Overview

The main purpose of our project is to analyse the association rules produced by applying the apriori algorithm on the data collected from retail data of store purchases. This is done using the following steps:

* Firstly, we went through the provided problem statement and the requirements to complete this project.
* Removed all the rows with unknown values.
* Removed all the transactions which are cancelled.
* Removed the records which are not actual transactions.
* Converted description of items to integers for easy computation.
* Converted the given data into transactions and wrote it to a CSV file.
* Read each rows in the CSV file generated as transactions for further process.
* We used apriori function to generate candidate item sets by lowering the support by a lot and and frequent item sets by giving appropriate support value.
* We used the same apriori function to generate association rule as well by changing the “target” to “rules”.
* We installed arulesViz package and used “plot” to visualize the item sets and rules generated.
* We used “subset” and “sort” to get the rules which have a lift greater than 10 and less than 10.
* Then used “head” to filter 10 rules each for lift > 10, Lift < 10.
* We used apriori function again to generate all the rules and sorted them by “confidence”.
* Visualized all 100 rules generated by using apriori.

File Description

**apriori.r**: This file contains the code which applies apriori algorithm and generates frequent item sets and association rules. It also contains code to visualize the generated outputs.

**market\_basket\_transactions.csv:** This file contains the processed data set which are ready to be read as transactions.

Division of Work

Sai krishna : Wrote code to visualize the item sets, output with different lift values and all the rules. Also coded to generate rules with different support and confidence combination and extracted rules with lift>10, lift<10.

Deepika : Done the pre-processing of data set provided. Coded to generate different item sets and visualize the rules generated.

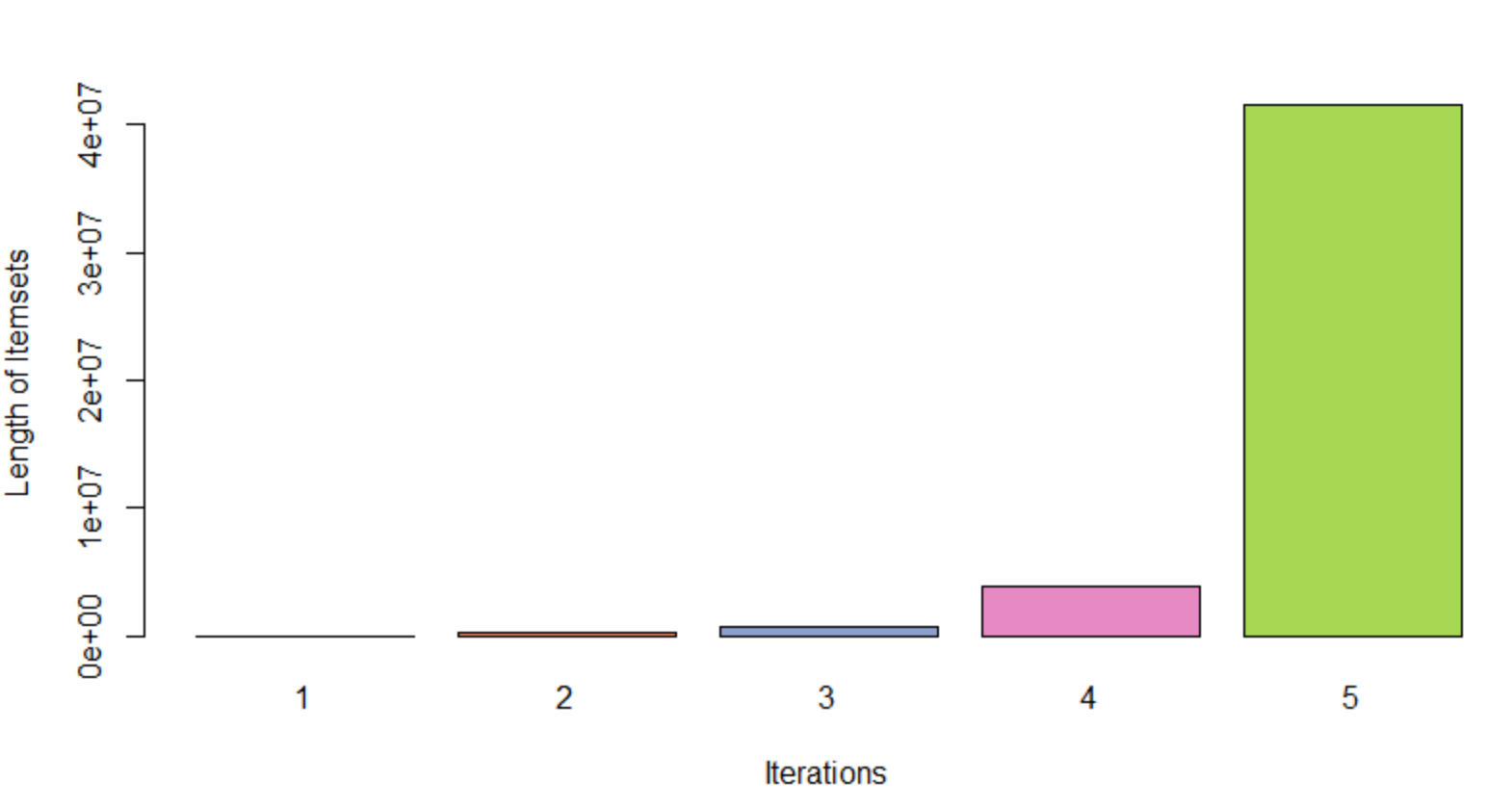
Problem Faced

* We faced problems while pre-processing the data and it was a time-consuming job.
* We faced problem while implementing “subset” function to get rules with lift>10.
* Also faced memory issues while running the apriori algorithm and increased min\_sup value to overcome it.

Analysis

**Candidate itemsets for each iteration:**

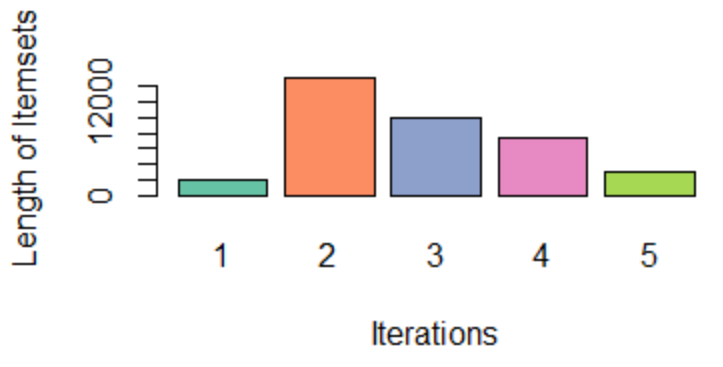
We have tested for all the min\_sup values from 0.0001 and faced memory issues upto 0.0004. Hence we took the min\_sup=0.0005. For all the candidate itemsets generation we took the same min\_sup but different conf values i.e, 0.005,0.007 & 0.008. However, the plots are same for all the three.



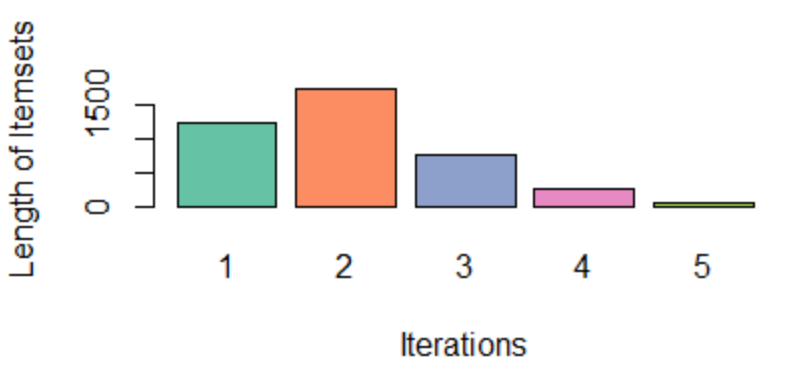
|  |
| --- |
| From the above data we can see that, as the number of iterations increases the number of candidate itemsets generated  are also increasing. |
| |  | | --- | |  | |

**Frequent itemsets for each iteration**

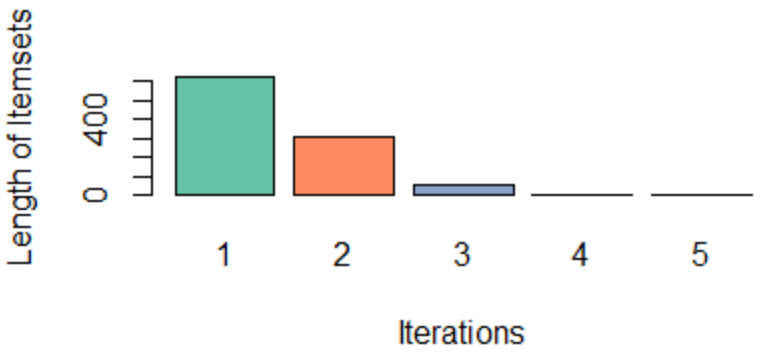
**i) min\_sup=0.002, min\_conf=0.5**



**ii) min\_sup=0.005, min\_conf=0.7**



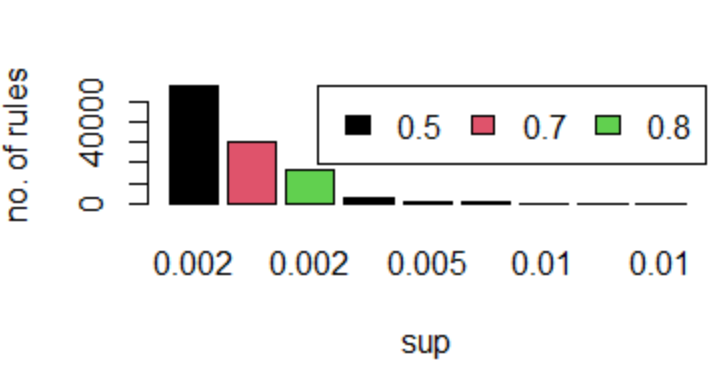
**iii) min\_sup=0.01, min\_conf=0.8**



From above we can see that if the min\_sup is less then number of frequent itemsets generated are increased more in 2nd iteration and then decreased gradually in next iterations. If we increase the min\_sup value then number of itemsets generated are highest in first iteration and they are gradually decreased in next iterations. Also, as the min\_sup value is increasing the number of frequent items generated is decreasing.

**Number of rules with different min support and confidence combinations:**

Support on xaxis vs no of rules on yaxis with each confidence differently colored



From the above graph we can say that as min\_sup and conf increase the number of rules generated decrease as the number of frequent itemsets generated are decreased(seen in frequent itemset plots).

**Top 10 Rules with lift>10:**

**[1] ",FRENCH BLUE METAL DOOR SIGN 0,FRENCH BLUE METAL DOOR SIGN 7,FRENCH BLUE METAL DOOR SIGN 5->FRENCH BLUE METAL DOOR SIGN 9"**

**[2] ",LANDMARK FRAME COVENT GARDEN->LANDMARK FRAME OXFORD STREET"**

**[3] ",LANDMARK FRAME OXFORD STREET->LANDMARK FRAME COVENT GARDEN"**

**[4] ",FRENCH BLUE METAL DOOR SIGN 0,FRENCH BLUE METAL DOOR SIGN 7->FRENCH BLUE METAL DOOR SIGN 9"**

**[5] ",FRENCH BLUE METAL DOOR SIGN 0,FRENCH BLUE METAL DOOR SIGN 2,FRENCH BLUE METAL DOOR SIGN 7->FRENCH BLUE METAL DOOR SIGN 9"**

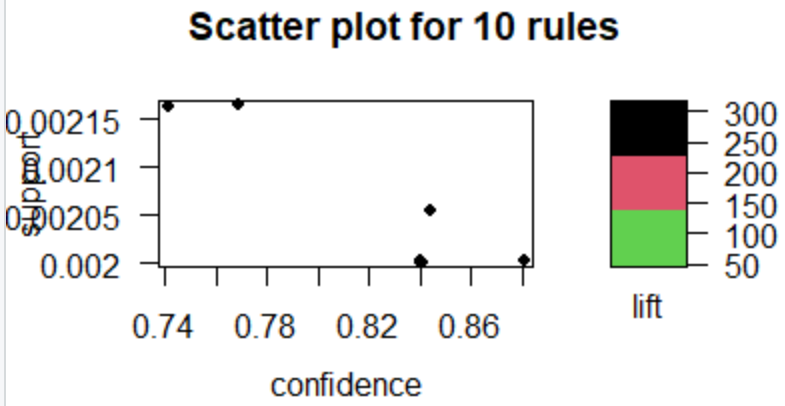
**[6] ",FRENCH BLUE METAL DOOR SIGN 1,FRENCH BLUE METAL DOOR SIGN 0,FRENCH BLUE METAL DOOR SIGN 7->FRENCH BLUE METAL DOOR SIGN 9"**

**[7] ",FRENCH BLUE METAL DOOR SIGN 1,FRENCH BLUE METAL DOOR SIGN 0,FRENCH BLUE METAL DOOR SIGN 2,FRENCH BLUE METAL DOOR SIGN 7->FRENCH BLUE METAL DOOR SIGN 9"**

**[8] ",FRENCH BLUE METAL DOOR SIGN 4,FRENCH BLUE METAL DOOR SIGN 7,FRENCH BLUE METAL DOOR SIGN 5,FRENCH BLUE METAL DOOR SIGN 3->FRENCH BLUE METAL DOOR SIGN 9"**

**[9] ",FRENCH BLUE METAL DOOR SIGN 1,FRENCH BLUE METAL DOOR SIGN 7,FRENCH BLUE METAL DOOR SIGN 5,FRENCH BLUE METAL DOOR SIGN 3->FRENCH BLUE METAL DOOR SIGN 9"**

**[10] ",FRENCH BLUE METAL DOOR SIGN 1,FRENCH BLUE METAL DOOR SIGN 4,FRENCH BLUE METAL DOOR SIGN 7,FRENCH BLUE METAL DOOR SIGN 3->FRENCH BLUE METAL DOOR SIGN 9"**



**Top 10 Rules with lift<10:**

**[1] ",WOODLAND CHARLOTTE BAG,LUNCH BAG SUKI DESIGN->LUNCH BAG BLACK SKULL."**

**[2] ",RED RETROSPOT CHARLOTTE BAG,LUNCH BAG CARS BLUE->LUNCH BAG SPACEBOY DESIGN"**

**[3] ",WOODLAND CHARLOTTE BAG,LUNCH BAG BLACK SKULL.->LUNCH BAG PINK POLKADOT"**

**[4] ",JUMBO BAG APPLES,LUNCH BAG APPLE DESIGN,LUNCH BAG RED RETROSPOT->LUNCH BAG PINK POLKADOT"**

**[5] ",RED RETROSPOT CHARLOTTE BAG,JUMBO BAG STRAWBERRY,JUMBO STORAGE BAG SUKI->JUMBO BAG RED RETROSPOT"**

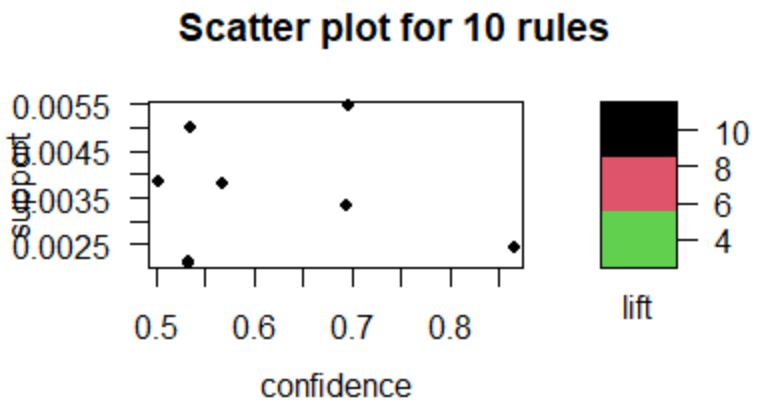
**[6] ",LUNCH BAG SUKI DESIGN,LUNCH BAG APPLE DESIGN,JUMBO STORAGE BAG SUKI->LUNCH BAG RED RETROSPOT"**

**[7] ",LUNCH BAG DOLLY GIRL DESIGN,LUNCH BAG CARS BLUE,LUNCH BAG BLACK SKULL.->LUNCH BAG RED RETROSPOT"**

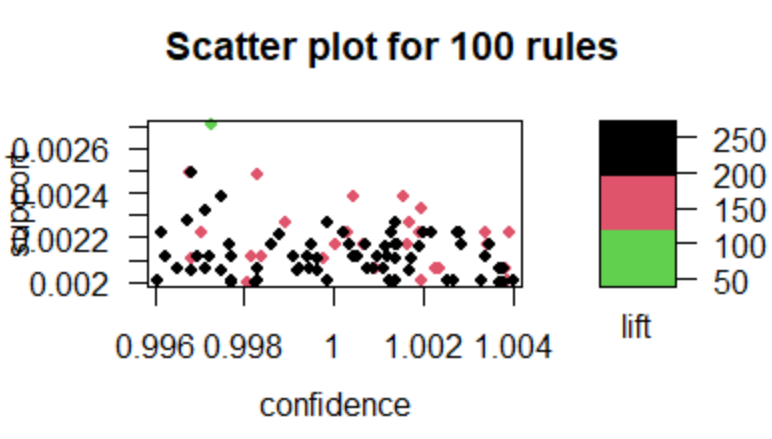
**[8] ",SMALL DOLLY MIX DESIGN ORANGE BOWL,LUNCH BAG RED RETROSPOT->LUNCH BAG SPACEBOY DESIGN"**

**[9] ",LUNCH BAG PINK POLKADOT,CHILDRENS APRON SPACEBOY DESIGN->LUNCH BAG SPACEBOY DESIGN"**

**[10] ",LUNCH BAG BLACK SKULL.,JUMBO BAG VINTAGE DOILY,LUNCH BAG RED RETROSPOT->LUNCH BAG SPACEBOY DESIGN"**



**Vizualizing top 100 rules as per decreasing order of confidence**



support vs confidence with coloring different range of lifts