**Assignment 06**

Hands on Frequent Pattern Mining

**Name: Dean D’souza**

**H.U. ID: 168424**

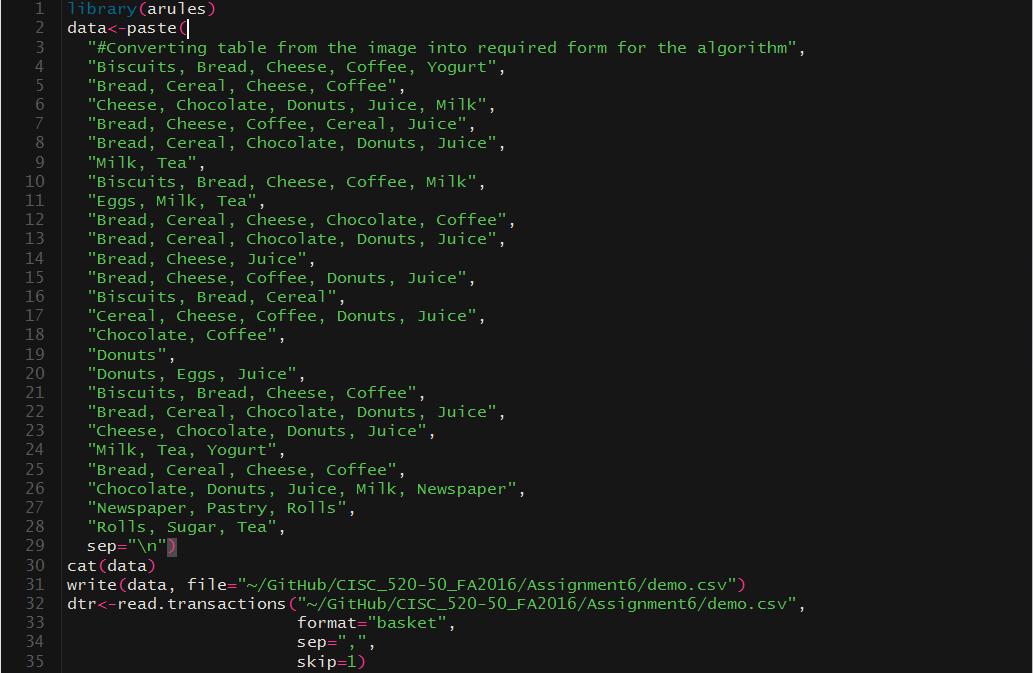
# Solutions:

**Consider the following transaction database:**

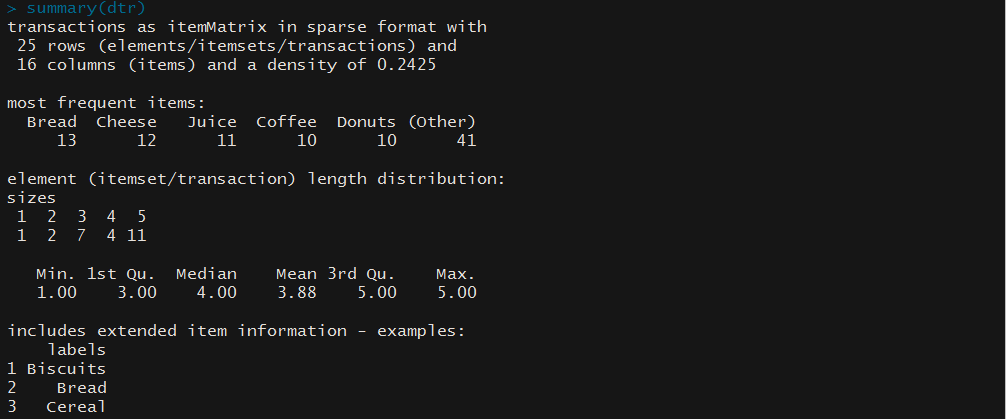
1. **Use the *Apriori algorithm* to find the frequent itemsets in D.**
2. **Extract all the strong association rules (has min conf say, 70%)**

For this assignment, I chose to perform the tasks in R using R studio and the package “arules” which contains a default apriori method.

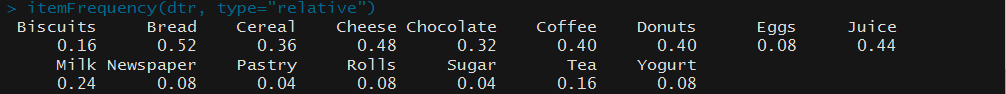
Firstly, the data needs to be converted to an appropriate form for use in R studio, and the code is as follows:

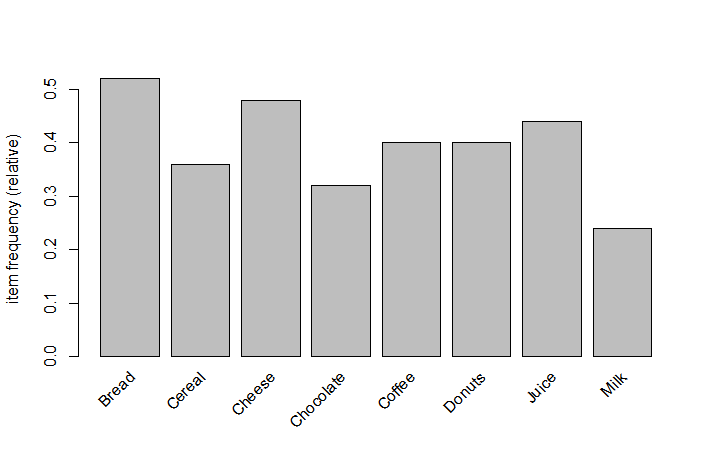
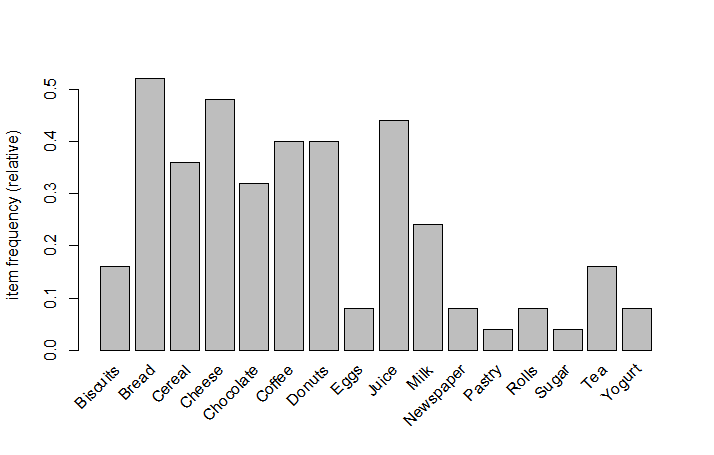


After preparing the data, summary() function was called to get a few statistics (and to verify that it was in proper form, inspect() was also used but not shown here to save space):

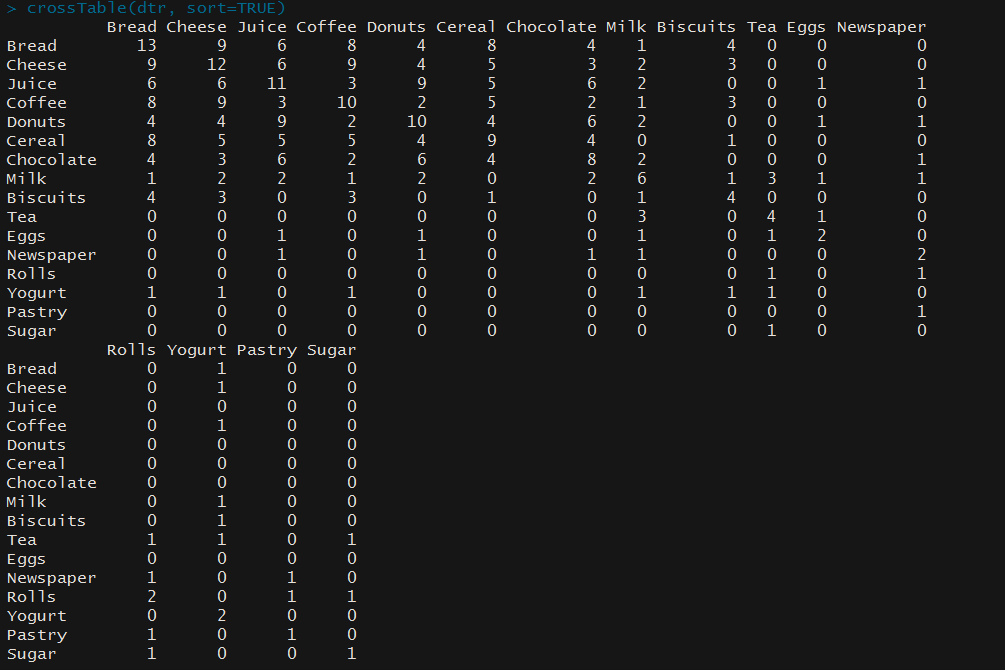


We then obtain the list of most frequent items:



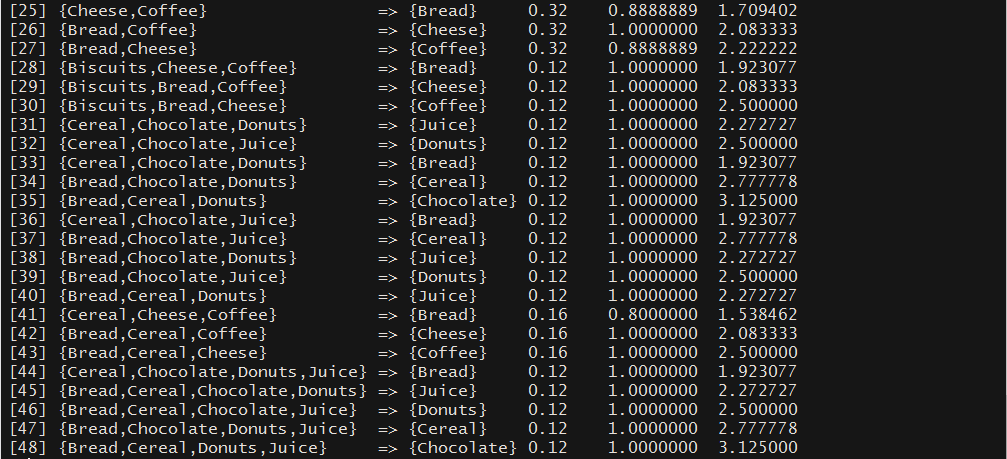
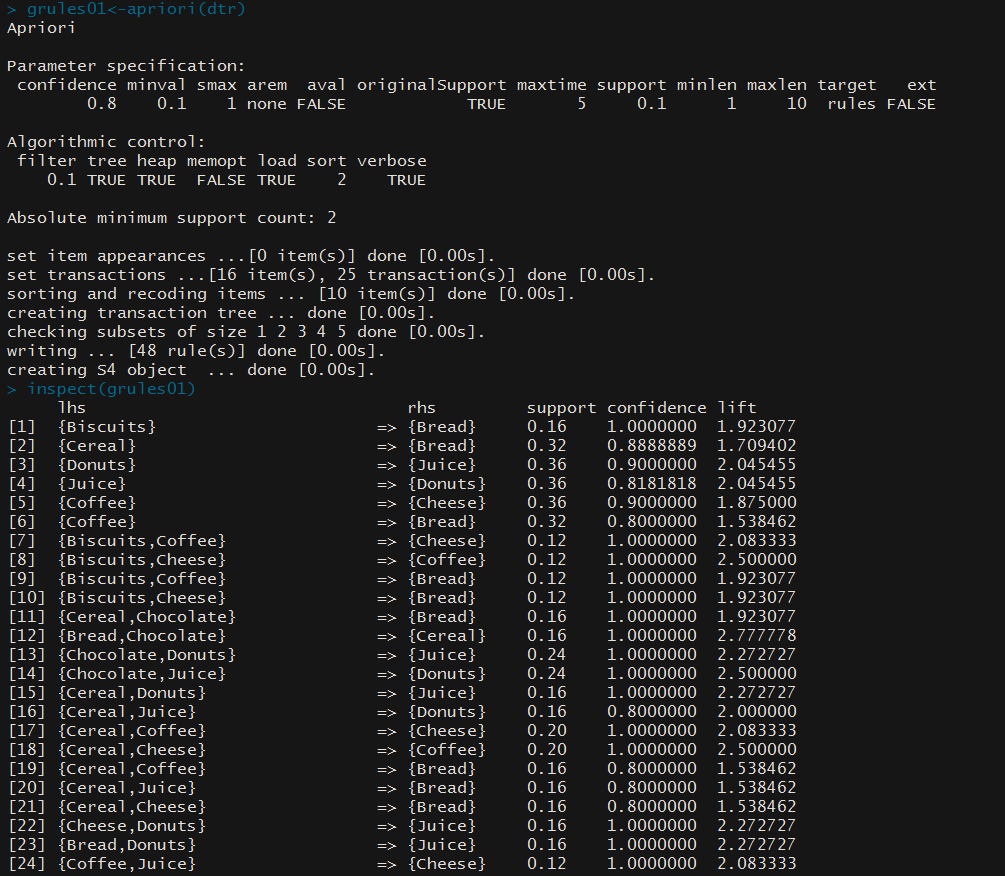


The plot above shows all items which have a support higher than 0.2. We use the crossTable() function to get an idea of frequent pairs:

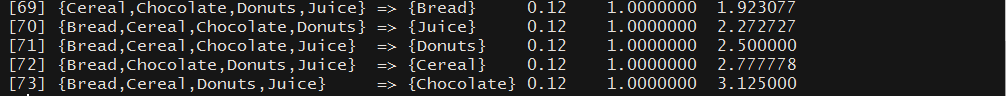
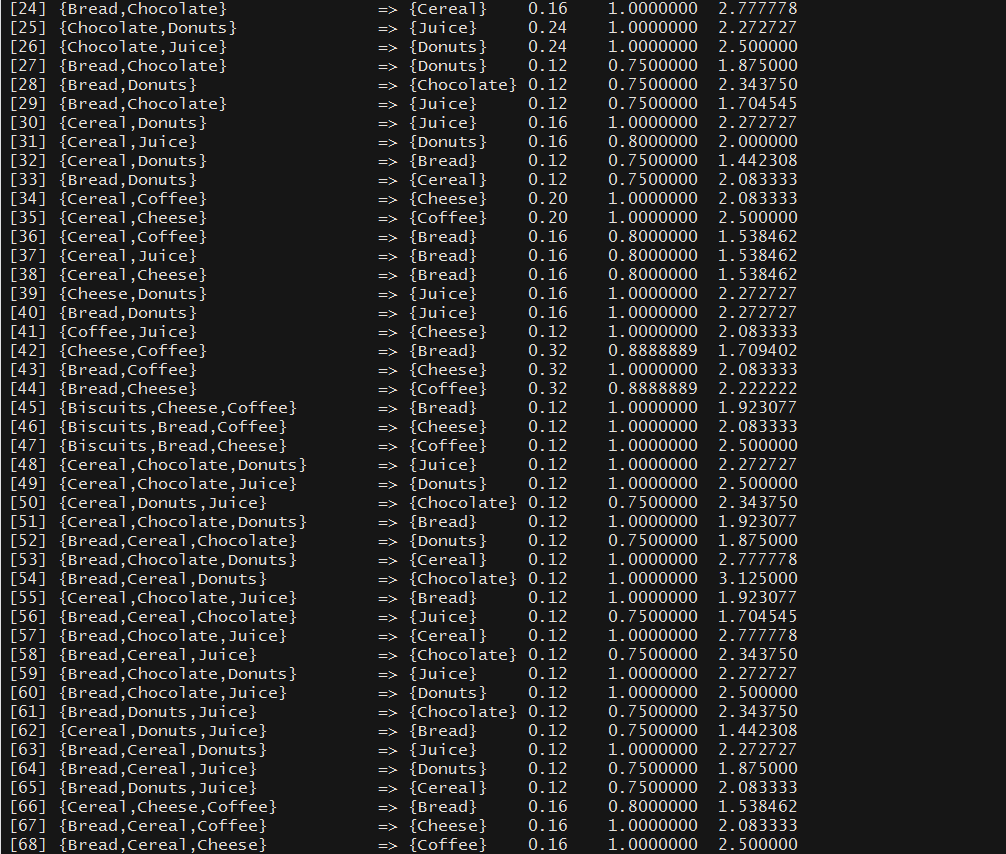
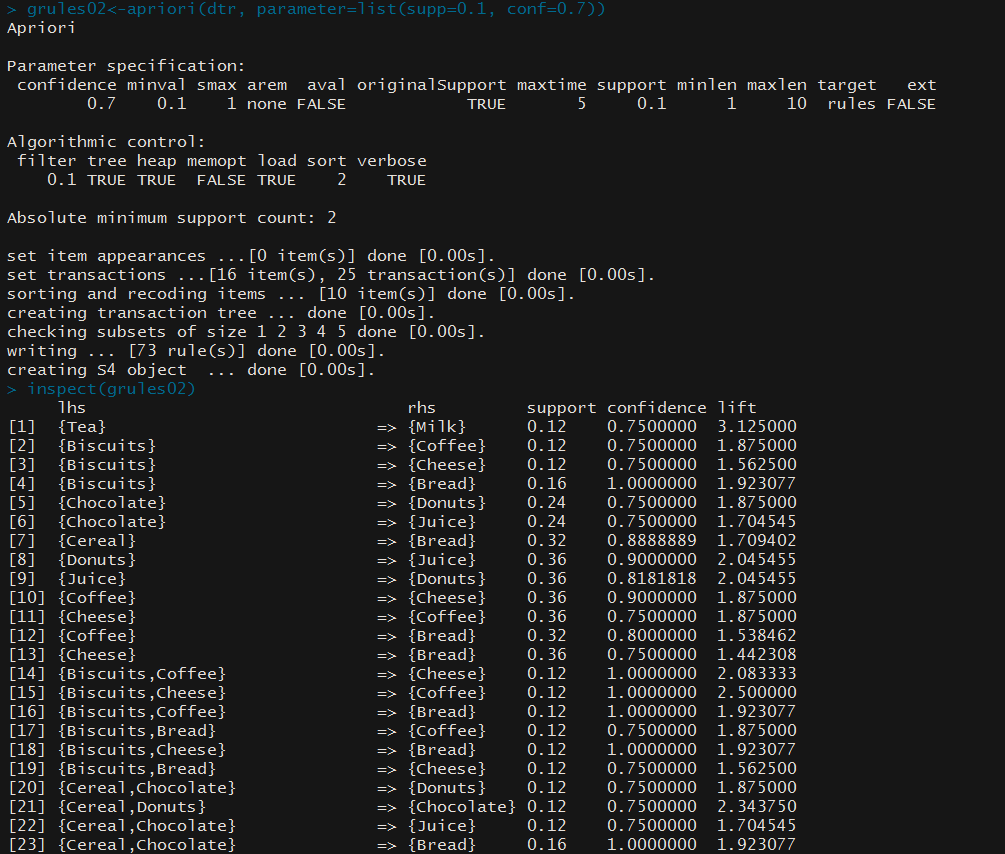


As we can see above, the most frequent pairs are {Bread, Cheese}, {Cheese, Coffee} and {Juice, Donuts}. We follow this step by starting to obtain the rules (Please skip to third Pass for final answer):

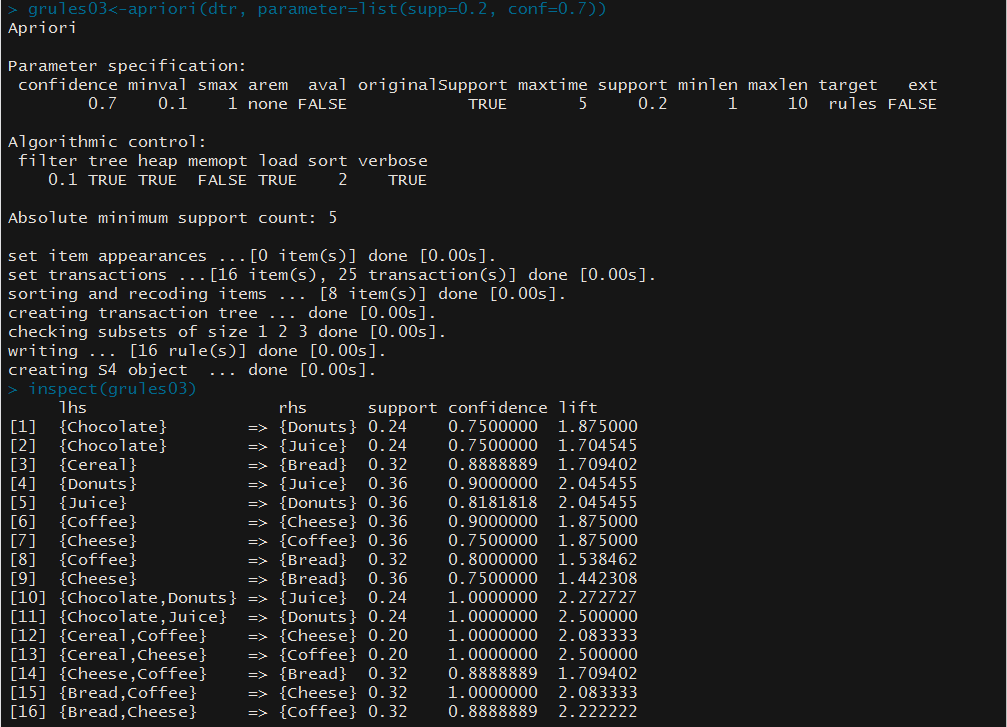
**1st Pass** (Support and Confidence Not Specified) which gives 48 rules:



**2nd Pass** (Support=0.1 and Confidence=0.7) which gives 73 rules:



**3rd Pass** (Support =0.2 and Confidence= 0.7) which gives 16 rules:



The above screenshot is small enough to get a good set of important rules, which include:

{Donuts} => {Juice}

{Coffee} => {Cheese}

{Chocolate, Donuts} => {Juice}

{Chocolate, Juice} => {Donuts}

{Cereal, Coffee} => {Cheese}

{Cereal, Cheese} => {Coffee}

{Bread, Coffee} => {Cheese}

These Rules were chosen from the list as they not only have good support and confidence of 1, but also because they have lift (which is the % chance of buying the item in L.H.S.) higher than 2 (except for {Coffee}=>{Cheese}).

# References:

[1] “Data Mining: Association Analysis” Lecture Slides, Stephen Penn, DM, PMP, Harrisburg University of Science and Technology, ANLY-510, Summer 2016

[2] “Market Basket Analysis with R”, Deepanshu Bhalla, ListenData, (<http://www.listendata.com/2015/12/market-basket-analysis-with-r.html>)