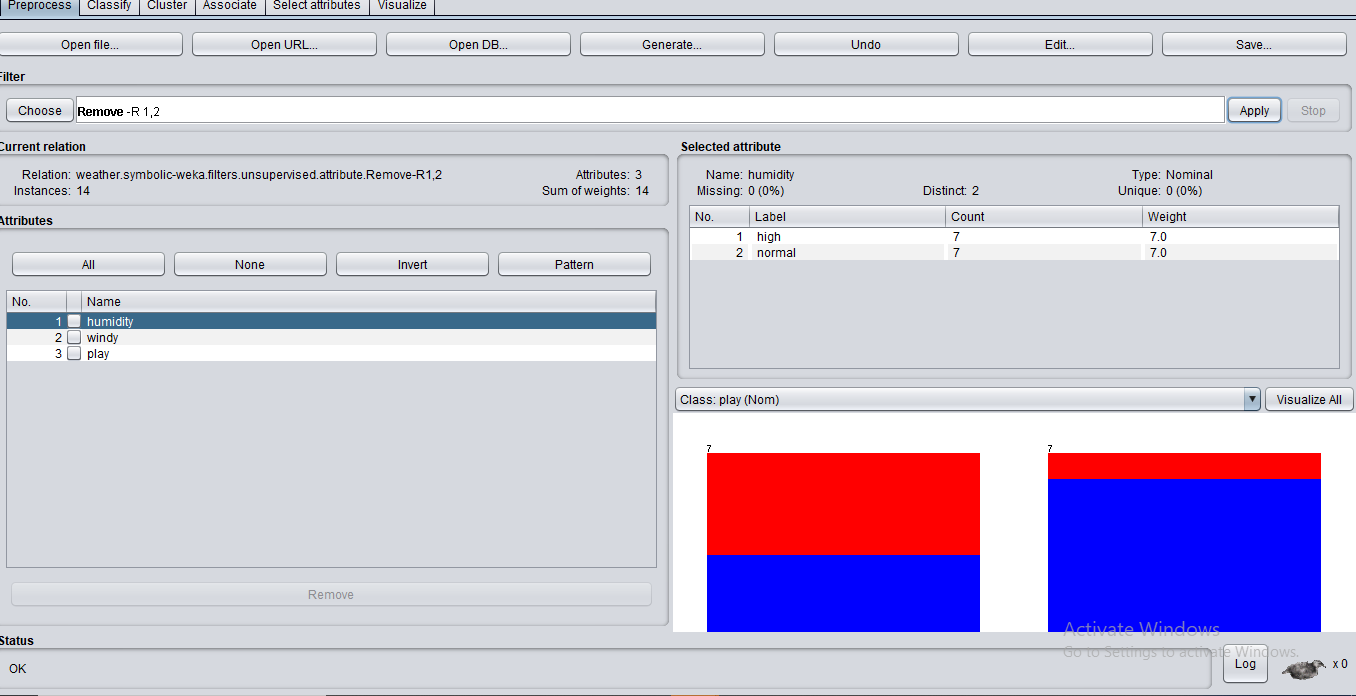
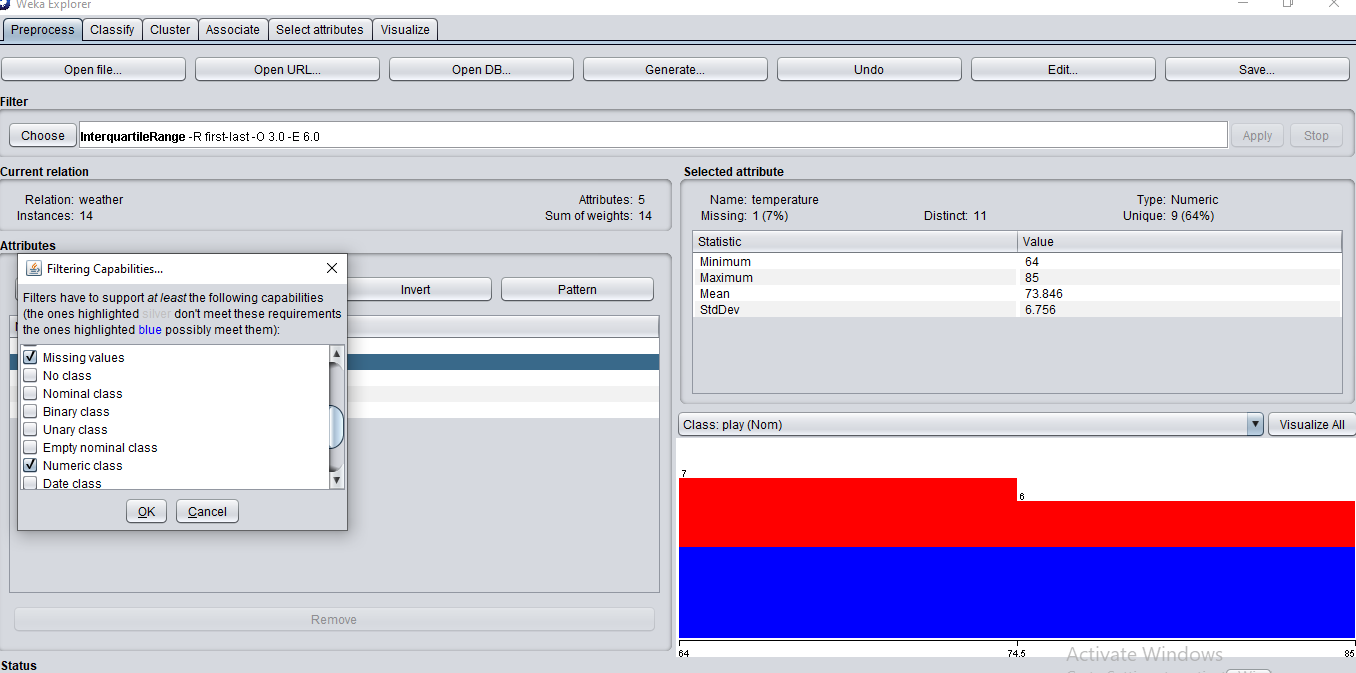
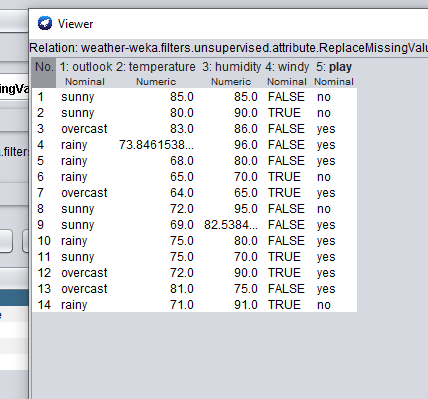
**Question No.1(Removing 2 attributes):**



We can remove attributes for various reasons, some of them are dropping / removing the attributes if they are not giving any meaningful information, we want to work on only a few attributes. We can say that we are doing dimensionality reduction.

**Question No.2 (Handling Missing Data):**

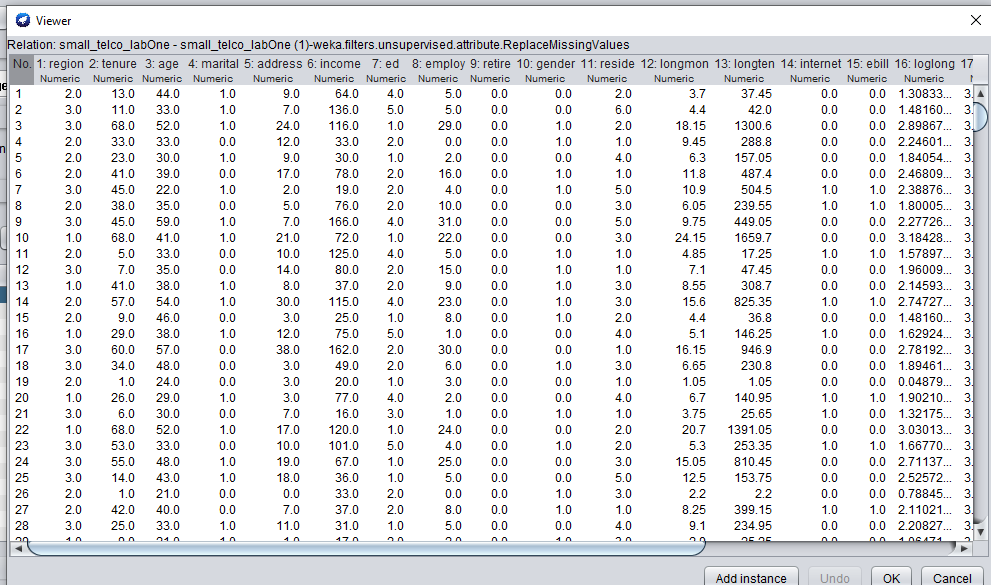


****

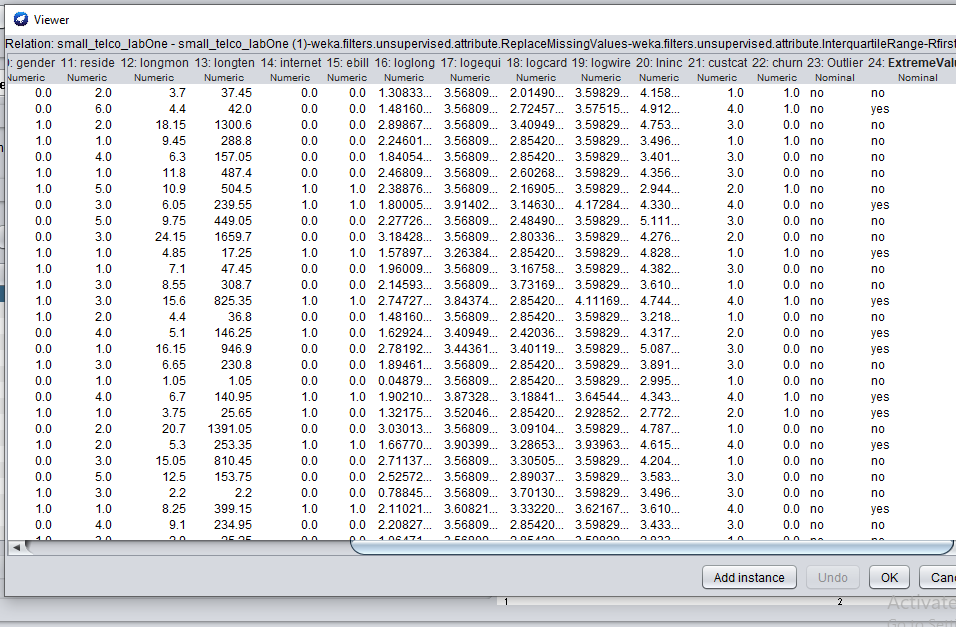
Filling in missing values is important in different cases. Handling missing values is important as the result might change because of the missing values and there is a high probability that it may provide the wrong result. Avoiding it can also reduce the accuracy of the model. In this case, the numeric data is replaced by mean (by default) but it could be done by mode etc.

**Question No.3 (Handling Outliers and Extreme values):**

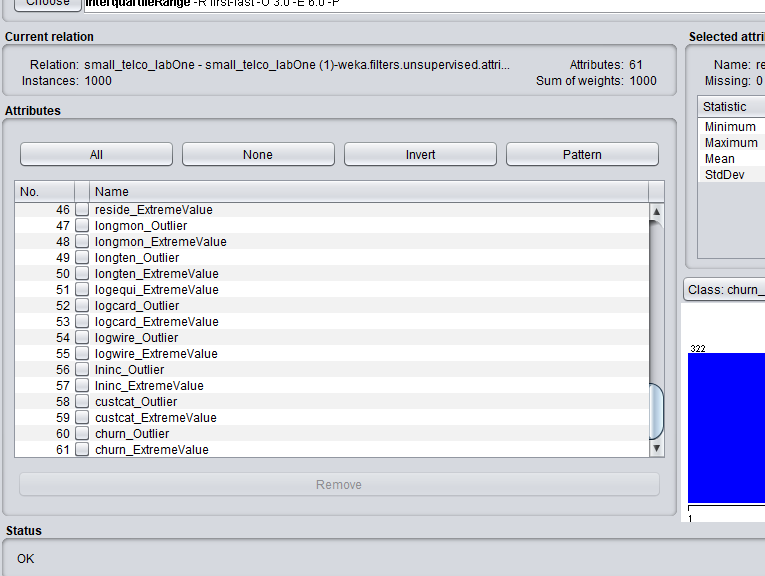
Filling in missing values is important in different cases. Handling missing values is important as the result might change because of the missing values and there is a high probability that it may provide the wrong result. Avoiding it can also reduce the accuracy of the model. In this case, the numeric data is replaced by mean (by default) but it could be done by mode etc.



Outliers

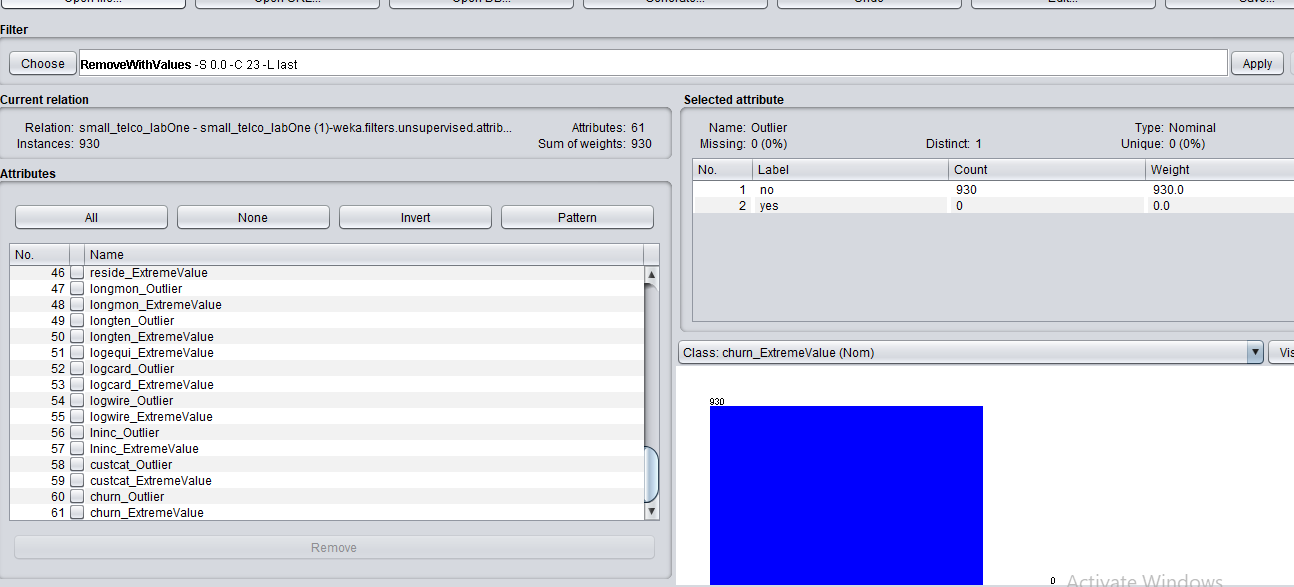


By applying this filter of handling outliers (FALSE) , it is known that whether the specific value is an outlier or not and whether the value is the extreme value(minimum or maximum) value or not. Handling outliers is also important as with outliers the model can result in wrong prediction.



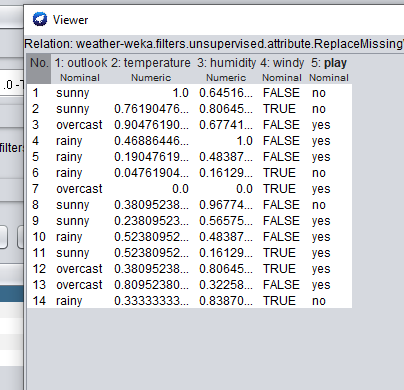
In this, the attributes/values with max “no” are removed

**Question No.4 (Remove with Values):**

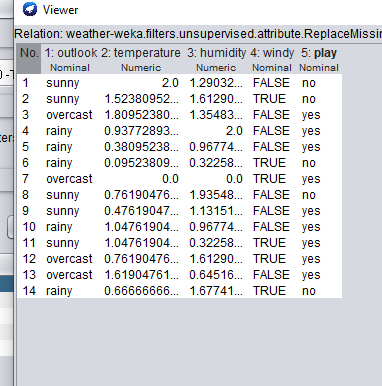
****

We have remove the outliers by detecting the values. For Example, if it finds the value “yes” in the outlier than it will remove that instance.

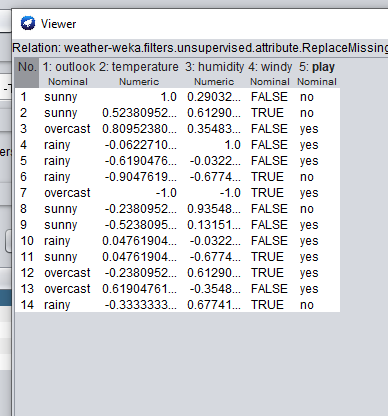
**Question No.5 (Using Filters to Perform Normalization):**



The range given id between (1 and 0) so it is normalizing all the values so that they are in the specified range. We are doing this so that we can handle and work on the value efficiently and accordingly as the values will be in a specified range.



The same thing is happening here but now the range is between 0 and 2



The same thing is happening here but now the range is between -1 and 2

The data is actually beinf standardized so that it is easy for us to work on it.