D212: Data Mining II

Performance Assessment

Task 3

Logan Rosemeyer

Western Governor's University

D212: Data Mining II

Dr. Keiona Middleton

March 12, 2023

Contents

A1: Proposal of Question

My question is: What are the top 3 groups of products that are bought together?

A2: Defined Goal

The goal of the data analysis is to determine the top 3 groups of products that are bought together.

B1: Explanation of Market Basket

Market basket analysis is a technique used to better understand customer purchasing patterns. Market basket analysis works by looking for combinations of items that occur together frequently in transactions (Li, 2017).

B2: Transaction Example

One example of a transaction is someone buying Logitech M510 Wireless mouse, HP 63 Ink, HP 65 ink, nonda USB C to USB Adapter, 10ft iPhone Charger Cable, HP 902XL ink, Creative Pebble 2.0 Speakers, Cleaning Gel Universal Dust Cleaner, Micro Center 32GB Memory card, YUNSONG 3pack 6ft Nylon Lightning Cable, TopMate C5 Laptop Cooler pad, Apple USB-C Charger cable, HyperX Cloud Stinger Headset, TONOR USB Gaming Microphone, Dust-Off Compressed Gas 2 pack, 3A USB Type C Cable 3 pack 6FT, HOVAMP iPhone charger, SanDisk Ultra 128GB card, FEEL2NICE 5 pack 10ft Lightning cable, FEIYOLD Blue light Blocking Glasses.

B3: Market Basket Assumption

One assumption of market basket analysis is that any subset of a frequent itemset must also be frequent. For example if itemset {x, y, z} is frequent, then itemset {x, z} must also be frequent. (Singh Chauhan, 2019)

C1: Transforming the Dataset

A copy of the cleaned dataset has been included in the submission.

C2: Code Execution

Here is the code used to generate association rules with the Apriori algorithm

```
library(arules)
basket <- as(list_data, "transactions")
basket <- as(basket, "matrix")
str(basket)
dim(basket)
basket

arules <- apriori(basket, control=list(verbose=F), parameter=list(supp=0.008, conf=0.4, minlen=2))

redundant_r <- is.redundant(arules)
refined_arules <- arules[!redundant_r]
inspect(head(sort(refined_arules, by="lift", decreasing=T), 20))

summary(refined_arules)
summary(head(sort(refined_arules, by="lift", decreasing=T), 3))
```

Here is the execution of the rules.

```
> inspect(head(sort(refined_arules, by="lift", decreasing=T), 20))
     lhs                                                rhs                                  support confidence   coverage     lift count
[1]  {Nylon Braided Lightning to USB cable,
      SanDisk Ultra 64GB card}                       => {VIVO Dual LCD Monitor Desk mount} 0.008665511  0.5118110 0.01693108 2.939582    65
[2]  {Apple Pencil,
      SanDisk Ultra 64GB card}                       => {VIVO Dual LCD Monitor Desk mount} 0.008932142  0.4466667 0.01999733 2.565426    67
[3]  {SanDisk Ultra 64GB card,
      Screen Mom Screen Cleaner kit}                 => {VIVO Dual LCD Monitor Desk mount} 0.009732036  0.4424242 0.02199707 2.541060    73
[4]  {Dust-Off Compressed Gas 2 pack,
      SanDisk Ultra 64GB card}                       => {VIVO Dual LCD Monitor Desk mount} 0.017064391  0.4169381 0.04092788 2.394681   128
[5]  {10ft iPHone Charger Cable 2 Pack,
      Screen Mom Screen Cleaner kit}                 => {Dust-Off Compressed Gas 2 pack}   0.008532196  0.5614035 0.01519797 2.355194    64
[6]  {Apple Lightning to Digital AV Adapter,
      USB 2.0 Printer cable}                         => {Apple Pencil}                     0.009065458  0.4121212 0.02199707 2.293265    68
[7]  {Nylon Braided Lightning to USB cable,
      SanDisk Ultra 64GB card}                       => {Dust-Off Compressed Gas 2 pack}   0.009198773  0.5433071 0.01693108 2.279277    69
[8]  {Apple Pencil,
      SanDisk Ultra 64GB card}                       => {Dust-Off Compressed Gas 2 pack}   0.010131982  0.5066667 0.01999733 2.125563    76
[9]  {FEIYOLD Blue light Blocking Glasses,
      HP 61 ink}                                     => {Dust-Off Compressed Gas 2 pack}   0.008265565  0.5040650 0.01639781 2.114649    62
[10] {SanDisk Ultra 64GB card,
      Screen Mom Screen Cleaner kit}                 => {Dust-Off Compressed Gas 2 pack}   0.011065191  0.5030303 0.02199707 2.110308    83
[11] {FEIYOLD Blue light Blocking Glasses,
      Screen Mom Screen Cleaner kit}                 => {Dust-Off Compressed Gas 2 pack}   0.008532196  0.5000000 0.01706439 2.097595    64
[12] {HP 61 ink,
      SanDisk Ultra 64GB card}                       => {Dust-Off Compressed Gas 2 pack}   0.010931876  0.4739884 0.02306359 1.988472    82
[13] {HP 61 ink,
      Stylus Pen for iPad}                           => {Dust-Off Compressed Gas 2 pack}   0.009332089  0.4697987 0.01986402 1.970895    70
[14] {Nylon Braided Lightning to USB cable,
      Screen Mom Screen Cleaner kit}                 => {Dust-Off Compressed Gas 2 pack}   0.011065191  0.4689266 0.02359685 1.967236    83
[15] {10ft iPHone Charger Cable 2 Pack}              => {Dust-Off Compressed Gas 2 pack}   0.023063592  0.4564644 0.05052660 1.914955   173
[16] {Stylus Pen for iPad,
      VIVO Dual LCD Monitor Desk mount}              => {Dust-Off Compressed Gas 2 pack}   0.011465138  0.4550265 0.02519664 1.908923    86
[17] {FEIYOLD Blue light Blocking Glasses,
      VIVO Dual LCD Monitor Desk mount}              => {Dust-Off Compressed Gas 2 pack}   0.010265298  0.4476744 0.02293028 1.878079    77
[18] {Anker USB C to HDMI Adapter,
      VIVO Dual LCD Monitor Desk mount}              => {Dust-Off Compressed Gas 2 pack}   0.009332089  0.4458599 0.02093054 1.870467    70
[19] {Screen Mom Screen Cleaner kit,
      VIVO Dual LCD Monitor Desk mount}              => {Dust-Off Compressed Gas 2 pack}   0.015731236  0.4436090 0.03546194 1.861024   118
[20] {HP 61 ink,
      Screen Mom Screen Cleaner kit}                 => {Dust-Off Compressed Gas 2 pack}   0.013998134  0.4356846 0.03212905 1.827780   105
```

Here is the summary of the rules.

```
> summary(refined_arules)
set of 31 rules

rule length distribution (lhs + rhs):sizes
 2  3
 4 27

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  2.000   3.000   3.000   2.871   3.000   3.000

summary of quality measures:
    support           confidence        coverage            lift           count
 Min.   :0.008266   Min.   :0.4013   Min.   :0.01520   Min.   :1.683   Min.   : 62.0
 1st Qu.:0.009132   1st Qu.:0.4181   1st Qu.:0.02046   1st Qu.:1.780   1st Qu.: 68.5
 Median :0.010932   Median :0.4436   Median :0.02293   Median :1.909   Median : 82.0
 Mean   :0.013185   Mean   :0.4525   Mean   :0.02987   Mean   :2.007   Mean   : 98.9
 3rd Qu.:0.014865   3rd Qu.:0.4719   3rd Qu.:0.03433   3rd Qu.:2.120   3rd Qu.:111.5
 Max.   :0.040928   Max.   :0.5614   Max.   :0.09825   Max.   :2.940   Max.   :307.0

mining info:
   data ntransactions support confidence
 basket          7501   0.008        0.4
                                                                                                                call
 apriori(data = basket, parameter = list(supp = 0.008, conf = 0.4, minlen = 2), control = list(verbose = F))
```

