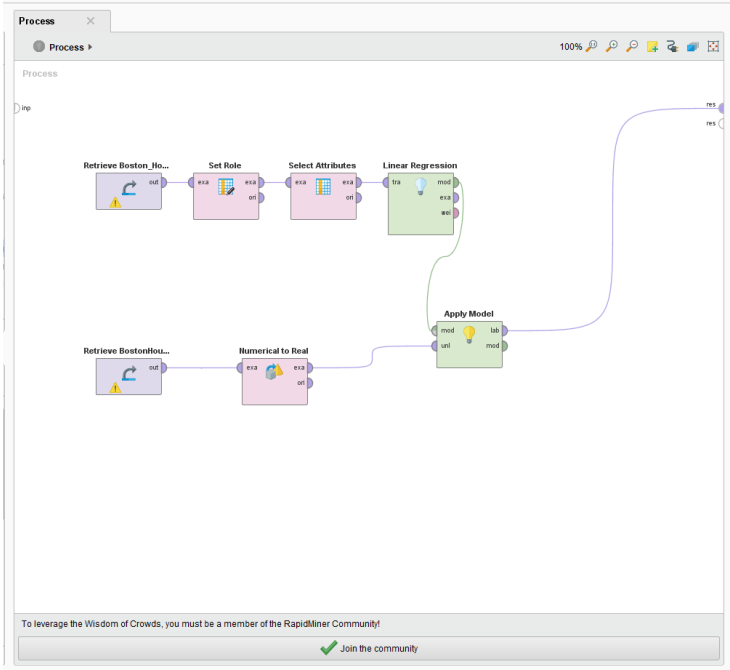
Answer the following questions

Predict New House using given predictors



a) Fit a multiple linear regression model to predict MEDV using CRIM, CHAS and RM as predictors.

- Write the equation for predicting the median house price from the predictors in the model.

Attribute	Coefficient
CRIM	-0.261
CHAS	3.763
RM	8.278
(Intercept)	-28.811

CRIM=-0.261

CHAS=3.763

RM=8.278

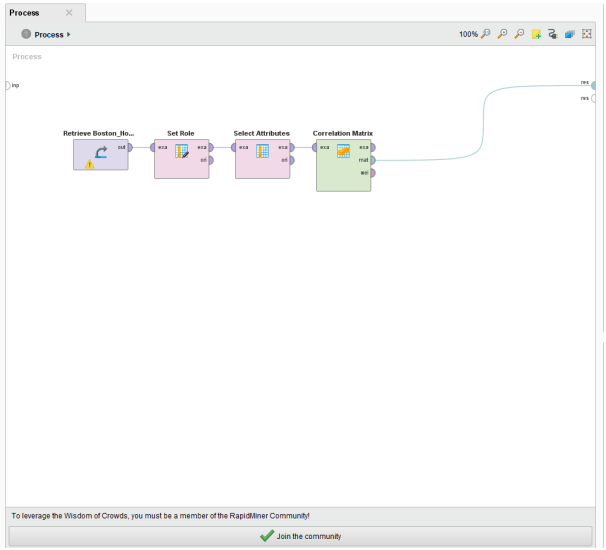
intercept=-28.811

predictio...	RM	CRIM	ZN	INDUS	CHAS
20.806	6	0.200	0	7	0

y = intercept + RM + CRIM + CHAS
y = -28.881 + 8.278(6) - 0.261(0.2) + 3.763(0)
y= 20.7348

Search for possible multicollinearity

b) There are several variables that measure levels of industrialization, which are expected to be positively correlated. These include INDUS, NOX (pollution), and TAX. **We expect a positive relationship between NOX (nitric oxides concentration, a pollutant), INDUS (proportion of non-retail business acres per town) and TAX (tax rate),** because areas that have a high proportion of non-retail businesses tend to have higher taxes and more pollution. These 3 predictors are likely to measure the same thing.

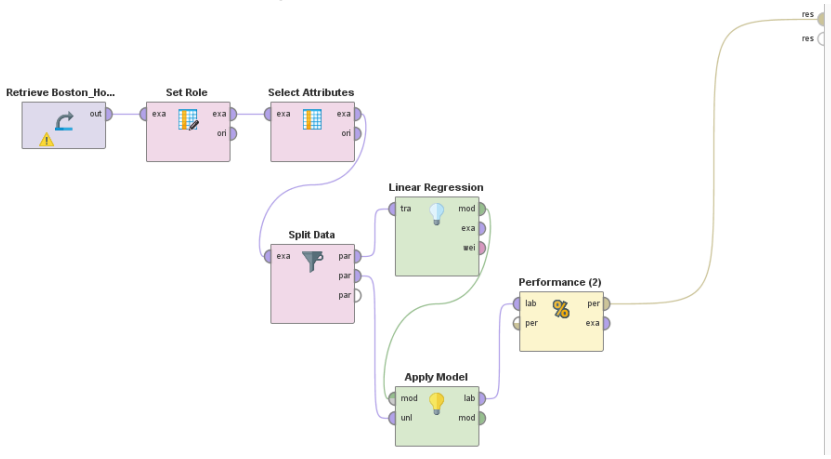


Attributes	INDUS	NOX	TAX
INDUS	1	0.764	0.721
NOX	0.764	1	0.668
TAX	0.721	0.668	1

- NOX and INDUS pair high correlated
- remove INDUS

Reduce the number of predictors & Propose Your Best Model

c) Use a feature selection mechanism to reduce the remaining predictors (from previous step). Run each model separately using the training/testing datasets. Then, give the best model in terms of regression equation. What is RMSE ?



RMSE = 5.236