In [51]:
```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
```

In [52]:
```
df = pd.read_csv('churn3.csv')
```

In [53]:
```
df.head()
```

Out[53]:

	Unnamed: 0	customer_id	vintage	age	gender	dependents	occupation	city	customer_nw
0	0	1	2101	66	1.0	0.0	1.0	187.0	
1	1	2	2348	35	1.0	0.0	1.0	NaN	
2	2	4	2194	31	1.0	0.0	0.0	146.0	
3	3	5	2329	90	NaN	NaN	1.0	1020.0	
4	4	6	1579	42	1.0	2.0	1.0	1494.0	

In [54]:
```
# checking for missing values
df.isnull().sum()
```

Out[54]:
```
Unnamed: 0                0
customer_id               0
vintage                   0
age                       0
gender                  525
dependents             2463
occupation               80
city                    803
customer_nw_category      0
branch_code               0
churn                     0
dtype: int64
```

In [55]:
```
# checking the distribution of Target Variable
df['churn'].value_counts()
```

Out[55]:
```
0    23122
1     5260
Name: churn, dtype: int64
```

In [56]:
```
X = df.drop(columns='churn', axis=1)
y = df['churn']
```

In [59]:
```
X_train, X_test, Y_train, Y_test = train_test_split(X, y, test_size=0.2, strat
```

In [45]:
```
from sklearn.metrics import accuracy_score
```

In [46]:
```
#Train test
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)
```

In [60]:
```
# printing out train and test sets

print('X_train : ')
print(X_train.head())
print('')
print('X_test : ')
print(X_test.head())
print('')
print('y_train : ')
print(y_train.head())
print('')
print('y_test : ')
print(y_test.head())
```

```
X_train :
       Unnamed: 0  customer_id  vintage  age  gender  dependents  occupation
\
24249       24249        25876     1994   26     1.0         0.0         1.0
16252       16252        17341     2037   38     1.0         0.0         1.0
2968         2968         3165     2043   56     1.0         2.0         1.0
27254       27254        29100     1948   41     0.0         0.0         1.0
92             92           96     2424   29     0.0         0.0         0.0

         city  customer_nw_category  branch_code
24249  1540.0                     3         1301
16252   551.0                     3          550
2968     15.0                     2          237
27254  1096.0                     3          578
92      409.0                     2           26

X_test :
       Unnamed: 0  customer_id  vintage  age  gender  dependents  occupation
\
387           387          416     2133   47     1.0         0.0         1.0
17340       17340        18504     2269   41     1.0         2.0         0.0
5285         5285         5626     1781   42     0.0         0.0         0.0
2220         2220         2380     2212   54     1.0         0.0         0.0
12902       12902        13777     2085   40     0.0         0.0         1.0

         city  customer_nw_category  branch_code
387     363.0                     2           31
17340   395.0                     2         2212
5285    146.0                     2          312
2220   1181.0                     3          235
12902  1214.0                     3          787

y_train :
27778    0
22768    0
25158    0
27845    0
9807     0
Name: churn, dtype: int64

y_test :
1430     0
15898    0
9356     0
14591    0
2098     0
Name: churn, dtype: int64
```

In [ ]:

In [ ]: