

**COLLEGE OF COMPUTER STUDIES**  
**ITE 401 – Data Mining and Warehousing**

|                                                                       |                    |                                       |
|-----------------------------------------------------------------------|--------------------|---------------------------------------|
| Name: Villareal, Russell Stephen C.                                   | Date: May 03, 2023 |                                       |
| Section: IT32S3                                                       | Program: BSIT      | Instructor: Ms. Paula Jean C. Mendoza |
| <b>Assessment Task: Finals Laboratory Activity 1 Association Rule</b> |                    |                                       |

**Instructions:**

- Using the data mining tool Rapidminer, construct the Association model on your data by creating your own Process Model.
- Analyze the problem provided.
- Make use of appropriate data that are properly formatted.
- Provide screenshots of your work. Include screenshots of your step-by-step process in generating your model.
- Make sure that the images are clear and readable with your name.

**Mining Frequent Patterns, Associations, and Correlations**  
**Performing Market Basket Analysis**

1. Use the following data about customer purchases. See the link provided for the excel file.
2. Identify the buying patterns of the customers utilizing the Association Rule.
3. Utilize or Rapidminer Application to produce the association rule.

Here are the steps needed to be completed.

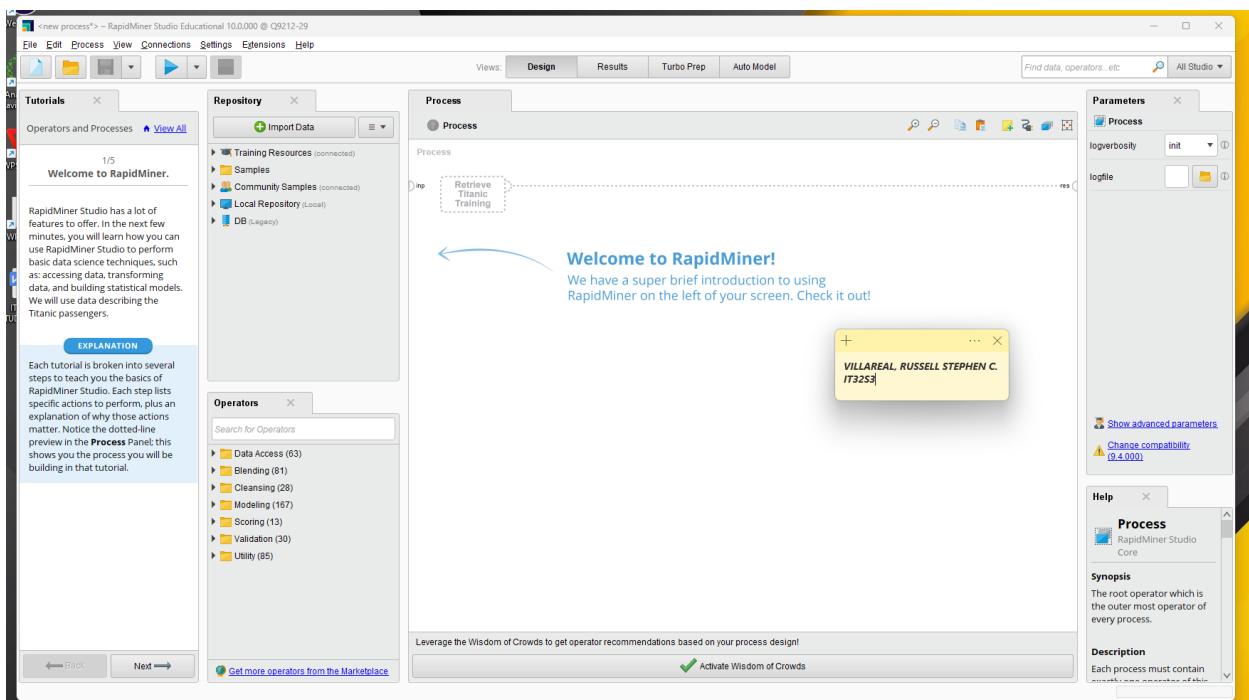
- Import the excel file from the link provided, or you may convert the excel file to its latest version if it is not supported by the Rapidminer
- Choose the correct sheet (Numerical Data) and make note of the column header labels in the file (customerId, itemId, itemCount)
- Load the Dataset in your data mining tool, use appropriate techniques for association and analyze the results of your generated association rule.
- Please note that the original data set contains 76 types of items and 89 customers with different combinations of items purchased (details can be found in the EXCEL file). However, the itemCount is generated using a random generator function of (=RANDBETWEEN (1,10)) from Excel.
- Click the FP Growth operator and change the value in minimum support value from 0.95 to 0.0001; otherwise, there will be no association rules produced. Other values remain unchanged.
- The value of minimum confidence is 0.8
- Your results will depend on the minimum confidence and support chosen in the FP-Growth, and the Create Association rules operators respectively. For this example, a minimum support of 0.0001 has been used, and a minimum confidence of 0.8.

## Questions:

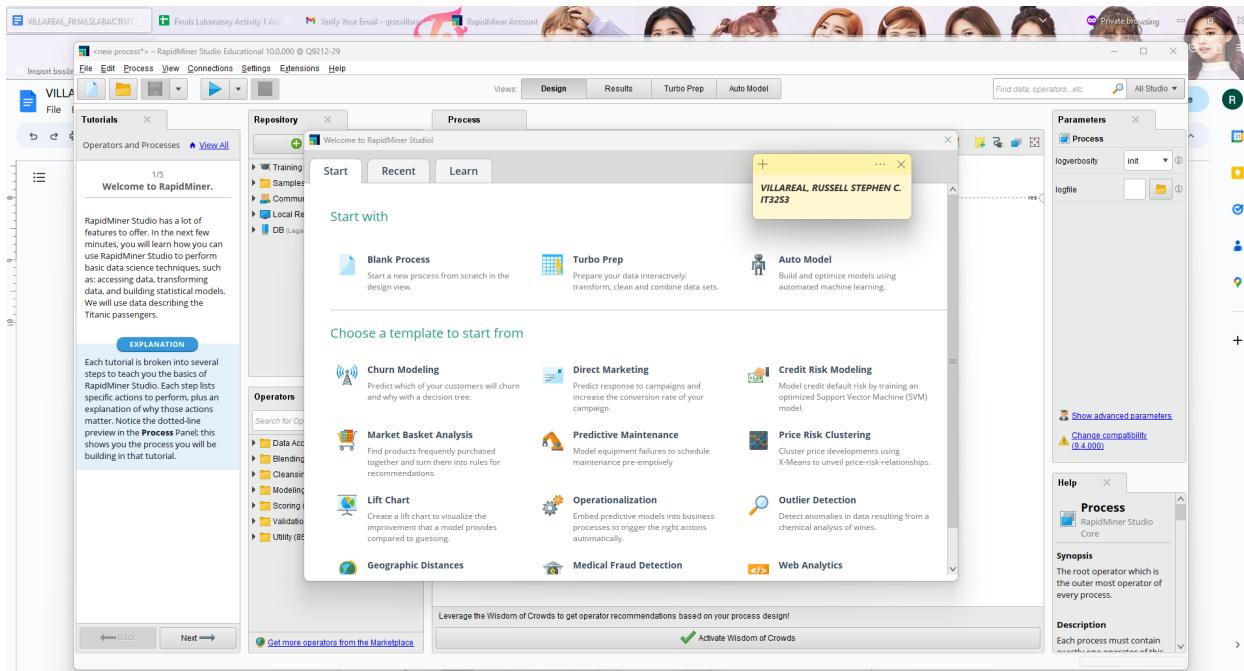
After performing your laboratory activity, answer the following questions to the best of your ability.

1. Write your conclusion based on the process of creating association rules using the RapidMiner.
2. Explain the result of your created process model. Discuss the support and the confidence obtained from the frequent item sets.

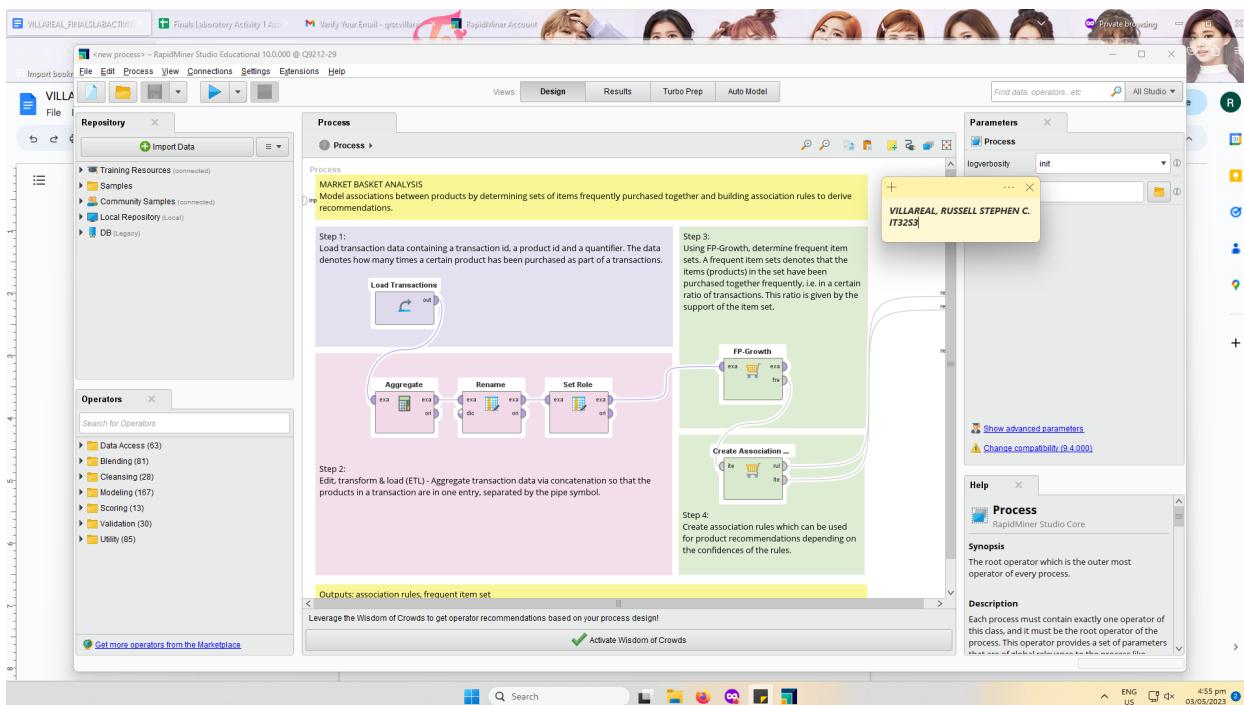
## Step by Step Process:



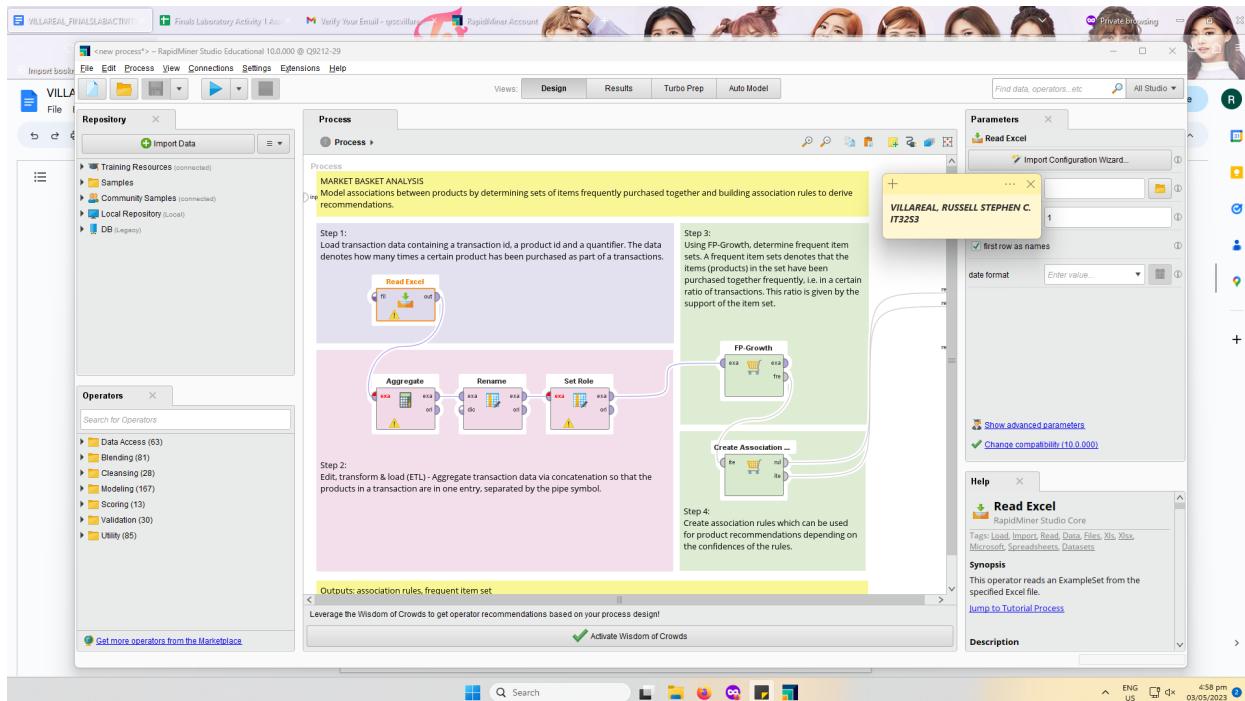
**Description:** Use the RapidMiner program to carry out Market Basket Analysis. Start by launching the program and going to the File tab. Click Open Process after that.



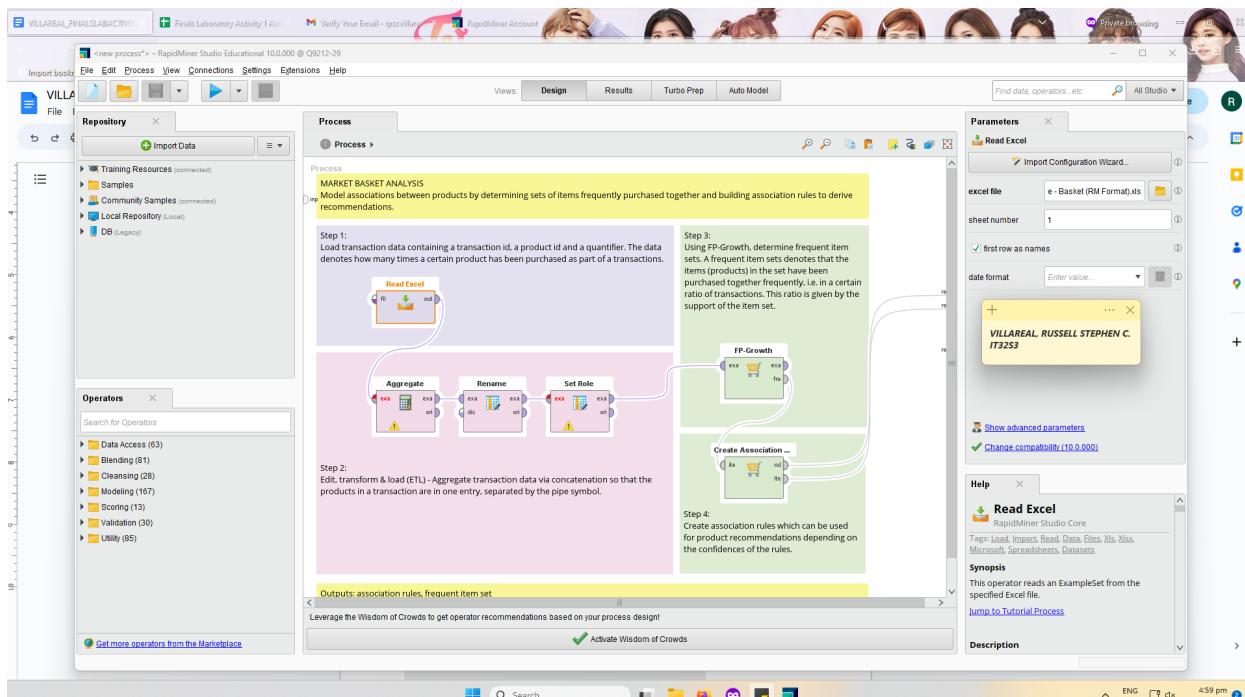
**Description:** The list of templates is displayed in a new window that opens. Analyze the market basket.



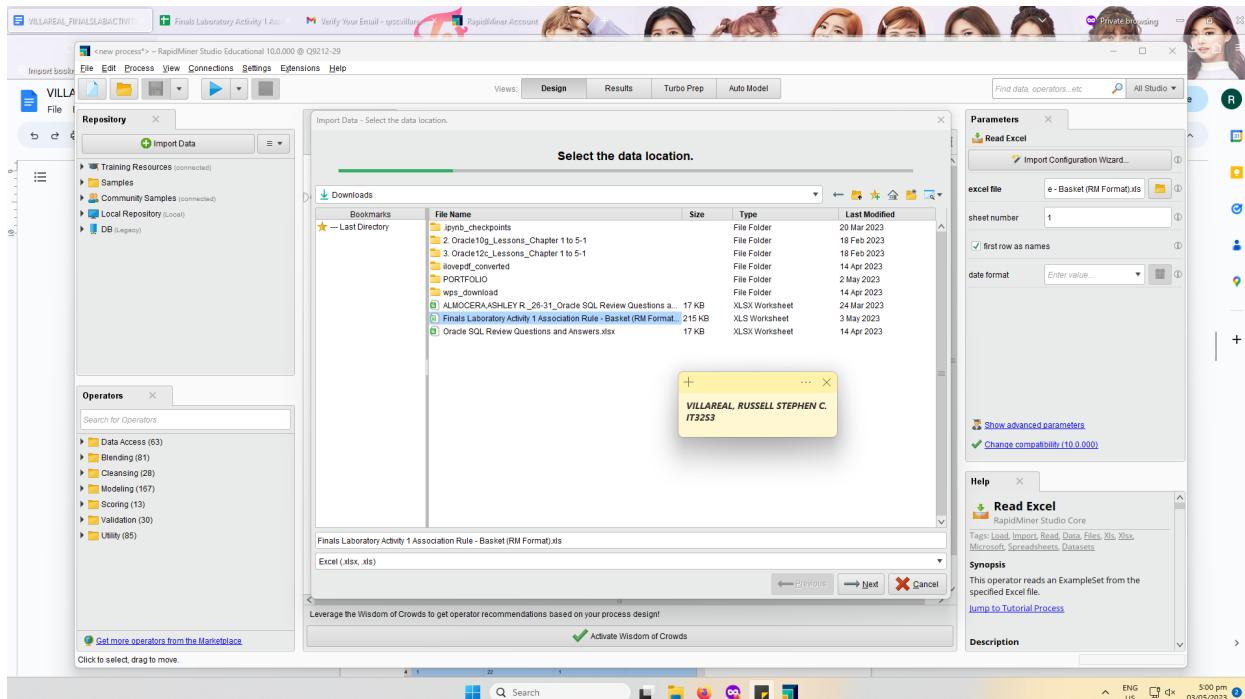
**Description:** After selecting the aforementioned template, the full Market Basket Analysis procedure emerges.



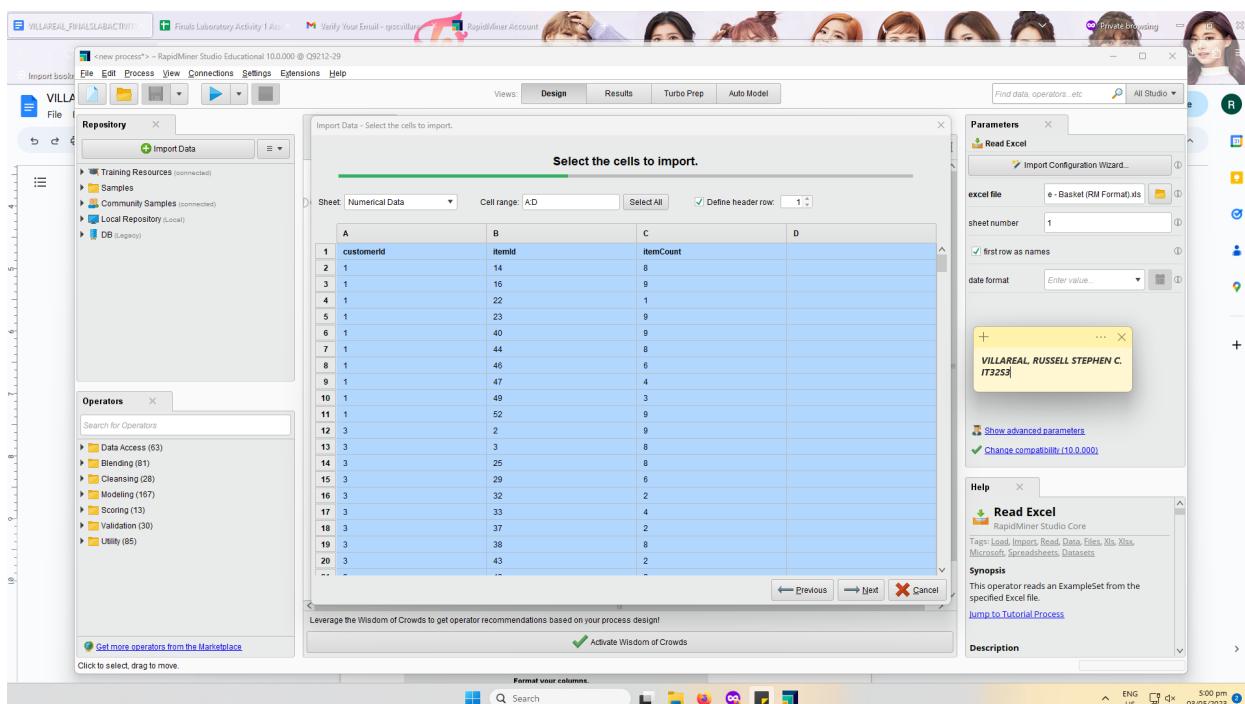
**Description:** By right-clicking the first process and selecting Replace Operator, you can change it from Load Transaction to Read Excel. Click the Data Access folder from there, followed by the Files and Read folders. Select Read Excel from the list of available Read processes.



**Description:** To access the.xls file containing the dataset for this task, use the Read Excel process.



**Description:** An additional window will open. Choose the Basket (RM Format) file.click Next after xls.



**Description:** then choose the Numerical Data sheet, which is the required dataset. Just click Next after that.

**Format your columns.**

customerid itemid itemCount

|    |    |   |
|----|----|---|
| 1  | 14 | 8 |
| 2  | 16 | 9 |
| 3  | 22 | 1 |
| 4  | 23 | 9 |
| 5  | 40 | 9 |
| 6  | 44 | 8 |
| 7  | 46 | 6 |
| 8  | 47 | 4 |
| 9  | 49 | 3 |
| 10 | 52 | 9 |
| 11 | 2  | 9 |
| 12 | 3  | 8 |
| 13 | 25 | 8 |
| 14 | 29 | 6 |
| 15 | 32 | 2 |
| 16 | 33 | 4 |
| 17 | 37 | 2 |
| 18 | 38 | 8 |
| 19 | 43 | ? |

Leverage the Wisdom of Crowds to get operator recommendations based on your process design!

Activate Wisdom of Crowds

no problems.

Import Data - Format your columns.

Import Data - Format your columns.

Parameters

excel file e-Basket (RM Format).xls

sheet number 1

first row as names

date format Enter value...

Help

Read Excel

Description

**Description:** Since there are no missing values and the values are in numerical data with an integer type, you can simply click Finish.

Process

MARKET BASKET ANALYSIS

Edit Parameter List: aggregation attributes

Step 1: Load transaction data. Denotes how many items are in a transaction.

aggregation attribute aggregation functions

ItemCount count

VILLAREAL, RUSSELL STEPHEN C. IT3253

Step 2: Edit, transform & load products in a transaction.

Add Entry Remove Entry Apply Cancel

Outputs: association rules, frequent item set

Leverage the Wisdom of Crowds to get operator recommendations based on your process design!

Activate Wisdom of Crowds

Parameters

use default aggregation

aggregation attributes Edit List (1)...

group by attributes Select Attributes...

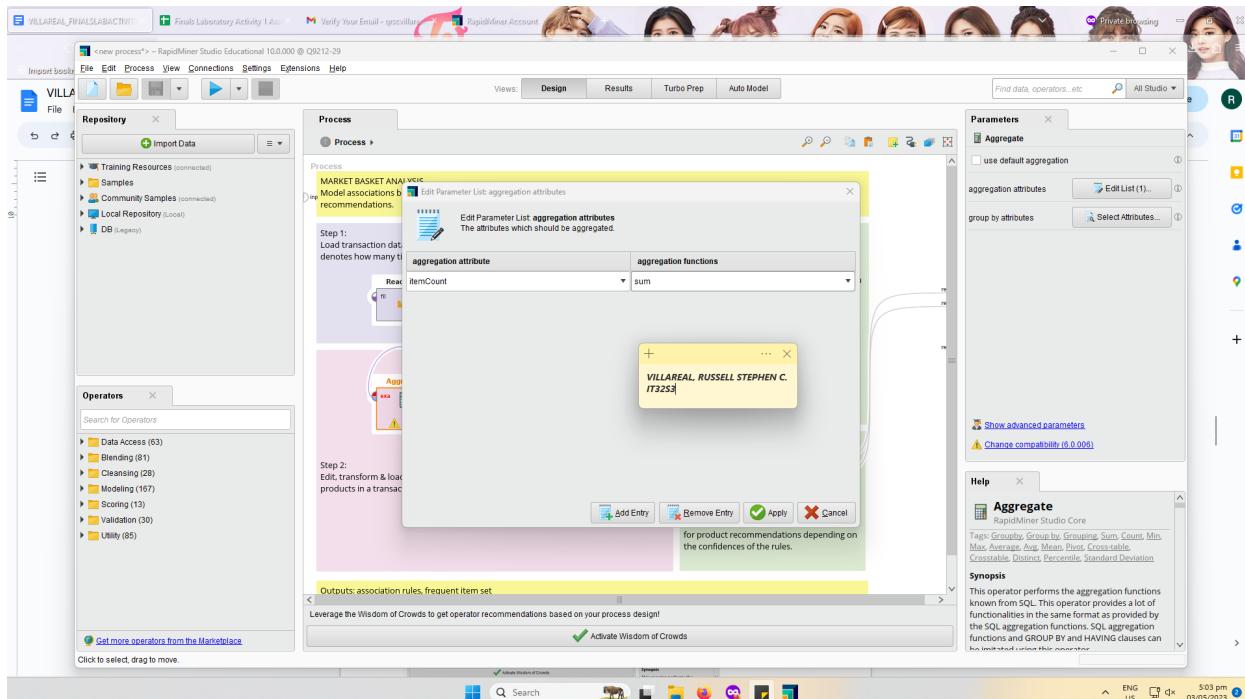
Help

Aggregate

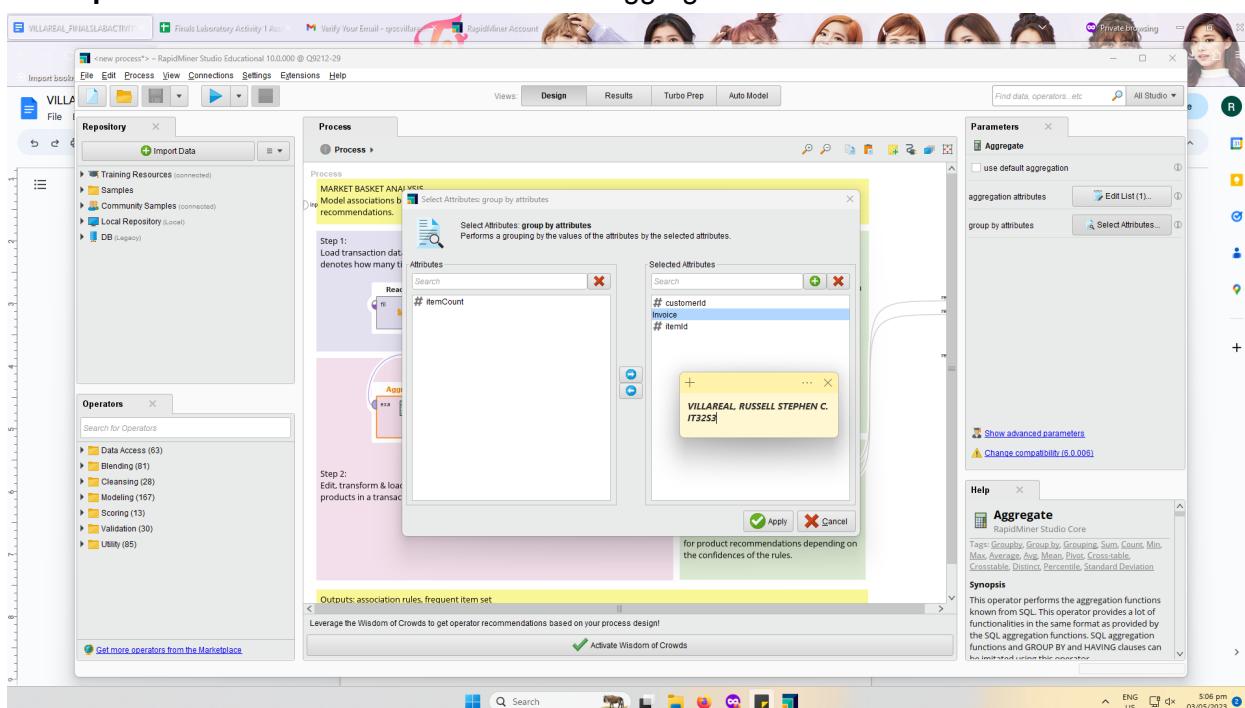
Synopsis

This operator performs the aggregation functions known from SQL. This operator provides a lot of functionalities in the same format as provided by the SQL aggregation functions. SQL aggregation functions and GROUP BY and HAVING clauses can be introduced using their operators.

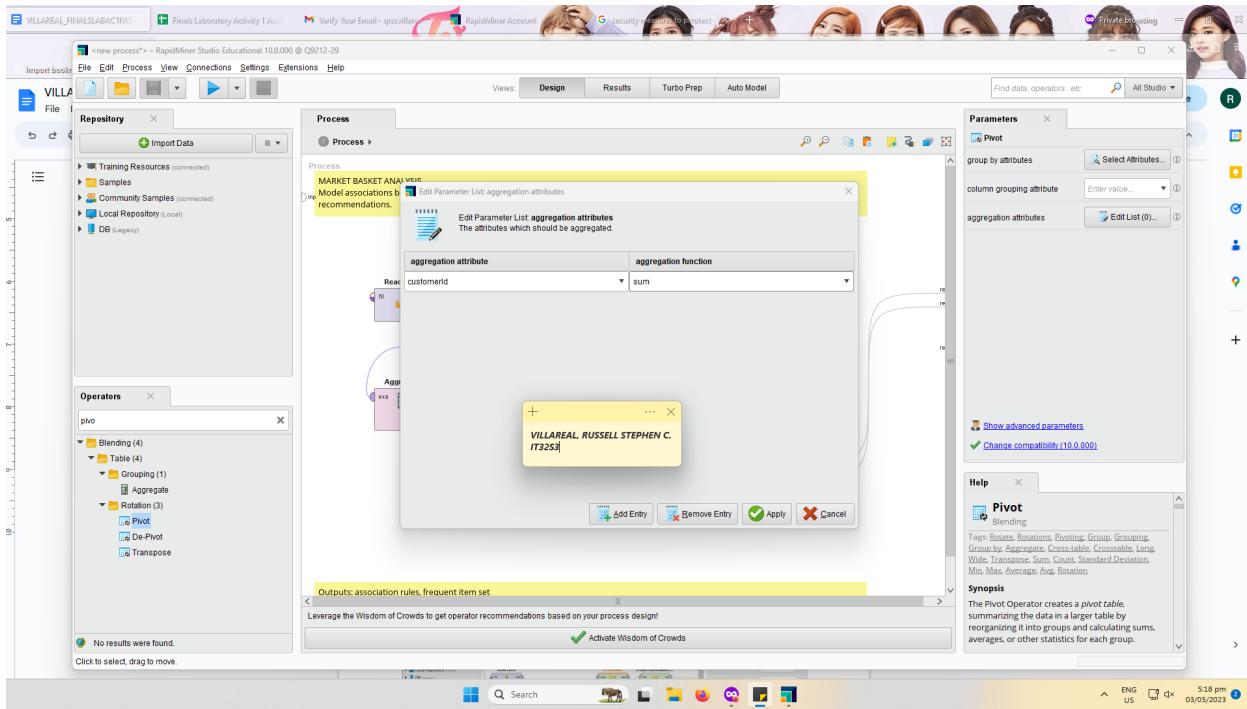
**Description:** Click the Aggregate process twice. The aggregation attribute should then be changed to itemCount.



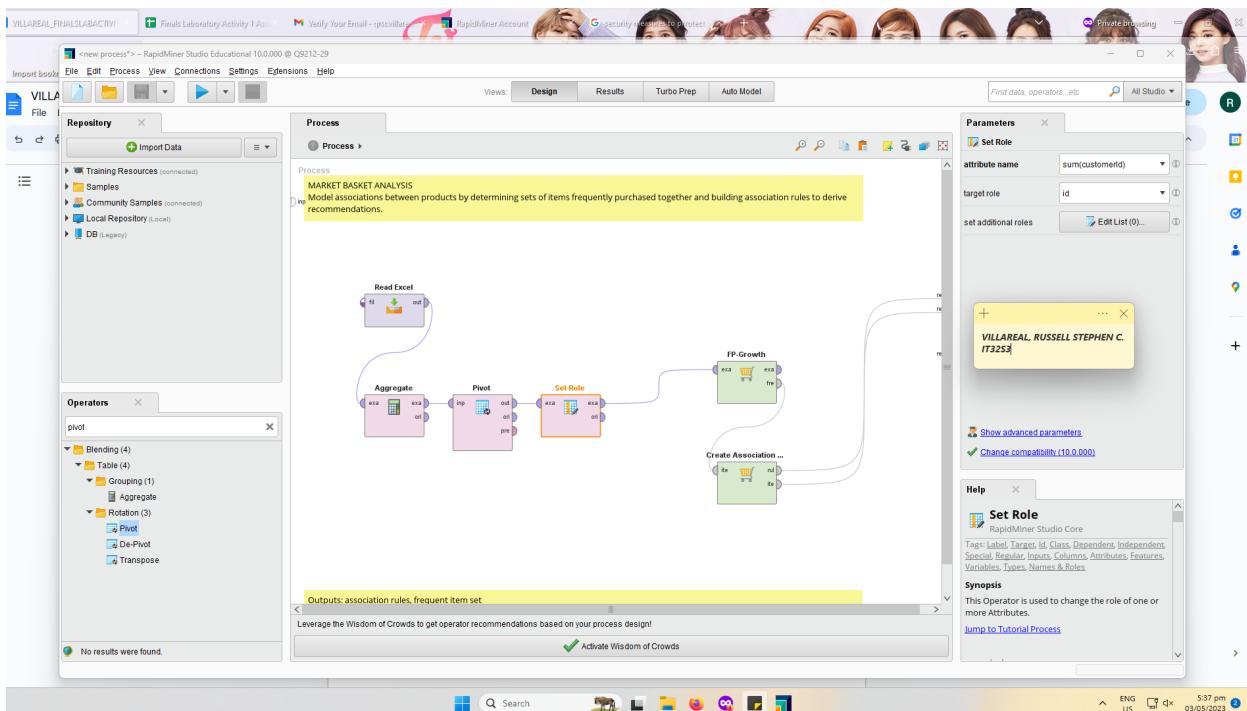
**Description:** Sum should also be used as the aggregation function.



