

# Machine Learning - Association Rules

*Bryana Benson*

*Conner Bryan*

Use Association Rules to do a market basket analysis for Superstore Sales. As a result of your analysis, what advice would you give to Superstore Sales?

## Libraries

```
library(readxl)
library(readr)
library(arules)
library(arulesViz)
```

## Data Import

```
#import Superstore Sales from directory
sales <- readxl::read_excel("sales.xls")
#Preview first few rows of data
head(sales)
```

```
## # A tibble: 6 x 21
##   `Row ID` `Order ID` `Order Date`      `Ship Date`      `Ship Mode`
##     <dbl> <chr>      <dtm>          <dtm>          <chr>
## 1       1 CA-2016-1~ 2016-11-08 00:00:00 2016-11-11 00:00:00 Second Cla~
## 2       2 CA-2016-1~ 2016-11-08 00:00:00 2016-11-11 00:00:00 Second Cla~
## 3       3 CA-2016-1~ 2016-06-12 00:00:00 2016-06-16 00:00:00 Second Cla~
## 4       4 US-2015-1~ 2015-10-11 00:00:00 2015-10-18 00:00:00 Standard C~
## 5       5 US-2015-1~ 2015-10-11 00:00:00 2015-10-18 00:00:00 Standard C~
## 6       6 CA-2014-1~ 2014-06-09 00:00:00 2014-06-14 00:00:00 Standard C~
## # ... with 16 more variables: `Customer ID` <chr>, `Customer Name` <chr>,
## #   Segment <chr>, Country <chr>, City <chr>, State <chr>, `Postal
## #   Code` <dbl>, Region <chr>, `Product ID` <chr>, Category <chr>,
## #   `Sub-Category` <chr>, `Product Name` <chr>, Sales <dbl>,
## #   Quantity <dbl>, Discount <dbl>, Profit <dbl>
```

## File Prep

Create a transaction file with the Order ID and Product Name columns.

```
#Extract Order ID and Product Name columns
t <- sales[,c(2,17)]

#Create a CSV file called "transactions"
readr::write_csv(t, "transactions.csv")

#Convert to transaction file
trans <- arules::read.transactions("transactions.csv", sep = ",", format = "single", cols = c(1,2) )

summary(trans)
```

```
## transactions as itemMatrix in sparse format with
## 2977 rows (elements/itemsets/transactions) and
```

```

## 1801 columns (items) and a density of 0.001093895
##
## most frequent items:
##           Staple envelope           Staples
##                30                29
##           Easy-staple paper       Staple remover
##                24                13
## KI Adjustable-Height Table       (Other)
##                12                5757
##
## element (itemset/transaction) length distribution:
## sizes
##      1      2      3      4      5      6      7      8      9     10     11     14
## 1511  746  353  196   88   35   33   7    4    1    2    1
##
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.00    1.00    1.00    1.97    2.00   14.00
##
## includes extended item information - examples:
##                                     labels
## 1 "While you Were Out" Message Book, One Form per Page
## 2           #10- 4 1/8" x 9 1/2" Recycled Envelopes
## 3           #10- 4 1/8" x 9 1/2" Security-Tint Envelopes
##
## includes extended transaction information - examples:
##      transactionID
## 1 CA-2014-100090
## 2 CA-2014-100293
## 3 CA-2014-100391

```

```

#Vector of transation sizes
arules::size(head(trans))

```

```

## [1] 2 1 1 4 2 1

```

```

#View first few rows of data in list format
arules::LIST(head(trans))

```

```

## $`CA-2014-100090`
## [1] "Hon 2111 Invitation Series Corner Table"
## [2] "Wilson Jones Ledger-Size, Piano-Hinge Binder, 2\"", Blue"
##
## $`CA-2014-100293`
## [1] "Xerox 1887"
##
## $`CA-2014-100391`
## [1] "Strathmore Photo Frame Cards"
##
## $`CA-2014-100678`
## [1] "Cameo Buff Policy Envelopes"
## [2] "DMI Arturo Collection Mission-style Design Wood Chair"
## [3] "Kensington Expert Mouse Optical USB Trackball for PC or Mac"
## [4] "Prang Dustless Chalk Sticks"
##
## $`CA-2014-100706`
## [1] "Case Logic 2.4GHz Wireless Keyboard"

```

```
## [2] "Ultra Door Push Plate"
##
## $`CA-2014-100867`
## [1] "RCA Visys Integrated PBX 8-Line Router"
```

## Association Rule Mining

### Support

- The proportion of the times an item set occurs in the dataset
- Vary until you get a response with a small set of rules

### Confidence

- Probability that a rule is correct

### Lift

- The ratio by which the confidence of a rule exceeds the expected confidence
  - 1 indicates items are independent and there is no relationship
  - Look for lift > 1

## Identify Frequent Item Sets

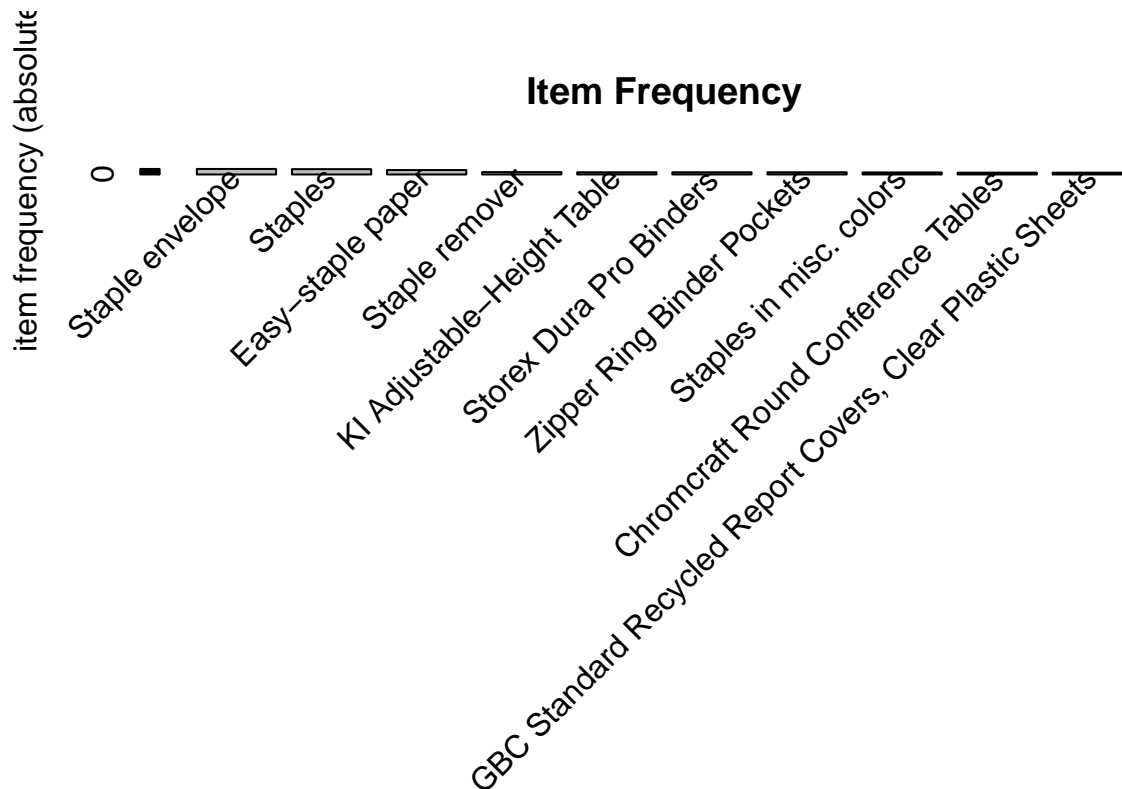
```
frequentItems <- arules::eclat(trans, parameter = list(supp = 0.08, maxlen = 5))
```

```
## Eclat
##
## parameter specification:
## tidLists support minlen maxlen          target  ext
## FALSE      0.08      1      5 frequent itemsets FALSE
##
## algorithmic control:
## sparse sort verbose
##      7      -2      TRUE
##
## Absolute minimum support count: 238
##
## eclat - zero frequent items
```

```
arules::inspect(head(trans))
```

##	items	transactionID
## [1]	{Hon 2111 Invitation Series Corner Table,	
##	Wilson Jones Ledger-Size, Piano-Hinge Binder, 2", Blue}	CA-2014-100090
## [2]	{Xerox 1887}	CA-2014-100293
## [3]	{Strathmore Photo Frame Cards}	CA-2014-100391
## [4]	{Cameo Buff Policy Envelopes,	
##	DMI Arturo Collection Mission-style Design Wood Chair,	
##	Kensington Expert Mouse Optical USB Trackball for PC or Mac,	
##	Prang Dustless Chalk Sticks}	CA-2014-100678
## [5]	{Case Logic 2.4GHz Wireless Keyboard,	
##	Ultra Door Push Plate}	CA-2014-100706
## [6]	{RCA Visys Integrated PBX 8-Line Router}	CA-2014-100867

```
itemFrequencyPlot(trans, topN = 10, type = "absolute", main = "Item Frequency")
```



### Apriori Algorithm: Identify Rules

```
rules <- apriori(trans, parameter = list(supp = 0.06, conf = .7, maxlen = 3))
```

```
## Apriori
##
## Parameter specification:
## confidence minval smax arem aval originalSupport maxtime support minlen
##          0.7    0.1    1 none FALSE                TRUE     5    0.06    1
## maxlen target  ext
##          3  rules FALSE
##
## Algorithmic control:
## filter tree heap memopt load sort verbose
##    0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 178
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[1801 item(s), 2977 transaction(s)] done [0.00s].
## sorting and recoding items ... [0 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 done [0.00s].
## writing ... [0 rule(s)] done [0.00s].
```

```
## creating S4 object ... done [0.00s].  
rules_conf <- sort(rules, by = "confidence", decreasing = TRUE)  
inspect(head(rules_conf))
```