Homework 3

March 8, 2020

1 Recitation Problems - Chapter 5

1.1 Exercise 2 - Support & Confidence

1.1.1 a. Compute the support for itemsets {e}, {b,d}, and {b,d,e} by treating each transaction ID as a market basket. Note: 10 transaction in total.

Support Count for {e} = # of transactions containing e = 8

Support for {e} = Support count/ total transactions = $(\frac{8}{10})$ = **0.8**

Support Count for {b,d} = # of transactions containing b & d = 2

Support for {b,d} = = $(\frac{2}{10})$ = **0.2**

Support Count for {b,d,e} = # of transactions containing b, d, & e = 2

Support for {b,d,e} = $(\frac{2}{10})$ = **0.2**

1.1.2 b. Use results from (a) to compute confidence for association rules {b,d} –> {e} and {e} –> {b,d}. Is confidence a symmetric measure?

Confidence for {b,d} –> {e} = Support for {b,d,e} / Support for {b,d} = $(\frac{0.2}{0.2})$ = **1**

Confidence for {e} –> {b,d} = Support for {b,d,e} / Support for {e} = $(\frac{0.2}{0.8})$ = **0.25**

Based on the above, **confidence is not a symmetric measure**.

1.1.3 c. Repeat (a) by treating Customer ID as a market basket. Note: 5 total transactions

Support Count for {e} = # of transactions containing e = 4

Support for {e} = Support count/ total transactions = $(\frac{4}{5})$ = **0.8**

Support Count for {b,d} = # of transactions containing b & d = 5

Support for {b,d} = $(\frac{5}{5})$ = **1**

Support Count for {b,d,e} = # of transactions containing b, d, & e = 4

Support for {b,d,e} = $(\frac{4}{5})$ = **0.8**

1.1.4   d. Use results from (c) to repeat (b).

Confidence for {b,d} –> {e} = Support for {b,d,e} / Support for {b,d} = $(\frac{0.8}{1})$ = **0.8**

Confidence for {e} –> {b,d} = Support for {b,d,e} / Support for {e} = $(\frac{0.8}{0.8})$ = **1**

1.1.5   e. Suppose s_1 and c_1 are support and confidence values for rule r when treating Transaction ID as a market basket and s_2 and c_2 are support and confidence values for rule r when treating Customer ID as a market basket.   Discuss whether there is a relationship between s_1 & s_2 or c_1 & c_2.

I do not think there is any significant relationship between either of two. The only thing that remains equal among the two is that the support for {e} is 0.8. Everything else is different.

1.2   Exercise 6 - Frequent Itemset Generation/Rule Generation

1.2.1   a. What is the maximum # of association rules that can be extracted from Table 5.21 (including minsup > 0)?

Total # of association rules = 3^d - \$2^{d+1}\$ + 1 = 3^6 - 2^7 + 1 = **602 rules**

1.2.2   b. What is the max size of frequent itemsets that can be extracted (minsup > 0)?

Since the largest itemset shown in the table is 4, then **4 is the max size**.

1.2.3   c. Write an expression for the max # of size-3 itemsets that can be derived from this dataset.

$\binom{n}{k} = \binom{6}{3} = 20$

1.2.4   d. Find an itemset (2+) that has the largest support.

Items: Milk, Beer, Diapers, Bread, Butter, and Cookies. Note: Support for {Milk, Butter} = {Butter, Milk} so removing repeated calculations saves time. _____

Support for {Milk, Beer} = $(\frac{1}{10})$ = 0.1

Support for {Milk, Diapers} = $(\frac{3}{10})$ = 0.3

Support for {Milk, Bread} = $(\frac{2}{10})$ = 0.2

Support for {Milk, Butter} = $(\frac{2}{10})$ = 0.2

Support for {Milk, Cookies} = $(\frac{1}{10})$ = 0.1 _____ Support for {Beer, Diapers} = $(\frac{3}{10})$ = 0.3

Support for {Beer, Bread} = 0

Support for {Beer, Butter} = 0

Support for {Beer, Cookies} = $(\frac{2}{10})$ = 0.2 _____ Support for {Diapers, Bread} = $(\frac{3}{10})$ = 0.3

Support for {Diapers, Butter} = $(\frac{3}{10})$ = 0.3

Support for {Diapers, Cookies} = ($\frac{2}{10}$) = 0.2 _____ Support for {Bread, Butter} = ($\frac{5}{10}$) = 0.5

Support for {Bread, Cookies} = ($\frac{1}{10}$) = 0.1 _____ Support for {Butter, Cookies} = ($\frac{1}{10}$) = 0.1

{Bread, Butter} has the largest support count.

1.2.5 e. Find a pair of items, *a* & *b*, such that the rules {a} –> {b} & {b} –> {a} have the same confidence.

Since I have all the support calculated for all pairs, I just need to calculate all the confidence for all rules.

Support {Milk} = 0.5; Support {Beer} = 0.4; Support {Diapers} = 0.6; Support {Bread} = 0.5; Support {Butter} = 0.5; Support {Cookies} = 0.4 _____ Confidence for {Milk} –> {Beer} = Support for {Milk, Beer} / Support for {Milk} = ($\frac{0.1}{0.5}$) = 0.2

Confidence for {Beer} –> {Milk} = Support for {Milk, Beer} / Support for {Beer} = ($\frac{0.1}{0.4}$) = 0.25 _____ Confidence for {Milk} –> {Diapers} = Support for {Milk, Diapers} / Support for {Milk} = ($\frac{0.3}{0.5}$) = 0.6

Confidence for {Diapers} –> {Milk} = Support for {Milk, Diapers} / Support for {Diapers} = ($\frac{0.3}{0.6}$) = 0.5 _____ Confidence for {Milk} –> {Bread} = Support for {Milk, Bread} / Support for {Milk} = ($\frac{0.2}{0.5}$) = 0.4

Confidence for {Bread} –> {Milk} = Support for {Milk, Bread} / Support for {Bread} = ($\frac{0.2}{0.5}$) = 0.4

Since it says find *a pair*, I can stop here. **{Milk, Bread}**.

1.3 Exercise 8 - Candidate Generation

Consider the following frequent 3-itemsets:

{1,2,3}, {1,2,4}, {1,2,5}, {1,3,4}, {1,3,5}, {2,3,4}, {2,3,5}, {3,4,5}

1.3.1 a. List all candidate 4-itemsets obtained by a candidate generation procedure using the F_{k-1} x F_1 merging strategy.

From the list given above, the frequent 1-itemsets would then be {1}, {2}, {3}, {4}, & {5}. The book mentioned that a candicate generation procedure is *complete* if it does not omit a frequent itemset and it does not contain duplicates. To ensure no duplicate candidates are generated, we must make sure the items within each frequent itemset are sorted in their lexicographic order. Since, in our case, they are numbers it will be easy to verify order.

Candidate Generation: 1. {1,2,3} + {4} = {1,2,3,4} 2. {1,2,3} + {5} = {1,2,3,5} 3. {1,2,4} + {5} = {1,2,4,5} 4. {1,3,4} + {5} = {1,3,4,5} 5. {2,3,4} + {5} = {2,3,4,5}

To understand what was accomplished above, a walk-through example is needed. For the first frequent 3-itemset, only two of the frequent 1-itemset could be added to generate a candidate since they contain items that are greater than the ones in the 3-itemset. That is why {1,2,3} is matched with {4} & {5}.

The frequent 3-itemset $\{1,2,5\}$ is skipped because it neither of the frequent 1-itemsets are greater than its current items. The biggest 1-itemset is $\{5\}$, but the 3-itemset already contains 5 in it. Same for $\{1,3,5\}$, $\{2,3,5\}$ & $\{3,4,5\}$.

Thus, the generated candidates are: **$\{1,2,3,4\}$, $\{1,2,3,5\}$, $\{1,2,4,5\}$, $\{1,3,4,5\}$, & $\{2,3,4,5\}$**

1.3.2   b. List all candidate 4-itemsets obtained by the candidate generation procedure used in the *Apriori.*

The Apriori algorithm uses the F_{k-1} x F_{k-1}, which only merges a pair of frequent (k-1)-itemsets if their first k-2 items are identical (lexicographic order included). Since we are forming a 4-itemset, the first k-2 = 4-2 = 2 positions must be the same in order to merge.

Candidate Generation Comparison: 1. $\{1,2,3\}$ vs $\{1,2,4\}$ - The first 2 position are the same, thus we merge the two. $\{1,2,3,4\}$ 2. $\{1,2,3\}$ vs $\{1,2,5\}$ - The first 2 positions are the same, thus we merge. $\{1,2,3,5\}$ 3. $\{1,2,3\}$ vs $\{1,3,4\}$ - The first 2 positions are not the same, so no merge done. 4. $\{1,2,3\}$ vs $\{1,3,5\}$ - No merge. 5. $\{1,2,3\}$ vs $\{2,3,4\}$ - No merge . 6. $\{1,2,4\}$ vs $\{1,2,5\}$ - Merge. $\{1,2,4,5\}$. 7. $\{2,3,4\}$ vs $\{2,3,5\}$ - Merge. $\{2,3,4,5\}$

Comparing this method to the previous shown in (a), only 4 candidates were generated instead of 5. These are **$\{1,2,3,4\}$, $\{1,2,3,5\}$, $\{1,2,4,5\}$ & $\{2,3,4,5\}$**.

1.3.3   c. List all candidate 4-itemsets that survive the candidate pruning step of the *Apriori* alg.

From my understanding, the candidate pruning just eliminates the candidates that contain infrequent subsets. So we begin to compare each of the candidates from (b):

Candidate Pruning: 1. $\{1,2,3,4\}$ - Check if $\{1,2,3\}$ is a frequent itemset $\rightarrow$ Yes. Check if $\{1,2,4\}$ is a frequent itemset $\rightarrow$ Yes. Check if $\{1,3,4\}$ is a frequent itemset $\rightarrow$ Yes. Check if $\{2,3,4\}$ is a frequent itemset $\rightarrow$ Yes. So this 4-itemset is fine. 2. $\{1,2,3,5\}$ - Check $\{1,2,3\}$ $\rightarrow$ Yes. Check $\{1,2,5\}$ $\rightarrow$ Yes. Check $\{1,3,5\}$ $\rightarrow$ Yes. Check $\{2,3,5\}$ $\rightarrow$ Yes. So this 4-itemset is fine. 3. $\{1,2,4,5\}$ - Check $\{1,2,4\}$ $\rightarrow$ Yes. Check $\{1,2,5\}$ $\rightarrow$ Yes. Check $\{1,4,5\}$ $\rightarrow$ No. This 4-itemset is out. 4. $\{2,3,4,5\}$ - Check $\{2,3,4\}$ $\rightarrow$ Yes. Check $\{2,3,5\}$ $\rightarrow$ Yes. Check $\{2,4,5\}$ $\rightarrow$ No. This 4-itemset is out.

Thus, the 4-itemsets that survived the pruning are: **$\{1,2,3,4\}$ & $\{1,2,3,5\}$**

1.4 Exercise 9 - Going thru Apriori

1.4.1 a. Draw itemset lattice representing Table 5.22
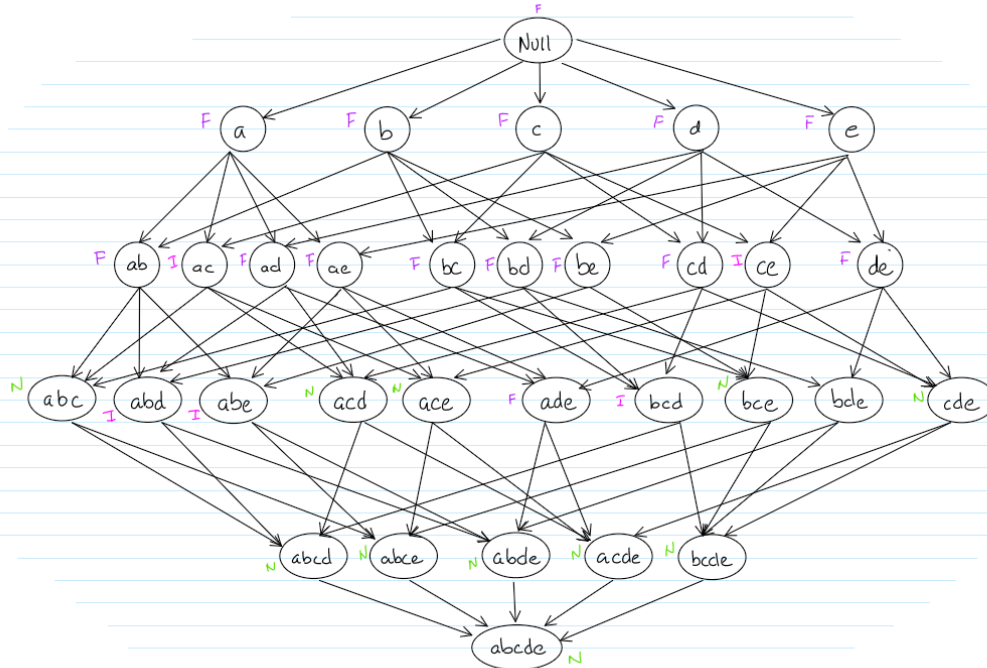
1.4.2 b. What is the percentage of freq. itemsets with respect to all itemsets?

There are a total of 32 itemsets within the lattice graph. Out of those 32, 16 are frequent(when I took picture, I missed showing {bde} as F). Thus, $\frac{16}{32} = 0.5 = \mathbf{50\%}$.

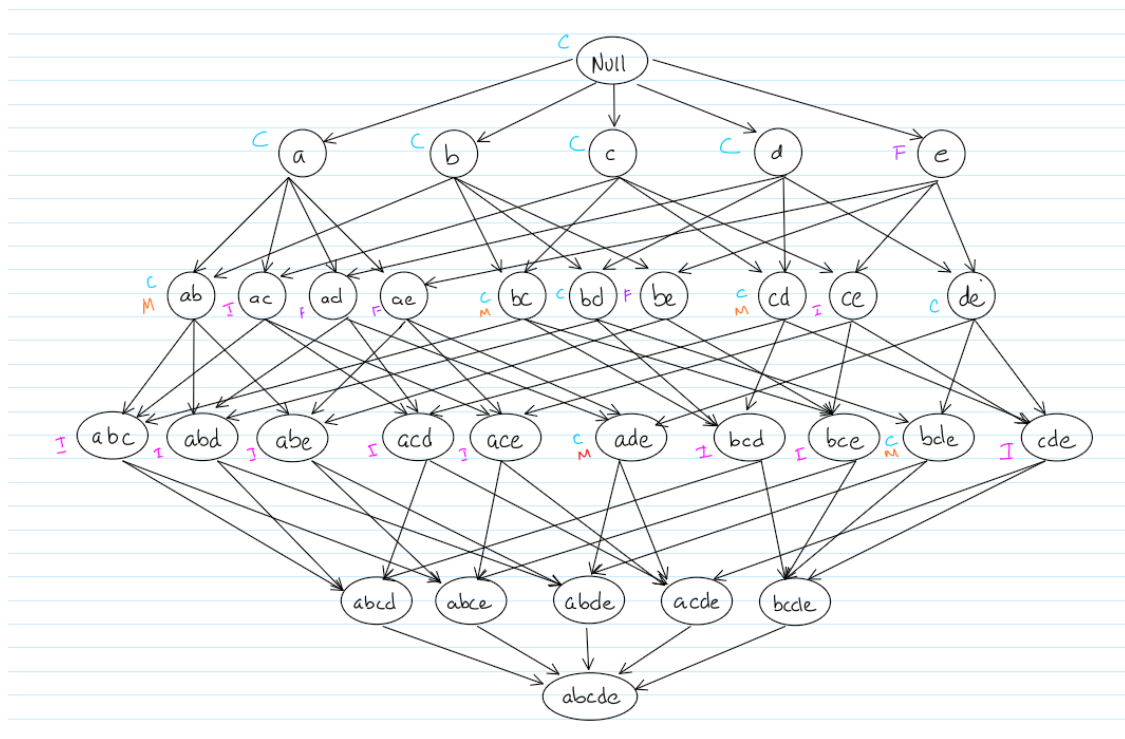### 1.4.3 c. What is pruning ratio of the Apriori Alg. of this data set?

There are 11 itemsets classified as N. Thus, $\frac{11}{32} = .345 = \mathbf{34.5\%}$

1.4.4 d. What is the false alarm rate?

There are 5 itemsets classified as I. Thus, $\frac{5}{32} = 0.156 = \mathbf{15.6\%}$.

1.5 Exercise 12 -

1.5.1 Label same lattice but including M (max. freq. itemset), C (closed freq. itemset), N, and I.

1.6 Exercise 13 - Contingency Table

1.6.1 a. Draw contingency table for Table 5.23

Rule: {b} -> {c}

	C	$\overline{C}$
B	3	4
$\overline{B}$	2	1

Rule: {a} -> {d}

	D	$\overline{D}$
A	4	1
$\overline{A}$	5	0

Rule: {b} -> {d}

	D	$\overline{D}$

	D	$\overline{D}$
B	6	1
$\overline{B}$	3	0

Rule: {e} -> {c}

	C	$\overline{C}$
E	2	4
$\overline{E}$	3	1

Rule: {c} -> {a}

	A	$\overline{A}$
C	2	3
$\overline{C}$	3	2

1.6.2   b. Compute the following measurements based on (a)

Rules	Support	Rank
{b} -> {c}	$\frac{3}{10} = 0.3$	3
{a} -> {d}	$\frac{4}{10} = 0.4$	2
{b} -> {d}	$\frac{6}{10} = 0.6$	1
{e} -> {c}	$\frac{2}{10} = 0.2$	4
{c} -> {a}	$\frac{2}{10} = 0.2$	4

Rules	Confidence	Rank
{b} -> {c}	$\frac{3}{7} = 0.4$	3
{a} -> {d}	$\frac{4}{5} = 0.8$	2
{b} -> {d}	$\frac{6}{7} = 0.9$	1
{e} -> {c}	$\frac{2}{6} = 0.3$	5
{c} -> {a}	$\frac{2}{5} = 0.4$	4

Rules	Interest/Lift	Rank
{b} -> {c}	$\frac{\frac{3}{10}}{\frac{7}{10}} * \frac{5}{10} = 0.21$	3
{a} -> {d}	$\frac{\frac{4}{10}}{\frac{5}{10}} * \frac{9}{10} = 0.72$	2
{b} -> {d}	$\frac{\frac{6}{10}}{\frac{7}{10}} * \frac{9}{10} = 0.77$	1

Rules	Interest/Lift	Rank
{e} -> {c}	$\frac{\frac{2}{10}}{\frac{6}{10}} * \frac{5}{10} = 0.16$	5
{c} -> {a}	$\frac{\frac{2}{10}}{\frac{5}{10}} * \frac{5}{10} = 0.20$	4

Because it's getting harder to write the formulas, just calculating and showing results seems to be fair…

Rules	IS	Rank
{b} -> {c}	0.51	3
{a} -> {d}	0.59	2
{b} -> {d}	0.75	1
{e} -> {c}	0.36	5
{c} -> {a}	0.40	4

Rules	Klosgen	Rank
{b} -> {c}	-0.039	2
{a} -> {d}	-0.063	4
{b} -> {d}	-0.033	1
{e} -> {c}	-0.075	5
{c} -> {a}	-0.045	3

Rules	Odds Ratio	Rank
{b} -> {c}	0.375	2
{a} -> {d}	0	4
{b} -> {d}	0	4
{e} -> {c}	0.167	3
{c} -> {a}	0.444	1

1.7 Exercise 20 - Support, Interest, and Correlation

1.7.1 a. Compute support, interest, and correlation for {A, B} from Table 1. Also, comput the confidence for rules {A} -> {B} & {B} -> {A}

Support for {A,B} $= \frac{9}{100} = 0.09$

Interest for {A,B} $= \frac{\frac{9}{100}}{\frac{10}{100} * \frac{10}{100}} = 9$

ϕ for {A,B} $= \frac{s(A,B) - s(A)s(B)}{sq.root(s(A) * (1 - s(A)) * s(B) * (1 - s(B)))} = 0.89$

Confidence for {A}->{B} $= \frac{\frac{9}{100}}{\frac{10}{100}} = 0.9$

Confidence for {B}–>{A} $= \frac{\frac{9}{100}}{\frac{10}{100}} = 0.9$

1.7.2 b. Repeat (a) for Table 2.

Support for {A,B} $= \frac{89}{100} = 0.89$

Interest for {A,B} $= \frac{\frac{89}{100}}{\frac{90}{100} * \frac{90}{100}} = 1.09$

ϕ for {A,B} $= \frac{s(A,B) - s(A)s(B)}{sq.root(s(A)*(1-s(A))*s(B)*(1-s(B)))} = 0.89$

Confidence for {A}–>{B} $= \frac{\frac{89}{100}}{\frac{90}{100}} = 0.98$

Confidence for {B}–>{A} $= \frac{\frac{89}{100}}{\frac{90}{100}} = 0.98$

1.7.3 c. What conclusions can we draw from (a) & (b)?

It seems that 3 out of the four measurements are known to be invariant, meaning they do not consider absences. These are support, interest, and confidence. Correlaton, however, takes into account absences and the actual presence of an item within a transaction. Also, for confidence, I noticed that for both (a) and (b), the confidence was equal for C(A,B) = C(B,A). Order did not matter, thus A & B are symmetric, which is rare for confidence since it is considered a asymmetric value.

