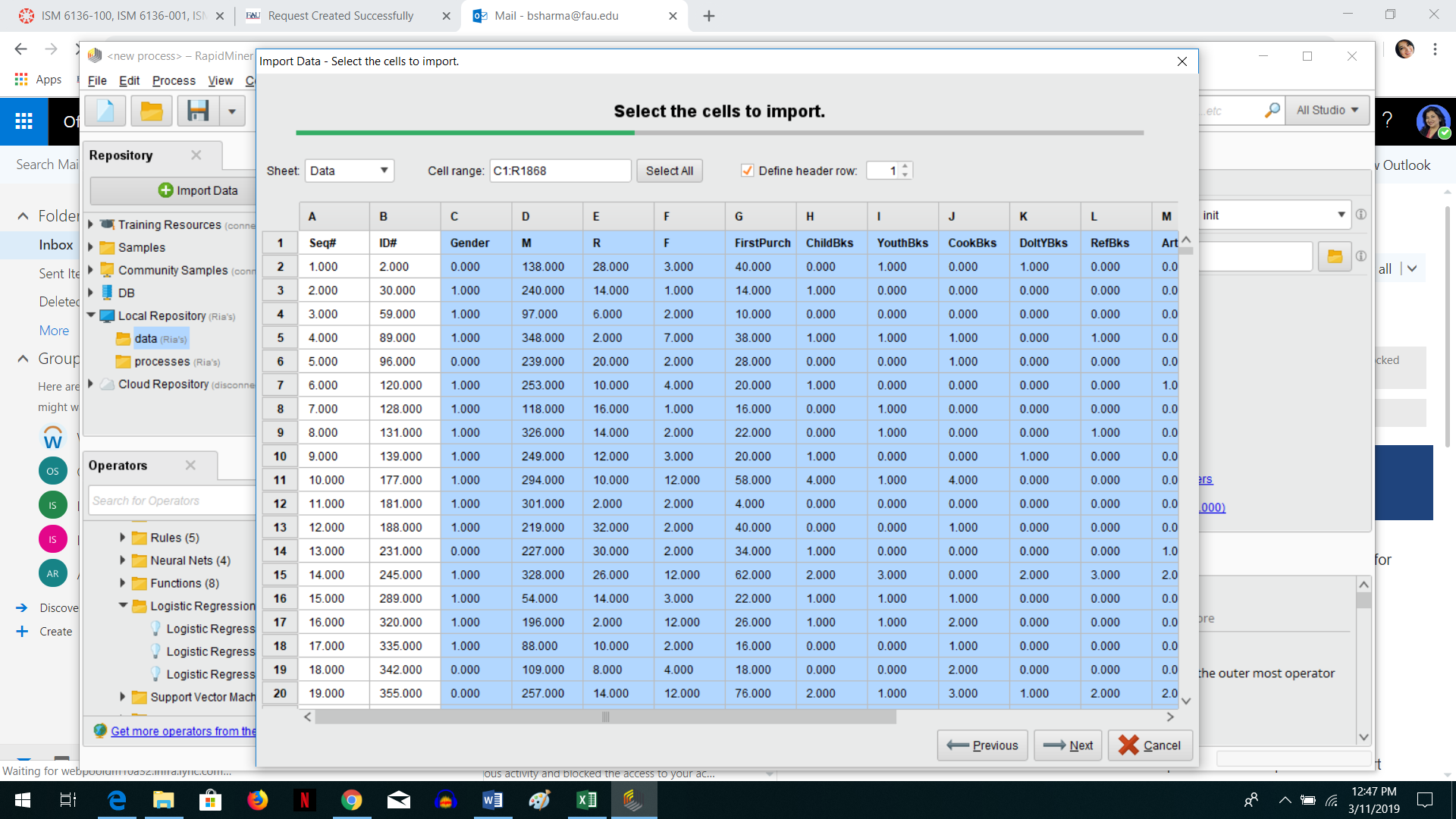
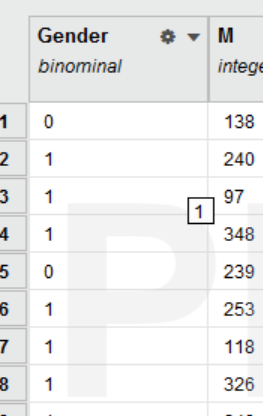
**Data mining on Charles Book Club – using RapidMiner**

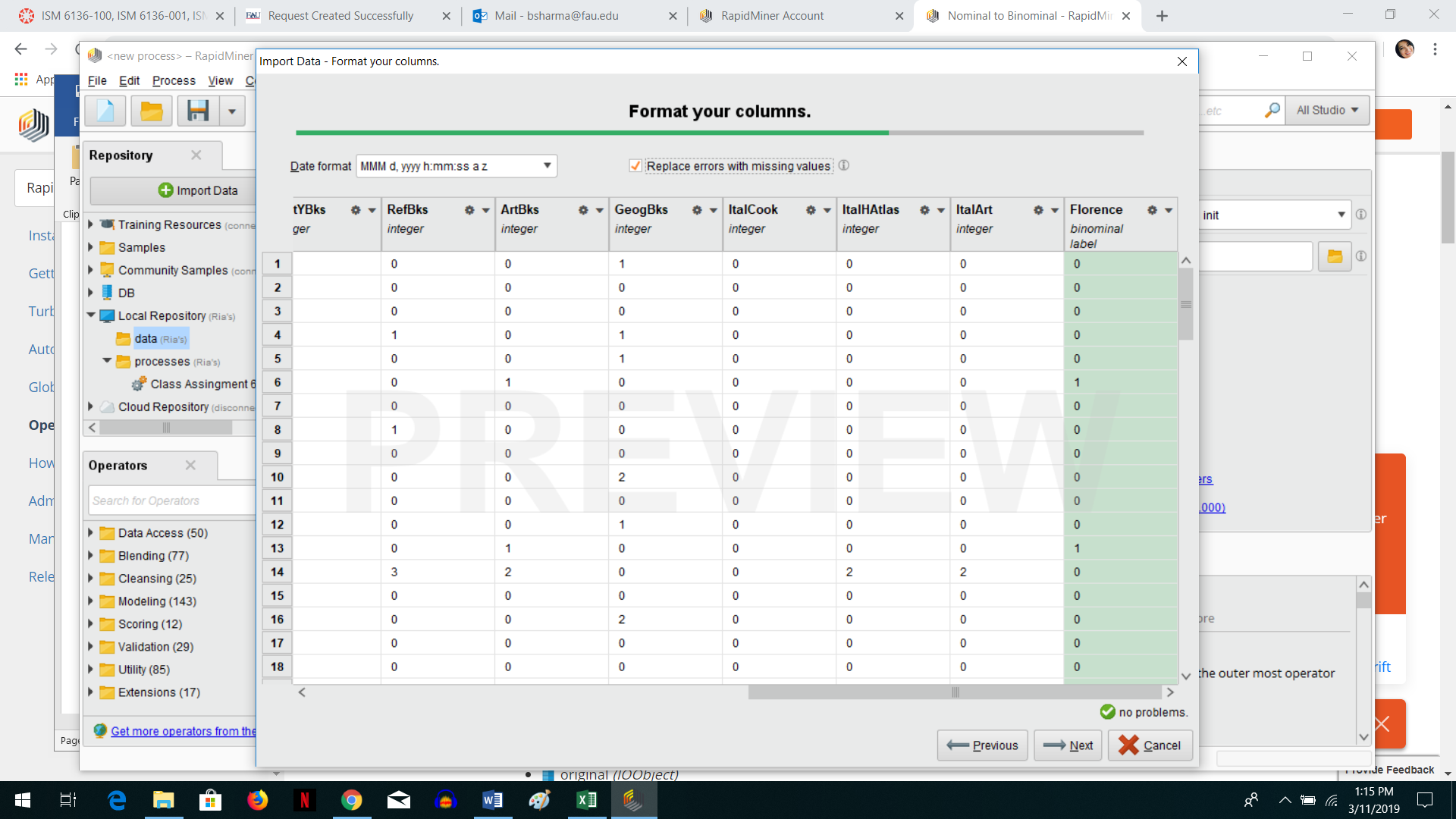
1. First import the dataset and then select the cells that need to be used for datamining.



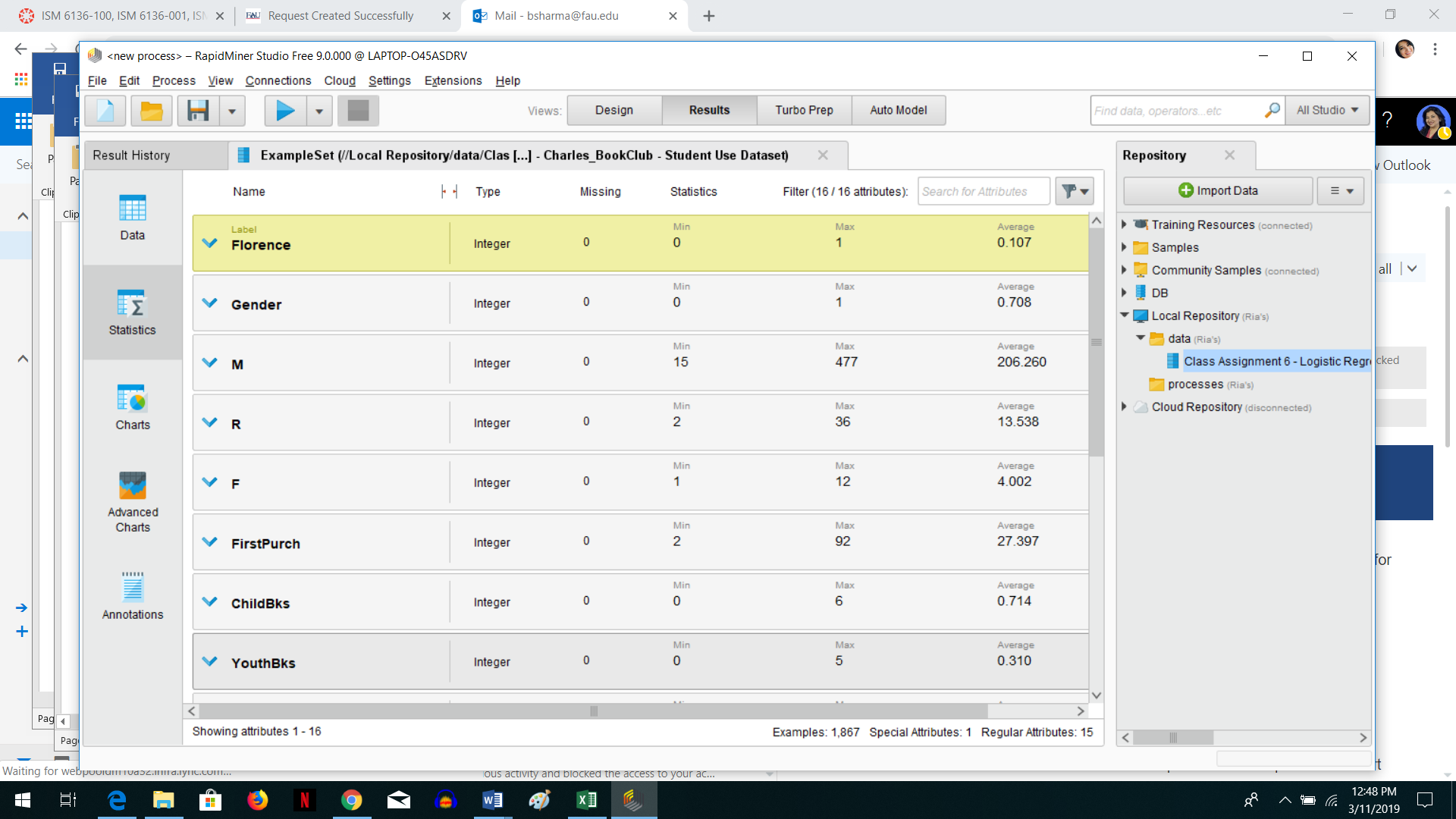
2. Change the Role to Label for Florence – Also check for any Categorical Variables – if there are then their TYPE needs to be changed to ‘binominal’ if they are 1 or 0 OR ‘polynomial’ if they are more than 2 categories.

In this dataset all the predictor variables are ‘numerical continuous’ (except Gender and Florence) and not categories so we do not need to change their types and they can stay as Integer – only Florence and Gender changes to ‘binomial’

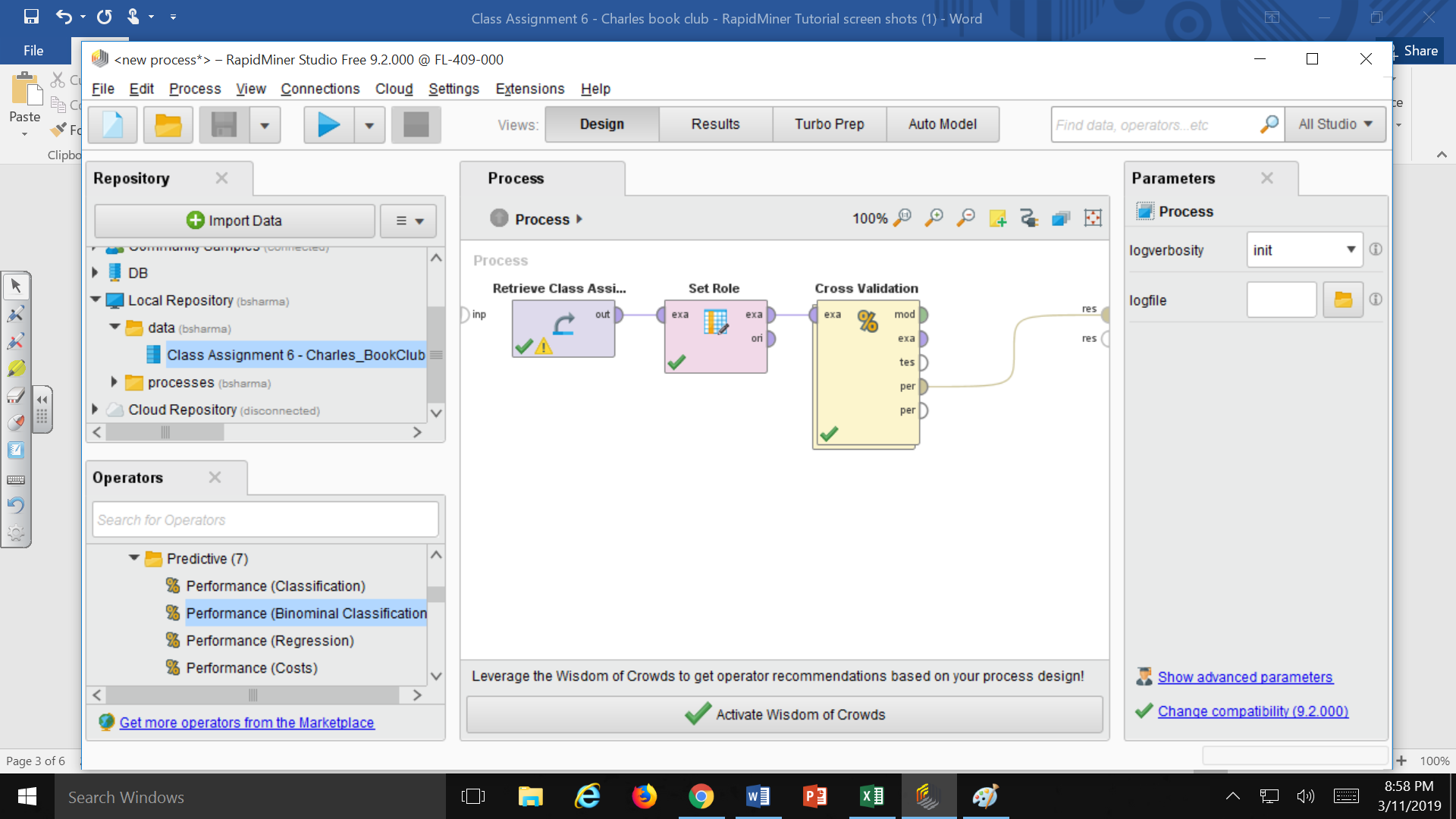




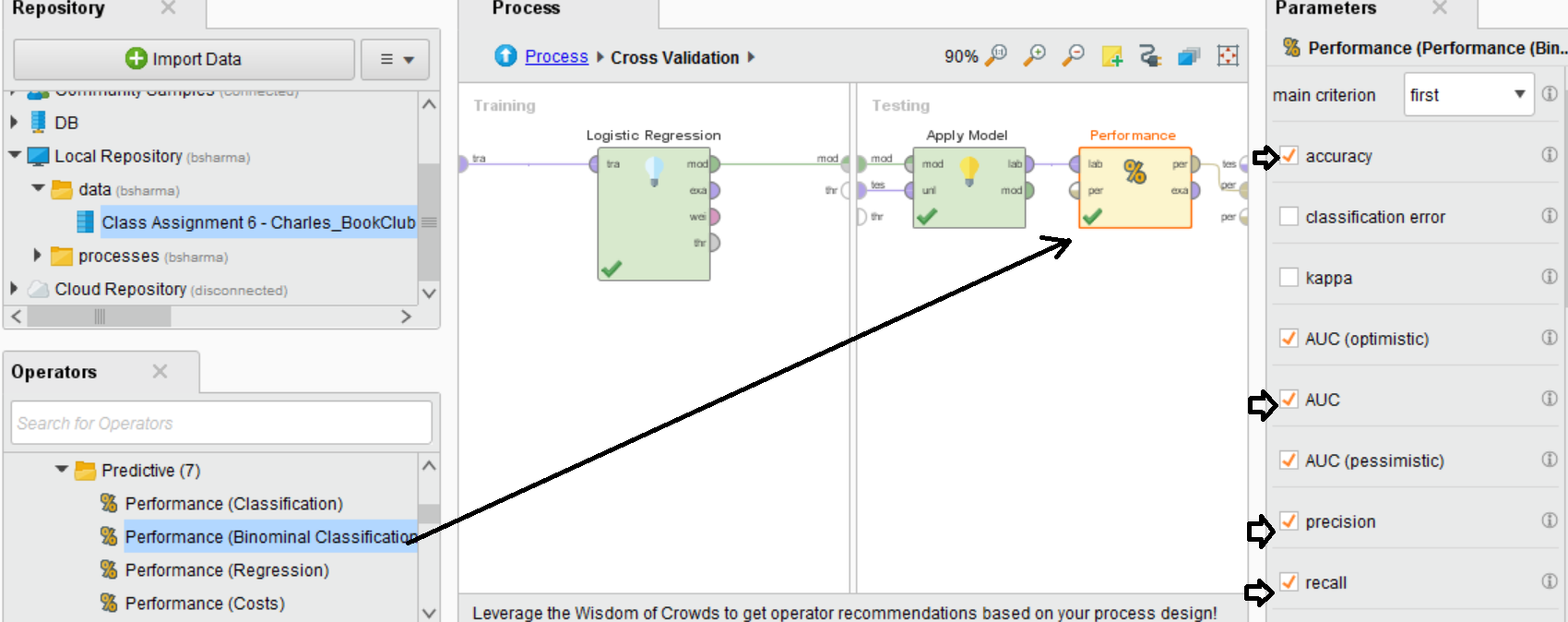
3. Check if there are any missing values under statistics – if there are then you will need to add the ‘Replace Missing Values Operator’ in Design.



4. Build the design as shown (remember the ‘Set Role’ operator will be added if you get an error on the ‘Cross Validation’ operator).

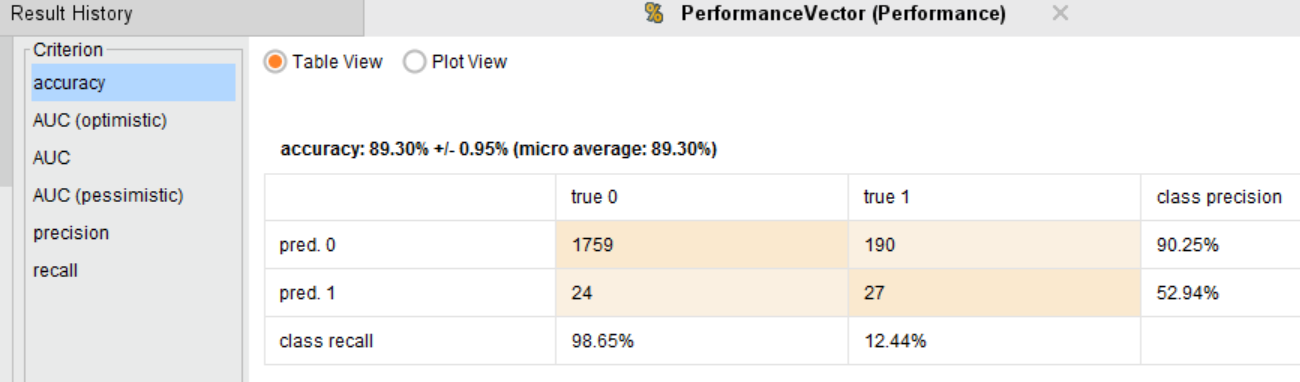


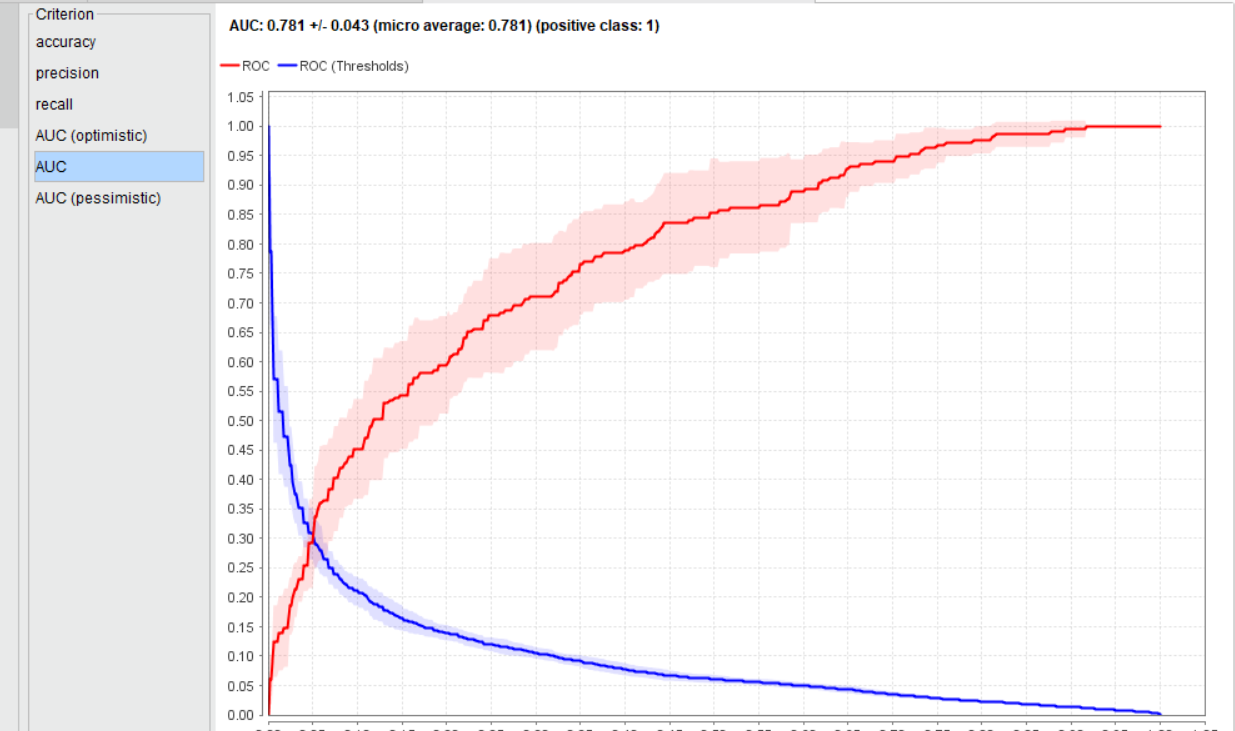
Inside view of the ‘Cross’validation’ operator – here you have to add ‘Logistic Regression’ operator. You can change the settings on the right side menu to get a different model.



5. Then you can ‘run’ it and see the confusion matrix results and the AUC, ROC results

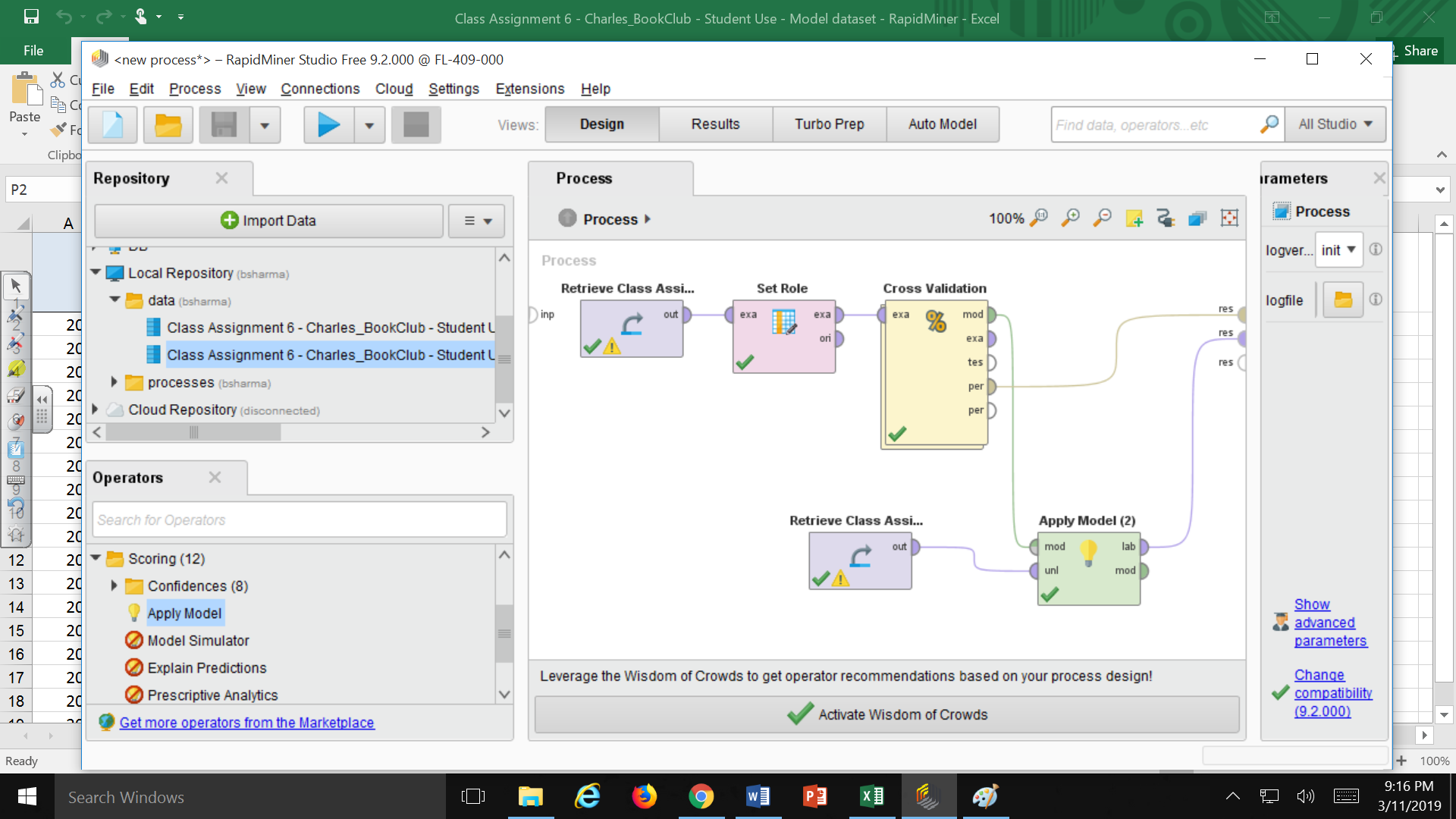
THIS IS JUST AN EXAMPLE ….You might get different results for different models.





<https://docs.rapidminer.com/latest/studio/operators/validation/performance/predictive/performance_binominal_classification.html>

6. Then add the New Data operator on the ‘best model’ process and run it to see the final prediction results.



**HELP on Settings Info : https://docs.rapidminer.com/latest/studio/operators/modeling/predictive/logistic\_regression/logistic\_regression.html**