CS 634

MIDTERM PROJECT

Topic:

APRIORI ALGORITHM IMPLEMENTATION

Instructor:

Dr. Jason Wang

Submitted By:

IKJYOT SINGH GUJRAL

(isg6)

Definition

The Apriori Algorithm is an influential algorithm for mining frequent itemsets for Boolean association rules. Apriori uses a "bottom up" approach, where frequent subsets are extended one item at a time (a step known as candidate generation, and groups of candidates are tested against the data. Apriori is designed to operate on database containing transactions (for example, collections of items bought by customers, or details of a website frequentation).

The Algorithm

STEP 1: Scan the transaction data base to get the support of S each 1-itemset, compare S with min_sup, and get a support of 1-itemsets, L1

STEP 2: Use $L\ k-1$ join $L\ k-1$ to generate a set of candidate k-itemsets. And use Apriori property to prune the unfrequented k-itemsets from this set

STEP 3: Scan the transaction database to get the support S of each candidate k-itemset in the find set, compare S with min_sup, and get a set of frequent k-itemsets $L\ k$

STEP 4: The candidate set = Null

STEP 5: For each frequent itemset 1, generate all nonempty subsets of 1

STEP 6: For every nonempty subset s of 1, output the rule “s=>(1-s)” if confidence C of the rule “s=>(1-s)” (=support s of 1/support S of s)’ min_conf NO YES

Limitations

Apriori algorithm can be very slow and the bottleneck is candidate generation.
For example, if the transaction DB has 104 frequent 1-itemsets, they will generate 107 candidate 2-itemsets even after employing the downward closure.
To compute those with sup more than min sup, the database need to be scanned at every level. It needs (n +1) scans, where n is the length of the longest pattern.

Methods to Improve Apriori's Efficiency

• Hash-based itemset counting: A k-itemset whose corresponding hashing bucket count is below the threshold cannot be frequent
• Transaction reduction: A transaction that does not contain any frequent k-itemset is useless in subsequent scans
• Partitioning: Any itemset that is potentially frequent in DB must be frequent in at least one of the partitions of DB.
• Sampling: mining on a subset of given data, lower support threshold + a method to determine the completeness
• Dynamic itemset counting: add new candidate itemsets only when all of their subsets are estimated to be frequent

Apriori Advantages/Disadvantages

Advantages
• Uses large itemset property • Easily parallelized • Easy to implement
Disadvantages
• Assumes transaction database is memory resident. • Requires many database scans

Implementation Details

In this project, The Apriori Algorithms has been implemented using Python Programming Language. The reason to choose Python is its simplicity, readability and flexibility. Python allows the user to implement complex logics using simple codes.

The entire project has been divided into two files:

1. The first file is responsible to generate transaction .csv files out of .txt files containing list of 10 items. There are 5 .csv files that are generated, each containing 20 transactions.
2. The second file itself is divided in two code blocks:
    a. Generate frequent itemsets for a database of 20 transactions based on the user defined support value.
    b. Generate Association rules out of the frequent itemsets based on the user defined confidence value.

Source Code: generate.py

Procedure

1. The user enters the name of the transactions database, he/she wants to perform the Apriori algorithm on. First, the user chooses tr_db1.csv.

```
 <1 internal call>
Following is a list of transactions files. Choose the one you want to implement Apriori
 Algorithm on:
tr_db1.csv
tr_db2.csv
tr_db3.csv
tr_db4.csv
tr_db5.csv
Enter the name of the file you want to process: tr_db1.csv
```

2. Hitting Enter displays all the transactions in tr_db1.csv database. Next, the user is asked to enter a support and a confidence value.

```
**************************** File:  tr_db1.csv  ****************************

***************** TRANSACTIONS IN  tr_db1.csv  ARE AS FOLLOWS *****************

['pants', 'scotch tape', 'magnet', 'radio', 'headphones', 'stockings']
['stockings', 'pants', 'headphones', 'magnet', 'sand paper']
['magnet', 'pants', 'scotch tape']
['headphones', 'radio', 'phone', 'sand paper', 'mop']
['thermostat', 'scotch tape', 'mop', 'phone']
['thermostat', 'scotch tape']
['sand paper', 'magnet']
['headphones', 'mop', 'sand paper', 'magnet']
['mop', 'headphones']
['pants', 'magnet', 'phone']
['sand paper', 'phone']
['phone', 'thermostat']
['phone', 'scotch tape', 'radio', 'thermostat', 'headphones', 'stockings']
['thermostat', 'phone']
['stockings', 'magnet', 'thermostat', 'headphones', 'radio']
['sand paper', 'phone', 'pants', 'thermostat', 'radio', 'stockings']
['phone', 'mop', 'radio', 'stockings', 'headphones']
['phone', 'pants', 'headphones', 'stockings', 'magnet', 'thermostat']
['headphones', 'stockings', 'mop']
['scotch tape', 'phone', 'sand paper']

************************ TRANSACTIONS ARE ABOVE ****************************

Please Enter Support Value in % : 20
Please Enter Confidence Value in % : 50
```

3. Hitting enter then displays the user defined attributes and finds out all the frequent itemsets. The frequent itemsets here range from a length of 1 to 4 items.

```
Please Enter Support Value in % : 20
Please Enter Confidence Value in % : 50
********** USER DEFINED ATTRIBUTES BELOW  **********

        SUPPORT:  20 %                      CONFIDENCE:  50 %

********** USER DEFINED ATTRIBUTES ABOVE **********

********** FREQUENT ITEMS BELOW **********

[['radio'], ['phone'], ['scotch tape'], ['headphones'], ['thermostat'], ['pants'],
 ['sand paper'], ['magnet'], ['mop'], ['stockings'], ['mop', 'headphones'],
 ['radio', 'headphones'], ['phone', 'stockings'], ['radio', 'phone'], ['magnet',
 'pants'], ['stockings', 'headphones'], ['magnet', 'headphones'], ['pants',
 'stockings'], ['thermostat', 'phone'], ['magnet', 'stockings'], ['phone',
 'headphones'], ['radio', 'stockings'], ['phone', 'sand paper'], ['thermostat',
 'stockings'], ['radio', 'stockings', 'headphones'], ['magnet', 'stockings',
 'headphones'], ['stockings', 'magnet', 'pants', 'headphones'], ['headphones',
 'magnet', 'pants', 'stockings']]

********** FREQUENT ITEMS ABOVE **********

********** ASSOCIATION RULES BELOW **********
```

4. Below the frequent item sets, a list of all the association rules that can be created out of the frequent itemsets, are listed. Only those rules are displayed which satisfy the user defined support and confidence values.

```
************************ ASSOCIATION RULES BELOW ****************************

['mop']  ->  ['headphones']  Support:  25.0  Confidence:  83.3
['headphones']  ->  ['mop']  Support:  25.0  Confidence:  50.0
['radio']  ->  ['headphones']  Support:  25.0  Confidence:  83.3
['headphones']  ->  ['radio']  Support:  25.0  Confidence:  50.0
['stockings']  ->  ['phone']  Support:  20.0  Confidence:  50.0
['radio']  ->  ['phone']  Support:  20.0  Confidence:  66.7
['magnet']  ->  ['pants']  Support:  25.0  Confidence:  62.5
['pants']  ->  ['magnet']  Support:  25.0  Confidence:  83.3
['stockings']  ->  ['headphones']  Support:  35.0  Confidence:  87.5
['headphones']  ->  ['stockings']  Support:  35.0  Confidence:  70.0
['magnet']  ->  ['headphones']  Support:  25.0  Confidence:  62.5
['headphones']  ->  ['magnet']  Support:  25.0  Confidence:  50.0
['pants']  ->  ['stockings']  Support:  20.0  Confidence:  66.7
['stockings']  ->  ['pants']  Support:  20.0  Confidence:  50.0
['thermostat']  ->  ['phone']  Support:  30.0  Confidence:  75.0
['phone']  ->  ['thermostat']  Support:  30.0  Confidence:  54.5
['magnet']  ->  ['stockings']  Support:  20.0  Confidence:  50.0
['stockings']  ->  ['magnet']  Support:  20.0  Confidence:  50.0
['radio']  ->  ['stockings']  Support:  25.0  Confidence:  83.3
['stockings']  ->  ['radio']  Support:  25.0  Confidence:  62.5
['sand paper']  ->  ['phone']  Support:  20.0  Confidence:  57.1
['thermostat']  ->  ['stockings']  Support:  20.0  Confidence:  50.0
['stockings']  ->  ['thermostat']  Support:  20.0  Confidence:  50.0
['radio']  ->  ['stockings', 'headphones']  Support:  20.0  Confidence:  66.7
['stockings']  ->  ['radio', 'headphones']  Support:  20.0  Confidence:  50.0
['radio', 'stockings']  ->  ['headphones']  Support:  20.0  Confidence:  80.0
```

```
['radio']  ->  ['stockings', 'headphones']  Support:  20.0  Confidence:  66.7
['stockings']  ->  ['radio', 'headphones']  Support:  20.0  Confidence:  50.0
['radio', 'stockings']  ->  ['headphones']  Support:  20.0  Confidence:  80.0
['radio', 'headphones']  ->  ['stockings']  Support:  20.0  Confidence:  80.0
['stockings', 'headphones']  ->  ['radio']  Support:  20.0  Confidence:  57.1
['magnet']  ->  ['stockings', 'headphones']  Support:  20.0  Confidence:  50.0
['stockings']  ->  ['magnet', 'headphones']  Support:  20.0  Confidence:  50.0
['magnet', 'stockings']  ->  ['headphones']  Support:  20.0  Confidence:  100.0
['magnet', 'headphones']  ->  ['stockings']  Support:  20.0  Confidence:  80.0
['stockings', 'headphones']  ->  ['magnet']  Support:  20.0  Confidence:  57.1

************************ ASSOCIATION RULES ABOVE ****************************

Process finished with exit code 0
```

5. Once we are done with the first database, we move on to the next database i.e. tr_db2.csv.

```
 <1 internal call>
Following is a list of transactions files. Choose the one you want to implement Apriori
 Algorithm on:
tr_db1.csv
tr_db2.csv
tr_db3.csv
tr_db4.csv
tr_db5.csv
Enter the name of the file you want to process: tr_db2.csv
```

6. Like before, all the transactions in tr_db2.csv database are displayed and the user is asked to give a support and a confidence value.

```
*************************** File:  tr_db2.csv  ***************************

***************** TRANSACTIONS IN  tr_db2.csv  ARE AS FOLLOWS ****************

['picture frame', 'playing card', 'lip gloss', 'pen', 'headphones', 'bananas']
['chocolate', 'picture frame', 'lip gloss', 'pen', 'playing card']
['bananas', 'chocolate', 'playing card', 'lip gloss']
['headphones', 'pen', 'pencil', 'piano', 'lip gloss', 'bananas']
['piano', 'pen', 'bananas', 'chocolate', 'headphones', 'lip gloss']
['pencil', 'pen']
['bananas', 'playing card', 'headphones']
['pen', 'playing card', 'piano', 'pencil']
['headphones', 'bananas']
['bananas', 'headphones', 'picture frame', 'pencil']
['sticky note', 'chocolate', 'playing card']
['pen', 'headphones', 'lip gloss']
['pencil', 'picture frame', 'headphones', 'bananas', 'pen', 'sticky note']
['headphones', 'pencil', 'lip gloss', 'chocolate', 'playing card', 'picture frame']
['playing card', 'chocolate']
['playing card', 'sticky note', 'picture frame', 'chocolate']
['bananas', 'pencil', 'sticky note', 'chocolate', 'pen']
['sticky note', 'chocolate']
['chocolate', 'picture frame', 'piano', 'playing card', 'lip gloss', 'sticky note']
['pen', 'playing card', 'picture frame', 'headphones']

********************** TRANSACTIONS ARE ABOVE ****************************

Please Enter Support Value in % : 30
Please Enter Confidence Value in % : 60
```

7. Once the user input is over, the program calculates all the frequent itemsets that satisfy the user defined support. Also, all the association rules that can be generated out of these frequent itemsets and satisfy the user defined confidence value, are displayed. As compared to the previous database, the number of association rules have reduced as both support and confidence have been increased.

```
Please Enter Support Value in % : 30
Please Enter Confidence Value in % : 60
******************** USER DEFINED ATTRIBUTES BELOW  **********************

         SUPPORT:  30 %                      CONFIDENCE:  60 %

********************** USER DEFINED ATTRIBUTES ABOVE **********************

************************** FREQUENT ITEMS BELOW ***************************

[['pen'], ['pencil'], ['picture frame'], ['playing card'], ['chocolate'], ['bananas'], ['lip gloss'],
 ['headphones'], ['sticky note'], ['headphones', 'bananas'], ['chocolate', 'playing card'], ['picture
 frame', 'playing card'], ['pen', 'headphones']]

************************** FREQUENT ITEMS ABOVE ***************************

************************ ASSOCIATION RULES BELOW **************************

['headphones']  ->  ['bananas']  Support:  35.0  Confidence:  70.0
['bananas']  ->  ['headphones']  Support:  35.0  Confidence:  77.8
['chocolate']  ->  ['playing card']  Support:  35.0  Confidence:  70.0
['playing card']  ->  ['chocolate']  Support:  35.0  Confidence:  63.6
['picture frame']  ->  ['playing card']  Support:  30.0  Confidence:  75.0
['pen']  ->  ['headphones']  Support:  30.0  Confidence:  60.0
['headphones']  ->  ['pen']  Support:  30.0  Confidence:  60.0

************************ ASSOCIATION RULES ABOVE **************************

Process finished with exit code 0
```

8. Now we take the 3rd database i.e. tr_db3.csv.

```
 <1 internal call>
Following is a list of transactions files. Choose the one you want to implement Apriori
 Algorithm on:
tr_db1.csv
tr_db2.csv
tr_db3.csv
tr_db4.csv
tr_db5.csv
Enter the name of the file you want to process: tr_db3.csv
```

9. All the transaction in this database are displayed and the support and confidence values are entered.

```
**************************** File:  tr_db3.csv  ****************************

***************** TRANSACTIONS IN  tr_db3.csv  ARE AS FOLLOWS ****************

['shirt', 'tv', 'tooth picks', 'scotch tape']
['ring', 'tv']
['scotch tape', 'tv', 'seat belt', 'tooth picks', 'pillow', 'shirt']
['seat belt', 'cup', 'pillow', 'tv', 'ring']
['ring', 'scotch tape', 'pillow', 'packing peanuts', 'shirt']
['pillow', 'seat belt', 'sticky note']
['tooth picks', 'pillow', 'cup', 'seat belt', 'packing peanuts']
['cup', 'sticky note', 'packing peanuts', 'pillow']
['packing peanuts', 'shirt']
['shirt', 'seat belt', 'cup', 'sticky note', 'packing peanuts']
['tv', 'scotch tape', 'pillow', 'sticky note', 'ring']
['shirt', 'scotch tape']
['cup', 'ring', 'pillow', 'shirt']
['packing peanuts', 'pillow', 'scotch tape', 'tv']
['pillow', 'cup']
['shirt', 'seat belt', 'ring']
['tv', 'cup', 'tooth picks', 'pillow']
['ring', 'cup']
['scotch tape', 'tv', 'packing peanuts', 'sticky note']
['sticky note', 'pillow']

************************ TRANSACTIONS ARE ABOVE ****************************

Please Enter Support Value in % : 40
Please Enter Confidence Value in % : 85
********************* USER DEFINED ATTRIBUTES BELOW  ***********************

         SUPPORT:  40 %                        CONFIDENCE:  85 %
```

10. Surprisingly, there are only single item frequent itemsets and also there are no association rules that satisfy the user defined support of 40% and confidence of 85%.

```
********************* USER DEFINED ATTRIBUTES BELOW  ***********************

         SUPPORT:  40 %                       CONFIDENCE:  85 %

********************** USER DEFINED ATTRIBUTES ABOVE ***********************

************************** FREQUENT ITEMS BELOW ****************************

[['cup'], ['pillow'], ['tv'], ['shirt']]

************************** FREQUENT ITEMS ABOVE ****************************

************************ ASSOCIATION RULES BELOW ***************************

No Rules that satisfy the conditions

************************ ASSOCIATION RULES ABOVE ***************************

Process finished with exit code 0
```

11. We now move to the 4th database i.e. tr_db4.csv.

```
Following is a list of transactions files. Choose the one you want to implement Apriori Algorithm on:
tr_db1.csv
tr_db2.csv
tr_db3.csv
tr_db4.csv
tr_db5.csv
Enter the name of the file you want to process: tr_db4.csv
```

12. All transactions in tr_db4.csv database are displayed and user provides support and confidence values.

```
************************** File:  tr_db4.csv  ***************************

***************** TRANSACTIONS IN  tr_db4.csv  ARE AS FOLLOWS ****************

['key chain', 'rubber duck']
['rubber duck', 'helmet', 'cookie jar', 'rusty nail', 'bottle']
['sailboat', 'bottle', 'cookie jar', 'key chain', 'helmet']
['bottle', 'pants', 'sailboat', 'rubber duck']
['cookie jar', 'rusty nail', 'pants', 'bottle', 'twezzer']
['rubber duck', 'twezzer', 'helmet']
['sailboat', 'helmet']
['key chain', 'twezzer', 'bottle', 'helmet', 'rusty nail']
['rusty nail', 'pants']
['clock', 'pants', 'bottle']
['helmet', 'twezzer']
['sailboat', 'cookie jar', 'pants', 'rubber duck']
['helmet', 'rusty nail', 'cookie jar']
['sailboat', 'pants', 'key chain', 'twezzer']
['sailboat', 'pants', 'cookie jar', 'clock', 'rubber duck']
['twezzer', 'cookie jar']
['clock', 'helmet', 'cookie jar', 'sailboat', 'pants']
['bottle', 'pants', 'rusty nail', 'key chain']
['helmet', 'rubber duck', 'sailboat', 'twezzer']
['pants', 'bottle', 'cookie jar']

************************ TRANSACTIONS ARE ABOVE ***************************

Please Enter Support Value in % : 20
Please Enter Confidence Value in % : 60
********************* USER DEFINED ATTRIBUTES BELOW  ***********************

        SUPPORT:  20 %                    CONFIDENCE:  60 %
```

13. Now we get frequent item sets whose length ranges from 1 to 3 items per set. We see that the confidence value of 60 % does cut off a few association rules that might have been generated from the frequent itemsets.

```
        SUPPORT:  20 %                    CONFIDENCE:  60 %

********** USER DEFINED ATTRIBUTES ABOVE **********

************************* FREQUENT ITEMS BELOW *************************

[['rusty nail'], ['sailboat'], ['rubber duck'], ['twezzer'], ['key chain'], ['bottle'], ['cookie jar'], ['pants'],
 ['helmet'], ['pants', 'cookie jar'], ['bottle', 'rusty nail'], ['sailboat', 'rubber duck'], ['cookie jar',
 'helmet'], ['sailboat', 'cookie jar'], ['pants', 'bottle'], ['twezzer', 'helmet'], ['cookie jar', 'bottle'],
 ['sailboat', 'helmet'], ['pants', 'sailboat'], ['pants', 'cookie jar', 'sailboat'], ['sailboat', 'cookie jar',
 'helmet'], ['pants', 'sailboat', 'cookie jar'], ['cookie jar', 'sailboat', 'helmet']]

************************* FREQUENT ITEMS ABOVE *************************

************************ ASSOCIATION RULES BELOW ************************

['rusty nail']  ->  ['bottle']  Support:  20.0  Confidence:  66.7
['bottle']  ->  ['pants']  Support:  25.0  Confidence:  62.5
['sailboat']  ->  ['pants']  Support:  25.0  Confidence:  62.5

********************** ASSOCIATION RULES ABOVE ************************

Process finished with exit code 0
```

14. The last database is tr_db5.csv.

```
Following is a list of transactions files. Choose the one you want to implement Apriori
 Algorithm on:
tr_db1.csv
tr_db2.csv
tr_db3.csv
tr_db4.csv
tr_db5.csv
Enter the name of the file you want to process: tr_db5.csv

************************** File:  tr_db5.csv  **************************
```

15. All transactions in this database are listed and the user is asked to enter a support and confidence value. Note that the support value is the minimum of all the support values entered. We can expect to see a lot of frequent itemsets.

```
****************** TRANSACTIONS IN  tr_db5.csv  ARE AS FOLLOWS ****************

['sponge', 'shoe lace', 'sailboat', 'floor']
['lip gloss', 'CD']
['sponge', 'CD', 'doll', 'sailboat']
['hanger', 'sailboat', 'lip gloss', 'shoe lace', 'twezzers', 'floor']
['shoe lace', 'floor', 'washing machine']
['sponge', 'twezzers', 'sailboat']
['CD', 'twezzers']
['shoe lace', 'sponge', 'twezzers', 'sailboat', 'washing machine']
['floor', 'sailboat', 'sponge', 'lip gloss', 'CD', 'washing machine']
['sailboat', 'sponge', 'washing machine', 'doll', 'floor']
['sponge', 'washing machine', 'CD']
['floor', 'CD', 'lip gloss', 'sailboat', 'doll', 'twezzers']
['doll', 'lip gloss', 'hanger', 'shoe lace']
['CD', 'sponge']
['lip gloss', 'shoe lace', 'floor', 'twezzers']
['hanger', 'washing machine']
['floor', 'twezzers']
['hanger', 'sailboat', 'doll', 'floor', 'twezzers']
['shoe lace', 'sponge', 'CD']
['washing machine', 'twezzers']

************************ TRANSACTIONS ARE ABOVE ****************************

Please Enter Support Value in % : 15
Please Enter Confidence Value in % : 60
```

16. As expected, we get many frequent item sets whose length ranges from 1 to 4 items per set.

```
******************** USER DEFINED ATTRIBUTES BELOW  **********************

        SUPPORT:  15 %                    CONFIDENCE:  60 %

********************** USER DEFINED ATTRIBUTES ABOVE **********************

************************* FREQUENT ITEMS BELOW ****************************

[['shoe lace'], ['doll'], ['twezzers'], ['washing machine'], ['CD'], ['floor'], ['sponge'], ['sailboat'],
 ['hanger'], ['lip gloss'], ['lip gloss', 'floor'], ['sailboat', 'lip gloss'], ['washing machine',
 'floor'], ['twezzers', 'floor'], ['sponge', 'washing machine'], ['CD', 'sponge'], ['shoe lace',
 'sailboat'], ['twezzers', 'lip gloss'], ['sailboat', 'floor'], ['shoe lace', 'twezzers'], ['sponge',
 'sailboat'], ['shoe lace', 'sponge'], ['shoe lace', 'lip gloss'], ['sailboat', 'washing machine'],
 ['twezzers', 'sailboat'], ['CD', 'sailboat'], ['sponge', 'floor'], ['CD', 'lip gloss'], ['shoe lace',
 'floor'], ['doll', 'floor'], ['doll', 'sailboat'], ['CD', 'lip gloss', 'floor', 'sailboat'], ['shoe lace',
 'sponge', 'floor', 'sailboat'], ['sponge', 'washing machine', 'sailboat'], ['shoe lace', 'twezzers',
 'floor'], ['shoe lace', 'lip gloss', 'floor', 'twezzers'], ['sailboat', 'doll', 'floor'], ['sailboat',
 'floor', 'lip gloss', 'twezzers'], ['twezzers', 'floor', 'sailboat', 'lip gloss'], ['shoe lace',
 'sailboat', 'sponge'], ['shoe lace', 'sponge', 'sailboat'], ['shoe lace', 'sailboat', 'twezzers',
 'floor'], ['sailboat', 'floor', 'twezzers'], ['sailboat', 'lip gloss', 'twezzers'], ['twezzers', 'floor',
 'sailboat', 'doll'], ['twezzers', 'floor', 'lip gloss'], ['sailboat', 'twezzers', 'floor', 'doll'], ['shoe
 lace', 'floor', 'twezzers'], ['CD', 'sailboat', 'floor', 'lip gloss'], ['sponge', 'floor', 'sailboat'],
 ['shoe lace', 'floor', 'sailboat', 'sponge'], ['sailboat', 'twezzers', 'floor'], ['shoe lace', 'floor',
 'lip gloss'], ['shoe lace', 'sailboat', 'floor', 'twezzers'], ['shoe lace', 'twezzers', 'floor',
 'sailboat'], ['sailboat', 'sponge', 'washing machine', 'floor'], ['shoe lace', 'twezzers', 'lip gloss'],
 ['shoe lace', 'lip gloss', 'twezzers'], ['shoe lace', 'sailboat', 'floor', 'sponge'], ['shoe lace',
 'sailboat', 'floor'], ['sponge', 'washing machine', 'sailboat', 'floor'], ['shoe lace', 'twezzers',
 'floor', 'lip gloss'], ['sailboat', 'floor', 'doll'], ['sailboat', 'washing machine', 'floor', 'sponge'],
 ['sailboat', 'lip gloss', 'floor', 'twezzers'], ['shoe lace', 'lip gloss', 'floor'], ['shoe lace',
 'floor', 'sailboat'], ['lip gloss', 'floor', 'twezzers'], ['sailboat', 'floor', 'sponge'], ['sailboat',
 'lip gloss', 'floor'], ['twezzers', 'sailboat', 'lip gloss'], ['sailboat', 'floor', 'lip gloss'], ['CD',
 'lip gloss', 'floor', 'sailboat'], ['sailboat', 'floor', 'lip gloss', 'twezzers'], ['twezzers', 'floor',
```

```
'lip gloss', 'floor', 'sailboat'], ['sailboat', 'floor', 'lip gloss', 'twezzers'], ['twezzers', 'floor',
'sailboat', 'lip gloss'], ['sailboat', 'floor', 'twezzers'], ['sailboat', 'twezzers', 'floor', 'lip
gloss'], ['sailboat', 'sponge', 'floor'], ['twezzers', 'floor', 'lip gloss'], ['sailboat', 'twezzers',
'floor', 'doll'], ['CD', 'sailboat', 'floor', 'lip gloss'], ['shoe lace', 'floor', 'sailboat', 'sponge'],
['shoe lace', 'sailboat', 'floor', 'twezzers'], ['sailboat', 'shoe lace', 'floor', 'sponge'], ['sailboat',
'sponge', 'washing machine', 'floor'], ['sailboat', 'shoe lace', 'floor', 'twezzers'], ['shoe lace',
'sailboat', 'sponge', 'floor'], ['sponge', 'washing machine', 'sailboat', 'floor'], ['shoe lace', 'floor',
'sailboat', 'twezzers'], ['sailboat', 'floor', 'doll', 'twezzers'], ['shoe lace', 'twezzers', 'floor',
'lip gloss'], ['sailboat', 'floor', 'doll'], ['sailboat', 'washing machine', 'floor', 'sponge'],
['sailboat', 'lip gloss', 'floor', 'twezzers'], ['shoe lace', 'floor', 'lip gloss', 'twezzers'],
['sailboat', 'floor', 'lip gloss'], ['CD', 'lip gloss', 'floor', 'sailboat'], ['sailboat', 'twezzers',
'floor', 'lip gloss'], ['sponge', 'washing machine', 'sailboat', 'floor'], ['sailboat', 'floor', 'lip
gloss', 'twezzers'], ['CD', 'sailboat', 'floor', 'lip gloss'], ['shoe lace', 'twezzers', 'floor', 'lip
gloss'], ['sailboat', 'washing machine', 'floor', 'sponge'], ['shoe lace', 'floor', 'lip gloss',
'twezzers']]

************************* FREQUENT ITEMS ABOVE ****************************
```

17. Also, we see that many association rules are generated, that satisfy the user defined support and confidence values.

```
************************* ASSOCIATION RULES BELOW ****************************

['lip gloss']  ->  ['floor']  Support:  20.0  Confidence:  66.7
['CD']  ->  ['sponge']  Support:  25.0  Confidence:  62.5
['sailboat']  ->  ['floor']  Support:  30.0  Confidence:  66.7
['floor']  ->  ['sailboat']  Support:  30.0  Confidence:  66.7
['sponge']  ->  ['sailboat']  Support:  30.0  Confidence:  66.7
['sailboat']  ->  ['sponge']  Support:  30.0  Confidence:  66.7
['doll']  ->  ['floor']  Support:  15.0  Confidence:  60.0
['doll']  ->  ['sailboat']  Support:  20.0  Confidence:  80.0
['sponge', 'washing machine']  ->  ['sailboat']  Support:  15.0  Confidence:  75.0
['washing machine', 'sailboat']  ->  ['sponge']  Support:  15.0  Confidence:  100.0
['doll']  ->  ['sailboat', 'floor']  Support:  15.0  Confidence:  60.0
['sailboat', 'doll']  ->  ['floor']  Support:  15.0  Confidence:  75.0
['doll', 'floor']  ->  ['sailboat']  Support:  15.0  Confidence:  100.0
['sailboat', 'twezzers']  ->  ['floor']  Support:  15.0  Confidence:  60.0
['floor', 'twezzers']  ->  ['sailboat']  Support:  15.0  Confidence:  60.0
['twezzers', 'floor']  ->  ['lip gloss']  Support:  15.0  Confidence:  60.0
['twezzers', 'lip gloss']  ->  ['floor']  Support:  15.0  Confidence:  100.0
['floor', 'lip gloss']  ->  ['twezzers']  Support:  15.0  Confidence:  75.0
['sponge', 'floor']  ->  ['sailboat']  Support:  15.0  Confidence:  100.0
['sailboat', 'twezzers']  ->  ['floor']  Support:  15.0  Confidence:  60.0
['twezzers', 'floor']  ->  ['sailboat']  Support:  15.0  Confidence:  60.0
['doll']  ->  ['sailboat', 'floor']  Support:  15.0  Confidence:  60.0
['sailboat', 'doll']  ->  ['floor']  Support:  15.0  Confidence:  75.0
['floor', 'doll']  ->  ['sailboat']  Support:  15.0  Confidence:  100.0
['lip gloss', 'floor']  ->  ['twezzers']  Support:  15.0  Confidence:  75.0
['lip gloss', 'twezzers']  ->  ['floor']  Support:  15.0  Confidence:  100.0
['floor', 'twezzers']  ->  ['lip gloss']  Support:  15.0  Confidence:  60.0
['floor', 'sponge']  ->  ['sailboat']  Support:  15.0  Confidence:  100.0
['sailboat', 'lip gloss']  ->  ['floor']  Support:  15.0  Confidence:  100.0
['lip gloss', 'floor']  ->  ['sailboat']  Support:  15.0  Confidence:  75.0
```

```
['lip gloss', 'floor']  ->  ['sailboat']  Support:  15.0  Confidence:  75.0
['sailboat', 'lip gloss']  ->  ['floor']  Support:  15.0  Confidence:  100.0
['floor', 'lip gloss']  ->  ['sailboat']  Support:  15.0  Confidence:  75.0
['sailboat', 'twezzers']  ->  ['floor']  Support:  15.0  Confidence:  60.0
['floor', 'twezzers']  ->  ['sailboat']  Support:  15.0  Confidence:  60.0
['sponge', 'floor']  ->  ['sailboat']  Support:  15.0  Confidence:  100.0
['twezzers', 'floor']  ->  ['lip gloss']  Support:  15.0  Confidence:  60.0
['twezzers', 'lip gloss']  ->  ['floor']  Support:  15.0  Confidence:  100.0
['floor', 'lip gloss']  ->  ['twezzers']  Support:  15.0  Confidence:  75.0
['doll']  ->  ['sailboat', 'floor']  Support:  15.0  Confidence:  60.0
['sailboat', 'doll']  ->  ['floor']  Support:  15.0  Confidence:  75.0
['floor', 'doll']  ->  ['sailboat']  Support:  15.0  Confidence:  100.0
['sailboat', 'lip gloss']  ->  ['floor']  Support:  15.0  Confidence:  100.0
['floor', 'lip gloss']  ->  ['sailboat']  Support:  15.0  Confidence:  75.0

************************ ASSOCIATION RULES ABOVE ****************************

Process finished with exit code 0
```

Conclusion

This concludes are testing of databases. We see that for different transaction and different values of support and confidence, the number of association rules can vary a lot. For the same database, reducing the support and confidence values will always increase the number of association rules that satisfy those values.

All Test Results