

**Business Intelligence Technologies: Practical 3: Clustering**

**Weighting: 4% Due date: 23.59pm Friday 25<sup>th</sup> November 2022**

Carry out cluster analysis using k-means on a subset of the Absenteeism database. Explain your processes, including data pre-processing choices and rationale for parameters used for the clustering algorithm. Evaluate the results, highlighting and explaining any significant findings.

**Assessment Criteria:**

Correct and fully documented processes showing sound understanding of principles of clustering: 2 marks

Documentation and evaluation of results: 2 marks

## Absenteeism at work Data Set

**Abstract:** The database was created with records of absenteeism at work from July 2007 to July 2010 at a courier company in Brazil. Total number of instances: 740

**Source:** Creators original owner and donors: Andrea Martiniano (1), Ricardo Pinto Ferreira (2), and Renato Jose Sassi (3).

### Attribute Information:

1. Individual identification (ID)
2. Reason for absence (ICD).

Absences attested by the International Code of Diseases (ICD) stratified into 21 categories (I to XXI) as follows:

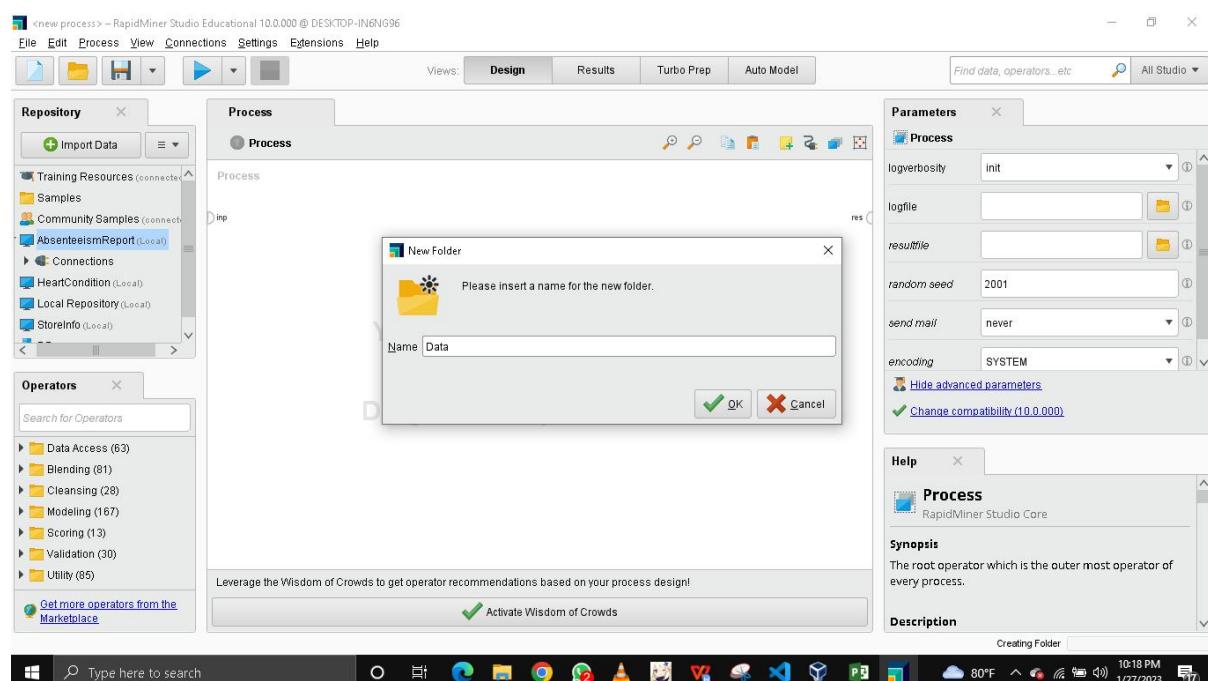
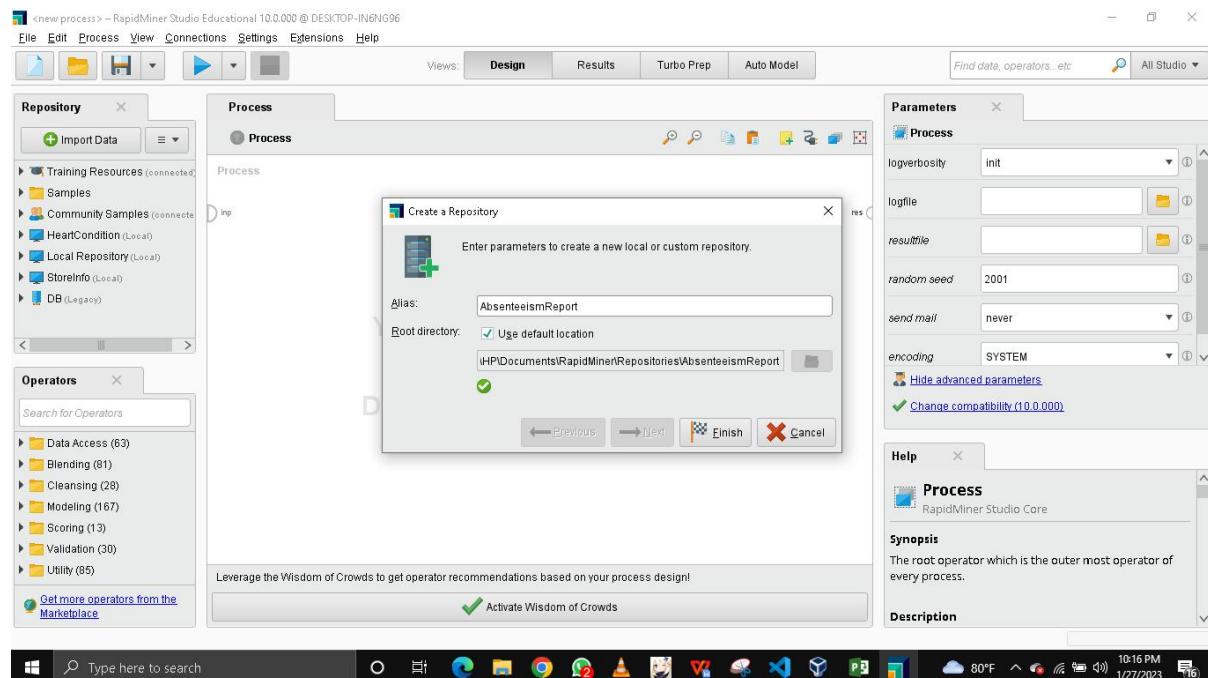
- I Certain infectious and parasitic diseases
- II Neoplasms
- III Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
- IV Endocrine, nutritional and metabolic diseases
- V Mental and behavioural disorders
- VI Diseases of the nervous system
- VII Diseases of the eye and adnexa
- VIII Diseases of the ear and mastoid process
- IX Diseases of the circulatory system
- X Diseases of the respiratory system
- XI Diseases of the digestive system
- XII Diseases of the skin and subcutaneous tissue
- XIII Diseases of the musculoskeletal system and connective tissue
- XIV Diseases of the genitourinary system
- XV Pregnancy, childbirth and the puerperium
- XVI Certain conditions originating in the perinatal period
- XVII Congenital malformations, deformations and chromosomal abnormalities
- XVIII Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
- XIX Injury, poisoning and certain other consequences of external causes
- XX External causes of morbidity and mortality
- XXI Factors influencing health status and contact with health services.

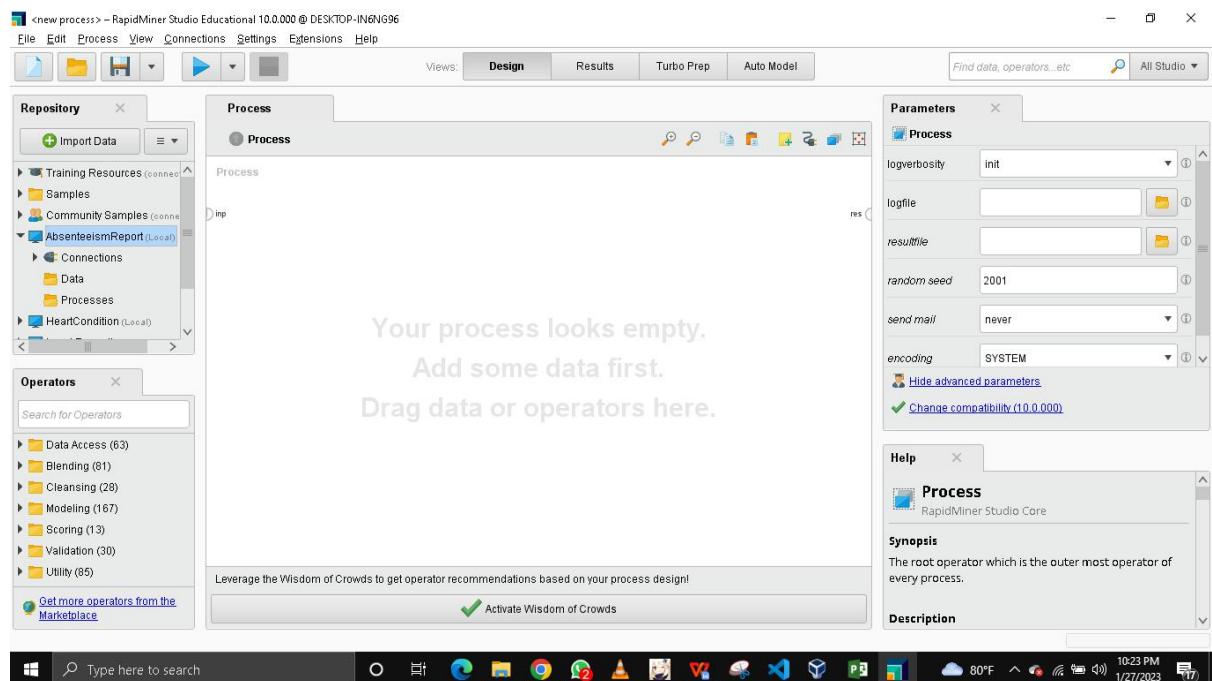
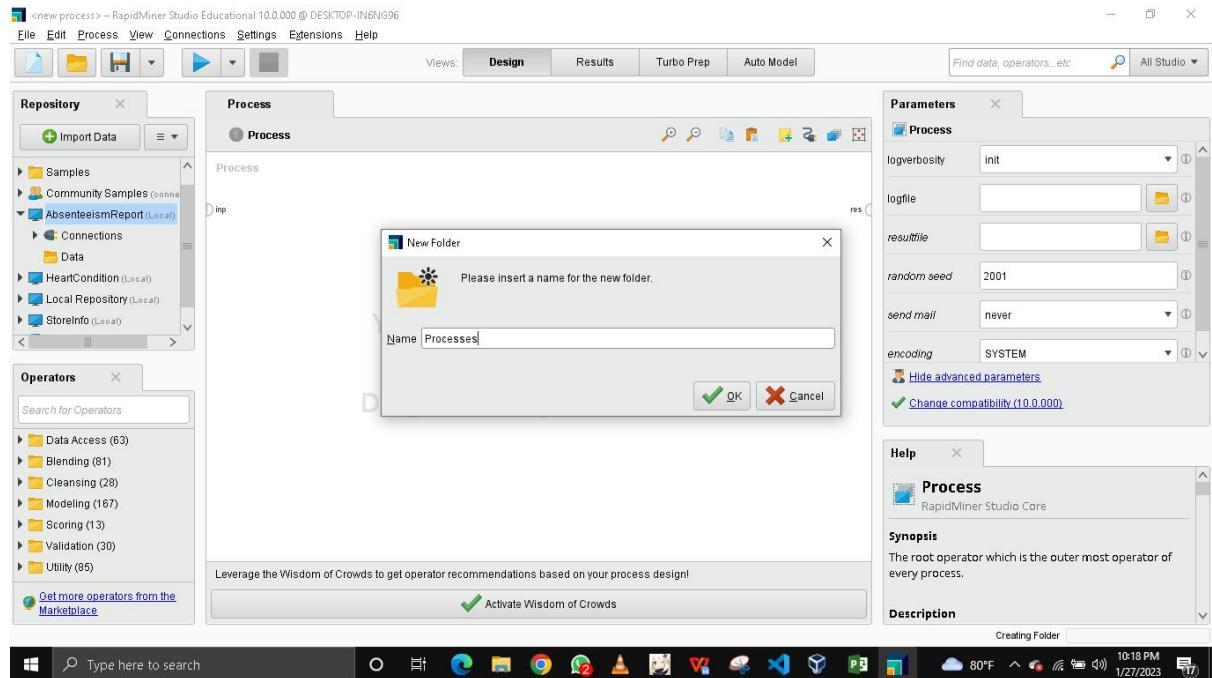
And 7 categories without (CID) patient follow-up (22), medical consultation (23), blood donation (24), laboratory examination (25), unjustified absence (26), physiotherapy (27), dental consultation (28).

3. Month of absence
4. Day of the week (Monday (2), Tuesday (3), Wednesday (4), Thursday (5), Friday (6))
5. Seasons (summer (1), autumn (2), winter (3), spring (4))
6. Transportation expense
7. Distance from Residence to Work (kilometers)
8. Service time (years)
9. Age
10. Work load Average/day
11. Hit target
12. Disciplinary failure (yes=1; no=0)
13. Education (high school (1), graduate (2), postgraduate (3), master and doctor (4))
14. Son (number of children)
15. Social drinker (yes=1; no=0)
16. Social smoker (yes=1; no=0)
17. Pet (number of pet)
18. Weight
19. Height
20. Body mass index
21. Absenteeism time in hours

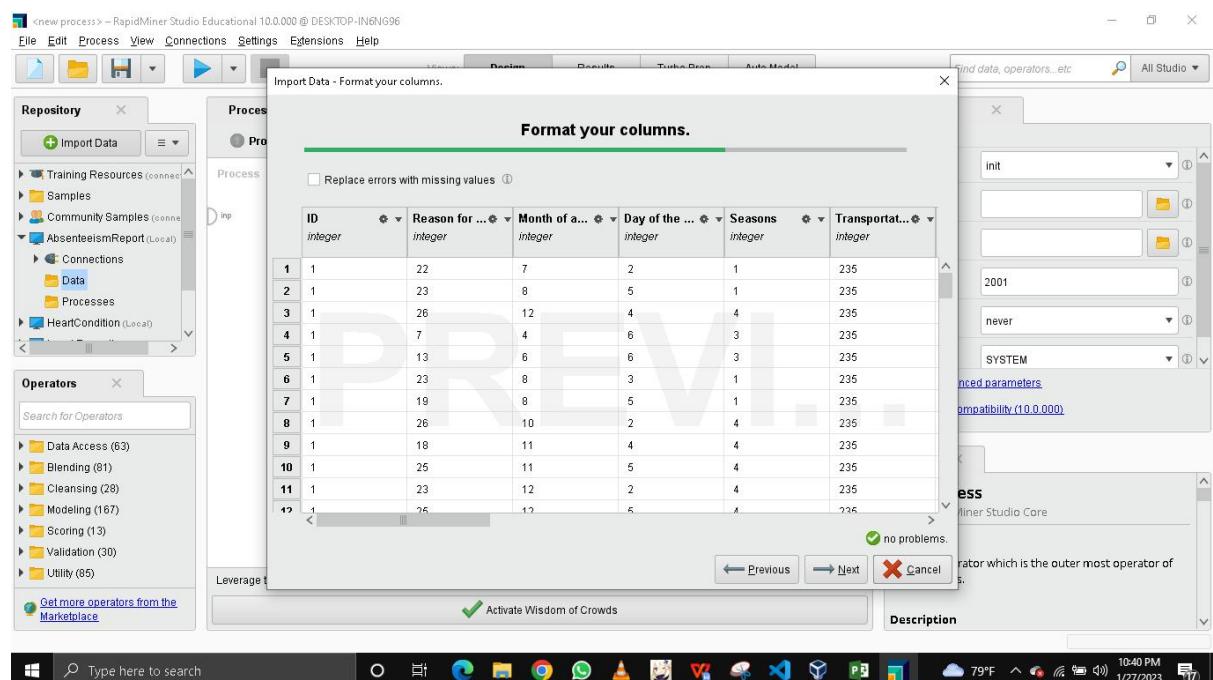
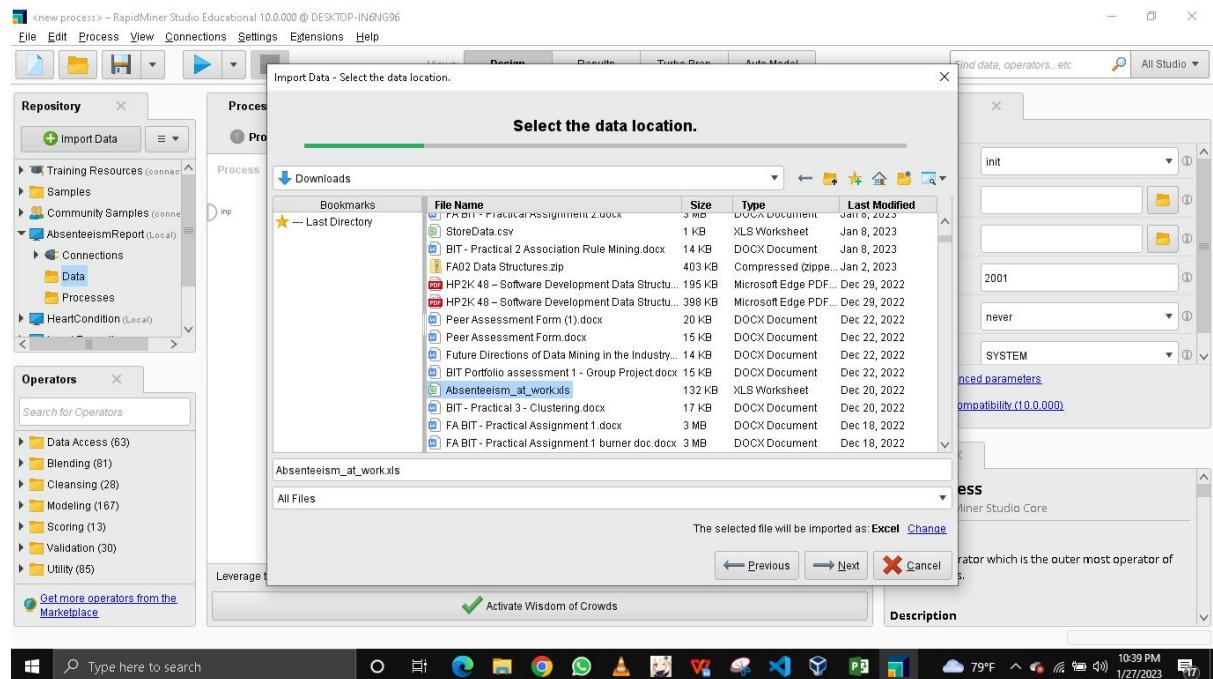
## Practical Documentation

1. Created a Repository labeled “AbsenteeismReport” and created two subfolders named “Data” and “Processes” respectively





2. Imported the Absenteeism\_at\_work table into the Data folder of the AbsenteeismReport Repository within the table's default attributes unchanged as shown below



<new process> – RapidMiner Studio Educational 10.0.000 @ DESKTOP-IN6NG96

File Edit Process View Connections Settings Extensions Help

Repository X Import Data

Process X

Import Data - Format your columns.

Format your columns.

Replace errors with missing values

|    | Transportat... * | Distance fr... * | Service time * | Age *   | Work load ... * | Hit target * |
|----|------------------|------------------|----------------|---------|-----------------|--------------|
|    | integer          | integer          | integer        | integer | integer         | integer      |
| 1  | 235              | 11               | 14             | 37      | 239554          | 97           |
| 2  | 235              | 11               | 14             | 37      | 205917          | 92           |
| 3  | 235              | 11               | 14             | 37      | 261306          | 97           |
| 4  | 235              | 11               | 14             | 37      | 326452          | 96           |
| 5  | 235              | 11               | 14             | 37      | 377550          | 94           |
| 6  | 235              | 11               | 14             | 37      | 265615          | 94           |
| 7  | 235              | 11               | 14             | 37      | 265615          | 94           |
| 8  | 235              | 11               | 14             | 37      | 265017          | 88           |
| 9  | 235              | 11               | 14             | 37      | 284031          | 97           |
| 10 | 235              | 11               | 14             | 37      | 284031          | 97           |
| 11 | 235              | 11               | 14             | 37      | 236629          | 93           |
| 12 | 235              | 11               | 14             | 37      | 236629          | 93           |

no problems.

Previous Next Cancel

Activate Wisdom of Crowds Description

Leverage

Find data, operators...etc All Studio

init

2001

never

SYSTEM

anced parameters

compatibility (10.0.000)

ESS

RapidMiner Studio Core

operator which is the outer most operator of

10:40 PM 1/27/2023

<new process> – RapidMiner Studio Educational 10.0.000 @ DESKTOP-IN6NG96

File Edit Process View Connections Settings Extensions Help

Repository X Import Data

Process X

Import Data - Format your columns.

Format your columns.

Replace errors with missing values

|    | Hit target * | Disciplinar... * | Education * | Son *   | Social drin... * | Social smo... * |
|----|--------------|------------------|-------------|---------|------------------|-----------------|
|    | integer      | integer          | integer     | integer | integer          | integer         |
| 1  | 97           | 0                | 3           | 1       | 0                | 0               |
| 2  | 92           | 0                | 3           | 1       | 0                | 0               |
| 3  | 97           | 0                | 3           | 1       | 0                | 0               |
| 4  | 96           | 0                | 3           | 1       | 0                | 0               |
| 5  | 94           | 0                | 3           | 1       | 0                | 0               |
| 6  | 94           | 0                | 3           | 1       | 0                | 0               |
| 7  | 94           | 0                | 3           | 1       | 0                | 0               |
| 8  | 88           | 0                | 3           | 1       | 0                | 0               |
| 9  | 97           | 0                | 3           | 1       | 0                | 0               |
| 10 | 97           | 0                | 3           | 1       | 0                | 0               |
| 11 | 93           | 0                | 3           | 1       | 0                | 0               |
| 12 | 92           | 0                | 3           | 1       | 0                | 0               |

no problems.

Previous Next Cancel

Activate Wisdom of Crowds Description

Leverage

Find data, operators...etc All Studio

init

2001

never

SYSTEM

anced parameters

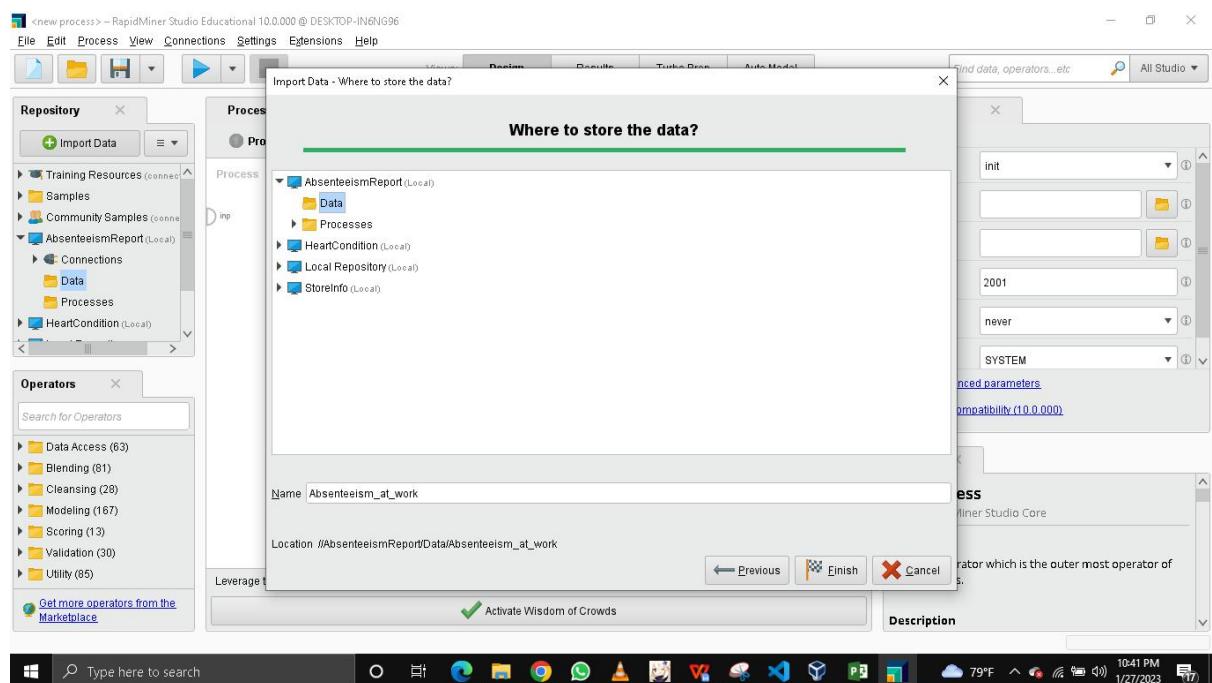
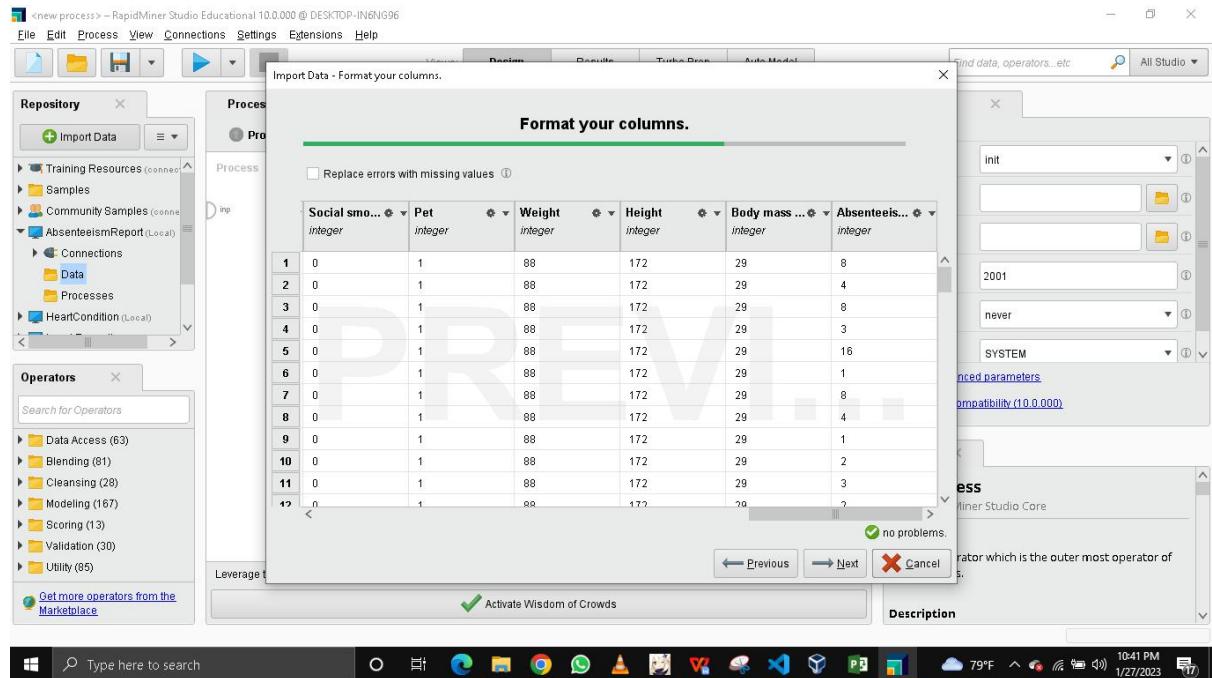
compatibility (10.0.000)

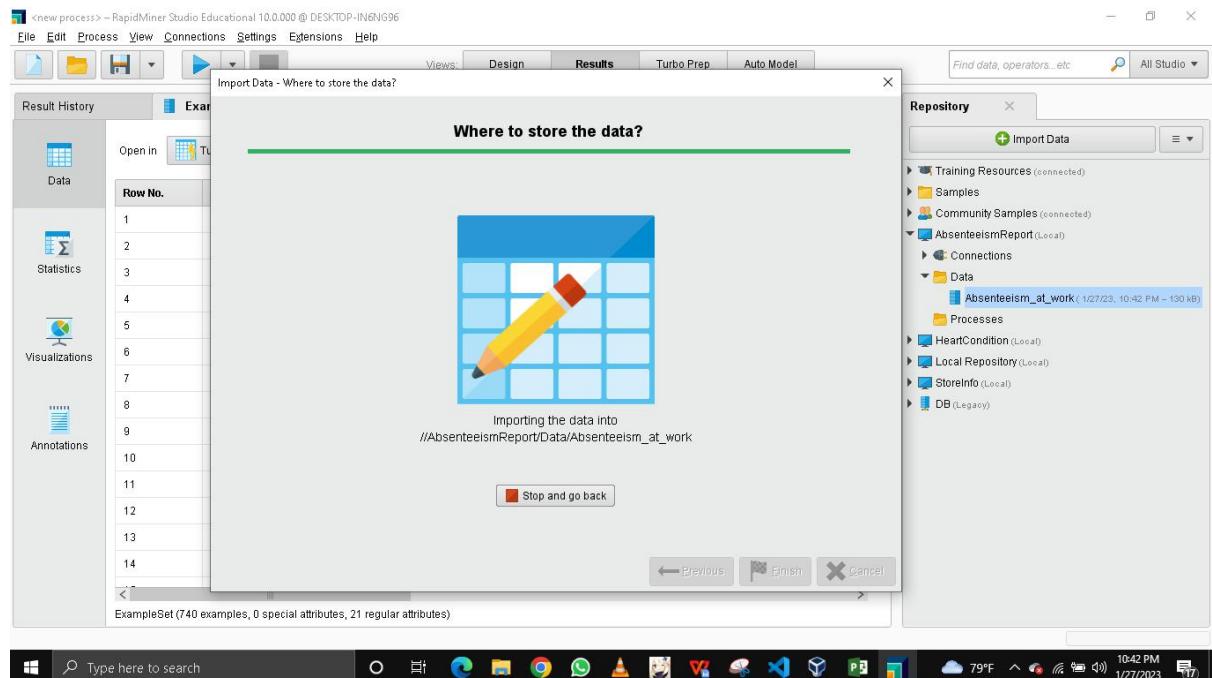
ESS

RapidMiner Studio Core

operator which is the outer most operator of

10:41 PM 1/27/2023





Below are the results of the completed import with the ranges of the values within the Absenteeism\_at\_work table shown for further detail

| Row No. | ID | Reason for ... | Month of ab... | Day of the w... | Seasons | Transportat... | Distance fro... | Service |
|---------|----|----------------|----------------|-----------------|---------|----------------|-----------------|---------|
| 1       | 1  | 22             | 7              | 2               | 1       | 235            | 11              | 14      |
| 2       | 1  | 23             | 8              | 5               | 1       | 235            | 11              | 14      |
| 3       | 1  | 26             | 12             | 4               | 4       | 235            | 11              | 14      |
| 4       | 1  | 7              | 4              | 6               | 3       | 235            | 11              | 14      |
| 5       | 1  | 13             | 6              | 6               | 3       | 235            | 11              | 14      |
| 6       | 1  | 23             | 8              | 3               | 1       | 235            | 11              | 14      |
| 7       | 1  | 19             | 8              | 5               | 1       | 235            | 11              | 14      |
| 8       | 1  | 26             | 10             | 2               | 4       | 235            | 11              | 14      |
| 9       | 1  | 18             | 11             | 4               | 4       | 235            | 11              | 14      |
| 10      | 1  | 25             | 11             | 5               | 4       | 235            | 11              | 14      |
| 11      | 1  | 23             | 12             | 2               | 4       | 235            | 11              | 14      |
| 12      | 1  | 25             | 12             | 5               | 4       | 235            | 11              | 14      |
| 13      | 1  | 14             | 12             | 3               | 2       | 235            | 11              | 14      |
| 14      | 1  | 13             | 1              | 3               | 2       | 235            | 11              | 14      |

ExampleSet (740 examples, 0 special attributes, 21 regular attributes)

Repository

Absenteeism\_at\_work

Data table  
Number of examples = 740  
21 attributes:

| Role            | Name      | Type          | Range | Missing |
|-----------------|-----------|---------------|-------|---------|
| ID              | # integer | = [1 - 36]    | = 0   |         |
| Reason for...   | # integer | = [0 - 28]    | = 0   |         |
| Month of ab...  | # integer | = [0 - 12]    | = 0   |         |
| Day of the w... | # integer | = [2 - 6]     | = 0   |         |
| Seasons         | # integer | = [1 - 4]     | = 0   |         |
| Transportat...  | # integer | = [118 - 388] | = 0   |         |

**RapidMiner Studio Educational 10.0.000 @ DESKTOP-IN6NG96**

File Edit Process View Connections Settings Extensions Help

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

**Result History** ExampleSet //AbsenteeismReport/Data/Absenteeism\_at\_work

Data Statistics Visualizations Annotations

Open in Turbo Prep Auto Model Filter (740 / 740 examples): all

| Distance from home | Service time | Age | Work load at work | Hit target | Disciplinary action | Education | Son | Social drinker |
|--------------------|--------------|-----|-------------------|------------|---------------------|-----------|-----|----------------|
| 11                 | 14           | 37  | 239554            | 97         | 0                   | 3         | 1   | 0              |
| 11                 | 14           | 37  | 205917            | 92         | 0                   | 3         | 1   | 0              |
| 11                 | 14           | 37  | 261306            | 97         | 0                   | 3         | 1   | 0              |
| 11                 | 14           | 37  | 326452            | 96         | 0                   | 3         | 1   | 0              |
| 11                 | 14           | 37  | 377550            | 94         | 0                   | 3         | 1   | 0              |
| 11                 | 14           | 37  | 265615            | 94         | 0                   | 3         | 1   | 0              |
| 11                 | 14           | 37  | 265615            | 94         | 0                   | 3         | 1   | 0              |
| 11                 | 14           | 37  | 265017            | 88         | 0                   | 3         | 1   | 0              |
| 11                 | 14           | 37  | 284031            | 97         | 0                   | 3         | 1   | 0              |
| 11                 | 14           | 37  | 284031            | 97         | 0                   | 3         | 1   | 0              |
| 11                 | 14           | 37  | 236629            | 93         | 0                   | 3         | 1   | 0              |
| 11                 | 14           | 37  | 236629            | 93         | 0                   | 3         | 1   | 0              |
| 11                 | 14           | 37  | 236629            | 93         | 0                   | 3         | 1   | 0              |
| 11                 | 14           | 37  | 330061            | 100        | 0                   | 3         | 1   | 0              |

ExampleSet (740 examples, 0 special attributes, 21 regular attributes)

**Repository**

Import Data

- Training Resources (connected)
- Samples
- Community Samples (connected)
- AbsenteeismReport (Local)
  - Connections
  - Data

Absenteeism\_at\_work

Data table Number of examples = 740 21 attributes:

| Role               | Name      | Type             | Range | Missing |
|--------------------|-----------|------------------|-------|---------|
| Transportation     | # integer | =[18 - 380]      | = 0   |         |
| Distance from home | # integer | = [5 - 52]       | = 0   |         |
| Service time       | # integer | = [1 - 29]       | = 0   |         |
| Age                | # integer | = [27 - 58]      | = 0   |         |
| Work load at work  | # integer | = [205917 - ...] | = 0   |         |
| Hit target         | # integer | = [81 - 100]     | = 0   |         |

Press 'F3' for focus.



**RapidMiner Studio Educational 10.0.000 @ DESKTOP-IN6NG96**

File Edit Process View Connections Settings Extensions Help

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

**Result History** ExampleSet //AbsenteeismReport/Data/Absenteeism\_at\_work

Data Statistics Visualizations Annotations

Open in Turbo Prep Auto Model Filter (740 / 740 examples): all

| tion | Son | Social drinker | Social smok... | Pet | Weight | Height | Body mass i... | Absenteeis... |
|------|-----|----------------|----------------|-----|--------|--------|----------------|---------------|
| 1    | 0   | 0              | 1              | 88  | 172    | 29     | 8              |               |
| 1    | 0   | 0              | 1              | 88  | 172    | 29     | 4              |               |
| 1    | 0   | 0              | 1              | 88  | 172    | 29     | 8              |               |
| 1    | 0   | 0              | 1              | 88  | 172    | 29     | 3              |               |
| 1    | 0   | 0              | 1              | 88  | 172    | 29     | 16             |               |
| 1    | 0   | 0              | 1              | 88  | 172    | 29     | 1              |               |
| 1    | 0   | 0              | 1              | 88  | 172    | 29     | 8              |               |
| 1    | 0   | 0              | 1              | 88  | 172    | 29     | 4              |               |
| 1    | 0   | 0              | 1              | 88  | 172    | 29     | 1              |               |
| 1    | 0   | 0              | 1              | 88  | 172    | 29     | 2              |               |
| 1    | 0   | 0              | 1              | 88  | 172    | 29     | 3              |               |
| 1    | 0   | 0              | 1              | 88  | 172    | 29     | 2              |               |
| 1    | 0   | 0              | 1              | 88  | 172    | 29     | 4              |               |
| 1    | 0   | 0              | 1              | 88  | 172    | 29     | 1              |               |

ExampleSet (740 examples, 0 special attributes, 21 regular attributes)

**Repository**

Import Data

- Training Resources (connected)
- Samples
- Community Samples (connected)
- AbsenteeismReport (Local)
  - Connections
  - Data

Absenteeism\_at\_work

Data table Number of examples = 740 21 attributes:

| Role            | Name      | Type         | Range | Missing |
|-----------------|-----------|--------------|-------|---------|
| Hit target      | # integer | = [81 - 100] | = 0   |         |
| Disciplinary... | # integer | = [0 - 1]    | = 0   |         |
| Education       | # integer | = [1 - 4]    | = 0   |         |
| Son             | # integer | = [0 - 4]    | = 0   |         |
| Social drin...  | # integer | = [0 - 1]    | = 0   |         |
| Social smo...   | # integer | = [0 - 1]    | = 0   |         |

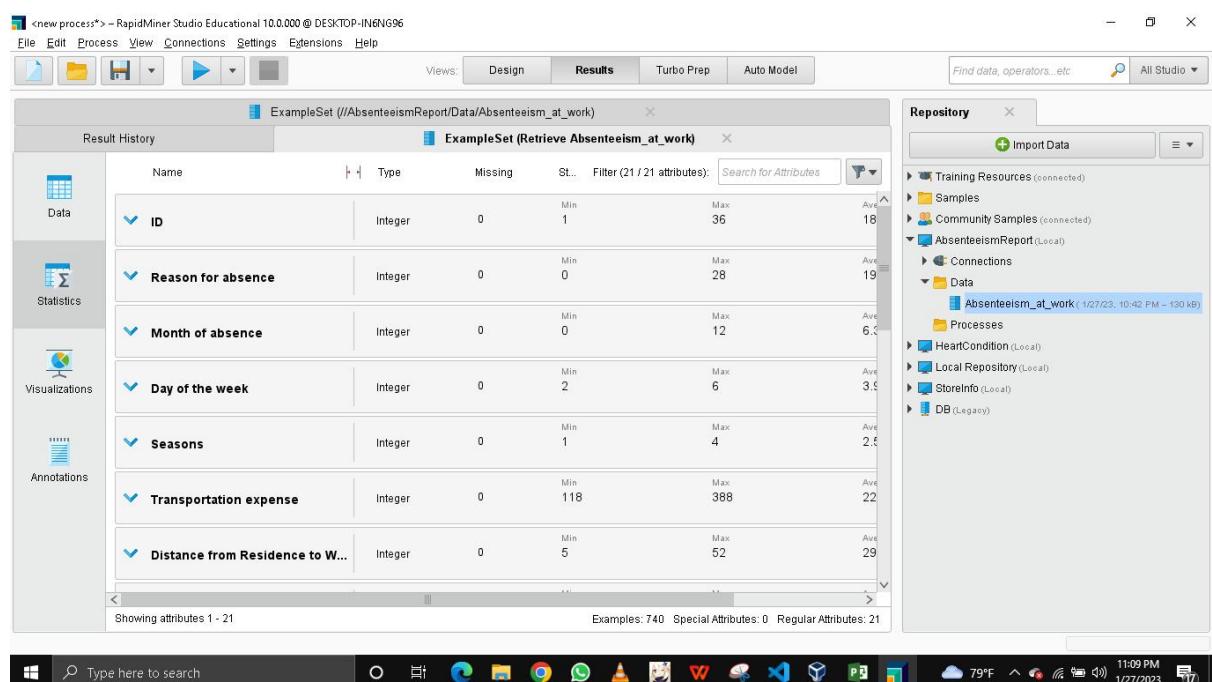
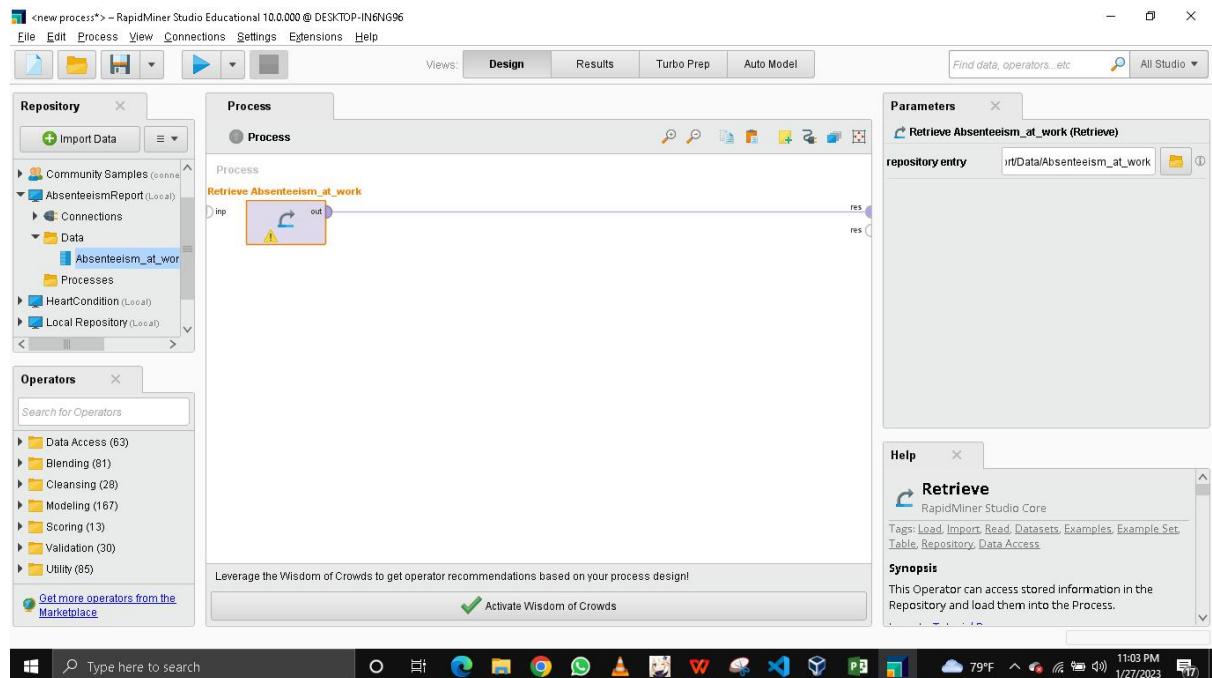
Press 'F3' for focus.



The screenshot shows the RapidMiner Studio interface. The main window has tabs for 'Views', 'Design', 'Results', 'Turbo Prep', and 'Auto Model'. The 'Results' tab is active, displaying the 'Absenteeism\_at\_work' dataset. The dataset table has columns: 'tion', 'Son', 'Social drinker', 'Social smok...', 'Pet', 'Weight', 'Height', 'Body mass i...', and 'Absentees...'. Below the table, it says 'ExampleSet (740 examples, 0 special attributes, 21 regular attributes)'. To the right, the 'Repository' panel shows the dataset in the 'AbsenteeismReport' folder. The Windows taskbar at the bottom shows the date and time as 1/27/2023 at 10:56 PM.

3. Retrieved the Absenteeism\_at\_work table into the Process panel and checked for any missing values; saving the process as “01-import Absenteeism\_at\_work table” in the Processes subfolder in the AbsenteeismReport Repository

The screenshot shows the RapidMiner Studio interface with the 'Process' tab active. On the left, the 'Repository' panel shows the 'AbsenteeismReport' folder containing the 'Absenteeism\_at\_work' dataset. The 'Operators' panel lists various data access operators like Data Access, Blending, Cleansing, etc. In the center, the 'Process' panel shows a single 'Retrieve Absenteeism\_at\_work' operator with an 'inp' port connected to its input and an 'out' port connected to its output. The 'Parameters' panel shows the 'repository entry' set to 'it\\Data\\Absenteeism\_at\_work'. The 'Help' panel provides a synopsis of the Retrieve operator. The Windows taskbar at the bottom shows the date and time as 1/27/2023 at 10:59 PM.



<new process\*> – RapidMiner Studio Educational 10.0.000 @ DESKTOP-INGNG96

File Edit Process View Connections Settings Extensions Help

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

Result History ExampleSet (Retrieve Absenteeism\_at\_work)

| Name                            | Type    | Missing | St..       | Filter (21 / 21 attributes): | Search for Attributes |
|---------------------------------|---------|---------|------------|------------------------------|-----------------------|
| Distance from Residence to Work | Integer | 0       | Min 5      | Max 52                       | Ave 29                |
| Service time                    | Integer | 0       | Min 1      | Max 29                       | Ave 12                |
| Age                             | Integer | 0       | Min 27     | Max 58                       | Ave 36                |
| Work load Average/day           | Integer | 0       | Min 205917 | Max 378884                   | Ave 27                |
| Hit target                      | Integer | 0       | Min 81     | Max 100                      | Ave 94                |
| Disciplinary failure            | Integer | 0       | Min 0      | Max 1                        | Ave 0.0               |
| Education                       | Integer | 0       | Min 1      | Max 4                        | Ave 1.2               |

Showing attributes 1 - 21 Examples: 740 Special Attributes: 0 Regular Attributes: 21

Repository Import Data

- Training Resources (connected)
- Samples
- Community Samples (connected)
- AbsenteeismReport (Local)
  - Connections
  - Data
    - Absenteeism\_at\_work (1/27/23, 10:42 PM – 130 kB)
    - Processes
- HeartCondition (Local)
- Local Repository (Local)
- StoreInfo (Local)
- DB (Legacy)

11:10 PM 1/27/2023

<new process\*> – RapidMiner Studio Educational 10.0.000 @ DESKTOP-INGNG96

File Edit Process View Connections Settings Extensions Help

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

Result History ExampleSet (Retrieve Absenteeism\_at\_work)

| Name           | Type    | Missing | St..    | Filter (21 / 21 attributes): | Search for Attributes |
|----------------|---------|---------|---------|------------------------------|-----------------------|
| Education      | Integer | 0       | Min 1   | Max 4                        | Ave 1.2               |
| Son            | Integer | 0       | Min 0   | Max 4                        | Ave 1.0               |
| Social drinker | Integer | 0       | Min 0   | Max 1                        | Ave 0.5               |
| Social smoker  | Integer | 0       | Min 0   | Max 1                        | Ave 0.0               |
| Pet            | Integer | 0       | Min 0   | Max 8                        | Ave 0.7               |
| Weight         | Integer | 0       | Min 56  | Max 108                      | Ave 79                |
| Height         | Integer | 0       | Min 163 | Max 196                      | Ave 17                |

Showing attributes 1 - 21 Examples: 740 Special Attributes: 0 Regular Attributes: 21

Repository Import Data

- Training Resources (connected)
- Samples
- Community Samples (connected)
- AbsenteeismReport (Local)
  - Connections
  - Data
    - Absenteeism\_at\_work (1/27/23, 10:42 PM – 130 kB)
    - Processes
- HeartCondition (Local)
- Local Repository (Local)
- StoreInfo (Local)
- DB (Legacy)

11:10 PM 1/27/2023

<new process\*> – RapidMiner Studio Educational 10.0.000 @ DESKTOP-IN6NG96

File Edit Process View Connections Settings Extensions Help

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

**Result History**

**ExampleSet (Retrieve Absenteeism\_at\_work)**

| Name                      | Type    | Missing | St...   | Filter (21 / 21 attributes): | Search for Attributes |
|---------------------------|---------|---------|---------|------------------------------|-----------------------|
| Social drinker            | Integer | 0       | Min 0   | Max 1                        | Ave 0.5               |
| Social smoker             | Integer | 0       | Min 0   | Max 1                        | Ave 0.0               |
| Pet                       | Integer | 0       | Min 0   | Max 8                        | Ave 0.1               |
| Weight                    | Integer | 0       | Min 56  | Max 108                      | Ave 79                |
| Height                    | Integer | 0       | Min 163 | Max 196                      | Ave 176               |
| Body mass index           | Integer | 0       | Min 19  | Max 38                       | Ave 26                |
| Absenteeism time in hours | Integer | 0       | Min 0   | Max 120                      | Ave 6.5               |

Showing attributes 1 - 21 Examples: 740 Special Attributes: 0 Regular Attributes: 21

**Repository**

- Import Data
- Training Resources (connected)
- Samples
- Community Samples (connected)
- AbsenteeismReport (Local)
  - Connections
  - Data
    - Absenteeism\_at\_work (1/27/23, 10:42 PM – 130 kB)
  - Processes
- HeartCondition (Local)
- Local Repository (Local)
- StoreInfo (Local)
- DB (Legacy)

Windows Taskbar: Type here to search, Start button, Icons (Calculator, Mail, etc.), Cloud icon, 79°F, 11:11 PM, 1/27/2023, Battery icon

<new process\*> – RapidMiner Studio Educational 10.0.000 @ DESKTOP-IN6NG96

File Edit Process View Connections Settings Extensions Help

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

**Repository**

- Import Data
- Community Samples (connected)
- AbsenteeismReport (Local)
  - Connections
  - Data
    - Absenteeism\_at\_work
  - Processes
- HeartCondition (Local)
- Local Repository (Local)
- StoreInfo (Local)

**Process**

**Repository Browser**

Select a repository location.

**Operators**

Search for Operators

- Data Access (63)
- Blending (61)
- Cleansing (28)
- Modeling (167)
- Scoring (13)
- Validation (30)
- Utility (85)

Get more operators from the Marketplace

**Process**

Retrieve Absenteeism\_at\_work

inp

**Retrievers**

retrieve Absenteeism\_at\_work (Retrieve)

repository entry //AbsenteeismReport/Data/Absenteeism\_at\_work

**Process**

Name: 01-import Absenteeism\_at\_work table

Location: //AbsenteeismReport/Processes/01-import Absenteeism\_at\_work/table

OK Cancel

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

Activate Wisdom of Crowds

**Retrieve**

RapidMiner Studio Core

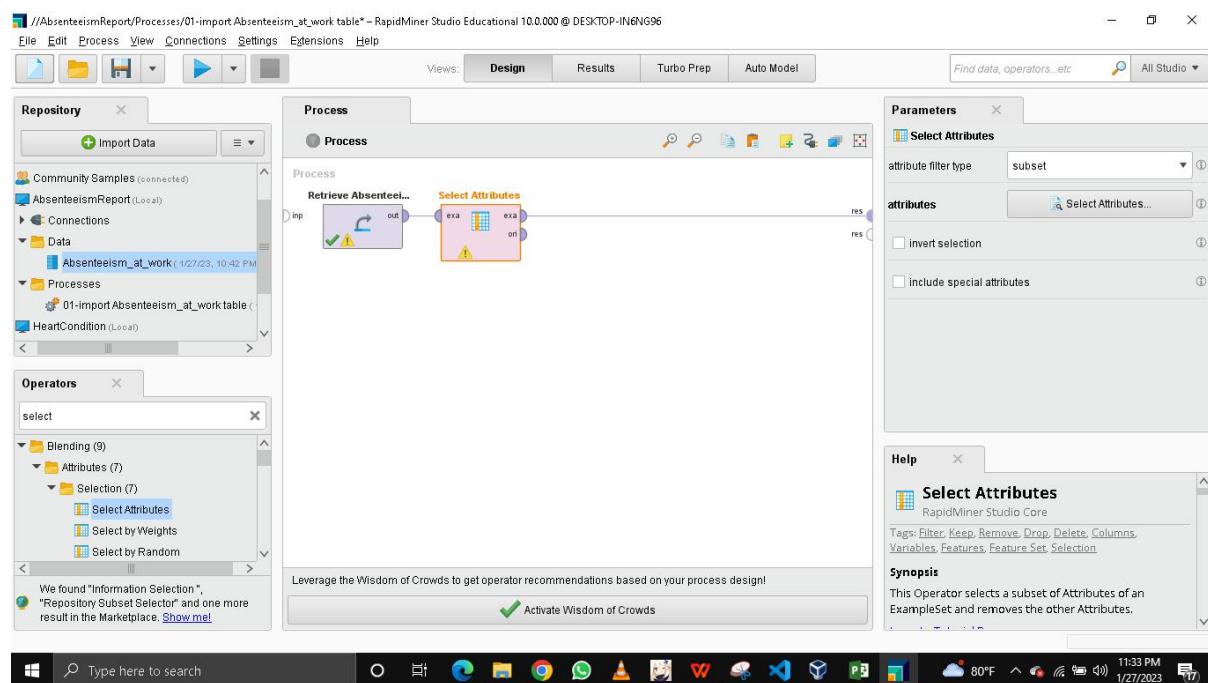
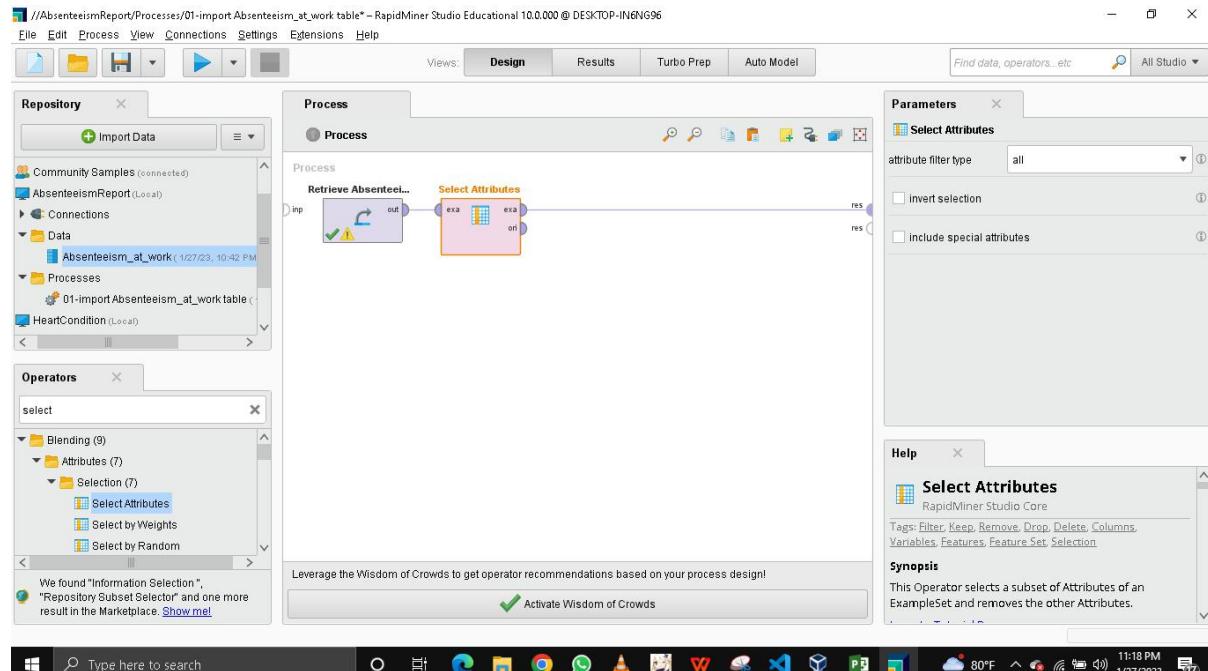
Add Import Read Datasets Examples Example Set Repository Data Access

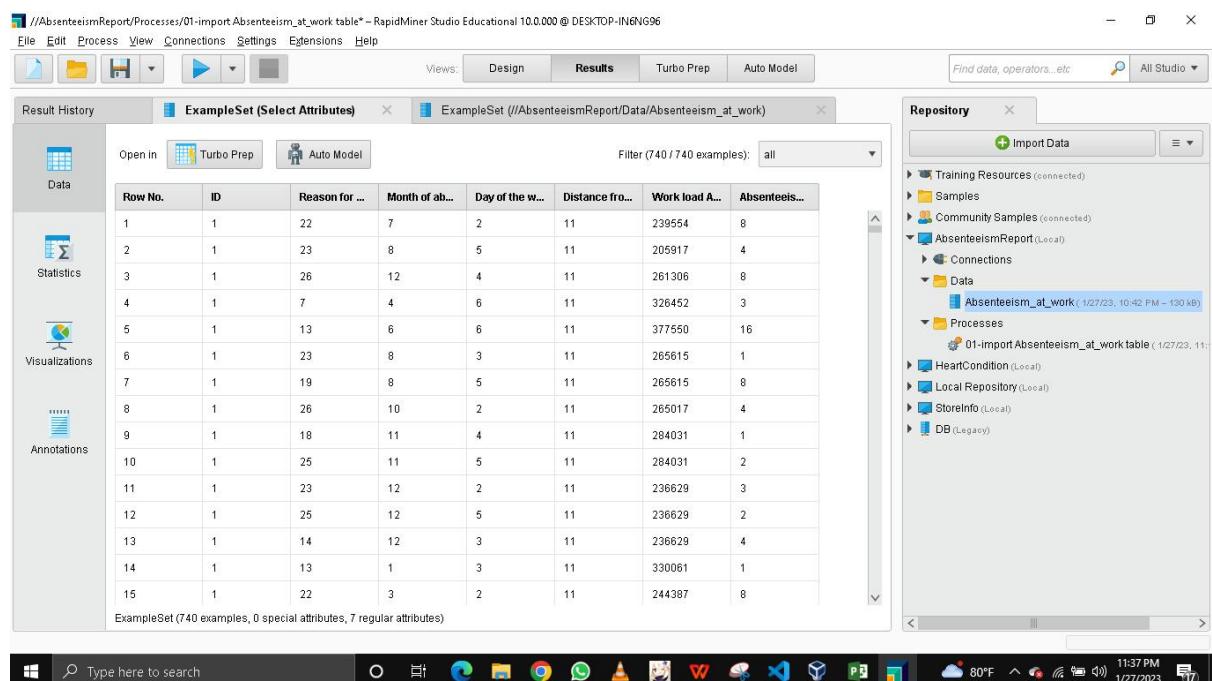
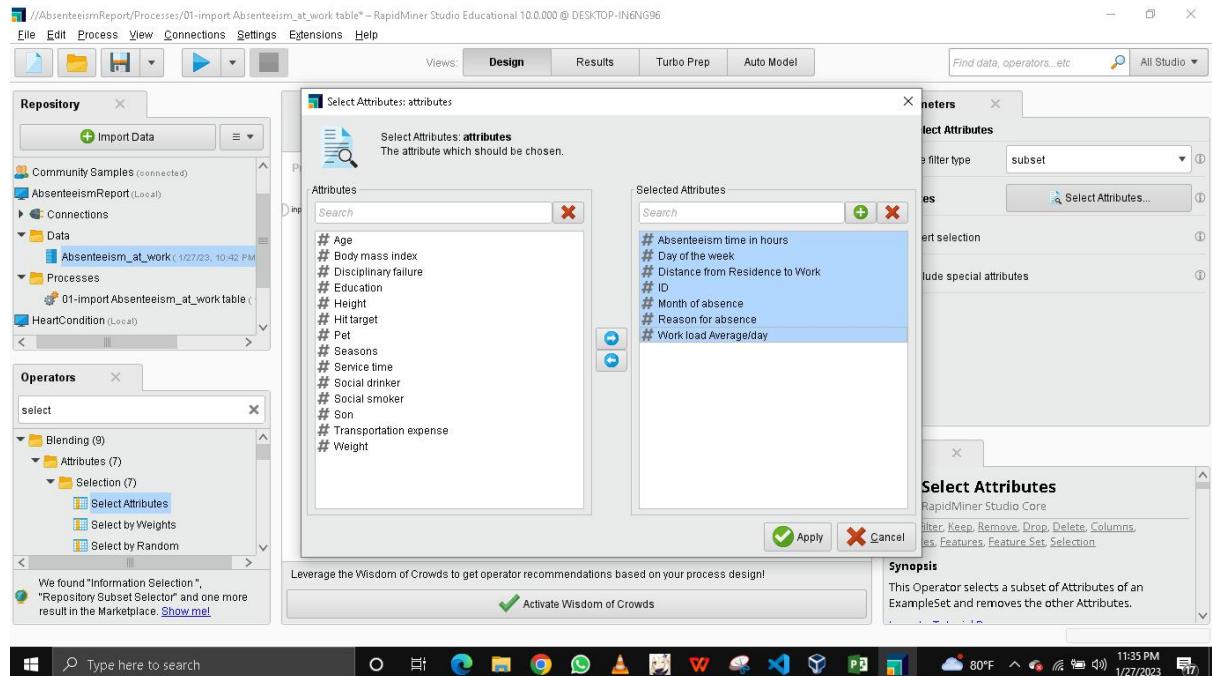
**Synopsis**

This Operator can access stored information in the Repository and load them into the Process.

Windows Taskbar: Type here to search, Start button, Icons (Calculator, Mail, etc.), Cloud icon, 80°F, 11:13 PM, 1/27/2023, Battery icon

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





//AbsenteeismReport/Processes/01-importAbsenteeism\_at\_work\_table - RapidMiner Studio Educational 10.0.000 @ DESKTOP-IN6NG96

File Edit Process View Connections Settings Extensions Help

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

**Result History**

**ExampleSet (Select Attributes)**

| Name                            | Type    | Missing | Statist...            | Filter (7 / 7 attributes) | Search for Attributes |
|---------------------------------|---------|---------|-----------------------|---------------------------|-----------------------|
| ID                              | Integer | 0       | Min 1 Max 36          | Average 18.01             |                       |
| Reason for absence              | Integer | 0       | Min 0 Max 28          | Average 19.21             |                       |
| Month of absence                | Integer | 0       | Min 0 Max 12          | Average 6.324             |                       |
| Day of the week                 | Integer | 0       | Min 2 Max 6           | Average 3.915             |                       |
| Distance from Residence to W... | Integer | 0       | Min 5 Max 52          | Average 29.63             |                       |
| Work load Average/day           | Integer | 0       | Min 205917 Max 378884 | Average 2714              |                       |
| Absenteeism time in hours       | Integer | 0       | Min 0 Max 120         | Average 6.924             |                       |

Showing attributes 1 - 7 Examples: 740 Special Attributes: 0 Regular Attributes: 7

**Repository**

- Import Data
- Training Resources (connected)
- Samples
- Community Samples (connected)
- AbsenteeismReport (Local)
  - Connections
  - Data
    - Absenteeism\_at\_work (1/27/23, 10:42 PM – 190 kB)
  - Processes
    - 01-importAbsenteeism\_at\_work\_table (1/27/23, 11:14 PM – 1 kB)
- HeartCondition (Local)
- Local Repository (Local)
- StoreInfo (Local)
- DB (Legacy)



//AbsenteeismReport/Processes/01-importAbsenteeism\_at\_work\_table - RapidMiner Studio Educational 10.0.000 @ DESKTOP-IN6NG96

File Edit Process View Connections Settings Extensions Help

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

**Repository**

**Repository Browser**

Select a repository location.

- Community Samples (connected)
- AbsenteeismReport (Local)
  - Connections
  - Data
    - Absenteeism\_at\_work (1/27/23, 10:42 PM – 190 kB)
  - Processes
    - 01-importAbsenteeism\_at\_work\_table (1/27/23, 11:14 PM – 1 kB)
- HeartCondition (Local)
- Local Repository (Local)
- StoreInfo (Local)

**Operators**

select

Blending (9)

Attributes (7)

Selection (7)

- Select Attributes
- Select by Weights
- Select by Random

We found "Information Selection", "Repository Subset Selector" and one more result in the Marketplace. Show me!

**Parameters**

Select Attributes

filter type: subset

Attributes: Select Attributes...

subset selection

Include special attributes

**Operators**

Name: 02-select attributes

Location: //AbsenteeismReport/Processes/02-select attributes

OK Cancel

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

Activate Wisdom of Crowds

**Select Attributes**

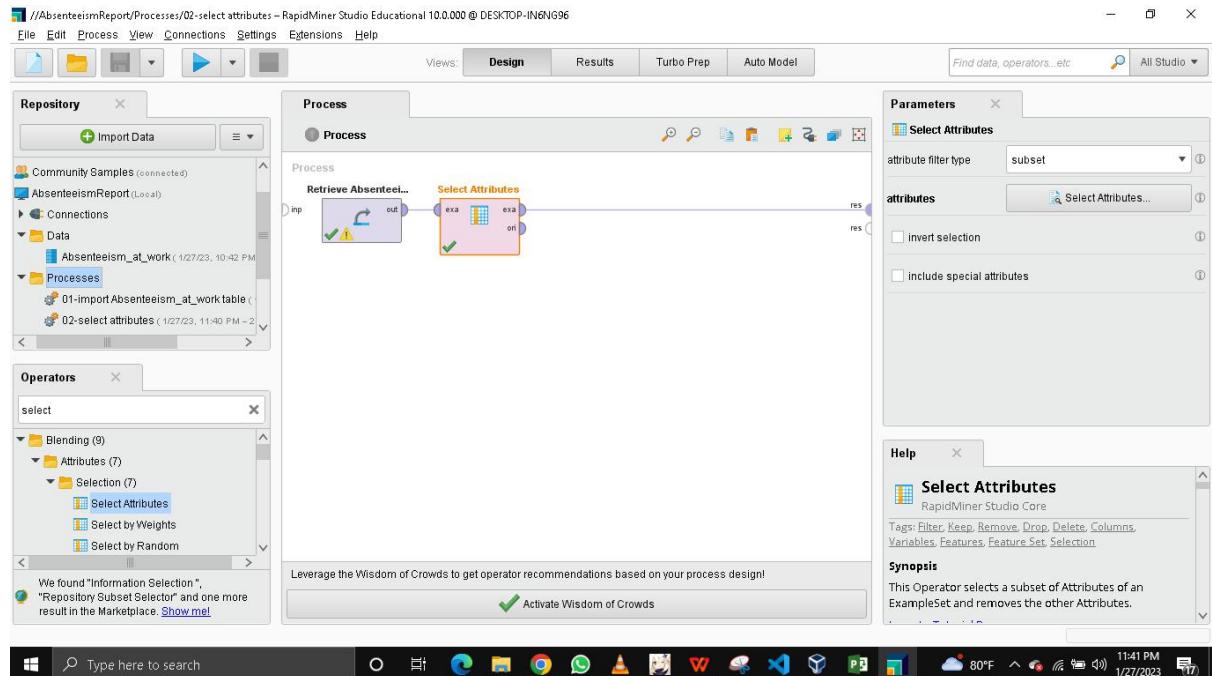
RapidMiner Studio Core

Filter, Keep, Remove, Drop, Delete, Columns, Row, Features, Feature Set, Selection

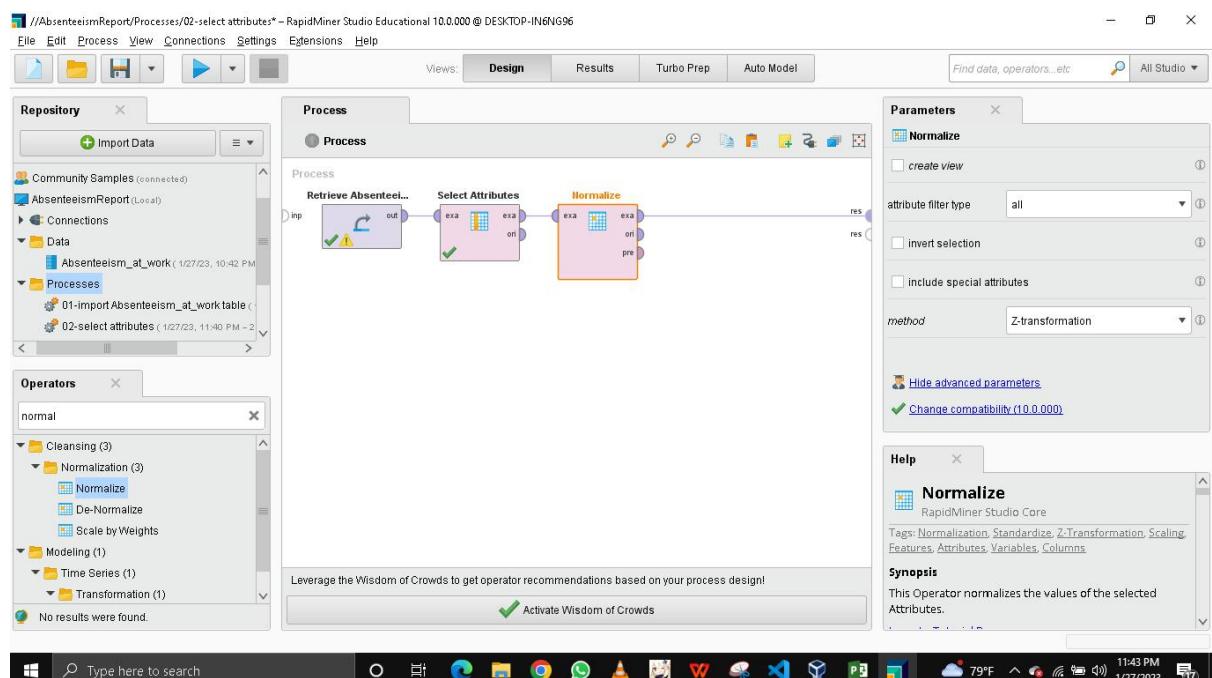
**Synopsis**

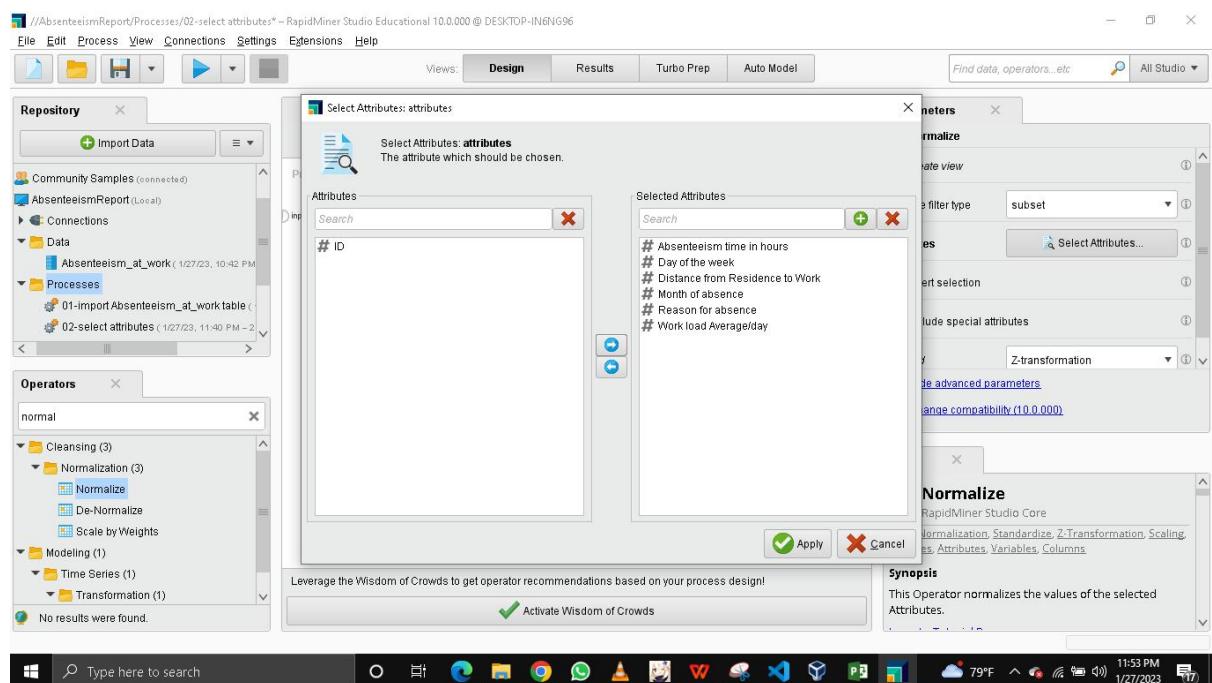
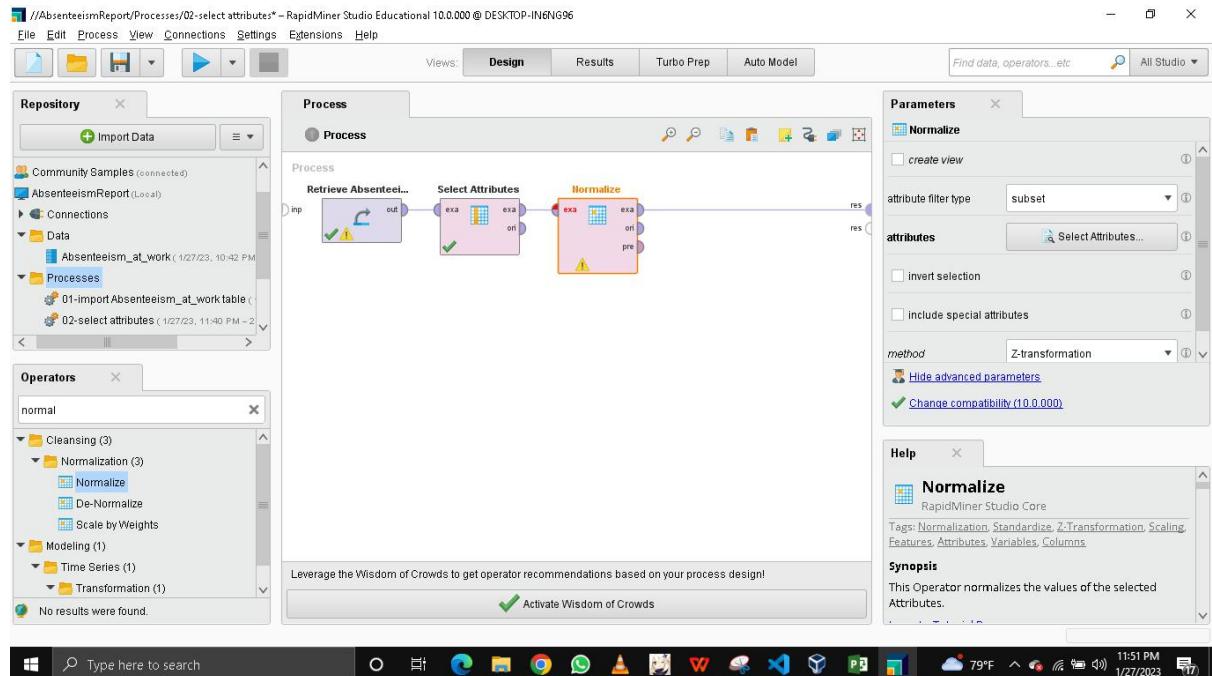
This Operator selects a subset of Attributes of an ExampleSet and removes the other Attributes.

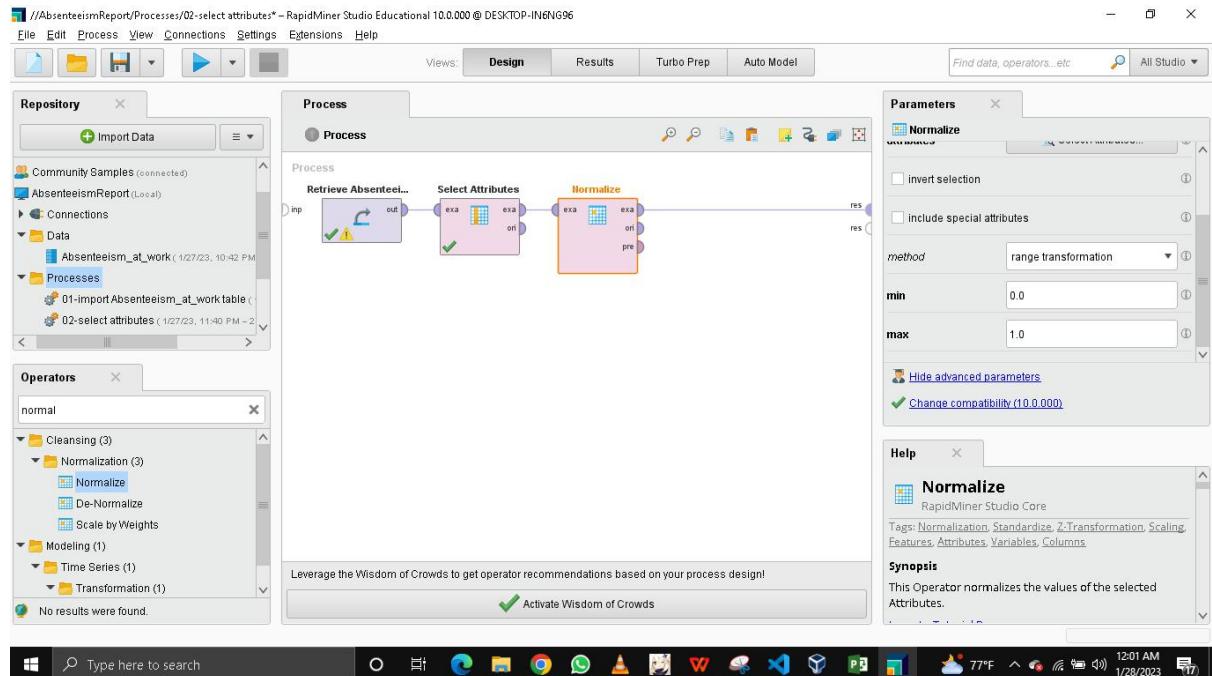




5. Used the “Normalize” operator to normalize all attributes to values of similar significance for analysis and saved the process as “03-attribute normalization”

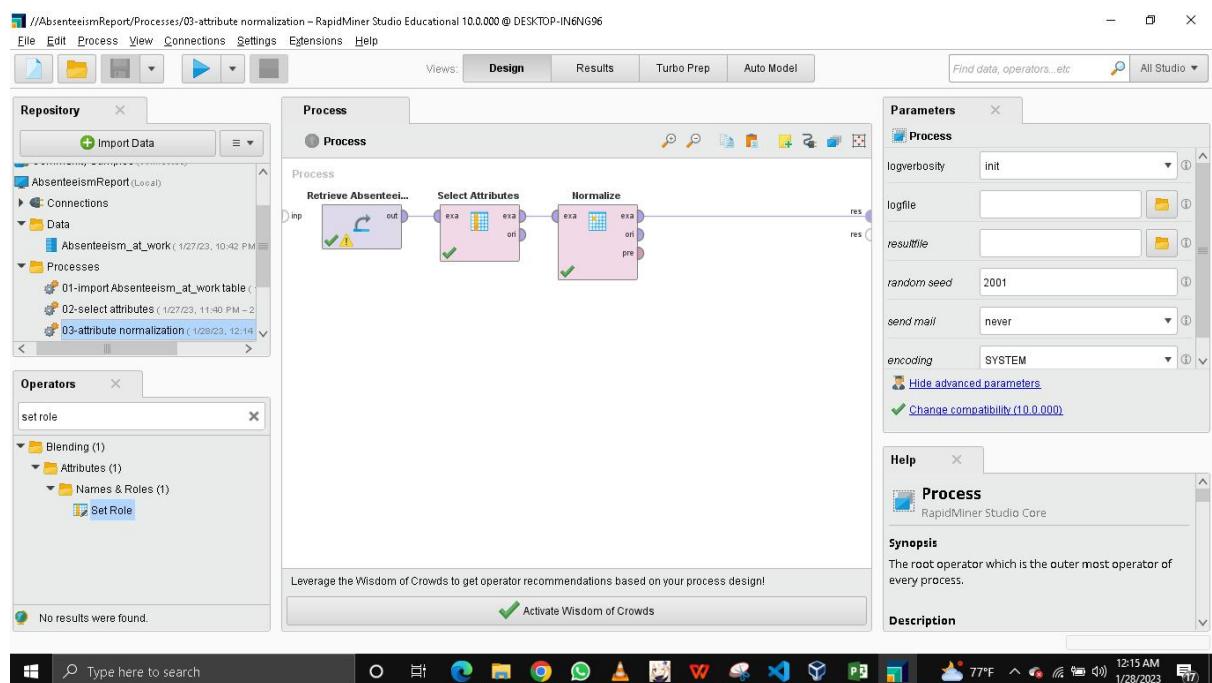
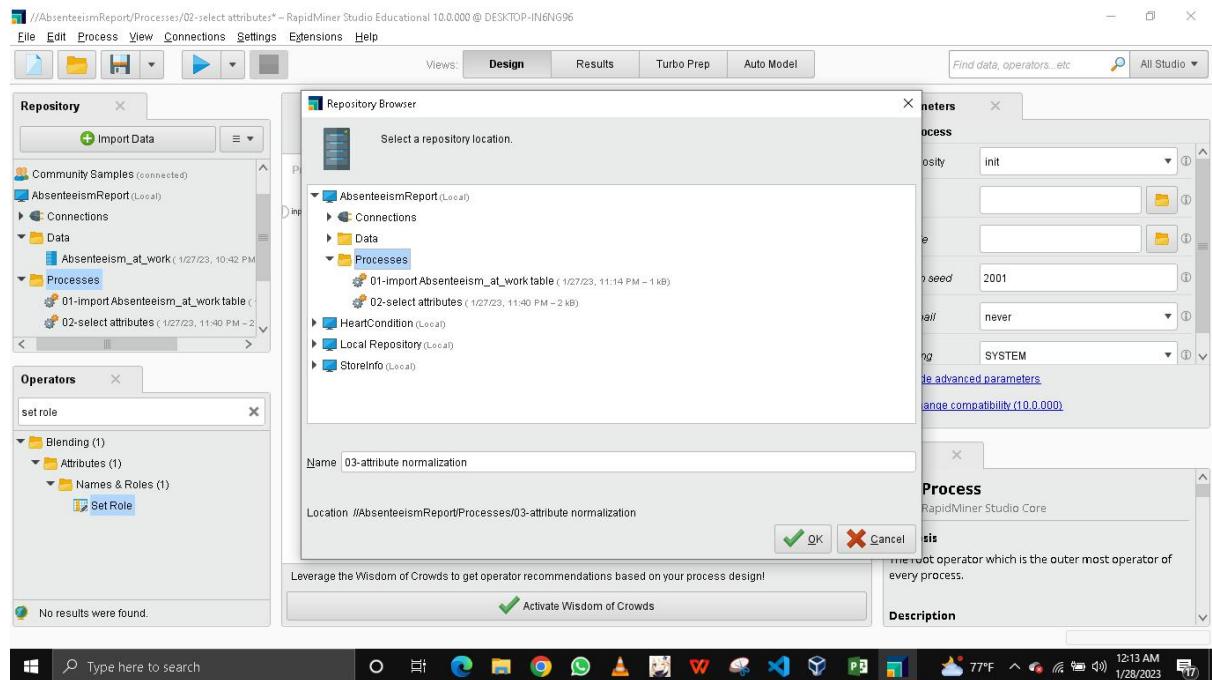




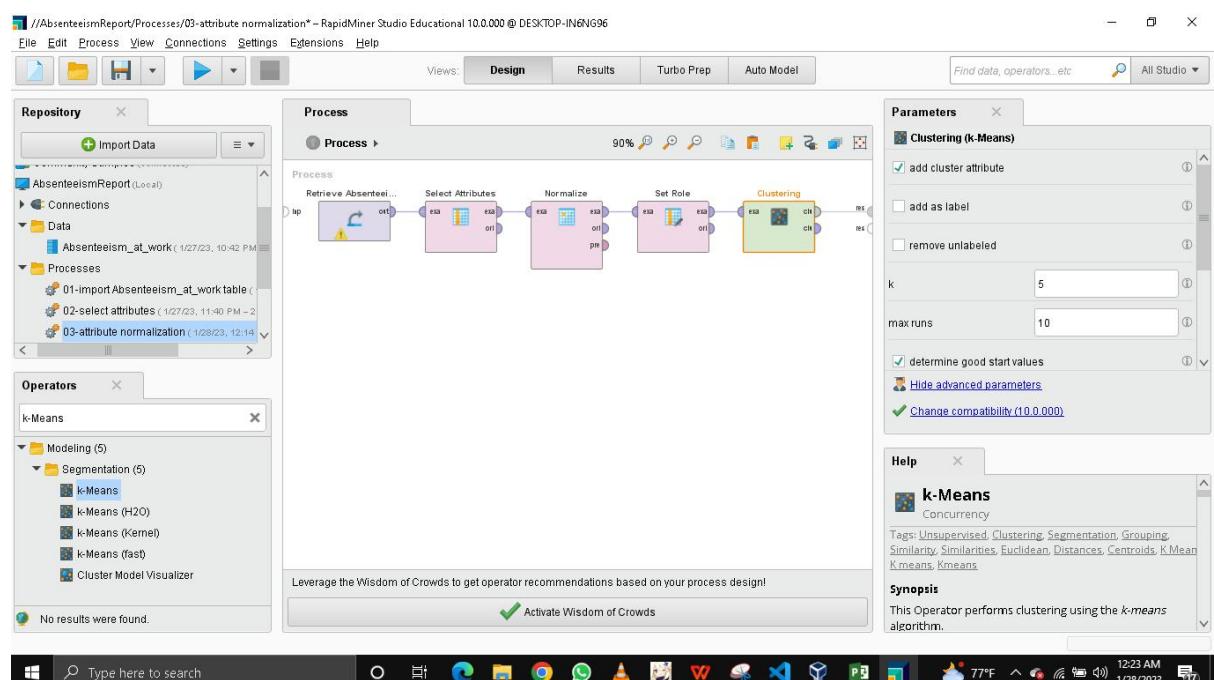
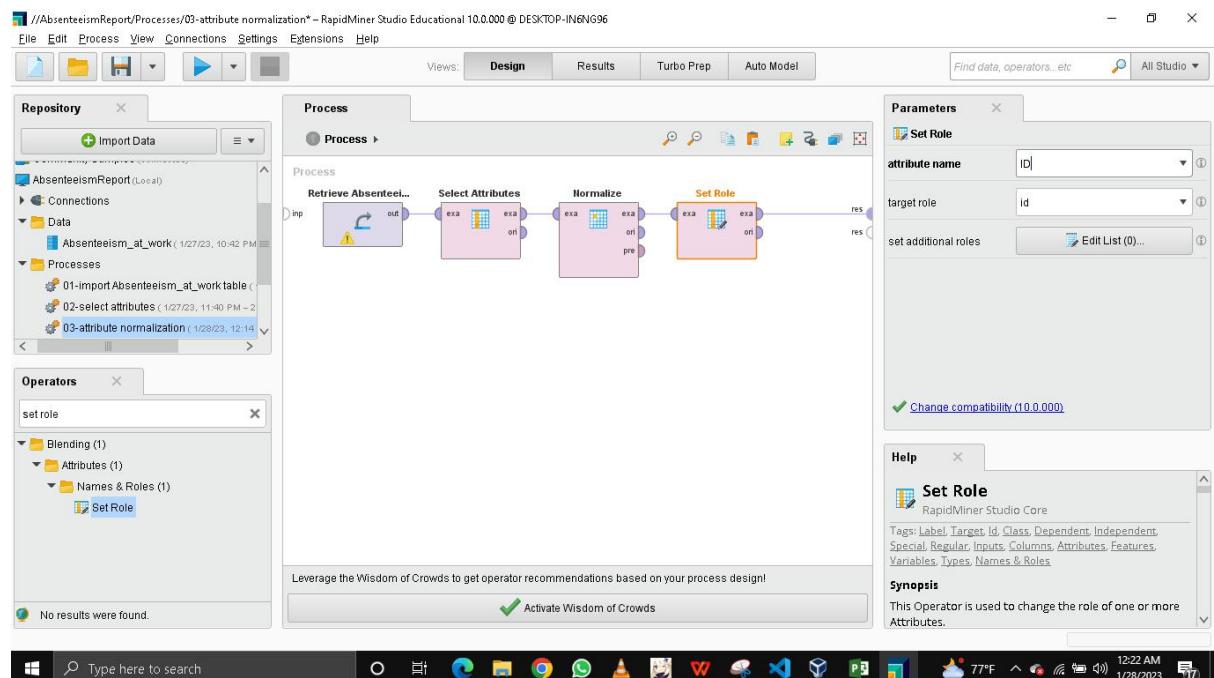


The screenshot shows the RapidMiner Studio interface with the following components:

- Result History:** Displays the "ExampleSet (Normalize)" tab, which shows the normalized data from the "Absenteeism\_at\_work" table. The table includes columns: Row No., Reason for..., Month of ab..., Day of the w..., Distance fro..., Work load A..., Absenteeis..., and ID. The data consists of 15 rows of normalized numerical values.
- Repository:** Shows the same tree structure as the first screenshot, indicating the state of the project after normalization.
- Operating System Taskbar:** Shows the Windows taskbar with various pinned icons and the system tray indicating the date and time as 1/28/2023 at 12:01 AM.



6. Defined the “ID” column within the Absenteeism\_at\_work table as the identifier using the “Set Role” operator to prevent cluster anonymity when using the k-Means “Clustering” operator to generate results and saved the process as “04-k-Means clustering”



//AbsenteeismReport/Processes/03-attribute normalization\* – RapidMiner Studio Educational 10.0.000 @ DESKTOP-IN6NG96

File Edit Process View Connections Settings Extensions Help

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

**Repository**

- Import Data
- AbsenteeismReport (Local)
  - Connections
  - Data Absenteeism\_at\_work (1/27/23, 10:42 PM)
  - Processes
    - 01-import Absenteeism\_at\_work table (1/27/23, 11:40 PM - 2 kB)
    - 02-select attributes (1/27/23, 11:40 PM - 2 kB)
    - 03-attribute normalization (1/28/23, 12:14 AM - 3 kB)

**Process**

Process

```

    graph LR
        A[Retrieve Absenteeism] --> B[Select Attributes]
        B --> C[Normalize]
        C --> D[Set Role]
        D --> E[Clustering]
    
```

90%

**Parameters**

Clustering (k-Means)

- add cluster attribute
- add as label
- remove unlabeled

k: 5

max runs: 10

determine good start values

Hide advanced parameters

Change compatibility (10.0.000)

**Operators**

k-Means

- Modeling (5)
  - Segmentation (5)
    - k-Means
    - k-Means (H2O)
    - k-Means (Kernel)
    - k-Means (fast)
    - Cluster Model Visualizer

No results were found.

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

Activate Wisdom of Crowds

Windows Taskbar: Type here to search, File Explorer, Task View, Edge, Chrome, WhatsApp, VLC, Word, Excel, Powerpoint, FileZilla, 77°F, 12:25 AM, 1/28/2023

//AbsenteeismReport/Processes/03-attribute normalization\* – RapidMiner Studio Educational 10.0.000 @ DESKTOP-IN6NG96

File Edit Process View Connections Settings Extensions Help

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

**Result History**

ExampleSet (/AbsenteeismReport/Data/Absenteeism\_at\_work)

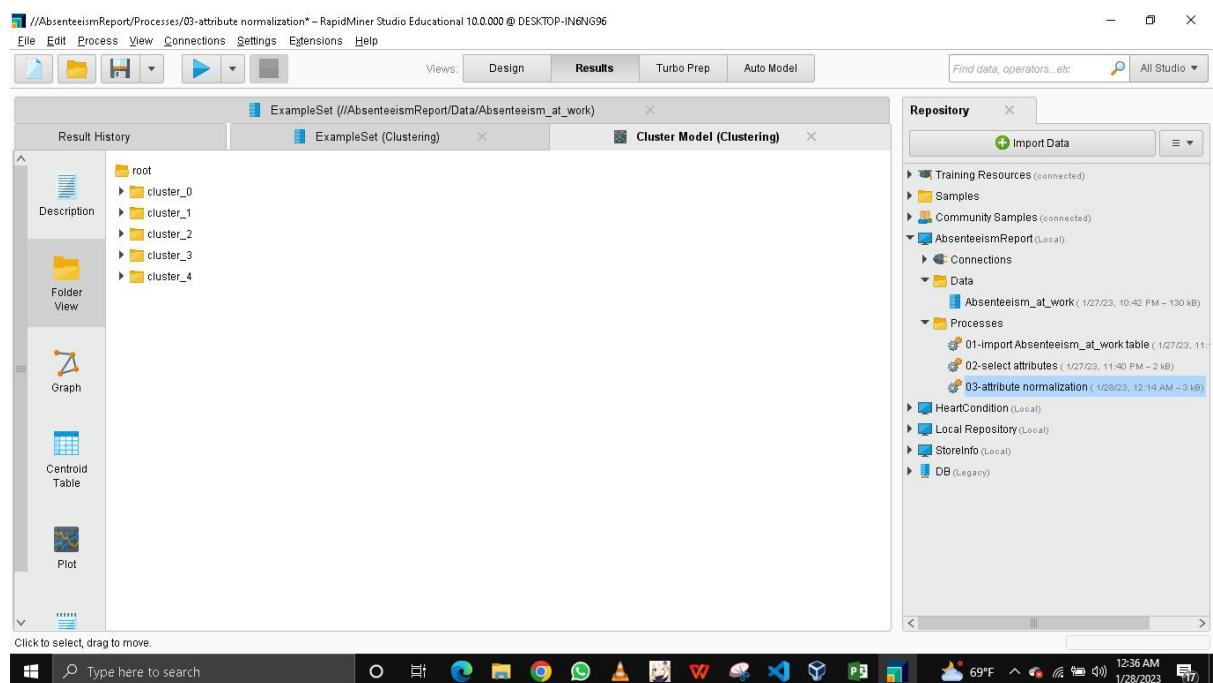
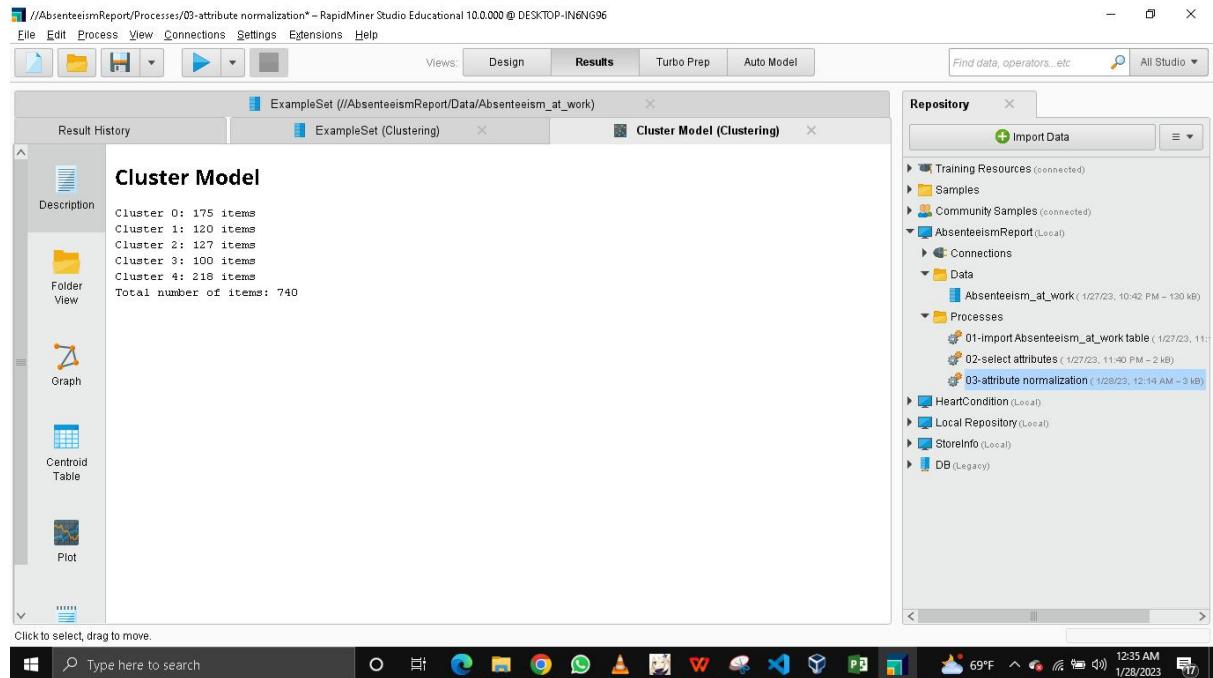
**ExampleSet (Clustering)**

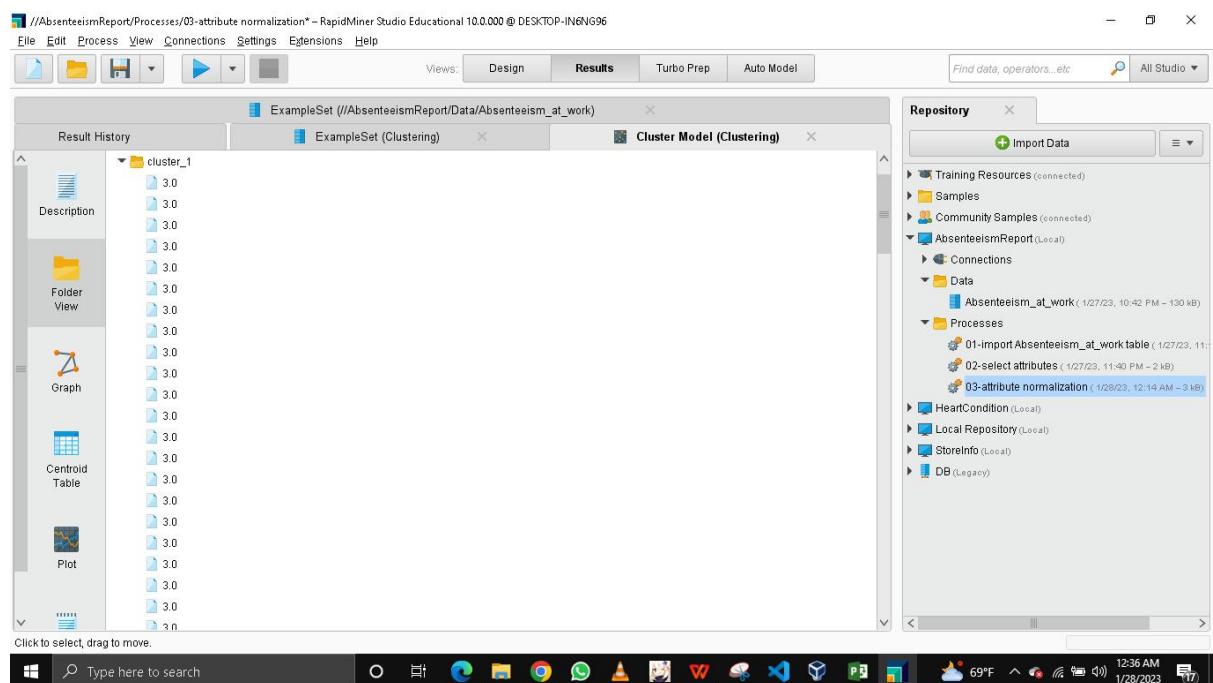
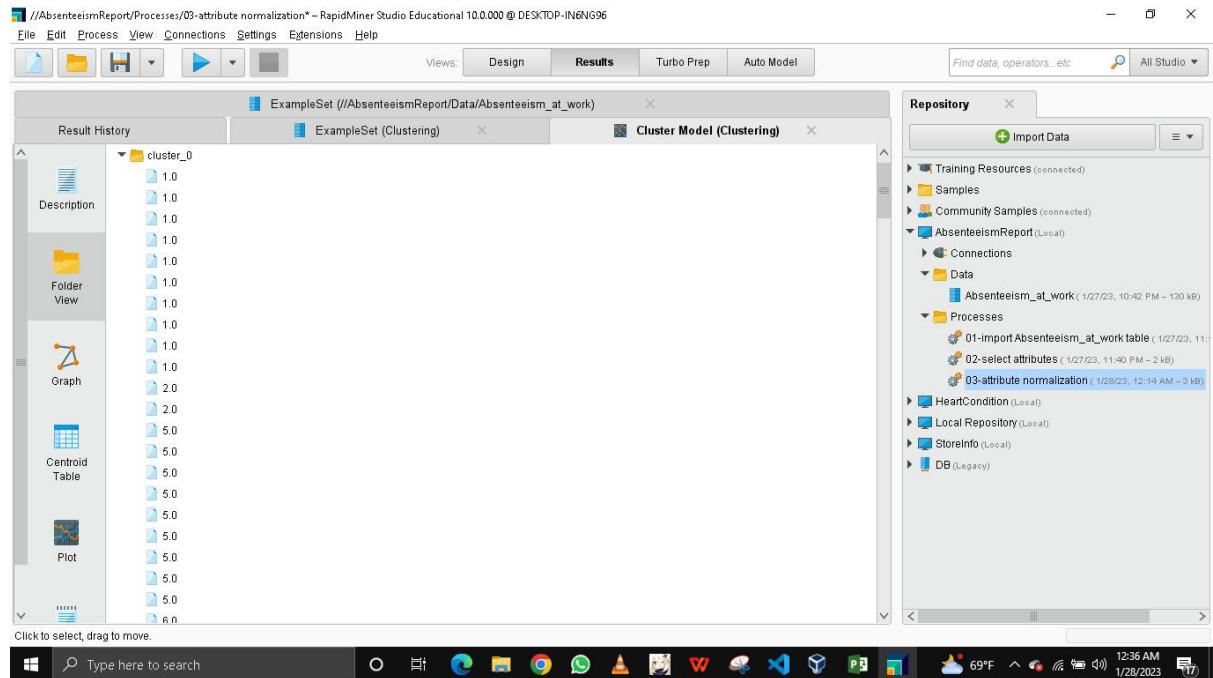
| Row No. | ID | cluster   | Reason for ... | Month of ab... | Day of the w... | Distance fro... | Work load A... | Absente |
|---------|----|-----------|----------------|----------------|-----------------|-----------------|----------------|---------|
| 1       | 1  | cluster_0 | 0.786          | 0.583          | 0               | 0.128           | 0.194          | 0.067   |
| 2       | 1  | cluster_4 | 0.821          | 0.667          | 0.750           | 0.128           | 0              | 0.033   |
| 3       | 1  | cluster_4 | 0.929          | 1              | 0.500           | 0.128           | 0.320          | 0.067   |
| 4       | 1  | cluster_2 | 0.250          | 0.333          | 1               | 0.128           | 0.697          | 0.025   |
| 5       | 1  | cluster_4 | 0.464          | 0.500          | 1               | 0.128           | 0.992          | 0.133   |
| 6       | 1  | cluster_0 | 0.821          | 0.667          | 0.250           | 0.128           | 0.345          | 0.008   |
| 7       | 1  | cluster_4 | 0.679          | 0.667          | 0.750           | 0.128           | 0.345          | 0.067   |
| 8       | 1  | cluster_0 | 0.929          | 0.933          | 0               | 0.128           | 0.342          | 0.033   |
| 9       | 1  | cluster_4 | 0.643          | 0.917          | 0.500           | 0.128           | 0.452          | 0.008   |
| 10      | 1  | cluster_4 | 0.893          | 0.917          | 0.750           | 0.128           | 0.452          | 0.017   |
| 11      | 1  | cluster_0 | 0.821          | 1              | 0               | 0.128           | 0.178          | 0.025   |
| 12      | 1  | cluster_4 | 0.893          | 1              | 0.750           | 0.128           | 0.178          | 0.017   |
| 13      | 1  | cluster_0 | 0.500          | 1              | 0.250           | 0.128           | 0.178          | 0.033   |

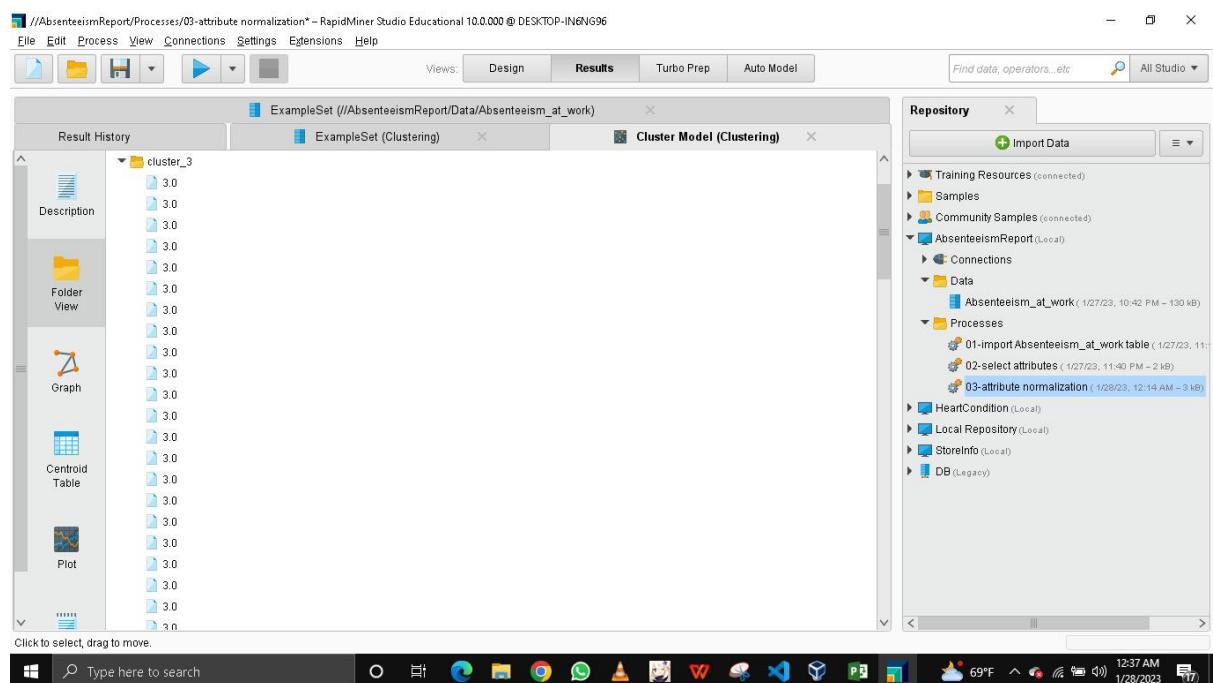
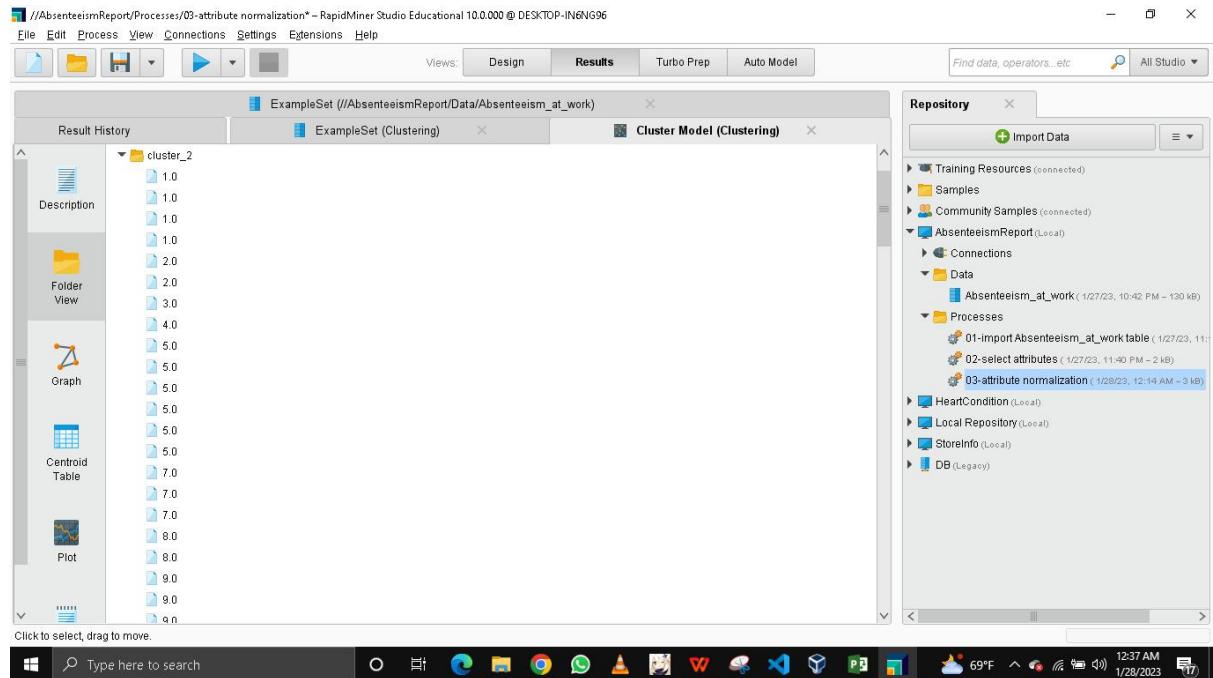
ExampleSet (740 examples, 2 special attributes, 6 regular attributes)

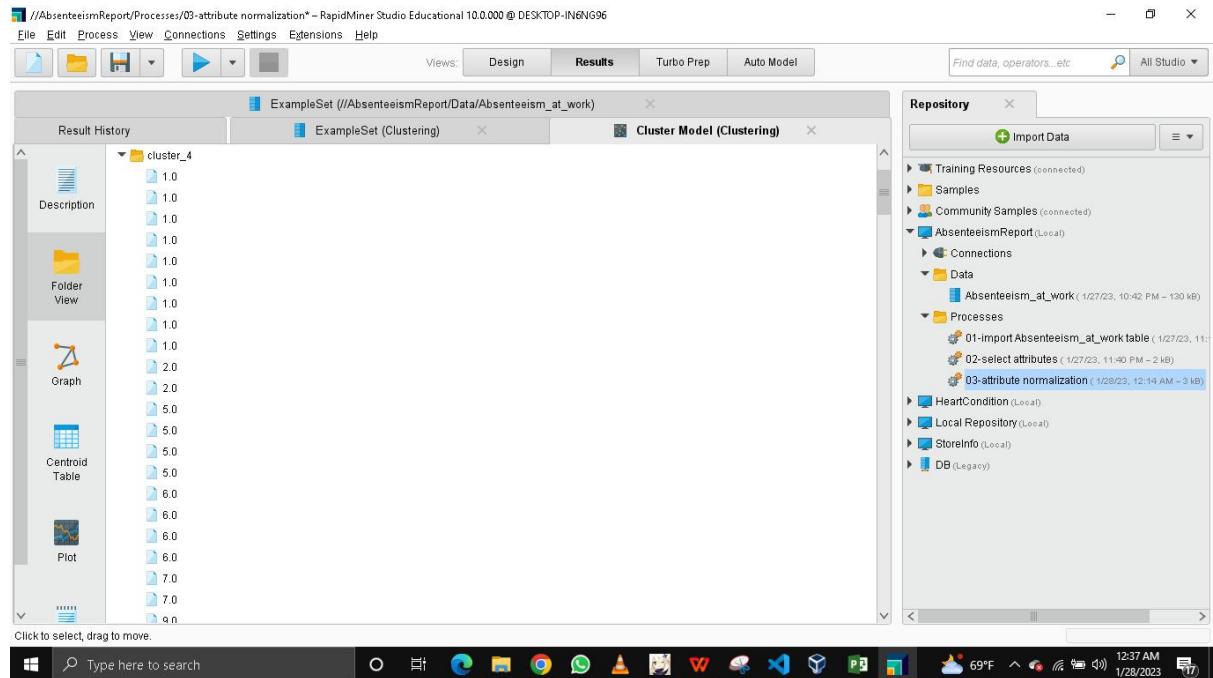
Click to select, drag to move.

Windows Taskbar: Type here to search, File Explorer, Task View, Edge, Chrome, WhatsApp, VLC, Word, Excel, Powerpoint, FileZilla, 69°F, 12:35 AM, 1/28/2023









Shown below are the results of the clustering process represented as graphical data for visualization

