M3T4file.R

christiancobollogomez

2021-12-26

```r
## Module 3 Task 4

#install.packages("arules")
#install.packages("arulesViz")

library(corrplot)
```

```
## corrplot 0.92 loaded
```

```r
library(caret)
```

```
## Loading required package: ggplot2
```

```
## Loading required package: lattice
```

```r
library(gbm)
```

```
## Loaded gbm 2.1.8
```

```r
library(reshape2) # Heatmap with melt()
library(ggplot2)

library(DataExplorer) # quick EDA package

library(arules)
```

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

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

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

```r
library(arulesViz)

set.seed(123)


## Load, prepare and visualize the data:

# Electronidex's data contains rows that represent single transactions with the
# purchased item(s) being separated by commas, which is also called a 'basket'
# format. RStudio does not inherently understand transactional data. If the
# dataset is uploaded using the read.csv () function, RStudio will try to create
```

```
# variables (remember your dataset doesn't have variables) and you will
# encounter problems with your analysis. Therefore, we need to upload the
# ElectronidexTransactions.csv through the read.transactions() function.

Data<-read.transactions("ElectronidexTransactions2017.csv", format = "basket",
                        sep=",", rm.duplicates=TRUE)
```

Warning in readLines(file, encoding = encoding): incomplete final line found on
'ElectronidexTransactions2017.csv'

distribution of transactions with duplicates:
items
1   2
191 10

```
# The read.transactions() function changes the dataset into a sparse matrix.
# It makes each row represent a transaction and creates columns for each item
# that a customer might purchase. Electronidex sells 125 items, so the sparse
# 3 matrix creates 125 columns. It also changes the data to binary.
# (1=item purchased in that transaction OR 0=no purchase.)



## We put comments to not show all the long data.
## we could algo inspect a sample.

# inspect(Data) # We can view the transactions.
length (Data) # Number of transactions.
```

[1] 9835

```
# size (Data) # Number of items per transaction
# LIST(Data) # Lists the transactions by conversion (LIST must be capitalized)
itemLabels(Data)# To see the item labels. in this case, all the different products
```

[1] "1TB Portable External Hard Drive"
[2] "2TB Portable External Hard Drive"
[3] "3-Button Mouse"
[4] "3TB Portable External Hard Drive"
[5] "5TB Desktop Hard Drive"
[6] "Acer Aspire"
[7] "Acer Desktop"
[8] "Acer Monitor"
[9] "Ailihen Stereo Headphones"
[10] "Alienware Laptop"
[11] "AOC Monitor"
[12] "APIE Bluetooth Headphone"
[13] "Apple Earpods"
[14] "Apple MacBook Air"
[15] "Apple MacBook Pro"
[16] "Apple Magic Keyboard"
[17] "Apple TV"
[18] "Apple Wired Keyboard"
[19] "Apple Wireless Keyboard"
[20] "ASUS 2 Monitor"
[21] "ASUS Chromebook"
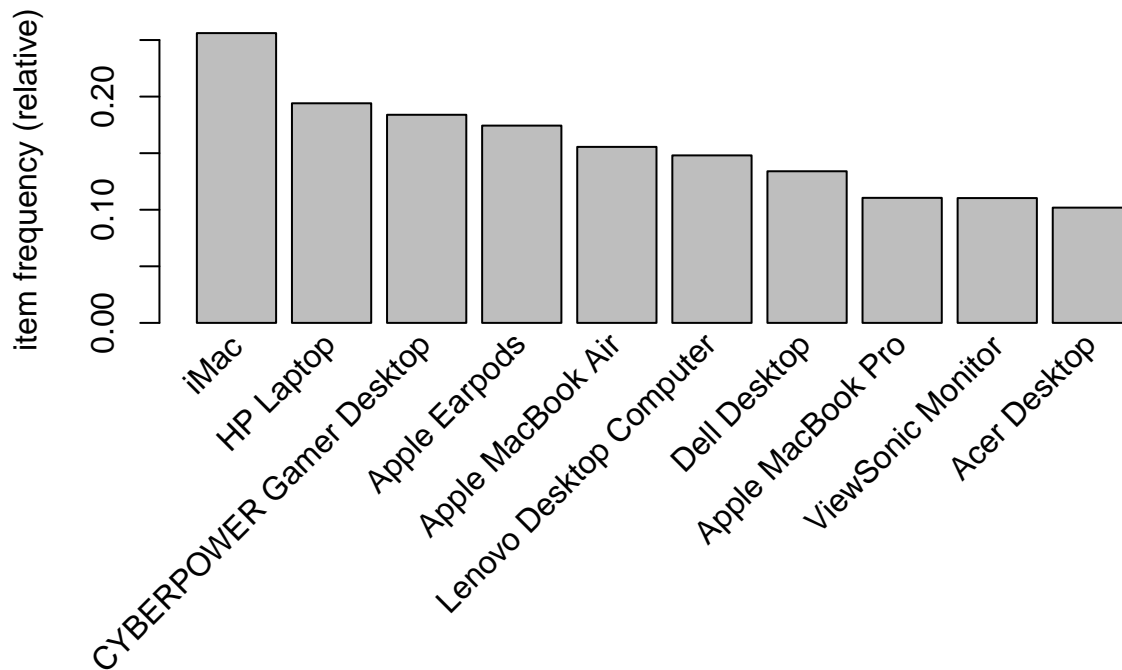##   [22] "ASUS Desktop"

```
##  [23] "ASUS Monitor"
##  [24] "Audio Cable"
##  [25] "Backlit LED Gaming Keyboard"
##  [26] "Belkin Mouse Pad"
##  [27] "Bose Companion Speakers"
##  [28] "Brother Printer"
##  [29] "Brother Printer Toner"
##  [30] "Cambridge Bluetooth Speaker"
##  [31] "Canon Ink"
##  [32] "Canon Office Printer"
##  [33] "Computer Game"
##  [34] "Cyber Acoustics"
##  [35] "CYBERPOWER Gamer Desktop"
##  [36] "Dell 2 Desktop"
##  [37] "Dell Desktop"
##  [38] "Dell KM117 Wireless Keyboard & Mouse"
##  [39] "Dell Laptop"
##  [40] "Dell Monitor"
##  [41] "Dell Wired Keyboard"
##  [42] "DOSS Touch Wireless Bluetooth"
##  [43] "DYMO Label Manker"
##  [44] "DYMO Labeling Tape"
##  [45] "EagleTec Wireless Combo Keyboard and Mouse"
##  [46] "Eluktronics Pro Gaming Laptop"
##  [47] "Epson Black Ink"
##  [48] "Epson Printer"
##  [49] "Etekcity Power Extension Cord Cable"
##  [50] "Ethernet Cable"
##  [51] "Fire HD Tablet"
##  [52] "Fire TV Stick"
##  [53] "Full Motion Monitor Mount"
##  [54] "Gaming Mouse Professional"
##  [55] "Generic Black 3-Button"
##  [56] "Google Home"
##  [57] "Halter Acrylic Monitor Stand"
##  [58] "Halter Mesh Metal Monitor Stand"
##  [59] "HDMI Adapter"
##  [60] "HDMI Cable 6ft"
##  [61] "Height-Adjustable Standing Desk"
##  [62] "HP Black & Tri-color Ink"
##  [63] "HP Desktop"
##  [64] "HP Laptop"
##  [65] "HP Monitor"
##  [66] "HP Notebook Touchscreen Laptop PC"
##  [67] "HP USB Keyboard"
##  [68] "HP Wireless Mouse"
##  [69] "HP Wireless Printer"
##  [70] "iMac"
##  [71] "Intel Desktop"
##  [72] "iPad"
##  [73] "iPad Pro"
##  [74] "iPhone Charger Cable"
##  [75] "JBL Splashproof Portable Bluetooth Speaker"
##  [76] "Kensington Headphones"
```

```
##  [77] "Kindle"
##  [78] "Koss Home Headphones"
##  [79] "Large Mouse Pad"
##  [80] "Lenovo Desktop Computer"
##  [81] "LG Monitor"
##  [82] "LG Touchscreen Laptop"
##  [83] "Logitech 3-button Mouse"
##  [84] "Logitech ClearChat Headset"
##  [85] "Logitech Desktop MK120 Mouse and keyboard Combo"
##  [86] "Logitech Keyboard"
##  [87] "Logitech MK270 Wireless Keyboard and Mouse Combo"
##  [88] "Logitech MK360 Wireless Keyboard and Mouse Combo"
##  [89] "Logitech MK550 Wireless Wave Keyboard and Mouse Combo"
##  [90] "Logitech Multimedia Speakers"
##  [91] "Logitech Stereo Headset"
##  [92] "Logitech Wireless Keyboard"
##  [93] "Logitech Wireless Mouse"
##  [94] "Mackie CR Speakers"
##  [95] "Microsoft Basic Optical Mouse"
##  [96] "Microsoft Headset"
##  [97] "Microsoft Office Home and Student 2016"
##  [98] "Microsoft Wireless Comfort Keyboard and Mouse"
##  [99] "Microsoft Wireless Desktop Keyboard and Mouse"
## [100] "Monster Beats By Dr Dre"
## [101] "Multi Media Stand"
## [102] "Otium Wireless Sports Bluetooth Headphone"
## [103] "Panasonic In-Ear Headphone"
## [104] "Panasonic On-Ear Stereo Headphones"
## [105] "PC Gaming Headset"
## [106] "Philips Flexible Earhook Headphone"
## [107] "Redragon Gaming Mouse"
## [108] "Rii LED Gaming Keyboard & Mouse Combo"
## [109] "Rii LED Keyboard"
## [110] "Rokono Mini Speaker"
## [111] "Roku Express"
## [112] "Samsung Charging Cable"
## [113] "Samsung Galaxy Tablet"
## [114] "Samsung Monitor"
## [115] "Sceptre Monitor"
## [116] "Slim 2TB Portable External Hard Drive"
## [117] "Slim Wireless Mouse"
## [118] "Smart Light Bulb"
## [119] "Sonos"
## [120] "USB Cable"
## [121] "VGA Monitor Cable"
## [122] "ViewSonic Monitor"
## [123] "Wireless Portable Mouse"
## [124] "XIBERIA Gaming Headset"
## [125] "Zombie Gaming Headset"
## Here we show in order the top 10 most sold products.
itemFrequencyPlot(Data, topN=10)
```
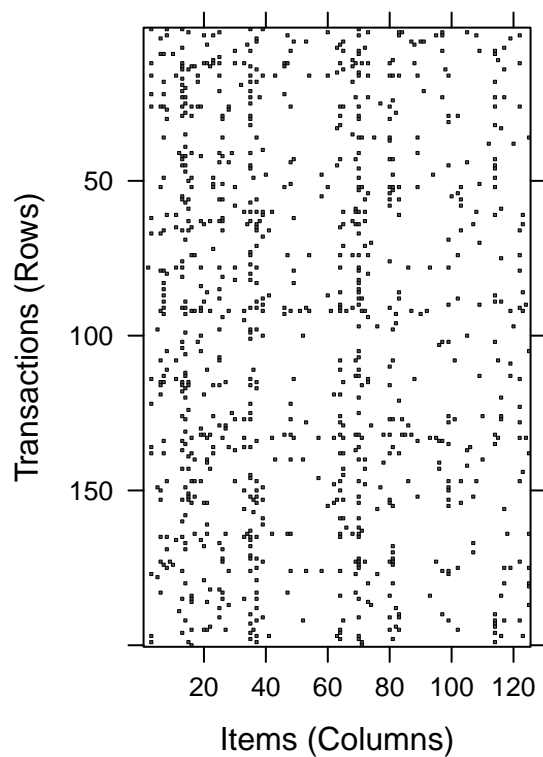
```
image(sample(Data,200))
```

```
RulesName<-apriori(Data, parameter = list(supp = 0.04, conf = 0.30, maxtime=10))
```

```
## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support minlen
##         0.3    0.1    1 none FALSE            TRUE      10    0.04      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: 393
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[125 item(s), 9835 transaction(s)] done [0.00s].
## sorting and recoding items ... [32 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 done [0.00s].
## writing ... [8 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].
```

```
RulesName
```

```
## set of 8 rules
```

```
inspect(RulesName)
```

```
##     lhs                         rhs          support    confidence coverage
## [1] {ViewSonic Monitor}      => {HP Laptop} 0.04799187 0.4350230  0.1103203
## [2] {ViewSonic Monitor}      => {iMac}      0.04941535 0.4479263  0.1103203
## [3] {Dell Desktop}           => {HP Laptop} 0.04494154 0.3353566  0.1340112
## [4] {Dell Desktop}           => {iMac}      0.05460092 0.4074355  0.1340112
## [5] {CYBERPOWER Gamer Desktop} => {iMac}     0.05673615 0.3084577  0.1839349
## [6] {Lenovo Desktop Computer} => {HP Laptop} 0.04616167 0.3118132  0.1480427
## [7] {Lenovo Desktop Computer} => {iMac}      0.05876970 0.3969780  0.1480427
## [8] {HP Laptop}              => {iMac}      0.07554652 0.3892090  0.1941027
##     lift       count
```

```
## [1] 2.241200 472
## [2] 1.748851 486
## [3] 1.727728 442
## [4] 1.590762 537
## [5] 1.204320 558
## [6] 1.606434 454
## [7] 1.549932 578
## [8] 1.519599 743
```

```
summary(RulesName)
```

```
## set of 8 rules
##
## rule length distribution (lhs + rhs):sizes
## 2
## 8
##
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       2       2       2       2       2       2
##
## summary of quality measures:
##     support          confidence         coverage           lift
##  Min.   :0.04494   Min.   :0.3085   Min.   :0.1103   Min.   :1.204
##  1st Qu.:0.04753   1st Qu.:0.3295   1st Qu.:0.1281   1st Qu.:1.542
##  Median :0.05201   Median :0.3931   Median :0.1410   Median :1.599
##  Mean   :0.05427   Mean   :0.3790   Mean   :0.1453   Mean   :1.649
##  3rd Qu.:0.05724   3rd Qu.:0.4143   3rd Qu.:0.1570   3rd Qu.:1.733
##  Max.   :0.07555   Max.   :0.4479   Max.   :0.1941   Max.   :2.241
##      count
##  Min.   :442.0
##  1st Qu.:467.5
##  Median :511.5
##  Mean   :533.8
##  3rd Qu.:563.0
##  Max.   :743.0
##
## mining info:
##   data ntransactions support confidence
##   Data          9835    0.04        0.3
##                                                                  call
##   apriori(data = Data, parameter = list(supp = 0.04, conf = 0.3, maxtime = 10))
```

```
## We order the Rules by confidence and support. The top one has near 45% of
## confidence. Also, observe that the higher support is a 7.5%.
```

```
inspect(sort(RulesName,by="confidence"))
```

```
##     lhs                          rhs            support    confidence coverage
## [1] {ViewSonic Monitor}       => {iMac}        0.04941535 0.4479263  0.1103203
## [2] {ViewSonic Monitor}       => {HP Laptop}   0.04799187 0.4350230  0.1103203
## [3] {Dell Desktop}            => {iMac}        0.05460092 0.4074355  0.1340112
## [4] {Lenovo Desktop Computer} => {iMac}        0.05876970 0.3969780  0.1480427
## [5] {HP Laptop}               => {iMac}        0.07554652 0.3892090  0.1941027
## [6] {Dell Desktop}            => {HP Laptop}   0.04494154 0.3353566  0.1340112
## [7] {Lenovo Desktop Computer} => {HP Laptop}   0.04616167 0.3118132  0.1480427
```

```
## [8] {CYBERPOWER Gamer Desktop} => {iMac}      0.05673615 0.3084577  0.1839349
##     lift     count
## [1] 1.748851 486
## [2] 2.241200 472
## [3] 1.590762 537
## [4] 1.549932 578
## [5] 1.519599 743
## [6] 1.727728 442
## [7] 1.606434 454
## [8] 1.204320 558
```

```
inspect(sort(RulesName,by="support"))
```

```
##     lhs                          rhs            support    confidence coverage
## [1] {HP Laptop}               => {iMac}         0.07554652 0.3892090  0.1941027
## [2] {Lenovo Desktop Computer} => {iMac}         0.05876970 0.3969780  0.1480427
## [3] {CYBERPOWER Gamer Desktop} => {iMac}        0.05673615 0.3084577  0.1839349
## [4] {Dell Desktop}            => {iMac}         0.05460092 0.4074355  0.1340112
## [5] {ViewSonic Monitor}       => {iMac}         0.04941535 0.4479263  0.1103203
## [6] {ViewSonic Monitor}       => {HP Laptop}    0.04799187 0.4350230  0.1103203
## [7] {Lenovo Desktop Computer} => {HP Laptop}    0.04616167 0.3118132  0.1480427
## [8] {Dell Desktop}            => {HP Laptop}    0.04494154 0.3353566  0.1340112
##     lift     count
## [1] 1.519599 743
## [2] 1.549932 578
## [3] 1.204320 558
## [4] 1.590762 537
## [5] 1.748851 486
## [6] 2.241200 472
## [7] 1.606434 454
## [8] 1.727728 442
```

```
## We check that there are not redundant rules.
is.redundant(RulesName)
```
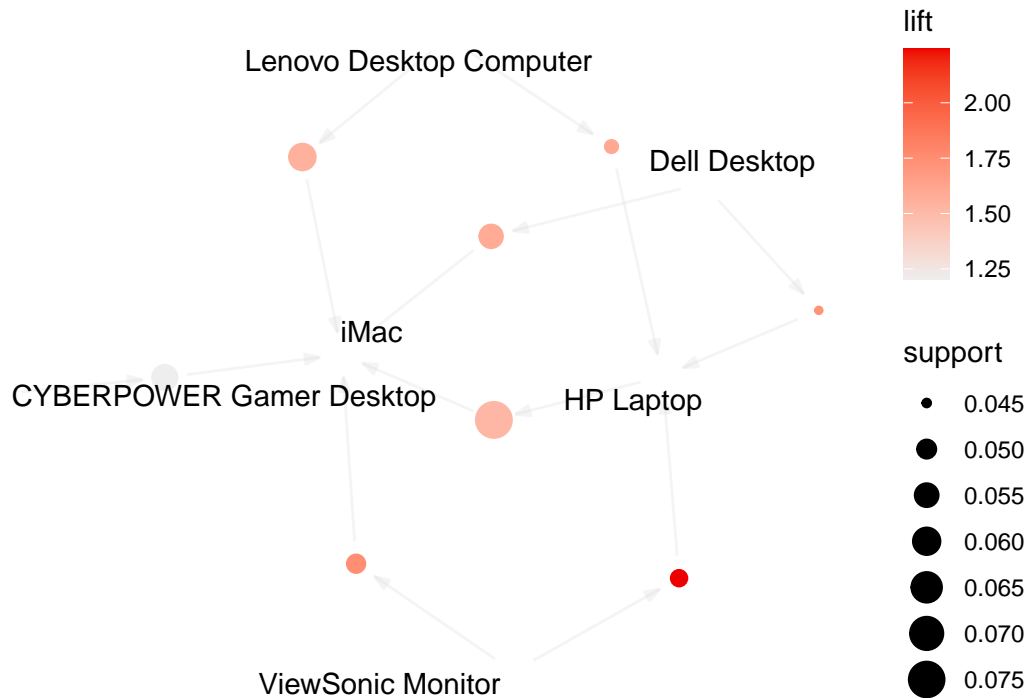
```
## [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

