Q4: Association Rules for Cosmetics

*Fall, 2022*

G Park

2022-12-04

# Question 4: Cosmetics Purchase (20 points)

The file Cosmetics.csv contains 1000 transaction information about cosmetics (1: purchased; 0: no purchased)

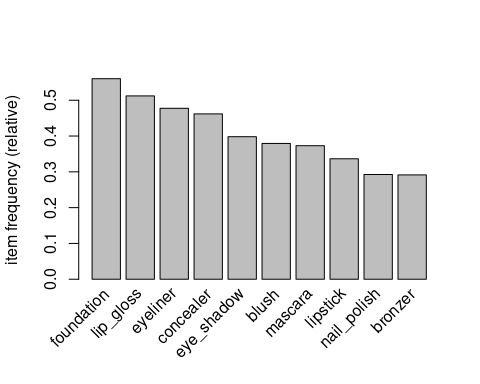
## Draw an item frequency plot and answer which two are the most popular items. (2 points)

The 2 most popular items are foundation and lip gross.

I build an item frequency plot to show the most popular items.

## transactions as itemMatrix in sparse format with  
## 957 rows (elements/itemsets/transactions) and  
## 14 columns (items) and a density of 0.33   
##   
## most frequent items:  
## foundation lip\_gloss eyeliner concealer eye\_shadow (Other)   
## 536 490 457 442 381 2080   
##   
## element (itemset/transaction) length distribution:  
## sizes  
## 1 2 3 4 5 6 7 8 9 10 11 12 13   
## 67 116 166 158 156 107 79 53 17 18 16 3 1   
##   
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.0 3.0 4.0 4.6 6.0 13.0   
##   
## includes extended item information - examples:  
## labels  
## 1 bag  
## 2 blush  
## 3 bronzer  
##   
## includes extended transaction information - examples:  
## transactionID  
## 1 c1  
## 2 c10  
## 3 c100

## bag blush bronzer brushes concealer   
## 0.056 0.379 0.292 0.156 0.462   
## eye\_shadow eyebrow\_pencils eyeliner foundation lip\_gloss   
## 0.398 0.044 0.478 0.560 0.512   
## lip\_liner lipstick mascara nail\_polish   
## 0.245 0.336 0.373 0.293



Most Frequently Bought Items

## Build an association rule model and set the support value as 0.01 and the confidence value as 0.1. Based on your association rule results, show the first eight rules and sort by their lift values and ensure to interpret your results. (9 point)

## Apriori  
##   
## Parameter specification:  
## confidence minval smax arem aval originalSupport maxtime support minlen  
## 0.8 0.1 1 none FALSE TRUE 5 0.01 2  
## 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: 9   
##   
## set item appearances ...[0 item(s)] done [0.00s].  
## set transactions ...[14 item(s), 957 transaction(s)] done [0.00s].  
## sorting and recoding items ... [14 item(s)] done [0.00s].  
## creating transaction tree ... done [0.00s].  
## checking subsets of size 1 2 3 4 5 6 7 8 9 10 done [0.00s].  
## writing ... [4433 rule(s)] done [0.00s].  
## creating S4 object ... done [0.00s].

## set of 4433 rules  
##   
## rule length distribution (lhs + rhs):sizes  
## 2 3 4 5 6 7 8 9 10   
## 3 53 338 862 1324 1149 549 140 15   
##   
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.0 5.0 6.0 6.2 7.0 10.0   
##   
## summary of quality measures:  
## support confidence coverage lift count   
## Min. :0.01 Min. :0.80 Min. :0.01 Min. :1.4 Min. : 10   
## 1st Qu.:0.01 1st Qu.:0.88 1st Qu.:0.01 1st Qu.:2.1 1st Qu.: 12   
## Median :0.02 Median :0.94 Median :0.02 Median :2.5 Median : 15   
## Mean :0.02 Mean :0.93 Mean :0.02 Mean :2.7 Mean : 21   
## 3rd Qu.:0.02 3rd Qu.:1.00 3rd Qu.:0.03 3rd Qu.:3.2 3rd Qu.: 23   
## Max. :0.34 Max. :1.00 Max. :0.40 Max. :6.2 Max. :321   
##   
## mining info:  
## data ntransactions support confidence  
## trans 957 0.01 0.8  
## call  
## apriori(data = trans, parameter = list(support = 0.01, confidence = 0.8, minlen = 2))

| rules | support | confidence | coverage | lift | count |
| --- | --- | --- | --- | --- | --- |
| {bronzer,eyeliner,lip\_liner,mascara,nail\_polish} => {brushes} | 0.026 | 0.96 | 0.027 | 6.2 | 25 |
| {bronzer,concealer,eyeliner,lip\_liner,mascara,nail\_polish} => {brushes} | 0.026 | 0.96 | 0.027 | 6.2 | 25 |
| {bronzer,eye\_shadow,eyeliner,lip\_liner,nail\_polish} => {brushes} | 0.024 | 0.96 | 0.025 | 6.2 | 23 |
| {bronzer,concealer,eye\_shadow,eyeliner,lip\_liner,nail\_polish} => {brushes} | 0.023 | 0.96 | 0.024 | 6.1 | 22 |
| {bronzer,eye\_shadow,eyeliner,lip\_liner,mascara,nail\_polish} => {brushes} | 0.022 | 0.95 | 0.023 | 6.1 | 21 |
| {bronzer,concealer,eye\_shadow,eyeliner,lip\_liner,mascara,nail\_polish} => {brushes} | 0.022 | 0.95 | 0.023 | 6.1 | 21 |
| {blush,bronzer,eyeliner,lip\_liner,mascara,nail\_polish} => {brushes} | 0.018 | 0.94 | 0.019 | 6.1 | 17 |
| {blush,bronzer,concealer,eyeliner,lip\_liner,mascara,nail\_polish} => {brushes} | 0.018 | 0.94 | 0.019 | 6.1 | 17 |

I use the first rule to illustrate the meaning of these rules.

**Support:** First bronzer,eyeliner,lip\_liner,mascara,nail\_polish => brushes constitute 3% of sales.

**Confidence:** When a customer buys bronzer, eyeliner, lip\_liner, mascara, nail\_polish, there is a 96% chance of buying brushes brushes.

**Lift:** Having bronzer, eyeliner, lip\_liner, mascara, nail\_polish in the shopping basket raises the probability that a customer will buy brushes by a factor of 6.2.

## Now, in your 4.2 results, you can see “support”, “confidence”, and “lift”. Please explain the meaning of support, confidence, and lift and discuss their implications in the setting of association rules. (9 point)

Support:

Support is the percentage of groups that contain all of the items listed in the association rule. In the first rule above, the support is 0.03, meaning that 3% of the transactions contained the items listed on the right (bronzer,eyeliner,lip\_liner,mascara,nail\_polish => brushes).

Implication of support: The support shows the volume of transactions that a product or groups of products as a percent of total transactions. When combined with other confidence and lift, we can focus on selling products that generate biggers sales. In other words, support allows us to filter for meaningful rules. It would not be very useful to have a confidence of 99% but that constitutes only0.001% of sales.

Confidence:

Confidence is the proportion of times that a customer buys item X given that she/he buys item Y. In our first rule, given that the customer buys bronzer,eyeliner,lip\_liner,mascara,nail\_polish => brushes then that customer bought brushes 96% of the time.

The implication of confidence: We can target customers buying bronzer,eyeliner,lip\_liner,mascara,nail\_polish ith bushes as they have a 96% chance of buying. Alternatively, in a store, these products could be placed close to each other.

Lift:

The lift value captures rule importance. Usually, its the confidence of a rule divided by the support of a product. It is the rise in probability of the purchase of product Y once we know that product X is in the basket. Again, the implication is that it helps managers decide on product placements in stores.