2 Practicum Problems

2.1 Problem 1 - Online Retail

```
[1]: #All necessary imports
     import pandas as pd
     from mlxtend.preprocessing import TransactionEncoder
     from mlxtend.frequent_patterns import apriori
     from mlxtend.frequent_patterns import association_rules

     import numpy as np
     from scipy.stats import zscore
     from sklearn.metrics import jaccard_score
```

```
[77]: #Creating a panda dataframe
      online_retail_data = pd.read_excel('http://archive.ics.uci.edu/ml/
       ↪machine-learning-databases/00352/Online%20Retail.xlsx')
      #pd.read_excel('OnlineRetail.xlsx', index_col=0, header=0)
```

```
[78]: #See the online retail data
      online_retail_data.head()
```

```
[78]:    InvoiceNo StockCode                          Description  Quantity  \
     0     536365    85123A   WHITE HANGING HEART T-LIGHT HOLDER         6
```

```
1   536365    71053                 WHITE METAL LANTERN         6
2   536365   84406B    CREAM CUPID HEARTS COAT HANGER           8
3   536365   84029G KNITTED UNION FLAG HOT WATER BOTTLE         6
4   536365   84029E      RED WOOLLY HOTTIE WHITE HEART.         6


         InvoiceDate  UnitPrice  CustomerID         Country
0 2010-12-01 08:26:00      2.55    17850.0  United Kingdom
1 2010-12-01 08:26:00      3.39    17850.0  United Kingdom
2 2010-12-01 08:26:00      2.75    17850.0  United Kingdom
3 2010-12-01 08:26:00      3.39    17850.0  United Kingdom
4 2010-12-01 08:26:00      3.39    17850.0  United Kingdom
```

[89]:
```python
#Clean up data

#Remove the white spaces from the description section
online_retail_data['Description'] = online_retail_data['Description'].str.
 ↪strip()

#Remove rows without invoice #
online_retail_data.dropna(axis=0, subset=['InvoiceNo'], inplace=True)
online_retail_data['InvoiceNo'] = online_retail_data['InvoiceNo'].astype('str')
#online_retail_data = online_retail_data[~online_retail_data['InvoiceNo'].str.
 ↪contains('C')]
online_retail_data
```

[89]:
```
       InvoiceNo StockCode                          Description  Quantity  \
0         536365    85123A  WHITE HANGING HEART T-LIGHT HOLDER         6
1         536365     71053                 WHITE METAL LANTERN         6
2         536365    84406B       CREAM CUPID HEARTS COAT HANGER         8
3         536365    84029G KNITTED UNION FLAG HOT WATER BOTTLE         6
4         536365    84029E       RED WOOLLY HOTTIE WHITE HEART.         6
...          ...       ...                                  ...       ...
541904    581587     22613         PACK OF 20 SPACEBOY NAPKINS        12
541905    581587     22899          CHILDREN'S APRON DOLLY GIRL         6
541906    581587     23254         CHILDRENS CUTLERY DOLLY GIRL         4
541907    581587     23255     CHILDRENS CUTLERY CIRCUS PARADE         4
541908    581587     22138        BAKING SET 9 PIECE RETROSPOT         3

               InvoiceDate  UnitPrice  CustomerID         Country
0      2010-12-01 08:26:00       2.55    17850.0  United Kingdom
1      2010-12-01 08:26:00       3.39    17850.0  United Kingdom
2      2010-12-01 08:26:00       2.75    17850.0  United Kingdom
3      2010-12-01 08:26:00       3.39    17850.0  United Kingdom
4      2010-12-01 08:26:00       3.39    17850.0  United Kingdom
...                    ...        ...        ...             ...
541904 2011-12-09 12:50:00       0.85    12680.0          France
541905 2011-12-09 12:50:00       2.10    12680.0          France
```

```
541906 2011-12-09 12:50:00        4.15     12680.0        France
541907 2011-12-09 12:50:00        4.15     12680.0        France
541908 2011-12-09 12:50:00        4.95     12680.0        France

[541909 rows x 8 columns]
```

[91]:
```python
#Consolidate the items into 1 transaction per row
basket = (online_retail_data[online_retail_data['Country'] =="France"]
        .groupby(['InvoiceNo', 'Description'])['Quantity']
        .sum().unstack().reset_index().fillna(0)
        .set_index('InvoiceNo'))
basket
```

[91]:
```
Description  10 COLOUR SPACEBOY PEN  12 COLOURED PARTY BALLOONS  \
InvoiceNo
536370                         0.0                         0.0
536852                         0.0                         0.0
536974                         0.0                         0.0
537065                         0.0                         0.0
537463                         0.0                         0.0
...                            ...                         ...
C579532                        0.0                         0.0
C579562                        0.0                         0.0
C580161                        0.0                         0.0
C580263                        0.0                         0.0
C581316                        0.0                         0.0

Description  12 EGG HOUSE PAINTED WOOD  12 MESSAGE CARDS WITH ENVELOPES  \
InvoiceNo
536370                           0.0                              0.0
536852                           0.0                              0.0
536974                           0.0                              0.0
537065                           0.0                              0.0
537463                           0.0                              0.0
...                              ...                              ...
C579532                          0.0                              0.0
C579562                          0.0                              0.0
C580161                          0.0                              0.0
C580263                          0.0                              0.0
C581316                          0.0                              0.0

Description  12 PENCIL SMALL TUBE WOODLAND  \
InvoiceNo
536370                             0.0
536852                             0.0
536974                             0.0
537065                             0.0
```

```
537463                                         0.0
…                                              …
C579532                                        0.0
C579562                                        0.0
C580161                                        0.0
C580263                                        0.0
C581316                                        0.0


Description  12 PENCILS SMALL TUBE RED RETROSPOT  12 PENCILS SMALL TUBE SKULL  \
InvoiceNo
536370                                    0.0                            0.0
536852                                    0.0                            0.0
536974                                    0.0                            0.0
537065                                    0.0                            0.0
537463                                    0.0                            0.0
…                                         …                              …
C579532                                   0.0                            0.0
C579562                                   0.0                            0.0
C580161                                   0.0                            0.0
C580263                                   0.0                            0.0
C581316                                   0.0                            0.0


Description  12 PENCILS TALL TUBE POSY  12 PENCILS TALL TUBE RED RETROSPOT  \
InvoiceNo
536370                          0.0                                  0.0
536852                          0.0                                  0.0
536974                          0.0                                  0.0
537065                          0.0                                  0.0
537463                          0.0                                  0.0
…                               …                                    …
C579532                         0.0                                  0.0
C579562                         0.0                                  0.0
C580161                         0.0                                  0.0
C580263                         0.0                                  0.0
C581316                         0.0                                  0.0


Description  12 PENCILS TALL TUBE WOODLAND  …  WRAP VINTAGE PETALS  DESIGN  \
InvoiceNo                                    …
536370                              0.0  …                            0.0
536852                              0.0  …                            0.0
536974                              0.0  …                            0.0
537065                              0.0  …                            0.0
537463                              0.0  …                            0.0
…                                   …    …                            …
C579532                             0.0  …                            0.0
C579562                             0.0  …                            0.0
C580161                             0.0  …                            0.0
```

```
C580263                              0.0  …                          0.0
C581316                              0.0  …                          0.0


Description  YELLOW COAT RACK PARIS FASHION  YELLOW GIANT GARDEN THERMOMETER  \
InvoiceNo
536370                                  0.0                             0.0
536852                                  0.0                             0.0
536974                                  0.0                             0.0
537065                                  0.0                             0.0
537463                                  0.0                             0.0
…                                       …                               …
C579532                                 0.0                             0.0
C579562                                 0.0                             0.0
C580161                                 0.0                             0.0
C580263                                 0.0                             0.0
C581316                                 0.0                             0.0


Description  YELLOW SHARK HELICOPTER  ZINC  STAR T-LIGHT HOLDER  \
InvoiceNo
536370                          0.0                       0.0
536852                          0.0                       0.0
536974                          0.0                       0.0
537065                          0.0                       0.0
537463                          0.0                       0.0
…                               …                         …
C579532                         0.0                       0.0
C579562                         0.0                       0.0
C580161                         0.0                       0.0
C580263                         0.0                       0.0
C581316                         0.0                       0.0


Description  ZINC FOLKART SLEIGH BELLS  ZINC HERB GARDEN CONTAINER  \
InvoiceNo
536370                            0.0                          0.0
536852                            0.0                          0.0
536974                            0.0                          0.0
537065                            0.0                          0.0
537463                            0.0                          0.0
…                                 …                            …
C579532                           0.0                          0.0
C579562                           0.0                          0.0
C580161                           0.0                          0.0
C580263                           0.0                          0.0
C581316                           0.0                          0.0


Description  ZINC METAL HEART DECORATION  ZINC T-LIGHT HOLDER STAR LARGE  \
InvoiceNo
```

```
536370                                   0.0                        0.0
536852                                   0.0                        0.0
536974                                   0.0                        0.0
537065                                   0.0                        0.0
537463                                   0.0                        0.0
…                                        …                          …
C579532                                  0.0                        0.0
C579562                                  0.0                        0.0
C580161                                  0.0                        0.0
C580263                                  0.0                        0.0
C581316                                  0.0                        0.0

Description  ZINC T-LIGHT HOLDER STARS SMALL
InvoiceNo
536370                                        0.0
536852                                        0.0
536974                                        0.0
537065                                        0.0
537463                                        0.0
…                                             …
C579532                                       0.0
C579562                                       0.0
C580161                                       0.0
C580263                                       0.0
C581316                                       0.0

[461 rows x 1564 columns]
```

```python
def encode_units(x):
    if x <= 0:
        return 0
    if x >= 1:
        return 1

basket_sets = basket.applymap(encode_units)
basket_sets.drop('POSTAGE', inplace=True, axis=1)

basket.head()
```

```
[95]: Description  10 COLOUR SPACEBOY PEN  12 COLOURED PARTY BALLOONS  \
      InvoiceNo
      536370                         0.0                        0.0
      536852                         0.0                        0.0
      536974                         0.0                        0.0
      537065                         0.0                        0.0
      537463                         0.0                        0.0
```

```
Description  12 EGG HOUSE PAINTED WOOD  12 MESSAGE CARDS WITH ENVELOPES  \
InvoiceNo
536370                            0.0                              0.0
536852                            0.0                              0.0
536974                            0.0                              0.0
537065                            0.0                              0.0
537463                            0.0                              0.0


Description  12 PENCIL SMALL TUBE WOODLAND  \
InvoiceNo
536370                          0.0
536852                          0.0
536974                          0.0
537065                          0.0
537463                          0.0


Description  12 PENCILS SMALL TUBE RED RETROSPOT  12 PENCILS SMALL TUBE SKULL  \
InvoiceNo
536370                                 0.0                            0.0
536852                                 0.0                            0.0
536974                                 0.0                            0.0
537065                                 0.0                            0.0
537463                                 0.0                            0.0


Description  12 PENCILS TALL TUBE POSY  12 PENCILS TALL TUBE RED RETROSPOT  \
InvoiceNo
536370                        0.0                                0.0
536852                        0.0                                0.0
536974                        0.0                                0.0
537065                        0.0                                0.0
537463                        0.0                                0.0


Description  12 PENCILS TALL TUBE WOODLAND  …  WRAP VINTAGE PETALS  DESIGN  \
InvoiceNo                                   …
536370                           0.0  …                          0.0
536852                           0.0  …                          0.0
536974                           0.0  …                          0.0
537065                           0.0  …                          0.0
537463                           0.0  …                          0.0


Description  YELLOW COAT RACK PARIS FASHION  YELLOW GIANT GARDEN THERMOMETER  \
InvoiceNo
536370                             0.0                               0.0
536852                             0.0                               0.0
536974                             0.0                               0.0
537065                             0.0                               0.0
537463                             0.0                               0.0
```

```
Description  YELLOW SHARK HELICOPTER  ZINC  STAR T-LIGHT HOLDER  \
InvoiceNo
536370                            0.0                     0.0
536852                            0.0                     0.0
536974                            0.0                     0.0
537065                            0.0                     0.0
537463                            0.0                     0.0

Description  ZINC FOLKART SLEIGH BELLS  ZINC HERB GARDEN CONTAINER  \
InvoiceNo
536370                         0.0                          0.0
536852                         0.0                          0.0
536974                         0.0                          0.0
537065                         0.0                          0.0
537463                         0.0                          0.0

Description  ZINC METAL HEART DECORATION  ZINC T-LIGHT HOLDER STAR LARGE  \
InvoiceNo
536370                           0.0                              0.0
536852                           0.0                              0.0
536974                           0.0                              0.0
537065                           0.0                              0.0
537463                           0.0                              0.0

Description  ZINC T-LIGHT HOLDER STARS SMALL
InvoiceNo
536370                              0.0
536852                              0.0
536974                              0.0
537065                              0.0
537463                              0.0

[5 rows x 1564 columns]
```

```python
#Find the frequent itemsets for all transactions from France
frequent_itemsets = apriori(basket_sets, min_support=0.05, use_colnames=True)
frequent_itemsets
```

```
[96]:    support                                         itemsets
     0   0.060738               (4 TRADITIONAL SPINNING TOPS)
     1   0.082430               (ALARM CLOCK BAKELIKE GREEN)
     2   0.086768               (ALARM CLOCK BAKELIKE PINK)
     3   0.080260                (ALARM CLOCK BAKELIKE RED)
     4   0.058568               (ASSORTED COLOUR MINI CASES)
     ..       …                                            …
     73  0.086768  (SET/20 RED RETROSPOT PAPER NAPKINS, SET/6 RED…
```

16

```
74  0.104121  (SET/6 RED SPOTTY PAPER CUPS, SET/6 RED SPOTTY…
75  0.054230  (ALARM CLOCK BAKELIKE PINK, ALARM CLOCK BAKELI…
76  0.058568  (PLASTERS IN TIN CIRCUS PARADE, PLASTERS IN TI…
77  0.084599  (SET/20 RED RETROSPOT PAPER NAPKINS, SET/6 RED…

[78 rows x 2 columns]
```

