# Retail with Market Basket Analysis

Caio di Felice Cunha

## Retail with Market Basket Analysis

In this project, the objective is to work with a data mining technique, the Market Basket Analysis. The question is simple:

In a store or supermarket, if Customer A buys a banana, does he also buy an apple? If true, what is the strength of this relationship? If we have another Customer B similar to Customer A, what are the chances that when he buys a banana he will also buy an apple? How does this affect our Marketing campaigns?

## Stage 1 - Import the libraries and connect with SQL

Data Source:

https://cran.r-project.org/web/packages/arules/index.html

https://cran.r-project.org/web/packages/arulesViz/index.html

I uploaded this file and insert into MySQL Database Server

```
# Loading Packages
  library(dplyr)
  library(arules) #Package used for association rules
  library(arulesViz) # Package aimed at visualizing association rules
  library(htmlwidgets)
  library(writexl)
  library(RMySQL)
  options(warn=-1)

## Conection MySQL
con = dbConnect(
  MySQL(),
  user = "root",
  password = a,
  bname = "market_basket_analysis_db",
  host = h)
```

## Stage 2 - Selecting the data properly and eliminating unecessary rows

```
qry <- 'SELECT * FROM market_basket_analysis_db.dataset_bd3;'

# Load and explore the dataset
data <- dbGetQuery(con, qry)
# dim(data)
# head(data)
# summary(data)
# str(data)

# A smart way to solve the problem in the dataset
# Separate even lines from odd lines
even_lines <- seq(2, nrow(data), 2)
odd_lines <- seq(1, nrow(data), 2)

# We separate the data and then use the dataset with the even lines
(valid data lines)
df1 <- data[even_lines, ]
# head(df1)
df2 <- data[odd_lines, ]
# head(df2)
```

## Stage 3 - Check the white spaces

```
# Check if we have missing values represented by white space
which(nchar(trimws(df1$Item01))==0)

## integer(0)

which(nchar(trimws(df1$Item02))==0)

# Check if we have missing values represented by whitespace (using
regular expression)
grepl("^\\s*$", df1$Item02)
```

## Stage 4 - 1st Tough choice

Now we have our first difficult decision. We will only take the columns that have both Item01 and Item02 filled in, as our intention is to understand the Market Basket, therefore, if there is only 1 product, the analysis will be very difficult to give any productive result.

```
# Number of distinct items
n_distinct(df1)

## [1] 5176

# Let's work only with records where item 2 was not null
df1_two <- df1[!grepl("^\\s*$", df1$Item02), ]
```

```
# Number of distinct items
n_distinct(df1_two)

## [1] 5066
```

## Stage 5 - Preparing the dataset

The second difficult decision is to define which column of products to work with. Decided on the first 06 columns

```
# Prepare the package by converting the variables to factor type
# (variables we will use from now on)
#View(df1_two)
package <- df1_two
package$Item01 <- as.factor(package$Item01)
package$Item02 <- as.factor(package$Item02)
package$Item03 <- as.factor(package$Item03)
package$Item04 <- as.factor(package$Item04)
package$Item05 <- as.factor(package$Item05)
package$Item06 <- as.factor(package$Item06)
# summary(package)
# View(package)
# str(package)

package_split <- split(package$Item01,
                       package$Item02,
                       package$Item03,
                       package$Item04,
                       package$Item05,
                       package$Item06,
                      drop = FALSE)

#View(package_split)

# Transactions
transactions <- as(package_split, "transactions")
```

The variables must be converted to a factor so that we can perform the analysis

*product_rules1*

```
# Let's check the rules of a product: Dust-Off Compressed Gas 2 pack
product_rules1 <- apriori(transactions,
                          parameter = list(conf = 0.5, minlen = 3),
                          appearance = list(rhs = "Dust-Off Compressed
Gas 2 pack", default = "lhs"))
```

```
## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support
minlen
##         0.5    0.1    1 none FALSE            TRUE       5     0.1
3
##  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: 11
##
## set item appearances ...[1 item(s)] done [0.00s].
## set transactions ...[104 item(s), 117 transaction(s)] done [0.00s].
## sorting and recoding items ... [41 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 [0.00s].
## writing ... [2312 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].
```

```
# Inspection of the rules
inspect(head(sort(product_rules1, by = "confidence"), 5))
```

```
##     lhs                                       rhs
support confidence  coverage     lift count
## [1] {Apple Pencil,
##      HP 62XL Tri-Color ink}            => {Dust-Off Compressed Gas 2
pack} 0.1025641          1 0.1025641 1.746269    12
## [2] {3A USB Type C Cable 3 pack 6FT,
##      Screen Mom Screen Cleaner kit}    => {Dust-Off Compressed Gas 2
pack} 0.1196581          1 0.1196581 1.746269    14
## [3] {Logitech M510 Wireless mouse,
##      TP-Link AC1750 Smart WiFi Router} => {Dust-Off Compressed Gas 2
pack} 0.1111111          1 0.1111111 1.746269    13
## [4] {Anker USB C to HDMI Adapter,
##      TP-Link AC1750 Smart WiFi Router} => {Dust-Off Compressed Gas 2
pack} 0.1025641          1 0.1025641 1.746269    12
## [5] {Screen Mom Screen Cleaner kit,
##      TP-Link AC1750 Smart WiFi Router} => {Dust-Off Compressed Gas 2
pack} 0.1452991          1 0.1452991 1.746269    17
```

Here you can se that the ones who bought Apple Pencil and HP 62XL Tri-Color ink, also bought a Dust-Off Compressed Gas 2 pack

*product_rules2*

```
# Let's check the rules of a product: HP 61 ink
product_rules2 <- apriori(transactions,
```

```
                          parameter = list(minlen = 3, conf = 0.5),
                          appearance = list(rhs = "HP 61 ink",default =
"lhs"))

## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support
minlen
##         0.5    0.1    1 none FALSE            TRUE       5     0.1
3
##  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: 11
##
## set item appearances ...[1 item(s)] done [0.00s].
## set transactions ...[104 item(s), 117 transaction(s)] done [0.00s].
## sorting and recoding items ... [41 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 [0.00s].
## writing ... [2183 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].

# Inspection of the rules
inspect(head(sort(product_rules2, by = "confidence"), 5))

##     lhs                                          rhs           support
confidence  coverage     lift count
## [1] {Nylon Braided Lightning to USB cable,
##      TP-Link AC1750 Smart WiFi Router}      => {HP 61 ink} 0.1025641
1.0000000 0.1025641 2.127273    12
## [2] {Nylon Braided Lightning to USB cable,
##      TP-Link AC1750 Smart WiFi Router,
##      VIVO Dual LCD Monitor Desk mount}      => {HP 61 ink} 0.1025641
1.0000000 0.1025641 2.127273    12
## [3] {Dust-Off Compressed Gas 2 pack,
##      Nylon Braided Lightning to USB cable,
##      TP-Link AC1750 Smart WiFi Router}      => {HP 61 ink} 0.1025641
1.0000000 0.1025641 2.127273    12
## [4] {Dust-Off Compressed Gas 2 pack,
##      Nylon Braided Lightning to USB cable,
##      TP-Link AC1750 Smart WiFi Router,
##      VIVO Dual LCD Monitor Desk mount}      => {HP 61 ink} 0.1025641
1.0000000 0.1025641 2.127273    12
## [5] {FEIYOLD Blue light Blocking Glasses,
```

```
##      Nylon Braided Lightning to USB cable}  => {HP 61 ink} 0.1452991
0.9444444 0.1538462 2.009091    17
```

*product_rules3*

```
# Let's check the rules of a product: VIVO Dual LCD Monitor Desk mount
product_rules3 <- apriori(transactions,
                          parameter = list(minlen = 3, conf = 0.5),
                          appearance = list(rhs = "VIVO Dual LCD Monitor
Desk mount", default = "lhs"))

## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support
minlen
##         0.5    0.1    1 none FALSE            TRUE       5     0.1
3
##  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: 11
##
## set item appearances ...[1 item(s)] done [0.00s].
## set transactions ...[104 item(s), 117 transaction(s)] done [0.00s].
## sorting and recoding items ... [41 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 [0.00s].
## writing ... [2375 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].

# Inspection of the rules
inspect(head(sort(product_rules3, by = "confidence"), 5))

##     lhs                                          rhs
support confidence  coverage     lift count
## [1] {Logitech M510 Wireless mouse,
##      TP-Link AC1750 Smart WiFi Router}       => {VIVO Dual LCD Monitor
Desk mount} 0.1111111          1 0.1111111 2.207547    13
## [2] {Anker USB C to HDMI Adapter,
##      TP-Link AC1750 Smart WiFi Router}       => {VIVO Dual LCD Monitor
Desk mount} 0.1025641          1 0.1025641 2.207547    12
## [3] {Nylon Braided Lightning to USB cable,
##      TP-Link AC1750 Smart WiFi Router}       => {VIVO Dual LCD Monitor
Desk mount} 0.1025641          1 0.1025641 2.207547    12
## [4] {Apple Lightning to Digital AV Adapter,
##      TP-Link AC1750 Smart WiFi Router}       => {VIVO Dual LCD Monitor
Desk mount} 0.1111111          1 0.1111111 2.207547    13
```

```
## [5] {Moread HDMI to VGA Adapter,
##      SanDisk Ultra 64GB card}                => {VIVO Dual LCD Monitor
Desk mount} 0.1025641          1 0.1025641 2.207547    12
```

