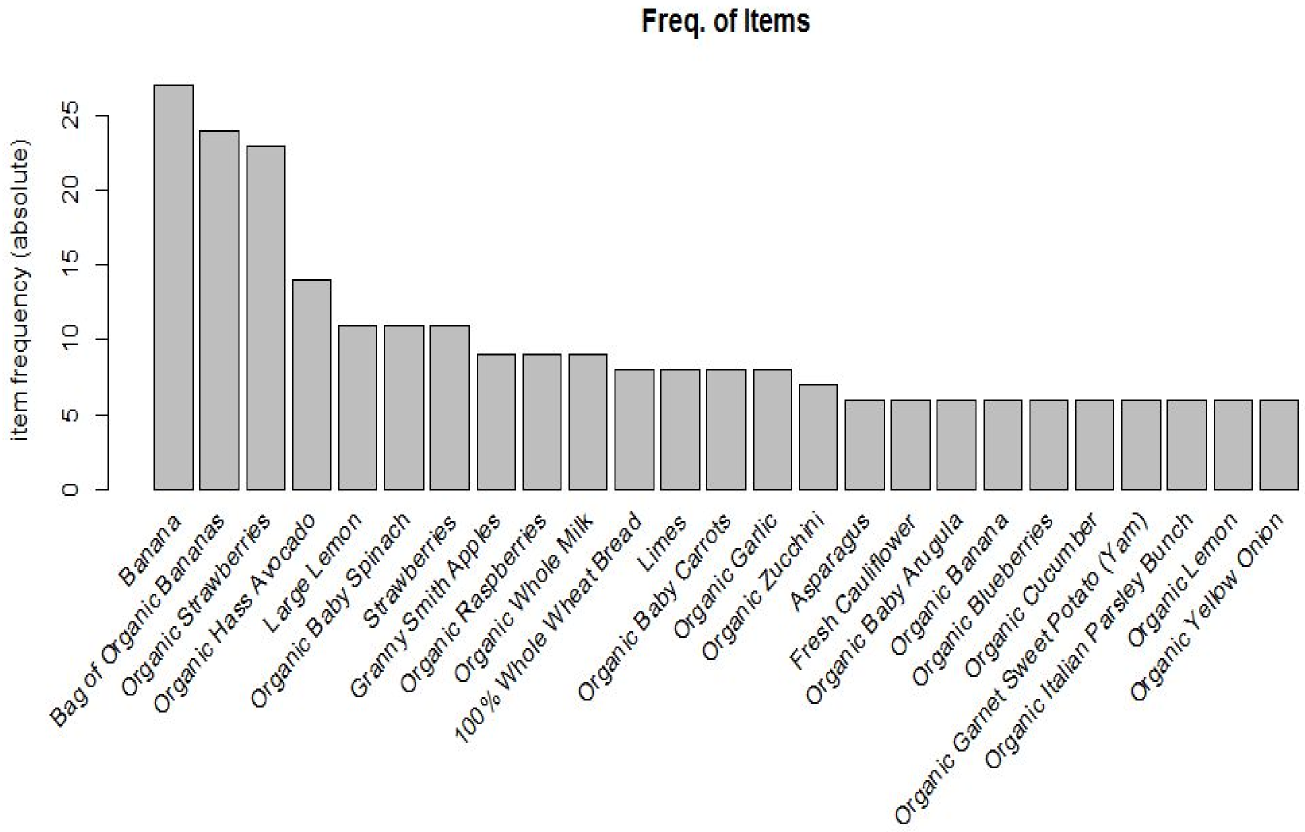
INSTACART MARKET BASKET ANALYSIS

Sathyaraj Natesan - SXN180006
Praveen Ramani - PXR170005
Pratik kamath - PXK170010
Nikhil Peshwe - NXP180001

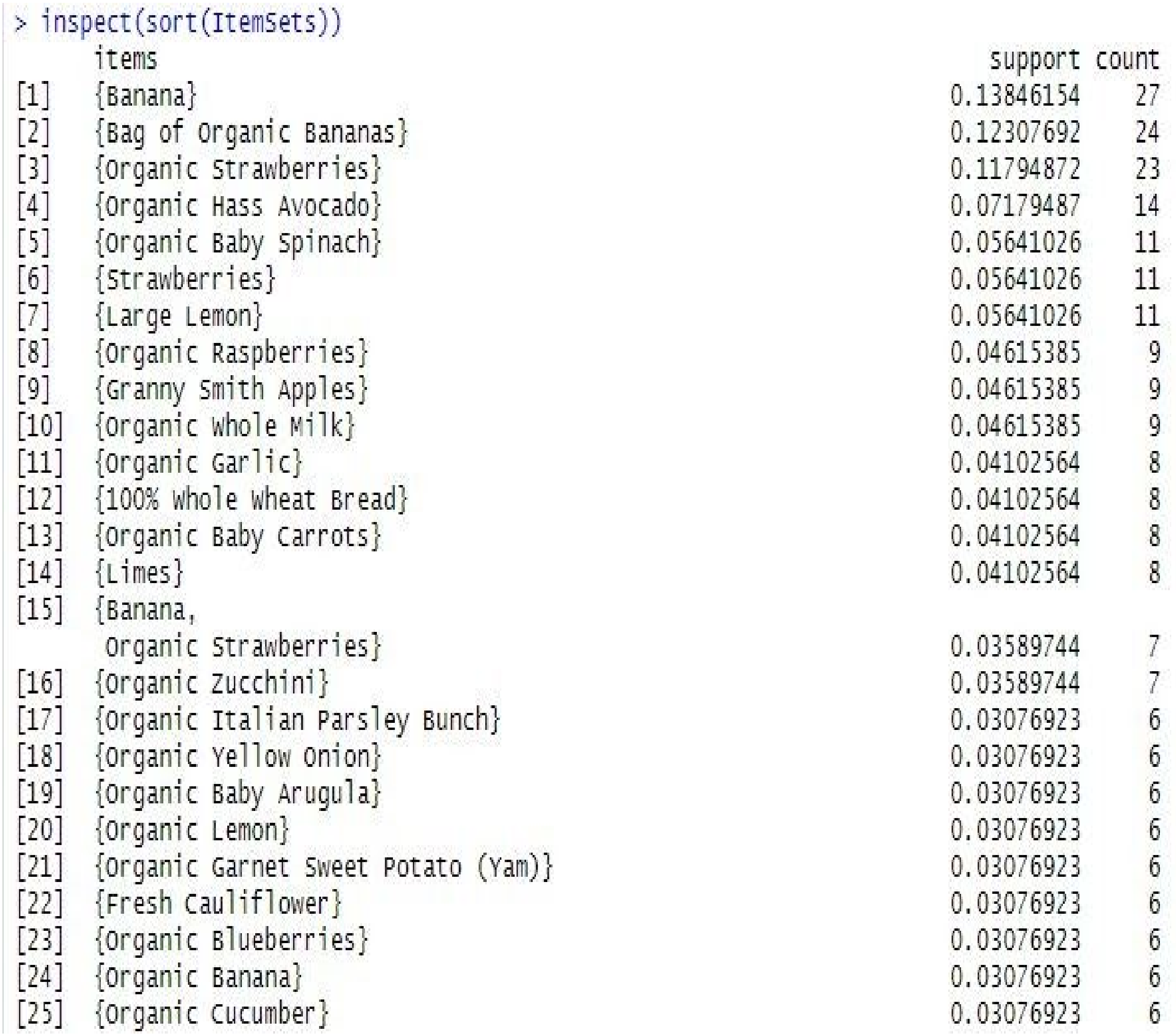
1. Frequent itemsets for products in orders dataset

The below graph depicts the frequency of top 25 items in descending order.



In the data file “output_products_train.csv”, we are fetching 2000 rows of data and then, we merge the two datafiles “order_products_train.csv” and “products.csv” on “product_id” to get product name from the merged data.

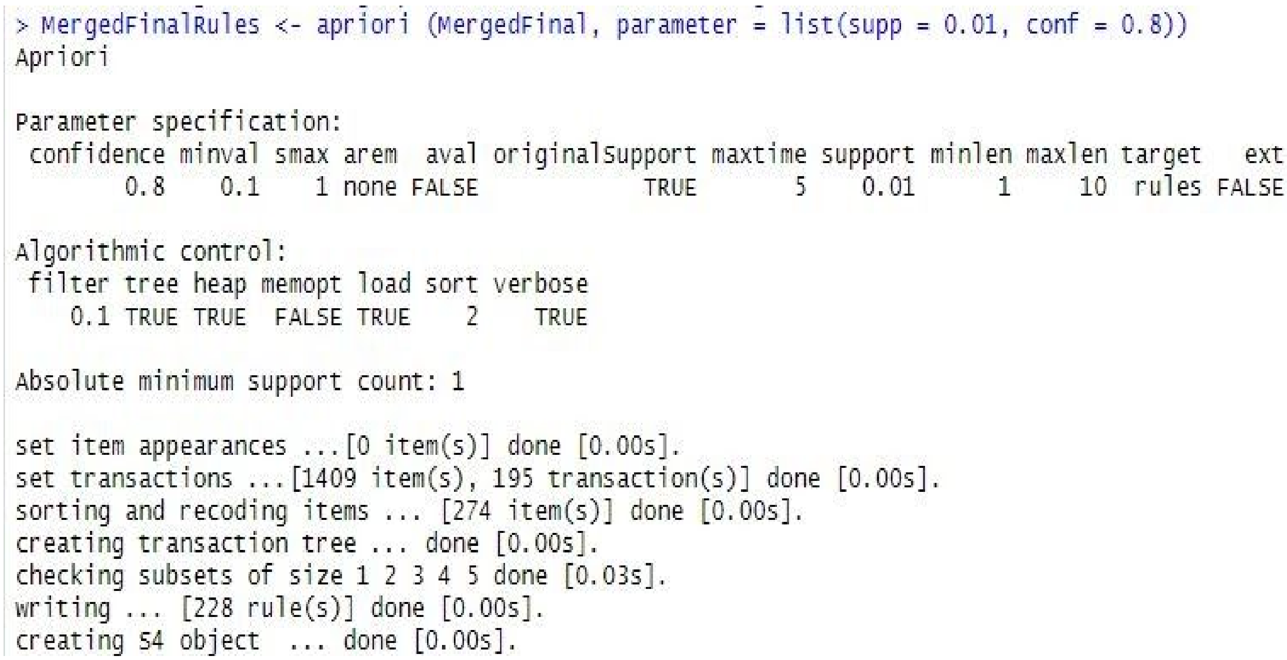
We mine the frequent itemsets using “eclat” with support of 0.01 and max length of 8.
The below screenshot depicts part of the output of **inspect(sort(ItemSets))**



```
> inspect(sort(ItemSets))
     items                                   support count
[1]  {Banana}                             0.13846154    27
[2]  {Bag of Organic Bananas}             0.12307692    24
[3]  {Organic Strawberries}               0.11794872    23
[4]  {Organic Hass Avocado}               0.07179487    14
[5]  {Organic Baby Spinach}               0.05641026    11
[6]  {Strawberries}                       0.05641026    11
[7]  {Large Lemon}                        0.05641026    11
[8]  {Organic Raspberries}                0.04615385     9
[9]  {Granny Smith Apples}                0.04615385     9
[10] {Organic Whole Milk}                 0.04615385     9
[11] {Organic Garlic}                     0.04102564     8
[12] {100% Whole Wheat Bread}             0.04102564     8
[13] {Organic Baby Carrots}               0.04102564     8
[14] {Limes}                              0.04102564     8
[15] {Banana,
      Organic Strawberries}               0.03589744     7
[16] {Organic Zucchini}                   0.03589744     7
[17] {Organic Italian Parsley Bunch}      0.03076923     6
[18] {Organic Yellow Onion}               0.03076923     6
[19] {Organic Baby Arugula}               0.03076923     6
[20] {Organic Lemon}                      0.03076923     6
[21] {Organic Garnet Sweet Potato (Yam)}  0.03076923     6
[22] {Fresh Cauliflower}                  0.03076923     6
[23] {Organic Blueberries}                0.03076923     6
[24] {Organic Banana}                     0.03076923     6
[25] {Organic Cucumber}                   0.03076923     6
```

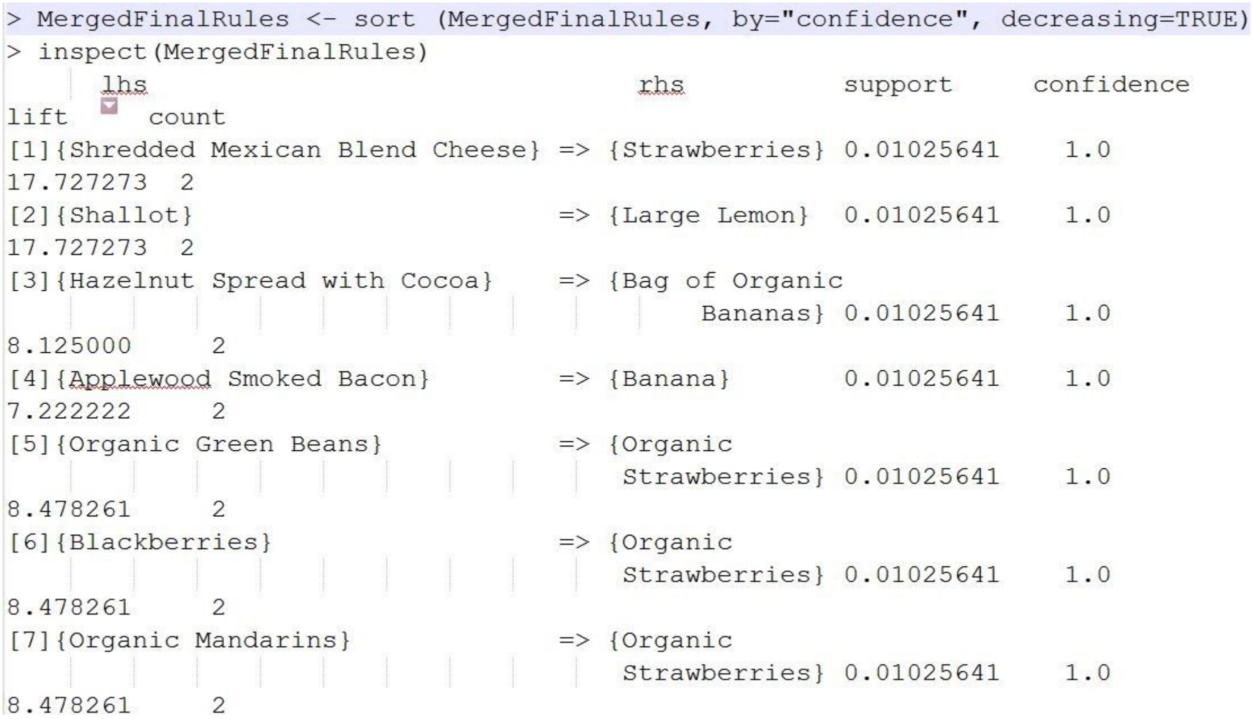
2.Association Rules for products in orders dataset

The below screenshot depicts, association rules generation using “apriori” function.



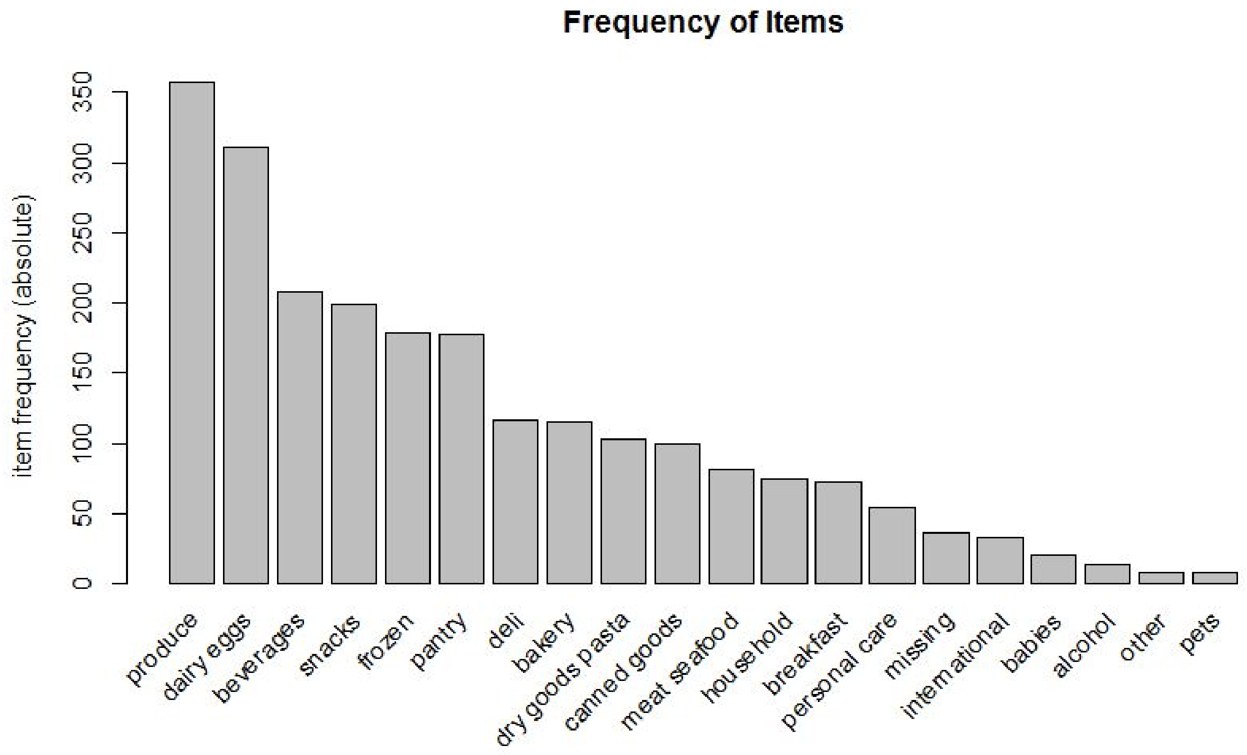
```
> MergedFinalRules <- apriori (MergedFinal, parameter = list(supp = 0.01, conf = 0.8))
Apriori

Parameter specification:
 confidence minval smax arem  aval originalSupport maxtime support minlen maxlen target   ext
        0.8    0.1    1 none FALSE            TRUE       5    0.01      1     10  rules FALSE

Algorithmic control:
 filter tree heap memopt load sort verbose
    0.1 TRUE TRUE  FALSE TRUE    2    TRUE

Absolute minimum support count: 1

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[1409 item(s), 195 transaction(s)] done [0.00s].
sorting and recoding items ... [274 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3 4 5 done [0.03s].
writing ... [228 rule(s)] done [0.00s].
creating S4 object  ... done [0.00s].
```

The generated association rules data have been sorted based on confidence values.Refer the details below for more information.



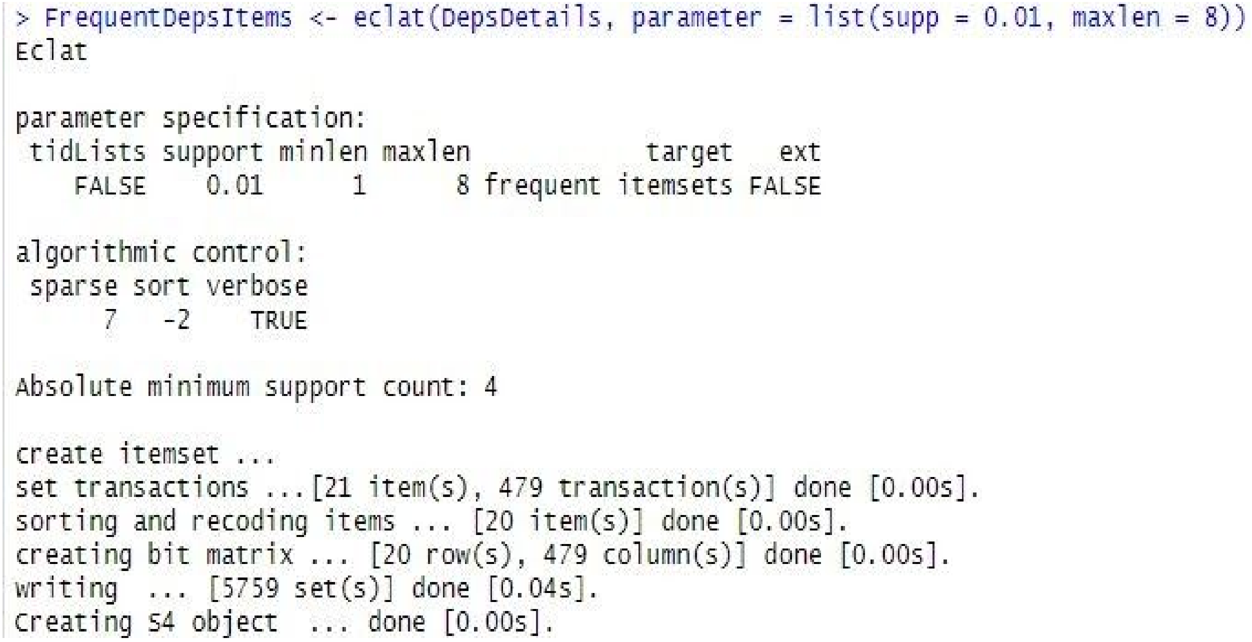
```
> MergedFinalRules <- sort (MergedFinalRules, by="confidence", decreasing=TRUE)
> inspect(MergedFinalRules)
      lhs                                 rhs          support     confidence
lift      count
[1]{Shredded Mexican Blend Cheese} => {Strawberries} 0.01025641    1.0
17.727273  2
[2]{Shallot}                       => {Large Lemon}  0.01025641    1.0
17.727273  2
[3]{Hazelnut Spread with Cocoa}    => {Bag of Organic
                                             Bananas} 0.01025641    1.0
8.125000     2
[4]{Applewood Smoked Bacon}        => {Banana}       0.01025641    1.0
7.222222     2
[5]{Organic Green Beans}           => {Organic
                                        Strawberries} 0.01025641    1.0
8.478261     2
[6]{Blackberries}                  => {Organic
                                        Strawberries} 0.01025641    1.0
8.478261     2
[7]{Organic Mandarins}             => {Organic
                                        Strawberries} 0.01025641    1.0
8.478261     2
```

3.Frequent itemsets for departments in orders dataset(which department have the highest number of orders)

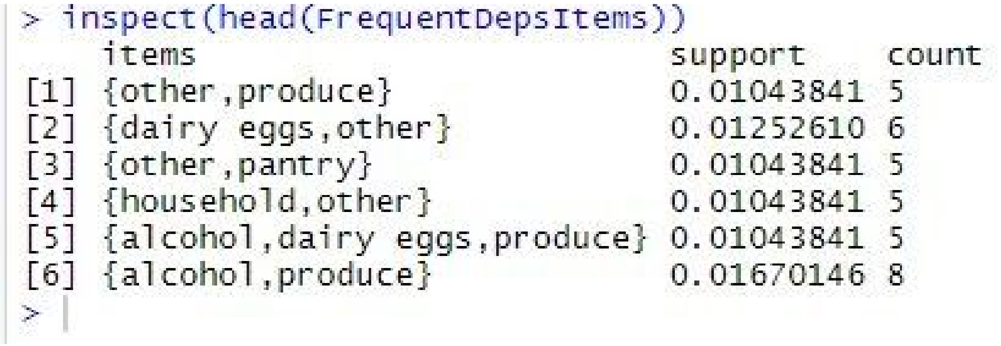


From the picture above, we can see the top 20 departments which have the highest number of orders.For this, we are merging two files namely “departments.csv” and “products.csv” to get department name.

We mine the frequent itemsets using “eclat” with support of 0.01 and max length of 8.
The below screenshot depicts part of the output of **inspect(head(FrequentDepsItems))**



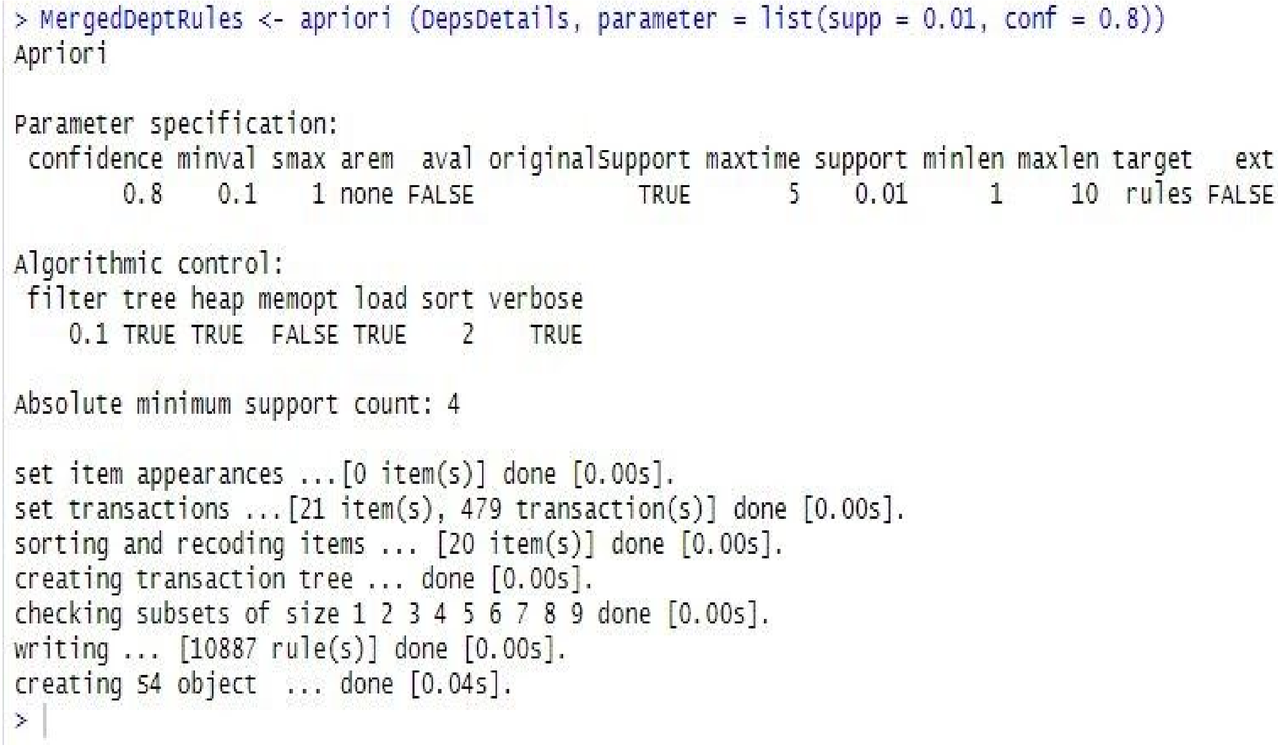
```
> FrequentDepsItems <- eclat(DepsDetails, parameter = list(supp = 0.01, maxlen = 8))
Eclat

parameter specification:
 tidLists support minlen maxlen            target   ext
    FALSE    0.01      1      8 frequent itemsets FALSE

algorithmic control:
 sparse sort verbose
      7   -2    TRUE

Absolute minimum support count: 4

create itemset ...
set transactions ...[21 item(s), 479 transaction(s)] done [0.00s].
sorting and recoding items ... [20 item(s)] done [0.00s].
creating bit matrix ... [20 row(s), 479 column(s)] done [0.00s].
writing  ... [5759 set(s)] done [0.04s].
Creating S4 object  ... done [0.00s].
```



```
> inspect(head(FrequentDepsItems))
    items                          support    count
[1] {other,produce}                0.01043841 5
[2] {dairy eggs,other}             0.01252610 6
[3] {other,pantry}                 0.01043841 5
[4] {household,other}              0.01043841 5
[5] {alcohol,dairy eggs,produce}   0.01043841 5
[6] {alcohol,produce}              0.01670146 8
> 
```

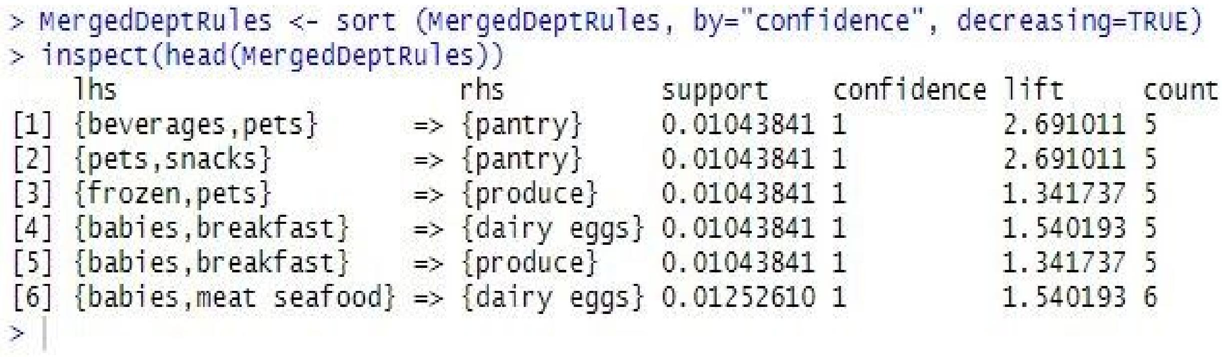
4.Association rules for departments in orders dataset

The below screenshot depicts, association rules generation using “apriori” function.



```
> MergedDeptRules <- apriori (DepsDetails, parameter = list(supp = 0.01, conf = 0.8))
Apriori

Parameter specification:
 confidence minval smax arem  aval originalSupport maxtime support minlen maxlen target   ext
        0.8    0.1    1 none FALSE            TRUE       5    0.01      1     10  rules FALSE

Algorithmic control:
 filter tree heap memopt load sort verbose
    0.1 TRUE TRUE  FALSE TRUE    2    TRUE

Absolute minimum support count: 4

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[21 item(s), 479 transaction(s)] done [0.00s].
sorting and recoding items ... [20 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3 4 5 6 7 8 9 done [0.00s].
writing ... [10887 rule(s)] done [0.00s].
creating S4 object  ... done [0.04s].
> 
```

The generated association rules data have been sorted based on confidence values.Refer the details below for more information.



```
> MergedDeptRules <- sort (MergedDeptRules, by="confidence", decreasing=TRUE)
> inspect(head(MergedDeptRules))
    lhs                        rhs          support    confidence lift     count
[1] {beverages,pets}        => {pantry}     0.01043841 1          2.691011 5
[2] {pets,snacks}           => {pantry}     0.01043841 1          2.691011 5
[3] {frozen,pets}           => {produce}    0.01043841 1          1.341737 5
[4] {babies,breakfast}      => {dairy eggs} 0.01043841 1          1.540193 5
[5] {babies,breakfast}      => {produce}    0.01043841 1          1.341737 5
[6] {babies,meat seafood} => {dairy eggs} 0.01252610 1          1.540193 6
> 
```