**DATA WAREHOUSING AND DATA MINING LAB**

**EXPERIMENT 3 : Implement Apriori algorithm to find frequent itemsets and generate association rules  
Aim:**

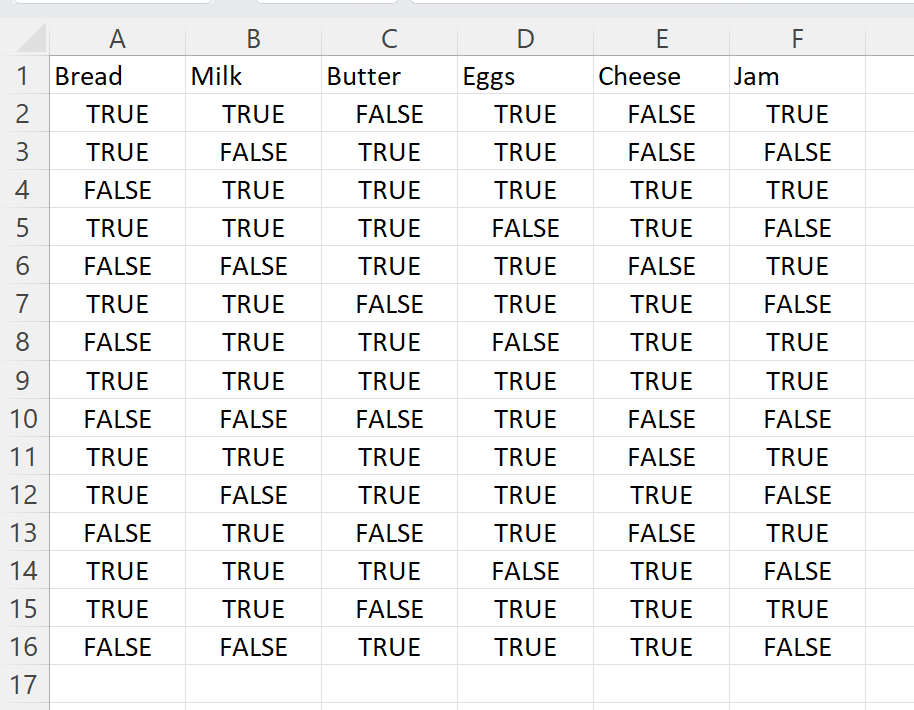
To find **frequent itemsets** and generate **association rules** using the **Apriori algorithm** in the Orange Data Mining Tool.

**Description :**

The program implements the **Apriori algorithm,** a popular method for mining **frequent itemsets** and generating **association rules** from transaction data. The algorithm identifies patterns in datasets by examining item combinations that frequently appear together, using a **minimum support** threshold. It then generates **association rules** such as **“if item A is bought, item B is likely to be bought”**, based on a **minimum confidence** threshold. The Apriori algorithm is commonly used in **market basket analysis** to uncover hidden relationships between products.

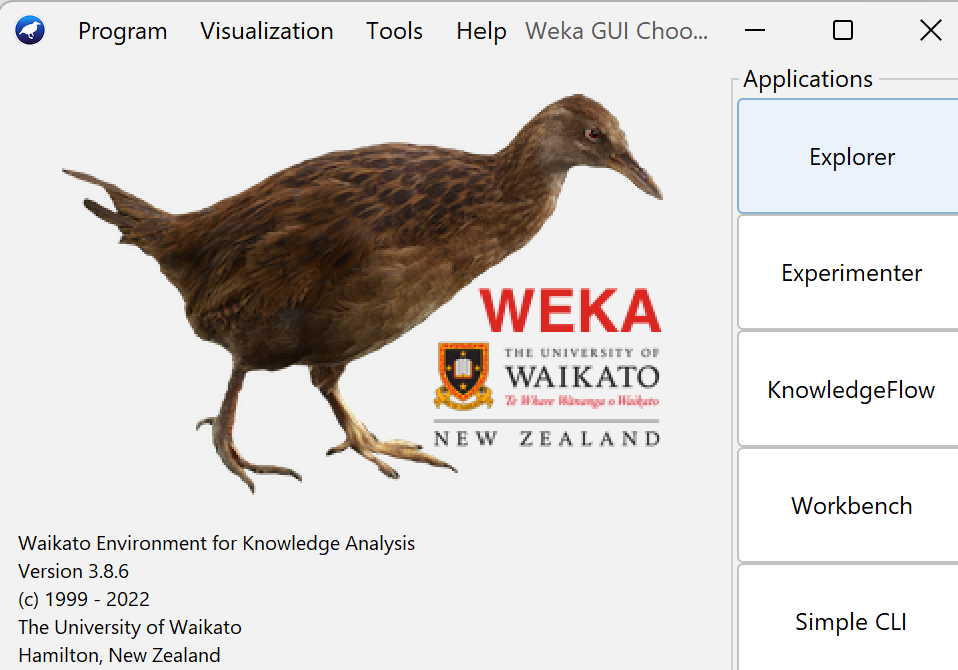
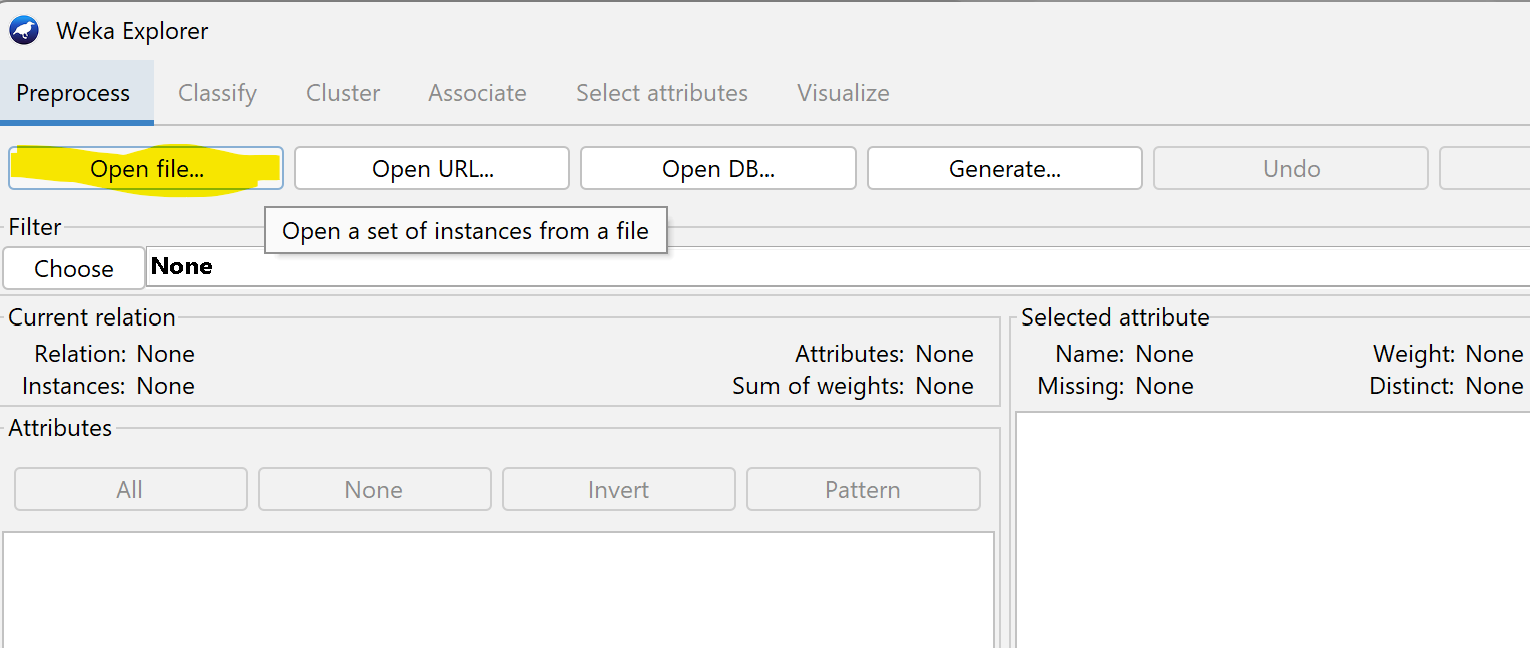
**EXECUTION USING WEKA :**

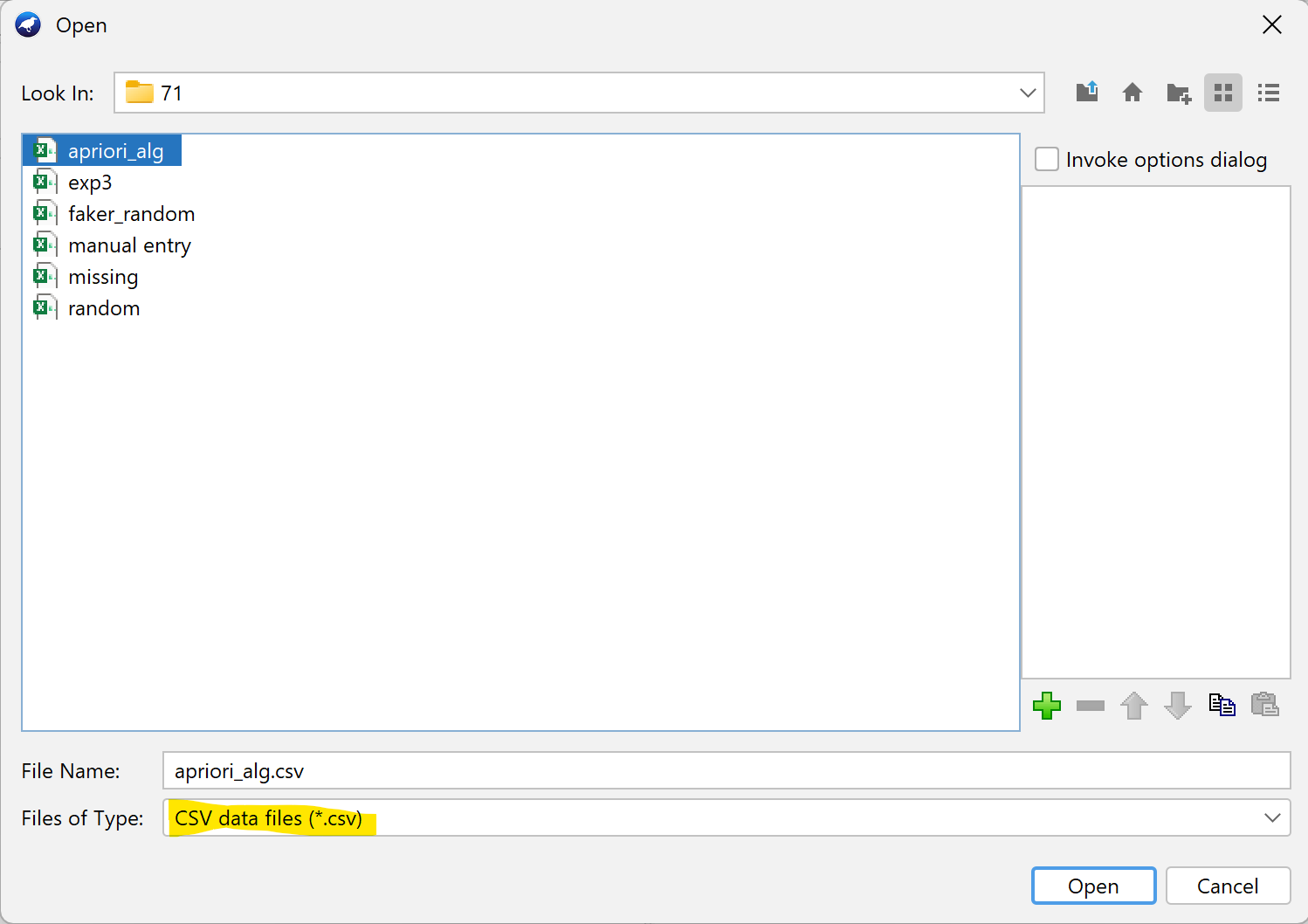
**Step 1 : Create a CSV file with the following Data Set**

****

**Step 2: Download WEKA Tool**

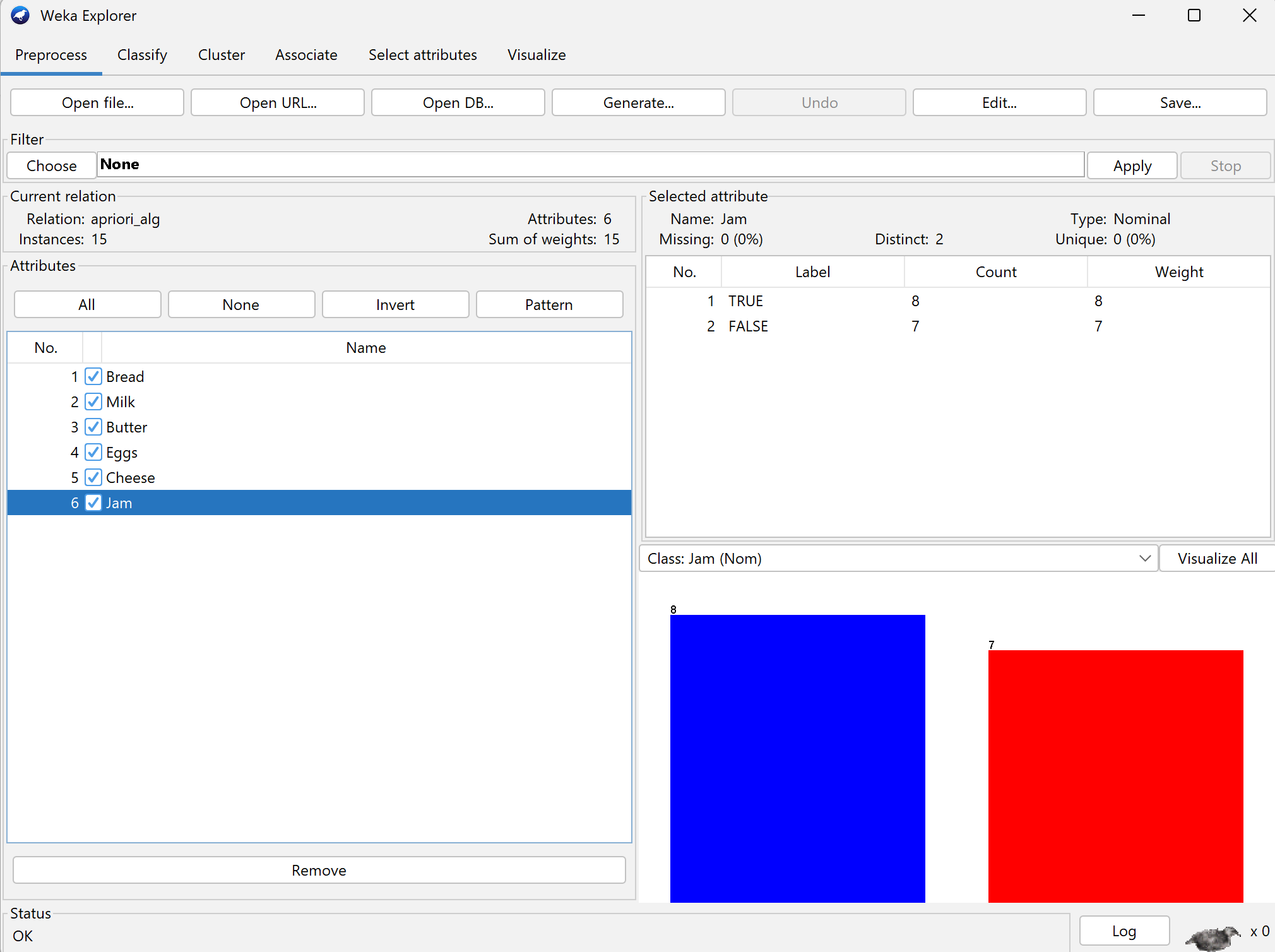
* Open WEKA GUI Chooser.
* Click on Explorer.
* Click Open File, and choose your CSV file (apriori\_alg.csv)

** **

****

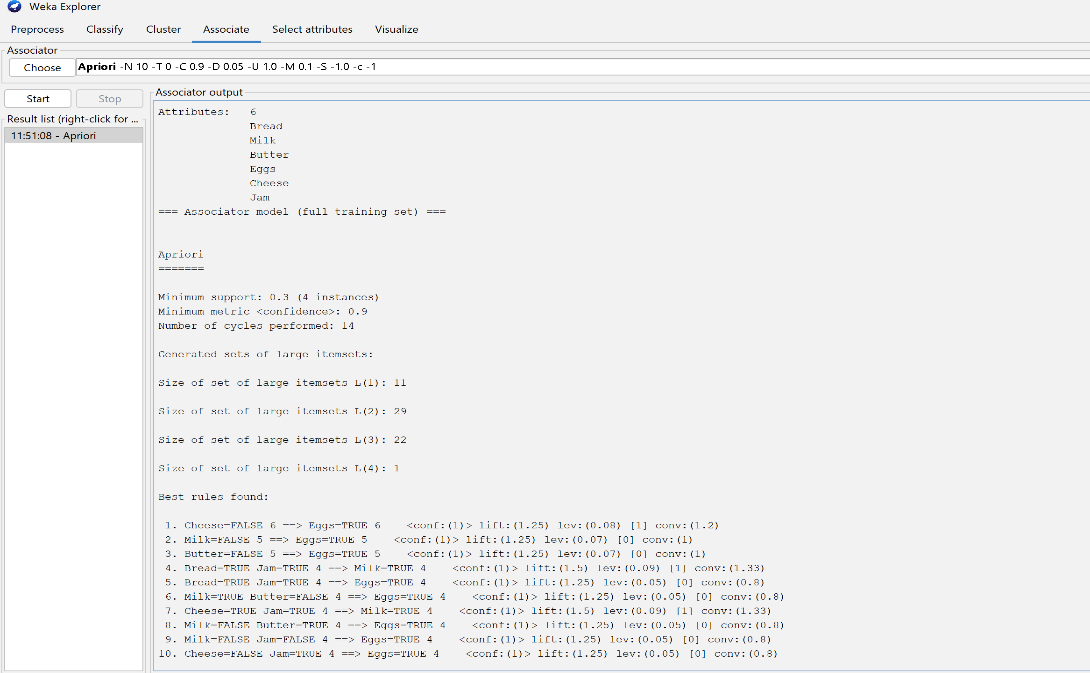
**Step 3 : Select all columns**

All items (Bread, Milk, Butter, Eggs, Cheese, Jam) should be categorical with values like "True" or "False", not numerical.

****

**Step 4 : After loading the dataset, go to the “Associate” tab at the top.**

**In the “Associator” box, select Apriori from the list. Click Start.**

****

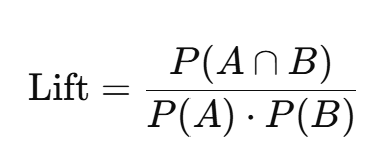
**Rule**: e.g., Milk=Yes ⇒ Bread=Yes means if Milk is bought, Bread is also likely to be bought.

**Confidence**: The probability that the RHS (right-hand side) item is bought when the LHS (left-hand side) item is bought.

conf: (1) means the rule is correct 100% of the time.

**Support**: The proportion of transactions in which both items in the rule appear.

**Lift**: How much more often the LHS and RHS appear together than if they were independent.  
**Formula**:



**Leverage**: The difference between the observed co-occurrence and what would be expected if they were independent.  
**Formula**:



**Conviction**: How strongly the presence of LHS predicts the absence of RHS (or vice versa); reflects the **implication** strength.

* Conviction = 1: No implication , Conviction > 1: Stronger implication