The LHS stands for antecedent while rhs stands for consequent

## Stage 6 - Change the metrics to make the model better

*product_rules1*

```
# Let's double-check the product rules: Dust-Off Compressed Gas 2 pack,
# changing one of the metrics
product_rules1 <- apriori(transactions,
                          parameter = list(minlen = 3, supp = 0.2, conf
= 0.5, target = "rules"),
                          appearance = list(rhs = "Dust-Off Compressed
Gas 2 pack", default = "lhs"))

## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support
minlen
##         0.5    0.1    1 none FALSE            TRUE       5     0.2
3
##  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: 23
##
## set item appearances ...[1 item(s)] done [0.00s].
## set transactions ...[104 item(s), 117 transaction(s)] done [0.00s].
## sorting and recoding items ... [23 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 done [0.00s].
## writing ... [38 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].

# Inspection of the rules
inspect(head(sort(product_rules1, by = "confidence"), 5))

##     lhs                                          rhs
support confidence  coverage     lift count
## [1] {Logitech M510 Wireless mouse,
##      Screen Mom Screen Cleaner kit}         => {Dust-Off Compressed
Gas 2 pack} 0.2222222          1 0.2222222 1.746269    26
## [2] {Anker USB C to HDMI Adapter,
##      Screen Mom Screen Cleaner kit}         => {Dust-Off Compressed
```

```
Gas 2 pack} 0.2307692          1 0.2307692 1.746269    27
## [3] {Screen Mom Screen Cleaner kit,
##      VIVO Dual LCD Monitor Desk mount}      => {Dust-Off Compressed
Gas 2 pack} 0.2991453          1 0.2991453 1.746269    35
## [4] {Nylon Braided Lightning to USB cable,
##      Screen Mom Screen Cleaner kit}         => {Dust-Off Compressed
Gas 2 pack} 0.2649573          1 0.2649573 1.746269    31
## [5] {Anker USB C to HDMI Adapter,
##      Screen Mom Screen Cleaner kit,
##      VIVO Dual LCD Monitor Desk mount}      => {Dust-Off Compressed
Gas 2 pack} 0.2136752          1 0.2136752 1.746269    25

# Filter out redundant rules
product_rules1_clean <- product_rules1[!is.redundant(product_rules1)]

# Inspection of the rules
inspect(head(sort(product_rules1_clean, by = "confidence"), 5))

##     lhs                                           rhs
support confidence  coverage     lift count
## [1] {Logitech M510 Wireless mouse,
##      Screen Mom Screen Cleaner kit}          => {Dust-Off Compressed
Gas 2 pack} 0.2222222    1.00000 0.2222222 1.746269    26
## [2] {Anker USB C to HDMI Adapter,
##      Screen Mom Screen Cleaner kit}          => {Dust-Off Compressed
Gas 2 pack} 0.2307692    1.00000 0.2307692 1.746269    27
## [3] {Screen Mom Screen Cleaner kit,
##      VIVO Dual LCD Monitor Desk mount}       => {Dust-Off Compressed
Gas 2 pack} 0.2991453    1.00000 0.2991453 1.746269    35
## [4] {Nylon Braided Lightning to USB cable,
##      Screen Mom Screen Cleaner kit}          => {Dust-Off Compressed
Gas 2 pack} 0.2649573    1.00000 0.2649573 1.746269    31
## [5] {Apple Lightning to Digital AV Adapter,
##      Screen Mom Screen Cleaner kit}          => {Dust-Off Compressed
Gas 2 pack} 0.2649573    0.96875 0.2735043 1.691698    31

# Sumário
summary(product_rules1_clean)

## set of 27 rules
##
## rule length distribution (lhs + rhs):sizes
##  3  4
## 23  4
##
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   3.000   3.000   3.000   3.148   3.000   4.000
##
## summary of quality measures:
##     support         confidence        coverage           lift
##  Min.   :0.2051   Min.   :0.6154   Min.   :0.2222   Min.   :1.075
```

```
##  1st Qu.:0.2222   1st Qu.:0.7742   1st Qu.:0.2521   1st Qu.:1.352
##  Median :0.2308   Median :0.9062   Median :0.2735   Median :1.583
##  Mean   :0.2384   Mean   :0.8710   Mean   :0.2770   Mean   :1.521
##  3rd Qu.:0.2479   3rd Qu.:0.9655   3rd Qu.:0.3034   3rd Qu.:1.686
##  Max.   :0.2991   Max.   :1.0000   Max.   :0.3675   Max.   :1.746
##      count
##  Min.   :24.00
##  1st Qu.:26.00
##  Median :27.00
##  Mean   :27.89
##  3rd Qu.:29.00
##  Max.   :35.00
##
## mining info:
##          data ntransactions support confidence
##  transactions           117     0.2        0.5
##
call
##  apriori(data = transactions, parameter = list(minlen = 3, supp = 0.2,
conf = 0.5, target = "rules"), appearance = list(rhs = "Dust-Off
Compressed Gas 2 pack", default = "lhs"))
```

