**EXPERIMENT NO. 5**

Aim: To implement Data Discretization and Visualization.

Requirement: Windows OS and Weka Tool.

Theory:

Data Discretization: Data discretization refers to a method of converting a huge number of data values into smaller ones so that the evaluation and management of data become easy. In other words, data discretization is a method of converting attributes values of continuous data into a finite set of intervals with minimum data loss. There are two forms of data discretization first is supervised discretization, and the second is unsupervised discretization. Supervised discretization refers to a method in which the class data is used. Unsupervised discretization refers to a method depending upon the way which operation proceeds. It means it works on the top-down splitting strategy and bottom-up merging strategy.

Now, we can understand this concept with the help of an example

Suppose we have an attribute of Age with the given values

|  |  |
| --- | --- |
| Age | 1,5,9,4,7,11,14,17,13,18, 19,31,33,36,42,44,46,70,74,78,77 |

**Table before Discretization**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Attribute | Age | Age | Age | Age |
|  | 1,5,4,9,7 | 11,14,17,13,18,19 | 31,33,36,42,44,46 | 70,74,77,78 |
| After Discretization | Child | Young | Mature | Old |

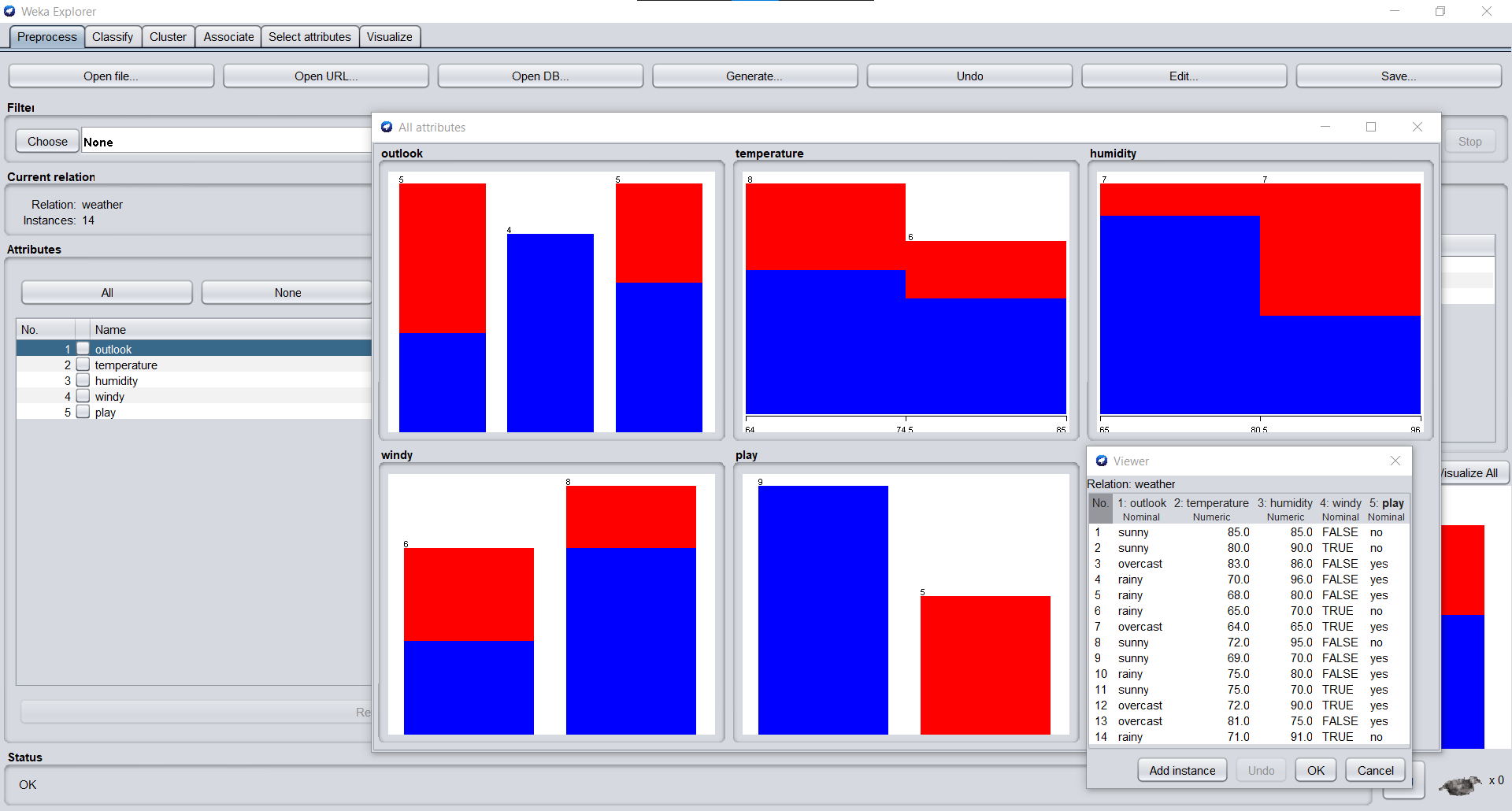
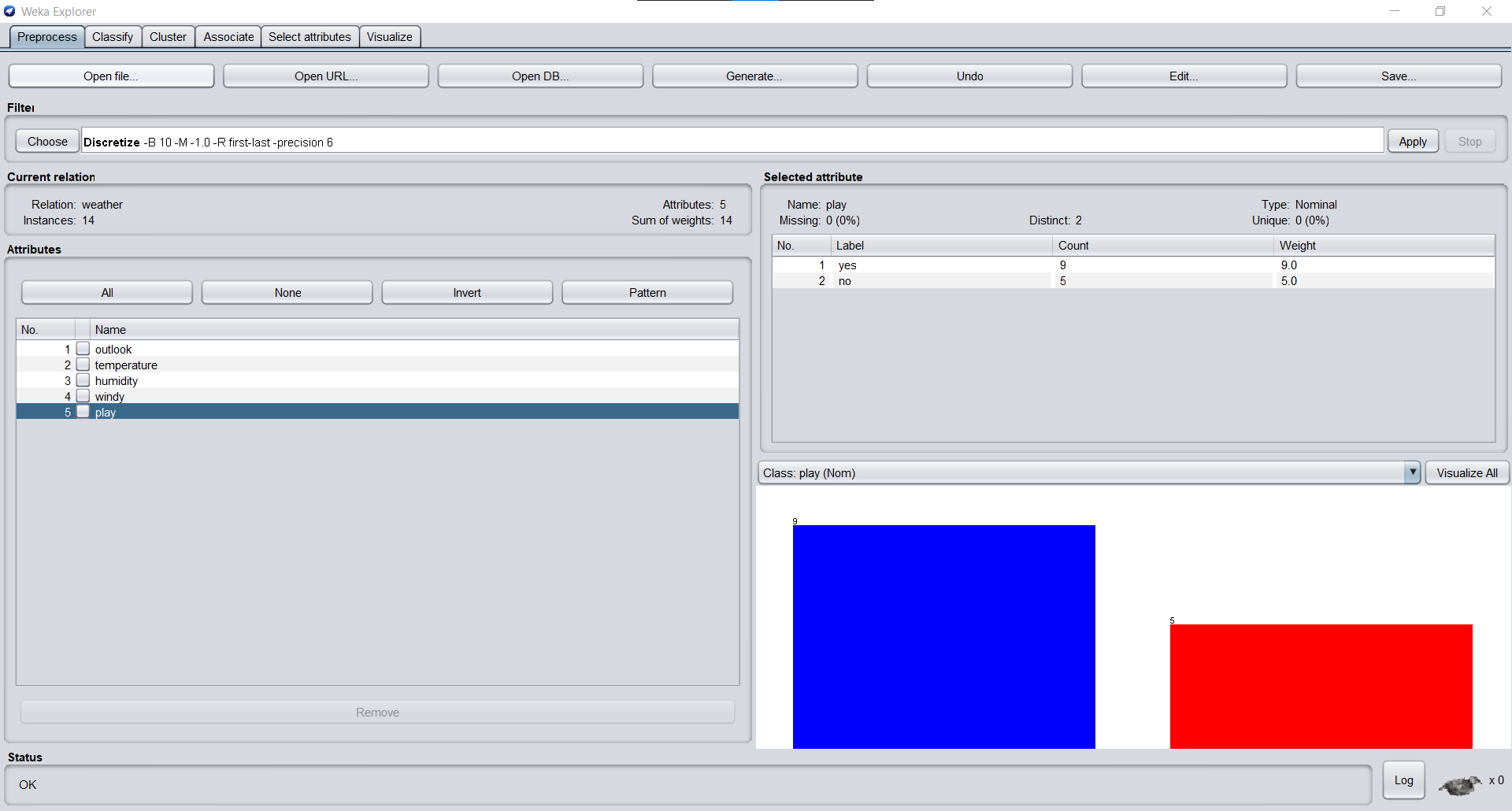
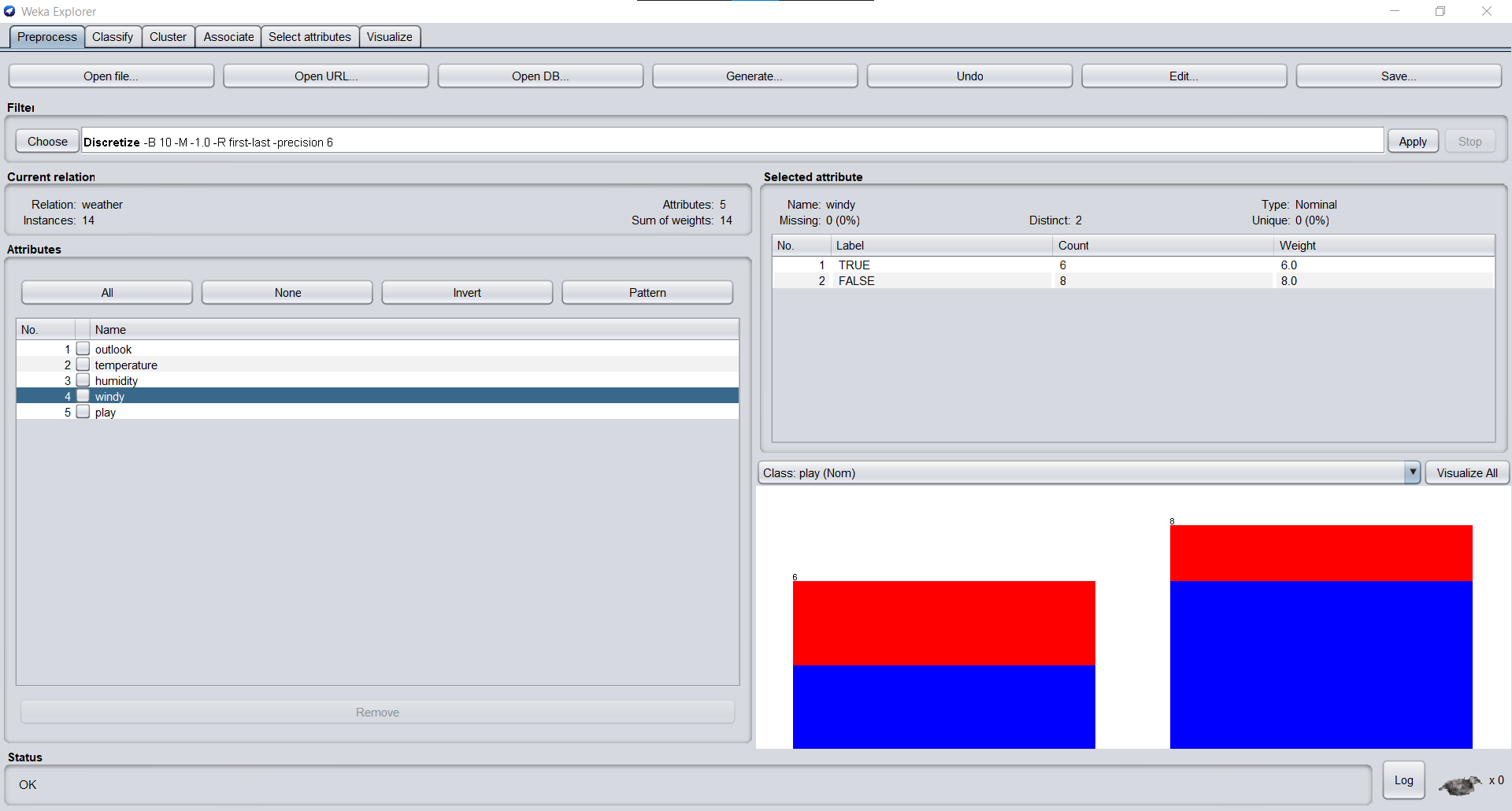
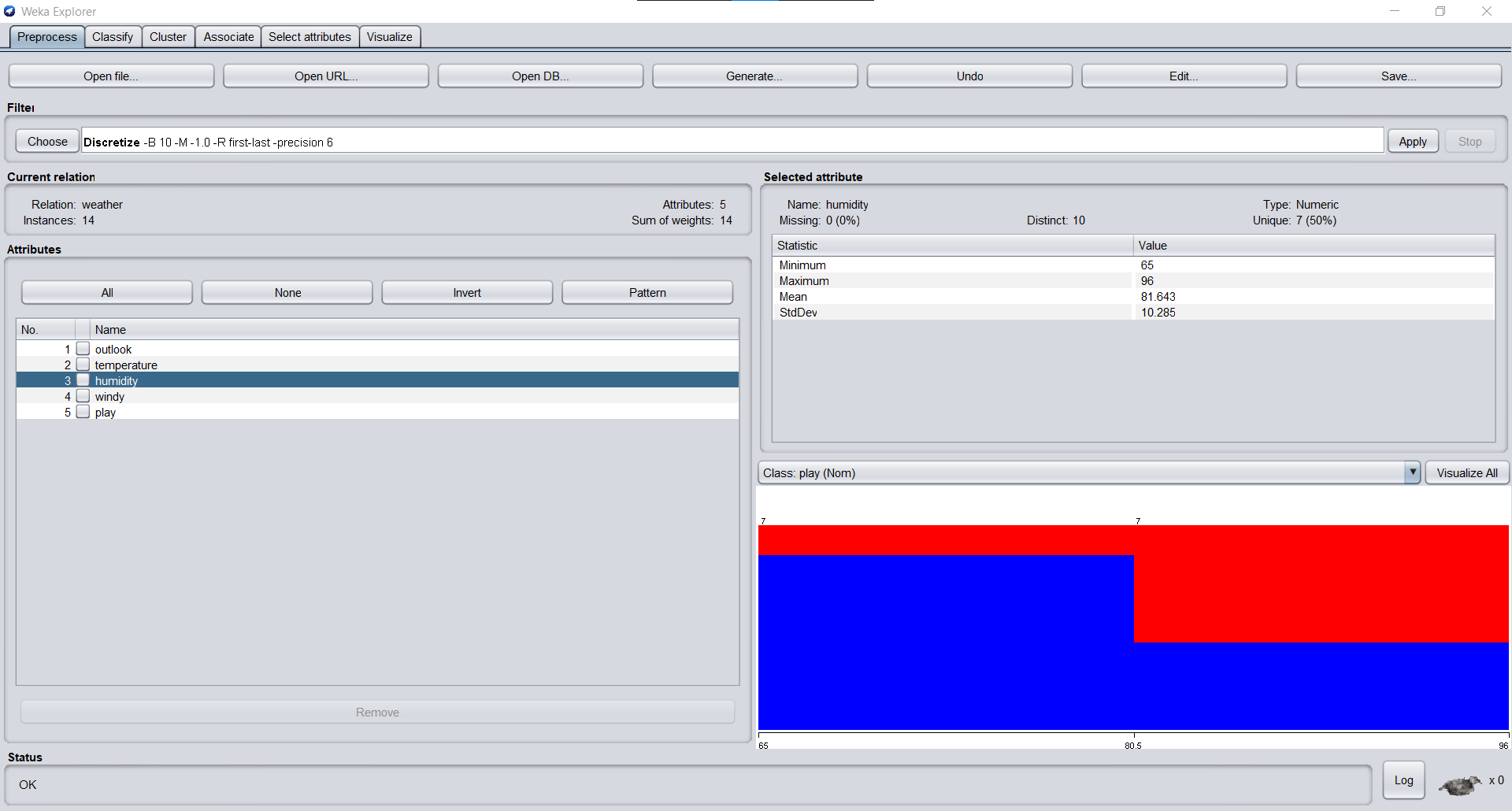
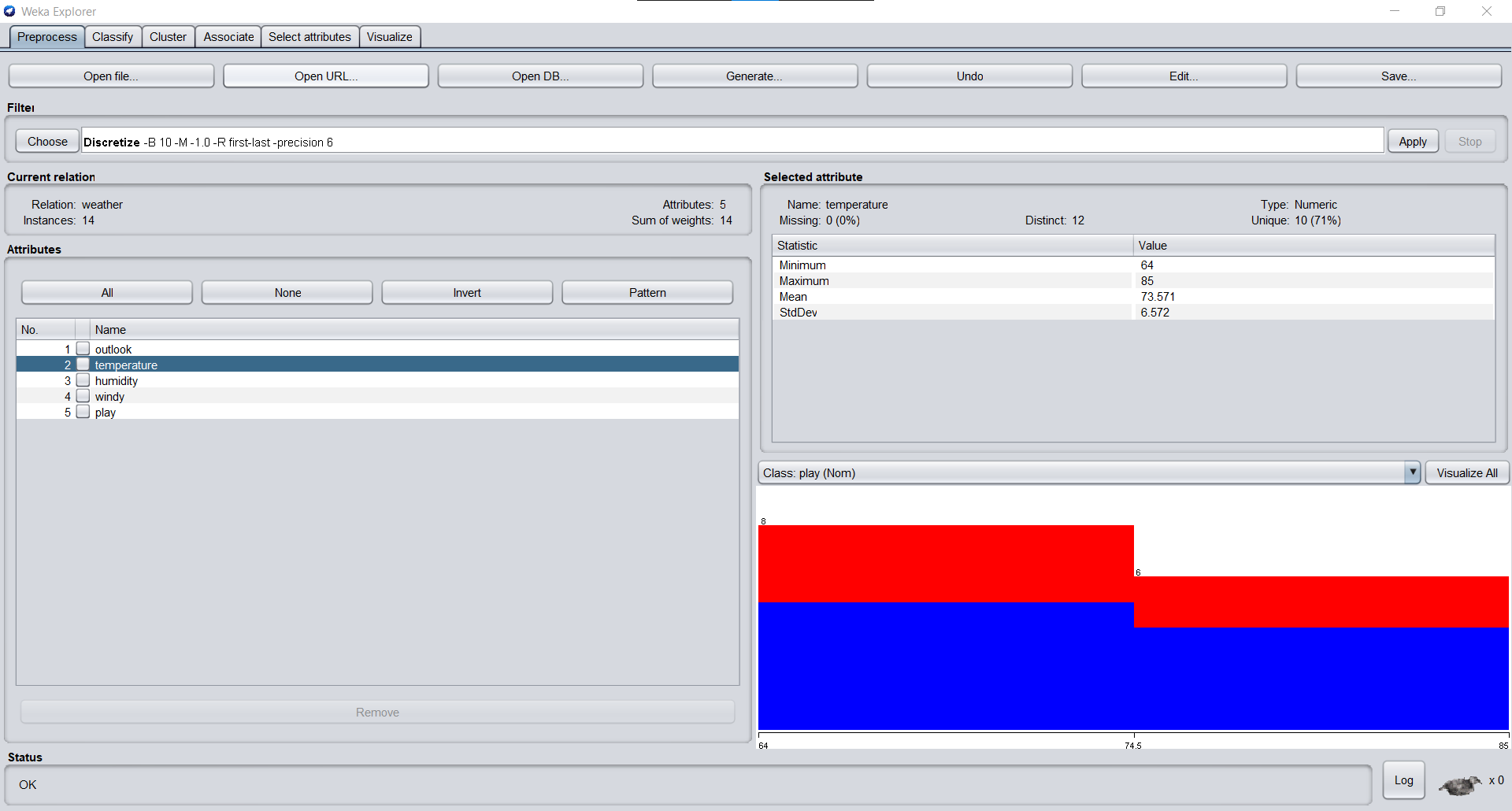
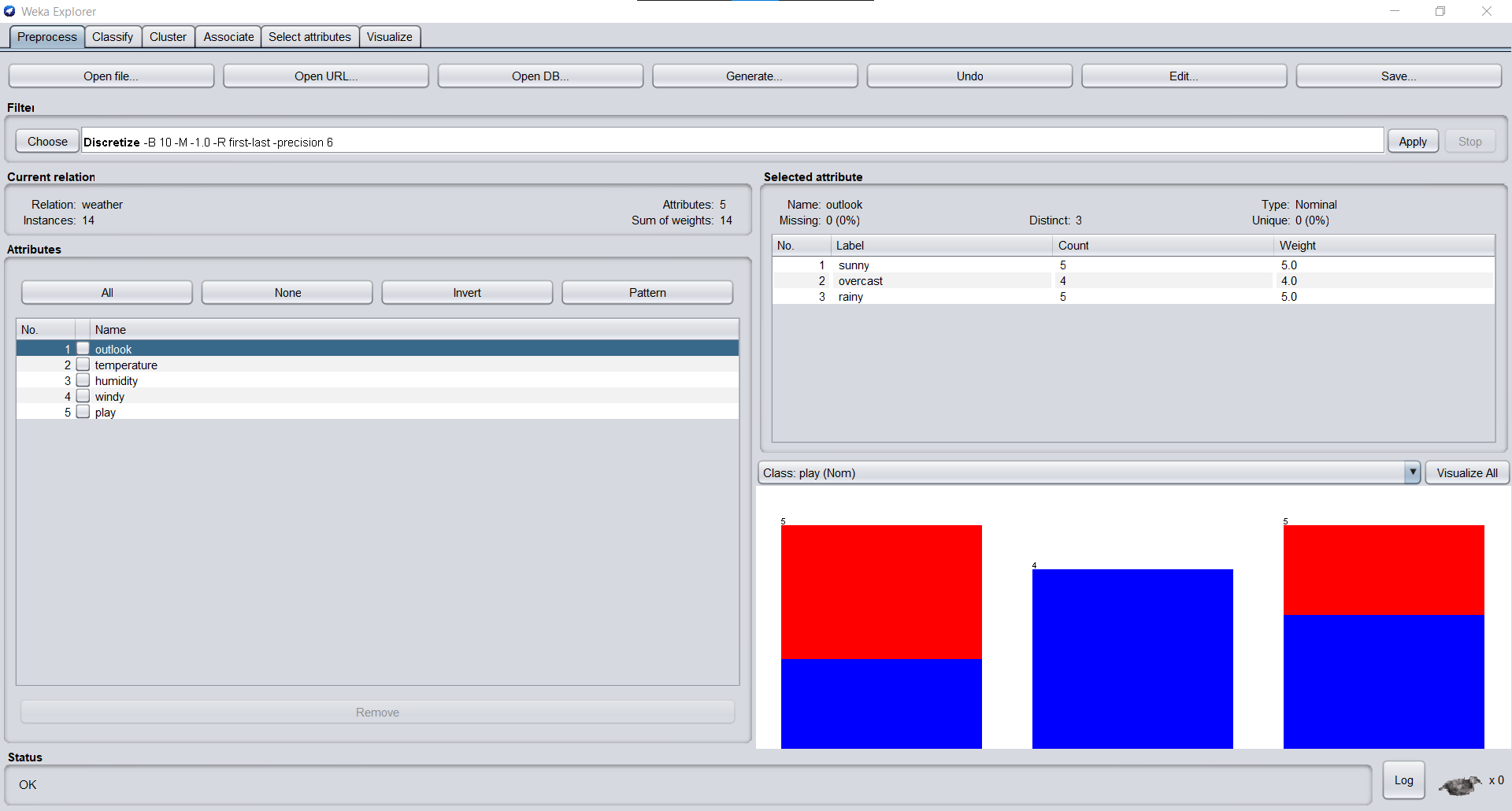
**Table after Discretization**

Data Visualization: Data visualization is the process of translating large data sets and metrics into charts, graphs and other visuals. The resulting visual representation of data makes it easier to identify and share real-time trends, outliers, and new insights about the information represented in the data.

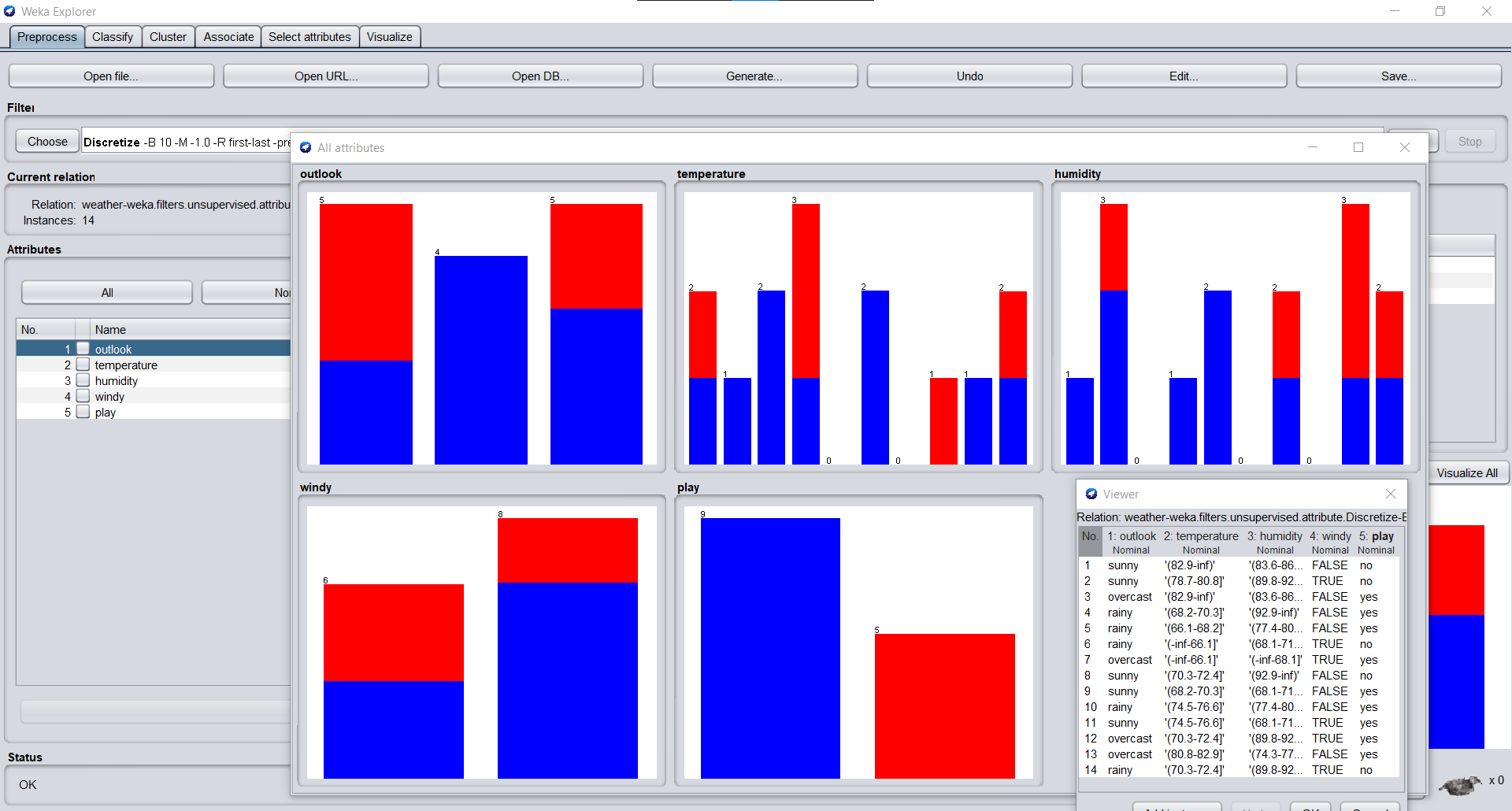
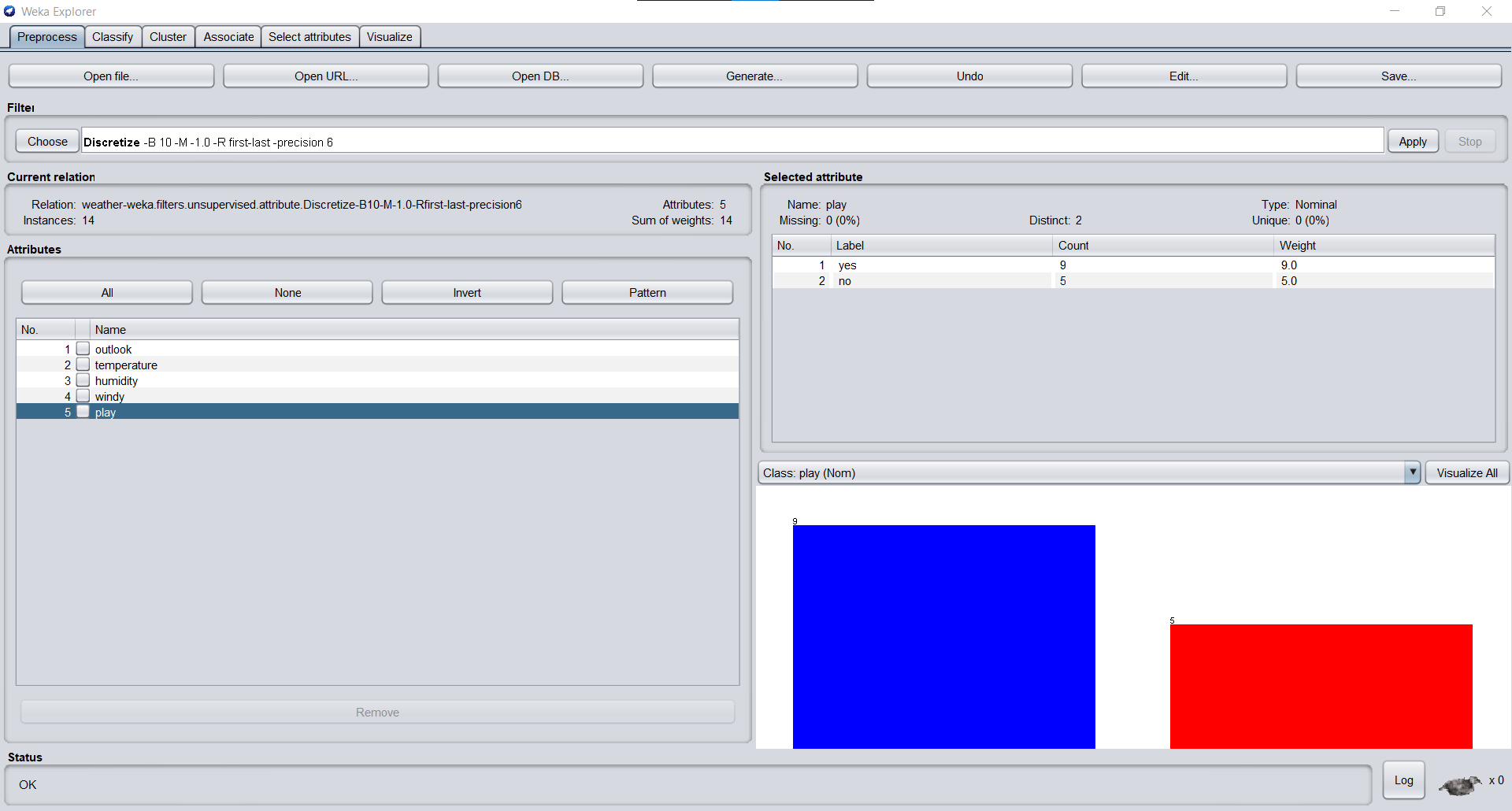
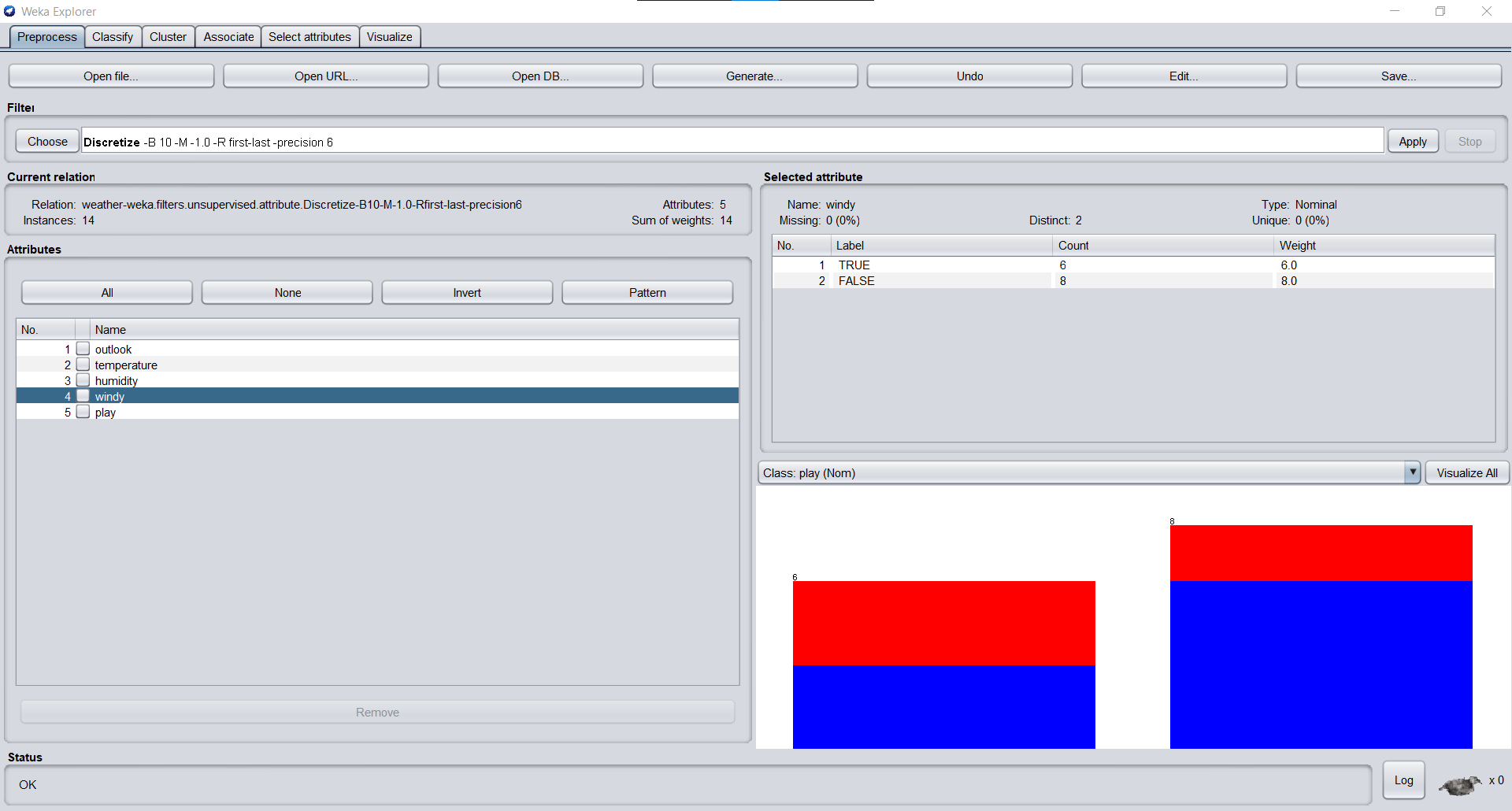
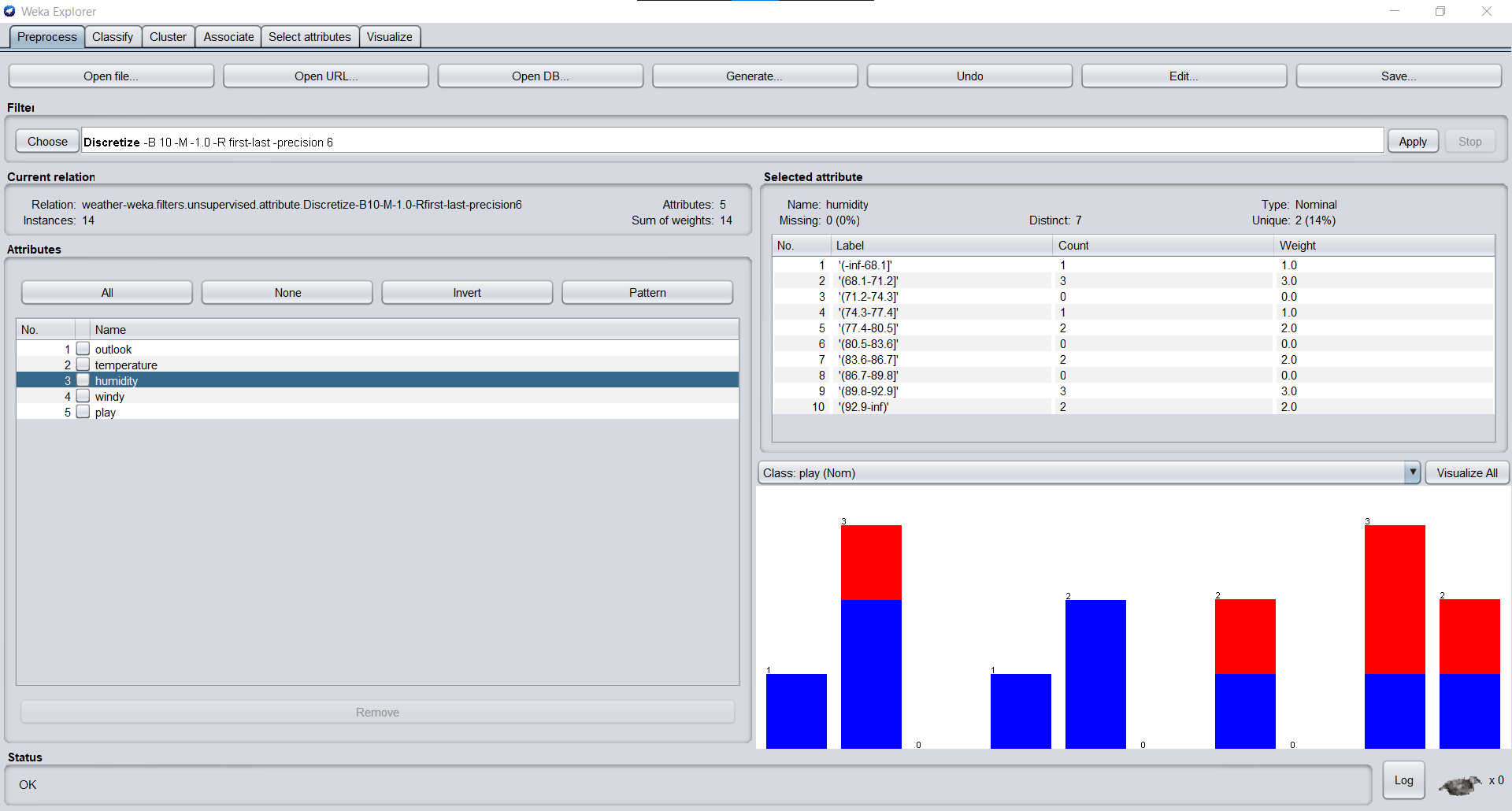
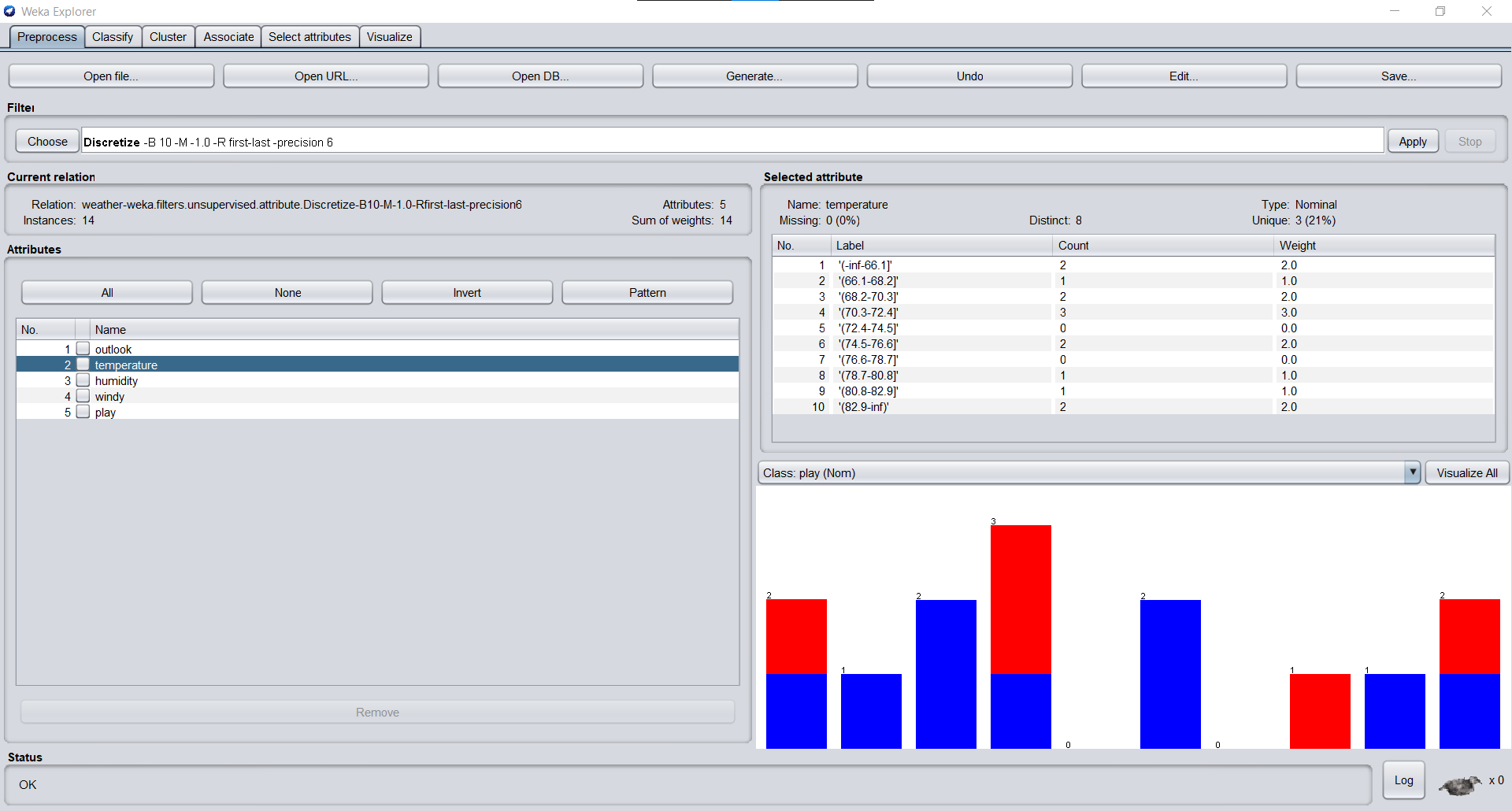
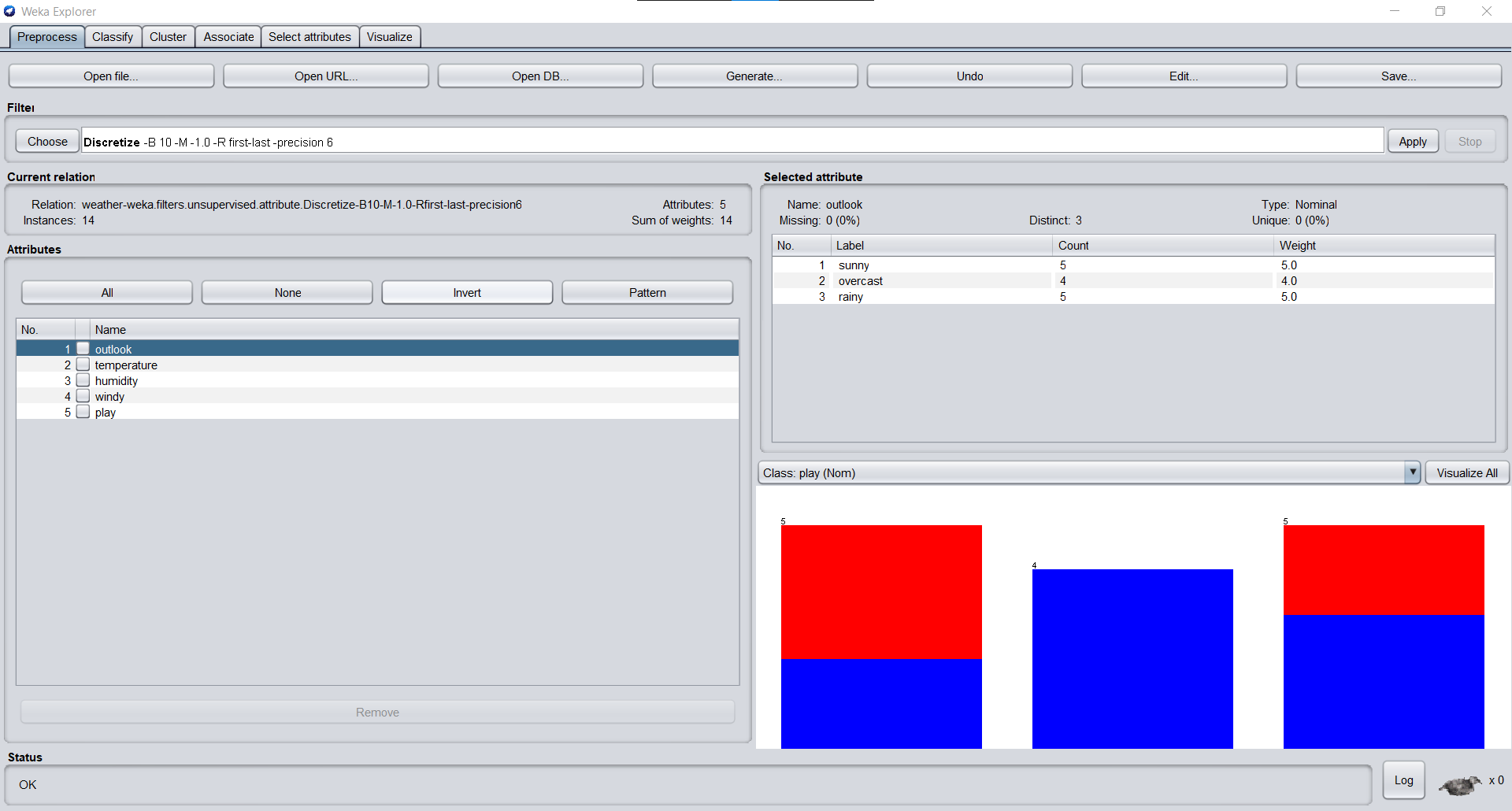
A dashboard is an information visualization tool. It helps you monitor events or activities at a glance by providing insights on one or more pages or screens. Unlike an infographic, which presents a static graphical representation, a dashboard conveys real-time information by pulling complex data points directly from large data sets. An interactive dashboard makes it easy to sort, filter, or drill into different types of data as needed. Data science techniques can be used to identify what is happening, why it's happening, and what will happen next at speed.

As the amount of big data increases, more people are using data visualization tools to access insights on their computer and on mobile devices. Dashboards are used by business people, data analysts, and data scientists to make data-driven business decisions.

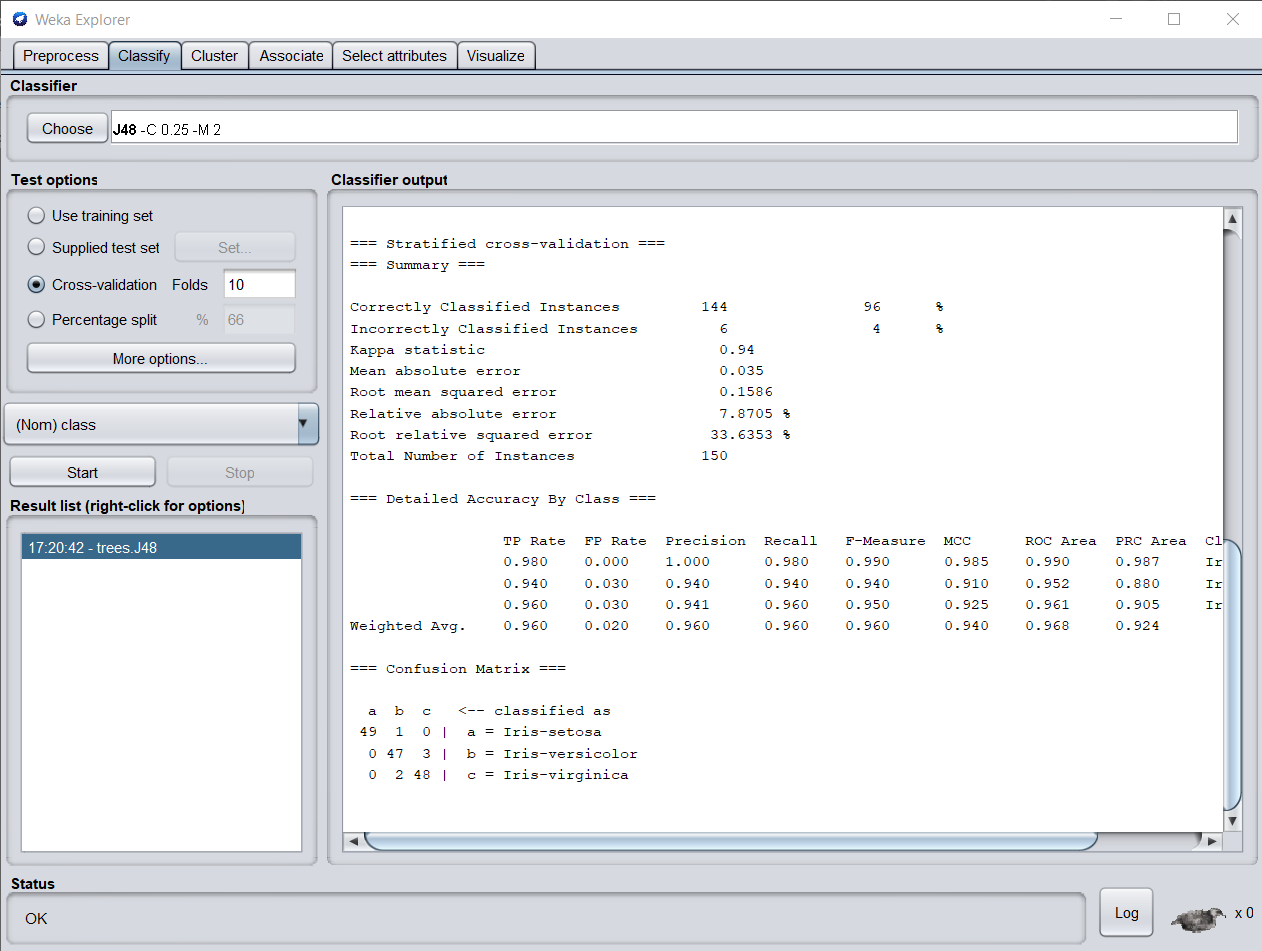
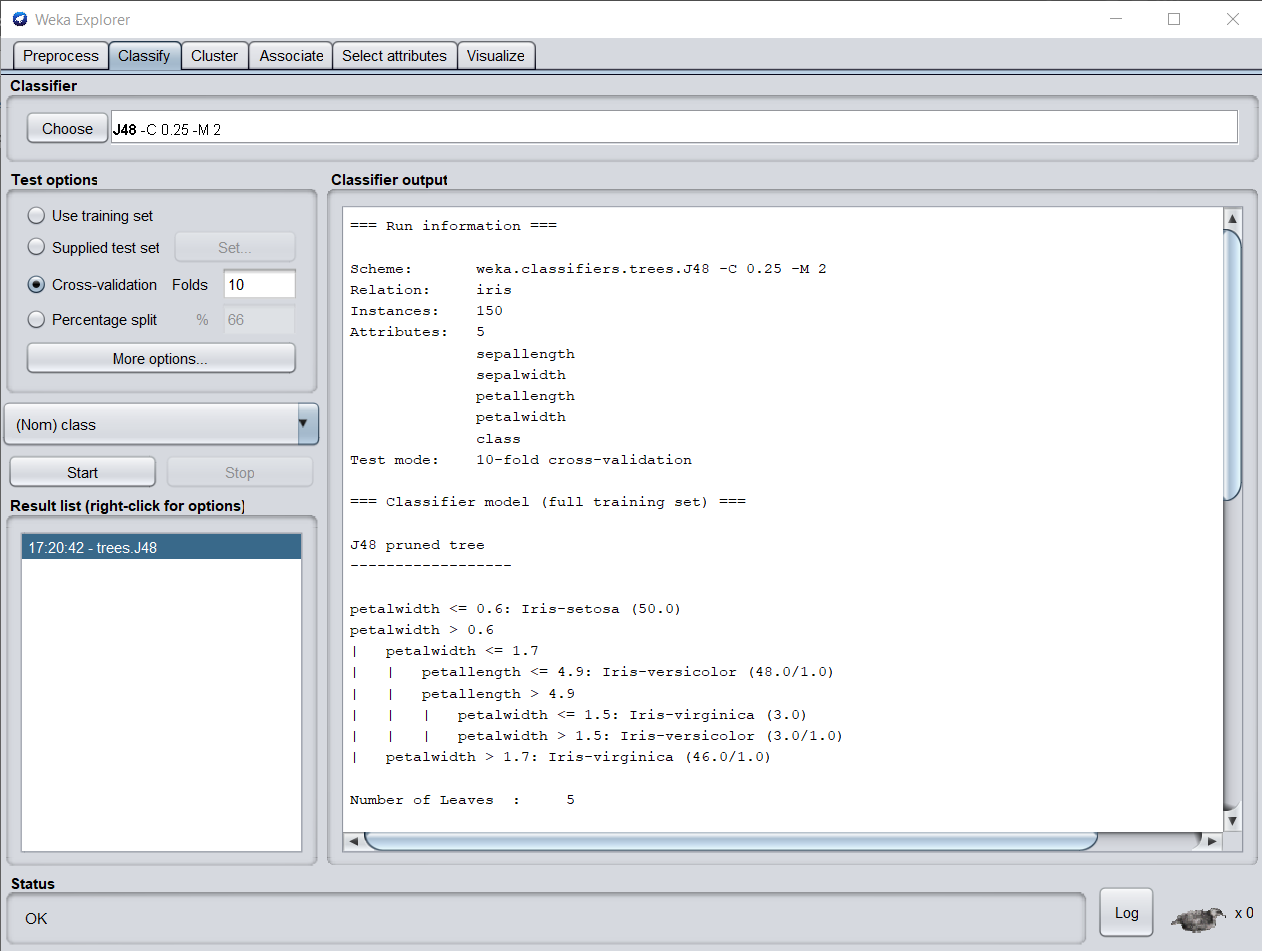
Weather Data before Discretization:

****

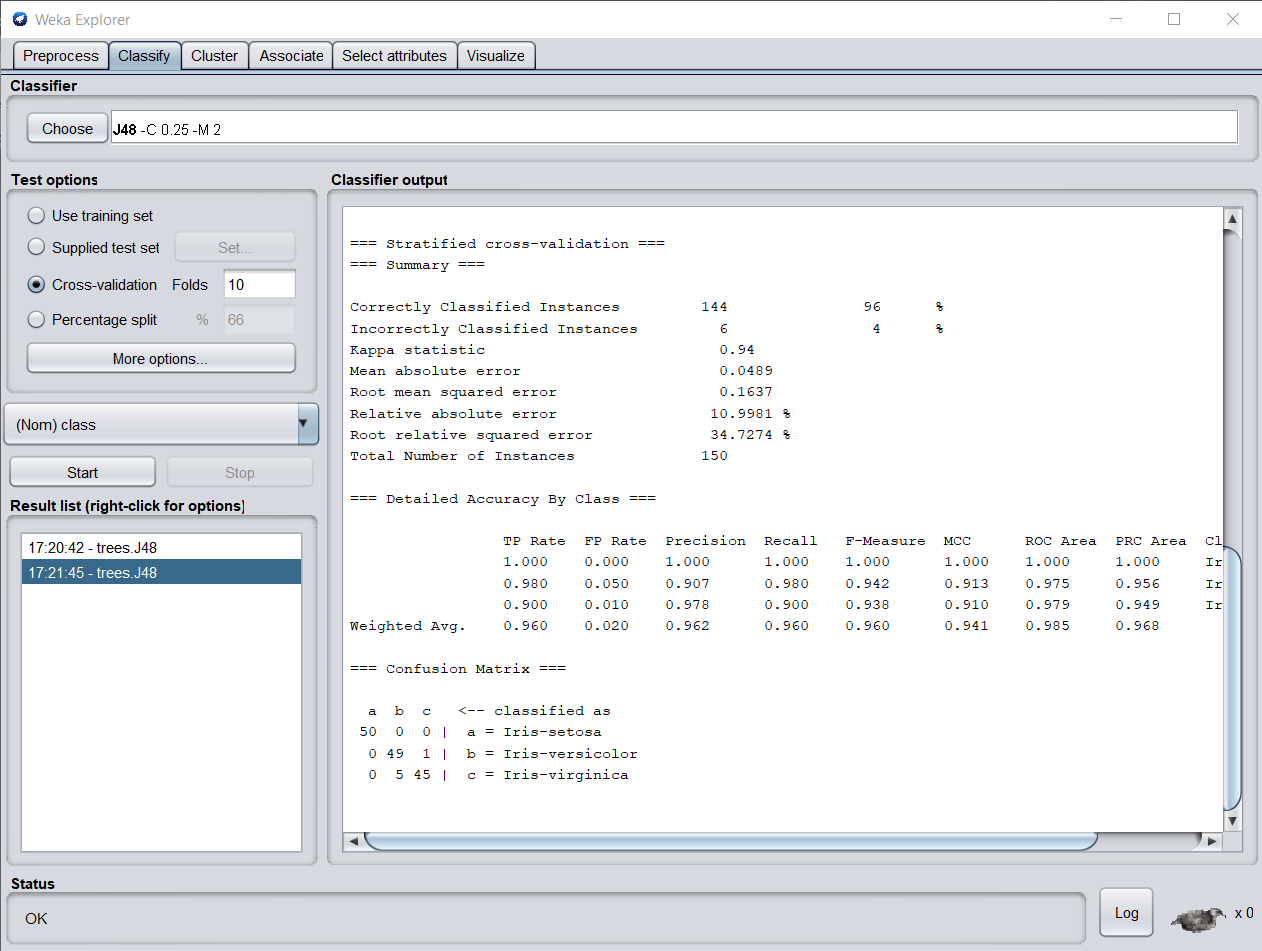
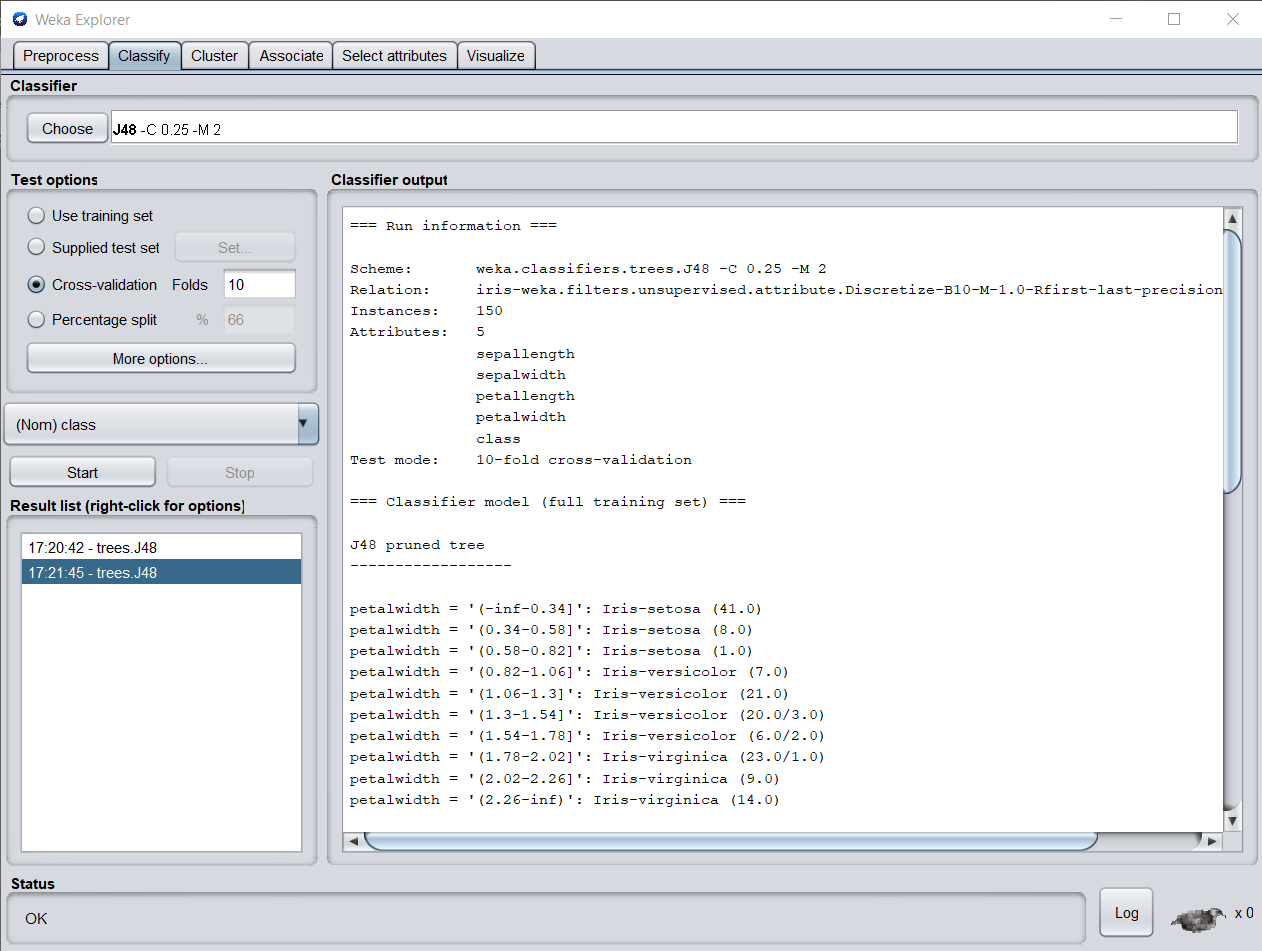
Weather Data after Discretization:

****

J48 Tree classification on iris data set before discretization:



J48 Tree classification on iris data set after discretization:



Conclusion: We have successfully implemented Data Discretization and Visualization on weather Data Set using Weka Tool.