

Linear Regression Using Rapid Miner

Summary and Research



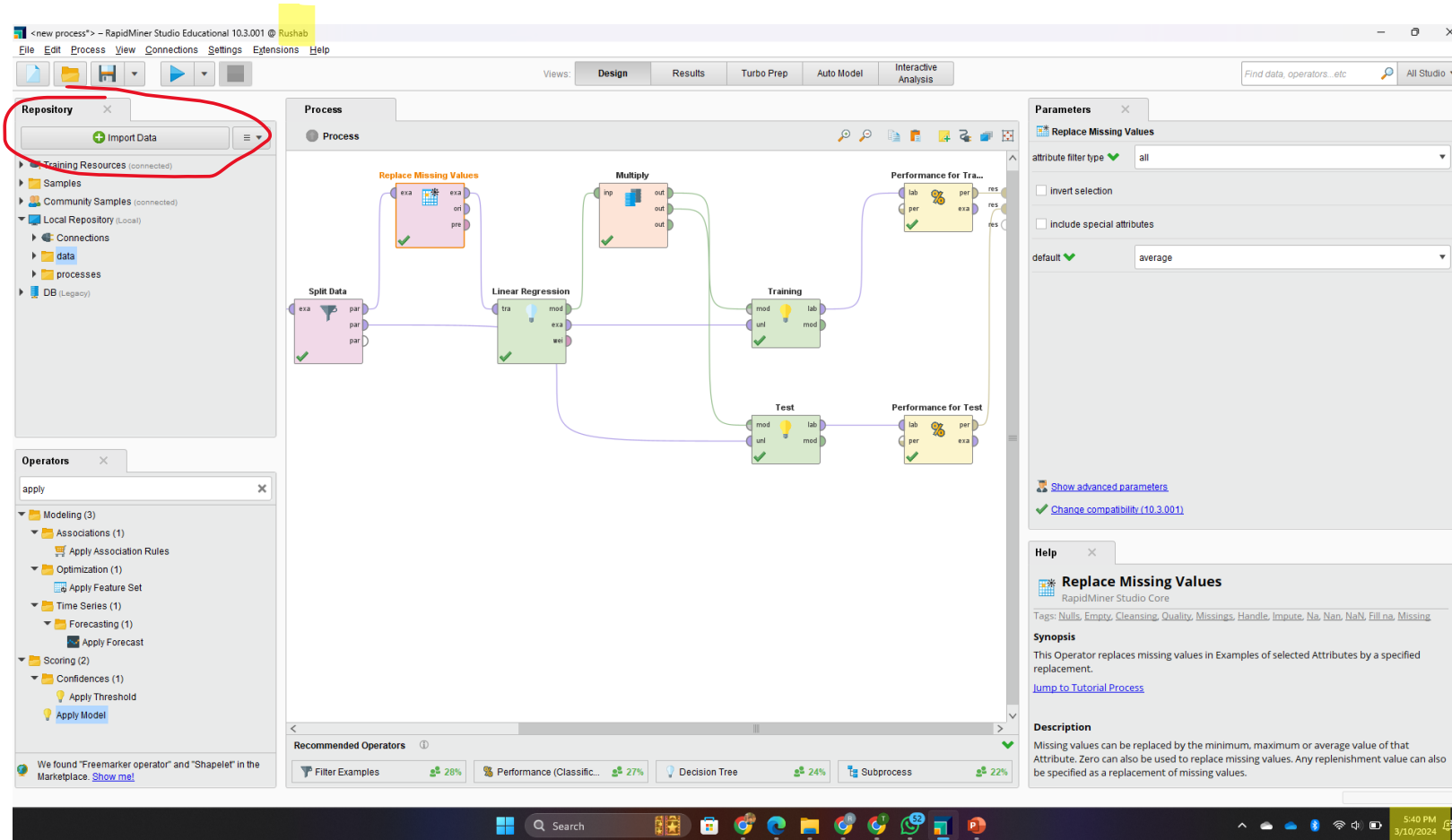
Weather Conditions in World War Two

Goal: Is there a relationship between the daily minimum and maximum temperature? Can you predict the maximum temperature given the minimum temperature?

Link:

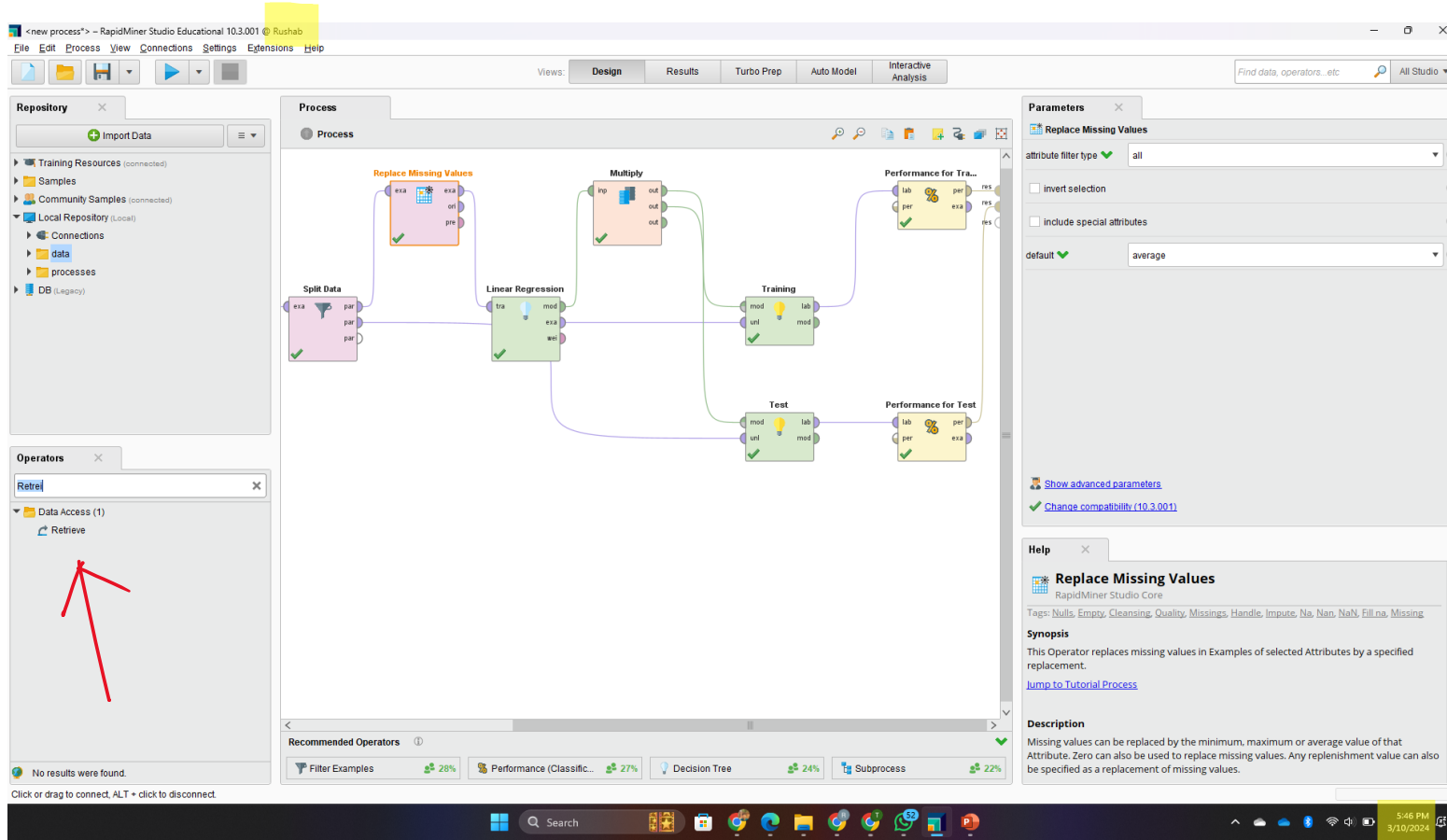
<https://www.kaggle.com/datasets/sm80/weatherww2?resource=download>

Adding the Data into Rapid Miner



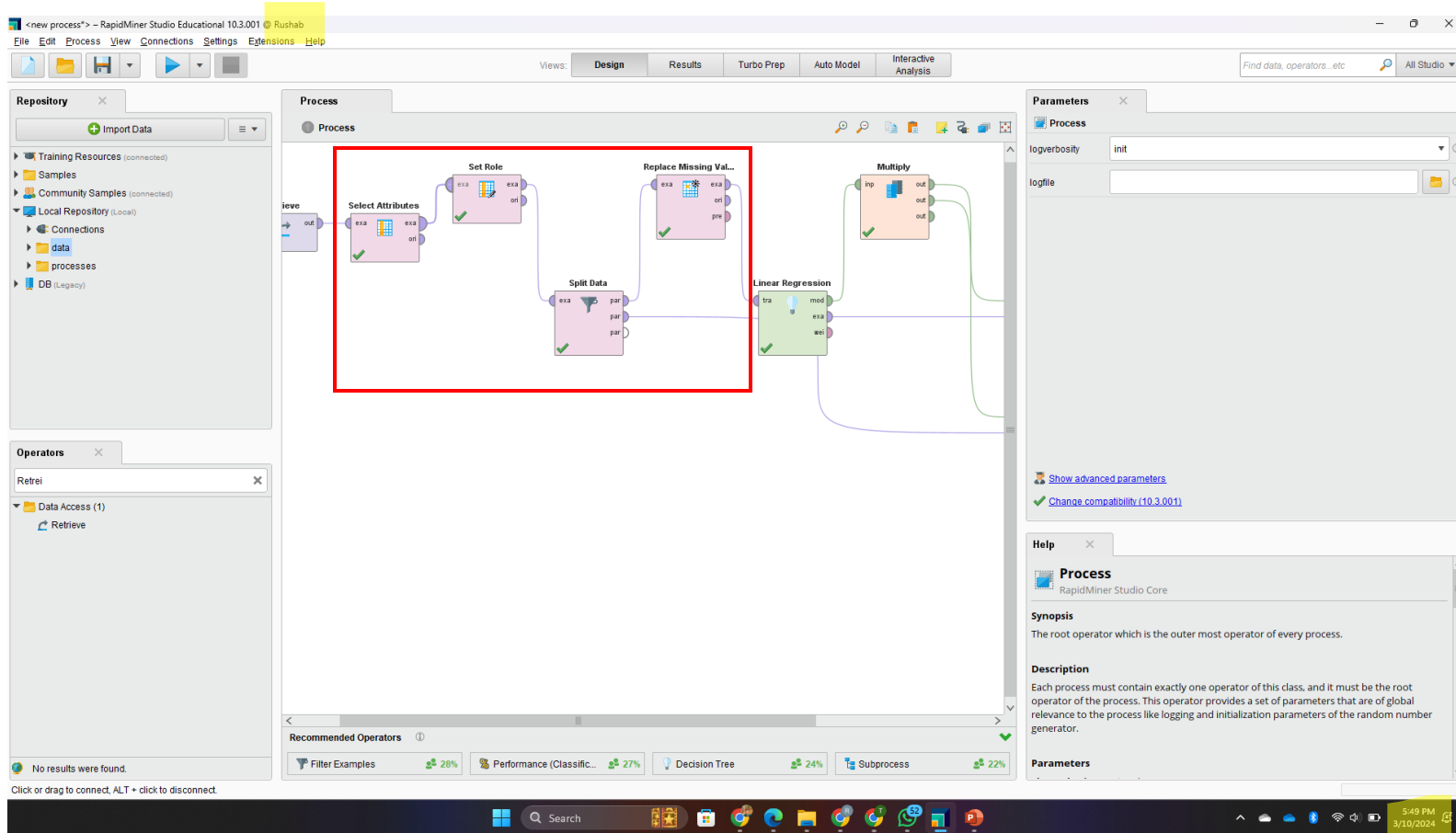
Click on Import
Data>Computer>
Select the file
required

Retrieving Data into Design Area



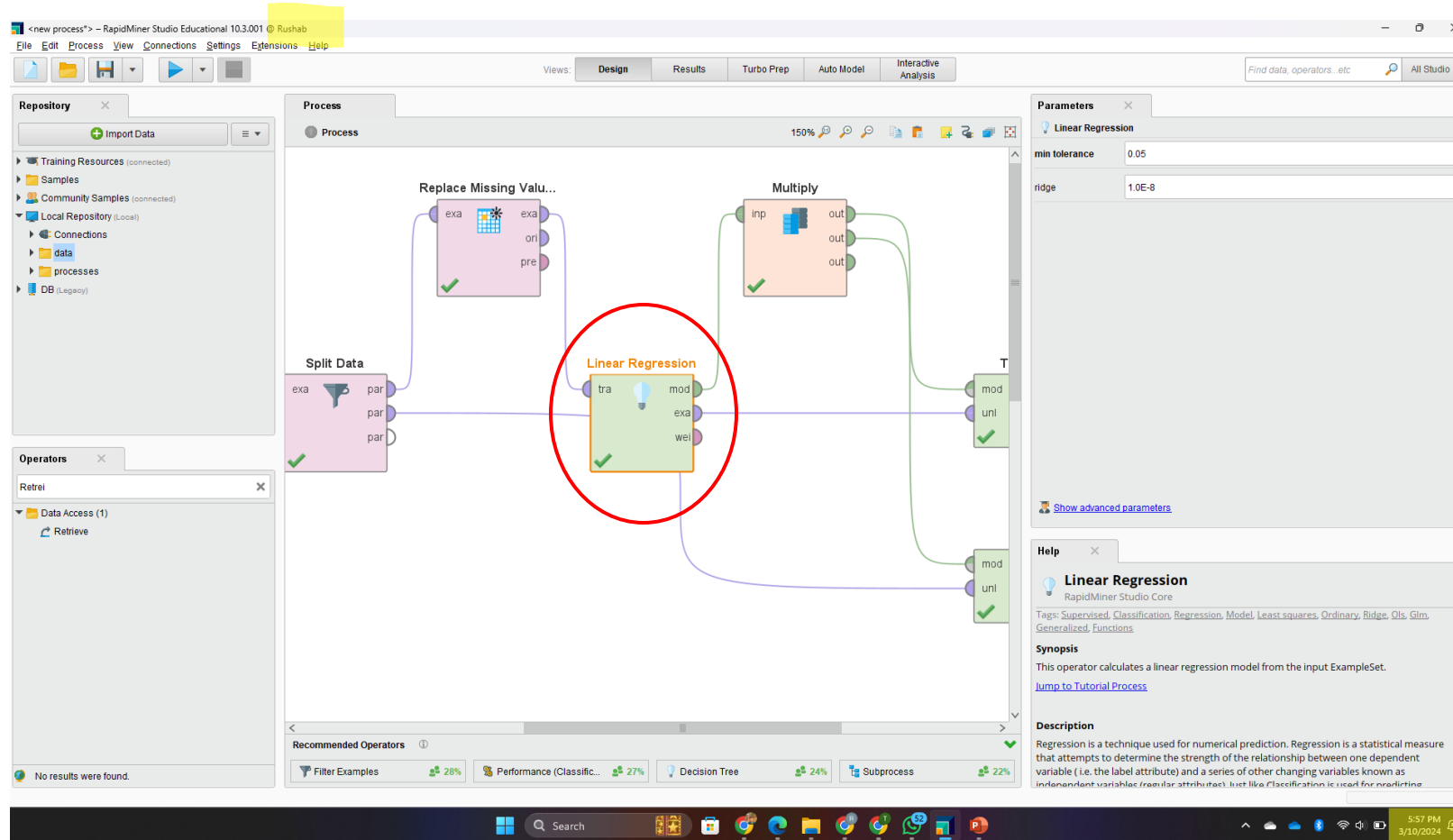
- In the Operators Panel search for Retrieve and drag it to the Design Area.
- Then double-click on the retrieve panel and select your file added in Local Repository

Adding Filters (Attributes, Split Data, Replacing Missing Values)



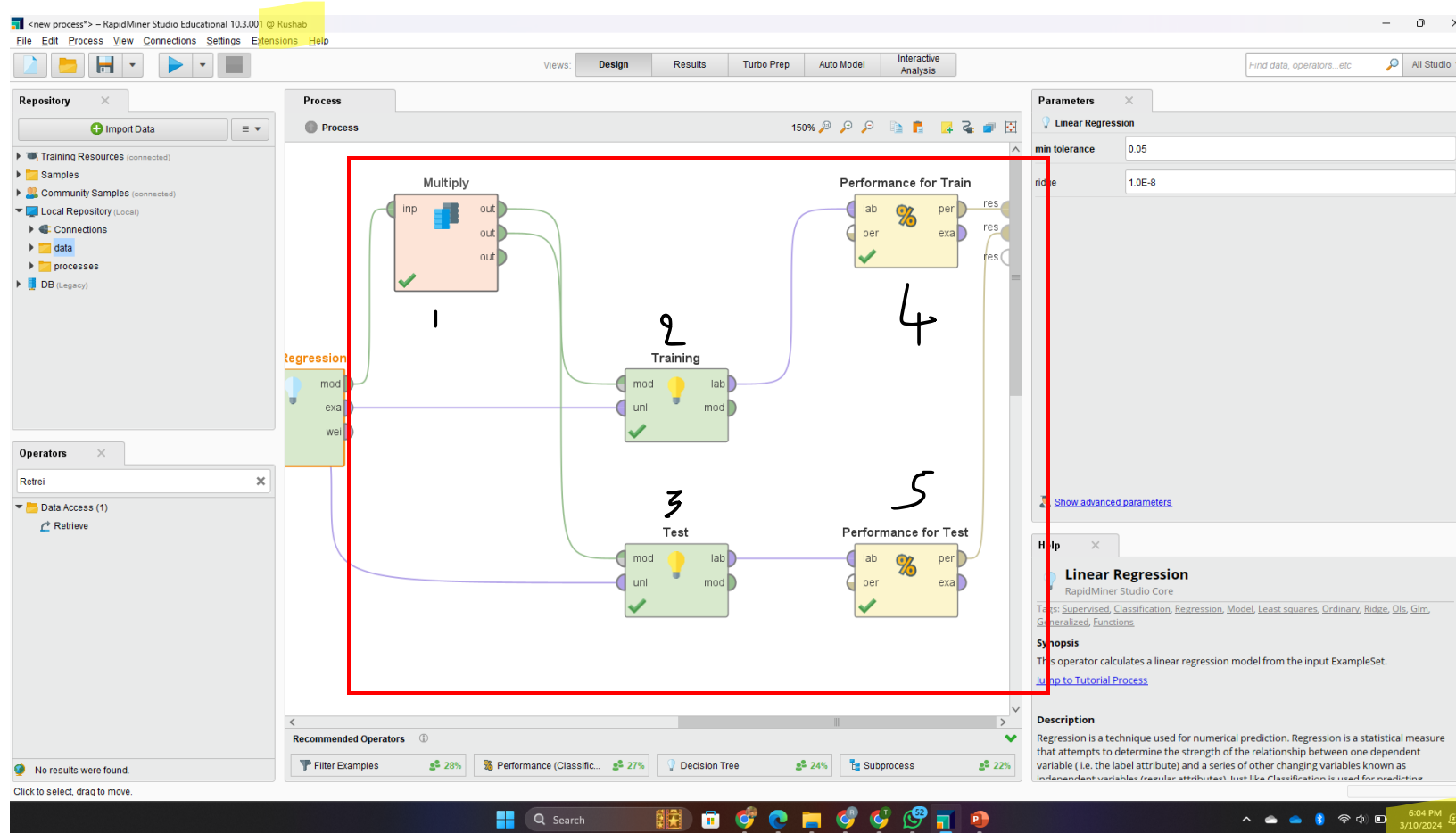
- **Attribute:** Selecting the required columns from the dataset for prediction.
- **Set Role:** To give a certain role/label to attributes
- **Split Data:** Split data into training and testing.
- **Replace Missing Values:** Since linear regression for missing/null values is not possible we replace them with average/zero/minimum/maximum.

Adding Linear Regression in the model



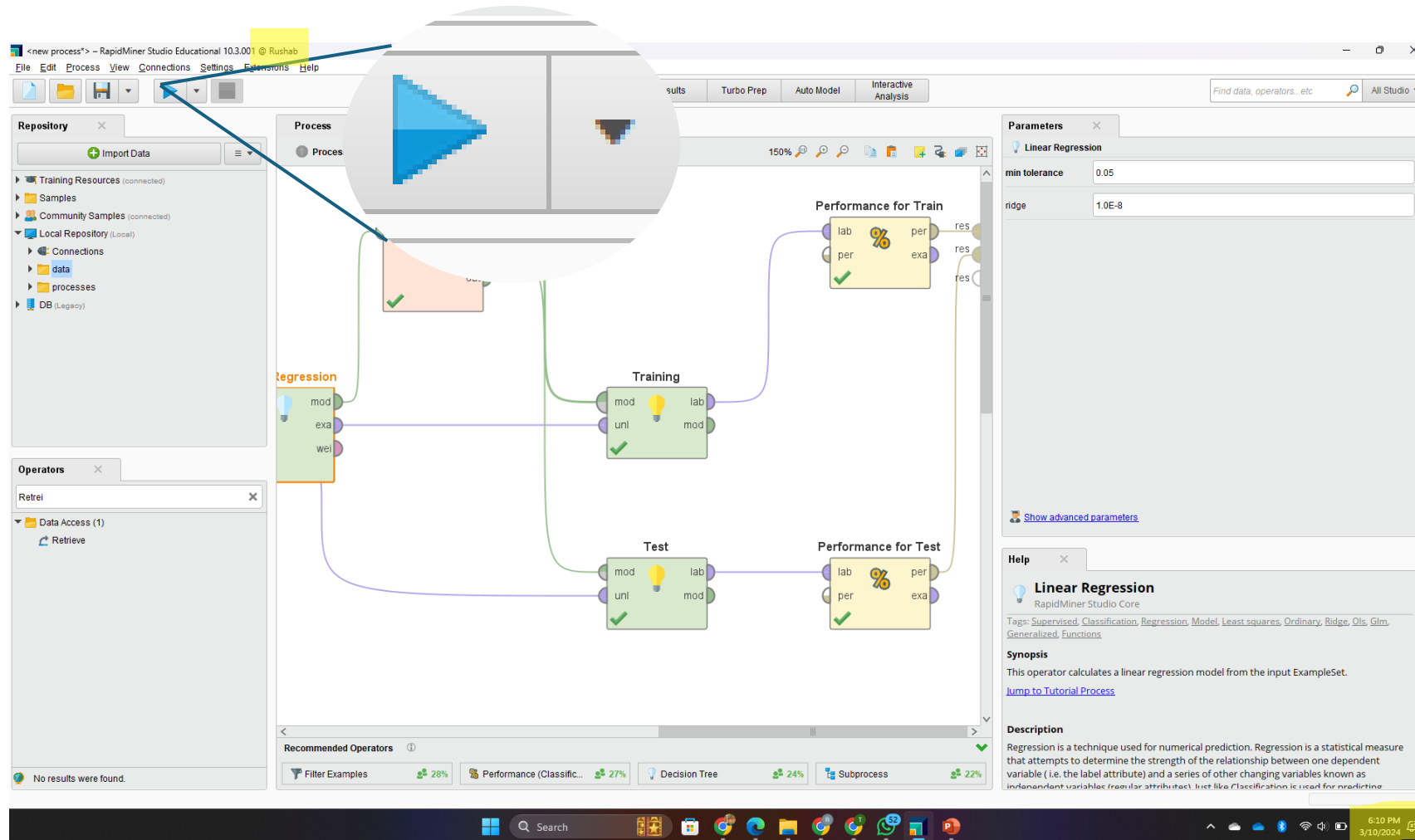
Then we add the Linear Regression Operator to the model for our prediction.

Adding to model outputs for Training and Test Data



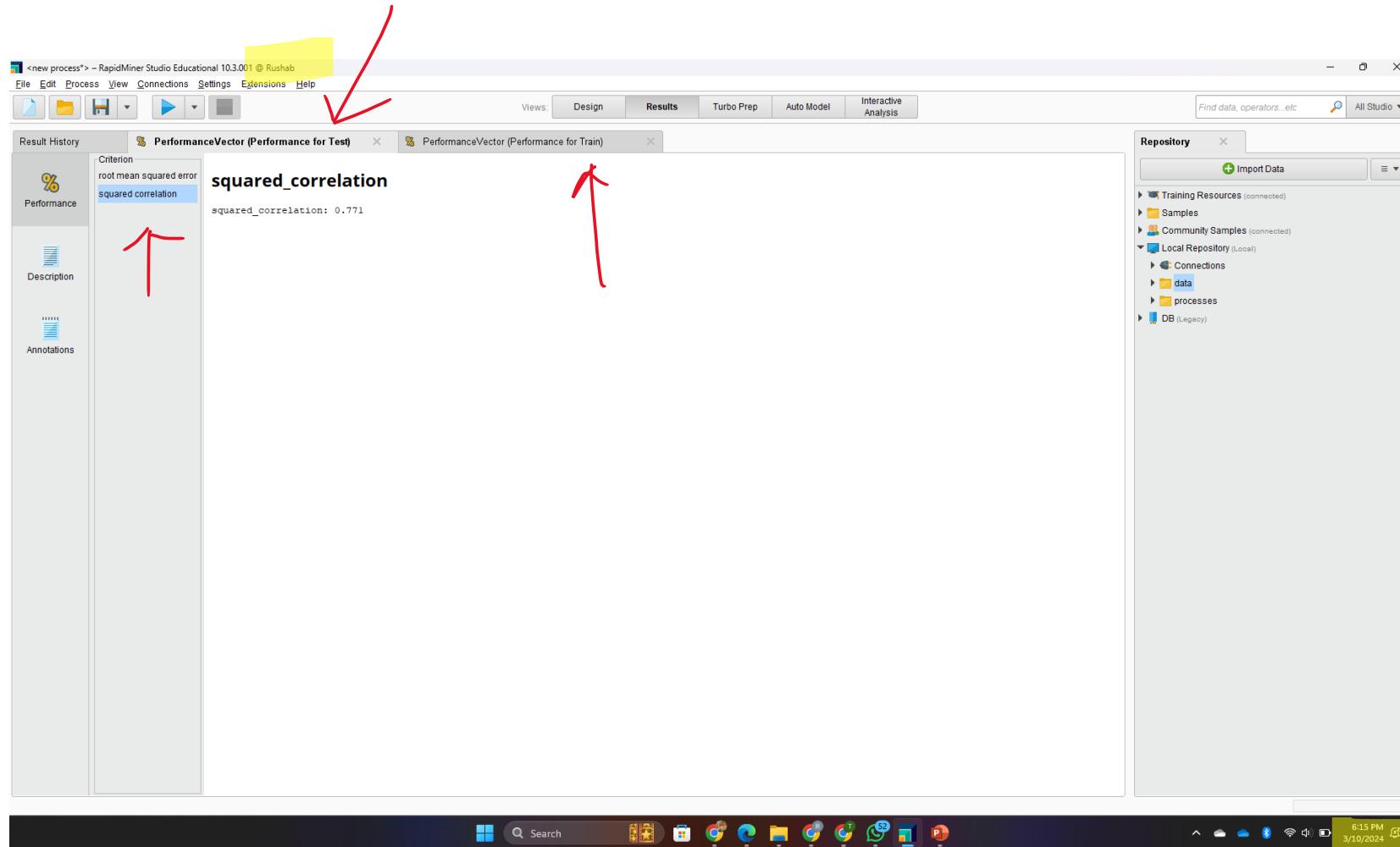
Then we separate the **linear regression** model to our Test and Training Data. Linear Regression containing the **training** data and other model output have **test** data.

Running the Model



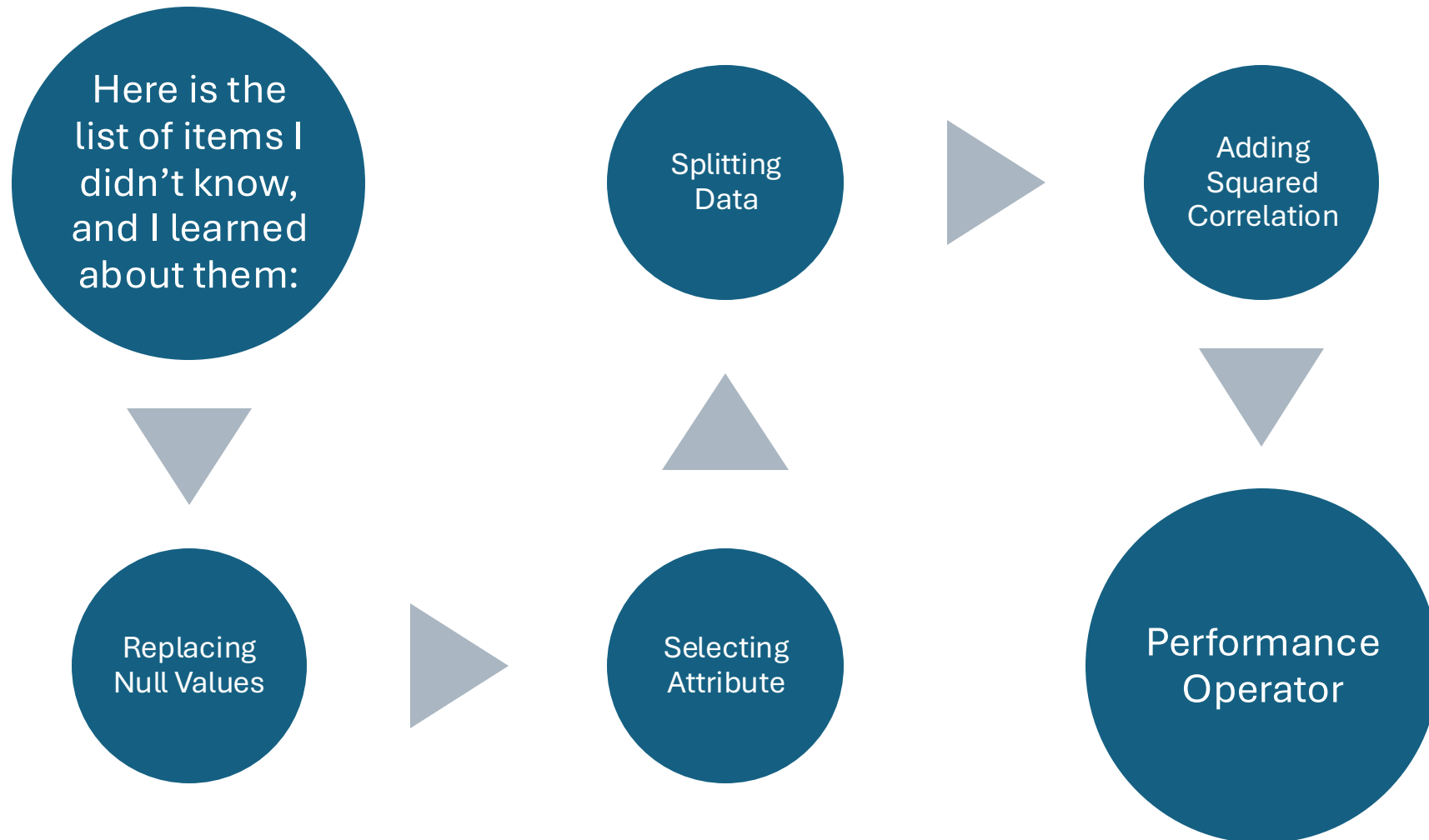
- To Run press the “Blue” Play button.
- It will first show any errors and if the model is errorless you will be prompted to Result Screen

Output



This is the result screen for both test data and training data

Research Item





Replacing Null Values

- There are N number of ways we can deal with Null values and the way I learnt is using the “Replace Missing Values” Operator which helps replace them with Zero/Average/Min/Max.

```
mirror_mod = modifier_ob.  
Set mirror object to mirror  
mirror_mod.mirror_object  
operation == "MIRROR_X":  
mirror_mod.use_x = True  
mirror_mod.use_y = False  
mirror_mod.use_z = False  
operation == "MIRROR_Y":  
mirror_mod.use_x = False  
mirror_mod.use_y = True  
mirror_mod.use_z = False  
operation == "MIRROR_Z":  
mirror_mod.use_x = False  
mirror_mod.use_y = False  
mirror_mod.use_z = True
```

```
selection at the end -add  
mirror_ob.select= 1  
modifier_ob.select=1  
context.scene.objects.active  
("Selected" + str(modifier_ob.  
mirror_ob.select = 0  
= bpy.context.selected_objects  
data.objects[one.name].select  
print("please select exactly
```

--- OPERATOR CLASSES ---

```
types.Operator):  
X mirror to the selected  
object.mirror_mirror_x"  
mirror X"
```

```
context):  
context.active_object is not
```

Selecting Attribute

When I was trying the regression at first I used the whole dataset which led to a software crash. Then after learning about the “Selecting Attribute” panel I just selected the columns needed.



Splitting Data

- Since the data I used was around 100000 columns it was hard to train that huge dataset. So the “Split Data” operator helped me rationalize the data between Training and Test which helped in better understanding the genuineness of the model.



Performance Operator

To show the statistical performance of regression model.



Squared Correlation

- Squared Correlation aka RSqaure Value. Helps determine the prediction strength of the model.
- The less the RSquare value better the model.



Thank
you

