

Data Mining Tool Usage - Weka

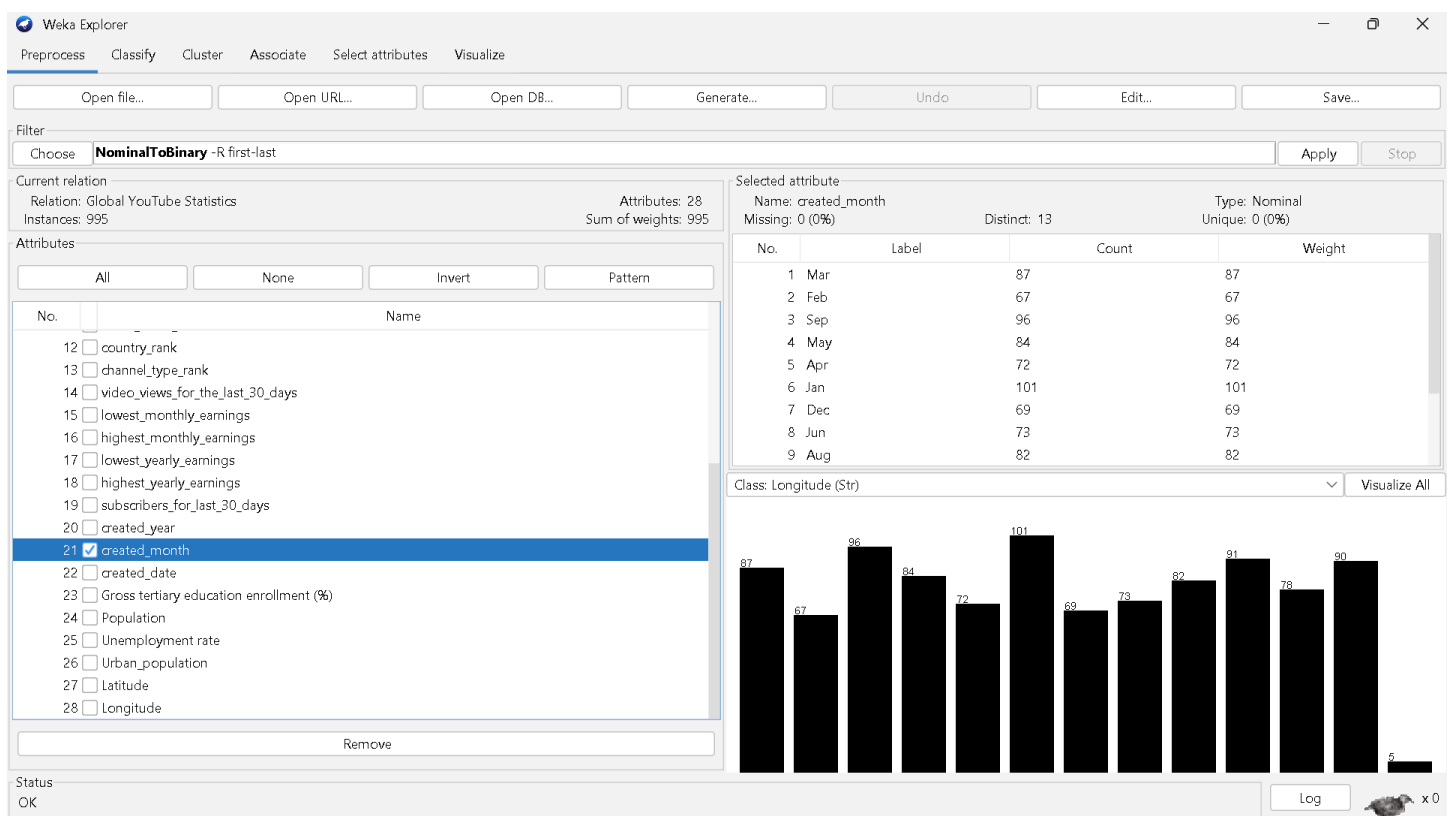
Weka is a collection of machine learning algorithms for data mining tasks. It contains tools for data preparation, classification, regression, clustering, association rules mining, and visualization.

Data Preprocessing

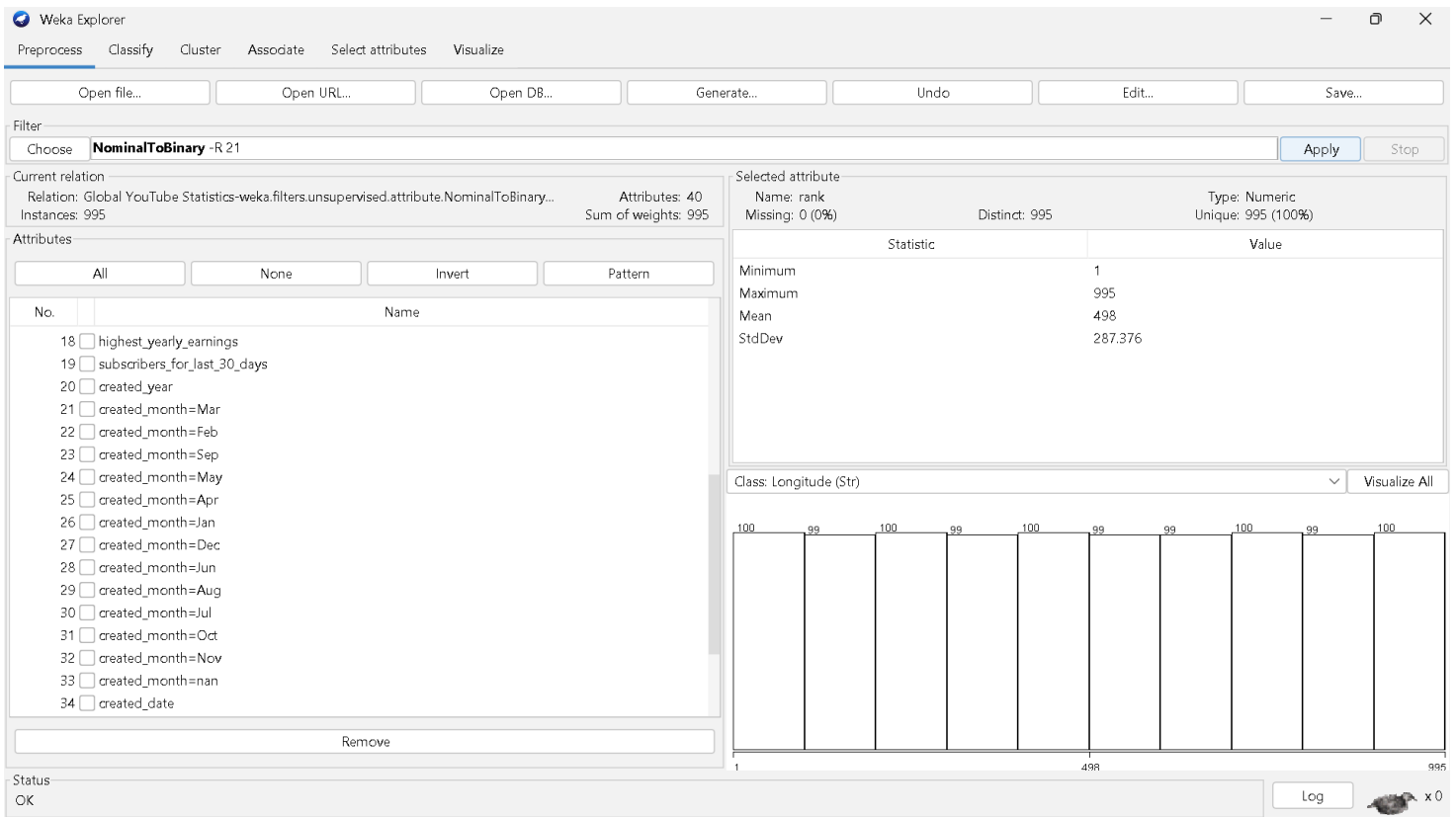
1. Convert Nominal attribute (created_month) to Binary attribute

Some machine learning models, such as decision trees or support vector machines, are designed to work with binary attributes. Converting nominal attributes to binary allows the application of these models without additional preprocessing.

Before conversion:

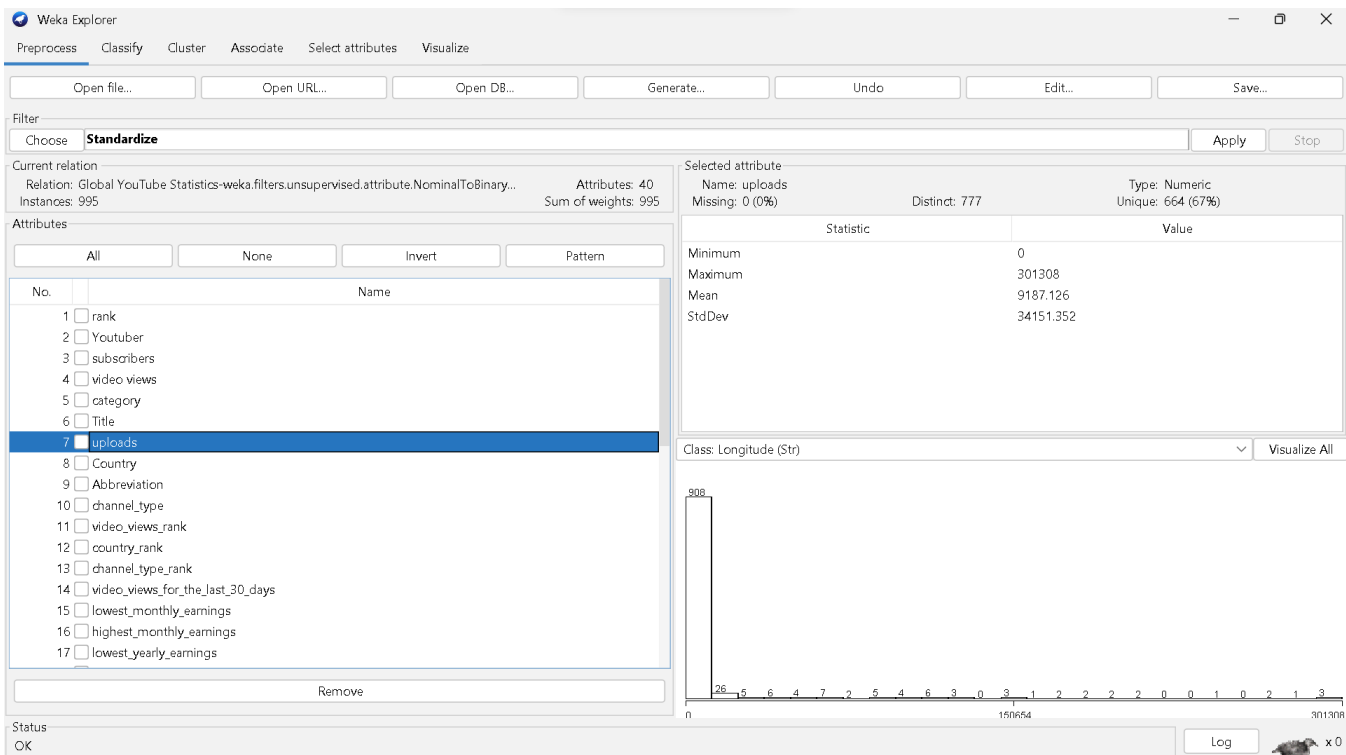


After conversion:



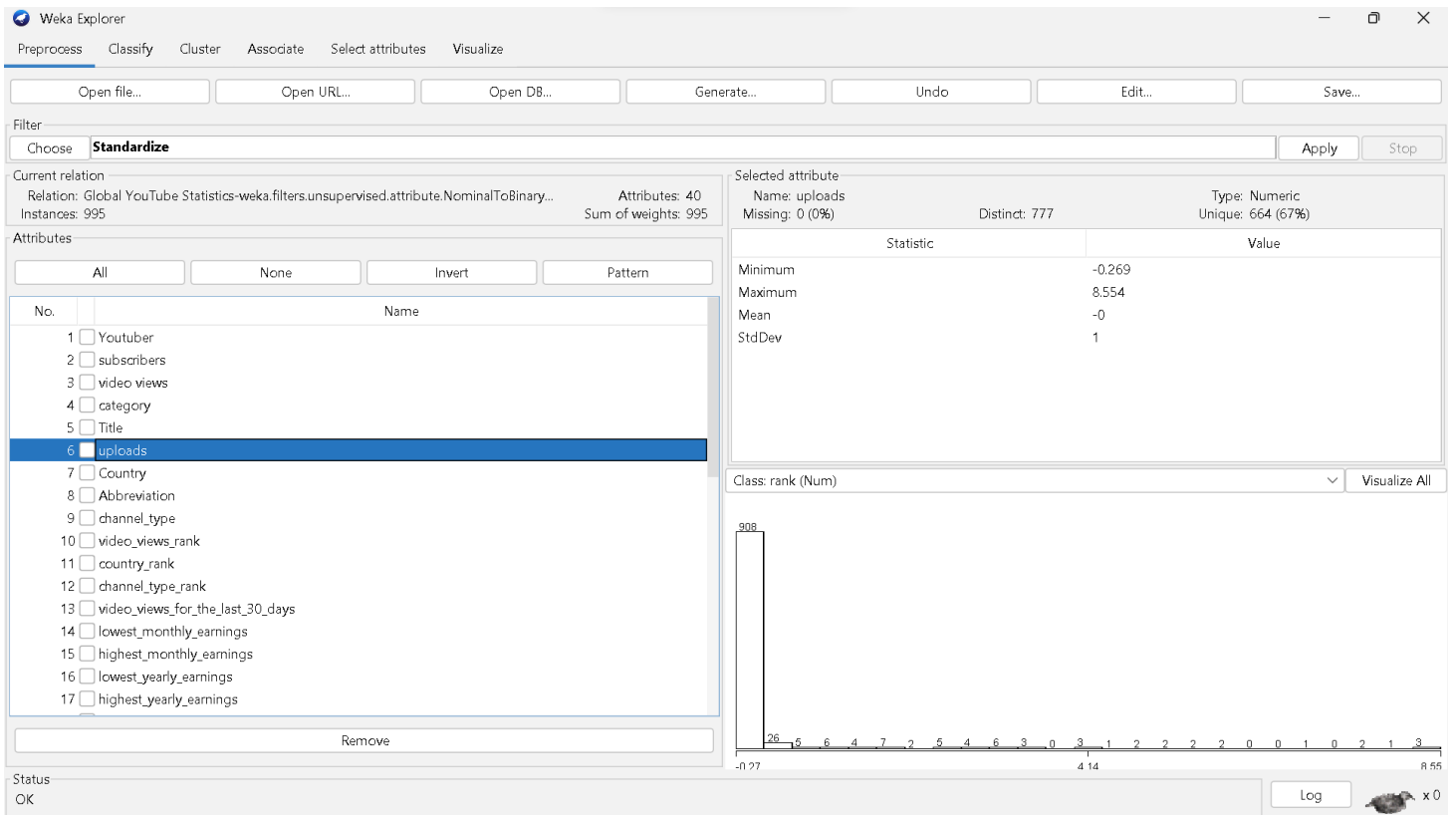
As visible in the screenshot above, new binary attributes have been created to represent the created_month of the channel after conversion using the **unsupervised.attribute.NominalToBinary** filter.

2. Standardize Numeric Attributes

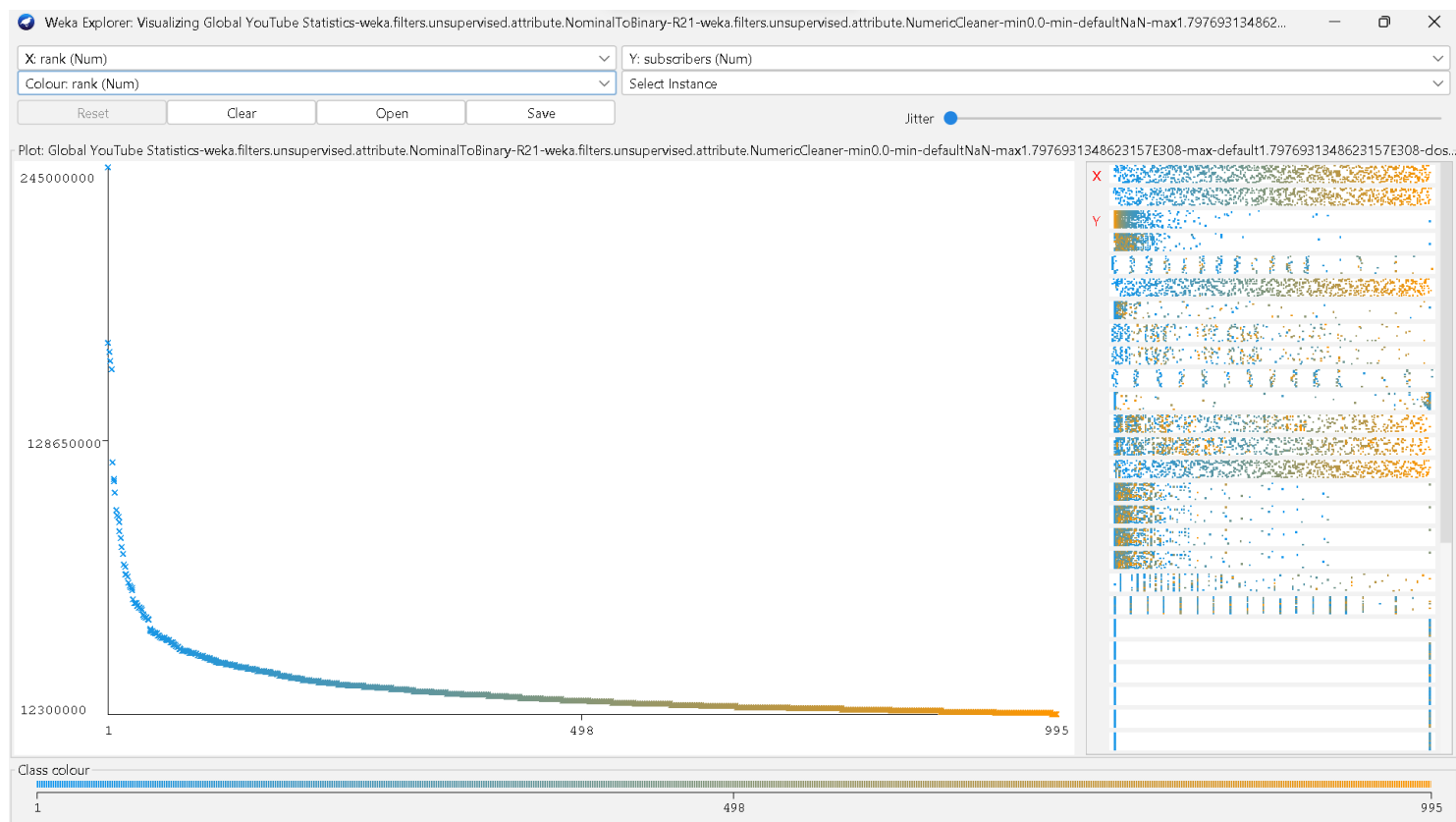


We use the **unsupervised.attribute.standardize** filter to standardize the attribute values of all Numeric attributes in the dataset. Above screenshot shows the stats of the attribute *uploads* before standardization.

After standardization, the *uploads* attribute has a mean of 0 and standard deviation of 1 as shown in the below screenshot.



Data Visualization



The above screenshot shows the plot of the ***subscribers*** attribute against rank of the YouTube channel. A clear relationship exists between the number of subscribers and the ranking of the YouTube channel.