```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

df = pd.read_csv('telecom_churn.csv')

df
```

```
        customer_id telecom_partner gender  age              state
city  \
0                 1    Reliance Jio      F   25          Karnataka
Kolkata
1                 2    Reliance Jio      F   55            Mizoram
Mumbai
2                 3        Vodafone      F   57  Arunachal Pradesh
Delhi
3                 4            BSNL      M   46         Tamil Nadu
Kolkata
4                 5            BSNL      F   26            Tripura
Delhi
...             ...             ...    ...  ...                ...
...
243548       243549          Airtel      F   28            Mizoram
Kolkata
243549       243550    Reliance Jio      F   52              Assam
Kolkata
243550       243551    Reliance Jio      M   59            Tripura
Kolkata
243551       243552            BSNL      M   49     Madhya Pradesh
Kolkata
243552       243553            BSNL      F   37          Telangana
Hyderabad

        pincode date_of_registration  num_dependents  estimated_salary
\
0        755597           2020-01-01               4            124962

1        125926           2020-01-01               2            130556

2        423976           2020-01-01               0            148828

3        522841           2020-01-01               1             38722

4        740247           2020-01-01               2             55098

...         ...                  ...             ...               ...

243548   110295           2023-05-03               3            130580

243549   713481           2023-05-03               0             82393
```

```
243550   520218           2023-05-03               4              51298

243551   387744           2023-05-03               2              83981

243552   139086           2023-05-04               0             144297


        calls_made  sms_sent  data_used  churn
0               44        45       -361      0
1               62        39       5973      0
2               49        24        193      1
3               80        25       9377      1
4               78        15       1393      0
...            ...       ...        ...    ...
243548          28         9       4102      0
243549          80        45       7521      0
243550          26         4       6547      0
243551          80        15       1125      0
243552          61         7       3384      0

[243553 rows x 14 columns]
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 243553 entries, 0 to 243552
Data columns (total 14 columns):
 #   Column                Non-Null Count   Dtype 
---  ------                --------------   ----- 
 0   customer_id           243553 non-null  int64 
 1   telecom_partner       243553 non-null  object
 2   gender                243553 non-null  object
 3   age                   243553 non-null  int64 
 4   state                 243553 non-null  object
 5   city                  243553 non-null  object
 6   pincode               243553 non-null  int64 
 7   date_of_registration  243553 non-null  object
 8   num_dependents        243553 non-null  int64 
 9   estimated_salary      243553 non-null  int64 
 10  calls_made            243553 non-null  int64 
 11  sms_sent              243553 non-null  int64 
 12  data_used             243553 non-null  int64 
 13  churn                 243553 non-null  int64 
dtypes: int64(9), object(5)
memory usage: 26.0+ MB
```

```
df = pd.read_csv('telecom_churn_data.csv')
```

```
df
```

```
      customerID  gender  SeniorCitizen Partner Dependents  tenure  \
0     7590-VHVEG  Female              0     Yes         No       1
1     5575-GNVDE    Male              0      No         No      34
2     3668-QPYBK    Male              0      No         No       2
3     7795-CFOCW    Male              0      No         No      45
4     9237-HQITU  Female              0      No         No       2
...          ...     ...            ...     ...        ...     ...
7038  6840-RESVB    Male              0     Yes        Yes      24
7039  2234-XADUH  Female              0     Yes        Yes      72
7040  4801-JZAZL  Female              0     Yes        Yes      11
7041  8361-LTMKD    Male              1     Yes         No       4
7042  3186-AJIEK    Male              0      No         No      66

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

     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
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
...               ...                        ...            ...
...
7038              Yes               Mailed check          84.80
1990.5
7039              Yes    Credit card (automatic)         103.20
7362.9
7040              Yes           Electronic check          29.60
346.45
7041              Yes               Mailed check          74.40
306.6
7042              Yes  Bank transfer (automatic)         105.65
6844.5

     Churn
0       No
1       No
2      Yes
3       No
4      Yes
...    ...
7038    No
7039    No
7040    No
7041   Yes
7042    No
```

```
[7043 rows x 21 columns]
```

```
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
```

Replacing blanks with 0 as they were casuing trouble in analysis

```
df['TotalCharges'] = df['TotalCharges'].replace(' ','0')
```

Changing the type of TotalCharges column from object to float

```
df['TotalCharges'] = df['TotalCharges'].astype('float')
```

```
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   float64
 20  Churn             7043 non-null   object 
dtypes: float64(2), int64(2), object(17)
memory usage: 1.1+ MB
```

```
df.isnull().sum()
```

```
customerID          0
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
```

```
df.describe()
```

```
       SeniorCitizen       tenure  MonthlyCharges  TotalCharges
count    7043.000000  7043.000000     7043.000000   7043.000000
mean        0.162147    32.371149       64.761692   2279.734304
std         0.368612    24.559481       30.090047   2266.794470
min         0.000000     0.000000       18.250000      0.000000
25%         0.000000     9.000000       35.500000    398.550000
50%         0.000000    29.000000       70.350000   1394.550000
75%         0.000000    55.000000       89.850000   3786.600000
max         1.000000    72.000000      118.750000   8684.800000
```

```
df.duplicated().sum()
```

```
0
```

```
df['customerID'].duplicated().sum()
```

```
0
```

Converting SeniorCitizen table values where 1 return Yes and where 0 return No

```
def conv(value):
    if value == 1:
        return 'Yes'
    else:
        return 'No'

df['SeniorCitizen'] = df['SeniorCitizen'].apply(conv)


df
```

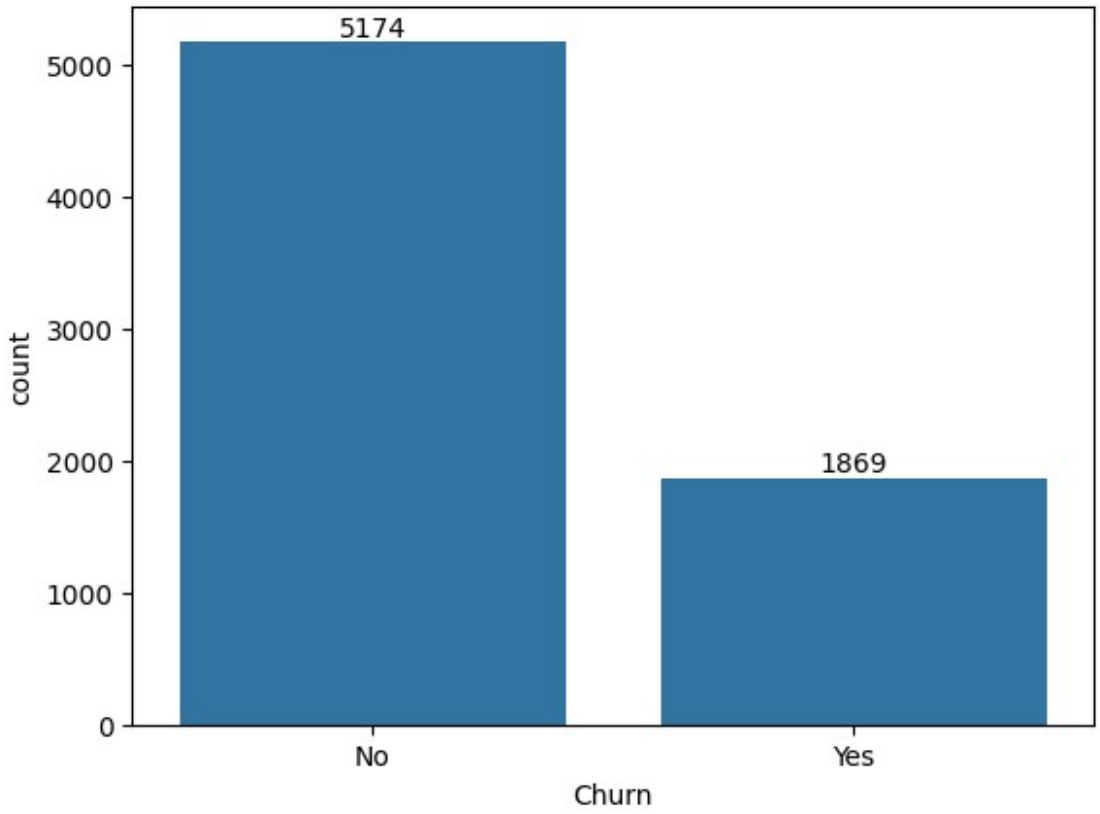
```
      customerID  gender SeniorCitizen Partner Dependents  tenure  \
0     7590-VHVEG  Female            No     Yes         No       1
1     5575-GNVDE    Male            No      No         No      34
2     3668-QPYBK    Male            No      No         No       2
3     7795-CFOCW    Male            No      No         No      45
4     9237-HQITU  Female            No      No         No       2
...          ...     ...           ...     ...        ...     ...
7038  6840-RESVB    Male            No     Yes        Yes      24
```

```
7039  2234-XADUH  Female            No     Yes        Yes      72
7040  4801-JZAZL  Female            No     Yes        Yes      11
7041  8361-LTMKD    Male           Yes     Yes         No       4
7042  3186-AJIEK    Male            No      No         No      66

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

     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-
```