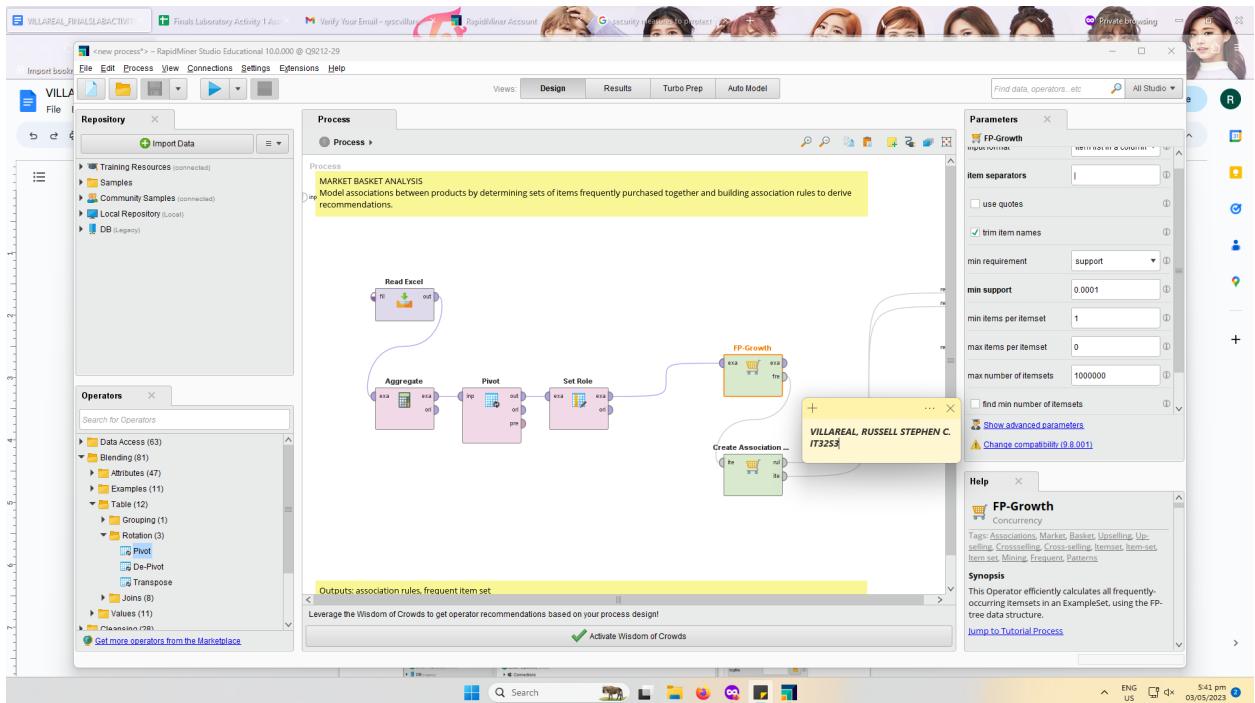
**Description:** then select the Select button next to the group of attributes under the Parameters of Aggregate procedure. As the Attribute, pick itemCount.



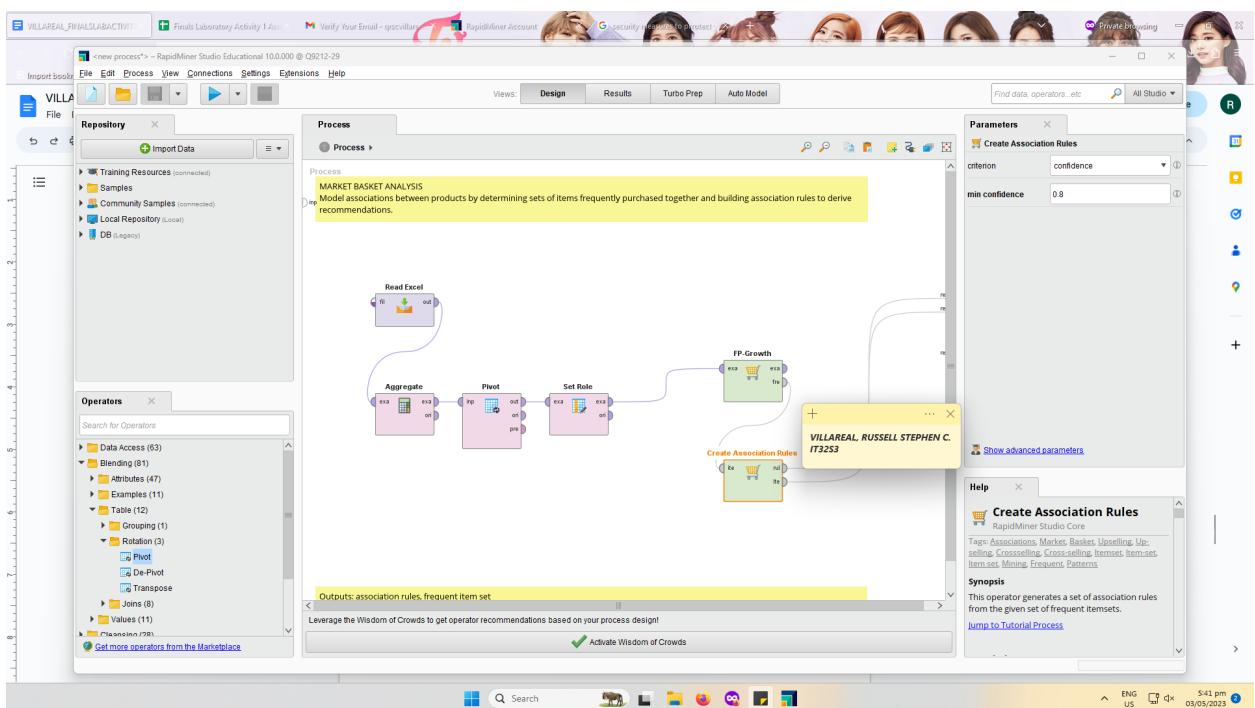
**Description:** The customerId can then be chosen from the dropdown list by double-clicking the Pivot operator.



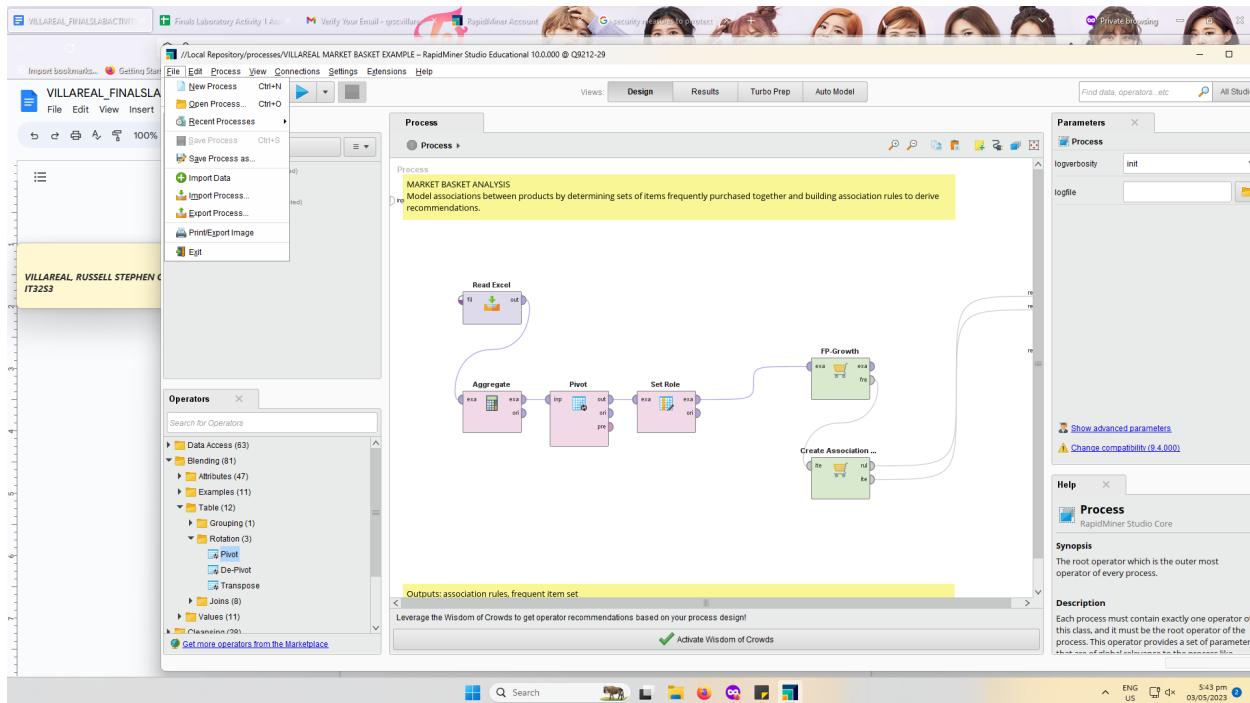
**Description:** Choose the customerId and set the target role of the Set Role operator to id.



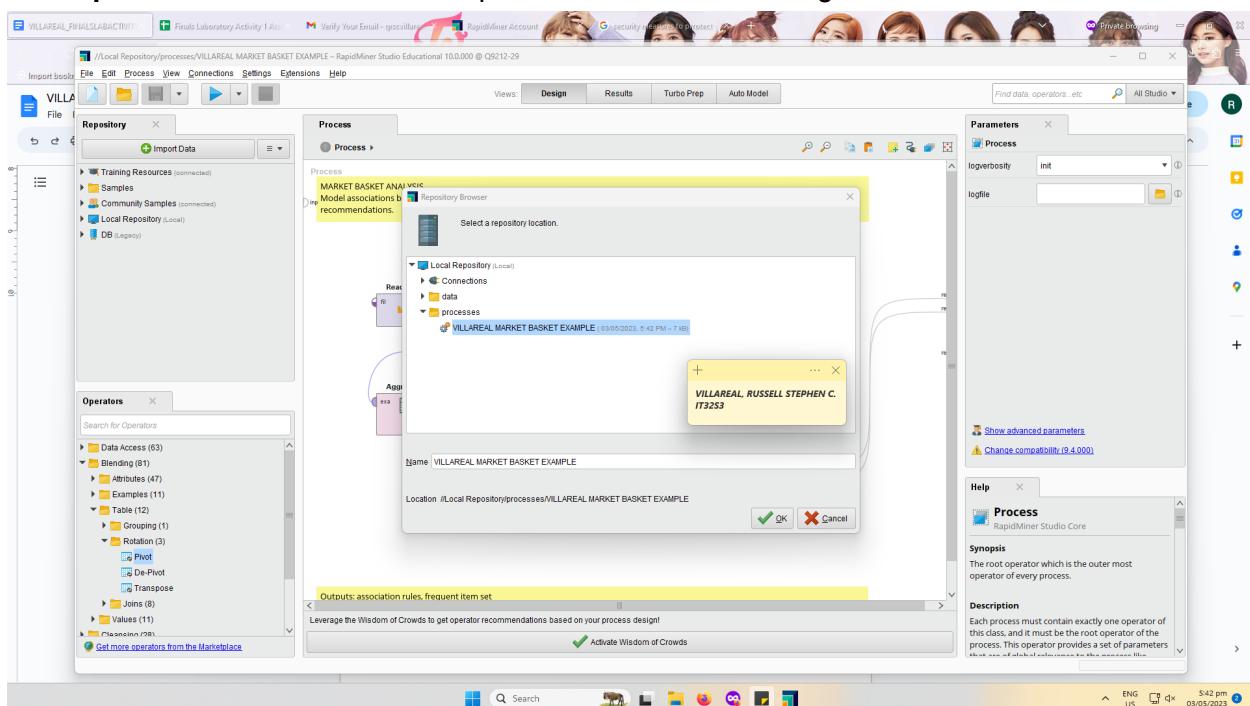
**Description:** The FP-Growth operator was double-clicked, and the min support was changed to 0.0001.



**Description:** Change the minimum confidence for the Create Association Rules operator to 0.8.



**Description:** Make sure to save the process first before running it.



**Description:** In the processes folder of the Local Repository, save the procedure as a Market Basket example.

The screenshot shows the RapidMiner Studio interface with a table titled "AssociationRules (Create Association Rules)". The table has columns: No., Premises, Conclusion, Support, Confidence, LaPlace, and Gain. The data consists of 23 rows of association rules. A sidebar on the left shows navigation options like Data, Graph, Description, and Annotations. A tooltip in the bottom right corner displays the details for rule number 23: "VILLAREAL, RUSSELL STEPHEN C. IT3253".

| No. | Premises | Conclusion | Support | Confidence | LaPlace | Gain   |
|-----|----------|------------|---------|------------|---------|--------|
| 1   | 85       | 7          | 0.003   | 1          | 1       | -0.003 |
| 2   | 77       | 1          | 0.003   | 1          | 1       | -0.003 |
| 3   | 75       | 4          | 0.003   | 1          | 1       | -0.003 |
| 4   | 87       | 4          | 0.003   | 1          | 1       | -0.003 |
| 5   | 64       | 9          | 0.003   | 1          | 1       | -0.003 |
| 6   | 75       | 32         | 0.003   | 1          | 1       | -0.003 |
| 7   | 87       | 51         | 0.003   | 1          | 1       | -0.003 |
| 8   | 84       | 14         | 0.003   | 1          | 1       | -0.003 |
| 9   | 77       | 45         | 0.003   | 1          | 1       | -0.003 |
| 10  | 85       | 73         | 0.003   | 1          | 1       | -0.003 |
| 11  | 7,21     | 8          | 0.003   | 1          | 1       | -0.003 |
| 12  | 8,27     | 7          | 0.003   | 1          | 1       | -0.003 |
| 13  | 8,76     | 7          | 0.003   | 1          | 1       | -0.003 |
| 14  | 3,17     | 7          | 0.003   | 1          | 1       | -0.003 |
| 15  | 3,45     | 7          | 0.003   | 1          | 1       | -0.003 |
| 16  | 10,53    | 7          | 0.003   | 1          | 1       | -0.003 |
| 17  | 7,71     | 10         | 0.003   | 1          | 1       | -0.003 |
| 18  | 10,71    | 7          | 0.003   | 1          | 1       | -0.003 |
| 19  | 7,2      | 20         | 0.003   | 1          | 1       | -0.003 |
| 20  | 7,20     | 2          | 0.003   | 1          | 1       | -0.003 |
| 21  | 4,35     | 7          | 0.003   | 1          | 1       | -0.003 |
| 22  | 7,79     | 4          | 0.003   | 1          | 1       | -0.003 |
| 23  | 4,79     | 7          | 0.003   | 1          | 1       | -0.003 |

**Description:** This is the table-formatted result of the association rule.

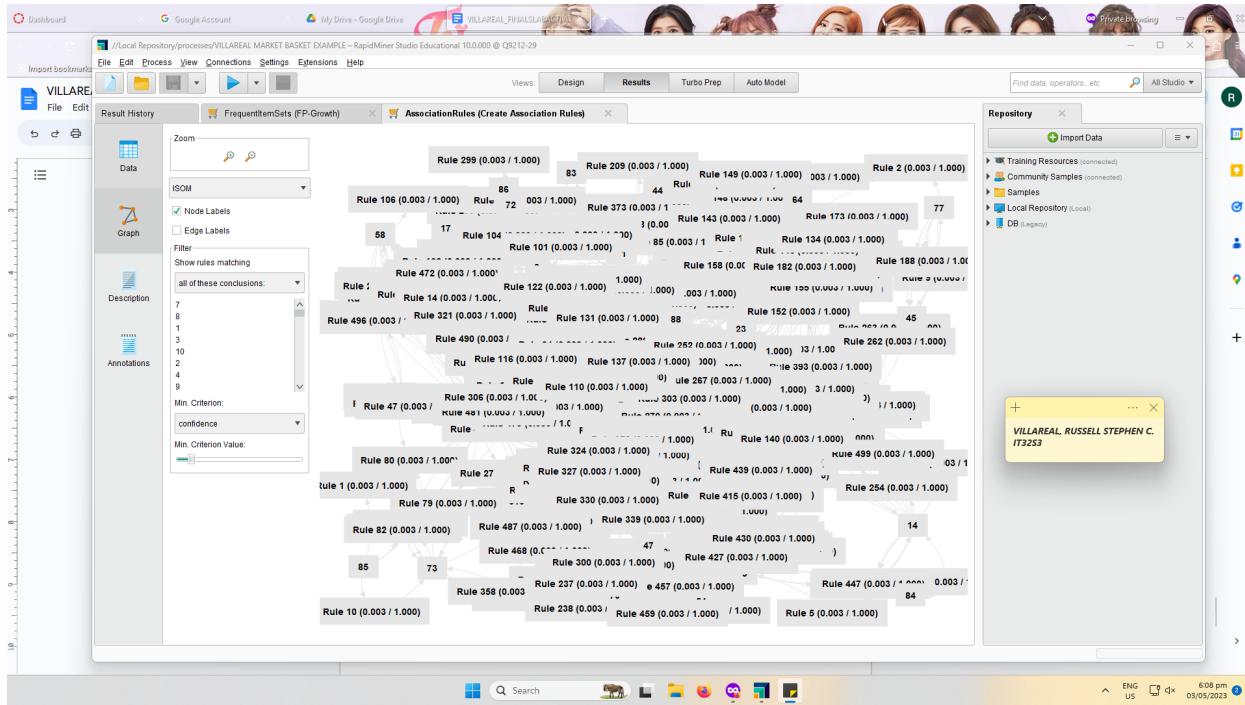
The screenshot shows the RapidMiner Studio interface with a list of association rules under the "AssociationRules" tab. The rules are listed in a vertical format, each showing a premise and a conclusion with a confidence value of 1.000. A tooltip in the bottom right corner displays the details for rule number 23: "VILLAREAL, RUSSELL STEPHEN C. IT3253".

```

Association Rules
[85] --> [7] (confidence: 1.000)
[77] --> [1] (confidence: 1.000)
[77] --> [13] (confidence: 1.000)
[87] --> [14] (confidence: 1.000)
[84] --> [9] (confidence: 1.000)
[75] --> [32] (confidence: 1.000)
[87] --> [51] (confidence: 1.000)
[84] --> [14] (confidence: 1.000)
[84] --> [13] (confidence: 1.000)
[85] --> [73] (confidence: 1.000)
[7, 21] --> [8] (confidence: 1.000)
[8, 27] --> [7] (confidence: 1.000)
[8, 76] --> [7] (confidence: 1.000)
[3, 17] --> [7] (confidence: 1.000)
[3, 45] --> [7] (confidence: 1.000)
[10, 71] --> [1] (confidence: 1.000)
[7, 71] --> [10] (confidence: 1.000)
[10, 71] --> [7] (confidence: 1.000)
[7, 2] --> [20] (confidence: 1.000)
[7, 20] --> [2] (confidence: 1.000)
[4, 35] --> [7] (confidence: 1.000)
[4, 35] --> [1] (confidence: 1.000)
[4, 79] --> [7] (confidence: 1.000)
[7, 9] --> [18] (confidence: 1.000)
[7, 6] --> [63] (confidence: 1.000)
[6, 63] --> [7] (confidence: 1.000)
[7, 6] --> [73] (confidence: 1.000)
[3, 31] --> [2] (confidence: 1.000)
[7, 41] --> [32] (confidence: 1.000)
[32, 81] --> [7] (confidence: 1.000)
[7, 44] --> [83] (confidence: 1.000)
[7, 83] --> [44] (confidence: 1.000)
[44, 83] --> [7] (confidence: 1.000)
[12, 71] --> [7] (confidence: 1.000)
[1, 33] --> [66] (confidence: 1.000)
[7, 66] --> [33] (confidence: 1.000)
[33, 66] --> [7] (confidence: 1.000)
[7, 23] --> [28] (confidence: 1.000)
[29, 23] --> [7] (confidence: 1.000)
[75, 24] --> [71] (confidence: 1.000)

```

**Description:** A textual description is the end result of the association rule.



**Description:** This is the graph that the association rule produced. This enables us to see the data visually.

1. Write your conclusion based on the process of creating association rules using the RapidMiner
- 
- A method for locating connections or links between variables or items in sizable datasets is called association rule mining. Popular data mining software called RapidMiner makes it simple to create association rules. To get insights into the data, the procedure entails importing and cleaning the dataset, building itemsets, and generating association rules using statistical measures. The rules are then filtered based on certain criteria before being visualized and interpreted. A user-friendly interface and potent algorithms are provided by RapidMiner to analyze massive datasets and find hidden patterns and relationships.

2. Explain the result of your created process model. Discuss the support and the confidence obtained from the frequent item sets.
- 
- Market-basket analysis is a practical method for comprehending frequently bought goods. It entails creating many-to-many relationships between goods and baskets or transactions. Transactions are represented as instances in an ExampleSet, while items are represented using binary characteristics. A minimal support threshold is used to identify frequent-itemsets, or groups of products that are commonly bought or sold together. The amount of support for an item is determined by dividing the number of times it appears in the ExampleSet by the total number of examples. A great example of common transactions data is the "Samples/data/Transactions" data set in RapidMiner.

*Honor Pledge:*

***"I affirm that I have not given or received any unauthorized help on this assignment, and  
that this work is my own."***