

Business Intelligence Technologies: Practical 2: Association Rule Mining

Weighting: 4% Due date:

Produce association rules for the storedata.csv dataset. Document and explain your data preparation and modelling processes (including choice and rationale for parameters used). Evaluate the results, highlighting and explaining any significant findings.

Assessment Criteria:

Correct and fully documented processes: 2 marks

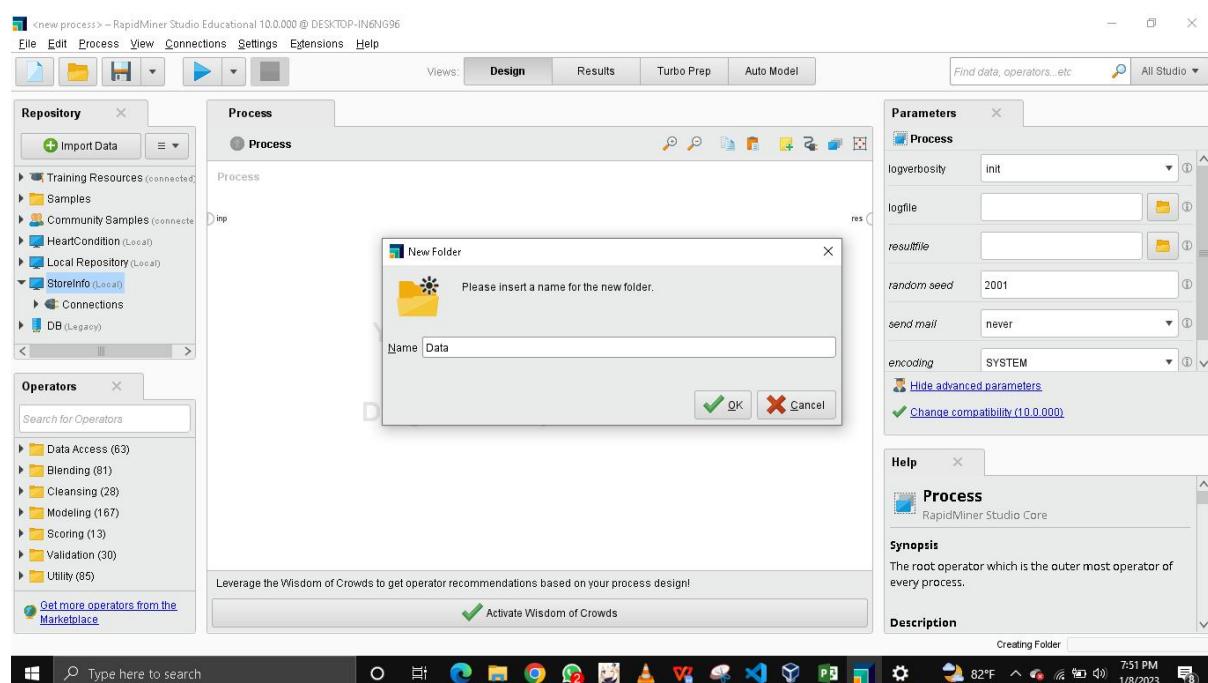
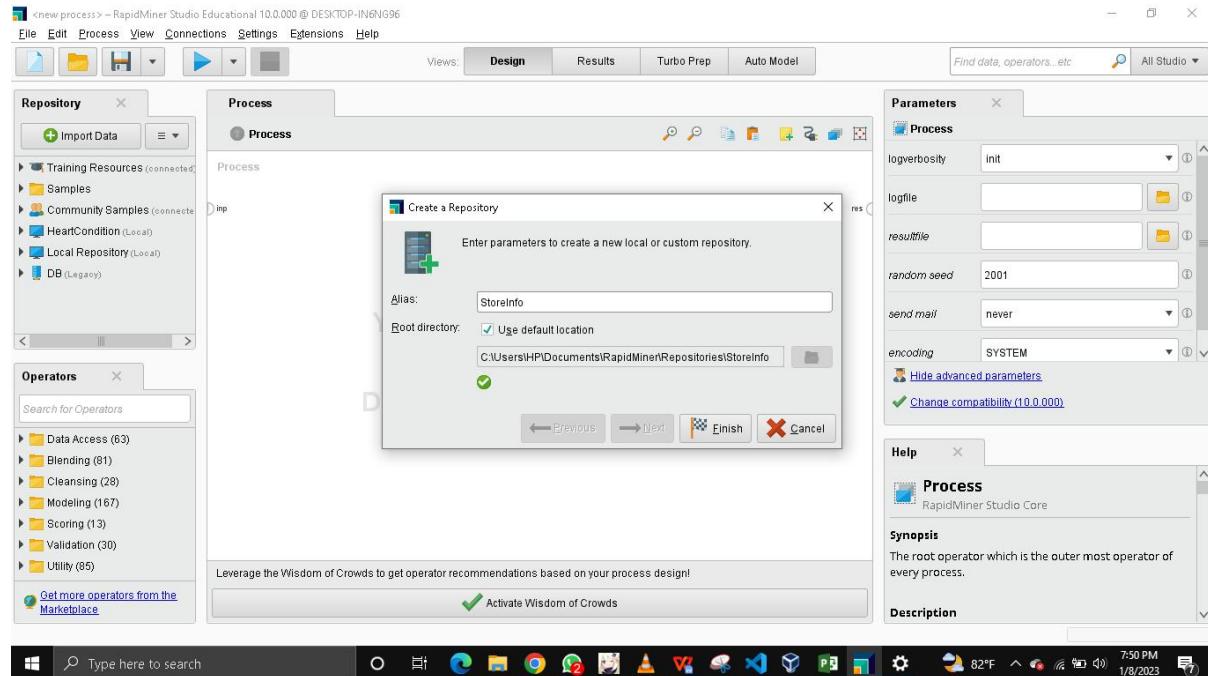
Explanation demonstrating sound understanding of principles of association rule mining, and evaluation of results: 2 marks

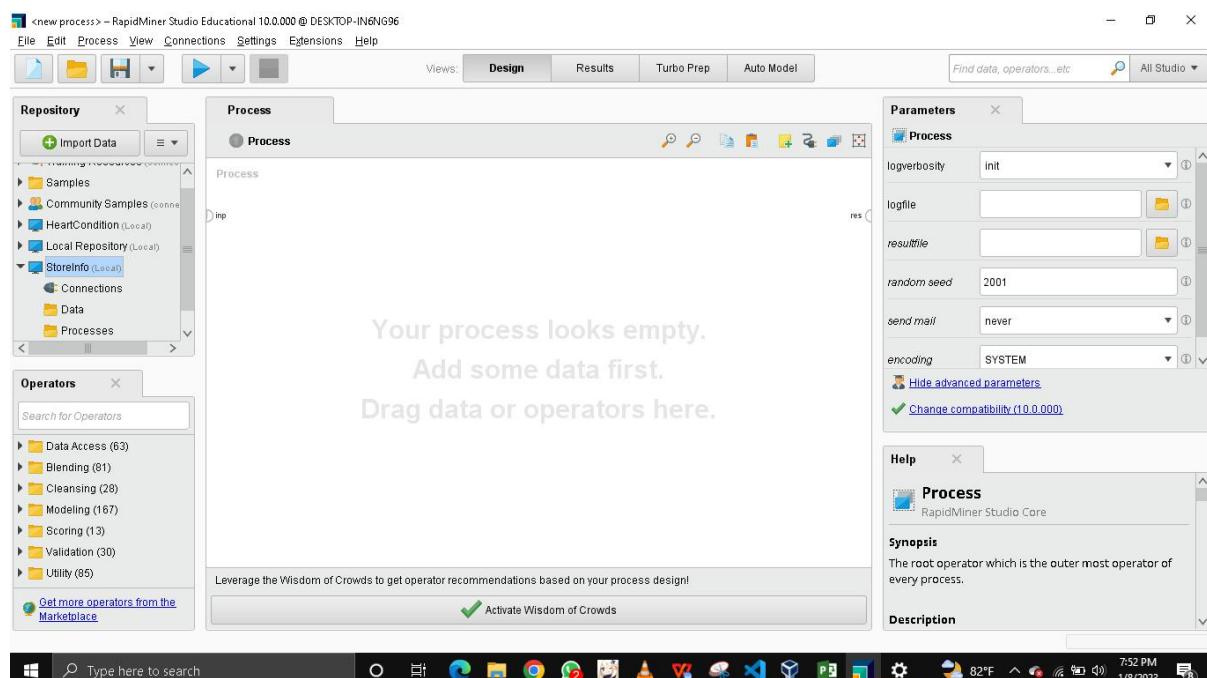
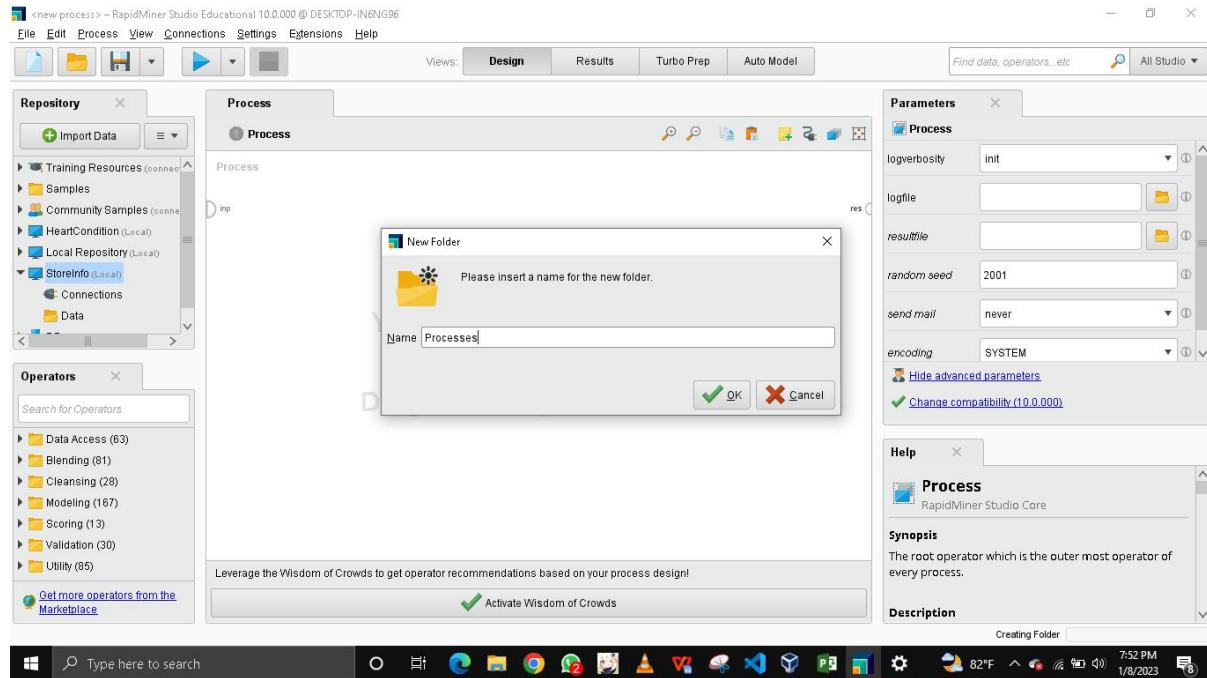
Storedata variables:

1. customerID
2. gender
3. purchaseDate
4. TID
5. flour
6. sugar
7. cornflakes
8. grapes
9. apples
10. butter
11. Milk

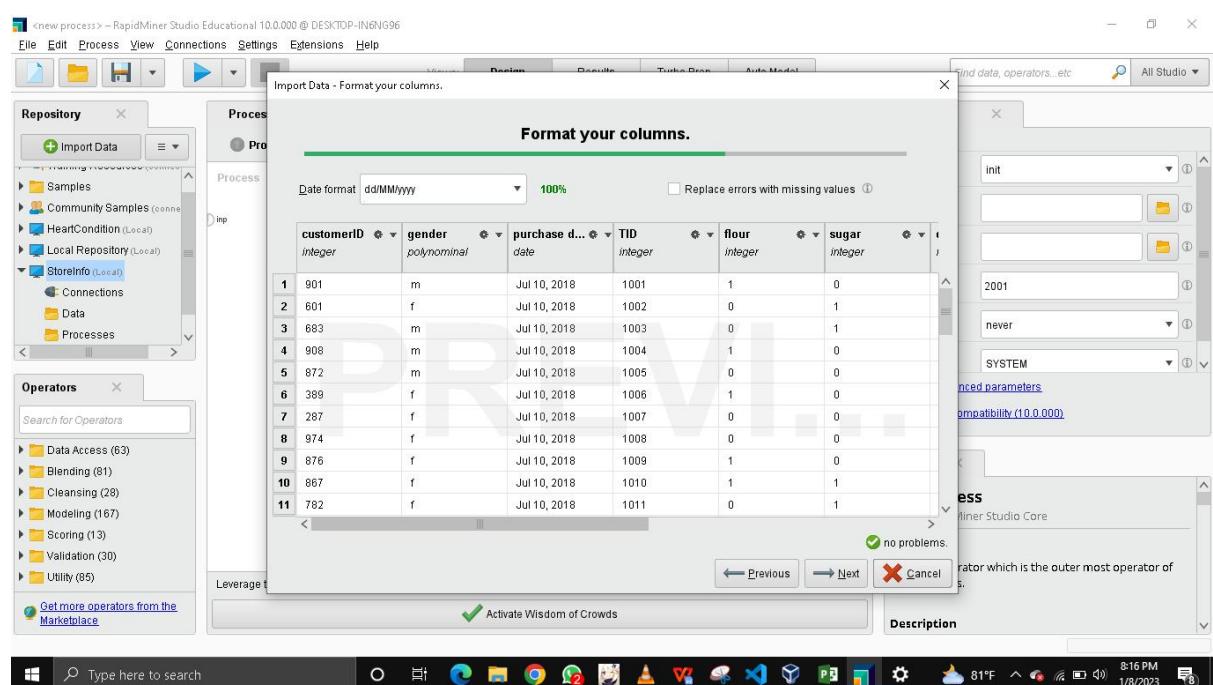
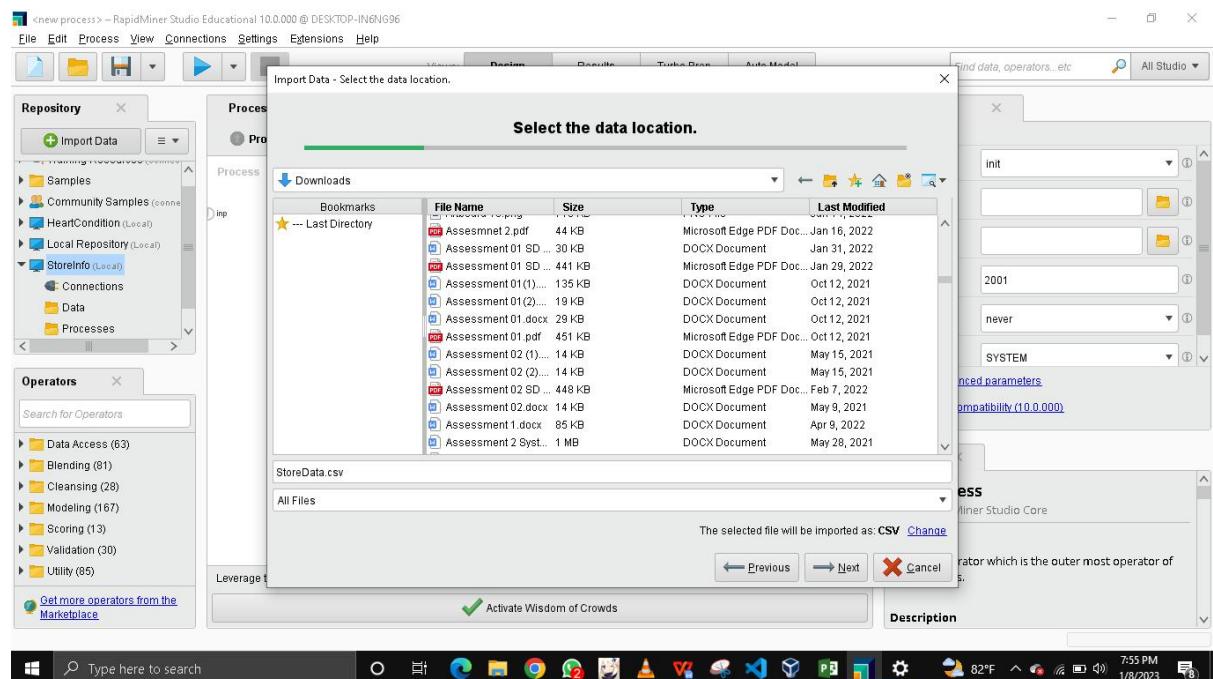
Practical Documentation

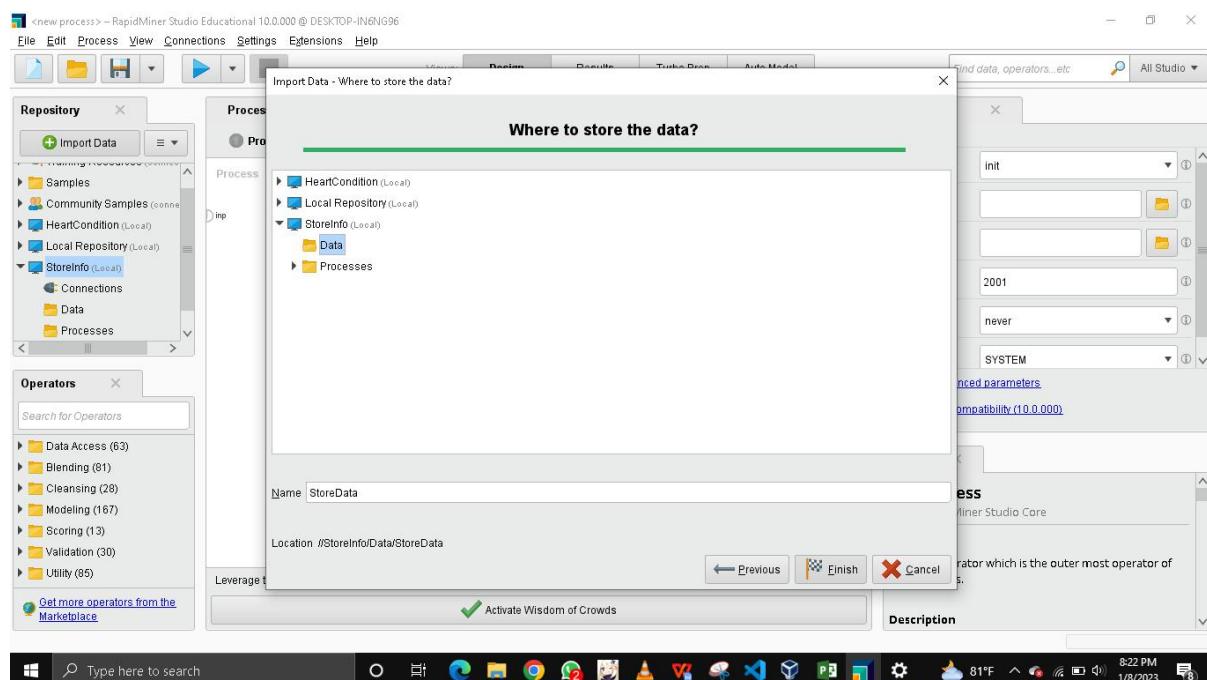
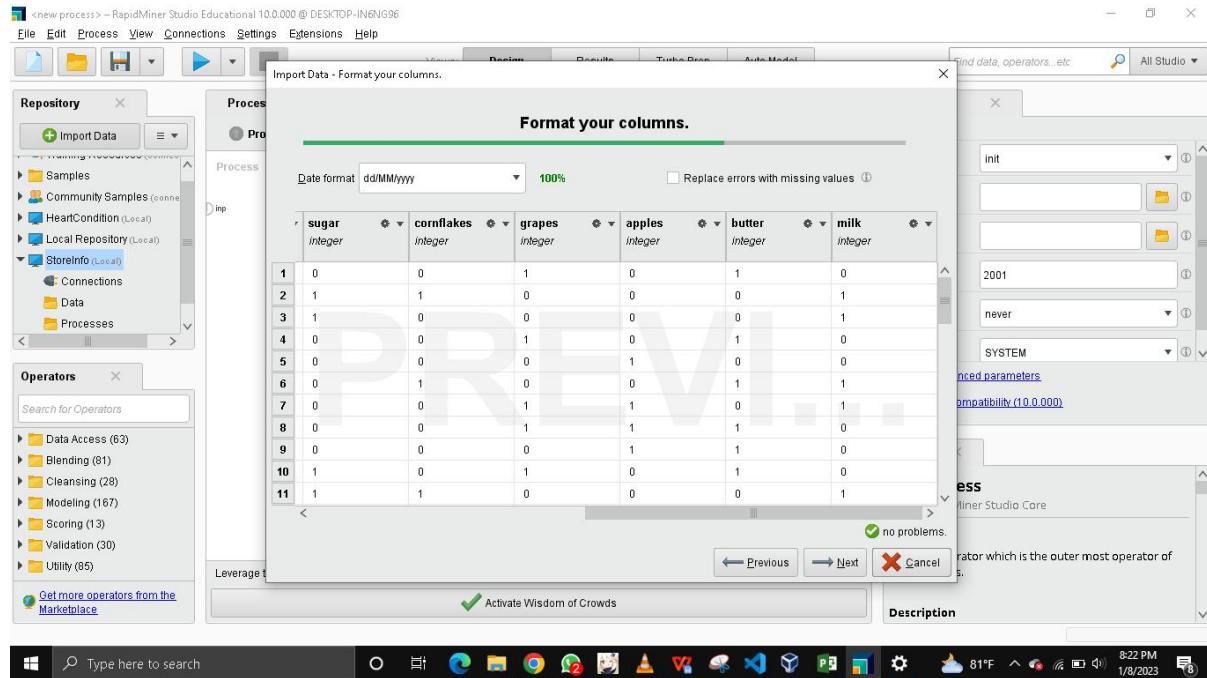
1. Created a Repository labeled "StoreInfo" and created two subfolders named "Data" and "Processes" respectively

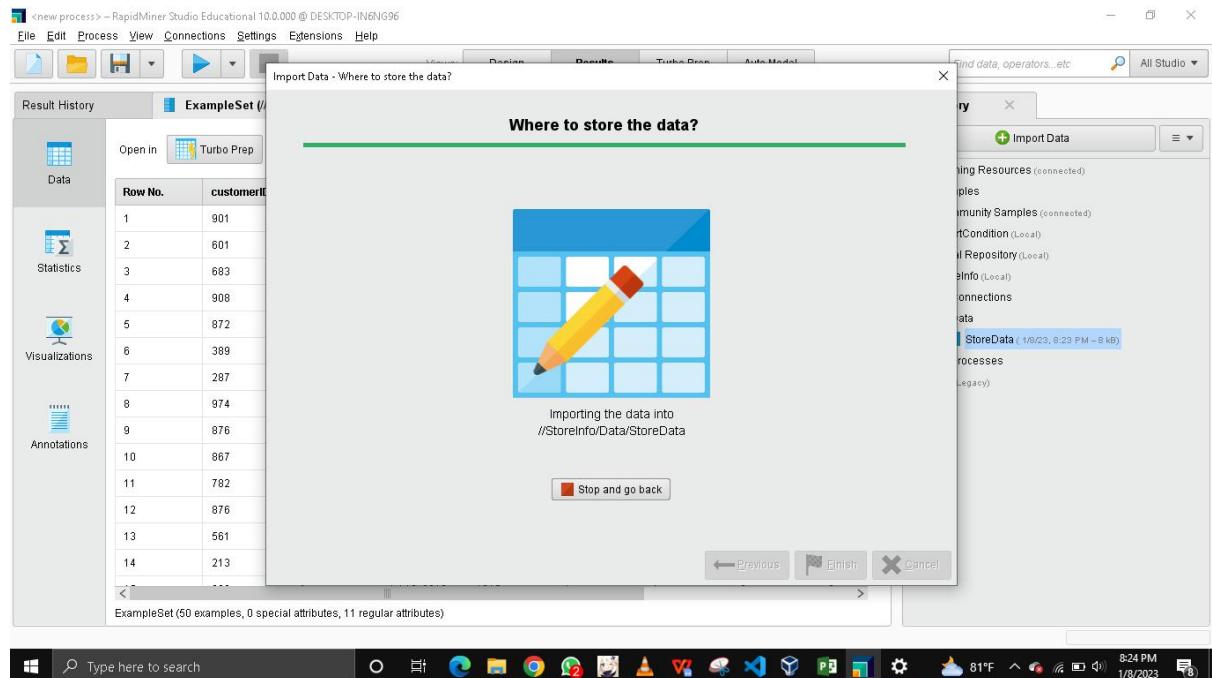




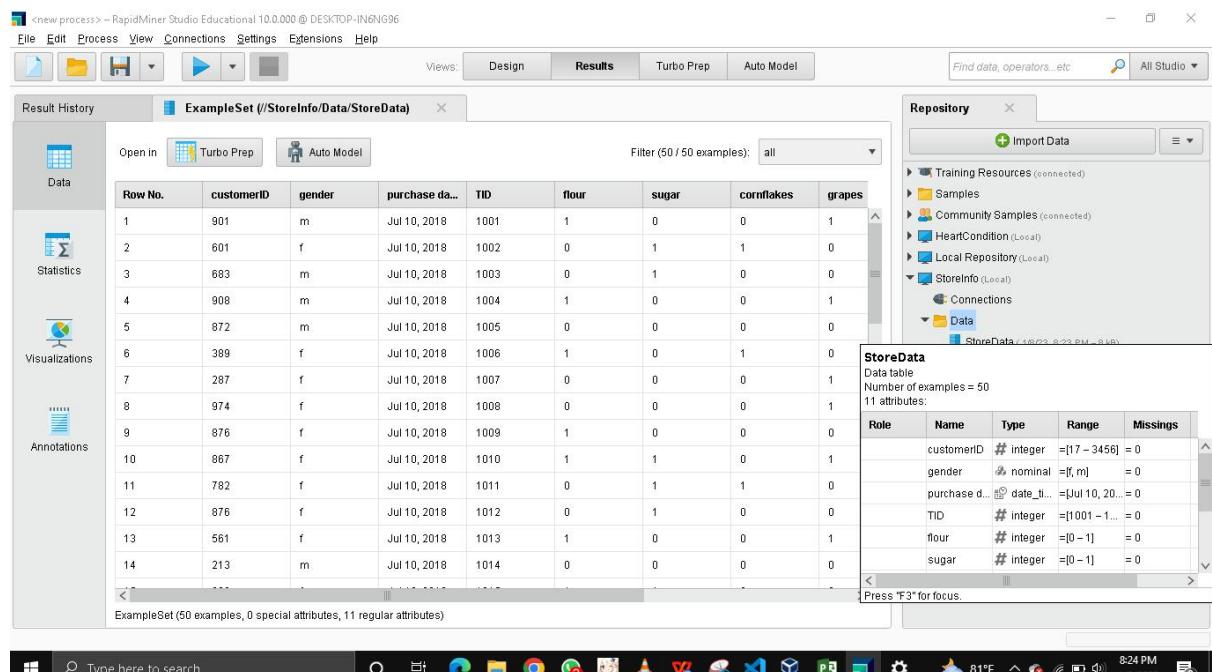
2. Imported the StoreData table into the Data folder of the StoreInfo Repository within the table's default attributes unchanged as shown below



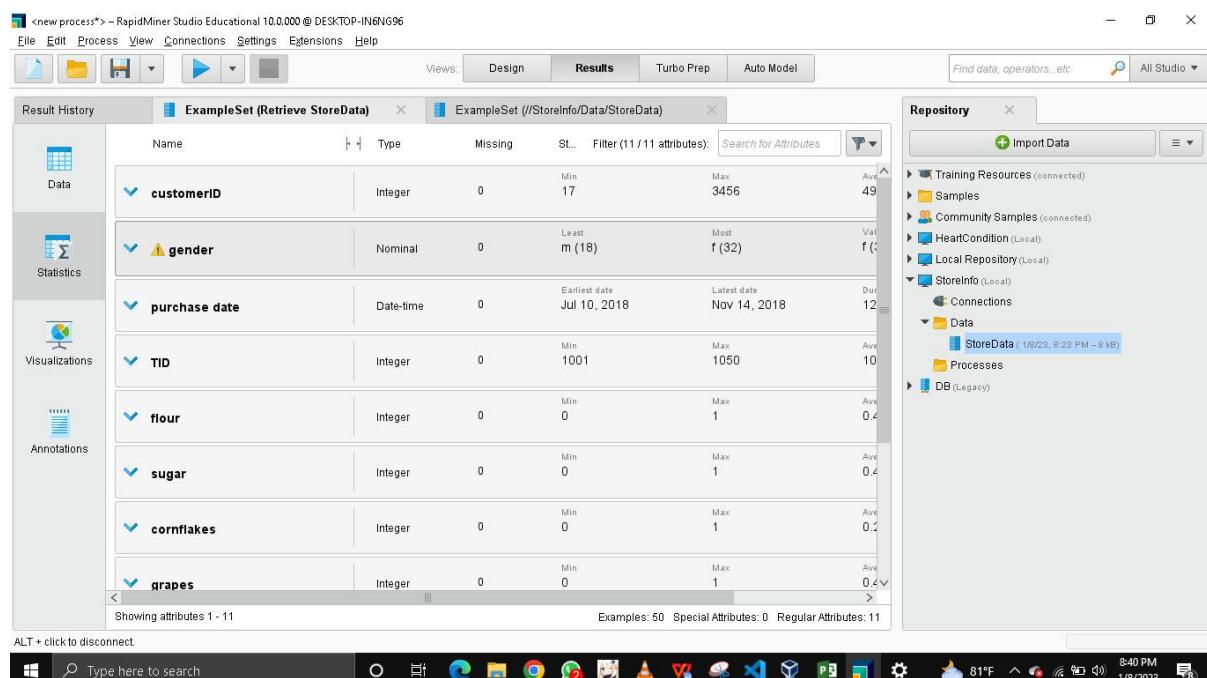
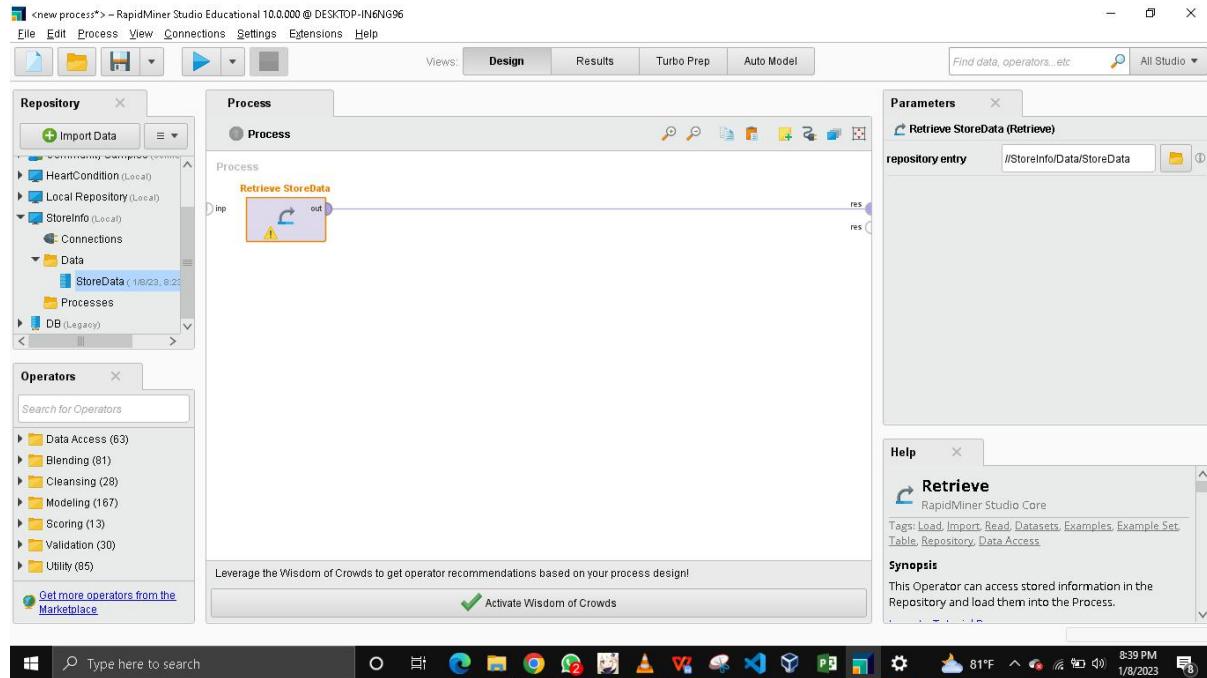




Below are the results of the completed import with the ranges of the values within the StoreData table shown for further detail



2. Retrieved the StoreData table into the Process panel and checked for any missing values; saving the process as “01-import StoreData table” in the Processes subfolder in the StoreInfo Repository



Results View Screenshot:

The screenshot shows the 'Results' tab selected in the top navigation bar. The main content area displays a table of attributes from the 'ExampleSet (Retrieve StoreData)' dataset. The table includes columns for Name, Type, Missing, Min, Max, and Ave. The attributes listed are flour, sugar, cornflakes, grapes, apples, butter, and milk, all of which are Integer type and have 0 missing values, with a range from 0 to 1 and an average of 0.5.

Repository Panel Screenshot:

The 'Repository' panel on the right shows a tree structure of available resources. It includes 'Training Resources' (connected), 'Samples', 'Community Samples' (connected), 'HeartCondition (Local)', 'Local Repository (Local)', and 'StoreInfo (Local)'. Under 'StoreInfo (Local)', there are 'Connections', 'Data' (with 'StoreData' selected), 'Processes', and 'DB (Legacy)'.

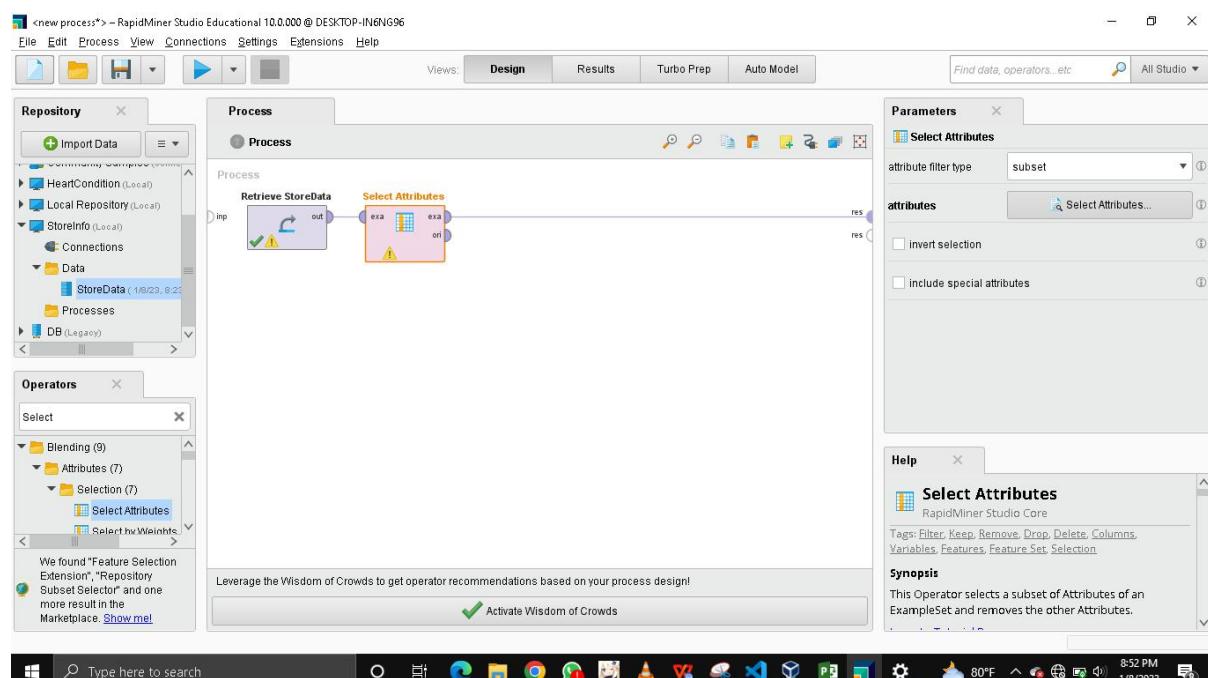
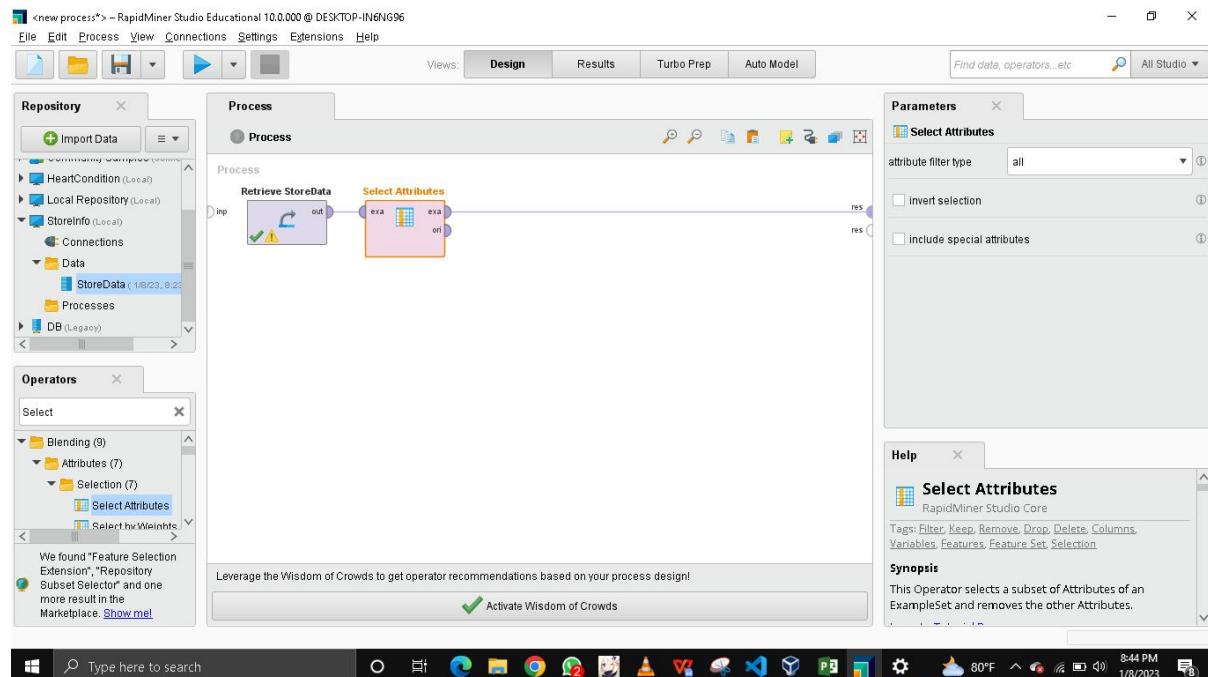
Design View Screenshot:

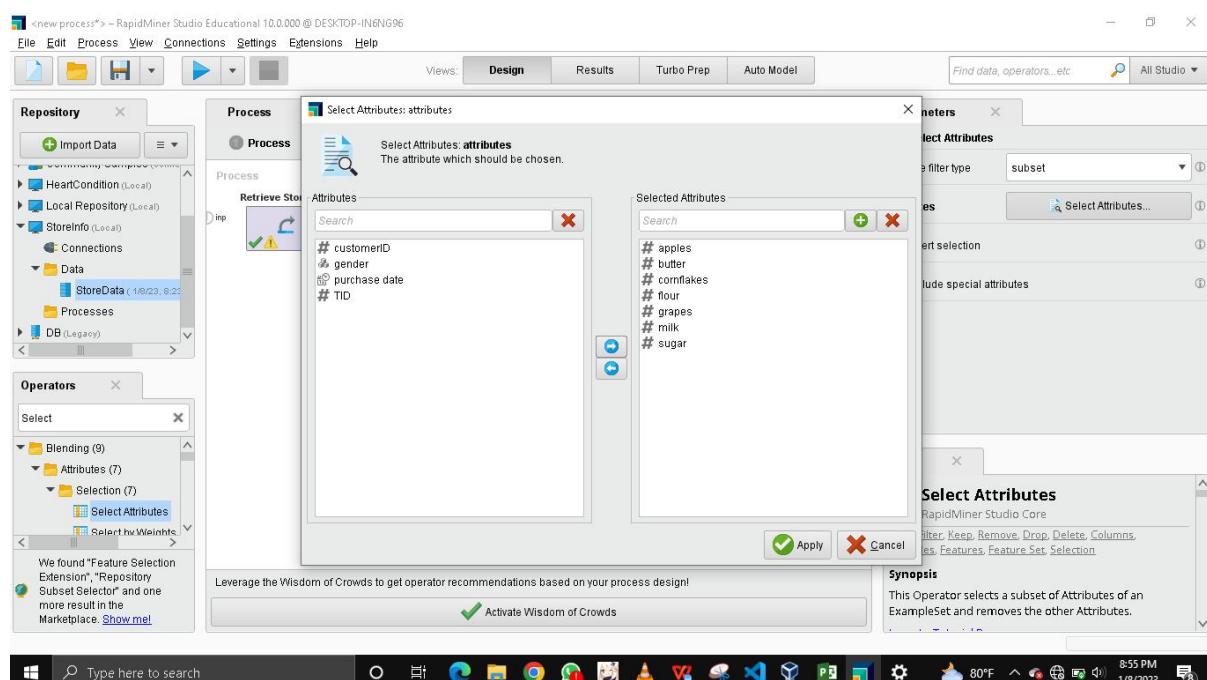
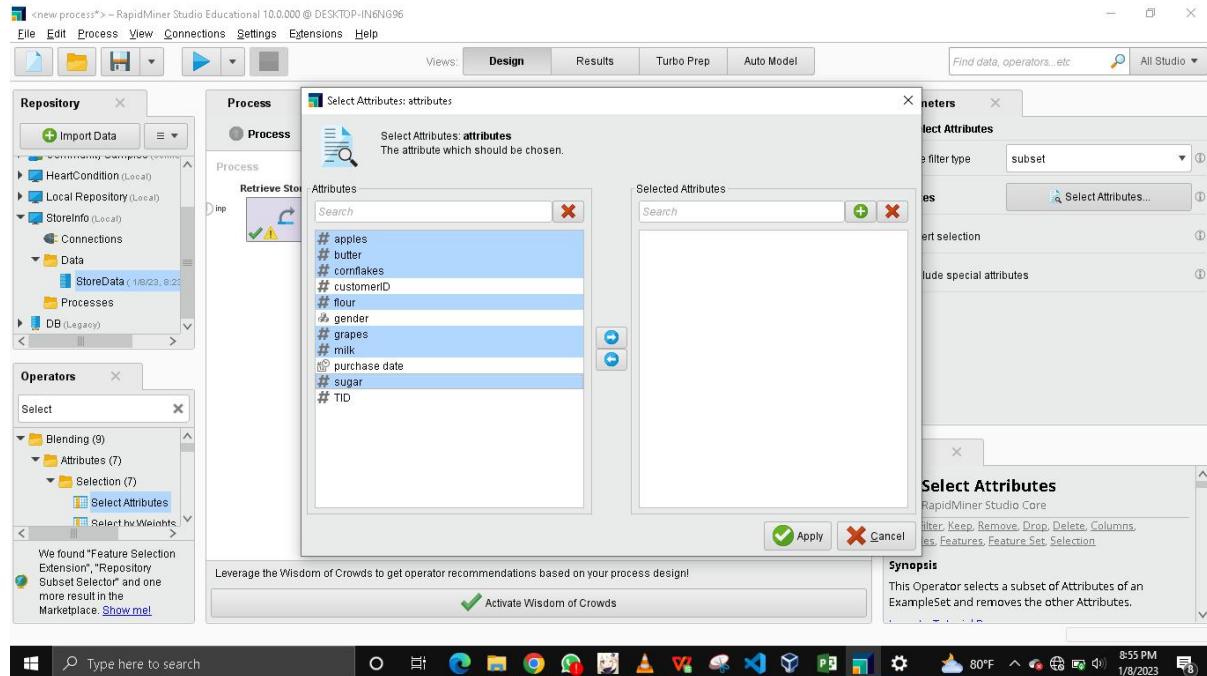
The screenshot shows the 'Design' tab selected in the top navigation bar. A 'Repository Browser' dialog box is open, prompting to 'Select a repository location.' The 'Process' tab is selected in the dialog. The 'Process' panel on the left shows a 'Retrieve StoreData' operator selected. The 'Parameters' panel on the right shows the following configuration for the process:

- Process: init
- seed: 2001
- log: never

The 'Description' panel at the bottom states: "The root operator which is the outer most operator of every process."

3. Used the “Select Attributes” operator to select only binomial data types to be analyzed and saved the process as “02-select attributes”





The screenshot shows the RapidMiner Studio interface with the following details:

- Title Bar:** //StoreInfo/Processes/01-import StoreData table* – RapidMiner Studio Educational 10.0.0.0 @ DESKTOP-IN6VG96
- Menu Bar:** File, Edit, Process, View, Connections, Settings, Extensions, Help
- Toolbar:** Includes icons for Open, Save, Run, Stop, and Preferences.
- Views:** Design, Results, Turbo Prep, Auto Model
- Search Bar:** Find data, operators...etc.
- Result History:** Shows ExampleSet (Select Attributes) and ExampleSet (/StoreInfo/Data/StoreData).
- Data View:** A table titled "ExampleSet (50 examples, 0 special attributes, 7 regular attributes)" with columns: Row No., flour, sugar, cornflakes, grapes, apples, butter, milk. The data rows are as follows:

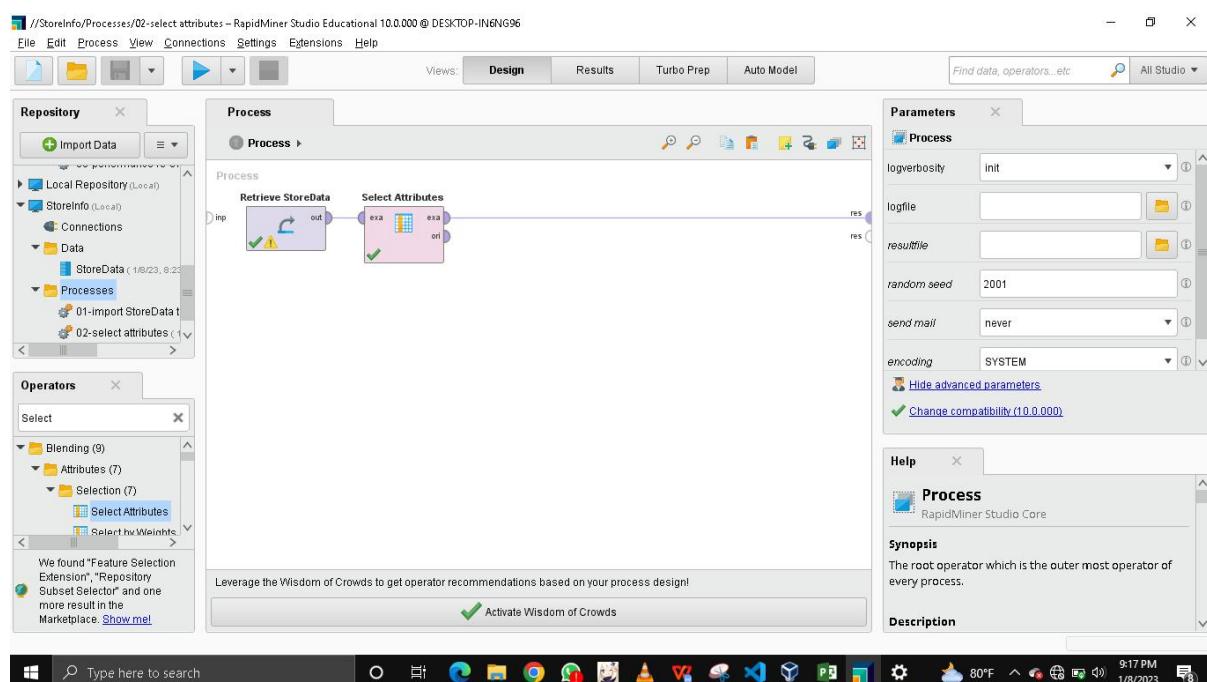
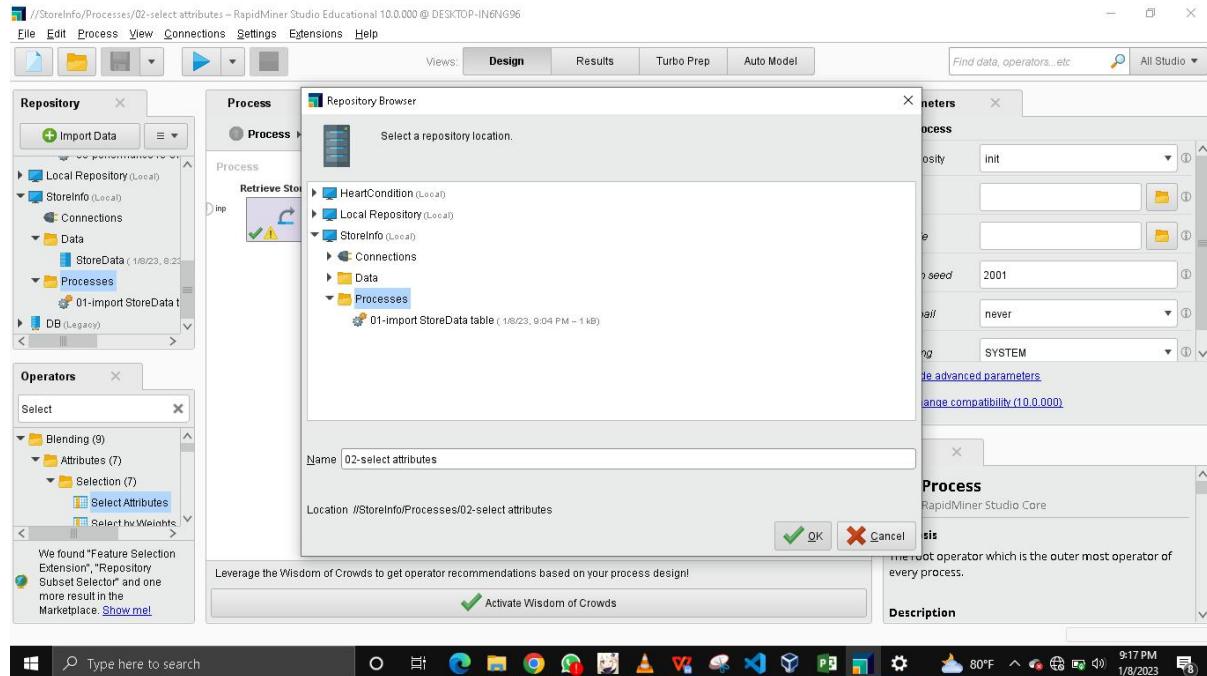
Row No.	flour	sugar	cornflakes	grapes	apples	butter	milk
1	1	0	0	1	0	1	0
2	0	1	1	0	0	0	1
3	0	1	0	0	0	0	1
4	1	0	0	1	0	1	0
5	0	0	0	0	1	0	0
6	1	0	1	0	0	1	1
7	0	0	0	1	1	0	1
8	0	0	0	1	1	1	0
9	1	0	0	0	1	1	0
10	1	1	0	1	0	1	0
11	0	1	1	0	0	0	1
12	0	1	0	0	0	0	1
13	1	0	0	1	0	1	0
14	0	0	0	0	1	0	0
15	1	1	0	0	0	1	1

- Repository:** Shows a tree structure with Training Resources, Samples, Community Samples, HeartCondition (Local), Connections, Data, Processes, Local Repository (Local), StoreInfo (Local), Data (StoreData table), Processes (01-import StoreData table), and DB (Legacy).

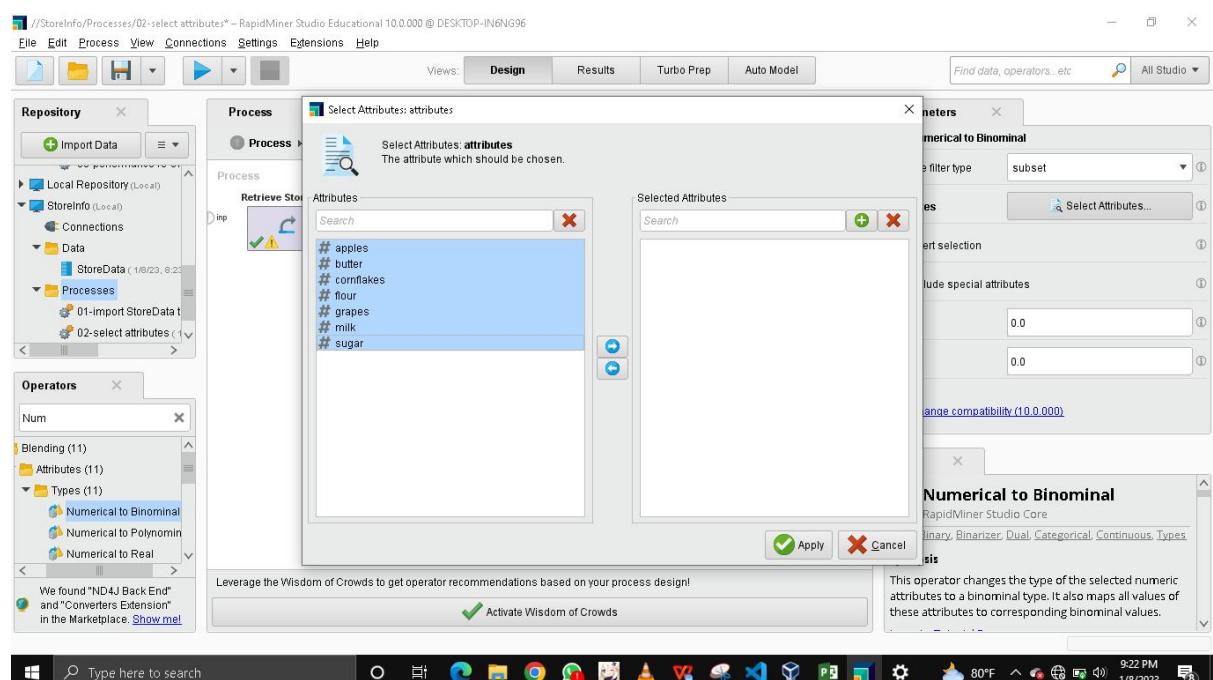
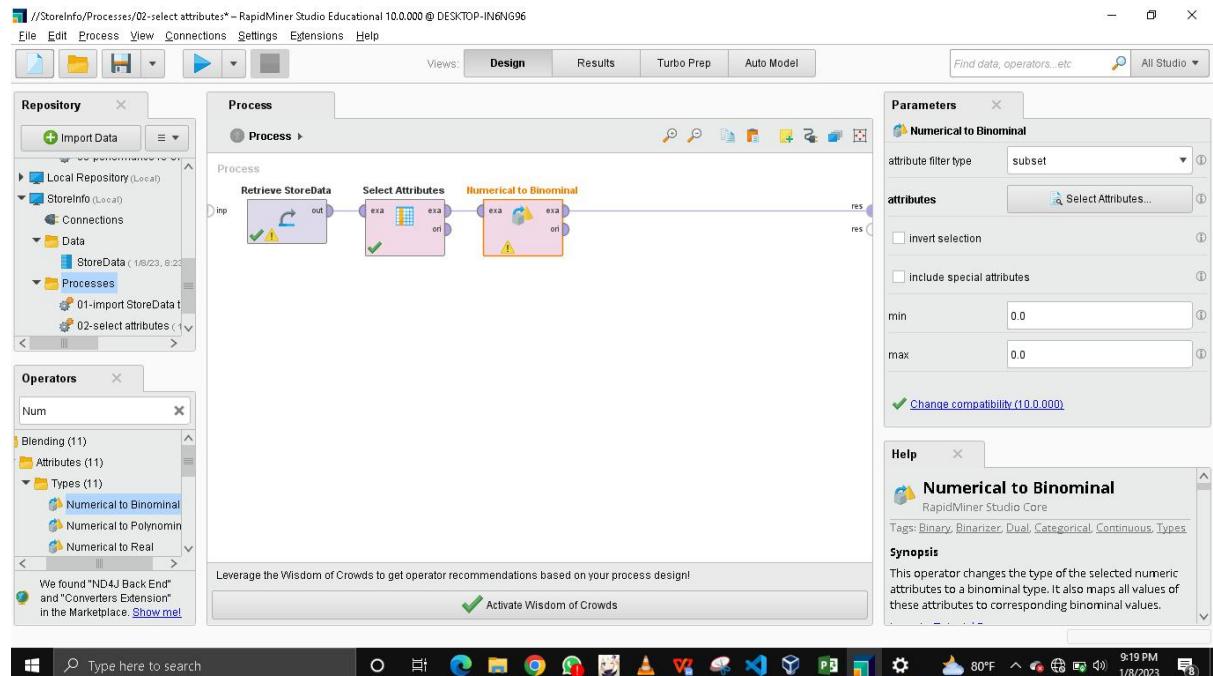
The screenshot shows the RapidMiner Studio interface with the following details:

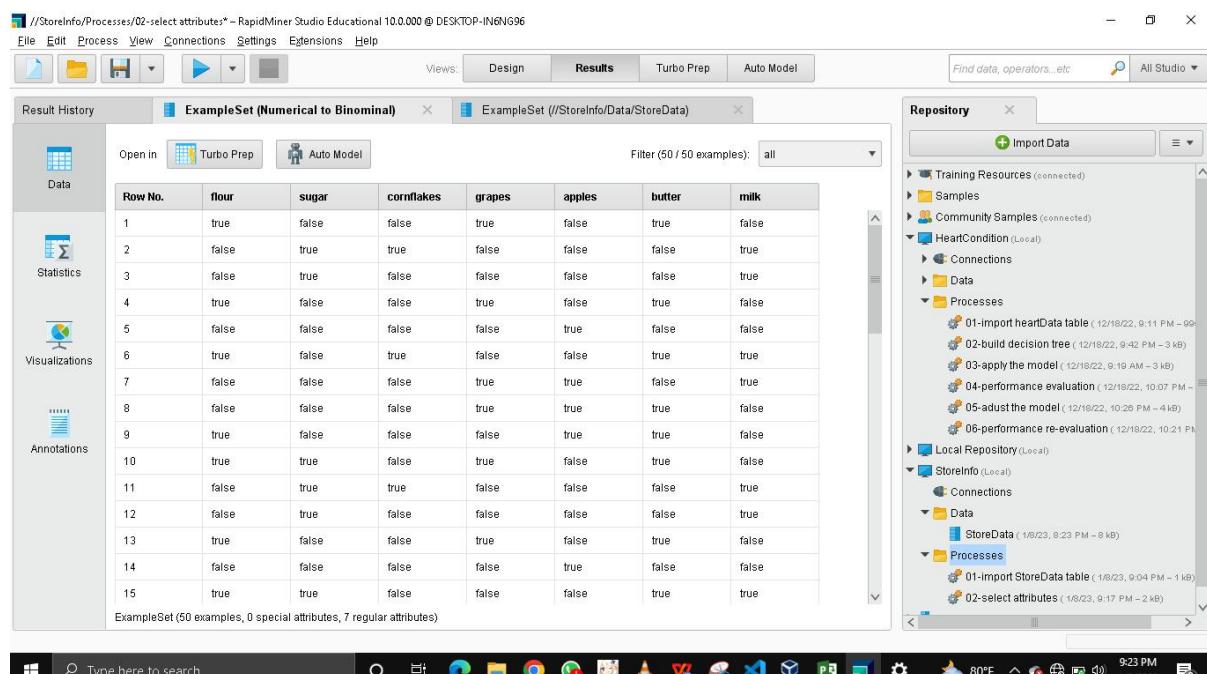
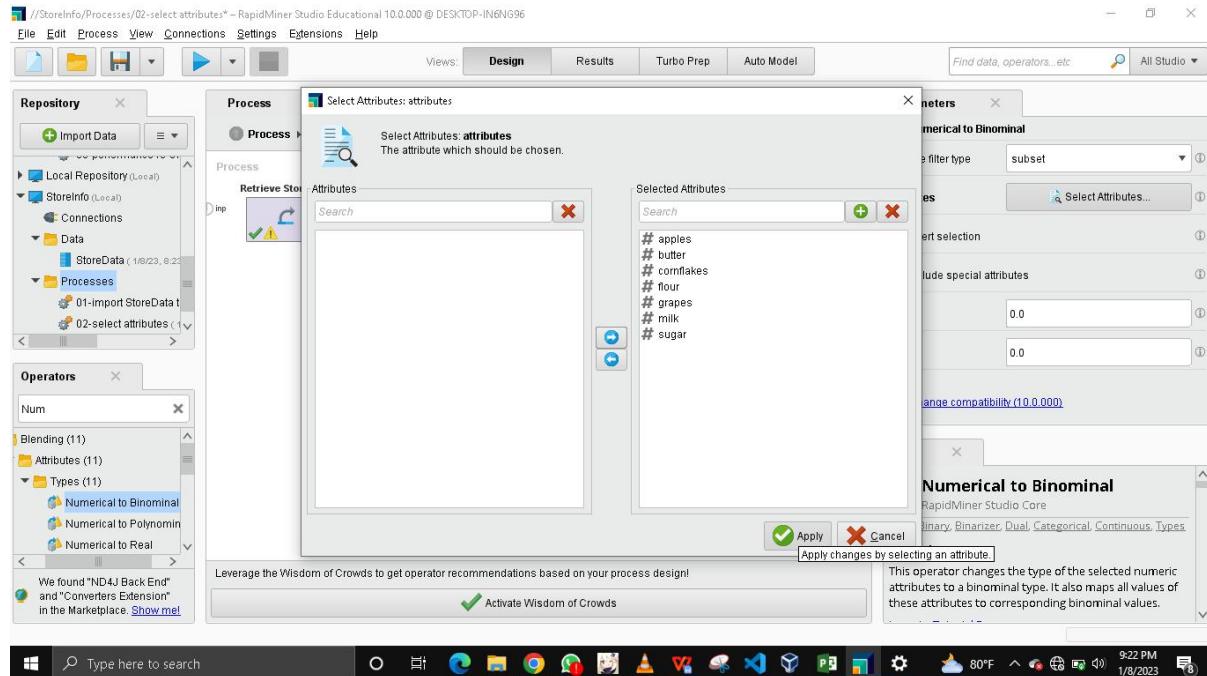
- Title Bar:** //StoreInfo/Processes/01-import.StoreData table* – RapidMiner Studio Educational 10.0.0.0 @ DESKTOP-INENG96
- Menu Bar:** File Edit Process View Connections Settings Extensions Help
- Toolbar:** Includes icons for File, Folder, History, Play, Stop, and Save.
- Views:** Design, Results, Turbo Prep, Auto Model.
- Result History:** Shows the current process flow: ExampleSet (Select Attributes) → ExampleSet (//StoreInfo/Data/StoreData).
- Data Tab:** Displays a table of attributes:

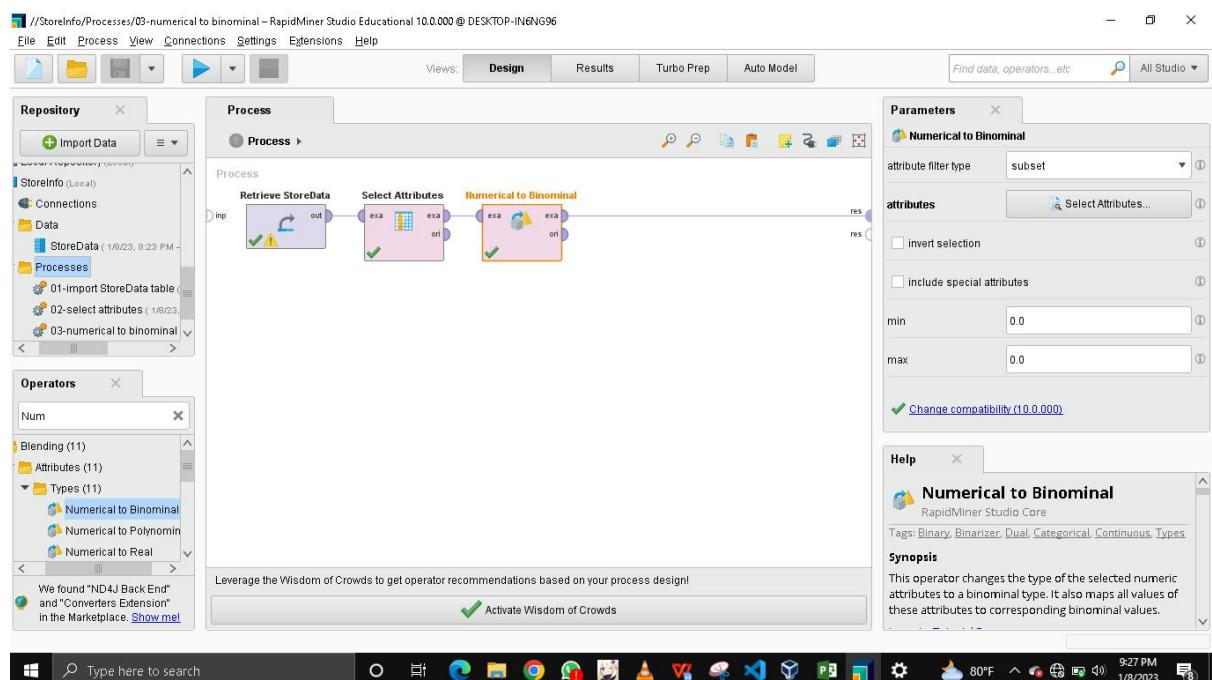
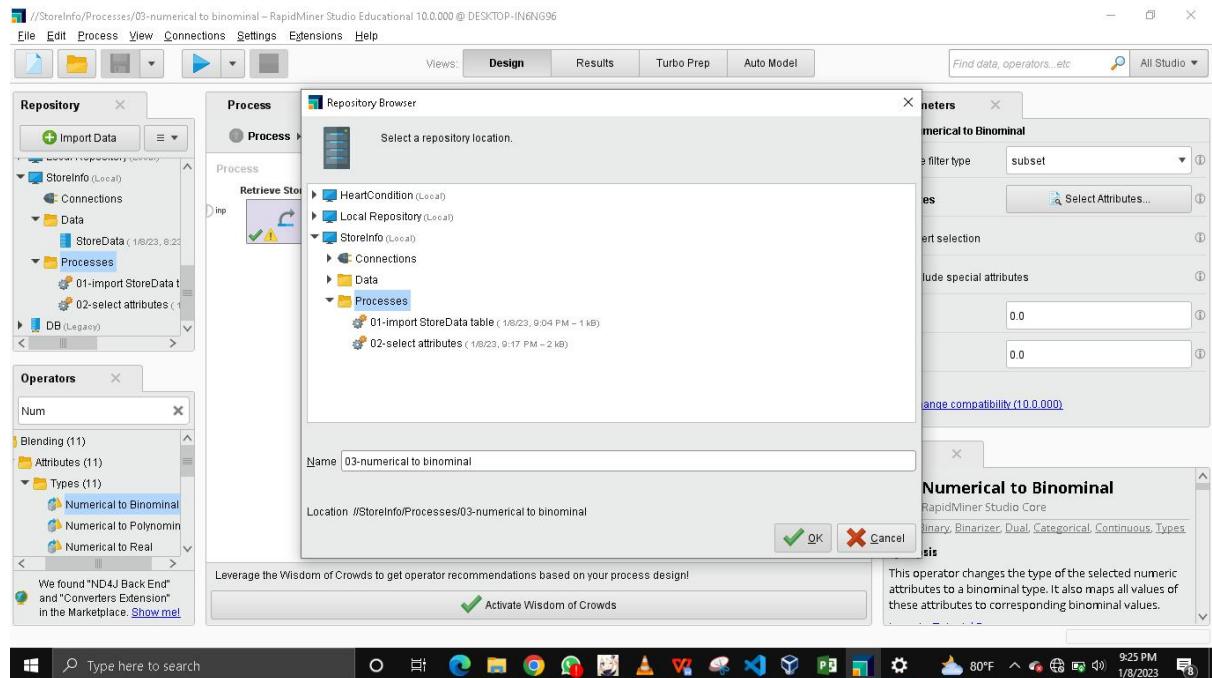
Name	Type	Missing	Statist...	Filter (7 / 7 attributes):	Search for Attributes
flour	Integer	0	Min 0	Max 1	Average 0.400
sugar	Integer	0	Min 0	Max 1	Average 0.440
cornflakes	Integer	0	Min 0	Max 1	Average 0.260
grapes	Integer	0	Min 0	Max 1	Average 0.400
apples	Integer	0	Min 0	Max 1	Average 0.440
butter	Integer	0	Min 0	Max 1	Average 0.520
milk	Integer	0	Min 0	Max 1	Average 0.520
- Annotations Tab:** Shows a list of annotations.
- Visualizations Tab:** Shows a list of visualizations.
- Repository Tab:** Shows the project structure:
 - Training Resources (connected)
 - Samples
 - Community Samples (connected)
 - HeartCondition (Local)
 - Connections
 - Data
 - Processes
 - Local Repository (Local)
 - StoreInfo (Local)
 - Connections
 - Data
 - StoreData (1/8/23, 9:23 PM ~ 8 kB)
 - Processes
 - 01-import.StoreData table (1/8/23, 9:04 PM ~ 1 kB)
 - DB (Legacy)
- Bottom Status Bar:** Showing attributes 1 - 7, Examples: 50, Special Attributes: 0, Regular Attributes: 7.
- System Icons:** Windows Start button, Search bar, Taskbar with various application icons (File Explorer, Edge, Google Chrome, etc.), and system status icons (Battery, Network, Volume, etc.).
- System Clock:** 9:10 PM, 1/8/2023.



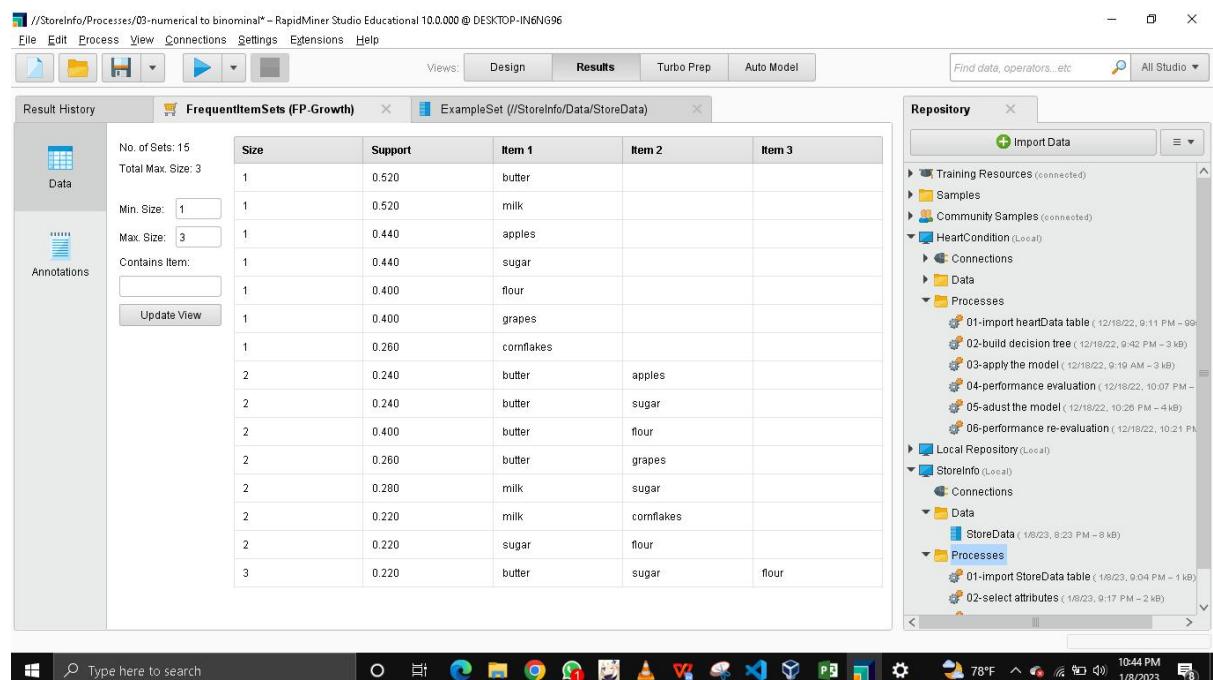
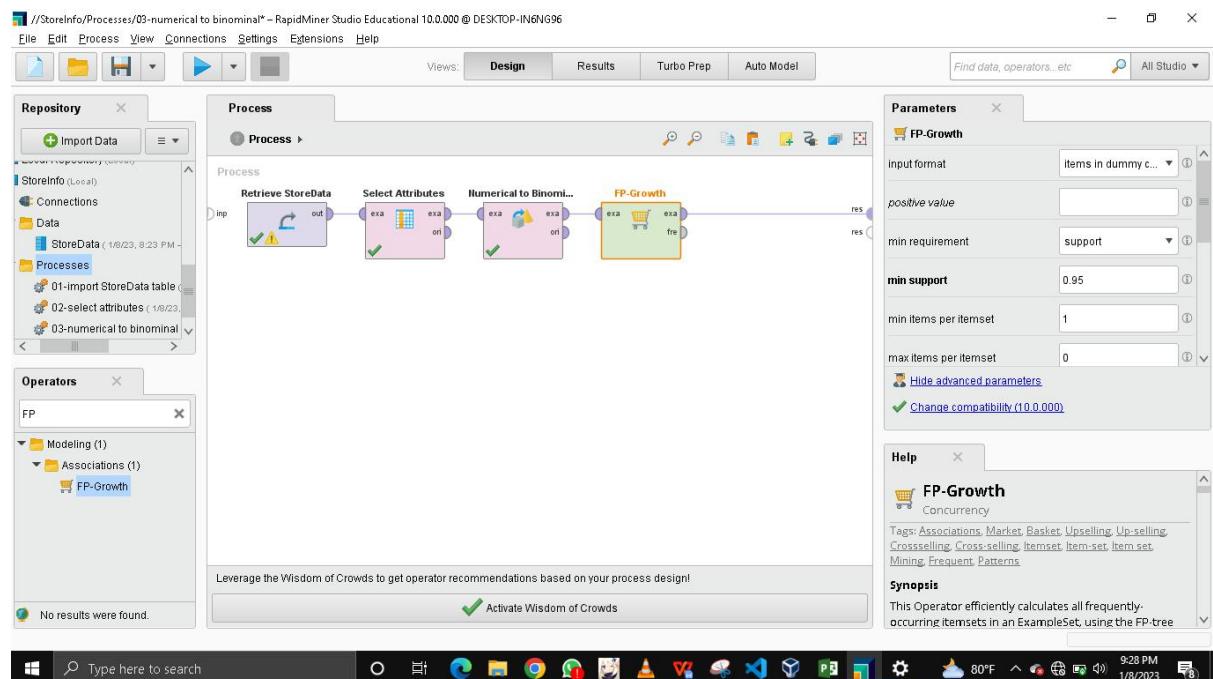
4. Used the “Numerical to Binominal” operator to convert all data into binomial data only for analysis and saved the process as “03-numerical to binominal”

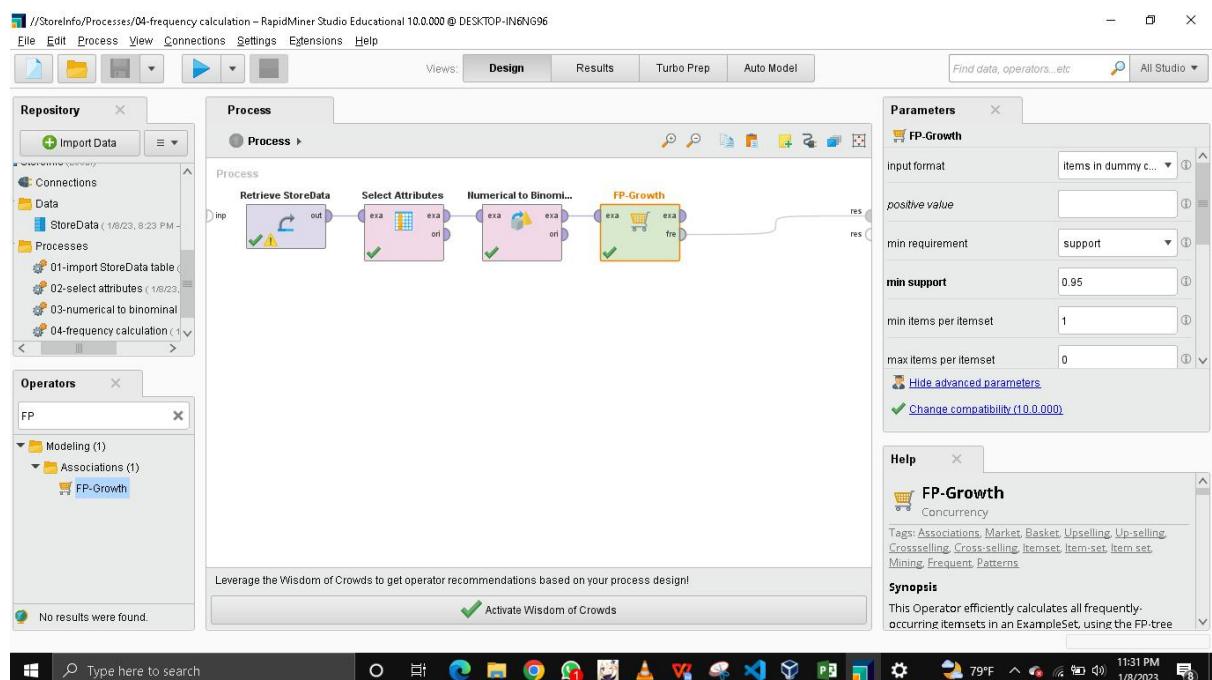
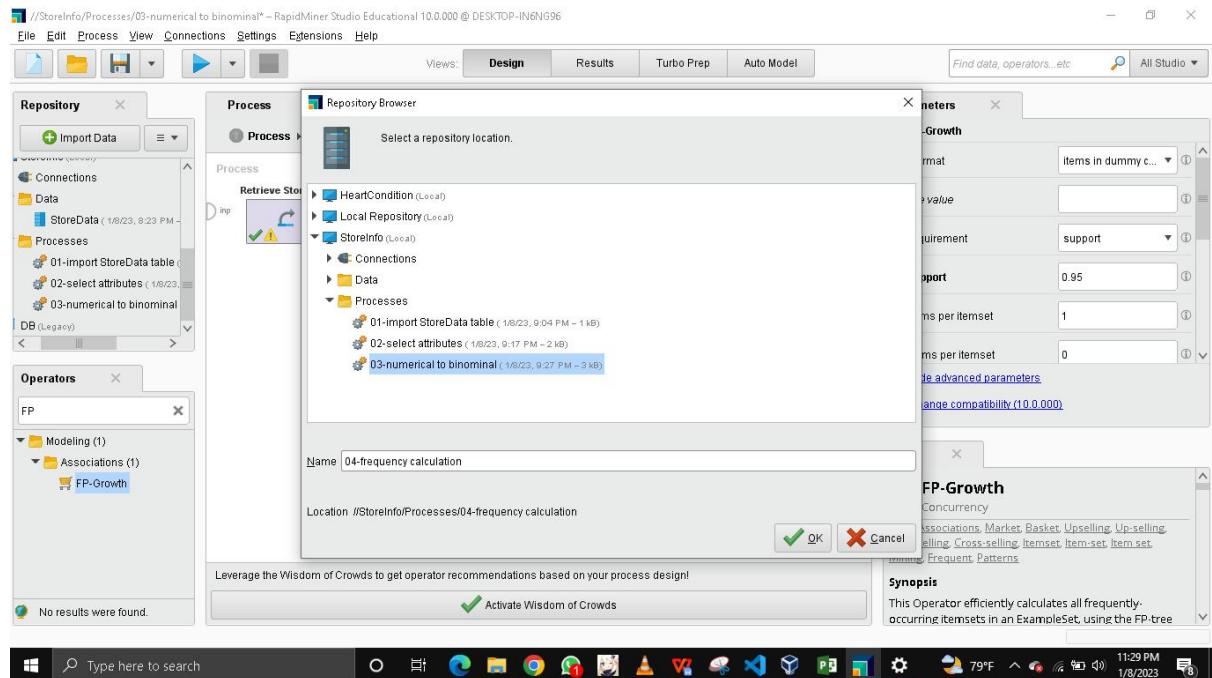




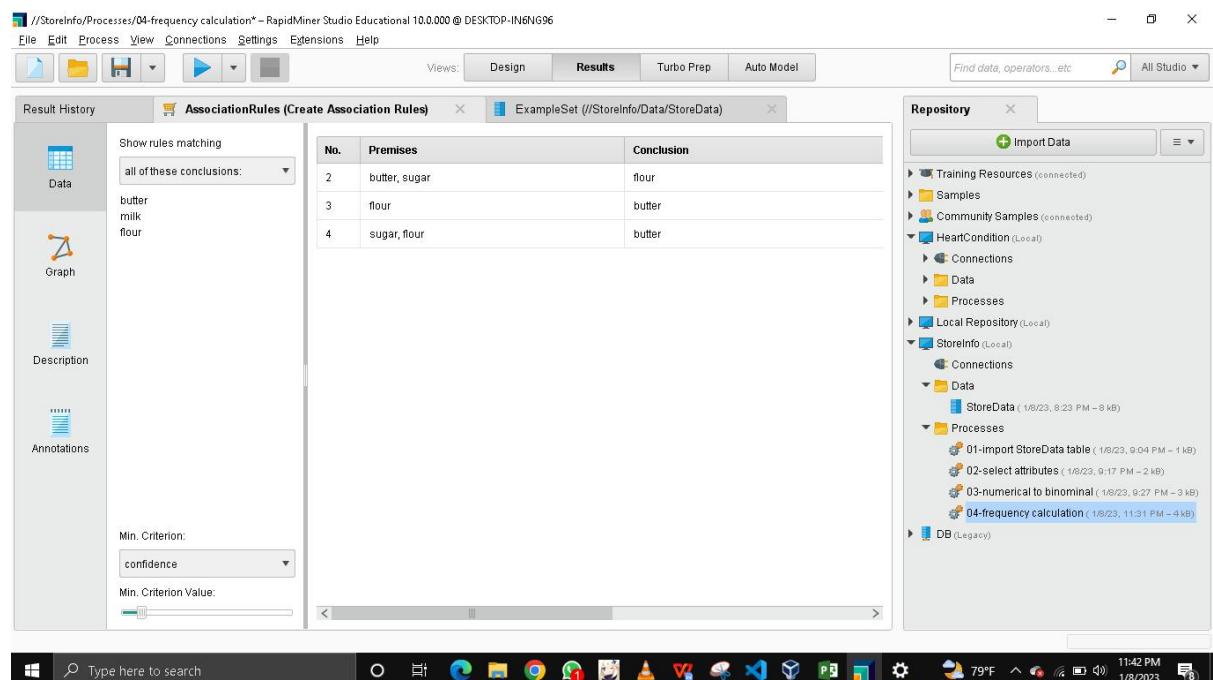
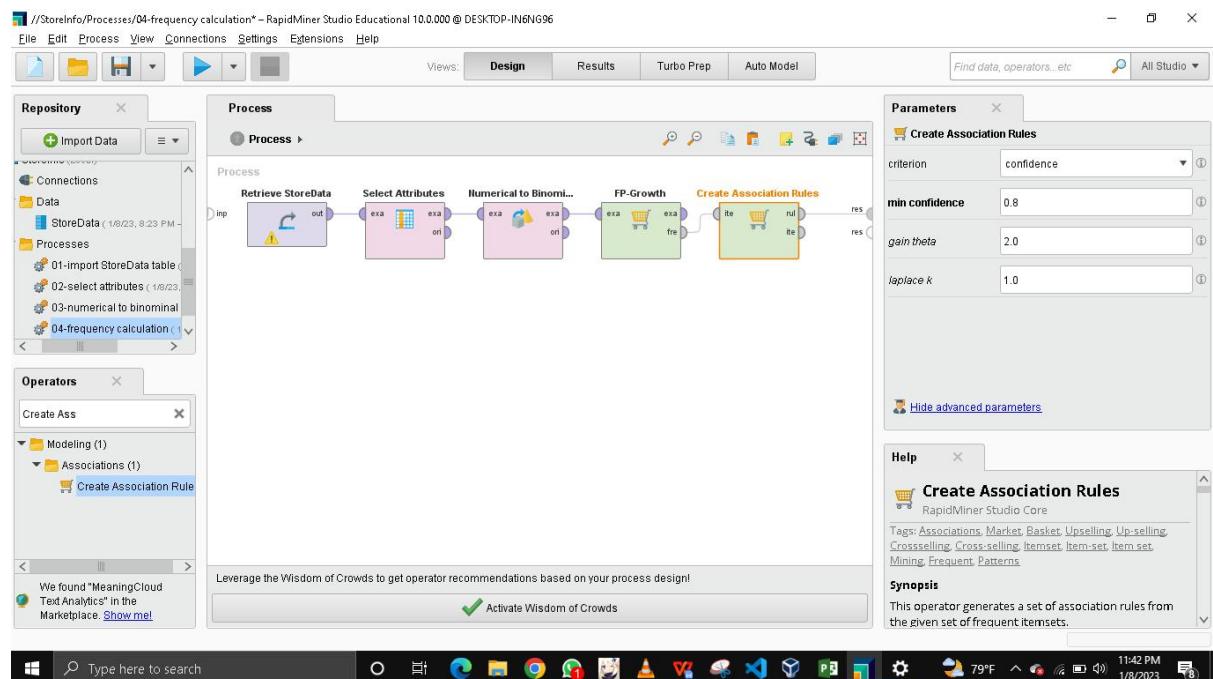


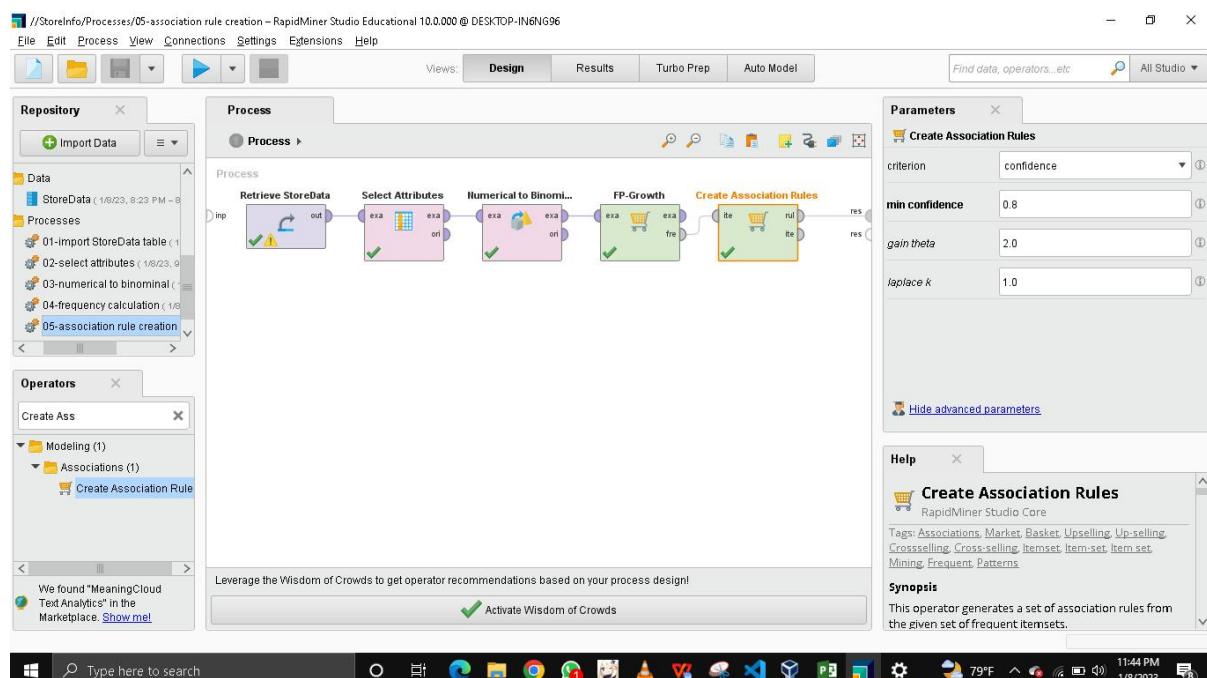
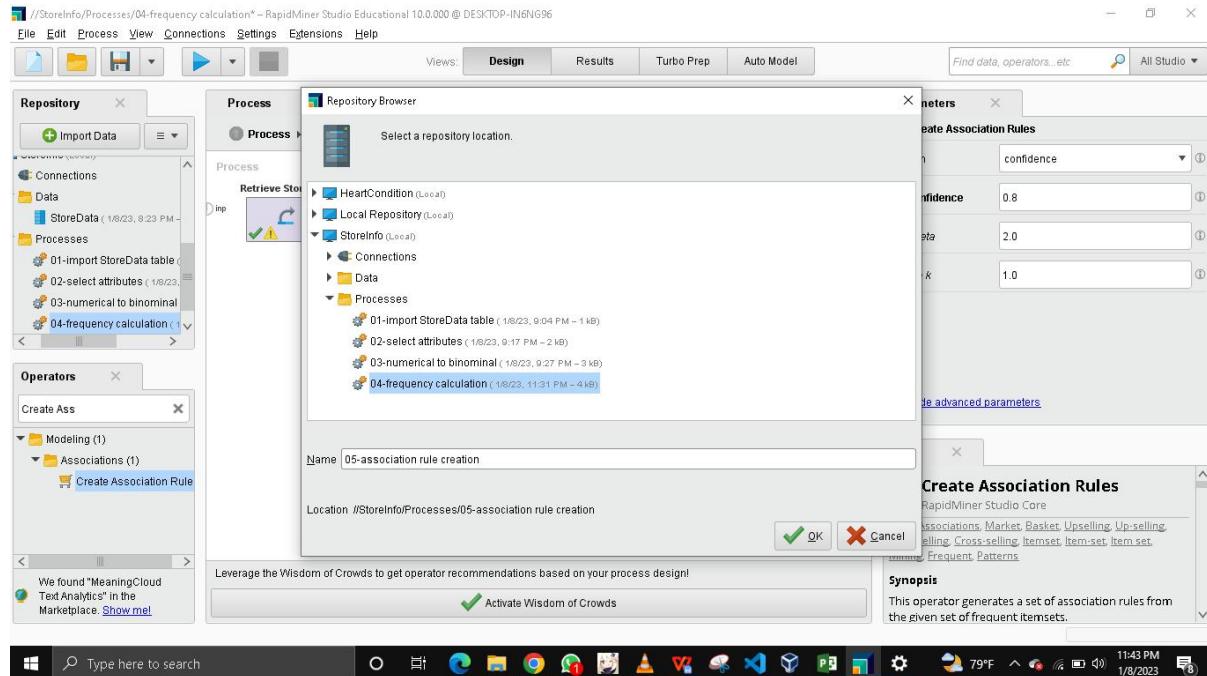
5. Used the “FP-Growth” operator to determine the frequency of each item and sets of items purchased together and saved the process as “04-frequency calculation”



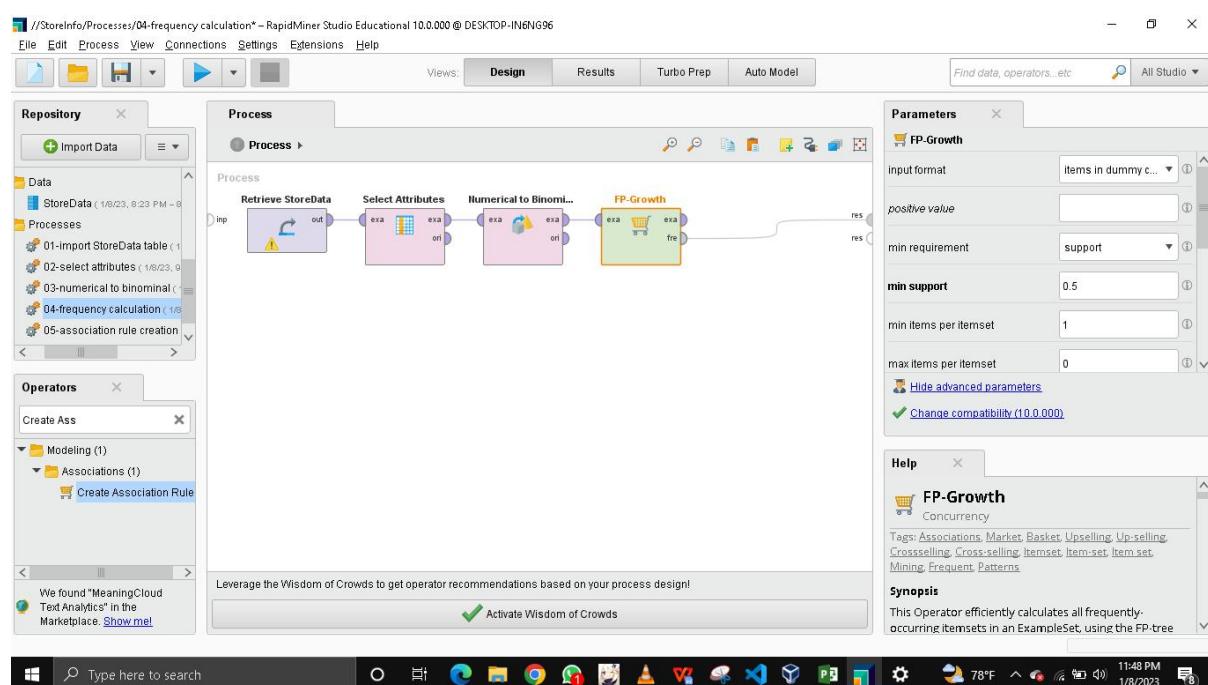
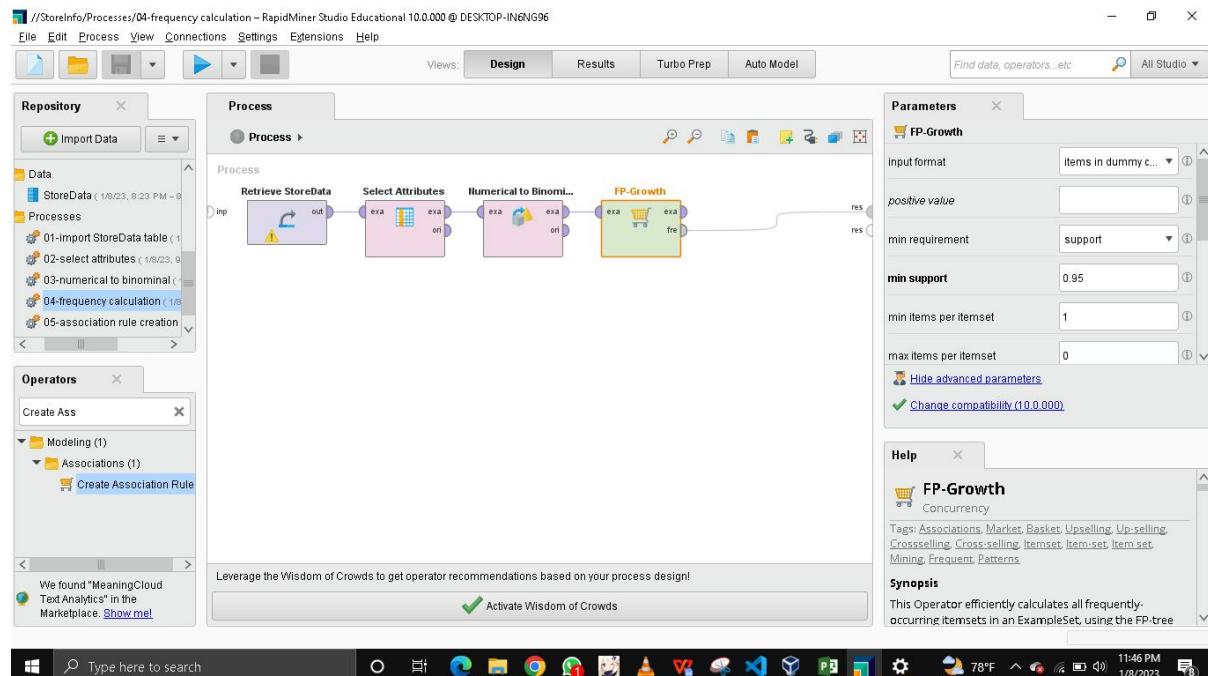


6. Used the “Create Association Rules” operator to create association rules for each of the frequented item sets and saved the process as “05-association rule creation”





7. Used the previously saved “04-frequency calculation” process and edited to lower the default minimum support value from 0.95 to 0.5 in the “FP-Growth” operator to improve the results of item set frequency by generating more two and three item - item sets and saved the process as “06-frequency recalculations”



//StoreInfo/Processes/04-frequency calculation* – RapidMiner Studio Educational 10.0.000 @ DESKTOP-IN8NG96

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model Find data, operators...etc All Studio

Result History **FrequentItemSets (FP-Growth)** ExampleSet (/StoreInfo/Data/StoreData)

No. of Sets: 25 Total Max. Size: 3

Min. Size: 1 Max. Size: 3 Contains Item: Update View

Size	Support	Item 1	Item 2	Item 3
1	0.520	butter		
1	0.520	milk		
1	0.440	apples		
1	0.440	sugar		
1	0.400	flour		
1	0.400	grapes		
1	0.260	cornflakes		
2	0.140	butter	milk	
2	0.240	butter	apples	
2	0.240	butter	sugar	
2	0.400	butter	flour	
2	0.260	butter	grapes	
2	0.160	milk	apples	
2	0.280	milk	sugar	
2	0.120	milk	grapes	
2	0.220	milk	cornflakes	

Repository Import Data

- Training Resources (connected)
- Samples
- Community Samples (connected)
- HeartCondition (Local)
 - Connections
 - Data
 - Processes
- Local Repository (Local)
- StoreInfo (Local)
 - Connections
 - Data
 - StoreData (1/8/23, 8:23 PM - 8 kB)
 - Processes
 - 01-import StoreData table (1/8/23, 9:04 PM - 1 kB)
 - 02-select attributes (1/8/23, 9:17 PM - 2 kB)
 - 03-numerical to binomial (1/8/23, 9:27 PM - 3 kB)
 - 04-frequency calculation (1/8/23, 11:31 PM - 4 kB)
 - 05-association rule creation (1/8/23, 11:43 PM - 5 kB)
- DB (Legacy)

11:48 PM 1/8/2023

//StoreInfo/Processes/04-frequency calculation* – RapidMiner Studio Educational 10.0.000 @ DESKTOP-IN8NG96

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model Find data, operators...etc All Studio

Result History **FrequentItemSets (FP-Growth)** ExampleSet (/StoreInfo/Data/StoreData)

No. of Sets: 25 Total Max. Size: 3

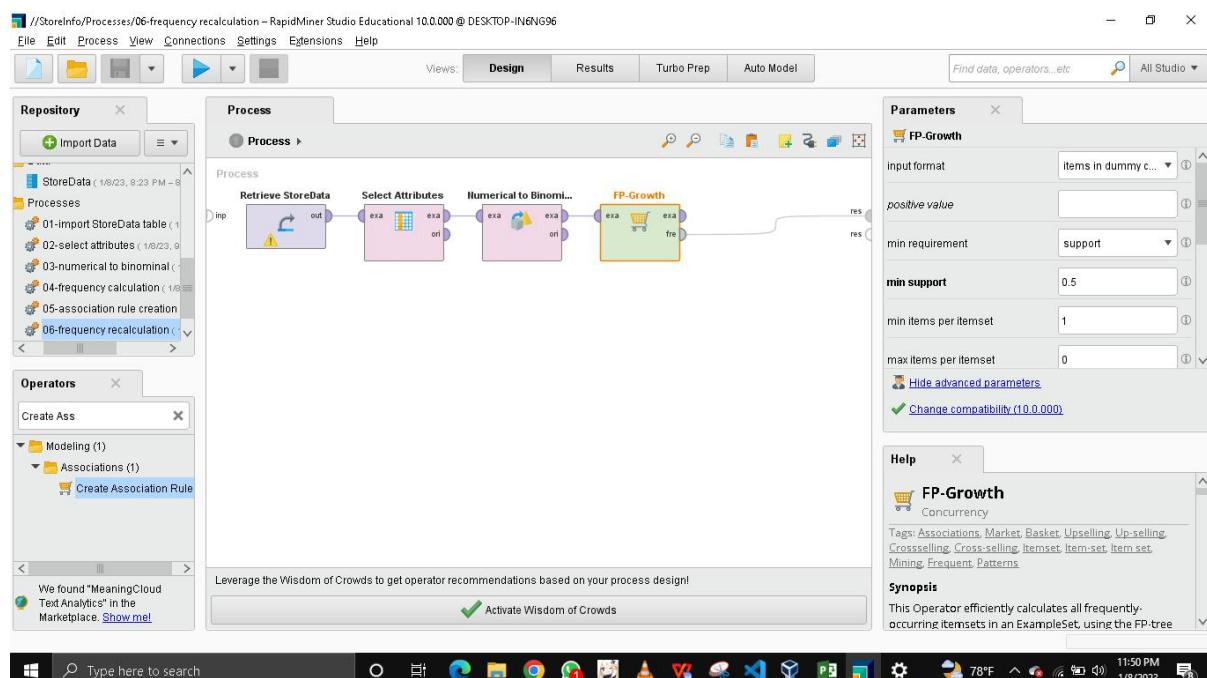
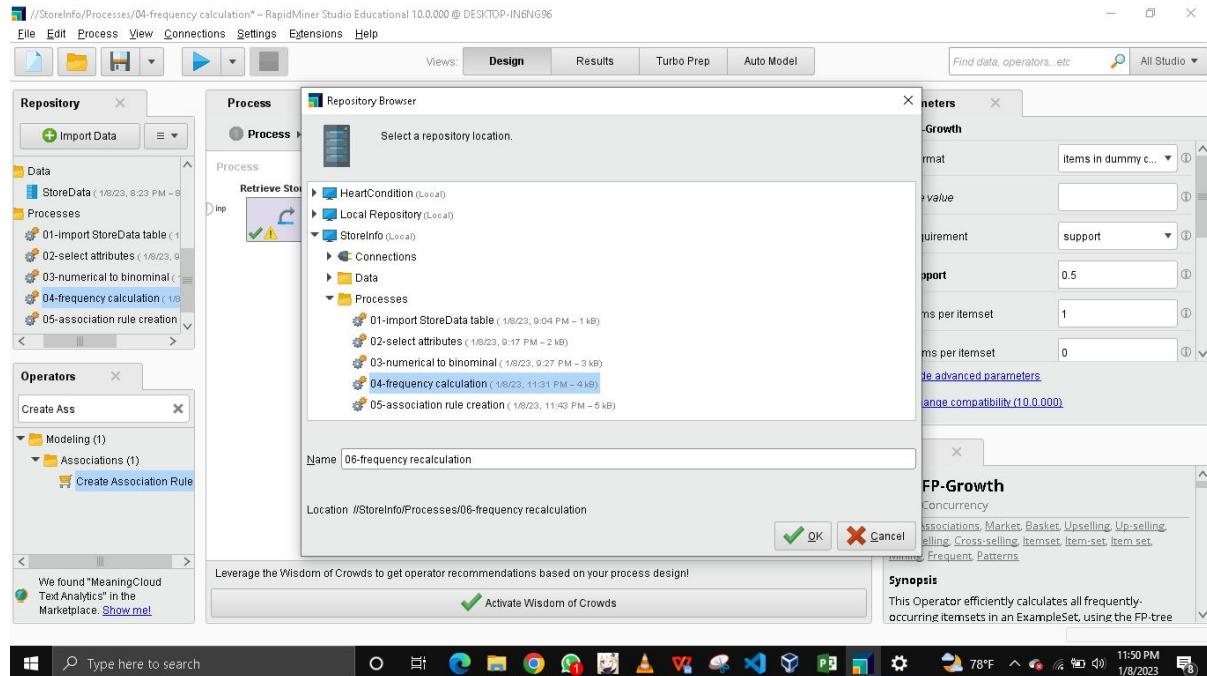
Min. Size: 1 Max. Size: 3 Contains Item: Update View

Size	Support	Item 1	Item 2	Item 3
2	0.240	butter	sugar	
2	0.400	butter	flour	
2	0.260	butter	grapes	
2	0.160	milk	apples	
2	0.280	milk	sugar	
2	0.120	milk	grapes	
2	0.220	milk	cornflakes	
2	0.120	apples	sugar	
2	0.140	apples	flour	
2	0.180	apples	grapes	
2	0.220	sugar	flour	
2	0.180	flour	grapes	
3	0.120	butter	apples	sugar
3	0.140	butter	apples	flour
3	0.220	butter	sugar	flour
3	0.180	butter	flour	grapes

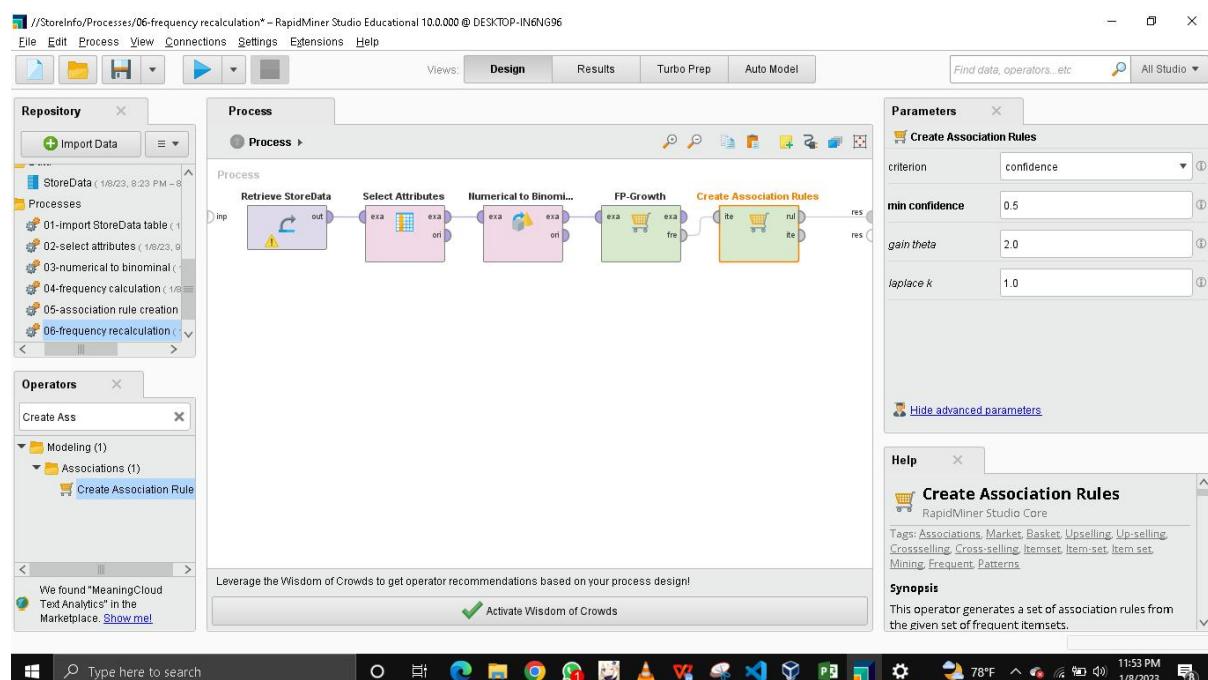
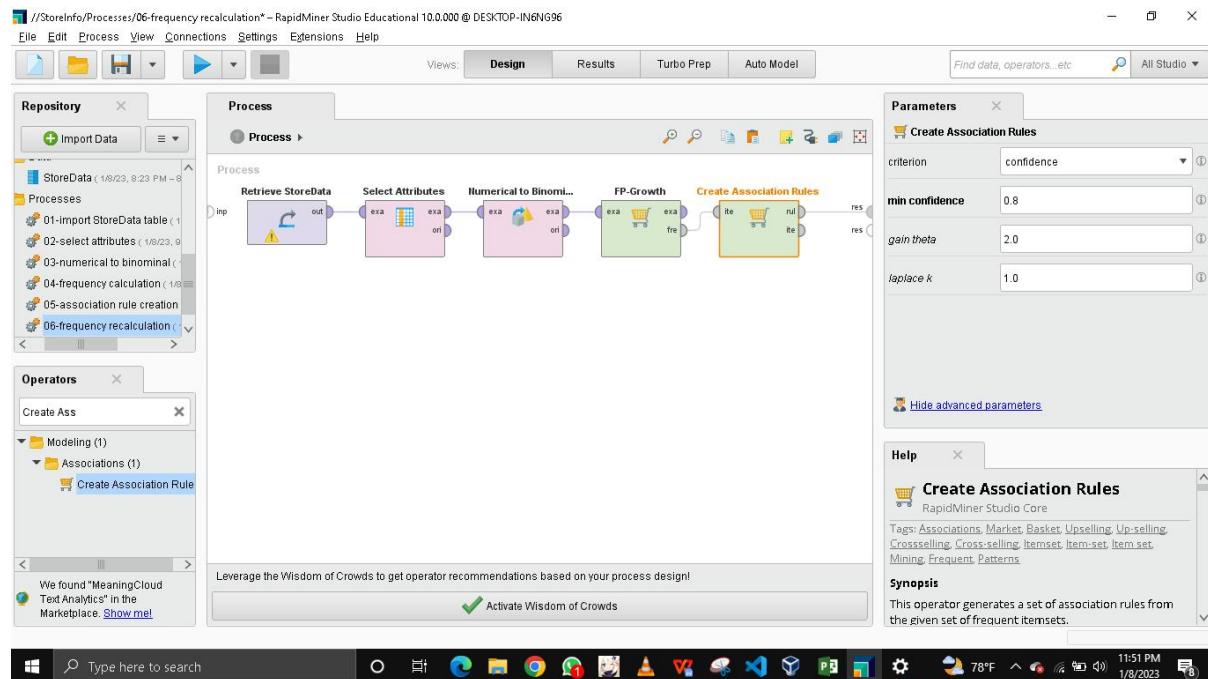
Repository Import Data

- Training Resources (connected)
- Samples
- Community Samples (connected)
- HeartCondition (Local)
 - Connections
 - Data
 - Processes
- Local Repository (Local)
- StoreInfo (Local)
 - Connections
 - Data
 - StoreData (1/8/23, 8:23 PM - 8 kB)
 - Processes
 - 01-import StoreData table (1/8/23, 9:04 PM - 1 kB)
 - 02-select attributes (1/8/23, 9:17 PM - 2 kB)
 - 03-numerical to binomial (1/8/23, 9:27 PM - 3 kB)
 - 04-frequency calculation (1/8/23, 11:31 PM - 4 kB)
 - 05-association rule creation (1/8/23, 11:43 PM - 5 kB)
- DB (Legacy)

11:49 PM 1/8/2023



8. Used the “06-frequency recalculation” process and added the “Create Association Rules” operator, this time decreasing the minimum confidence value from 0.8 to 0.5 in the operator to increase the number of association rules generated for each of the aforementioned frequented item sets and saved the process as “07-association rule recreation”



The screenshot shows the RapidMiner Studio interface with the following details:

- Title Bar:** //StoreInfo/Processes/06-frequency recalculation* - RapidMiner Studio Educational 10.0.000 @ DESKTOP-IN6NG96
- Menu Bar:** File, Edit, Process, View, Connections, Settings, Extensions, Help
- Toolbar:** Views: Design, Results, Turbo Prep, Auto Model
- Search Bar:** Find data, operators, etc.
- Result History:** Shows the 'AssociationRules' process (Create Association Rules) and the 'ExampleSet (/StoreInfo/Data/StoreData)'.
- Data View (Left):** Contains icons for Data, Graph, Description, and Annotations. The Data icon is selected.
- AssociationRules Process View:** Shows the rule table with columns: No., Premises, and Conclusion. The rules listed are:
 - No. 9: flour - Conclusion: sugar
 - No. 10: flour - Conclusion: butter, sugar
 - No. 11: butter, flour - Conclusion: sugar
 - No. 12: butter, apples - Conclusion: flour
 - No. 13: sugar - Conclusion: milk
 - No. 14: grapes - Conclusion: butter
 - No. 15: butter, grapes - Conclusion: flour
 - No. 16: butter - Conclusion: flour
 - No. 17: cornflakes - Conclusion: milk
 - No. 18: butter, sugar - Conclusion: flour
 - No. 19: flour - Conclusion: butter
 - No. 20: apples, sugar - Conclusion: butter
 - No. 21: apples, flour - Conclusion: butter
 - No. 22: sugar, flour - Conclusion: butter
 - No. 23: flour, grapes - Conclusion: butter
- Min. Criterion:** confidence
- Min. Criterion Value:** 0.000000
- Repository View (Right):** Shows a tree structure of available resources:
 - Training Resources (connected)
 - Samples
 - Community Samples (connected)
 - HeartCondition (Local)
 - Connections
 - Data
 - Processes
 - Local Repository (Local)
 - StoreInfo (Local)
 - Connections
 - Data
 - StoreData (1/8/23, 8:23 PM - 6 kB)
 - Processes
 - 01-import StoreData table (1/8/23, 9:04 PM - 1 kB)
 - 02-select attributes (1/8/23, 9:17 PM - 2 kB)
 - 03-numerical to binomial (1/8/23, 9:27 PM - 3 kB)
 - 04-frequency calculation (1/8/23, 11:31 PM - 4 kB)
 - 05-association rule creation (1/8/23, 11:43 PM - 5 kB)
 - 06-frequency recalculation (1/8/23, 11:50 PM - 4 kB)
 - DB (Legacy)

