**DATA WAREHOUSING AND DATA MINING LAB**

**Experiment 5: Mining Frequent Patterns using FP-Growth Algorithm**

**Aim:**  
 To apply the FP-Growth algorithm on a retail dataset to identify frequent patterns and generate association rules.

**Description:**  
 FP-Growth (Frequent Pattern Growth) is an advanced pattern mining algorithm used to find frequent itemsets without generating candidate sets. It is faster and more memory-efficient than the Apriori algorithm. In this experiment, a transactional dataset is used, and association rules are generated based on a minimum support and confidence. The outcome reveals patterns like “If a customer buys milk, they also buy bread” with confidence scores. The data format must be in **True/False** or **Yes/No** format for each product column per transaction.

**Orange Tool :**

Orange enables efficient implementation of algorithms like FP-Growth without any coding. Its visual workflow makes it easy to explore data, discover hidden patterns, and analyze results quickly and clearly.

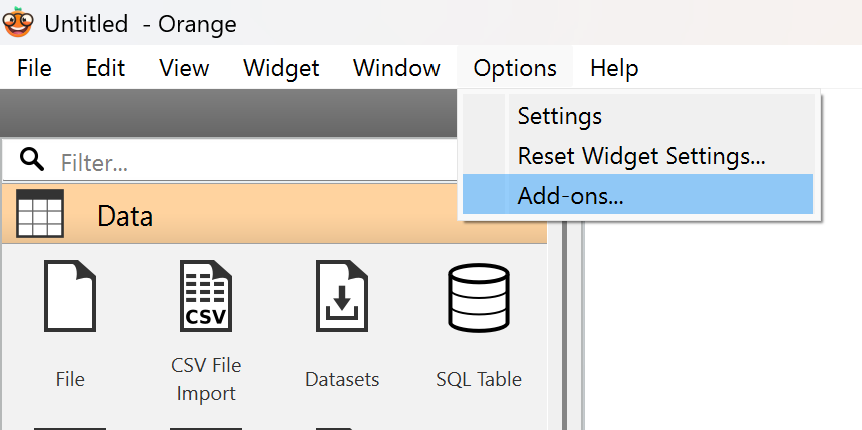
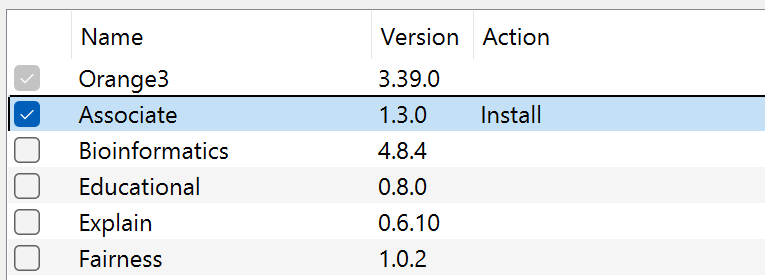
**Step 1: Create a CSV file with the following Data Set**

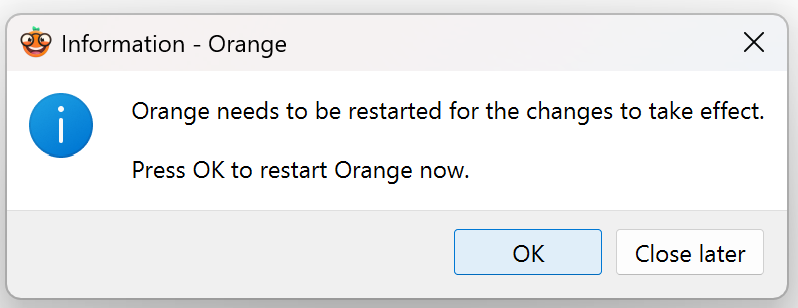


**Note:** For Apriori to work in Orange, your dataset must have attributes in **Yes/No** format.

**Step 2: Download Orange Tool**

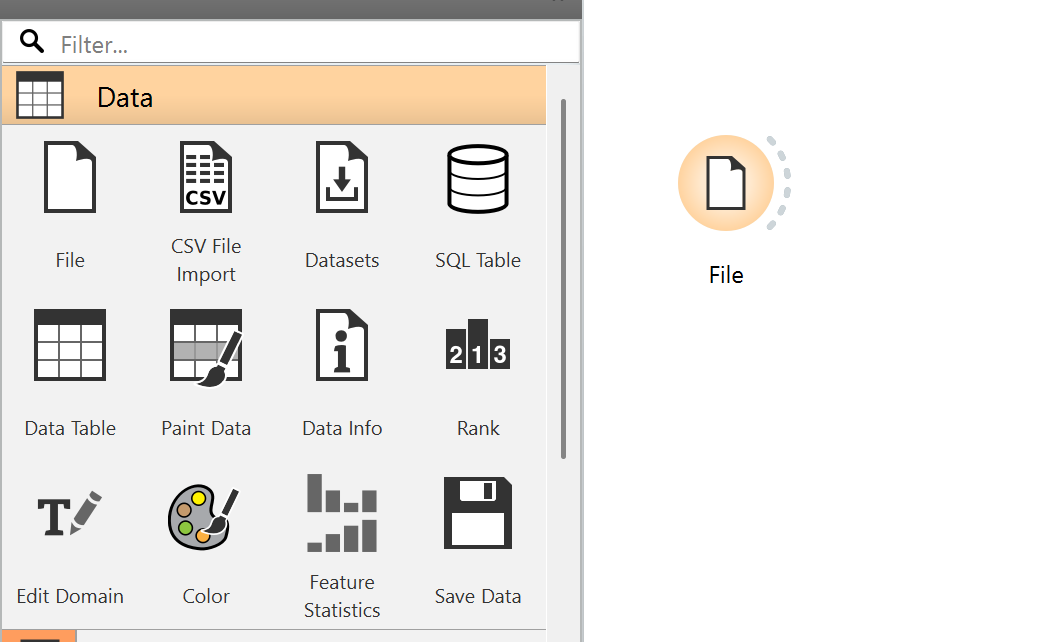
* Open Orange
* Click Options from the top menu
* Select Add-ons...
* Scroll and check Associate (this includes Apriori algorithm)
* Click Install Selected
* Restart Orange when prompted

