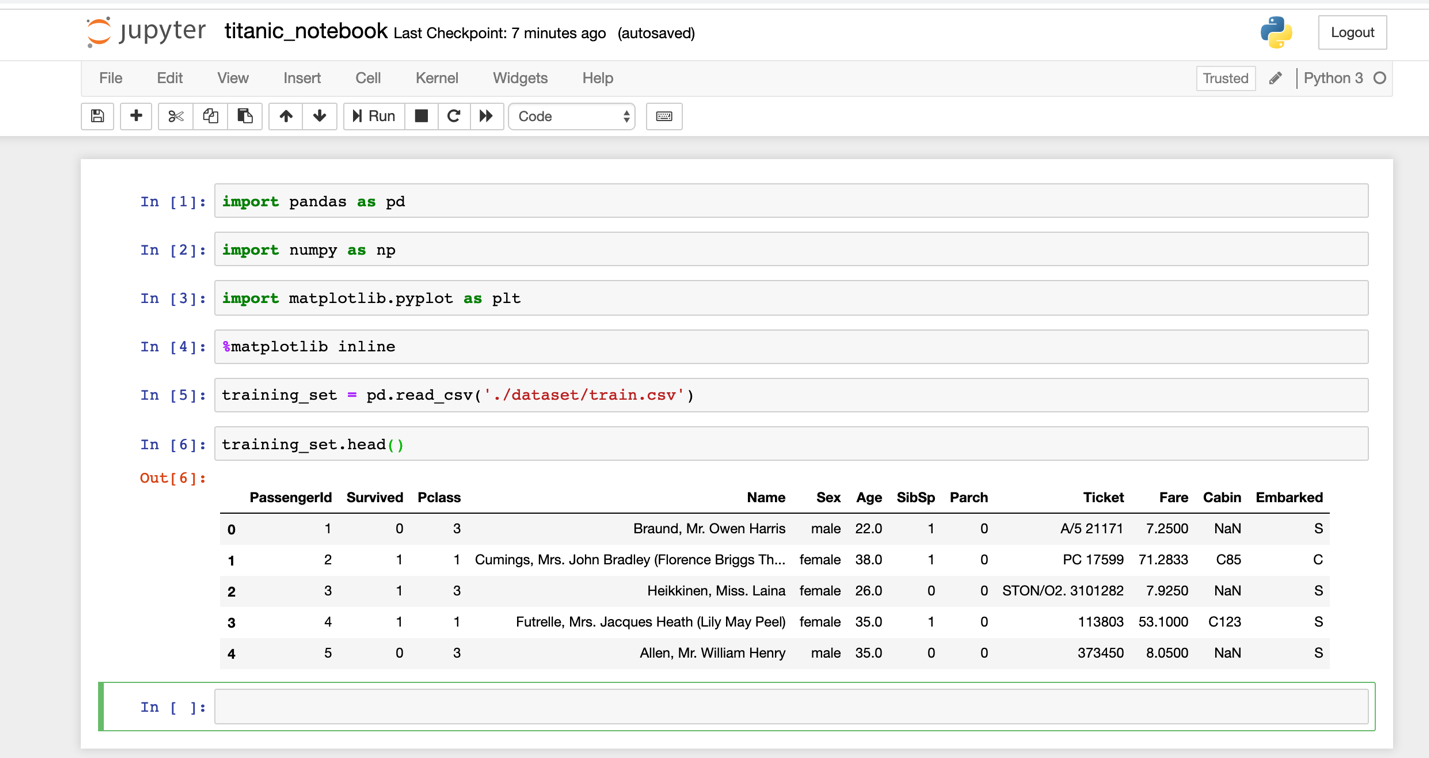
# Requerimientos técnicos sobre Mac OS

Lo primero que se debe tener en cuenta es instalar las siguientes dependencias:

* brew install python3
* pip3 install jupyter jupyter\_kernel\_gateway
* pip3 install matplotlib
* pip3 install pandas
* pip3 install sklearn

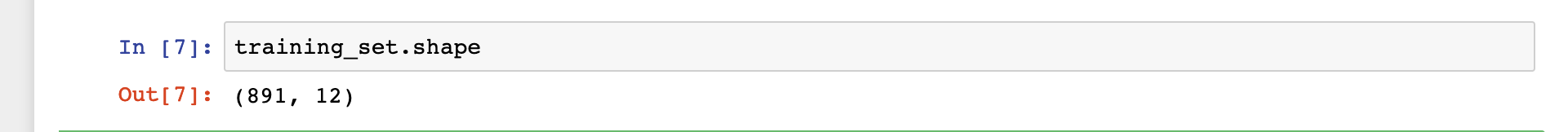
# Procesamiento del conjunto de datos.

Lo primero que se debe realizer es leer el dataset , para lo cual se empleara la función read\_csv de pandas, la cual cargará el archive train.csv. Con la función head se podrán pre-visualizar algunas de las filas del conjunto de datos como se puede apreciar en la siguiente imagen.

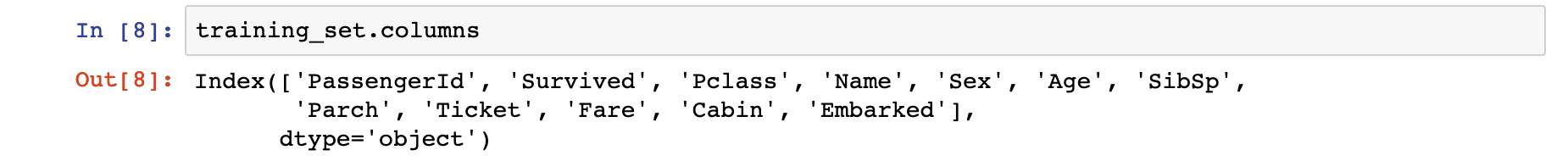


## Información básica del conjunto de datos

La estructura training\_set (pandas.core.frame.DataFrame) tiene varias propiedades entre ellas shape, la cual muestra el número filas y columnas. Por ejemplo la siguiente imagen indica que tiene 891 filas y 12 columnas.



Si se quiere averiguar los títulos de las columnas se puede acceder a la propiedad columns



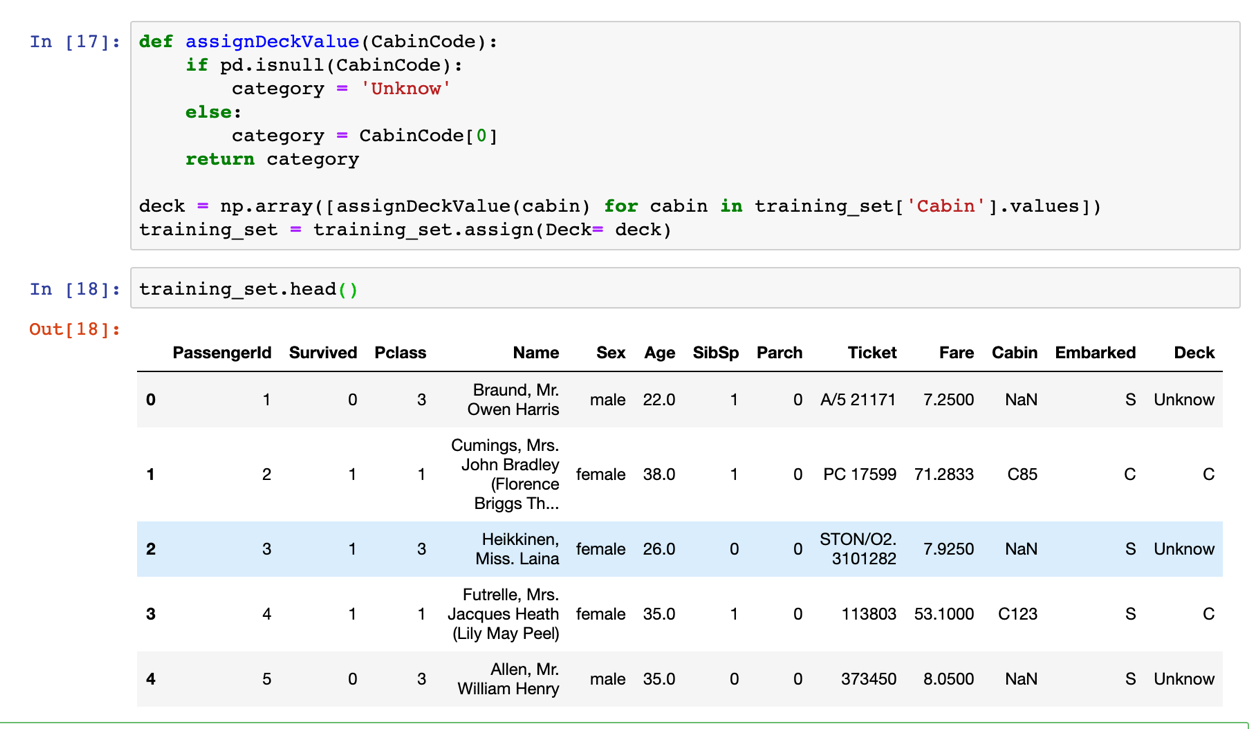
Si se quiere averiguar el numero de valores nulos por columna se debe llamar a la función isnull y luego sumarizar todos los valores con la función sum



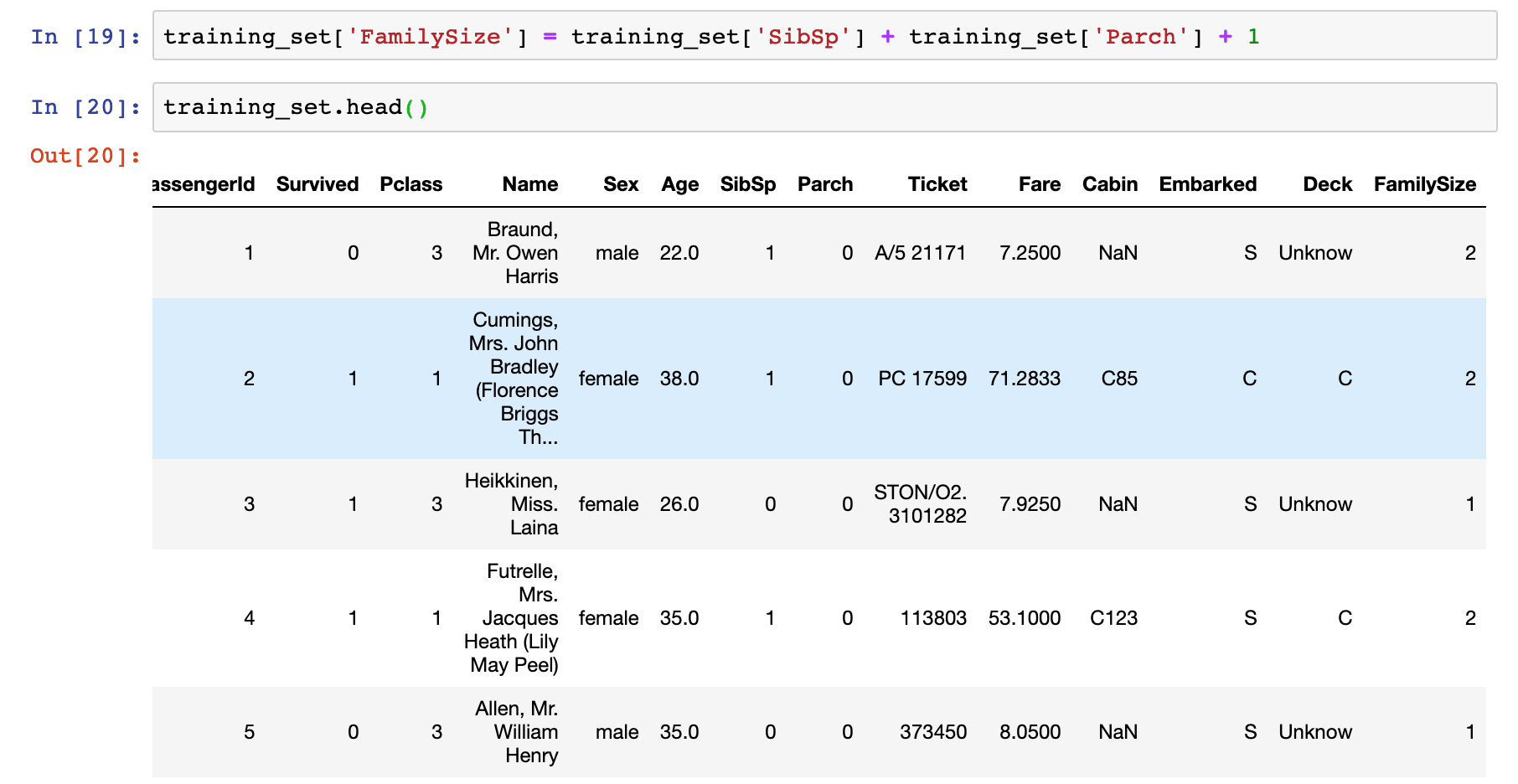
En la imagen anterior se puede apreciar que en la columna Age hay 177 registros nulos de la misma forma en la columna Cabin 687.

# Preparando el conjunto de datos

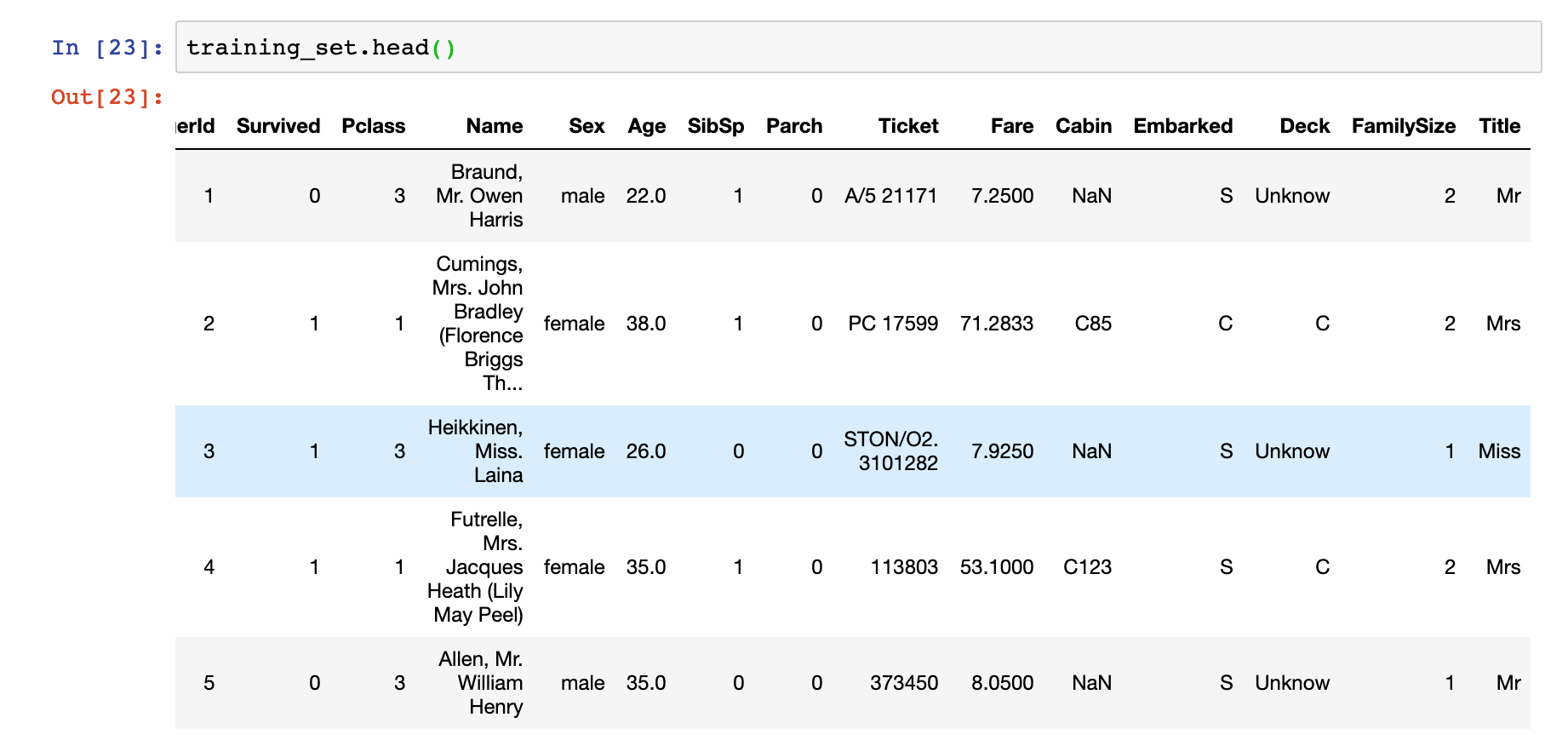
Se deben crear nuevas clases, en este caso se creará una nueva columna Deck, la cual nos indica si la columna



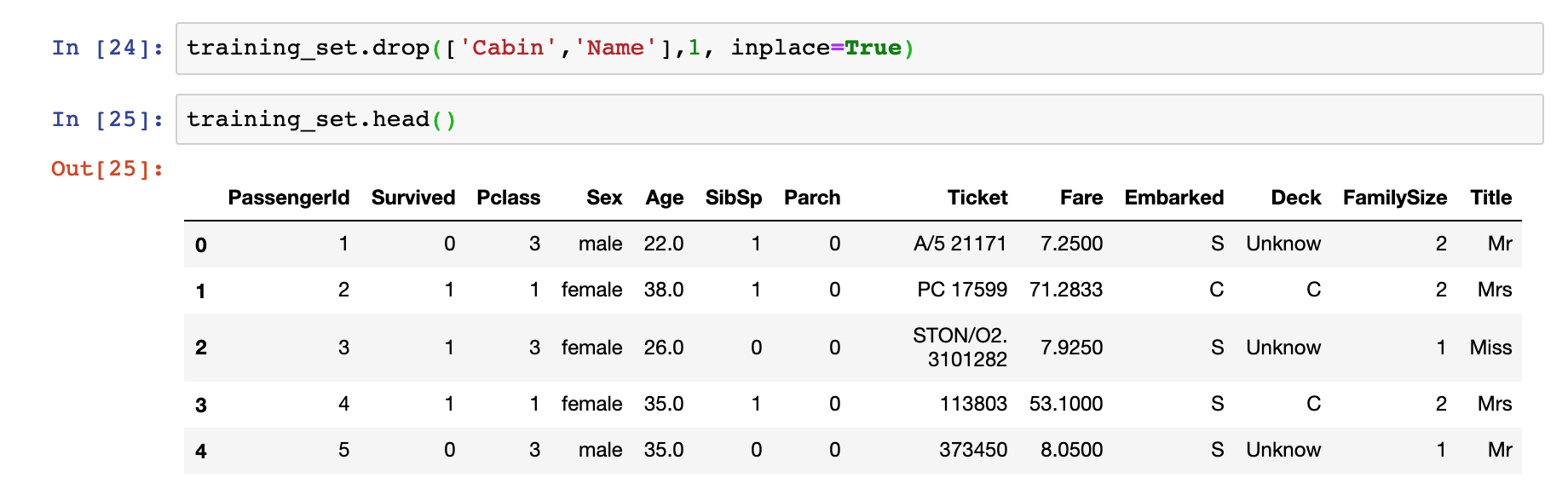
Ahora se procede a crear la clase Family size, la cual indicará el número de hermanos y padres en conjunto.



Ahora se creará la clase Title la cual indicara si es Sr, Ms, entre otros.

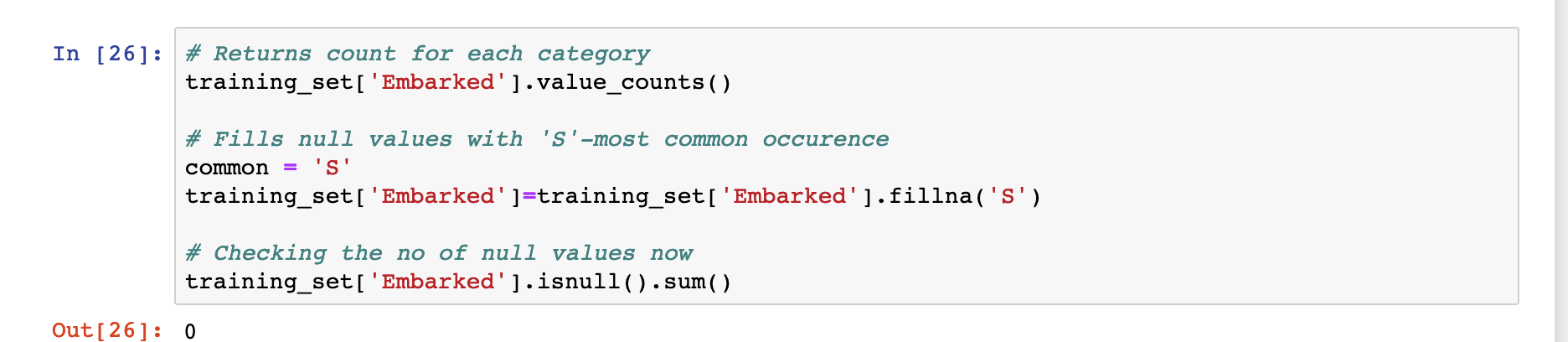


Ahora que se proceso la información como se requiere se proce a borrar las columnas Cabin y Name

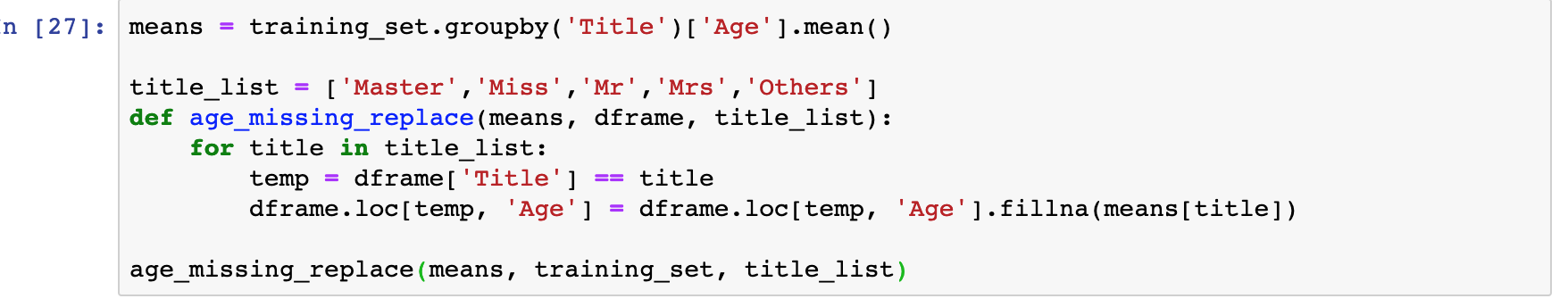


# Controlando valores Faltantes

Para la clase Embarked, faltan altunos valores, los valores que puede tomar esta son C=Cherbourg, Q=Queenstown, S = Southampton, en este caso se procede para los valores faltantes asignar S que es el mas comun. Luego se realiza una contabilidad para verificar que no hay valores nulos para esta columna.

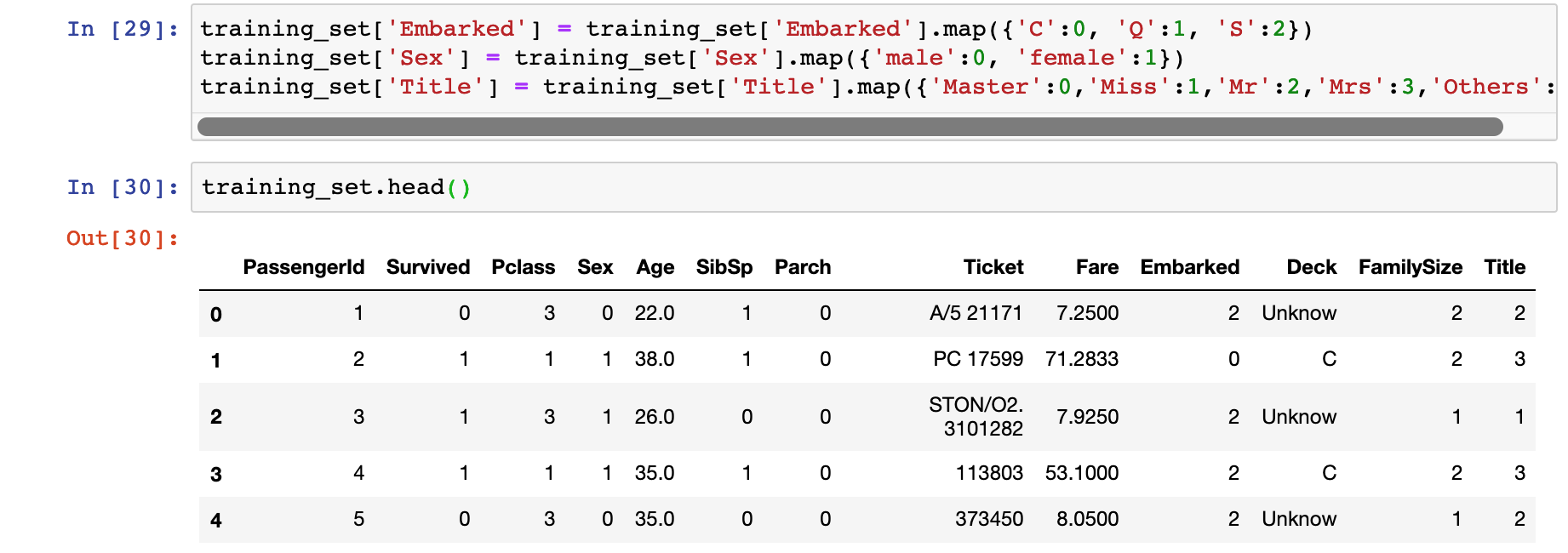


Para la clase Age, se realiza un procedimiento similar, estableciendo el valor medio a partir del titulo.

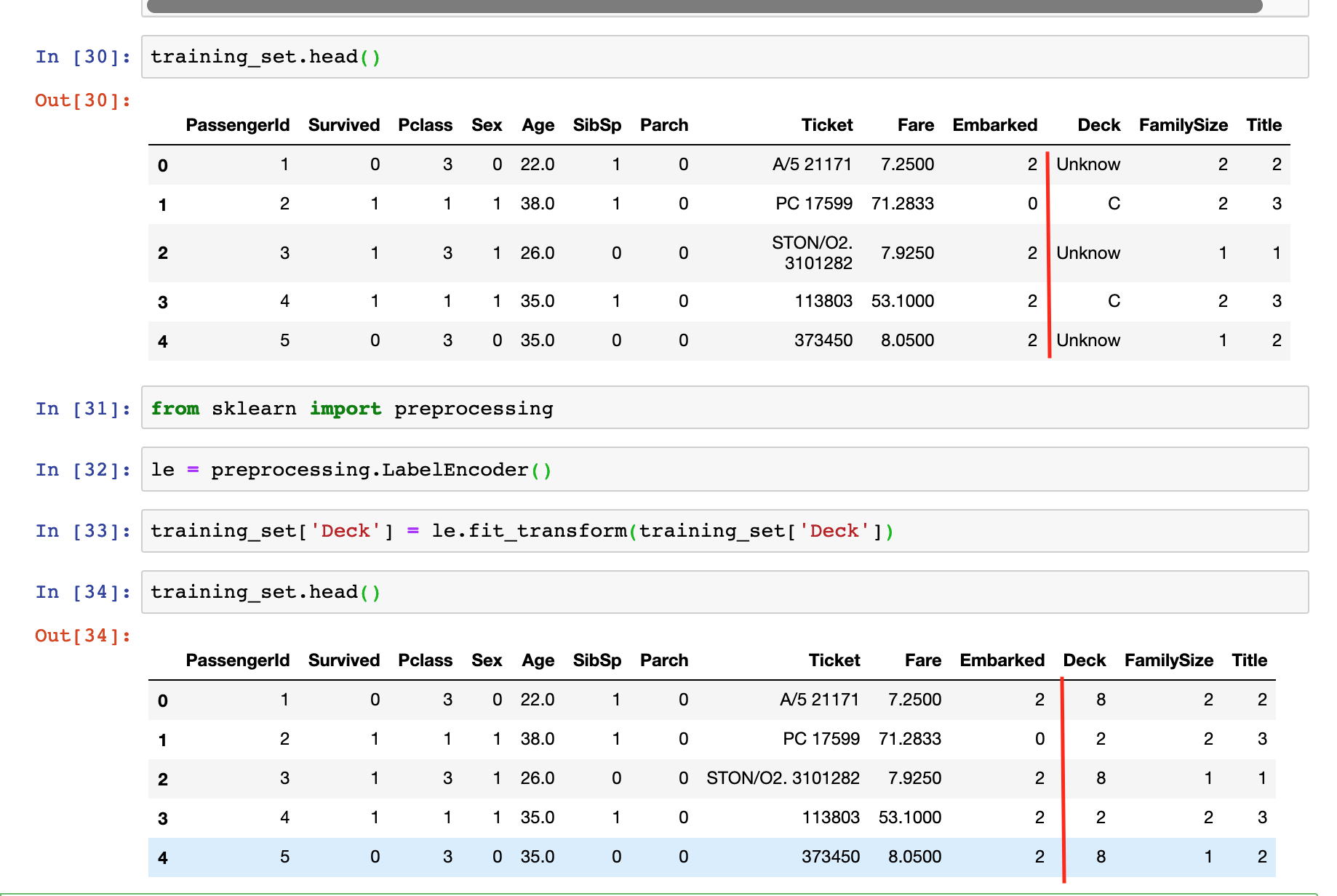


# Codificación de características categóricas.

Muchos algoritmos de aprendizaje automático no pueden admitir valores categóricos sin convertirse en valores numéricos. Para lo cual se encarga de usar la función map la cual asignara un valor numeric a cada columna dependiendo cómo se haya definido.



Asignar valores de forma manual, no es la mejor opción ya que se puede dejar datos de forma inconsistente y el tiempo empleado puede ser alto depende del volumen de datos a procesar. En este caso se empleara LabelEncoder de sklear para realizar la conversión de la columna Deck. En la siguiente imagen se aprecia la asignación de valores numéricos por categoría de forma aleatoria a la columna Deck.



Después de realizar este paso ya se encuentra lista la información sin ningún valor faltante.