C3: Association Rules Table

Here is a screenshot of values for the support, lift, and confidence of the association rules table.

```
> inspect(head(sort(refined_arules, by="lift", decreasing=T), 20))
     lhs                                                rhs                                 support confidence   coverage     lift count
[1]  {Nylon Braided Lightning to USB cable,
      SanDisk Ultra 64GB card}                       => {VIVO Dual LCD Monitor Desk mount} 0.008665511  0.5118110 0.01693108 2.939582    65
[2]  {Apple Pencil,
      SanDisk Ultra 64GB card}                       => {VIVO Dual LCD Monitor Desk mount} 0.008932142  0.4466667 0.01999733 2.565426    67
[3]  {SanDisk Ultra 64GB card,
      Screen Mom Screen Cleaner kit}                 => {VIVO Dual LCD Monitor Desk mount} 0.009732036  0.4424242 0.02199707 2.541060    73
[4]  {Dust-Off Compressed Gas 2 pack,
      SanDisk Ultra 64GB card}                       => {VIVO Dual LCD Monitor Desk mount} 0.017064391  0.4169381 0.04092788 2.394681   128
[5]  {10ft iPHone Charger Cable 2 Pack,
      Screen Mom Screen Cleaner kit}                 => {Dust-Off Compressed Gas 2 pack}   0.008532196  0.5614035 0.01519797 2.355194    64
[6]  {Apple Lightning to Digital AV Adapter,
      USB 2.0 Printer cable}                         => {Apple Pencil}                     0.009065458  0.4121212 0.02199707 2.293265    68
[7]  {Nylon Braided Lightning to USB cable,
      SanDisk Ultra 64GB card}                       => {Dust-Off Compressed Gas 2 pack}   0.009198773  0.5433071 0.01693108 2.279277    69
[8]  {Apple Pencil,
      SanDisk Ultra 64GB card}                       => {Dust-Off Compressed Gas 2 pack}   0.010131982  0.5066667 0.01999733 2.125563    76
[9]  {FEIYOLD Blue light Blocking Glasses,
      HP 61 ink}                                     => {Dust-Off Compressed Gas 2 pack}   0.008265565  0.5040650 0.01639781 2.114649    62
[10] {SanDisk Ultra 64GB card,
      Screen Mom Screen Cleaner kit}                 => {Dust-Off Compressed Gas 2 pack}   0.011065191  0.5030303 0.02199707 2.110308    83
[11] {FEIYOLD Blue light Blocking Glasses,
      Screen Mom Screen Cleaner kit}                 => {Dust-Off Compressed Gas 2 pack}   0.008532196  0.5000000 0.01706439 2.097595    64
[12] {HP 61 ink,
      SanDisk Ultra 64GB card}                       => {Dust-Off Compressed Gas 2 pack}   0.010931876  0.4739884 0.02306359 1.988472    82
[13] {HP 61 ink,
      Stylus Pen for iPad}                           => {Dust-Off Compressed Gas 2 pack}   0.009332089  0.4697987 0.01986402 1.970895    70
[14] {Nylon Braided Lightning to USB cable,
      Screen Mom Screen Cleaner kit}                 => {Dust-Off Compressed Gas 2 pack}   0.011065191  0.4689266 0.02359685 1.967236    83
[15] {10ft iPHone Charger Cable 2 Pack}              => {Dust-Off Compressed Gas 2 pack}   0.023063592  0.4564644 0.05052660 1.914955   173
[16] {Stylus Pen for iPad,
      VIVO Dual LCD Monitor Desk mount}              => {Dust-Off Compressed Gas 2 pack}   0.011465138  0.4550265 0.02519664 1.908923    86
[17] {FEIYOLD Blue light Blocking Glasses,
      VIVO Dual LCD Monitor Desk mount}              => {Dust-Off Compressed Gas 2 pack}   0.010265298  0.4476744 0.02293028 1.878079    77
[18] {Anker USB C to HDMI Adapter,
      VIVO Dual LCD Monitor Desk mount}              => {Dust-Off Compressed Gas 2 pack}   0.009332089  0.4458599 0.02093054 1.870467    70
[19] {Screen Mom Screen Cleaner kit,
      VIVO Dual LCD Monitor Desk mount}              => {Dust-Off Compressed Gas 2 pack}   0.015731236  0.4436090 0.03546194 1.861024   118
[20] {HP 61 ink,
      Screen Mom Screen Cleaner kit}                 => {Dust-Off Compressed Gas 2 pack}   0.013998134  0.4356846 0.03212905 1.827780   105
```

As well as a summary of the rules.

```
> summary(refined_arules)
set of 31 rules

rule length distribution (lhs + rhs):sizes
 2  3
 4 27

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  2.000   3.000   3.000   2.871   3.000   3.000

summary of quality measures:
    support           confidence        coverage            lift           count
 Min.   :0.008266   Min.   :0.4013   Min.   :0.01520   Min.   :1.683   Min.   : 62.0
 1st Qu.:0.009132   1st Qu.:0.4181   1st Qu.:0.02046   1st Qu.:1.780   1st Qu.: 68.5
 Median :0.010932   Median :0.4436   Median :0.02293   Median :1.909   Median : 82.0
 Mean   :0.013185   Mean   :0.4525   Mean   :0.02987   Mean   :2.007   Mean   : 98.9
 3rd Qu.:0.014865   3rd Qu.:0.4719   3rd Qu.:0.03433   3rd Qu.:2.120   3rd Qu.:111.5
 Max.   :0.040928   Max.   :0.5614   Max.   :0.09825   Max.   :2.940   Max.   :307.0

mining info:
   data ntransactions support confidence
 basket          7501   0.008        0.4
                                                                                                                                call
 apriori(data = basket, parameter = list(supp = 0.008, conf = 0.4, minlen = 2), control = list(verbose = F))
```

C4: Top Three Rules

Here are the top three rules by lift.

```
> inspect(head(sort(refined_arules, by="lift", decreasing=T), 20))
     lhs                                          rhs                                  support confidence   coverage     lift count
[1]  {Nylon Braided Lightning to USB cable,
      SanDisk Ultra 64GB card}                 => {VIVO Dual LCD Monitor Desk mount} 0.008665511  0.5118110 0.01693108 2.939582    65
[2]  {Apple Pencil,
      SanDisk Ultra 64GB card}                 => {VIVO Dual LCD Monitor Desk mount} 0.008932142  0.4466667 0.01999733 2.565426    67
[3]  {SanDisk Ultra 64GB card,
      Screen Mom Screen Cleaner kit}           => {VIVO Dual LCD Monitor Desk mount} 0.009732036  0.4424242 0.02199707 2.541060    73
```

And here is a summary of the top three rules.

```
> summary(head(sort(refined_arules, by="lift", decreasing=T), 3))
set of 3 rules

rule length distribution (lhs + rhs):sizes
3
3

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

summary of quality measures:
    support           confidence        coverage            lift           count
 Min.   :0.008666   Min.   :0.4424   Min.   :0.01693   Min.   :2.541   Min.   :65.00
 1st Qu.:0.008799   1st Qu.:0.4445   1st Qu.:0.01846   1st Qu.:2.553   1st Qu.:66.00
 Median :0.008932   Median :0.4467   Median :0.02000   Median :2.565   Median :67.00
 Mean   :0.009110   Mean   :0.4670   Mean   :0.01964   Mean   :2.682   Mean   :68.33
 3rd Qu.:0.009332   3rd Qu.:0.4792   3rd Qu.:0.02100   3rd Qu.:2.753   3rd Qu.:70.00
 Max.   :0.009732   Max.   :0.5118   Max.   :0.02200   Max.   :2.940   Max.   :73.00

mining info:
   data ntransactions support confidence
 basket          7501   0.008        0.4
                                                                                                                call
 apriori(data = basket, parameter = list(supp = 0.008, conf = 0.4, minlen = 2), control = list(verbose = F))
```

D1: Significance of Support, Lift, and Confidence Summary

Overall, I would say the support numbers are decent. The mean support of our 31 rules is 0.013185, which means that the items listed in that rule are purchased in 1.3% of transactions. The max support is 4.09%. Confidence numbers are also decent, with a mean of 0.4525 and a max of 0.5614. This means that on average, if a transaction contains the items on the left, then the item on the right also be included in the transaction 45.25% of the time. The lift numbers are very good. A lift of more than 1 suggests that the presence of the items on the left hand side increases the probability that the item on the right hand side will also occur in the transaction. The minimum lift is 1.683 and the mean lift is over 2. This tells me that my lift numbers are very good.

D2: Practical Significance of Findings

I think my findings are very significant for the organization. The top four rules by lift all have VIVO Dual LCD Monitor Desk mount. This is a pretty good indication that this particular product is bought with multiple other products. The same can be said with the Dust-Off Compressed Gas 2 pack. While the rules don't have as high of a lift, there are more rules that include the compressed gas on the right hand side.

D3: Course of Action

I believe that the next course of action for the organization is to create deals for consumers that include the compressed gas and the monitor desk mount. It is clear that these two items are most frequently bought with other items. The organization should next look at which groups of items with be most profitable when included in a bundle or a sale together, and put that sale into place.

E: Panopto Recording

The Panopto recording can be found using the link below and has also been included in the submission of this task.

https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=201292b1-f1ec-438e-8c9b-afc5013e1fae

F: Sources for Third-Party Code

Apriori: Mining Associations with the Apriori Algorithm. RDocumentation. Retrieved March 15, 2023. https://www.rdocumentation.org/packages/arules/versions/1.7-5/topics/apriori

G: Sources

Li, Susan. *A Gentle Introduction on Market Basket Analysis – Association Rules*. Towards Data Science. September 24, 2017. Retrieved March 13, 2023. https://towardsdatascience.com/a-gentle-introduction-on-market-basket-analysis-association-rules-fa4b986a40ce

Singh Chauhan, Nagesh. *Market Basket Analysis: A Tutorial.* KDnuggets. December 24, 2019. Retrieved March 13, 2023. https://www.kdnuggets.com/2019/12/market-basket-analysis.html