lab-1

October 2, 2024

[ ]:

[2]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

[3]:
```
df=pd.read_csv("C:/5th_semester/machine_learning/ml_lab/Lab_01/Data_sets/
 telecom_customer_churn.csv")
```

[4]:
```
df
```

[4]:

	Customer ID	Gender	Age	Married	Number of Dependents	City	\
0	0002-ORFBO	Female	37	Yes	0	Frazier Park	
1	0003-MKNFE	Male	46	No	0	Glendale	
2	0004-TLHLJ	Male	50	No	0	Costa Mesa	
3	0011-IGKFF	Male	78	Yes	0	Martinez	
4	0013-EXCHZ	Female	75	Yes	0	Camarillo	
…	…	…	…	…	…	…	
7038	9987-LUTYD	Female	20	No	0	La Mesa	
7039	9992-RRAMN	Male	40	Yes	0	Riverbank	
7040	9992-UJOEL	Male	22	No	0	Elk	
7041	9993-LHIEB	Male	21	Yes	0	Solana Beach	
7042	9995-HOTOH	Male	36	Yes	0	Sierra City	

	Zip Code	Latitude	Longitude	Number of Referrals	… \
0	93225	34.827662	-118.999073	2	…
1	91206	34.162515	-118.203869	0	…
2	92627	33.645672	-117.922613	0	…
3	94553	38.014457	-122.115432	1	…
4	93010	34.227846	-119.079903	3	…

```
…       …       …         …                  … …
7038    91941 32.759327 -116.997260              0 …
7039    95367 37.734971 -120.954271              1 …
7040    95432 39.108252 -123.645121              0 …
7041    92075 33.001813 -117.263628              5 …
7042    96125 39.600599 -120.636358              1 …

      Payment Method Monthly Charge Total Charges Total Refunds \
0       Credit Card         65.60       593.30          0.00
1       Credit Card         -4.00       542.40         38.33
2   Bank Withdrawal         73.90       280.85          0.00
3   Bank Withdrawal         98.00      1237.85          0.00
4       Credit Card         83.90       267.40          0.00
…               …           …           …             …
7038    Credit Card         55.15       742.90          0.00
7039 Bank Withdrawal        85.10      1873.70          0.00
7040    Credit Card         50.30        92.75          0.00
7041    Credit Card         67.85      4627.65          0.00
7042 Bank Withdrawal        59.00      3707.60          0.00

Total Extra Data Charges Total Long Distance Charges Total Revenue \
0                         0                   381.51       974.81
1                        10                    96.21       610.28
2                         0                   134.60       415.45
3                         0                   361.66      1599.51
4                         0                    22.14       289.54
…                       …                      …           …
7038                      0                   606.84      1349.74
7039                      0                   356.40      2230.10
7040                      0                    37.24       129.99
7041                      0                   142.04      4769.69
7042                      0                     0.00      3707.60
```

```
     Customer Status Churn Category                Churn Reason
0             Stayed            NaN                         NaN
1             Stayed            NaN                         NaN
2            Churned     Competitor Competitor had better
                                    devices
3          Churned Dissatisfaction      Product
                                        dissatisfaction
4          Churned Dissatisfaction       Network reliability
…              …             …                          …
7038          Stayed            NaN                         NaN
7039       Churned Dissatisfaction      Product
                                        dissatisfaction
7040          Joined            NaN                         NaN
7041          Stayed            NaN                         NaN
7042          Stayed            NaN                         NaN
[7043 rows x 38 columns]
```

[5]:
```
df.isnull().sum()
```

[5]:
```
Customer ID                        0
Gender                             0
Age                                0
Married                            0
Number of Dependents               0
City                               0
Zip Code                           0
Latitude                           0
Longitude                          0
Number of Referrals                0
Tenure in Months                   0
Offer                           3877
Phone Service                      0
Avg Monthly Long Distance Charges  682
Multiple Lines                     682
Internet Service                   0
Internet Type                   1526
Avg Monthly GB Download         1526
Online Security                 1526
Online Backup                   1526
Device Protection Plan          1526
Premium Tech Support            1526
Streaming TV                    1526
Streaming Movies                1526
Streaming Music                 1526
Unlimited Data                  1526
Contract                           0
```

```
Paperless Billing                  0
Payment Method                     0
Monthly Charge                     0
Total Charges                      0
Total Refunds                      0
Total Extra Data Charges           0
Total Long Distance Charges        0
Total Revenue                      0
Customer Status                    0
 Churn Category   5174 Churn Reason
 5174 dtype: int64
```

[7]:
```
df.columns
```

[7]:
```
Index(['Customer ID', 'Gender', 'Age', 'Married', 'Number of
Dependents',
          'City', 'Zip Code', 'Latitude', 'Longitude', 'Number of
          Referrals',
          'Tenure in Months', 'Offer', 'Phone Service',
          'Avg Monthly Long Distance Charges', 'Multiple Lines',
          'Internet Service', 'Internet Type', 'Avg Monthly GB
          Download',
          'Online Security', 'Online Backup', 'Device Protection Plan',
          'Premium Tech Support', 'Streaming TV', 'Streaming Movies',
          'Streaming Music', 'Unlimited Data', 'Contract', 'Paperless
          Billing',
  'Payment Method', 'Monthly Charge', 'Total Charges', 'Total Refunds',
       'Total Extra Data Charges', 'Total Long Distance Charges',
           'Total Revenue', 'Customer Status', 'Churn Category', 'Churn
          Reason'], dtype='object')
```

[15]:
```
df.fillna(df.mean(), inplace=True)
df.fillna(df.mode().iloc[0], inplace=True)
print(df.isnull().sum())
```

```
Age                                        0
Number of Dependents                       0
Zip Code                                   0
Latitude                                   0
Longitude                                  0
                                          ..
Churn Reason_Poor expertise of online      0
support
Churn Reason_Poor expertise of phone       0
support
Churn Reason_Price too high                0
Churn Reason_Product dissatisfaction       0
```

```
Churn Reason_Service dissatisfaction      0
Length: 8211, dtype: int64
```

[29]:
```
df.sample(5)
```

[29]:
```
     Age Number of Dependents Zip Code Latitude  Longitude \
4079  57                   0   93261 35.809921 -119.127437
5995  57                   0   92283 32.852947 -114.850784
386   38                   0   96007 40.448632 -122.306657
4944  69                   0   95345 37.581496 -119.972762
6870  54                   0   90740 33.754620 -118.071128

      Number of Referrals Tenure in Months \
4079                   0                3
5995                   0               19
386                    0                9
4944                   4               70
6870                   0               39
      Avg Monthly Long Distance Charges Avg Monthly GB Download \
4079                             47.50                    22.0
5995                             29.55                    13.0
386                              32.57                    13.0
4944                             40.94                     8.0
6870                             36.93                    12.0
      Monthly Charge … Churn Reason_Lack of self-service on Website
      \
4079        107.95 …                                          False
5995         78.70 …                                         False
386          80.55 …                                         False
4944         88.55 …                                         False
6870         81.40 …                                         False
      Churn Reason_Limited range of services \
4079                                   False
5995                                   False
386                                    False
4944                                   False
6870                                   False

      Churn Reason_Long distance charges Churn Reason_Moved \
4079                               False              False
5995                               False              False
386                                False              False
4944                               False              False
6870                               False              False
      Churn Reason_Network reliability \
4079                             False
5995                             False
```

```
386                          False
4944                         False
6870                         False

     Churn Reason_Poor expertise of online support \
4079                                       False
5995                                       False
386                                        False
4944                                       False
6870                                       False

     Churn Reason_Poor expertise of phone support \
4079                                     False
5995                                     False
386                                      False
4944                                     False
6870                                     False

     Churn Reason_Price too high Churn Reason_Product
     dissatisfaction \
4079                     False                              False
5995                     False                              False
386                      False                              False
4944                     False                              False
6870                     False                              False
     Churn Reason_Service dissatisfaction
4079                               False
5995                               False
386                                False
4944                               False
6870                               False

     [5 rows x 8211 columns]
```

[31]:
```
numerical_cols=df.select_dtypes(np.number)
numerical_cols.head()
```

[31]:
```
  Age Number of Dependents Zip Code Latitude  Longitude \
0   370     93225 34.827662 -118.999073
1   460     91206 34.162515 -118.203869
2   500     92627 33.645672 -117.922613
3   780     94553 38.014457 -122.115432
4   750     93010 34.227846 -119.079903

 Number of Referrals Tenure in Months Avg Monthly Long Distance Charges \
    0                 2   9     42.39
```

```
1                  0   9     10.69
2                  0   4     33.65
3                  1   13    27.82
4                  3   3     7.38

   Avg Monthly GB Download Monthly Charge Total Charges Total
   Refunds \
0                  16.0           65.6        593.30         0.00
1                  10.0           -4.0        542.40        38.33
2                  30.0           73.9        280.85         0.00
3                   4.0           98.0       1237.85         0.00
4                  11.0           83.9        267.40         0.00
   Total Extra Data Charges Total Long Distance Charges Total
   Revenue
0                         0                      381.51       974.81
1                        10                       96.21       610.28
2                         0                      134.60       415.45
3                         0                      361.66      1599.51
4                         0                       22.14       289.54
```

[ ]:

[25]:
```
from sklearn.preprocessing import OneHotEncoder
from sklearn.preprocessing import LabelEncoder
from sklearn.compose import ColumnTransformer
from sklearn.preprocessing import OrdinalEncoder
```

[ ]:
```
transformer=ColumnTransformer(transformer[
    ("tnrf1",OneHotEncoder(drop="first" ),['City','Gender','Payment Method']),

])
```