Market Basket Analysis

Uche Kalu

4/8/2022

Market Basket Analysis is also called **Association Rules**.

It is used to uncover links between items by large retailers. It works by searching for combinations of items that happen in transactions together

It uses such info to recognize customer purchasing patterns, to identify who customers are, to understand why you buy certain items and to know which products are purchased jointly so as to know which products to promote

Using Market Basket Analysis, we want to find out which consumer items from this hypothetical business dataset are bought together by customers

```
mydata <- read.csv("Cosmetics.csv", header = T, colClasses = "factor")
```

```


str(mydata)
```

Lets access our business data

```
## 'data.frame':    1000 obs. of  14 variables:
##  $ Bag            : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 2 ...
##  $ Blush          : Factor w/ 2 levels "No","Yes": 2 1 2 1 2 1 2 1 1 2 ...
##  $ Nail.Polish    : Factor w/ 2 levels "No","Yes": 2 2 1 2 1 1 2 2 1 2 ...
##  $ Brushes        : Factor w/ 2 levels "No","Yes": 2 1 1 2 1 1 2 2 1 2 ...
##  $ Concealer      : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 2 2 1 2 1 ...
##  $ Eyebrow.Pencils: Factor w/ 2 levels "No","Yes": 1 1 2 1 1 1 1 1 1 1 ...
##  $ Bronzer        : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 1 2 2 1 1 ...
##  $ Lip.liner      : Factor w/ 2 levels "No","Yes": 2 2 2 1 2 1 2 1 1 1 ...
##  $ Mascara        : Factor w/ 2 levels "No","Yes": 2 1 2 1 2 1 2 2 1 2 ...
##  $ Eye.shadow     : Factor w/ 2 levels "No","Yes": 1 1 2 1 2 1 2 2 1 2 ...
##  $ Foundation     : Factor w/ 2 levels "No","Yes": 1 2 2 2 1 1 2 1 2 1 ...
##  $ Lip.Gloss      : Factor w/ 2 levels "No","Yes": 1 2 2 1 2 1 2 2 1 1 ...
##  $ Lipstick       : Factor w/ 2 levels "No","Yes": 1 1 2 1 2 1 1 1 2 1 ...
##  $ Eyeliner       : Factor w/ 2 levels "No","Yes": 2 1 1 2 1 2 1 1 1 1 ...
```

```
summary(mydata)
```

```
##   Bag      Blush     Nail.Polish Brushes   Concealer Eyebrow.Pencils Bronzer
##  No :946   No :637   No :720     No :851   No :558   No :958         No :721
##  Yes: 54   Yes:363   Yes:280     Yes:149   Yes:442   Yes: 42         Yes:279
##  Lip.liner Mascara   Eye.shadow Foundation Lip.Gloss Lipstick  Eyeliner
##  No :766   No :643   No :619    No :464    No :510   No :678   No :543
##  Yes:234   Yes:357   Yes:381    Yes:536    Yes:490   Yes:322   Yes:457
```

```
yrdata <- apriori(mydata)
```

This is the function used to create Market Basket Analysis

```
## Warning in apriori(mydata): Mining stopped (maxlen reached). Only patterns up to
## a length of 10 returned!
```

The number of rules per item is displayed - 3 rules with 1 item, 85 rules in 2 items, 10739 rules with 5 items. also displays total number of rules- 68880 rules (which is to much)

```
summary(yrdata)
```

```
## set of 68880 rules
##
## rule length distribution (lhs + rhs):sizes
##     1     2     3     4     5     6     7     8     9    10
##     3    85   942  4350 10739 17062 18066 11996  4665   972
##
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   6.000   7.000   6.542   8.000  10.000
##
## summary of quality measures:
##     support          confidence        coverage           lift
##  Min.   :0.1000   Min.   :0.8000   Min.   :0.1000   Min.   :0.8781
##  1st Qu.:0.1150   1st Qu.:0.8667   1st Qu.:0.1250   1st Qu.:1.0389
##  Median :0.1370   Median :0.9453   Median :0.1490   Median :1.1565
##  Mean   :0.1583   Mean   :0.9259   Mean   :0.1718   Mean   :1.2019
##  3rd Qu.:0.1770   3rd Qu.:0.9821   3rd Qu.:0.1930   3rd Qu.:1.2438
##  Max.   :0.9580   Max.   :1.0000   Max.   :1.0000   Max.   :3.5714
##      count
##  Min.   :100.0
##  1st Qu.:115.0
##  Median :137.0
##  Mean   :158.3
##  3rd Qu.:177.0
##  Max.   :958.0
##
## mining info:
##    data ntransactions support confidence                   call
##  mydata          1000     0.1        0.8 apriori(data = mydata)
```

Lets reduce the number of Rules by using only with specified parameter values *Here we display d rules of all columns which are now values in lhs column(left hand side) which is 'IF' and rhs column(right hand side) which is 'Then' but they are all = NO (i.e which items was not bought - If Nail.Polish was not bought in lhs, Then Brushes was not bought in rhs). Since we want to see which item was bought, dis rule is not helpful.*

```
inspect(myrules)
```

```
##      lhs                                   rhs                  support
## [1]  {Nail.Polish=No}                   => {Brushes=No}         0.720
## [2]  {Brushes=No}                       => {Nail.Polish=No}     0.720
## [3]  {Lip.liner=No}                     => {Bag=No}             0.732
## [4]  {Lip.liner=No}                     => {Eyebrow.Pencils=No} 0.734
## [5]  {Brushes=No}                       => {Bag=No}             0.817
## [6]  {Bag=No}                           => {Brushes=No}         0.817
## [7]  {Brushes=No}                       => {Eyebrow.Pencils=No} 0.820
## [8]  {Eyebrow.Pencils=No}               => {Brushes=No}         0.820
## [9]  {Bag=No}                           => {Eyebrow.Pencils=No} 0.909
## [10] {Eyebrow.Pencils=No}               => {Bag=No}             0.909
## [11] {Bag=No, Lip.liner=No}             => {Eyebrow.Pencils=No} 0.703
## [12] {Eyebrow.Pencils=No, Lip.liner=No} => {Bag=No}             0.703
## [13] {Bag=No, Brushes=No}               => {Eyebrow.Pencils=No} 0.789
## [14] {Brushes=No, Eyebrow.Pencils=No}   => {Bag=No}             0.789
## [15] {Bag=No, Eyebrow.Pencils=No}       => {Brushes=No}         0.789
##      confidence coverage lift     count
## [1]  1.0000000  0.720    1.175088 720
## [2]  0.8460635  0.851    1.175088 720
## [3]  0.9556136  0.766    1.010162 732
## [4]  0.9582245  0.766    1.000234 734
## [5]  0.9600470  0.851    1.014849 817
## [6]  0.8636364  0.946    1.014849 817
## [7]  0.9635723  0.851    1.005817 820
## [8]  0.8559499  0.958    1.005817 820
## [9]  0.9608879  0.946    1.003015 909
## [10] 0.9488518  0.958    1.003015 909
## [11] 0.9603825  0.732    1.002487 703
## [12] 0.9577657  0.734    1.012437 703
## [13] 0.9657283  0.817    1.008067 789
## [14] 0.9621951  0.820    1.017120 789
## [15] 0.8679868  0.909    1.019961 789
```

Finding interesting rules- 1 (i.e rules that will display Yes in the lhs and rhs columns) *We see that Foundation column has highest values of Yes among all columns(as such it is d most popular item bought) and Eyebrow.Pencils has highest value for NO*

```
summary(mydata)
```

```
##   Bag      Blush     Nail.Polish Brushes   Concealer Eyebrow.Pencils Bronzer
##  No :946   No :637   No :720     No :851   No :558   No :958         No :721
##  Yes: 54   Yes:363   Yes:280     Yes:149   Yes:442   Yes: 42         Yes:279
##  Lip.liner Mascara   Eye.shadow Foundation Lip.Gloss Lipstick  Eyeliner
##  No :766   No :643   No :619    No :464   No :510   No :678   No :543
##  Yes:234   Yes:357   Yes:381    Yes:536   Yes:490   Yes:322   Yes:457
```

```
myrules <- apriori(mydata, parameter = list(minlen = 2, maxlen = 3,
                                            conf = 0.7),
                   appearance = list(rhs = c("Foundation=Yes"),
                                     default = "lhs"))
```

```
## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support minlen
##         0.7    0.1    1 none FALSE            TRUE       5     0.1      2
##  maxlen target  ext
##       3  rules TRUE
##
## Algorithmic control:
##  filter tree heap memopt load sort verbose
##     0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 100
##
## set item appearances ...[1 item(s)] done [0.00s].
## set transactions ...[28 item(s), 1000 transaction(s)] done [0.00s].
## sorting and recoding items ... [26 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 done [0.01s].
## writing ... [16 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].
```

This displays rules with Yes values(ie If Lip.Gloss was bought in lhs, Then Foundation was also bought in rhs) in lhs and rhs columns but there are still rules with No values displayed in lhs so dis rules is not of interest to us

```
inspect(myrules)
```

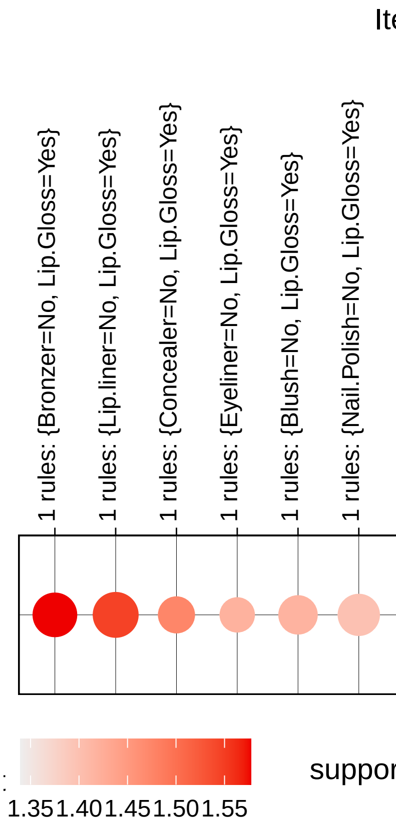
```
##      lhs                                   rhs              support confidence
## [1]  {Lip.Gloss=Yes}                    => {Foundation=Yes} 0.356   0.7265306
## [2]  {Lip.Gloss=Yes, Lipstick=Yes}      => {Foundation=Yes} 0.116   0.7341772
## [3]  {Mascara=Yes, Lip.Gloss=Yes}       => {Foundation=Yes} 0.130   0.7182320
## [4]  {Eye.shadow=Yes, Lip.Gloss=Yes}    => {Foundation=Yes} 0.146   0.7263682
## [5]  {Lip.Gloss=Yes, Eyeliner=No}       => {Foundation=Yes} 0.200   0.7604563
## [6]  {Concealer=No, Lip.Gloss=Yes}      => {Foundation=Yes} 0.215   0.7904412
## [7]  {Eye.shadow=No, Lip.Gloss=Yes}     => {Foundation=Yes} 0.210   0.7266436
## [8]  {Blush=No, Lip.Gloss=Yes}          => {Foundation=Yes} 0.237   0.7596154
## [9]  {Mascara=No, Lip.Gloss=Yes}        => {Foundation=Yes} 0.226   0.7313916
## [10] {Lip.Gloss=Yes, Lipstick=No}       => {Foundation=Yes} 0.240   0.7228916
## [11] {Nail.Polish=No, Lip.Gloss=Yes}    => {Foundation=Yes} 0.267   0.7500000
## [12] {Bronzer=No, Lip.Gloss=Yes}        => {Foundation=Yes} 0.295   0.8452722
## [13] {Lip.liner=No, Lip.Gloss=Yes}      => {Foundation=Yes} 0.310   0.8288770
## [14] {Brushes=No, Lip.Gloss=Yes}        => {Foundation=Yes} 0.313   0.7417062
## [15] {Bag=No, Lip.Gloss=Yes}            => {Foundation=Yes} 0.335   0.7282609
## [16] {Eyebrow.Pencils=No, Lip.Gloss=Yes} => {Foundation=Yes} 0.345   0.7278481
##      coverage lift     count
## [1]  0.490    1.355468 356
## [2]  0.158    1.369734 116
## [3]  0.181    1.339985 130
## [4]  0.201    1.355164 146
## [5]  0.263    1.418762 200
## [6]  0.272    1.474704 215
## [7]  0.289    1.355678 210
## [8]  0.312    1.417193 237
## [9]  0.309    1.364537 226
## [10] 0.332    1.348678 240
## [11] 0.356    1.399254 267
## [12] 0.349    1.577000 295
## [13] 0.374    1.546412 310
## [14] 0.422    1.383780 313
## [15] 0.460    1.358696 335
## [16] 0.474    1.357926 345
```

Graphs and Charts

```
library(arulesViz)
```

```
## Warning: package 'arulesViz' was built under R version 4.1.2
```

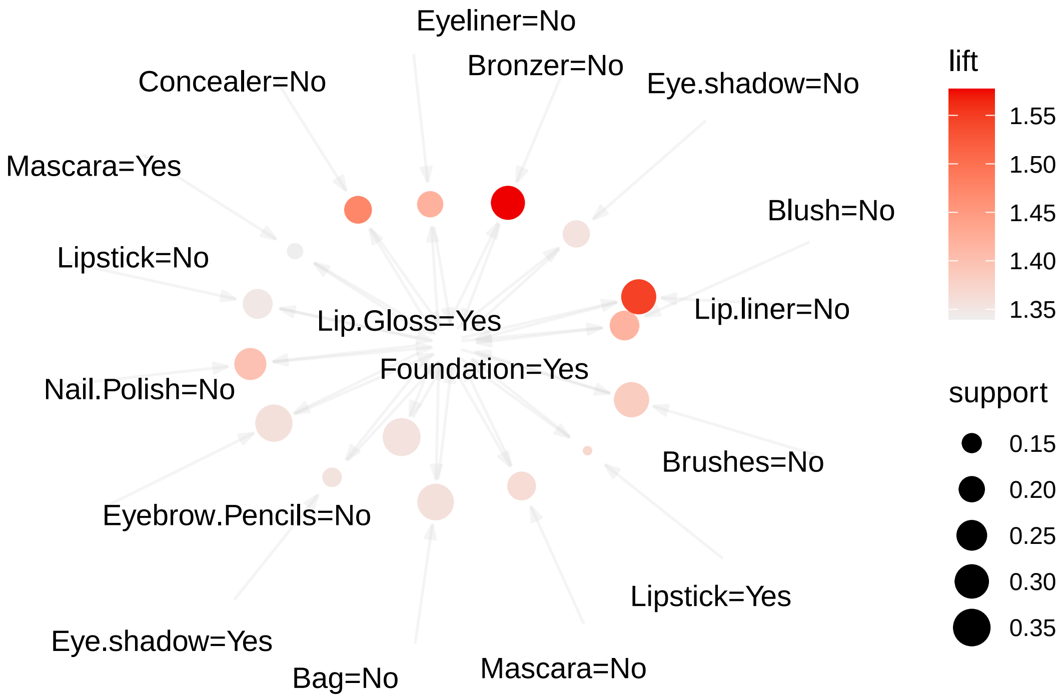
```
plot(myrules, method = "grouped")
```



Displaying scatterplot chart of d rules using confidence, support and lift values

```
plot(myrules, method = "graph", control = list(type = "items"))
```

```
## Available control parameters (with default values):
## layout    =  stress
## circular  =  FALSE
## ggraphdots    =  NULL
## edges     =  <environment>
## nodes     =  <environment>
## nodetext  =  <environment>
## colors    =  c("#EE0000FF", "#EEEEEEFF")
## engine    =  ggplot2
## max   =  100
## verbose   =  FALSE
```



Finding interesting rules - 2 (ie displaying only rules with Yes values in lhs and rhs columns)
Lets list all the rules in lhs and rhs columns that have value- Yes

```
myrules <- apriori(mydata, parameter = list(minlen = 2, maxlen = 3,
                                           conf = 0.5),
                   appearance = list(rhs = c("Foundation=Yes"), lhs = c("Bag=Yes", "Blush=Yes", "Nail.Po
```

```
## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support minlen
##         0.5    0.1    1 none FALSE            TRUE       5     0.1      2
##  maxlen target  ext
##       3  rules TRUE
##
## Algorithmic control:
##  filter tree heap memopt load sort verbose
##     0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 100
##
## set item appearances ...[14 item(s)] done [0.00s].
## set transactions ...[14 item(s), 1000 transaction(s)] done [0.00s].
## sorting and recoding items ... [12 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 done [0.00s].
## writing ... [21 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].
```

This displays only rules with Yes values in lhs and rhs columns (ie If people bought Lipstick Then they also bought Foundation etc)

```
inspect(myrules)
```

```
##      lhs                                rhs              support confidence
## [1]  {Lipstick=Yes}                  => {Foundation=Yes} 0.167   0.519
## [2]  {Nail.Polish=Yes}               => {Foundation=Yes} 0.143   0.511
## [3]  {Blush=Yes}                     => {Foundation=Yes} 0.192   0.529
## [4]  {Mascara=Yes}                   => {Foundation=Yes} 0.192   0.538
## [5]  {Eye.shadow=Yes}                => {Foundation=Yes} 0.211   0.554
## [6]  {Eyeliner=Yes}                  => {Foundation=Yes} 0.238   0.521
## [7]  {Lip.Gloss=Yes}                 => {Foundation=Yes} 0.356   0.727
## [8]  {Concealer=Yes}                 => {Foundation=Yes} 0.231   0.523
## [9]  {Lip.Gloss=Yes, Lipstick=Yes}   => {Foundation=Yes} 0.116   0.734
## [10] {Blush=Yes, Mascara=Yes}        => {Foundation=Yes} 0.101   0.549
## [11] {Blush=Yes, Eye.shadow=Yes}     => {Foundation=Yes} 0.100   0.549
## [12] {Blush=Yes, Lip.Gloss=Yes}      => {Foundation=Yes} 0.119   0.669
## [13] {Blush=Yes, Concealer=Yes}      => {Foundation=Yes} 0.115   0.523
## [14] {Mascara=Yes, Eye.shadow=Yes}   => {Foundation=Yes} 0.166   0.517
## [15] {Mascara=Yes, Lip.Gloss=Yes}    => {Foundation=Yes} 0.130   0.718
## [16] {Concealer=Yes, Mascara=Yes}    => {Foundation=Yes} 0.107   0.525
## [17] {Eye.shadow=Yes, Lip.Gloss=Yes} => {Foundation=Yes} 0.146   0.726
## [18] {Concealer=Yes, Eye.shadow=Yes} => {Foundation=Yes} 0.104   0.517
## [19] {Lip.Gloss=Yes, Eyeliner=Yes}   => {Foundation=Yes} 0.156   0.687
## [20] {Concealer=Yes, Eyeliner=Yes}   => {Foundation=Yes} 0.152   0.512
## [21] {Concealer=Yes, Lip.Gloss=Yes}  => {Foundation=Yes} 0.141   0.647
##      coverage lift  count
## [1]  0.322    0.968 167
## [2]  0.280    0.953 143
## [3]  0.363    0.987 192
## [4]  0.357    1.003 192
## [5]  0.381    1.033 211
## [6]  0.457    0.972 238
## [7]  0.490    1.355 356
## [8]  0.442    0.975 231
## [9]  0.158    1.370 116
## [10] 0.184    1.024 101
## [11] 0.182    1.025 100
## [12] 0.178    1.247 119
## [13] 0.220    0.975 115
## [14] 0.321    0.965 166
## [15] 0.181    1.340 130
## [16] 0.204    0.979 107
## [17] 0.201    1.355 146
## [18] 0.201    0.965 104
## [19] 0.227    1.282 156
## [20] 0.297    0.955 152
## [21] 0.218    1.207 141
```

```
which(myredun)
```

Finding and Displaying Redundant rules

```
##                 {Foundation=Yes,Lipstick=Yes}
##                                             1
##              {Nail.Polish=Yes,Foundation=Yes}
##                                             2
##                    {Blush=Yes,Foundation=Yes}
##                                             3
##                  {Mascara=Yes,Foundation=Yes}
##                                             4
##               {Eye.shadow=Yes,Foundation=Yes}
##                                             5
##                 {Foundation=Yes,Eyeliner=Yes}
##                                             6
##                {Foundation=Yes,Lip.Gloss=Yes}
##                                             7
##                {Concealer=Yes,Foundation=Yes}
##                                             8
##   {Foundation=Yes,Lip.Gloss=Yes,Lipstick=Yes}
##                                             9
##        {Blush=Yes,Mascara=Yes,Foundation=Yes}
##                                            10
##     {Blush=Yes,Eye.shadow=Yes,Foundation=Yes}
##                                            11
##      {Blush=Yes,Foundation=Yes,Lip.Gloss=Yes}
##                                            12
##      {Blush=Yes,Concealer=Yes,Foundation=Yes}
##                                            13
##   {Mascara=Yes,Eye.shadow=Yes,Foundation=Yes}
##                                            14
##    {Mascara=Yes,Foundation=Yes,Lip.Gloss=Yes}
##                                            15
##    {Concealer=Yes,Mascara=Yes,Foundation=Yes}
##                                            16
## {Eye.shadow=Yes,Foundation=Yes,Lip.Gloss=Yes}
##                                             17
## {Concealer=Yes,Eye.shadow=Yes,Foundation=Yes}
##                                             18
##   {Foundation=Yes,Lip.Gloss=Yes,Eyeliner=Yes}
##                                             19
##   {Concealer=Yes,Foundation=Yes,Eyeliner=Yes}
##                                             20
##  {Concealer=Yes,Foundation=Yes,Lip.Gloss=Yes}
##                                             21
```

```r
yrrules <- myrules[!myredun]  # removes d redundant rules

yrrules <- sort(yrrules, by = "lift")
```

```r
inspect(yrrules)
```

Removing Redundant rules