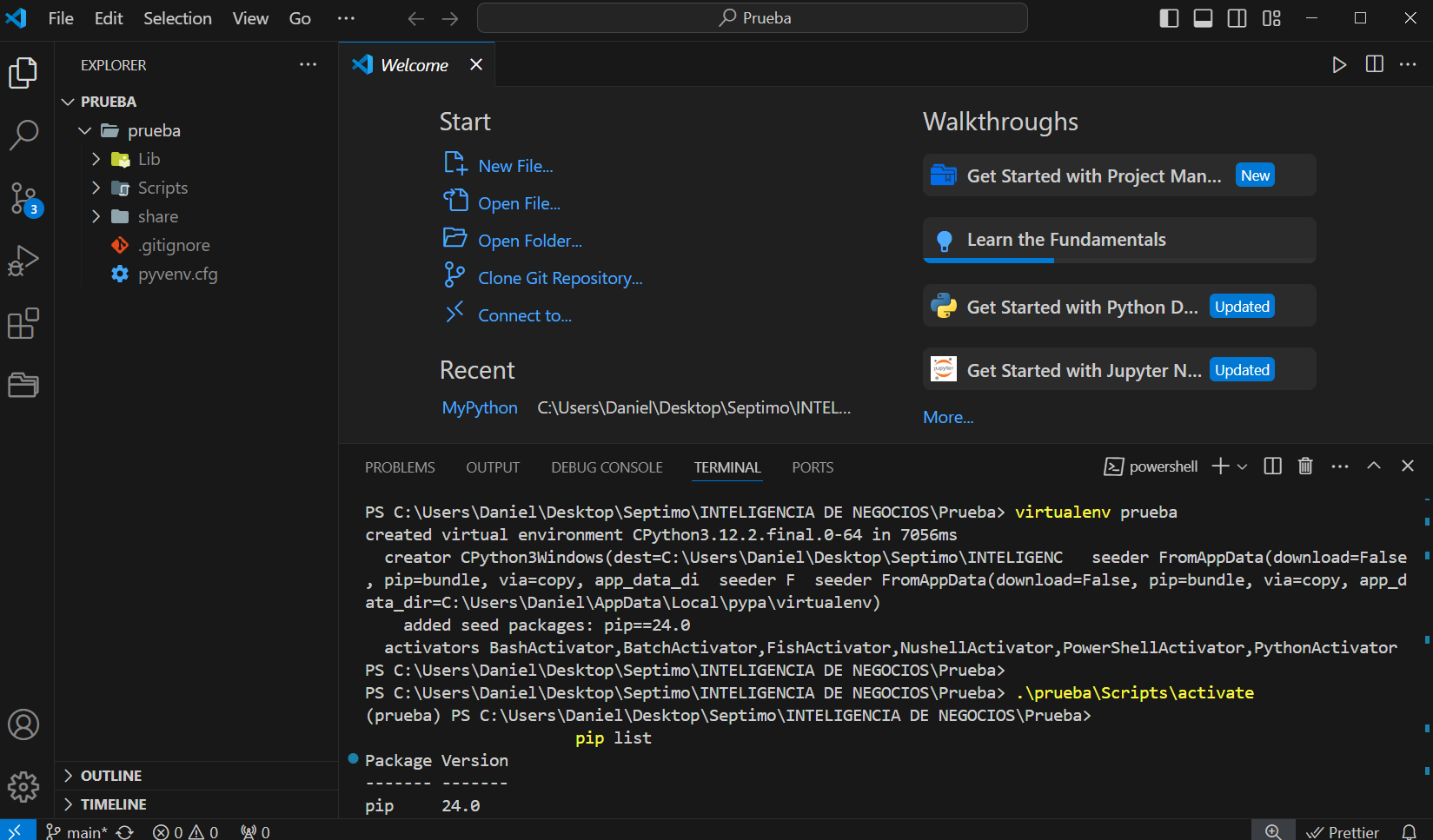
Prueba N1

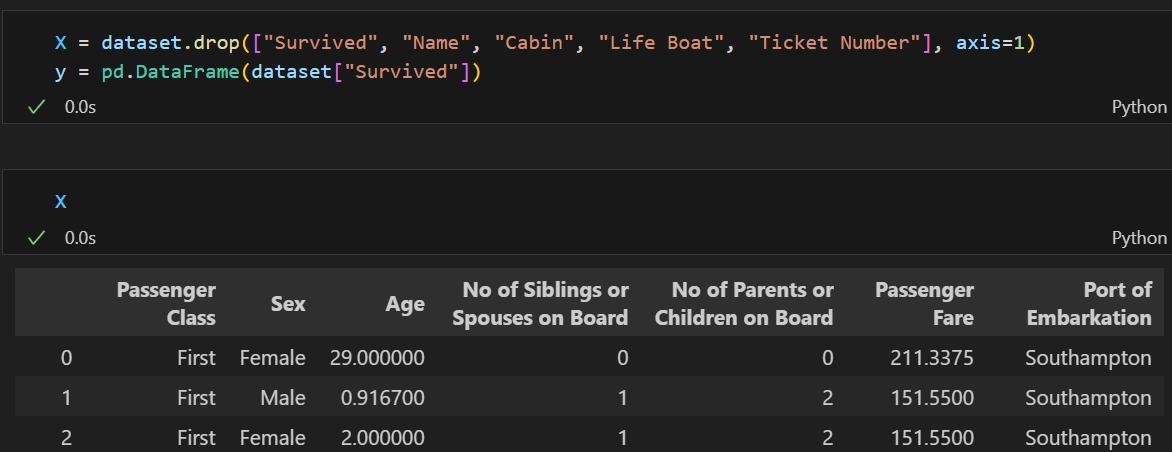
Nombre: Daniel Galarza

Fecha: 09/04/2024

1. Creación del ambiente llamado Prueba
2. Use el [dataset](https://sistemaseducaciononline.uta.edu.ec/mod/folder/view.php?id=51763) Titanic.xlsx
3. Aplique el algoritmo de clasificación

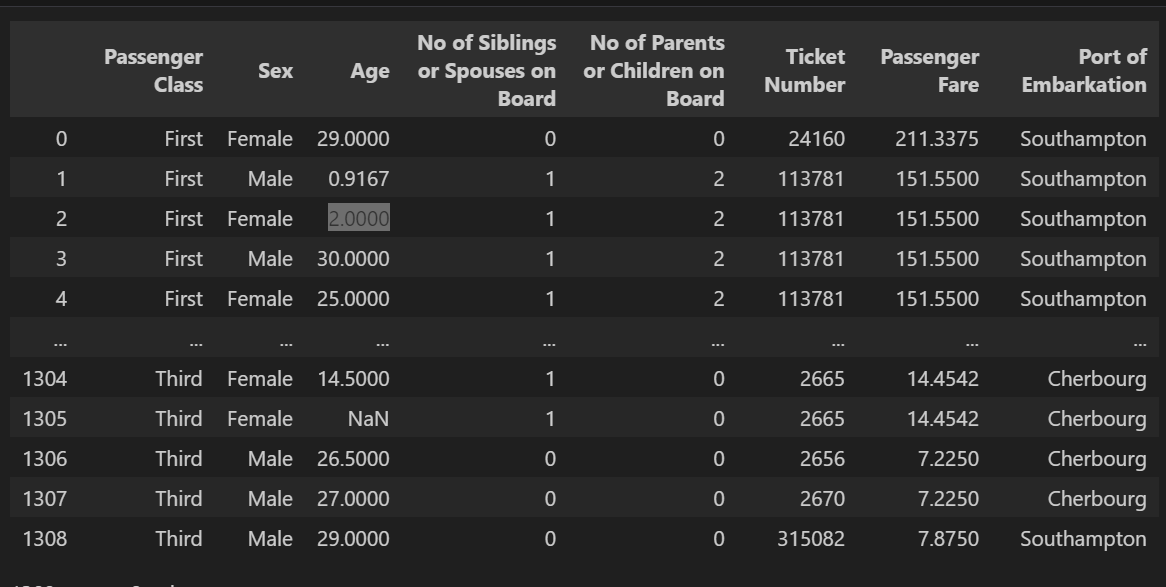
Elimino las variables independientes que no me sirven

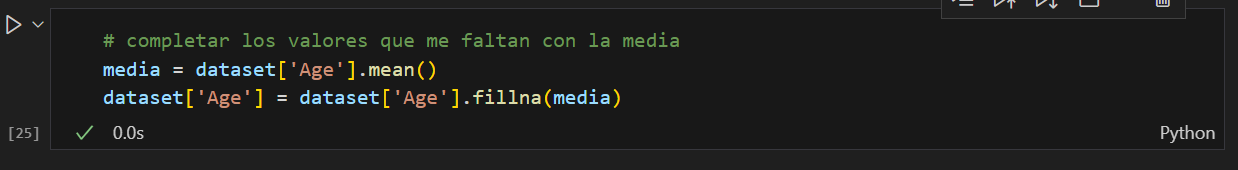
Los elimino porque el Name no me sirve, la cabina tampoco y el bote salvavidas tiene demasiados valores nulos que no podre controlar solo usando la media.

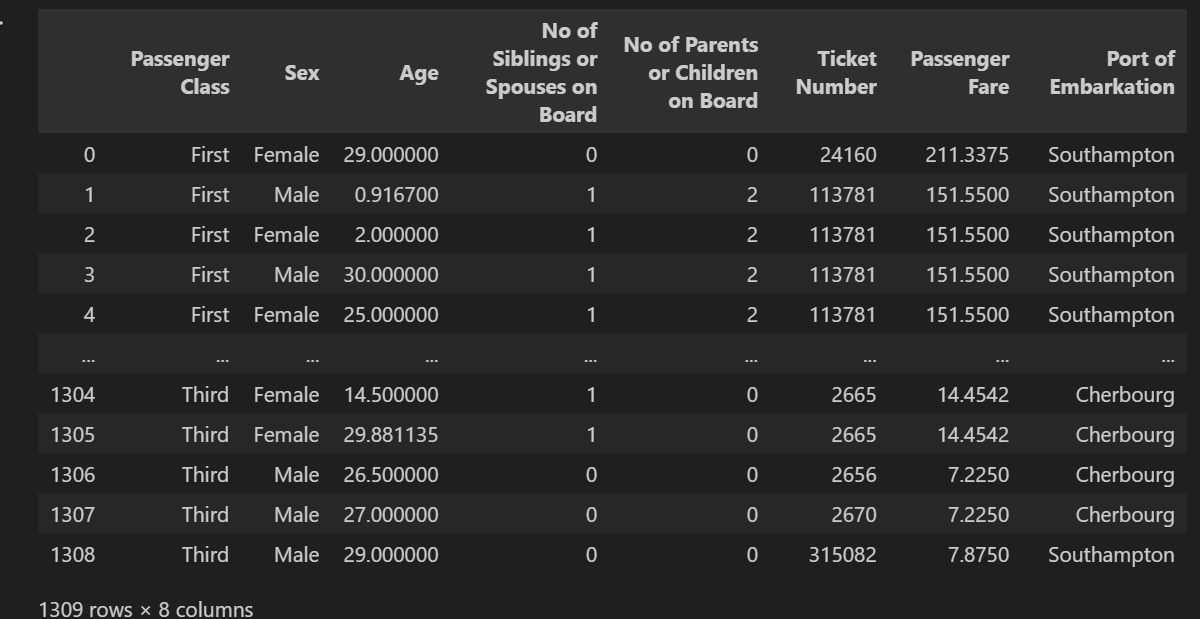


Verificamos que existen valores nulos

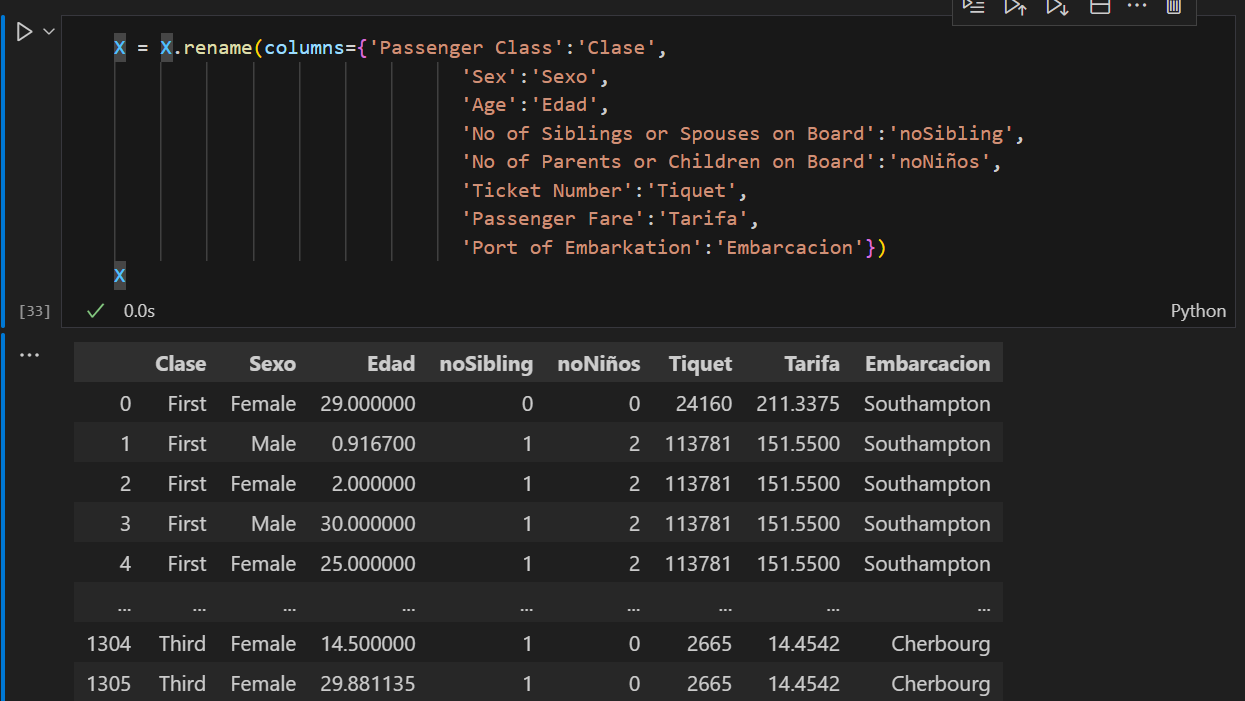
Llenamos con la media esos valores para no que no exista sesgo



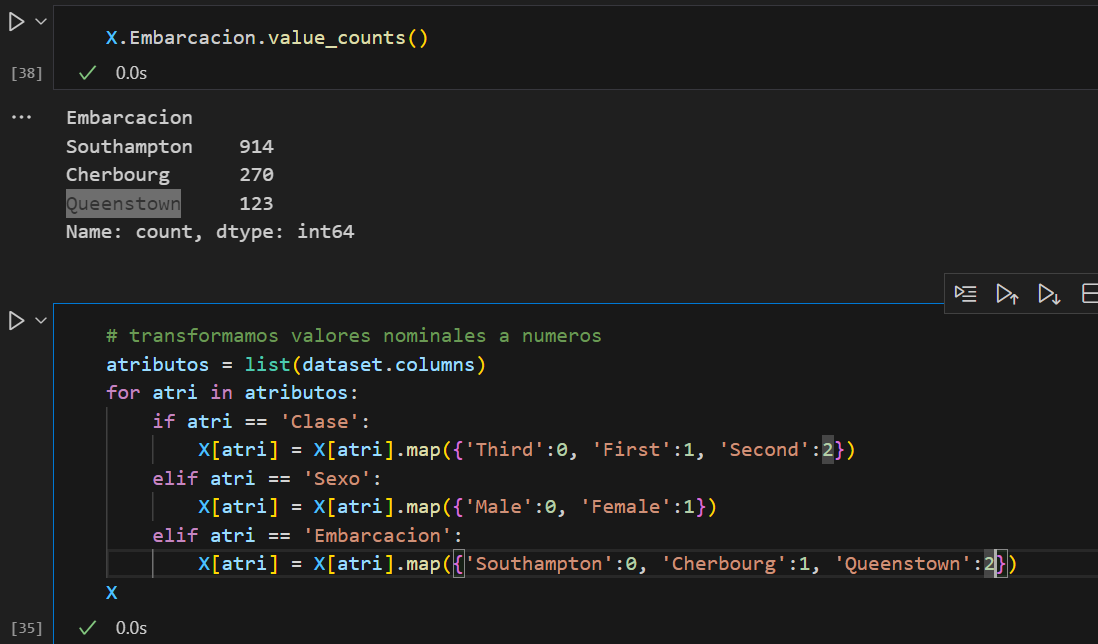


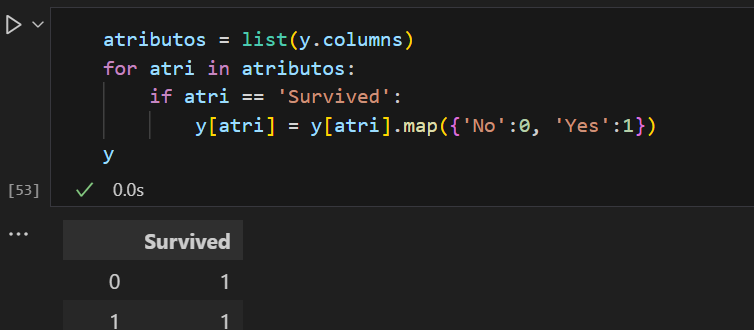


Renombrar etiquetas



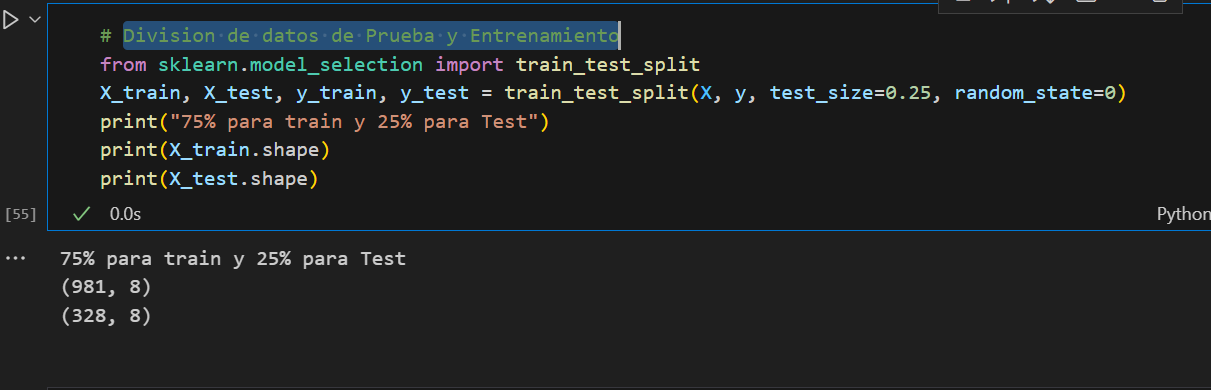
Transformamos valores nominales a números





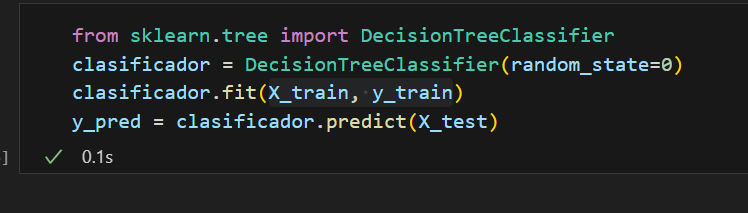
División de datos de Prueba y Entrenamiento:

Aplicamos el algoritmo con un 25% de datos solamente sean Test y el resto con 75 de entrenamiento.

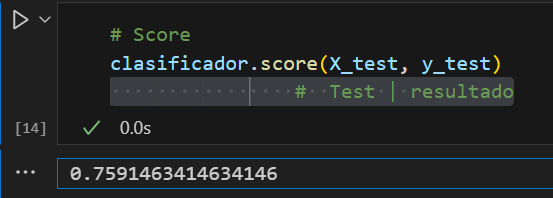


1. Aplique el algoritmo de clasificación

Para ello usamos las variables de x en entrenamiento y y en entrenamiento dado eso calculamos la predicción.



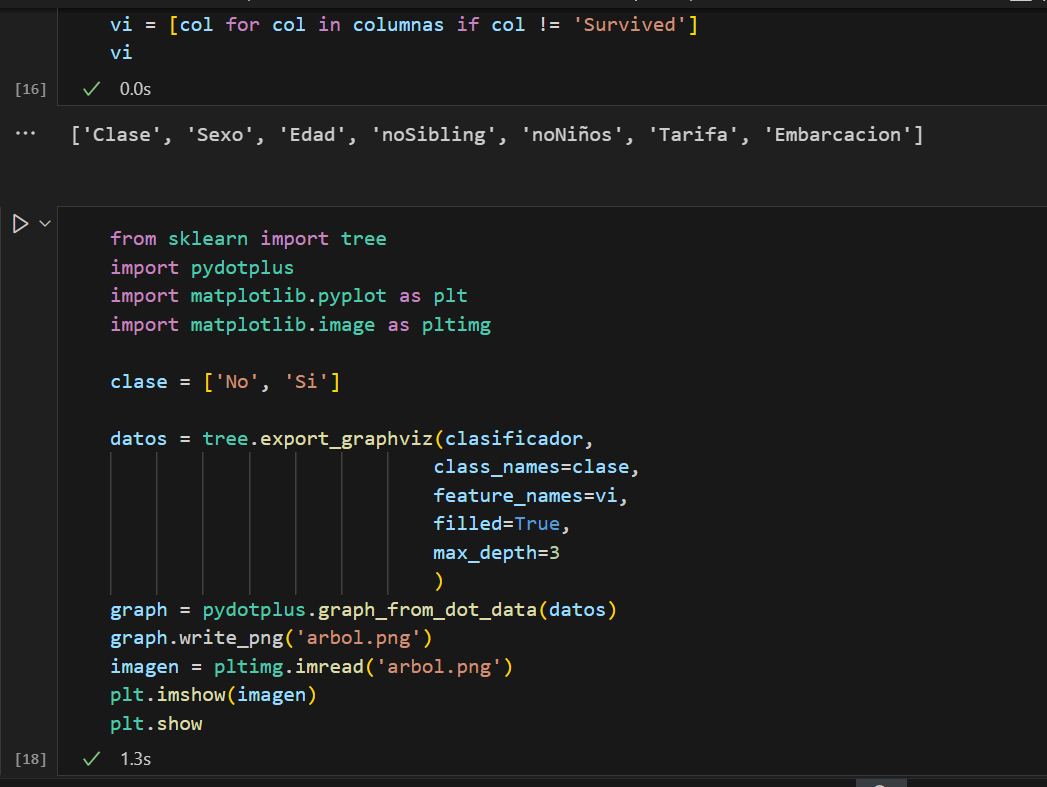
4. Determine el score

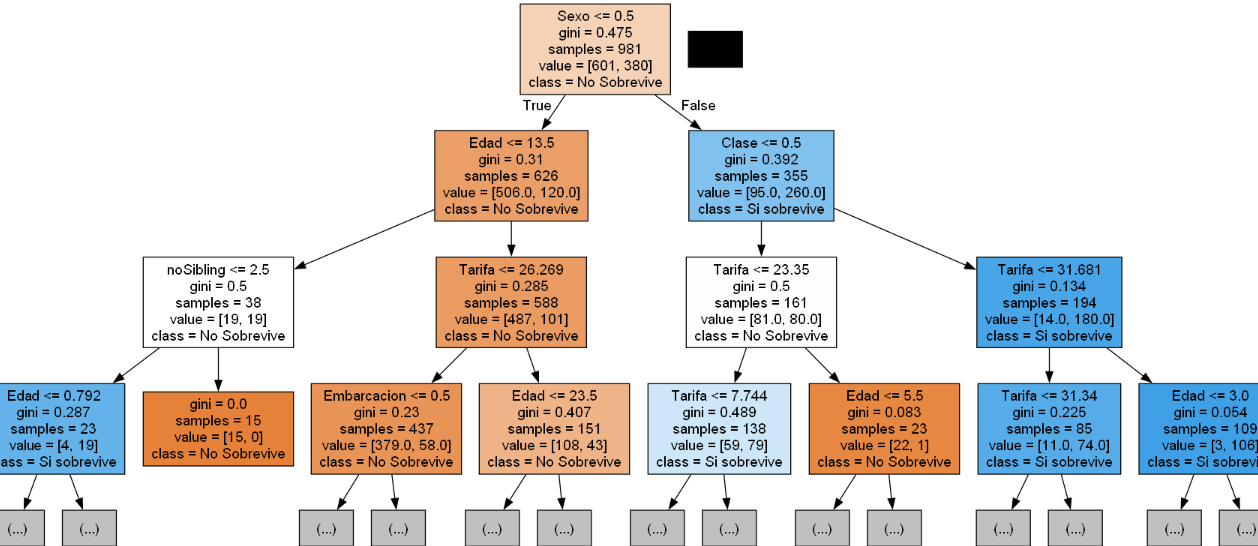


5. Genere la Matriz de Confusión

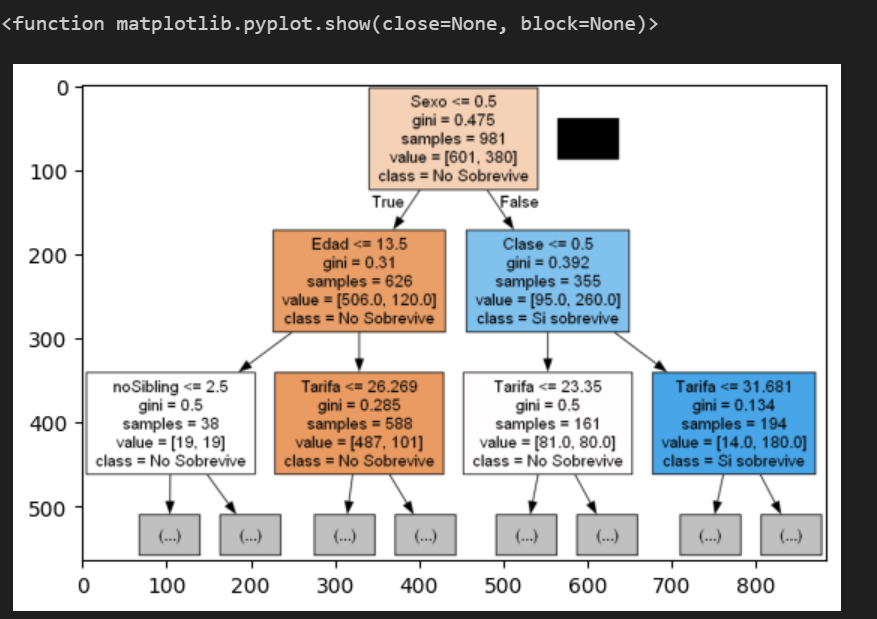
6. Grafique el árbol

Obtención de variables independientes pata crear el árbol y creación del árbol usando max\_depth para hacerlo más pequeño usando 3.



Me sale un árbol grande:  


Por ello aplicaremos una reducción con un max\_depth = 2:



7. Utilice Weka y compare el resultado obtenido en Python