# Assignment 3

## *Market Basket Analysis - Groceries Data*

Mohamed Megahed

4/18/2020

```r
# Libraries Loading
library(dplyr)

library(arules)

library(arulesViz)

# Loading Dataset into a sparse matrix
groceries <- read.transactions("C:/E/MSBA/Spring 2020/Adv Data Mining/Assignmen
t 3/groceries_v2.csv", sep = ",")

summary(groceries)

## transactions as itemMatrix in sparse format with
##  9834 rows (elements/itemsets/transactions) and
##  169 columns (items) and a density of 0.0260911
##
## most frequent items:
##       whole milk other vegetables       rolls/buns              soda
##             2513             1902             1809             1715
##           yogurt          (Other)
##             1372            34051
##
## element (itemset/transaction) length distribution:
## sizes
##    1    2    3    4    5    6    7    8    9   10   11   12   13   14   15
16
## 2159 1643 1299 1005  854  645  545  438  350  246  182  117   78   77   55
46
##   17   18   19   20   21   22   23   24   26   27   28   29   32
##   29   14   14    9   11    4    6    1    1    1    1    3    1
##
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   2.000   3.000   4.409   6.000  32.000
##
## includes extended item information - examples:
##             labels
## 1 abrasive cleaner
## 2 artif. sweetener
## 3   baby cosmetics
```
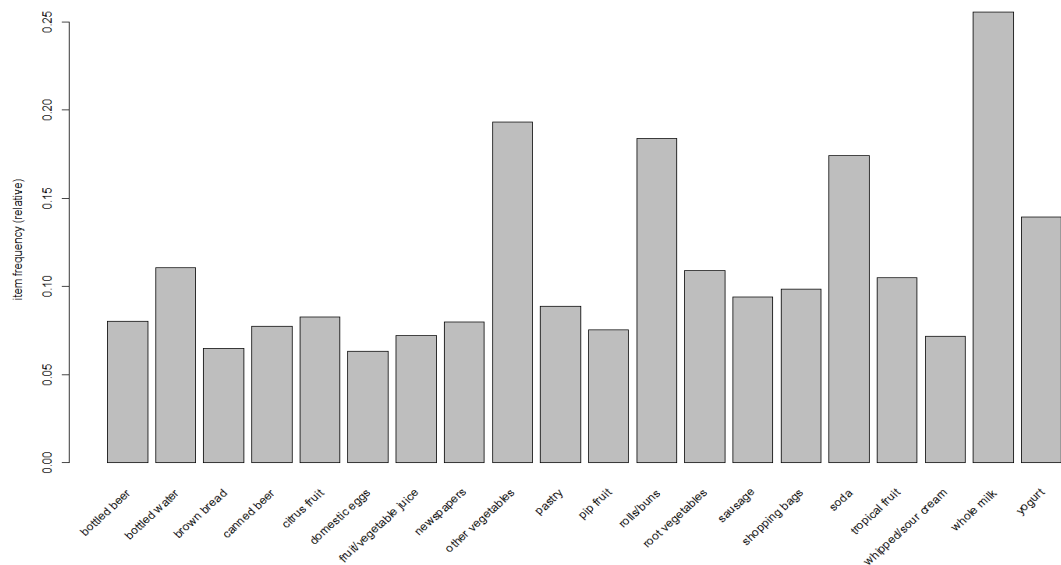
## Examine the frequency of items
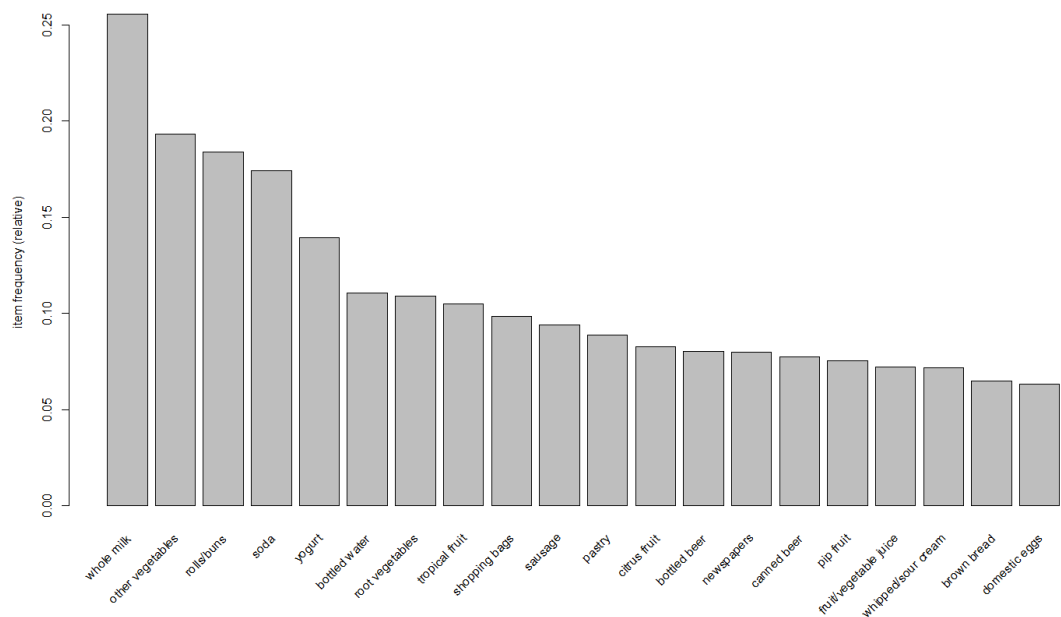
```
itemFrequency(groceries[, 1:5])
```

```
## abrasive cleaner artif. sweetener    baby cosmetics         baby food
##      0.0035590807     0.0032540167      0.0006101281      0.0001016880
##             bags
##      0.0004067521
```

## Plot the frequency of items

```
itemFrequencyPlot(groceries, support = 0.06)
```



```
itemFrequencyPlot(groceries, topN = 20)
```

## Set better support and confidence levels to learn more rules

```
Association_rules <- apriori(groceries, parameter = list(support = 0.01, confid
ence = 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: 98
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[169 item(s), 9834 transaction(s)] done [0.00s].
## sorting and recoding items ... [88 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 done [0.00s].
## writing ... [15 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].
Association_rules
## set of 15 rules

summary(Association_rules)

## set of 15 rules
##
## rule length distribution (lhs + rhs):sizes
##  3
## 15
##
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       3       3       3       3       3       3
## summary of quality measures:
##     support           confidence           lift            count
##  Min.   :0.01007   Min.   :0.5000   Min.   :1.984   Min.   : 99.0
##  1st Qu.:0.01174   1st Qu.:0.5151   1st Qu.:2.036   1st Qu.:115.5
##  Median :0.01230   Median :0.5245   Median :2.203   Median :121.0
##  Mean   :0.01316   Mean   :0.5411   Mean   :2.300   Mean   :129.4
##  3rd Qu.:0.01403   3rd Qu.:0.5718   3rd Qu.:2.432   3rd Qu.:138.0
##  Max.   :0.02227   Max.   :0.5862   Max.   :3.031   Max.   :219.0
## mining info:
##       data ntransactions support confidence
##  groceries          9834    0.01        0.5
```

In the final section of the summary() output, we receive mining information, telling us about how the rules were chosen. Here, we see that the groceries data, which contained 9,834 transactions, was used to construct rules with a minimum support of 0.01 and minimum confidence of 0.5.

## Look at the first five rules

```
inspect(Association_rules[1:5])
```

```
##     lhs                                    rhs            support    confide
nce
## [1] {curd,yogurt}                       => {whole milk} 0.01006711 0.58235
29
## [2] {butter,other vegetables}           => {whole milk} 0.01149075 0.57360
41
## [3] {domestic eggs,other vegetables}    => {whole milk} 0.01230425 0.55251
14
## [4] {whipped/sour cream,yogurt}         => {whole milk} 0.01088062 0.52450
98
## [5] {other vegetables,whipped/sour cream} => {whole milk} 0.01464308 0.50704
23
##     lift     count
## [1] 2.278893  99
## [2] 2.244657 113
## [3] 2.162116 121
## [4] 2.052539 107
## [5] 1.984184 144
```

*The first rule can be read in plain language as, "if a customer buys curd,yogurt, they will also buy whole milk." With support of 0.010 and confidence of 0.582.*

## Sorting the rules based on confidence:

```
Association_rules_Sorted <- sort(Association_rules, by='confidence', decreasing
= TRUE)
summary(Association_rules_Sorted)
```

```
## set of 15 rules
##
## rule length distribution (lhs + rhs):sizes
##  3
## 15
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       3       3       3       3       3       3
##
## summary of quality measures:
##     support          confidence          lift          count
##  Min.   :0.01007   Min.   :0.5000   Min.   :1.984   Min.   : 99.0
##  1st Qu.:0.01174   1st Qu.:0.5151   1st Qu.:2.036   1st Qu.:115.5
##  Median :0.01230   Median :0.5245   Median :2.203   Median :121.0
##  Mean   :0.01316   Mean   :0.5411   Mean   :2.300   Mean   :129.4
##  3rd Qu.:0.01403   3rd Qu.:0.5718   3rd Qu.:2.432   3rd Qu.:138.0
##  Max.   :0.02227   Max.   :0.5862   Max.   :3.031   Max.   :219.0
## mining info:
##      data ntransactions support confidence
##  groceries          9834    0.01        0.5
```

## The Association rules with there support and confidence.

```
inspect(Association_rules_Sorted)
```

```
##       lhs                                            rhs                  support
## [1]  {citrus fruit,root vegetables}             => {other vegetables} 0.01037218
## [2]  {root vegetables,tropical fruit}           => {other vegetables} 0.01230425
## [3]  {curd,yogurt}                              => {whole milk}       0.01006711
## [4]  {butter,other vegetables}                  => {whole milk}       0.01149075
## [5]  {root vegetables,tropical fruit}           => {whole milk}       0.01199919
## [6]  {root vegetables,yogurt}                   => {whole milk}       0.01454139
## [7]  {domestic eggs,other vegetables}           => {whole milk}       0.01230425
## [8]  {whipped/sour cream,yogurt}                => {whole milk}       0.01088062
## [9]  {rolls/buns,root vegetables}               => {whole milk}       0.01271100
## [10] {other vegetables,pip fruit}               => {whole milk}       0.01352451
## [11] {tropical fruit,yogurt}                    => {whole milk}       0.01515152
## [12] {other vegetables,yogurt}                  => {whole milk}       0.02226968
## [13] {other vegetables,whipped/sour cream}      => {whole milk}       0.01464308
## [14] {rolls/buns,root vegetables}               => {other vegetables} 0.01220256
## [15] {root vegetables,yogurt}                   => {other vegetables} 0.01291438
##      confidence lift     count
## [1]  0.5862069  3.030893 102
## [2]  0.5845411  3.022280 121
## [3]  0.5823529  2.278893  99
## [4]  0.5736041  2.244657 113
## [5]  0.5700483  2.230742 118
## [6]  0.5629921  2.203130 143
## [7]  0.5525114  2.162116 121
## [8]  0.5245098  2.052539 107
## [9]  0.5230126  2.046679 125
## [10] 0.5175097  2.025146 133
## [11] 0.5173611  2.024564 149
## [12] 0.5128806  2.007030 219
## [13] 0.5070423  1.984184 144
## [14] 0.5020921  2.595990 120
## [15] 0.5000000  2.585174 127
```

### Sorting Association rules by lift, Support, Confidence

```
inspect(sort(Association_rules, by = "lift")[1:5])
```

```
##      lhs                                  rhs                  support
## [1] {citrus fruit,root vegetables}   => {other vegetables} 0.01037218
## [2] {root vegetables,tropical fruit} => {other vegetables} 0.01230425
## [3] {rolls/buns,root vegetables}     => {other vegetables} 0.01220256
## [4] {root vegetables,yogurt}         => {other vegetables} 0.01291438
## [5] {curd,yogurt}                    => {whole milk}       0.01006711
##      confidence lift     count
## [1] 0.5862069  3.030893 102
## [2] 0.5845411  3.022280 121
## [3] 0.5020921  2.595990 120
## [4] 0.5000000  2.585174 127
## [5] 0.5823529  2.278893  99
```

*These rules are more interesting than that we looked at previously. The first rule, with a lift of about 3.030, implies that people who buy citrus fruit,root vegetables are nearly three times more likely to buy other vegetables than the typical customers.*