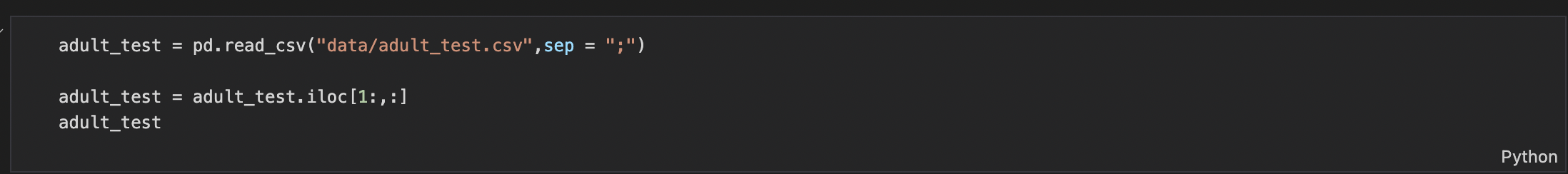
Graphical user interface

Description automatically generated

Text

Description automatically generated



Graphical user interface

Description automatically generated

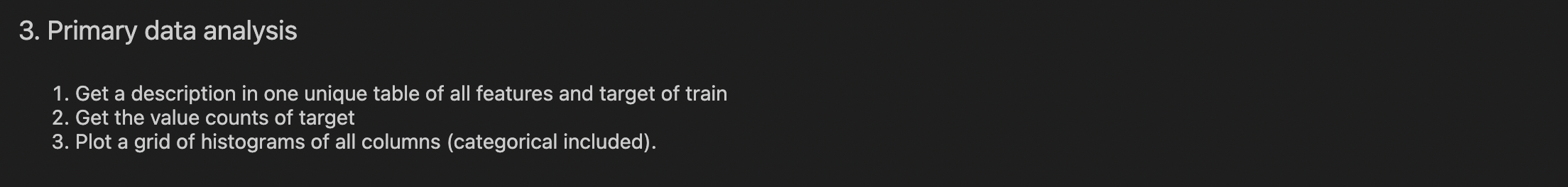
Shape

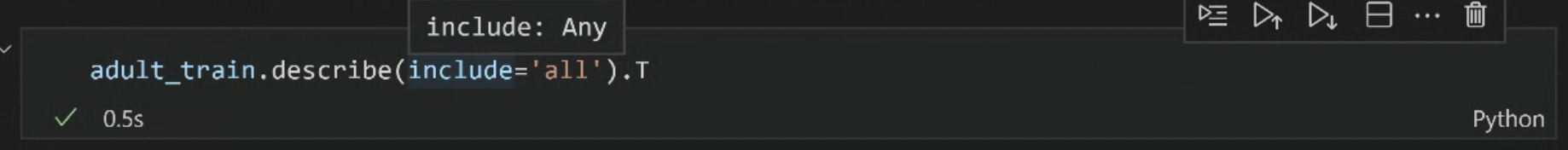
Description automatically generated with medium confidence

Binarizamos train y test para las columnas que necesitamos

Graphical user interface, text

Description automatically generated





Graphical user interface, application

Description automatically generated

Los datos no están balanceados

Establecemos el tamaño de la figura y el número de filas a columnas

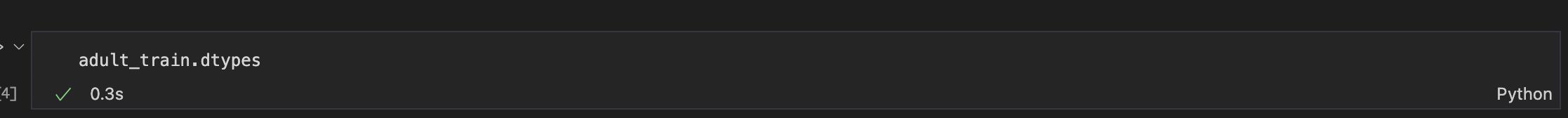
Generamos con un bucle un grid de histogramas

A screenshot of a computer

Description automatically generated

Text

Description automatically generated



La edad no está como entero, le cambiamos el tipo

A screenshot of a computer

Description automatically generated with medium confidence

Nos quedamos con las columnas tipo object y numéricas en listas separadas

Text

Description automatically generated

Graphical user interface

Description automatically generated

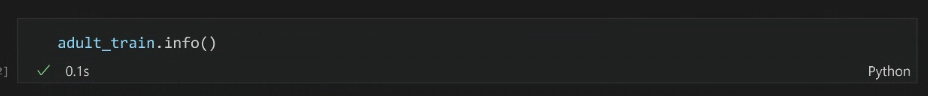
Nos quedamos con el primer valor de la moda, ya que puede haber varios

Rellenamos valores null

Text

Description automatically generated

Comprobamos que no hay valores null



Text

Description automatically generated

Tenemos las mismas columnas, pero podemos tener valores en train y test diferentes

Text

Description automatically generated

Sacamos los valores que tenemos en las columnas train que no tenemos en test

Graphical user interface, text

Description automatically generated

Comprobamos el número de columnas y vemos que los datos de test tienen una columna menos.

Graphical user interface

Description automatically generated

Text

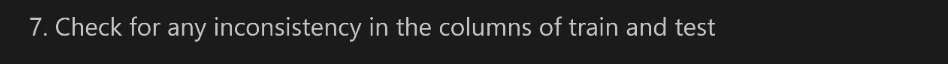
Description automatically generated

Con esta otra solución las columnas nuevas que pueda haber en test son ignoradas.

¡ ¡ Deben ser las mismas y en el mismo orden!!

Text

Description automatically generated



Graphical user interface, application

Description automatically generated

Graphical user interface, application

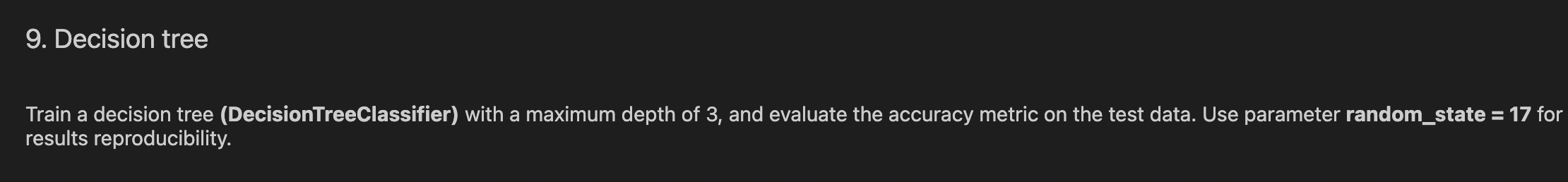
Description automatically generated

A partir de aquí empezamos la división de datos en train y test y entrenamos el modelo.

Eliminamos la columna Target

Graphical user interface, text, application

Description automatically generated



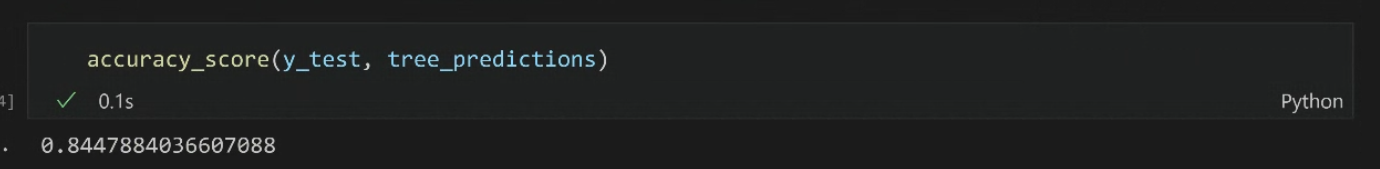
Graphical user interface, text, application

Description automatically generated



Graphical user interface, text

Description automatically generated



Graphical user interface, text

Description automatically generated with medium confidence

Probamos con una serie de combinaciones de parámetros, para maximizar el accuracy, que es nuestra métrica elegida.

Text

Description automatically generated

Text

Description automatically generated

Una vez que sabemos la mejor combinación de parámetros hacemos la predicción

Graphical user interface, text

Description automatically generated

Text

Description automatically generated

Tenemos un 85% con los valores de test