**ASSIGNMENT-1**

**Take one DATASET as example and Load them as shown in below :**

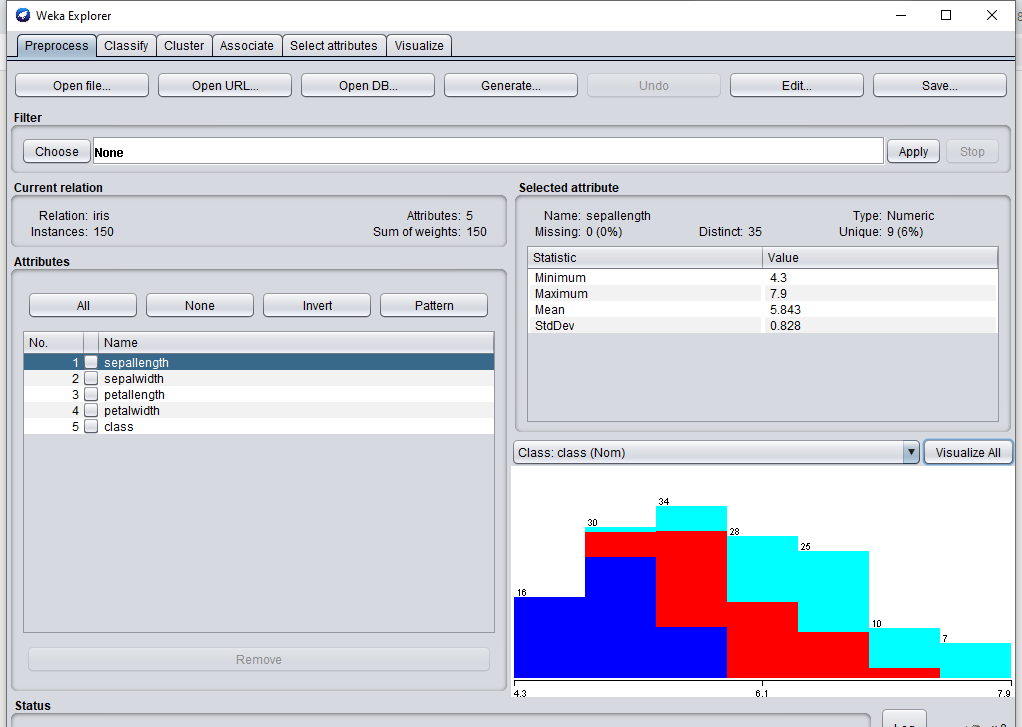
**FOR IRIS DATASET:**

Initialy open the iris .arff file from Weka and start explore. By exploring them you can visualize the following elements as shown them bellow.

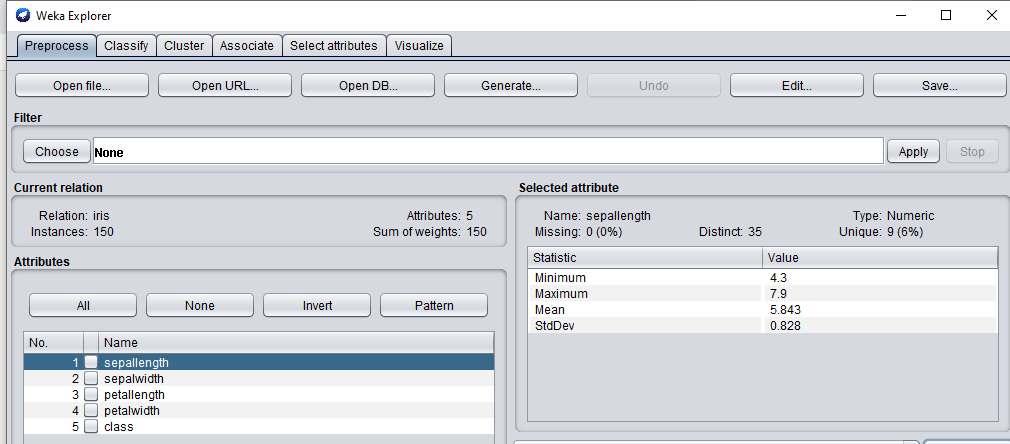
1. List the attribute names and their types

|  |  |
| --- | --- |
| **Attribute Name** | **Attribute Type** |
| sepalength | Numeric |
| sepalwidth | Numeric |
| petalength | Numeric |
| petalwidth | Numeric |
| class | Nominal |

The iris dataset consists of sepalength and sepalwidth, petalength, petalwidth, class are the attributes of the dataset. And the Attributes are followed by their types like Numeric, Nominal are types of the attributes of the Dataset.

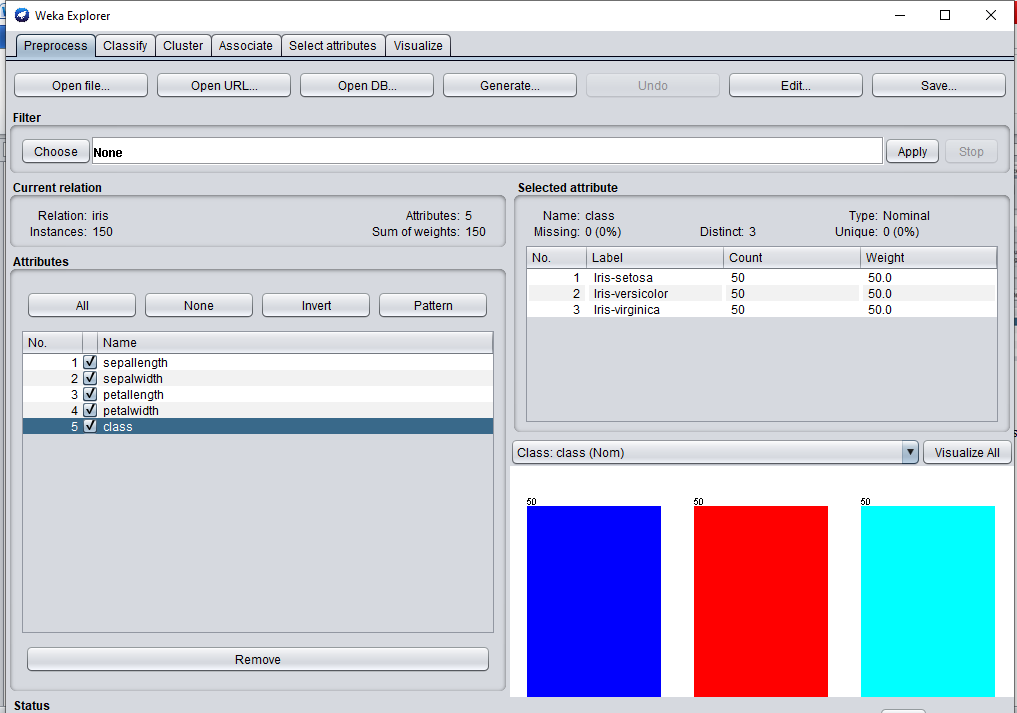


1. The Number of records in the dataset are 150 as shown below.

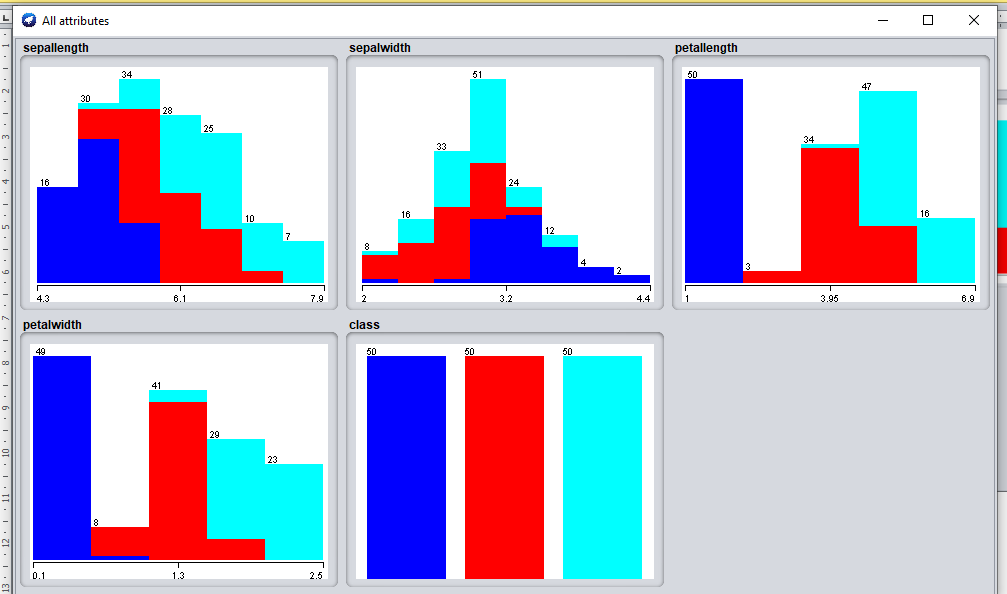


Therefore there are 150 records.

1. Identify the class attribute in the dataset : class



1. Graphical histogram representation of all attributes against class attribute



1. Determine the number of records for each class:



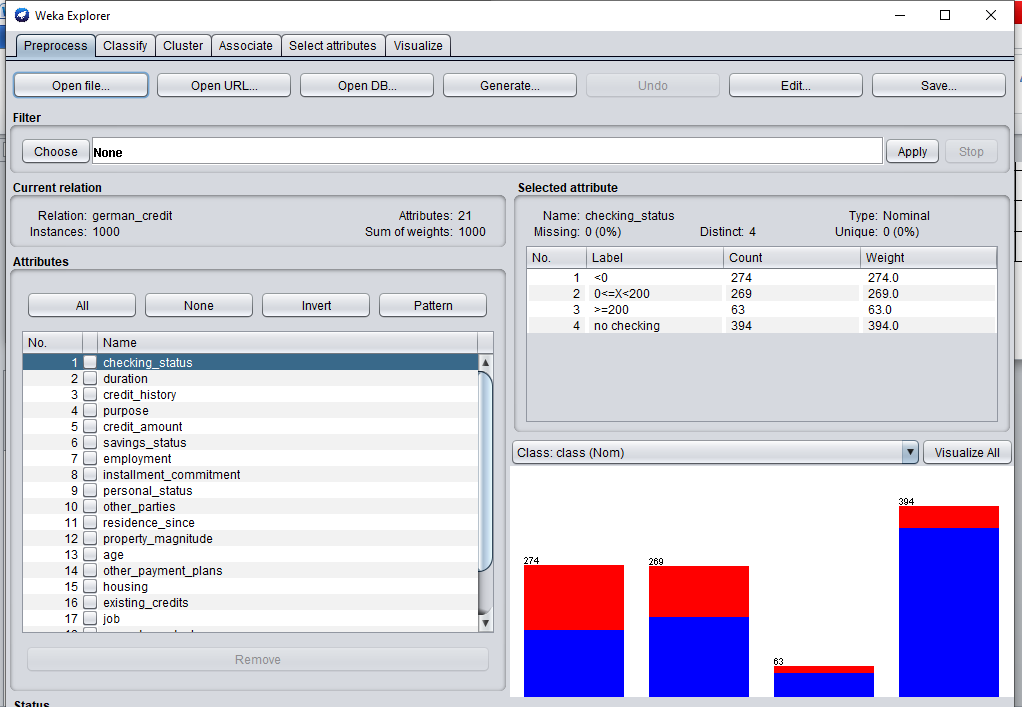
* Iris-setosa – 50 records
* Iris-virginica – 50 records
* Iris-versicolor – 50 records

**FOR GERMAN CREDIT DATASET :**

1. List the attribute names and their types:

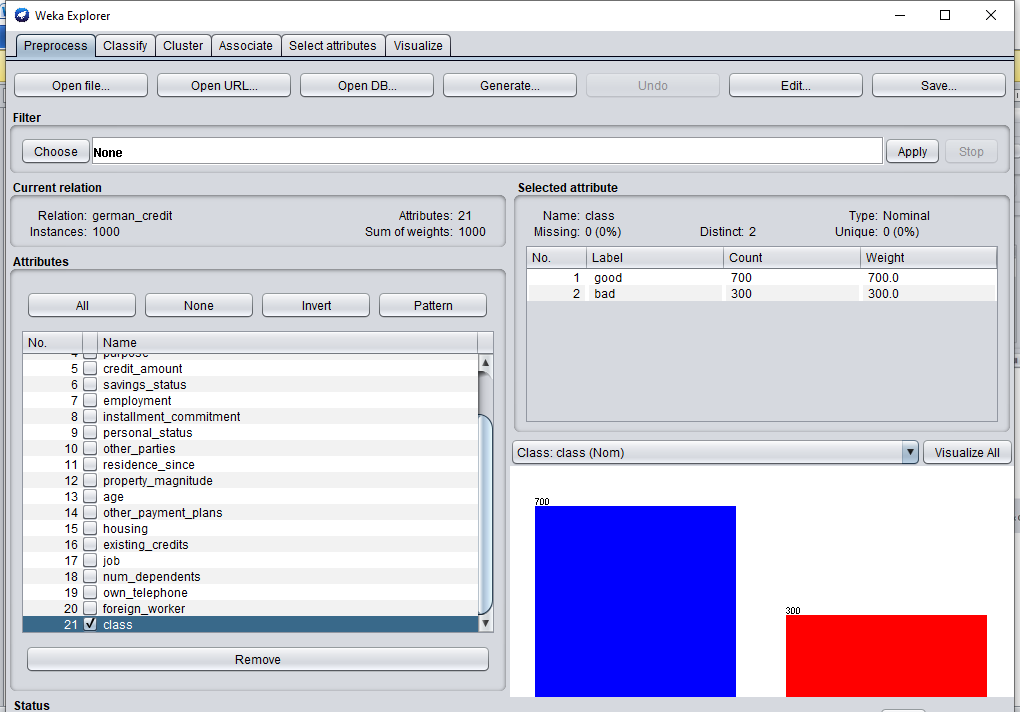
|  |  |
| --- | --- |
| **Attribute Name** | **Attribute Type** |
| credit\_history | Nominal |
| purpose | Nominal |
| credit\_amount | Numeric |
| serving\_status | Nominal |
| employment | Nominal |
| installment\_commitment | Numeric |
| personal\_status | Nominal |
| other\_parties | Nominal |
| residence\_since | Numeric |
| property\_magnitude | Nominal |
| age | Numeric |
| duration | Numeric |
| housing | Nominal |
| existing\_credits | Numeric |
| job | Nominal |
| other\_payment\_plans | Nominal |
| checking\_status | Nominal |
| num\_dependents | Numeric |
| own\_telephone | Nominal |
| foreign\_worker | Nominal |
| class | Nominal |

1. Number of records in the dataset

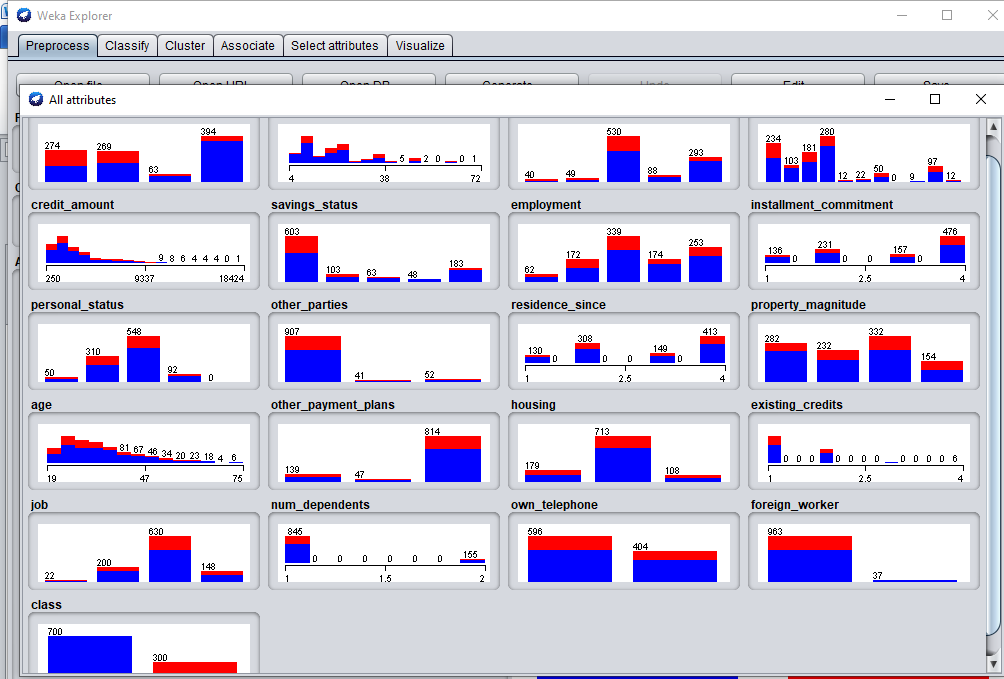


* Therefore there are 1000 Records

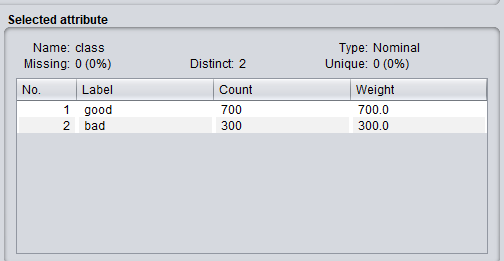
1. Identify the class attribute in the dataset : class



1. Graphical histogram representation of all attributes against class attribute



1. Determine the number of records for each class:



* Good – 700 records
* Bad – 300 records