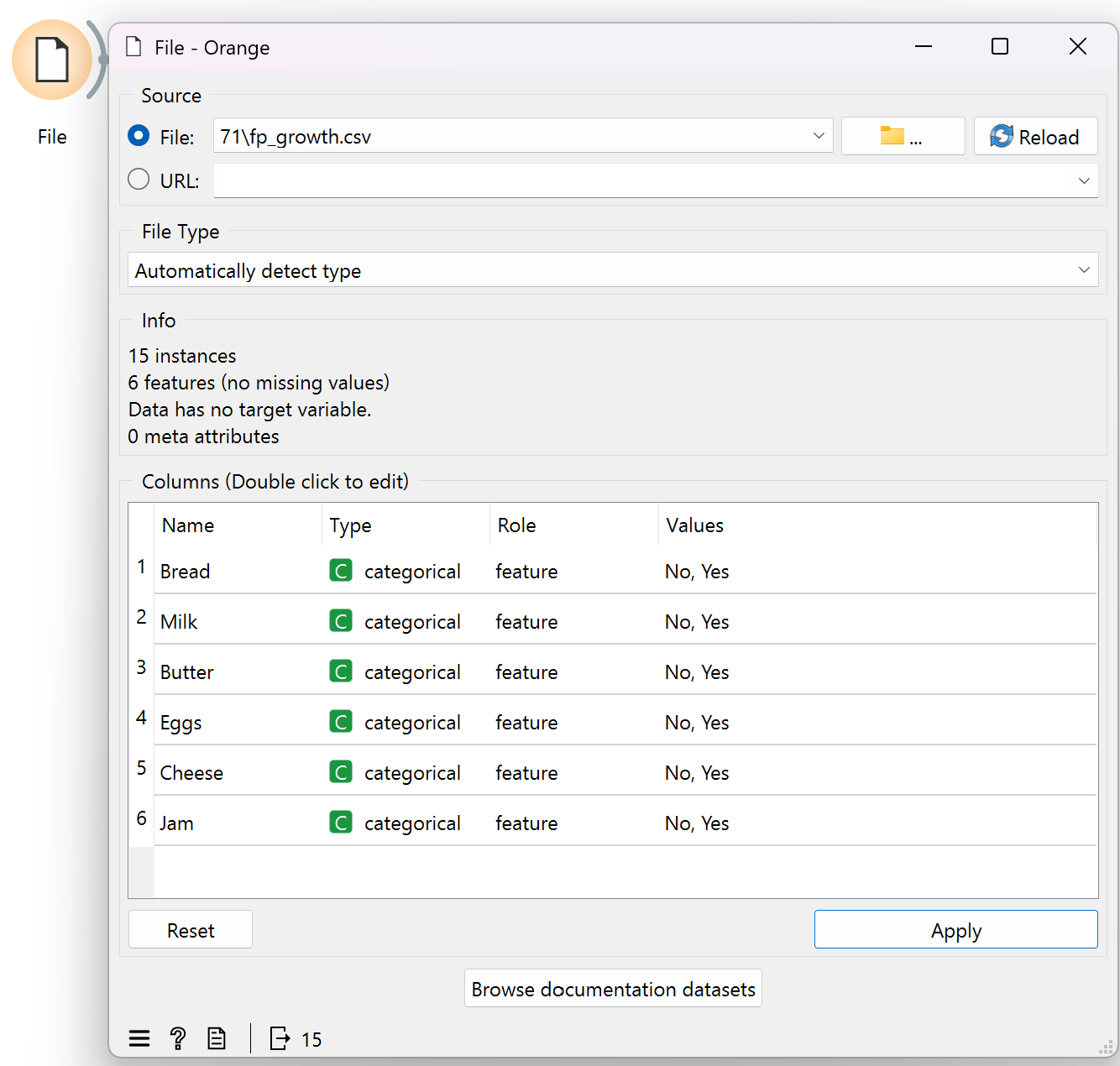
 



**Step 3: Load Data into Orange**

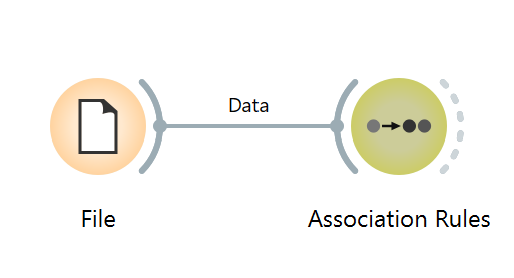
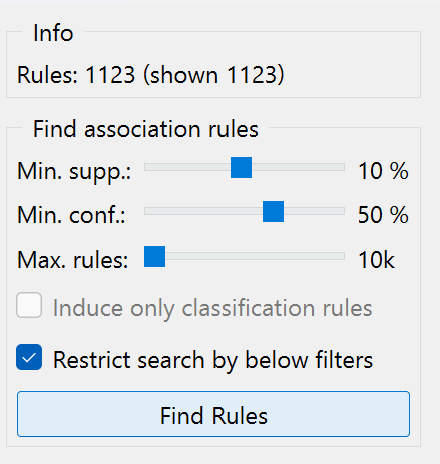
* Open Orange, Click New
* Drag and drop the “File” widget onto the canvas
* Double-click it and browse your Yes/No .csv dataset
* Click Apply (if it’s not clickable, check that the dataset has valid Yes/No entries)



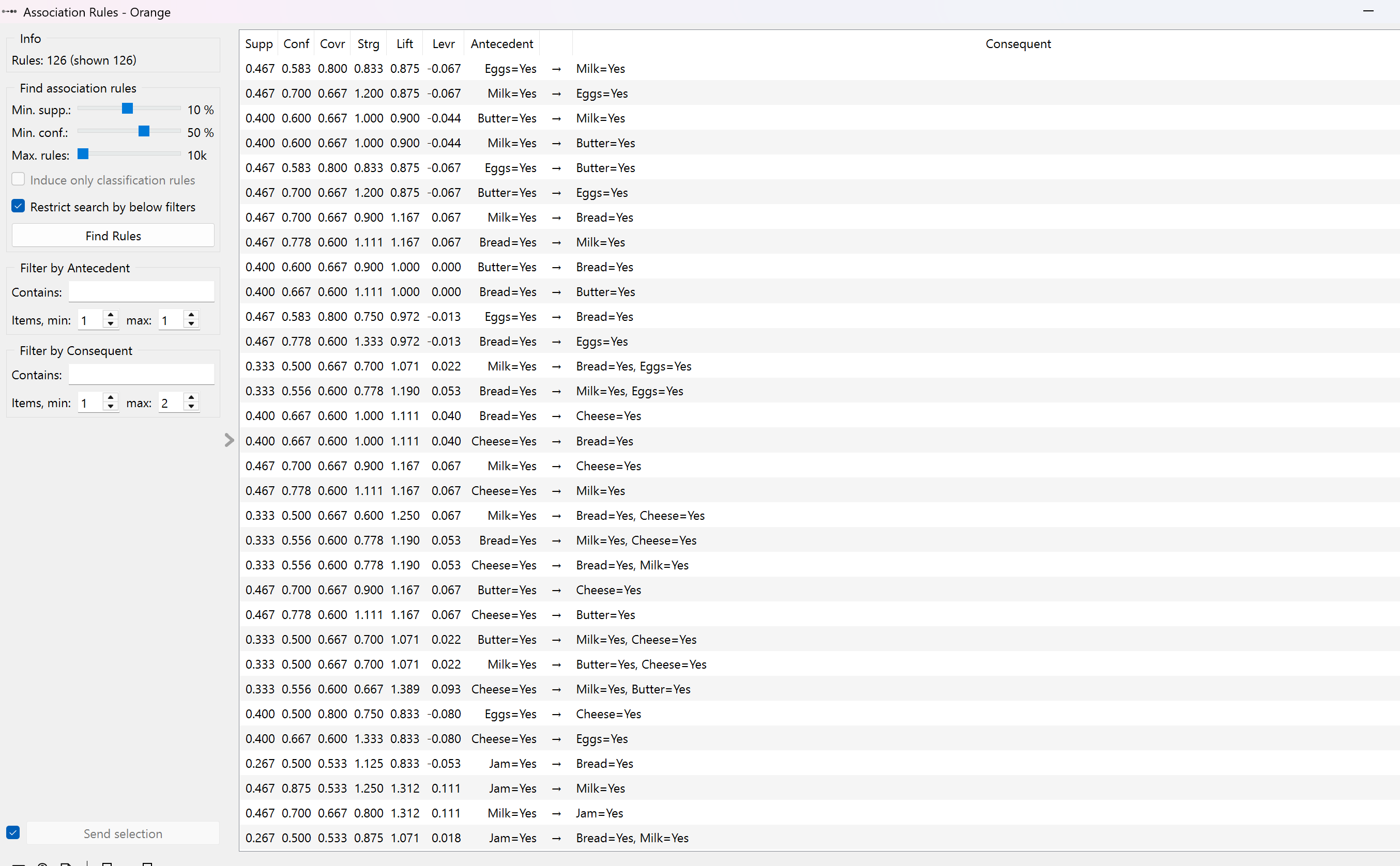


**Step 4: Add and Connect the Associate Widget**

* From the left sidebar, scroll down to the Associate section (now available after the add-on)
* Drag the Associate rules widget onto the canvas
* Connect File → Associate rules
* Double-click Associate to open it
* Set Min. supp. to 10%
* Set Min. conf. to 50%
* click Find Rules

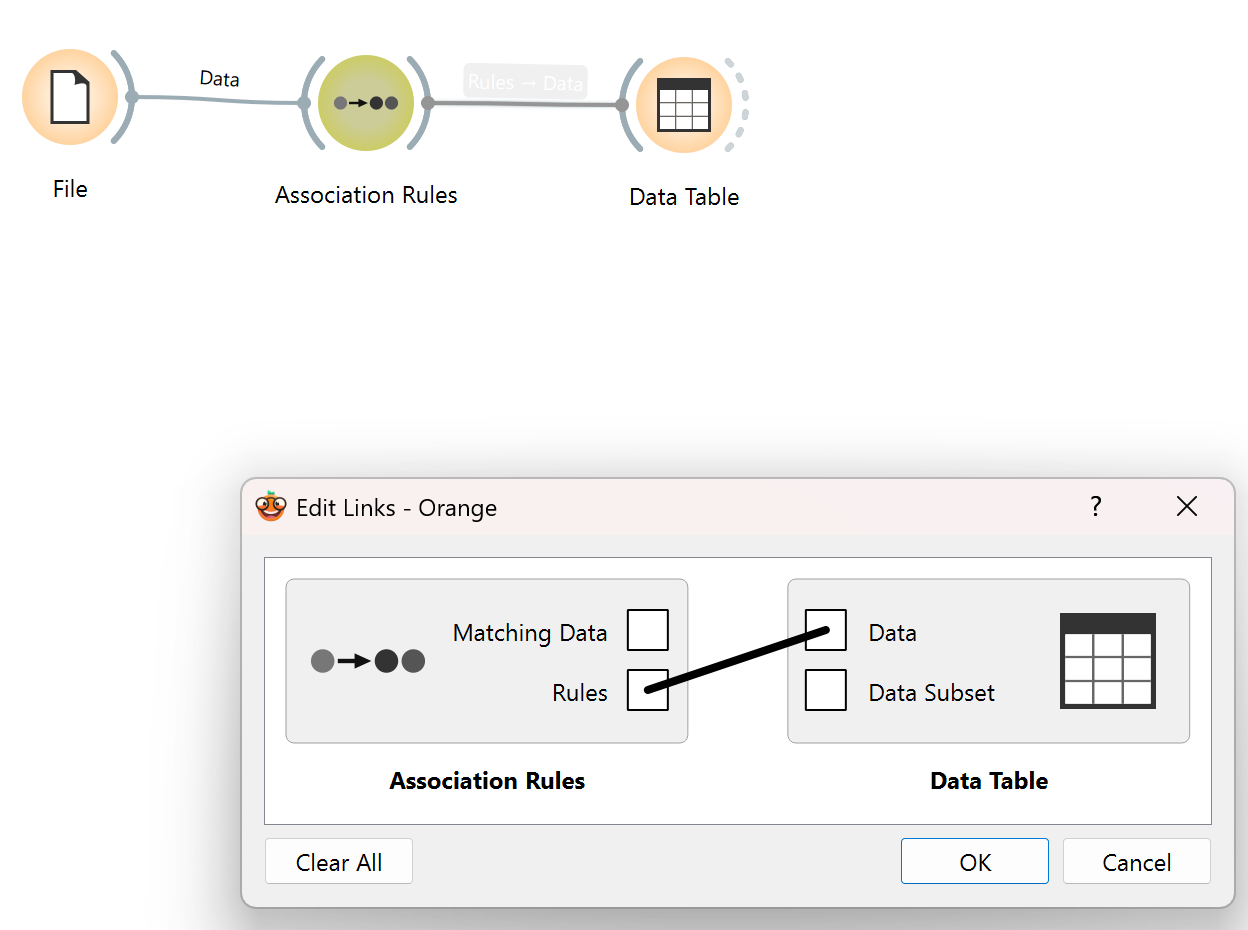
**This is the Expected Output**

****

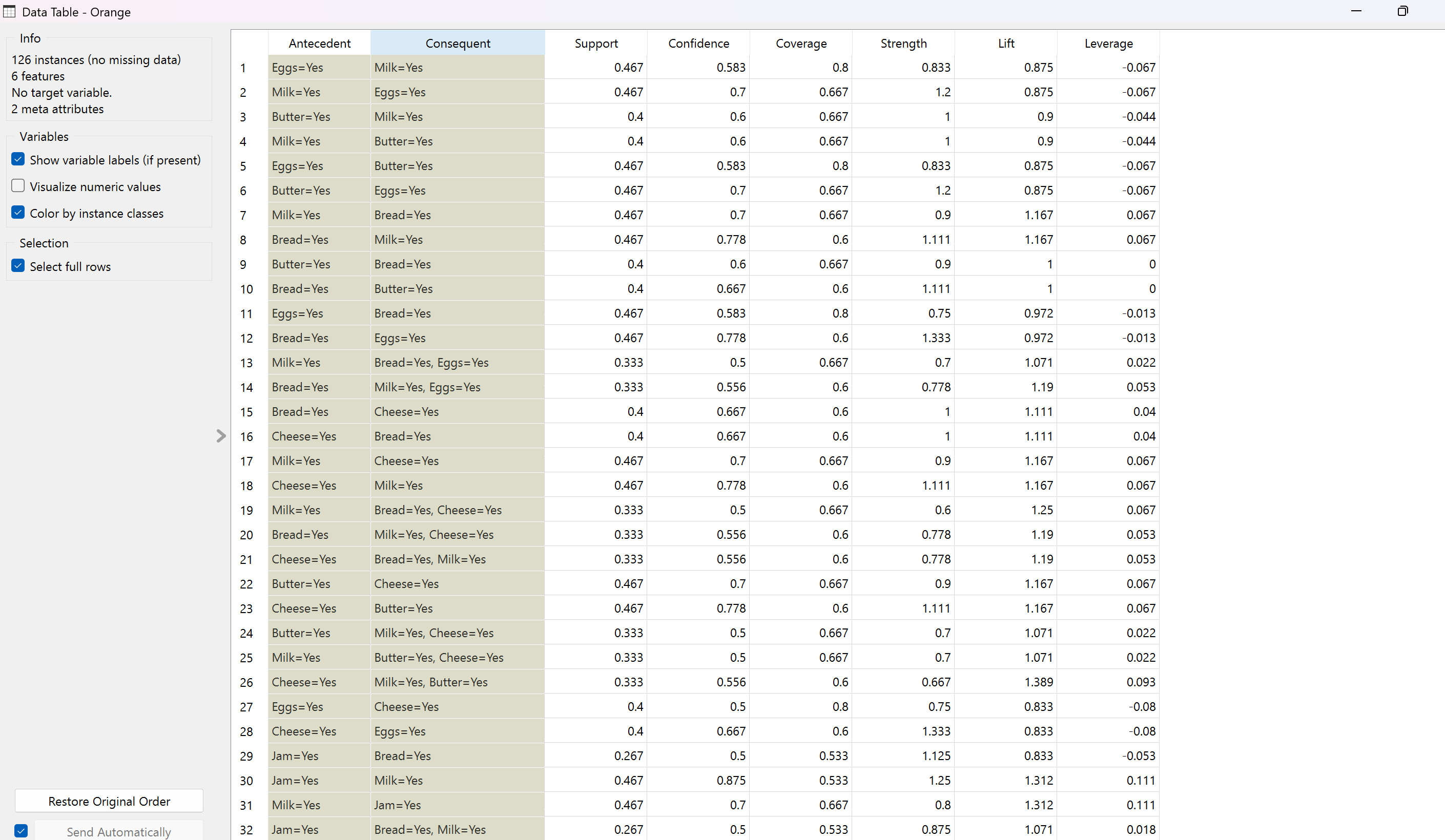
* **Support:**The percentage of transactions that contain a particular itemset. Higher support means the itemset is more frequent.
* **Confidence:**The likelihood that item Y is also purchased when item X is bought. It measures the reliability of the rule.
* **Coverage:**Also called antecedent support — it shows how often the if-part (LHS) of the rule appears in the data.
* **Strength:**Indicates how strong or significant the rule is, often reflected by confidence or other metrics.
* **Lift:**Measures how much more likely item Y is bought when item X is bought, compared to them being bought independently. A lift > 1 indicates a strong association**.**
* **Leverage:**Shows the difference between the observed frequency of X and Y appearing together and the expected frequency if they were independent**.**

**Step 5 : Click on a Data Table to analyze it in detail.**

* Add the “Data Table” widget to display selected rules.
* Choose: Rules ➝ Data
* Rules ➝ Data lets you **view the list of rules** with support, confidence, lift, etc.
* Click OK



* Click on Data Table and the expected output is as follows



**NOTE :** In this experiment, we used Orange, a user-friendly visual data mining tool, to implement the FP-Growth algorithm for mining frequent patterns from a retail dataset. Orange simplifies complex data mining tasks using a drag-and-drop interface with interactive widgets.

We used the following widgets:

* File: To load the retail dataset (formatted with Yes/No or binary values).
* Associate: Applied the FP-Growth algorithm to extract frequent itemsets and generate strong association rules.
* Data Table: Displayed the resulting frequent patterns or rules for easy interpretation.

The workflow was:  
 File → Associate → Data Table

In the Associate widget, we selected "Association Rules" to view the rules generated by FP-Growth based on the given support and confidence thresholds. These rules help identify strong relationships between items in the dataset.