d212_task3.R

sigsp

2022-09-27

```r
## Author: Stephen E. Porter
## Title: D212 Task 3 Market Basket Analysis
## Course: WGU D212: Data Mining II
## Instructor: Dr.Keiona Middleton

# Libraries
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5      v purrr   0.3.4
## v tibble  3.1.5      v dplyr   1.0.7
## v tidyr   1.1.4      v stringr 1.4.0
## v readr   2.0.2      v forcats 0.5.1
```

```
## Warning: package 'ggplot2' was built under R version 4.1.1
```

```
## Warning: package 'tibble' was built under R version 4.1.1
```

```
## Warning: package 'tidyr' was built under R version 4.1.1
```

```
## Warning: package 'readr' was built under R version 4.1.1
```

```
## Warning: package 'dplyr' was built under R version 4.1.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(arules)
```

```
## Warning: package 'arules' was built under R version 4.1.3
```

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

```
## Warning: package 'Matrix' was built under R version 4.1.3
```

```
##
## Attaching package: 'Matrix'
```

```
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
```

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

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

```
## The following objects are masked from 'package:base':
##
##      abbreviate, write
library(RColorBrewer)

## Warning: package 'RColorBrewer' was built under R version 4.1.3
# Import CSV as data frame
df <- read.csv(file = 'C:/WGU/D212 Data Mining II/teleco_market_basket.csv')

# Checking for nulls
sapply(df, function(x) sum(is.na(x)))

## Item01 Item02 Item03 Item04 Item05 Item06 Item07 Item08 Item09 Item10 Item11
##      0      0      0      0      0      0      0      0      0      0      0
## Item12 Item13 Item14 Item15 Item16 Item17 Item18 Item19 Item20
##      0      0      0      0      0      0      0      0      0
str(df)

## 'data.frame':    15002 obs. of  20 variables:
##  $ Item01: chr  "" "Logitech M510 Wireless mouse" "" "Apple Lightning to Digital AV Adapter" ...
##  $ Item02: chr  "" "HP 63 Ink" "" "TP-Link AC1750 Smart WiFi Router" ...
##  $ Item03: chr  "" "HP 65 ink" "" "Apple Pencil" ...
##  $ Item04: chr  "" "nonda USB C to USB Adapter" "" "" ...
##  $ Item05: chr  "" "10ft iPHone Charger Cable" "" "" ...
##  $ Item06: chr  "" "HP 902XL ink" "" "" ...
##  $ Item07: chr  "" "Creative Pebble 2.0 Speakers" "" "" ...
##  $ Item08: chr  "" "Cleaning Gel Universal Dust Cleaner" "" "" ...
##  $ Item09: chr  "" "Micro Center 32GB Memory card" "" "" ...
##  $ Item10: chr  "" "YUNSONG 3pack 6ft Nylon Lightning Cable" "" "" ...
##  $ Item11: chr  "" "TopMate C5 Laptop Cooler pad" "" "" ...
##  $ Item12: chr  "" "Apple USB-C Charger cable" "" "" ...
##  $ Item13: chr  "" "HyperX Cloud Stinger Headset" "" "" ...
##  $ Item14: chr  "" "TONOR USB Gaming Microphone" "" "" ...
##  $ Item15: chr  "" "Dust-Off Compressed Gas 2 pack" "" "" ...
##  $ Item16: chr  "" "3A USB Type C Cable 3 pack 6FT" "" "" ...
##  $ Item17: chr  "" "HOVAMP iPhone charger" "" "" ...
##  $ Item18: chr  "" "SanDisk Ultra 128GB card" "" "" ...
##  $ Item19: chr  "" "FEEL2NICE 5 pack 10ft Lighning cable" "" "" ...
##  $ Item20: chr  "" "FEIYOLD Blue light Blocking Glasses" "" "" ...
glimpse(df)

## Rows: 15,002
## Columns: 20
## $ Item01 <chr> "", "Logitech M510 Wireless mouse", "", "Apple Lightning to Dig~
## $ Item02 <chr> "", "HP 63 Ink", "", "TP-Link AC1750 Smart WiFi Router", "", ""~
## $ Item03 <chr> "", "HP 65 ink", "", "Apple Pencil", "", "", "", "", "Morea~
## $ Item04 <chr> "", "nonda USB C to USB Adapter", "", "", "", "", "", "", "~
## $ Item05 <chr> "", "10ft iPHone Charger Cable", "", "", "", "", "", "", "A~
## $ Item06 <chr> "", "HP 902XL ink", "", "", "", "", "", "", "", "", "", ""~
## $ Item07 <chr> "", "Creative Pebble 2.0 Speakers", "", "", "", "", "", ""~
## $ Item08 <chr> "", "Cleaning Gel Universal Dust Cleaner", "", "", "", "", ~
## $ Item09 <chr> "", "Micro Center 32GB Memory card", "", "", "", "", "", ""~
## $ Item10 <chr> "", "YUNSONG 3pack 6ft Nylon Lightning Cable", "", "", "", ~
```

```
## $ Item11 <chr> "", "TopMate C5 Laptop Cooler pad", "", "", "", "", "", "", "",~
## $ Item12 <chr> "", "Apple USB-C Charger cable", "", "", "", "", "", "", "", ""~
## $ Item13 <chr> "", "HyperX Cloud Stinger Headset", "", "", "", "", "", "", "",~
## $ Item14 <chr> "", "TONOR USB Gaming Microphone", "", "", "", "", "", "", "", ~
## $ Item15 <chr> "", "Dust-Off Compressed Gas 2 pack", "", "", "", "", "", "", "~
## $ Item16 <chr> "", "3A USB Type C Cable 3 pack 6FT", "", "", "", "", "", "", "~
## $ Item17 <chr> "", "HOVAMP iPhone charger", "", "", "", "", "", "", "", ""~
## $ Item18 <chr> "", "SanDisk Ultra 128GB card", "", "", "", "", "", "", "", "",~
## $ Item19 <chr> "", "FEEL2NICE 5 pack 10ft Lighning cable", "", "", "", "", "",~
## $ Item20 <chr> "", "FEIYOLD Blue light Blocking Glasses", "", "", "", "", "", ~
```

