MARKETING & RETAIL ANALYTICS ASSIGNMENT 2

MARKET BASKET ANALYSIS

SUBMITTED BY :-

ABHISHEK RANJAN

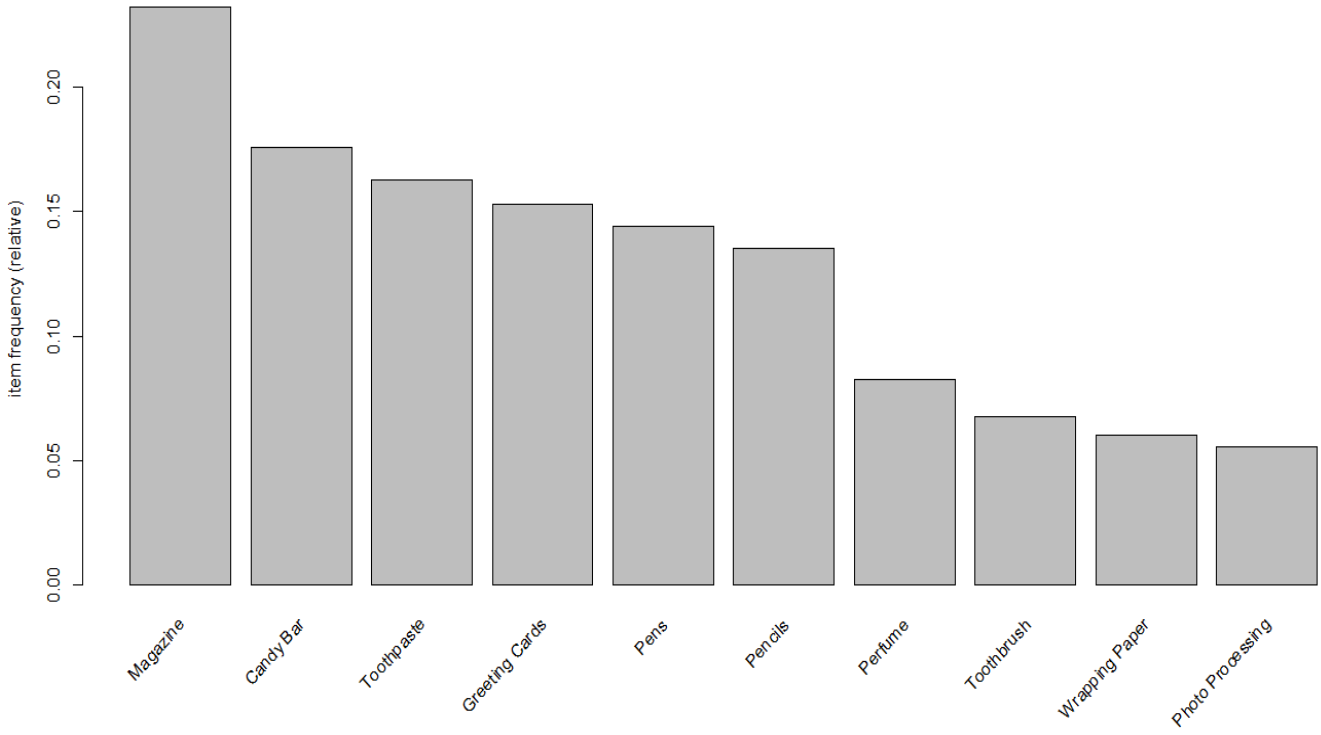
ADARSH SHRIVASTAVA

#Reading the transactions from data set

```
trans = read.transactions("C:/URI/MRA Assignment/MarketBasketAssignment_1.csv", format = "single", sep = "," , cols = c("Transaction", "Product"), rm.duplicates=TRUE)
```

#Item frequency Plot

```
itemFrequencyPlot(trans, topN=10)
```



#Get Rules from Apriori with min support & Confidence as 0.0

rules <- apriori(trans, parameter=list(supp=0.00, conf=0.0, target="rules",minlen = 2))

```
Apriori

Parameter specification:
 confidence minvalsmaxaremavaloriginalSupportmaxtime support minlenmaxlen
target   ext
          0    0.1    1 none FALSE            TRUE       5       0      2
10  rules FALSE

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

Absolute minimum support count: 0

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[17 item(s), 6726 transaction(s)] done [0.00s].
sorting and recoding items ... [17 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3 4 5 6 7 8 9 done [0.02s].
writing ... [666434 rule(s)] done [0.46s].
creating S4 object  ... done [0.38s].
```

