

In [20]:

```
#import libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
from sklearn import preprocessing
from sklearn import preprocessing
warnings.filterwarnings('ignore')
```

In [2]:

```
dataset = pd.read_csv("C:\\Users\\admin\\Desktop\\purchase.csv")
```

In [3]:

```
dataset.shape
```

Out[3]:

```
(27, 4)
```

In [5]:

```
dataset.isnull()
```

Out[5]:

	Country	Age	Salary	Purchased
0	False	False	False	False
1	False	False	False	False
2	False	False	False	False
3	False	True	True	False
4	False	False	False	False
5	False	True	False	False
6	False	False	True	False
7	False	False	False	False
8	False	False	False	False
9	False	False	False	False
10	False	False	False	False
11	False	False	False	False
12	False	True	True	False
13	False	False	False	False
14	False	True	False	False
15	False	False	True	False
16	False	False	False	False
17	False	False	False	False
18	False	False	False	False
19	False	False	False	False
20	False	False	False	False
21	False	True	True	False
22	False	False	False	False
23	False	True	False	False
24	False	False	True	False
25	False	False	False	False
26	False	False	False	False

In [6]:

```
dataset.isnull().sum()
```

Out[6]:

```
Country      0  
Age          6  
Salary       6  
Purchased    0  
dtype: int64
```

In [7]:

```
dataset["Age"].fillna("30",inplace=True)
```

In [8]:

```
dataset
```

Out[8]:

	Country	Age	Salary	Purchased
0	France	44	72000.0	No
1	Spain	56	48000.0	Yes
2	Germany	34	54000.0	No
3	France	30	NaN	No
4	Germany	44	58000.0	Yes
5	France	30	61000.0	Yes
6	Spain	35	NaN	No
7	France	50	79000.0	Yes
8	Germany	37	83000.0	No
9	France	44	72000.0	No
10	Spain	56	48000.0	Yes
11	Germany	34	54000.0	No
12	France	30	NaN	No
13	Germany	44	58000.0	Yes
14	France	30	61000.0	Yes
15	Spain	35	NaN	No
16	France	50	79000.0	Yes
17	Germany	37	83000.0	No
18	France	44	72000.0	No
19	Spain	56	48000.0	Yes
20	Germany	34	54000.0	No
21	France	30	NaN	No
22	Germany	44	58000.0	Yes
23	France	30	61000.0	Yes
24	Spain	35	NaN	No
25	France	50	79000.0	Yes
26	Germany	37	83000.0	No

In [9]:

```
dataset["Salary"].fillna("0.0",inplace=True)
```

In [10]:

```
dataset
```

Out[10]:

	Country	Age	Salary	Purchased
0	France	44	72000	No
1	Spain	56	48000	Yes
2	Germany	34	54000	No
3	France	30	0.0	No
4	Germany	44	58000	Yes
5	France	30	61000	Yes
6	Spain	35	0.0	No
7	France	50	79000	Yes
8	Germany	37	83000	No
9	France	44	72000	No
10	Spain	56	48000	Yes
11	Germany	34	54000	No
12	France	30	0.0	No
13	Germany	44	58000	Yes
14	France	30	61000	Yes
15	Spain	35	0.0	No
16	France	50	79000	Yes
17	Germany	37	83000	No
18	France	44	72000	No
19	Spain	56	48000	Yes
20	Germany	34	54000	No
21	France	30	0.0	No
22	Germany	44	58000	Yes
23	France	30	61000	Yes
24	Spain	35	0.0	No
25	France	50	79000	Yes
26	Germany	37	83000	No

In [13]:

```
dataset.isnull().sum()
```

Out[13]:

```
Country      0
Age          0
Salary       0
Purchased    0
dtype: int64
```

In [24]:

```
#import libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
from sklearn import preprocessing
from sklearn import preprocessing

X=dataset.iloc[:, :-1].values
X
```

Out[24]:

```
array([[0, 44.0, 72000.0],
       [2, 56.0, 48000.0],
       [1, 34.0, 54000.0],
       [0, '30', '0.0'],
       [1, 44.0, 58000.0],
       [0, '30', 61000.0],
       [2, 35.0, '0.0'],
       [0, 50.0, 79000.0],
       [1, 37.0, 83000.0],
       [0, 44.0, 72000.0],
       [2, 56.0, 48000.0],
       [1, 34.0, 54000.0],
       [0, '30', '0.0'],
       [1, 44.0, 58000.0],
       [0, '30', 61000.0],
       [2, 35.0, '0.0'],
       [0, 50.0, 79000.0],
       [1, 37.0, 83000.0],
       [0, 44.0, 72000.0],
       [2, 56.0, 48000.0],
       [1, 34.0, 54000.0],
       [0, '30', '0.0'],
       [1, 44.0, 58000.0],
       [0, '30', 61000.0],
       [2, 35.0, '0.0'],
       [0, 50.0, 79000.0],
       [1, 37.0, 83000.0]], dtype=object)
```

In [25]:

```
label_encoder = preprocessing.LabelEncoder()
X[:,0] = label_encoder.fit_transform(X[:,0])
X
```

Out[25]:

```
array([[0, 44.0, 72000.0],
       [2, 56.0, 48000.0],
       [1, 34.0, 54000.0],
       [0, '30', '0.0'],
       [1, 44.0, 58000.0],
       [0, '30', 61000.0],
       [2, 35.0, '0.0'],
       [0, 50.0, 79000.0],
       [1, 37.0, 83000.0],
       [0, 44.0, 72000.0],
       [2, 56.0, 48000.0],
       [1, 34.0, 54000.0],
       [0, '30', '0.0'],
       [1, 44.0, 58000.0],
       [0, '30', 61000.0],
       [2, 35.0, '0.0'],
       [0, 50.0, 79000.0],
       [1, 37.0, 83000.0],
       [0, 44.0, 72000.0],
       [2, 56.0, 48000.0],
       [1, 34.0, 54000.0],
       [0, '30', '0.0'],
       [1, 44.0, 58000.0],
       [0, '30', 61000.0],
       [2, 35.0, '0.0'],
       [0, 50.0, 79000.0],
       [1, 37.0, 83000.0]], dtype=object)
```

In [26]:

```
onehotencoder = OneHotEncoder(categorical_features=[0])
X=onehotencoder.fit_transform(X)
X
dummy=pd.DataFrame(X.toarray())
dummy.columns=['France', 'Germany', 'Spain', '3', '4']
dummy = dummy.iloc[:, :-2]
dummy=dummy.astype(int)
dummy
```

**NameError**

Traceback (most recent call last)

&lt;ipython-input-26-5cb8a56bd257&gt; in &lt;module&gt;

```
----> 1 onehotencoder = OneHotEncoder(categorical_features=[0])
      2 X=onehotencoder.fit_transform(X)
      3 X
      4 dummy=pd.DataFrame(X.toarray())
      5 dummy.columns=['France', 'Germany', 'Spain', '3', '4']
```

**NameError:** name 'OneHotEncoder' is not defined

In [ ]: