COLLEGE CODE:4212
REGISTER NUMBER:421221243029

CUSTOMER CHURN PREDICTION

DATA ANALYTICS WITH COGNOS:GROUP2

PHASE:3

This phase involves in designing of the steps that defining in each phase of the previous documentation this involves importing necessary functions, data processing and so on in this phase we have to begin our project by loading and preprocessing the dataset.

The IBM suggests using the jupyter notebook for loading and preprocess the dataset:

Here for this project title we need to define the loading the libraries, understand the data and visualize the missing values.

For this certain inputs are defined for this project.in this phase each of the input lines of the project is given as follows:

phase3

October 17, 2023

[1]:
```
import pandas as pd
import numpy as np
import missingno as msno
```

[2]:
```
df = pd.read_csv('WA_Fn-UseC_-Telco-Customer-Churn.csv')
```

[3]:
```
df.head()
```

[3]:
```
   customerID  gender  SeniorCitizen Partner Dependents  tenure PhoneService  \
0  7590-VHVEG  Female              0     Yes         No       1           No
1  5575-GNVDE    Male              0      No         No      34          Yes
2  3668-QPYBK    Male              0      No         No       2          Yes
3  7795-CFOCW    Male              0      No         No      45           No
4  9237-HQITU  Female              0      No         No       2          Yes

      MultipleLines InternetService OnlineSecurity  … DeviceProtection  \
0  No phone service             DSL             No  …               No
1                No             DSL            Yes  …              Yes
2                No             DSL            Yes  …               No
3  No phone service             DSL            Yes  …              Yes
4                No     Fiber optic             No  …               No

  TechSupport StreamingTV StreamingMovies        Contract PaperlessBilling  \
0          No          No              No  Month-to-month              Yes
1          No          No              No        One year               No
2          No          No              No  Month-to-month              Yes
3         Yes          No              No        One year               No
4          No          No              No  Month-to-month              Yes

              PaymentMethod MonthlyCharges  TotalCharges Churn
0          Electronic check          29.85         29.85    No
1              Mailed check          56.95        1889.5    No
2              Mailed check          53.85        108.15   Yes
3  Bank transfer (automatic)          42.30       1840.75    No
4          Electronic check          70.70        151.65   Yes

[5 rows x 21 columns]
```

[4]:
```
df.shape
```

[4]:
```
(7043, 21)
```

[5]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 21 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   customerID        7043 non-null   object
 1   gender            7043 non-null   object
 2   SeniorCitizen     7043 non-null   int64
 3   Partner           7043 non-null   object
 4   Dependents        7043 non-null   object
 5   tenure            7043 non-null   int64
 6   PhoneService      7043 non-null   object
 7   MultipleLines     7043 non-null   object
 8   InternetService   7043 non-null   object
 9   OnlineSecurity    7043 non-null   object
 10  OnlineBackup      7043 non-null   object
 11  DeviceProtection  7043 non-null   object
 12  TechSupport       7043 non-null   object
 13  StreamingTV       7043 non-null   object
 14  StreamingMovies   7043 non-null   object
 15  Contract          7043 non-null   object
 16  PaperlessBilling  7043 non-null   object
 17  PaymentMethod     7043 non-null   object
 18  MonthlyCharges    7043 non-null   float64
 19  TotalCharges      7043 non-null   object
 20  Churn             7043 non-null   object
dtypes: float64(1), int64(2), object(18)
memory usage: 1.1+ MB
```

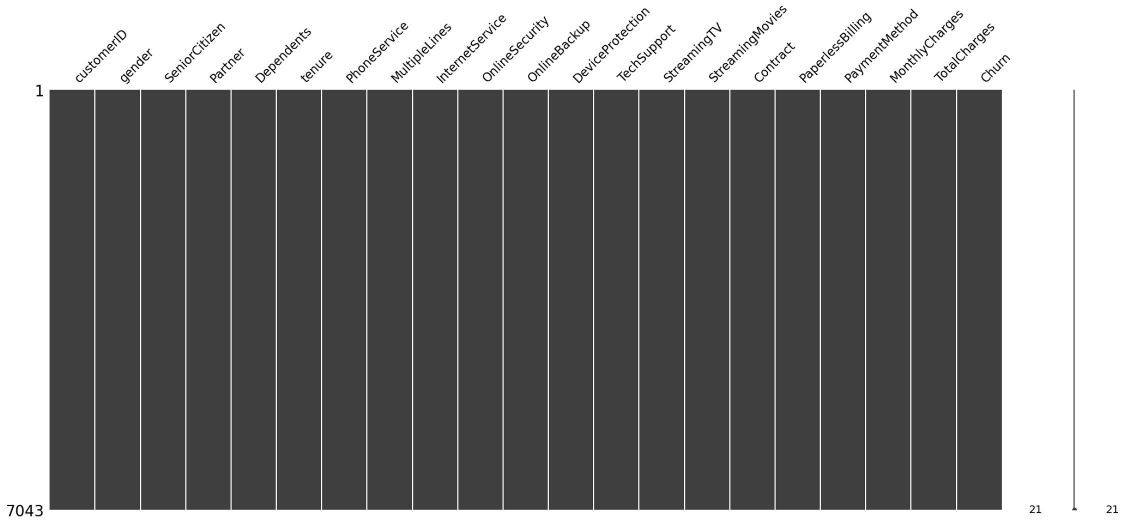
[6]:
```
df.columns.values
```

[6]:
```
array(['customerID', 'gender', 'SeniorCitizen', 'Partner', 'Dependents',
       'tenure', 'PhoneService', 'MultipleLines', 'InternetService',
       'OnlineSecurity', 'OnlineBackup', 'DeviceProtection',
       'TechSupport', 'StreamingTV', 'StreamingMovies', 'Contract',
       'PaperlessBilling', 'PaymentMethod', 'MonthlyCharges',
       'TotalCharges', 'Churn'], dtype=object)
```

[7]:
```
df.dtypes
```

```
[7]: customerID           object
     gender               object
     SeniorCitizen         int64
     Partner              object
     Dependents           object
     tenure                int64
     PhoneService         object
     MultipleLines        object
     InternetService      object
     OnlineSecurity       object
     OnlineBackup         object
     DeviceProtection     object
     TechSupport          object
     StreamingTV          object
     StreamingMovies      object
     Contract             object
     PaperlessBilling     object
     PaymentMethod        object
     MonthlyCharges      float64
     TotalCharges         object
     Churn                object
     dtype: object
```

```
[8]: msno.matrix(df);
```



```
[9]: df = df.drop(['customerID'], axis = 1)
     df.head()
```

[9]:
```
   gender  SeniorCitizen Partner Dependents  tenure PhoneService  \
0  Female              0     Yes         No       1           No
1    Male              0      No         No      34          Yes
2    Male              0      No         No       2          Yes
3    Male              0      No         No      45           No
4  Female              0      No         No       2          Yes

      MultipleLines InternetService OnlineSecurity OnlineBackup  \
0  No phone service             DSL             No          Yes
1                No             DSL            Yes           No
2                No             DSL            Yes          Yes
3  No phone service             DSL            Yes           No
4                No     Fiber optic             No           No

  DeviceProtection TechSupport StreamingTV StreamingMovies        Contract  \
0               No          No          No              No  Month-to-month
1              Yes          No          No              No        One year
2               No          No          No              No  Month-to-month
3              Yes         Yes          No              No        One year
4               No          No          No              No  Month-to-month

  PaperlessBilling              PaymentMethod  MonthlyCharges TotalCharges  \
0              Yes           Electronic check           29.85        29.85
1               No               Mailed check           56.95       1889.5
2              Yes               Mailed check           53.85       108.15
3               No  Bank transfer (automatic)           42.30      1840.75
4              Yes           Electronic check           70.70       151.65

  Churn
0    No
1    No
2   Yes
3    No
4   Yes
```

[10]:
```
df['TotalCharges'] = pd.to_numeric(df.TotalCharges, errors='coerce')
df.isnull().sum()
```

[10]:
```
gender               0
SeniorCitizen        0
Partner              0
Dependents           0
tenure               0
PhoneService         0
MultipleLines        0
InternetService      0
OnlineSecurity       0
```

```
OnlineBackup         0
DeviceProtection     0
TechSupport          0
StreamingTV          0
StreamingMovies      0
Contract             0
PaperlessBilling     0
PaymentMethod        0
MonthlyCharges       0
TotalCharges        11
Churn                0
dtype: int64
```

[11]:

```
df[np.isnan(df['TotalCharges'])]
```

[11]:

```
      gender  SeniorCitizen Partner Dependents  tenure PhoneService  \
488   Female              0     Yes        Yes       0           No
753     Male              0      No        Yes       0          Yes
936   Female              0     Yes        Yes       0          Yes
1082    Male              0     Yes        Yes       0          Yes
1340  Female              0     Yes        Yes       0           No
3331    Male              0     Yes        Yes       0          Yes
3826    Male              0     Yes        Yes       0          Yes
4380  Female              0     Yes        Yes       0          Yes
5218    Male              0     Yes        Yes       0          Yes
6670  Female              0     Yes        Yes       0          Yes
6754    Male              0      No        Yes       0          Yes

         MultipleLines InternetService       OnlineSecurity  \
488   No phone service             DSL                  Yes
753                 No              No  No internet service
936                 No             DSL                  Yes
1082               Yes              No  No internet service
1340  No phone service             DSL                  Yes
3331                No              No  No internet service
3826               Yes              No  No internet service
4380                No              No  No internet service
5218                No              No  No internet service
6670               Yes             DSL                   No
6754               Yes             DSL                  Yes

             OnlineBackup     DeviceProtection          TechSupport  \
488                    No                  Yes                  Yes
753   No internet service  No internet service  No internet service
936                   Yes                  Yes                   No
1082  No internet service  No internet service  No internet service
1340                  Yes                  Yes                  Yes
```

```
3331  No internet service  No internet service  No internet service
3826  No internet service  No internet service  No internet service
4380  No internet service  No internet service  No internet service
5218  No internet service  No internet service  No internet service
6670                  Yes                  Yes                  Yes
6754                  Yes                   No                  Yes

              StreamingTV      StreamingMovies  Contract PaperlessBilling  \
488                   Yes                   No  Two year              Yes
753   No internet service  No internet service  Two year               No
936                   Yes                  Yes  Two year               No
1082  No internet service  No internet service  Two year               No
1340                  Yes                   No  Two year               No
3331  No internet service  No internet service  Two year               No
3826  No internet service  No internet service  Two year               No
4380  No internet service  No internet service  Two year               No
5218  No internet service  No internet service  One year              Yes
6670                  Yes                   No  Two year               No
6754                   No                   No  Two year              Yes

                  PaymentMethod  MonthlyCharges  TotalCharges Churn
488   Bank transfer (automatic)           52.55           NaN    No
753                Mailed check           20.25           NaN    No
936                Mailed check           80.85           NaN    No
1082               Mailed check           25.75           NaN    No
1340    Credit card (automatic)           56.05           NaN    No
3331               Mailed check           19.85           NaN    No
3826               Mailed check           25.35           NaN    No
4380               Mailed check           20.00           NaN    No
5218               Mailed check           19.70           NaN    No
6670               Mailed check           73.35           NaN    No
6754  Bank transfer (automatic)           61.90           NaN    No
```

[12]:

```
df[df['tenure'] == 0].index
```

[12]:

```
Int64Index([488, 753, 936, 1082, 1340, 3331, 3826, 4380, 5218, 6670, 6754],
dtype='int64')
```

[13]:

```
df.drop(labels=df[df['tenure'] == 0].index, axis=0, inplace=True)
df[df['tenure'] == 0].index
```

[13]:

```
Int64Index([], dtype='int64')
```

[14]:

```
df.fillna(df["TotalCharges"].mean())
```

[14]:

```
      gender  SeniorCitizen Partner Dependents  tenure PhoneService  \
0     Female              0     Yes         No       1           No
```

```
1       Male              0      No         No      34           Yes
2       Male              0      No         No       2           Yes
3       Male              0      No         No      45            No
4     Female              0      No         No       2           Yes
…      …            …     …        …     …          …
7038    Male              0     Yes        Yes      24           Yes
7039  Female              0     Yes        Yes      72           Yes
7040  Female              0     Yes        Yes      11            No
7041    Male              1     Yes         No       4           Yes
7042    Male              0      No         No      66           Yes

         MultipleLines InternetService OnlineSecurity OnlineBackup  \
0     No phone service             DSL             No          Yes
1                   No             DSL            Yes           No
2                   No             DSL            Yes          Yes
3     No phone service             DSL            Yes           No
4                   No     Fiber optic             No           No
…                …             …            …          …
7038               Yes             DSL            Yes           No
7039               Yes     Fiber optic             No          Yes
7040  No phone service             DSL            Yes           No
7041               Yes     Fiber optic             No           No
7042                No     Fiber optic            Yes           No

     DeviceProtection TechSupport StreamingTV StreamingMovies        Contract  \
0                  No          No          No              No  Month-to-month
1                 Yes          No          No              No        One year
2                  No          No          No              No  Month-to-month
3                 Yes         Yes          No              No        One year
4                  No          No          No              No  Month-to-month
…               …         …         …             …             …
7038              Yes         Yes         Yes             Yes        One year
7039              Yes          No         Yes             Yes        One year
7040               No          No          No              No  Month-to-month
7041               No          No          No              No  Month-to-month
7042              Yes         Yes         Yes             Yes        Two year

     PaperlessBilling              PaymentMethod  MonthlyCharges  \
0                 Yes           Electronic check           29.85
1                  No               Mailed check           56.95
2                 Yes               Mailed check           53.85
3                  No  Bank transfer (automatic)           42.30
4                 Yes           Electronic check           70.70
…               …                        …             …
7038              Yes               Mailed check           84.80
7039              Yes    Credit card (automatic)          103.20
7040              Yes           Electronic check           29.60
```

```
7041              Yes               Mailed check           74.40
7042              Yes  Bank transfer (automatic)          105.65

      TotalCharges Churn
0            29.85    No
1          1889.50    No
2           108.15   Yes
3          1840.75    No
4           151.65   Yes
…            …   …
7038       1990.50    No
7039       7362.90    No
7040        346.45    No
7041        306.60   Yes
7042       6844.50    No

[7032 rows x 20 columns]
```

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

```
[15]: gender              0
      SeniorCitizen       0
      Partner             0
      Dependents          0
      tenure              0
      PhoneService        0
      MultipleLines       0
      InternetService     0
      OnlineSecurity      0
      OnlineBackup        0
      DeviceProtection    0
      TechSupport         0
      StreamingTV         0
      StreamingMovies     0
      Contract            0
      PaperlessBilling    0
      PaymentMethod       0
      MonthlyCharges      0
      TotalCharges        0
      Churn               0
      dtype: int64
```

```
[16]: df["SeniorCitizen"]= df["SeniorCitizen"].map({0: "No", 1: "Yes"})
      df.head()
```

```
[16]:    gender SeniorCitizen Partner Dependents  tenure PhoneService  \
      0  Female            No     Yes         No       1           No
```

```
1    Male            No      No         No      34          Yes
2    Male            No      No         No       2          Yes
3    Male            No      No         No      45           No
4  Female            No      No         No       2          Yes

      MultipleLines InternetService OnlineSecurity OnlineBackup  \
0  No phone service             DSL             No          Yes
1                No             DSL            Yes           No
2                No             DSL            Yes          Yes
3  No phone service             DSL            Yes           No
4                No     Fiber optic             No           No

  DeviceProtection TechSupport StreamingTV StreamingMovies        Contract  \
0               No          No          No              No  Month-to-month
1              Yes          No          No              No        One year
2               No          No          No              No  Month-to-month
3              Yes         Yes          No              No        One year
4               No          No          No              No  Month-to-month

  PaperlessBilling              PaymentMethod  MonthlyCharges  TotalCharges  \
0              Yes           Electronic check           29.85         29.85
1               No               Mailed check           56.95       1889.50
2              Yes               Mailed check           53.85        108.15
3               No  Bank transfer (automatic)           42.30       1840.75
4              Yes           Electronic check           70.70        151.65

  Churn
0    No
1    No
2   Yes
3    No
4   Yes
```

[17]:

```
df["InternetService"].describe(include=['object', 'bool'])
```

[17]:

```
count             7032
unique               3
top       Fiber optic
freq              3096
Name: InternetService, dtype: object
```

[18]:

```
numerical_cols = ['tenure', 'MonthlyCharges', 'TotalCharges']
df[numerical_cols].describe()
```

[18]:

```
            tenure  MonthlyCharges  TotalCharges
count  7032.000000     7032.000000   7032.000000
mean     32.421786       64.798208   2283.300441
```

```
std      24.545260       30.085974   2266.771362
min       1.000000       18.250000     18.800000
25%       9.000000       35.587500    401.450000
50%      29.000000       70.350000   1397.475000
75%      55.000000       89.862500   3794.737500
max      72.000000      118.750000   8684.800000
```