In [27]:
```
import pandas as pd
import numpy as np
```

In [28]:
```
df = pd.read_csv(r'C:\Users\Churn_test.csv')
```

In [29]:
```
import os
os.getcwd()
```

Out[29]:
```
'C:\\Users\\user'
```

In [30]:
```
df
```

Out[30]:

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines
0	8879-ZKJOF	Female	0	No	No	41	Yes	No
1	0201-MIBOL	Female	1	No	No	66	Yes	Yes
2	1600-DILPE	Female	0	No	No	12	Yes	No
3	8601-QACRS	Female	0	No	No	5	Yes	Yes
4	7919-ZODZZ	Female	0	Yes	Yes	10	Yes	No
...	...	...	...	...	...	...	...	...
1404	5130-IEKQT	Male	1	No	No	25	Yes	Yes
1405	4452-ROHMO	Female	0	No	No	15	Yes	No
1406	6164-HAQTX	Male	0	No	No	71	No	No phone service
1407	3982-DQLUS	Male	1	Yes	Yes	65	Yes	Yes
1408	9874-QLCLH	Female	0	Yes	Yes	17	Yes	Yes

1409 rows × 20 columns

In [31]:
```
df2 = pd.read_csv(r'C:\Users\Churn_train.csv')
```

In [32]:
```
df2
```

Out[32]:

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines
0	5442-PPTJY	Male	0	Yes	Yes	12	Yes	No
1	6261-RCVNS	Female	0	No	No	42	Yes	No
2	2176-OSJUV	Male	0	Yes	No	71	Yes	Yes
3	6161-ERDGD	Male	0	Yes	Yes	71	Yes	Yes
4	2364-UFROM	Male	0	No	No	30	Yes	No
...	...	...	...	...	...	...	...	...
5629	0781-LKXBR	Male	1	No	No	9	Yes	Yes
5630	3507-GASNP	Male	0	No	Yes	60	Yes	No
5631	8868-WOZGU	Male	0	No	No	28	Yes	Yes
5632	1251-KRREG	Male	0	No	No	2	Yes	Yes
5633	5840-NVDCG	Female	0	Yes	Yes	16	Yes	No

5634 rows × 21 columns

In [33]:
```
df2.describe()
```

Out[33]:

	SeniorCitizen	tenure	MonthlyCharges
count	5634.000000	5634.000000	5634.000000
mean	0.161874	32.277955	64.779127
std	0.368368	24.555211	30.104993
min	0.000000	0.000000	18.250000
25%	0.000000	9.000000	35.400000
50%	0.000000	29.000000	70.375000
75%	0.000000	55.000000	89.850000
max	1.000000	72.000000	118.650000

In [34]:
```
type(df2)
```

Out[34]:
```
pandas.core.frame.DataFrame
```

In [35]:
```
df2.dtypes
```

Out[35]:
```
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
```

In [36]:
```
df2.head(4)
```

Out[36]:

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	Ir
0	5442-PPTJY	Male	0	Yes	Yes	12	Yes	No	
1	6261-RCVNS	Female	0	No	No	42	Yes	No	
2	2176-OSJUV	Male	0	Yes	No	71	Yes	Yes	
3	6161-ERDGD	Male	0	Yes	Yes	71	Yes	Yes	

4 rows × 21 columns

In [37]:
```
df2.columns
```

Out[37]:
```
Index(['customerID', 'gender', 'SeniorCitizen', 'Partner', 'Dependents',
       'tenure', 'PhoneService', 'MultipleLines', 'InternetService',
       'OnlineSecurity', 'OnlineBackup', 'DeviceProtection', 'TechSupport',
       'StreamingTV', 'StreamingMovies', 'Contract', 'PaperlessBilling',
       'PaymentMethod', 'MonthlyCharges', 'TotalCharges', 'Churn'],
      dtype='object')
```

In [38]:
```
df2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5634 entries, 0 to 5633
Data columns (total 21 columns):
 #   Column            Non-Null Count  Dtype  
---  ------            --------------  -----  
 0   customerID        5634 non-null   object 
 1   gender            5634 non-null   object 
 2   SeniorCitizen     5634 non-null   int64  
 3   Partner           5634 non-null   object 
 4   Dependents        5634 non-null   object 
 5   tenure            5634 non-null   int64  
 6   PhoneService      5634 non-null   object 
 7   MultipleLines     5634 non-null   object 
 8   InternetService   5634 non-null   object 
 9   OnlineSecurity    5634 non-null   object 
 10  OnlineBackup      5634 non-null   object 
 11  DeviceProtection  5634 non-null   object 
 12  TechSupport       5634 non-null   object 
 13  StreamingTV       5634 non-null   object 
 14  StreamingMovies   5634 non-null   object 
 15  Contract          5634 non-null   object 
 16  PaperlessBilling  5634 non-null   object 
 17  PaymentMethod     5634 non-null   object 
 18  MonthlyCharges    5634 non-null   float64
 19  TotalCharges      5634 non-null   object 
 20  Churn             5634 non-null   object 
dtypes: float64(1), int64(2), object(18)
memory usage: 924.5+ KB
```

In [39]:
```
num_male=sum(df2['gender'] == 'Male')
```

In [40]:
```
num_male
```

Out[40]: 2838

In [41]:
```
num_female=sum(df2['gender'] == 'Female')
num_female
```

Out[41]: 2796

In [42]:
```
num_equal=sum(df2['gender'] == 'Female')
num_equal
```

Out[42]: 2796

In [43]:
```
df2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5634 entries, 0 to 5633
Data columns (total 21 columns):
 #   Column            Non-Null Count  Dtype  
---  ------            --------------  -----  
 0   customerID        5634 non-null   object 
 1   gender            5634 non-null   object 
 2   SeniorCitizen     5634 non-null   int64  
 3   Partner           5634 non-null   object 
 4   Dependents        5634 non-null   object 
 5   tenure            5634 non-null   int64  
 6   PhoneService      5634 non-null   object 
 7   MultipleLines     5634 non-null   object 
 8   InternetService   5634 non-null   object 
 9   OnlineSecurity    5634 non-null   object 
 10  OnlineBackup      5634 non-null   object 
 11  DeviceProtection  5634 non-null   object 
 12  TechSupport       5634 non-null   object 
 13  StreamingTV       5634 non-null   object 
 14  StreamingMovies   5634 non-null   object 
 15  Contract          5634 non-null   object 
 16  PaperlessBilling  5634 non-null   object 
 17  PaymentMethod     5634 non-null   object 
 18  MonthlyCharges    5634 non-null   float64
 19  TotalCharges      5634 non-null   object 
 20  Churn             5634 non-null   object 
dtypes: float64(1), int64(2), object(18)
memory usage: 924.5+ KB
```

In [44]:
```
df2.columns
```

Out[44]:
```
Index(['customerID', 'gender', 'SeniorCitizen', 'Partner', 'Dependents',
       'tenure', 'PhoneService', 'MultipleLines', 'InternetService',
       'OnlineSecurity', 'OnlineBackup', 'DeviceProtection', 'TechSupport',
       'StreamingTV', 'StreamingMovies', 'Contract', 'PaperlessBilling',
       'PaymentMethod', 'MonthlyCharges', 'TotalCharges', 'Churn'],
      dtype='object')
```

In [45]:
```
max_monthlycharge=max(df2['MonthlyCharges'])
```

In [46]:
```
max_monthlycharge
```

Out[46]:
```
118.65
```

In [47]:
```
import matplotlib.pyplot as plt
```

In [48]:
```
import seaborn as sns
```

In [49]:
```
df2['Churn'].value_counts()
```

Out[49]:
```
No     4113
Yes    1521
Name: Churn, dtype: int64
```

In [50]:
```
df2.columns
```

Out[50]:
```
Index(['customerID', 'gender', 'SeniorCitizen', 'Partner', 'Dependents',
       'tenure', 'PhoneService', 'MultipleLines', 'InternetService',
       'OnlineSecurity', 'OnlineBackup', 'DeviceProtection', 'TechSupport',
       'StreamingTV', 'StreamingMovies', 'Contract', 'PaperlessBilling',
       'PaymentMethod', 'MonthlyCharges', 'TotalCharges', 'Churn'],
      dtype='object')
```

In [26]:
```
df2.columns('Contract')
```

```
---------------------------------------------------------------------------
TypeError                                 Traceback (most recent call last)
Input In [26], in <cell line: 1>()
----> 1 df2.columns('Contract')

TypeError: 'Index' object is not callable
```

In [ ]:
```
df2.head(2)
```

In [ ]:
```
min_tenure= min(df2['tenure'])
max_tenure= max(df2['tenure'])
min_tenure
max_tenure
```

In [ ]:
```
max_tenure
```

In [ ]:
```
sns.pairplot(df2);
```

In [51]:
```
df2.columns
```

Out[51]:
```
Index(['customerID', 'gender', 'SeniorCitizen', 'Partner', 'Dependents',
       'tenure', 'PhoneService', 'MultipleLines', 'InternetService',
       'OnlineSecurity', 'OnlineBackup', 'DeviceProtection', 'TechSupport',
       'StreamingTV', 'StreamingMovies', 'Contract', 'PaperlessBilling',
       'PaymentMethod', 'MonthlyCharges', 'TotalCharges', 'Churn'],
      dtype='object')
```

In [56]:
```
df4 = df2[['customerID', 'Churn', index= False] ]
```

```
  Input In [56]
    df4 = df2[['customerID', 'Churn', index= False] ]
                                                   ^

SyntaxError: invalid syntax
```

In [57]:
```
df4
```

Out[57]:

	customerID	Churn
0	5442-PPTJY	No
1	6261-RCVNS	Yes
2	2176-OSJUV	No
3	6161-ERDGD	No
4	2364-UFROM	No
...	...	...
5629	0781-LKXBR	Yes
5630	3507-GASNP	No
5631	8868-WOZGU	Yes
5632	1251-KRREG	Yes
5633	5840-NVDCG	No

5634 rows × 2 columns

In [64]:
```
df5 = df4.head(1409)
```

In [65]:
```
df5
```

Out[65]:

	customerID	Churn
0	5442-PPTJY	No
1	6261-RCVNS	Yes
2	2176-OSJUV	No
3	6161-ERDGD	No
4	2364-UFROM	No
...	...	...
1404	7838-LAZFO	No
1405	9975-SKRNR	No
1406	9483-GCPWE	No
1407	0674-EYYZV	No
1408	2511-ALLCS	No

1409 rows × 2 columns

In [66]:
```
df5.to_csv('customer churn prediction.csv', index = False, header = False)
```

In [ ]:
```

```

In [ ]:

In [ ]:

In [ ]: