To clear all previous work from the R environment

rm(list=ls(all=T))

Importing the required libraries

library(arules)

library(arulesViz)

Importing the data in the format of a 'transactions' class – as an item matrix

item\_matrix<-read.transactions('MBAdata3.csv',format="basket",sep = ',',rm .duplicates = T,header = TRUE)

Checking the class of the imported data

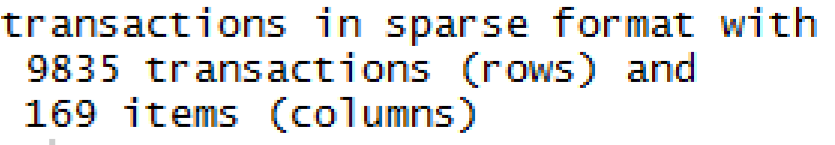
class(item\_matrix)

The data was imported in a ‘transactions’ format (as an item matrix)

**Question 1: Show summary stats, and frequency count of items in the dataset.**

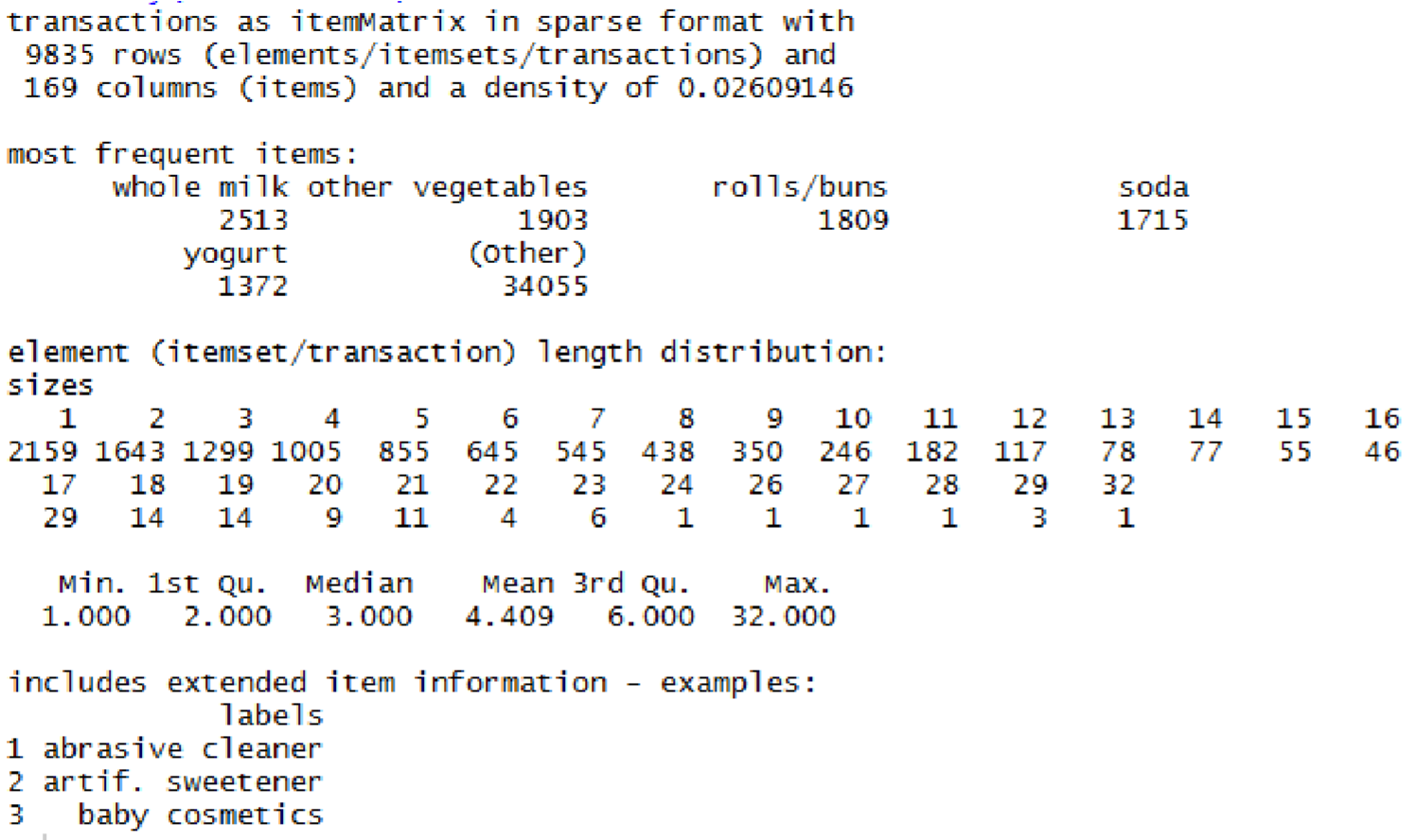
Checking the number of transactions and items in the data

item\_matrix



There are 9835 different transactions and 169 different items in the data set

**Summary statistics**



* **Density:** It is the percentage of non-empty cells in the item matrix which is given by,

Total number of items purchased / Total number of possible items in the item matrix.

2.6% of the cells in the item matrix are non – empty.

The total number of items that are purchased can be calculated as:

9835 \* 169 \*0.0260 = 43214.99

The total number of items purchased in 9835 transactions are 43,215 (includes duplica te items)

* **Most frequent items:** The top 5 items that have been purchased the most are whole milk, other vegetables, rolls/buns, soda and yogurt (in descending order).The numbers below each item in the summary show the total number of times that a particular item was purchased (total number of times an item has appeared in the data set).

* **Element length distribution:** Out of all the transactions, 2159 transactions have only 1 item, 1643 transactions have 2 items and so on till the largest transaction where 1 transaction has 32 items (row 1218 in the data set).This indicates that most customers purchase a small number of items in each transaction.

* **Descriptive statistics:** The minimum number of items purchased by a customer in a single transaction is 1, the maximum number of items purchased by a customer in a single transaction is 32. On an average, a customer purchases 3 items in a single transaction.

**Item frequency**

Finding the frequency of each item in the data set

freq<-itemFrequency(item\_matrix,type='absolute',weighted=FALSE)

Converting 'freq' into a data frame

freq\_df<-as.data.frame(freq)

Assigning column names to the data frame

freq\_df <- cbind(Item = rownames(freq\_df), freq\_df) rownames(freq\_df) <- 1:nrow(freq\_df)

Sorting the data frame in descending order of the item frequency

freq\_df<-freq\_df[order(-freq\_df$freq),]

View(freq\_df)

Displaying the first 10 items from the above created data frame

View(head(freq\_df,10))



* Whole milk is the most bought product
* Sound storage medium and baby food are the least bought products (both have been purchased only once)

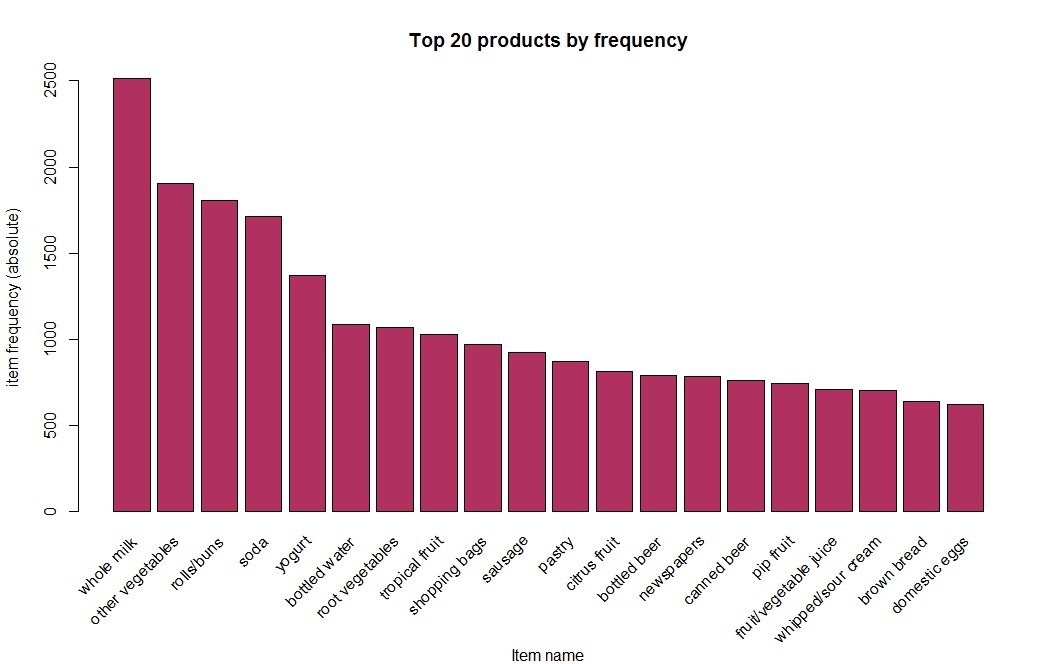
**Question 2: Visualise in at least 2 different charts item frequency, other characteristics of the dataset you think is important.**

Top 20 most frequent (most bought) products from the data set

itemFrequencyPlot(item\_matrix,topN=20,type="absolute",col="maroon",main="T op 20 products by frequency")

|  |
| --- |
| Clearly whole milk is the most bought product.    Shopping bags are also among the top 20 (in fact top 10) products which indicates that most customers do not carry/ miss out on carrying their own shopping bags. |

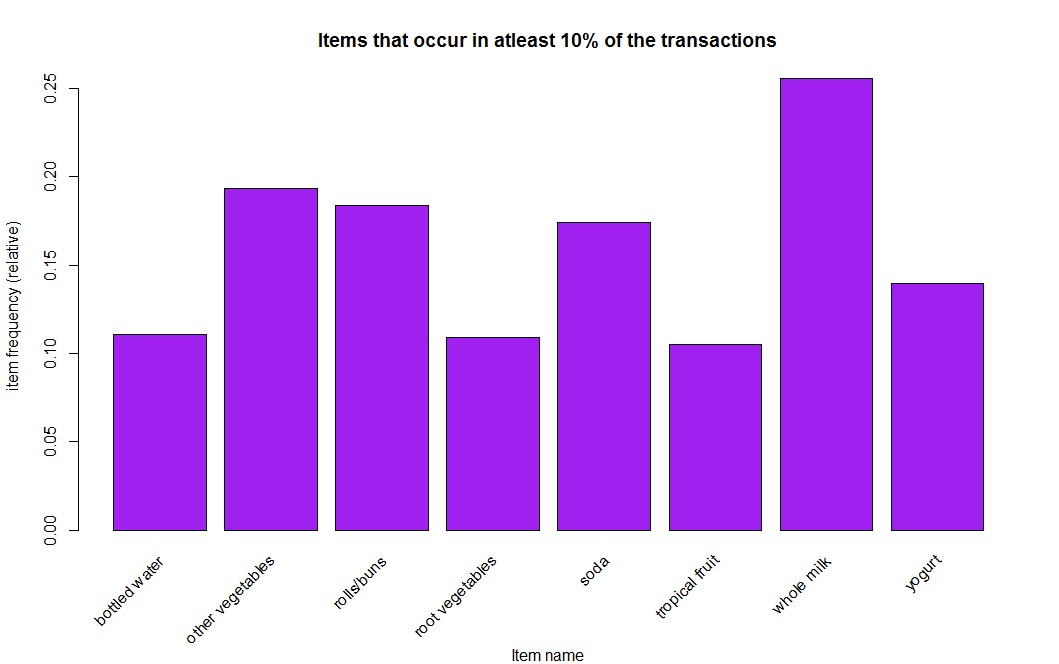
Item frequency plot by support



|  |
| --- |
| The 8 items in the plot occur in at least 10% (983 or more) of the total number of transactions. |

itemFrequencyPlot(item\_matrix, support = 0.1,col="purple",xlab="Item name"

,main="Items that occur in atleast 10% of the transactions")



**Question 3: How many total number of rules were found in the dataset (use Support= 0.0001**

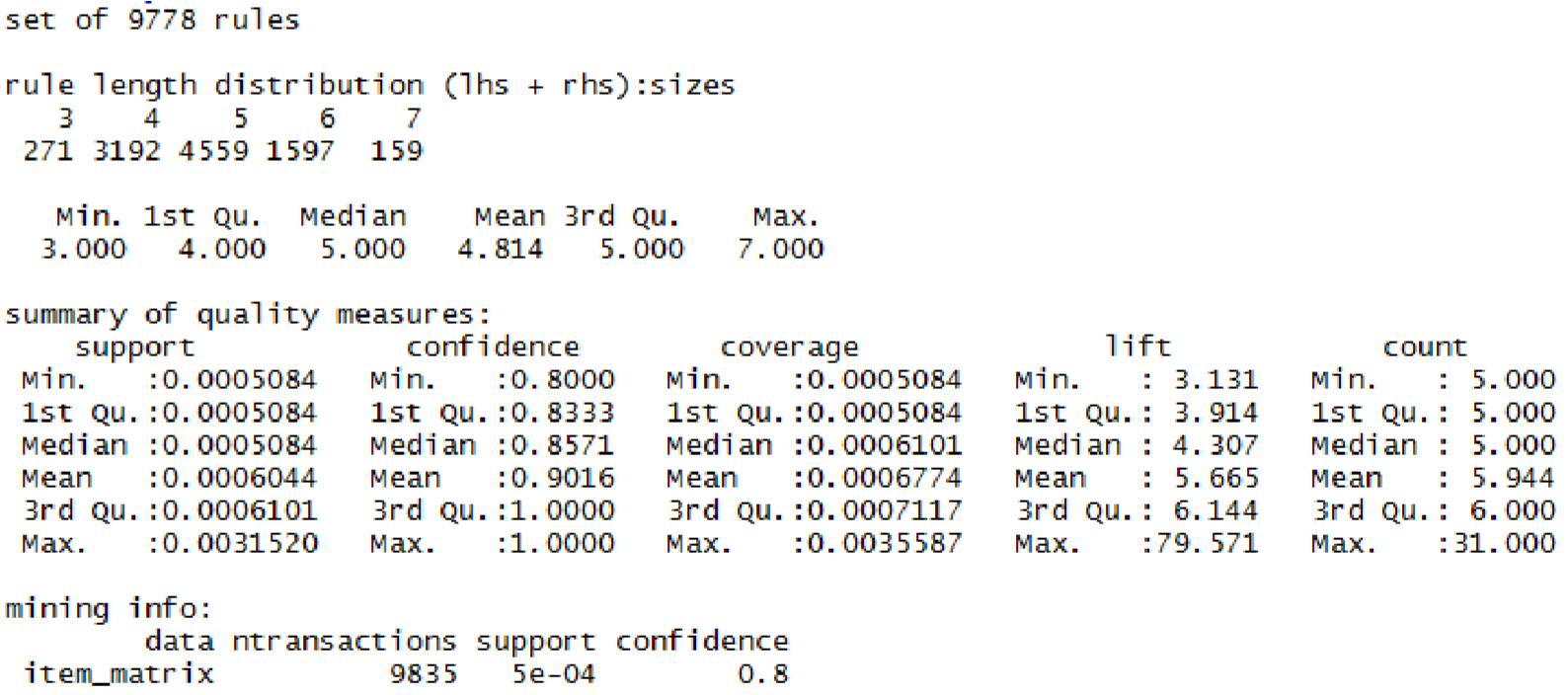
**& confidence = 0.8 as cut-off)**

Forming the association rules using the apriori algorithm

mba<- apriori(item\_matrix, parameter = list(supp = 0.0005, conf = 0.8))

Sorting the generated rules by confidence asso\_rules<-sort(mba,by='confidence',decreasing = TRUE)

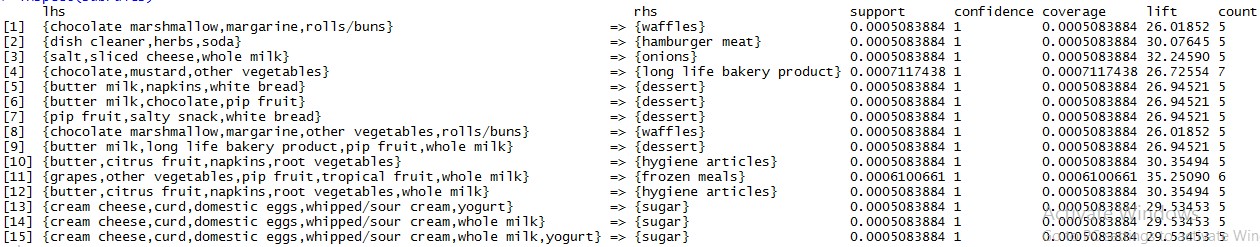
Model summary summary(asso\_rules)



With a support of 0.0005 and confidence of 0.8 a total of 9778 rules were found.

There are 271 rules with 3 items (antecedent + consequent), 3192 rules with 4 items, 4559 Rules with 5 items, 1597 rules with 6 items and 159 rules with 7 items.

**Question 4: What is the combination of lift and confidence values to get a subset of 15**



**rulesets?**

Subsetting the rules

subrules <- subset(asso\_rules,lift>25.3 & confidence>0.86) subrules



inspect(subrules)

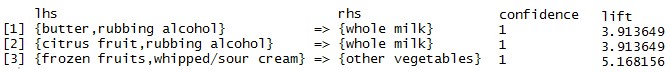
* From the summary of the apriori model, the maximum confidence value is 1 and maximum lift value is 79.57
* The combination of lift and confidence values that returns a subset of 15 rule sets is Lift > 25.3 and confidence > 0.86

**Question 5: For the top 3 popular Consequents what are the Antecedents & what are the lift & confidence values?**

Extracting the top 3 Consequents and respective antecedents.

inspect(asso\_rules[1:3])

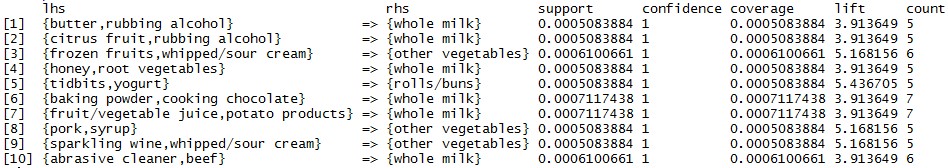
The below output shows the top three consequents(rhs) with their respective antecedents(lhs), lift and confidence values



**Question 6: What are the top 10 rules? Visualise them if possible. Also write in ruleset format & in explain what these rules mean?**

Sorting the generated rules by confidence asso\_rules<-sort(mba,by='confidence',decreasing = TRUE)

inspect(asso\_rules[1:10])

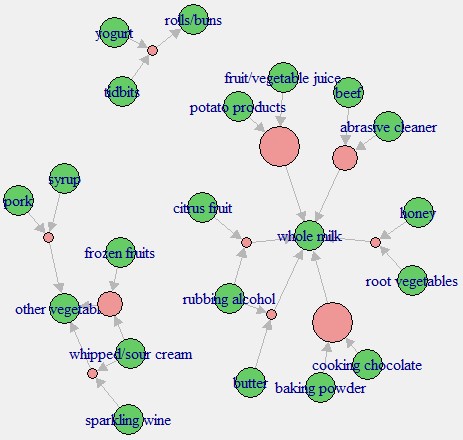
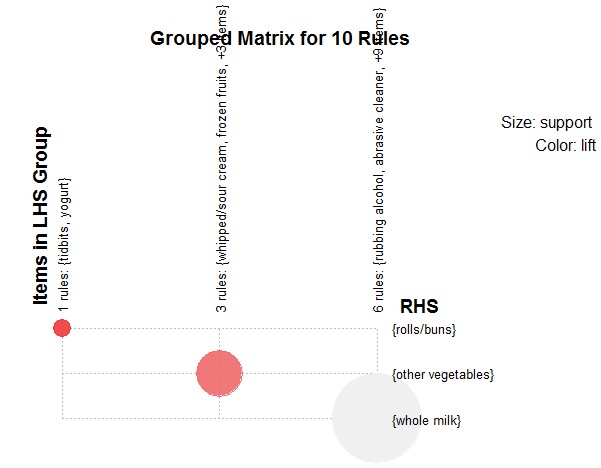


Extracting the top 10 rules

Visualizing the top 10 rules by confidence using the ‘grouped’ and ‘graph’ method

plot(asso\_rules[1:10], method = "grouped")

plot(asso\_rules[1:10],method="graph",engine="interactive",shading = NA)



|  |  |  |
| --- | --- | --- |
| The above grouped matrix graph shows the names of the consequents (horizontal) and all  their respective antecedents (vertical) for the top  10 rules. It also shows that there is 1 rule with roll/buns, 3 rules with other vegetables and 6 rules with whole milk as the consequents. |  | The above graph shows the names of the 3 consequents that have appeared in the top 10 rules and all their respective consequents with the help of arrows.  For eg: Yogurt and tidbits together are the antecedents for rolls/buns. |

Explaining the top 10 rules

**Rule 1**: {butter,rubbing alcohol} => {whole milk}

Support: In 0.05% of the total transactions butter, rubbing alcohol and whole milk are present together

Coverage(support of the antecedent): In 0.05% of the total transactions butter and rubbing alcohol are present together

Confidence: 100% of the customers who bought butter and rubbing alcohol together also bought whole milk

**Rule 2:** {citrus fruit,rubbing alcohol} => {whole milk}

Support: In 0.05% of the total transactions citrus fruit, rubbing alcohol and whole milk are present together

Coverage: In 0.05% of the total transactions citrus fruit and rubbing alcohol are present together

Confidence: 100% of the customers who bought citrus fruit and rubbing alcohol together also bought whole milk

