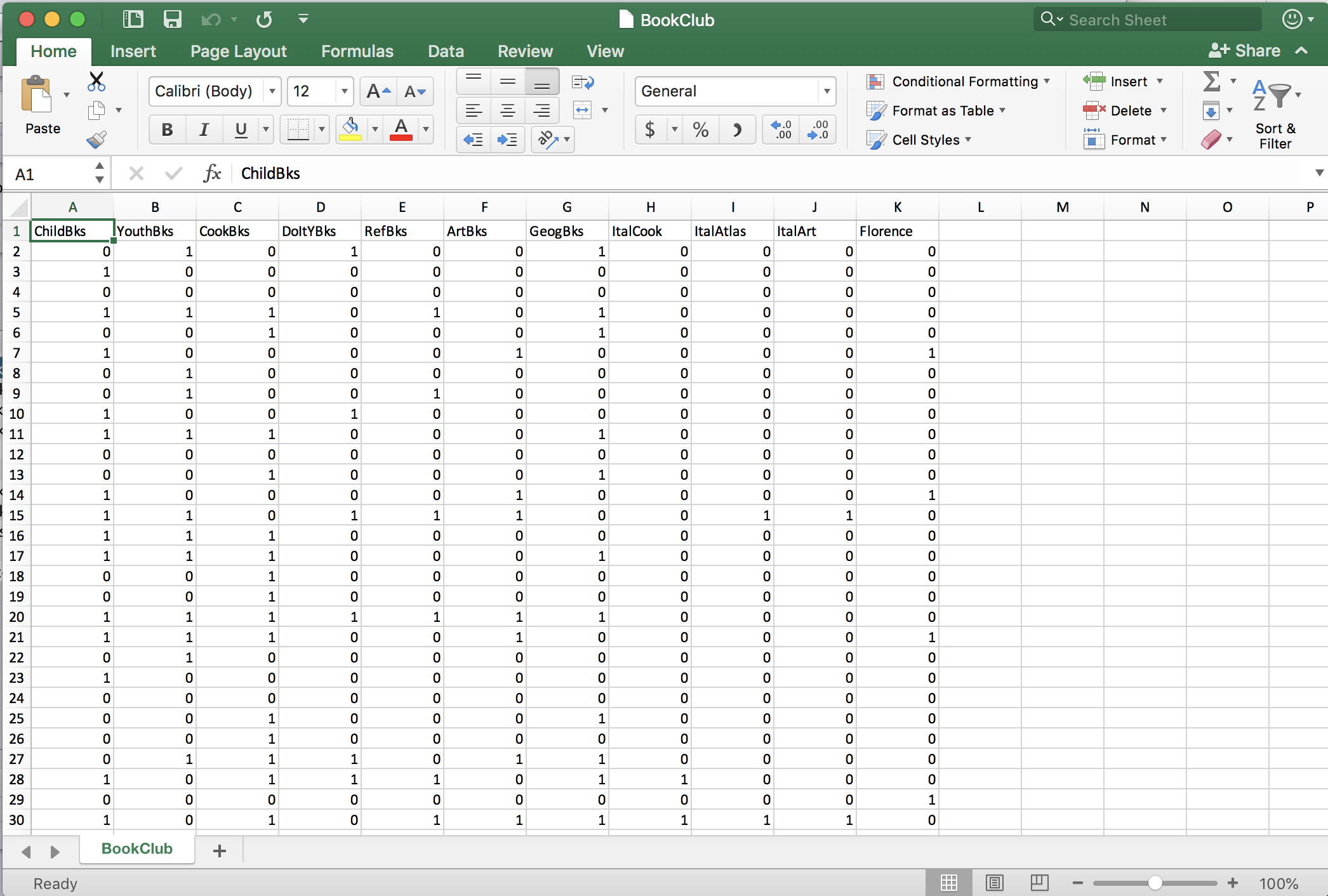
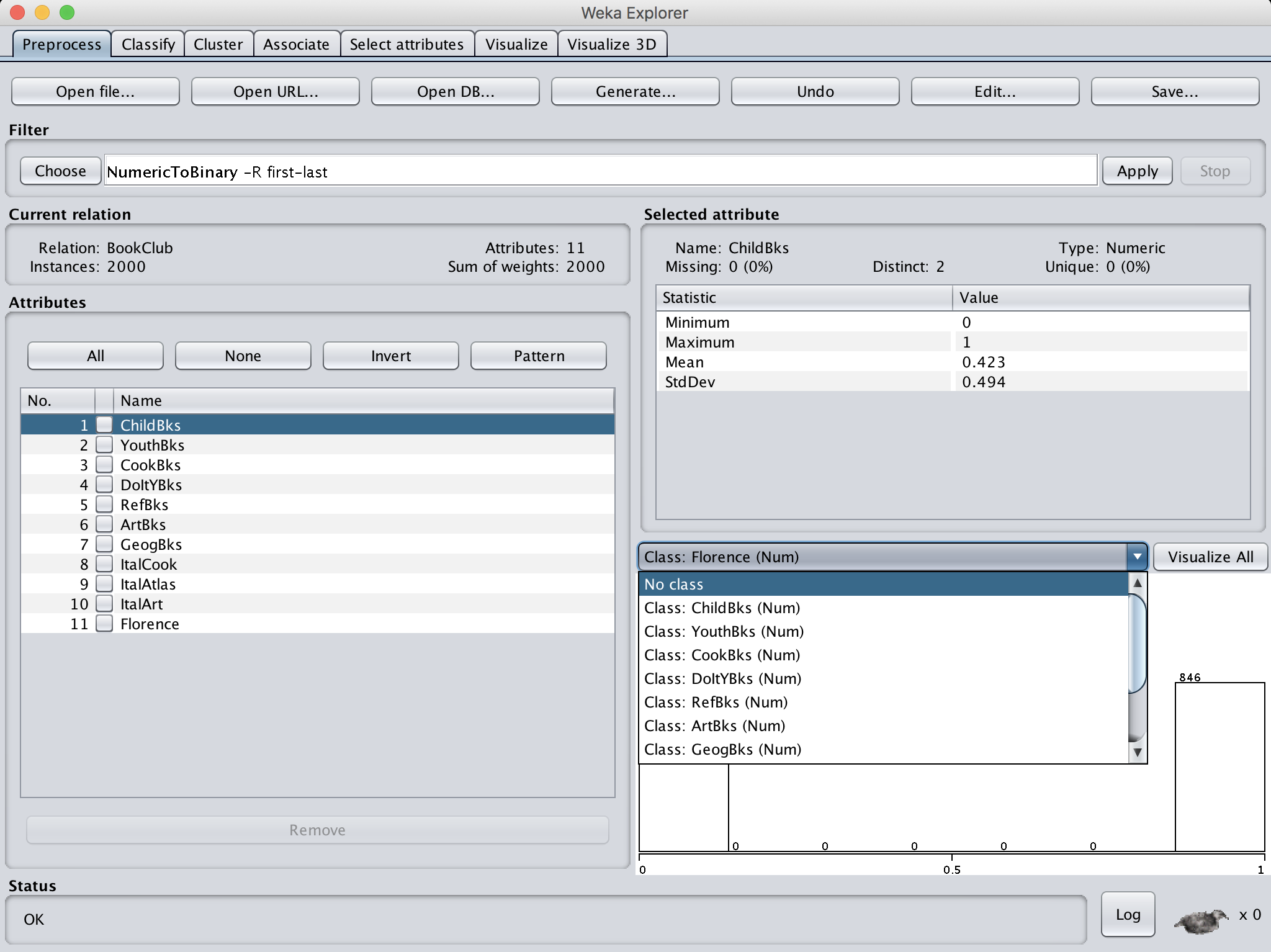
ARM

1. Open “BookClub.xls” to see how the data looks like. It has 11 attributes (all numeric) and 2000 records.



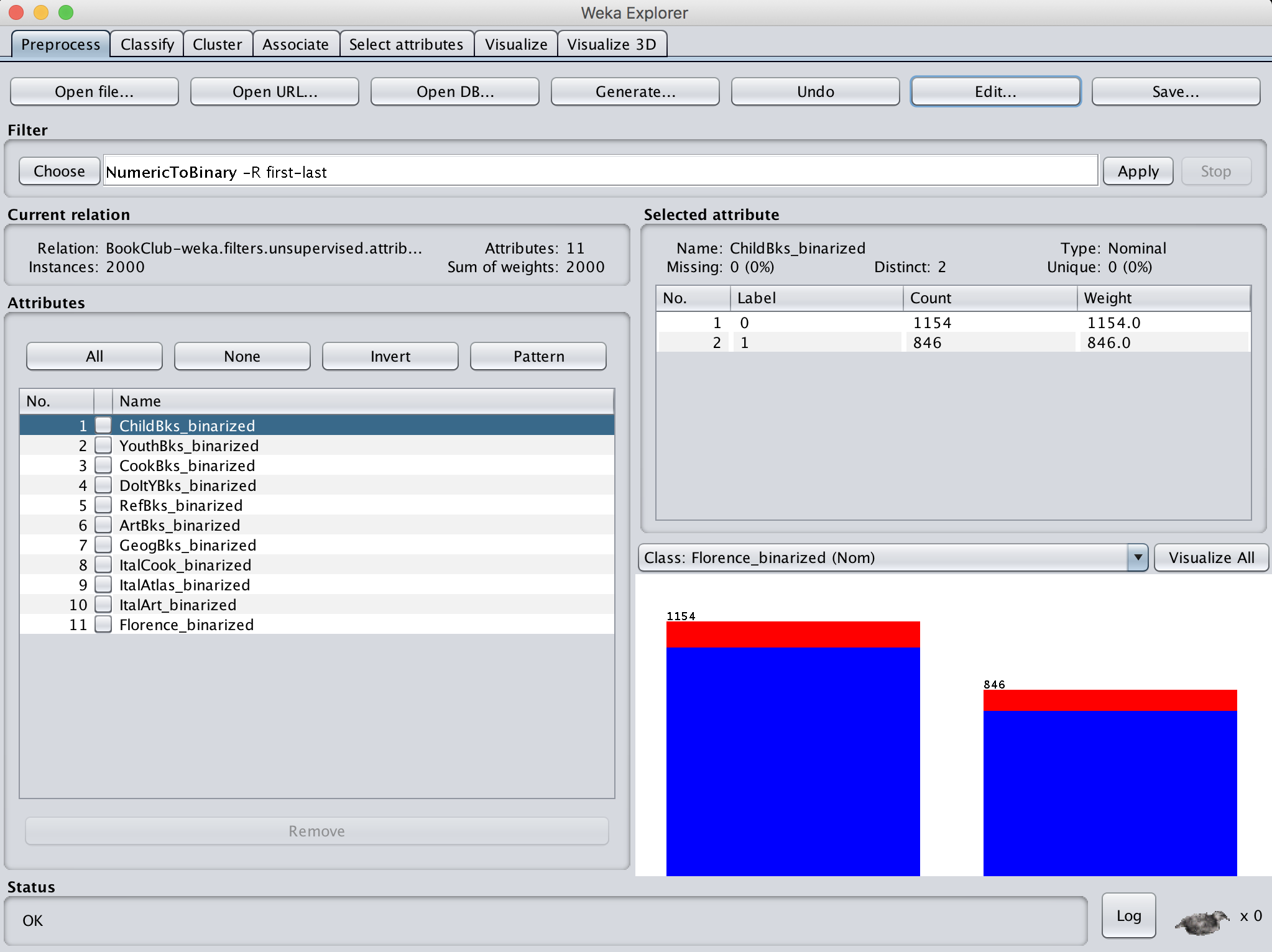
1. BookClub.csv is a comma separated version of BookClub.xls that can be read into Weka.

Since all attributes are numeric, ARM cannot be applied. So convert numeric attributes to binary using “NumericToBinary” filter (Preprocess->Filters->Unsupervised->Attribute->NumericToBinary)

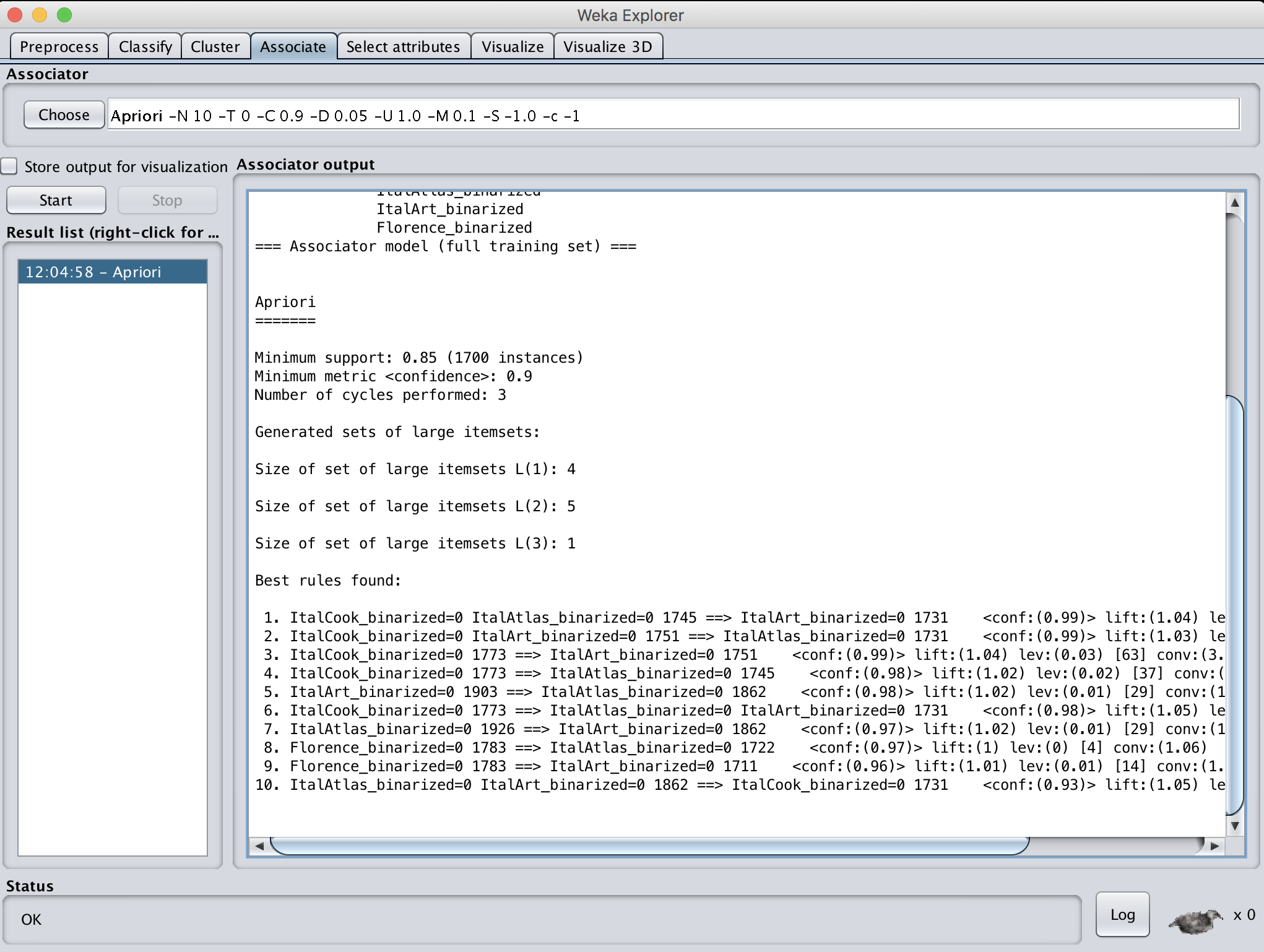


Before you click “Apply” to apply “NumericToBinary” filter, ensure that you want to convert from first attribute to last, and click “no class” in order to convert all attributes to binary.

1. Now all attributes are converted to binary.

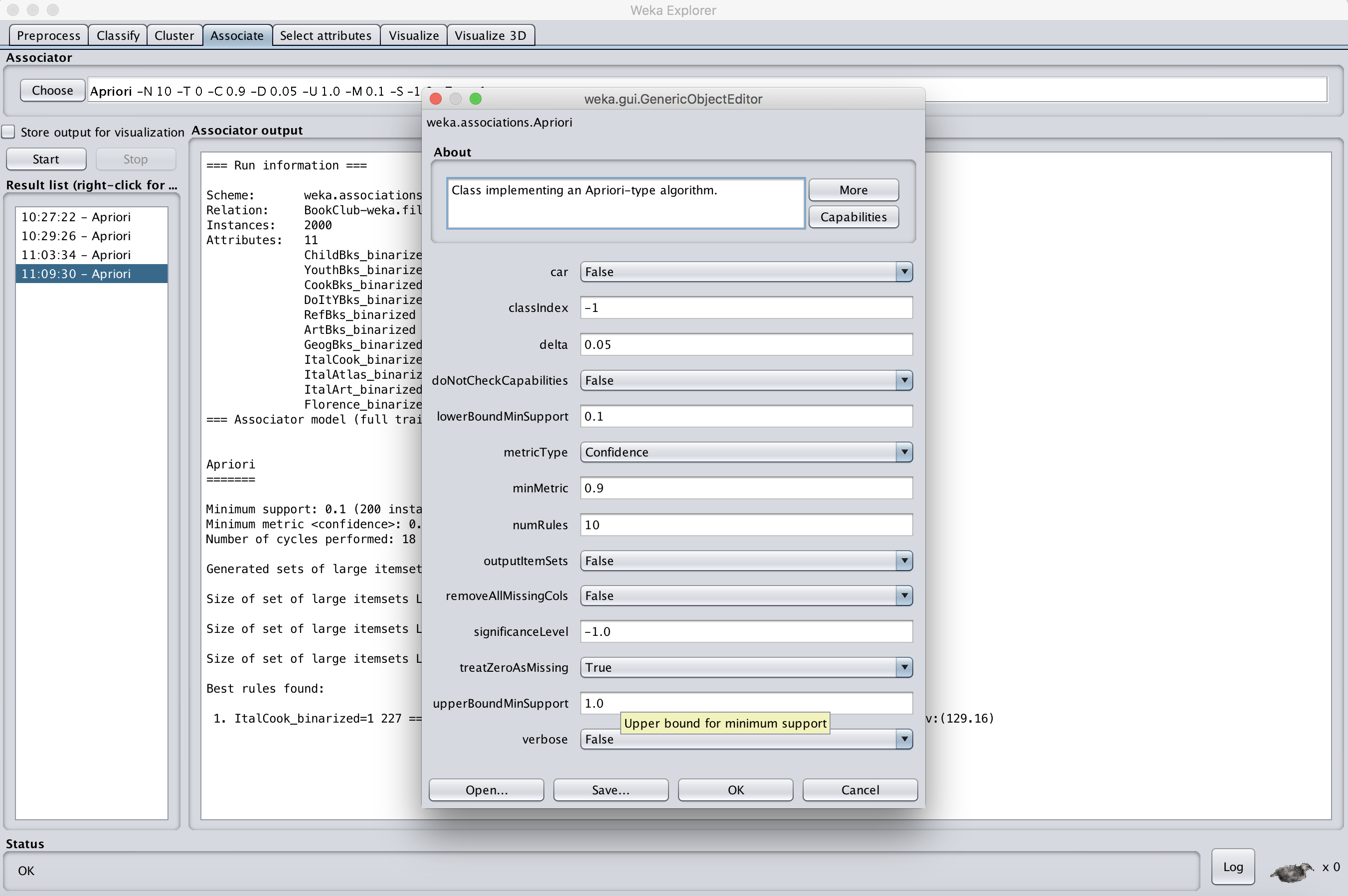


1. Now move to “Associate” tab, then “Apriori” and “FPGrowth” are active to choose. Click Apriori, use default parameter values and click Start.



Try to understand the rules. These rules are based on 1 (purchased) and 0 (non-purchased) together. Since the value of 0 dominates the data with more occurrences, most strong rules found are with 0. That is, 100% of customers who did NOT purchase book A also NOT purchase book B. We are more interested in a pattern states customers who purchase A also purchase B, so we need to further preprocess to find these purchase rules (not non-purchasing rules).

Or you can set treatZeroAsMissing to true to ignore all these zeroes.



1. Load “BookClubOne.csv” into Weka, follow the same preprocessing step (numericToBinary) and association rules mining to find patterns.

Now we have found a rule

“ItalCook\_binarized=1 227 ==> CookBks\_binarized=1 227 <conf:(1)> lift:(2.32) lev:(0.06) [129] conv:(129.16)

