

```
In [2]: import numpy as np
import pandas as pd
```

```
In [3]: dataset = pd.read_csv("bank.csv")
```

```
In [4]: dataset
```

```
Out[4]:
```

	age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous
0	59	admin.	married	secondary	no	2343	yes	no	unknown	5	may	1042	1	-1	0
1	56	admin.	married	secondary	no	45	no	no	unknown	5	may	1467	1	-1	0
2	41	technician	married	secondary	no	1270	yes	no	unknown	5	may	1389	1	-1	0
3	55	services	married	secondary	no	2476	yes	no	unknown	5	may	579	1	-1	0
4	54	admin.	married	tertiary	no	184	no	no	unknown	5	may	673	2	-1	0
...
11157	33	blue-collar	single	primary	no	1	yes	no	cellular	20	apr	257	1	-1	0
11158	39	services	married	secondary	no	733	no	no	unknown	16	jun	83	4	-1	0
11159	32	technician	single	secondary	no	29	no	no	cellular	19	aug	156	2	-1	0
11160	43	technician	married	secondary	no	0	no	yes	cellular	8	may	9	2	172	5
11161	34	technician	married	secondary	no	0	no	no	cellular	9	jul	628	1	-1	0

11162 rows × 17 columns

```
In [6]: dataset.isnull().any()
```

```
Out[6]: age          False
job            False
marital        False
education      False
default        False
balance        False
housing        False
loan           False
contact        False
day            False
month          False
duration       False
campaign       False
pdays        False
previous      False
poutcome      False
deposit       False
dtype: bool
```

```
In [23]: dataset["default"].unique()
```

```
Out[23]: array(['no', 'yes'], dtype=object)
```

```
In [24]: x = dataset.iloc[:,16].values
y = dataset.iloc[:,16].values
```

```
In [25]: x
```

```
Out[25]: array([[59, 'admin.', 'married', ..., -1, 0, 'unknown'],
 [56, 'admin.', 'married', ..., -1, 0, 'unknown'],
 [41, 'technician', 'married', ..., -1, 0, 'unknown'],
 ...,
 [32, 'technician', 'single', ..., -1, 0, 'unknown'],
 [43, 'technician', 'married', ..., 172, 5, 'failure'],
 [34, 'technician', 'married', ..., -1, 0, 'unknown']], dtype=object)
```

```
In [26]: y
```

```
Out[26]: array(['yes', 'yes', 'yes', ..., 'no', 'no', 'no'], dtype=object)
```

```
In [29]: lst=[1,2,3,4,6,7,8,10,15,16]
for i in lst:
    name=dataset.columns[i]
    print(name)
    print(dataset[name].unique())

job
['admin.' 'technician' 'services' 'management' 'retired' 'blue-collar'
 'unemployed' 'entrepreneur' 'housemaid' 'unknown' 'self-employed'
 'student']
marital
['married' 'single' 'divorced']
education
['secondary' 'tertiary' 'primary' 'unknown']
default
['no' 'yes']
housing
['yes' 'no']
loan
['no' 'yes']
contact
['unknown' 'cellular' 'telephone']
month
['may' 'jun' 'jul' 'aug' 'oct' 'nov' 'dec' 'jan' 'feb' 'mar' 'apr' 'sep']
poutcome
['unknown' 'other' 'failure' 'success']
deposit
['yes' 'no']
```

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

```
In [31]: ct = ColumnTransformer([("oh",OneHotEncoder(),[1,2,3,4,6,7,8,10,15])], remainder = "passthrough")
x = ct.fit_transform(x)
```

```
In [32]: x
```

```
Out[32]: array([[1.0, 0.0, 0.0, ..., 1, -1, 0],
 [1.0, 0.0, 0.0, ..., 1, -1, 0],
 [0.0, 0.0, 0.0, ..., 1, -1, 0],
 ...,
 [0.0, 0.0, 0.0, ..., 2, -1, 0],
 [0.0, 0.0, 0.0, ..., 2, 172, 5],
 [0.0, 0.0, 0.0, ..., 1, -1, 0]], dtype=object)
```

```
In [33]: x.shape
```

```
Out[33]: (11162, 51)
```

```
In [41]: y.shape
```

```
Out[41]: (11162,)
```

```
In [42]: from sklearn.model_selection import train_test_split
x_train, x_test,y_train,y_test =train_test_split(x,y,test_size= 0.2,random_state =1)
```

```
In [43]: x.shape
```

```
Out[43]: (11162, 51)
```

```
In [44]: x_train.shape
```

```
Out[44]: (8929, 51)
```

```
In [45]: x_test.shape
```

```
Out[45]: (2233, 51)
```

```
In [46]: y_train.shape
```

```
Out[46]: (8929,)
```

```
In [47]: y_test.shape
```

```
Out[47]: (2233,)
```

```
In [49]: from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
x_train = sc.fit_transform(x_train)# permanent transformation
x_test = sc.transform(x_test)# temporary
```

```
In [50]: x_train
```

```
Out[50]: array([[ -0.36918047,  2.17933868, -0.17145904, ..., -0.55002237,
                -0.48107659, -0.3612238 ],
                [ -0.36918047,  2.17933868, -0.17145904, ...,  2.73405131,
                -0.48107659, -0.3612238 ],
                [ -0.36918047,  2.17933868, -0.17145904, ..., -0.55002237,
                -0.48107659, -0.3612238 ],
                ...,
                [ -0.36918047, -0.45885479,  5.83229683, ..., -0.55002237,
                -0.48107659, -0.3612238 ],
                [ 2.7087023 , -0.45885479, -0.17145904, ..., -0.18512529,
                 3.11825242,  4.40298199],
                [ -0.36918047,  2.17933868, -0.17145904, ..., -0.55002237,
                -0.48107659, -0.3612238 ]])
```

```
In [53]: x_test
```

```
Out[53]: array([[ 2.7087023 , -0.45885479, -0.17145904, ..., -0.55002237,
                -0.48107659, -0.3612238 ],
                [ -0.36918047,  2.17933868, -0.17145904, ..., -0.18512529,
                -0.48107659, -0.3612238 ],
                [ -0.36918047,  2.17933868, -0.17145904, ..., -0.18512529,
                -0.48107659, -0.3612238 ],
                ...,
                [ -0.36918047, -0.45885479,  5.83229683, ..., -0.55002237,
                -0.48107659, -0.3612238 ],
                [ -0.36918047, -0.45885479, -0.17145904, ...,  0.17977178,
                -0.48107659, -0.3612238 ],
                [ -0.36918047, -0.45885479, -0.17145904, ...,  0.54466886,
                -0.48107659, -0.3612238 ]])
```

```
In [ ]:
```