**Instacart Market Basket Analysis**

**Lokeswari Umakanthan (LXU190000)**

**Dataset :** The Dataset is retrieved from the Kaggle Instacart Market Basket Challenge <https://www.kaggle.com/c/instacart-market-basket-analysis/overview>.

The dataset has three set of files describing customer’s orders and transactions. The 3 Datasets are:

1. Order\_products\_prior.csv

2. Products.csv

3. Departments.csv

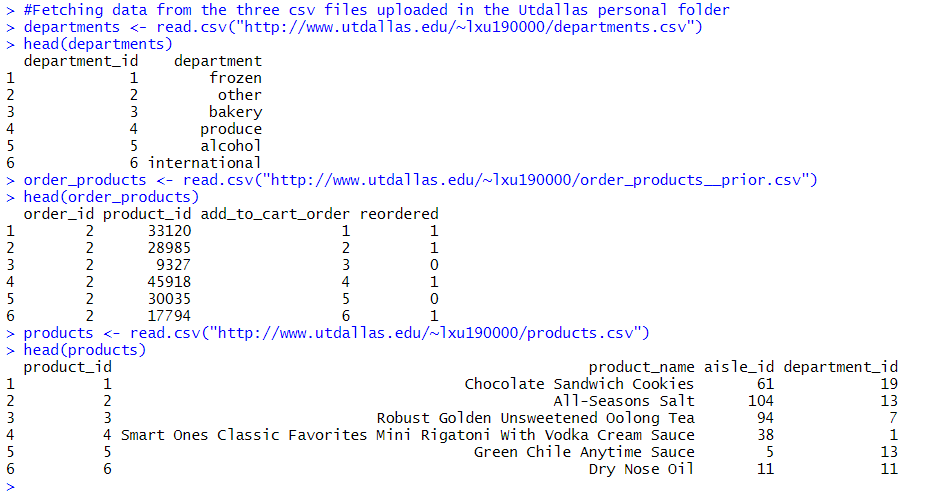
Packages required:

• Arules

• Plyr

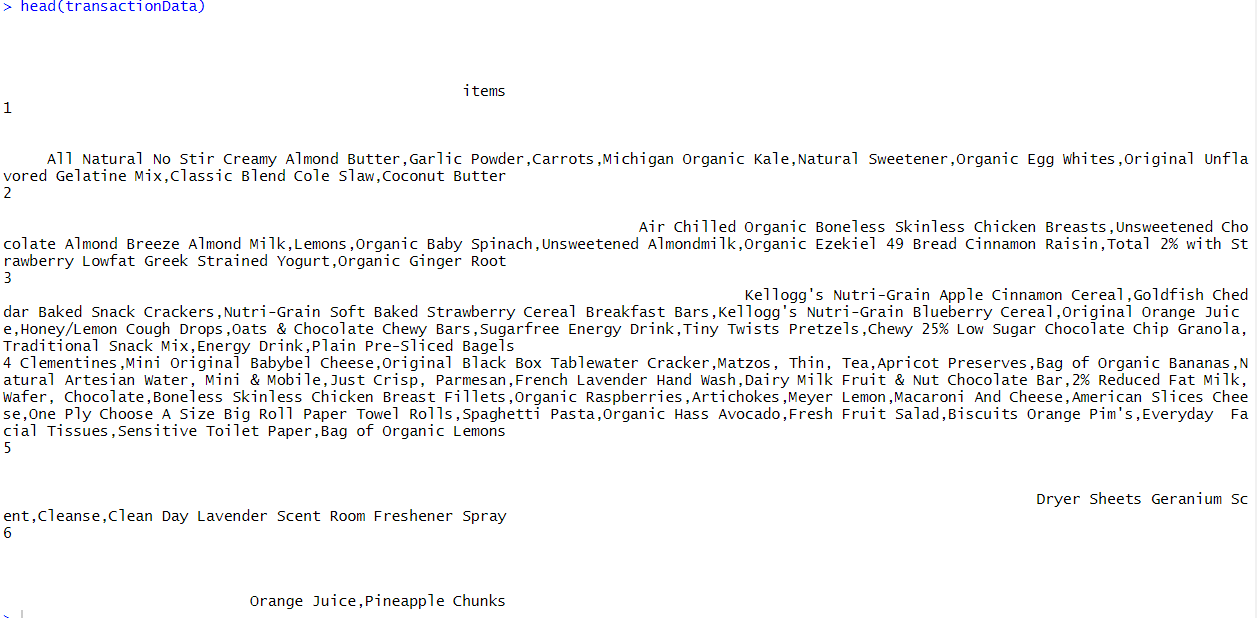
Fetching Data:

* The data of all the three files are fetched using read.csv(“Path to the file”) function.
* The head function retrieves the top 6 rows of each dataframe.

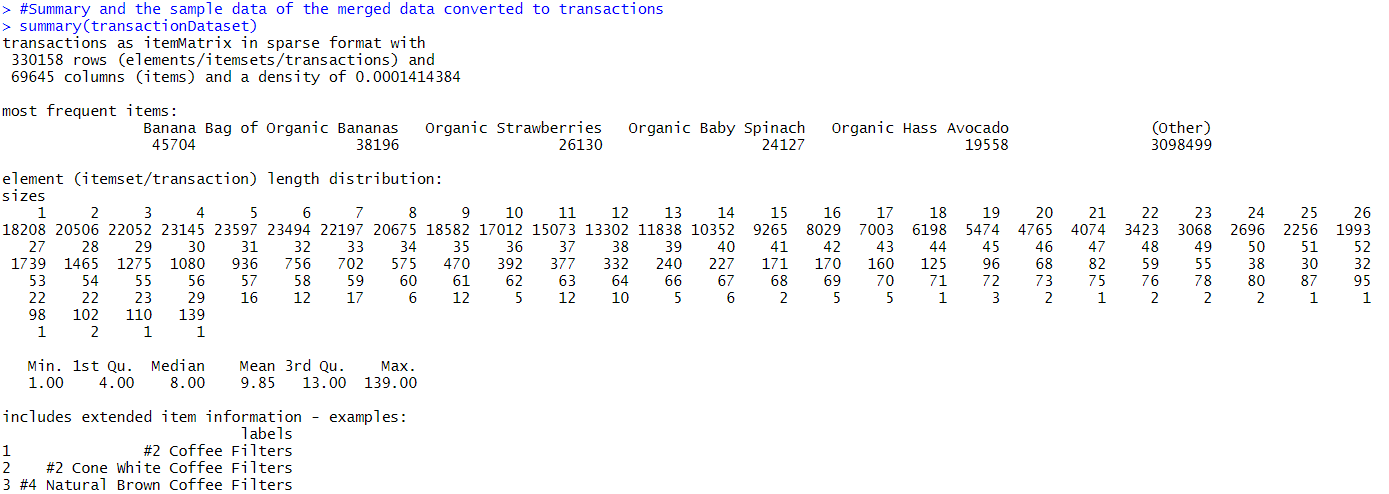


Merging two Dataframes (order\_products and products):

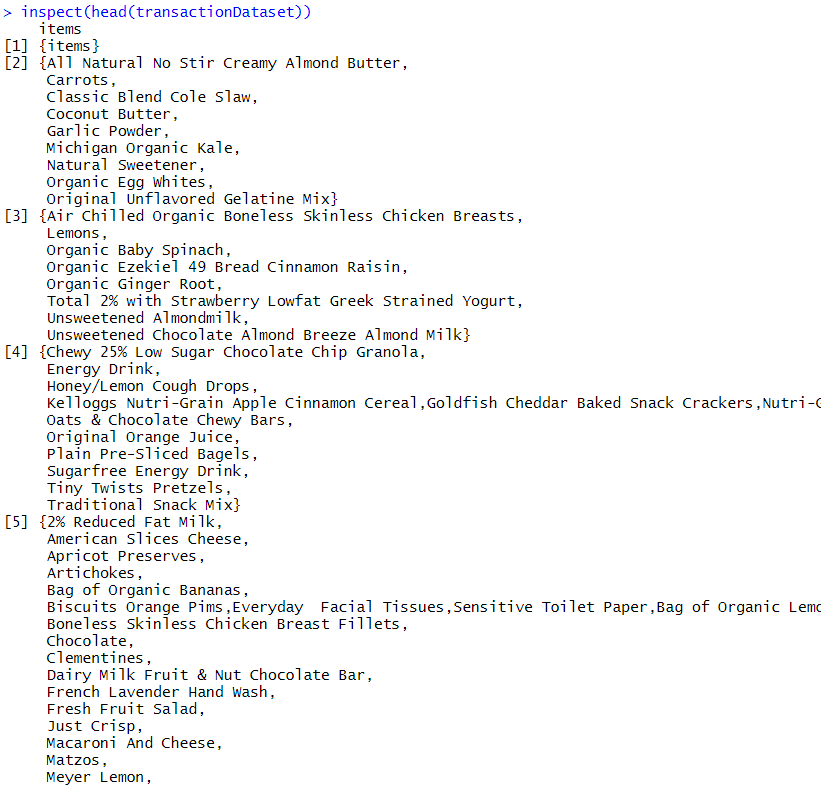
* Grouping the data using Order\_id and merging the products that’s having the same Order\_id.
* The merge function is used to join the two tables. In the Join Table the ddply() function is used to group and merge the products in the same cell for a single order ID. The resulted grouped and merged table is stored in transactionData.
* Top 6 of the sample transactionData is as below:



* The transactionData is written in local system, uploaded to Personal UtDallas website and read from the public URL. The summary of transactionDataset is as below:

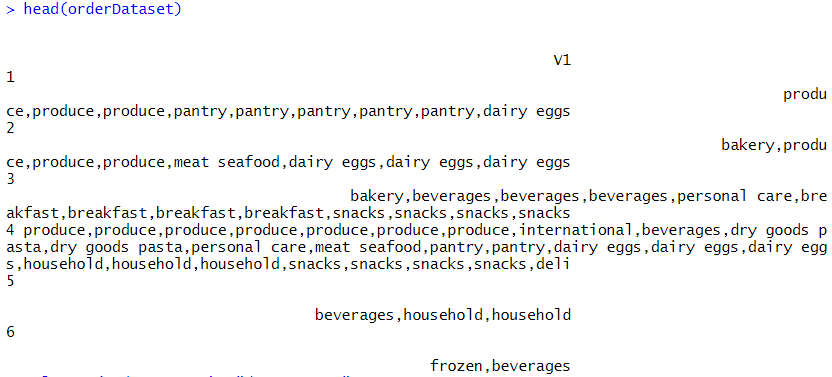


* Inspecting transactionDataset gives the list of items in each cells. It compiles the list of items based on their orderID.

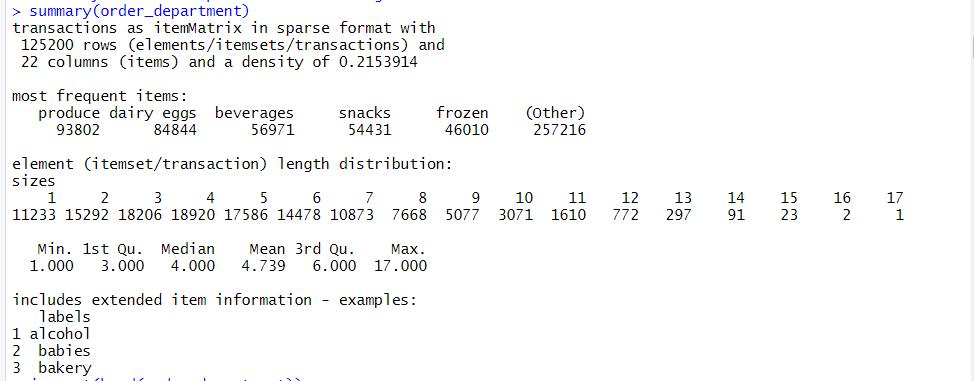


Merging two Dataframes (order\_products and departments):

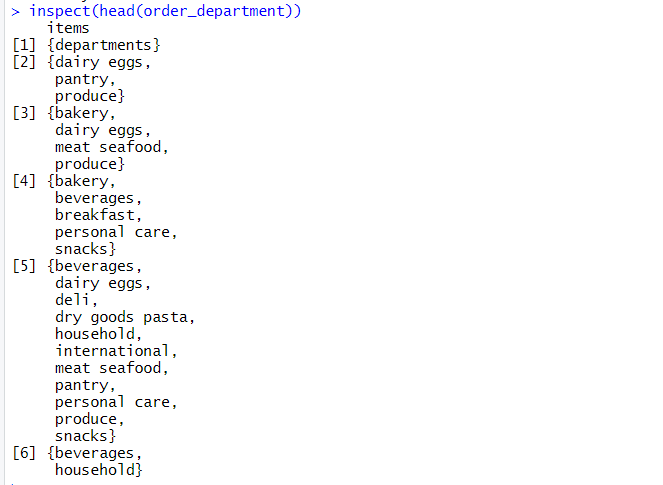
* Grouping the data using Order\_id and merging the departments that’s having the same Order\_id.
* The merge function is used to join the two tables. In the Join Table the ddply() function is used to group and merge the departments in the same cell for a single order ID.
* The resulted grouped and merged table is stored in orderDataset .
* Top 6 of the sample orderDataset is as below:



* The orderDataset is written in local system, uploaded to Personal UtDallas website and read from the public URL. The summary of order\_department is as below:

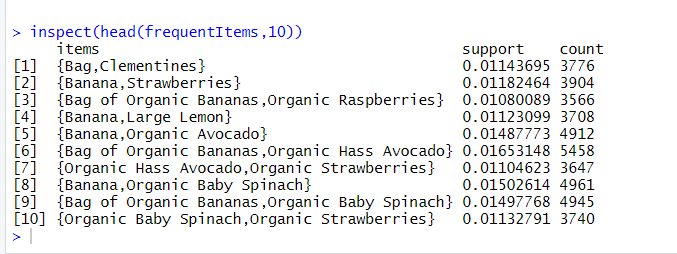


* Inspecting order\_department gives the list of items in each cells. It compiles the list of items based on their orderID.

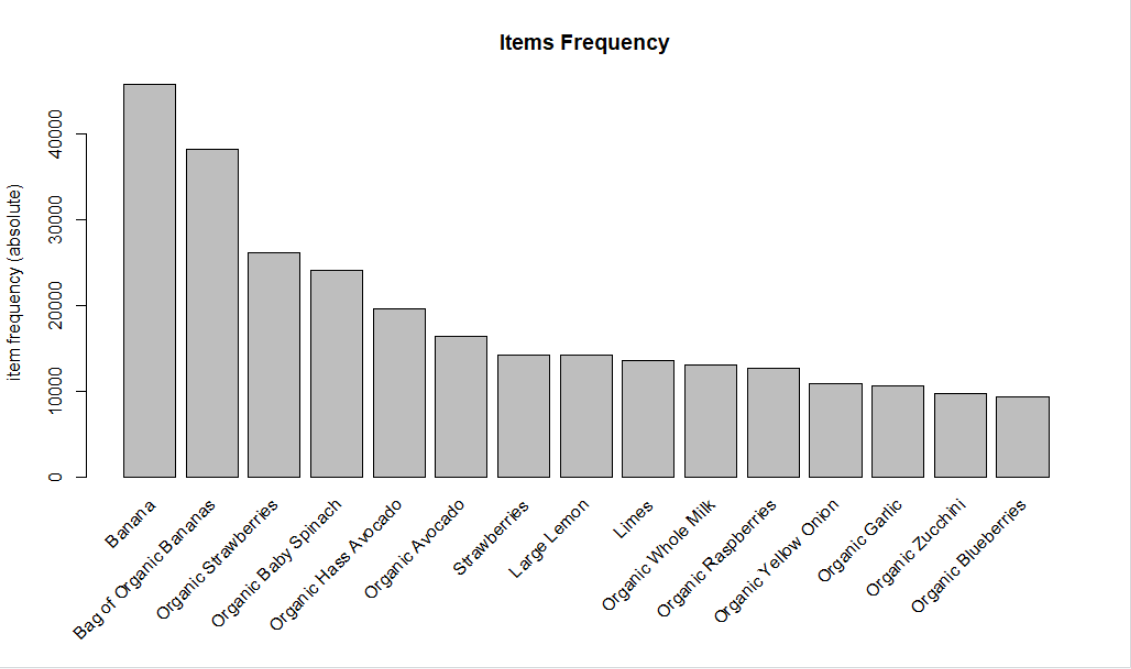


**1.Frequent itemsets for products in transactionDataset. Output product names :**

* Fetching top frequent itemsets from the transactionDataset (Joined Data of order\_products and products). The Frequent data items are fetched by using eclat() function and giving arguments as minSupport = 0.01 and maxLength = 10. The sample frequency count of the first 6 order itemset is shown below:

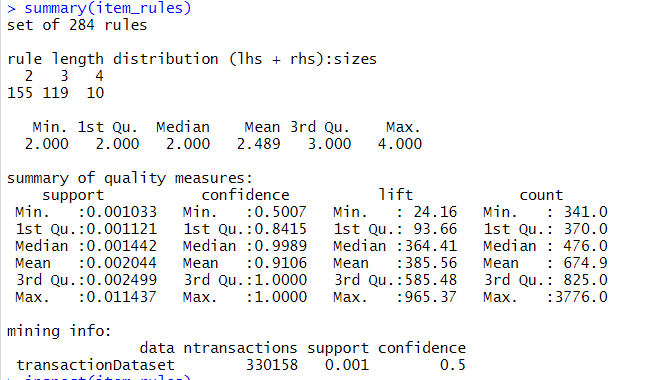


* Frequent item plot using itemfrequencyPlot() in the transactionDataset. Fetching top 15 frequent items and their count of occurrences.

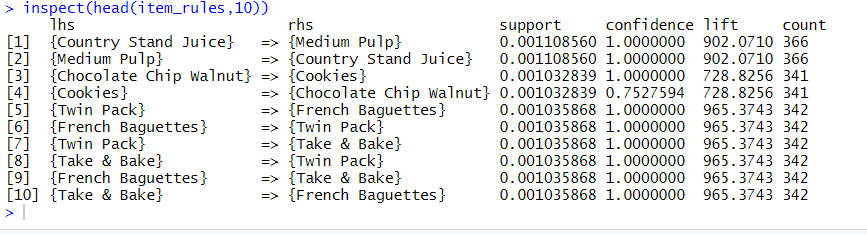


**2. Association rules for products in orders dataset. Output product names :**

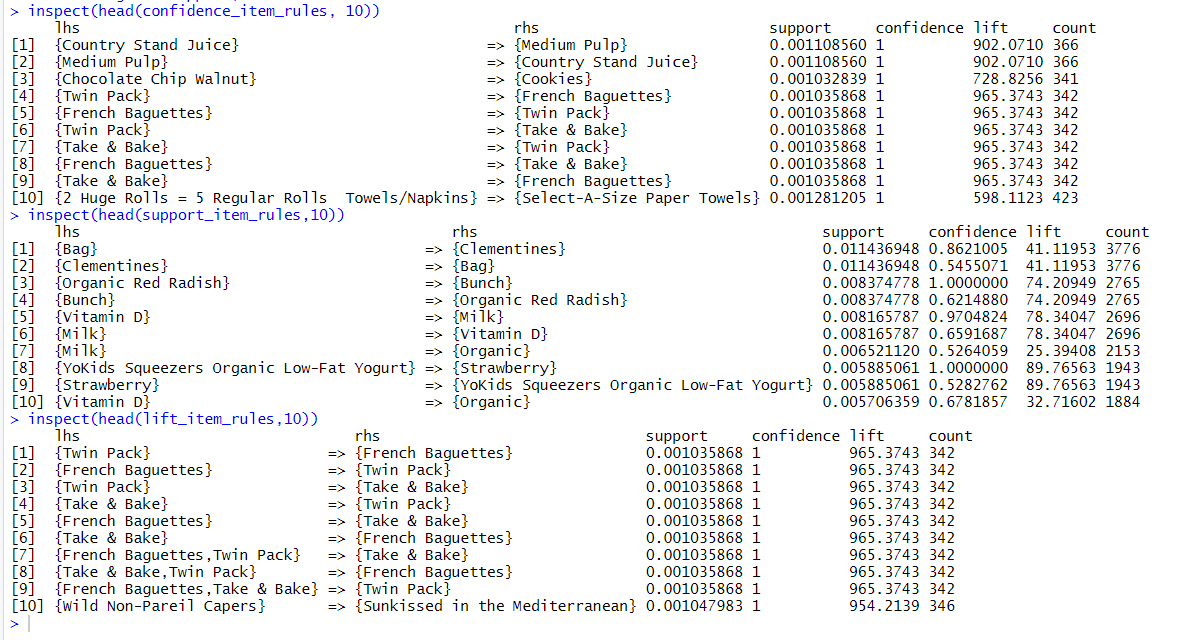
* Using apriori() function of arules package, fetching the association rules for the transactionDataset. The parameters given for the apriori() function is:
* Support = 0.001 and Confidence = 0.5



* Getting the inspect summary of the sample 6 association rules.

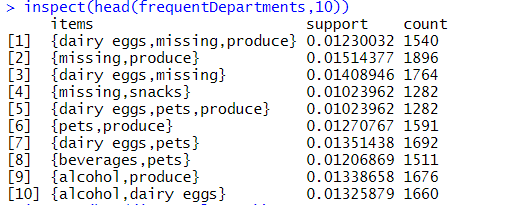


* Getting the inspect summary of the data obtained by sorting the association rules by confidence, support and lift.