```
## We represent visually with a plot and a graph.
plot(RulesName)
```

Scatter plot for 8 rules
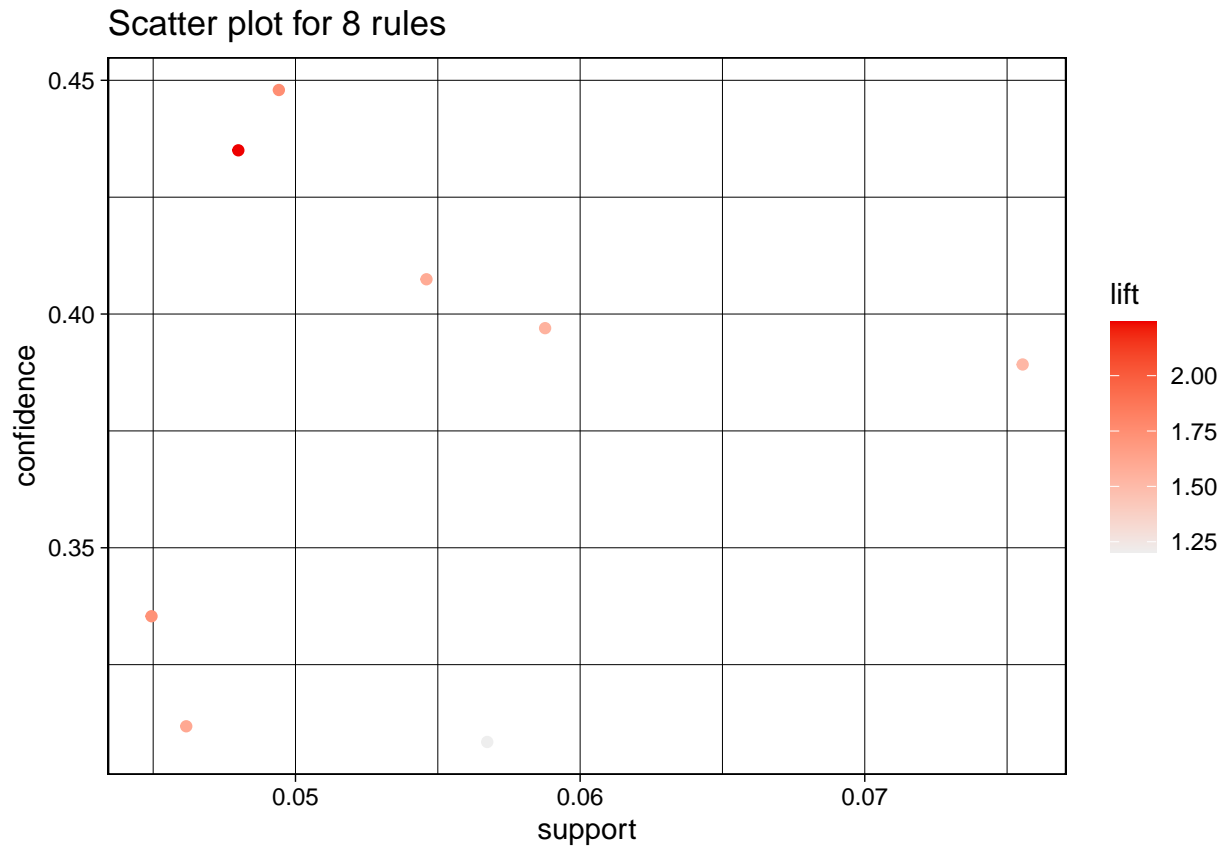

