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

In [2]:
```
dataset = pd.read_csv(r"C:\Users\hp\Downloads\Telegram Desktop\bank.csv")
```

In [3]:
```
dataset
```

Out[3]:

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

11162 rows × 17 columns

In [4]:
```
dataset.head(5)
```

Out[4]:

	age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	poutcome	deposit
0	59	admin.	married	secondary	no	2343	yes	no	unknown	5	may	1042	1	-1	0	unknown	yes
1	56	admin.	married	secondary	no	45	no	no	unknown	5	may	1467	1	-1	0	unknown	yes
2	41	technician	married	secondary	no	1270	yes	no	unknown	5	may	1389	1	-1	0	unknown	yes
3	55	services	married	secondary	no	2476	yes	no	unknown	5	may	579	1	-1	0	unknown	yes
4	54	admin.	married	tertiary	no	184	no	no	unknown	5	may	673	2	-1	0	unknown	yes

In [5]:
```
dataset.isnull().sum()
```

Out[5]:
```
age          0
job          0
marital      0
education    0
default      0
balance      0
housing      0
loan         0
contact      0
day          0
month        0
duration     0
campaign     0
pdays        0
previous     0
poutcome     0
deposit      0
dtype: int64
```

In [6]:
```
x = dataset.iloc[:,0:16]
y = dataset.iloc[:,16:17]
```

In [7]:
```
x
```

Out[7]:

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

11162 rows × 16 columns

In [8]:
```
y
```

Out[8]:

	deposit
0	yes
1	yes
2	yes
3	yes
4	yes
...	...
11157	no
11158	no
11159	no
11160	no
11161	no

11162 rows × 1 columns

In [9]:
```
dataset["job"].unique()
```

Out[9]:
```
array(['admin.', 'technician', 'services', 'management', 'retired',
       'blue-collar', 'unemployed', 'entrepreneur', 'housemaid',
       'unknown', 'self-employed', 'student'], dtype=object)
```

In [11]:
```
dataset["marital"].unique()
```

Out[11]:
```
array(['married', 'single', 'divorced'], dtype=object)
```

In [13]:
```
dataset["education"].unique()
```

Out[13]:
```
array(['secondary', 'tertiary', 'primary', 'unknown'], dtype=object)
```

In [14]:
```
dataset["contact"].unique()
```

Out[14]:
```
array(['unknown', 'cellular', 'telephone'], dtype=object)
```

In [15]:
```
dataset["poutcome"].unique()
```

Out[15]:
```
array(['unknown', 'other', 'failure', 'success'], dtype=object)
```

In [16]:
```
from sklearn.preprocessing import OneHotEncoder
from sklearn.compose import ColumnTransformer
```

In [17]:
```
ct = ColumnTransformer([("oh",OneHotEncoder(),[1,2,3,4,6,7,8,10,15])], remainder = "passthrough")
x = ct.fit_transform(x)
```

In [18]:
```
x
```

Out[18]:
```
array([[  1.,   0.,   0., ...,   1.,  -1.,   0.],
       [  1.,   0.,   0., ...,   1.,  -1.,   0.],
       [  0.,   0.,   0., ...,   1.,  -1.,   0.],
       ...,
       [  0.,   0.,   0., ...,   2.,  -1.,   0.],
       [  0.,   0.,   0., ...,   2., 172.,   5.],
       [  0.,   0.,   0., ...,   1.,  -1.,   0.]])
```

In [20]:
```
x.shape
```

Out[20]:
```
(11162, 51)
```

In [19]:
```
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 [21]:
```
x.shape
```

Out[21]:
```
(11162, 51)
```

In [22]:
```
x_train.shape
```

Out[22]:
```
(8929, 51)
```

In [23]:
```
x_test.shape
```

Out[23]:
```
(2233, 51)
```

In [24]:
```
y_train.shape
```

Out[24]:
```
(8929, 1)
```

In [25]:
```
y_test.shape
```

Out[25]:
```
(2233, 1)
```

In [26]:
```
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
x_train = sc.fit_transform(x_train)
x_test = sc.transform(x_test)
```

In [27]:
```
x_train
```

Out[27]:
```
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 [28]:
```
x_test
```

Out[28]:
```
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 ]])
```