[99]:
```
#Find the item with the largest support
max_itemset = frequent_itemsets.max()
max_itemset
```

[99]:
```
support                                           0.160521
itemsets    (SET/20 RED RETROSPOT PAPER NAPKINS, SET/6 RED…
dtype: object
```

[106]:
```
#Create association rules based on the frequent_itemsets
rules = association_rules(frequent_itemsets, metric="lift", min_threshold=1)
rules
```

[106]:
```
                                        antecedents  \
0                        (ALARM CLOCK BAKELIKE PINK)
1                       (ALARM CLOCK BAKELIKE GREEN)
2                       (ALARM CLOCK BAKELIKE GREEN)
3                         (ALARM CLOCK BAKELIKE RED)
4                        (ALARM CLOCK BAKELIKE PINK)
5                         (ALARM CLOCK BAKELIKE RED)
6                        (CHILDRENS CUTLERY SPACEBOY)
7                      (CHILDRENS CUTLERY DOLLY GIRL)
8                                (SPACEBOY LUNCH BOX)
9                              (DOLLY GIRL LUNCH BOX)
10                            (LUNCH BAG APPLE DESIGN)
11                            (LUNCH BAG RED RETROSPOT)
12                            (LUNCH BAG APPLE DESIGN)
13                         (LUNCH BAG SPACEBOY DESIGN)
14                            (LUNCH BAG RED RETROSPOT)
15                         (LUNCH BAG SPACEBOY DESIGN)
16                                 (LUNCH BAG WOODLAND)
17                         (LUNCH BAG SPACEBOY DESIGN)
18                     (LUNCH BOX WITH CUTLERY RETROSPOT)
19              (STRAWBERRY LUNCH BOX WITH CUTLERY)
20                     (PLASTERS IN TIN CIRCUS PARADE)
21                          (PLASTERS IN TIN SPACEBOY)
22                  (PLASTERS IN TIN WOODLAND ANIMALS)
23                     (PLASTERS IN TIN CIRCUS PARADE)
24                  (PLASTERS IN TIN WOODLAND ANIMALS)
25                          (PLASTERS IN TIN SPACEBOY)
26                     (ROUND SNACK BOXES SET OF 4 FRUITS)
```

```
27                    (ROUND SNACK BOXES SET OF4 WOODLAND)
28                    (SET/20 RED RETROSPOT PAPER NAPKINS)
29                          (SET/6 RED SPOTTY PAPER CUPS)
30                    (SET/20 RED RETROSPOT PAPER NAPKINS)
31                        (SET/6 RED SPOTTY PAPER PLATES)
32                          (SET/6 RED SPOTTY PAPER CUPS)
33                        (SET/6 RED SPOTTY PAPER PLATES)
34  (ALARM CLOCK BAKELIKE PINK, ALARM CLOCK BAKELI…
35  (ALARM CLOCK BAKELIKE PINK, ALARM CLOCK BAKELI…
36  (ALARM CLOCK BAKELIKE GREEN, ALARM CLOCK BAKEL…
37                          (ALARM CLOCK BAKELIKE PINK)
38                           (ALARM CLOCK BAKELIKE RED)
39                         (ALARM CLOCK BAKELIKE GREEN)
40  (PLASTERS IN TIN WOODLAND ANIMALS, PLASTERS IN…
41  (PLASTERS IN TIN SPACEBOY, PLASTERS IN TIN CIR…
42  (PLASTERS IN TIN WOODLAND ANIMALS, PLASTERS IN…
43                      (PLASTERS IN TIN CIRCUS PARADE)
44                    (PLASTERS IN TIN WOODLAND ANIMALS)
45                          (PLASTERS IN TIN SPACEBOY)
46  (SET/20 RED RETROSPOT PAPER NAPKINS, SET/6 RED…
47  (SET/20 RED RETROSPOT PAPER NAPKINS, SET/6 RED…
48  (SET/6 RED SPOTTY PAPER CUPS, SET/6 RED SPOTTY…
49                    (SET/20 RED RETROSPOT PAPER NAPKINS)
50                          (SET/6 RED SPOTTY PAPER CUPS)
51                        (SET/6 RED SPOTTY PAPER PLATES)
```

	consequents	antecedent support
0	(ALARM CLOCK BAKELIKE GREEN)	0.086768
1	(ALARM CLOCK BAKELIKE PINK)	0.082430
2	(ALARM CLOCK BAKELIKE RED)	0.082430
3	(ALARM CLOCK BAKELIKE GREEN)	0.080260
4	(ALARM CLOCK BAKELIKE RED)	0.086768
5	(ALARM CLOCK BAKELIKE PINK)	0.080260
6	(CHILDRENS CUTLERY DOLLY GIRL)	0.058568
7	(CHILDRENS CUTLERY SPACEBOY)	0.060738
8	(DOLLY GIRL LUNCH BOX)	0.106291
9	(SPACEBOY LUNCH BOX)	0.084599
10	(LUNCH BAG RED RETROSPOT)	0.106291
11	(LUNCH BAG APPLE DESIGN)	0.130152
12	(LUNCH BAG SPACEBOY DESIGN)	0.106291
13	(LUNCH BAG APPLE DESIGN)	0.101952
14	(LUNCH BAG SPACEBOY DESIGN)	0.130152
15	(LUNCH BAG RED RETROSPOT)	0.101952
16	(LUNCH BAG SPACEBOY DESIGN)	0.099783
17	(LUNCH BAG WOODLAND)	0.101952
18	(STRAWBERRY LUNCH BOX WITH CUTLERY)	0.121475
19	(LUNCH BOX WITH CUTLERY RETROSPOT)	0.104121

```
20                    (PLASTERS IN TIN SPACEBOY)          0.143167
21               (PLASTERS IN TIN CIRCUS PARADE)          0.117137
22               (PLASTERS IN TIN CIRCUS PARADE)          0.145336
23             (PLASTERS IN TIN WOODLAND ANIMALS)         0.143167
24                    (PLASTERS IN TIN SPACEBOY)          0.145336
25             (PLASTERS IN TIN WOODLAND ANIMALS)         0.117137
26             (ROUND SNACK BOXES SET OF4 WOODLAND)       0.091106
27             (ROUND SNACK BOXES SET OF 4 FRUITS)        0.134490
28                  (SET/6 RED SPOTTY PAPER CUPS)         0.112798
29             (SET/20 RED RETROSPOT PAPER NAPKINS)       0.117137
30                 (SET/6 RED SPOTTY PAPER PLATES)        0.112798
31             (SET/20 RED RETROSPOT PAPER NAPKINS)       0.108460
32                 (SET/6 RED SPOTTY PAPER PLATES)        0.117137
33                  (SET/6 RED SPOTTY PAPER CUPS)         0.108460
34                  (ALARM CLOCK BAKELIKE GREEN)          0.062907
35                   (ALARM CLOCK BAKELIKE RED)           0.062907
36                  (ALARM CLOCK BAKELIKE PINK)           0.067245
37   (ALARM CLOCK BAKELIKE GREEN, ALARM CLOCK BAKEL…       0.086768
38   (ALARM CLOCK BAKELIKE PINK, ALARM CLOCK BAKELI…       0.080260
39   (ALARM CLOCK BAKELIKE PINK, ALARM CLOCK BAKELI…       0.082430
40                    (PLASTERS IN TIN SPACEBOY)          0.086768
41             (PLASTERS IN TIN WOODLAND ANIMALS)         0.075922
42               (PLASTERS IN TIN CIRCUS PARADE)          0.088937
43   (PLASTERS IN TIN WOODLAND ANIMALS, PLASTERS IN…       0.143167
44   (PLASTERS IN TIN SPACEBOY, PLASTERS IN TIN CIR…       0.145336
45   (PLASTERS IN TIN WOODLAND ANIMALS, PLASTERS IN…       0.117137
46                 (SET/6 RED SPOTTY PAPER PLATES)        0.086768
47                  (SET/6 RED SPOTTY PAPER CUPS)         0.086768
48             (SET/20 RED RETROSPOT PAPER NAPKINS)       0.104121
49   (SET/6 RED SPOTTY PAPER CUPS, SET/6 RED SPOTTY…       0.112798
50   (SET/20 RED RETROSPOT PAPER NAPKINS, SET/6 RED…       0.117137
51   (SET/20 RED RETROSPOT PAPER NAPKINS, SET/6 RED…       0.108460
```

	consequent support	support	confidence	lift	leverage	conviction
0	0.082430	0.062907	0.725000	8.795395	0.055754	3.336620
1	0.086768	0.062907	0.763158	8.795395	0.055754	3.855869
2	0.080260	0.067245	0.815789	10.164296	0.060629	4.992873
3	0.082430	0.067245	0.837838	10.164296	0.060629	5.658351
4	0.080260	0.062907	0.725000	9.033108	0.055943	3.344508
5	0.086768	0.062907	0.783784	9.033108	0.055943	4.223698
6	0.060738	0.054230	0.925926	15.244709	0.050673	12.680043
7	0.058568	0.054230	0.892857	15.244709	0.050673	8.786696
8	0.084599	0.060738	0.571429	6.754579	0.051745	2.135936
9	0.106291	0.060738	0.717949	6.754579	0.051745	3.168606
10	0.130152	0.056399	0.530612	4.076871	0.042565	1.853155
11	0.106291	0.056399	0.433333	4.076871	0.042565	1.577134
12	0.101952	0.054230	0.510204	5.004342	0.043393	1.833514

13	0.106291	0.054230	0.531915	5.004342	0.043393	1.909288
14	0.101952	0.056399	0.433333	4.250355	0.043130	1.584790
15	0.130152	0.056399	0.553191	4.250355	0.043130	1.946803
16	0.101952	0.054230	0.543478	5.330712	0.044057	1.967152
17	0.099783	0.054230	0.531915	5.330712	0.044057	1.923191
18	0.104121	0.054230	0.446429	4.287574	0.041582	1.618361
19	0.121475	0.054230	0.520833	4.287574	0.041582	1.833443
20	0.117137	0.075922	0.530303	4.527217	0.059152	1.879645
21	0.143167	0.075922	0.648148	4.527217	0.059152	2.435209
22	0.143167	0.086768	0.597015	4.170059	0.065961	2.126215
23	0.145336	0.086768	0.606061	4.170059	0.065961	2.169531
24	0.117137	0.088937	0.611940	5.224157	0.071913	2.275071
25	0.145336	0.088937	0.759259	5.224157	0.071913	3.550142
26	0.134490	0.054230	0.595238	4.425883	0.041977	2.138318
27	0.091106	0.054230	0.403226	4.425883	0.041977	1.523011
28	0.117137	0.086768	0.769231	6.566952	0.073555	3.825741
29	0.112798	0.086768	0.740741	6.566952	0.073555	3.422064
30	0.108460	0.086768	0.769231	7.092308	0.074534	3.863341
31	0.112798	0.086768	0.800000	7.092308	0.074534	4.436009
32	0.108460	0.104121	0.888889	8.195556	0.091417	8.023861
33	0.117137	0.104121	0.960000	8.195556	0.091417	22.071584
34	0.082430	0.054230	0.862069	10.458258	0.049045	6.652386
35	0.080260	0.054230	0.862069	10.740913	0.049181	6.668113
36	0.086768	0.054230	0.806452	9.294355	0.048395	4.718366
37	0.067245	0.054230	0.625000	9.294355	0.048395	2.487346
38	0.062907	0.054230	0.675676	10.740913	0.049181	2.889371
39	0.062907	0.054230	0.657895	10.458258	0.049045	2.739196
40	0.117137	0.058568	0.675000	5.762500	0.048405	2.716503
41	0.145336	0.058568	0.771429	5.307889	0.047534	3.739154
42	0.143167	0.058568	0.658537	4.599778	0.045835	2.509297
43	0.088937	0.058568	0.409091	4.599778	0.045835	1.541799
44	0.075922	0.058568	0.402985	5.307889	0.047534	1.547831
45	0.086768	0.058568	0.500000	5.762500	0.048405	1.826464
46	0.108460	0.084599	0.975000	8.989500	0.075188	35.661605
47	0.117137	0.084599	0.975000	8.323611	0.074435	35.314534
48	0.112798	0.084599	0.812500	7.203125	0.072854	4.731743
49	0.104121	0.084599	0.750000	7.203125	0.072854	3.583514
50	0.086768	0.084599	0.722222	8.323611	0.074435	3.287636
51	0.086768	0.084599	0.780000	8.989500	0.075188	4.151055

[115]:
```python
#let's narrow down the rules using the max lift
#The max lift here is:
rules['lift'].max()
```

[115]: 15.244708994708995

```
[118]: #Let's see the information regarding the lift equal to the max
       rules[(rules['lift'] == rules['lift'].max())]
```

```
[118]:                    antecedents                       consequents  \
       6    (CHILDRENS CUTLERY SPACEBOY)    (CHILDRENS CUTLERY DOLLY GIRL)
       7  (CHILDRENS CUTLERY DOLLY GIRL)    (CHILDRENS CUTLERY SPACEBOY)

          antecedent support   consequent support   support   confidence       lift  \
       6            0.058568             0.060738   0.05423     0.925926  15.244709
       7            0.060738             0.058568   0.05423     0.892857  15.244709

          leverage   conviction
       6  0.050673    12.680043
       7  0.050673     8.786696
```

```
[119]: #Let's narrow down the rules using the max confidence
       rules['confidence'].max()
```

```
[119]: 0.975
```

```
[120]: #Let's see the information regarding the lift equal to the max
       rules[(rules['confidence'] == rules['confidence'].max())]
```

```
[120]:                                      antecedents  \
       46   (SET/20 RED RETROSPOT PAPER NAPKINS, SET/6 RED…
       47   (SET/20 RED RETROSPOT PAPER NAPKINS, SET/6 RED…

                              consequents   antecedent support   consequent support  \
       46   (SET/6 RED SPOTTY PAPER PLATES)             0.086768             0.108460
       47     (SET/6 RED SPOTTY PAPER CUPS)             0.086768             0.117137

            support   confidence       lift   leverage   conviction
       46  0.084599       0.975   8.989500   0.075188    35.661605
       47  0.084599       0.975   8.323611   0.074435    35.314534
```

```
[123]: #Since the Antecedent isn't clearly displayed, I had to specifically call it
       rules[(rules['confidence'] == rules['confidence'].max())].iloc[0, 0]
```

```
[123]: frozenset({'SET/20 RED RETROSPOT PAPER NAPKINS',
                  'SET/6 RED SPOTTY PAPER CUPS'})
```

```
[124]: rules[(rules['confidence'] == rules['confidence'].max())].iloc[1, 0]
```

```
[124]: frozenset({'SET/20 RED RETROSPOT PAPER NAPKINS',
                  'SET/6 RED SPOTTY PAPER PLATES'})
```

From the above, the rules with the max lift are Rule 6 & 7. The antecedents and consequents are shown above. 1. Rule 6: Antecedents = (CHILDRENS CUTLERY SPACEBOY), Consequents =

(CHILDRENS CUTLERY DOLLY GIRL) 2. Rule 7: Antecedents = (CHILDRENS CUTLERY DOLLY GIRL), Consequents = (CHILDRENS CUTLERY SPACEBOY)

From the above, the rules with the max confidence are Rule 46 & 47. The antecedents and consequents are shown above. 1. Rule 46: Antecedents = (SET/20 RED RETROSPOT PAPER NAPKINS, SET/6 RED SPOTTY PAPER CUPS), Consequents = (SET/6 RED SPOTTY PAPER PLATES) 2. Rule 47: Antecedents = (SET/20 RED RETROSPOT PAPER NAPKINS, SET/6 RED SPOTTY PAPER PLATES), Consequents = (SET/6 RED SPOTTY PAPER CUPS)

As seen from above, the rule with the highest confidence is not the same as the rule with the highest lift. This is the due to the difference between confidence and lift. Confidence is based purely on the support of the itemset, while lift depends on confidence of the rule and the support of the consequent. Because of this, the results are different.

2.2 Problem 2 - Extended Bakery

```python
#Creating a panda dataframe
bakery = pd.read_csv("bakery.csv", header=0)
bakery.head()
```

[2]:

	Transaction Number	Chocolate Cake	Lemon Cake	Casino Cake	Opera Cake	\
0	1	0	0	0	0	
1	2	0	0	0	0	
2	3	0	0	0	1	
3	4	0	0	0	0	
4	5	0	0	0	0	

	Strawberry Cake	Truffle Cake	Chocolate Eclair	Coffee Eclair	\
0	0	0	0	0	
1	0	0	0	1	
2	0	0	0	0	
3	0	1	0	0	
4	0	0	1	0	

	Vanilla Eclair	…	Lemon Lemonade	Raspberry Lemonade	Orange Juice	\
0	0	…	0	0	0	
1	0	…	0	0	0	
2	0	…	0	0	1	
3	0	…	0	0	0	
4	0	…	0	0	1	

	Green Tea	Bottled Water	Hot Coffee	Chocolate Coffee	\
0	0	0	0	0	
1	0	0	1	0	
2	0	0	0	0	
3	0	0	0	0	
4	0	0	0	0	

```
     Vanilla Frappuccino  Cherry Soda  Single Espresso
0                      0            0                0
1                      0            0                0
2                      0            0                0
3                      1            0                0
4                      0            0                0

[5 rows x 51 columns]
```

[3]:
```
#Calculate correlation across the entire dataset
correlation = bakery.corr()
correlation.head()
```

[3]:
```
                    Transaction Number  Chocolate Cake  Lemon Cake  \
Transaction Number            1.000000       -0.002249    0.001587
Chocolate Cake               -0.002249        1.000000   -0.030612
Lemon Cake                    0.001587       -0.030612    1.000000
Casino Cake                  -0.006502        0.401565   -0.030430
Opera Cake                   -0.001879       -0.042011   -0.029621


                    Casino Cake  Opera Cake  Strawberry Cake  Truffle Cake  \
Transaction Number    -0.006502   -0.001879        -0.008055      0.005930
Chocolate Cake         0.401565   -0.042011        -0.037961     -0.032316
Lemon Cake            -0.030430   -0.029621        -0.030727     -0.026778
Casino Cake            1.000000   -0.037037        -0.029364     -0.029421
Opera Cake            -0.037037    1.000000        -0.041277     -0.030118


                    Chocolate Eclair  Coffee Eclair  Vanilla Eclair  …  \
Transaction Number         -0.000774      -0.002197        0.002456  …
Chocolate Cake             -0.010160      -0.059381       -0.009168  …
Lemon Cake                  0.003175      -0.048956       -0.001108  …
Casino Cake                -0.003114      -0.050073       -0.000807  …
Opera Cake                 -0.006484      -0.056926       -0.010784  …


                    Lemon Lemonade  Raspberry Lemonade  Orange Juice  \
Transaction Number       -0.002323            0.001561     -0.000950
Chocolate Cake           -0.039086           -0.039584     -0.041921
Lemon Cake               -0.026731           -0.030828     -0.025465
Casino Cake              -0.034711           -0.031449     -0.037973
Opera Cake               -0.035087           -0.036938     -0.033854


                    Green Tea  Bottled Water  Hot Coffee  Chocolate Coffee  \
Transaction Number  -0.001296      -0.002288    0.001999          0.003000
Chocolate Cake      -0.033524      -0.029809   -0.055919          0.485566
Lemon Cake          -0.023632      -0.030244   -0.047898         -0.032172
Casino Cake         -0.029165      -0.025159   -0.047462          0.398779
Opera Cake          -0.028021      -0.033523   -0.050908         -0.043968
```

23

```
                    Vanilla Frappuccino   Cherry Soda   Single Espresso
Transaction Number            -0.001759      0.004348          0.004523
Chocolate Cake                -0.037013     -0.030846         -0.031562
Lemon Cake                    -0.024287     -0.019121         -0.022823
Casino Cake                   -0.033103     -0.027864         -0.026831
Opera Cake                    -0.040874     -0.028131         -0.029624
```

```
[5 rows x 51 columns]
```

[4]: ```python
#The correlation between 'Chocolate Cake' & 'Choclate Coffee' is:
correlation.loc['Chocolate Cake','Chocolate Coffee']
```

[4]: 0.48556649252846507

[5]: ```python
#To understand if an item is symmetric or asymmetric, we need to utilize the
 ↪same measure twice, but with the items
#rerranged as so M(A,B) =? M(B,A)
correlation.loc['Chocolate Coffee', 'Chocolate Cake']
```

[5]: 0.48556649252846507

Because Correlation('Chocolate Cake','Chocolate Coffee') == Correlation(Chocolate Coffee', 'Chocolate Cake'), we can say that the items are symmetric binary values. As for the question in regards to the items being as association rules, I would answer yes. The correlation coefficient for binary and for association rules are hardly different.

For binary variables, correlation can be defined as:

$\phi = \frac{f_{11}*f_{00}-f_{01}*f_{10}}{sq.root(f_{1+}*f_{+1}*f_{0+}*f_{+0})}$

In terms of support measures:

$\phi = \frac{s(A,B)-s(A)*s(B)}{sq.root(s(A)*(1-s(A))*s(B)*(1-s(B)))}$

From this, we can see that the numerators are identical, thus the correlation coefficient with the support measures can be seen as a normalized version of the Pearson's version. Because of this, I would say that the association rules for the above items would have the same values for ϕ.

[ ]: