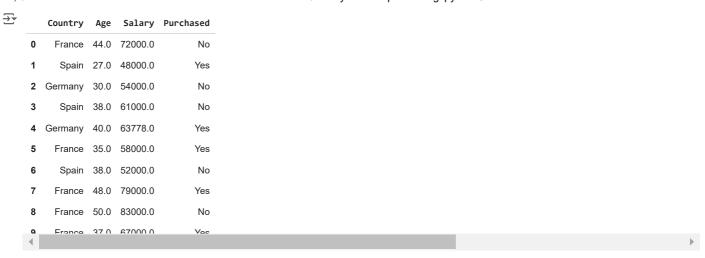
df

```
import numpy as np
import pandas as pd
df=pd.read_csv("/content/pre-process_datasample.csv")
₹
         Country
                  Age
                       Salary Purchased
     0
                      72000 0
          France 44.0
                                      Nο
     1
           Spain 27.0
                      48000.0
                                     Yes
     2 Germany
                 30.0
                      54000.0
                                      No
     3
                 38.0
                      61000.0
                                      No
           Spain
        Germany
                 40.0
                         NaN
                                     Yes
                      58000.0
     5
          France
                 35.0
                                     Yes
     6
           Spain NaN 52000.0
                                      No
     7
          France
                 48.0
                      79000.0
                                     Yes
     8
                 50.0
                      83000.0
            NaN
                                      No
          France 37.0 67000.0
Double-click (or enter) to edit
df.info()
<pr
    RangeIndex: 10 entries, 0 to 9
    Data columns (total 4 columns):
     # Column
                    Non-Null Count Dtype
         Country
                    9 non-null
                                    object
                    9 non-null
                                    float64
         Age
         Salary
                    9 non-null
                                    float64
         Purchased 10 non-null
                                    object
    dtypes: float64(2), object(2)
    memory usage: 448.0+ bytes
df.Country.mode()
₹
        Country
     0
        France
    4
df.Country.mode()[0]
    4
type(df.Country.mode())
<del>_</del>
      pandas.core.series.Series
      def __init__(data=None, index=None, dtype: Dtype | None=None, name=None, copy: bool | None=None,
      fastpath: bool=False) -> None
          THUEN IS HOL MOHE, the resulting series is rethuened with the thuen values.
      dtype : str, numpy.dtype, or ExtensionDtype, optional
          Data type for the output Series. If not specified, this will be
          inferred from `data`.
          See the :ref:`user guide <basics.dtypes>` for more usages.
      name : Hashable, default None
df.Country.fillna(df.Country.mode()[0],inplace=True)
df.Age.fillna(df.Age.median(),inplace=True)
df.Salary.fillna(round(df.Salary.mean()),inplace=True)
```



pd.get\_dummies(df.Country)



updated\_dataset=pd.concat([pd.get\_dummies(df.Country),df.iloc[:,[1,2,3]]],axis=1)

updated\_dataset

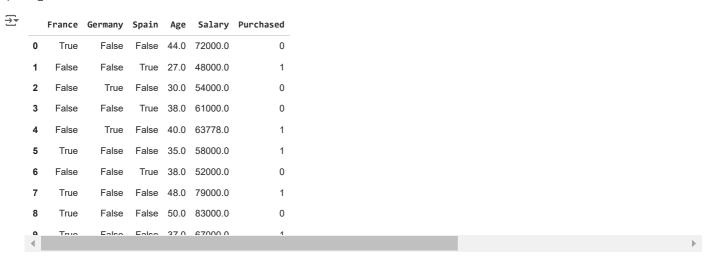


df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10 entries, 0 to 9
Data columns (total 4 columns):
                Non-Null Count Dtype
 # Column
                10 non-null
                                object
     Country
                10 non-null
                                float64
     Age
                                float64
     Salary
                10 non-null
     Purchased 10 non-null
                                object
dtypes: float64(2), object(2)
memory usage: 448.0+ bytes
```

updated\_dataset.Purchased.replace(['No','Yes'],[0,1],inplace=True)

updated\_dataset



Start coding or generate with AI.