**Lab: How to Create and Apply and Test a Model for a Decision Tree**

Download the excel sheet for the data set named as “customer-churn-data” from model. Follow the steps to prepare the data for the training set.

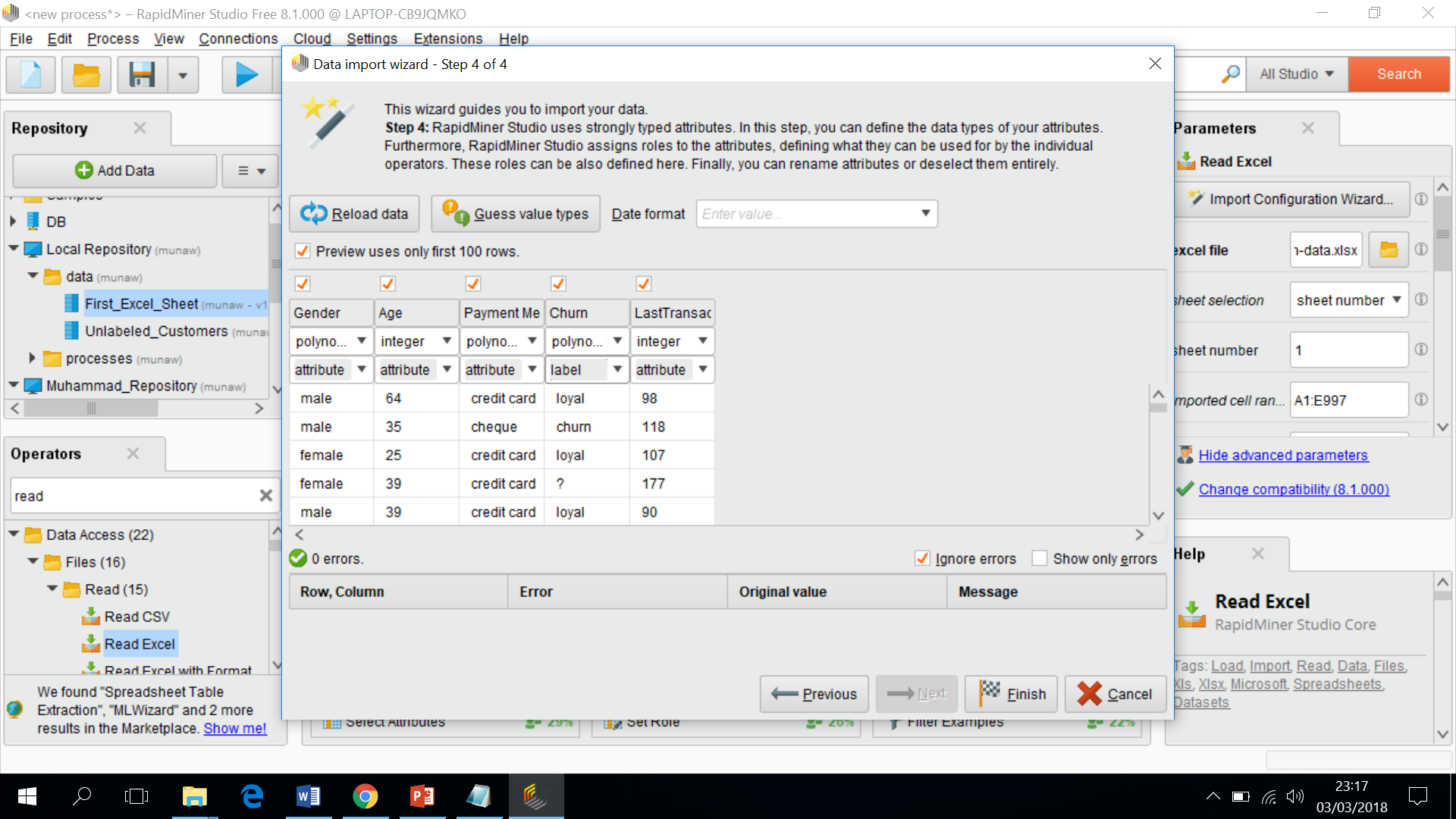
1) First, we do the data preparation and the create a model as you did in the previous tutorials.

2) Retrieve the data from the repository or using read excel operator to load the “customer-churn-data”

OR

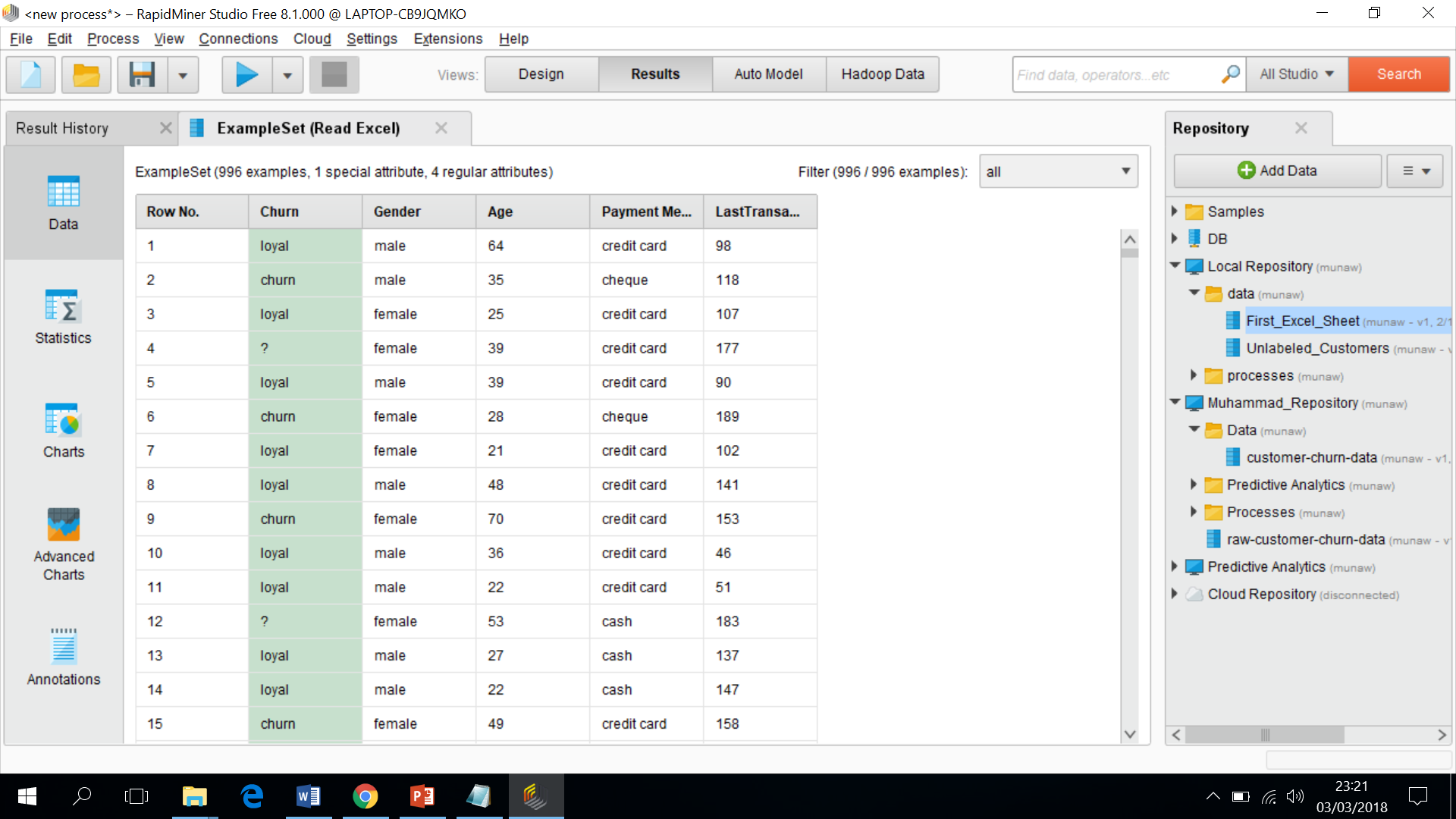
Drag the operator “Read Excel” into the process window and load the data file “customer-churn-data”.

3) Click the button on the parameters pane “Import Configuration visit” and select the “Next button” until the following image appears as mentioned below



4) Change the option, “attribute” for the Churn to “label” as highlighted with an arrow.

5) Connect the “Read Excel” Operator port to the output port and execute the process and the following window will appear as mentioned below



6) We need to follow some steps before the development of the decision tree, the term “label” means that we are interested in this attribute, called as “attribute of interest” for the modelling.

7) You already know that “?” marks means that the data is missing about the loyalty of the customer, Now filter out the missing values as mentioned before.

8) “Filter example” operator is used for this purpose and drags this operator into the process window.

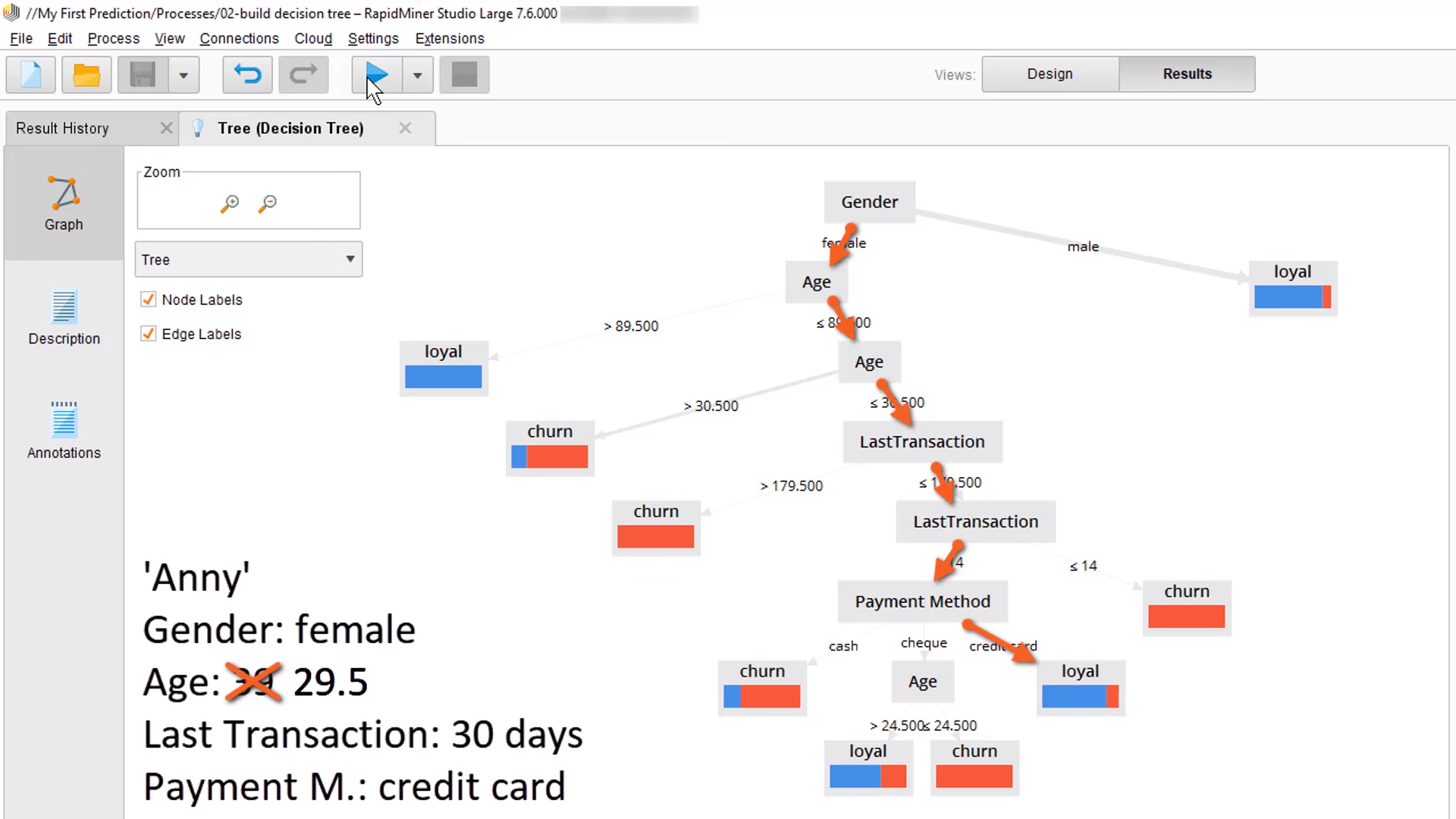
9) Click on “Add Filter” button on the parameters pane and select the attribute “Churn” and the option “ is not missing”.

10) Check and Analyse the statistics that rapid miner provided you after this execution.

11) Now we can build a decision tree, select the operator “Decision Tree” from the operator window and drags into the process window.

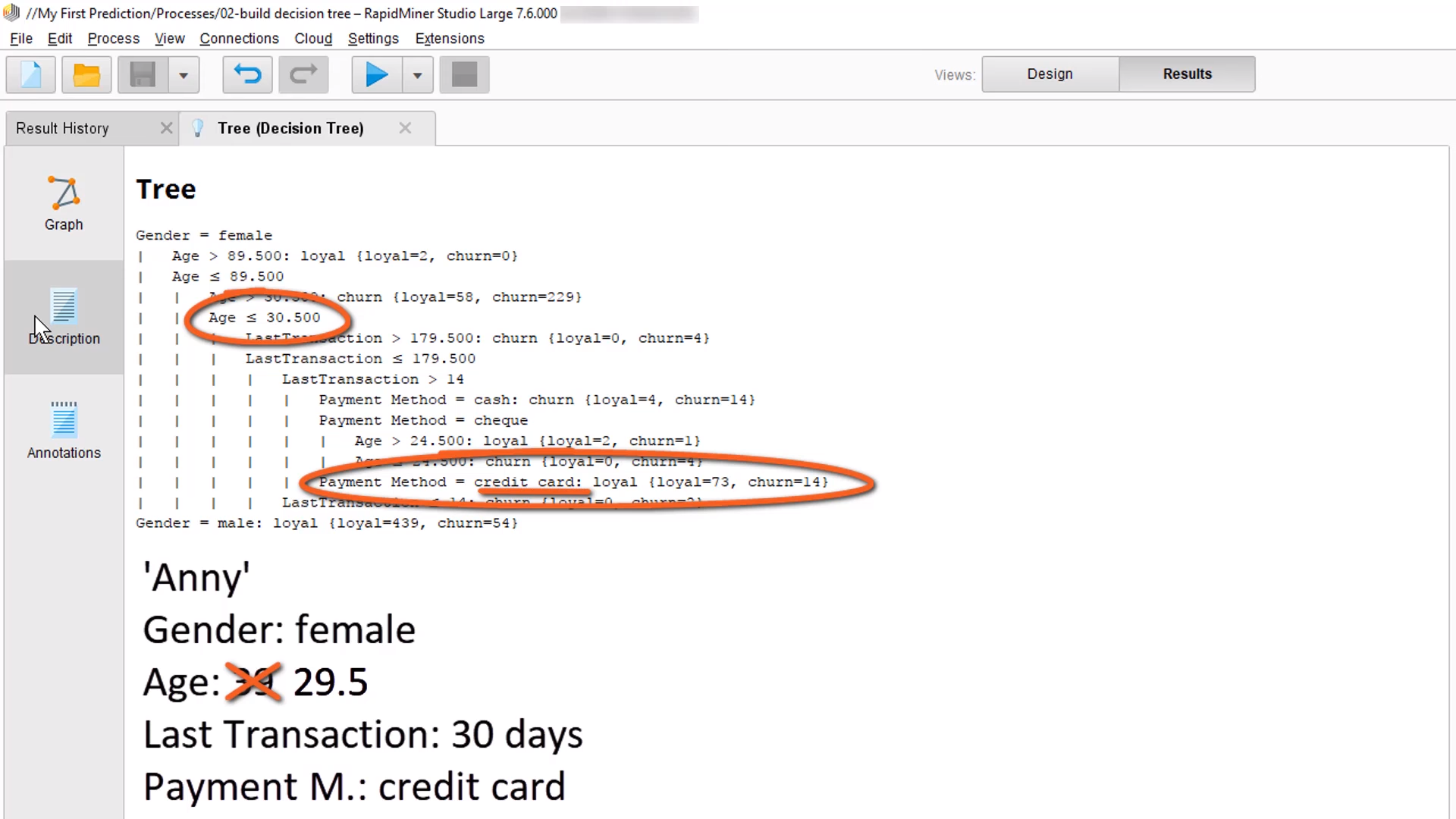
12) Now save all of your work into your folder.

13) Run the process and you will get a decision tree as you have seen before as mentioned below



Let us understand the customer “Anny” that she joined a month ago. We can get understanding that she will be a loyal or churn customer.

14) If you click on the description, then the following window appears that describes about the behaviour of the customer.

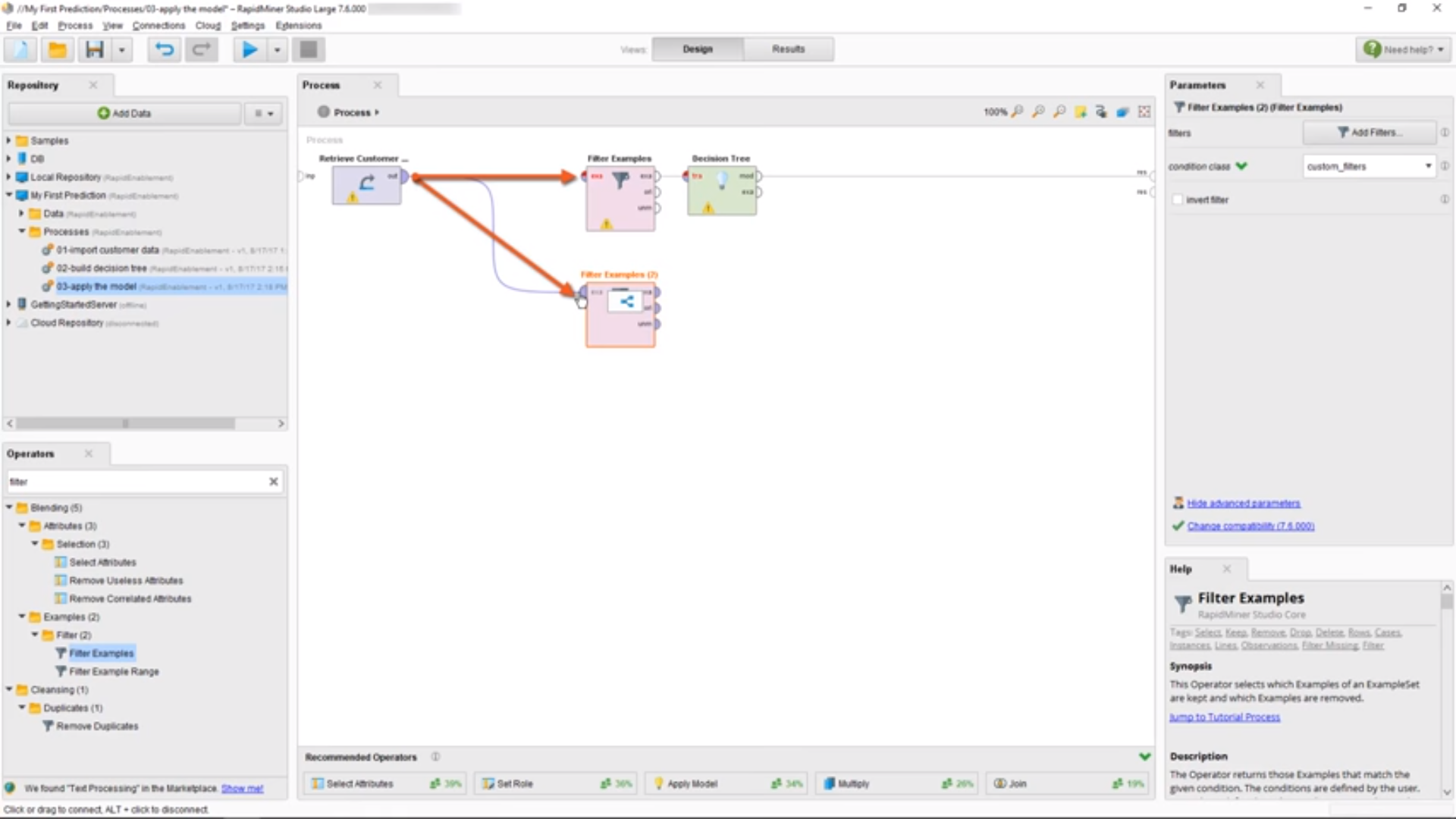


15) Now we apply the model. Copy the model from the repository pane and paste with a different name, called as “Apply the model”. You can split your work so that you could trace error in case of any problems.

16) For a single customer, check the decision tree for his loyalty or churn behaviour, it is quite difficult task.

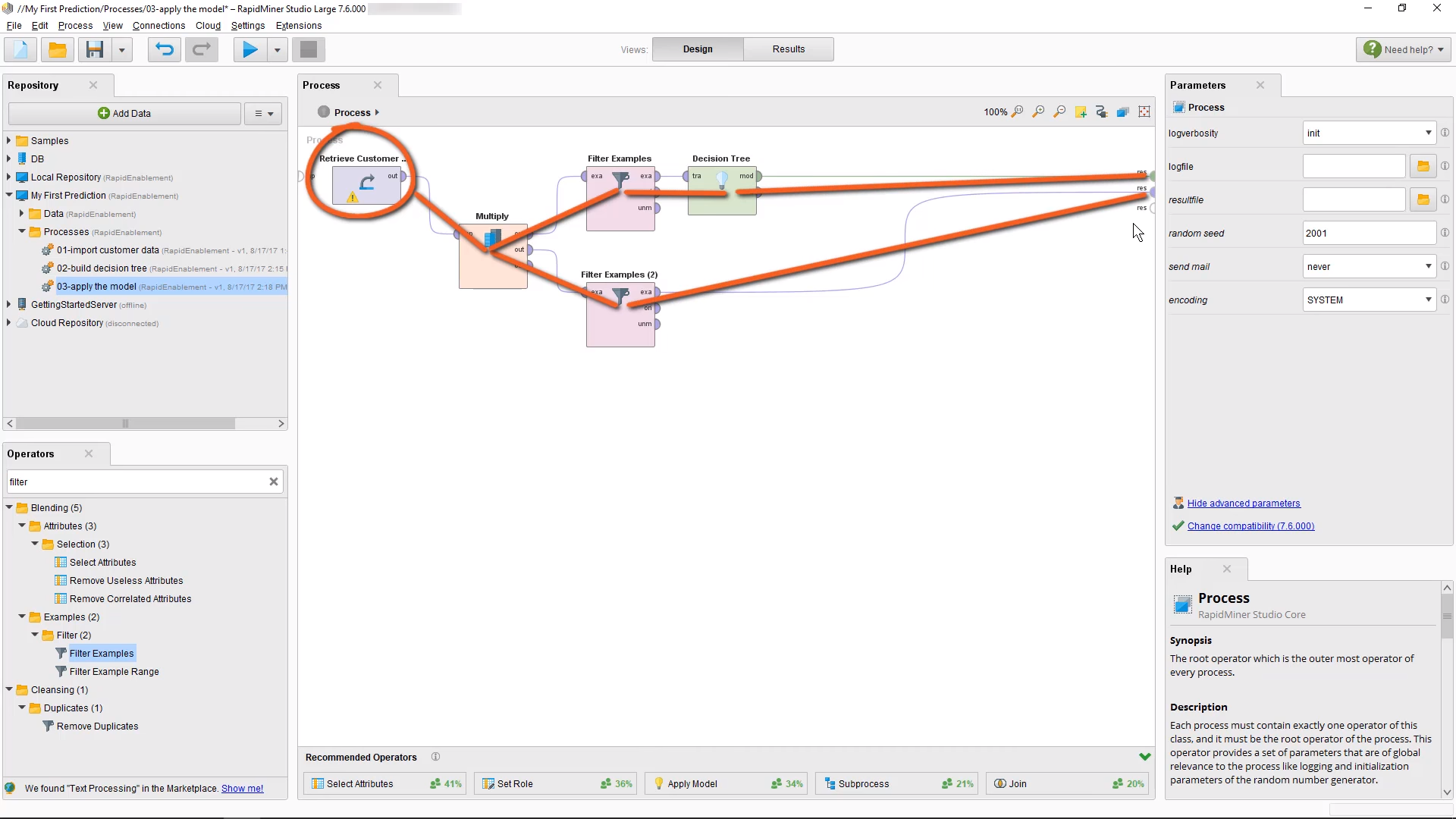
17) Now we apply the model on missing data for the Churn attribute in the customer’s data.

18) In rapid miner, one input can connect with one output port, How to get input data for two different operators? we use “multiply” operator to perform the following action as mentioned below

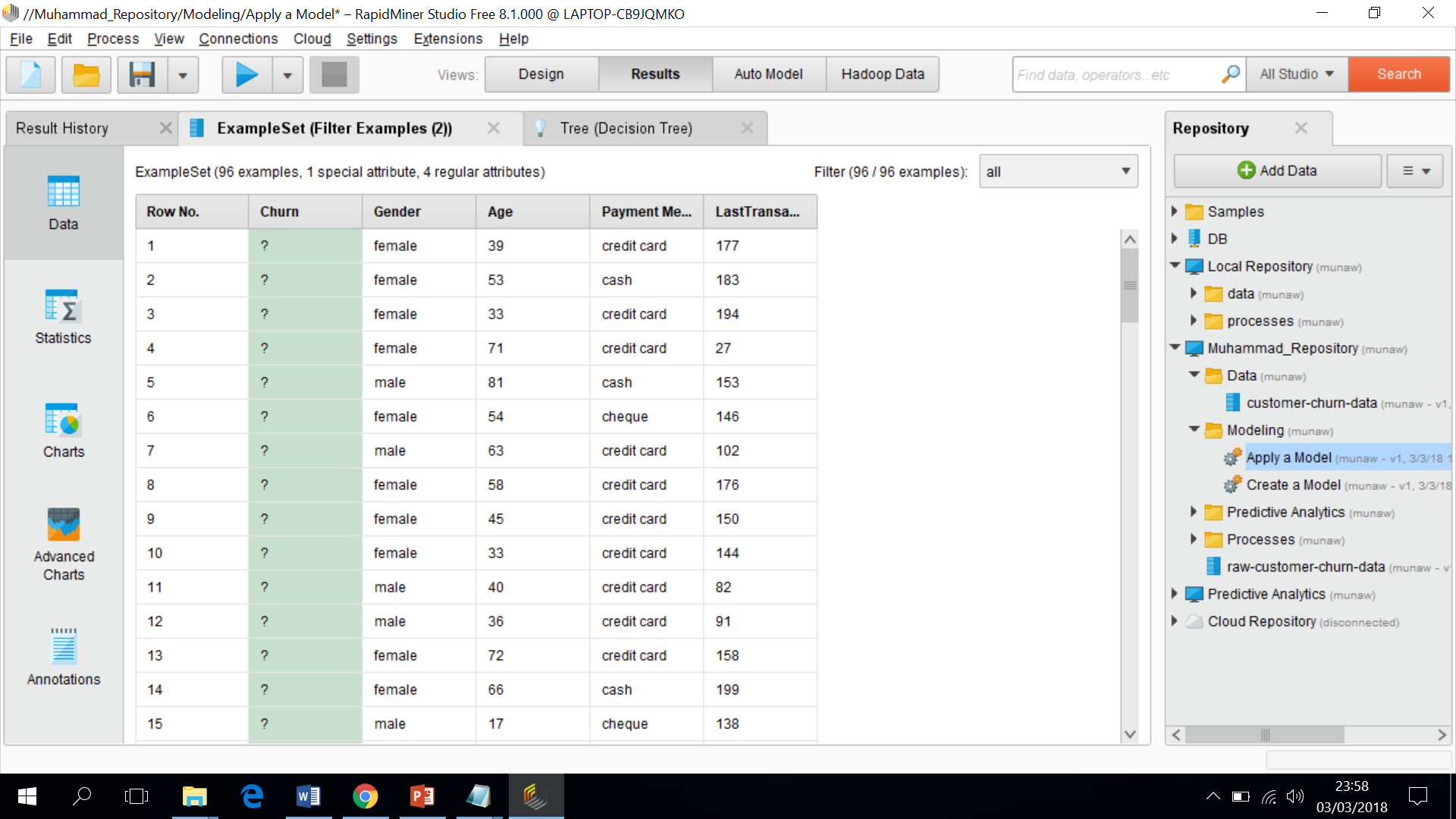


19) Use “Multiply” operator from the Operator pane and drag into process window.

20) For second “Filter example” operator, select the button “Add Filters” and select the attribute “Churn” with an option “is missing”. Join the Second Filter example **port (exa)** to the output port. We divide the data into two parts as shown in the figure below.



21) We obtain the data for the 96 examples as shown below

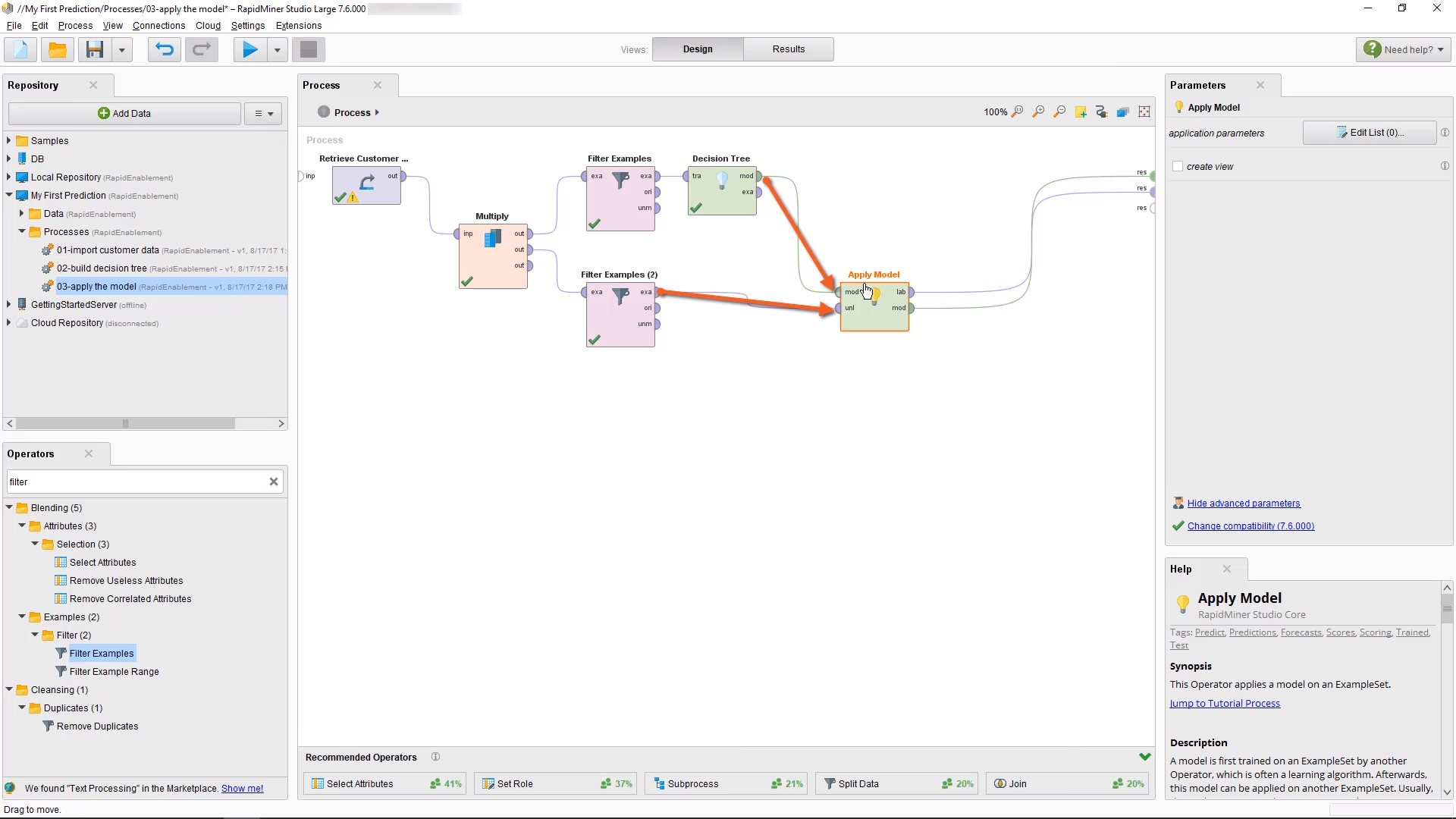


And the other tab shows the decision tree diagram for the customers with missing values.

22) Now we need to apply the model to understand the behaviour of customers in the next step.

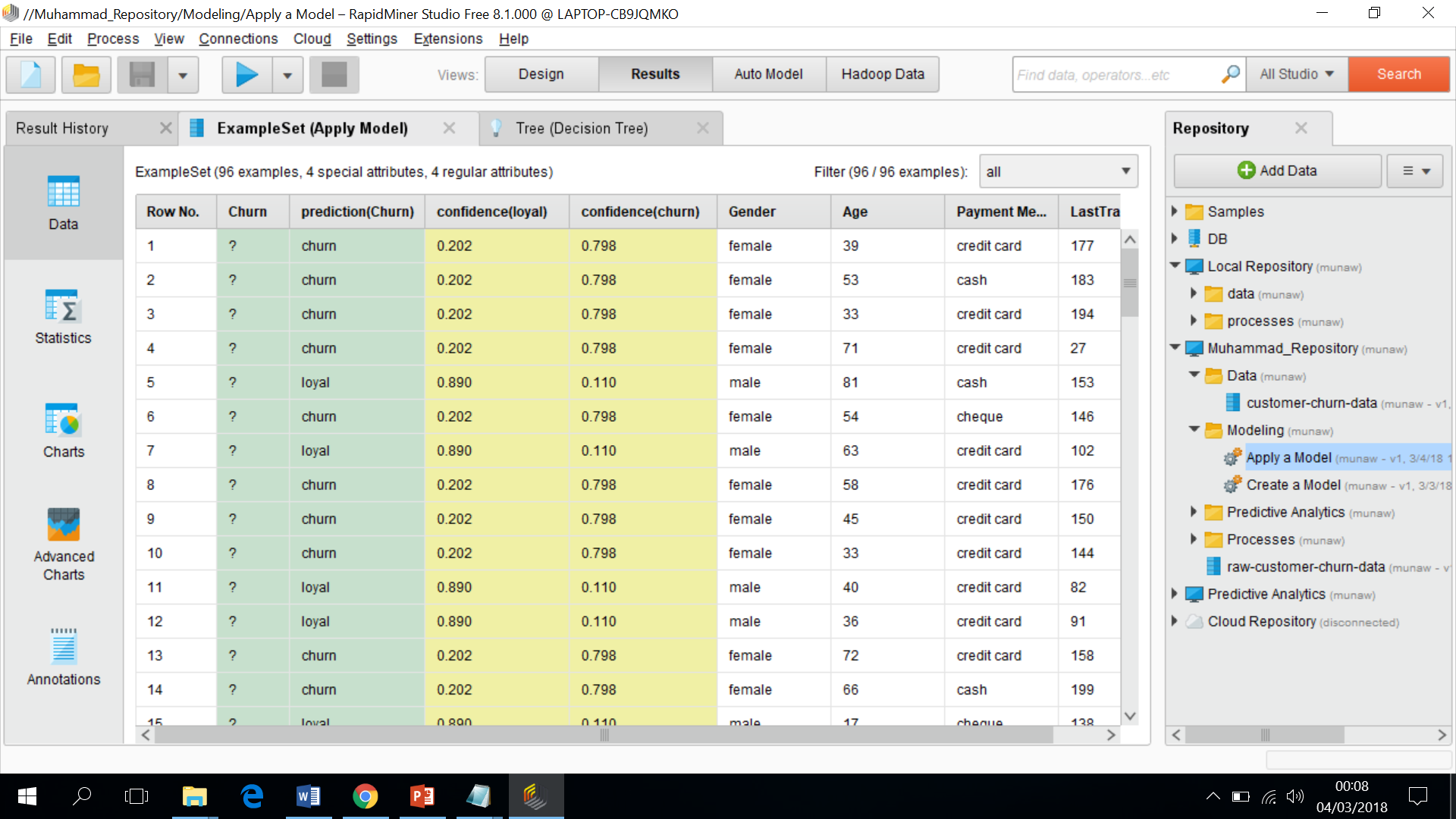
23) Select the operator “Apply Model” and drags into process window.

24) Drags on the lower wire to join with both operators and the following window appears as



25) One port in the “Apply Model” shows the label port and other exhibits the model port. Save the entire process before running it.

26) Execute it now, finally, we got the results for the customers with missing data values in the data set as mentioned below



27) We got three new attributes, Prediction, Confidence (loyal) and Confidence (Churn) in the above window.

28) You have successfully predicted the results based on the missed customer data and now we test the model.

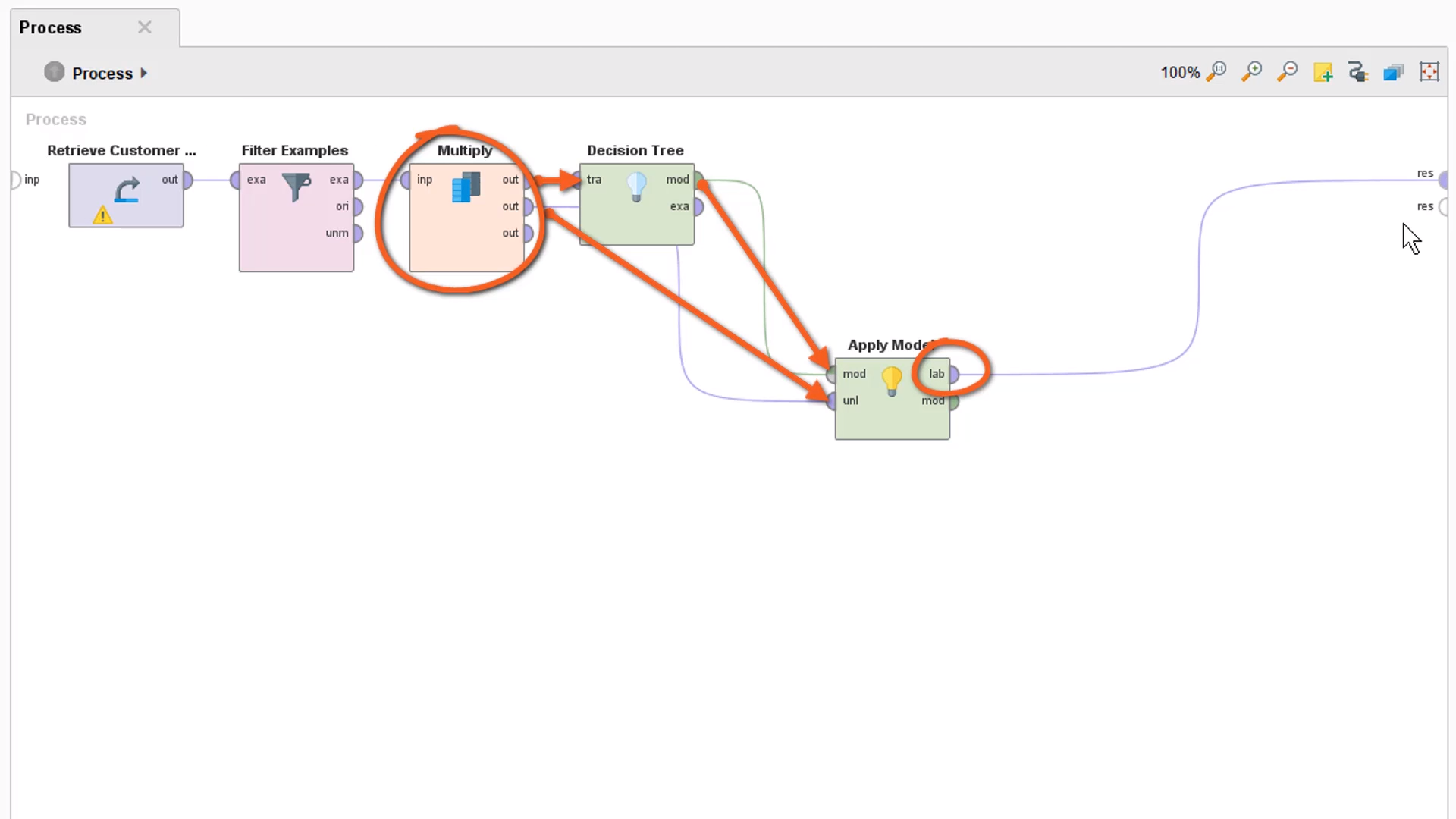
29) The first step is to train the model and then test the model because we do not know the performance of our system.

30) Select the operator “Apply Model” and connect the decision tree port to the “Apply Model” operator.

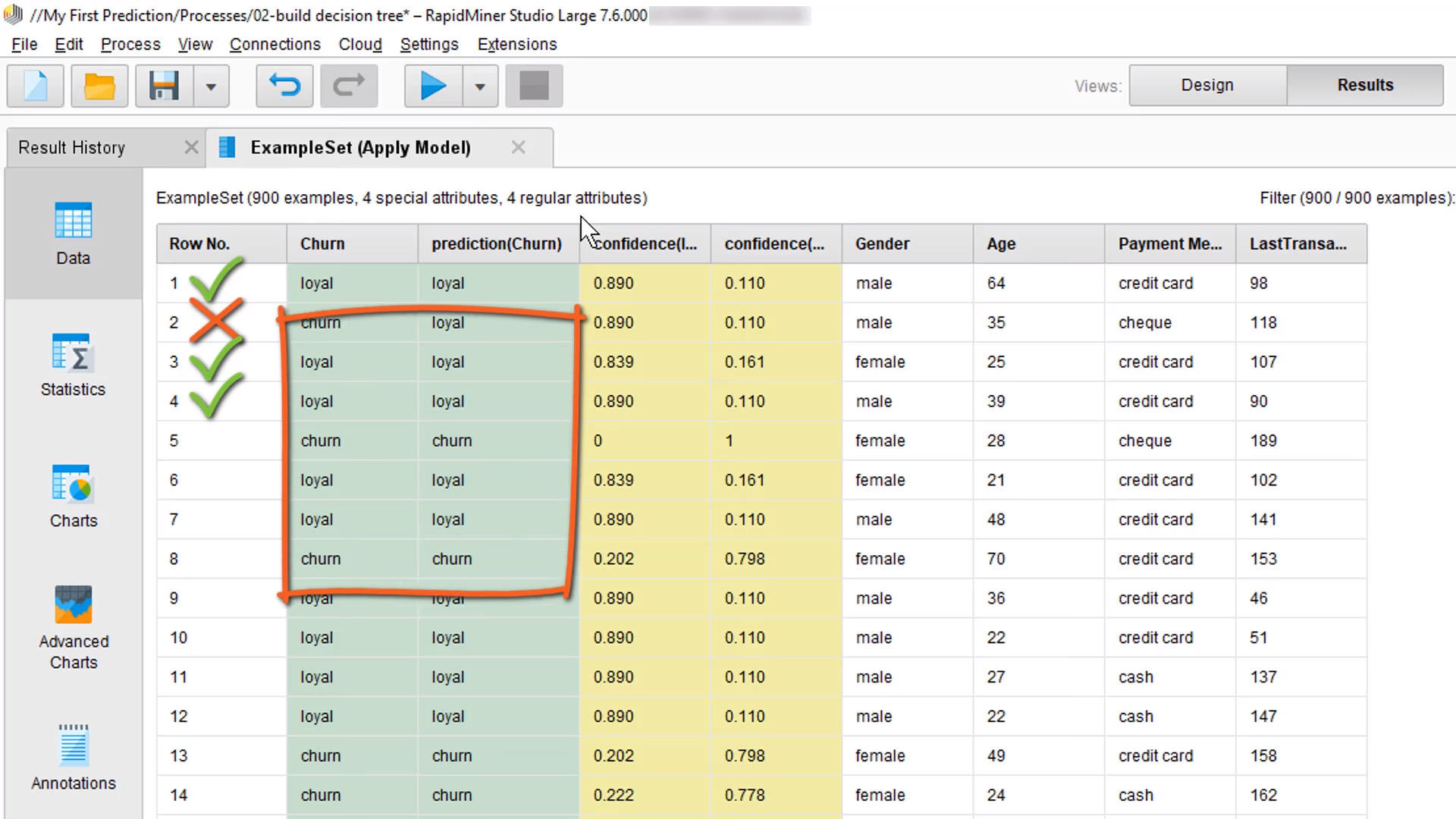
31) In “Apply Model” operator, the second port is called as “unl” is used for the prediction of data.

32) Port “exa” from the “Filter Example” operator connect with port “unl” of the operator “Apply Model” as shown in the next screen shot.

33) Use “Multiply” operator to get data on both operators and the process window should like this as mentioned below



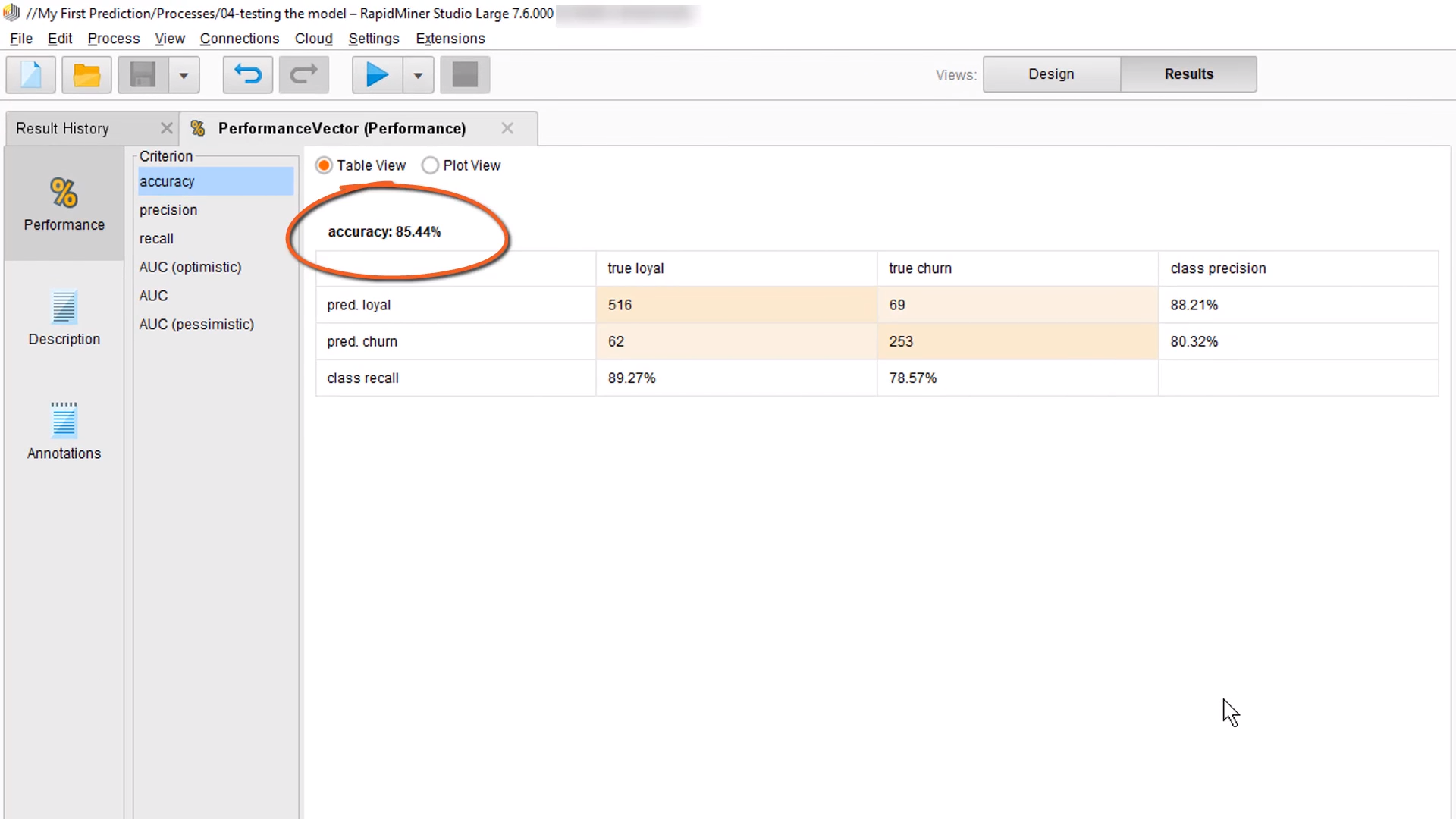
34) Execute the model and you will see the results as mentioned below



This shows the data present in the data file and after prediction. Some predictions are correct and some are not correct.

35) In order to measure the precise performance, we use the operator, “Performance” in the “Segmentation” folder in the operators pane and drags to the end of the process.

36) Before the final execution, save the process. Execute the process and you will get the following window as mentioned blow



You got the accuracy of 85.44%.

37) What we did so far, For example, we provided 100 questions to the students and tested their answers based on the model.

* If you like to explore further, you can watch a youtube video as mentioned below
* <https://www.youtube.com/watch?v=AduZ6TurpDI>
* <https://www.youtube.com/watch?v=DMMzs5X-BNg>