```python
import numpy as np
import pandas as pd
df=pd.read_csv("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  Germany  50.0   83000.0       No
9   France  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        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()
```

```
0    France
dtype: object
```

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

```
'France'
```

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

```
pandas.core.series.Series
```

```python
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)
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  Germany  50.0  83000.0       No
9   France  37.0  67000.0      Yes
```

pd.get_dummies(df.Country)

```
   France  Germany  Spain
0       1        0      0
1       0        0      1
2       0        1      0
3       0        0      1
4       0        1      0
5       1        0      0
6       0        0      1
7       1        0      0
8       0        1      0
9       1        0      0
```

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

updated_dataset

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

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

updated_dataset
```

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