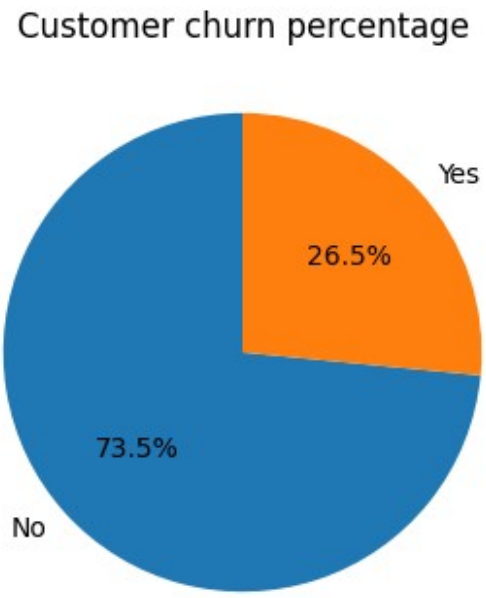
```
to-month
7041               No          No          No               No  Month-
to-month
7042              Yes         Yes         Yes              Yes
Two year

     PaperlessBilling              PaymentMethod MonthlyCharges
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
...               ...                        ...            ...
...
7038              Yes               Mailed check          84.80
1990.50
7039              Yes    Credit card (automatic)         103.20
7362.90
7040              Yes           Electronic check          29.60
346.45
7041              Yes               Mailed check          74.40
306.60
7042              Yes  Bank transfer (automatic)         105.65
6844.50

      Churn
0        No
1        No
2       Yes
3        No
4       Yes
...     ...
7038     No
7039     No
7040     No
7041    Yes
7042     No

[7043 rows x 21 columns]

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   object 
 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   float64
 20  Churn             7043 non-null   object 
dtypes: float64(2), int64(1), object(18)
memory usage: 1.1+ MB
```

```
ax = sns.countplot(x = 'Churn', data = df)
ax.bar_label(ax.containers[0])
plt.show()
```

