Project 3: Association Rules (Data Mining)

Name – Navtejinder Singh Brar UTA ID – 1001874259

Name – Vamsi Krishna Pentakota UTA ID – 1001758105

Description: We are given with a dataset named "online-retail" from a UK based retail store which contains around 18485 transactions and it is an excel file. First, we imported this dataset into RStudio and then pre-processed the data. In the pre-preprocessing, first we removed all the rows that had missing values and after that we removed all the observations whose invoice number were starting with "C" then, we removed all the observations that contained the words provided in the project description from the column named "Description" and then we converted our dataset into transactions. After converting the dataset into transactions, we wrote that transactional data into a file and used that file to complete the tasks of the project. With that transaction file we found out the item sets i.e., both frequent (support is kept as 0.008, 0.011 and 0.015 as allocated to us in the project description and we are using max-length as 3, min length is kept 1) and candidate item sets (we are using support as 0.0001 for candidate). Moreover, we then applied the apriori algorithm to the transactions and found out all the rules (we are using support as 0.008, 0.011 and 0.015, confidence as 0.5, 0.7 and 0.8 and max length is 5, min length is kept 1). In addition to this, we have also filtered the rules that had lift both less than and greater than 10. Moreover, we have visualized all of the above that is Frequent and candidate item sets, Rules generated by apriori and rules filtered using lift and minimum confidence and many more. We will discuss all of the above in detail and do analysis and comparisons as we move forward with the report.

The picture shown below gives the summary about the transactions we are using in this project that is we have a total of 18485 rows that are invoice numbers and 7793 columns that are items (unique items).

```
> summary(tr)
transactions as itemMatrix in sparse format with
 18485 rows (elements/itemsets/transactions) and
 7793 columns (items) and a density of 0.002277027

most frequent items:
WHITE HANGING HEART T-LIGHT HOLDER          REGENCY CAKESTAND 3 TIER           JUMBO BAG RED RETROSPOT
                              1761                              1533                              1418
                     PARTY BUNTING      ASSORTED COLOUR BIRD ORNAMENT                            (Other)
                              1268                              1240                            320794

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   20   21   22   23   24   25
1556  849  762  773  744  703  644  643  655  586  598  535  494  513  550  521  453  441  482  410  387  312  304  262  241
  26   27   28   29   30   31   32   33   34   35   36   37   38   39   40   41   42   43   44   45   46   47   48   49   50
 250  229  216  224  211  160  164  135  139  139  102  115   86  113   91   92   87   89   66   60   69   61   63   54   49
  51   52   53   54   55   56   57   58   59   60   61   62   63   64   65   66   67   68   69   70   71   72   73   74   75
  64   42   42   46   44   37   28   38   32   27   27   18   24   25   20   26   24   22   16   20   18   14   15   16   11
  76   77   78   79   80   81   82   83   84   85   86   87   88   89   90   91   92   93   94   95   96   97   98   99  100
  15   12    7    9   14   15   12    9    9   10   11   14    8    7    4    7   10    6    4    4    3    6    5    2    4
 101  102  103  104  105  106  107  108  109  110  111  112  113  114  116  117  118  120  121  122  123  125  126  127  131
   2    4    4    3    2    2    6    3    4    3    2    1    3    1    3    3    3    1    2    2    1    3    2    2    1
 132  133  134  140  141  142  143  145  146  147  149  154  157  168  171  177  178  180  202  204  228  236  249  250  285
   1    2    1    1    2    2    1    1    2    1    1    3    2    2    2    1    1    1    1    1    1    1    1    1    1
 320  400  419
   1    1    1

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1.00    5.00   13.00   17.74   23.00  419.00

includes extended item information - examples:
                    labels
1                1 HANGER
2     10 COLOUR SPACEBOY PEN
3 12 COLOURED PARTY BALLOONS
```

Figure 1.

Division of Labor:

Name	Topics covered	Time Spent
Navtejinder Singh Brar	Pre-processing, 1st and 2nd	20 hours
Vamsi Krishna Pentakota	Pre-processing, 3rd and 4th	20 hours

Candidate (min_sup of 0.0001) and frequent Item sets (min_sup of 0.008, 0.011 and 0.015) min length = 1 and max length = 3 confidence =1:

For Candidate Item sets: We are using the minimum support as 0.0001 to generate the candidate item sets and we are getting around 18 million item sets (18945213 item sets exactly). The number of item sets are increasing in each iteration that is in 1st iteration the no. of item sets is 3733 (order 1), in 2nd the no. of item sets is 973878 (order 2), in 3rd the no. of item sets is 17967602 (order 3) .

```
> summary(itemsets)
set of 18945213 itemsets

most frequent items:
WHITE HANGING HEART T-LIGHT HOLDER           LUNCH BAG RED RETROSPOT   SET OF 3 CAKE TINS PANTRY DESIGN
                            280868                            226132                             204722
      SET OF 4 PANTRY JELLY MOULDS          REGENCY CAKESTAND 3 TIER                            (Other)
                            200589                            198560                           54743424

element (itemset/transaction) length distribution:sizes
       1        2        3
    3733   973878 17967602

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.000   3.000   3.000   2.948   3.000   3.000

summary of quality measures:
    support          transIdenticalToItemsets     count
 Min.   :0.0001082   Min.   :0.00e+00         Min.   :   2.000
 1st Qu.:0.0001082   1st Qu.:0.00e+00         1st Qu.:   2.000
 Median :0.0001082   Median :0.00e+00         Median :   2.000
 Mean   :0.0001469   Mean   :8.00e-09         Mean   :   2.716
 3rd Qu.:0.0001623   3rd Qu.:0.00e+00         3rd Qu.:   3.000
 Max.   :0.0952664   Max.   :3.57e-03         Max.   :1761.000

includes transaction ID lists: FALSE

mining info:
 data ntransactions support confidence
   tr         18485   1e-04          1
```

Figure 2

For Frequent Item sets (min_sup = 0.008): For the minimum support of 0.008 we are generating a total of 1061 item sets. The number of item sets are decreasing in every iteration as that is 1st iteration the no. of item sets is 651 (order 1), in 2nd the no. of item sets is 335 (order 2) and in 3rd the no. of item sets is 75 (order 3).

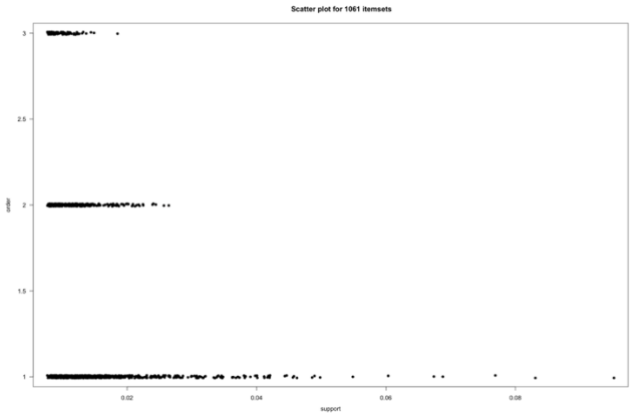


Figure 3

For Frequent Item sets (min_sup = 0.011): For the minimum support of 0.011 we are generating a total of 594 item sets. The number of item sets are decreasing in every iteration as that is 1st iteration the no. of item sets is 438 (order 1), in 2nd the no. of item sets is 139 (order 2) and in 3rd the no. of item sets is 17 (order 3).

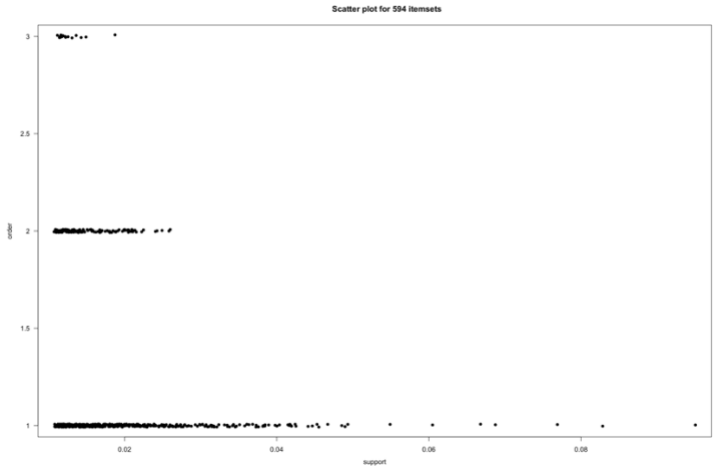


Figure 4

For Frequent Item sets (min_sup = 0.015): For the minimum support of 0.015 we are generating a total of 305 item sets. The number of item sets are decreasing in every iteration as that is 1st iteration the no. of item sets is 256(order 1), in 2nd the no. of item sets is 48 (order 2) and in 3rd the no. of item sets is 1 (order 3).

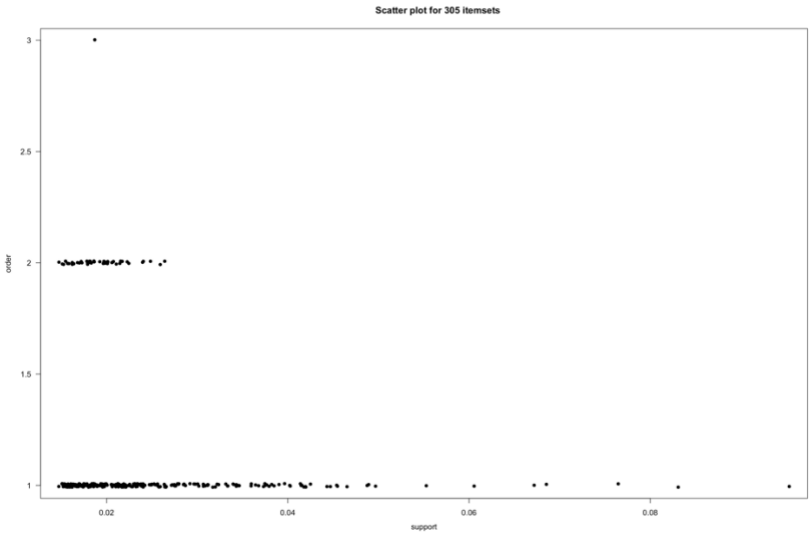
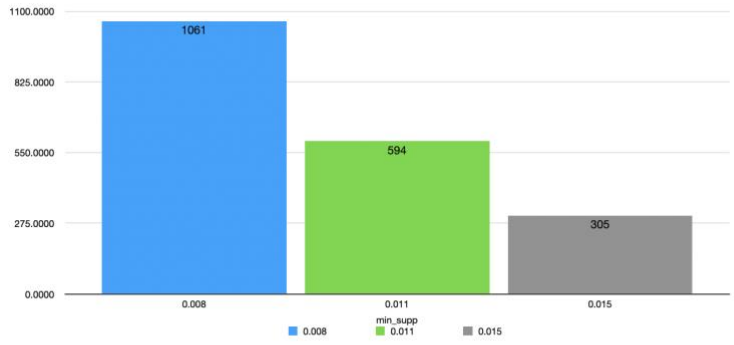


Figure 5

What we can analyze?

We can analyze that the number of candidate item sets are increasing drastically in every iteration whereas, the number of Frequent item sets are decreasing in every iteration. Moreover, for frequent item sets the number of item sets are decreasing as we increase the minimum support.



	0.008	0.011	0.015
Item sets	1061	594	305

Figure 6

Generating rules with Apriori algorithm using provided minimum support (0.008, 0.011 and 0.015) and minimum confidence (0.5, 0.7 and 0.8):

So, here we are going to discuss 9 different combinations of minimum support and minimum confidence which we have generated. The minimum length (1) and maximum length (5) is kept constant for all the combinations.

1) min_sup = 0.008 and min_conf = 0.5: A total of 335 rules were generated using these parameters. Around, 33 rules had lift less than 10 and 302 rules had lift greater than 10. The mean lift was 27.683 and the minimum lift was 6.336 and the maximum lift was 103.954. The no. of rules for rule length (lhs + rhs) 2, 3 and 4 are 164, 159 and 12. Plotting done at [row 1, column 1] of Figure 7.

2) min_sup = 0.011 and min_conf = 0.5: A total of 119 rules were generated using these parameters. Around, 21 rules had lift less than 10 and 98 rules had lift greater than 10. The mean lift was 21.528 and the minimum lift was 6.336 and the maximum lift was 83.643. The no. of rules for rule length (lhs + rhs) 2, 3 and 4 are 69, 46 and 4. . Plotting done at [row 1, column 2] of Figure 7.

3) min_sup = 0.015 and min_conf = 0.5: A total of 31 rules were generated using these parameters. Around, 10 rules had lift less than 10 and 21 rules had lift greater than 10. The mean lift was 15.578 and the minimum lift was 6.336 and the maximum lift was 26.366. The no. of rules for rule length (lhs + rhs) 2 and 3 are 28 and 3. . Plotting done at [row 1, column 3] of Figure 7.

4) min_sup = 0.008 and min_conf = 0.7: A total of 99 rules were generated using these parameters. Around, 4 rules had lift less than 10 and 95 rules had lift greater than 10. The mean lift was 49.981 and the minimum lift was 7.616 and the maximum lift was 103.954 .The no. of rules for rule length (lhs + rhs) 2, 3 and 4 are 55, 37 and 7. . Plotting done at [row 2, column 1] of Figure 7.

5) min_sup = 0.011 and min_conf = 0.7: A total of 35 rules were generated using these parameters. Around, 1 rules had lift less than 10 and 34 rules had lift greater than 10. The mean lift was 38.731 and the minimum lift was 7.616 and the maximum lift was 83.643.The no. of rules for rule length (lhs + rhs) 2, 3 and 4 are 19, 13 and 3. . Plotting done at [row 2, column 2] of Figure 7.

6) min_sup = 0.015 and min_conf = 0.7: A total of 7 rules were generated using these parameters. Around, 0 rules had lift less than 10 and 7 rules had lift greater than 10. The mean lift was 23.03 and the minimum lift was 20.17 and the maximum lift was 26.37.The no. of rules for rule length (lhs + rhs) 2 and 3 are 4 and 3. . Plotting done at [row 2, column 3] of Figure 7.

7) min_sup = 0.008 and min_conf = 0.8: A total of 49 rules were generated using these parameters. Around, 0 rules had lift less than 10 and 49 rules had lift greater than 10. The mean lift was 72.14 and the minimum lift was 10.56 and the maximum lift was 103.95. The no. of rules for rule length (lhs + rhs) 2, 3 and 4 are 32, 15 and 2. Plotting done at [row 3, column 1] of Figure 7.

8) min_sup = 0.011 and min_conf = 0.8: A total of 17 rules were generated using these parameters. Around, 0 rules had lift less than 10 and 17 rules had lift greater than 10. The mean lift was 48.67 and the minimum lift was 21.51 and the maximum lift was 83.64.The no. of rules for rule length (lhs + rhs) 2, 3 and 4 are 7,8 and 2. . Plotting done at [row 3, column 2] of Figure 7.

9) min_sup = 0.015 and min_conf = 0.8: A total of 3 rules were generated using these parameters. Around, 0 rules had lift less than 10 and 3 rules had lift greater than 10. The mean lift was 24.18 and the

minimum lift was 21.95 and the maximum lift was 26.37. The no. of rules for rule length (lhs + rhs) 2 and 3 are 1 and 2. . Plotting done at [row 3, column 3] of Figure 7.

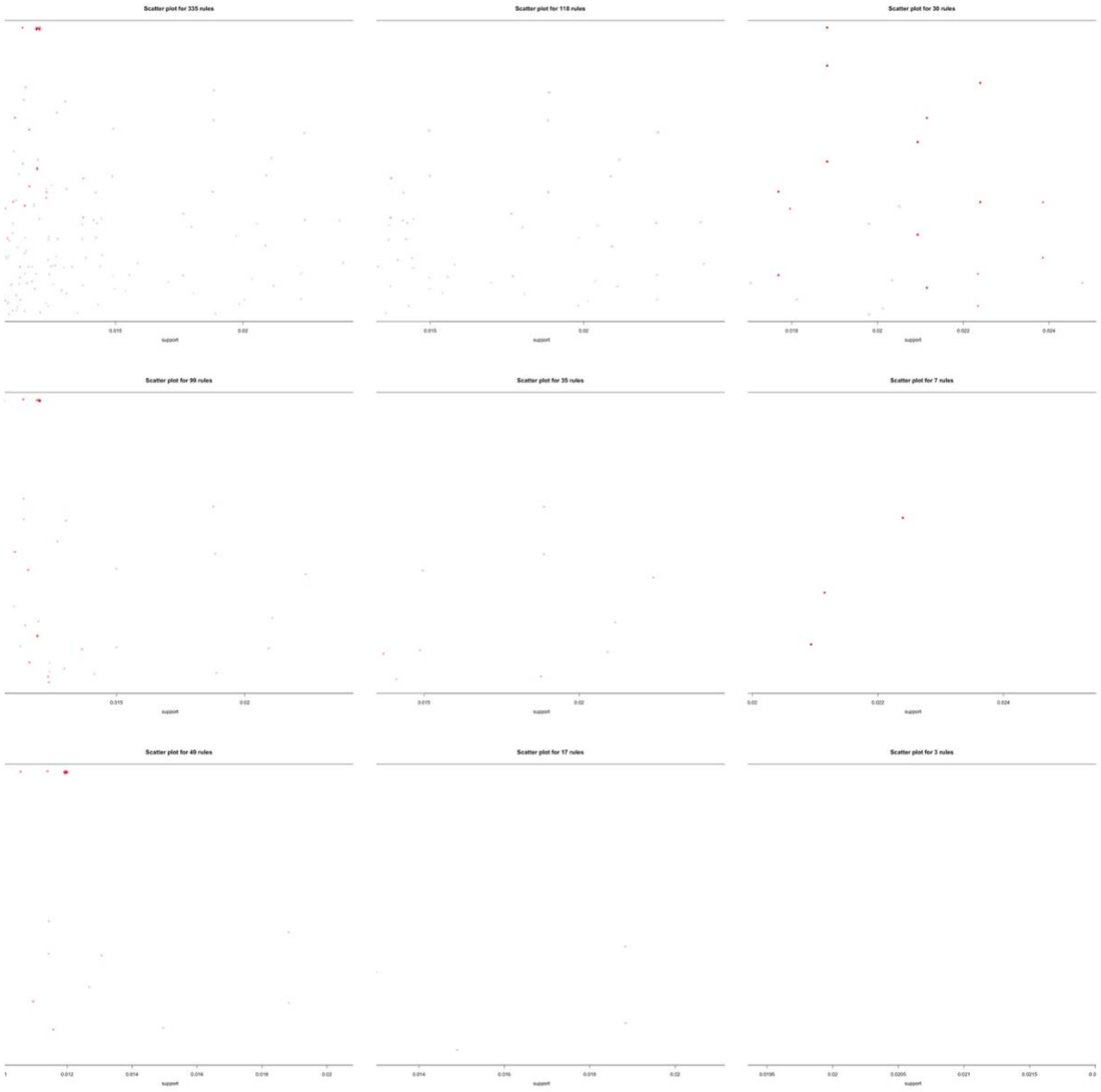


Figure 7

Filter 10 rules for lift > 10 and lift < 10:

1) min_sup = 0.008 and min_conf = 0.5: There are a total of 33 rules that have lift less than 10 and rules that have lift greater than 10 are 302. Rules that have a lift less than 10 have only one rule with order of 4 whereas, lift > 10 has 11 order 4 rules(here order is lhs+rhs).

```
> inspect(lift.subset1[1:10])
     lhs                                  rhs                                 support     confidence coverage   lift
[1]  {TEA TIME PARTY BUNTING}             => {PARTY BUNTING}                    0.008060590 0.5265018  0.01530971 7.675383
[2]  {REGENCY TEAPOT ROSES}               => {REGENCY CAKESTAND 3 TIER}         0.008926156 0.5288462  0.01687855 6.376857
[3]  {SET OF TEA COFFEE SUGAR TINS PANTRY} => {SET OF 3 CAKE TINS PANTRY DESIGN} 0.009899919 0.5479042  0.01806870 9.939165
[4]  {CANDLEHOLDER PINK HANGING HEART}    => {WHITE HANGING HEART T-LIGHT HOLDER} 0.012442521 0.7255521  0.01714904 7.616030
[5]  {JUMBO BAG SCANDINAVIAN BLUE PAISLEY} => {JUMBO BAG RED RETROSPOT}         0.012550717 0.5843829  0.02147687 7.617995
[6]  {JUMBO STORAGE BAG SKULLS}           => {JUMBO BAG RED RETROSPOT}          0.011035975 0.5087282  0.02169326 6.631763
[7]  {JUMBO BAG WOODLAND ANIMALS}         => {JUMBO BAG RED RETROSPOT}          0.012334325 0.5217391  0.02364079 6.801374
[8]  {JUMBO BAG SPACEBOY DESIGN}          => {JUMBO BAG RED RETROSPOT}          0.012550717 0.5536993  0.02266703 7.218005
[9]  {PINK REGENCY TEACUP AND SAUCER}     => {REGENCY CAKESTAND 3 TIER}         0.014931025 0.5454545  0.02737355 6.577122
[10] {JUMBO BAG PINK VINTAGE PAISLEY}     => {JUMBO BAG RED RETROSPOT}          0.015363808 0.5409524  0.02840141 7.051837
```

Figure 8

```
> inspect(lift.subset11[1:10])
     lhs                          rhs                          support     confidence coverage    lift      count
[1]  {HERB MARKER BASIL}       => {HERB MARKER THYME}       0.008006492 0.8969697  0.008926156 101.72077 148
[2]  {HERB MARKER THYME}       => {HERB MARKER BASIL}       0.008006492 0.9079755  0.008817961 101.72077 148
[3]  {HERB MARKER BASIL}       => {HERB MARKER PARSLEY}     0.008060590 0.9030303  0.008926156 103.04022 149
[4]  {HERB MARKER PARSLEY}     => {HERB MARKER BASIL}       0.008060590 0.9197531  0.008763863 103.04022 149
[5]  {HERB MARKER BASIL}       => {HERB MARKER ROSEMARY}    0.008222883 0.9212121  0.008926156 101.36075 152
[6]  {HERB MARKER ROSEMARY}    => {HERB MARKER BASIL}       0.008222883 0.9047619  0.009088450 101.36075 152
[7]  {HERB MARKER BASIL}       => {HERB MARKER MINT}        0.008006492 0.8969697  0.008926156  97.53226 148
[8]  {HERB MARKER MINT}        => {HERB MARKER BASIL}       0.008006492 0.8705882  0.009196646  97.53226 148
[9]  {REGENCY TEA PLATE PINK}  => {REGENCY TEA PLATE GREEN} 0.009629429 0.8900000  0.010819583  68.26411 178
[10] {REGENCY TEA PLATE GREEN} => {REGENCY TEA PLATE PINK}  0.009629429 0.7385892  0.013037598  68.26411 178
```

Figure 9

2) min_sup = 0.011 and min_conf = 0.5: There are a total of 21 rules that have lift less than 10 and rules that have lift greater than 10 are 98. Rules that have a lift less than 10 have only one rule with order of 4 whereas, lift > 10 has 11 order 3 rules(here order is lhs+rhs).

```
> inspect(lift.subset2[1:10])
     lhs                                  rhs                                 support    confidence coverage   lift
[1]  {CANDLEHOLDER PINK HANGING HEART}    => {WHITE HANGING HEART T-LIGHT HOLDER} 0.01244252 0.7255521  0.01714904 7.616030
[2]  {JUMBO BAG SCANDINAVIAN BLUE PAISLEY} => {JUMBO BAG RED RETROSPOT}         0.01255072 0.5843829  0.02147687 7.617995
[3]  {JUMBO STORAGE BAG SKULLS}           => {JUMBO BAG RED RETROSPOT}          0.01103598 0.5087282  0.02169326 6.631763
[4]  {JUMBO BAG WOODLAND ANIMALS}         => {JUMBO BAG RED RETROSPOT}          0.01233433 0.5217391  0.02364079 6.801374
[5]  {JUMBO BAG SPACEBOY DESIGN}          => {JUMBO BAG RED RETROSPOT}          0.01255072 0.5536993  0.02266703 7.218005
[6]  {PINK REGENCY TEACUP AND SAUCER}     => {REGENCY CAKESTAND 3 TIER}         0.01493103 0.5454545  0.02737355 6.577122
[7]  {JUMBO BAG PINK VINTAGE PAISLEY}     => {JUMBO BAG RED RETROSPOT}          0.01536381 0.5409524  0.02840141 7.051837
[8]  {JUMBO  BAG BAROQUE BLACK WHITE}     => {JUMBO BAG RED RETROSPOT}          0.01704084 0.5585106  0.03051123 7.280726
[9]  {JUMBO BAG STRAWBERRY}               => {JUMBO BAG RED RETROSPOT}          0.01979984 0.6354167  0.03116040 8.283270
[10] {GREEN REGENCY TEACUP AND SAUCER}    => {REGENCY CAKESTAND 3 TIER}         0.01812280 0.5368590  0.03375710 6.473476
```

Figure 10

```
> inspect(lift.subset21[1:10])
     lhs                                rhs                                support    confidence coverage   lift     count
[1]  {SET 3 RETROSPOT TEA}           => {SUGAR}                           0.01195564 1.0000000  0.01195564 83.64253 221
[2]  {SUGAR}                         => {SET 3 RETROSPOT TEA}             0.01195564 1.0000000  0.01195564 83.64253 221
[3]  {SET 3 RETROSPOT TEA}           => {COFFEE}                          0.01195564 1.0000000  0.01195564 63.08874 221
[4]  {COFFEE}                        => {SET 3 RETROSPOT TEA}             0.01195564 0.7542662  0.01585069 63.08874 221
[5]  {SUGAR}                         => {COFFEE}                          0.01195564 1.0000000  0.01195564 63.08874 221
[6]  {COFFEE}                        => {SUGAR}                           0.01195564 0.7542662  0.01585069 63.08874 221
[7]  {SET/6 RED SPOTTY PAPER CUPS}   => {SET/6 RED SPOTTY PAPER PLATES}   0.01157695 0.8230769  0.01406546 51.57484 214
[8]  {SET/6 RED SPOTTY PAPER PLATES} => {SET/6 RED SPOTTY PAPER CUPS}     0.01157695 0.7254237  0.01595889 51.57484 214
[9]  {SHED}                          => {KEY FOB}                         0.01141466 1.0000000  0.01141466 60.40850 211
[10] {KEY FOB}                       => {SHED}                            0.01141466 0.6895425  0.01655396 60.40850 211
```

Figure 11

3) min_sup = 0.015 and min_conf = 0.5: There are a total of 10 rules that have lift less than 10 and rules that have lift greater than 10 are 21. All the rules that have lift < 10 are of order 2. In contrast, rules that have lift > 10 almost contains 3, order 3 rules.

```
> inspect(lift.subset3[1:10])
     lhs                                  rhs                                   support    confidence coverage   lift
[1]  {JUMBO BAG PINK VINTAGE PAISLEY}   => {JUMBO BAG RED RETROSPOT}            0.01536381 0.5409524  0.02840141 7.051837
[2]  {JUMBO  BAG BAROQUE BLACK WHITE}   => {JUMBO BAG RED RETROSPOT}            0.01704084 0.5585106  0.03051123 7.280726
[3]  {JUMBO BAG STRAWBERRY}             => {JUMBO BAG RED RETROSPOT}            0.01979984 0.6354167  0.03116040 8.283270
[4]  {GREEN REGENCY TEACUP AND SAUCER}  => {REGENCY CAKESTAND 3 TIER}           0.01812280 0.5368590  0.03375710 6.473476
[5]  {RED HANGING HEART T-LIGHT HOLDER} => {WHITE HANGING HEART T-LIGHT HOLDER} 0.02050311 0.6579861  0.03116040 6.906799
[6]  {ROSES REGENCY TEACUP AND SAUCER}  => {REGENCY CAKESTAND 3 TIER}           0.02012443 0.5254237  0.03830133 6.335589
[7]  {JUMBO STORAGE BAG SUKI}           => {JUMBO BAG RED RETROSPOT}            0.02034082 0.5620329  0.03619151 7.326642
[8]  {JUMBO BAG PINK POLKADOT}          => {JUMBO BAG RED RETROSPOT}            0.02591290 0.6141026  0.04219638 8.005420
[9]  {LUNCH BAG WOODLAND}               => {LUNCH BAG RED RETROSPOT}            0.01979984 0.5176803  0.03824723 8.536415
[10] {LUNCH BAG PINK POLKADOT}          => {LUNCH BAG RED RETROSPOT}            0.02477685 0.5585366  0.04436029 9.210124
     count
```

Figure 12

```
> inspect(lift.subset31[1:10])
     lhs                                  rhs                                   support    confidence coverage   lift     count
[1]  {PINK REGENCY TEACUP AND SAUCER}   => {GREEN REGENCY TEACUP AND SAUCER}  0.02239654 0.8181818  0.02737355 24.23733 414
[2]  {GREEN REGENCY TEACUP AND SAUCER}  => {PINK REGENCY TEACUP AND SAUCER}   0.02239654 0.6634615  0.03375710 24.23733 414
[3]  {PINK REGENCY TEACUP AND SAUCER}   => {ROSES REGENCY TEACUP AND SAUCER}  0.02115229 0.7727273  0.02737355 20.17495 391
[4]  {ROSES REGENCY TEACUP AND SAUCER}  => {PINK REGENCY TEACUP AND SAUCER}   0.02115229 0.5522599  0.03830133 20.17495 391
[5]  {ALARM CLOCK BAKELIKE PINK}        => {ALARM CLOCK BAKELIKE GREEN}       0.01558020 0.5680473  0.02742764 15.79001 288
[6]  {ALARM CLOCK BAKELIKE PINK}        => {ALARM CLOCK BAKELIKE RED}         0.01796051 0.6548323  0.02742764 16.22597 332
[7]  {DOLLY GIRL LUNCH BOX}             => {SPACEBOY LUNCH BOX}               0.01769002 0.6770186  0.02612929 21.76468 327
[8]  {SPACEBOY LUNCH BOX}               => {DOLLY GIRL LUNCH BOX}             0.01769002 0.5686957  0.03110630 21.76468 327
[9]  {GREEN REGENCY TEACUP AND SAUCER}  => {ROSES REGENCY TEACUP AND SAUCER}  0.02629159 0.7788462  0.03375710 20.33471 486
[10] {ROSES REGENCY TEACUP AND SAUCER}  => {GREEN REGENCY TEACUP AND SAUCER}  0.02629159 0.6864407  0.03830133 20.33471 486
```

Figure 13

4) min_sup = 0.008 and min_conf = 0.7: There are a total of 4 rules that have lift less than 10 and rules that have lift greater than 10 are 95. Lift < 10 has no order 4 rules whereas, lift > 10 has 7 order 4 rules.

```
> inspect(lift.subset4)
    lhs                                  rhs                                   support     confidence coverage    lift     count
[1] {CANDLEHOLDER PINK HANGING HEART}  => {WHITE HANGING HEART T-LIGHT HOLDER} 0.012442521  0.7255521  0.01714904 7.616030   230
[2] {JUMBO BAG PINK POLKADOT,
     JUMBO BAG PINK VINTAGE PAISLEY}   => {JUMBO BAG RED RETROSPOT}            0.008439275  0.7090909  0.01190154 9.243685   156
[3] {JUMBO  BAG BAROQUE BLACK WHITE,
     JUMBO BAG PINK POLKADOT}          => {JUMBO BAG RED RETROSPOT}            0.008331079  0.7230047  0.01152286 9.425065   154
[4] {JUMBO SHOPPER VINTAGE RED PAISLEY,
     JUMBO STORAGE BAG SUKI}           => {JUMBO BAG RED RETROSPOT}            0.008547471  0.7281106  0.01173925 9.491625   158
```

Figure 14

```
> inspect(lift.subset41[1:10])
     lhs                          rhs                          support     confidence coverage    lift      count
[1]  {HERB MARKER BASIL}       => {HERB MARKER THYME}       0.008006492 0.8969697  0.008926156 101.72077 148
[2]  {HERB MARKER THYME}       => {HERB MARKER BASIL}       0.008006492 0.9079755  0.008817961 101.72077 148
[3]  {HERB MARKER BASIL}       => {HERB MARKER PARSLEY}     0.008060590 0.9030303  0.008926156 103.04022 149
[4]  {HERB MARKER PARSLEY}     => {HERB MARKER BASIL}       0.008060590 0.9197531  0.008763863 103.04022 149
[5]  {HERB MARKER BASIL}       => {HERB MARKER ROSEMARY}    0.008222883 0.9212121  0.008926156 101.36075 152
[6]  {HERB MARKER ROSEMARY}    => {HERB MARKER BASIL}       0.008222883 0.9047619  0.009088450 101.36075 152
[7]  {HERB MARKER BASIL}       => {HERB MARKER MINT}        0.008006492 0.8969697  0.008926156  97.53226 148
[8]  {HERB MARKER MINT}        => {HERB MARKER BASIL}       0.008006492 0.8705882  0.009196646  97.53226 148
[9]  {REGENCY TEA PLATE PINK}  => {REGENCY TEA PLATE GREEN} 0.009629429 0.8900000  0.010819583  68.26411 178
[10] {REGENCY TEA PLATE GREEN} => {REGENCY TEA PLATE PINK}  0.009629429 0.7385892  0.013037598  68.26411 178
```

Figure 15

5) min_sup = 0.011 and min_conf = 0.7: There are a total of 1 rule that have lift less than 10 and rules that have lift greater than 10 are 34. Lift < 10 has only one rule whereas, lift > 10 has 34 rules.

```
> inspect(lift.subset5)
    lhs                                rhs                                 support    confidence coverage   lift    count
[1] {CANDLEHOLDER PINK HANGING HEART} => {WHITE HANGING HEART T-LIGHT HOLDER} 0.01244252 0.7255521  0.01714904 7.61603 230
```

Figure 16

```
> inspect(lift.subset51[1:10])
     lhs                              rhs                              support    confidence coverage   lift     count
[1]  {SET 3 RETROSPOT TEA}         => {SUGAR}                          0.01195564 1.0000000  0.01195564 83.64253 221
[2]  {SUGAR}                       => {SET 3 RETROSPOT TEA}            0.01195564 1.0000000  0.01195564 83.64253 221
[3]  {SET 3 RETROSPOT TEA}         => {COFFEE}                         0.01195564 1.0000000  0.01195564 63.08874 221
[4]  {COFFEE}                      => {SET 3 RETROSPOT TEA}            0.01195564 0.7542662  0.01585069 63.08874 221
[5]  {SUGAR}                       => {COFFEE}                         0.01195564 1.0000000  0.01195564 63.08874 221
[6]  {COFFEE}                      => {SUGAR}                          0.01195564 0.7542662  0.01585069 63.08874 221
[7]  {SET/6 RED SPOTTY PAPER CUPS} => {SET/6 RED SPOTTY PAPER PLATES} 0.01157695 0.8230769  0.01406546 51.57484 214
[8]  {SET/6 RED SPOTTY PAPER PLATES} => {SET/6 RED SPOTTY PAPER CUPS} 0.01157695 0.7254237  0.01595889 51.57484 214
[9]  {SHED}                        => {KEY FOB}                        0.01141466 1.0000000  0.01141466 60.40850 211
[10] {PAINTED METAL PEARS ASSORTED} => {ASSORTED COLOUR BIRD ORNAMENT} 0.01233433 0.7169811  0.01720314 10.68822 228
```

Figure 17

6) min_sup = 0.015 and min_conf = 0.7: There are a total of 0 rules that have lift less than 10 and rules that have lift greater than 10 are 7.

```
> inspect(lift.subset61)
    lhs                                rhs                                  support    confidence coverage   lift     count
[1] {PINK REGENCY TEACUP AND SAUCER}    => {GREEN REGENCY TEACUP AND SAUCER}  0.02239654 0.8181818  0.02737355 24.23733 414
[2] {PINK REGENCY TEACUP AND SAUCER}    => {ROSES REGENCY TEACUP AND SAUCER}  0.02115229 0.7727273  0.02737355 20.17495 391
[3] {GREEN REGENCY TEACUP AND SAUCER}   => {ROSES REGENCY TEACUP AND SAUCER}  0.02629159 0.7788462  0.03375710 20.33471 486
[4] {GARDENERS KNEELING PAD CUP OF TEA} => {GARDENERS KNEELING PAD KEEP CALM} 0.02093589 0.7413793  0.02823911 21.99743 387
[5] {GREEN REGENCY TEACUP AND SAUCER,
     PINK REGENCY TEACUP AND SAUCER}    => {ROSES REGENCY TEACUP AND SAUCER}  0.01882608 0.8405797  0.02239654 21.94649 348
[6] {PINK REGENCY TEACUP AND SAUCER,
     ROSES REGENCY TEACUP AND SAUCER}   => {GREEN REGENCY TEACUP AND SAUCER}  0.01882608 0.8900256  0.02115229 26.36558 348
[7] {GREEN REGENCY TEACUP AND SAUCER,
     ROSES REGENCY TEACUP AND SAUCER}   => {PINK REGENCY TEACUP AND SAUCER}   0.01882608 0.7160494  0.02629159 26.15844 348
```

Figure 18

7) min_sup = 0.008 and min_conf = 0.8: There are a total of 0 rules that have lift less than 10 and rules that have lift greater than 10 are 49.

```
> inspect(lift.subset71[1:10])
    lhs                          rhs                          support     confidence coverage    lift      count
[1]  {HERB MARKER BASIL}      => {HERB MARKER THYME}       0.008006492 0.8969697  0.008926156 101.72077 148
[2]  {HERB MARKER THYME}      => {HERB MARKER BASIL}       0.008006492 0.9079755  0.008817961 101.72077 148
[3]  {HERB MARKER BASIL}      => {HERB MARKER PARSLEY}     0.008060590 0.9030303  0.008926156 103.04022 149
[4]  {HERB MARKER PARSLEY}    => {HERB MARKER BASIL}       0.008060590 0.9197531  0.008763863 103.04022 149
[5]  {HERB MARKER BASIL}      => {HERB MARKER ROSEMARY}    0.008222883 0.9212121  0.008926156 101.36075 152
[6]  {HERB MARKER ROSEMARY}   => {HERB MARKER BASIL}       0.008222883 0.9047619  0.009088450 101.36075 152
[7]  {HERB MARKER BASIL}      => {HERB MARKER MINT}        0.008006492 0.8969697  0.008926156  97.53226 148
[8]  {HERB MARKER MINT}       => {HERB MARKER BASIL}       0.008006492 0.8705882  0.009196646  97.53226 148
[9]  {REGENCY TEA PLATE PINK} => {REGENCY TEA PLATE GREEN} 0.009629429 0.8900000  0.010819583  68.26411 178
[10] {REGENCY TEA PLATE PINK} => {REGENCY TEA PLATE ROSES} 0.009467136 0.8750000  0.010819583  55.39170 175
```

Figure 19

8) min_sup = 0.011 and min_conf = 0.8: There are a total of 0 rules that have lift < 10 and rules that has lift > 10 are 17.

```
> inspect(lift.subset81[1:10])
     lhs                                rhs                                support    confidence coverage   lift     count
[1]  {SET 3 RETROSPOT TEA}           => {SUGAR}                            0.01195564 1.0000000  0.01195564 83.64253 221
[2]  {SUGAR}                         => {SET 3 RETROSPOT TEA}              0.01195564 1.0000000  0.01195564 83.64253 221
[3]  {SET 3 RETROSPOT TEA}           => {COFFEE}                           0.01195564 1.0000000  0.01195564 63.08874 221
[4]  {SUGAR}                         => {COFFEE}                           0.01195564 1.0000000  0.01195564 63.08874 221
[5]  {SET/6 RED SPOTTY PAPER CUPS}   => {SET/6 RED SPOTTY PAPER PLATES}    0.01157695 0.8230769  0.01406546 51.57484 214
[6]  {SHED}                          => {KEY FOB}                          0.01141466 1.0000000  0.01141466 60.40850 211
[7]  {PINK REGENCY TEACUP AND SAUCER} => {GREEN REGENCY TEACUP AND SAUCER} 0.02239654 0.8181818  0.02737355 24.23733 414
[8]  {SET 3 RETROSPOT TEA,SUGAR}     => {COFFEE}                           0.01195564 1.0000000  0.01195564 63.08874 221
[9]  {COFFEE,SET 3 RETROSPOT TEA}    => {SUGAR}                            0.01195564 1.0000000  0.01195564 83.64253 221
[10] {COFFEE,SUGAR}                  => {SET 3 RETROSPOT TEA}              0.01195564 1.0000000  0.01195564 83.64253 221
```

Figure 20

9) min_sup = 0.015 and min_conf = 0.8: There are a total of 0 rules that have lift < 10 and rules that has lift > 10 are 3.

```
> inspect(lift.subset91)
    lhs                                rhs                                support    confidence coverage   lift     count
[1] {PINK REGENCY TEACUP AND SAUCER}  => {GREEN REGENCY TEACUP AND SAUCER} 0.02239654 0.8181818  0.02737355 24.23733 414
[2] {GREEN REGENCY TEACUP AND SAUCER,
     PINK REGENCY TEACUP AND SAUCER}  => {ROSES REGENCY TEACUP AND SAUCER} 0.01882608 0.8405797  0.02239654 21.94649 348
[3] {PINK REGENCY TEACUP AND SAUCER,
     ROSES REGENCY TEACUP AND SAUCER} => {GREEN REGENCY TEACUP AND SAUCER} 0.01882608 0.8900256  0.02115229 26.36558 348
```

Figure 21

Visualize top 100 rules based on min_conf in descending order:

1) min_sup = 0.008 and min_conf = 0.5:

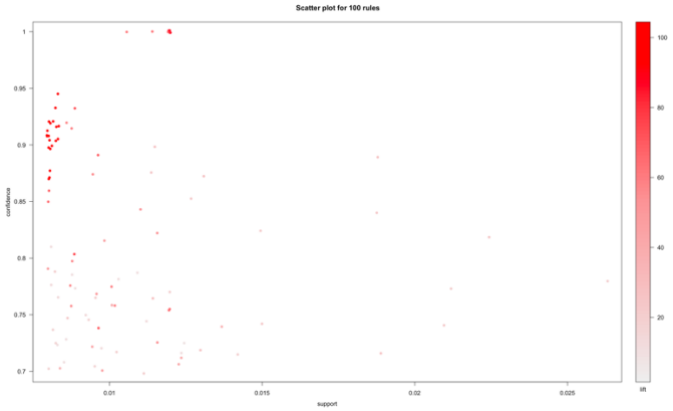


Figure 22

2) min_sup = 0.011 and min_conf = 0.7:

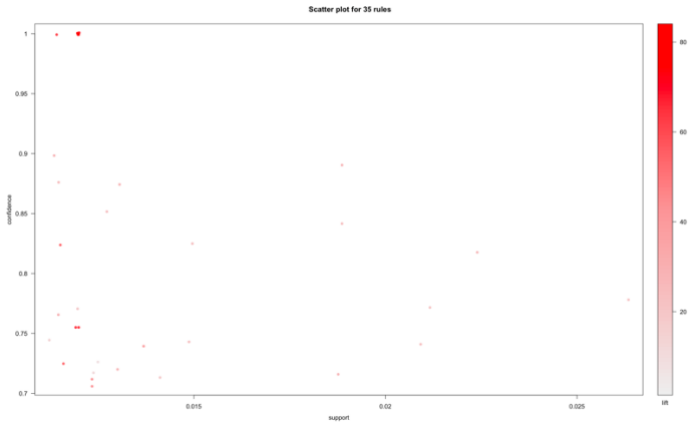


Figure 23

3) min_sup = 0.015 and min_conf = 0.8:

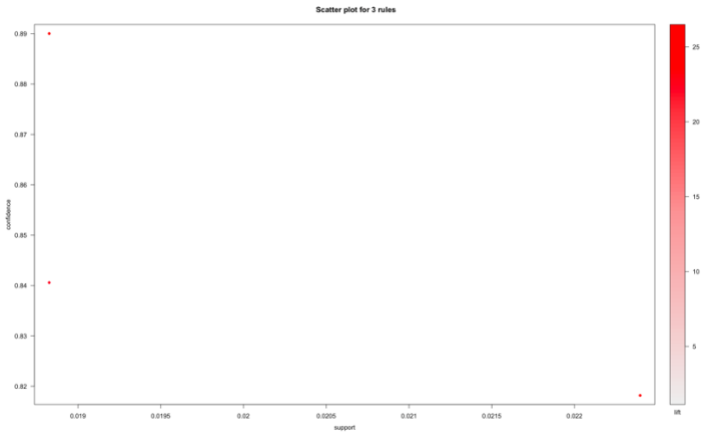


Figure 24

Visualize rules with lift<10:

1) min_sup = 0.008 and min_conf = 0.5:

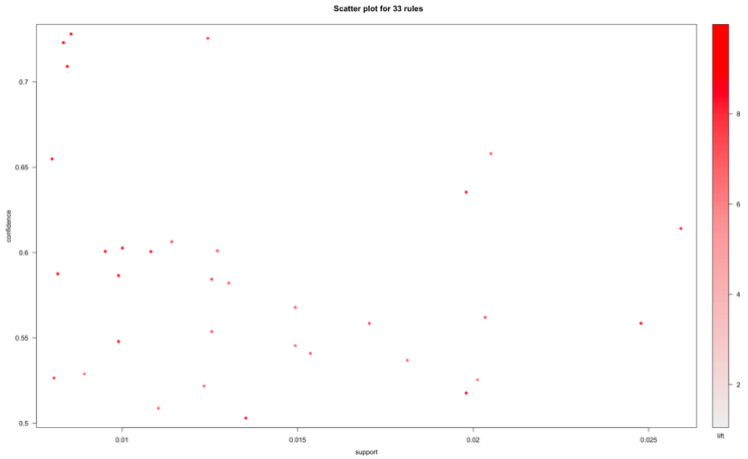


Figure 25

2) min_sup = 0.011 and min_conf = 0.7:

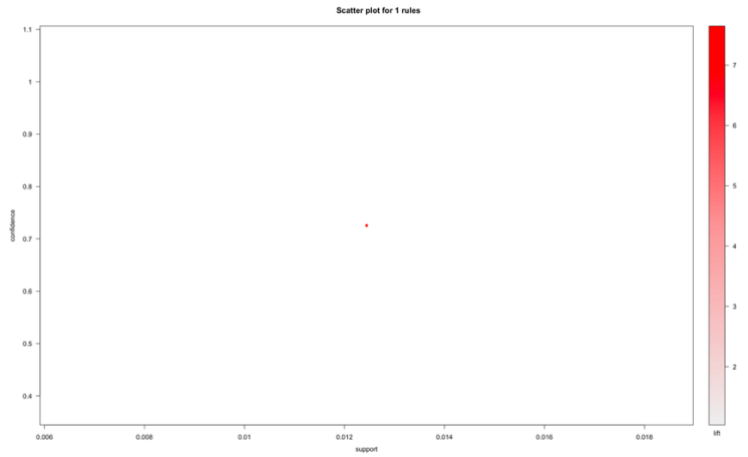


Figure 26

3) min_sup = 0.015 and min_conf = 0.8: There are 0 rules

Visualize rules with lift>10:

1) min_sup = 0.008 and min_conf = 0.5:

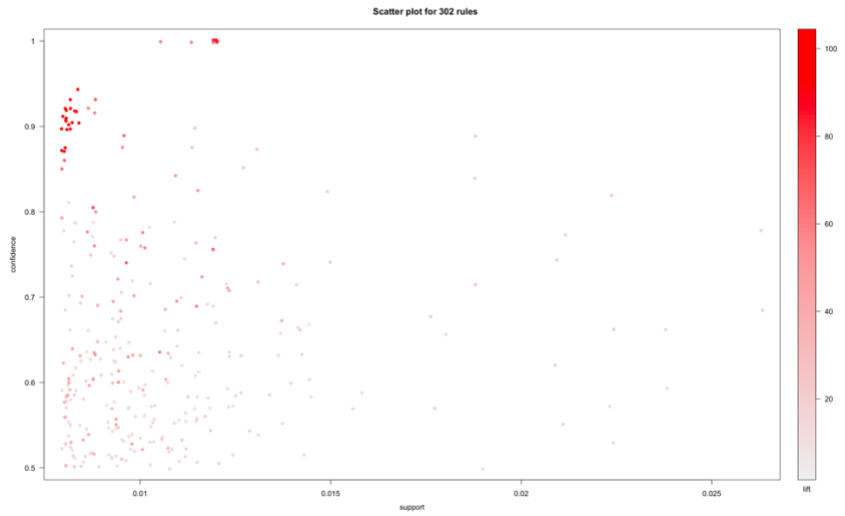


Figure 27

2) min_sup = 0.011 and min_conf = 0.7:

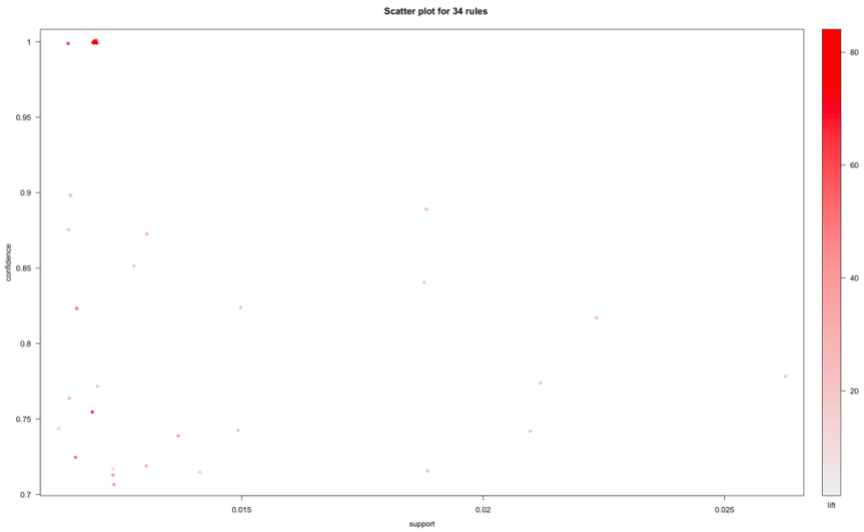


Figure 28

3) min_sup = 0.015 and min_conf = 0.8:

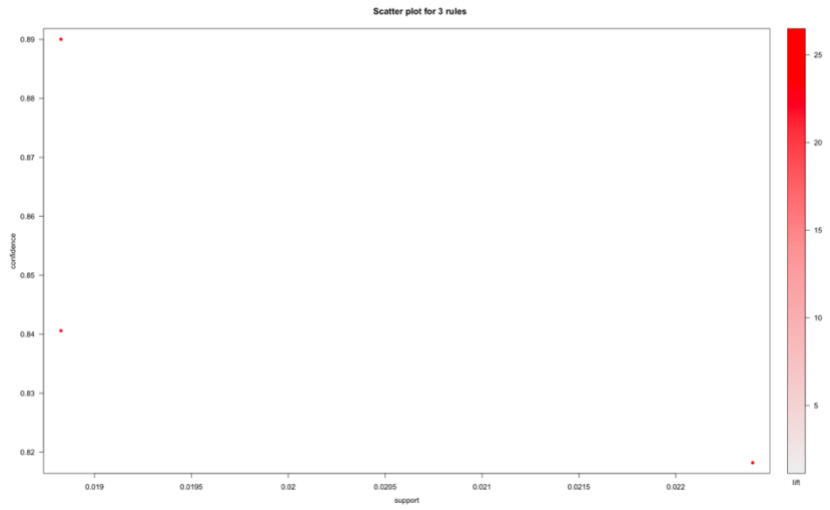


Figure 29

Three problems encountered and solved:

1) The number of candidate item sets generated were around 274 million and the laptop was crashing every time we were running it. So, we encountered this my increasing the memory usage to 10GB.
2) Converting the data frame into transactions was a challenge.
3) Visualizing the rules was a challenge we overcame that with using arulesviz functions.