telecomm-churn-analysis

November 11, 2024

[15]:
```
pip install pandas
```

```
Requirement already satisfied: pandas in d:\anaconda3\lib\site-packages (2.2.2)
Requirement already satisfied: numpy>=1.26.0 in d:\anaconda3\lib\site-packages
(from pandas) (1.26.4)
Requirement already satisfied: python-dateutil>=2.8.2 in d:\anaconda3\lib\site-
packages (from pandas) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in d:\anaconda3\lib\site-packages
(from pandas) (2024.1)
Requirement already satisfied: tzdata>=2022.7 in d:\anaconda3\lib\site-packages
(from pandas) (2023.3)
Requirement already satisfied: six>=1.5 in d:\anaconda3\lib\site-packages (from
python-dateutil>=2.8.2->pandas) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

[16]:
```
pip install matplotlib
```

```
Requirement already satisfied: matplotlib in d:\anaconda3\lib\site-packages
(3.8.4)
Requirement already satisfied: contourpy>=1.0.1 in d:\anaconda3\lib\site-
packages (from matplotlib) (1.2.0)
Requirement already satisfied: cycler>=0.10 in d:\anaconda3\lib\site-packages
(from matplotlib) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in d:\anaconda3\lib\site-
packages (from matplotlib) (4.51.0)
Requirement already satisfied: kiwisolver>=1.3.1 in d:\anaconda3\lib\site-
packages (from matplotlib) (1.4.4)
Requirement already satisfied: numpy>=1.21 in d:\anaconda3\lib\site-packages
(from matplotlib) (1.26.4)
Requirement already satisfied: packaging>=20.0 in d:\anaconda3\lib\site-packages
(from matplotlib) (23.2)
Requirement already satisfied: pillow>=8 in d:\anaconda3\lib\site-packages (from
matplotlib) (10.3.0)
Requirement already satisfied: pyparsing>=2.3.1 in d:\anaconda3\lib\site-
packages (from matplotlib) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in d:\anaconda3\lib\site-
packages (from matplotlib) (2.9.0.post0)
Requirement already satisfied: six>=1.5 in d:\anaconda3\lib\site-packages (from
```

```
python-dateutil>=2.7->matplotlib) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

[17]:
```
pip install numpy
```

```
Requirement already satisfied: numpy in d:\anaconda3\lib\site-packages (1.26.4)
Note: you may need to restart the kernel to use updated packages.
```

[18]:
```
pip install seaborn
```

```
Requirement already satisfied: seaborn in d:\anaconda3\lib\site-packages
(0.13.2)
Requirement already satisfied: numpy!=1.24.0,>=1.20 in d:\anaconda3\lib\site-
packages (from seaborn) (1.26.4)
Requirement already satisfied: pandas>=1.2 in d:\anaconda3\lib\site-packages
(from seaborn) (2.2.2)
Requirement already satisfied: matplotlib!=3.6.1,>=3.4 in d:\anaconda3\lib\site-
packages (from seaborn) (3.8.4)
Requirement already satisfied: contourpy>=1.0.1 in d:\anaconda3\lib\site-
packages (from matplotlib!=3.6.1,>=3.4->seaborn) (1.2.0)
Requirement already satisfied: cycler>=0.10 in d:\anaconda3\lib\site-packages
(from matplotlib!=3.6.1,>=3.4->seaborn) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in d:\anaconda3\lib\site-
packages (from matplotlib!=3.6.1,>=3.4->seaborn) (4.51.0)
Requirement already satisfied: kiwisolver>=1.3.1 in d:\anaconda3\lib\site-
packages (from matplotlib!=3.6.1,>=3.4->seaborn) (1.4.4)
Requirement already satisfied: packaging>=20.0 in d:\anaconda3\lib\site-packages
(from matplotlib!=3.6.1,>=3.4->seaborn) (23.2)
Requirement already satisfied: pillow>=8 in d:\anaconda3\lib\site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (10.3.0)
Requirement already satisfied: pyparsing>=2.3.1 in d:\anaconda3\lib\site-
packages (from matplotlib!=3.6.1,>=3.4->seaborn) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in d:\anaconda3\lib\site-
packages (from matplotlib!=3.6.1,>=3.4->seaborn) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in d:\anaconda3\lib\site-packages
(from pandas>=1.2->seaborn) (2024.1)
Requirement already satisfied: tzdata>=2022.7 in d:\anaconda3\lib\site-packages
(from pandas>=1.2->seaborn) (2023.3)
Requirement already satisfied: six>=1.5 in d:\anaconda3\lib\site-packages (from
python-dateutil>=2.7->matplotlib!=3.6.1,>=3.4->seaborn) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

[20]:
```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
```

[25]:
```
df = pd.read_csv(r'C:/Users/asus/Downloads/Customer Churn.csv')
df
```

[25]:

	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]
```

[27]:
```
df.head()
```

[27]:
```
   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]
```

[29]:
```
df.tail()
```

[29]:
```
      customerID  gender  SeniorCitizen Partner Dependents  tenure  \
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  …  \
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  \
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  \
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
7038    No
7039    No
7040    No
7041   Yes
7042    No

[5 rows x 21 columns]
```

[31]:
```
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 tenure is 0 and total charges are recorded

[35]:
```
df["TotalCharges"] = df["TotalCharges"].replace(" ","0")
```

#coverting the data type for TotalCharge column to float

[36]:
```
df["TotalCharges"] = df["TotalCharges"].astype("float")
```

[38]:
```
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
```

[40]:
```
df.isnull()
```

[40]:

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	\
0	False	False	False	False	False	False	
1	False	False	False	False	False	False	
2	False	False	False	False	False	False	
3	False	False	False	False	False	False	
4	False	False	False	False	False	False	
…	…	…	…	…	…	…	
7038	False	False	False	False	False	False	
7039	False	False	False	False	False	False	
7040	False	False	False	False	False	False	
7041	False	False	False	False	False	False	
7042	False	False	False	False	False	False	

	PhoneService	MultipleLines	InternetService	OnlineSecurity	…	\
0	False	False	False	False	…	
1	False	False	False	False	…	
2	False	False	False	False	…	
3	False	False	False	False	…	
4	False	False	False	False	…	
…	…	…	…	…	…	
7038	False	False	False	False	…	
7039	False	False	False	False	…	
7040	False	False	False	False	…	

```
7041         False          False             False           False  …
7042         False          False             False           False  …

      DeviceProtection  TechSupport  StreamingTV  StreamingMovies  Contract  \
0                False        False        False            False     False
1                False        False        False            False     False
2                False        False        False            False     False
3                False        False        False            False     False
4                False        False        False            False     False
…                …            …            …                …         …
7038             False        False        False            False     False
7039             False        False        False            False     False
7040             False        False        False            False     False
7041             False        False        False            False     False
7042             False        False        False            False     False

      PaperlessBilling  PaymentMethod  MonthlyCharges  TotalCharges  Churn
0                False          False           False         False  False
1                False          False           False         False  False
2                False          False           False         False  False
3                False          False           False         False  False
4                False          False           False         False  False
…                …              …               …             …      …
7038             False          False           False         False  False
7039             False          False           False         False  False
7040             False          False           False         False  False
7041             False          False           False         False  False
7042             False          False           False         False  False

[7043 rows x 21 columns]
```

[45]:
```
df.isnull().sum()
```

[45]:
```
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
```

[48]:
```
df.isnull().sum().sum()
```

[48]:
```
0
```

[49]:
```
df.describe()
```

[49]:

	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

[50]:
```
df.duplicated()
```

[50]:
```
0       False
1       False
2       False
3       False
4       False
        …
7038    False
7039    False
7040    False
7041    False
7042    False
Length: 7043, dtype: bool
```

[51]:
```
df.duplicated().sum()
```

[51]:
```
0
```

[52]:
```
df['customerID'].duplicated().sum()
```

[52]:
```
0
```

[53]:
```
def conv(value):
    if value == 1:
        return "Yes"
    else:
        return "No"
```

#converted 0 and 1 values of SeniorCitizen column to Yes and No

[54]:
```
df['SeniorCitizen'] = df['SeniorCitizen'].apply(conv)
```

[57]:
```
df
```

[57]:

	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	

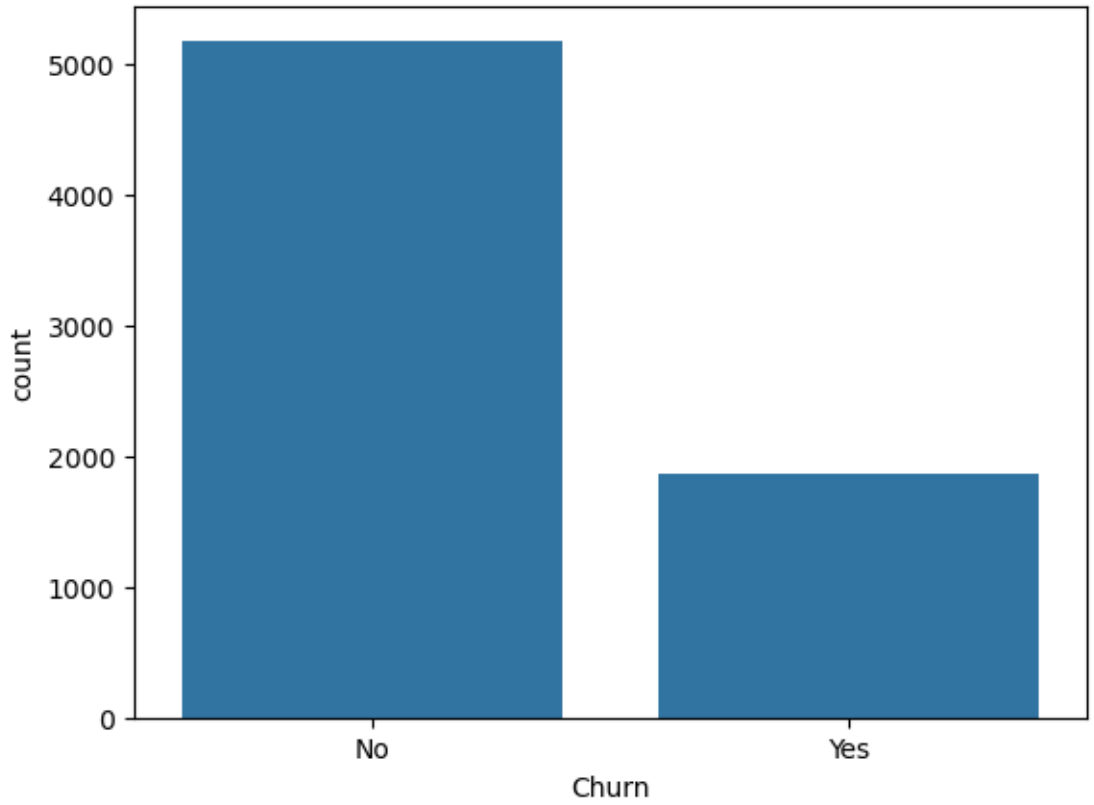
```
7040               No          No          No              No  Month-to-month
7041               No          No          No              No  Month-to-month
7042              Yes         Yes         Yes             Yes        Two year

     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]
```

[58]:
```
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
```

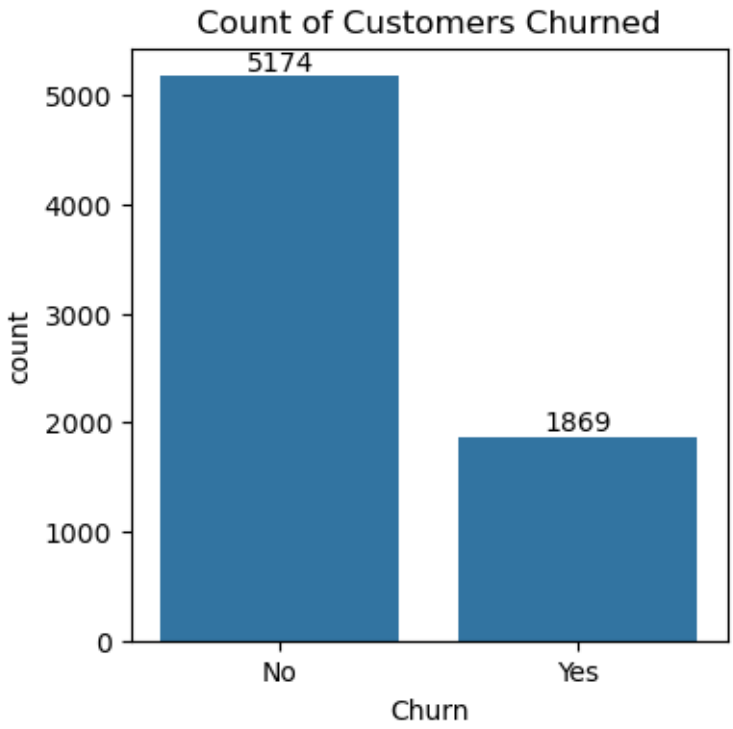
```
 8   InternetService   7043 non-null   object
 9   OnlineSecurity    7043 non-null   object
 10  OnlineBackup      7043 non-null   object
 11  DeviceProtection  7043 non-null   object
 12  TechSupport       7043 non-null   object
 13  StreamingTV       7043 non-null   object
 14  StreamingMovies   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
```

[60]:
```
sns.countplot(x = 'Churn', data = df)
plt.show()
```



[90]:
```
plt.figure(figsize = (4,4))
ax = sns.countplot(x = 'Churn', data = df)
```

```
ax.bar_label(ax.containers[0])
plt.title("Count of Customers Churned")
plt.show()
```



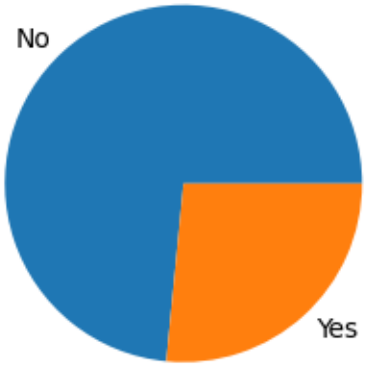
[68]:
```
gb = df.groupby("Churn").agg({'Churn':'count'})
```

[69]:
```
gb
```

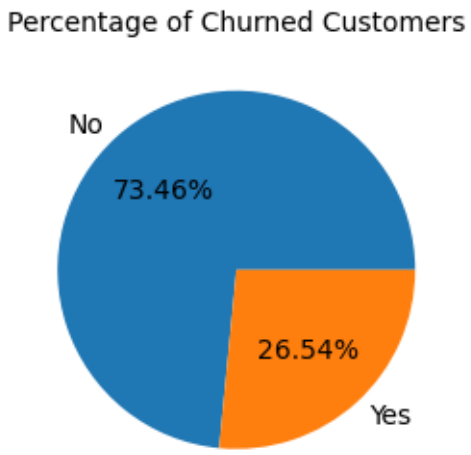
[69]:
```
        Churn
Churn
No       5174
Yes      1869
```

[80]:
```
plt.figure(figsize = (3,4))
plt.pie(gb['Churn'])
plt.show()
```

[82]:
```
plt.figure(figsize = (3,4))
plt.pie(gb['Churn'], labels = gb.index)
plt.show()
```

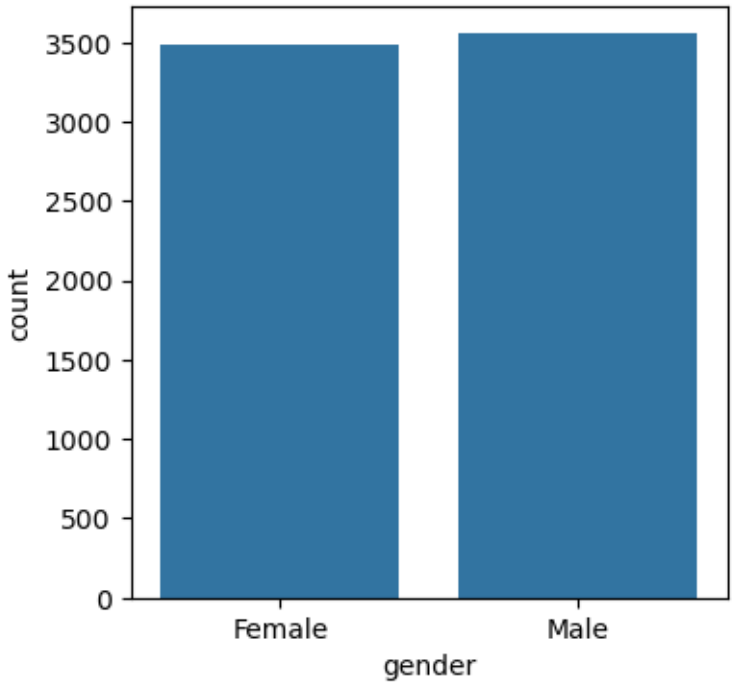


[84]:
```
plt.figure(figsize = (3,4))
plt.pie(gb['Churn'], labels = gb.index, autopct= "%1.2f%%")
plt.title("Percentage of Churned Customers", fontsize = 10)
plt.show()
```

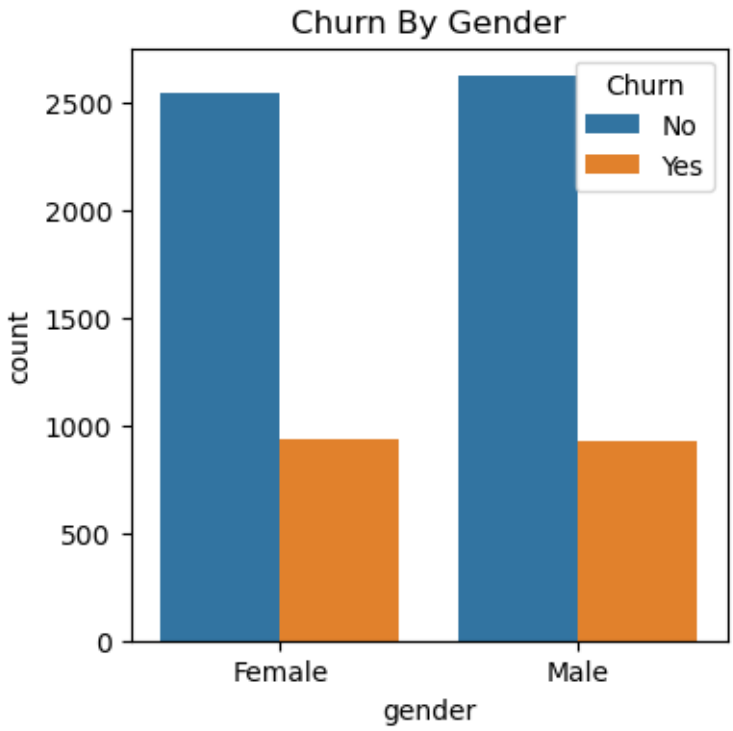


#From the above pie chart we can conclude that 26.54% of the customers have already churned out #Now let's explore the reason behind it.

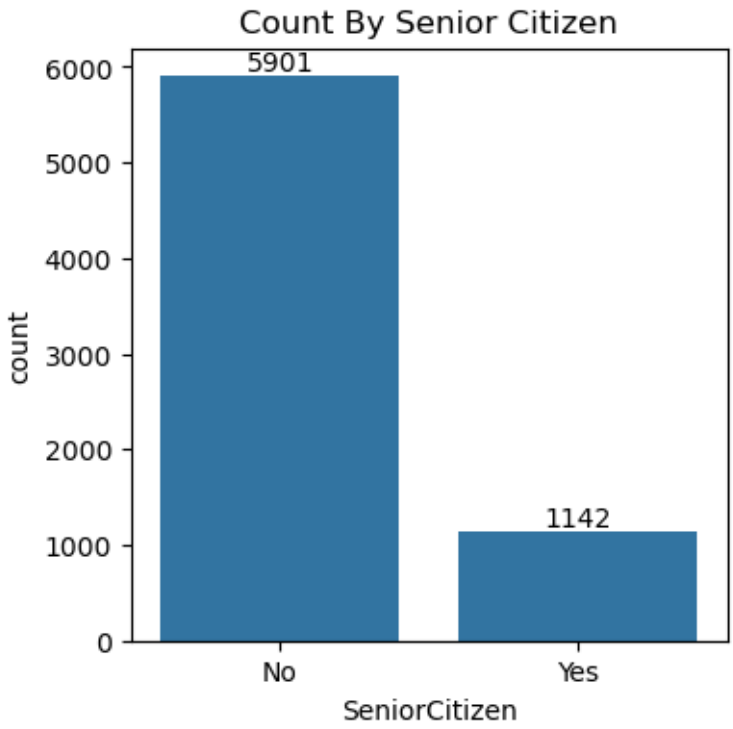
[93]:
```
plt.figure(figsize = (4,4))
sns.countplot(x = "gender", data = df)
plt.show()
```



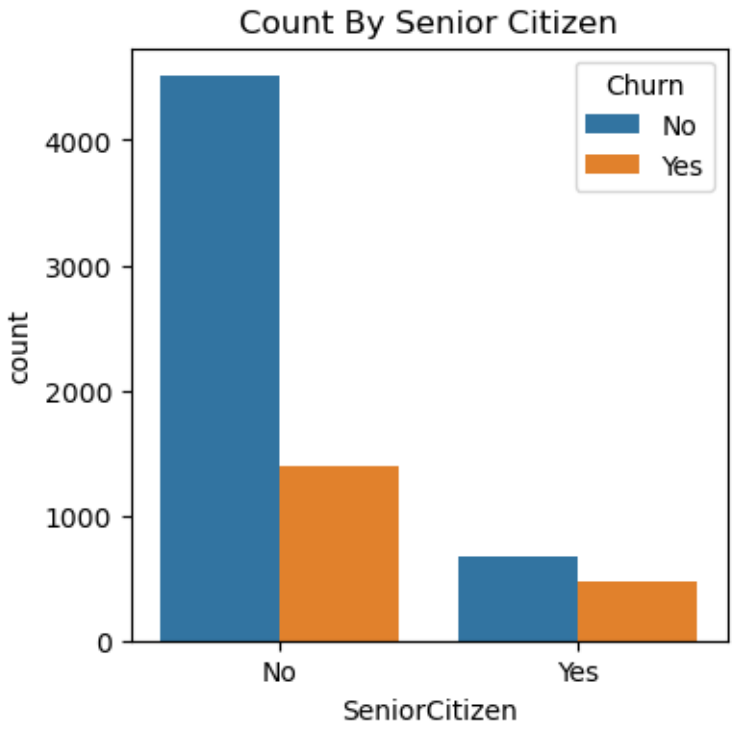
[98]:
```
plt.figure(figsize = (4,4))
sns.countplot(x = "gender", data = df, hue = "Churn")
plt.title("Churn By Gender")
plt.show()
```



[104]:
```
plt.figure(figsize = (4,4))
ax = sns.countplot(x = "SeniorCitizen", data = df)
ax.bar_label(ax.containers[0])
plt.title("Count By Senior Citizen")
plt.show()
```



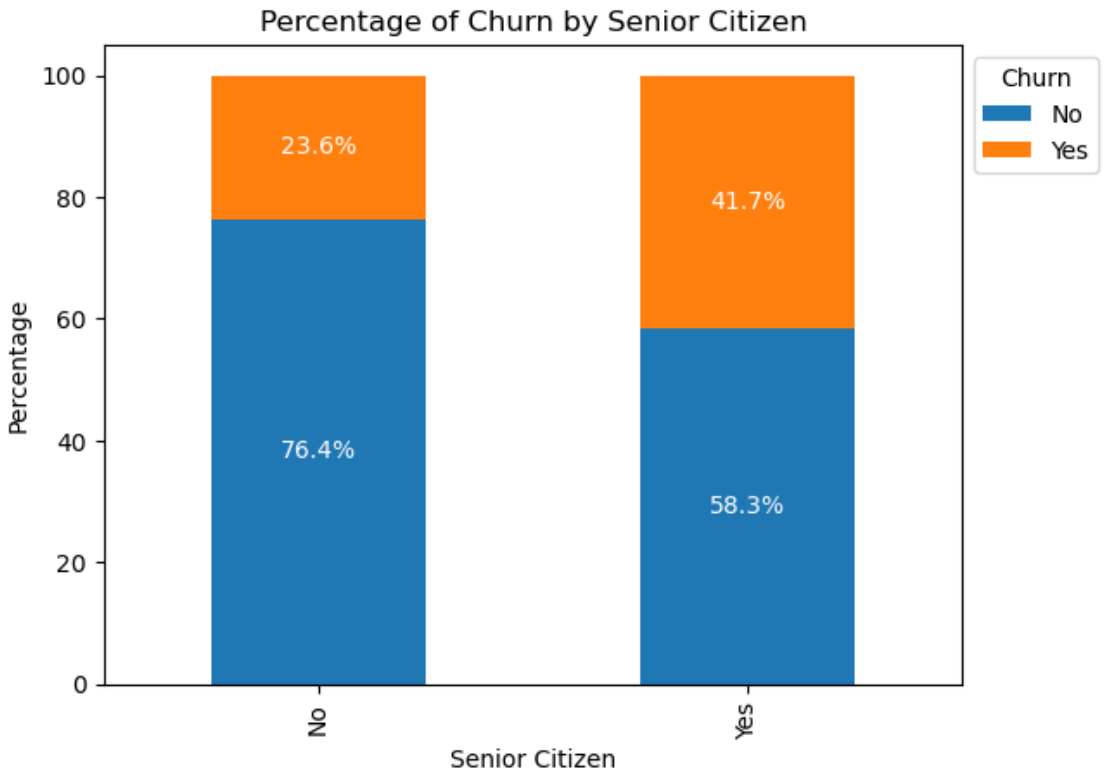
[108]:
```
plt.figure(figsize = (4,4))
sns.countplot(x = "SeniorCitizen", data = df, hue = "Churn")
plt.title("Count By Senior Citizen")
plt.show()
```



[111]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Sample data (use your actual DataFrame)
# df = pd.DataFrame(...)

# Calculate the counts for each group and normalize them
count_data = df.groupby(['SeniorCitizen', 'Churn']).size().unstack()
percent_data = count_data.apply(lambda x: x / x.sum() * 100, axis=1)

# Plot stacked bar chart
plt.figure(figsize = (4,4))
percent_data.plot(kind='bar', stacked=True, color=['#1f77b4', '#ff7f0e']) #␣
 ↪Customize colors if needed
plt.title("Percentage of Churn by Senior Citizen")
plt.xlabel("Senior Citizen")
plt.ylabel("Percentage")
plt.legend(title="Churn", loc="upper left", bbox_to_anchor=(1, 1))

# Add percentages on top of each segment
for i in range(percent_data.shape[0]):
    for j in range(percent_data.shape[1]):
```

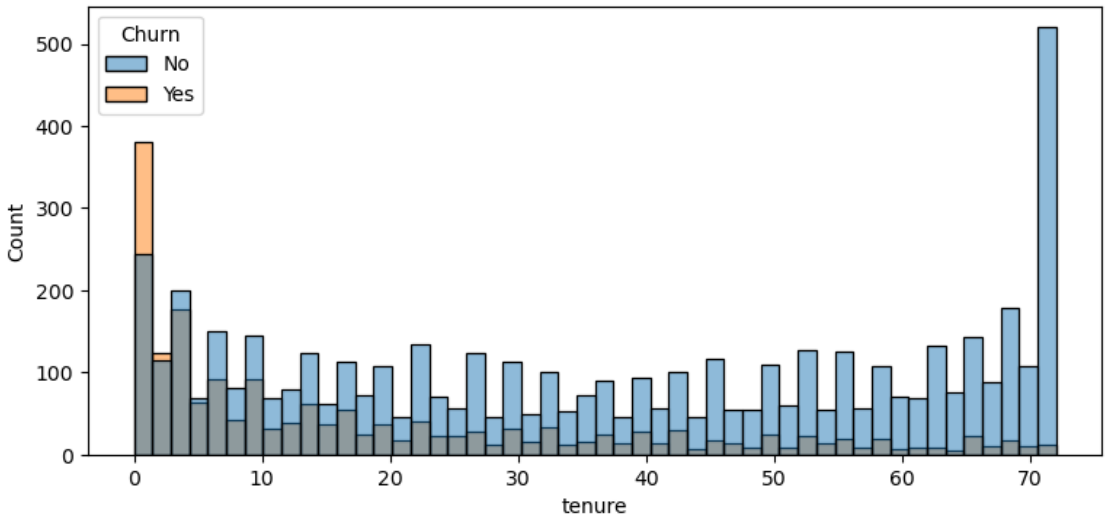
```
        plt.text(i, percent_data.iloc[i, :j+1].sum() - percent_data.iloc[i, j]/
↪2,
                 f"{percent_data.iloc[i, j]:.1f}%", ha="center", va="center",␣
↪color="white")

plt.show()
```

```
<Figure size 400x400 with 0 Axes>
```



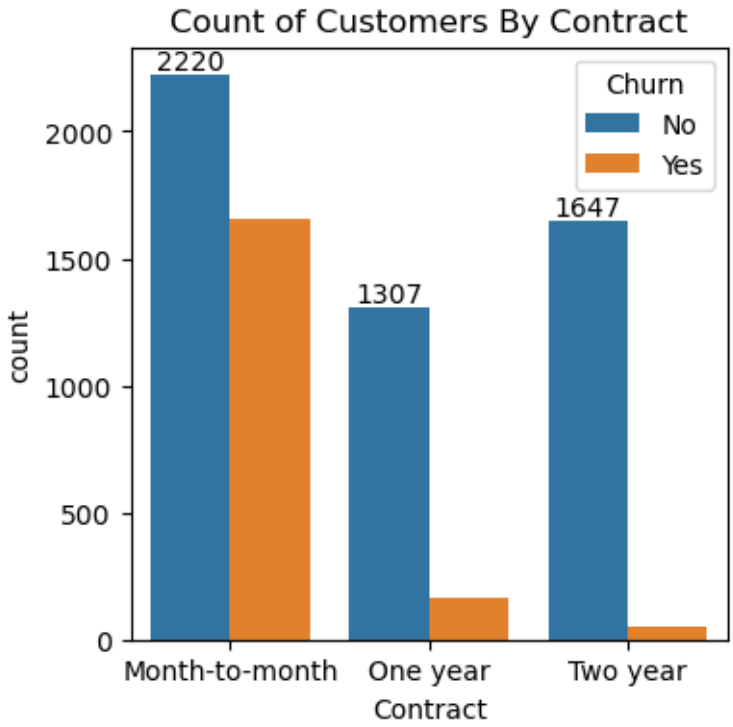
#Comparatively a greater percentage of Senior Citizens have churned out

[118]:
```
plt.figure(figsize = (9,4))
sns.histplot(x= "tenure", data = df, bins = 50, hue = "Churn")
plt.show()
```



#People who used the services for longer period have stayed but the people who used the services only for 1 or 2 months have churned out.

[124]:
```
plt.figure(figsize = (4,4))
ax = sns.countplot(x = "Contract", data = df, hue = "Churn")
ax.bar_label(ax.containers[0])
plt.title("Count of Customers By Contract")
plt.show()
```



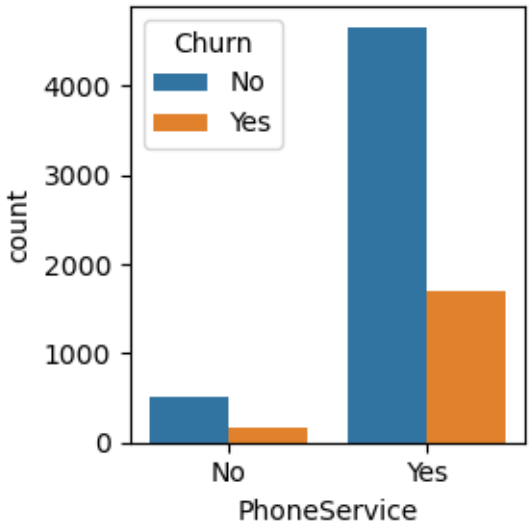
#People who have month to month contract are more likely to churn than those who have longer contracts.

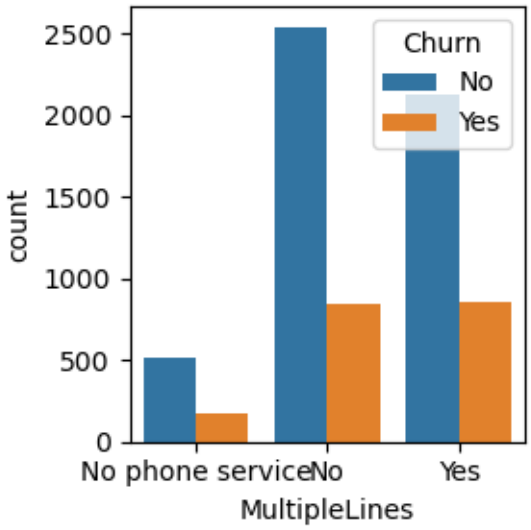
[125]:
```
df.columns.values
```

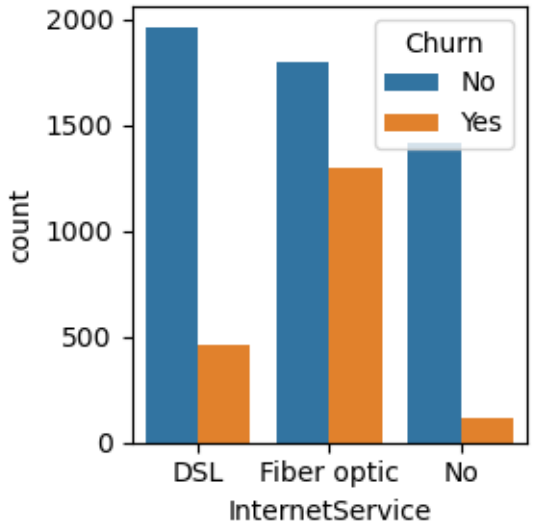
[125]:
```
array(['customerID', 'gender', 'SeniorCitizen', 'Partner', 'Dependents',
       'tenure', 'PhoneService', 'MultipleLines', 'InternetService',
       'OnlineSecurity', 'OnlineBackup', 'DeviceProtection',
       'TechSupport', 'StreamingTV', 'StreamingMovies', 'Contract',
       'PaperlessBilling', 'PaymentMethod', 'MonthlyCharges',
       'TotalCharges', 'Churn'], dtype=object)
```

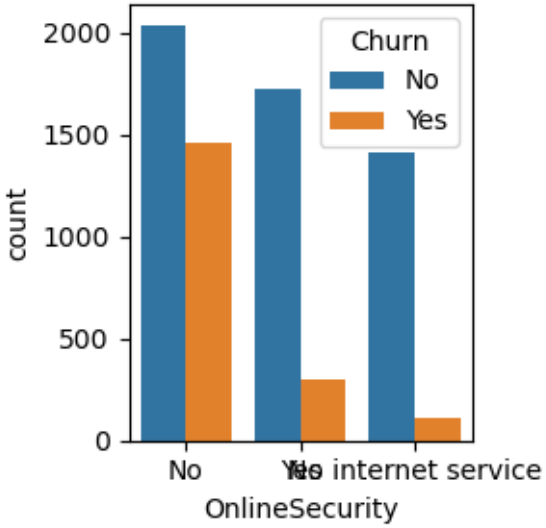
[130]:
```
plt.figure(figsize = (3,3))
sns.countplot(x = "PhoneService", data =df, hue = "Churn")
plt.tight_layout()
plt.show()
plt.figure(figsize = (3,3))
sns.countplot(x = "MultipleLines", data =df, hue = "Churn")
plt.tight_layout()
plt.show()
plt.figure(figsize = (3,3))
sns.countplot(x = "InternetService", data =df, hue = "Churn")
plt.tight_layout()
plt.show()
plt.figure(figsize = (3,3))
sns.countplot(x = "OnlineSecurity", data =df, hue = "Churn")
plt.tight_layout()
plt.show()
plt.figure(figsize = (3,3))
sns.countplot(x = "OnlineBackup", data =df, hue = "Churn")
plt.tight_layout()
plt.show()
plt.figure(figsize = (3,3))
sns.countplot(x = "DeviceProtection", data =df, hue = "Churn")
plt.tight_layout()
plt.show()
plt.figure(figsize = (3,3))
sns.countplot(x = "TechSupport", data =df, hue = "Churn")
plt.tight_layout()
plt.show()
plt.figure(figsize = (3,3))
sns.countplot(x = "StreamingTV", data =df, hue = "Churn")
plt.tight_layout()
plt.show()
plt.figure(figsize = (3,3))
```

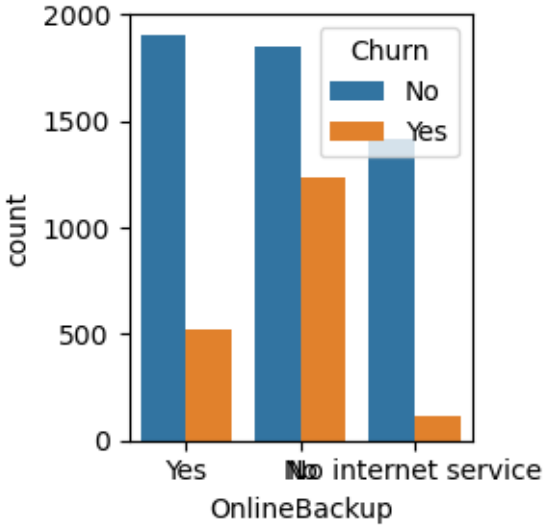
```
sns.countplot(x = "StreamingMovies", data =df, hue = "Churn")
plt.tight_layout()
plt.show()
```

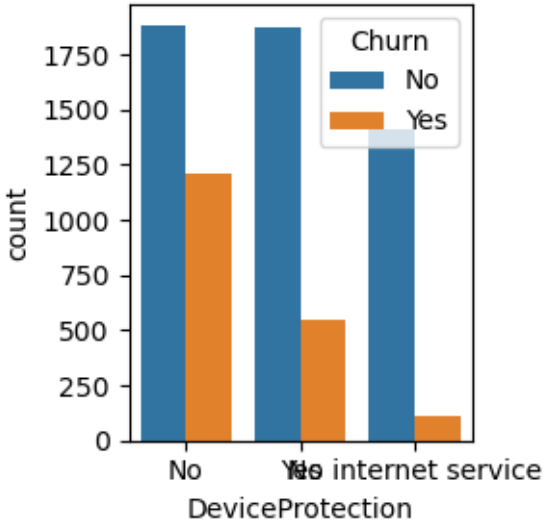


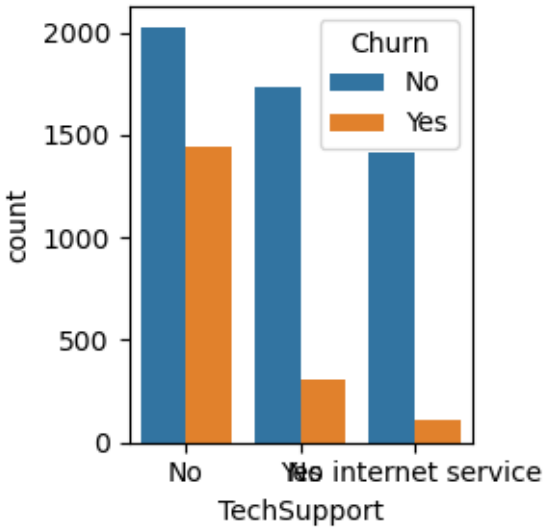


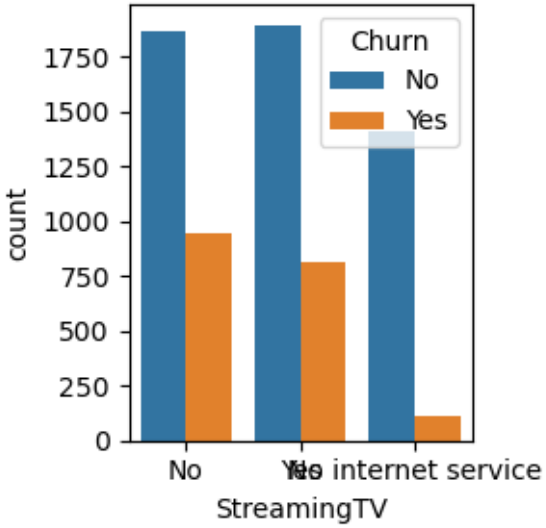

2000
1500
1000
500
0
count
Churn
No
Yes
DSL
Fiber optic
No
InternetService

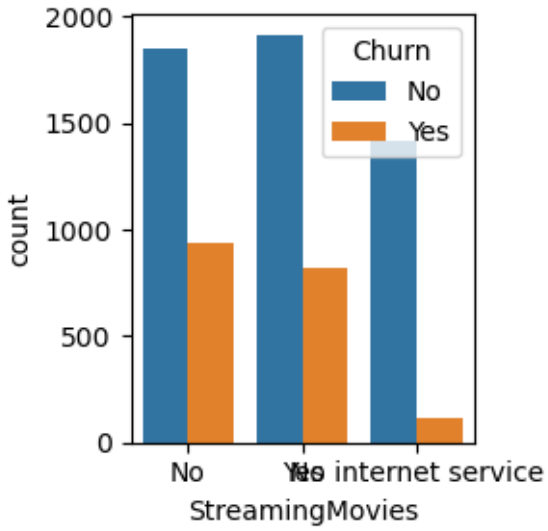

2000
1500
1000
500
0
count
Churn
No
Yes
No
No internet service
OnlineSecurity


2000
1500
1000
500
0
count
Churn
No
Yes
Yes
No internet service
OnlineBackup


1750
1500
1250
1000
750
500
250
0
count
Churn
No
Yes
No
Yes
No internet service
DeviceProtection

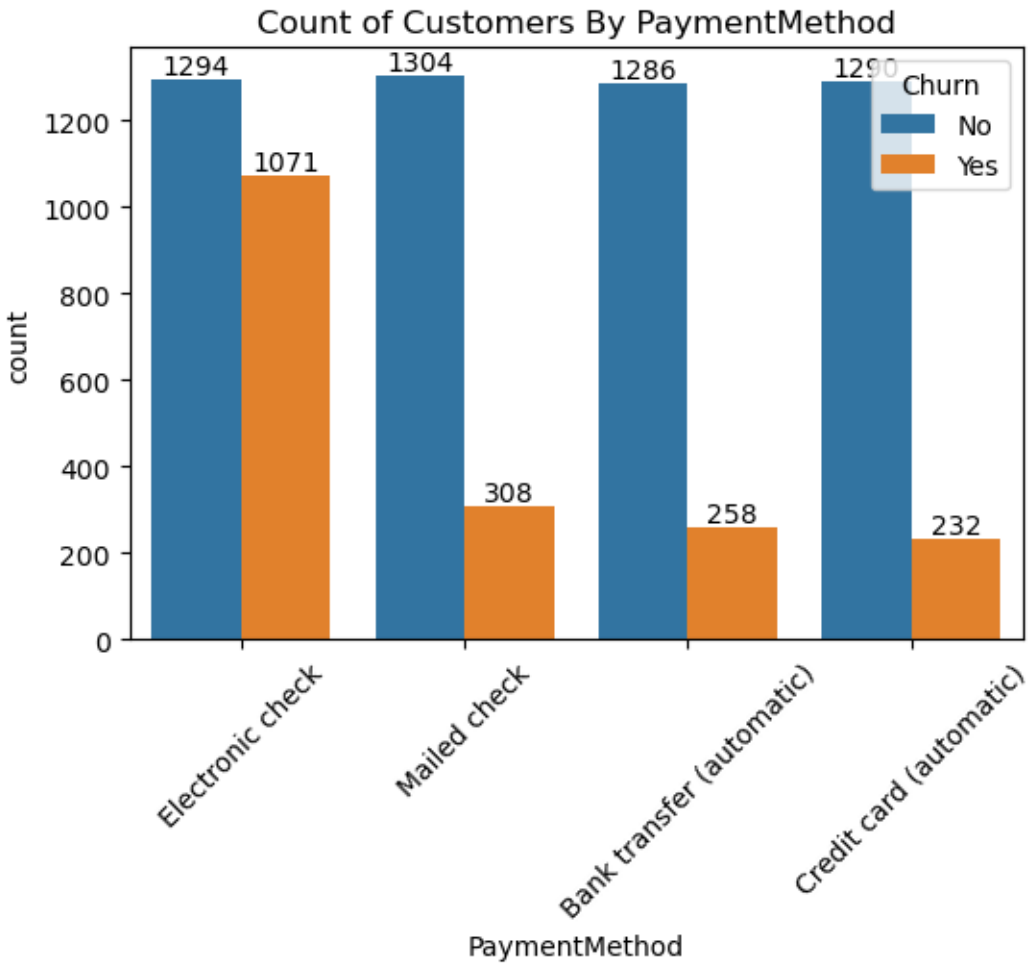

Churn
No
Yes
2000
1500
1000
500
0
count
No
Yes
No internet service
TechSupport


Churn
No
Yes
1750
1500
1250
1000
750
500
250
0
count
No
Yes
No internet service
StreamingTV



#The majority of the customers who do not churn tend to have services like PhoneSrvice, InternetService (particularly DSL), and Online security enabled. For Services like Online Backup, Techsupport, and streaminTV, churn rates are noticeably higher when these services are not used or not available.

[136]:
```
plt.figure(figsize = (6,4))
ax = sns.countplot(x = "PaymentMethod", data = df, hue = "Churn")
ax.bar_label(ax.containers[0])
ax.bar_label(ax.containers[1])
plt.title("Count of Customers By PaymentMethod")
plt.xticks(rotation = 45)
plt.show()
```



#Customer is likely to Churn when he is using electronic check as a Payment Method

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]: