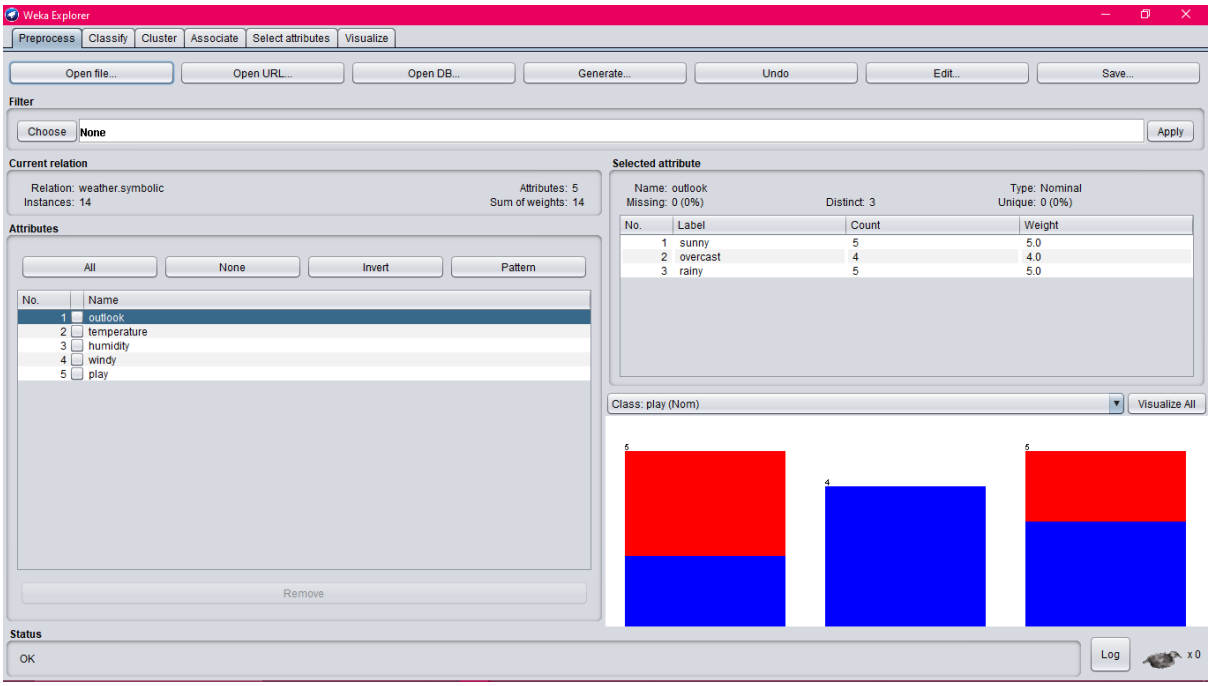
DVA Lab 1
Kush Munot A-47 A4 batch

Q1. Press the Explorer button on the main panel and load the weather dataset and answer the

following questions

1. How many instances are there in the dataset?
   Ans: 14



2. State the names of the attributes along with their types and values.

   Ans:

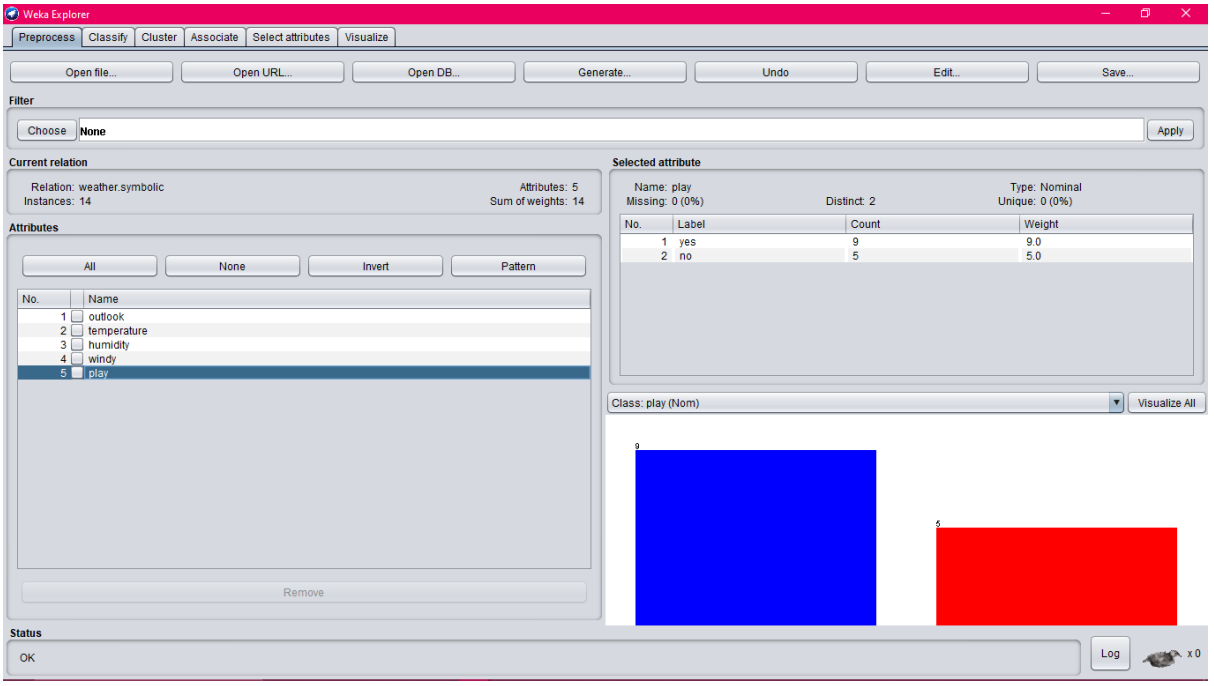
Attribute	Type	Value
Outlook	Nominal	3
Temperature	Nominal	3
Humidity	Nominal	2
Windy	Nominal	2
Play	Nominal	2

3. What is the class attribute?

Ans: The class attribute is play.

4. How will you determine how many instances of each class are present in the data

Ans: By checking the count of the class

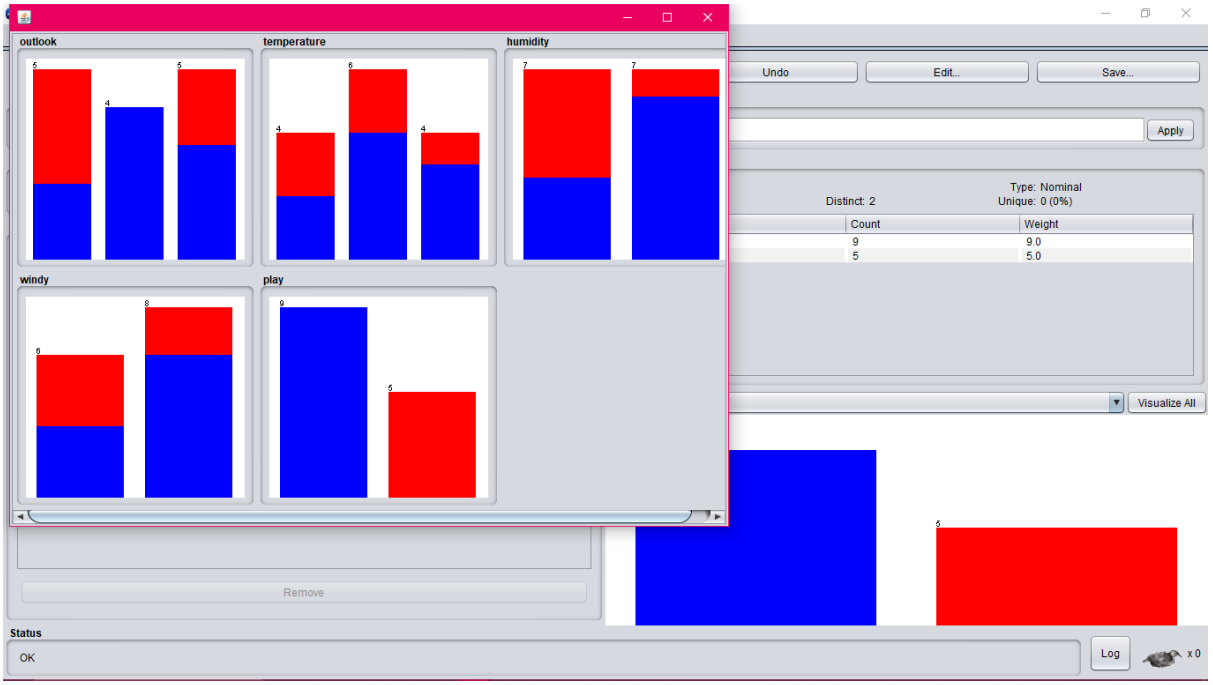


5. What happens with the Visualize All button is pressed?

   Ans: All the attributes are visualized against the class attribute.

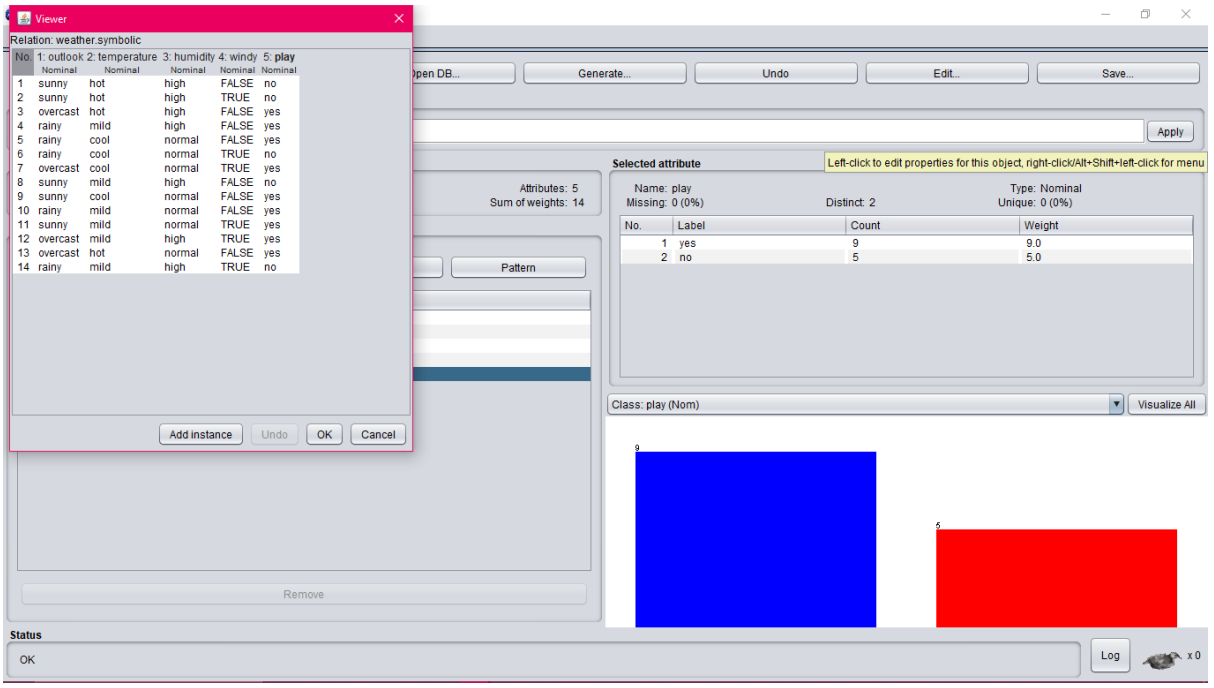
   The first bar of outlook indicates that on a sunny day how many play(yes) and how many do not play(no). The yes and no are differentiated by the different colours. Blue indicates yes and red indicates no.

   Similarly we can interpret the remaining attributes.



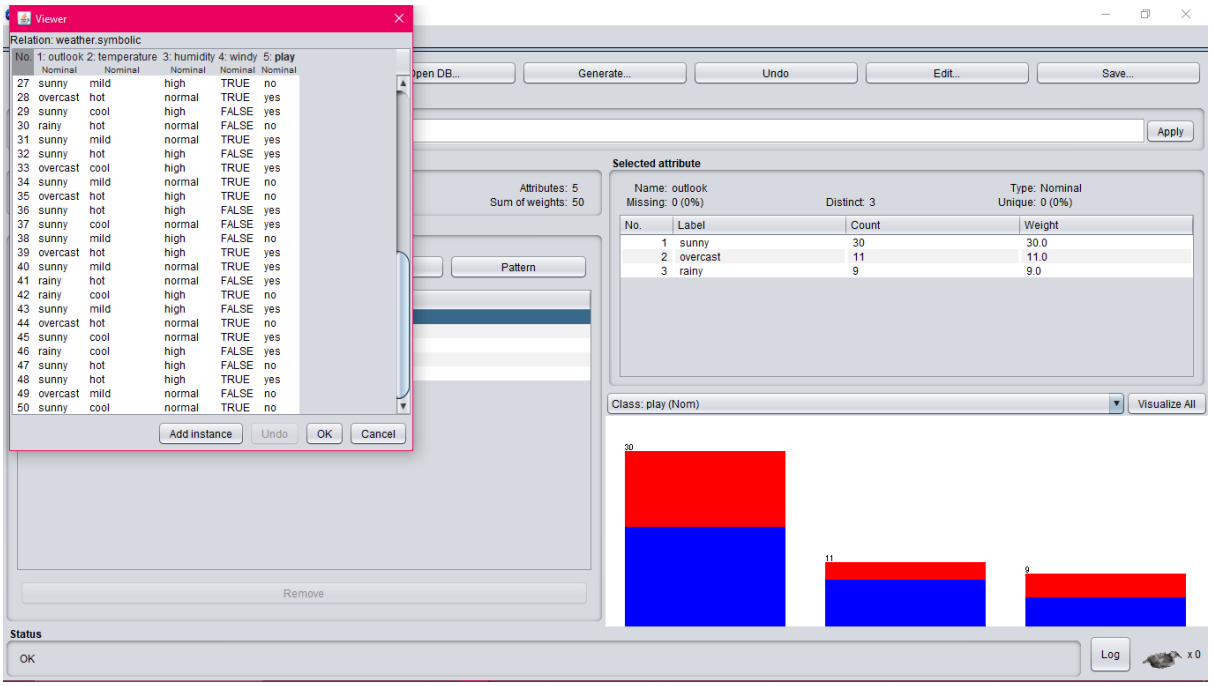
6. How will you view the instances in the dataset? How will you save the changes?

Ans: By pressing the edit button we can view the dataset. We can change the data and press OK to save the changes.



7. Now, extend the dataset to include 50 instances in total.

Ans:

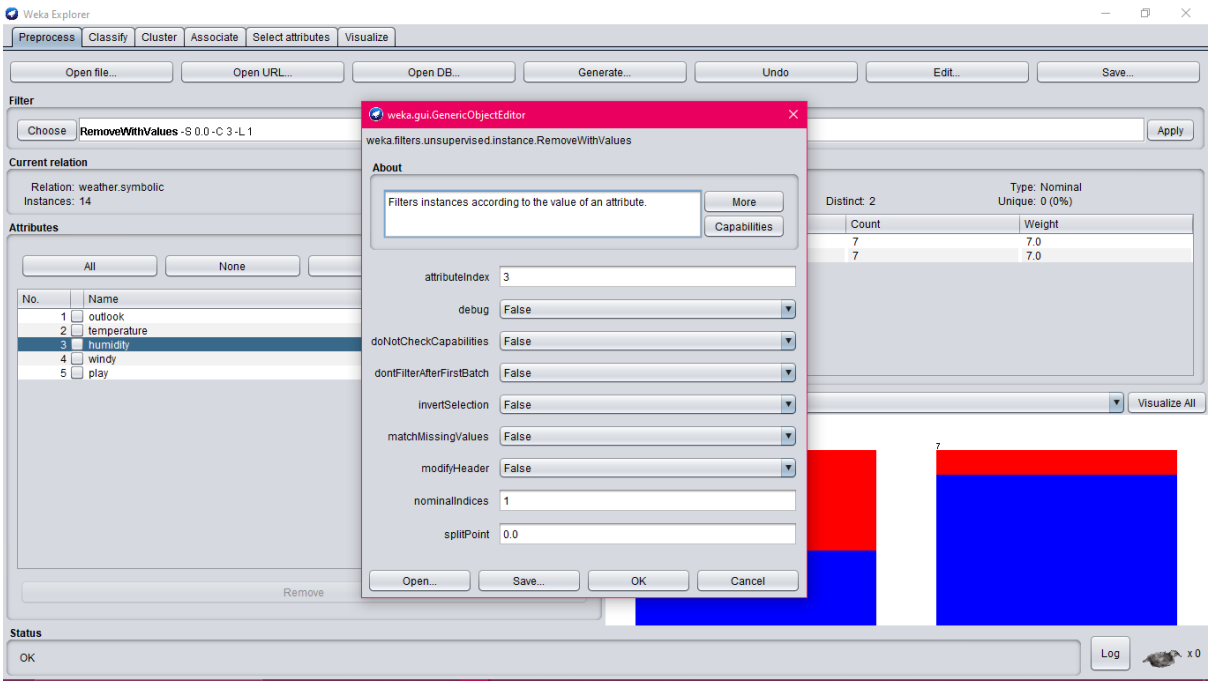


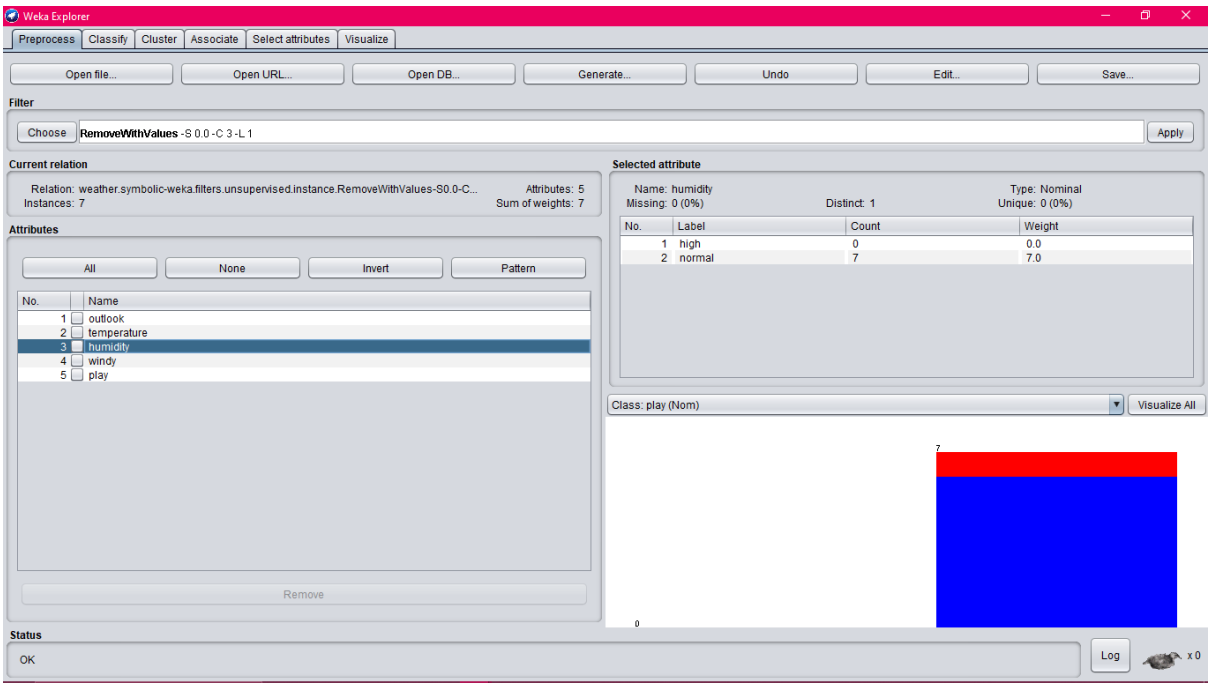
Q2. Do as directed to apply Filter

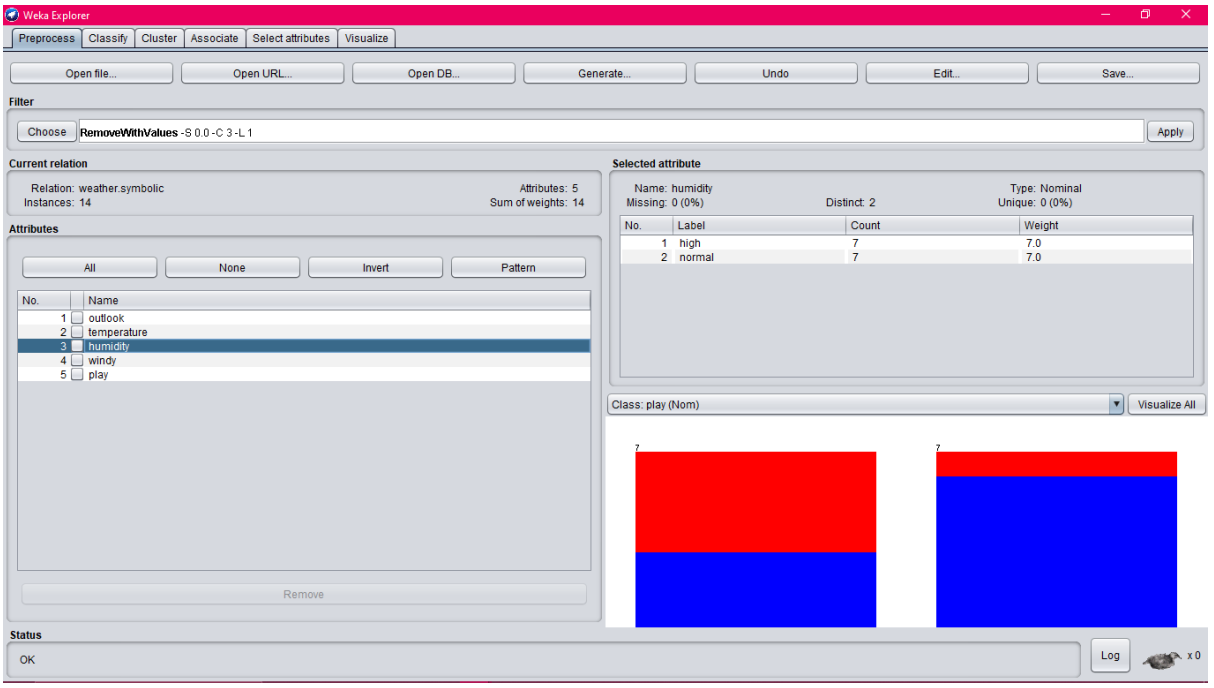
1. Use the unsupervised filter RemoveWithValues to remove all instances where the

attribute ‘humidity’ has the value ‘high’? Undo the effect of the filter.

Ans:

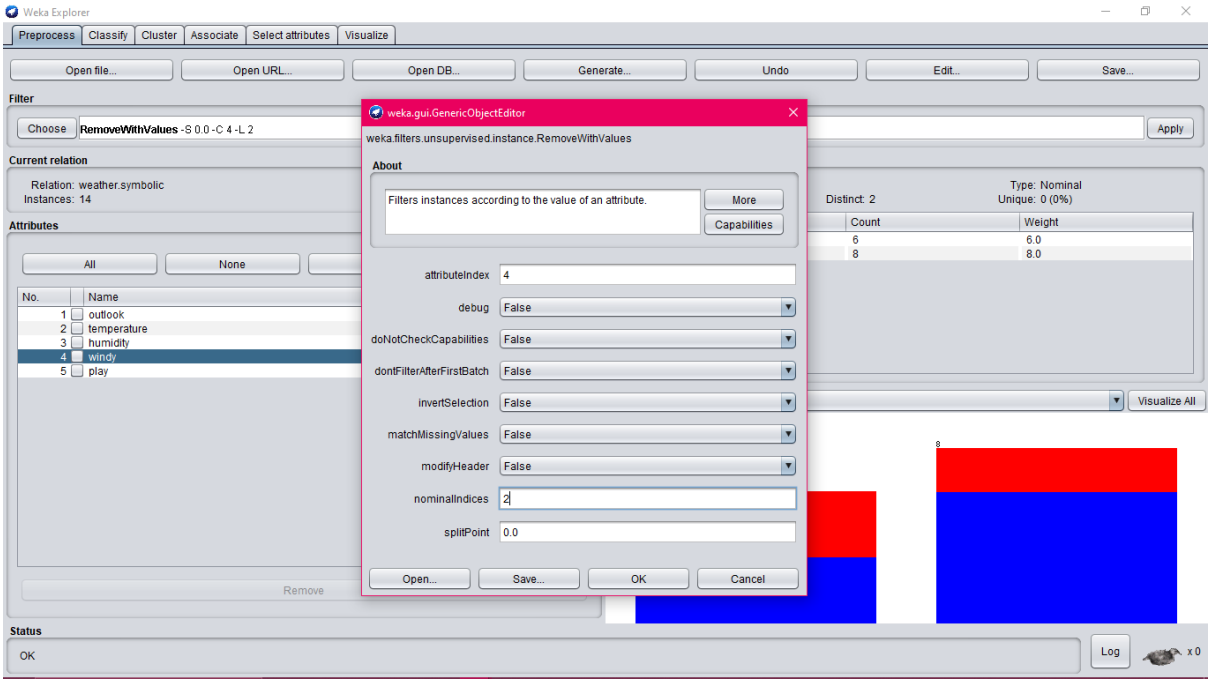


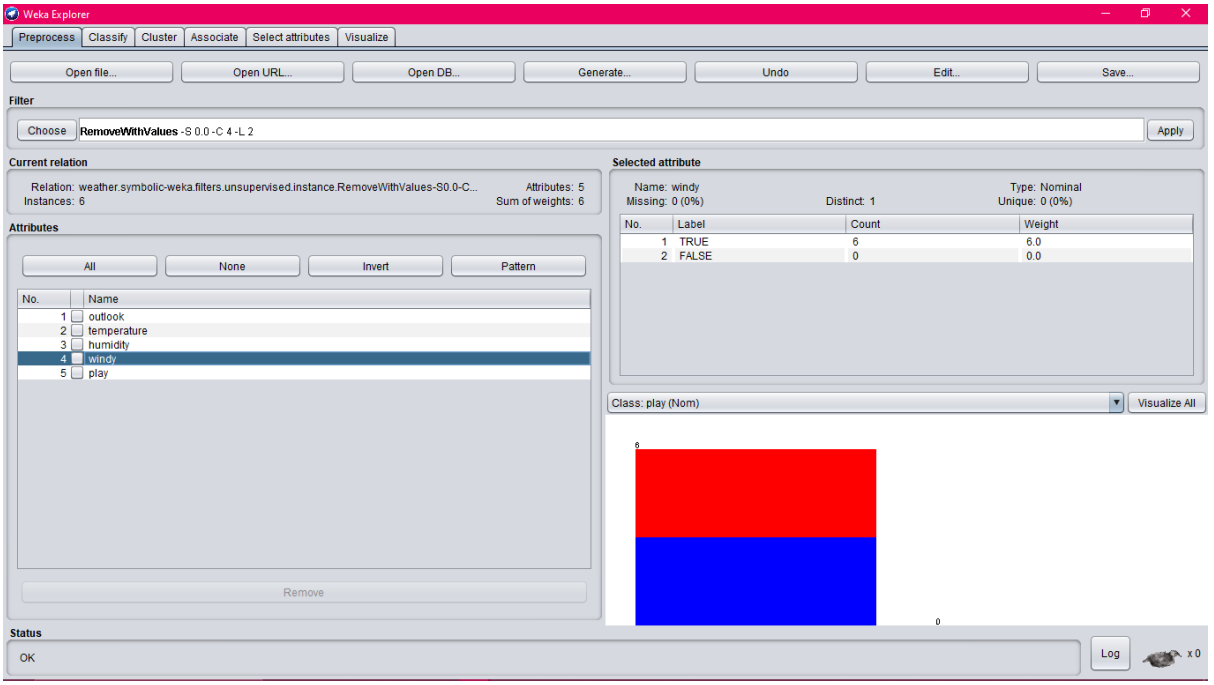




2. Remove the ‘FALSE’ instances of windy attribute and undo the effect.

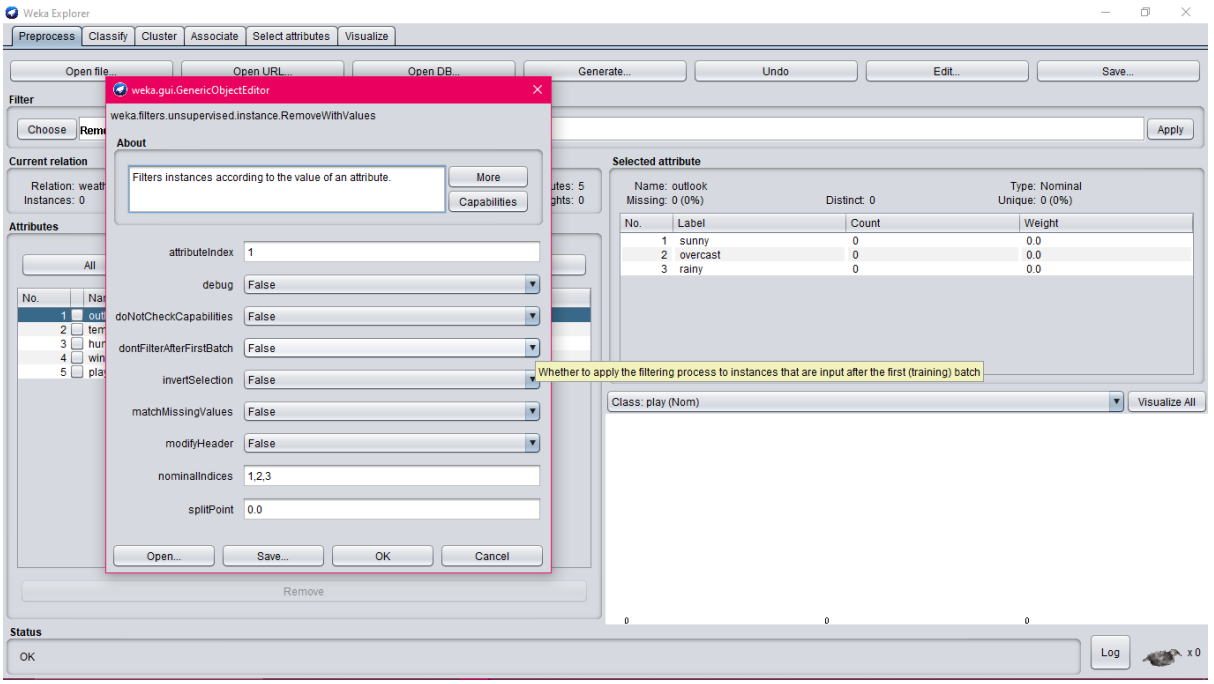
Ans:





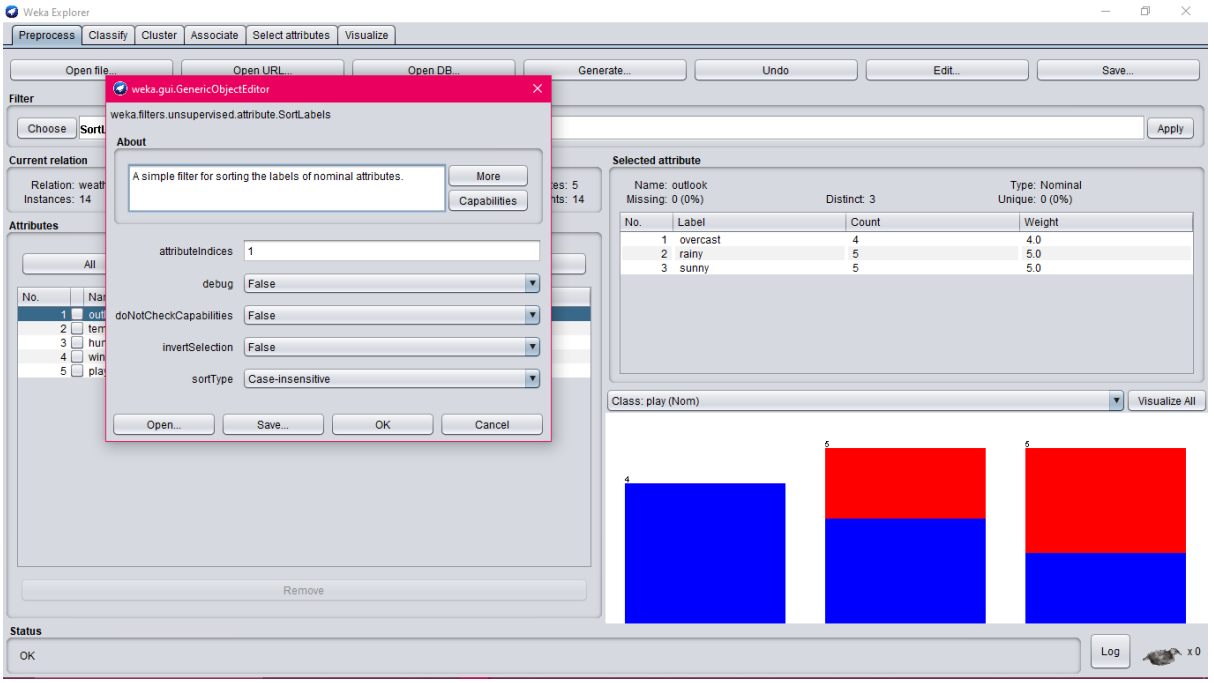
3. Remove the attribute outlook and undo the effect.

Ans:

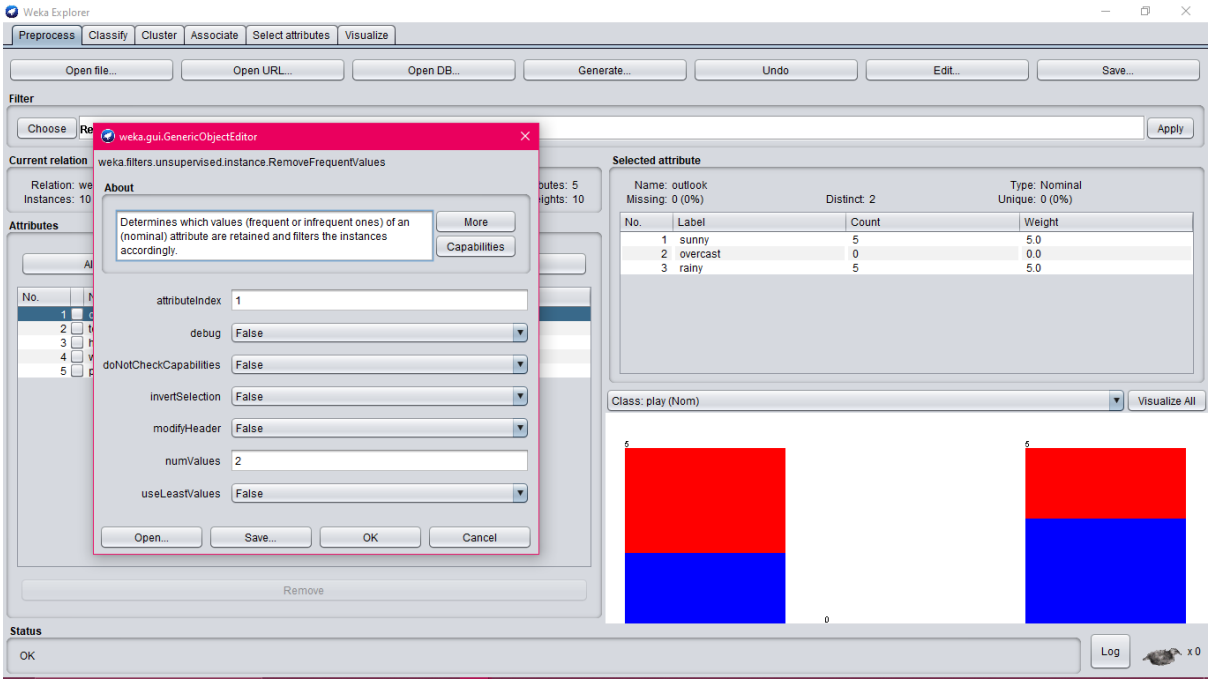


4. Experiment with different filters and report their effects.

Ans: a) I used sortLabels filter on attributes and it sorted the values of that attribute based upon its values



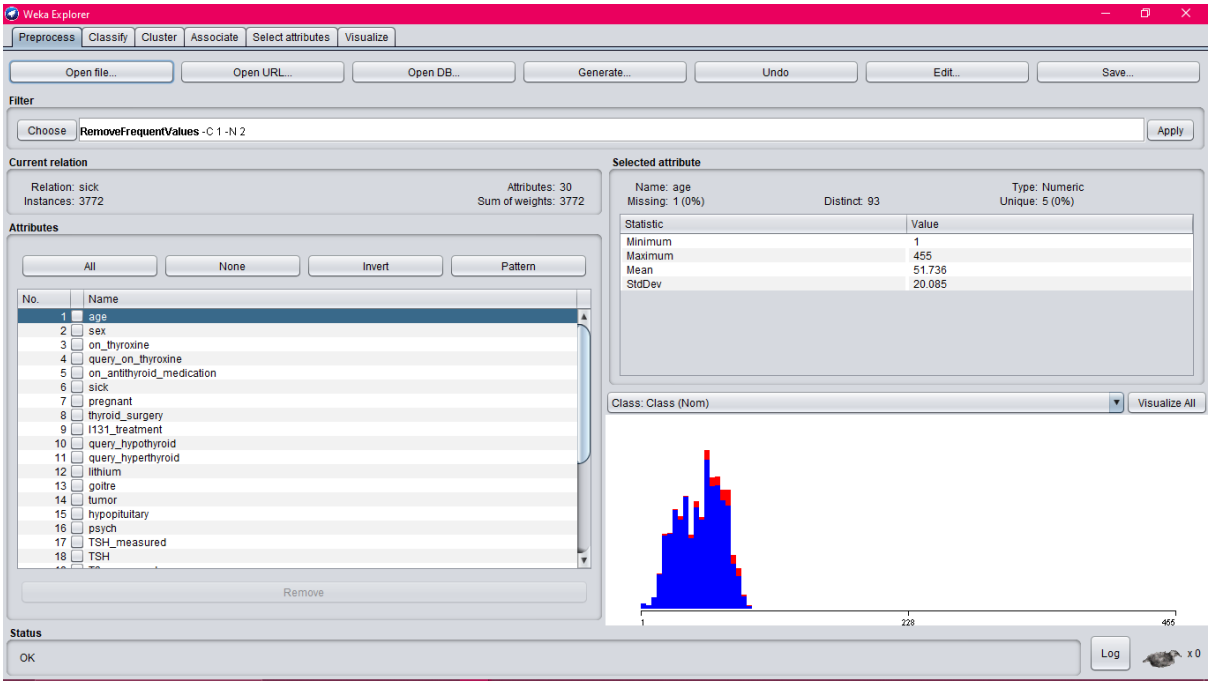
b) RemoveFrequentValues removes all instances of the given value



Q3. Application of Discretization Filters [use sick.arff dataset]

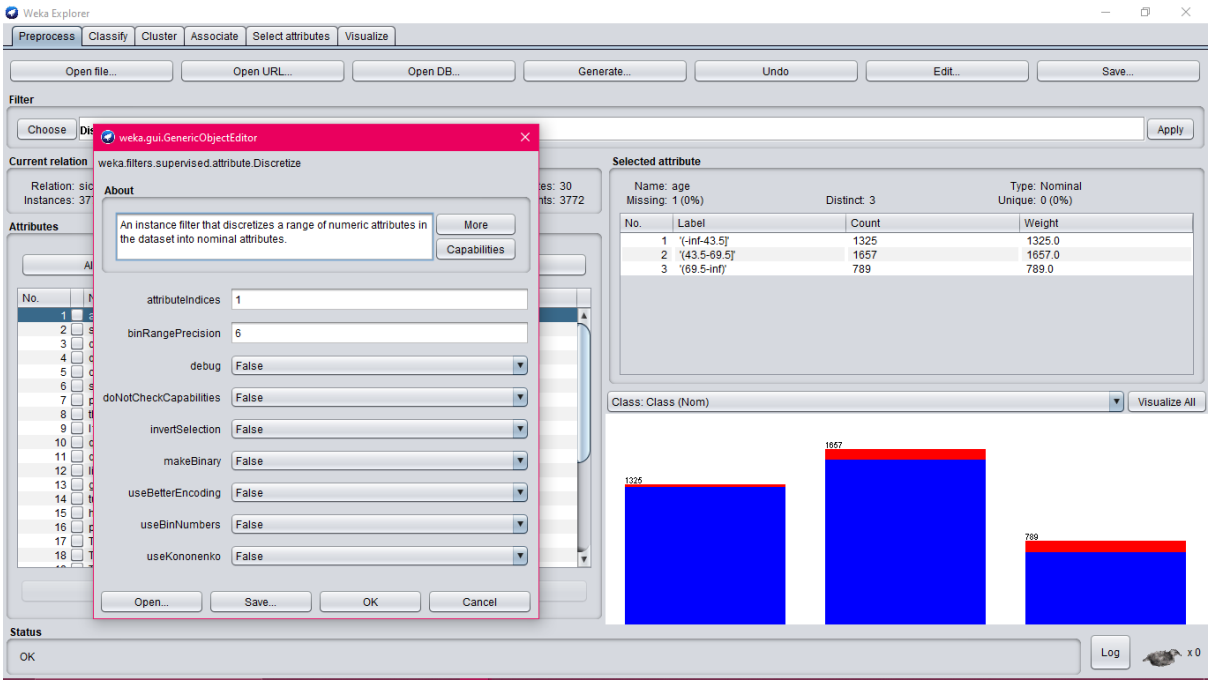
1. Load the 'sick.arff' dataset.

Ans:



2. Apply the supervised discretization filter on different attributes.

Ans:



3. What is the effect of this filter on the attributes?

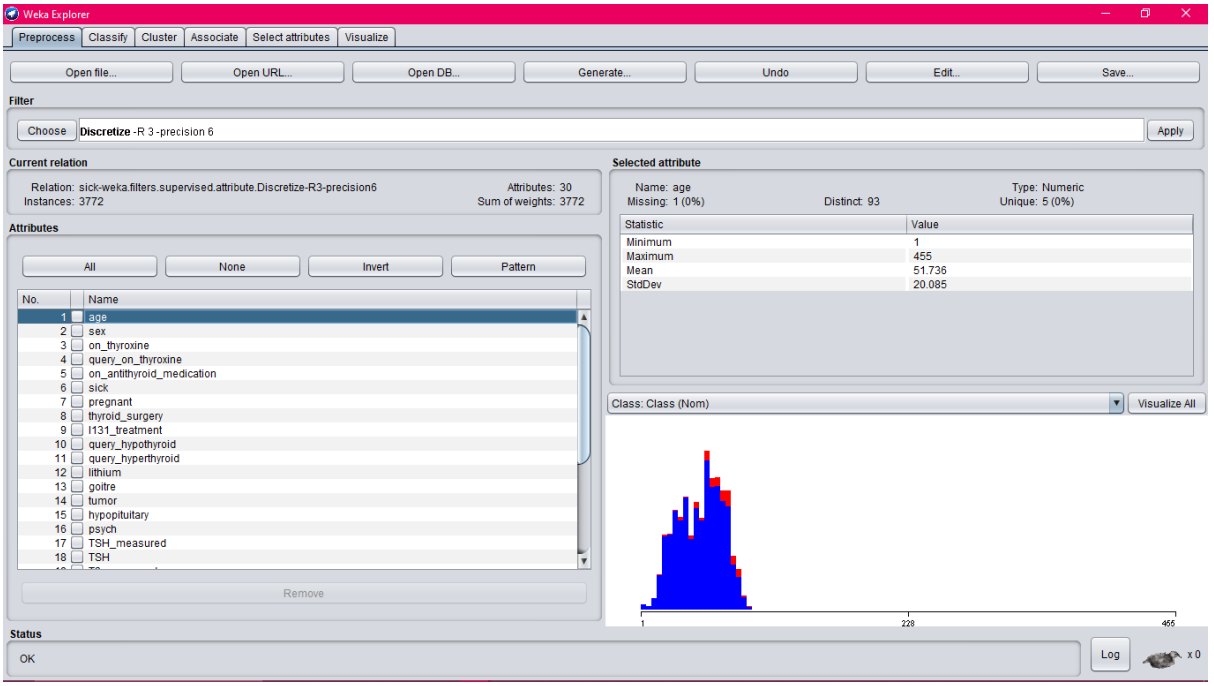
Ans: Before discretization distinct values were show to be 93 whereas afterwards they were converted to 3 values.

4. How many distinct ranges have been created for each attribute?

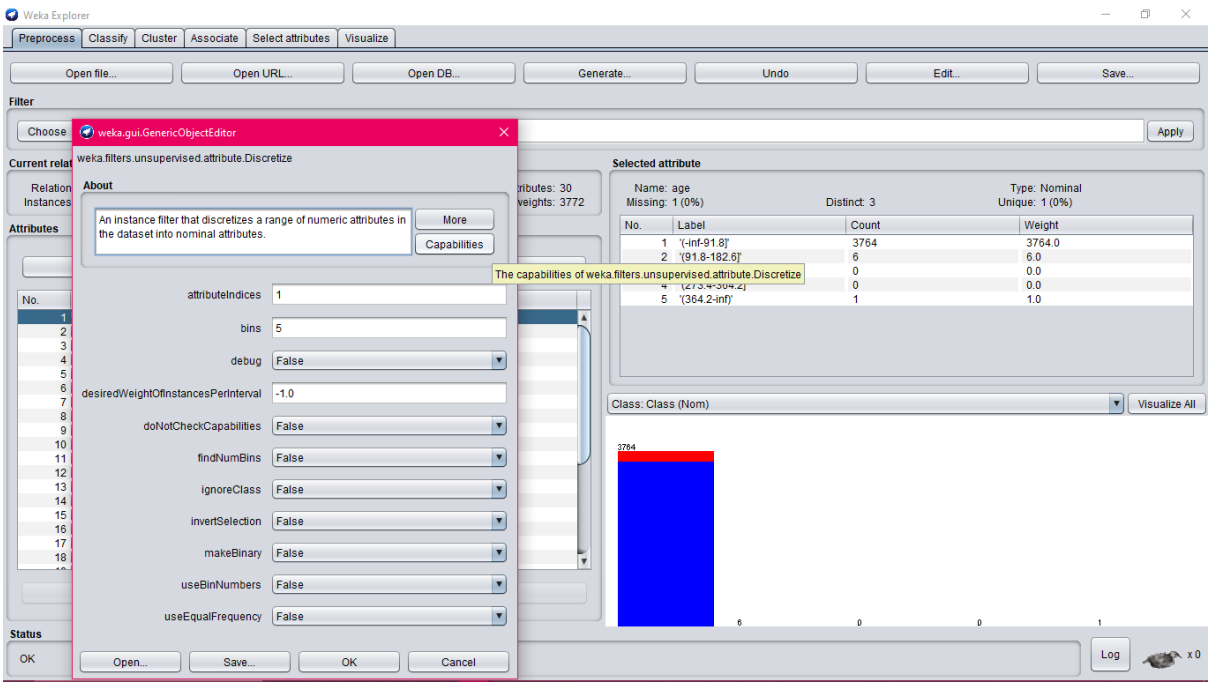
Ans: For age attribute, 3 distinct ranges have been created.

5. Undo the filter applied in the previous step.

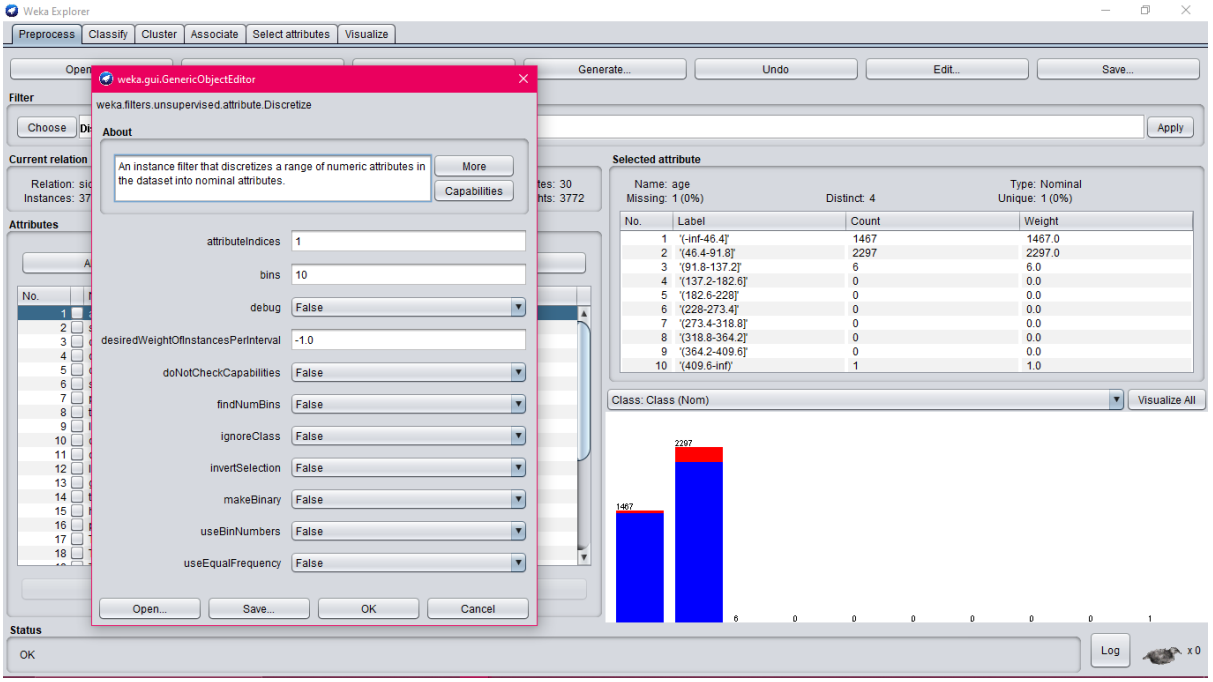
Ans:



6. Apply the unsupervised discretization filter. [Use equal-width binning approach]
   1. In this step, set 'bins'=5



   2. In this step, set 'bins'=10

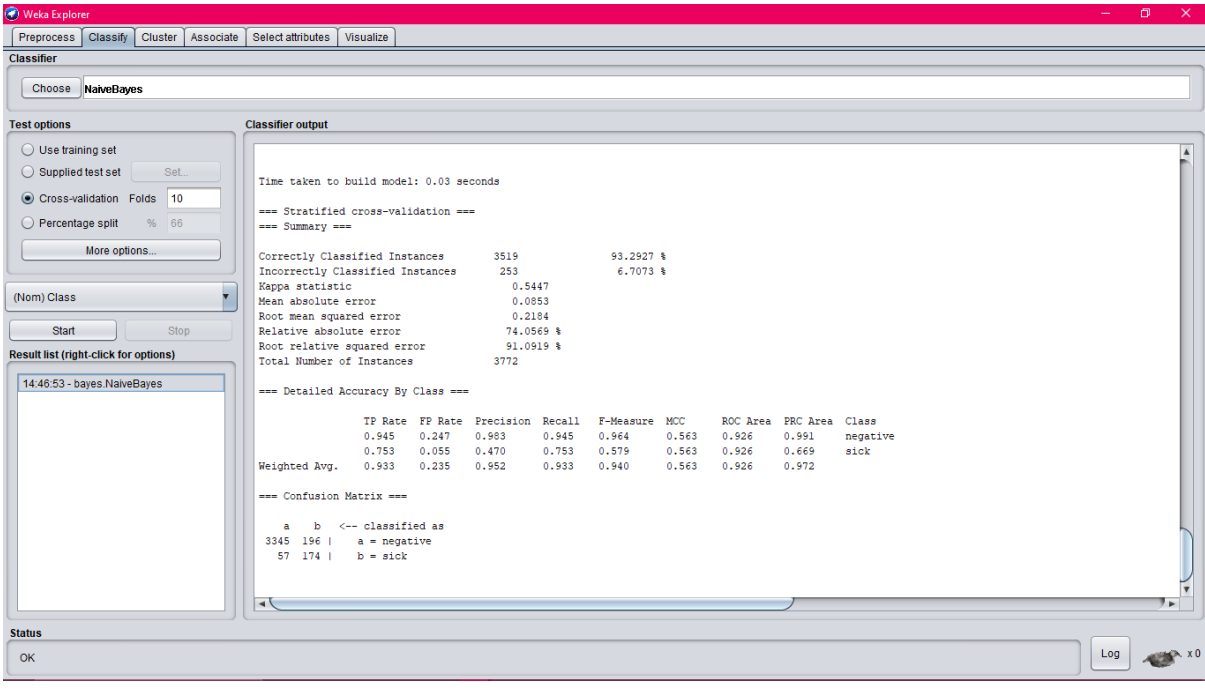


3. What is the effect of the unsupervised filter on the dataset?

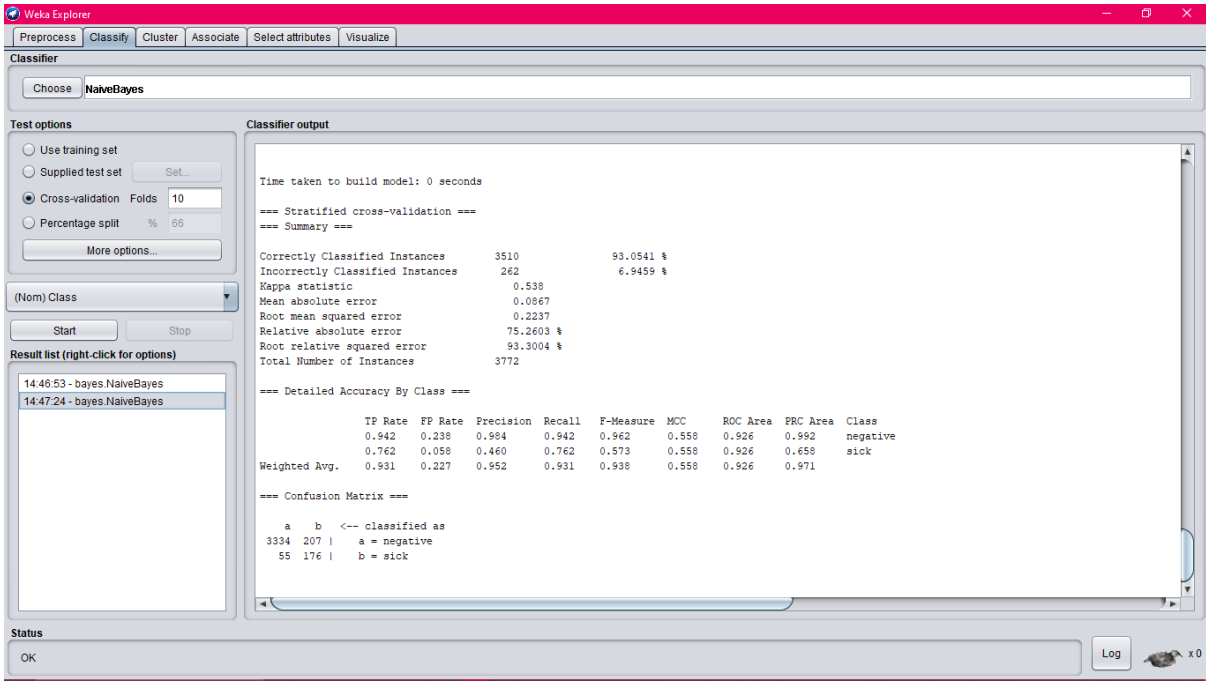
   The filter discretizes a range of numeric attributes in the dataset into nominal attributes.

7. Run the the Naive Bayes classifier after apply the following filters

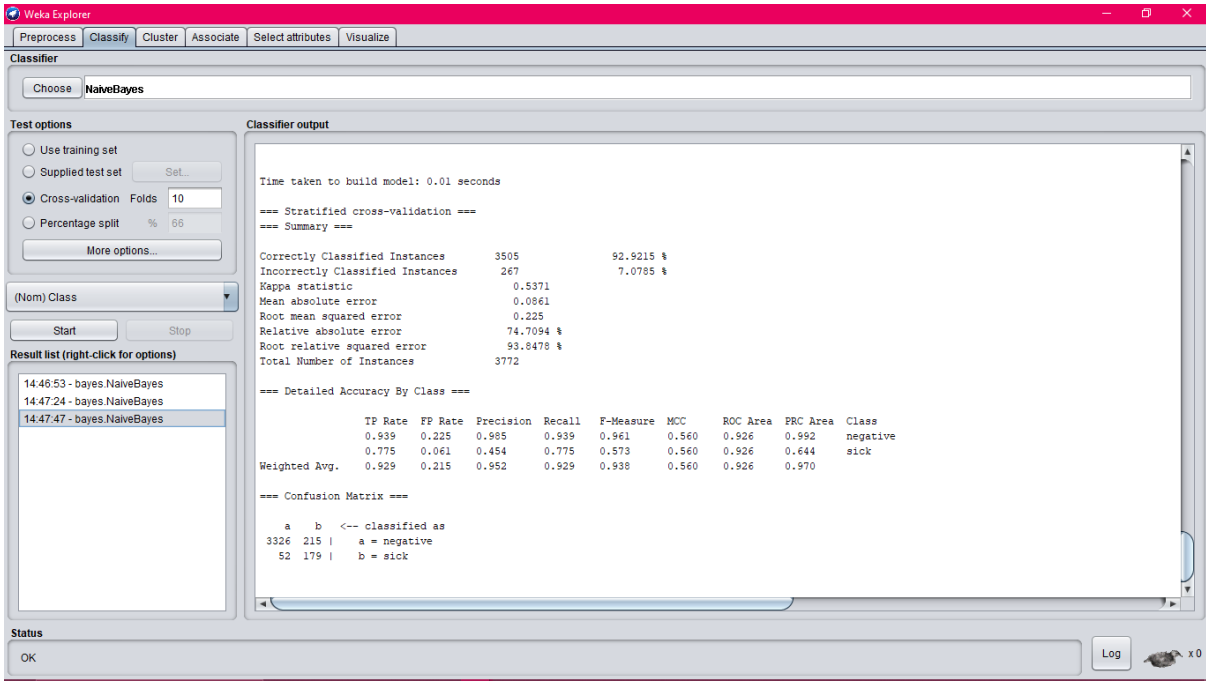
   1. Unsupervised discretized with 'bins'=5



   2. Unsupervised discretized with 'bins'=10



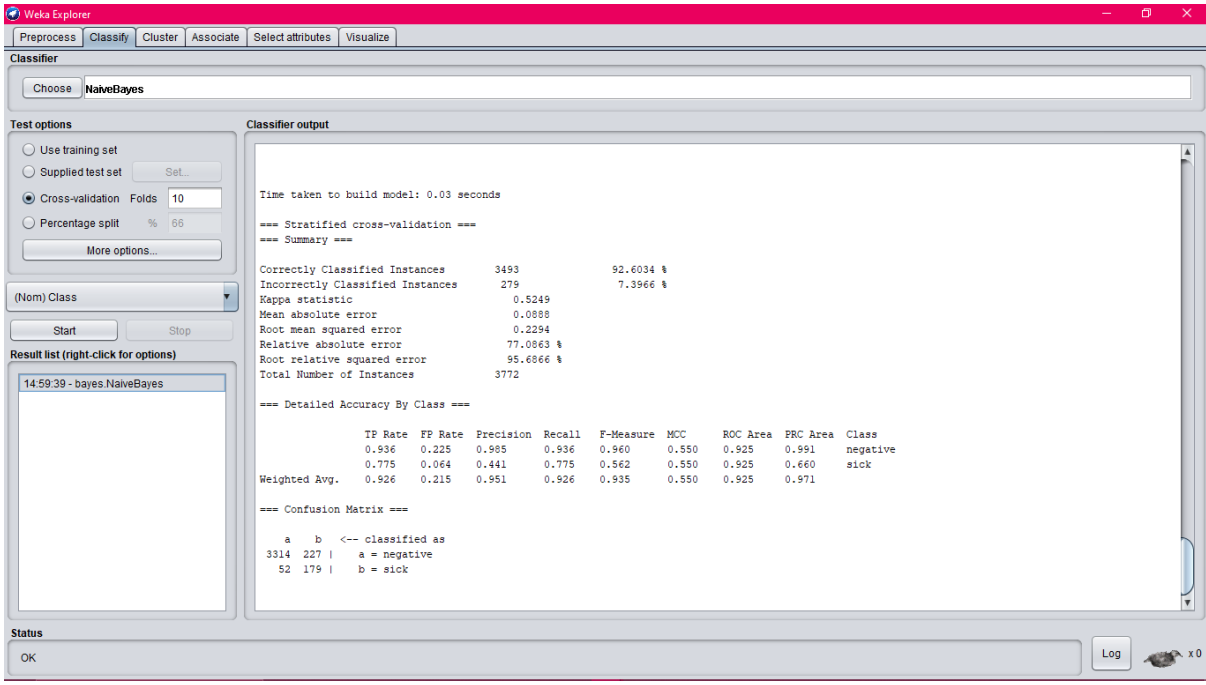
4. Unsupervised discretized with 'bins''=20.



8. Compare the accuracy of the following cases

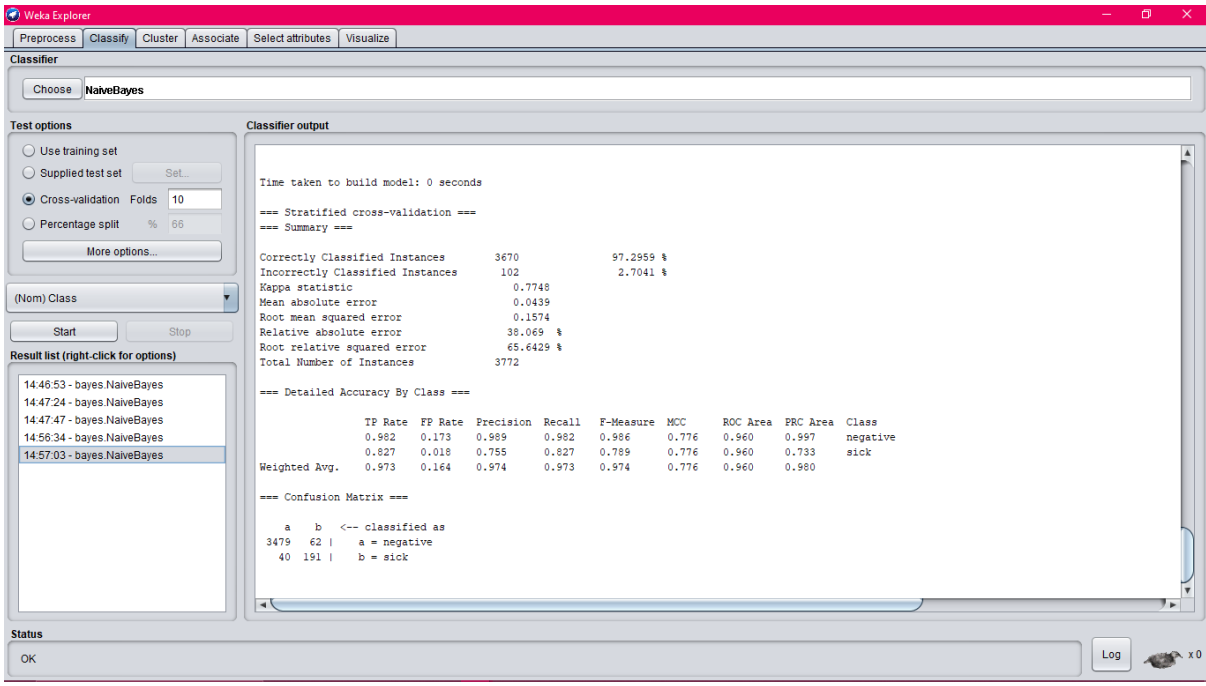
    1. Naive Bayes without discretization filters

    Accuracy = 92.6034%



2. Naive Bayes with a supervised discretization filter

Accuracy = 97.2959%



3. Naive Bayes with an unsupervised discretization filter with different values for

the 'bins attributes.

Bin = 5 93.2927%

Bin = 10 93.0541%

Bin = 20 92.9215 %

This shows that as the number of bins increase, the accuracy decreases.

9. Repeat steps 6 to 8 using equal-frequency binning approach and present your

conclusion.