# Problem statement

An organization wanted to mine association rules of frequently bought items from its stores and suggest some recommendations to its customers. As a data scientist, you are required to recognize patterns from the available data and evaluate efficacy of methods to obtain patterns. Your activities should include - performing various activities pertaining to the data such as, preparing the dataset for analysis; investigating the relationships in the data set with visualization; identify frequent patterns; formulate association rules and evaluate quality of rules.

# Proposed Solution

We will be using Apriori Algorithm to mine Association rules to identify underlying relations between different items. In this case, we are considering data from a store where customers can buy variety of items. Usually, there is a pattern in what the customers buy. More profit can be generated if the relationship between the items purchased in different transactions can be identified.

For instance, if item A and B are bought together more frequently then several steps can be taken to maximize the sales/profit. Below are some of the suggested steps to do so.

1. A and B can be placed together. So, when a customer buys one of the products, he doesn't have to go far away to buy the other product or has increased tendency to buy the other product.
2. People who buy one of the products can be targeted through an advertisement campaign/attractive pricing to buy the other.
3. Collective discounts can be offered on these products.
4. Both A and B can be packaged together.

# Input

The input file used is Dataset.xlsx. It contains 7500 transactions of various items. To apply apriori algorithm on the dataset, we need to set 4 parameters to start with.

1. Minimum support = 0.0045
2. Minimum confidence = 20% = 0.2
3. Minimum length = 2 to create 2 itemset rules
4. Minimum lift = 3

# Output

The output is the set of association rules with 2 itemset. This is written into an excel sheet, Output.xlsx, for easy readability and analysis.

# Observations

From the first rule, we can see that chicken and light cream are commonly bought together. This means that people who bought chicken also bought light cream. One of the reasons could be that they use light cream to cook the chicken using a popularly used recipe.

The support value for the first rule is 0.0045. This number is calculated by dividing the number of transactions containing both chicken and light cream divided by the total number of transactions. The confidence level for the rule is 0.2905 which shows that out of all the transactions that contain light cream, 29.05% of the transactions also contain chicken. Finally, the lift of 4.84 means that chicken is 4.84 times more likely to be bought by the customers who buy light cream compared to the default likelihood of the sale of chicken.

Similarly, the second rule states that mushroom cream sauce and escalope are bought frequently. The support for mushroom cream sauce is 0.0057. The confidence for this rule is 0.3006 which means that out of all the transactions containing mushroom, 30.06% of the transactions are likely to contain escalope as well. Finally, lift of 3.79 shows that the escalope is 3.79 more likely to be bought by the customers that buy mushroom cream sauce, compared to its default sale