*product_rules2*

```
# Let's double-check the product rules: HP 61 ink,
# changing one of the metrics
product_rules2 <- apriori(transactions,
                          parameter = list(minlen = 3, supp = 0.2, conf
= 0.5, target = "rules"),
                          appearance = list(rhs = "HP 61 ink", default =
"lhs"))
```

```
## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support
minlen
##         0.5    0.1    1 none FALSE            TRUE       5     0.2
3
##  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: 23
##
## set item appearances ...[1 item(s)] done [0.00s].
## set transactions ...[104 item(s), 117 transaction(s)] done [0.00s].
## sorting and recoding items ... [23 item(s)] done [0.00s].
```

```
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 done [0.00s].
## writing ... [35 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].

# Inspection of the rules
inspect(head(sort(product_rules2, by = "confidence"), 5))

##     lhs                                             rhs           support
confidence  coverage     lift count
## [1] {Apple Lightning to Digital AV Adapter,
##      SAMSUNG EVO 32GB card}                      => {HP 61 ink} 0.2478632
0.8787879 0.2820513 1.869421    29
## [2] {Apple Lightning to Digital AV Adapter,
##      SanDisk 128GB Ultra microSDXC card}         => {HP 61 ink} 0.2136752
0.8620690 0.2478632 1.833856    25
## [3] {Apple Lightning to Digital AV Adapter,
##      Cat8 Ethernet Cable,
##      SAMSUNG EVO 32GB card}                      => {HP 61 ink} 0.2136752
0.8620690 0.2478632 1.833856    25
## [4] {Anker USB C to HDMI Adapter,
##      SAMSUNG EVO 32GB card}                      => {HP 61 ink} 0.2051282
0.8571429 0.2393162 1.823377    24
## [5] {Apple Lightning to Digital AV Adapter,
##      Nylon Braided Lightning to USB cable,
##      SAMSUNG EVO 32GB card}                      => {HP 61 ink} 0.2051282
0.8571429 0.2393162 1.823377    24

# Filter out redundant rules
product_rules2_clean <- product_rules2[!is.redundant(product_rules2)]

# Inspection of the rules
inspect(head(sort(product_rules2_clean, by = "confidence"), 5))

##     lhs                                             rhs           support
confidence  coverage     lift count
## [1] {Apple Lightning to Digital AV Adapter,
##      SAMSUNG EVO 32GB card}                      => {HP 61 ink} 0.2478632
0.8787879 0.2820513 1.869421    29
## [2] {Apple Lightning to Digital AV Adapter,
##      SanDisk 128GB Ultra microSDXC card}         => {HP 61 ink} 0.2136752
0.8620690 0.2478632 1.833856    25
## [3] {Anker USB C to HDMI Adapter,
##      SAMSUNG EVO 32GB card}                      => {HP 61 ink} 0.2051282
0.8571429 0.2393162 1.823377    24
## [4] {Nylon Braided Lightning to USB cable,
##      SAMSUNG EVO 32GB card}                      => {HP 61 ink} 0.2307692
0.8437500 0.2735043 1.794886    27
## [5] {Anker USB C to HDMI Adapter,
##      Apple Lightning to Digital AV Adapter,
```

```
##      Nylon Braided Lightning to USB cable}   => {HP 61 ink} 0.2051282
0.8275862 0.2478632 1.760502    24

# Summary
summary(product_rules2_clean)

## set of 30 rules
##
## rule length distribution (lhs + rhs):sizes
##  3  4
## 25  5
##
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   3.000   3.000   3.000   3.167   3.000   4.000
##
## summary of quality measures:
##     support         confidence        coverage           lift
##  Min.   :0.2051   Min.   :0.6304   Min.   :0.2393   Min.   :1.341
##  1st Qu.:0.2051   1st Qu.:0.7407   1st Qu.:0.2650   1st Qu.:1.576
##  Median :0.2179   Median :0.7777   Median :0.2821   Median :1.654
##  Mean   :0.2236   Mean   :0.7720   Mean   :0.2920   Mean   :1.642
##  3rd Qu.:0.2393   3rd Qu.:0.8042   3rd Qu.:0.3141   3rd Qu.:1.711
##  Max.   :0.2650   Max.   :0.8788   Max.   :0.3932   Max.   :1.869
##      count
##  Min.   :24.00
##  1st Qu.:24.00
##  Median :25.50
##  Mean   :26.17
##  3rd Qu.:28.00
##  Max.   :31.00
##
## mining info:
##          data ntransactions support confidence
##  transactions           117     0.2        0.5
##
call
##  apriori(data = transactions, parameter = list(minlen = 3, supp = 0.2,
conf = 0.5, target = "rules"), appearance = list(rhs = "HP 61 ink",
default = "lhs"))
```

*product_rules3*

```
# Let's double-check the product rules: VIVO Dual LCD Monitor Desk mount,
# changing one of the metrics
product_rules3 <- apriori(transactions,
                          parameter = list(minlen = 3, supp = 0.2, conf
= 0.5, target = "rules"),
                          appearance = list(rhs = "VIVO Dual LCD Monitor
Desk mount", default = "lhs"))

## Apriori
##
```

```
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support
minlen
##         0.5    0.1    1 none FALSE            TRUE       5     0.2
3
##  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: 23
##
## set item appearances ...[1 item(s)] done [0.00s].
## set transactions ...[104 item(s), 117 transaction(s)] done [0.00s].
## sorting and recoding items ... [23 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 done [0.00s].
## writing ... [34 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].

# Inspection of the rules
inspect(head(sort(product_rules3, by = "confidence"), 5))

##     lhs                                           rhs
support confidence  coverage     lift count
## [1] {Dust-Off Compressed Gas 2 pack,
##      SanDisk Ultra 64GB card}                 => {VIVO Dual LCD Monitor
Desk mount} 0.2307692  0.9642857 0.2393162 2.128706    27
## [2] {Apple Lightning to Digital AV Adapter,
##      Dust-Off Compressed Gas 2 pack,
##      SanDisk Ultra 64GB card}                 => {VIVO Dual LCD Monitor
Desk mount} 0.2136752  0.9615385 0.2222222 2.122642    25
## [3] {Apple Lightning to Digital AV Adapter,
##      Dust-Off Compressed Gas 2 pack,
##      Logitech M510 Wireless mouse}            => {VIVO Dual LCD Monitor
Desk mount} 0.2136752  0.9615385 0.2222222 2.122642    25
## [4] {Anker USB C to HDMI Adapter,
##      Apple Lightning to Digital AV Adapter,
##      Dust-Off Compressed Gas 2 pack}          => {VIVO Dual LCD Monitor
Desk mount} 0.2136752  0.9615385 0.2222222 2.122642    25
## [5] {Anker USB C to HDMI Adapter,
##      Dust-Off Compressed Gas 2 pack,
##      Nylon Braided Lightning to USB cable}    => {VIVO Dual LCD Monitor
Desk mount} 0.2307692  0.9310345 0.2478632 2.055303    27

# Filter out redundant rules
product_rules3_clean <- product_rules3[!is.redundant(product_rules3)]
```

```r
# Inspection of the rules
inspect(head(sort(product_rules3_clean, by = "confidence"), 5))
```

```
##     lhs                                             rhs
support confidence  coverage     lift count
## [1] {Dust-Off Compressed Gas 2 pack,
##      SanDisk Ultra 64GB card}                => {VIVO Dual LCD Monitor
Desk mount} 0.2307692  0.9642857 0.2393162 2.128706    27
## [2] {Apple Lightning to Digital AV Adapter,
##      Dust-Off Compressed Gas 2 pack,
##      Logitech M510 Wireless mouse}           => {VIVO Dual LCD Monitor
Desk mount} 0.2136752  0.9615385 0.2222222 2.122642    25
## [3] {Anker USB C to HDMI Adapter,
##      Apple Lightning to Digital AV Adapter,
##      Dust-Off Compressed Gas 2 pack}         => {VIVO Dual LCD Monitor
Desk mount} 0.2136752  0.9615385 0.2222222 2.122642    25
## [4] {Anker USB C to HDMI Adapter,
##      Dust-Off Compressed Gas 2 pack,
##      Nylon Braided Lightning to USB cable}   => {VIVO Dual LCD Monitor
Desk mount} 0.2307692  0.9310345 0.2478632 2.055303    27
## [5] {Anker USB C to HDMI Adapter,
##      Screen Mom Screen Cleaner kit}          => {VIVO Dual LCD Monitor
Desk mount} 0.2136752  0.9259259 0.2307692 2.044025    25
```

```r
# Summary
summary(product_rules3_clean)
```

```
## set of 30 rules
##
## rule length distribution (lhs + rhs):sizes
##  3  4
## 21  9
##
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     3.0     3.0     3.0     3.3     4.0     4.0
##
## summary of quality measures:
##     support         confidence        coverage           lift
##  Min.   :0.2051   Min.   :0.6047   Min.   :0.2222   Min.   :1.335
##  1st Qu.:0.2137   1st Qu.:0.7757   1st Qu.:0.2479   1st Qu.:1.712
##  Median :0.2308   Median :0.8469   Median :0.2650   Median :1.870
##  Mean   :0.2308   Mean   :0.8360   Mean   :0.2795   Mean   :1.845
##  3rd Qu.:0.2372   3rd Qu.:0.8929   3rd Qu.:0.3056   3rd Qu.:1.971
##  Max.   :0.2991   Max.   :0.9643   Max.   :0.3932   Max.   :2.129
##      count
##  Min.   :24.00
##  1st Qu.:25.00
##  Median :27.00
##  Mean   :27.00
##  3rd Qu.:27.75
##  Max.   :35.00
```

```
##
## mining info:
##          data ntransactions support confidence
##  transactions           117     0.2        0.5
##
call
##  apriori(data = transactions, parameter = list(minlen = 3, supp = 0.2,
conf = 0.5, target = "rules"), appearance = list(rhs = "VIVO Dual LCD
Monitor Desk mount", default = "lhs"))
```

We can visually see that the metrics have improved

## Stage 7 - Top 3 rules

```
# Top 3 rules
inspect(head(sort(product_rules1_clean, by = "support", decreasing =
TRUE), 1))

##     lhs                                      rhs
support confidence  coverage     lift count
## [1] {Screen Mom Screen Cleaner kit,
##      VIVO Dual LCD Monitor Desk mount} => {Dust-Off Compressed Gas 2
pack} 0.2991453          1 0.2991453 1.746269    35

inspect(head(sort(product_rules2_clean, by = "confidence", decreasing =
TRUE), 1))

##     lhs                                          rhs           support
confidence  coverage     lift count
## [1] {Apple Lightning to Digital AV Adapter,
##      SAMSUNG EVO 32GB card}                  => {HP 61 ink} 0.2478632
0.8787879 0.2820513 1.869421    29

inspect(head(sort(product_rules3_clean, by = "confidence", decreasing =
TRUE), 1))

##     lhs                                    rhs
support confidence  coverage     lift count
## [1] {Dust-Off Compressed Gas 2 pack,
##      SanDisk Ultra 64GB card}         => {VIVO Dual LCD Monitor Desk
mount} 0.2307692  0.9642857 0.2393162 2.128706    27
```

## Stage 8 - Saving the results

```
# We save the ruleset of the 3 products as a dataframe and then save it
to disk
df_product1 <- as(product_rules1_clean, "data.frame")
head(df_product1)

##
rules
## 1      {SanDisk Ultra 64GB card,VIVO Dual LCD Monitor Desk mount} =>
{Dust-Off Compressed Gas 2 pack}
```

```
## 2 {Apple Lightning to Digital AV Adapter,SanDisk Ultra 64GB card} =>
{Dust-Off Compressed Gas 2 pack}
## 3        {SAMSUNG EVO 32GB card,VIVO Dual LCD Monitor Desk mount} =>
{Dust-Off Compressed Gas 2 pack}
## 4     {Apple Lightning to Digital AV Adapter,Cat8 Ethernet Cable} =>
{Dust-Off Compressed Gas 2 pack}
## 5    {Logitech M510 Wireless mouse,Screen Mom Screen Cleaner kit} =>
{Dust-Off Compressed Gas 2 pack}
## 6 {Logitech M510 Wireless mouse,VIVO Dual LCD Monitor Desk mount} =>
{Dust-Off Compressed Gas 2 pack}
##     support confidence  coverage     lift count
## 1 0.2307692  0.9310345 0.2478632 1.625836    27
## 2 0.2222222  0.8666667 0.2564103 1.513433    26
## 3 0.2136752  0.8928571 0.2393162 1.559168    25
## 4 0.2051282  0.6153846 0.3333333 1.074627    24
## 5 0.2222222  1.0000000 0.2222222 1.746269    26
## 6 0.2478632  0.9062500 0.2735043 1.582556    29

write_xlsx(df_product1, "df_product1.xlsx")

df_product2 <- as(product_rules2_clean, "data.frame")
head(df_product2)

##
rules
## 1 {Apple Lightning to Digital AV Adapter,SanDisk 128GB Ultra microSDXC
card} => {HP 61 ink}
