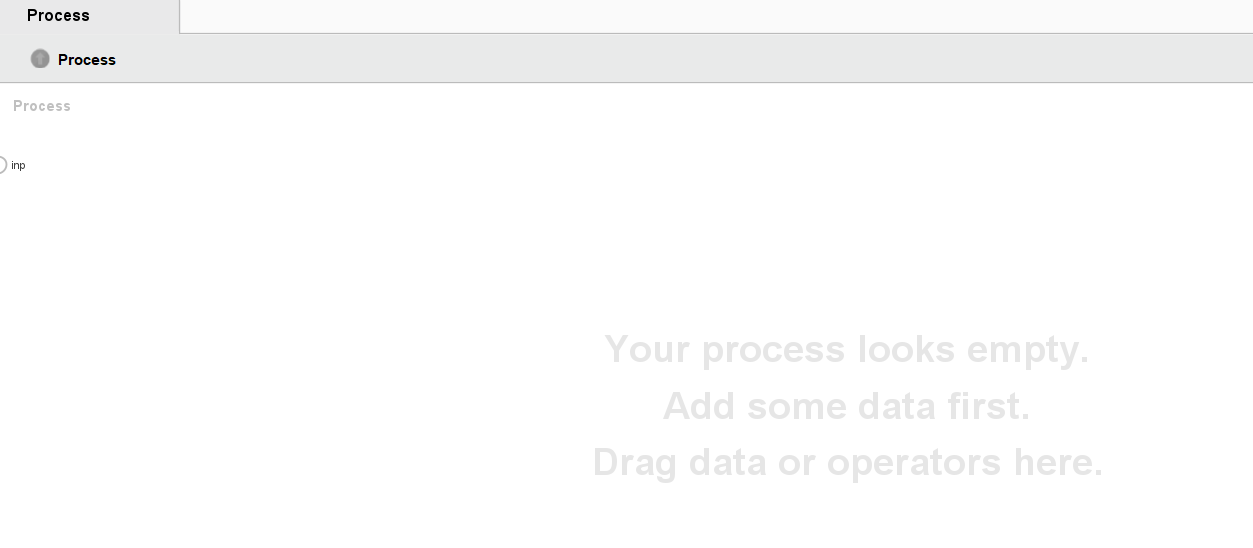
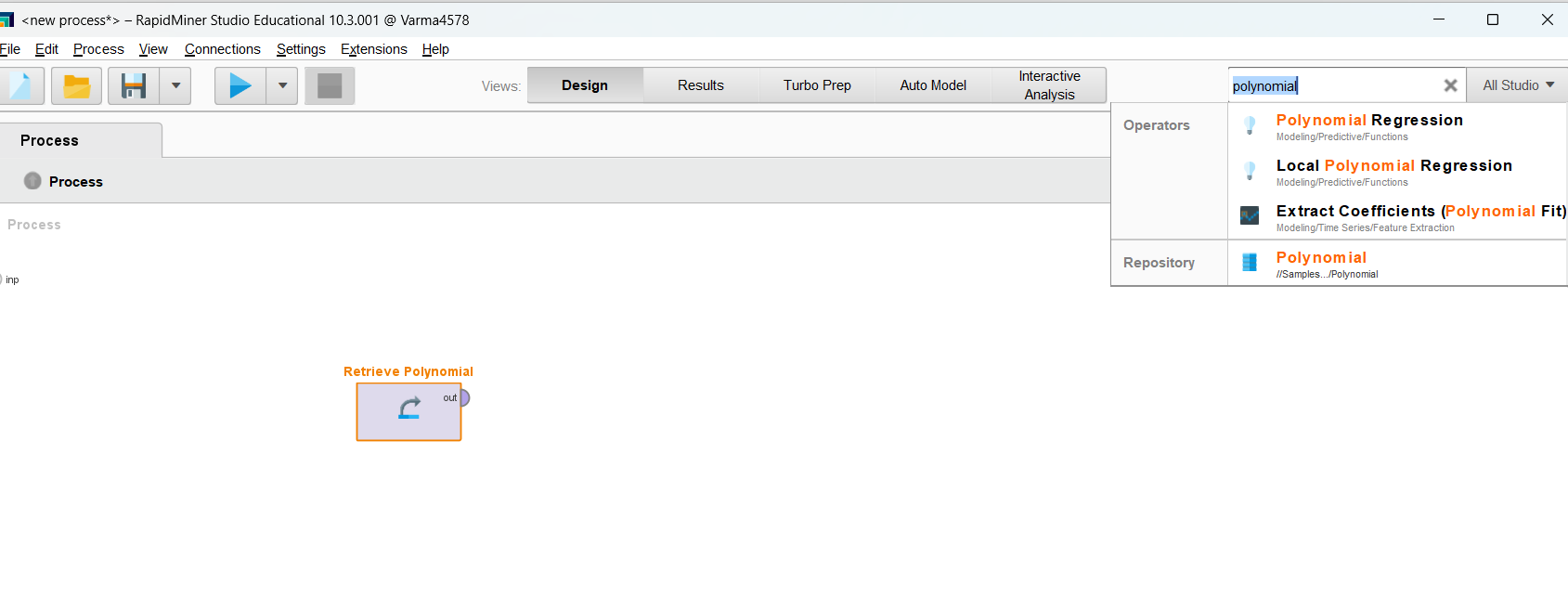
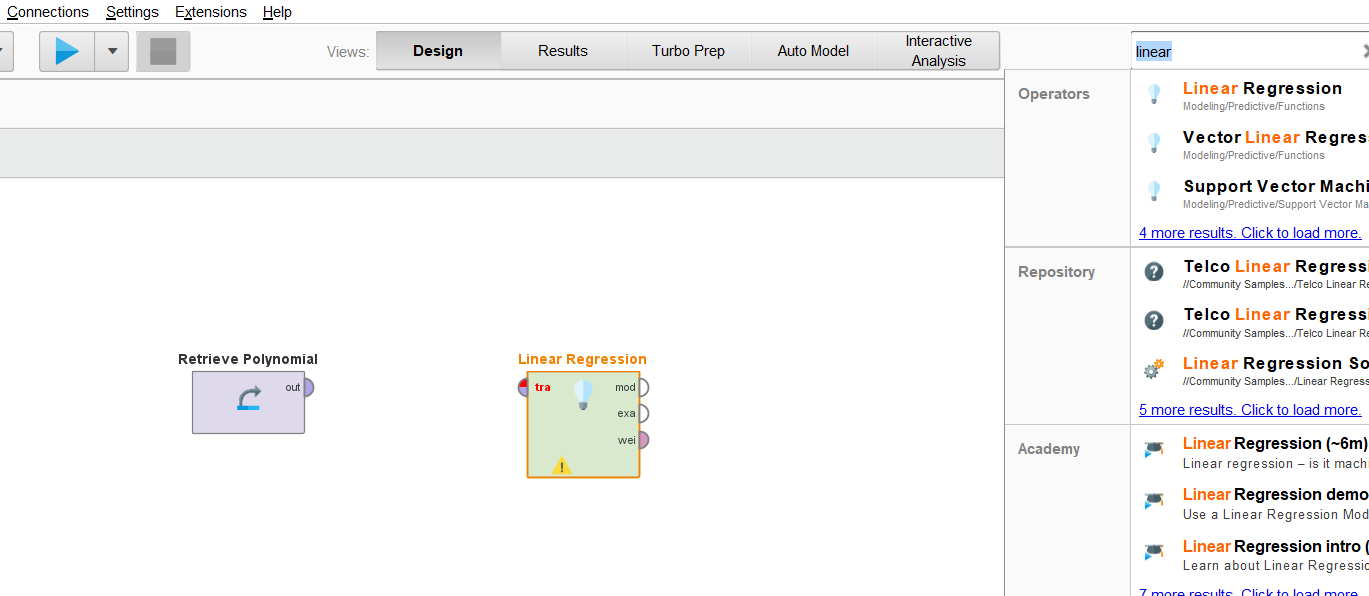
**Lab 6**

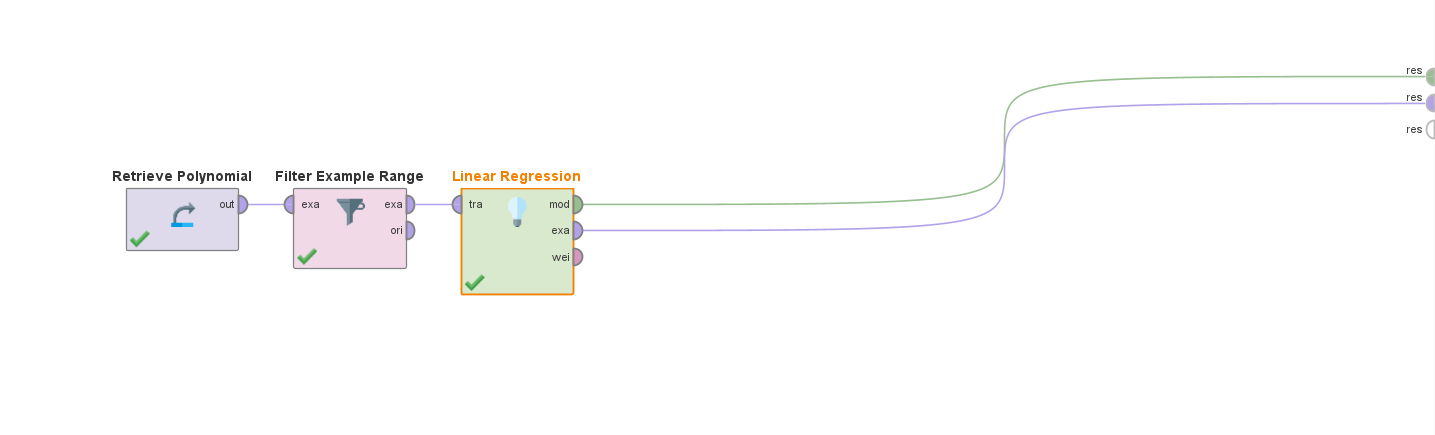
**Start with a blank process canvas. Expand Samples, data, and then drag the Polynomial dataset onto the process canvas.**

****

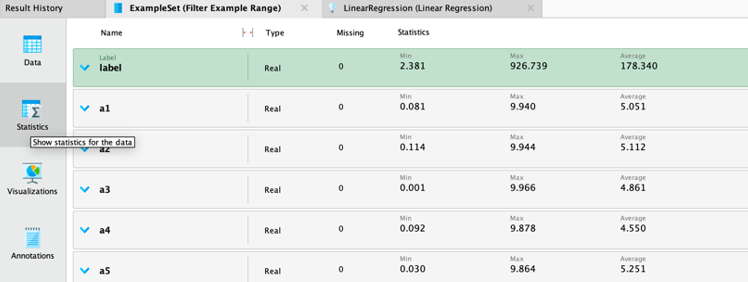
**Part I.** Regression ****

****

**Connect the dataset out to the Filter Example Range, this is accomplished by holding down the mouse and dragging the line from one half bubble (out) to the other half bubble (exa).**

****

A screenshot of a computer

Description automatically generated****A screen shot of a graph

Description automatically generatedA screen shot of a graph

Description automatically generatedThe coefficient, standard error, standard coefficient, and tolerance are crucial metrics in statistical analysis, especially in regression analysis, as they help interpret the relationship between independent variables (predictors) and dependent variables (outcome). Let's dissect their meanings in light of the given data:   
(A1, a2, a3) Coefficient: Keeping other variables constant, these values show the estimated impact of each independent variable on the dependent variable. When one unit changes in the first independent variable, for example, and the other variables remain constant, the dependent variable changes as well (a1 represents the coefficient of the first independent variable).  
The coefficient estimate's precision or variability is indicated by the standard error. A smaller standard error denotes more accurate estimations, which implies that the coefficient is probably closer.

The coefficient value that has been standardized by dividing it by its standard error is known as the standard coefficient. In particular, when the predictors are measured on different scales, it enables a comparison of the relative importance of each predictor variable in influencing the dependent variable.   
Tolerance is a measure of how well a predictor variable can be predicted by other independent variables in the model. It is the reciprocal of the Variance Inflation Factor (VIF). When a tolerance is near to 1, it means that the variable adds distinct information to the model and is not redundant with other predictors.   
There are a number of reasons why a4 and a5 might not be included in the model, including multicollinearity, insignificance, or lack of relevance. There is multicollinearity.

when there is a high degree of correlation between the independent variables, which causes unstable coefficient estimates and interpretive challenges. The term "insignificance" refers to the lack of significant contribution from the variables in explaining the variance in the dependent variable. A lack of relevance indicates that the variables are not necessary to include because they do not, either practically or theoretically, correspond with the desired outcome. Thus, it's possible that these variables were removed in order to enhance the interpretability, stability, and performance of the model.

A screen shot of a graph

Description automatically generated

**Part 2:** **Gaussian Process Model**

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screen shot of a graph

Description automatically generatedA screen shot of a graph

Description automatically generatedA screen shot of a graph

Description automatically generatedA screen shot of a graph

Description automatically generatedA screen shot of a graph

Description automatically generatedA screen shot of a graph

Description automatically generated

**Part 3 – Exit Rapid Miner**

Sucessfully exit from rapid minder