**Association Rules:**

* Try different values of support and confidence. Observe the change in number of rules for different support, confidence values.
* Change the minimum length in apriori algorithm.
* Visualize the obtained rules using different plots.

Support is nothing but frequency of Items that sold out and Support increases then decreases in the association rules.

Confidence is uniqueness of Items in the dataset and confidence increases then decreases in the association rules.

Minimum length in apriori algorithm is increases then decreases in the association rules.

* Lift is equal to 1 means no association between the items..
* Lift is greater than 1 then means high association between the items..
* Lift is less than 1 then means less association between the items.

**Q1 : Grocery Dataset.**

As I received cleaned dataset and am proceeding with the same dataset.

When I tried to run the model with different support, as usually increases in the support is causes decreases in the number of rules.

|  |  |
| --- | --- |
| **Support** | **Rules** |
| 0.002 | 408 |
| 0.003 | 265 |
| 0.004 | 180 |
| 0.005 | 148 |

When I tried to run the model with different confidence, as usually increases in the confidence is causes decreases in the number of rules.

|  |  |
| --- | --- |
| **Confidence** | **Rules** |
| 0.6 | 476 |
| 0.7 | 408 |
| 0.8 | 353 |
| 0.9 | 311 |

When I tried to run the model with different minimum lengths, as usually increases in the minimum length is causes decreases in the number of rules.

|  |  |
| --- | --- |
| **Min Length** | **Rules** |
| 2 | 408 |
| 3 | 297 |
| 4 | 70 |
| 5 | 0 |

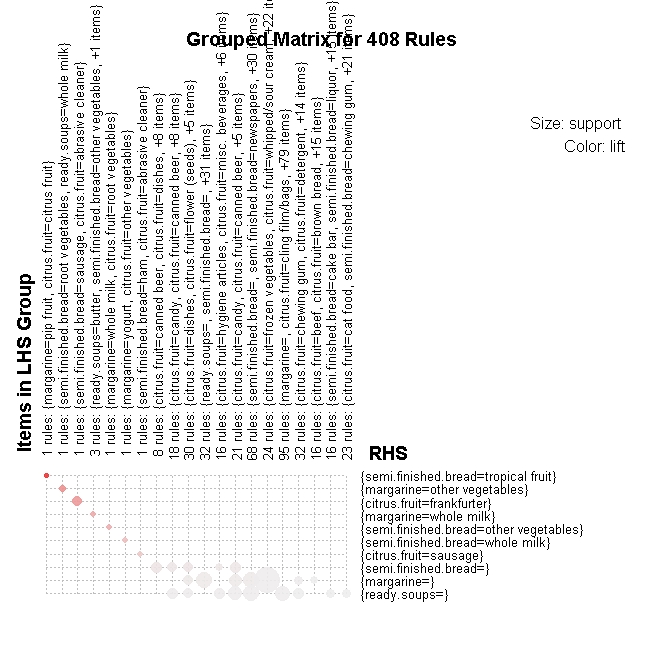
By looking all the data executing the model with 0.002 support , 60% confidence and minimum length 2 is generating 408 rules.

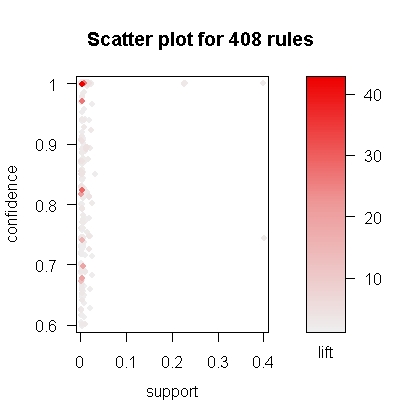
Please find the best association rules

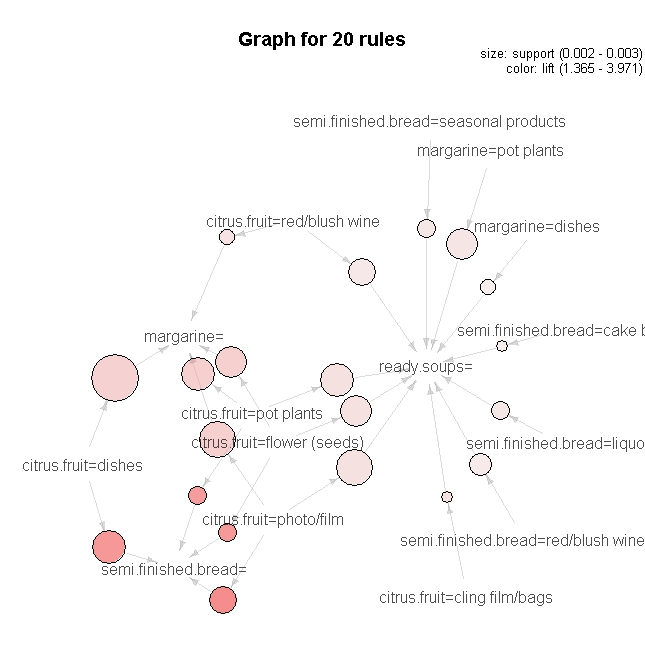
.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Rule No** | **Association ( LHS & RHS)** | **Support** | **Confidence** | **Coverage** | **Lift** | **Count** |
| 230 | {citrus.fruit=citrus fruit,margarine=pip fruit} =>  {semi.finished.bread=tropical fruit} | 0.002026806 | 1 | 0.002027 | 42.60445682 | 31 |
| 292 | {semi.finished.bread=root vegetables,ready.soups=whole milk} =>  {margarine=other vegetables} | 0.003988231 | 0.824324 | 0.004838 | 28.58966109 | 61 |
| 204 | {semi.finished.bread=other vegetables,ready.soups=butter} =>  {margarine=whole milk} | 0.002157568 | 0.970588 | 0.002223 | 26.65196958 | 33 |
| 62 | {semi.finished.bread=sausage} =>  {citrus.fruit=frankfurter} | 0.006472703 | 1 | 0.006473 | 26.37068966 | 99 |
| 290 | {citrus.fruit=root vegetables,margarine=whole milk} =>  {semi.finished.bread=other vegetables} | 0.003203661 | 0.816667 | 0.003923 | 20.14663978 | 49 |
| 66 | {ready.soups=butter} =>  {margarine=whole milk} | 0.004838182 | 0.698113 | 0.00693 | 19.16991294 | 74 |

Please find the Plots.







Please find the generated rules in the attached CSV file.

**Q2 : Books Dataset.**

Books dataset is in the form of dummy format, so converted into Matrix format as part of data cleansing.

When I tried to run the model with different support, as usually increases in the support is causes decreases in the number of rules.

|  |  |
| --- | --- |
| **Support** | **Rules** |
| 0.02 | 301 |
| 0.03 | 150 |
| 0.04 | 101 |
| 0.05 | 61 |

When I tried to run the model with different confidence, as usually increases in the confidence is causes decreases in the number of rules.

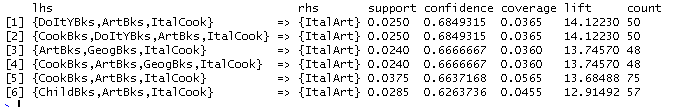
|  |  |
| --- | --- |
| **Confidence** | **Rules** |
| 0.5 | 455 |
| 0.6 | 301 |
| 0.7 | 203 |
| 0.8 | 170 |

When I tried to run the model with different minimum lengths, as usually increases in the minimum length is causes decreases in the number of rules.

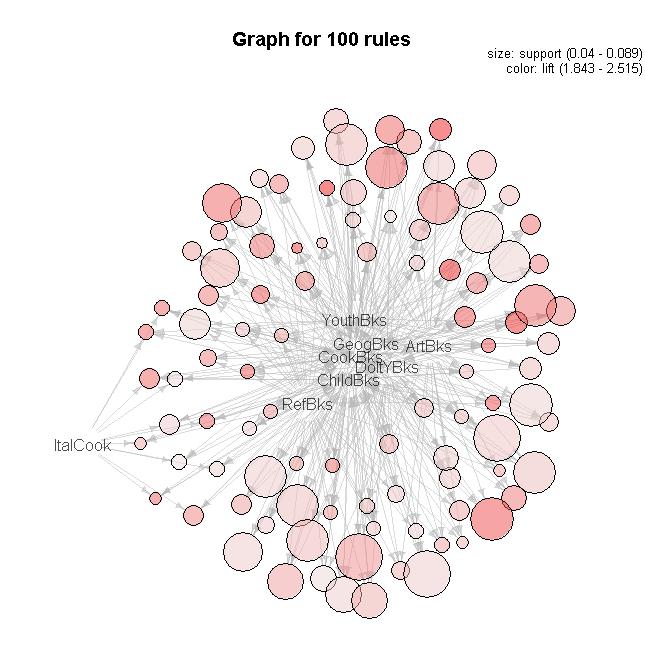
|  |  |
| --- | --- |
| **Min Length** | **Rules** |
| 3 | 419 |
| 4 | 301 |
| 5 | 124 |
| 6 | 19 |

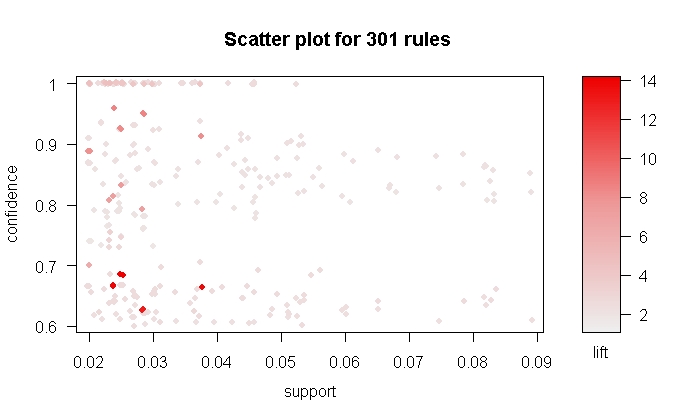
By looking all the data executing the model with 0.02 support , 60% confidence and minimum length 4 is generating 301 rules.

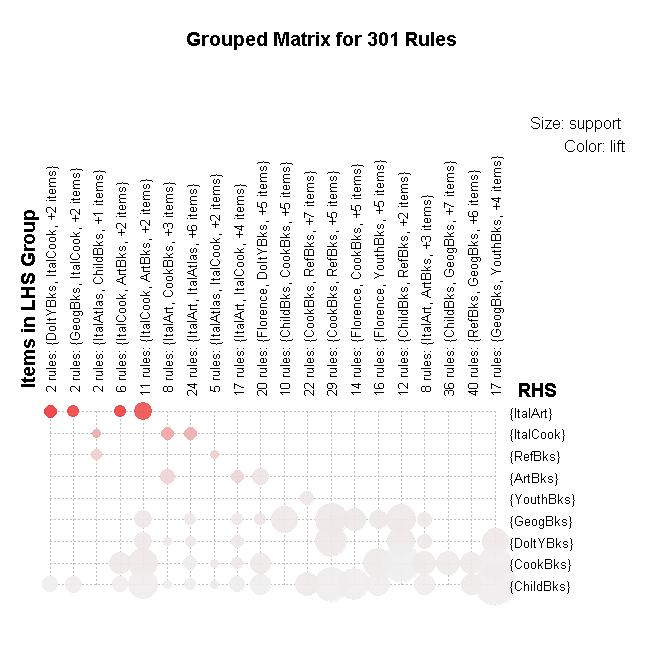
Please find the best association rules.

****

Please find the plots:

****

****

****

Please find the generated rules in the attached CSV file.

**Q3 : Movies Dataset.**

As I received cleaned dataset and am proceeding with the same dataset.

When I tried to run the model with different support, as usually increases in the support is causes decreases in the number of rules.

|  |  |
| --- | --- |
| **Support** | **Rules** |
| 0.02 | 358 |
| 0.1 | 358 |
| 0.2 | 78 |
| 0.3 | 76 |

When I tried to run the model with different confidence, as usually increases in the confidence is causes decreases in the number of rules.

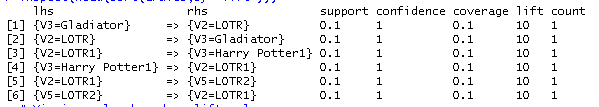
|  |  |
| --- | --- |
| **Confidence** | **Rules** |
| 0.5 | 378 |
| 0.6 | 358 |
| 0.7 | 344 |
| 0.8 | 340 |

When I tried to run the model with different minimum lengths, as usually increases in the minimum length is causes decreases in the number of rules.

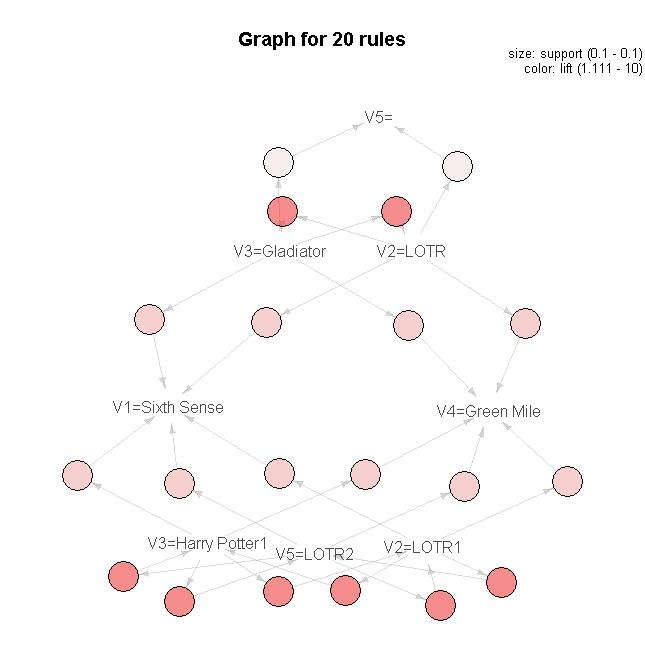
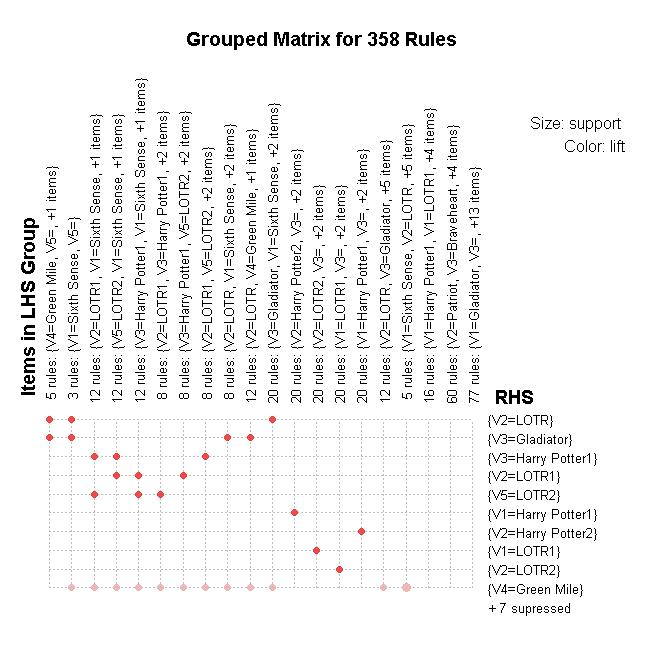
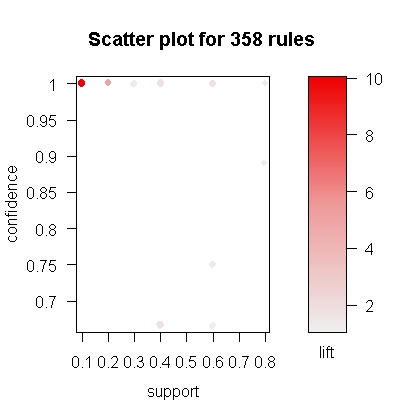
|  |  |
| --- | --- |
| **Min Length** | **Rules** |
| 1 | 362 |
| 2 | 358 |
| 3 | 296 |
| 4 | 151 |

By looking all the data executing the model with 0.02 support , 60% confidence and minimum length 2 is generating 358 rules.

Please find the best association rules.

****

Please find the plots:

****

Please find the generated rules in the attached CSV file.

**Q4: MyphoneData:**

As I received cleaned dataset and am proceeding with the same dataset. When I tried to run the model with different support, as usually increases in the support is causes decreases in the number of rules.

|  |  |
| --- | --- |
| **Support** | **Rules** |
| 0.08 | 34 |
| 0.09 | 34 |
| 0.1 | 14 |
| 0.2 | 04 |

When I tried to run the model with different confidence, as usually increases in the confidence is causes decreases in the number of rules.

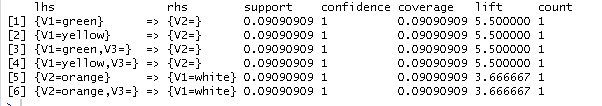
|  |  |
| --- | --- |
| **Confidence** | **Rules** |
| 0.5 | 45 |
| 0.6 | 34 |
| 0.7 | 31 |
| 0.8 | 31 |

When I tried to run the model with different minimum lengths, as usually increases in the minimum length is causes decreases in the number of rules.

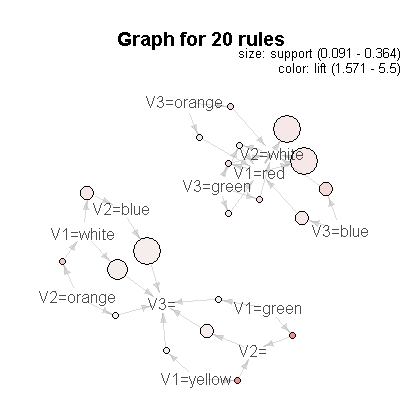
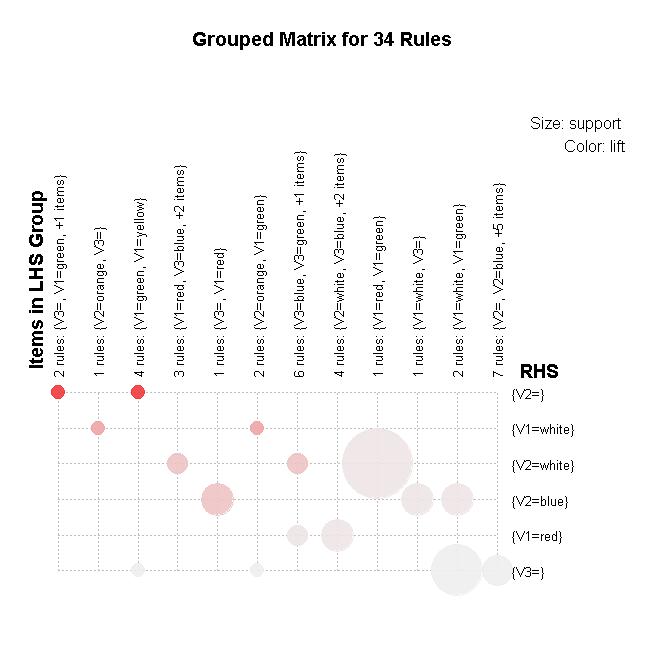
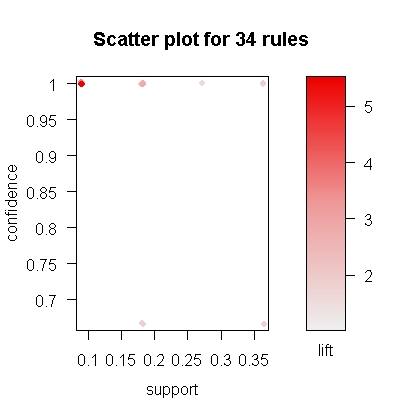
|  |  |
| --- | --- |
| **Min Length** | **Rules** |
| 1 | 35 |
| 2 | 34 |
| 3 | 16 |

By looking all the data executing the model with 0.09 support , 60% confidence and minimum length 2 is generating 34 rules.

Please find the best association rules.



Please find the plots:

****

Please find the generated rules in the attached CSV file.

**Q5: Transaction retail Data:**

**As I haven’t received cleaned dataset** and **Replaced NA values with Spaces**. If I remove the NA then it will delete the lot of data. That’s why I have replaced the NA with spaces. Now dataset is cleaned. When I tried to run the model with different support, as usually increases in the support is causes decreases in the number of rules.

|  |  |
| --- | --- |
| **Support** | **Rules** |
| 0.002 | 2891 |
| 0.003 | 828 |
| 0.004 | 445 |
| 0.005 | 295 |

When I tried to run the model with different confidence, as usually increases in the confidence is causes decreases in the number of rules.

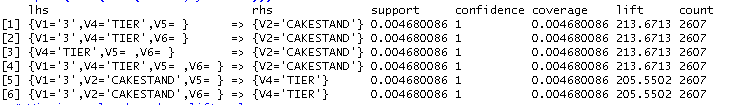
|  |  |
| --- | --- |
| **Confidence** | **Rules** |
| 0.6 | 455 |
| 0.7 | 421 |
| 0.8 | 398 |
| 0.9 | 377 |

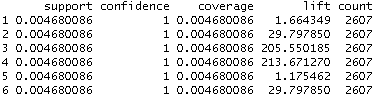
When I tried to run the model with different minimum lengths, as usually increases in the minimum length is causes decreases in the number of rules.

|  |  |
| --- | --- |
| **Min Length** | **Rules** |
| 2 | 1144 |
| 3 | 869 |
| 4 | 421 |
| 5 | 124 |

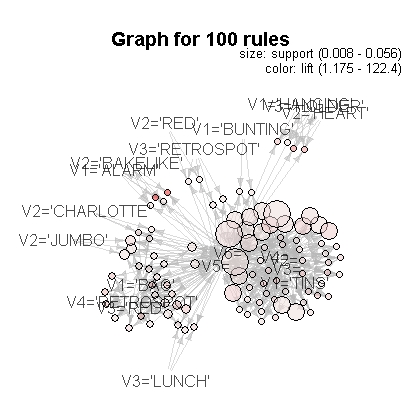
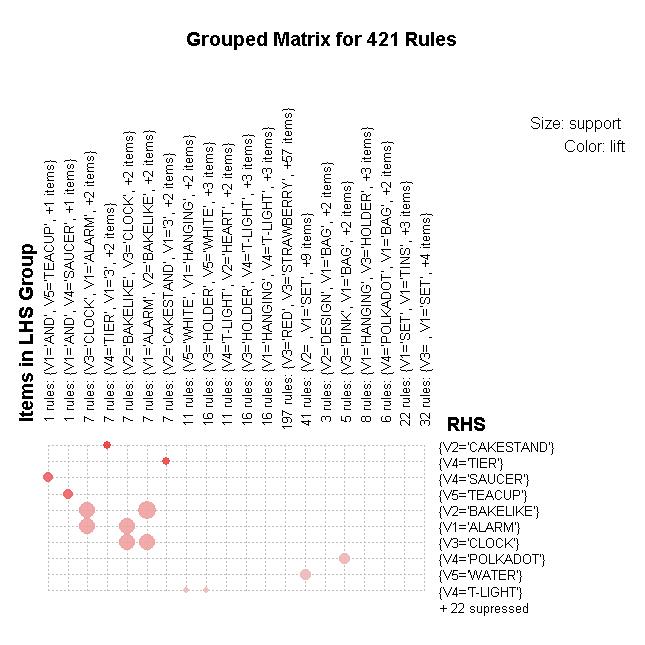
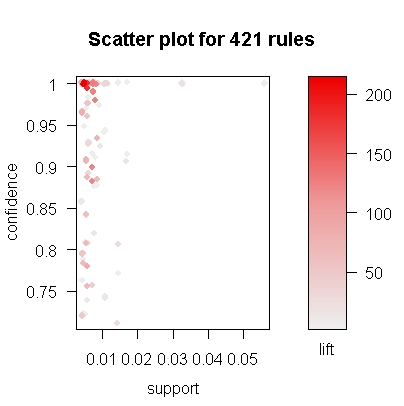
By looking all the data executing the model with 0.004 support, 70% confidence and minimum length 4 is generating 421 rules.

Please find the best association rules.





Please find the plots:

****

Please find the generated rules in the attached CSV file.