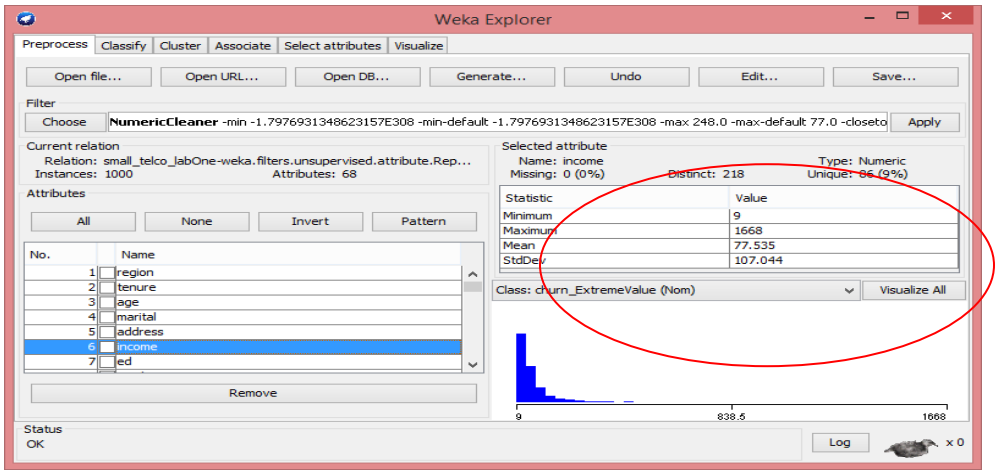
Lab Exercise One
Data Preprocessing with WEKA Explorer

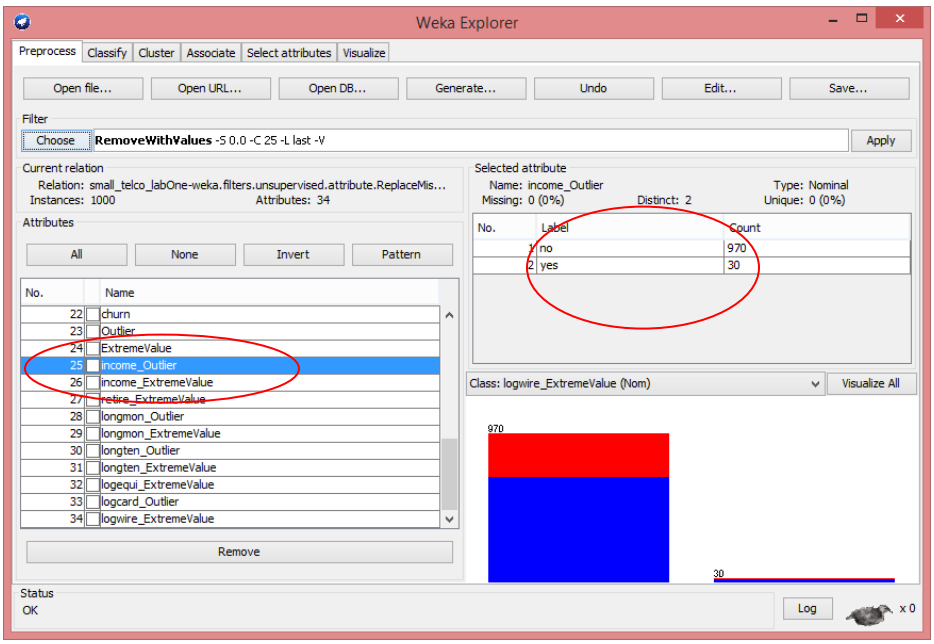
Using Filters to replace values

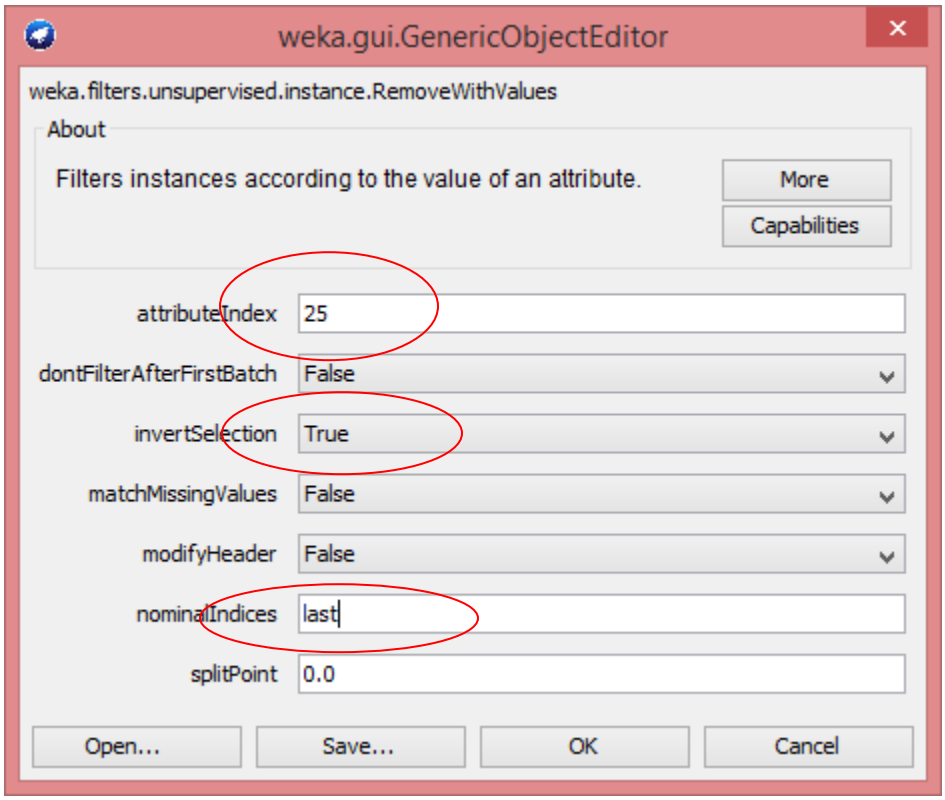
Unsupervised Attribute Filter – NumericCleaner: This filter replaces the values of numeric attributes that are too small, too large, or too close to a particular value with default values.

1. Instead of removing instances with outliers and extreme values, we could replace attributes values to a default values. Let's use **income** attribute as an example. Click income attribute, its statistic are shown on the right part of the window: the min. is 9, the max. is 1668 and mean is 77.535.

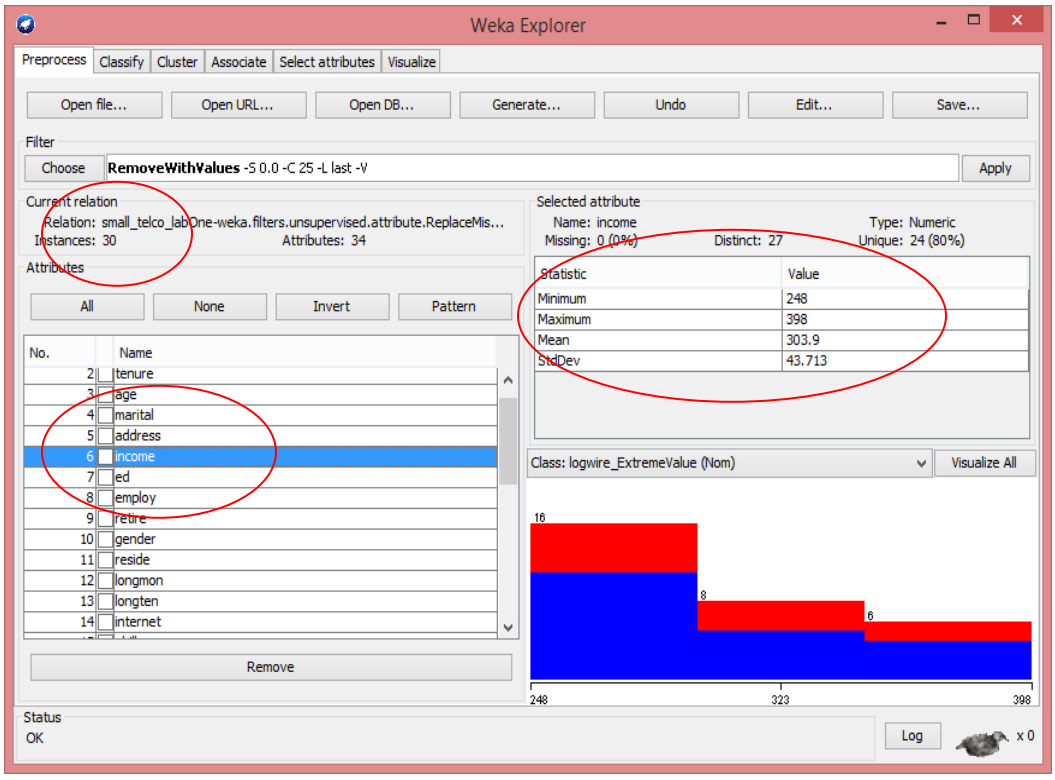


2. Perform an inverted **RemoveWithValues** filter with **income_Outlier** attribute.

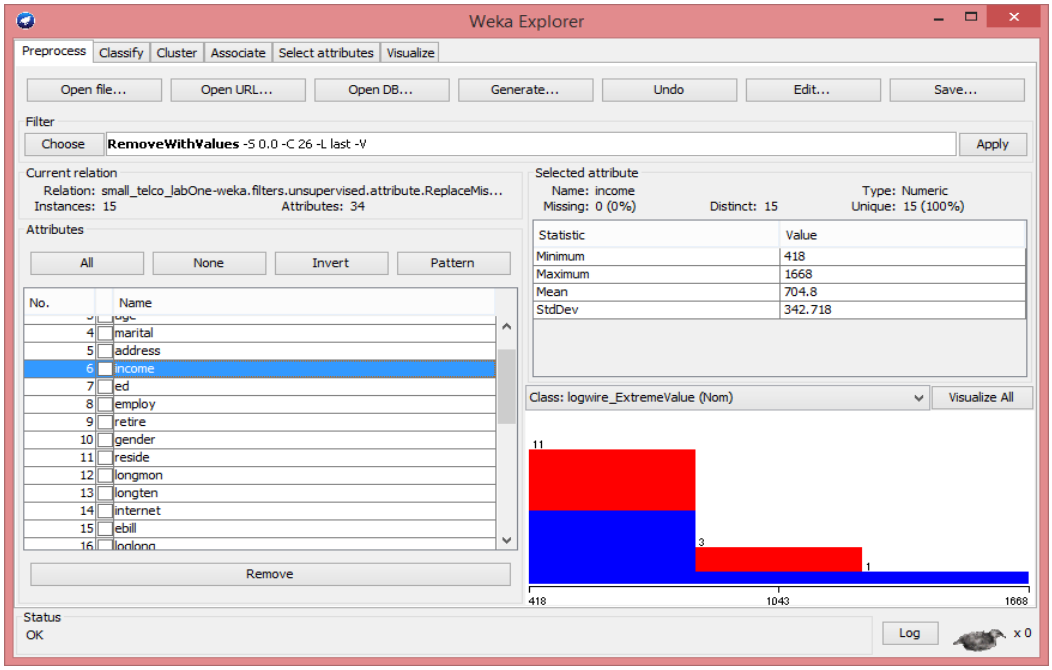




3. Check the **income** attribute of the remaining 30 instances of data. Its min. is 248, max. is 398.

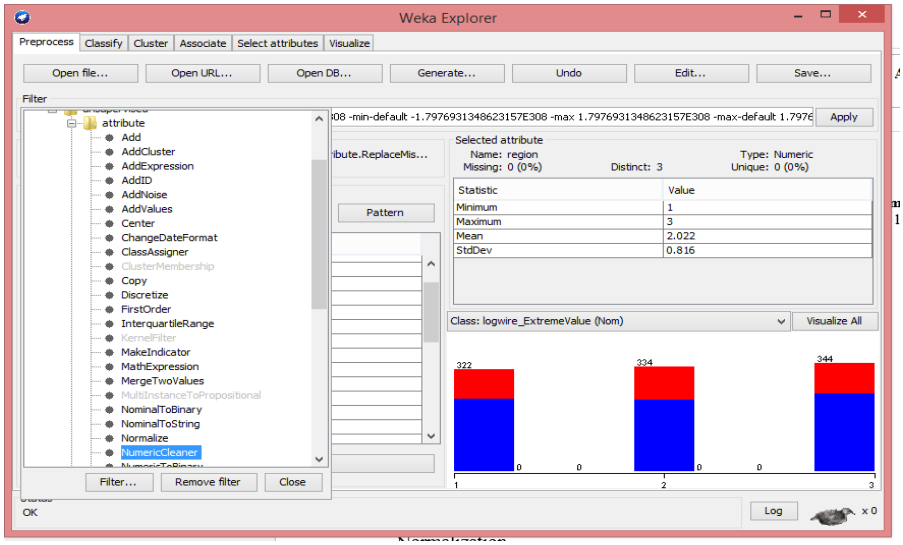


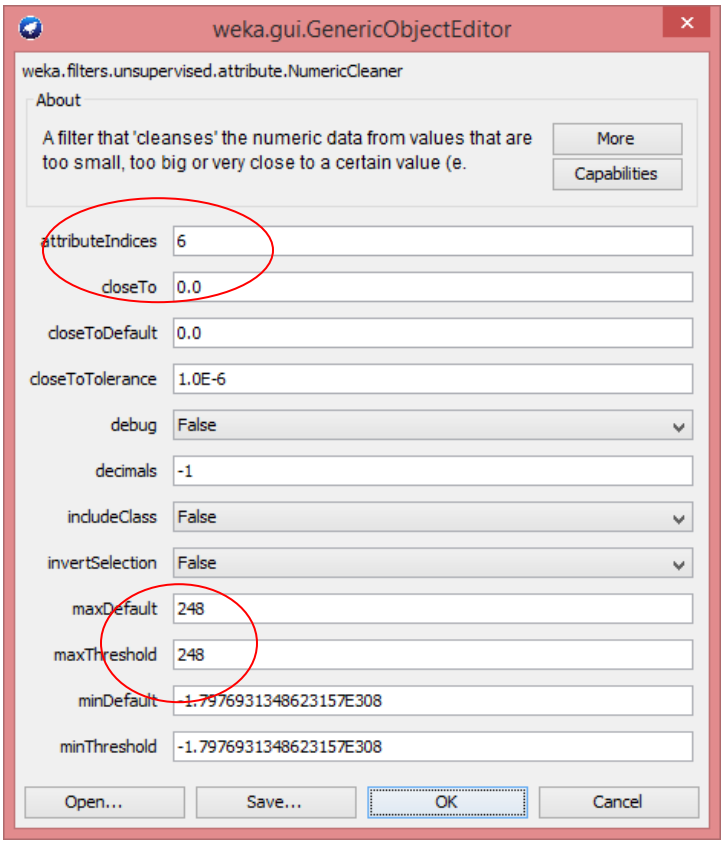
4. Click **Undo** button and repeat step 2 with **income_Extremevalue** attribue. Check the income attribute with the remaining 15 instances. Its min. is 418, the max. is 1668.



5. Click Undo button.

6. Then we are ready to perform the unsupervised attribute filter – **NumericCleaner** on all instances. Choose **NumericCleaner** filter from the drop-down list, then left-click the box of the filter to show the properties window.





7. Click Apply button to perform the filter, then select income attribute to see the statistic of the modified attribute. If you are fine with the result, save the dataset.