541 86  
## [16] {Housewares.Division,   
## Health.Products.Division,   
## Automotive.Division,   
## Personal.Electronics.Division,   
## Novelty.Gift.Division,   
## Jewelry.Division} => {Garden.Division} 0.01720688 0.7889908 2.899541 86  
## [17] {Housewares.Division,   
## Automotive.Division,   
## Personal.Electronics.Division,   
## Garden.Division} => {Novelty.Gift.Division} 0.02420968 0.6576087 2.890702 121  
## [18] {Housewares.Division,   
## Health.Products.Division,   
## Automotive.Division,   
## Personal.Electronics.Division,   
## Garden.Division} => {Novelty.Gift.Division} 0.02420968 0.6576087 2.890702 121  
## [19] {Housewares.Division,   
## Automotive.Division,   
## Novelty.Gift.Division} => {Garden.Division} 0.02781112 0.7853107 2.886017 139  
## [20] {Housewares.Division,   
## Health.Products.Division,   
## Automotive.Division,   
## Novelty.Gift.Division} => {Garden.Division} 0.02781112 0.7853107 2.886017 139  
## [21] {Automotive.Division,   
## Personal.Electronics.Division,   
## Garden.Division,   
## Jewelry.Division} => {Novelty.Gift.Division} 0.02400960 0.6521739 2.866812 120  
## [22] {Health.Products.Division,   
## Automotive.Division,   
## Personal.Electronics.Division,   
## Garden.Division,   
## Jewelry.Division} => {Novelty.Gift.Division} 0.02400960 0.6521739 2.866812 120  
## [23] {Housewares.Division,   
## Automotive.Division,   
## Garden.Division,   
## Jewelry.Division} => {Novelty.Gift.Division} 0.01920768 0.6315789 2.776281 96  
## [24] {Housewares.Division,   
## Health.Products.Division,   
## Automotive.Division,   
## Garden.Division,   
## Jewelry.Division} => {Novelty.Gift.Division} 0.01920768 0.6315789 2.776281 96  
## [25] {Housewares.Division,   
## Automotive.Division,   
## Garden.Division} => {Novelty.Gift.Division} 0.02781112 0.6123348 2.691688 139  
## [26] {Housewares.Division,   
## Health.Products.Division,   
## Automotive.Division,   
## Garden.Division} => {Novelty.Gift.Division} 0.02781112 0.6123348 2.691688 139  
## [27] {Housewares.Division,   
## Automotive.Division,   
## Personal.Electronics.Division} => {Garden.Division} 0.03681473 0.7301587 2.683333 184  
## [28] {Housewares.Division,   
## Health.Products.Division,   
## Automotive.Division,   
## Personal.Electronics.Division} => {Garden.Division} 0.03681473 0.7301587 2.683333 184  
## [29] {Automotive.Division,   
## Personal.Electronics.Division,   
## Garden.Division} => {Novelty.Gift.Division} 0.03541417 0.6103448 2.682941 177  
## [30] {Health.Products.Division,   
## Automotive.Division,   
## Personal.Electronics.Division,   
## Garden.Division} => {Novelty.Gift.Division} 0.03541417 0.6103448 2.682941 177  
## [31] {Housewares.Division,   
## Personal.Electronics.Division,   
## Garden.Division,   
## Jewelry.Division} => {Novelty.Gift.Division} 0.03821529 0.6082803 2.673865 191  
## [32] {Housewares.Division,   
## Health.Products.Division,   
## Personal.Electronics.Division,   
## Garden.Division,   
## Jewelry.Division} => {Novelty.Gift.Division} 0.03821529 0.6082803 2.673865 191  
## [33] {Housewares.Division,   
## Automotive.Division,   
## Personal.Electronics.Division} => {Novelty.Gift.Division} 0.03061224 0.6071429 2.668865 153  
## [34] {Housewares.Division,   
## Health.Products.Division,   
## Automotive.Division,   
## Personal.Electronics.Division} => {Novelty.Gift.Division} 0.03061224 0.6071429 2.668865 153  
## [35] {Personal.Electronics.Division,   
## Garden.Division,   
## Jewelry.Division} => {Novelty.Gift.Division} 0.05322129 0.6059226 2.663501 266  
## [36] {Health.Products.Division,   
## Personal.Electronics.Division,   
## Garden.Division,   
## Jewelry.Division} => {Novelty.Gift.Division} 0.05322129 0.6059226 2.663501 266  
## [37] {Housewares.Division,   
## Automotive.Division,   
## Personal.Electronics.Division,   
## Jewelry.Division} => {Novelty.Gift.Division} 0.02180872 0.6055556 2.661888 109  
## [38] {Housewares.Division,   
## Health.Products.Division,   
## Automotive.Division,   
## Personal.Electronics.Division,   
## Jewelry.Division} => {Novelty.Gift.Division} 0.02180872 0.6055556 2.661888 109  
## [39] {Housewares.Division,   
## Automotive.Division,   
## Personal.Electronics.Division,   
## Jewelry.Division} => {Garden.Division} 0.02601040 0.7222222 2.654167 130  
## [40] {Housewares.Division,   
## Health.Products.Division,   
## Automotive.Division,   
## Personal.Electronics.Division,   
## Jewelry.Division} => {Garden.Division} 0.02601040 0.7222222 2.654167 130

#bivariate relationship Scatterplot of all rules  
plot(rules, jitter = 0)

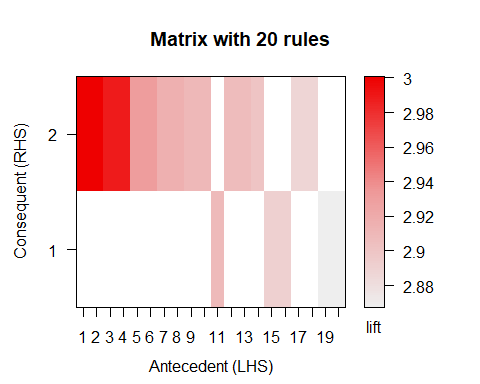


#Limit rules those with lift above 2.8  
subrules <- subset(rules, lift > 2.8)  
  
inspect(sort(subrules, by = "lift"))

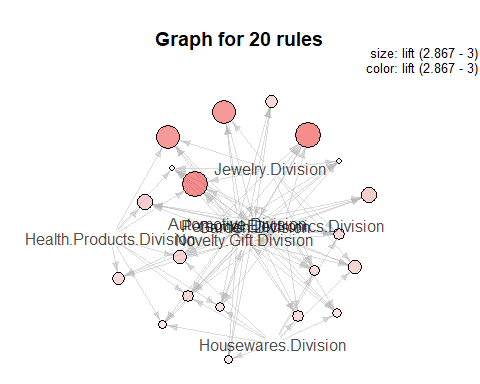
## lhs rhs support confidence lift count  
## [1] {Automotive.Division,   
## Personal.Electronics.Division,   
## Novelty.Gift.Division,   
## Jewelry.Division} => {Garden.Division} 0.02400960 0.8163265 3.000000 120  
## [2] {Health.Products.Division,   
## Automotive.Division,   
## Personal.Electronics.Division,   
## Novelty.Gift.Division,   
## Jewelry.Division} => {Garden.Division} 0.02400960 0.8163265 3.000000 120  
## [3] {Automotive.Division,   
## Novelty.Gift.Division,   
## Jewelry.Division} => {Garden.Division} 0.02781112 0.8128655 2.987281 139  
## [4] {Health.Products.Division,   
## Automotive.Division,   
## Novelty.Gift.Division,   
## Jewelry.Division} => {Garden.Division} 0.02781112 0.8128655 2.987281 139  
## [5] {Automotive.Division,   
## Personal.Electronics.Division,   
## Novelty.Gift.Division} => {Garden.Division} 0.03541417 0.7972973 2.930068 177  
## [6] {Health.Products.Division,   
## Automotive.Division,   
## Personal.Electronics.Division,   
## Novelty.Gift.Division} => {Garden.Division} 0.03541417 0.7972973 2.930068 177  
## [7] {Housewares.Division,   
## Automotive.Division,   
## Novelty.Gift.Division,   
## Jewelry.Division} => {Garden.Division} 0.01920768 0.7933884 2.915702 96  
## [8] {Housewares.Division,   
## Health.Products.Division,   
## Automotive.Division,   
## Novelty.Gift.Division,   
## Jewelry.Division} => {Garden.Division} 0.01920768 0.7933884 2.915702 96  
## [9] {Automotive.Division,   
## Novelty.Gift.Division} => {Garden.Division} 0.04401761 0.7913669 2.908273 220  
## [10] {Health.Products.Division,   
## Automotive.Division,   
## Novelty.Gift.Division} => {Garden.Division} 0.04401761 0.7913669 2.908273 220  
## [11] {Housewares.Division,   
## Automotive.Division,   
## Personal.Electronics.Division,   
## Garden.Division,   
## Jewelry.Division} => {Novelty.Gift.Division} 0.01720688 0.6615385 2.907976 86  
## [12] {Housewares.Division,   
## Health.Products.Division,   
## Automotive.Division,   
## Personal.Electronics.Division,   
## Garden.Division,   
## Jewelry.Division} => {Novelty.Gift.Division} 0.01720688 0.6615385 2.907976 86  
## [13] {Housewares.Division,   
## Automotive.Division,   
## Personal.Electronics.Division,   
## Novelty.Gift.Division} => {Garden.Division} 0.02420968 0.7908497 2.906373 121  
## [14] {Housewares.Division,   
## Health.Products.Division,   
## Automotive.Division,   
## Personal.Electronics.Division,   
## Novelty.Gift.Division} => {Garden.Division} 0.02420968 0.7908497 2.906373 121  
## [15] {Housewares.Division,   
## Automotive.Division,   
## Personal.Electronics.Division,   
## Novelty.Gift.Division,   
## Jewelry.Division} => {Garden.Division} 0.01720688 0.7889908 2.899541 86  
## [16] {Housewares.Division,   
## Health.Products.Division,   
## Automotive.Division,   
## Personal.Electronics.Division,   
## Novelty.Gift.Division,   
## Jewelry.Division} => {Garden.Division} 0.01720688 0.7889908 2.899541 86  
## [17] {Housewares.Division,   
## Automotive.Division,   
## Personal.Electronics.Division,   
## Garden.Division} => {Novelty.Gift.Division} 0.02420968 0.6576087 2.890702 121  
## [18] {Housewares.Division,   
## Health.Products.Division,   
## Automotive.Division,   
## Personal.Electronics.Division,   
## Garden.Division} => {Novelty.Gift.Division} 0.02420968 0.6576087 2.890702 121  
## [19] {Housewares.Division,   
## Automotive.Division,   
## Novelty.Gift.Division} => {Garden.Division} 0.02781112 0.7853107 2.886017 139  
## [20] {Housewares.Division,   
## Health.Products.Division,   
## Automotive.Division,   
## Novelty.Gift.Division} => {Garden.Division} 0.02781112 0.7853107 2.886017 139  
## [21] {Automotive.Division,   
## Personal.Electronics.Division,   
## Garden.Division,   
## Jewelry.Division} => {Novelty.Gift.Division} 0.02400960 0.6521739 2.866812 120  
## [22] {Health.Products.Division,   
## Automotive.Division,   
## Personal.Electronics.Division,   
## Garden.Division,   
## Jewelry.Division} => {Novelty.Gift.Division} 0.02400960 0.6521739 2.866812 120

#Matrix of top 20 Rules  
plot(subrules[1:20], method = "matrix", measure = "lift")

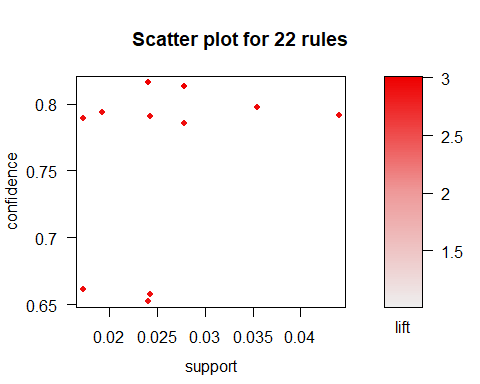
## Itemsets in Antecedent (LHS)  
## [1] "{Automotive.Division,Personal.Electronics.Division,Novelty.Gift.Division,Jewelry.Division}"   
## [2] "{Health.Products.Division,Automotive.Division,Personal.Electronics.Division,Novelty.Gift.Division,Jewelry.Division}"   
## [3] "{Automotive.Division,Novelty.Gift.Division,Jewelry.Division}"   
## [4] "{Health.Products.Division,Automotive.Division,Novelty.Gift.Division,Jewelry.Division}"   
## [5] "{Automotive.Division,Personal.Electronics.Division,Novelty.Gift.Division}"   
## [6] "{Health.Products.Division,Automotive.Division,Personal.Electronics.Division,Novelty.Gift.Division}"   
## [7] "{Housewares.Division,Automotive.Division,Novelty.Gift.Division,Jewelry.Division}"   
## [8] "{Housewares.Division,Health.Products.Division,Automotive.Division,Novelty.Gift.Division,Jewelry.Division}"   
## [9] "{Automotive.Division,Novelty.Gift.Division}"   
## [10] "{Health.Products.Division,Automotive.Division,Novelty.Gift.Division}"   
## [11] "{Housewares.Division,Automotive.Division,Personal.Electronics.Division,Garden.Division,Jewelry.Division}"   
## [12] "{Housewares.Division,Automotive.Division,Personal.Electronics.Division,Novelty.Gift.Division}"   
## [13] "{Housewares.Division,Health.Products.Division,Automotive.Division,Personal.Electronics.Division,Novelty.Gift.Division}"  
## [14] "{Housewares.Division,Automotive.Division,Personal.Electronics.Division,Novelty.Gift.Division,Jewelry.Division}"   
## [15] "{Housewares.Division,Automotive.Division,Personal.Electronics.Division,Garden.Division}"   
## [16] "{Housewares.Division,Health.Products.Division,Automotive.Division,Personal.Electronics.Division,Garden.Division}"   
## [17] "{Housewares.Division,Automotive.Division,Novelty.Gift.Division}"   
## [18] "{Housewares.Division,Health.Products.Division,Automotive.Division,Novelty.Gift.Division}"   
## [19] "{Automotive.Division,Personal.Electronics.Division,Garden.Division,Jewelry.Division}"   
## [20] "{Health.Products.Division,Automotive.Division,Personal.Electronics.Division,Garden.Division,Jewelry.Division}"   
## Itemsets in Consequent (RHS)  
## [1] "{Novelty.Gift.Division}" "{Garden.Division}"



#Graph of top 20 rules  
plot(subrules[1:20], method = "graph", measure = "lift")



#Scatterplot of subrules  
plot(subrules, jitter = 0)



**Portfolio Project: Cross-Selling Data Mining Using Association Rules**

For the Portfolio Project, I have analyzed the Cross selling dataset to determine cross selling association rules using R, Association rules basically helps identifying items that can be grouped/clustered together in transaction, or can be offered post-transaction via mail, coupons etc. Any retail industry looks forward to performs this kind of data mining procedure to identify buying patterns from historical customers and providing similar products help organization to increase sales, and in fact minimizes the search efforts of customer via upselling. Association rules works in 2 step process at first, it generates the rules and then further assessment of rule strength is done on result set. For this project exercise- We uploaded the CatalogCrossSell.csv as our dataset and used association rule-generating algorithm named-Apriori and inspected the data by judging the strength of rules using lift ratio.

At first, we identified items opted together, as this is first step in association rules to generate all the rules that would be candidates for indicating associations between items. After finding such combinations of single items, pairs of items, triplets of items, and so on, formed our transactions database. I have also created bar plot to understand the sale of individual items and to give bird eye view of sale. Then to determine strong correlation between the antecedent and consequent item sets, we executed apriori with minimum support as 0.01 and minimum confidence as 0.5. Form the rules object- I inspected the first 40 rules by sorting with lift ratio, and by viewing results, could translate from the rule that if- Sale in Automotive.Division, Personal.Electronics.Division, Novelty.Gift.Division, Jewelry.Division is done then with confidence of 81%, I can say that have potential selling Garden.Division as well with lift ratio of 3.0.

Additionally, I have further explored the result set by adding limiting rule logic to analyze all combinations which have lift ratio above 2.8, Actually Lift ratio is ideal way to judge the strength of any association results since this compares the confidence of the rule with a benchmark value. Hence analyzing the data set with various lift ratio provided detailed insights of correlation and possible groups, further using result set of lift ratio greater than 2.8, I have visualized data with Matrix, Graph and Scatterplot.