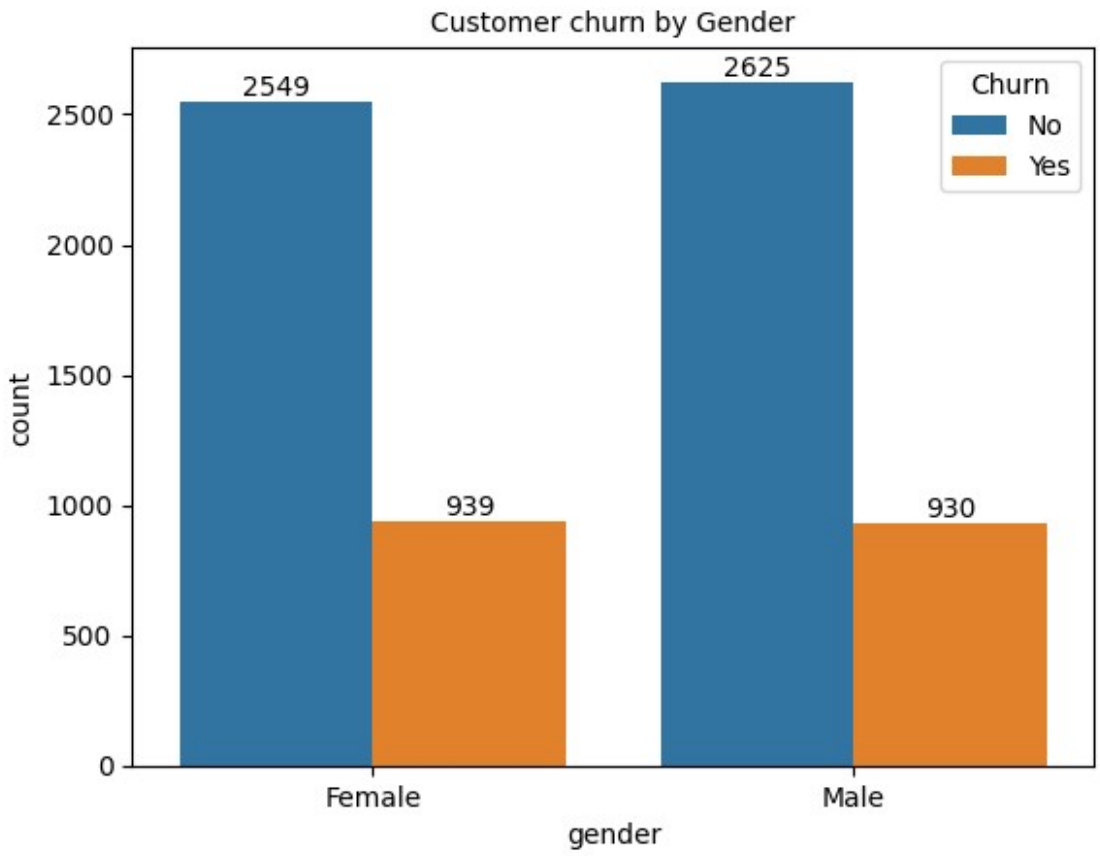



```
plt.figure(figsize=(4,4))
gb = df.groupby('Churn').agg({'Churn':'count'})
gb
plt.pie(gb['Churn'], startangle= 90, labels = gb.index, autopct=
'%1.1f%%')
plt.title("Customer churn percentage")
plt.show()
```

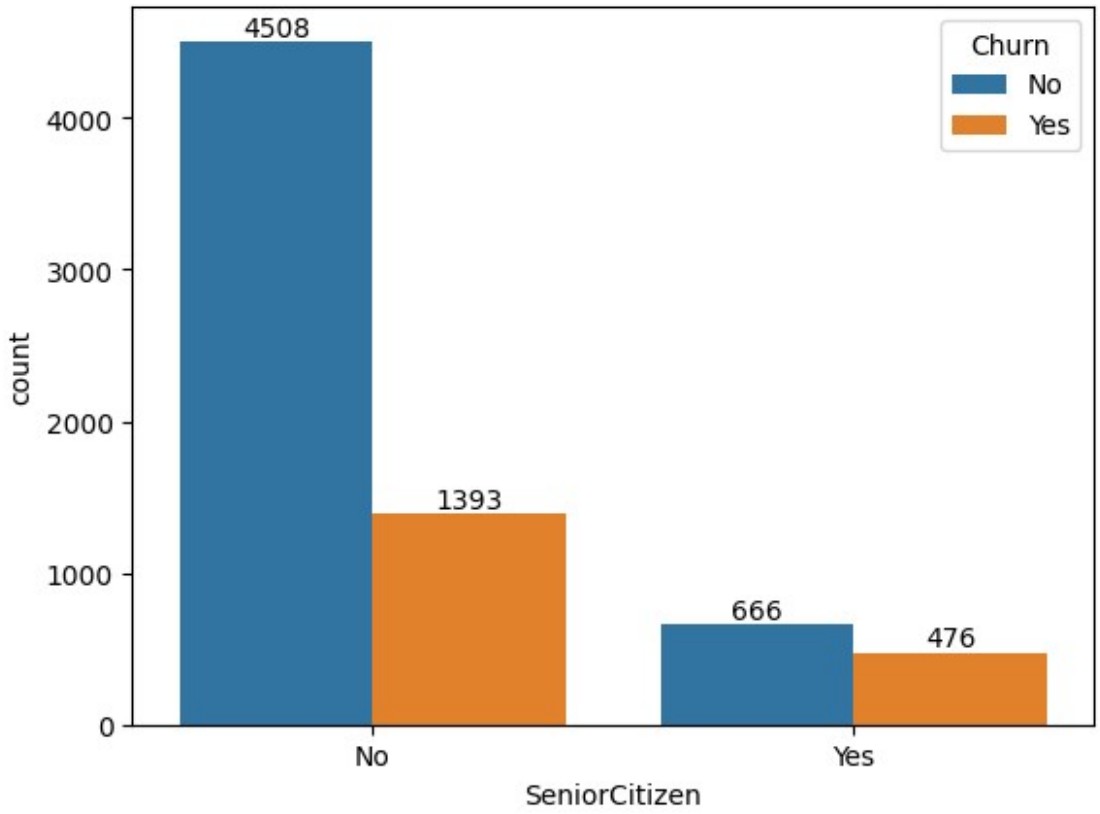


From the above pie chart we can see that 26.5% customer had churned out. Now exploring the reason behind this

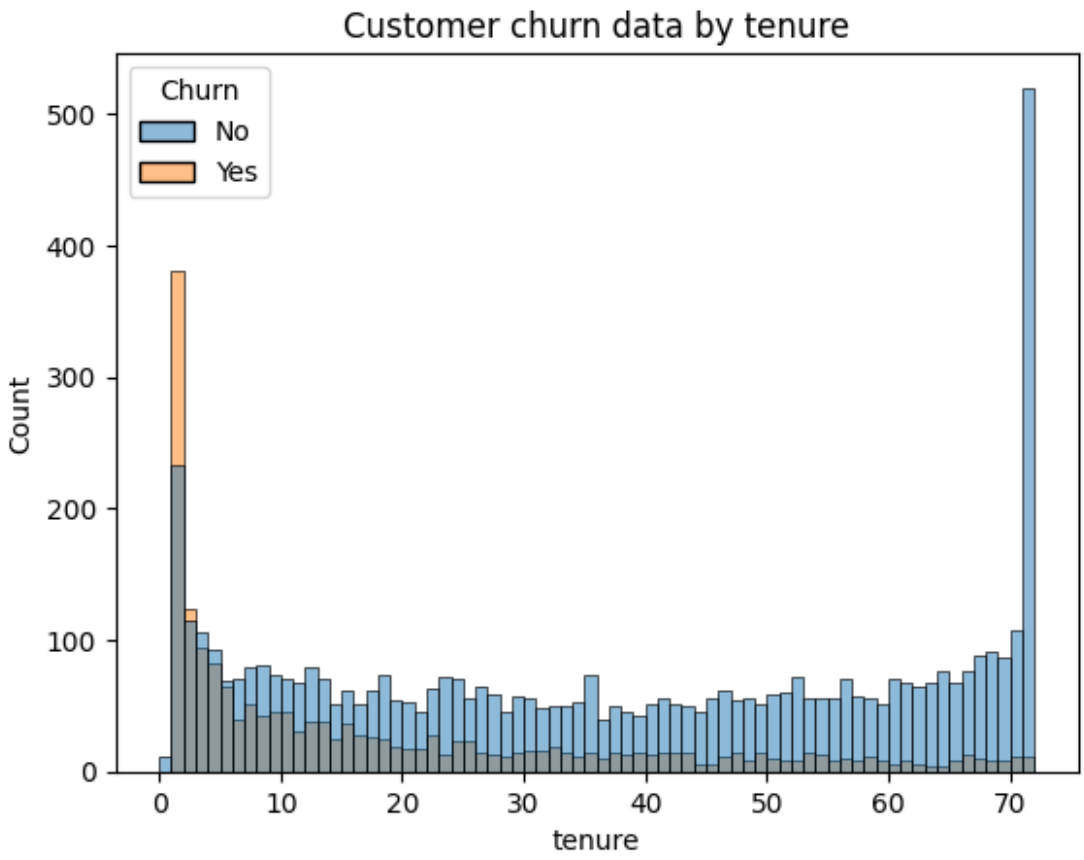
```
ay = sns.countplot(x = 'gender', data = df, hue = 'Churn')
for container in ay.containers:
    ay.bar_label(container)
plt.title("Customer churn by Gender", fontsize = 10)
plt.show()
```




```
az = sns.countplot(x = 'SeniorCitizen', data = df, hue = 'Churn')
for container in az.containers:
    az.bar_label(container)
```

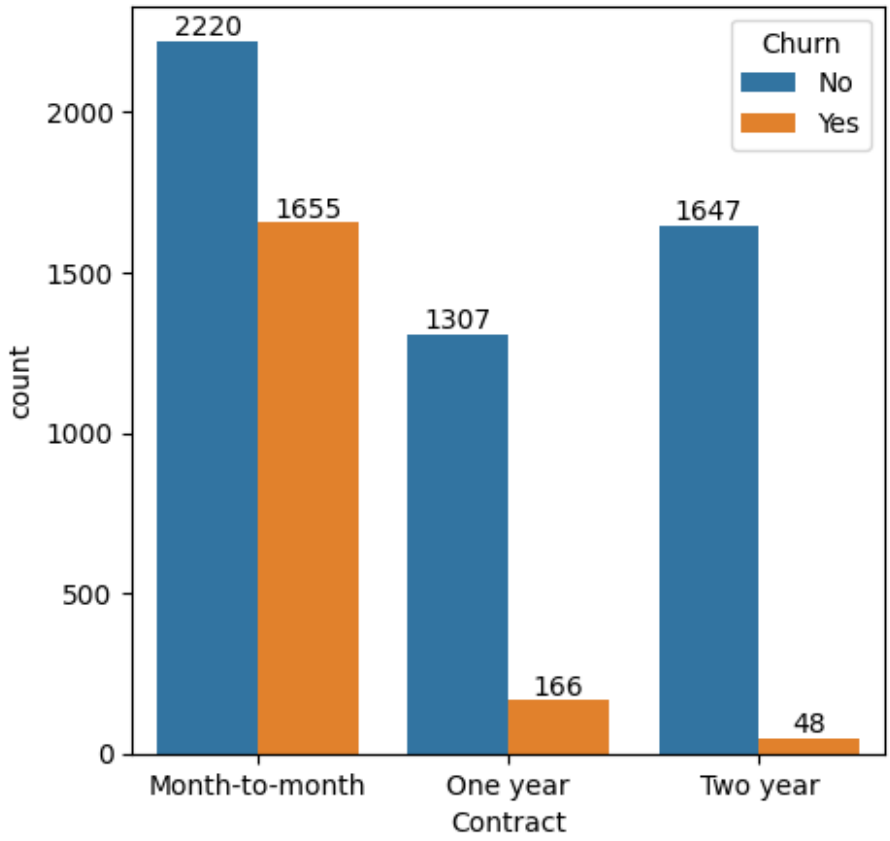



```
sns.histplot(x = 'tenure', data = df, hue = 'Churn', bins = 72)
plt.title("Customer churn data by tenure")
plt.show()
```



