**INF 354**

**INTELIGENCIA ARTIFICIAL**

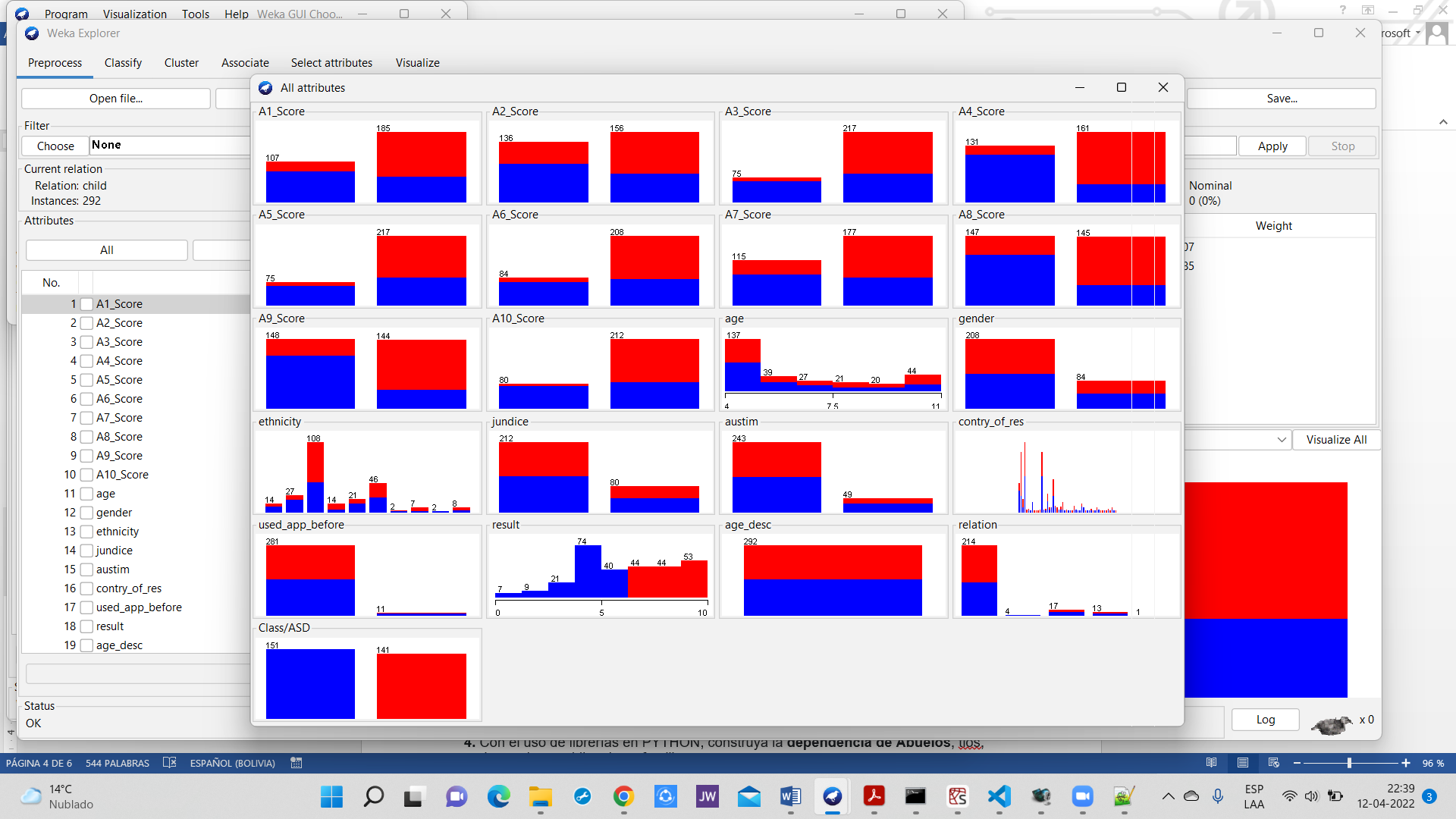
**GUTIERREZ MACHICADO GERMAN**

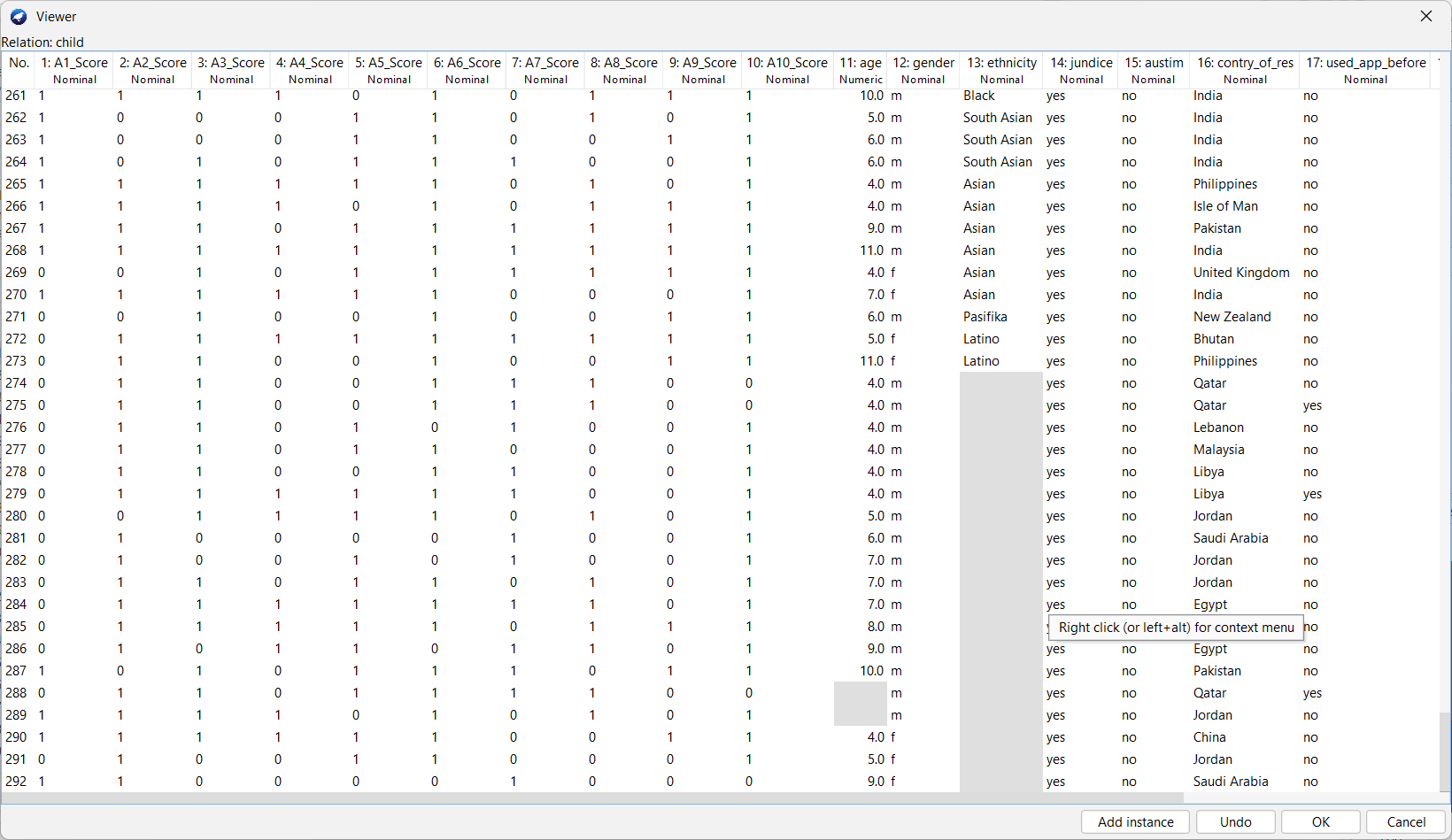
**CI: 6072590**

**12-04-2022**

**PUNTO 2.** Del dataset anterior realice en **WEKA, tres algoritmos de pre procesamiento**.

**1 primer algoritmo de pre procesamiento: ReplaceMissingValues**





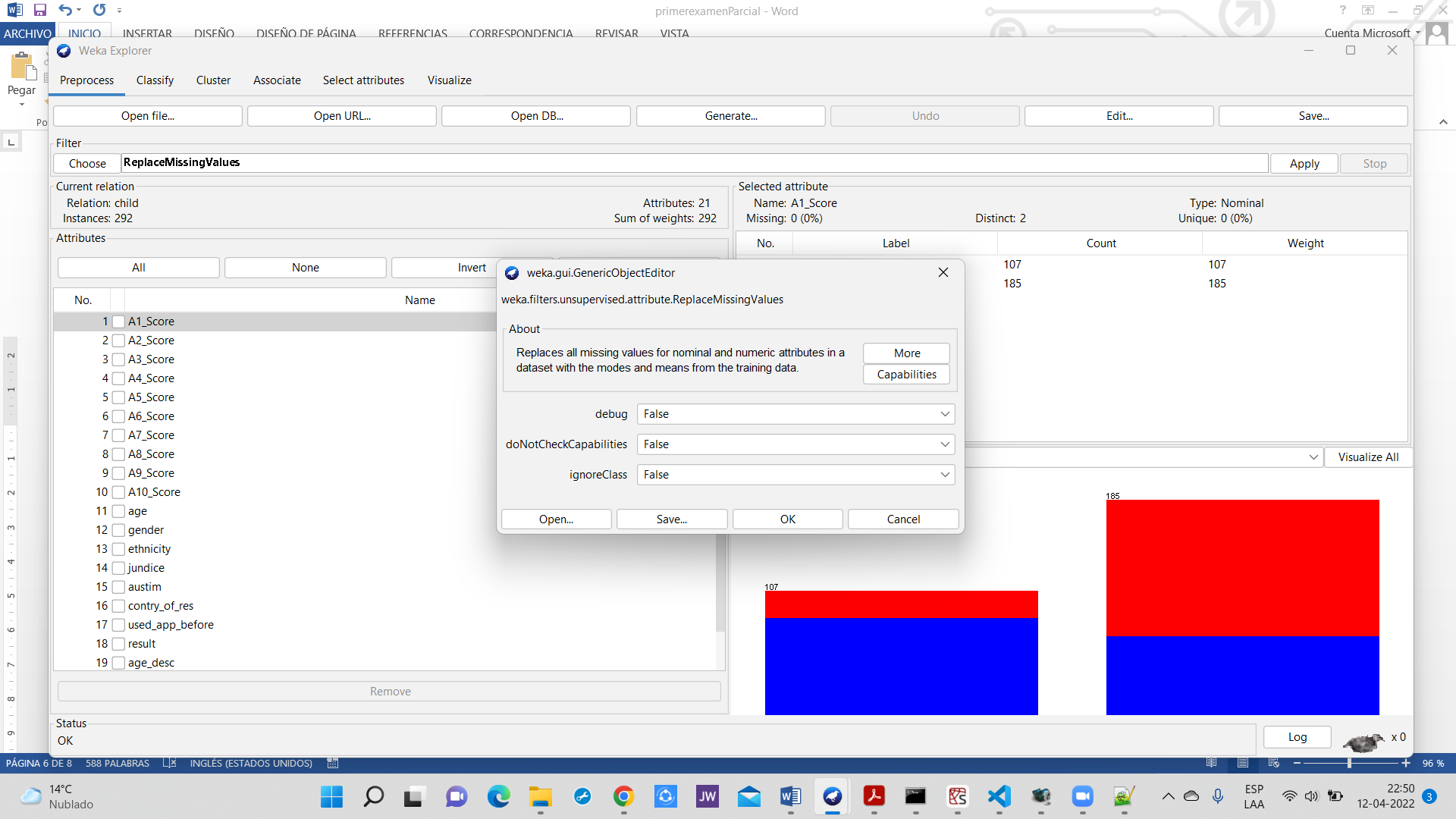
En la columna “age” se pueden observar 4 campos vacíos

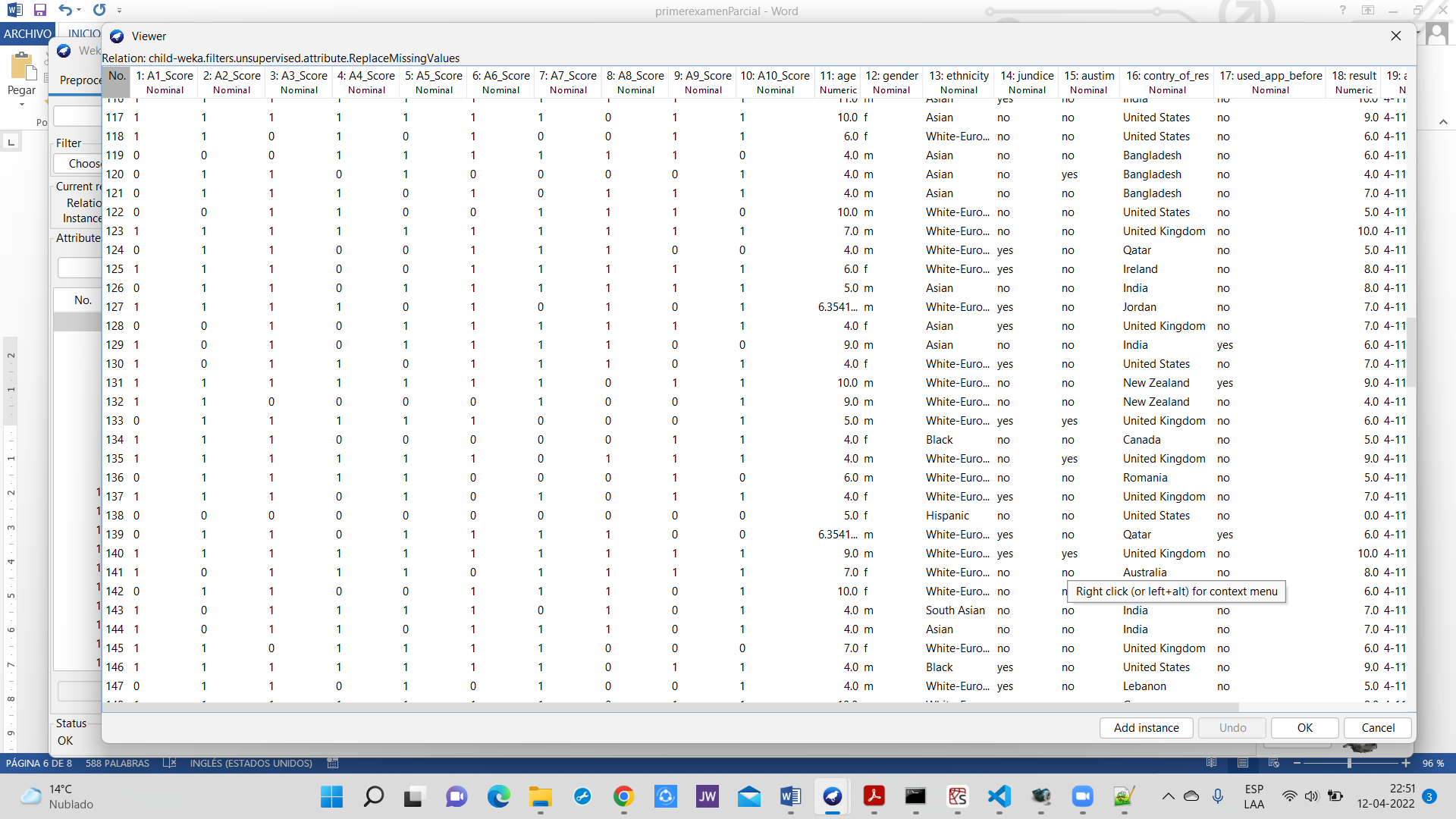
En la columna “ethnicity” se pueden observar 43 campos vacíos

En weka procedemos de la siguiente manera:

En la pestaña “**preprocessing**”

Filter > choose > filters > unsupervised > attribute > **replaceMissingValues** > apply





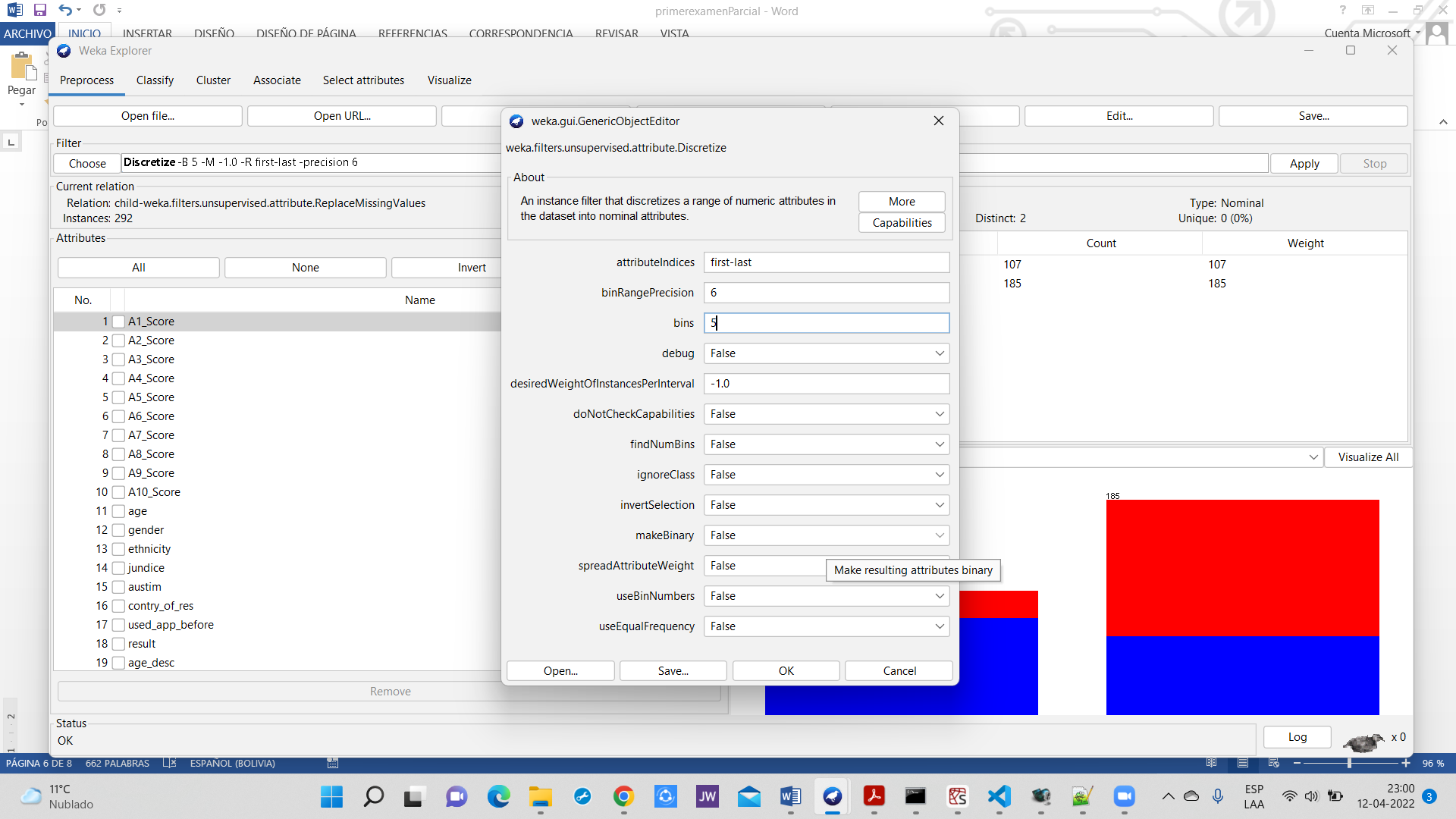
Podemos ver que todos los campos han sido rellenados, los 4 campos de “age” y los 43 campos de “ethnicity”

**2 segundo algoritmo de pre procesamiento: Discretización**

En la pestaña “**preprocessing**”

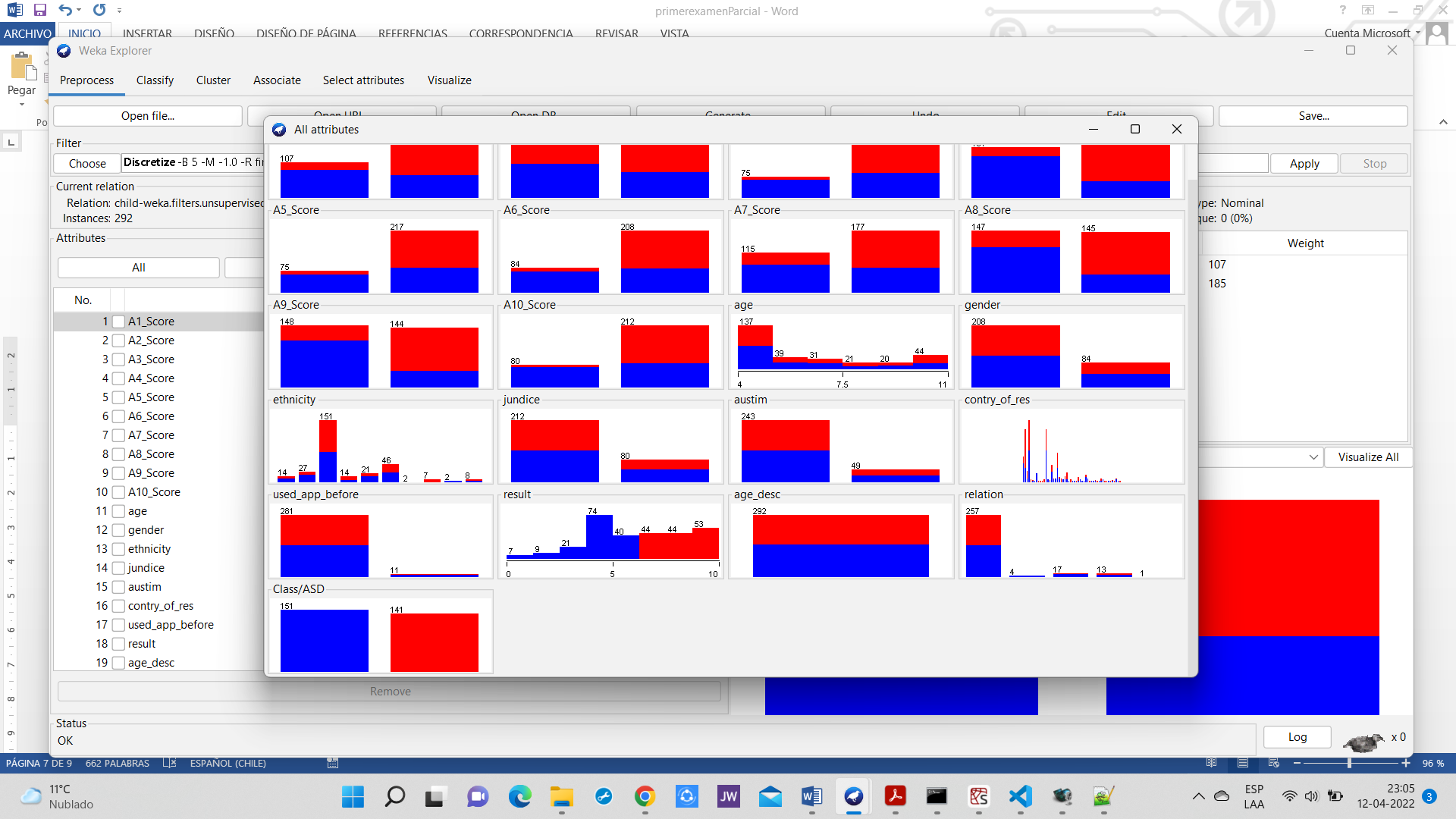
Filter > choose > filters > unsupervised > attribute > **Discretize** > apply

En “bins” colocamos el número de columnas iguales para todos. Dividimos los datos en grupos



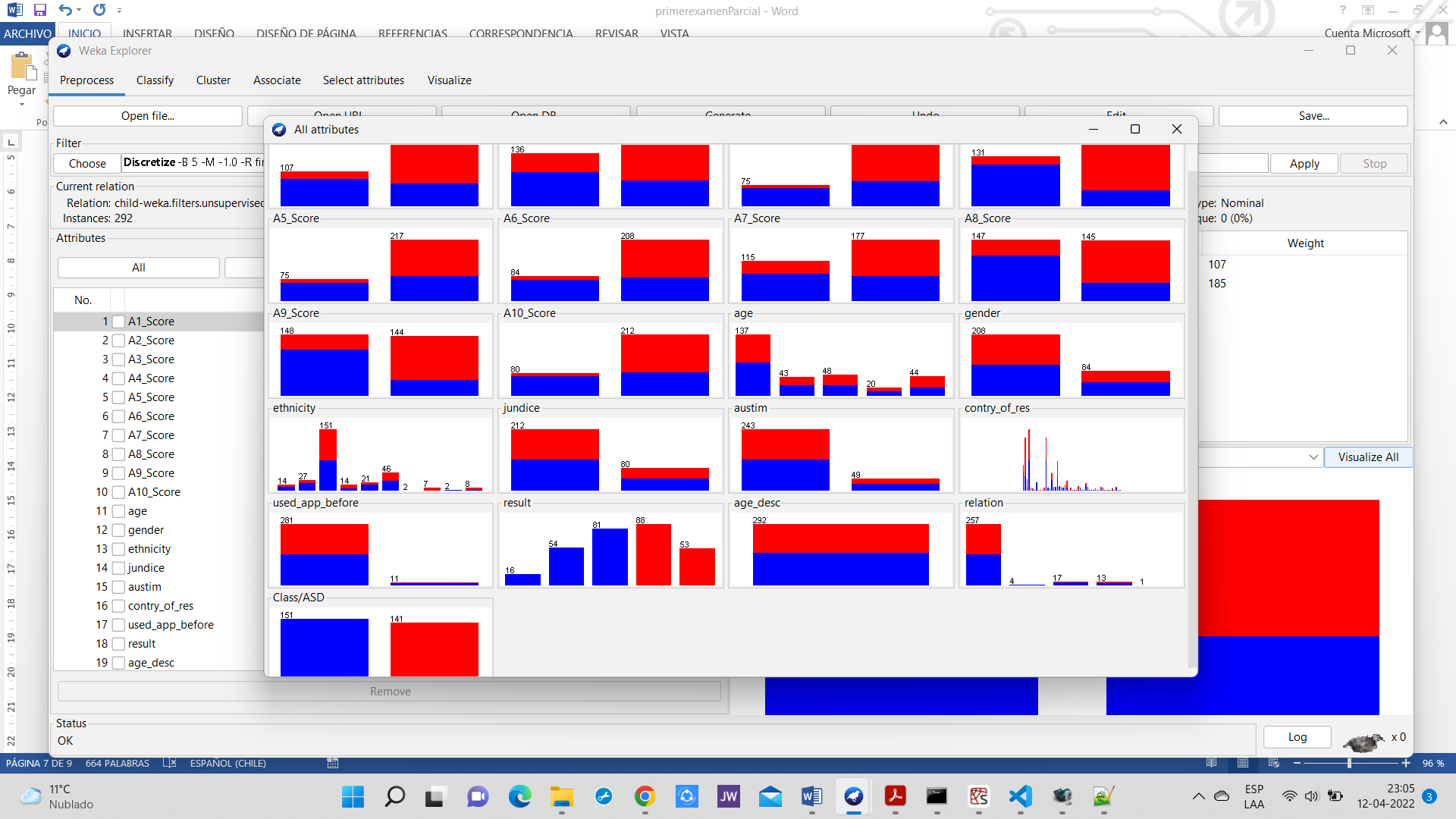
Si ponemos el “useEqualFrequency” en “true” se intentará igual las frecuencias. Es decir que se intentrá igualar la altura de las barras

Antes:



Después:

Con bins=5



Vemos que solamente las columnas con números fueron afectadas

**3 tercer algoritmo de pre procesamiento: Normalización**

En la pestaña “**preprocessing**”

Filter > choose > filters > unsupervised > attribute > **Normalize** > apply

La normalización de los datos es es el proceso de reescalar uno o más atributos a un rango entre 0 y 1.

La normalización es una buena técnica a utilizar cuando no se sabe la distribución de los datos o cuando sabes que la distribución es no Gaussiana (la curva campana)

