```python
import numpy as np
import pandas as pd
df=pd.read_csv("/content/pre-process_datasample.csv")
df
```

|   | Country | Age | Salary | Purchased |
|---|---------|-----|--------|-----------|
| 0 | France | 44.0 | 72000.0 | No |
| 1 | Spain | 27.0 | 48000.0 | Yes |
| 2 | Germany | 30.0 | 54000.0 | No |
| 3 | Spain | 38.0 | 61000.0 | No |
| 4 | Germany | 40.0 | NaN | Yes |
| 5 | France | 35.0 | 58000.0 | Yes |
| 6 | Spain | NaN | 52000.0 | No |
| 7 | France | 48.0 | 79000.0 | Yes |
| 8 | NaN | 50.0 | 83000.0 | No |
| 9 | France | 37.0 | 67000.0 | Yes |

Double-click (or enter) to edit

```python
df.info()
```

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

```python
df.Country.mode()
```

|   | Country |
|---|---------|
| 0 | France |

```python
df.Country.mode()[0]
```

```python
type(df.Country.mode())
```

```
pandas.core.series.Series
def __init__(data=None, index=None, dtype: Dtype | None=None, name=None, copy: bool | None=None,
fastpath: bool=False) -> None
    index is not None, the resulting Series is reindexed with the index 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
    The name to give to the Series.
```

```python
df.Country.fillna(df.Country.mode()[0],inplace=True)
```

```python
df.Age.fillna(df.Age.median(),inplace=True)
```

```python
df.Salary.fillna(round(df.Salary.mean()),inplace=True)
```

```python
df
```

|   | Country | Age | Salary | Purchased |
|---|---------|-----|--------|-----------|
| 0 | France | 44.0 | 72000.0 | No |
| 1 | Spain | 27.0 | 48000.0 | Yes |
| 2 | Germany | 30.0 | 54000.0 | No |
| 3 | Spain | 38.0 | 61000.0 | No |
| 4 | Germany | 40.0 | 63778.0 | Yes |
| 5 | France | 35.0 | 58000.0 | Yes |
| 6 | Spain | 38.0 | 52000.0 | No |
| 7 | France | 48.0 | 79000.0 | Yes |
| 8 | France | 50.0 | 83000.0 | No |
| 9 | France | 37.0 | 67000.0 | Yes |

```python
pd.get_dummies(df.Country)
```

|   | France | Germany | Spain |
|---|--------|---------|-------|
| 0 | True | False | False |
| 1 | False | False | True |
| 2 | False | True | False |
| 3 | False | False | True |
| 4 | False | True | False |
| 5 | True | False | False |
| 6 | False | False | True |
| 7 | True | False | False |
| 8 | True | False | False |
| 9 | True | False | False |

```python
updated_dataset=pd.concat([pd.get_dummies(df.Country),df.iloc[:,[1,2,3]]],axis=1)
```

```python
updated_dataset
```

|   | France | Germany | Spain | Age | Salary | Purchased |
|---|--------|---------|-------|-----|--------|-----------|
| 0 | True | False | False | 44.0 | 72000.0 | No |
| 1 | False | False | True | 27.0 | 48000.0 | Yes |
| 2 | False | True | False | 30.0 | 54000.0 | No |
| 3 | False | False | True | 38.0 | 61000.0 | No |
| 4 | False | True | False | 40.0 | 63778.0 | Yes |
| 5 | True | False | False | 35.0 | 58000.0 | Yes |
| 6 | False | False | True | 38.0 | 52000.0 | No |
| 7 | True | False | False | 48.0 | 79000.0 | Yes |
| 8 | True | False | False | 50.0 | 83000.0 | No |
| 9 | True | False | False | 37.0 | 67000.0 | Yes |

```python
df.info()
```

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

```python
updated_dataset.Purchased.replace(['No','Yes'],[0,1],inplace=True)
```

updated_dataset

| | France | Germany | Spain | Age | Salary | Purchased |
|---|---|---|---|---|---|---|
| 0 | True | False | False | 44.0 | 72000.0 | 0 |
| 1 | False | False | True | 27.0 | 48000.0 | 1 |
| 2 | False | True | False | 30.0 | 54000.0 | 0 |
| 3 | False | False | True | 38.0 | 61000.0 | 0 |
| 4 | False | True | False | 40.0 | 63778.0 | 1 |
| 5 | True | False | False | 35.0 | 58000.0 | 1 |
| 6 | False | False | True | 38.0 | 52000.0 | 0 |
| 7 | True | False | False | 48.0 | 79000.0 | 1 |
| 8 | True | False | False | 50.0 | 83000.0 | 0 |
| 9 | True | False | False | 37.0 | 67000.0 | 1 |

Start coding or generate with AI.