

Name: Vedant Puranik
Roll No: 43152
Class: BE-9
Batch No: R9

Assignment 7 Output

```
> library(arules)
> library(arulesViz)
Loading required package: grid
> library(tidyverse)
> library(readxl)
> library(plyr)
> online.data <- read_excel("Online Retail.xlsx")
> head(online.data)
# A tibble: 6 x 8
  InvoiceNo StockCode Description    Quantity InvoiceDate    UnitPrice CustomerID Country
  <chr>    <chr>    <chr>          <dbl> <dtm>          <dbl>    <dbl> <chr>
1 536365  85123A  WHITE HANGING ~    6 2010-12-01 08:26:00    2.55    17850
United~
2 536365  71053  WHITE METAL LA~    6 2010-12-01 08:26:00    3.39    17850 United~
3 536365  84406B  CREAM CUPID HE~    8 2010-12-01 08:26:00    2.75    17850
United~
4 536365  84029G  KNITTED UNION ~    6 2010-12-01 08:26:00    3.39    17850
United~
5 536365  84029E  RED WOOLLY HOT~    6 2010-12-01 08:26:00    3.39    17850
United~
6 536365  22752  SET 7 BABUSHKA~    2 2010-12-01 08:26:00    7.65    17850
United~
> online.data <- online.data[complete.cases(online.data),]
> online.data %>% mutate(Description=as.factor(Description))
# A tibble: 406,829 x 8
  InvoiceNo StockCode Description    Quantity InvoiceDate    UnitPrice CustomerID Country
  <chr>    <chr>    <fct>          <dbl> <dtm>          <dbl>    <dbl> <chr>
1 536365  85123A  WHITE HANGING~    6 2010-12-01 08:26:00    2.55    17850
United~
2 536365  71053  WHITE METAL L~    6 2010-12-01 08:26:00    3.39    17850 United~
3 536365  84406B  CREAM CUPID H~    8 2010-12-01 08:26:00    2.75    17850
United~
4 536365  84029G  KNITTED UNION~    6 2010-12-01 08:26:00    3.39    17850
United~
5 536365  84029E  RED WOOLLY HO~    6 2010-12-01 08:26:00    3.39    17850
United~
6 536365  22752  SET 7 BABUSHK~    2 2010-12-01 08:26:00    7.65    17850 United~
```

```

7 536365 21730 GLASS STAR FR~ 6 2010-12-01 08:26:00 4.25 17850 United~
8 536366 22633 HAND WARMER U~ 6 2010-12-01 08:28:00 1.85 17850
United~
9 536366 22632 HAND WARMER R~ 6 2010-12-01 08:28:00 1.85 17850
United~
10 536367 84879 ASSORTED COLO~ 32 2010-12-01 08:34:00 1.69 13047
United~

```

... with 406,819 more rows

```
> online.data$InvoiceDate <- as.Date(online.data$InvoiceDate)
```

```
> online.data$InvoiceNo <- as.numeric(online.data$InvoiceNo)
```

Warning message:

NAs introduced by coercion

```
> head(online.data)
```

A tibble: 6 x 8

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
	<dbl>	<chr>	<chr>	<dbl>	<date>	<dbl>	<dbl>	<chr>
1	536365	85123A	WHITE HANGING HEART ~	6	2010-12-01	2.55	17850	United K~
2	536365	71053	WHITE METAL LANTERN	6	2010-12-01	3.39	17850	United K~
3	536365	84406B	CREAM CUPID HEARTS C~	8	2010-12-01	2.75	17850	United K~
4	536365	84029G	KNITTED UNION FLAG H~	6	2010-12-01	3.39	17850	United K~
5	536365	84029E	RED WOOLLY HOTTIE WH~	6	2010-12-01	3.39	17850	United K~
6	536365	22752	SET 7 BABUSHKA NESTI~	2	2010-12-01	7.65	17850	United K~

```
> dim(online.data)
```

```
[1] 406829 8
```

```
> glimpse(online.data)
```

Rows: 406,829

Columns: 8

\$ InvoiceNo <dbl> 536365, 536365, 536365, 536365, 536365, 536365, 536365, 536366, 536366...

\$ StockCode <chr> "85123A", "71053", "84406B", "84029G", "84029E", "22752", "21730", "22...

\$ Description <chr> "WHITE HANGING HEART T-LIGHT HOLDER", "WHITE METAL LANTERN", "CREAM CU...

\$ Quantity <dbl> 6, 6, 8, 6, 6, 2, 6, 6, 6, 32, 6, 6, 8, 6, 6, 3, 2, 3, 3, 4, 4, 6, 3, ...

\$ InvoiceDate <date> 2010-12-01, 2010-12-01, 2010-12-01, 2010-12-01, 2010-12-01, 2010-12-0...

\$ UnitPrice <dbl> 2.55, 3.39, 2.75, 3.39, 3.39, 7.65, 4.25, 1.85, 1.85, 1.69, 2.10, 2.10...

```
$ CustomerID <dbl> 17850, 17850, 17850, 17850, 17850, 17850, 17850, 17850, 17850,
13047, ...
$ Country <chr> "United Kingdom", "United Kingdom", "United Kingdom", "United Kingdom"...
> transactionsFile <- ddply(online.data,c("InvoiceNo","InvoiceDate"),
+ function(online.data)paste(online.data$Description,
+ collapse=", "))
> transactionsFile$InvoiceNo <- NULL
> transactionsFile$InvoiceDate <- NULL
> colnames(transactionsFile) <- c("Items")
> head(transactionsFile)
```

Items

1

WHITE HANGING HEART T-LIGHT HOLDER,WHITE METAL LANTERN,CREAM CUPID
HEARTS COAT HANGER,KNITTED UNION FLAG HOT WATER BOTTLE,RED WOOLLY
HOTTIE WHITE HEART.,SET 7 BABUSHKA NESTING BOXES,GLASS STAR FROSTED
T-LIGHT HOLDER

2

HAND WARMER UNION JACK,HAND WARMER RED POLKA DOT

3

ASSORTED COLOUR BIRD ORNAMENT,POPPY'S PLAYHOUSE BEDROOM,POPPY'S
PLAYHOUSE KITCHEN,FELTCRAFT PRINCESS CHARLOTTE DOLL,IVORY KNITTED MUG
COSY,BOX OF 6 ASSORTED COLOUR TEASPOONS,BOX OF VINTAGE JIGSAW
BLOCKS,BOX OF VINTAGE ALPHABET BLOCKS,HOME BUILDING BLOCK WORD,LOVE
BUILDING BLOCK WORD,RECIPE BOX WITH METAL HEART,DOORMAT NEW ENGLAND

4

JAM MAKING SET WITH JARS,RED COAT RACK PARIS FASHION,YELLOW COAT RACK
PARIS FASHION,BLUE COAT RACK PARIS FASHION

5

BATH BUILDING BLOCK WORD

6 ALARM CLOCK BAKELIKE PINK,ALARM CLOCK BAKELIKE RED,ALARM CLOCK
BAKELIKE GREEN,PANDA AND BUNNIES STICKER SHEET,STARS GIFT
TAPE,INFLATABLE POLITICAL GLOBE,VINTAGE HEADS AND TAILS CARD GAME,SET/2
RED RETROSPOT TEA TOWELS,ROUND SNACK BOXES SET OF4
WOODLAND,SPACEBOY LUNCH BOX,LUNCH BOX I LOVE LONDON,CIRCUS PARADE
LUNCH BOX,CHARLOTTE BAG DOLLY GIRL DESIGN,RED TOADSTOOL LED NIGHT
LIGHT,SET 2 TEA TOWELS I LOVE LONDON,VINTAGE SEASIDE JIGSAW PUZZLES,MINI
JIGSAW CIRCUS PARADE,MINI JIGSAW SPACEBOY,MINI PAINT SET VINTAGE,POSTAGE

```
> transData <- read.transactions("OnlineDataTransactions.csv",format="basket",sep=",")
```

There were 50 or more warnings (use warnings() to see the first 50)

```
> transData
```

transactions in sparse format with

18839 transactions (rows) and

7887 items (columns)

```
> summary(transData)
transactions as itemMatrix in sparse format with
18839 rows (elements/itemsets/transactions) and
7887 columns (items) and a density of 0.002260834
```

most frequent items:

WHITE HANGING HEART T-LIGHT HOLDER	1798	1644	REGENCY CAKESTAND 3 TIER
JUMBO BAG RED RETROSPOT	1450	1282	PARTY BUNTING
ASSORTED COLOUR BIRD ORNAMENT	1249	328499	(Other)

element (itemset/transaction) length distribution:

sizes

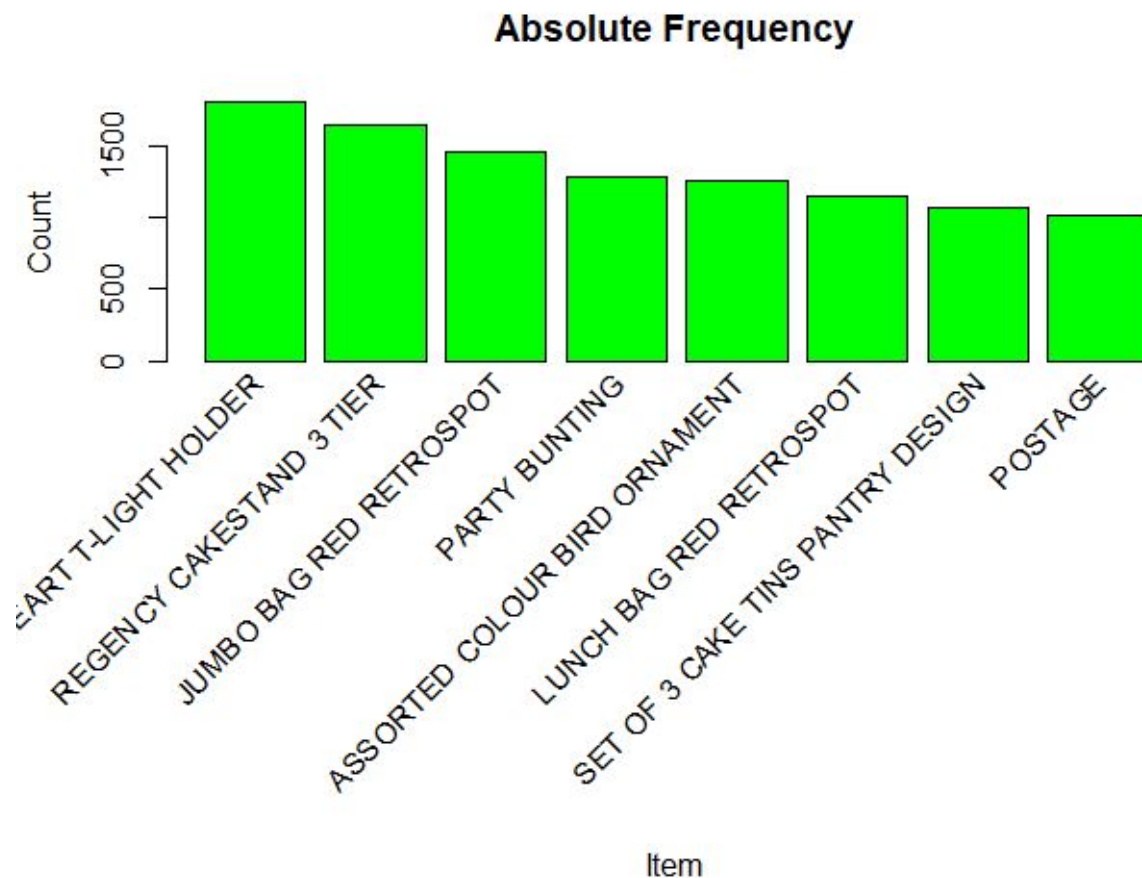
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1578	867	762	773	768	721	660	652	648	586	621	532	510	532	555	525	470	442	483
20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
425	396	319	310	276	241	255	230	218	223	215	173	163	143	146	139	112	118	89
39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57
117	96	97	89	93	67	66	68	65	61	64	53	67	43	42	50	43	37	31
58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76
40	30	27	28	18	26	25	20	27	25	25	15	20	20	13	16	16	12	16
77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95
12	7	9	14	15	12	8	9	11	11	14	8	6	5	6	12	6	4	4
96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114
3	6	5	2	4	2	5	4	3	2	2	6	3	4	3	2	1	3	1
116	117	118	120	121	122	123	125	126	127	131	132	133	134	140	141	142	143	145
4	3	3	1	2	2	1	3	2	2	1	1	2	1	1	2	2	1	1
146	147	150	154	157	168	171	177	178	180	202	204	228	236	249	250	285	320	400
2	1	1	3	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1
419																		
1																		

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.00	5.00	13.00	17.83	23.00	419.00

includes extended item information - examples:

```
labels
1      1 HANGER
2     10 COLOUR SPACEBOY PEN
3    12 COLOURED PARTY BALLOONS
```

```
> itemFrequencyPlot(transData,topN=8,type="absolute",col="green1",
+                   xlab="Item",ylab="Count",main="Absolute Frequency")
```



```
> rules <- apriori(transData,parameter=list(support=0.001,confidence=0.8))
Apriori
```

Parameter specification:

```
confidence minval smax arem  aval originalSupport maxtime support minlen maxlen target  ext
0.8 0.1 1 none FALSE TRUE 5 0.001 1 10 rules TRUE
```

Algorithmic control:

```
filter tree heap memopt load sort verbose
0.1 TRUE TRUE FALSE TRUE 2 TRUE
```

Absolute minimum support count: 18

```
set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[7887 item(s), 18839 transaction(s)] done [0.13s].
sorting and recoding items ... [2455 item(s)] done [0.01s].
creating transaction tree ... done [0.01s].
checking subsets of size 1 2 3 4 5 6 7 8 9 10 done [0.38s].
writing ... [116493 rule(s)] done [0.06s].
creating S4 object ... done [0.06s].
```

Warning message:

In apriori(transData, parameter = list(support = 0.001, confidence = 0.8)) :

Mining stopped (maxlen reached). Only patterns up to a length of 10 returned!

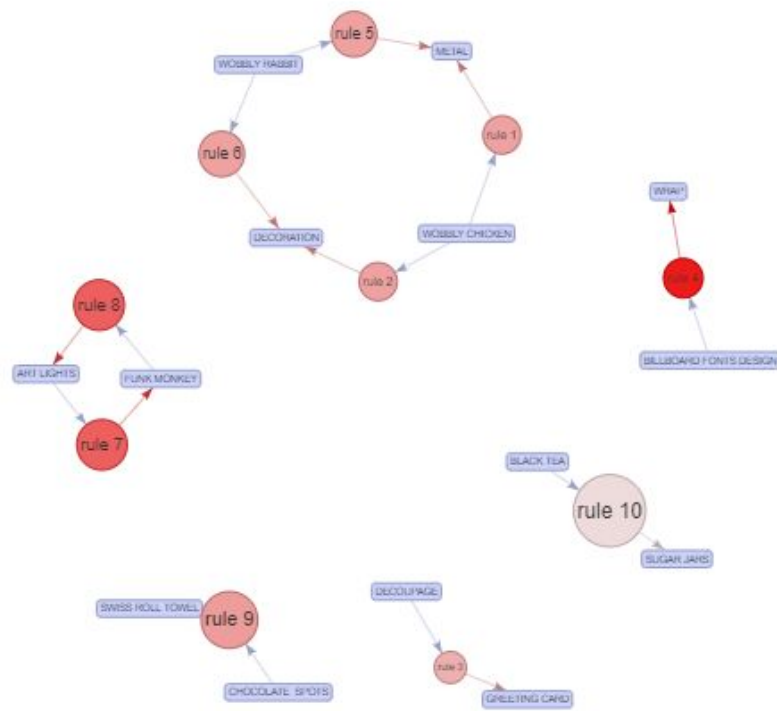
> inspect(rules[1:10,])

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{DOG LICENCE WALL ART}	=> {BICYCLE SAFETY WALL ART}	0.001061627	0.9090909	0.001167790	518.98072	20
[2]	{MIXED NUTS LIGHT GREEN BOWL}	=> {SMALL DOLLY MIX DESIGN ORANGE BOWL}	0.001061627	0.9090909	0.001167790	60.51719	20
[3]	{WOBBLY CHICKEN}	=> {METAL}	0.001486278	1.0000000			
	0.001486278	376.78000	28				
[4]	{WOBBLY CHICKEN}	=> {DECORATION}	0.001486278	1.0000000			
	0.001486278	376.78000	28				
[5]	{MIRRORED WALL ART LADIES}	=> {MIRRORED WALL ART GENTS}	0.001114709	0.8076923	0.001380116	543.43269	21
[6]	{DECOUPAGE}	=> {GREETING CARD}	0.001220872	1.0000000			
	0.001220872	330.50877	23				
[7]	{BILLBOARD FONTS DESIGN}	=> {WRAP}	0.001539360	1.0000000			
	0.001539360	607.70968	29				
[8]	{WRAP}	=> {BILLBOARD FONTS DESIGN}	0.001539360	0.9354839			
	0.001645523	607.70968	29				
[9]	{ENAMEL PINK TEA CONTAINER}	=> {ENAMEL PINK COFFEE CONTAINER}	0.001645523	0.8157895	0.002017092	326.99272	31
[10]	{LANDMARK FRAME COVENT GARDEN}	=> {LANDMARK FRAME OXFORD STREET}	0.001751685	0.8048780	0.002176336	344.61585	33

> top10rules <- head(rules,n=10,by='confidence')

> plot(top10rules,method="graph",engine="htmlwidget")

Select by id ▼



>