Importing the libraries

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

Importing the dataset

```
df = 
pd.read_csv('https://github.com/YBI-Foundation/Dataset/raw/main/Custom
er%20Purchase.csv')
```

Get info of dataframe

```
df.head()
```

```
   Customer ID  Age  Gender Education   Review Purchased
0         1021   30  Female    School  Average        No
1         1022   68  Female        UG     Poor        No
2         1023   70  Female        PG     Good        No
3         1024   72  Female        PG     Good        No
4         1025   16  Female        UG  Average        No
```

```
df.shape
```

```
(50, 6)
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50 entries, 0 to 49
Data columns (total 6 columns):
 #   Column       Non-Null Count  Dtype 
---  ------       --------------  ----- 
 0   Customer ID  50 non-null     int64 
 1   Age          50 non-null     int64 
 2   Gender       50 non-null     object
 3   Education    50 non-null     object
 4   Review       50 non-null     object
 5   Purchased    50 non-null     object
dtypes: int64(2), object(4)
memory usage: 2.5+ KB
```

```
df.columns
```

```
Index(['Customer ID', 'Age', 'Gender', 'Education', 'Review', 
'Purchased'], dtype='object')
```

get target variable seperated from dataframe

```
y=df['Purchased']
y
```

```
0      No
1      No
2      No
3      No
4      No
5     Yes
6      No
7     Yes
8      No
9     Yes
10    Yes
11    Yes
12     No
13     No
14    Yes
15     No
16    Yes
17    Yes
18     No
19    Yes
20    Yes
21     No
22    Yes
23     No
24    Yes
25     No
26     No
27     No
28     No
29    Yes
30     No
31    Yes
32    Yes
33    Yes
34     No
35    Yes
36    Yes
37    Yes
38     No
39     No
40     No
41    Yes
42    Yes
43     No
44     No
45    Yes
46     No
47    Yes
48    Yes
```

```
49     No
Name: Purchased, dtype: object
```

Encoding the Dependent Variable

```
#We used labelencoder for y is coz we need only single column in
result so that it can be predicted
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
y = le.fit_transform(y)
y
```

```
array([0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 1, 1, 0, 0, 1, 0, 1, 1, 0, 1, 1,
0,
       1, 0, 1, 0, 0, 0, 0, 1, 0, 1, 1, 1, 0, 1, 1, 1, 0, 0, 0, 1, 1,
0,
       0, 1, 0, 1, 1, 0])
```

Display Eencoded Categories

```
le.classes_
```

```
array(['No', 'Yes'], dtype=object)
```

Reverse integer array to categorical data

```
le.inverse_transform([0,1,1])
```

```
array(['No', 'Yes', 'Yes'], dtype=object)
```