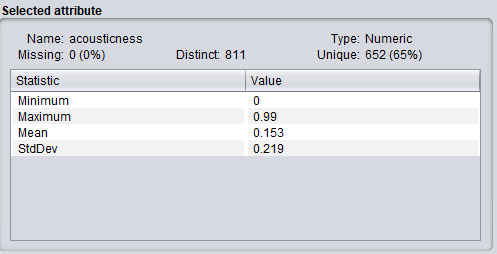
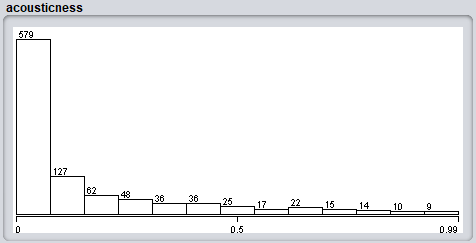
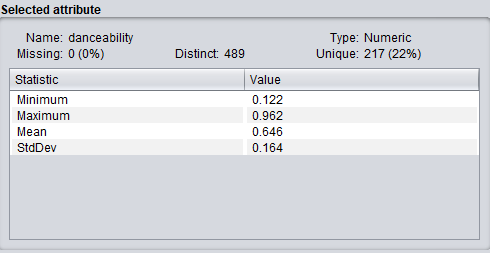
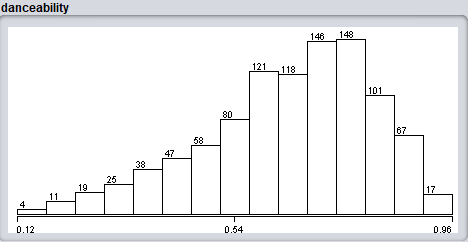
**Preprocessing Using Weka Tool**

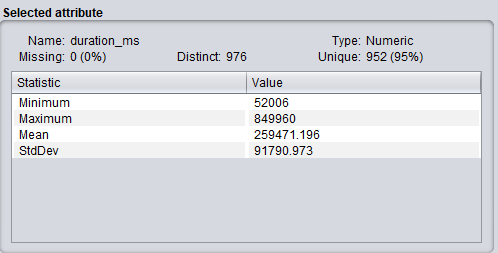
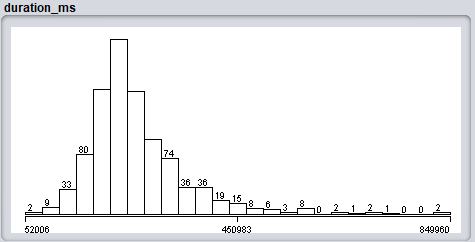
Data pre-processing is an important ste[p in the data mining](https://en.wikipedia.org/wiki/Data_mining) process. The phrase ["garbage in,](https://en.wikipedia.org/wiki/GIGO) [garbage out"](https://en.wikipedia.org/wiki/GIGO) is particularly applicable to data mining and [machine learning](https://en.wikipedia.org/wiki/Machine_learning) projects. Data-gathering methods are often loosely controlled, resulting in [out-of-range](https://en.wikipedia.org/w/index.php?title=Range_error&amp;action=edit&amp;redlink=1) values (e.g., Income:−100), impossible data combinations (e.g., Sex: Male, Pregnant: Yes), [missing values,](https://en.wikipedia.org/wiki/Missing_values) etc.

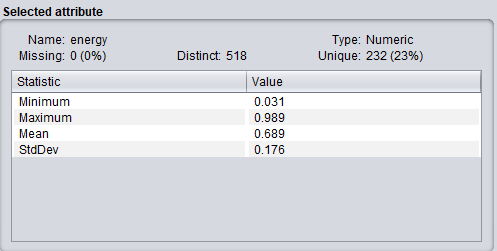
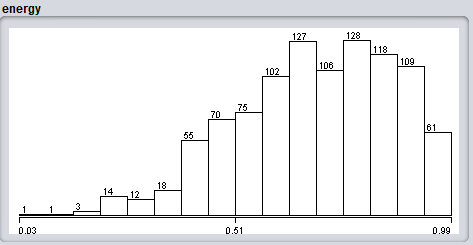
Here we have done preprocessing using **Weka** tool. In this, we first import the dataset & then visualize all the attributes of the data one by one. Such that we come to know about the attributes that which of them attribute is neither numeric nor nominal. After identifying those attributes, we removed those attributes before applying any algorithm such as decision tree, random forest, etc. So that we can get to that which algorithm will be suited best for to get the maximum percentage of correctly classified data.

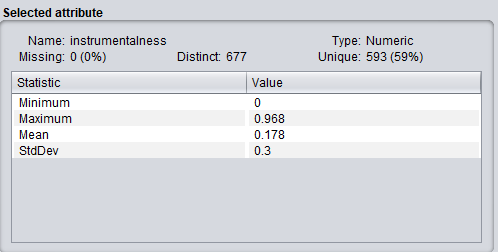
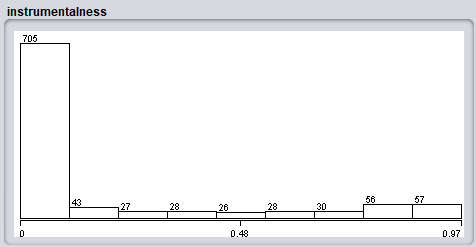
VISUALIZATION of all the Attributes using Weka Tool

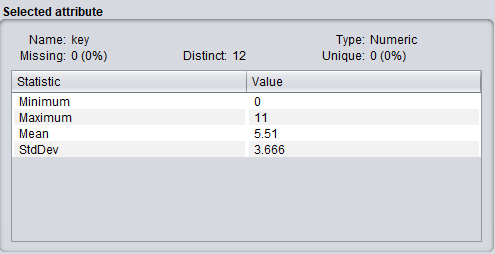
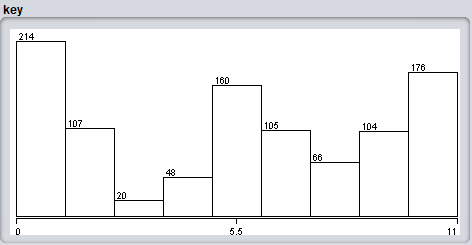


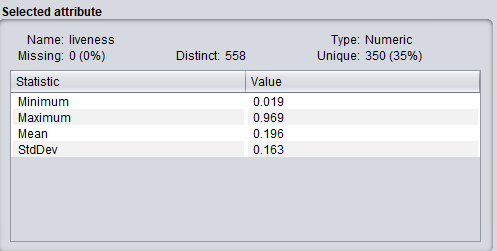
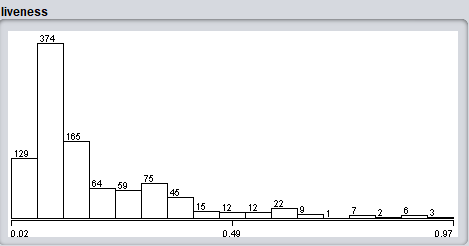


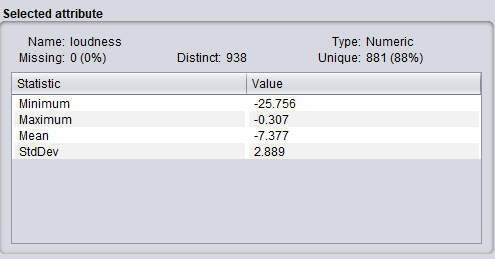
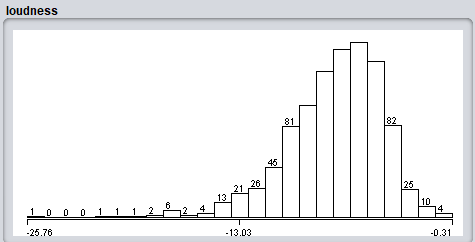


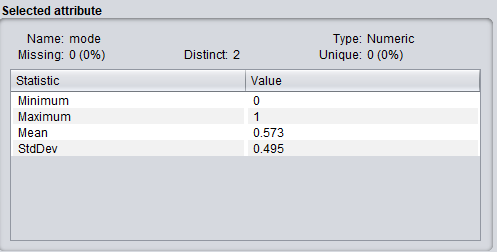
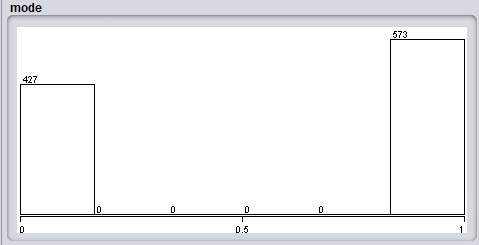


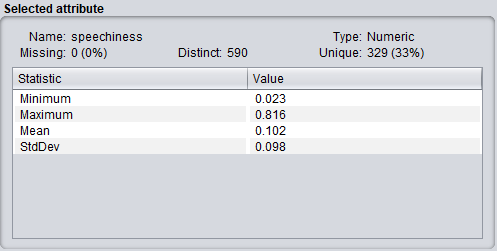
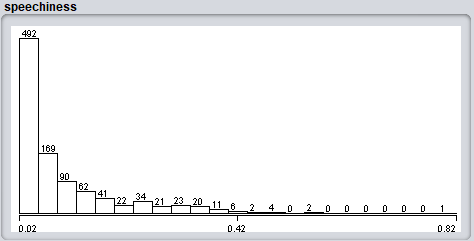


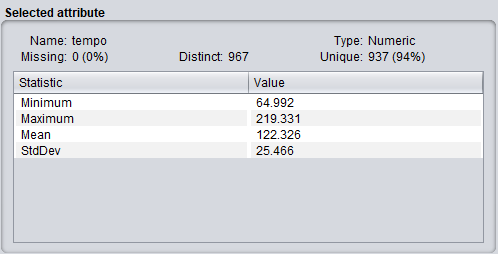
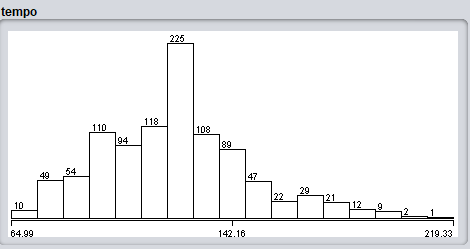


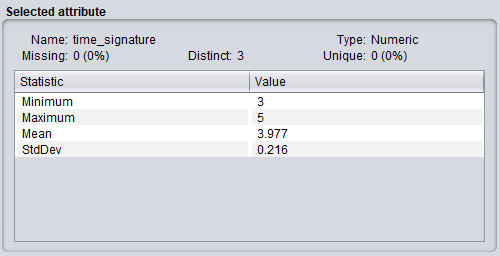
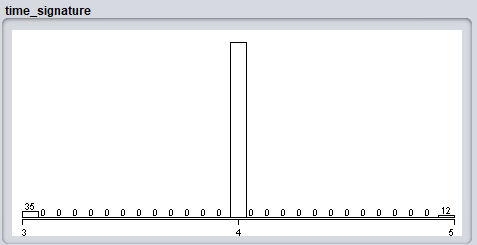


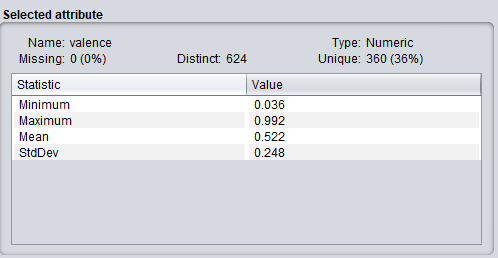
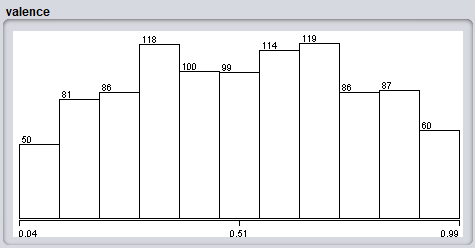












**Clustering using Weka Tool**

Clusterer Model

kMeans

======

Number of iterations: 3

Within cluster sum of squared errors: 497.6963609508743

Initial starting points (random):

Cluster 0: 0.00904,0.614,196352,0.907,0.000131,4,0.36,-2.536,0,0.143,122.906,4,0.354,1

Cluster 1: 0.0286,0.76,479547,0.391,0.00691,1,0.0582,-8.73,1,0.0353,116.873,4,0.965,1

Missing values globally replaced with mean/mode

Final cluster centroids:

Cluster#

Attribute Full Data 0 1

(1000.0) (427.0) (573.0)

======================================================

acousticness 0.153 0.1501 0.1551

danceability 0.646 0.647 0.6453

duration\_ms 259471.196 269529.3747 251975.8342

energy 0.6886 0.6932 0.6852

instrumentalness 0.178 0.1876 0.1708

key 5.51 6.5012 4.7714

liveness 0.1955 0.1969 0.1944

loudness -7.3768 -7.3926 -7.3651

mode 0.573 0 1

speechiness 0.1025 0.1025 0.1025

tempo 122.3263 123.7733 121.2481

time\_signature 3.977 3.9883 3.9686

valence 0.522 0.5322 0.5144

target 1 1 1

Time taken to build model (full training data) : 0.04 seconds

=== Model and evaluation on training set ===

Clustered Instances

1. 427 ( 43%)

1 573 ( 57%)

