In [1]:
```
import pandas as pd
```

In [2]:
```
data=pd.read_csv(r"C:\Users\DELL\Downloads\ML Feature Encoding Resource16931320640.csv")
```

In [3]:
```
data
```

Out[3]:

	age	workclass	fnlwgt	education	educational-num	marital-status	occupation	relationship	race	gender	capital-gain	capital-loss	hours-per-week	c
0	25	Private	226802	11th	7	Never-married	Machine-op-inspct	Own-child	Black	Male	0	0	40	U
1	38	Private	89814	HS-grad	9	Married-civ-spouse	Farming-fishing	Husband	White	Male	0	0	50	U
2	28	Local-gov	336951	Assoc-acdm	12	Married-civ-spouse	Protective-serv	Husband	White	Male	0	0	40	U
3	44	Private	160323	Some-college	10	Married-civ-spouse	Machine-op-inspct	Husband	Black	Male	7688	0	40	U
4	18	?	103497	Some-college	10	Never-married	?	Own-child	White	Female	0	0	30	U
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
48837	27	Private	257302	Assoc-acdm	12	Married-civ-spouse	Tech-support	Wife	White	Female	0	0	38	U
48838	40	Private	154374	HS-grad	9	Married-civ-spouse	Machine-op-inspct	Husband	White	Male	0	0	40	U
48839	58	Private	151910	HS-grad	9	Widowed	Adm-clerical	Unmarried	White	Female	0	0	40	U
48840	22	Private	201490	HS-grad	9	Never-married	Adm-clerical	Own-child	White	Male	0	0	20	U
48841	52	Self-emp-inc	287927	HS-grad	9	Married-civ-spouse	Exec-managerial	Wife	White	Female	15024	0	40	U

48842 rows × 15 columns

In [4]:
```
data["income"].unique()
```

Out[4]:
```
array(['<=50K', '>50K'], dtype=object)
```

In [14]:
```
data["income"].replace(['<=50K', '>50K'],[0,1],inplace=True)
```

In [15]:
```
data
```

Out[15]:

	age	workclass	fnlwgt	education	educational-num	marital-status	occupation	relationship	race	gender	capital-gain	capital-loss	hours-per-week	c
0	25	Private	226802	11th	7	Never-married	Machine-op-inspct	Own-child	Black	Male	0	0	40	U
1	38	Private	89814	HS-grad	9	Married-civ-spouse	Farming-fishing	Husband	White	Male	0	0	50	U
2	28	Local-gov	336951	Assoc-acdm	12	Married-civ-spouse	Protective-serv	Husband	White	Male	0	0	40	U
3	44	Private	160323	Some-college	10	Married-civ-spouse	Machine-op-inspct	Husband	Black	Male	7688	0	40	U
4	18	?	103497	Some-college	10	Never-married	?	Own-child	White	Female	0	0	30	U
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
48837	27	Private	257302	Assoc-acdm	12	Married-civ-spouse	Tech-support	Wife	White	Female	0	0	38	U
48838	40	Private	154374	HS-grad	9	Married-civ-spouse	Machine-op-inspct	Husband	White	Male	0	0	40	U
48839	58	Private	151910	HS-grad	9	Widowed	Adm-clerical	Unmarried	White	Female	0	0	40	U
48840	22	Private	201490	HS-grad	9	Never-married	Adm-clerical	Own-child	White	Male	0	0	20	U
48841	52	Self-emp-inc	287927	HS-grad	9	Married-civ-spouse	Exec-managerial	Wife	White	Female	15024	0	40	U

48842 rows × 15 columns

In [16]:
```
data["income"].replace([0,1],['<=50K', '>50K'],inplace=True)
```

In [17]:
```
data
```

Out[17]:

	age	workclass	fnlwgt	education	educational-num	marital-status	occupation	relationship	race	gender	capital-gain	capital-loss	hours-per-week	c
0	25	Private	226802	11th	7	Never-married	Machine-op-inspct	Own-child	Black	Male	0	0	40	U
1	38	Private	89814	HS-grad	9	Married-civ-spouse	Farming-fishing	Husband	White	Male	0	0	50	U
2	28	Local-gov	336951	Assoc-acdm	12	Married-civ-spouse	Protective-serv	Husband	White	Male	0	0	40	U
3	44	Private	160323	Some-college	10	Married-civ-spouse	Machine-op-inspct	Husband	Black	Male	7688	0	40	U
4	18	?	103497	Some-college	10	Never-married	?	Own-child	White	Female	0	0	30	U
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
48837	27	Private	257302	Assoc-acdm	12	Married-civ-spouse	Tech-support	Wife	White	Female	0	0	38	U
48838	40	Private	154374	HS-grad	9	Married-civ-spouse	Machine-op-inspct	Husband	White	Male	0	0	40	U
48839	58	Private	151910	HS-grad	9	Widowed	Adm-clerical	Unmarried	White	Female	0	0	40	U
48840	22	Private	201490	HS-grad	9	Never-married	Adm-clerical	Own-child	White	Male	0	0	20	U
48841	52	Self-emp-inc	287927	HS-grad	9	Married-civ-spouse	Exec-managerial	Wife	White	Female	15024	0	40	U

48842 rows × 15 columns

In [18]:
```
data.head(2)
```

Out[18]:

	age	workclass	fnlwgt	education	educational-num	marital-status	occupation	relationship	race	gender	capital-gain	capital-loss	hours-per-week	native country
0	25	Private	226802	11th	7	Never-married	Machine-op-inspct	Own-child	Black	Male	0	0	40	United State
1	38	Private	89814	HS-grad	9	Married-civ-spouse	Farming-fishing	Husband	White	Male	0	0	50	United State

In [11]:
```
data["workclass"].unique()
```

Out[11]:
```
array(['Private', 'Local-gov', '?', 'Self-emp-not-inc', 'Federal-gov',
       'State-gov', 'Self-emp-inc', 'Without-pay', 'Never-worked'],
      dtype=object)
```

In [13]:
```
from sklearn.preprocessing import LabelEncoder
```

In [20]:
```
encoder=LabelEncoder()
```

In [21]:
```
a=encoder.fit_transform(data["workclass"])
```

In [22]:
```
encoder.inverse_transform(a)
```

Out[22]:
```
array(['Private', 'Private', 'Local-gov', ..., 'Private', 'Private',
       'Self-emp-inc'], dtype=object)
```

In [19]:
```
data["gender"].unique()
```

Out[19]:
```
array(['Male', 'Female'], dtype=object)
```

In [23]:
```
z=pd.get_dummies(data["gender"])
```

In [24]:
```
z
```

Out[24]:

	Female	Male
0	False	True
1	False	True
2	False	True
3	False	True
4	True	False
...	...	...
48837	True	False
48838	False	True
48839	True	False
48840	False	True
48841	True	False

48842 rows × 2 columns

In [25]:
```
z=pd.from_dummies(data["gender"])
```

```
---------------------------------------------------------------------------
TypeError                                 Traceback (most recent call last)
Cell In[25], line 1
----> 1 z=pd.from_dummies(data["gender"])

File ~\anaconda3\Lib\site-packages\pandas\core\reshape\encoding.py:450, in from_dummies(data, sep, default_
category)
    447 from pandas.core.reshape.concat import concat
    449 if not isinstance(data, DataFrame):
--> 450     raise TypeError(
    451         "Expected 'data' to be a 'DataFrame'; "
    452         f"Received 'data' of type: {type(data).__name__}"
    453     )
    455 if data.isna().any().any():
    456     raise ValueError(
    457         "Dummy DataFrame contains NA value in column: "
    458         f"'{data.isna().any().idxmax()}'"
    459     )

TypeError: Expected 'data' to be a 'DataFrame'; Received 'data' of type: Series
```

In [26]:
```
z.idxmax(axis=1)
```

Out[26]:
```
0          Male
1          Male
2          Male
3          Male
4        Female
          ...
48837    Female
48838      Male
48839    Female
48840      Male
48841    Female
Length: 48842, dtype: object
```

In [27]:
```
from sklearn.preprocessing import OneHotEncoder
```

In [28]:
```
encoder1=OneHotEncoder()
```

In [31]:
```
df=pd.DataFrame(encoder1.fit_transform(data[["gender"]]))
```

In [32]:
```
df
```

Out[32]:

	0
0	(0, 1)\t1.0
1	(0, 1)\t1.0
2	(0, 1)\t1.0
3	(0, 1)\t1.0
4	(0, 0)\t1.0
...	...
48837	(0, 0)\t1.0
48838	(0, 1)\t1.0
48839	(0, 0)\t1.0
48840	(0, 1)\t1.0
48841	(0, 0)\t1.0

48842 rows × 1 columns

In [35]:
```
b=encoder1.fit_transform(data[["gender"]]).toarray()
```

In [36]:
```
encoder1.inverse_transform(b)
```

Out[36]:
```
array([['Male'],
       ['Male'],
       ['Male'],
       ...,
       ['Female'],
       ['Male'],
       ['Female']], dtype=object)
```

In [37]:
```
b
```

Out[37]:
```
array([[0., 1.],
       [0., 1.],
       [0., 1.],
       ...,
       [1., 0.],
       [0., 1.],
       [1., 0.]])
```

In [ ]:
```
pd.DataFrame
```