4. Association Rule Mining implementation in R and SAS EM

We begin with importing our dataset to both R programming and SAS Enterprise Miner. We will go through both approaches one at a time.

A. Association Rule Mining with R

Firstly, we take deeper into our data and its type using str function in R Output [1].

```
> str(transactions)
'data.frame':   25 obs. of  20 variables:
 $ SPIRITS           : Factor w/ 2 levels "NO","YES": 2 2 2 2 2 2 2 2 2 2 ...
 $ CANDY             : Factor w/ 2 levels "NO","YES": 2 2 2 2 2 2 2 2 2 2 ...
 $ BEVERAGES         : Factor w/ 2 levels "NO","YES": 2 2 2 2 2 2 2 2 2 2 ...
 $ HEALTH.AIDS       : Factor w/ 2 levels "NO","YES": 2 2 2 2 2 2 2 2 2 2 ...
 $ WINE              : Factor w/ 2 levels "NO","YES": 2 2 2 2 2 2 2 2 2 2 ...
 $ TOBACCO           : Factor w/ 2 levels "NO","YES": 2 2 2 2 2 2 2 2 2 1 ...
 $ HOUSEHOLD.CLEANING: Factor w/ 2 levels "NO","YES": 2 2 2 2 2 2 2 2 2 2 ...
 $ PERSONAL.CARE     : Factor w/ 2 levels "NO","YES": 2 2 2 2 2 2 2 2 2 2 ...
 $ BEAUTY.CARE       : Factor w/ 2 levels "NO","YES": 2 2 2 2 2 2 2 2 2 2 ...
 $ BEER              : Factor w/ 2 levels "NO","YES": 2 2 2 2 2 2 2 2 2 2 ...
 $ PET.SUPPLIES      : Factor w/ 2 levels "NO","YES": 2 2 2 2 2 2 2 2 2 2 ...
 $ AMERICAN.GREETINGS: Factor w/ 2 levels "NO","YES": 2 2 2 2 2 2 2 2 2 2 ...
 $ GENERAL.GROCERIES : Factor w/ 2 levels "NO","YES": 2 2 2 2 2 2 2 2 2 2 ...
 $ TOYS              : Factor w/ 2 levels "NO","YES": 2 2 2 2 2 2 2 2 2 2 ...
 $ BABY.CARE         : Factor w/ 2 levels "NO","YES": 1 2 2 2 2 2 2 2 2 2 ...
 $ SUNGLASSES        : Factor w/ 2 levels "NO","YES": 2 1 1 2 2 2 2 2 2 1 ...
 $ BATTERIES         : Factor w/ 2 levels "NO","YES": 2 1 1 2 2 2 2 1 2 2 ...
 $ BATH              : Factor w/ 2 levels "NO","YES": 2 2 1 1 2 2 2 2 2 2 ...
 $ BEDDING           : Factor w/ 2 levels "NO","YES": 2 2 2 1 2 1 2 2 2 2 ...
 $ GIFTS             : Factor w/ 2 levels "NO","YES": 1 1 1 1 1 1 1 1 1 1 ...
```

Output [1]

We compute sum of columns with item purchase status of “Yes” or “No” and we create a data frame to calculate the sum and frequency of occurrences of such instances as shown in Output [2].

```
> #colSums() function computes the sums of columns.
> YES <- colSums(transactions == "YES")
> YES
           SPIRITS              CANDY          BEVERAGES        HEALTH.AIDS               WINE            TOBACCO
                15                 19                 20                 18                 16                 14
HOUSEHOLD.CLEANING      PERSONAL.CARE        BEAUTY.CARE               BEER       PET.SUPPLIES AMERICAN.GREETINGS
                18                 16                 17                 15                 17                 18
 GENERAL.GROCERIES               TOYS          BABY.CARE         SUNGLASSES          BATTERIES               BATH
                15                 15                 13                 12                  9                 10
           BEDDING              GIFTS
                11                  1
> NO <-colSums(transactions=="NO")
> NO
           SPIRITS              CANDY          BEVERAGES        HEALTH.AIDS               WINE            TOBACCO
                10                  6                  5                  7                  9                 11
HOUSEHOLD.CLEANING      PERSONAL.CARE        BEAUTY.CARE               BEER       PET.SUPPLIES AMERICAN.GREETINGS
                 7                  9                  8                 10                  8                  7
 GENERAL.GROCERIES               TOYS          BABY.CARE         SUNGLASSES          BATTERIES               BATH
                10                 10                 12                 13                 16                 15
           BEDDING              GIFTS
                14                 24
```

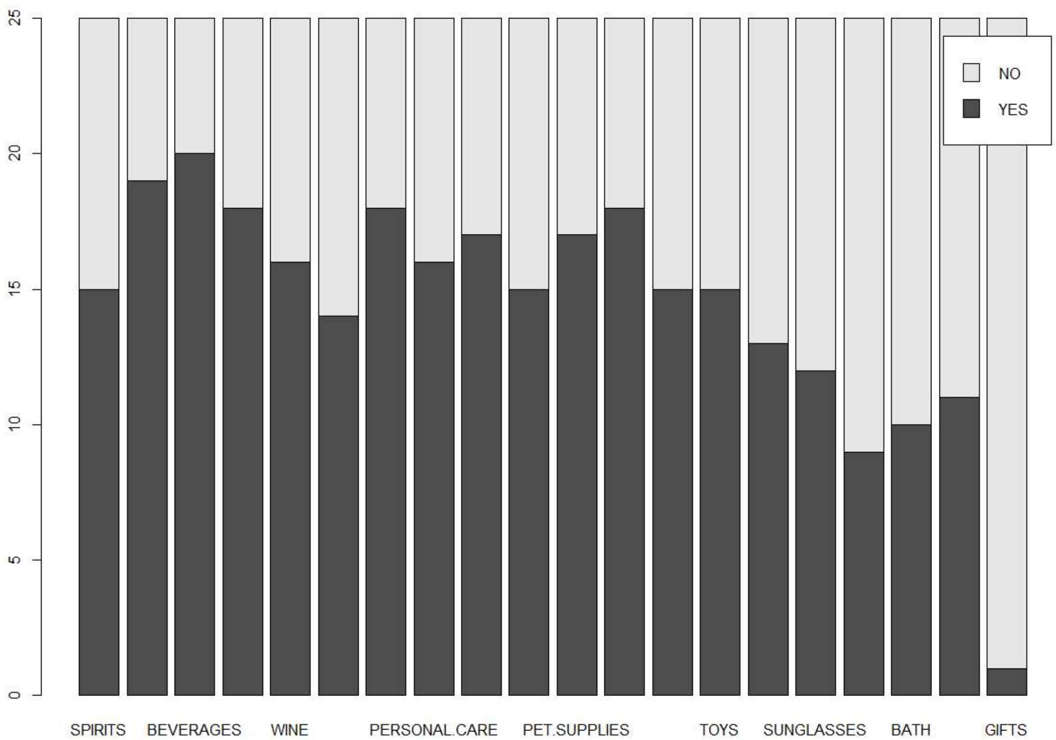
Output [2]

Now we combine both frequency into one data frame and view the data together as in output [3].

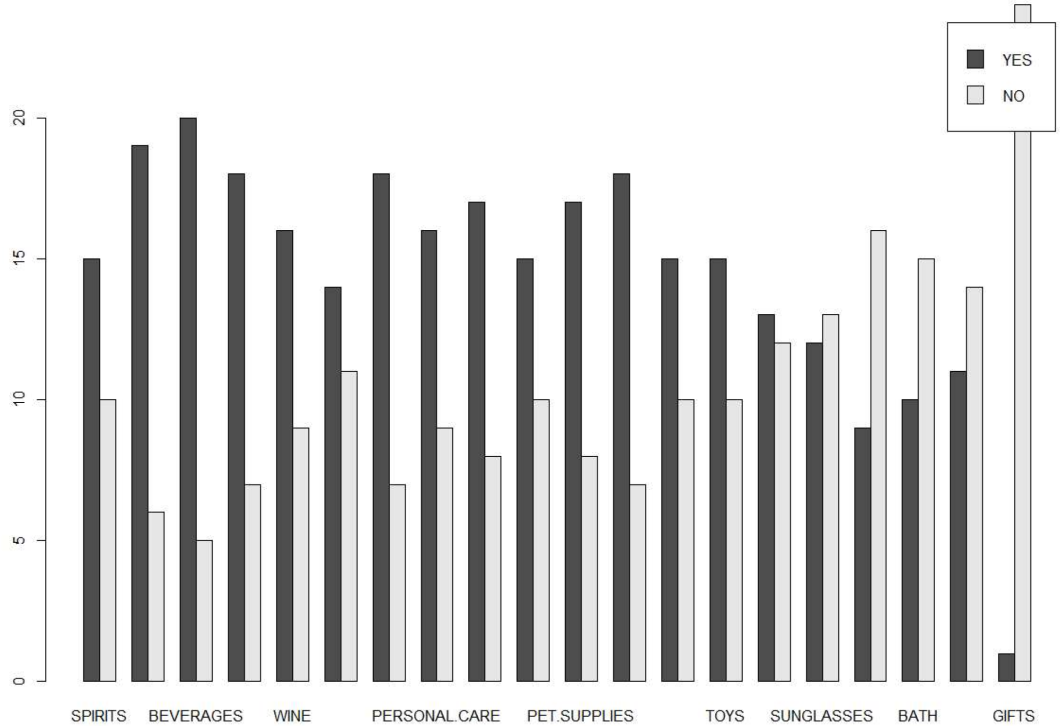
```
> visited <- rbind(YES,NO)
> visited
    SPIRITS CANDY BEVERAGES HEALTH.AIDS WINE TOBACCO HOUSEHOLD.CLEANING PERSONAL.CARE BEAUTY.CARE BEER PET.SUPPLIES
YES      15    19        20          18   16      14                 18            16          17   15           17
NO       10     6         5           7    9      11                  7             9           8   10            8
    AMERICAN.GREETINGS GENERAL.GROCERIES TOYS BABY.CARE SUNGLASSES BATTERIES BATH BEDDING GIFTS
YES                 18                15   15        13         12         9   10      11     1
NO                   7                10   10        12         13        16   15      14    24
```

Output [3]

We show our data statistics using bar plot, bar plot [1] shows the frequency of items and bar plot [2] shows the purchase frequency of the items in our data.



bar plot [1]



bar plot [2]

Using arules package and Apriori function call we create an association rule and out from first summary of rules we get 1 rule with 5 item and 226598 rules. Upon further inspection of rules, we get all the associations in our results, in our case we had to set the max limit to minimize overall rules creation.

We use apriori again to create rules with 95% confidence level as show in output [4].

```
> rules <- apriori(transactions, parameter =list(minlen=2,maxlen=3, conf = 0.95))
Apriori

Parameter specification:
 confidence minval smax arem  aval originalSupport maxtime support minlen maxlen target   ext
       0.95    0.1    1 none FALSE            TRUE       5     0.1      2      3  rules TRUE

Algorithmic control:
 filter tree heap memopt load sort verbose
    0.1 TRUE TRUE  FALSE TRUE    2    TRUE

Absolute minimum support count: 2
```

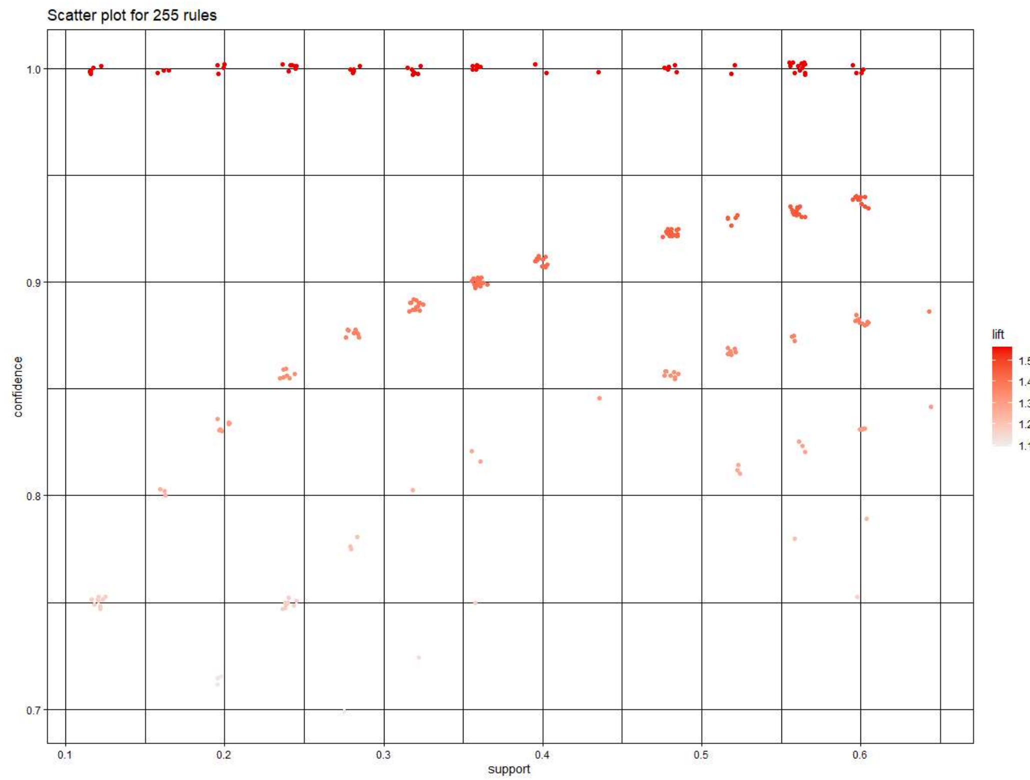
Output [4]

We will inspect the rules created and now we can see the list of items with related item association as in output [5].

```
> inspect(rules)
     lhs                           rhs                          support confidence coverage lift     count
[1]  {BEVERAGES=NO}             => {SPIRITS=NO}                 0.20    1          0.20     2.500000 5
[2]  {BEVERAGES=NO}             => {BEER=NO}                    0.20    1          0.20     2.500000 5
[3]  {BEVERAGES=NO}             => {TOBACCO=NO}                 0.20    1          0.20     2.272727 5
[4]  {BEVERAGES=NO}             => {BABY.CARE=NO}               0.20    1          0.20     2.083333 5
[5]  {BEVERAGES=NO}             => {BEDDING=NO}                 0.20    1          0.20     1.785714 5
[6]  {BEVERAGES=NO}             => {BATTERIES=NO}               0.20    1          0.20     1.562500 5
[7]  {BEVERAGES=NO}             => {GIFTS=NO}                   0.20    1          0.20     1.041667 5
[8]  {CANDY=NO}                 => {PERSONAL.CARE=NO}           0.24    1          0.24     2.777778 6
[9]  {CANDY=NO}                 => {WINE=NO}                    0.24    1          0.24     2.777778 6
[10] {CANDY=NO}                 => {SPIRITS=NO}                 0.24    1          0.24     2.500000 6
[11] {CANDY=NO}                 => {TOYS=NO}                    0.24    1          0.24     2.500000 6
[12] {CANDY=NO}                 => {GENERAL.GROCERIES=NO}       0.24    1          0.24     2.500000 6
[13] {CANDY=NO}                 => {BEER=NO}                    0.24    1          0.24     2.500000 6
[14] {CANDY=NO}                 => {BABY.CARE=NO}               0.24    1          0.24     2.083333 6
[15] {CANDY=NO}                 => {BEDDING=NO}                 0.24    1          0.24     1.785714 6
[16] {CANDY=NO}                 => {BATH=NO}                    0.24    1          0.24     1.666667 6
[17] {CANDY=NO}                 => {GIFTS=NO}                   0.24    1          0.24     1.041667 6
[18] {AMERICAN.GREETINGS=NO}    => {BATH=NO}                    0.28    1          0.28     1.666667 7
[19] {AMERICAN.GREETINGS=NO}    => {BATTERIES=NO}               0.28    1          0.28     1.562500 7
[20] {AMERICAN.GREETINGS=NO}    => {GIFTS=NO}                   0.28    1          0.28     1.041667 7
[21] {HOUSEHOLD.CLEANING=NO}    => {HEALTH.AIDS=NO}             0.28    1          0.28     3.571429 7
[22] {HEALTH.AIDS=NO}           => {HOUSEHOLD.CLEANING=NO}      0.28    1          0.28     3.571429 7
[23] {HOUSEHOLD.CLEANING=NO}    => {PET.SUPPLIES=NO}            0.28    1          0.28     3.125000 7
[24] {HOUSEHOLD.CLEANING=NO}    => {BEAUTY.CARE=NO}             0.28    1          0.28     3.125000 7
[25] {HOUSEHOLD.CLEANING=NO}    => {PERSONAL.CARE=NO}           0.28    1          0.28     2.777778 7
[26] {HOUSEHOLD.CLEANING=NO}    => {SPIRITS=NO}                 0.28    1          0.28     2.500000 7
[27] {HOUSEHOLD.CLEANING=NO}    => {TOYS=NO}                    0.28    1          0.28     2.500000 7
[28] {HOUSEHOLD.CLEANING=NO}    => {BEER=NO}                    0.28    1          0.28     2.500000 7
[29] {HOUSEHOLD.CLEANING=NO}    => {TOBACCO=NO}                 0.28    1          0.28     2.272727 7
[30] {HOUSEHOLD.CLEANING=NO}    => {BABY.CARE=NO}               0.28    1          0.28     2.083333 7
[31] {HOUSEHOLD.CLEANING=NO}    => {BEDDING=NO}                 0.28    1          0.28     1.785714 7
[32] {HOUSEHOLD.CLEANING=NO}    => {BATH=NO}                    0.28    1          0.28     1.666667 7
[33] {HOUSEHOLD.CLEANING=NO}    => {BATTERIES=NO}               0.28    1          0.28     1.562500 7
[34] {HOUSEHOLD.CLEANING=NO}    => {GIFTS=NO}                   0.28    1          0.28     1.041667 7
[35] {HEALTH.AIDS=NO}           => {PET.SUPPLIES=NO}            0.28    1          0.28     3.125000 7
[36] {HEALTH.AIDS=NO}           => {BEAUTY.CARE=NO}             0.28    1          0.28     3.125000 7
```

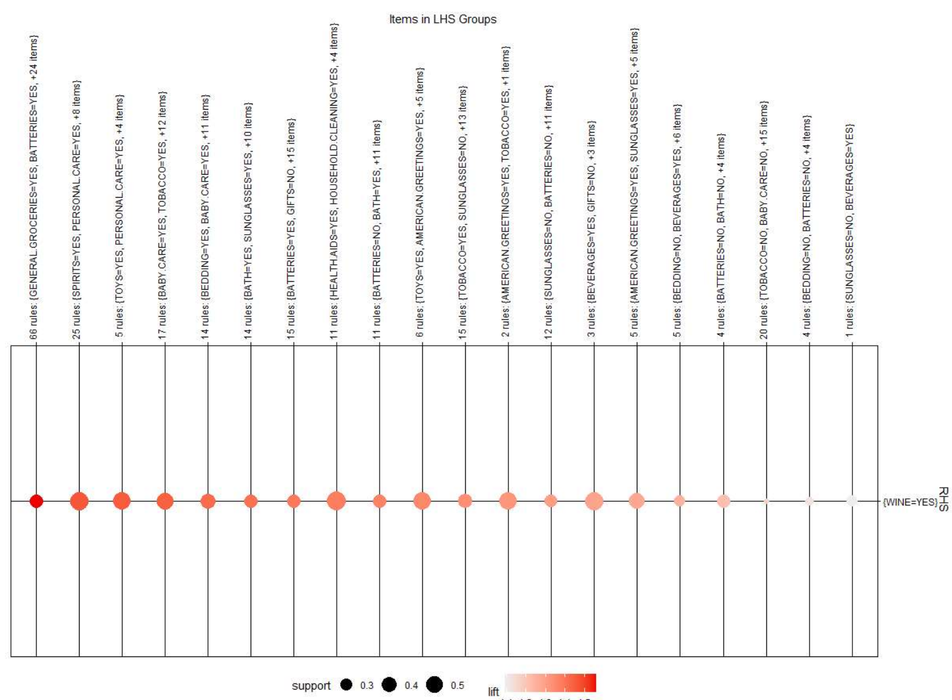
Output [5]

Plotting will enable us in finding custom association relation model of our data as show in scatter plot [1], here we are trying to find all rules associated to Wine with condition purchased is Yes.



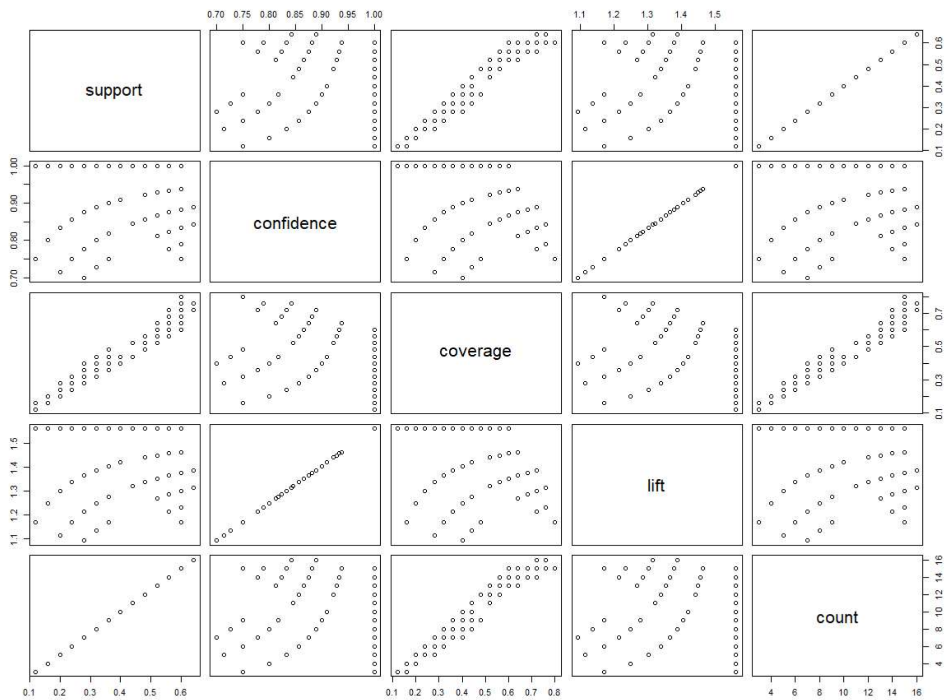
Scatter plot [1]

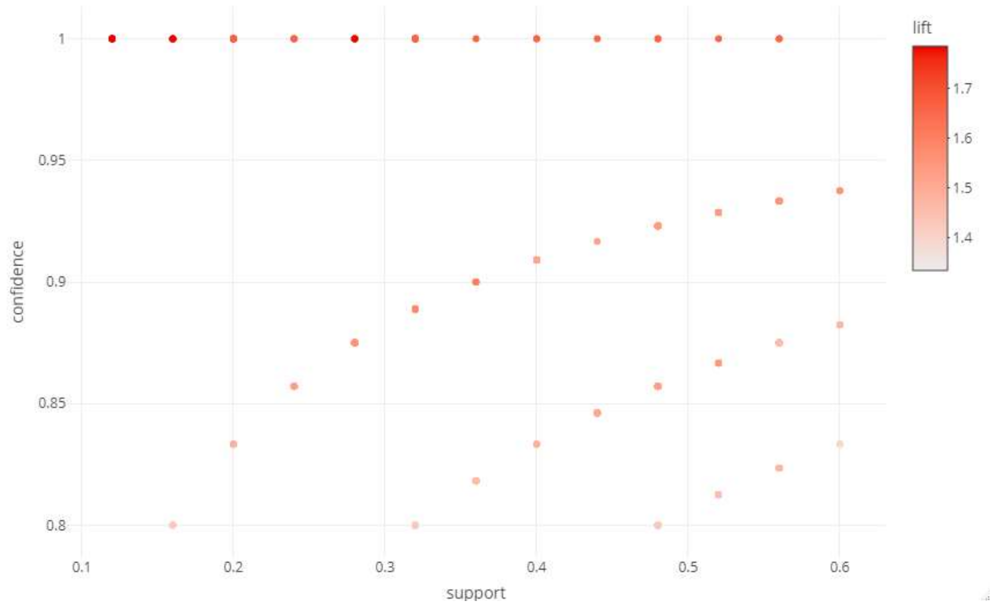
And the RHS and LHS rules quality and support measure are plotted together with lift parameter as show in plot [1].



Plot [1]

Now we test the same by creating 3rd rule for tabacco with purchase condition to Yes with apriori and show the results in statistics of confidence, support, coverage, lift and count.

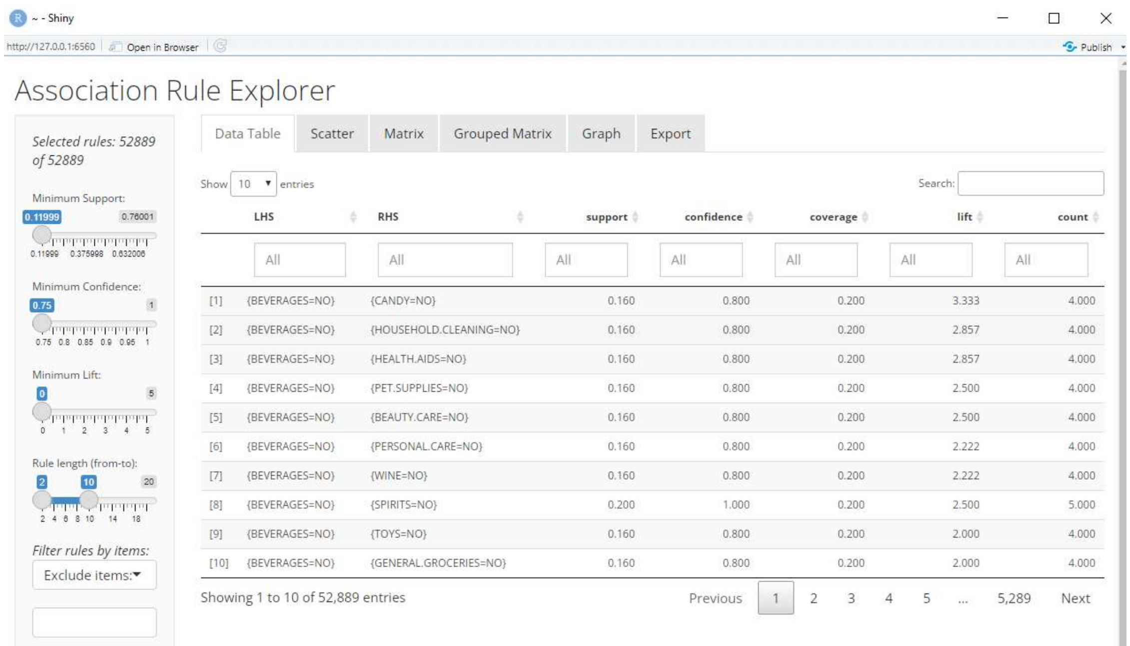




Scatter plot [2]

Scatter plot [2] show the confidence, support, and lift intercepts on our rule 3.

We can use shiny dashboard and explore our created rules or default rules association and show the relations of our data as shown in Figure [1].



	LHS	RHS	support	confidence	coverage	lift	count
[1]	{BEVERAGES=NO}	{CANDY=NO}	0.160	0.800	0.200	3.333	4.000
[2]	{BEVERAGES=NO}	{HOUSEHOLD.CLEANING=NO}	0.160	0.800	0.200	2.857	4.000
[3]	{BEVERAGES=NO}	{HEALTH.AIDS=NO}	0.160	0.800	0.200	2.857	4.000
[4]	{BEVERAGES=NO}	{PET.SUPPLIES=NO}	0.160	0.800	0.200	2.500	4.000
[5]	{BEVERAGES=NO}	{BEAUTY.CARE=NO}	0.160	0.800	0.200	2.500	4.000
[6]	{BEVERAGES=NO}	{PERSONAL.CARE=NO}	0.160	0.800	0.200	2.222	4.000
[7]	{BEVERAGES=NO}	{WINE=NO}	0.160	0.800	0.200	2.222	4.000
[8]	{BEVERAGES=NO}	{SPIRITS=NO}	0.200	1.000	0.200	2.500	5.000
[9]	{BEVERAGES=NO}	{TOYS=NO}	0.160	0.800	0.200	2.000	4.000
[10]	{BEVERAGES=NO}	{GENERAL.GROCERIES=NO}	0.160	0.800	0.200	2.000	4.000

Figure [1]

Appendix for Association Rule Mining in R

```r
transactions <-read.csv("C:/Users/Phani/Documents/Transactions.csv",header=T, colClasses="factor")
setwd("C:/Users/Phani/Documents")
names(transactions)
head(transactions)
tail(transactions)
summary(transactions)
str(transactions)
dim(transactions)
#colSums() function computes the sums of columns.
YES <- colSums(transactions == "YES")
YES
NO <-colSums(transactions=="NO")
NO
visited <- rbind(YES,NO)
visited
barplot(visited,legend=rownames(visited)) #Plot 1
barplot(visited, beside=T,legend=rownames(visited))# Plot 2
#arules package is a powerful tool for mining associative rules in transactional databases. The most
install.packages("arules") # install “arules” package.
library(arules) # activate “arules" package
#The apriori () function from the arule package implements the
#NOte that, by default, the apriori () function executes all the
#Usage of apriori () function
rules <- apriori(transactions, maxlen = 5)
summary(rules)
#The result tells you that there was 2 rules with 3 item and 255 rules
inspect(rules, maxlen=20)
#inspect function prints the internal representation of an R object
# when the max len parameter is NOt set, the algorithm continues
#set the minlen=2,maxlen=3 and confident = 0.95
rules <- apriori(transactions, parameter =list(minlen=2,maxlen=3, conf = 0.95))
summary(rules)
inspect(rules)
#Plotting could be the easiest way to find the most visited
rules <- apriori(transactions, parameter = list(minlen=2, maxlen=3,conf = 0.95), appearance= list(rhs=c("WINE=YES"),default="lhs"))
barplot(visited, beside=T,legend=rownames(rules))
#random variable selection
rules <- apriori(transactions,
                 parameter = list(minlen=2, maxlen=3,conf = 0.70),
                 appearance= list(rhs=c("WINE=YES"),default="lhs"))
summary(rules)
inspect(rules)
install.packages("arulesViz") # install “arulesViz"
library(arulesViz) # activate “arules" package
plot(rules)
plot(rules, jitter=0)
plot(rules, method="grouped")
plot(rules@quality)

rules3 <- apriori(transactions, parameter = list(minlen=2,maxlen=3, conf = 0.80), appearance =list(rhs=c("TOBACCO=YES","BEER=YES"),default="lhs"))
summary(rules3)
inspect(rules3)
barplot(visited, beside=T,legend=rownames(visited))
```

```r

library(plotly)
install.packages("arulesviz")
library(arulesviz)
plot(rules3, engine = "plotly")
#removing jitter
plot(rules3, engine = "plotly", jitter = 0)
plot(rules3, measure = c("support", "lift"), shading = "confidence", engine = "plotly")



rules2 <- apriori(transactions, parameter = list(minlen=2, maxlen=5,conf = 0.6),
                  appearance =list(rhs=c("BABY.CARE=YES"),lhs=c("BATTERIES=YES", "BEDDING=YES","GIFTS=YES", "CANDY=YES","BEVERAGES=YES",
                                   "HOUSEHOLD.CLEANING=YES","AMERICAN.GREETINGS=YES","SUNGLASSES=YES"),default="none"))
inspect(rules2)
rules_ex <-apriori(transactions,
                   parameter =list(minlen=2,maxlen=4,conf=0.75))
#Explore association rules using interactive manipulations and
ruleExplorer(rules_ex)
```

B. Association Rule Mining with SAS Enterprise Miner

Before to the initial step of importing data, we transform the data with only transaction and item set list as show below.

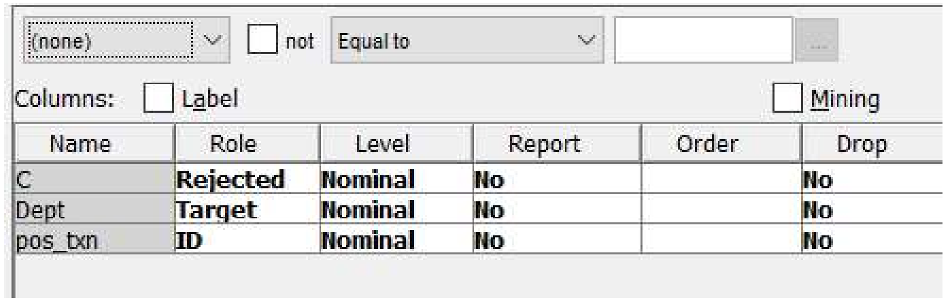
	A	B
1	pos_txn	Dept
2	16120100160021008773	HOSIERY
3	16120100160021008773	VITAMINS & HLTH AIDS
4	16120100160021008775	SMALL ELECTRICS
5	16120100160022004679	SPIRITS
6	16120100160023004356	SPORTS NUTRITION
7	16120100160023004356	SEASONAL
8	16120100160023004356	PET SUPPLIES
9	16120100160023004357	FITNESS/EXERCISE
10	16120100160023004358	FITNESS/EXERCISE
11	16120100160023004359	MISC CUSTOM SERVICES
12	16120100160023004359	TOBACCO
13	16120100160023004360	TOBACCO
14	16120100160023004361	GIRLS ACCESSORIES
15	16120100160023004361	TOYS

As our initial step we import our dataset to SAS EM library to perform association rule mining technique. Once import is complete, we will have to assigned role our input data to be transaction type as show in Figure [2].

Property	Value
Notes	
Train	
Variables	
Import File	C:\Users\Phani\Dow
Maximum Rows to Imp	1000000
Maximum Columns to	10000
Delimiter	,
Name Row	Yes
Number of Rows to Sk	0
Guessing Rows	500
File Location	Local
File Type	xlsx
Advanced Advisor	No
Rerun	No
Score	
Role	Transaction
Report	
Summarize	No

Figure 2

Now, set the variable roles of our dataset as shown in figure 3. Transaction ids to be ID and itemset to be Target variable.



Name	Role	Level	Report	Order	Drop
C	Rejected	Nominal	No		No
Dept	Target	Nominal	No		No
pos_txn	ID	Nominal	No		No

Figure 3

Our next step would be drag and drop association icon on to diagram platform and run association mining to get our rules formed by association of items in our data as shown in Figure 4.

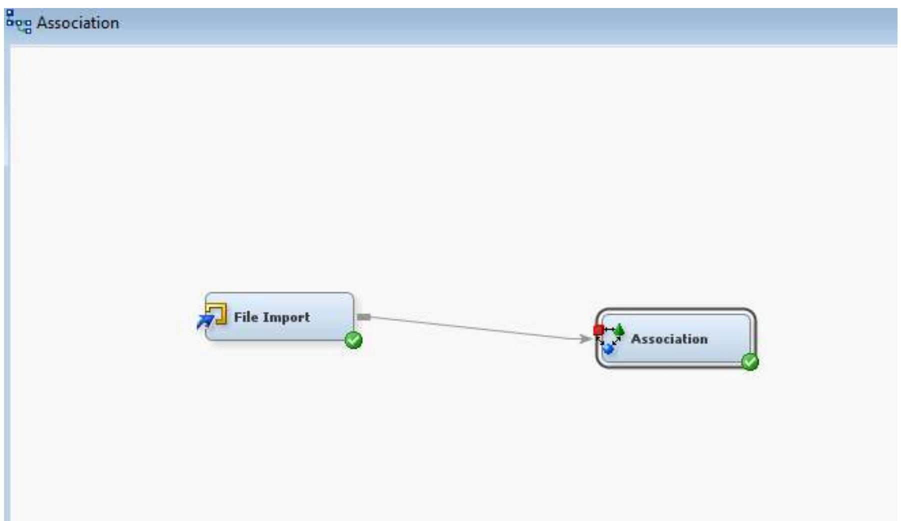
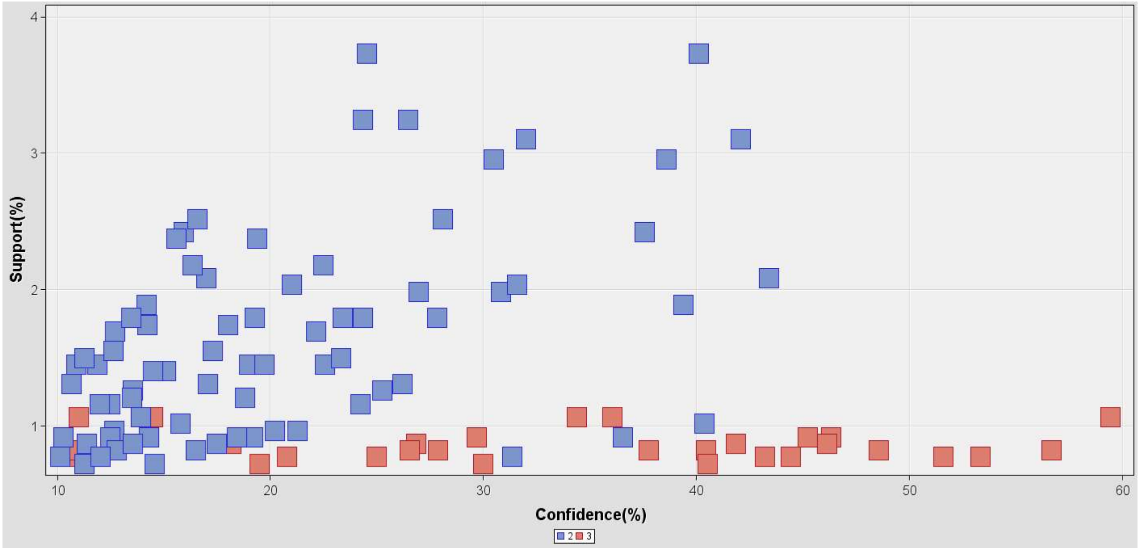


Figure 4

SAS Enterprise Miner now computes and create association rules among our itemset and we get our results as follows.

