

DATA WAREHOUSING AND DATA MINING-LABORATORY

ASSIGNMENT – 1

Name: Pranav mishra

REG NO: 12114762

ROLL NO: **RDOC09A55**

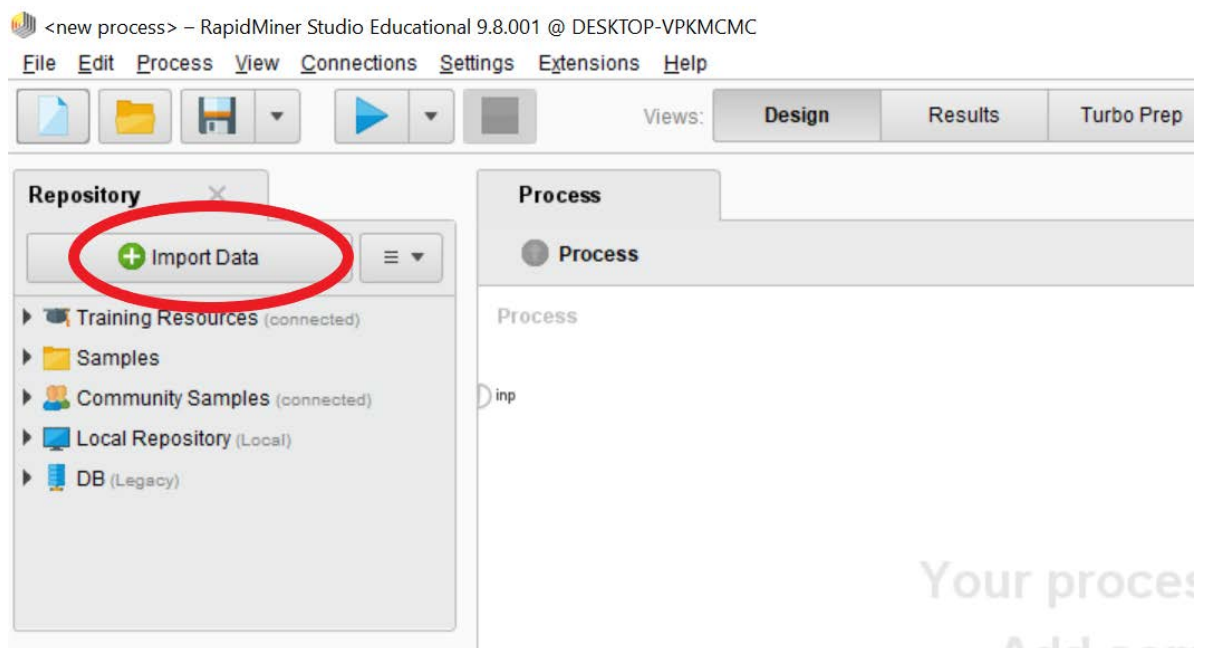
Course Code: CAP447

Teacher: Shilpa Course Title: Data Warehousing and Data Mining-Laboratory

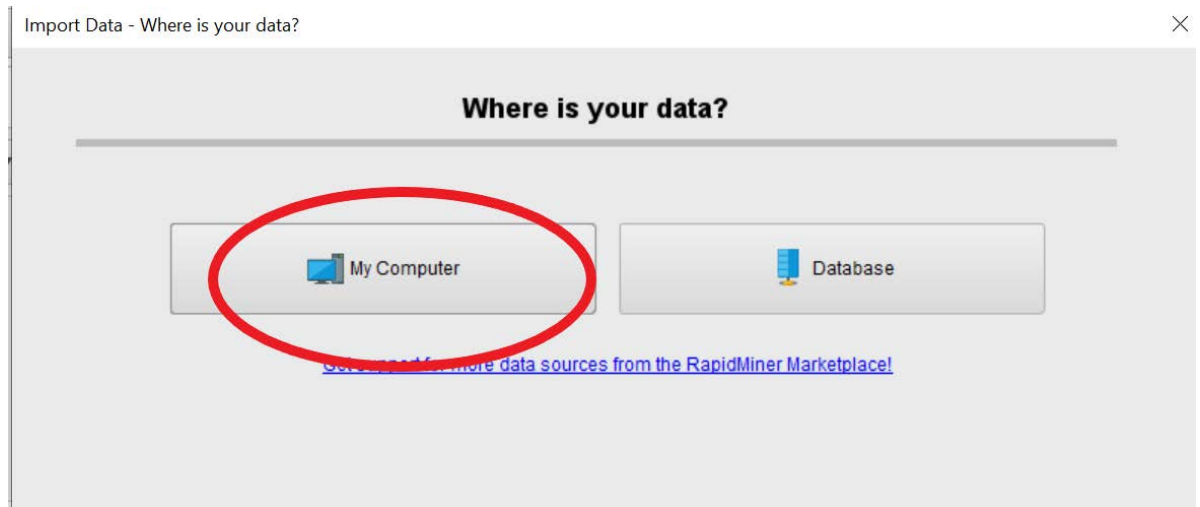
Q1. Perform the steps to import data from your local computer.

Answer: To import data from local computer have following steps:

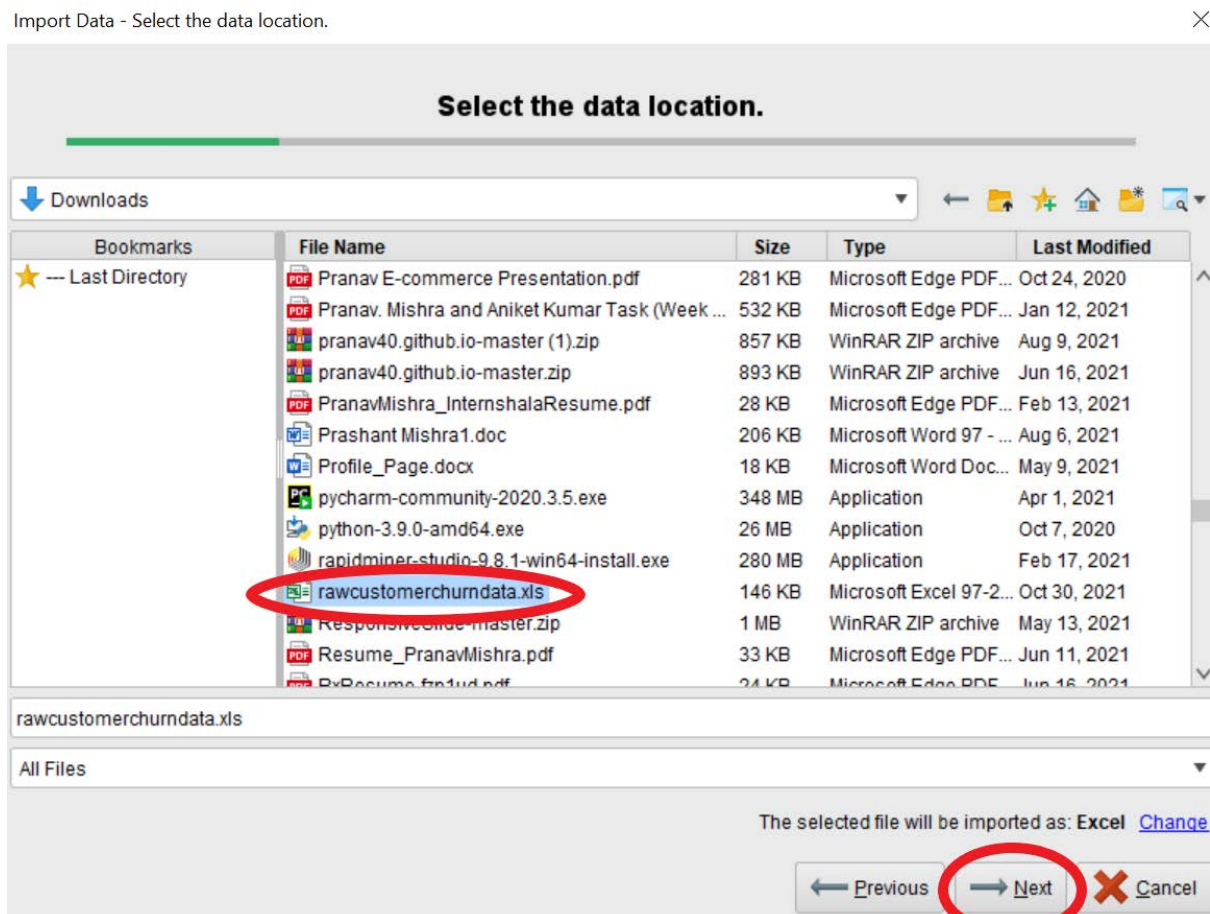
- Click on **Data Import Button**



- Select from where you want to import data?
Select the **My Computer Button** and then select the data(file).



- Select the exact location of data (file) and then press → **Next Button**
Here my data is present in **Downloads** and then press **Next Button**.



- Then it will ask to select cells which you want to import and click on **→Next Button**.
Here I select all cells from data(file) and then press to the **Next Button**.

Import Data - Select the cells to import.

Select the cells to import.

Sheet: customer data Cell range: A:F Select All ☒ Define header row: 1

	A	B	C	D	E	F
1	Name	Gender	Age	Payment Method	Churn	LastTransaction
2	Nicolas Garrett	male	64	credit card	loyal	98
3	Isaac Reyes	male	35	cheque	churn	118
4	Jaime Sullivan	female	25	credit card	loyal	107
5	Geraldine Miller	female	39	credit card		177
6	Curtis Frazier	m	39	credit card	loyal	90
7	Jeannie Palmer	female	28	cheque	churn	189
8	Phyllis Romero	female	21	credit card	loyal	102
9	Maxine Edwards	female		cheque	loyal	111
10	Marty Cohen		32	cheque	churn	50
11	Lionel Mendoza	male	48	credit card	loyal	141
12	Maureen Norman	female	70	credit card	churn	153
13	Santiago Cruz	male	36	credit card	loyal	46
14	Santiago Cruz	male	36	credit card	loyal	46
15	Nelson Davis	male	22	credit card	loyal	51
16	Josephine Owens	female	53	cash		183
17	Clarence Vaughn	male	27	cash	loyal	137
18	Jon Griffin	male	22	cash	loyal	147
19	Nettie Neal	female	49	credit card	churn	158
20	Belinda Reeves	female	24	cash	churn	162

← Previous → Next ✗ Cancel

- Will ask to change the format of data. If you want to change the format of data then you can change it by click on setting icon after change the format of data just click on **→Next Button**

Import Data - Select the cells to import.

Select the cells to import.

Sheet: customer data Cell range: A:F Select All ☒ Define head... 1

	A	B	C	D	E	F
263	Jaime Bass	male	35	credit card	loyal	131
264	Patty Wright	female	18	credit card		117
265	Bruce Russell	male	90	credit card	loyal	138
266	Carrie Stevenson	male	51	credit card	loyal	106
267	Ethel Robbins	male	27	cash		130
268	Angel Morgan	female	24	credit card	loyal	115
269	Neil Christensen	male	28	cash	loyal	94
270	Tracy Goodwin	female	43	cheque	churn	167
271	Percy Cain	female	21	credit card	churn	106
272	Krista Turner	female	75	credit card		207
273	Rickey Sutton	female	51	credit card		142
274	Roxanne Tucker	male	39	credit card		122
275	Jeff Morris	male	46	credit card	loyal	50
276	Sandy Barton	female	34	credit card	churn	160

Previous **Next** Cancel

- Save your data(file) where you want to save, can change file name, after select the location click on **Finish button** and the save the data(file).

Import Data - Where to store the data?

Where to store the data?

Local Repository (Local)

Name rawcustomerchurndata

Location //Local Repository/rawcustomerchurndata

Previous **Finish** Cancel

- Then you can see data is Successfully imported in **Local Repository folder** and select the data and drag and drop then perform the operation and execute the operation.

<new process> - RapidMiner Studio Educational 9.8.001 @ DESKTOP-VPKMCMC

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model Deployments

Find data, operators, etc. All Studio

Result History ExampleSet (/Local Repository/rawcustomerchurndata)

Open in Turbo Prep Auto Model

Filter (999 / 999 examples): all

Row No.	Name	Gender	Age	Payment Me...	Churn	LastTransa...
1	Nicolas Garrett	male	64	credit card	loyal	98
2	Isaac Reyes	male	35	cheque	churn	118
3	Jaime Sullivan	female	25	credit card	loyal	107
4	Geraldine Mill...	female	39	credit card	?	177
5	Curtis Frazier	m	39	credit card	loyal	90
6	Jeannie Pal...	female	28	cheque	churn	189
7	Phyllis Romero	female	21	credit card	loyal	102
8	Maxine Edwa...	female	?	cheque	loyal	111
9	Marty Cohen	?	32	cheque	churn	50
10	Lionel Mendo...	male	48	credit card	loyal	141
11	Maureen Nor...	female	70	credit card	churn	153
12	Santiago Cruz	male	36	credit card	loyal	46
13	Santiago Cruz	male	36	credit card	loyal	46
14	Nelson Davis	male	??	credit card	loyal	51

ExampleSet (999 examples, 0 special attributes, 6 regular attributes)

Repository

Import Data

- Training Resources (connected)
- Samples
- Community Samples
- Local Repository (Local)**
- Connections
- data
- processes
- Assignment (11/10/21 1:36)
- DataSet (11/8/21 11:02 AM)
- Dataset Record (11/8/21 10)
- Navie (11/17/21 11:28 AM)
- rawcustomerchurndata (11)
- DB (Legacy)

Q2. Perform following operations on your data you imported.

A. Remove duplicates

In our data there is many duplicates in Gender, Payment Method, Churn, Last Transaction etc.

<new process> - RapidMiner Studio Educational 9.8.001 @ DESKTOP-VPKMC MC

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model Deployments Find data, open

ExampleSet (//Local Repository/rawcustomerchurndata) ExampleSet (//Local Repository/rawcustomerchurndata)

Result History ExampleSet (Remove Duplicates)

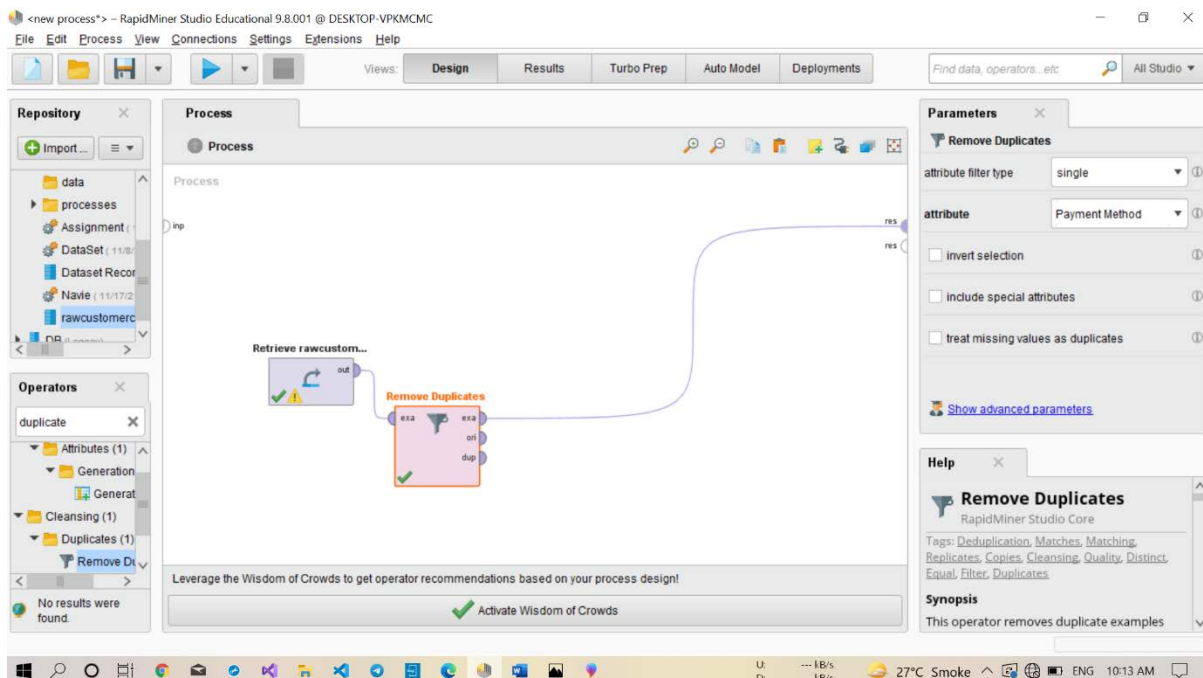
Open in Turbo Prep Auto Model Filter (998 / 998 examples): all

Row No.	Name	Gender	Age	Payment Me...	Churn	LastTransa...
1	Nicolas Garrett	male	64	credit card	loyal	98
2	Isaac Reyes	male	35	cheque	churn	118
3	Jaime Sullivan	female	25	credit card	loyal	107
4	Geraldine Mill...	female	39	credit card	?	177
5	Curtis Frazier	m	39	credit card	loyal	90
6	Jeannie Pal...	female	28	cheque	churn	189
7	Phyllis Romero	female	21	credit card	loyal	102
8	Maxine Edwa...	female	?	cheque	loyal	111
9	Marty Cohen	?	32	cheque	churn	50
10	Lionel Mendo...	male	48	credit card	loyal	141
11	Maureen Nor...	female	70	credit card	churn	153
12	Santiago Cruz	male	36	credit card	loyal	46
13	Nelson Davis	male	22	credit card	loyal	51

ExampleSet (998 examples, 0 special attributes, 6 regular attributes)

In parameters, select one attributes filter types to remove duplicates value from select attributes.

select the attribute **Payment Method**.



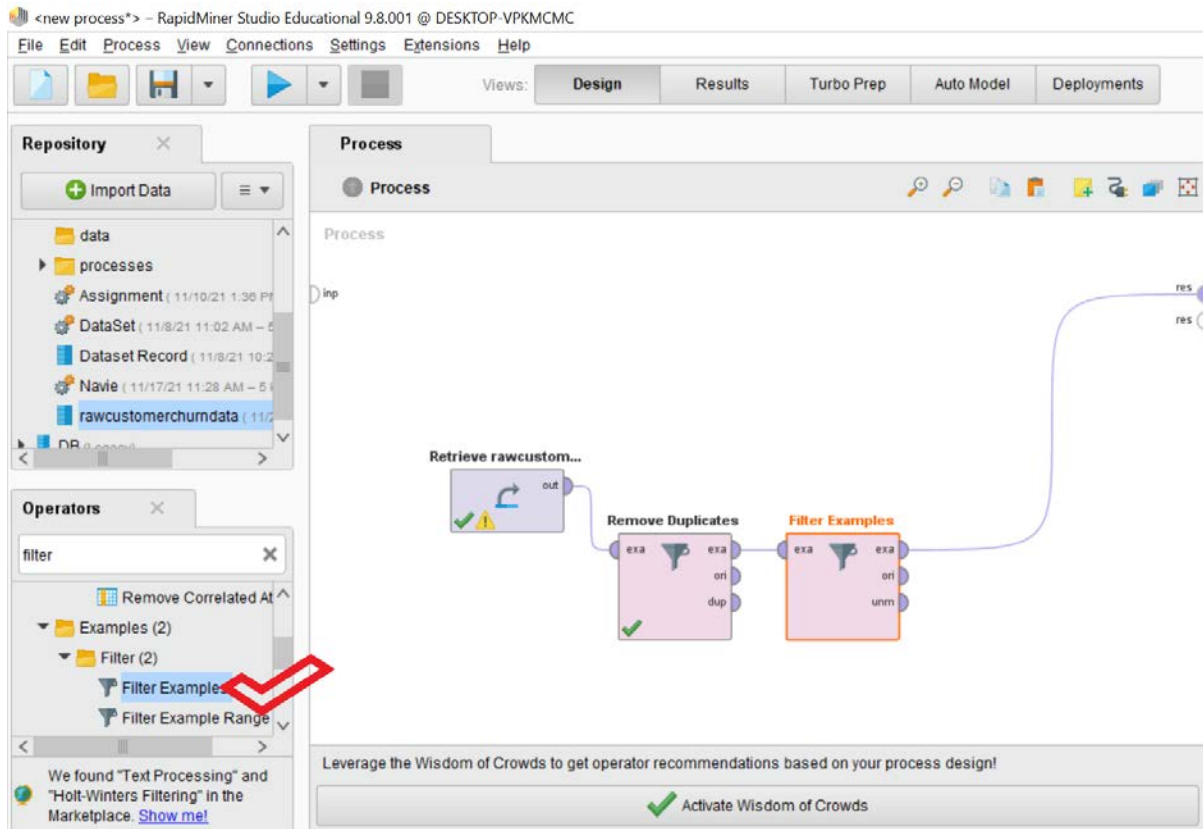
We can see in Payment Method all duplicate values removed.

The screenshot shows the 'Results' view of RapidMiner Studio. The 'ExampleSet (Remove Duplicates)' is selected, showing 3 examples. The data is displayed in a table with the following columns: Row No., Name, Gender, Age, Payment Me..., Churn, and LastTransa... The 'Payment Me...' column is highlighted with a red box, indicating that all duplicate values have been removed.

Row No.	Name	Gender	Age	Payment Me...	Churn	LastTransa...
1	Nicolas Garrett	male	64	credit card	loyal	98
2	Isaac Reyes	male	35	cheque	churn	118
3	Josephine O...	female	53	cash	?	183

B. Filter missing data

Missing data is filtered by using filter **Examples** operator then execute the filter operation then see the result of missing value.



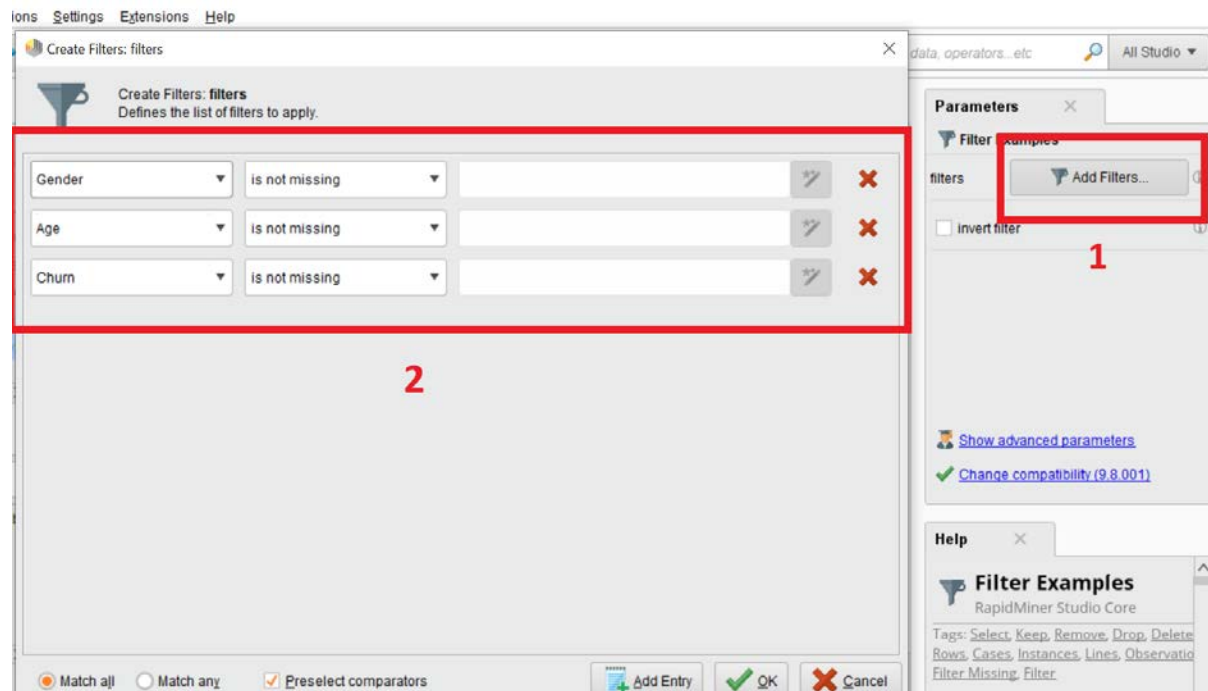
Here We can see there is three attributes I have select Gender, **Age** and **Churn** and perform the operation and got missing values in three attributes.

The screenshot shows the 'Results' pane of RapidMiner Studio. It displays a table of attributes with their types, missing values, and statistics. The 'Missing' column shows the number of missing values for each attribute. The attributes 'Gender', 'Age', and 'Churn' have missing values of 1, 1, and 95 respectively, which are highlighted with red boxes. The 'Filter (6 / 6 attributes)' dropdown is set to 'Search for Attributes'.

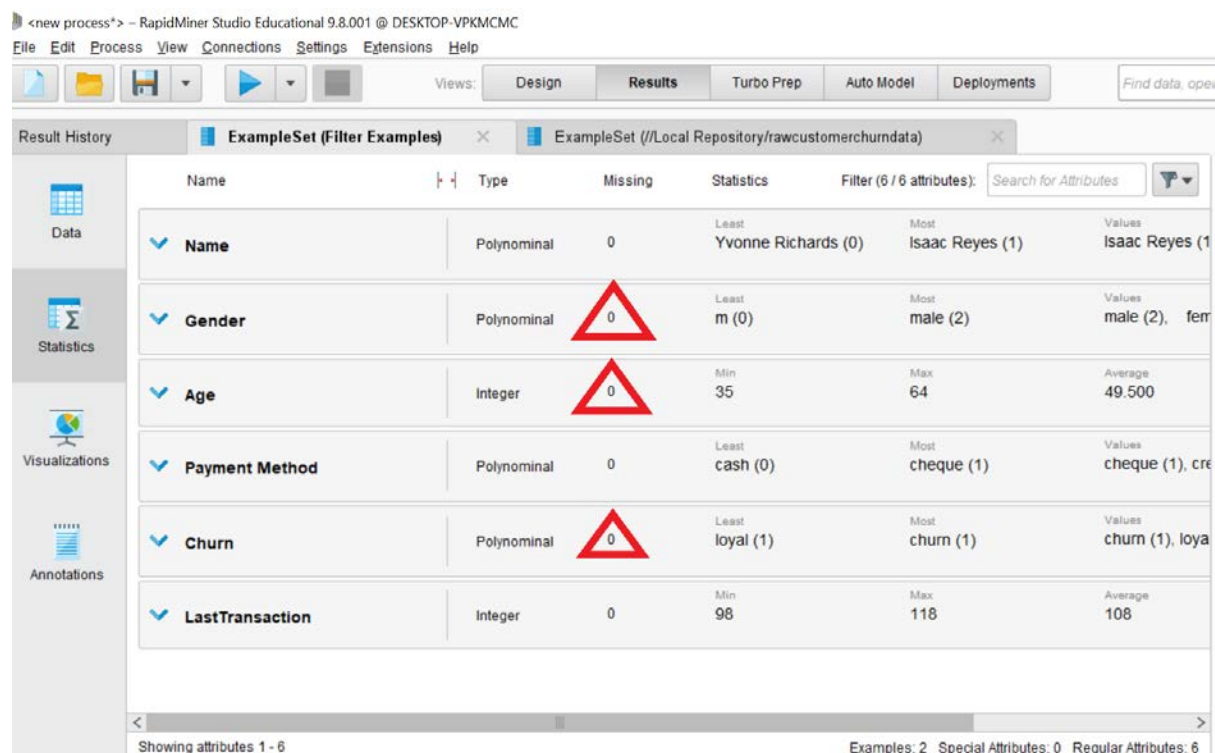
Name	Type	Missing	Statistics	Filter (6 / 6 attributes)
Name	Polynomial	0	Least Yvonne Richards (1) Most Santiago Cruz (2)	Search for Attributes
Gender	Polynomial	1	Least m (1) Most male (548)	
Age	Integer	1	Min 17 Max 91 Average 45.593	
Payment Method	Polynomial	0	Least cheque (70) Most credit card (650)	
Churn	Polynomial	95	Least churn (323) Most loyal (580)	
LastTransaction	Integer	0	Min 1 Max 223 Average 110.946	

Showing attributes 1 - 6 Examples: 999 Special Attributes: 0 Regular Attributes: 6

And then go to **Parameters**, then add attributes from which we want to filter missing values, after adding the attributes then go to the **OK Button** and press.



Now all the Missing values are filtered, data haven't missing values in the selected above attributes **gender, age and churn**.



c. Setting roles

When we use Set Role operator, firstly we can drag and drop the set operator and then select the attributes name from parameter, after select attribute name we select the target role which you want to give specific role.

Here I drag and drop the **Set Role Operator** then go to parameter and select attribute name is **Payment Method** and then select the target role **Label** then I execute the operation and see the result.

The screenshot displays the RapidMiner Studio interface. The main canvas shows a process flow starting with 'Retrieve rawcustom...', followed by 'Remove Duplicates', 'Filter Examples', and finally the 'Set Role' operator. The 'Set Role' operator is highlighted with a red box. The 'Parameters' panel on the right shows the configuration for the 'Set Role' operator: 'attribute name' is set to 'Payment Method' and 'target role' is set to 'label'. The 'Repository' panel on the left shows the data sources, including 'rawcustomchurndata'. The 'Operators' panel on the left shows the 'Set Role' operator selected under 'Names & Roles'. The 'Help' panel on the right provides information about the 'Set Role' operator, including its tags and synopsis.

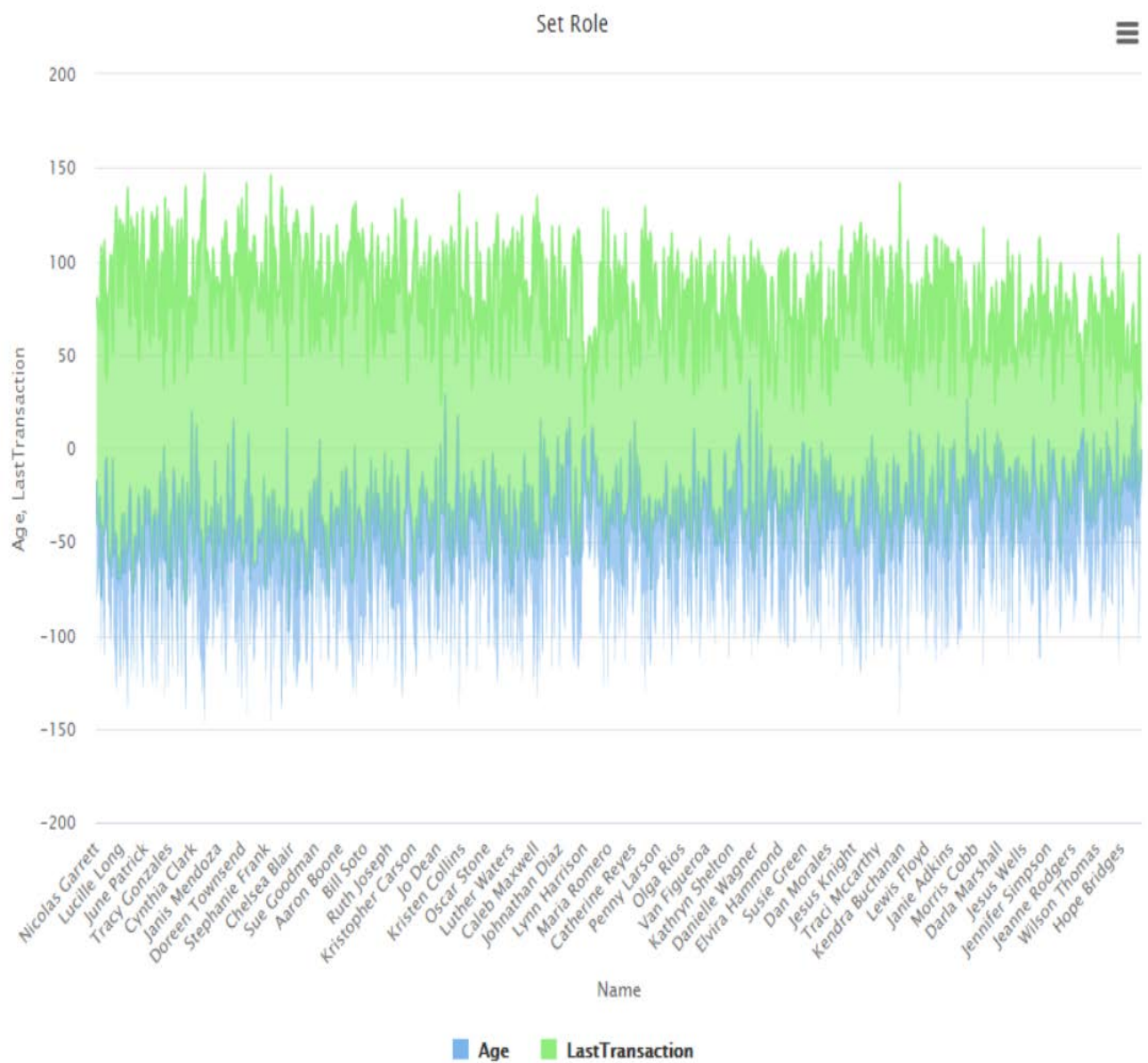
Output of set role operation. Payment method have different color showing different role between all attributes after perform the operation.

Row No.	Payment Me...	Name	Gender	Age	Churn	LastTransa...
1	credit card	Nicolas Garrett	male	64	loyal	98
2	cheque	Isaac Reyes	male	35	churn	118
3	credit card	Jaime Sullivan	female	25	loyal	107
4	credit card	Curtis Frazier	m	39	loyal	90
5	cheque	Jeannie Pal...	female	28	churn	189
6	credit card	Phyllis Romero	female	21	loyal	102
7	credit card	Lionel Mendo...	male	48	loyal	141
8	credit card	Maureen Nor...	female	70	churn	153
9	credit card	Santiago Cruz	male	36	loyal	46
10	credit card	Nelson Davis	male	22	loyal	51
11	cash	Clarence Vau...	male	27	loyal	137
12	cash	Jon Griffin	male	22	loyal	147
13	credit card	Nettie Neal	female	49	churn	158
14	cash	Belinda Reev...	female	24	churn	162
15	credit card	Taylor Murphy	male	45	loyal	55
16	credit card	Emmett James	male	45	loyal	160
17	cash	Paula Murray	female	66	churn	156
18	cash	Penny Reese	female	82	churn	177

ExampleSet (900 examples, 1 special attribute, 5 regular attributes)

D. Visualizing data using one of the graphs and bar charts.

Graphs: When I execute the all operation then I go to the **Visualizations** **Button** then I select the **Graph** and see the visualizing data in **Graph** format.



Bar chart: I select the **Bar chart** and see the visualizing data in **Bar** format.

