

Assignment No-3

CODE:-

```
# Imports
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import MinMaxScaler
from google.colab import files

# Upload dataset
uploaded = files.upload()
filename = next(iter(uploaded))
df = pd.read_csv(filename)

# Explore dataset
df.info()
print(df.isnull().sum())
print(df.describe())

# Handle missing values
df.fillna(df.median(numeric_only=True), inplace=True)
for col in df.select_dtypes(include=['object']).columns:
    df[col].fillna(df[col].mode()[0], inplace=True)

print(df.isnull().sum())

# Remove duplicates
df.drop_duplicates(inplace=True)
print(f'Duplicates removed, remaining: {df.duplicated().sum()}')

# Standardize categorical data example
if 'Gender' in df.columns:
    df['Gender'] = df['Gender'].str.strip().str.lower()
if 'SeniorCitizen' in df.columns:
    df['SeniorCitizen'] = df['SeniorCitizen'].replace({1: 'Yes', 0: 'No'})
```

```

# Convert data types
if 'Churn' in df.columns:
    df['Churn'] = df['Churn'].astype('category')
if 'CustomerID' in df.columns:
    df['CustomerID'] = df['CustomerID'].astype(str)

print(df.dtypes)

# Handle outliers using IQR for numeric columns only
Q1 = df.select_dtypes(include=np.number).quantile(0.25)
Q3 = df.select_dtypes(include=np.number).quantile(0.75)
IQR = Q3 - Q1

condition = ~((df.select_dtypes(include=np.number) < (Q1 - 1.5 * IQR)) |
              (df.select_dtypes(include=np.number) > (Q3 + 1.5 *
IQR))).any(axis=1)
df_no_outliers = df[condition]
print(f"Shape after removing outliers: {df_no_outliers.shape}")

# Feature engineering: create 'Tenure' if date column exists
if 'ContractStartDate' in df.columns:
    df['Tenure'] = pd.to_datetime(df['ContractStartDate'],
errors='coerce').apply(lambda x: (pd.to_datetime('today') - x).days // 30
if pd.notnull(x) else np.nan)
    df['Tenure'].fillna(df['Tenure'].median(), inplace=True)

# Scale numerical columns if they exist
num_cols = ['MonthlyCharges', 'TotalCharges']
for col in num_cols:
    if col in df.columns:
        scaler = MinMaxScaler()
        df[col] = scaler.fit_transform(df[[col]])

```

```
# Prepare features and target
if 'Churn' in df.columns:
    X = df.drop(columns=['Churn', 'CustomerID'], errors='ignore')
    y = df['Churn']

    # Split into train/test
    X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)

    print(f"Train shape: {X_train.shape}, Test shape: {X_test.shape}")

# Export cleaned data
df.to_csv('Cleaned_Telecom_Customer_Churn.csv', index=False)

# Download the cleaned file
files.download('Cleaned_Telecom_Customer_Churn.csv')
```

OUTPUT :-

```

Choose File Telecom_C_hurn.csv.csv
Telecom_Customer_Churn.csv.csv[0x0]: 1435200 bytes, last modified: 9/17/2025 - 100% done
Saving Telecom_Customer_Churn.csv.csv to Telecom_Customer_Churn.csv (2).csv
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7843 entries, 0 to 7842
Data columns (total 38 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Customer ID      7843 non-null    object  
 1   Gender            7843 non-null    object  
 2   Age               7843 non-null    int64  
 3   Married           7843 non-null    object  
 4   Number of Dependents  7843 non-null    int64  
 5   City              7843 non-null    object  
 6   Zip Code          7843 non-null    int64  
 7   Latitude          7843 non-null    float64 
 8   Longitude         7843 non-null    float64 
 9   Number of Referrals  7843 non-null    int64  
 10  Tenure in Months 7843 non-null    int64  
 11  Offer              3106 non-null    object  
 12  Phone Service     7843 non-null    object  
 13  Avg Monthly Long Distance Charges 6361 non-null    float64 
 14  Multiple Lines     6361 non-null    object  
 15  Internet Service   7843 non-null    object  
 16  Internet Type     5517 non-null    object  
 17  Avg Monthly GB Download 5517 non-null    float64 
 18  Online Security    5517 non-null    object  
 19  Online Backup       5517 non-null    object  
 20  Device Protection Plan 5517 non-null    object  
 21  Premium Tech Support 5517 non-null    object  
 22  Streaming TV        5517 non-null    object  
 23  Streaming Movies    5517 non-null    object  
 24  Streaming Music     5517 non-null    object  
 25  Unlimited Data      5517 non-null    object  
 26  Contract            7843 non-null    object  
 27  Paperless Billing   7843 non-null    object  
 28  Payment Method      7843 non-null    object  
 29  Monthly Charge      7843 non-null    float64 
 30  Total Charges       7843 non-null    float64 
 31  Total Refunds       7843 non-null    float64 
 32  Total Extra Data Charges 7843 non-null    float64 
 33  Total Long Distance Charges 7843 non-null    float64 
 34  Total Revenue        7843 non-null    float64 
 35  Customer Status      7843 non-null    object  
 36  Churn Category       1869 non-null    object  
 37  Churn Reason         1869 non-null    object 

20 churn category      1869 non-null    object  
21 Churn Reason        1869 non-null    object  
dtype: float64(9), int64(6), object(23)
memory usage: 2.0+ MB
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         682
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
count 7843.000000 7043.000000 7043.000000 7043.000000 \
mean 46.509726 0.468092 93486.070567 36.197455
std 16.760353 0.043893 1954.747545 7.449070
min 0.000000 0.000000 0.000000 0.000000
max 100.000000 100.000000 100.000000 100.000000

```

Untitled3.ipynb

```

    churn_reasn
dtype: int64
   count    Age  Number of Dependents  Zip Code  Latitude \
count  7843.000000      7843.000000  7843.000000 \
mean   46.599726      0.468692  93486.076567  36.197455
std    16.750352      0.962802  3856.767595  2.468929
min   16.000000      0.000000  80000.000000  32.166828
25%  32.000000      0.000000  92101.000000  33.996466
50%  46.000000      0.000000  93518.000000  36.205465
75%  60.000000      0.000000  95329.000000  38.161321
max  80.000000      9.000000  96159.000000  41.962127

   Longitude  Number of Referrals Tenure In Months \
count  7843.000000      7843.000000  7843.000000 \
mean -119.756684      1.951867  32.386767
std   2.154425      3.001190  24.542651
min  -124.303721      0.000000  1.000000
25%  -120.780099      0.000000  9.000000
50%  -119.595293      0.000000  29.000000
75%  -117.969795      3.000000  55.000000
max  -114.192001     11.000000  72.000000

Avg Monthly Long Distance Charges  Avg Monthly GB Download \
count  6301.000000      5517.000000  5517.000000 \
mean   25.420517      26.189958
std    14.200374      19.586585
min   1.000000      2.000000
25%  11.000000      18.000000
50%  25.600000      29.000000
75%  37.680000      30.000000
max  49.900000      85.000000

Monthly Charge  Total Charges  Total Refunds  Total Extra Data Charges \
count  7043.000000      7043.000000  7043.000000  7043.000000 \
mean   63.596131      2288.381264  1.962182  6.866713
std    31.204743      2266.220462  7.992614  25.104978
min  -10.000000      18.000000  0.000000  0.000000
25%  40.000000      400.150000  0.000000  0.000000
50%  78.050000      1304.550000  0.000000  0.000000
75%  89.750000      3786.600000  0.000000  0.000000
max 118.750000      8684.800000  49.700000  150.000000

```

Variables Terminal ✓ 10:48AM Python 3
20°C Light rain ↻ ENG IN ↻ 10:50:15 17-09-2025