MTH 443: Lab Problem Set 7

[1] The dataset "grocery_baskets.csv" gives data on 5400 transactions at a supermarket chain. Every row of the dataset is a transaction and lists the items for a single transaction.

(a) Find the top 25 items with respect to support of items.

(b) Use apriori algorithm to find the association rules for the following $(minsup, minconf)$ combinations:

(i) $minsup = 0.1; minconf = 0.6$

(ii) $minsup = 0.01; minconf = 0.2$

(iii) $minsup = 0.05; minconf = 0.3$

(iv) $minsup = 0.5; minconf = 0.7$

(c) Calculate LIFT of all the generated rules in **(b)** for the different combinations.

LIFT is defined as the ratio of confidence for the rule to the prior probability of having the rule prediction, i.e. $LIFT(A => B) = \frac{CONF(A=>B)}{SUPPORT(B)/N}$. *A lift value of 1 indicates that the two items are independent, while a value greater than 1 indicates a positive association. A value less than 1 indicates a negative association.*

R Package for Association Rule Mining: arules, arulesViz

Python: libraries available for apriori algorithm