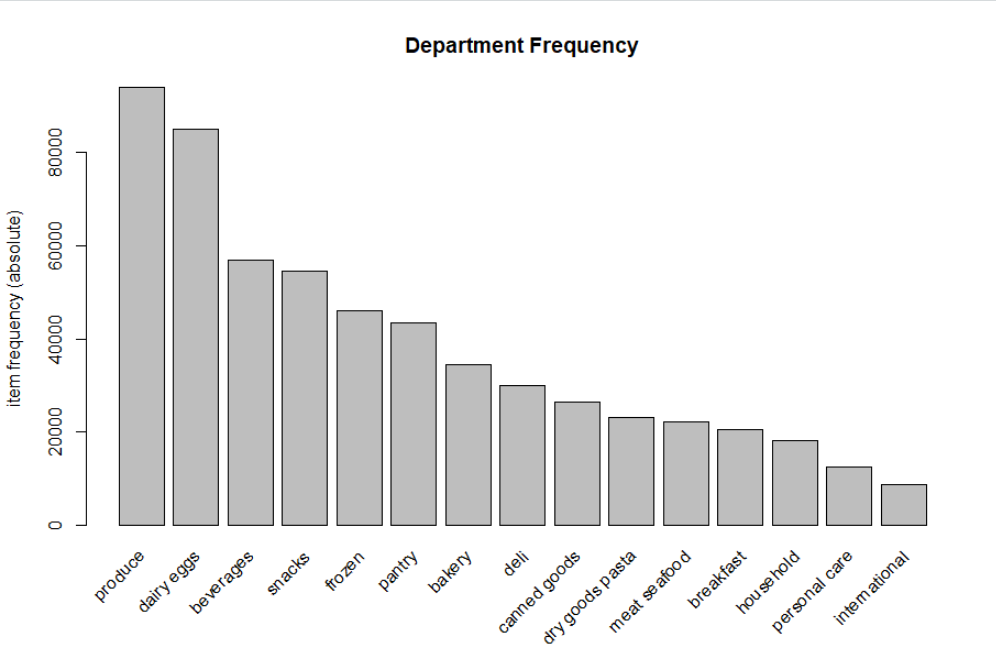


**3.Frequent departmentsets for department in order\_department Dataset. Output department names :**

* Fetching top frequent departmentsetsfrom the order\_department Dataset (Joined Data of order\_products and departments). The Frequent data items are fetched by using eclat() function and giving arguments as minSupport = 0.01 and maxLength = 10. The sample frequency count of the first 6 order itemset is shown below:

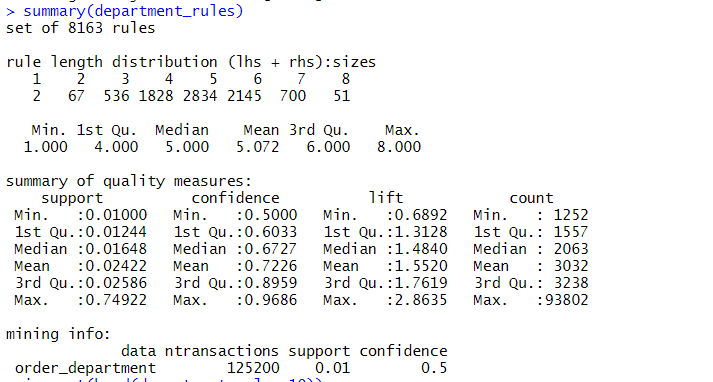


* Frequent item plot using itemfrequencyPlot() in the order\_department. Fetching top 15 frequent departments and their count of occurrences.

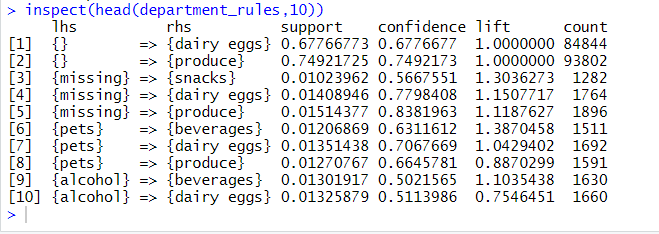


**4. Association rules for departments in order\_department dataset. Output department names:**

* Using apriori() function of arules package, fetching the association rules for the order\_department. The parameters given for the apriori() function is:
* Support = 0.001 and Confidence = 0.5



* Getting the inspect summary of the sample 6 association rules.



* Getting the inspect summary of the data obtained by sorting the association rules by confidence, support and lift.