```r
# Every other row in the dataframe is blank. Dropping blank columns
toDelete <- seq(0, nrow(df), 2)
df <- df[toDelete,]
glimpse(df)
```

```
## Rows: 7,501
## Columns: 20
## $ Item01 <chr> "Logitech M510 Wireless mouse", "Apple Lightning to Digital AV ~
## $ Item02 <chr> "HP 63 Ink", "TP-Link AC1750 Smart WiFi Router", "", "HP 65 ink~
## $ Item03 <chr> "HP 65 ink", "Apple Pencil", "", "", "Moread HDMI to VGA Adapte~
## $ Item04 <chr> "nonda USB C to USB Adapter", "", "", "", "HP 62XL Tri-Color in~
## $ Item05 <chr> "10ft iPHone Charger Cable", "", "", "", "Apple USB-C Charger c~
## $ Item06 <chr> "HP 902XL ink", "", "", "", "", "", "", "", "", "", "", "", "",~
## $ Item07 <chr> "Creative Pebble 2.0 Speakers", "", "", "", "", "", "", "", ""~
## $ Item08 <chr> "Cleaning Gel Universal Dust Cleaner", "", "", "", "", "", "", ~
## $ Item09 <chr> "Micro Center 32GB Memory card", "", "", "", "", "", "", "", ""~
## $ Item10 <chr> "YUNSONG 3pack 6ft Nylon Lightning Cable", "", "", "", "", "", ~
## $ Item11 <chr> "TopMate C5 Laptop Cooler pad", "", "", "", "", "", "", "", ""~
## $ Item12 <chr> "Apple USB-C Charger cable", "", "", "", "", "", "", "", "", ""~
## $ Item13 <chr> "HyperX Cloud Stinger Headset", "", "", "", "", "", "", "", ""~
## $ Item14 <chr> "TONOR USB Gaming Microphone", "", "", "", "", "", "", "", "", ~
## $ Item15 <chr> "Dust-Off Compressed Gas 2 pack", "", "", "", "", "", "", "", "~
## $ Item16 <chr> "3A USB Type C Cable 3 pack 6FT", "", "", "", "", "", "", "", "~
## $ Item17 <chr> "HOVAMP iPhone charger", "", "", "", "", "", "", "", "", "", ""~
## $ Item18 <chr> "SanDisk Ultra 128GB card", "", "", "", "", "", "", "", "", ""~
## $ Item19 <chr> "FEEL2NICE 5 pack 10ft Lighning cable", "", "", "", "", "", "",~
## $ Item20 <chr> "FEIYOLD Blue light Blocking Glasses", "", "", "", "", "", "", ~
```

```r
head(df, n=1)
```

```
##                            Item01    Item02   Item03                    Item04
## 2 Logitech M510 Wireless mouse HP 63 Ink HP 65 ink nonda USB C to USB Adapter
##                         Item05      Item06                      Item07
## 2 10ft iPHone Charger Cable HP 902XL ink Creative Pebble 2.0 Speakers
##                                     Item08                        Item09
## 2 Cleaning Gel Universal Dust Cleaner Micro Center 32GB Memory card
##                                        Item10                      Item11
## 2 YUNSONG 3pack 6ft Nylon Lightning Cable TopMate C5 Laptop Cooler pad
##                        Item12                       Item13
## 2 Apple USB-C Charger cable HyperX Cloud Stinger Headset
##                           Item14                        Item15
## 2 TONOR USB Gaming Microphone Dust-Off Compressed Gas 2 pack
##                                Item16                Item17                  Item18
## 2 3A USB Type C Cable 3 pack 6FT HOVAMP iPhone charger SanDisk Ultra 128GB card
```

```
##                                  Item19                                Item20
## 2 FEEL2NICE 5 pack 10ft Lighning cable FEIYOLD Blue light Blocking Glasses
```

```
write.csv(df, 'C:/WGU/D212 Data Mining II/churn_mba.csv', row.names = FALSE)

tr <- read.transactions('C:/WGU/D212 Data Mining II/churn_mba.csv',
                        format="basket", sep=',')
```
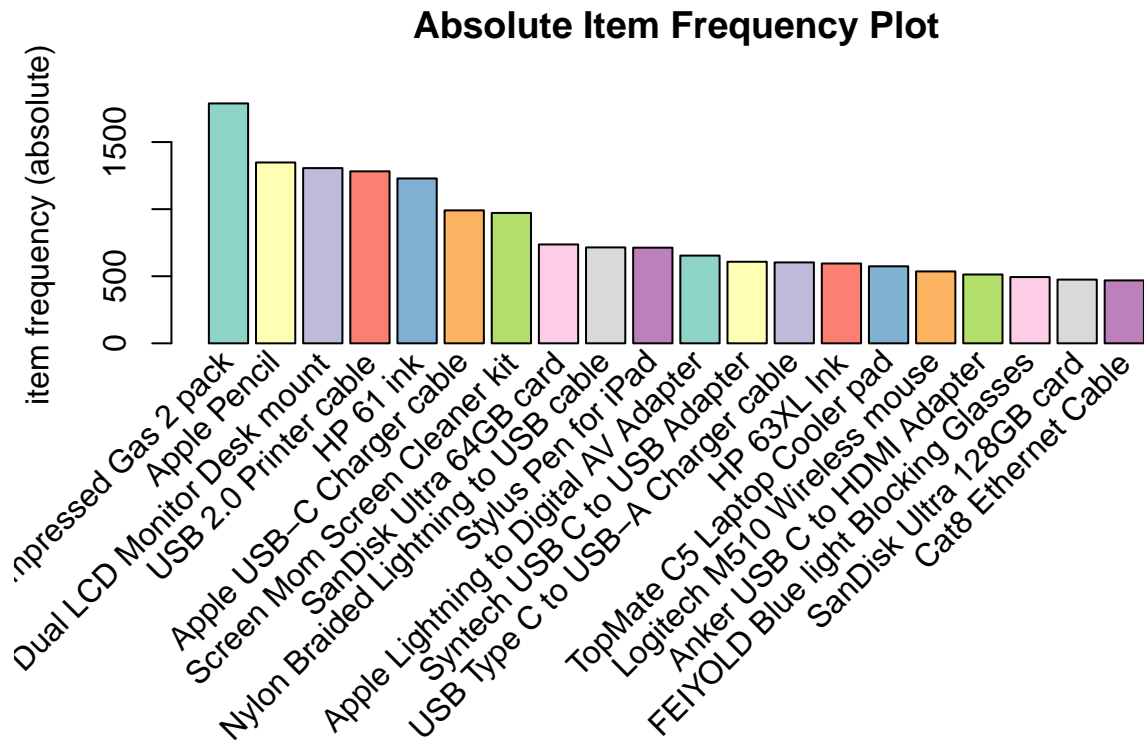
```
## Warning in asMethod(object): removing duplicated items in transactions
```

```
summary(tr)
```

```
## transactions as itemMatrix in sparse format with
##  7502 rows (elements/itemsets/transactions) and
##  139 columns (items) and a density of 0.02817282
##
## most frequent items:
##    Dust-Off Compressed Gas 2 pack                         Apple Pencil
##                              1788                                 1348
## VIVO Dual LCD Monitor Desk mount              USB 2.0 Printer cable
##                              1306                                 1282
##                        HP 61 ink                            (Other)
##                              1229                                22425
##
## element (itemset/transaction) length distribution:
## sizes
##     1    2    3    4    5    6    7    8    9   10   11   12   13   14   15   16
## 1754 1358 1044  816  667  493  391  324  259  139  102   67   40   22   17    4
##    18   19   20
##     1    2    2
##
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   2.000   3.000   3.916   5.000  20.000
##
## includes extended item information - examples:
##                              labels
## 1            10ft iPHone Charger Cable
## 2      10ft iPHone Charger Cable 2 Pack
## 3 3 pack Nylon Braided Lightning Cable
```

```
itemFrequencyPlot(tr, topN=20, type="absolute", col=brewer.pal(10,'Set3'),
                 main="Absolute Item Frequency Plot")
```

Absolute Item Frequency Plot


```r
all.rules <- apriori(tr, parameter=list(supp=0.01, conf=0.4))
```

```
## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support minlen
##         0.4    0.1    1 none FALSE            TRUE       5    0.01      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: 75
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[139 item(s), 7502 transaction(s)] done [0.00s].
