

CarrefourAssociationAnalysis

Joy Machuka

9/11/2021

Problem Definition

As a Data analyst at Carrefour Kenya working for the marketing department Carrefour you are to come up with relevant marketing strategies that will result in the higher no. of sales. Amongst the processes is checking for associations in the data.

Specifying the Question

Identify relationships between the Group of Items

Metric of Success

Identify the relationships between the Groups of Items provided

Context

Working on the Carrefour Kenya Sales data to identify the most relevant marketing strategies by identifying related groups of Items.

Experimental Design

Load the Data Check the Data Implement the Solution Challenge the Solution

Data Sourcing

```
library(arules)
```

```
## Loading required package: Matrix
```

```
##
```

```
## Attaching package: 'arules'
```

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

```
# Data to find associations
```

```
association <- read.transactions("http://bit.ly/SupermarketDatasetII")
```

```
## Warning in asMethod(object): removing duplicated items in transactions
```

```
head(association)
```

```
## transactions in sparse format with  
## 6 transactions (rows) and  
## 5729 items (columns)
```

```
dim(association)
```

```
## [1] 7501 5729
```

```
class(association)
```

```
## [1] "transactions"  
## attr(,"package")  
## [1] "arules"
```

```
#check first 10 items
```

```
inspect(association[1:10])
```

```
## items  
## [1] {cheese,energy,  
## drink,tomato,  
## fat,  
## flour,yams,cottage,  
## grapes,whole,  
## juice,frozen,  
## juice,low,  
## mix,green,  
## oil,  
## shrimp,almonds,avocado,vegetables,  
## smoothie,spinach,olive,  
## tea,honey,salad,mineral,  
## water,salmon,antioxydant,  
## weat,  
## yogurt,green}  
## [2] {burgers,meatballs,eggs}  
## [3] {chutney}  
## [4] {turkey,avocado}  
## [5] {bar,whole,  
## mineral,  
## rice,green,  
## tea,
```

```
##      water,milk,energy,
##      wheat}
## [6] {fat,
##      low,
##      yogurt}
## [7] {fries,
##      pasta,french,
##      wheat,
##      whole}
## [8] {cream,shallot,
##      soup,light}
## [9] {frozen,
##      tea,
##      vegetables,spaghetti,green}
## [10] {french,
##      fries}
```

```
# number of items in the first 10 itemset or transaction
size(head(association, 10))
```

```
## [1] 15  1  1  1  6  3  4  2  3  2
```

```
item <- as.data.frame(itemLabels(association))
colnames(item) <- "Item"
head(item,10)
```

```
##                                Item
## 1                                &
## 2                                accessories
## 3      accessories,antioxydant
## 4      accessories,champagne,fresh
## 5      accessories,champagne,protein
## 6      accessories,chocolate
## 7 accessories,chocolate,champagne,frozen
## 8      accessories,chocolate,frozen
## 9      accessories,chocolate,low
## 10     accessories,chocolate,pasta,salt
```

```
summary(association)
```

```
## transactions as itemMatrix in sparse format with
## 7501 rows (elements/itemsets/transactions) and
## 5729 columns (items) and a density of 0.0005421748
##
```

```
## most frequent items:
```

```
##      tea  wheat mineral      fat  yogurt (Other)
##      803    645    577    574    543    20157
```

```
##
```

```
## element (itemset/transaction) length distribution:
```

```
## sizes
```

```
##      1      2      3      4      5      6      7      8      9     10     11     12     13     15     16
## 1603 2007 1382  942  651  407  228  151   70   39   13    5    1    1    1
```

```
##
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.000   2.000   3.000   3.106   4.000   16.000
##
## includes extended item information - examples:
##                labels
## 1                &
## 2            accessories
## 3 accessories,antioxydant
```

```
print("Frequency of each item")
```

```
## [1] "Frequency of each item"
```

```
itemFrequency(association[,1:15], type="absolute")
```

```
##                &                accessories
##                371                10
##      accessories,antioxydant      accessories,champagne,fresh
##                1                1
##      accessories,champagne,protein      accessories,chocolate
##                1                1
## accessories,chocolate,champagne,frozen      accessories,chocolate,frozen
##                1                1
##      accessories,chocolate,low      accessories,chocolate,pasta,salt
##                1                1
##      accessories,chocolate,salt,green      accessories,cookies
##                1                5
##      accessories,cottage      accessories,escalope
##                2                1
##      accessories,french
##                13
```

```
cat ("Tab\n")
```

```
## Tab
```

```
print("Frequency of first 15 Items compared to Others")
```

```
## [1] "Frequency of first 15 Items compared to Others"
```

```
# How many times these items have appeared as compared to others.
round(itemFrequency(association[,1:15], type="relative")*100,2)
```

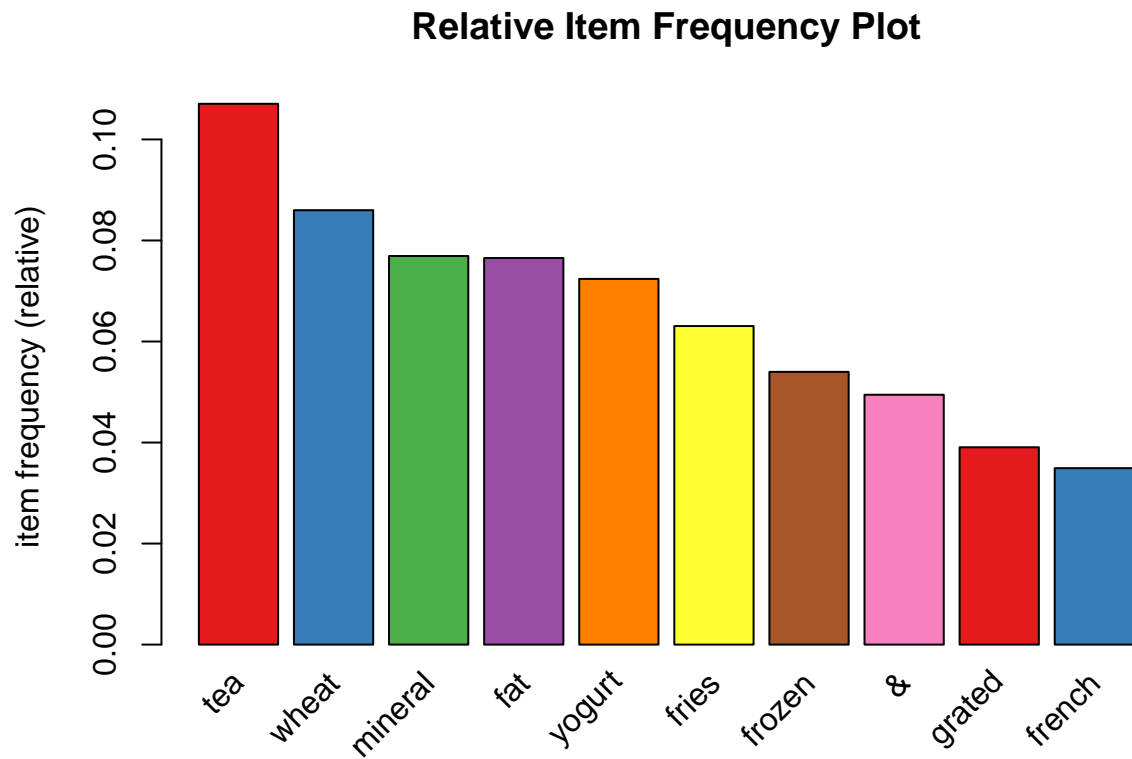
```
##                &                accessories
##                4.95                0.13
##      accessories,antioxydant      accessories,champagne,fresh
##                0.01                0.01
##      accessories,champagne,protein      accessories,chocolate
##                0.01                0.01
```

```
## accessories,chocolate,champagne,frozen      accessories,chocolate,frozen
##                                0.01           0.01
##                accessories,chocolate,low      accessories,chocolate,pasta,salt
##                                0.01           0.01
##                accessories,chocolate,salt,green accessories,cookies
##                                0.01           0.07
##                accessories,cottage             accessories,escalope
##                                0.03           0.01
##                accessories,french
##                                0.17
```

```
# Most common items
```

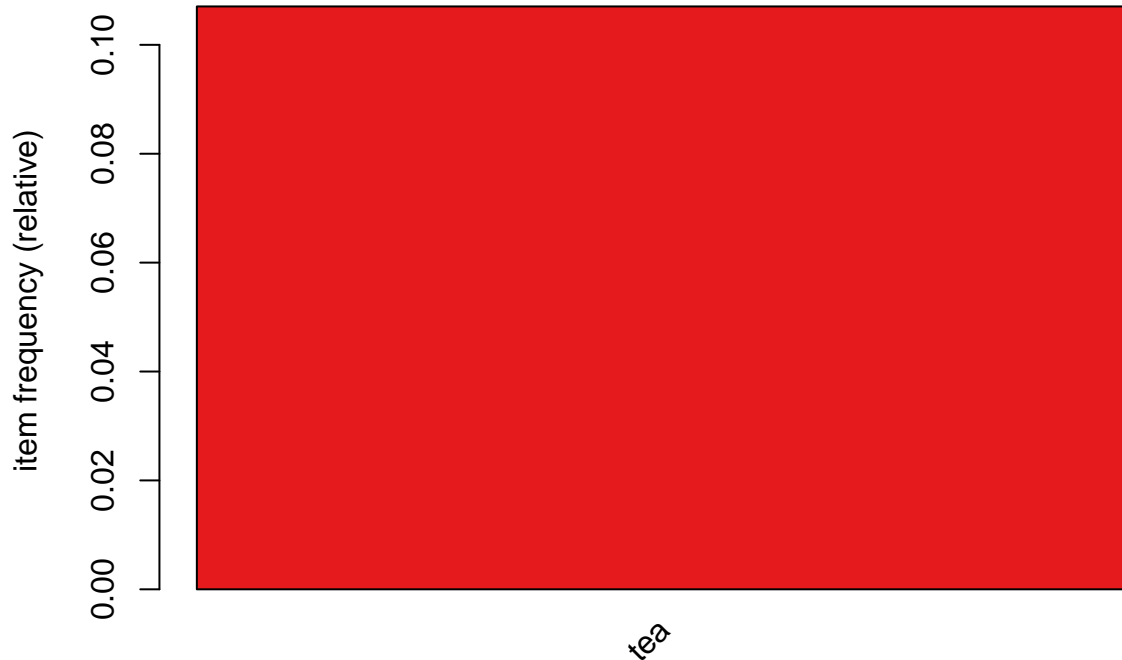
```
library(RColorBrewer)
```

```
itemFrequencyPlot(association, topN= 10, col=brewer.pal(8,'Set1'),main="Relative Item Frequency Plot")
```



```
itemFrequencyPlot(association, support= 0.1, col=brewer.pal(8,'Set1'),main="Relative Importance is at 1")
```

Relative Importance is at least 10%



```
# Build a Model based on association rules
# min_supp -> 0.001
# confidence -> 0.8
rules <- apriori(association, parameter = list(supp=0.001, conf=0.75))

## Apriori
##
## Parameter specification:
## confidence minval smax arem aval originalSupport maxtime support minlen
##          0.75   0.1   1 none FALSE                TRUE     5   0.001     1
## maxlen target  ext
##          10  rules TRUE
##
## Algorithmic control:
## filter tree heap memopt load sort verbose
##    0.1 TRUE TRUE  FALSE TRUE     2    TRUE
##
## Absolute minimum support count: 7
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[5729 item(s), 7501 transaction(s)] done [0.01s].
## sorting and recoding items ... [354 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 done [0.00s].
## writing ... [275 rule(s)] done [0.00s].
## creating S4 object ... done [0.00s].
```

```
rules
```

```
## set of 275 rules
```

```
summary(rules)
```

```
## set of 275 rules
```

```
##
```

```
## rule length distribution (lhs + rhs):sizes
```

```
## 2 3 4
```

```
## 110 145 20
```

```
##
```

```
## Min. 1st Qu. Median Mean 3rd Qu. Max.
```

```
## 2.000 2.000 3.000 2.673 3.000 4.000
```

```
##
```

```
## summary of quality measures:
```

##	support	confidence	coverage	lift
## Min.	:0.001067	Min. :0.7619	Min. :0.001067	Min. : 7.143
## 1st Qu.:	0.001200	1st Qu.:0.9211	1st Qu.:0.001200	1st Qu.: 11.630
## Median :	0.001600	Median :1.0000	Median :0.001600	Median : 13.068
## Mean :	0.002851	Mean :0.9603	Mean :0.003004	Mean : 22.589
## 3rd Qu.:	0.002666	3rd Qu.:1.0000	3rd Qu.:0.002800	3rd Qu.: 20.218
## Max. :	0.068391	Max. :1.0000	Max. :0.076523	Max. :613.718

```
## count
```

```
## Min. : 8.00
```

```
## 1st Qu.: 9.00
```

```
## Median : 12.00
```

```
## Mean : 21.39
```

```
## 3rd Qu.: 20.00
```

```
## Max. :513.00
```

```
##
```

```
## mining info:
```

```
## data ntransactions support confidence
```

```
## association 7501 0.001 0.75
```

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

##	lhs	rhs	support	confidence
## [1]	{cookies,low}	=> {yogurt}	0.001066524	1
## [2]	{cookies,low}	=> {fat}	0.001066524	1
## [3]	{extra}	=> {dark}	0.001066524	1
## [4]	{burgers,whole}	=> {wheat}	0.001199840	1
## [5]	{fries,escalope,pasta,mushroom}	=> {cream}	0.001066524	1

##	coverage	lift	count
## [1]	0.001066524	13.81400	8
## [2]	0.001066524	13.06794	8
## [3]	0.001066524	83.34444	8
## [4]	0.001199840	11.62946	9
## [5]	0.001066524	47.77707	8

```
rules<-sort(rules, by="confidence", decreasing=TRUE)
inspect(rules[1:10])
```

##	lhs	rhs	support	confidence
## [1]	{cookies,low}	=> {yogurt}	0.001066524	1
## [2]	{cookies,low}	=> {fat}	0.001066524	1
## [3]	{extra}	=> {dark}	0.001066524	1
## [4]	{burgers,whole}	=> {wheat}	0.001199840	1
## [5]	{fries,escalope,pasta,mushroom}	=> {cream}	0.001066524	1
## [6]	{fries,cookies,green}	=> {tea}	0.001333156	1
## [7]	{shrimp,whole}	=> {wheat}	0.001066524	1
## [8]	{rice,cake}	=> {wheat}	0.001333156	1
## [9]	{flour,green}	=> {weat}	0.001199840	1
## [10]	{rice,chocolate,french}	=> {wheat}	0.001066524	1

##	coverage	lift	count
## [1]	0.001066524	13.81400	8
## [2]	0.001066524	13.06794	8
## [3]	0.001066524	83.34444	8
## [4]	0.001199840	11.62946	9
## [5]	0.001066524	47.77707	8
## [6]	0.001333156	9.34122	10
## [7]	0.001066524	11.62946	8
## [8]	0.001333156	11.62946	10
## [9]	0.001199840	107.15714	9
## [10]	0.001066524	11.62946	8

The 5 most popular Items are:-

1. tea
2. wheat
3. mineal
4. fat
5. yoghurt

The 5 most popular have support at least 10% and 2 more items which are:-

1. green tea
2. milk

We have 4 rules with 100% confidence in the association. We can use this association to increase sales in Females customers in the Branch A and B buy identifying which products the Female Customers buy most This can also be done for Male Customers in Branch C

The Model can also be used to improve the Marketing and Sales of the Product Lines that are not popular with any of the customers in any Branch.