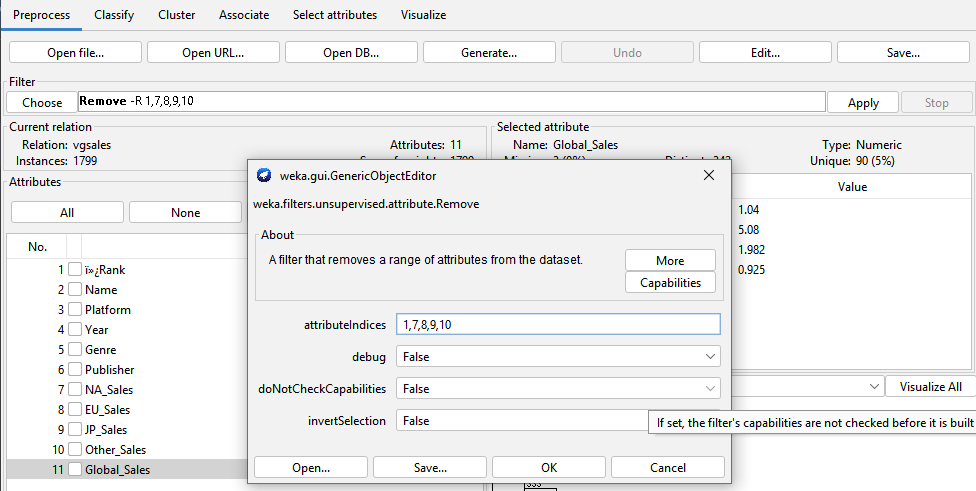
**EJERCICIO 5. WEKA**

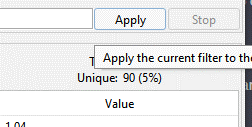
**1. Remoción de Atributos.**

Es una técnica de preprocesamiento que nos permite eliminar ciertas columnas del dataset, aquellas que no pueden resultar muy útiles para el análisis posterior.

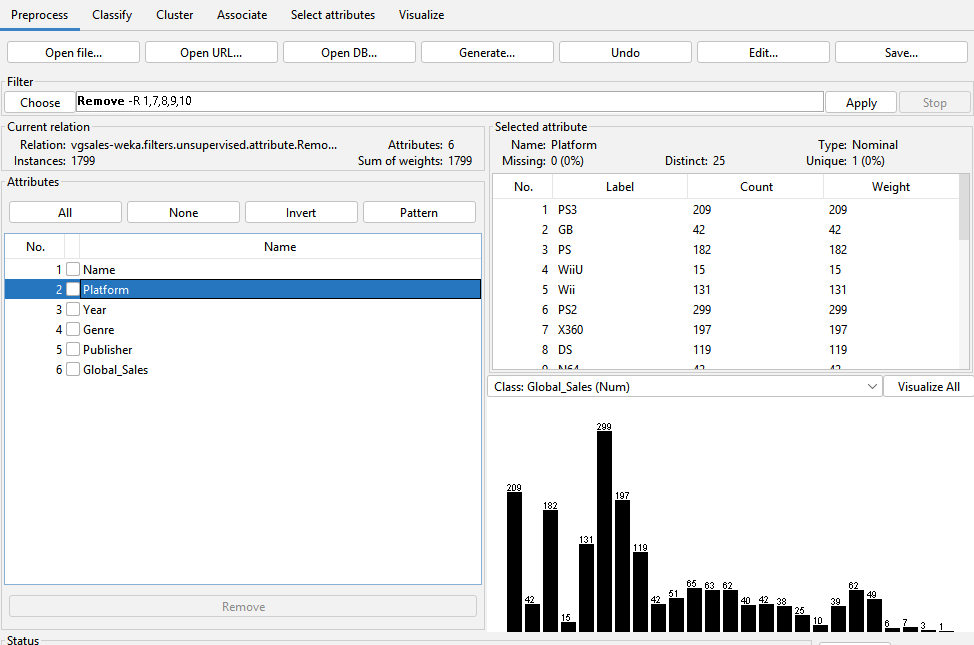
1. En Weka, una vez abierto el data set, seleccionamos el Filtro **Remove**, a continuación, hacemos click y en la ventana de dialogo escribimos las columnas (atributos) que deseamos eliminar.



1. Luego hacemos click en Apply, para ejecutar la acción.

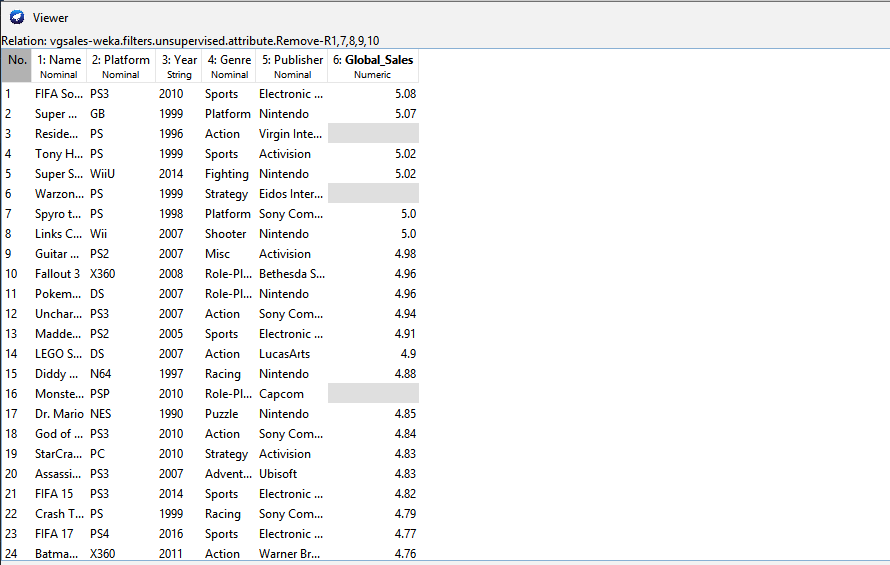


1. Luego se observa que se han eliminado las columnas seleccionadas y nos quedamos con aquellas de mayor interés. En este caso eliminamos las ventas por regiones (EU\_Sales, NA\_Sales..), ya que Global\_Sales cuenta con el total de estas ventas.

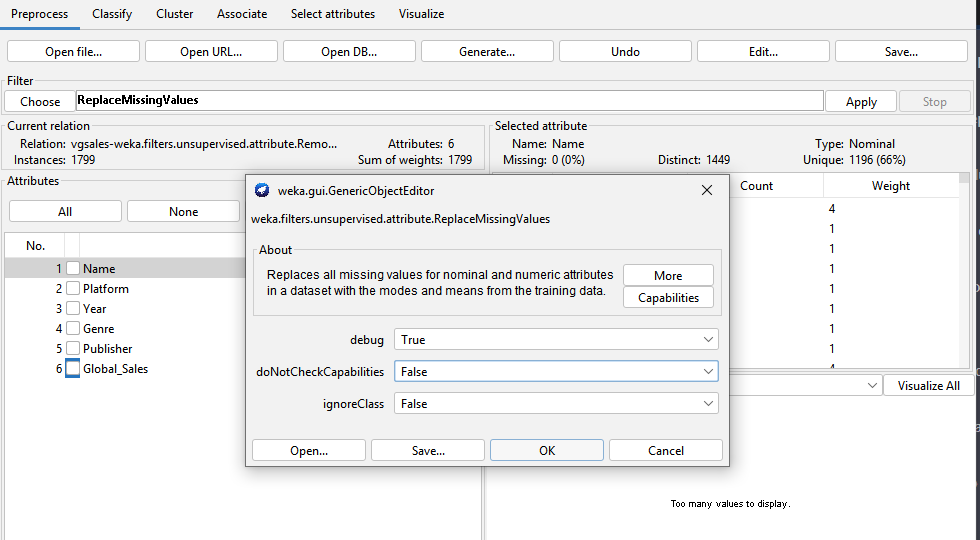


**2. Manejo de Datos Faltantes**

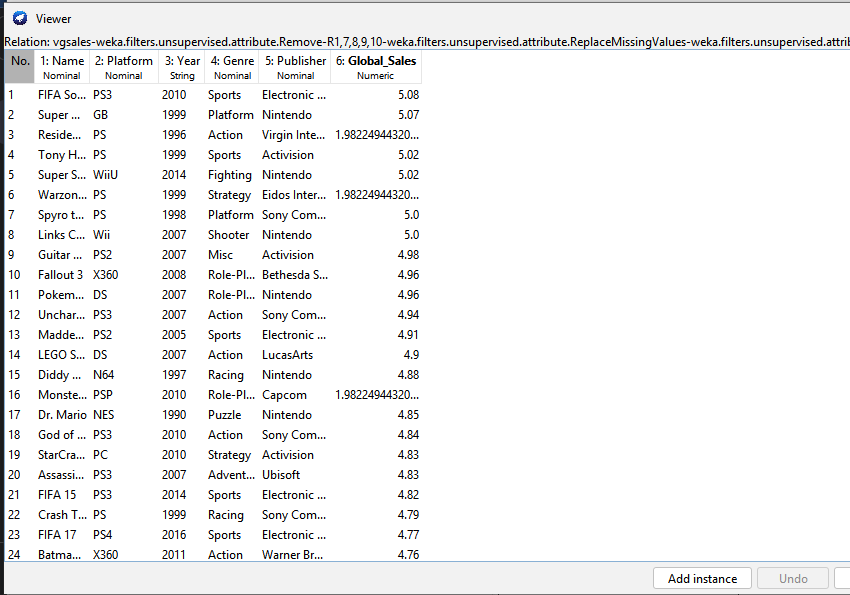
1. Si observamos los datos del dataset en Weka, podemos apreciar que existen datos vacíos o faltantes en la columna **Global\_Sales (ventas globales).** Entonces se va a rellenar estos datos para evitar cierta perdida de información. Lo ideal es rellenarlos con la media de los datos de esa columna.



1. Para rellenar estos campos vacíos, seleccionamos el filtro **ReplaceMissingValues** en la ventana de dialogo marcamos todo en True.



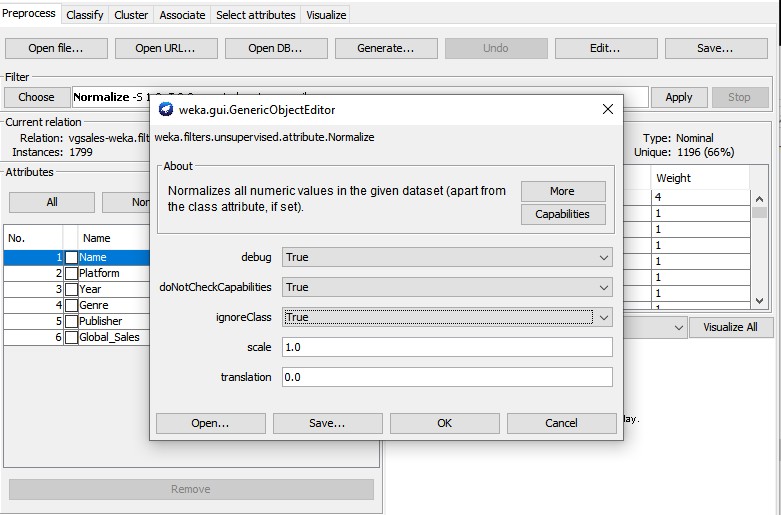
1. Una vez hecho click en **Apply**, observamos nuevamente los datos, y se observa que los datos que eran vacíos, se rellenaron con la media de esa columna.



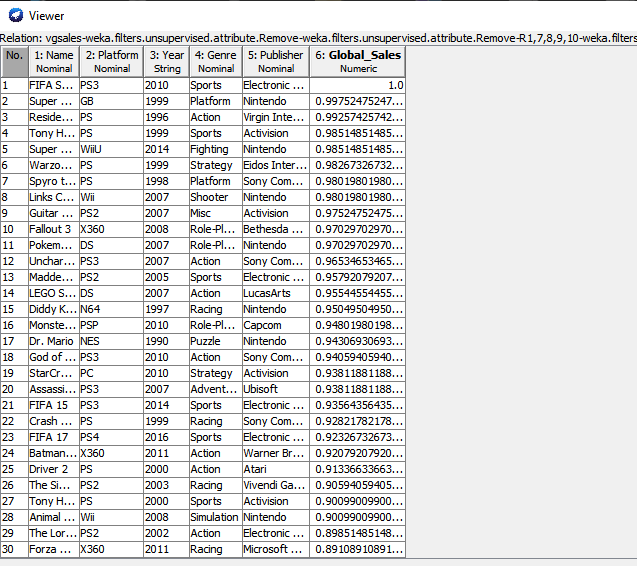
**3. Normalización y Escalado de datos.**

Es una técnica importante para asegurarse de que todos los datos tengan el mismo impacto en los algoritmos de aprendizaje automático. Una escala universal es de 0 a 1 o de -1 a 1.

1. En Weka, seleccionamos el filtro **Normalice**, en el cuadro de opciones colocamos todo en True y el resto lo dejamos por defecto como está.



1. Luego de hacer click en **Apply**, se ejecuta el proceso, y si visualizamos los datos vemos que la columna Global\_Sales ha sido normalizada en la escala [0,1].



**4.** Una vez realizadas las técnicas de preprocesamiento, podemos guardar los cambios realizados, hacemos click en **Save**, seleccionamos la ruta y un nombre para el dataset, el formato utilizado normalmente por Weka es **(.arff)**.