## sorting and recoding items ... [75 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 done [0.00s].
## writing ... [18 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].
```

```r
summary(all.rules)
```

```
## set of 18 rules
```

```
## 
## rule length distribution (lhs + rhs):sizes
##  2  3
##  4 14
## 
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.000   3.000   3.000   2.778   3.000   3.000
## 
## summary of quality measures:
##     support            confidence         coverage            lift
##  Min.   :0.01013   Min.   :0.4013   Min.   :0.01999   Min.   :1.684
##  1st Qu.:0.01116   1st Qu.:0.4175   1st Qu.:0.02399   1st Qu.:1.764
##  Median :0.01373   Median :0.4355   Median :0.03266   Median :1.845
##  Mean   :0.01621   Mean   :0.4414   Mean   :0.03737   Mean   :1.888
##  3rd Qu.:0.01706   3rd Qu.:0.4561   3rd Qu.:0.04049   3rd Qu.:1.954
##  Max.   :0.04092   Max.   :0.5067   Max.   :0.09824   Max.   :2.395
##      count
##  Min.   : 76.00
##  1st Qu.: 83.75
##  Median :103.00
##  Mean   :121.61
##  3rd Qu.:128.00
##  Max.   :307.00
## 
## mining info:
##  data ntransactions support confidence
##    tr           7502    0.01        0.4
##                                                     call
##  apriori(data = tr, parameter = list(supp = 0.01, conf = 0.4))
```

```
inspect(head(all.rules, by="lift", n=18))
```

```
##       lhs                                  rhs                                       support confide
## [1]  {Dust-Off Compressed Gas 2 pack,
##       SanDisk Ultra 64GB card}         => {VIVO Dual LCD Monitor Desk mount} 0.01706212  0.4169
## [2]  {Apple Pencil,
##       SanDisk Ultra 64GB card}         => {Dust-Off Compressed Gas 2 pack}   0.01013063  0.5060
## [3]  {SanDisk Ultra 64GB card,
##       Screen Mom Screen Cleaner kit}   => {Dust-Off Compressed Gas 2 pack}   0.01106372  0.5030
## [4]  {HP 61 ink,
##       SanDisk Ultra 64GB card}         => {Dust-Off Compressed Gas 2 pack}   0.01093042  0.4739
## [5]  {Nylon Braided Lightning to USB cable,
##       Screen Mom Screen Cleaner kit}   => {Dust-Off Compressed Gas 2 pack}   0.01106372  0.4689
## [6]  {10ft iPHone Charger Cable 2 Pack} => {Dust-Off Compressed Gas 2 pack}  0.02306052  0.4564
## [7]  {Stylus Pen for iPad,
##       VIVO Dual LCD Monitor Desk mount} => {Dust-Off Compressed Gas 2 pack}  0.01146361  0.4550
## [8]  {FEIYOLD Blue light Blocking Glasses,
##       VIVO Dual LCD Monitor Desk mount} => {Dust-Off Compressed Gas 2 pack}  0.01026393  0.4470
## [9]  {Screen Mom Screen Cleaner kit,
##       VIVO Dual LCD Monitor Desk mount} => {Dust-Off Compressed Gas 2 pack}  0.01572914  0.4430
## [10] {HP 61 ink,
##       Screen Mom Screen Cleaner kit}   => {Dust-Off Compressed Gas 2 pack}   0.01399627  0.4350
## [11] {SanDisk Ultra 64GB card,
##       VIVO Dual LCD Monitor Desk mount} => {Dust-Off Compressed Gas 2 pack}  0.01706212  0.4353
## [12] {Nylon Braided Lightning to USB cable,
```

```
##          VIVO Dual LCD Monitor Desk mount}      => {Dust-Off Compressed Gas 2 pack}   0.01199680   0.4300
## [13] {Apple Pencil,
##          Screen Mom Screen Cleaner kit}         => {Dust-Off Compressed Gas 2 pack}   0.01306318   0.4242
## [14] {FEIYOLD Blue light Blocking Glasses}      => {Dust-Off Compressed Gas 2 pack}   0.02759264   0.4190
## [15] {SanDisk Ultra 64GB card}                  => {Dust-Off Compressed Gas 2 pack}   0.04092242   0.4165
## [16] {Apple Pencil,
##          HP 61 ink}                             => {Dust-Off Compressed Gas 2 pack}   0.01346308   0.4050
## [17] {HP 61 ink,
##          VIVO Dual LCD Monitor Desk mount}      => {Dust-Off Compressed Gas 2 pack}   0.01586244   0.4047
## [18] {3A USB Type C Cable 3 pack 6FT}           => {Dust-Off Compressed Gas 2 pack}   0.01706212   0.4011
```