```r
plot(RulesName, method="graph", control=list(type="items"))
```

```
## Warning: Unknown control parameters: type

## 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
```

lift

2.00
1.75
1.50
1.25

support

0.045
0.050
0.055
0.060
0.065
0.070
0.075

Lenovo Desktop Computer
Dell Desktop
iMac
CYBERPOWER Gamer Desktop
HP Laptop
ViewSonic Monitor

```
## Conclusions and Insights:


# iMacs has the greater frequency, greater than 25%. As we can see from the
# A Priori Rules, it is usually bough with other products, like other computers
# or monitors. We may deduce that the customers are looking to complement
# its ecosystem by buying both Windows and MacOS computers.

# The trend described about make sense when we think on a company as a customer.
# For example, other  products (for example, from Apple) appear in the top
# frequencies but do not appear in the rules. Probably, because those products
# (macbooks or earpods) are more oriented to a final user, instead of a company
# buying hardware.

# Greater confidence values came with the ViewSonic Monitor (into iMac or HP
# Laptop) This makes a lot of sense. You cannot use a monitor without a computer.

# Lift is a metric that quantifies the relation between lhs and rhs.  All are
# greater than 1, which implies positive correlation. The greater values
# of the lift appear with the ViewSonic Monitor as lhs, which make sense according
# to the logic described above with the monitors.

# Best products to Blackwell can be precisely the ones in the top 10 solds, but
# specifically those that appear repeatedly in the rules, as you can play with
# those cross-selling patterns (iMac, HP Laptop, CYBERPOWER Gamer Desktop,
# Dell Desktop, ViewSonic Monitor,...).

# Notice that those rules, in general, are not very strong. As we said before,
# supports are above 4-8% and confidence is not greater than 45%, but in this
# context those numbers may be enough to reveal some trends.
```