## 2             {Nylon Braided Lightning to USB cable,VicTsing Wireless
mouse} => {HP 61 ink}
## 3            {Apple Lightning to Digital AV Adapter,VicTsing Wireless
mouse} => {HP 61 ink}
## 4            {Apple Lightning to Digital AV Adapter,SanDisk Ultra 64GB
card} => {HP 61 ink}
## 5                                  {Cat8 Ethernet Cable,SAMSUNG EVO 32GB
card} => {HP 61 ink}
## 6                         {Anker USB C to HDMI Adapter,SAMSUNG EVO 32GB
card} => {HP 61 ink}
##     support confidence  coverage     lift count
## 1 0.2136752  0.8620690 0.2478632 1.833856    25
## 2 0.2051282  0.7741935 0.2649573 1.646921    24
## 3 0.2051282  0.8000000 0.2564103 1.701818    24
## 4 0.2051282  0.8000000 0.2564103 1.701818    24
## 5 0.2222222  0.7878788 0.2820513 1.676033    26
## 6 0.2051282  0.8571429 0.2393162 1.823377    24

write_xlsx(df_product2, "df_product2.xlsx")

df_product3 <- as(product_rules3_clean, "data.frame")
head(df_product3)
```

```
##
rules
## 1       {Nylon Braided Lightning to USB cable,SanDisk Ultra 64GB card}
=> {VIVO Dual LCD Monitor Desk mount}
## 2      {Apple Lightning to Digital AV Adapter,SanDisk Ultra 64GB card}
=> {VIVO Dual LCD Monitor Desk mount}
## 3             {Dust-Off Compressed Gas 2 pack,SanDisk Ultra 64GB card}
=> {VIVO Dual LCD Monitor Desk mount}
## 4               {Dust-Off Compressed Gas 2 pack,SAMSUNG EVO 32GB card}
=> {VIVO Dual LCD Monitor Desk mount}
## 5  {Logitech M510 Wireless mouse,Nylon Braided Lightning to USB cable}
=> {VIVO Dual LCD Monitor Desk mount}
## 6 {Apple Lightning to Digital AV Adapter,Logitech M510 Wireless mouse}
=> {VIVO Dual LCD Monitor Desk mount}
##     support confidence  coverage     lift count
## 1 0.2051282  0.8888889 0.2307692 1.962264    24
## 2 0.2307692  0.9000000 0.2564103 1.986792    27
## 3 0.2307692  0.9642857 0.2393162 2.128706    27
## 4 0.2136752  0.8928571 0.2393162 1.971024    25
## 5 0.2136752  0.6756757 0.3162393 1.491586    25
## 6 0.2307692  0.7714286 0.2991453 1.702965    27

write_xlsx(df_product3, "df_product3.xlsx")
```

## Conclusion

As you can see, Basket Market Analysis is a powerful weapon. In the hands of those who know how to use it, it can greatly help all areas of the company, such as marketing and sales.

Every company that aims to have good sales of its products should use this method

## Disclaimer:

Disclaimer: a good part of this project was largely done in the Data Science Academy, Big Data Analytics with R and Microsoft Azure Machine Learning course (part of the Data Scientist training)

## End