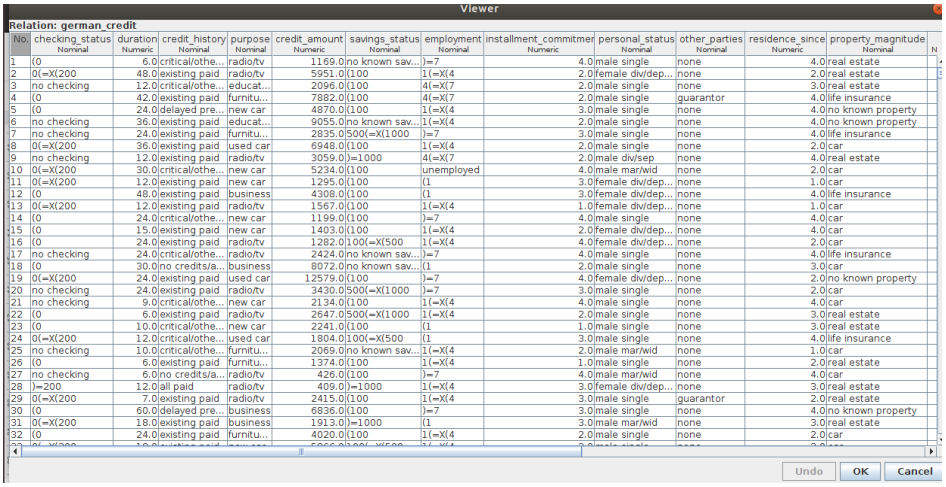
Estudiante: Sebastian Fabian Montes Mujica
CI 5977223

PRIMER EXAMEN PARCIAL

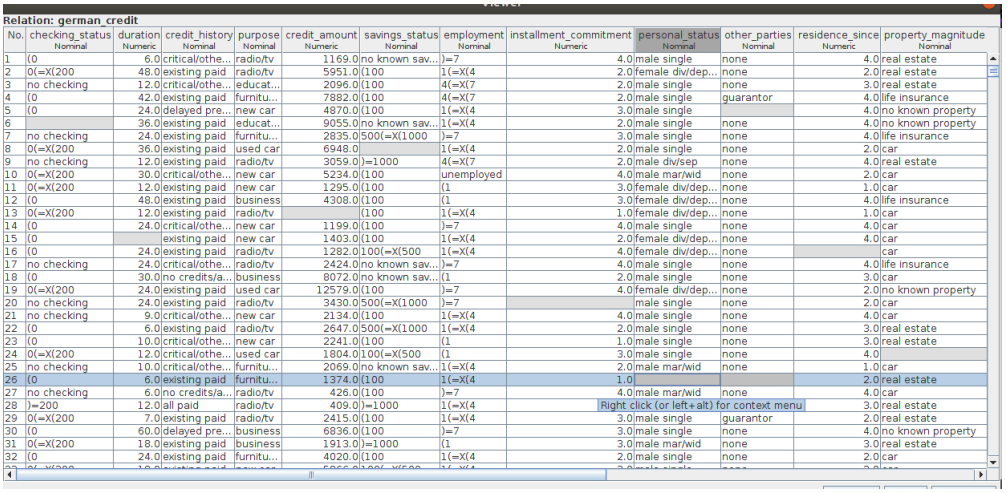
2. De un dataset de su tarea anterior en WEKA, realice tres algoritmos de preprocesamiento.

I. En caso de existir datos (atributos) faltantes en el dataset (instancia) durante el trabajo de preprocesamiento se realiza el llenado de estos datos mediante la función **ReplaceMissingValues** que funciona para atributos numéricos y nominales. Para la inserción de datos realiza el cálculo de la moda y la media en el dataset.

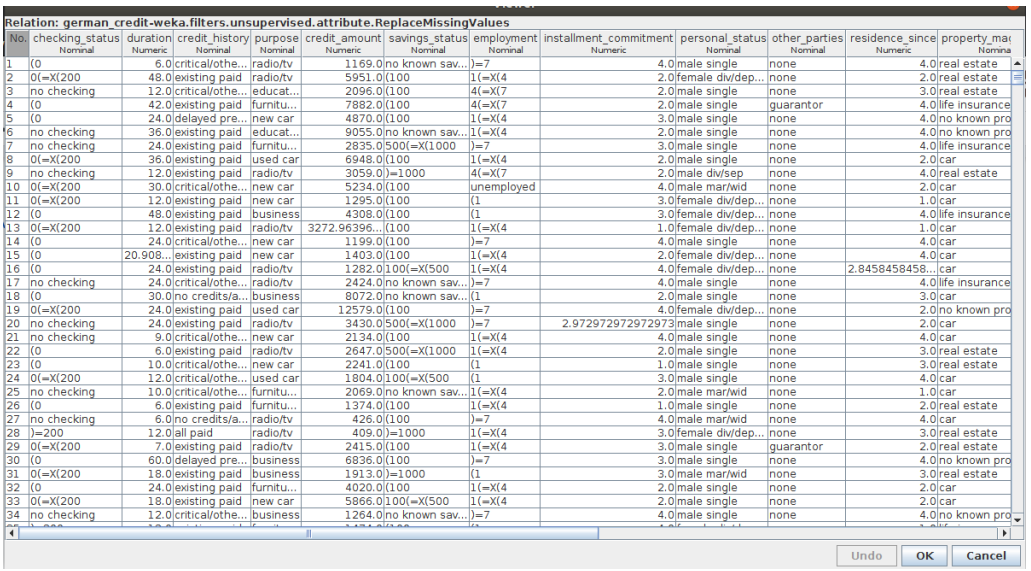
ejm: Dataset original (no contiene espacios vacíos)



Dataset editado con espacios vacíos



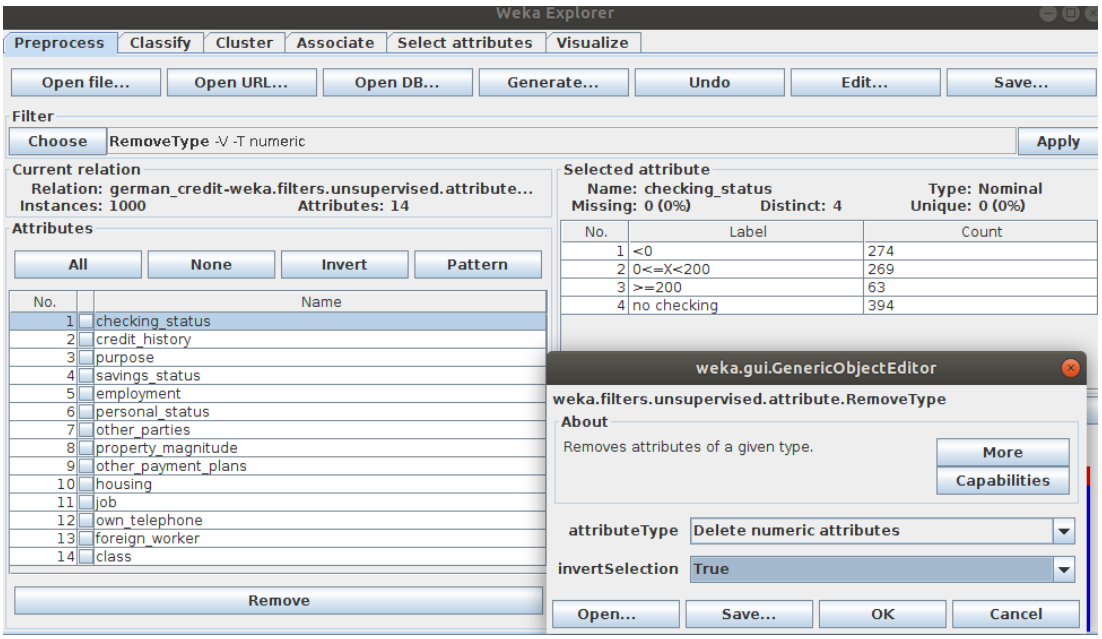
Dataset rellenado utilizando la herramienta de preprocesamiento



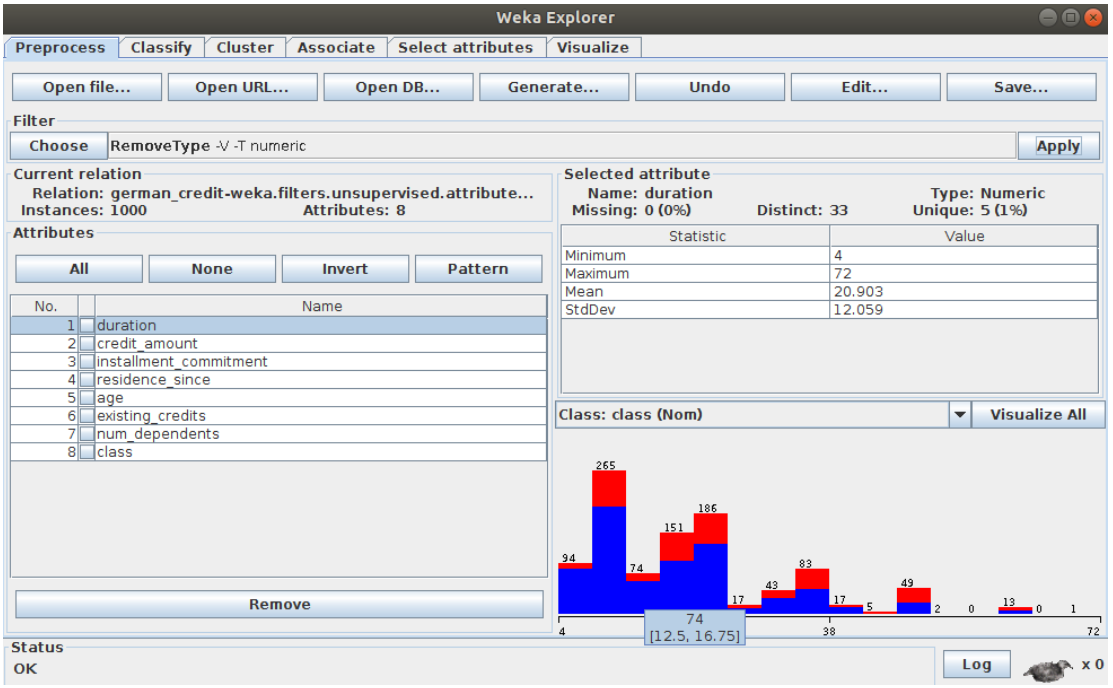
Esta operación afecta estadísticamente al resultado debido a que sustituye el elemento faltante con los valores de la media.

II. Si es necesario visualizar únicamente los atributos de tipo numérico o de tipo cadena para la comprensión más precisa de lo que se busca sin el ruido de los elementos no necesitados podemos utilizar la opción de atributo Remove Type que elimina los atributos no necesitados.

Dataset sin preprocesamiento

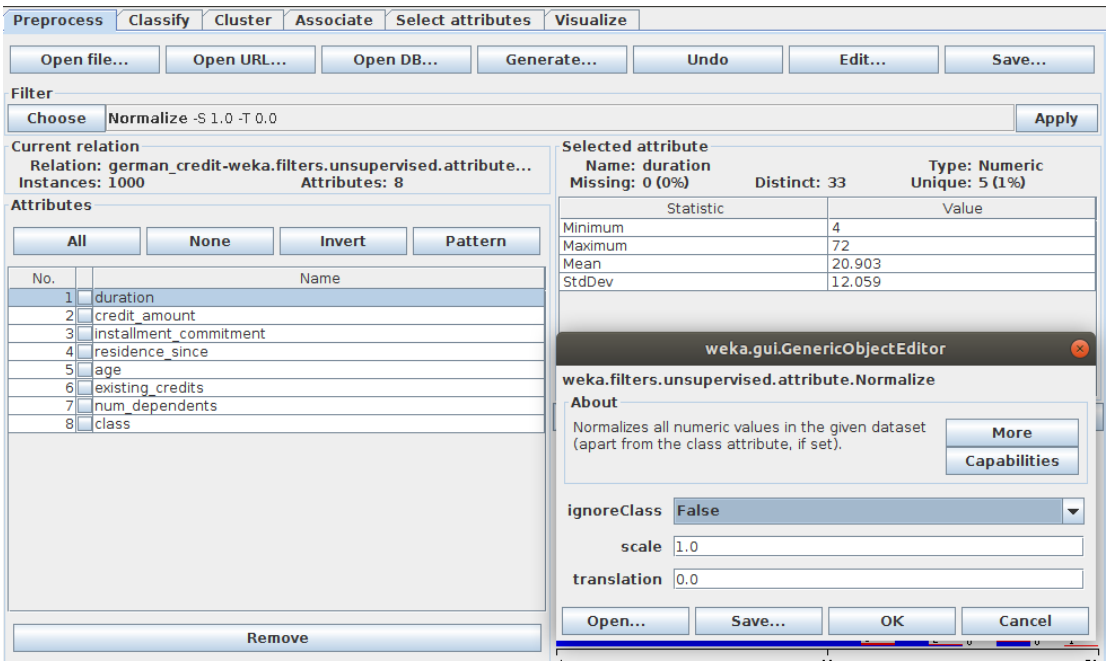


Dataset con preprocesamiento

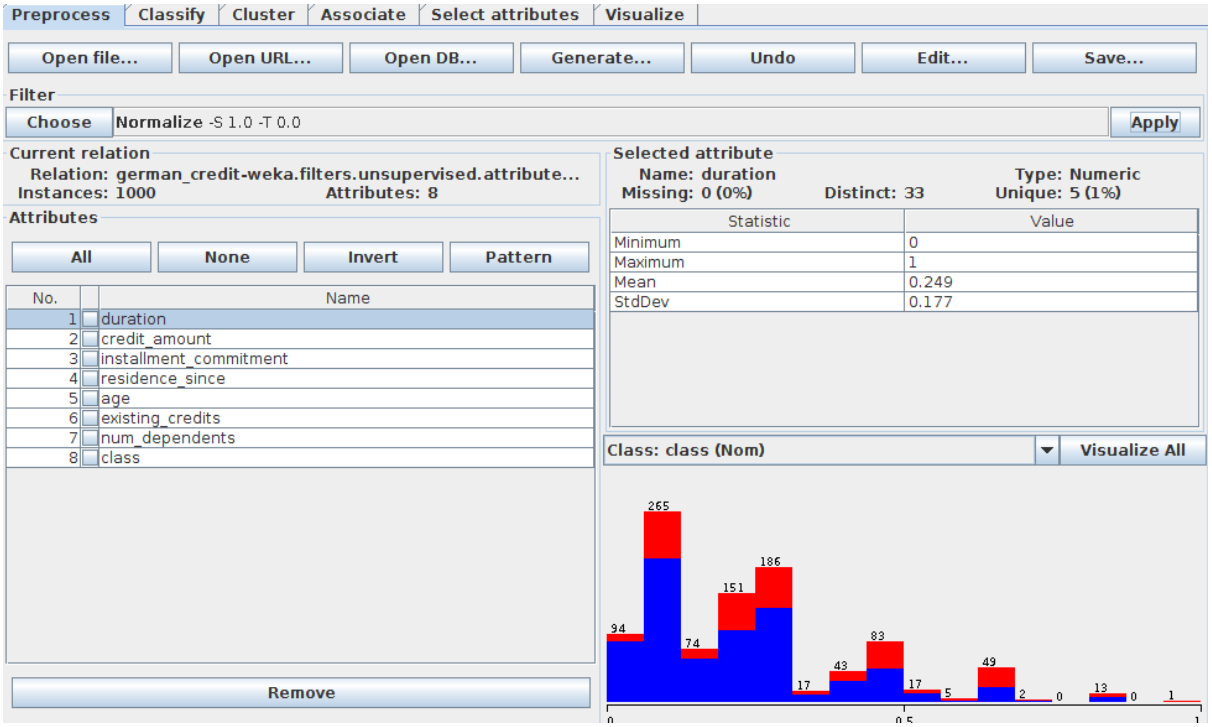


III. Con los datos numéricos obtenidos podemos realizar la normalización de los mismos. Para ello utilizaremos el preprocesamiento de atributos Normalize, que normaliza todos los valores numéricos del dataset, viene por defecto con los valores [0,1] de intervalo pero este puede ser modificado mediante la escala a otro intervalo por ejemplo 2 que sería igual al intervalo [-1,1]

Dataset sin preprocesamiento

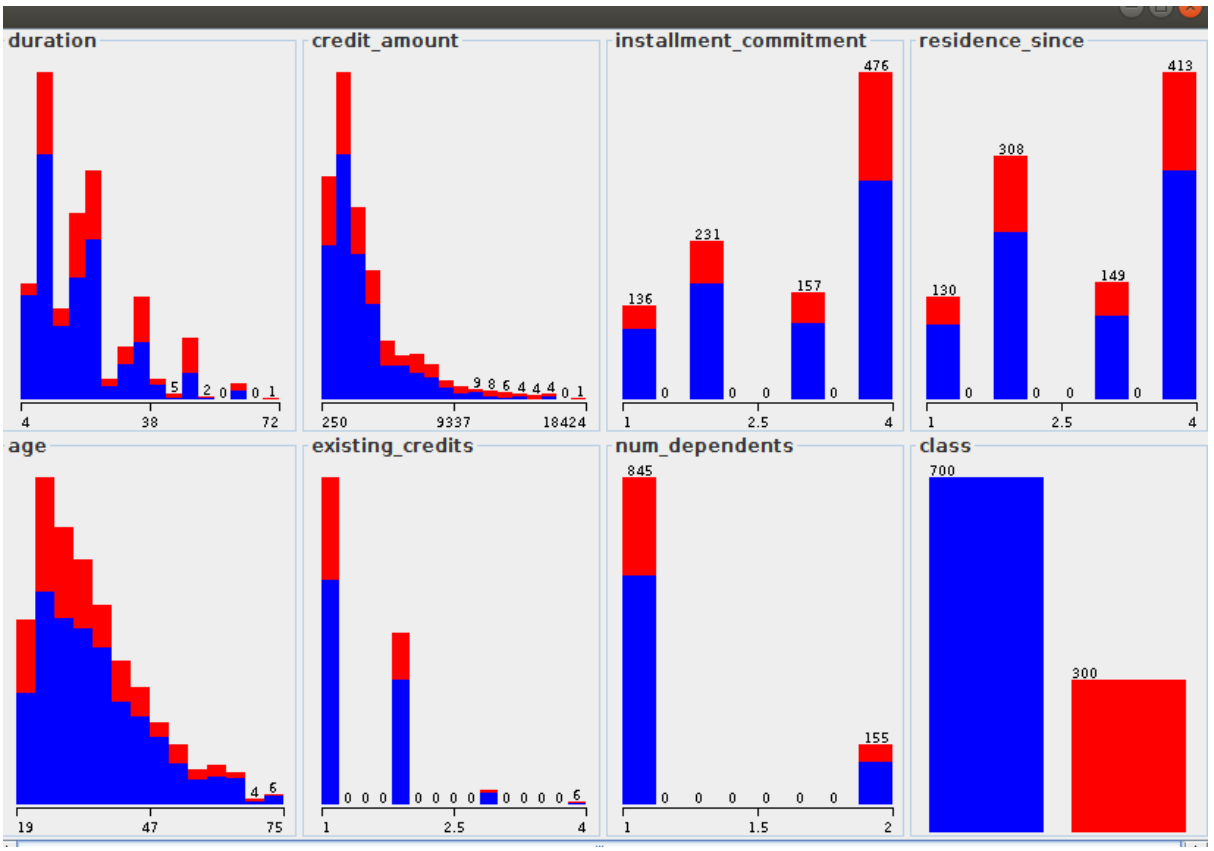


Dataset con preprocesamiento



Notemos que los datos datos de Mínimo y Máximo variaron debido a la normalización de Min=4 y Max=72 a 0,1 como se configuró previo a la normalización de la misma manera los datos de la media y la desviación standart se modificaron.

Resultados sin el preprocesmiento



Resultados obtenidos por el preprocesamiento