NOMBRE: SAMUEL ALVAREZ PANUNI

PRIMER PARCIAL INF354

1.- Seleccione una data set el cual debe ser tabular de al menos 1000 filas y 5 columnas.

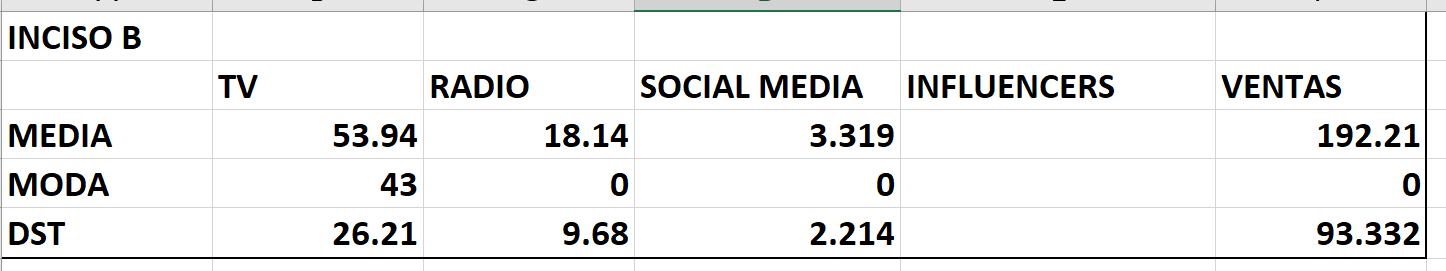
a.

b.- La media, la moda, la desviación estándar con el uso de numpy y pandas

Data Set de marketing y ventas descripción:

Hice estos datos para mis estudiantes en 'Marketing basado en datos' y 'Ciencia de datos para empresas'. Los datos contienen:

* Presupuesto de promoción televisiva (en millones)
* Presupuesto de promoción de redes sociales (en millones)
* Presupuesto de promoción radiofónica (en millones)
* Influencer: si la promoción colabora con Mega, Macro, Nano, Micro influencer
* Ventas (en millones)



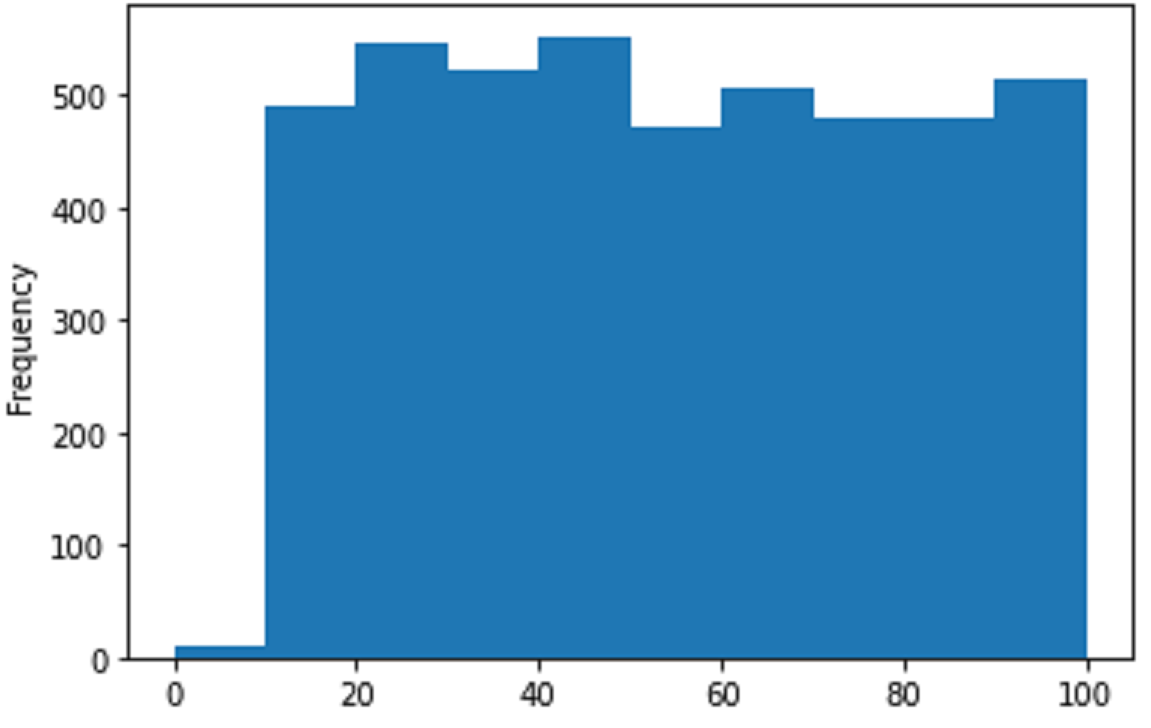
Como se puede ver en los datos la media nos indica el promedio del conjunto de valores dados

La moda nos indica el valor que más se repite dentro de los datos

La desviación estándar nos indica un índice numérico de la dispersión del conjunto de datos.

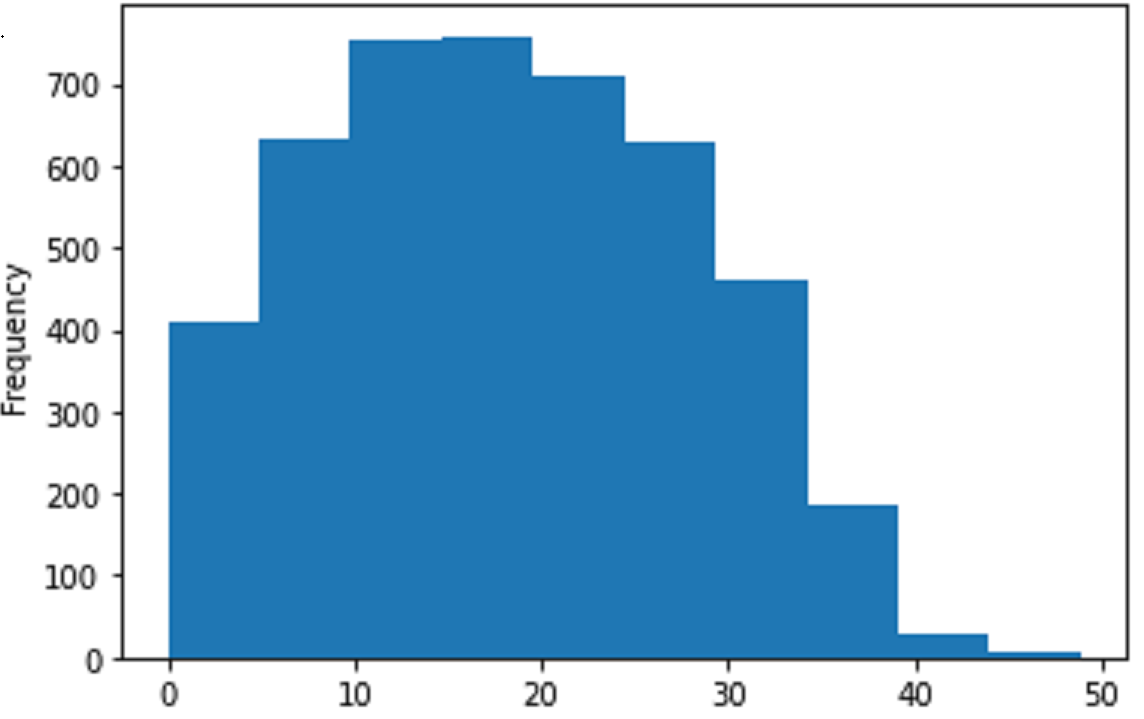
C.- GRAFIQUE LOS DATOS Y EXPLIQUE SU COMPORTAMIENTO (PHYTON)

Grafico del presupuesto en tv.



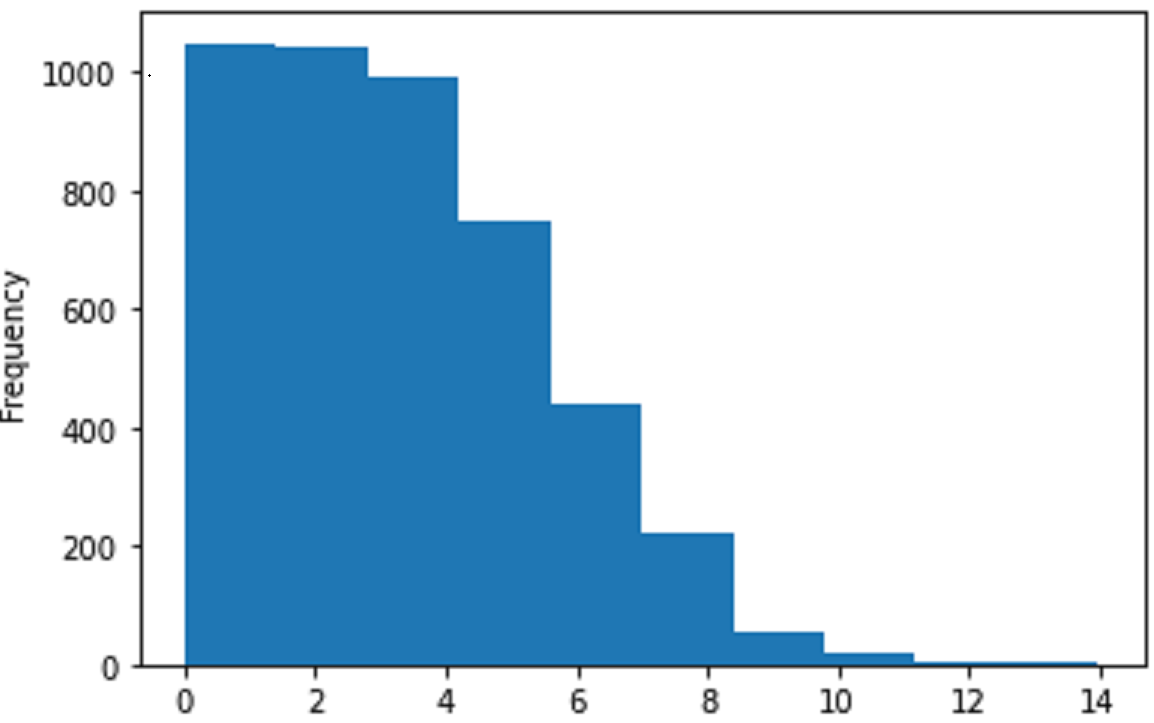
El grafico nos indica que el presupuesto de promoción televisiva tiende a ser entre los 20 a los 100 millones de dólares con una frecuencia de 500

Gráfico de presupuesto en radio.



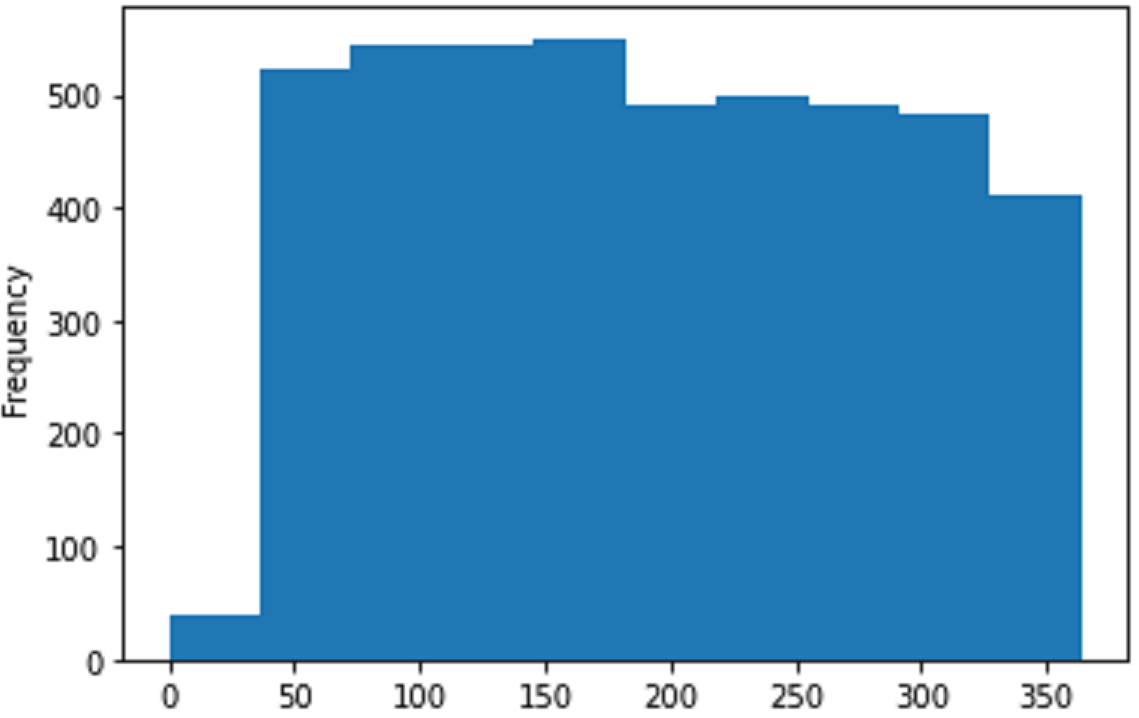
El grafico nos indica que el presupuesto de promoción en radiodifusión está en una frecuencia más alta, cuando la inversión está entre los 10 a 20 millones de dólares.

Grafico del presupuesto en medios sociales.



El grafico nos indica que el presupuesto de promoción en medios sociales está en una frecuencia más alta, cuando la inversión está entre los cero y cuatro millones de dólares.

Grafico del presupuesto en ventas.

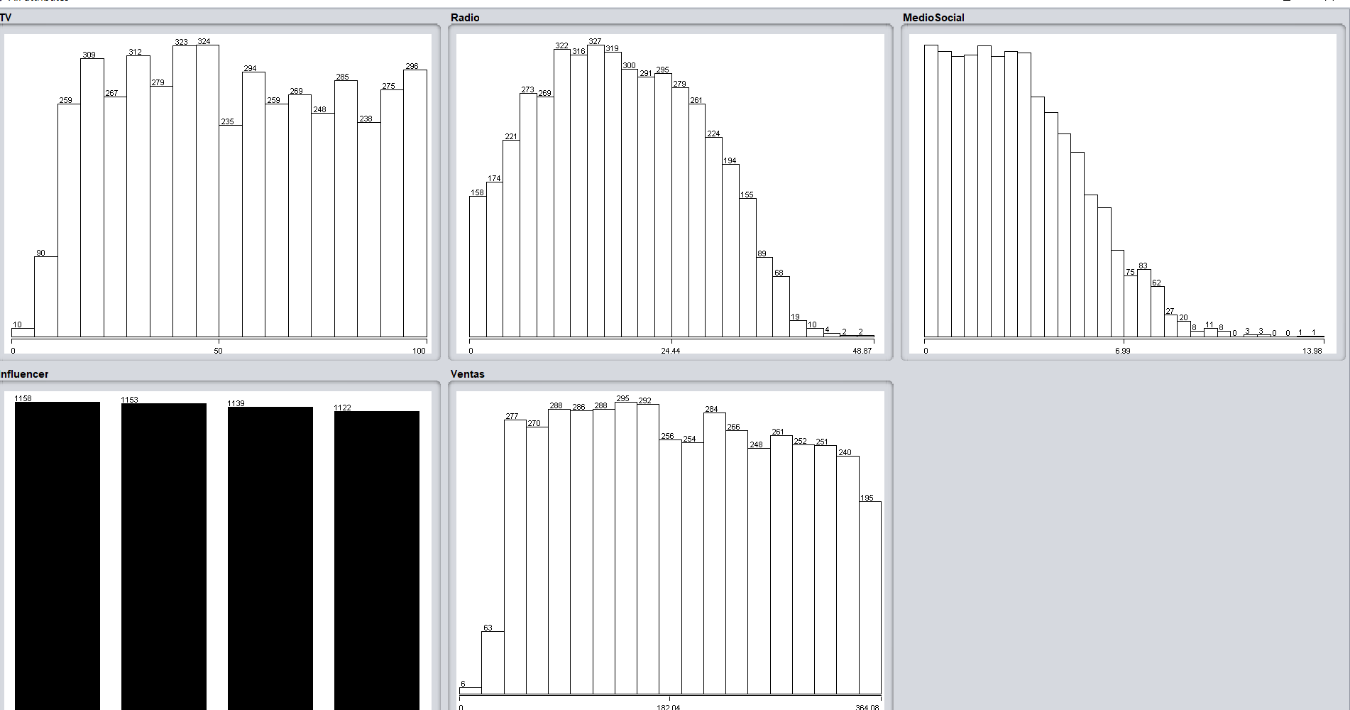


El grafico nos indica que el margen de ventas obtenidas en una frecuencia de 500 el margen está entre los 50 a 150 millones en ventas.

2.- DEL DATASET ANTERIOR en WEKA realice tres algoritmos de pre procesamiento

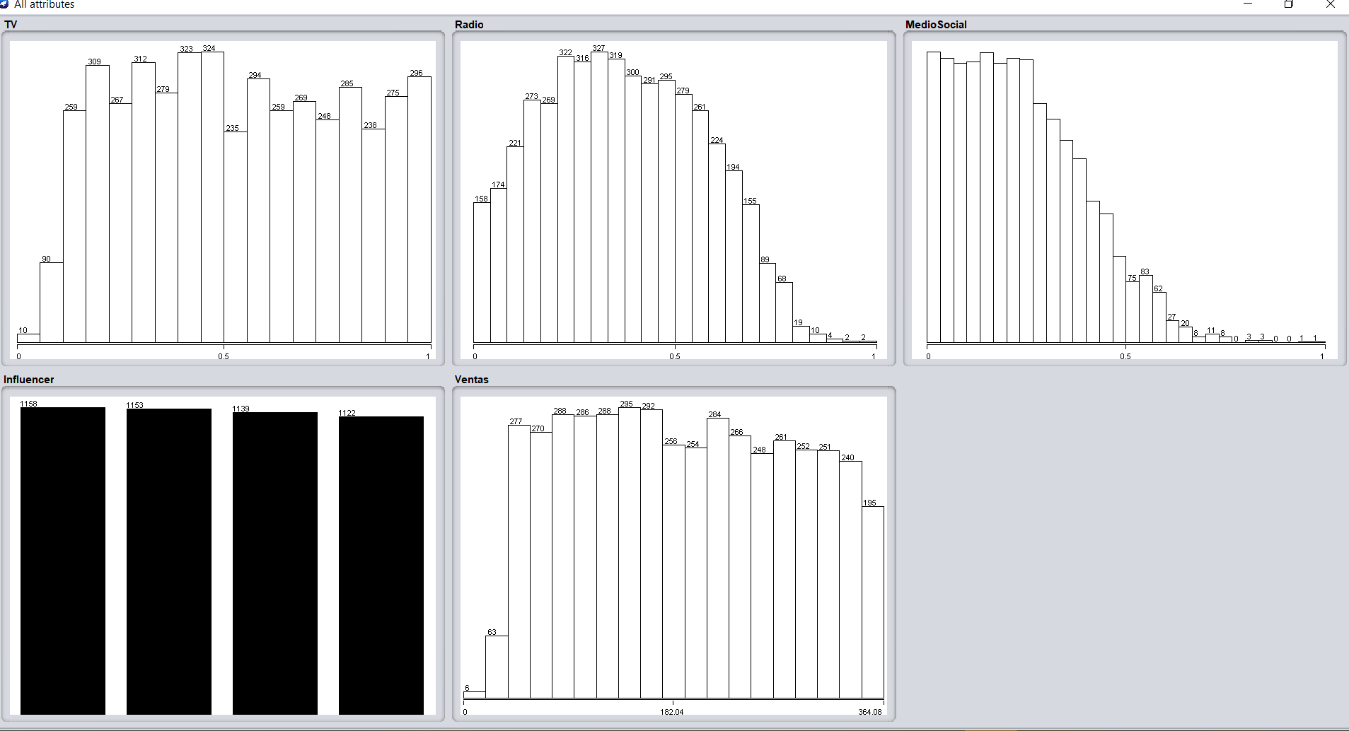
a.- algoritmo de pre procesamiento de normalización

antes de la normalización:



El grafico nos muestra los datos sin normalizar

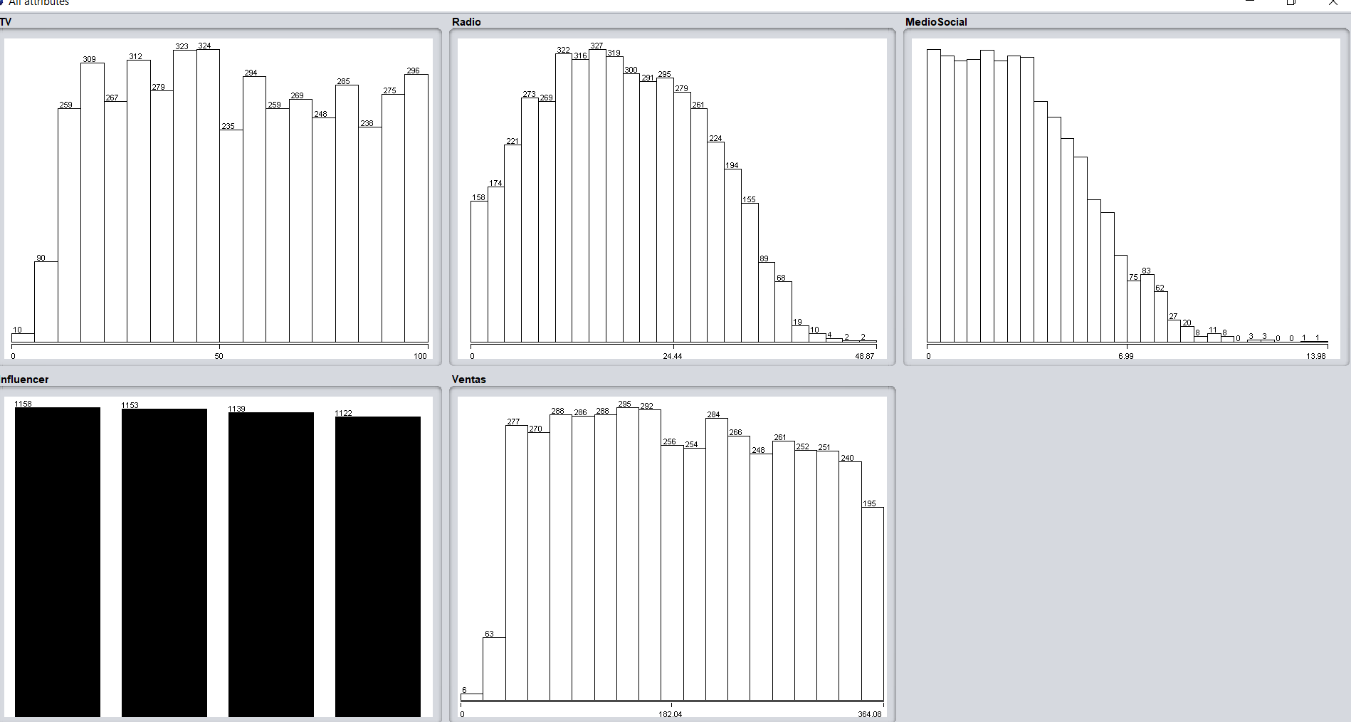
Después de la normalización



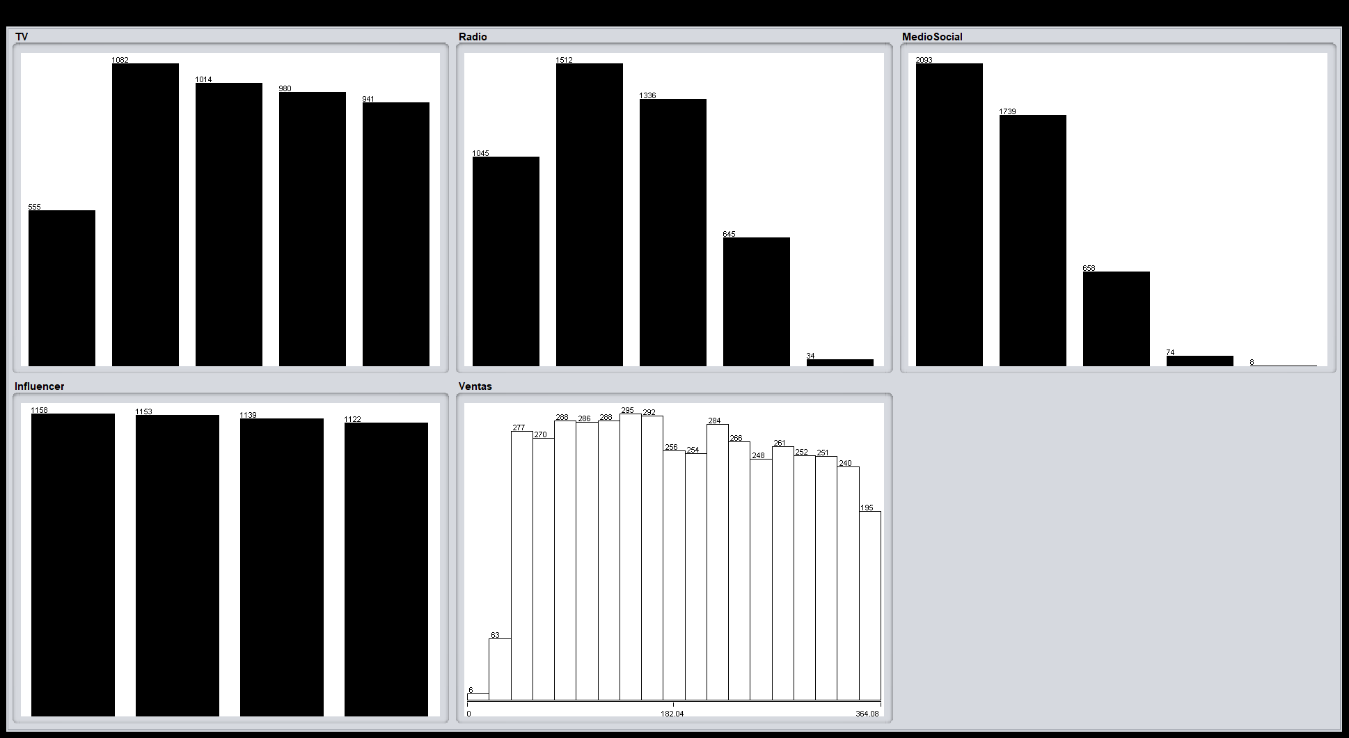
Se puede observar el grafico después de la normalización que no ha existido una gran variación más que la transformación de datos.

b.- algoritmo de pre procesamiento de discretizacion

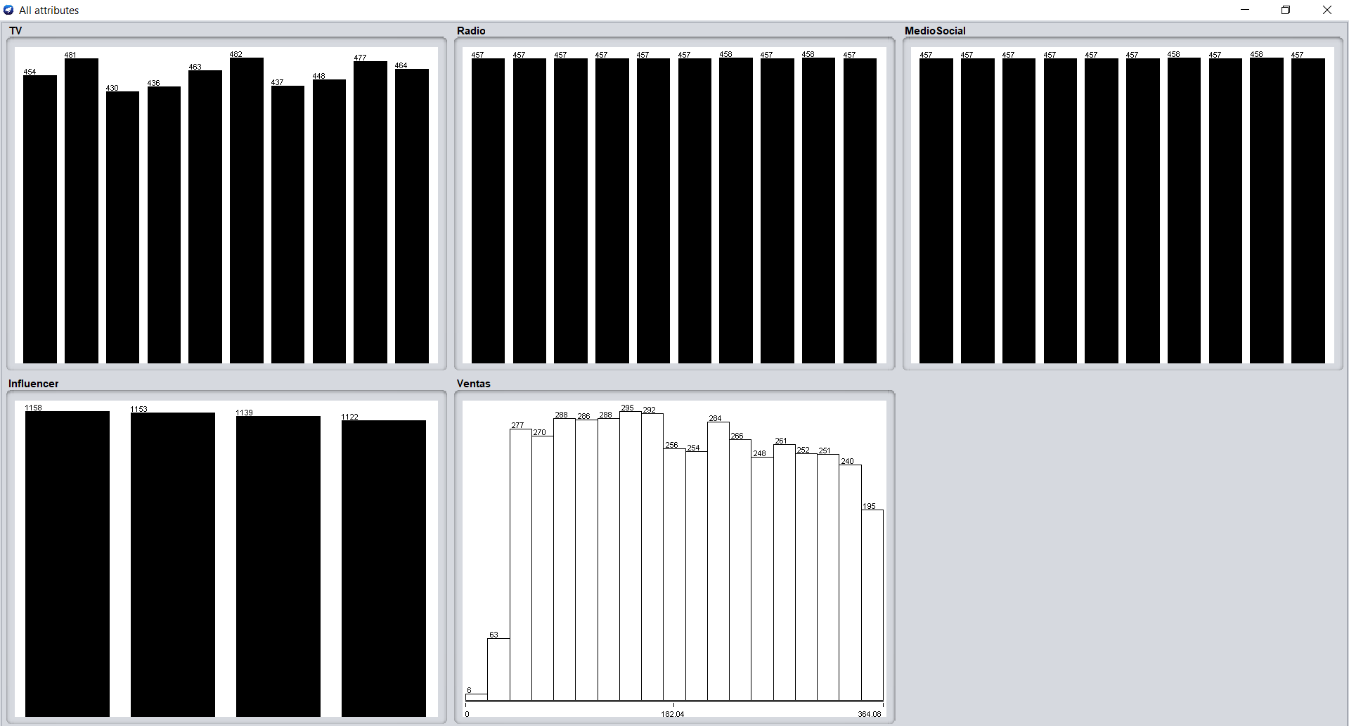
antes de la discretizacion



Después de la discretizacion



Lo que provoca la discretizacion es la división de las frecuencias

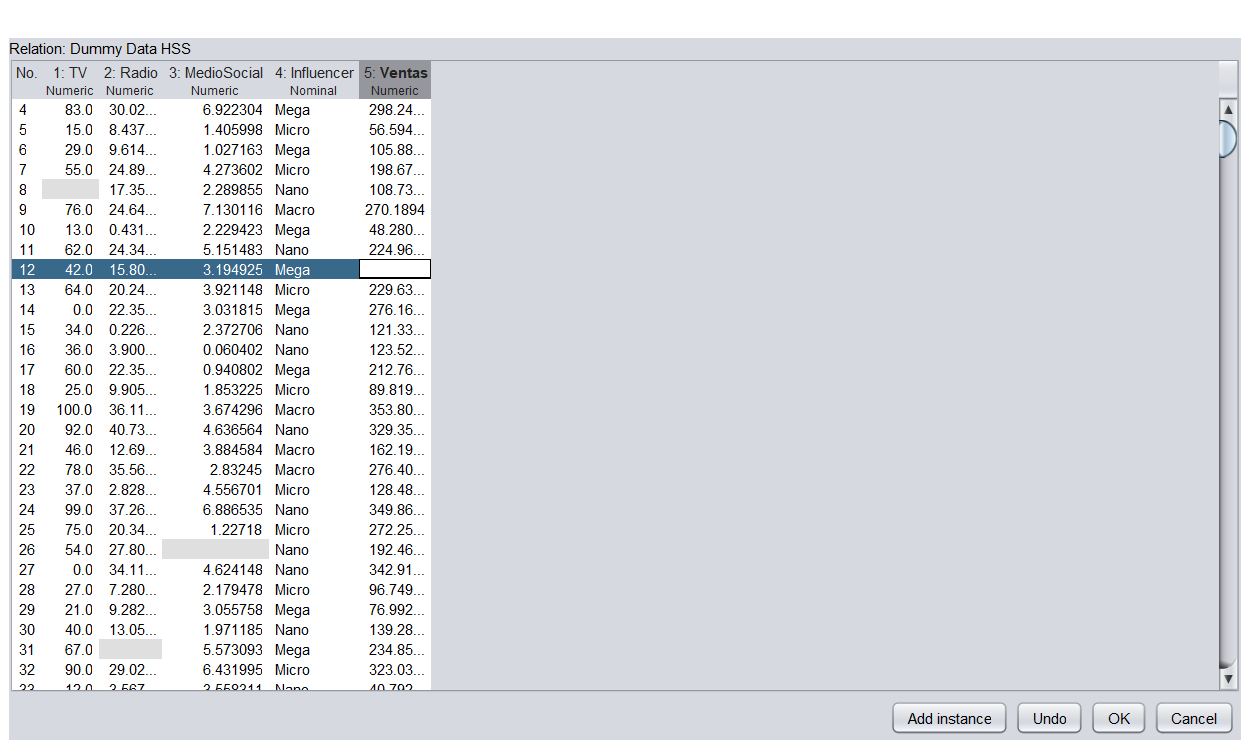


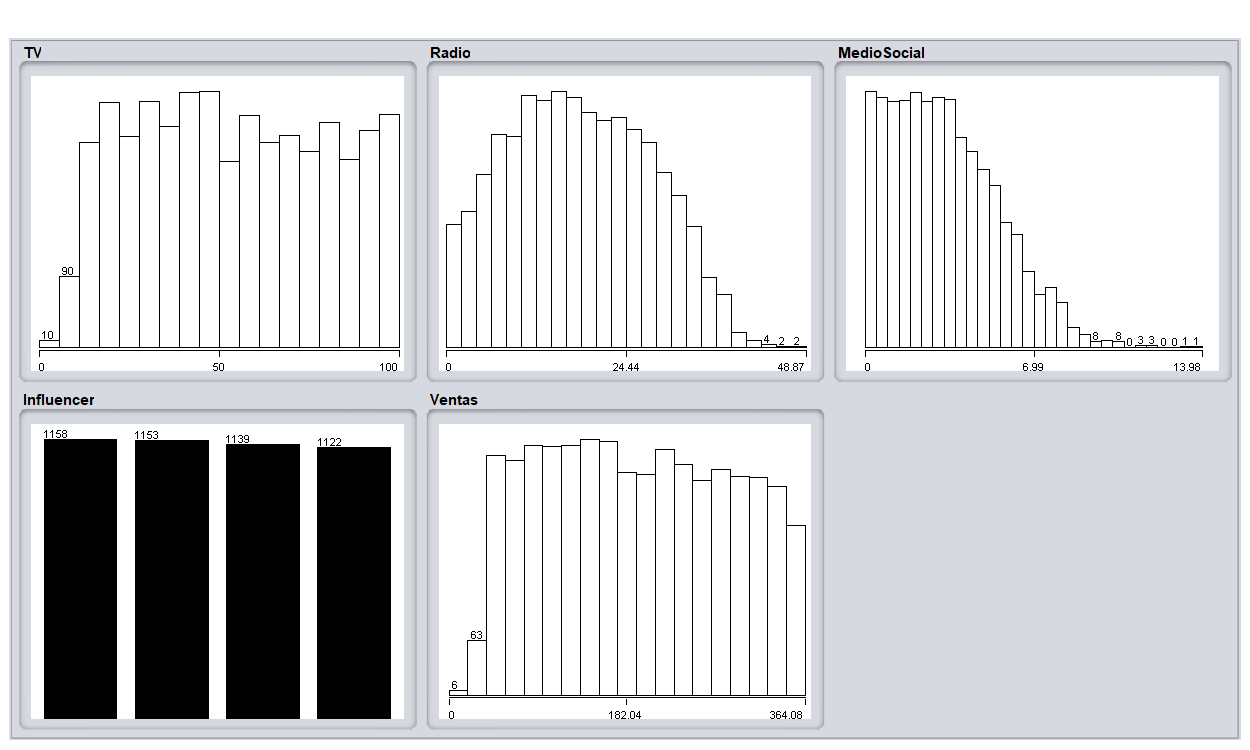
Por qué se realiza la discretizacion para tener tipos de datos ordenados para obtener mejores resultados

c.-

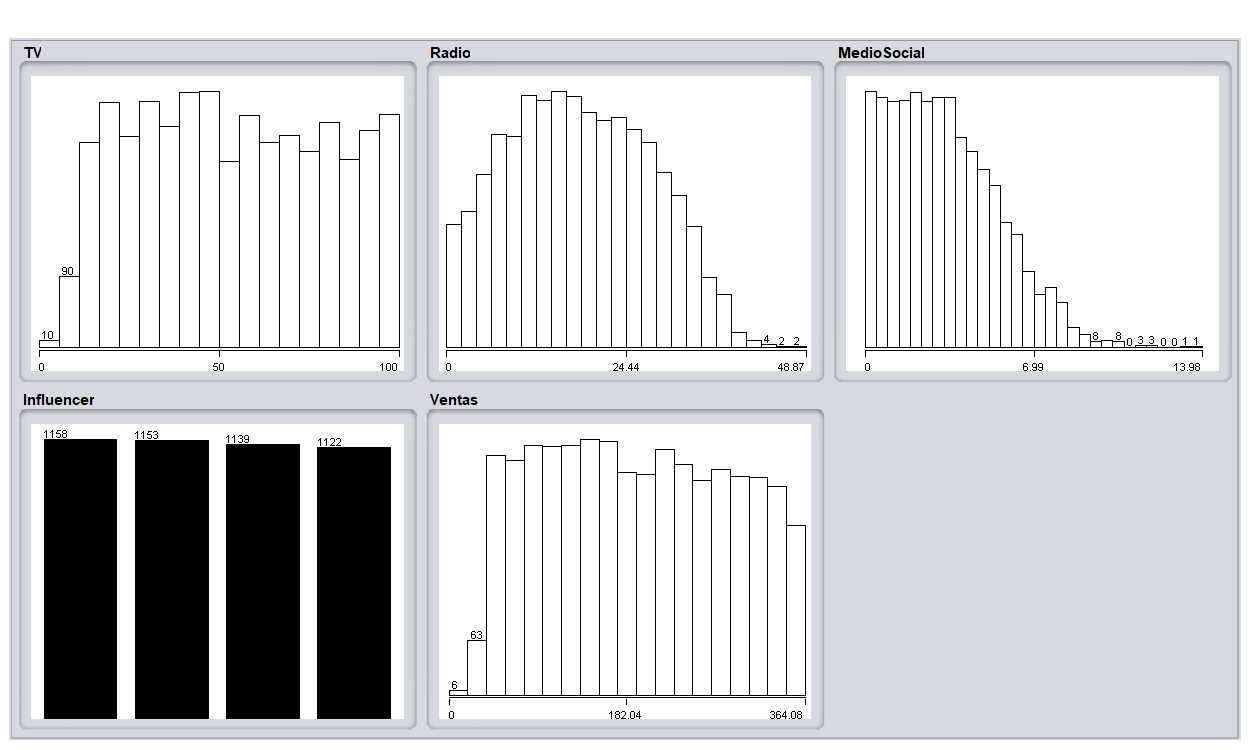
algoritmo de pre procesamiento de imputación

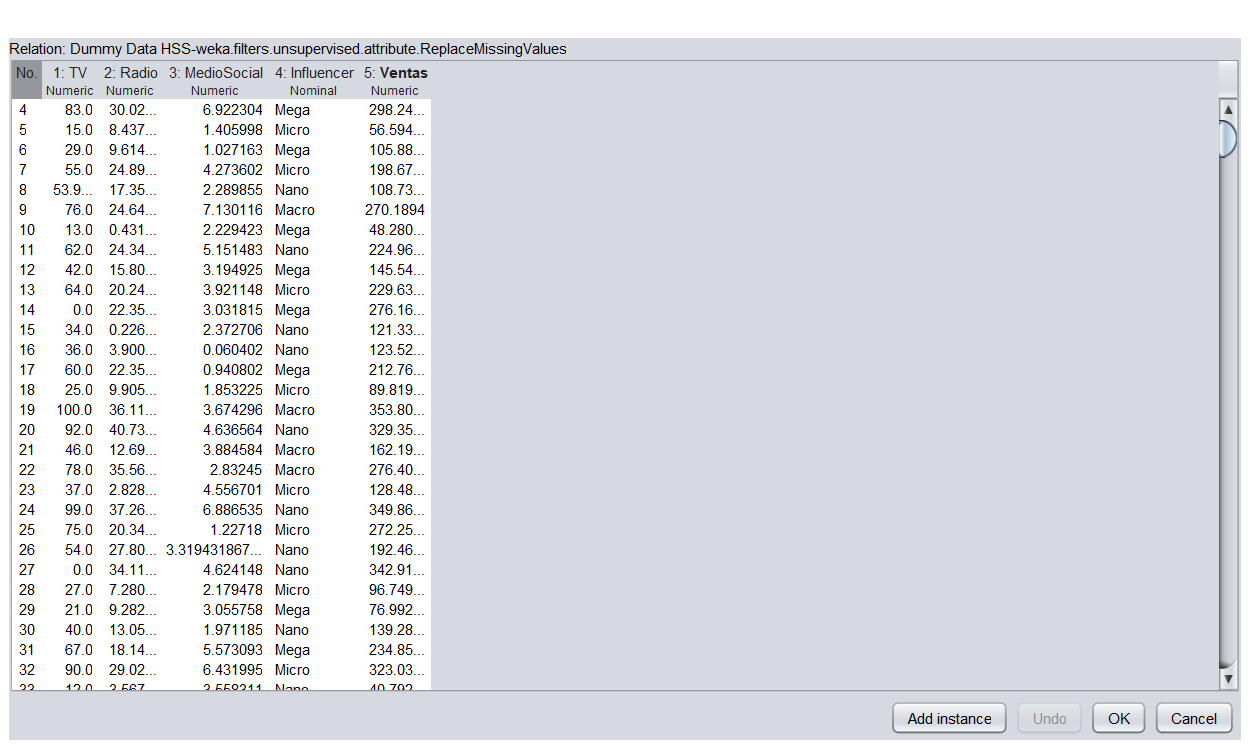
antes de la imputación





Después de la imputación

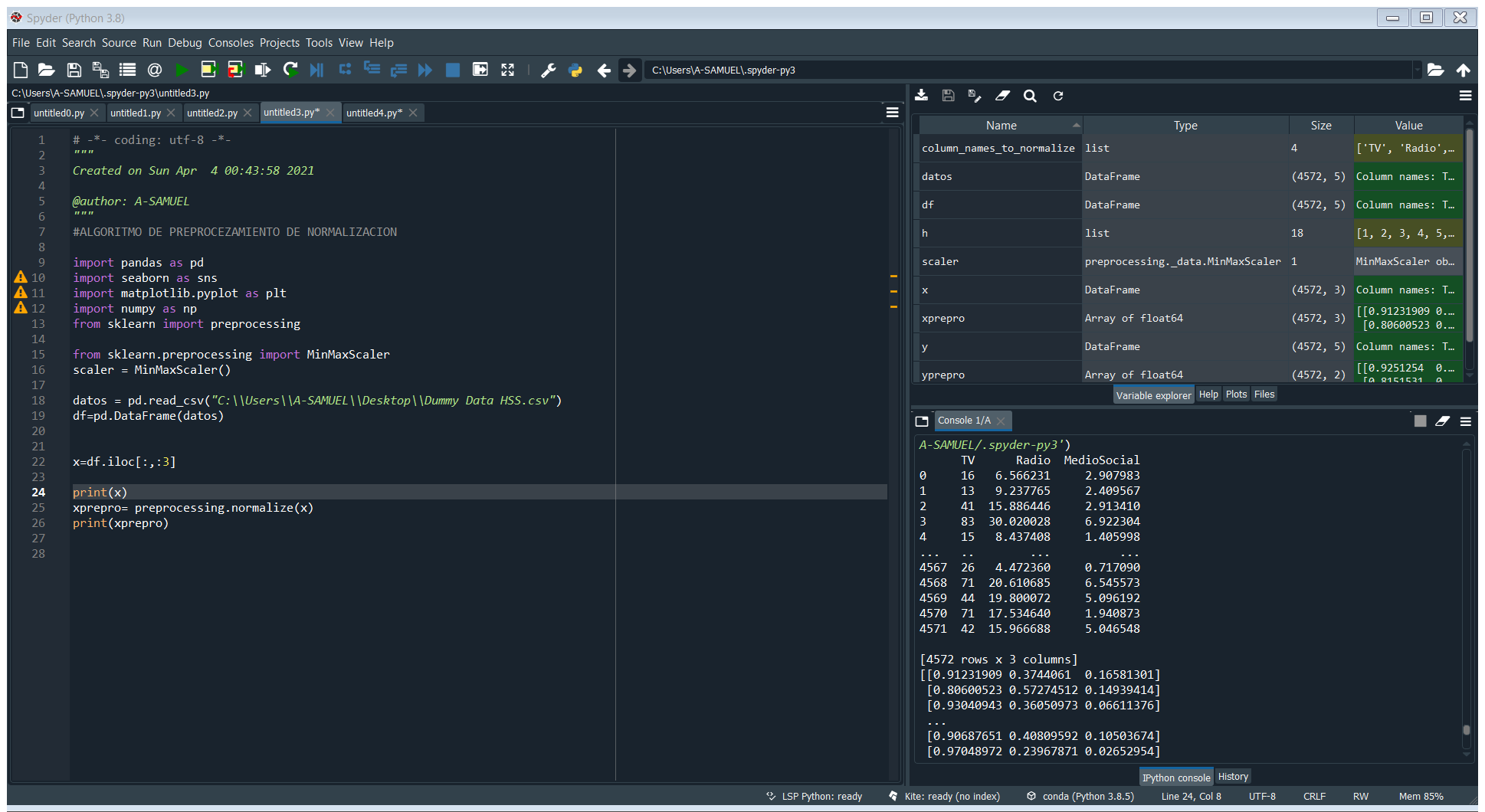




Lo que obtenemos al utilizar el algoritmo de imputación es completar los datos faltantes que puede existir en nuestra data set, la manera estándar de completar los datos faltantes de este algoritmo es obteniendo la media de cada campo.

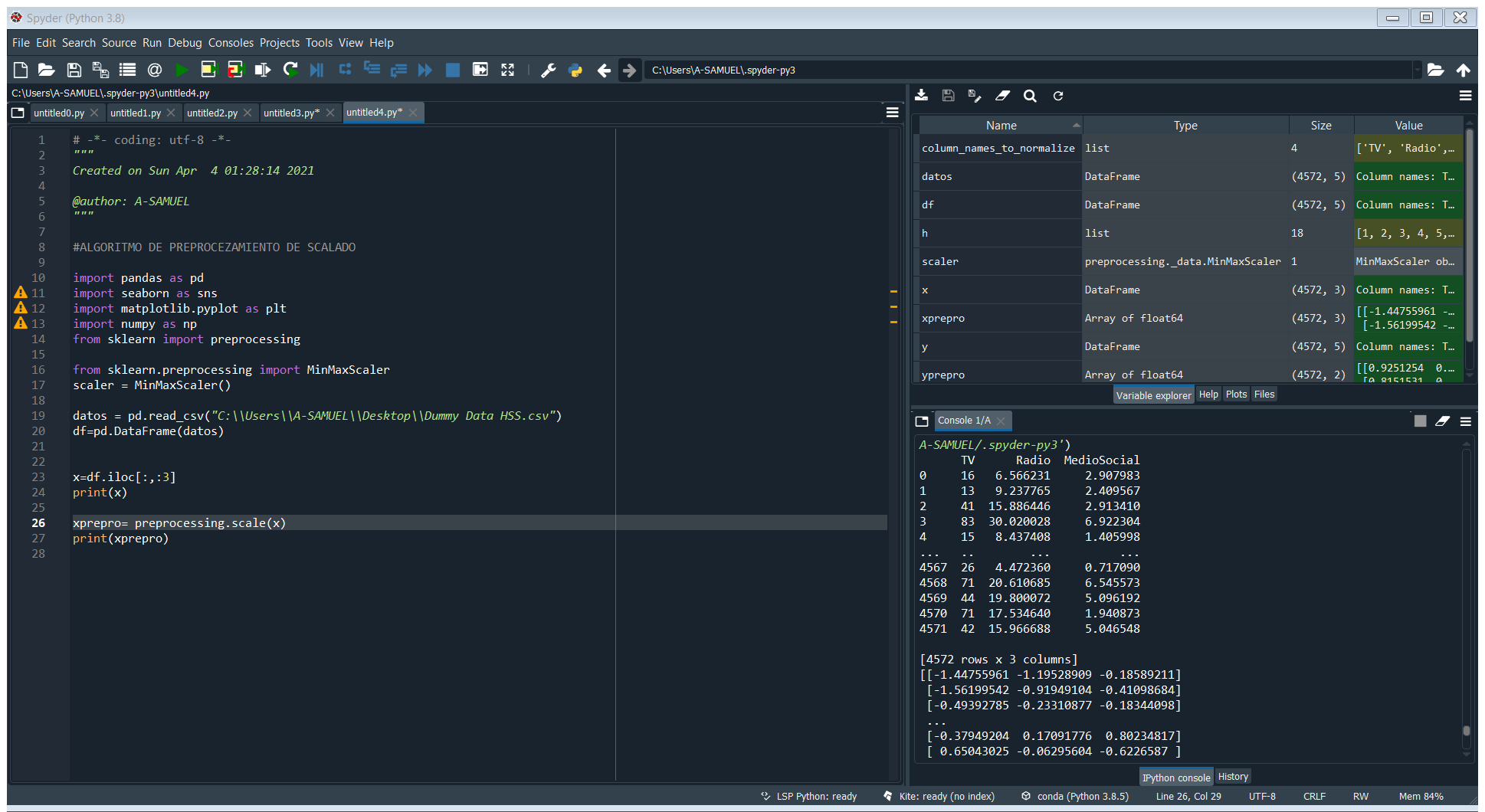
3.- Del data set anterior en Python, realice tres algoritmos de pre procesamiento

a.- algoritmo de pre procesamiento de normalización.

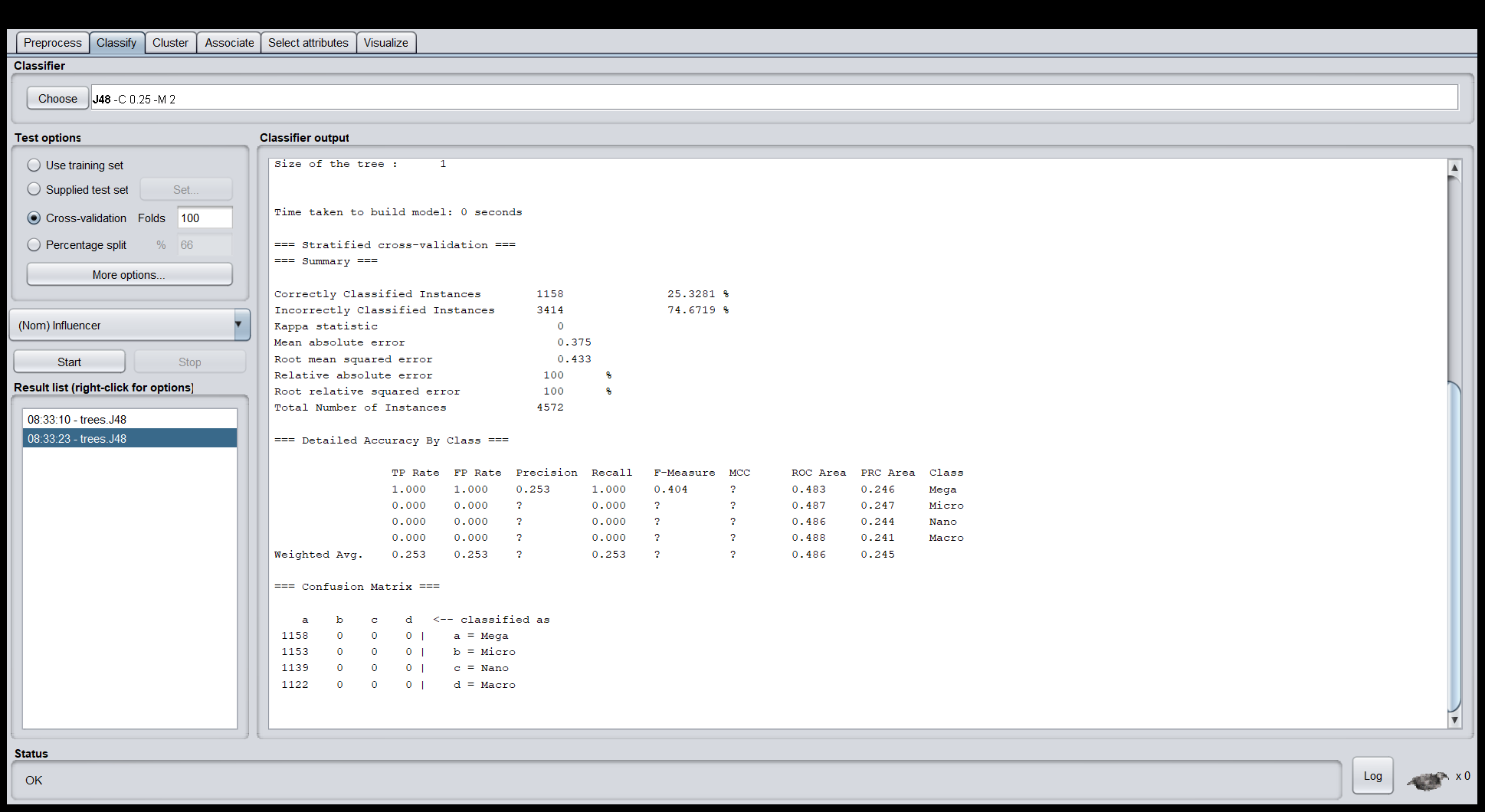


Como se puede observar los datos han sido normalizados.

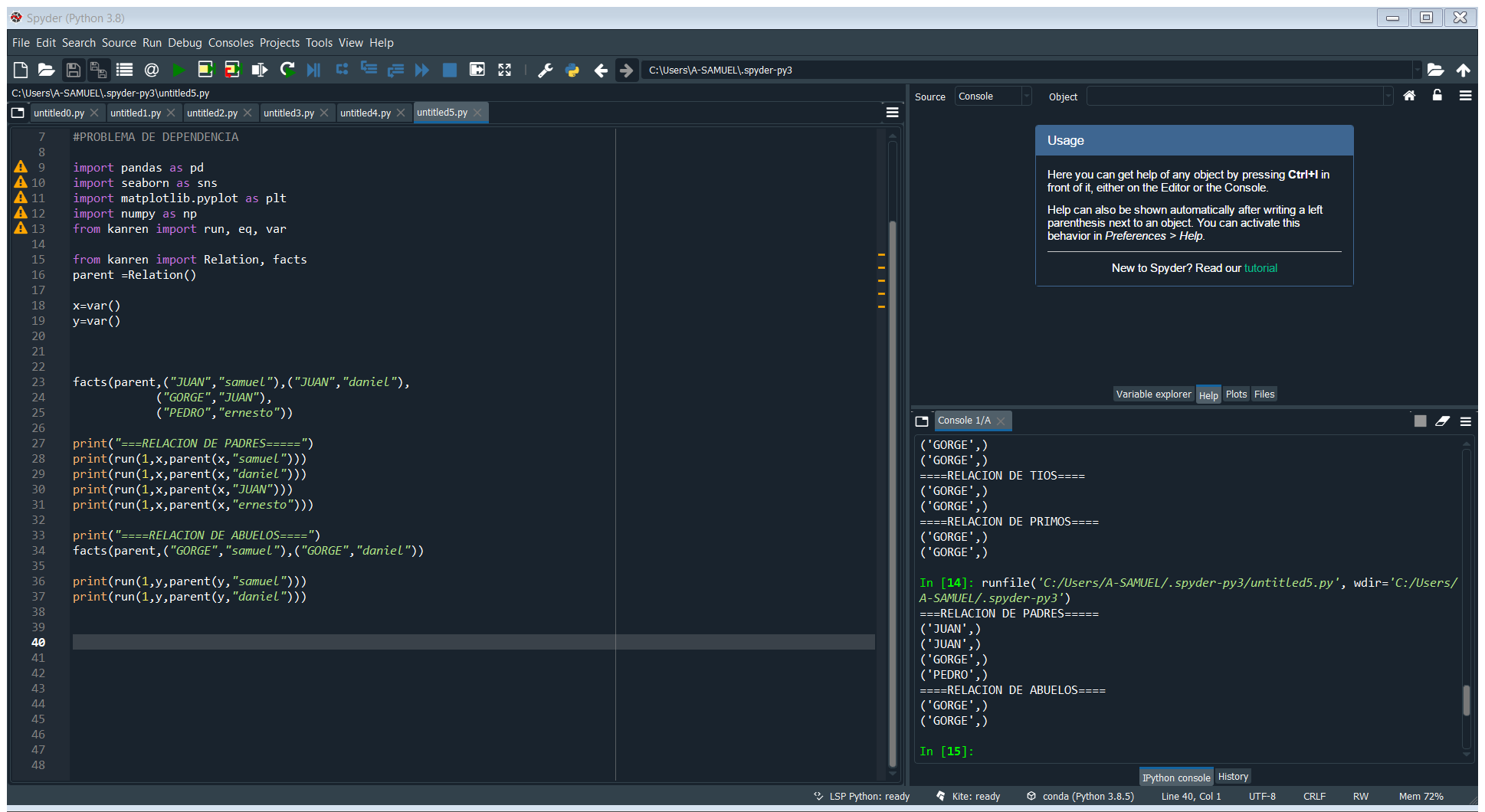
b.- Algoritmo de pre procesamiento de escalado



4.- del problema anterior en Python y weka realice un árbol de decisión basado en entropía, compare los resultados, esto debe repetirse al menos 10 veces y comparar las matrices de confusión y obtener la confiabilidad. Se recomienda comparar media y mediana de los resultados



5.- Con el uso de librerías en Python, construya la dependencia de abuelos, tíos, padres, primos, e hijos, de su familia.



El grafico nos muestra la dependencia de padres y abuelos.