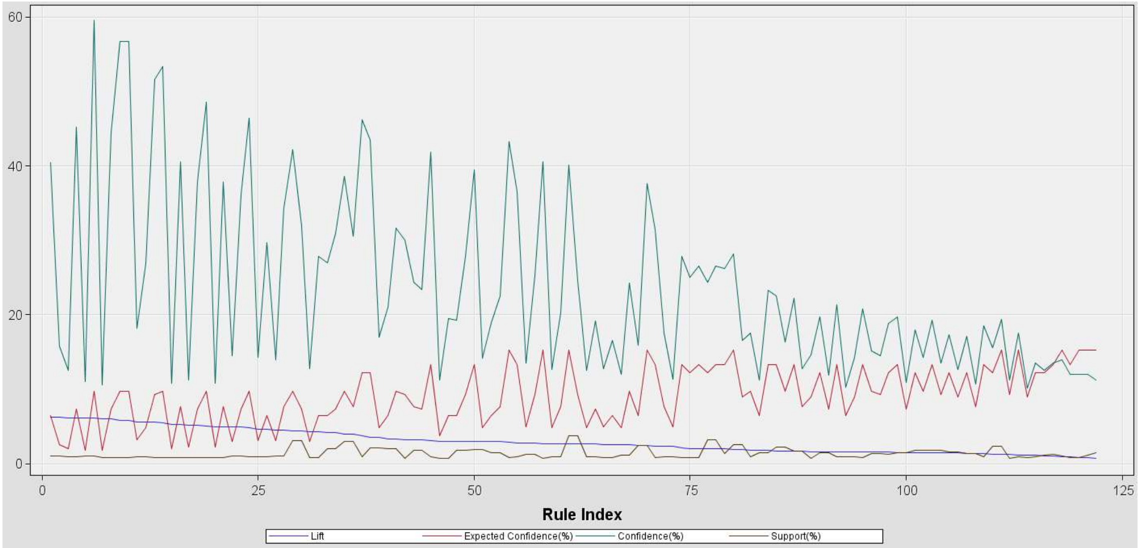


Output [6]

In output [6], we get the support of confident matrix table of individual rules which are created from our dataset.

Output [7]

Output [7], gives the RHS AND LHS comparison of individual item set and their confidence level.



Output 8

Output 8, show the line graph of confidence interval, support and lift of rules that got created and their rule index values.

Output

```
*------------------------------------------------------------*
User:                Phani
Date:                14 January 2022
Time:                14:22:13
*------------------------------------------------------------*
* Training Output
*------------------------------------------------------------*




Variable Summary

          Measurement    Frequency
Role         Level         Count

ID          NOMINAL          1
TARGET      NOMINAL          1




Association Report

              Expected
             Confidence   Confidence   Support            Transaction
Relations       (%)          (%)         (%)       Lift      Count      Rule                                          Left Hand of Rule

     2          6.44        40.38       1.02       6.27      21.00      MASS COSMETICS ==> BEAUTY CARE                MASS COSMETICS
     2          2.52        15.79       1.02       6.27      21.00      BEAUTY CARE ==> MASS COSMETICS                BEAUTY CARE
     3          2.03        12.50       0.92       6.14      19.00      PERSONAL CARE ==> HEALTH AIDS & BEAUTY CARE   PERSONAL CARE
     3          7.36        45.24       0.92       6.14      19.00      HEALTH AIDS & BEAUTY CARE ==> PERSONAL CARE   HEALTH AIDS & BEAUTY CARE
     3          9.69        59.46       1.07       6.14      22.00      PERSONAL CARE & HOUSEHOLD CLEANING ==> HEALTH AIDS   PERSONAL CARE & HOUSEHOLD CLEANING
     3          1.79        11.00       1.07       6.14      22.00      HEALTH AIDS ==> PERSONAL CARE & HOUSEHOLD CLEANING   HEALTH AIDS
     3          1.74        10.53       0.78       6.04      16.00      PERSONAL CARE ==> HEALTH AIDS & BEVERAGES     PERSONAL CARE
     3          7.36        44.44       0.78       6.04      16.00      HEALTH AIDS & BEVERAGES ==> PERSONAL CARE     HEALTH AIDS & BEVERAGES
     3          9.69        56.67       0.82       5.85      17.00      PERSONAL CARE & CANDY ==> HEALTH AIDS         PERSONAL CARE & CANDY
     3          9.69        56.67       0.82       5.85      17.00      HOUSEHOLD CLEANING & BEAUTY CARE ==> HEALTH AIDS   HOUSEHOLD CLEANING & BEAUTY CARE
     3          3.25        18.18       0.87       5.60      18.00      GENERAL GROCERIES ==> CANDY & BEVERAGES       GENERAL GROCERIES
     3          4.80        26.87       0.87       5.60      18.00      CANDY & BEVERAGES ==> GENERAL GROCERIES       CANDY & BEVERAGES
```

```
     3      9.30     51.61     0.78     5.55     16.00    SPIRITS & CANDY ==> WINE                           SPIRITS & CANDY
     3      9.69     53.33     0.78     5.50     16.00    PERSONAL CARE & BEVERAGES ==> HEALTH AIDS          PERSONAL CARE & BEVERAGES
     3      2.03     10.76     0.82     5.29     17.00    HOUSEHOLD CLEANING ==> HEALTH AIDS & BEAUTY CARE   HOUSEHOLD CLEANING
     3      7.66     40.48     0.82     5.29     17.00    HEALTH AIDS & BEAUTY CARE ==> HOUSEHOLD CLEANING   HEALTH AIDS & BEAUTY CARE
     3      2.18     11.18     0.82     5.13     17.00    PERSONAL CARE ==> HEALTH AIDS & CANDY              PERSONAL CARE
     3      7.36     37.78     0.82     5.13     17.00    HEALTH AIDS & CANDY ==> PERSONAL CARE              HEALTH AIDS & CANDY
     3      9.69     48.57     0.82     5.01     17.00    HOUSEHOLD CLEANING & CANDY ==> HEALTH AIDS         HOUSEHOLD CLEANING & CANDY
     3      2.18     10.76     0.82     4.94     17.00    HOUSEHOLD CLEANING ==> HEALTH AIDS & CANDY         HOUSEHOLD CLEANING
     3      7.66     37.78     0.82     4.94     17.00    HEALTH AIDS & CANDY ==> HOUSEHOLD CLEANING         HEALTH AIDS & CANDY
     3      2.96     14.47     1.07     4.90     22.00    PERSONAL CARE ==> HOUSEHOLD CLEANING & HEALTH AIDS   PERSONAL CARE
     3      7.36     36.07     1.07     4.90     22.00    HOUSEHOLD CLEANING & HEALTH AIDS ==> PERSONAL CARE   HOUSEHOLD CLEANING & HEALTH AIDS
     3      9.69     46.34     0.92     4.78     19.00    PERSONAL CARE & BEAUTY CARE ==> HEALTH AIDS        PERSONAL CARE & BEAUTY CARE
     3      6.44     29.69     0.92     4.61     19.00    PERSONAL CARE & HEALTH AIDS ==> BEAUTY CARE        PERSONAL CARE & HEALTH AIDS



                                                                                                              Rule     Rule
Right Hand of Rule                  Rule Item 1         Rule Item 2            Rule Item 3            Rule Item 4            Item 5   Index

BEAUTY CARE                         MASS COSMETICS      ===================>   BEAUTY CARE                                       1
MASS COSMETICS                      BEAUTY CARE         ===================>   MASS COSMETICS                                    2
HEALTH AIDS & BEAUTY CARE           PERSONAL CARE       ===================>   HEALTH AIDS            BEAUTY CARE                3
PERSONAL CARE                       HEALTH AIDS         BEAUTY CARE            ===================>   PERSONAL CARE              4
HEALTH AIDS                         PERSONAL CARE       HOUSEHOLD CLEANING     ===================>   HEALTH AIDS                6
PERSONAL CARE & HOUSEHOLD CLEANING  HEALTH AIDS         ===================>   PERSONAL CARE          HOUSEHOLD CLEANING         5
HEALTH AIDS & BEVERAGES             PERSONAL CARE       ===================>   HEALTH AIDS            BEVERAGES                  7
PERSONAL CARE                       HEALTH AIDS         BEVERAGES              ===================>   PERSONAL CARE              8
HEALTH AIDS                         PERSONAL CARE       CANDY                  ===================>   HEALTH AIDS                9
HEALTH AIDS                         HOUSEHOLD CLEANING  BEAUTY CARE            ===================>   HEALTH AIDS               10
CANDY & BEVERAGES                   GENERAL GROCERIES   ===================>   CANDY                  BEVERAGES                 11
GENERAL GROCERIES                   CANDY               BEVERAGES              ===================>   GENERAL GROCERIES         12
WINE                                SPIRITS             CANDY                  ===================>   WINE                      13
HEALTH AIDS                         PERSONAL CARE       BEVERAGES              ===================>   HEALTH AIDS               14
HEALTH AIDS & BEAUTY CARE           HOUSEHOLD CLEANING  ===================>   HEALTH AIDS            BEAUTY CARE               15
HOUSEHOLD CLEANING                  HEALTH AIDS         BEAUTY CARE            ===================>   HOUSEHOLD CLEANING        16
HEALTH AIDS & CANDY                 PERSONAL CARE       ===================>   HEALTH AIDS            CANDY                     17
PERSONAL CARE                       HEALTH AIDS         CANDY                  ===================>   PERSONAL CARE             18
HEALTH AIDS                         HOUSEHOLD CLEANING  CANDY                  ===================>   HEALTH AIDS               19
HEALTH AIDS & CANDY                 HOUSEHOLD CLEANING  ===================>   HEALTH AIDS            CANDY                     20
HOUSEHOLD CLEANING                  HEALTH AIDS         CANDY                  ===================>   HOUSEHOLD CLEANING        21
HOUSEHOLD CLEANING & HEALTH AIDS    PERSONAL CARE       ===================>   HOUSEHOLD CLEANING     HEALTH AIDS               22
PERSONAL CARE                       HOUSEHOLD CLEANING  HEALTH AIDS            ===================>   PERSONAL CARE             23
HEALTH AIDS                         PERSONAL CARE       BEAUTY CARE            ===================>   HEALTH AIDS               24
BEAUTY CARE                         PERSONAL CARE       HEALTH AIDS            ===================>   BEAUTY CARE               26




Rule Statistics

The MEANS Procedure

Variable    Label                          Minimum       Maximum          Mean
-------------------------------------------------------------------------------
EXP_CONF    Expected Confidence(%)       1.7441860    15.2131783     9.1311634
CONF        Confidence(%)               10.1265823    59.4594595    23.6447400
SUPPORT     Support(%)                   0.7267442     3.7306202     1.3685030
LIFT        Lift                         0.7409844     6.2672065     2.9782965
-------------------------------------------------------------------------------



Sequence Report

The FREQ Procedure

                       Relations

                                      Cumulative    Cumulative
SET_SIZE    Frequency     Percent      Frequency       Percent
---------------------------------------------------------------
       2           83       68.03             83         68.03
       3           39       31.97            122        100.00


*-------------------------------------------------------------*
* Score Output
*-------------------------------------------------------------*


*-------------------------------------------------------------*
```

Output 9

In our final output, (Output 9) we get variable summary, Association report of all items in dataset, RHS, LHS rule index. Also, some Rule Statistics such as The MEANS Procedure, Frequency Procedure reports are computed.

Upon more inspection we explore rules table we see 122 rules with relations in LHS vs RHS association rule format Output 10.

Rules Table

Rule Index	Relations	Expected Confidence(%)	Confidence(%)	Support(%)	Lift	Transaction Count	Rule	Left Hand of Rule	Right Hand of Rule	Rule Item 1	Rule Item 2	Rule Item 3	Rule Item 4	Rule Item 5	Transpose Rule
1	2	6.44	40.38	1.02	6.27	21.00	MASS C...	MASS C...	BEAUTY ...	MASS C...	======...	BEAUTY ...			1
2	2	2.52	15.79	1.02	6.27	21.00	BEAUTY ...	BEAUTY ...	MASS C...	BEAUTY ...	======...	MASS C...			1
3	3	2.03	12.50	0.92	6.14	19.00	PERSO...	PERSO...	HEALTH ...	PERSO...	======...	HEALTH ...	BEAUTY ...		1
4	3	7.36	45.24	0.92	6.14	19.00	HEALTH ...	HEALTH ...	PERSO...	HEALTH ...	BEAUTY ...	======...	PERSO...		1
6	3	9.69	59.46	1.07	6.14	22.00	PERSO...	PERSO...	HEALTH ...	PERSO...	HOUSEH...	======...	HEALTH ...		1
5	3	1.79	11.00	1.07	6.14	22.00	HEALTH ...	HEALTH ...	PERSO...	HEALTH ...	======...	PERSO...	HOUSEH...		1
7	3	1.74	10.53	0.78	6.04	16.00	PERSO...	PERSO...	HEALTH ...	PERSO...	======...	HEALTH ...	BEVERA...		1
8	3	7.36	44.44	0.78	6.04	16.00	HEALTH ...	HEALTH ...	PERSO...	HEALTH ...	BEVERA...	======...	PERSO...		1
9	3	9.69	56.67	0.82	5.85	17.00	PERSO...	PERSO...	HEALTH ...	PERSO...	CANDY	======...	HEALTH ...		1
10	3	9.69	56.67	0.82	5.85	17.00	HOUSEH...	HOUSEH...	HEALTH ...	HOUSEH...	BEAUTY ...	======...	HEALTH ...		1
11	3	3.25	18.18	0.87	5.60	18.00	GENERA...	GENERA...	CANDY ...	GENERA...	======...	CANDY	BEVERA...		1
12	3	4.80	26.87	0.87	5.60	18.00	CANDY ...	CANDY ...	GENERA...	CANDY	BEVERA...	======...	GENERA...		1
13	3	9.30	51.61	0.78	5.55	16.00	SPIRITS ...	SPIRITS ...	WINE	SPIRITS	CANDY	======...	WINE		1
14	3	9.69	53.33	0.78	5.50	16.00	PERSO...	PERSO...	HEALTH ...	PERSO...	BEVERA...	======...	HEALTH ...		1
15	3	2.03	10.76	0.82	5.29	17.00	HOUSEH...	HOUSEH...	HEALTH ...	HOUSEH...	======...	HEALTH ...	BEAUTY ...		1
16	3	7.66	40.48	0.82	5.29	17.00	HEALTH ...	HEALTH ...	HOUSEH...	HEALTH ...	BEAUTY ...	======...	HOUSEH...		1
17	3	2.18	11.18	0.82	5.13	17.00	PERSO...	PERSO...	HEALTH ...	PERSO...	======...	HEALTH ...	CANDY		1
18	3	7.36	37.78	0.82	5.13	17.00	HEALTH ...	HEALTH ...	PERSO...	HEALTH ...	CANDY	======...	PERSO...		1
19	3	9.69	48.57	0.82	5.01	17.00	HOUSEH...	HOUSEH...	HEALTH ...	HOUSEH...	CANDY	======...	HEALTH ...		1
20	3	2.18	10.76	0.82	4.94	17.00	HOUSEH...	HOUSEH...	HEALTH ...	HOUSEH...	======...	HEALTH ...	CANDY		1
21	3	7.66	37.78	0.82	4.94	17.00	HEALTH ...	HEALTH ...	HOUSEH...	HEALTH ...	CANDY	======...	HOUSEH...		1
22	3	2.96	14.47	1.07	4.90	22.00	PERSO...	PERSO...	HOUSEH...	PERSO...	======...	HOUSEH...	HEALTH ...		1
23	3	7.36	36.07	1.07	4.90	22.00	HOUSEH...	HOUSEH...	PERSO...	HOUSEH...	HEALTH ...	======...	PERSO...		1
24	3	9.69	46.34	0.92	4.78	19.00	PERSO...	PERSO...	HEALTH ...	PERSO...	BEAUTY ...	======...	HEALTH ...		1
26	3	6.44	29.69	0.92	4.61	19.00	PERSO...	PERSO...	BEAUTY ...	PERSO...	HEALTH ...	======...	BEAUTY ...		1
25	3	3.10	14.29	0.92	4.61	19.00	BEAUTY ...	BEAUTY ...	PERSO...	BEAUTY ...	======...	PERSO...	HEALTH ...		1
28	3	7.66	34.38	1.07	4.49	22.00	PERSO...	PERSO...	HOUSEH...	PERSO...	HEALTH ...	======...	HOUSEH...		1
27	3	3.10	13.92	1.07	4.49	22.00	HOUSEH...	HOUSEH...	PERSO...	HOUSEH...	======...	PERSO...	HEALTH ...		1
29	2	9.69	42.11	3.10	4.35	64.00	PERSO...	PERSO...	HEALTH ...	PERSO...	======...	HEALTH ...			1
30	2	7.36	32.00	3.10	4.35	64.00	HEALTH ...	HEALTH ...	PERSO...	HEALTH ...	======...	PERSO...			1
32	3	6.44	27.87	0.82	4.32	17.00	HOUSEH...	HOUSEH...	BEAUTY ...	HOUSEH...	HEALTH ...	======...	BEAUTY ...		1
31	3	2.96	12.78	0.82	4.32	17.00	BEAUTY ...	BEAUTY ...	HOUSEH...	BEAUTY ...	======...	HOUSEH...	HEALTH ...		1
33	2	6.44	26.97	1.99	4.19	41.00	PERSO...	PERSO...	BEAUTY ...	PERSO...	======...	BEAUTY ...			1
34	2	7.36	30.83	1.99	4.19	41.00	BEAUTY ...	BEAUTY ...	PERSO...	BEAUTY ...	======...	PERSO...			1
35	2	9.69	38.61	2.96	3.98	61.00	HOUSEH...	HOUSEH...	HEALTH ...	HOUSEH...	======...	HEALTH ...			1
36	2	7.66	30.50	2.96	3.98	61.00	HEALTH ...	HEALTH ...	HOUSEH...	HEALTH ...	======...	HOUSEH...			1

Output [10]

5. Conclusion and Output Comparison of R and SAS EM

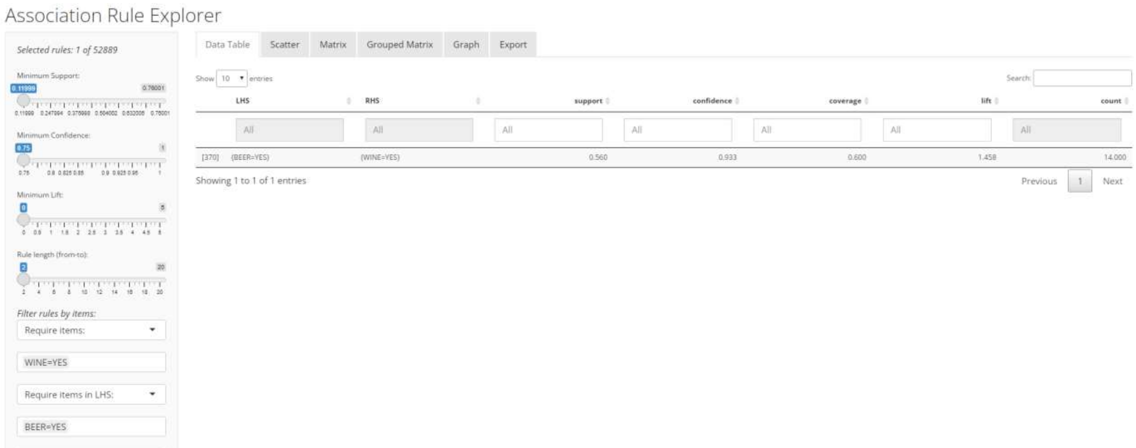
From R programming output we have rule explorer which enables us easy approach in finding associated rules between the selected items from our dataset. So, we take association rule SAS Enterprise miner and check the support, confidence, and lift of that rules with R programming computed rule explorer.

From the Rule Table Output [10], we pick WINE ==> BEER Output [11] as our rule to compare with R programming rule explorer we set WINE in LHS and BEER in RHS we get out initial support lift and confidence values. We are had change rule length to 2 since are testing only two items from our data set Output [12].

Rules Table

SET_SIZE	EXP_CONF	CONF	SUPPORT	LIFT	COUNT	RULE	_LHAND	_RHAND	ITEM1	ITEM2	ITEM3	ITEM4	ITEM5	index	Transpose
2	9.30	27.82	1.79	2.99	37.00	BEER ==> WINE	BEER	WINE	BEER	======...	WINE			49	

Output [11]



Output [12]

Comparing the default values of both R results and SAS Enterprise Miner results, we get differences in count, support, lift and confidence as well, if we could tune minimum support and minimum lift in rule explorer, we can achieve same results as of from SAS Miner. By this, we came to conclusion that R programming and SAS EM although may have difference in execution, but we may achieve similar statistical values.

References

1. Agrawal R, Imielinski T, Swami A. Mining association rules between sets of items in large databases. In: Proceedings of ACM SIGMOD conference on management of data (SIGMOD 1993). New York: ACM; 1993. p. 207216.
2. Fayyad U, Piatetsky-Shapiro G, Smyth P. Knowledge discovery and data mining: Towards a unifying framework. InProc. KDD-96: Second International Conference on Knowledge Discovery & Data Mining Menlo Park, CA: AAAI Press, 1996, pp.82–88.
3. Sanmiquel L, Rossell JM, Vintro C. Study of Spanish mining accidents using data mining techniques. Saf Sci. 2015;75:49–55.
4. Hussein, N., Alashqur, A. and Sowan, B., 2015. Using the interestingness measure lift to generate association rules. Journal of Advanced Computer Science & Technology, 4(1), p.156.
5. A. Agresti. Categorical Data Analysis. John Wiley & Sons, 1990.
6. https://data.world/zpencer/transaction-itemset/workspace/file?filename=transactions_by_dept.csv