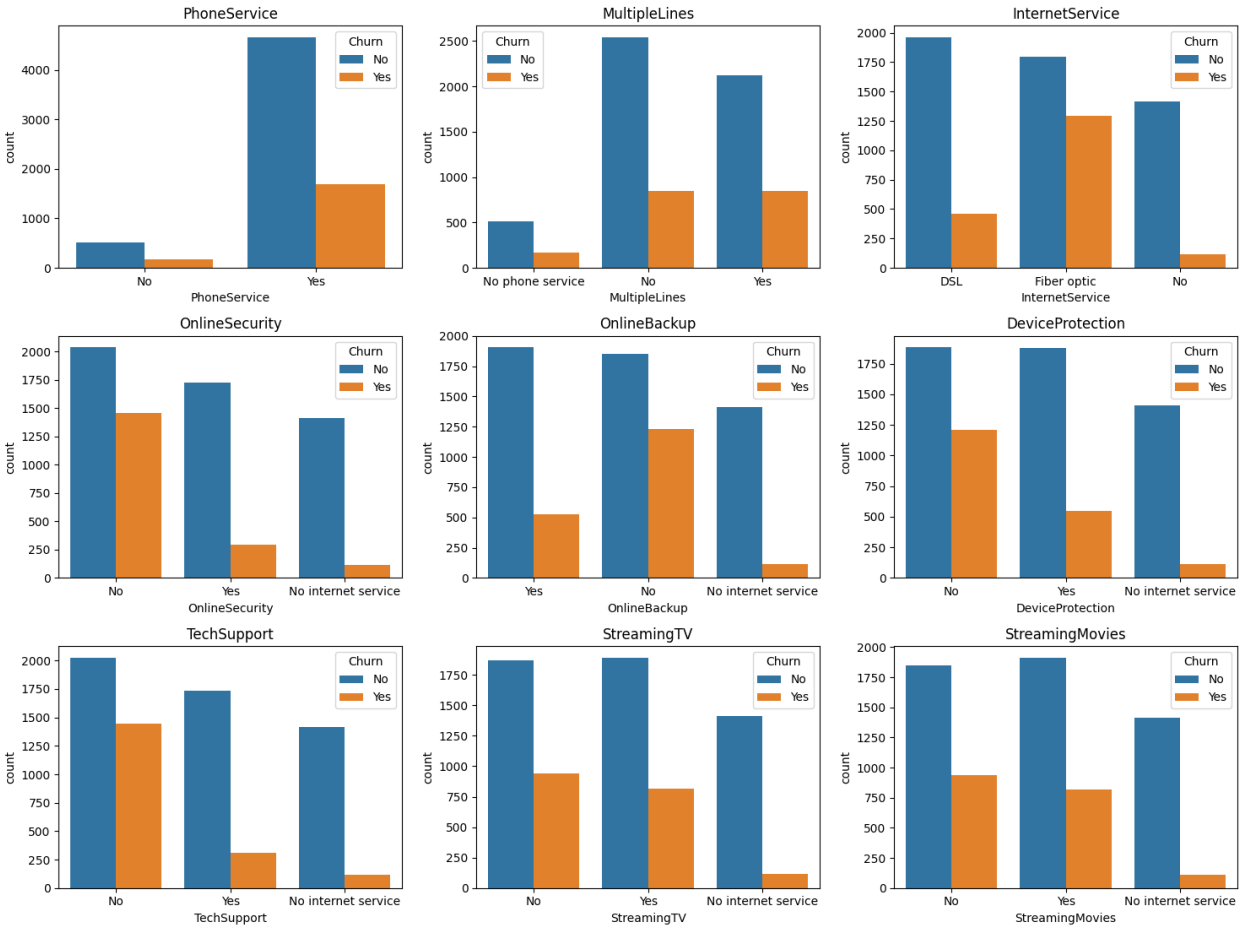
We can see that customer are churning out more in the first initial months, also many customers are retaining.

```
plt.figure(figsize = (5,5))
ab = sns.countplot(x = 'Contract', data = df, hue = 'Churn')
for container in ab.containers:
    ab.bar_label(container)
plt.show()
```




```
df.columns.values

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

# Create subplots (3 rows, 3 columns)
fig, axes = plt.subplots(nrows=3, ncols=3, figsize=(15, 12))
fig.suptitle("Countplots for Different Services", fontsize=16)

# Flatten the axes array for easy iteration
axes = axes.flatten()

# Creating countplots for each individual column
sns.countplot(x=df['PhoneService'], hue = df['Churn'], ax=axes[0])
axes[0].set_title("PhoneService")

sns.countplot(x=df['MultipleLines'],hue = df['Churn'], ax=axes[1])
axes[1].set_title("MultipleLines")
```

```
sns.countplot(x=df['InternetService'],hue = df['Churn'], ax=axes[2])
axes[2].set_title("InternetService")

sns.countplot(x=df['OnlineSecurity'],hue = df['Churn'], ax=axes[3])
axes[3].set_title("OnlineSecurity")

sns.countplot(x=df['OnlineBackup'],hue = df['Churn'], ax=axes[4])
axes[4].set_title("OnlineBackup")

sns.countplot(x=df['DeviceProtection'],hue = df['Churn'], ax=axes[5])
axes[5].set_title("DeviceProtection")

sns.countplot(x=df['TechSupport'],hue = df['Churn'], ax=axes[6])
axes[6].set_title("TechSupport")

sns.countplot(x=df['StreamingTV'],hue = df['Churn'], ax=axes[7])
axes[7].set_title("StreamingTV")

sns.countplot(x=df['StreamingMovies'],hue = df['Churn'], ax=axes[8])
axes[8].set_title("StreamingMovies")

# Adjust layout to prevent overlapping
plt.tight_layout(rect=[0, 0, 1, 0.96])
plt.show()
```


Countplots for Different Services
PhoneService
MultipleLines
InternetService
OnlineSecurity
OnlineBackup
DeviceProtection
TechSupport
StreamingTV
StreamingMovies
Churn
No
Yes
count
No phone service
No internet service
DSL
Fiber optic

Phone & Multiple Lines:

Customers with multiple lines churn more than those with a single line.

Internet Service:

Fiber optic users have the highest churn rate, likely due to pricing or service quality issues.

DSL users churn less, and customers without internet service churn the least.

Online Security & Backup:

Customers without security or backup services churn more than those with these services.

Device Protection & Tech Support:

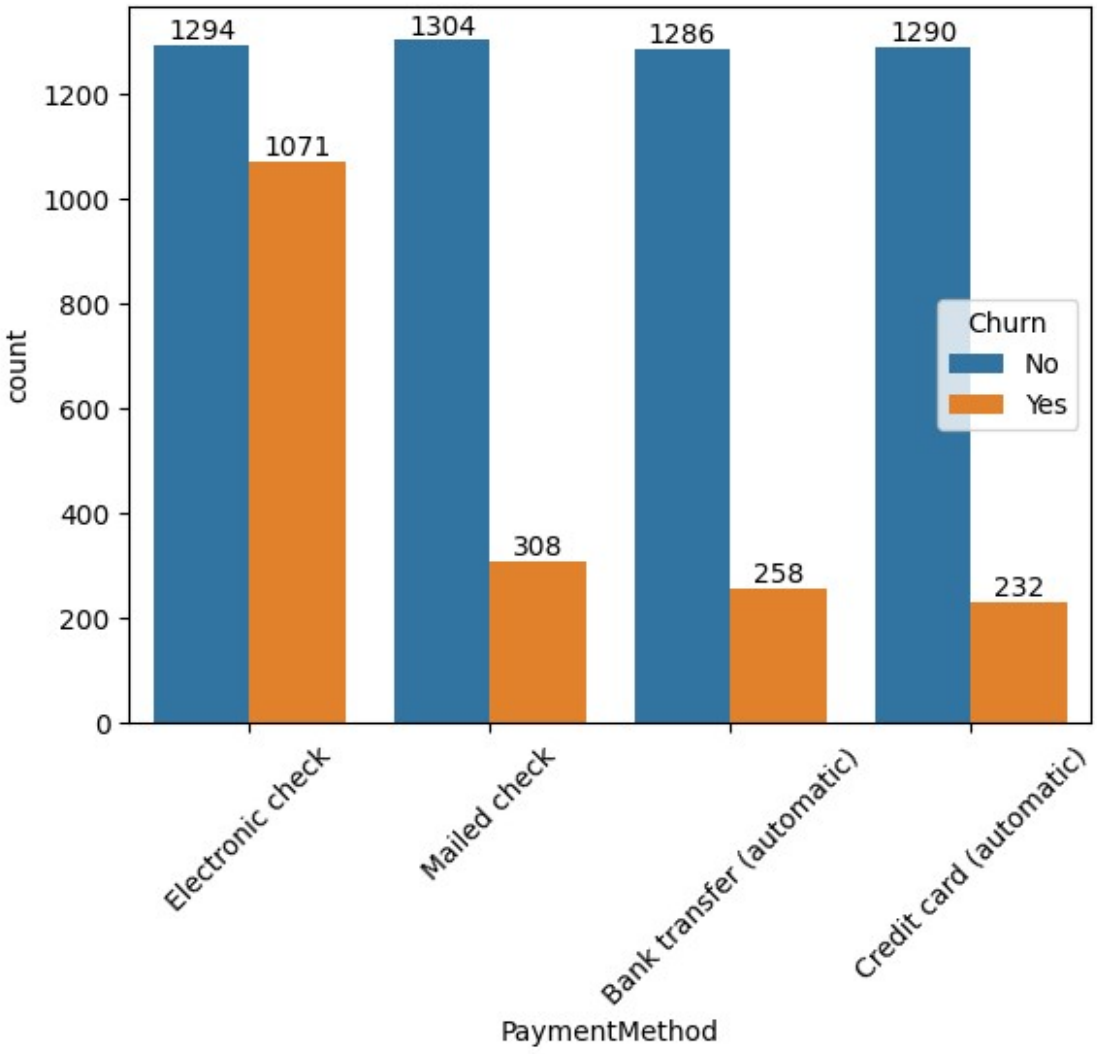
Lack of tech support leads to higher churn good support improves retention.

Device protection reduces churn, indicating its value in customer loyalty.

Streaming TV & Movies:

Customers without streaming services churn more bundling these could improve retention.

```
ac = sns.countplot(x = 'PaymentMethod', data = df, hue = 'Churn')
for container in ac.containers:
    ac.bar_label(container)
plt.xticks(rotation = 45)
plt.show()
```



Cutomers with Electronic check are more churning out followed with Mailed Check