In [1]:

```
import pandas as pd
import numpy as np
```

In [2]:

```
dataset=pd.read_csv(r"C:\Users\kotha\Downloads\bank.csv")
```

In [3]:

```
dataset
```

Out[3]:

	age	job	marital	education	default	balance	housing	loan	contact	day	mon
0	59	admin.	married	secondary	no	2343	yes	no	unknown	5	ma
1	56	admin.	married	secondary	no	45	no	no	unknown	5	ma
2	41	technician	married	secondary	no	1270	yes	no	unknown	5	ma
3	55	services	married	secondary	no	2476	yes	no	unknown	5	ma
4	54	admin.	married	tertiary	no	184	no	no	unknown	5	ma
...	...	...	...	...	...	...	...	...	...	...	
11157	33	blue-collar	single	primary	no	1	yes	no	cellular	20	a
11158	39	services	married	secondary	no	733	no	no	unknown	16	ju
11159	32	technician	single	secondary	no	29	no	no	cellular	19	au
11160	43	technician	married	secondary	no	0	no	yes	cellular	8	ma
11161	34	technician	married	secondary	no	0	no	no	cellular	9	

11162 rows × 17 columns

In [4]:

```
type(dataset)
```

Out[4]:

```
pandas.core.frame.DataFrame
```

In [5]:

```
dataset.head()
```

Out[5]:

	age	job	marital	education	default	balance	housing	loan	contact	day	month	c
0	59	admin.	married	secondary	no	2343	yes	no	unknown	5	may	
1	56	admin.	married	secondary	no	45	no	no	unknown	5	may	
2	41	technician	married	secondary	no	1270	yes	no	unknown	5	may	
3	55	services	married	secondary	no	2476	yes	no	unknown	5	may	
4	54	admin.	married	tertiary	no	184	no	no	unknown	5	may	

In [6]:

```
dataset.tail()
```

Out[6]:

	age	job	marital	education	default	balance	housing	loan	contact	day	mon
11157	33	blue-collar	single	primary	no	1	yes	no	cellular	20	a
11158	39	services	married	secondary	no	733	no	no	unknown	16	ju
11159	32	technician	single	secondary	no	29	no	no	cellular	19	au
11160	43	technician	married	secondary	no	0	no	yes	cellular	8	ma
11161	34	technician	married	secondary	no	0	no	no	cellular	9	

In [7]:

```
dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11162 entries, 0 to 11161
Data columns (total 17 columns):
 #   Column     Non-Null Count  Dtype 
---  ------     --------------  ----- 
 0   age        11162 non-null  int64 
 1   job        11162 non-null  object
 2   marital    11162 non-null  object
 3   education  11162 non-null  object
 4   default    11162 non-null  object
 5   balance    11162 non-null  int64 
 6   housing    11162 non-null  object
 7   loan       11162 non-null  object
 8   contact    11162 non-null  object
 9   day        11162 non-null  int64 
 10  month      11162 non-null  object
 11  duration   11162 non-null  int64 
 12  campaign   11162 non-null  int64 
 13  pdays      11162 non-null  int64 
 14  previous   11162 non-null  int64 
 15  poutcome   11162 non-null  object
 16  deposit    11162 non-null  object
dtypes: int64(7), object(10)
memory usage: 1.4+ MB
```

In [8]:

```
dataset.describe()
```

Out[8]:

	age	balance	day	duration	campaign	pdays	
count	11162.000000	11162.000000	11162.000000	11162.000000	11162.000000	11162.000000	1116
mean	41.231948	1528.538524	15.658036	371.993818	2.508421	51.330407	
std	11.913369	3225.413326	8.420740	347.128386	2.722077	108.758282	
min	18.000000	-6847.000000	1.000000	2.000000	1.000000	-1.000000	
25%	32.000000	122.000000	8.000000	138.000000	1.000000	-1.000000	
50%	39.000000	550.000000	15.000000	255.000000	2.000000	-1.000000	
75%	49.000000	1708.000000	22.000000	496.000000	3.000000	20.750000	
max	95.000000	81204.000000	31.000000	3881.000000	63.000000	854.000000	5

In [9]:

```
dataset.shape
```

Out[9]:

```
(11162, 17)
```

In [10]:

```
dataset["job"].unique()
```

Out[10]:

```
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 [12]:

```
dataset.isnull().sum
```

Out[12]:

```
<bound method DataFrame.sum of          age    job  marital  education  defa
ult  balance  housing   loan  \
0      False  False    False      False    False    False    False  False
1      False  False    False      False    False    False    False  False
2      False  False    False      False    False    False    False  False
3      False  False    False      False    False    False    False  False
4      False  False    False      False    False    False    False  False
...      ...    ...      ...        ...      ...      ...      ...    ...
11157  False  False    False      False    False    False    False  False
11158  False  False    False      False    False    False    False  False
11159  False  False    False      False    False    False    False  False
11160  False  False    False      False    False    False    False  False
11161  False  False    False      False    False    False    False  False

       contact    day  month  duration  campaign  pdays  previous  poutcome
\
0        False  False  False     False     False  False     False     False
1        False  False  False     False     False  False     False     False
2        False  False  False     False     False  False     False     False
3        False  False  False     False     False  False     False     False
4        False  False  False     False     False  False     False     False
...        ...    ...    ...       ...       ...    ...       ...       ...
11157    False  False  False     False     False  False     False     False
11158    False  False  False     False     False  False     False     False
11159    False  False  False     False     False  False     False     False
11160    False  False  False     False     False  False     False     False
11161    False  False  False     False     False  False     False     False

       deposit
0        False
1        False
2        False
3        False
4        False
...        ...
11157    False
11158    False
11159    False
11160    False
11161    False

[11162 rows x 17 columns]>
```

In [13]:

```
dataset.isnull().any
```

Out[13]:

```
<bound method DataFrame.any of          age    job  marital  education  defa
ult  balance  housing   loan  \
0      False  False    False      False    False    False    False  False
1      False  False    False      False    False    False    False  False
2      False  False    False      False    False    False    False  False
3      False  False    False      False    False    False    False  False
4      False  False    False      False    False    False    False  False
...      ...    ...      ...        ...      ...      ...      ...    ...
11157  False  False    False      False    False    False    False  False
11158  False  False    False      False    False    False    False  False
11159  False  False    False      False    False    False    False  False
11160  False  False    False      False    False    False    False  False
11161  False  False    False      False    False    False    False  False

       contact    day  month  duration  campaign  pdays  previous  poutcome
\
0        False  False  False     False     False  False     False     False
1        False  False  False     False     False  False     False     False
2        False  False  False     False     False  False     False     False
3        False  False  False     False     False  False     False     False
4        False  False  False     False     False  False     False     False
...        ...    ...    ...       ...       ...    ...       ...       ...
11157    False  False  False     False     False  False     False     False
11158    False  False  False     False     False  False     False     False
11159    False  False  False     False     False  False     False     False
11160    False  False  False     False     False  False     False     False
11161    False  False  False     False     False  False     False     False

       deposit
0        False
1        False
2        False
3        False
4        False
...        ...
11157    False
11158    False
11159    False
11160    False
11161    False

[11162 rows x 17 columns]>
```

In [14]:

```
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
dataset["job"]=le.fit_transform(dataset["job"])
dataset["marital"]=le.fit_transform(dataset["marital"])
dataset["education"]=le.fit_transform(dataset["education"])
dataset["housing"]=le.fit_transform(dataset["housing"])
dataset["loan"]=le.fit_transform(dataset["loan"])
dataset["month"]=le.fit_transform(dataset["month"])
dataset["contact"]=le.fit_transform(dataset["contact"])
dataset["poutcome"]=le.fit_transform(dataset["poutcome"])
dataset["deposit"]=le.fit_transform(dataset["deposit"])
dataset["default"]=le.fit_transform(dataset["default"])
```

In [15]:

```
dataset.head()
```

Out[15]:

	age	job	marital	education	default	balance	housing	loan	contact	day	month	duration
0	59	0	1	1	0	2343	1	0	2	5	8	1042
1	56	0	1	1	0	45	0	0	2	5	8	1467
2	41	9	1	1	0	1270	1	0	2	5	8	1389
3	55	7	1	1	0	2476	1	0	2	5	8	579
4	54	0	1	2	0	184	0	0	2	5	8	673

In [16]:

```
dataset["marital"].unique()
```

Out[16]:

```
array([1, 2, 0])
```

In [17]:

```
dataset["housing"].unique()
```

Out[17]:

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

In [18]:

```
dataset["education"].unique()
```

Out[18]:

```
array([1, 2, 0, 3])
```

In [19]:

```
dataset["job"].unique()
```

Out[19]:

```
array([ 0,  9,  7,  4,  5,  1, 10,  2,  3, 11,  6,  8])
```

In [20]:

```
dataset["loan"].unique()
```

Out[20]:

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

In [21]:

```
dataset["month"].unique()
```

Out[21]:

```
array([ 8,  6,  5,  1, 10,  9,  2,  4,  3,  7,  0, 11])
```

In [22]:

```
dataset["contact"].unique()
```

Out[22]:

```
array([2, 0, 1])
```

In [23]:

```
dataset["poutcome"].unique()
```

Out[23]:

```
array([3, 1, 0, 2])
```

In [24]:

```
dataset["deposit"].unique()
```

Out[24]:

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

In [25]:

```
dataset["default"].unique()
```

Out[25]:

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

In [26]:

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

Out[26]:

```
array([[59,  0,  1, ...,  0,  3,  1],
       [56,  0,  1, ...,  0,  3,  1],
       [41,  9,  1, ...,  0,  3,  1],
       ...,
       [32,  9,  2, ...,  0,  3,  0],
       [43,  9,  1, ...,  5,  0,  0],
       [34,  9,  1, ...,  0,  3,  0]], dtype=int64)
```

In [27]:

```
y
```

Out[27]:

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

In [31]:

```
from sklearn.preprocessing import OneHotEncoder
one=OneHotEncoder()
z=one.fit_transform(x[:,0:3:16]).toarray()
x=np.delete(x,3,axis=1)
x=np.concatenate((z,x),axis=1)
x
```

Out[31]:

```
array([[1., 0., 0., ..., 0., 3., 1.],
       [1., 0., 0., ..., 0., 3., 1.],
       [1., 0., 0., ..., 0., 3., 1.],
       ...,
       [1., 0., 0., ..., 0., 3., 0.],
       [1., 0., 0., ..., 5., 0., 0.],
       [1., 0., 0., ..., 0., 3., 0.]])
```

In [ ]: