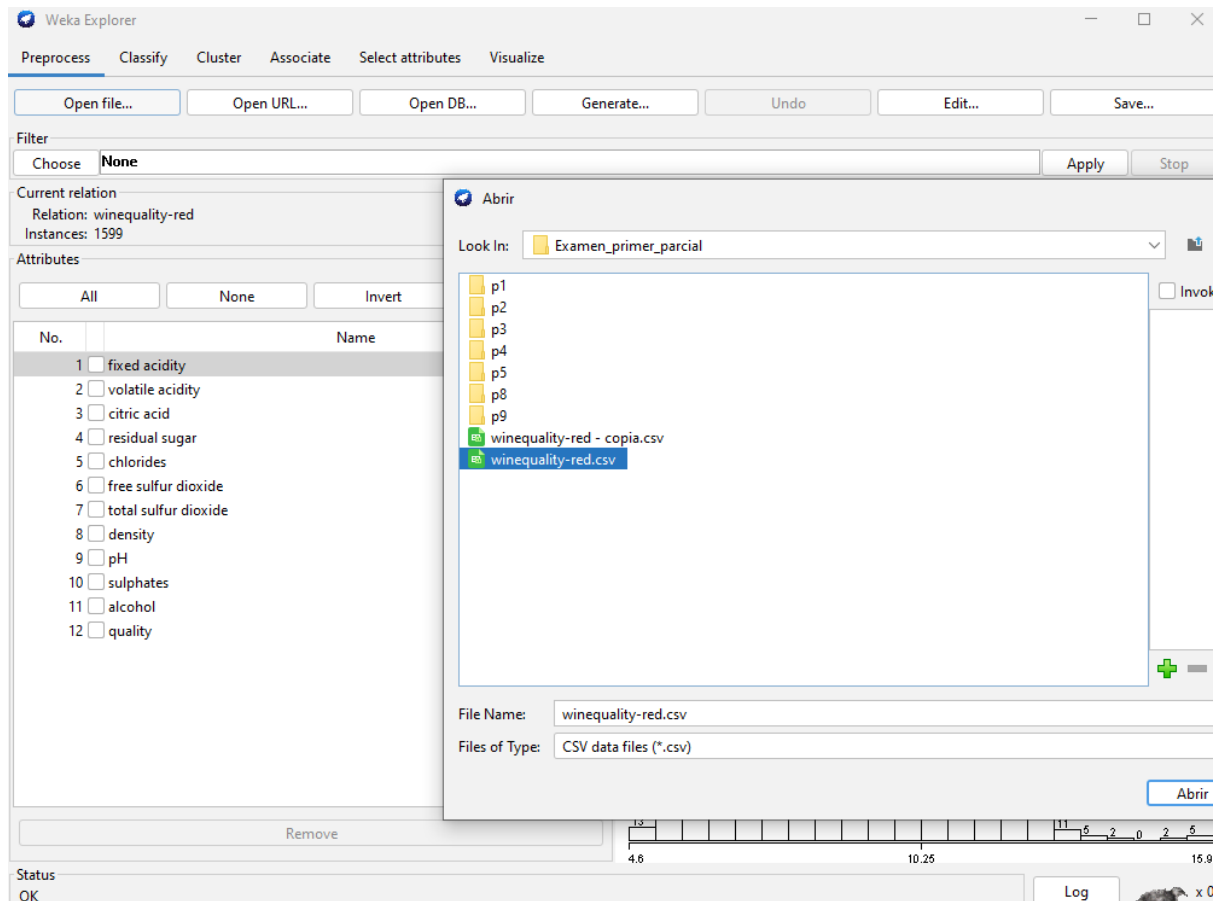


Univ. Diego Joel Condori Quispe  
12926609 L.P.

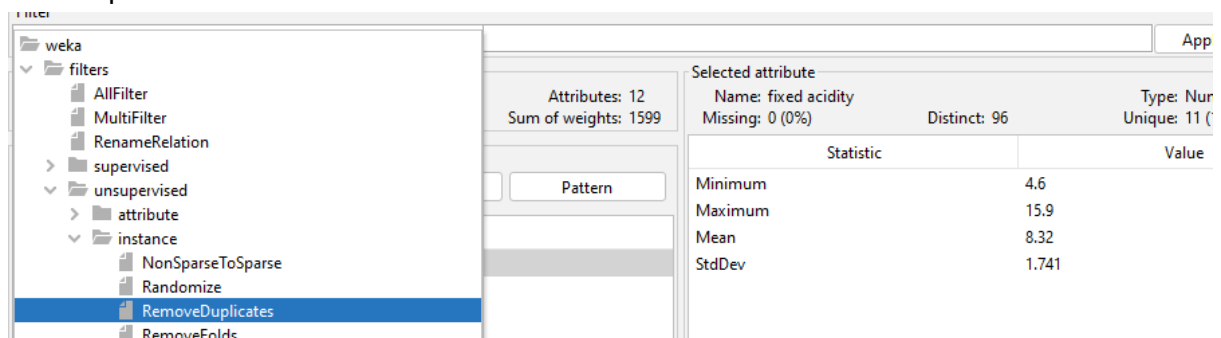
### 1. Preprocesamiento 1: Limpieza de datos para RemoveDuplicates()

Motivo para usar este preprocesamiento: Nos ayudará a evitar la redundancia en cada columna para cada variable categórica y mejorando la calidad de los datos.

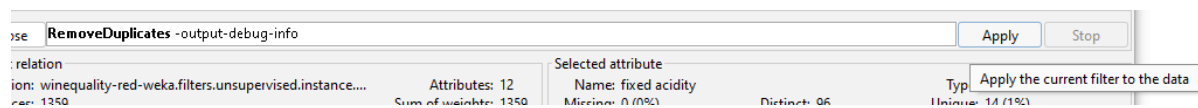
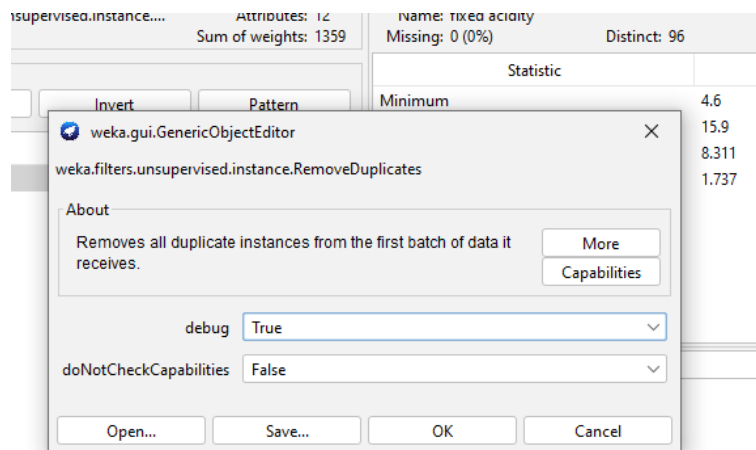
Abrimos WEKA, seleccionamos la opción explorer, luego open file para buscar nuestro archivo en la cual elegimos formato con extensión .csv:



Aplicaremos un filtro de preprocesamiento: RemoveDuplicates, para eliminar registros que sean duplicados en nuestro dataset.

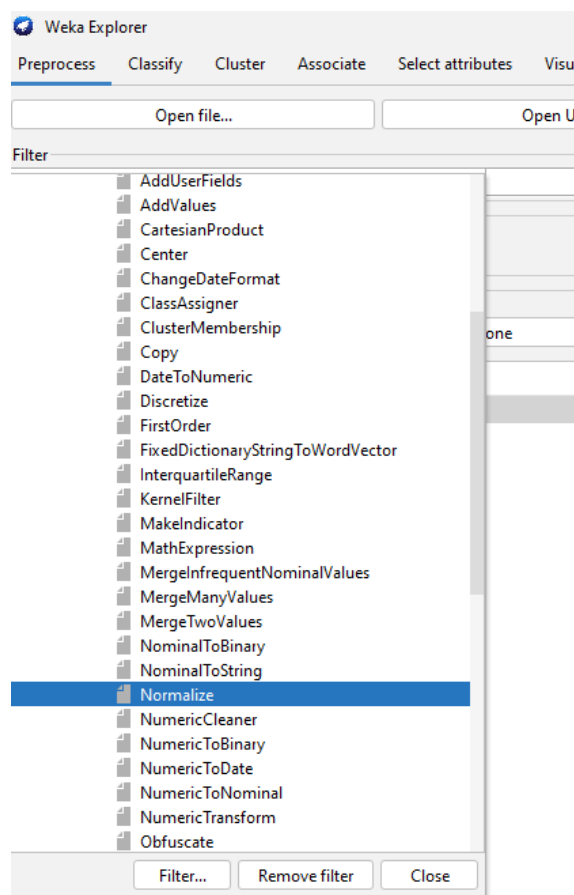


Para tener mas detalles le damos Debug en True:

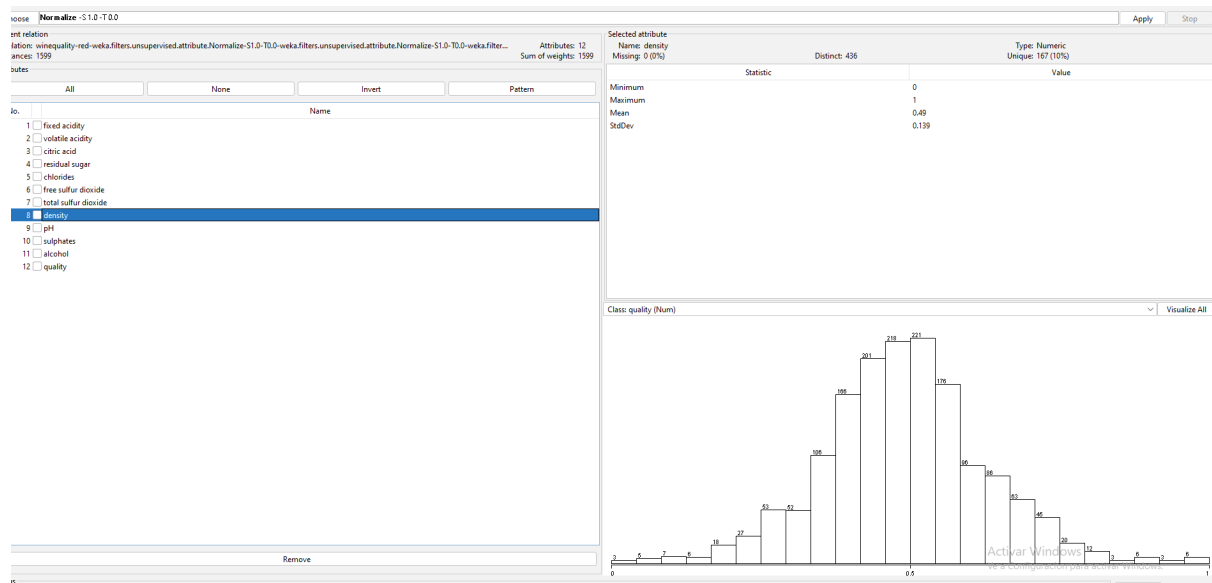


## 2. Preprocesamiento 2: Normalize

Nuevamente se elegirá un muestreo, nos vamos a filtros, unsupervised luego a attribute y ahi estara Normalize.

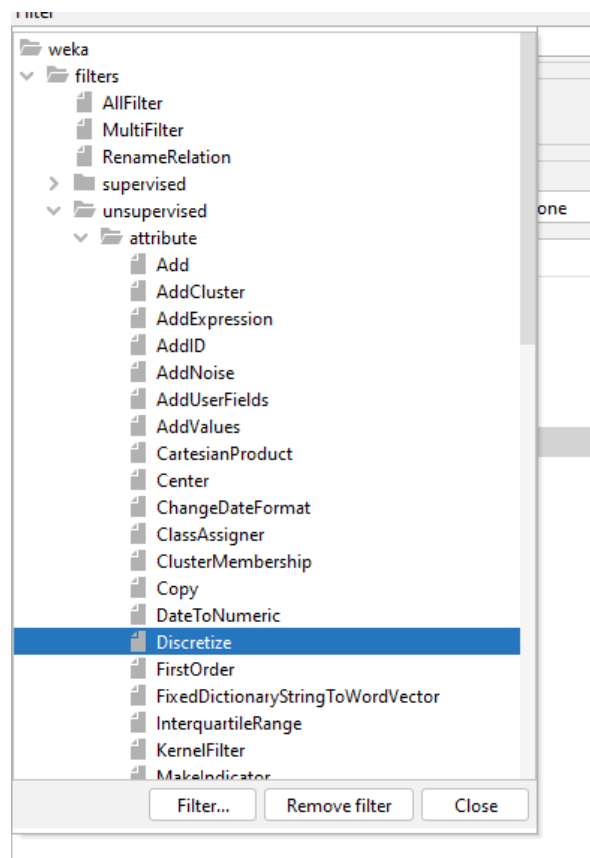


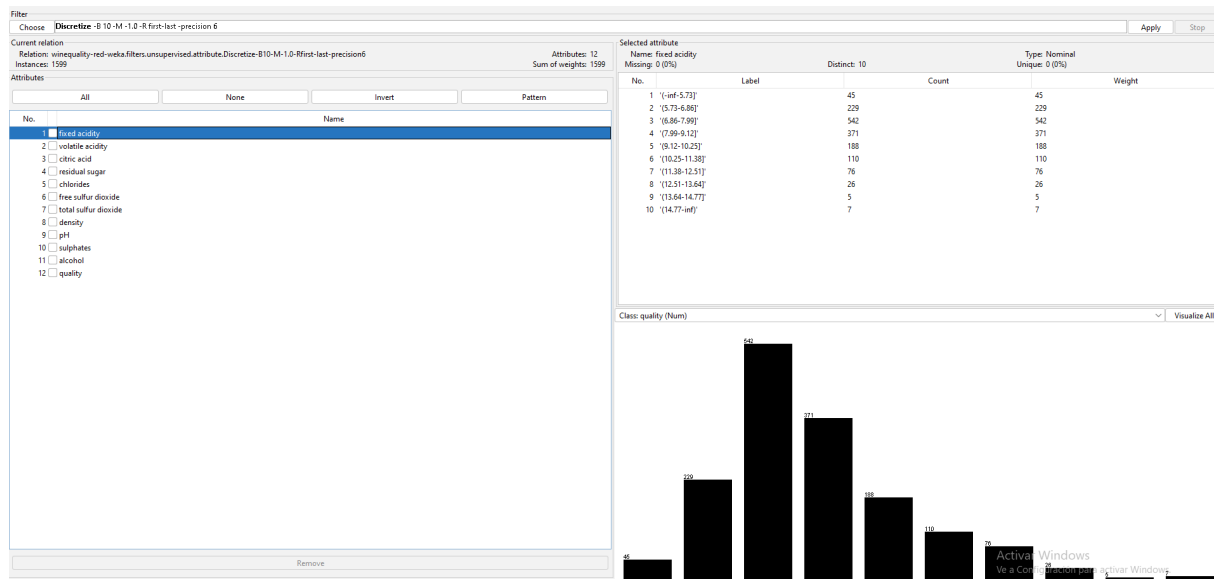
La normalización de los datos en su conjunto de calidad de vinos puede ayudar a mejorar la interpretación de los modelos, la comparabilidad de características y la eficacia de los algoritmos, lo que potencialmente puede mejorar su capacidad para determinar la calidad de los vinos de manera más precisa.



### 3. Preprocesamiento 3: Discretize

Nuevamente se elegirá un muestreo, nos vamos a filtros, unsupervised luego a attribute y ahi estara Discretize.

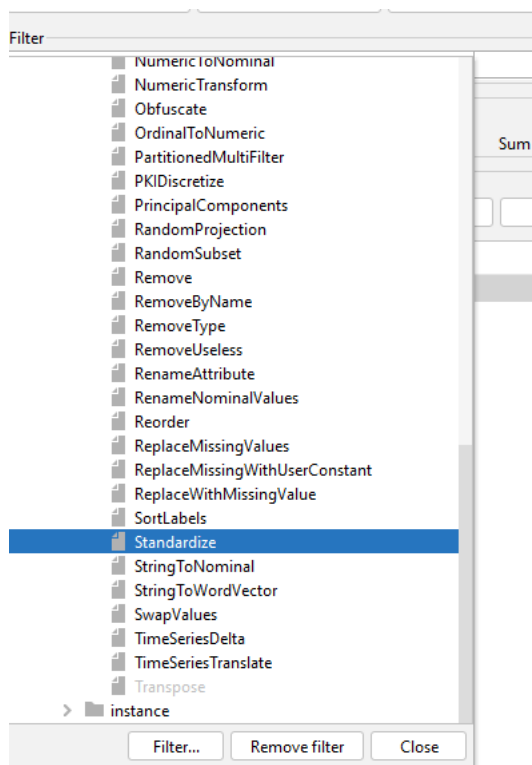


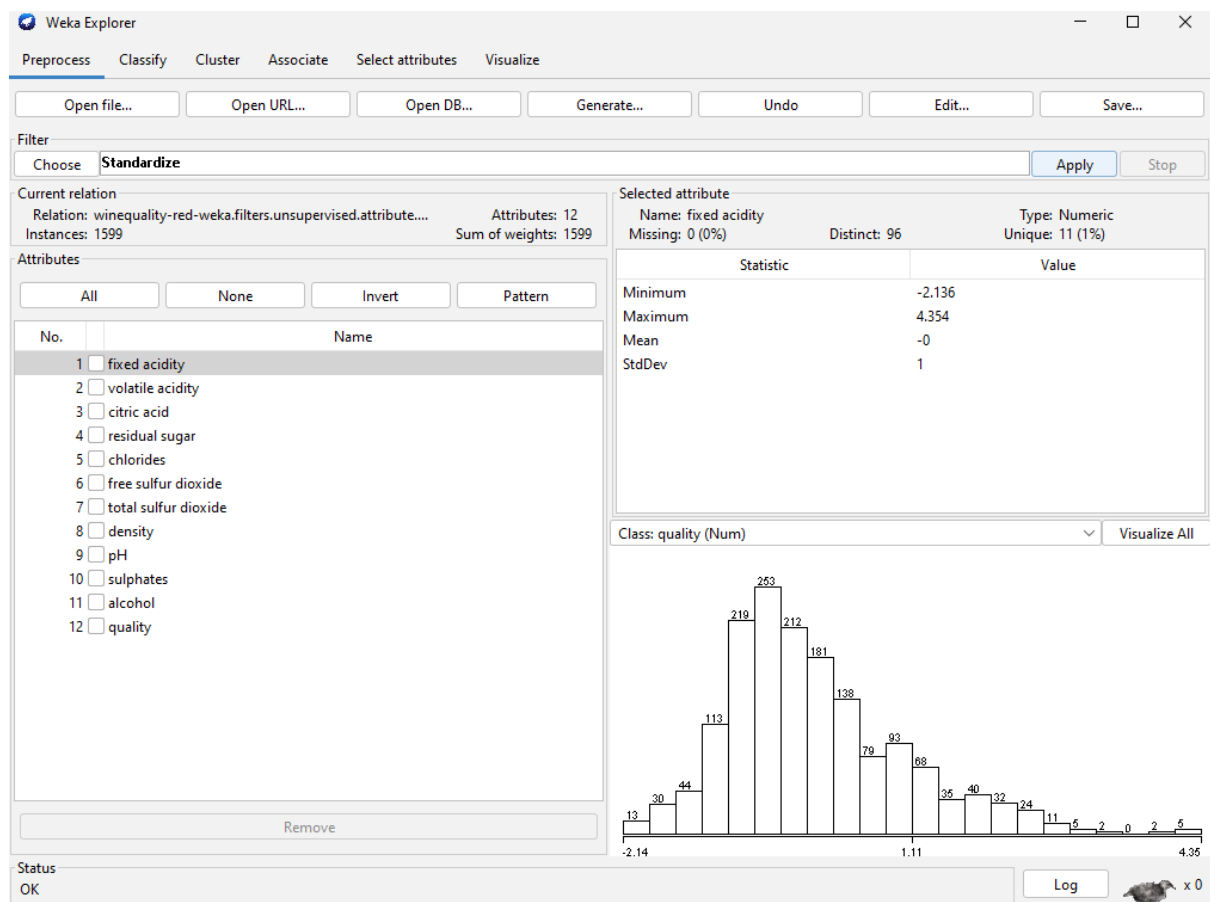


La discretización le permite analizar estas variables numéricas en términos de categorías, lo que puede facilitar la interpretación de cómo influyen en la calidad del vino. Sin embargo, es importante recordar que la discretización puede ser una simplificación y puede perder cierta información contenida en los valores numéricos originales.

#### 4. Preprocesamiento 4: Standarize

Nuevamente se elegirá un muestreo, nos vamos a filtros, unsupervised luego a attribute y ahi estara Standarize





El filtro Standardize en Weka se utiliza para estandarizar características numéricas en su conjunto de datos, ajustándolas para que tengan una media de 0 y una desviación estándar de 1. Esto facilita la comparación de características, mejora el rendimiento de algunos modelos de aprendizaje automático y facilita la interpretación en modelos lineales. Sin embargo, no es necesario para todos los algoritmos y no se aplica a la variable objetivo. Puede aplicarlo a las características numéricas para mejorar la comparabilidad y el análisis de su conjunto de datos de calidad de vinos rojos.