summary(rules)

```
set of 666434 rules

rule length distribution (lhs + rhs):sizes
     2      3      4      5      6      7      8      9
   272   2040   9520  30940  74256 136136 194480 218790

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  2.000   7.000   8.000   7.687   9.000   9.000

summary of quality measures:
    support           confidence          lift
 Min.   :0.000e+00   Min.   :0.0000   Min.   :  0.000
 1st Qu.:0.000e+00   1st Qu.:1.0000   1st Qu.:  6.543
 Median :0.000e+00   Median :1.0000   Median : 16.649
 Mean   :6.010e-06   Mean   :0.9719   Mean   : 43.742
 3rd Qu.:0.000e+00   3rd Qu.:1.0000   3rd Qu.: 34.142
 Max.   :4.609e-02   Max.   :1.0000   Max.   :305.727

mining info:
  data ntransactions support confidence
 trans          6726       0          0
```

 # Filtering rules with lift >1

```
filteredrules_lift<- subset(rules, lift > 1.0)
rules <- sort(filteredrules_lift, by="lift", decreasing = TRUE)
summary(rules)
```

```
set of 650404 rules

rule length distribution (lhs + rhs):sizes
     2      3      4      5      6      7      8      9
    48   1137   6645  26179  70140 133948 193688 218619

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  2.000   7.000   8.000   7.744   9.000   9.000

summary of quality measures:
    support            confidence          lift
 Min.   :0.000e+00   Min.   :0.00366   Min.   :  1.002
 1st Qu.:0.000e+00   1st Qu.:1.00000   1st Qu.:  6.941
 Median :0.000e+00   Median :1.00000   Median : 16.649
 Mean   :4.810e-06   Mean   :0.99584   Mean   : 44.820
 3rd Qu.:0.000e+00   3rd Qu.:1.00000   3rd Qu.: 34.142
 Max.   :4.609e-02   Max.   :1.00000   Max.   :305.727

mining info:
  data ntransactions support confidence
 trans          6726       0          0
```

 # Top rules item sets with minimum support = 1%

```
filteredrules_support<- subset(rules, support > 0.01)
rules <- sort(filteredrules_support, by="support", decreasing = TRUE)
summary(rules)
```

```
set of 41 rules

rule length distribution (lhs + rhs):sizes
2  3
20 21

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    2.0     2.0     3.0     2.5     3.0     3.0

summary of quality measures:
    support        confidence        lift
 Min.   :0.010   Min.   :0.07   Min.   :1.04
 1st Qu.:0.011   1st Qu.:0.22   1st Qu.:1.44
 Median :0.017   Median :0.30   Median :1.74
 Mean   :0.021   Mean   :0.29   Mean   :1.90
 3rd Qu.:0.030   3rd Qu.:0.35   3rd Qu.:2.44
 Max.   :0.046   Max.   :0.48   Max.   :3.06

mining info:
```

```
  data ntransactions support confidence
trans          6726       0          0
```

inspect(rules)

```
lhs                                rhs                support confidence lift
[1]  {Perfume}                  => {Toothbrush}       0.017   0.207      3.1
[2]  {Toothbrush}               => {Perfume}          0.017   0.253      3.1
[3]  {Bow}                      => {Toothbrush}       0.010   0.196      2.9
[4]  {Toothbrush}               => {Bow}              0.010   0.149      2.9
[5]  {Candy Bar,Magazine}       => {Greeting Cards}   0.017   0.431      2.8
[6]  {Magazine,Pencils}         => {Greeting Cards}   0.012   0.422      2.8
[7]  {Pencils,Toothpaste}       => {Candy Bar}        0.011   0.484      2.8
[8]  {Greeting Cards,Magazine}  => {Candy Bar}        0.017   0.460      2.6
[9]  {Magazine,Toothpaste}      => {Candy Bar}        0.013   0.443      2.5
[10] {Greeting Cards,Toothpaste} => {Candy Bar}       0.015   0.438      2.5
[11] {Magazine,Toothpaste}      => {Greeting Cards}   0.011   0.373      2.4
[12] {Greeting Cards,Magazine}  => {Pencils}          0.012   0.321      2.4
[13] {Candy Bar,Toothpaste}     => {Greeting Cards}   0.015   0.351      2.3
[14] {Magazine,Pencils}         => {Candy Bar}        0.010   0.365      2.1
[15] {Candy Bar,Magazine}       => {Toothpaste}       0.013   0.331      2.0
[16] {Candy Bar,Toothpaste}     => {Pencils}          0.011   0.265      2.0
[17] {Candy Bar,Greeting Cards} => {Toothpaste}       0.015   0.316      1.9
[18] {Candy Bar,Pencils}        => {Toothpaste}       0.011   0.314      1.9
[19] {Candy Bar,Magazine}       => {Pencils}          0.010   0.260      1.9
[20] {Greeting Cards,Magazine}  => {Toothpaste}       0.011   0.298      1.8
[21] {Greeting Cards,Pencils}   => {Magazine}         0.012   0.403      1.7
[22] {Candy Bar}                => {Greeting Cards}   0.046   0.262      1.7
[23] {Greeting Cards}           => {Candy Bar}        0.046   0.302      1.7
[24] {Candy Bar,Greeting Cards} => {Magazine}         0.017   0.374      1.6
[25] {Pencils}                  => {Candy Bar}        0.035   0.260      1.5
[26] {Candy Bar}                => {Pencils}          0.035   0.200      1.5
[27] {Toothpaste}               => {Candy Bar}        0.041   0.255      1.5
[28] {Candy Bar}                => {Toothpaste}       0.041   0.236      1.5
[29] {Pencils}                  => {Greeting Cards}   0.030   0.221      1.4
[30] {Greeting Cards}           => {Pencils}          0.030   0.196      1.4
[31] {Greeting Cards,Toothpaste} => {Magazine}        0.011   0.335      1.4
[32] {Candy Bar,Toothpaste}     => {Magazine}         0.013   0.319      1.4
[33] {Greeting Cards}           => {Toothpaste}       0.033   0.218      1.3
[34] {Toothpaste}               => {Greeting Cards}   0.033   0.205      1.3
[35] {Photo Processing}         => {Magazine}         0.017   0.298      1.3
[36] {Magazine}                 => {Photo Processing} 0.017   0.071      1.3
[37] {Candy Bar,Pencils}        => {Magazine}         0.010   0.297      1.3
[38] {Greeting Cards}           => {Magazine}         0.037   0.245      1.1
[39] {Magazine}                 => {Greeting Cards}   0.037   0.162      1.1
[40] {Pencils}                  => {Toothpaste}       0.023   0.168      1.0
[41] {Toothpaste}               => {Pencils}          0.023   0.140      1.0
```

Interpretation of Terms:

Lhs: Left hand side of the rule also called antecedent

Rhs: Right hand side of the rules also called consequent

Support : Frequencies of the itemset being analyzed . For example support of 0.1 would indicate that 1 in 100 customers have the itemset (purchased the itemset)

Confidence : Estimate of the probability for the consequent RHS given the antecedent/LHS.

Lift : The ratio of confidence to expected confidence or the change in probability given the antecedent. Generally, anything over 1 indicates significant relationship

Therefore, in line 1 in the above rules, we can interpret the results as those customers with an itemset of **Perfume** and **Tooth brush** with a lift of 3.07, which means there is a potential gain if we were to target these customers with a direct sales effort.

Business Pitch : Spend time with those customers in the itemset and you increase your sales efficiency and cross sell potential . Customers with higher cross sell ratio have a tendency to have a higher retention rate as well.

 # Remove duplicate rules

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

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

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

rules.pruned<- rules[!redundant]

rules<-rules.pruned

options(digits=2)

summary(rules)
```

```
set of 17 rules

rule length distribution (lhs + rhs):sizes
2  3
10  7

   Min.   1st Qu.  Median    Mean   3rd Qu.    Max.
    2.0     2.0     2.0      2.4      3.0      3.0

summary of quality measures:
    support         confidence        lift
 Min.   :0.010   Min.   :0.17   Min.   :1.04
 1st Qu.:0.012   1st Qu.:0.22   1st Qu.:1.45
 Median :0.017   Median :0.26   Median :2.07
 Mean   :0.022   Mean   :0.31   Mean   :2.04
 3rd Qu.:0.033   3rd Qu.:0.42   3rd Qu.:2.75
 Max.   :0.046   Max.   :0.48   Max.   :3.06

mining info:
```

```
  data ntransactions support confidence
trans          6726       0          0
```

inspect(rules)

```
lhsrhs                  support confidence lift
[1]  {Perfume}                  => {Toothbrush}     0.017   0.21       3.1
[2]  {Bow}                      => {Toothbrush}     0.010   0.20       2.9
[3]  {Candy Bar,Magazine}       => {Greeting Cards} 0.017   0.43       2.8
[4]  {Magazine,Pencils}         => {Greeting Cards} 0.012   0.42       2.8
[5]  {Pencils,Toothpaste}       => {Candy Bar}      0.011   0.48       2.8
[6]  {Magazine,Toothpaste}      => {Candy Bar}      0.013   0.44       2.5
[7]  {Greeting Cards,Toothpaste} => {Candy Bar}     0.015   0.44       2.5
[8]  {Magazine,Toothpaste}      => {Greeting Cards} 0.011   0.37       2.4
[9]  {Magazine,Pencils}         => {Candy Bar}      0.010   0.36       2.1
[10] {Candy Bar}                => {Greeting Cards} 0.046   0.26       1.7
[11] {Pencils}                  => {Candy Bar}      0.035   0.26       1.5
[12] {Toothpaste}               => {Candy Bar}      0.041   0.26       1.5
[13] {Pencils}                  => {Greeting Cards} 0.030   0.22       1.4
[14] {Greeting Cards}           => {Toothpaste}     0.033   0.22       1.3
[15] {Photo Processing}         => {Magazine}       0.017   0.30       1.3
[16] {Greeting Cards}           => {Magazine}       0.037   0.25       1.1
[17] {Pencils}                  => {Toothpaste}     0.023   0.17       1.0
```

 # Taking to top Rules with support > 1% & Lift > 1

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

rules <- rules[1:10]

inspect(rules)

```
lhsrhs                  support confidence lift
[1]  {Perfume}                  => {Toothbrush}     0.017   0.21       3.1
[2]  {Bow}                      => {Toothbrush}     0.010   0.20       2.9
[3]  {Candy Bar,Magazine}       => {Greeting Cards} 0.017   0.43       2.8
[4]  {Magazine,Pencils}         => {Greeting Cards} 0.012   0.42       2.8
[5]  {Pencils,Toothpaste}       => {Candy Bar}      0.011   0.48       2.8
[6]  {Magazine,Toothpaste}      => {Candy Bar}      0.013   0.44       2.5
[7]  {Greeting Cards,Toothpaste} => {Candy Bar}     0.015   0.44       2.5
[8]  {Magazine,Toothpaste}      => {Greeting Cards} 0.011   0.37       2.4
[9]  {Magazine,Pencils}         => {Candy Bar}      0.010   0.36       2.1
[10] {Candy Bar}                => {Greeting Cards} 0.046   0.26       1.7
```

Here we have increase the minimum threshold of support and confidence 0 to 1, this will filter out the rules whose confidence or support is less than 0.1. Also the rules are shown in decreasing order by lift, top 10 rules are shown above.

#Top 10 rules items sets from the above results with min confidence = 10%

```
filteredrules_confidence<- subset(rules, confidence > 0.1)

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

options(digits = 2)

inspect(rules)
```

```
lhs                                rhs              support confidence lift
[1]  {Perfume}                  => {Toothbrush}     0.017   0.21       3.1
[2]  {Bow}                      => {Toothbrush}     0.010   0.20       2.9
[3]  {Candy Bar,Magazine}       => {Greeting Cards} 0.017   0.43       2.8
[4]  {Magazine,Pencils}         => {Greeting Cards} 0.012   0.42       2.8
[5]  {Pencils,Toothpaste}       => {Candy Bar}      0.011   0.48       2.8
[6]  {Magazine,Toothpaste}      => {Candy Bar}      0.013   0.44       2.5
[7]  {Greeting Cards,Toothpaste} => {Candy Bar}     0.015   0.44       2.5
[8]  {Magazine,Toothpaste}      => {Greeting Cards} 0.011   0.37       2.4
[9]  {Magazine,Pencils}         => {Candy Bar}      0.010   0.36       2.1
[10] {Candy Bar}                => {Greeting Cards} 0.046   0.26       1.7
```

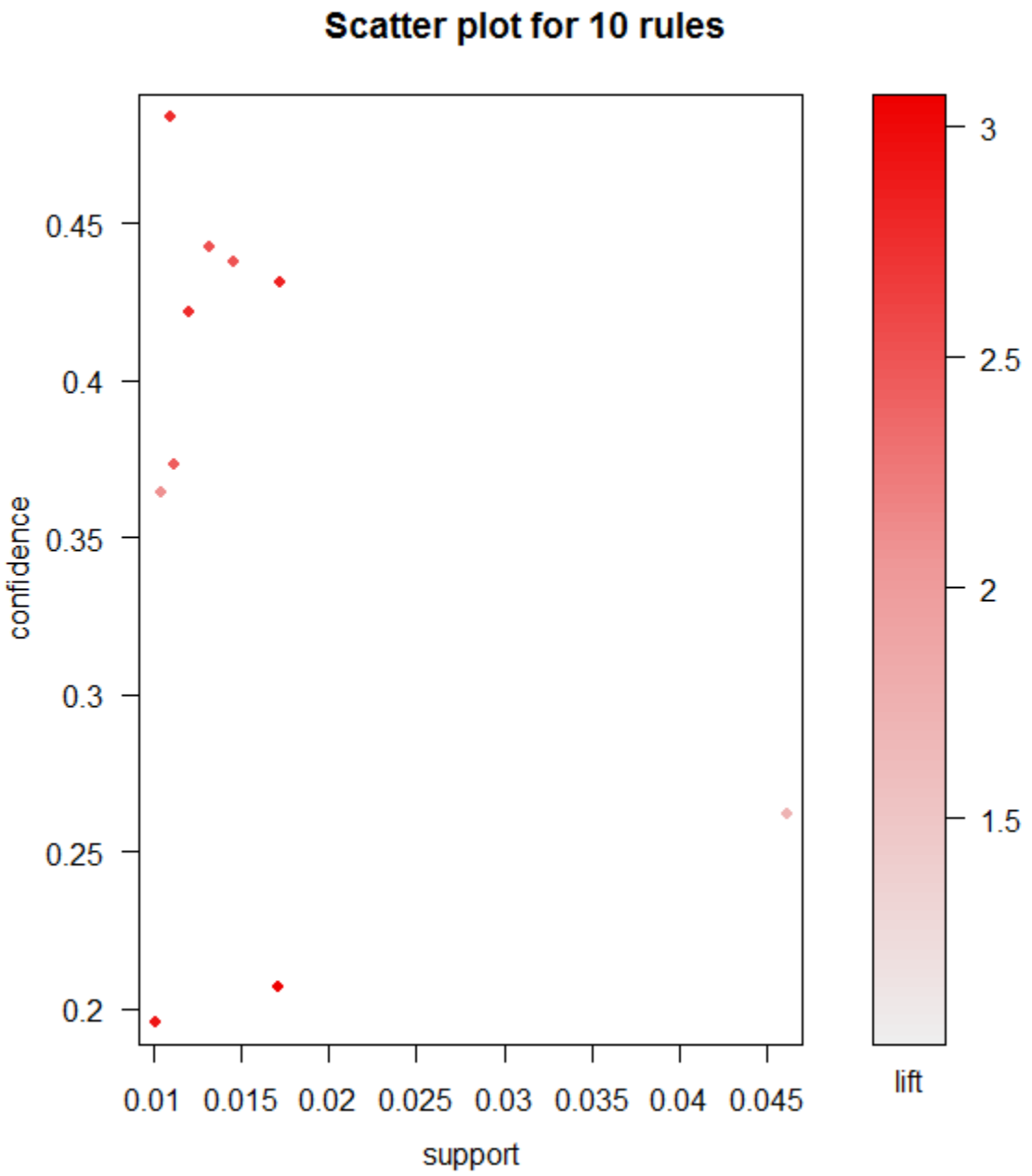
NOTE : These are the 10 rules denoting the items that are frequently bought together and can be used for marketing promotions or for user accessibility be placed together in the retail store. The marketing productivity will be increased in factors of the lift.

summary(rules)

```
set of 10 rules

rule length distribution (lhs + rhs):sizes
2 3
3 7

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    2.0     2.2     3.0     2.7     3.0     3.0

summary of quality measures:
    support         confidence        lift
 Min.   :0.010   Min.   :0.20   Min.   :1.72
 1st Qu.:0.011   1st Qu.:0.29   1st Qu.:2.45
 Median :0.013   Median :0.40   Median :2.64
 Mean   :0.016   Mean   :0.36   Mean   :2.55
 3rd Qu.:0.016   3rd Qu.:0.44   3rd Qu.:2.81
 Max.   :0.046   Max.   :0.48   Max.   :3.06

mining info:
  data ntransactions support confidence
 trans          6726       0          0
```

 # Plot the Rules

plot(rules)



Explanation of the chart: The plot shows all the association rules confidence vs. support vs. lift. The more closer the point is to upper right corner more is the respective support and confidence for that particular rule. Lift is demonstrated by the intensity of Red color of the point.

Using the analysis to drive business decision-making :

The output of the analysis reflects how frequently items co-occur in transactions. This is dependent upon the strength of association between the items, and way the retailers of the shop kept there in aisle. The final recommendation is that we can encourage cross selling above mentioned 10 rules by encouraging customers by giving discount or put in aisle in near to each other