**Rule 3:** {frozen fruits,whipped/sour cream} => {other vegetables}

Support: In 0.06% of the total transactions frozen fruits, whipped/sour cream and other vegetables are present together

Coverage: In 0.06% of the total transactions frozen fruits,whipped/sour cream are present together

Confidence: 100% of the customers who bought frozen fruits and whipped/sour cream together also bought other vegetables

**Rule 4:** {honey,root vegetables} => {whole milk}

Support: In 0.05% of the total transactions honey, root vegetables and whole milk are present together

Coverage: In 0.05% of the total transactions honey and root vegetables are present together

Confidence: 100% of the customers who bought honey and root vegetables together also bought whole milk

**Rule 5:** {tidbits,yogurt} => {rolls/buns}

Support: In 0.05% of the total transactions tidbits, yogurt and rolls/buns are present together

Coverage: In 0.05% of the total transactions tidbits and yogurt are present together

Confidence: 100% of the customers who bought tidbits and yogurt together also bought rolls/buns

**Rule 6:** {baking powder,cooking chocolate} => {whole milk}

Support: In 0.07% of the total transactions baking powder, cooking chocolate and whole milk are present together

Coverage: In 0.07% of the total transactions baking powder and cooking chocolate are present together

Confidence: 100% of the customers who bought baking powder and cooking chocolate together also bought whole milk

**Rule 7:** {fruit/vegetable juice,potato products} => {whole milk}

Support: In 0.07% of the total transactions fruit/vegetable juice, potato products and whole milk are present together

Coverage: In 0.07% of the total transactions fruit/vegetable juice and potato products are present together

Confidence: 100% of the customers who bought fruit/vegetable juice and potato products together also bought whole milk

**Rule 8:** {pork,syrup} => {other vegetables}

Support: In 0.05% of the total transactions pork, syrup and other vegetables are present together

Coverage: In 0.05% of the total transactions pork and syrup are present together

Confidence: 100% of the customers who bought pork and syrup together also bought other vegetables

**Rule 9:** {sparkling wine,whipped/sour cream} => {other vegetables}

Support: In 0.05% of the total transactions sparkling wine,whipped/sour cream and other vegetables are present together

Coverage: In 0.05% of the total transactions sparkling wine and whipped/sour cream are present together

Confidence: 100% of the customers who bought sparkling wine and whipped/sour cream together also bought other vegetables

**Rule 10:** {abrasive cleaner,beef} => {whole milk}

Support: In 0.06% of the total transactions abrasive cleaner, beef and whole milk are present together

Coverage: In 0.06% of the total transactions abrasive cleaner and beef are present together

Confidence: 100% of the customers who bought abrasive cleaner and beef together also bought whole milk

Lift: Since the lift for all 10 rules is > 1, it can be said that the rules have not been formed by chance and the antecedent and consequent are dependent on each other. The degree of dependence depends on how greater the lift value is than 1.

**Question 7: What are your recommendations to the retailer? Write some retail strategies as a result of your analysis.**

* Any product whose sales the retailer wants to increase (say domestic eggs which are at the last position among the top 20 most bought products) can be placed beside the whole milk or other vegetables rack since they are the two most bought products and most customers would visit the sections where they are placed. Hence any other product that is placed next to these racks would have a good chance of getting noticed and possibly getting purchased by most customers.

* Baby products don’t seem to sell too well, baby food has been purchased once and baby cosmetics six times. This could mean that most people don’t prefer buying baby products (especially food and cosmetics) from a regular grocery store and prefer buying it from a pharmacy or an exclusive baby products store instead. Or it could also mean that these products are not placed appropriately and can be placed next to a top selling/ most bought product so their visibility can be increased.

* Whole milk is one product that is quite popular among the customers, it is the most frequent product and also appears among the top 10 rules multiple times. Hence the retailer has to make sure that enough number of units of whole milk are kept in stock always. Whole milk is also one product that people will prefer to buy fresh. So a milk packet/carton that is even one day old may not sell. Hence the right number of units of whole milk need to be kept in stock on a daily basis such that the retailer neither falls short of them nor does he have too many of them.

* From rule 6, it seems like the customers are buying these products for baking purposes. The retailer can create a ‘baker’s combo’ by offering any other product that could be used for baking (say butter for example) along with these products(baking powder, cooking chocolate and whole milk) at a lesser price (bundle pricing). So anybody who bought these 3 items would be encouraged to buy butter as well.