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

In [2]:
```
# one hot encoder on data sets
```

In [44]:
```
bank_df=pd.read_csv(r"C:\Users\Mrityunjay\Desktop\Data science naresh it\Class note
bank_df
```

Out[44]:

| | age | job | marital | education | default | balance | housing | loan | contact | da |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 30 | unemployed | married | primary | no | 1787 | no | no | cellular | |
| **1** | 33 | services | married | secondary | no | 4789 | yes | yes | cellular | |
| **2** | 35 | management | single | tertiary | no | 1350 | yes | no | cellular | |
| **3** | 30 | management | married | tertiary | no | 1476 | yes | yes | unknown | |
| **4** | 59 | blue-collar | married | secondary | no | 0 | yes | no | unknown | |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| **4516** | 33 | services | married | secondary | no | -333 | yes | no | cellular | |
| **4517** | 57 | self-employed | married | tertiary | yes | -3313 | yes | yes | unknown | |
| **4518** | 57 | technician | married | secondary | no | 295 | no | no | cellular | |
| **4519** | 28 | blue-collar | married | secondary | no | 1137 | no | no | cellular | |
| **4520** | 44 | entrepreneur | single | tertiary | no | 1136 | yes | yes | cellular | |

4521 rows × 17 columns

In [4]:
```
cat_columns=bank_df.select_dtypes(include="object").columns
num_columns=bank_df.select_dtypes(exclude="object").columns
num_columns
```

Out[4]:
```
Index(['age', 'balance', 'day', 'duration', 'campaign', 'pdays', 'previous'], dtyp
e='object')
```

In [5]:
```
cat_columns
```

Out[5]:
```
Index(['job', 'marital', 'education', 'default', 'housing', 'loan', 'contact',
       'month', 'poutcome', 'y'],
      dtype='object')
```

In [6]:
```
bank_df["job"].unique()
```

Out[6]:
```
array(['unemployed', 'services', 'management', 'blue-collar',
       'self-employed', 'technician', 'entrepreneur', 'admin.', 'student',
       'housemaid', 'retired', 'unknown'], dtype=object)
```

In [7]:
```
df=pd.get_dummies(bank_df["job"],dtype="int")
df
```

Out[7]:

| | admin. | blue-collar | entrepreneur | housemaid | management | retired | self-employed | services |
|---|---|---|---|---|---|---|---|---|
| **0** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **1** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| **2** | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| **3** | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| **4** | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... |
| **4516** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| **4517** | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| **4518** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **4519** | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| **4520** | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |

4521 rows × 12 columns

In [8]:
```
len(bank_df["job"].unique())
```

Out[8]: 12

In [9]:
```
bank_df.shape
```

Out[9]: (4521, 17)

In [10]:
```
for i in cat_columns:
    print(bank_df[i].unique(),bank_df[i].nunique())
```

```
['unemployed' 'services' 'management' 'blue-collar' 'self-employed'
 'technician' 'entrepreneur' 'admin.' 'student' 'housemaid' 'retired'
 'unknown'] 12
['married' 'single' 'divorced'] 3
['primary' 'secondary' 'tertiary' 'unknown'] 4
['no' 'yes'] 2
['no' 'yes'] 2
['no' 'yes'] 2
['cellular' 'unknown' 'telephone'] 3
['oct' 'may' 'apr' 'jun' 'feb' 'aug' 'jan' 'jul' 'nov' 'sep' 'mar' 'dec'] 12
['unknown' 'failure' 'other' 'success'] 4
['no' 'yes'] 2
```

In [11]:
```
for i in cat_columns:
    df=pd.get_dummies(bank_df,dtype="int")
df
```

Out[11]:

| | age | balance | day | duration | campaign | pdays | previous | job_admin. | job_blue-collar | job |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 30 | 1787 | 19 | 79 | 1 | -1 | 0 | 0 | 0 | |
| **1** | 33 | 4789 | 11 | 220 | 1 | 339 | 4 | 0 | 0 | |
| **2** | 35 | 1350 | 16 | 185 | 1 | 330 | 1 | 0 | 0 | |
| **3** | 30 | 1476 | 3 | 199 | 4 | -1 | 0 | 0 | 0 | |
| **4** | 59 | 0 | 5 | 226 | 1 | -1 | 0 | 0 | 1 | |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| **4516** | 33 | -333 | 30 | 329 | 5 | -1 | 0 | 0 | 0 | |
| **4517** | 57 | -3313 | 9 | 153 | 1 | -1 | 0 | 0 | 0 | |
| **4518** | 57 | 295 | 19 | 151 | 11 | -1 | 0 | 0 | 0 | |
| **4519** | 28 | 1137 | 6 | 129 | 4 | 211 | 3 | 0 | 1 | |
| **4520** | 44 | 1136 | 3 | 345 | 2 | 249 | 7 | 0 | 0 | |

4521 rows × 53 columns

In [12]:
```
count=0
for i in cat_columns:
    count+=bank_df[i].nunique()
print(count+len(num_columns))
```

53

In [42]:
```
dfs=pd.DataFrame()
for i in cat_columns:
    df=pd.get_dummies(bank_df[i],dtype="int")
    df[f"{i}_encode"]=bank_df[i].values

    dfs=pd.concat([dfs,df],axis=1)
dfs
```

Out[42]:

| | admin. | blue-collar | entrepreneur | housemaid | management | retired | self-employed | services |
|---|---|---|---|---|---|---|---|---|
| **0** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **1** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| **2** | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| **3** | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| **4** | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... |
| **4516** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| **4517** | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| **4518** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **4519** | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| **4520** | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |

4521 rows × 56 columns

In [46]:
```
#number of columns lables
count
```

Out[46]: 46

In [ ]: