

## Importing the libraries

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

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

## Get information on 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.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.isnull().sum()
```

```
Customer ID    0
Age            0
Gender          0
Education       0
Review          0
Purchased       0
dtype: int64
```

```
df.columns
```

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

```
X=df[['Education', 'Review', 'Gender']]
```

## One Hot Encoder

```
from sklearn.preprocessing import OneHotEncoder  
oe=OneHotEncoder()
```

```
X_encoder= oe.fit_transform(X)
```

```
X
```

	Education	Review	Gender
0	School	Average	Female
1	UG	Poor	Female
2	PG	Good	Female
3	PG	Good	Female
4	UG	Average	Female
5	School	Average	Female
6	School	Good	Male
7	School	Poor	Female
8	UG	Average	Female
9	UG	Good	Male
10	UG	Good	Female
11	UG	Good	Male
12	School	Poor	Male
13	School	Average	Female
14	PG	Poor	Male
15	UG	Poor	Male
16	UG	Poor	Male
17	UG	Poor	Female
18	School	Good	Male
19	PG	Poor	Male
20	School	Average	Female
21	PG	Average	Male
22	PG	Poor	Female
23	School	Good	Female
24	PG	Average	Female
25	School	Good	Female
26	PG	Poor	Female
27	PG	Poor	Female
28	School	Poor	Male
29	UG	Average	Female
30	UG	Average	Male
31	School	Poor	Female
32	UG	Average	Male
33	PG	Good	Female
34	School	Average	Male
35	School	Poor	Male
36	UG	Good	Female
37	PG	Average	Male

38	School	Good	Female
39	PG	Poor	Male
40	School	Good	Male
41	PG	Good	Male
42	PG	Good	Female
43	PG	Poor	Male
44	UG	Average	Female
45	PG	Poor	Male
46	PG	Poor	Female
47	PG	Good	Female
48	UG	Good	Female
49	UG	Good	Female

X\_encoder.toarray()

```
[0., 0., 1., 0., 1., 0., 1., 0.],
[1., 0., 0., 1., 0., 0., 0., 1.],
[0., 1., 0., 0., 1., 0., 1., 0.],
[1., 0., 0., 0., 0., 1., 0., 1.],
[0., 1., 0., 0., 1., 0., 0., 1.],
[1., 0., 0., 0., 1., 0., 0., 1.],
[1., 0., 0., 0., 1., 0., 1., 0.],
[1., 0., 0., 0., 0., 1., 0., 1.],
[0., 0., 1., 1., 0., 0., 1., 0.],
[1., 0., 0., 0., 0., 1., 0., 1.],
[1., 0., 0., 0., 0., 1., 1., 0.],
[1., 0., 0., 0., 1., 0., 1., 0.],
[0., 0., 1., 0., 1., 0., 1., 0.],
[0., 0., 1., 0., 1., 0., 1., 0.])
```

```
X_encoder.shape
```

```
(50, 8)
```

## reverse encoding

```
oe.categories_
```

```
[array(['PG', 'School', 'UG'], dtype=object),
 array(['Average', 'Good', 'Poor'], dtype=object),
 array(['Female', 'Male'], dtype=object)]
```

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

```
array([['School', 'Average', 'Female']], dtype=object)
```

## Dropping first dummy Variable to avoid dummy trap

```
X=df[['Gender', 'Education', 'Review']]
```

```
ohe=OneHotEncoder(drop='first')
```

```
X=ohe.fit_transform(X)
```

```
ohe.categories_
```

```
[array(['Female', 'Male'], dtype=object),
 array(['PG', 'School', 'UG'], dtype=object),
 array(['Average', 'Good', 'Poor'], dtype=object)]
```

```
X.shape
```

```
(50, 5)
```

```
X.toarray()
```

