Telco Churn Analysis Project

[1]:
```python
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

[3]:
```python
df = pd.read_csv('customer Churn.csv')
df
```

[3]:
```
      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]
```

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

[5]:
```
   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]
```

```
[7]: 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 no total charges are recorded.

[10]:

```
df["TotalCharges"] = df["TotalCharges"].replace(" ","0")
df["TotalCharges"] = df["TotalCharges"].astype("float")
```

[12]:

```
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
```

[14]:

```
df.isnull()
```

[14]:

```
      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]
```

[16]:

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

[16]:

```
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
```

[18]:

```
df.describe()
```

[18]:

```
       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
```

[20]:

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

[20]:

```
0
```

[22]:

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

[22]:

```
0
```

[24]:

```
def conv(value):
    if value == 1:
        return "yes"
    else:
        return "no"


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

#converted 0 and 1 values of senior citizen to yes/no to make easier to understand.

[27]:

```
df.head()
```

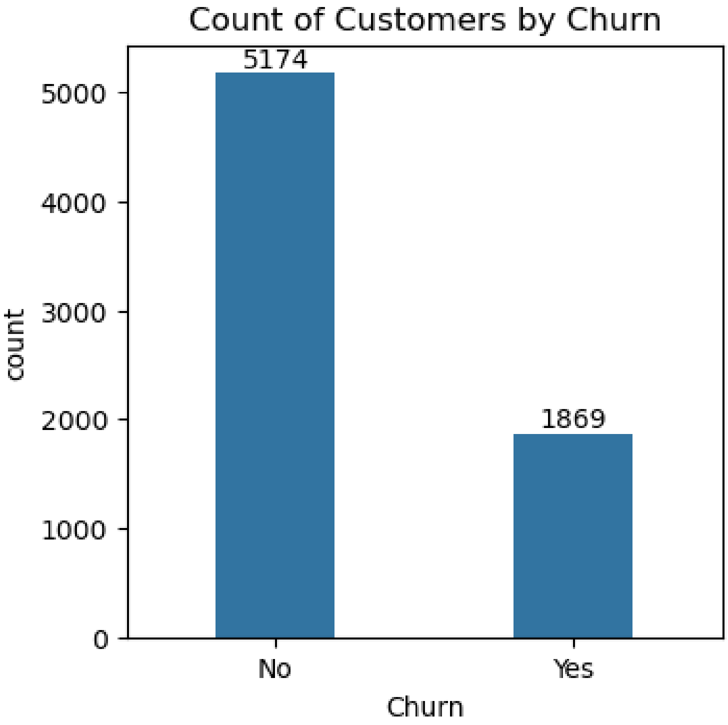
[27]:

```
   customerID  gender SeniorCitizen Partner Dependents  tenure PhoneService   \
0  7590-VHVEG  Female            no     Yes         No       1           No
1  5575-GNVDE    Male            no      No         No      34          Yes
2  3668-QPYBK    Male            no      No         No       2          Yes
3  7795-CFOCW    Male            no      No         No      45           No
4  9237-HQITU  Female            no      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        1889.50     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]
```

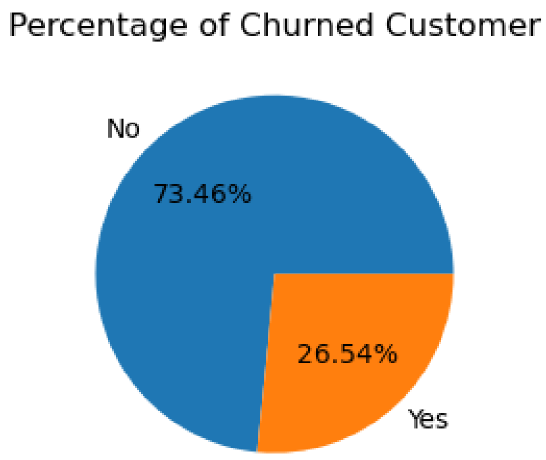
[29]:

```
plt.figure(figsize=(4,4))
ax = sns.countplot(x = 'Churn', data=df, width=0.4)
```

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



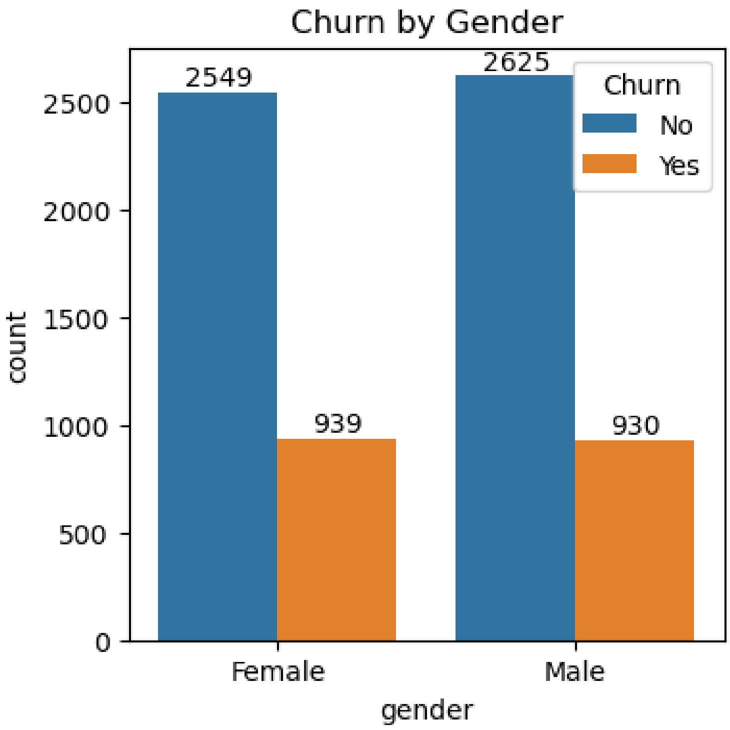
[31]:
```
plt.figure(figsize= (3,4))
gb = df.groupby("Churn").agg({'Churn' : "count"})
plt.pie(gb['Churn'], labels= gb.index, autopct = "%1.2f%%")
plt.title("Percentage of Churned Customer")
plt.show()
```



#from the given pie chart we can conclude that 26.54% of our customers have churned out. Let's explore the reason behind it.

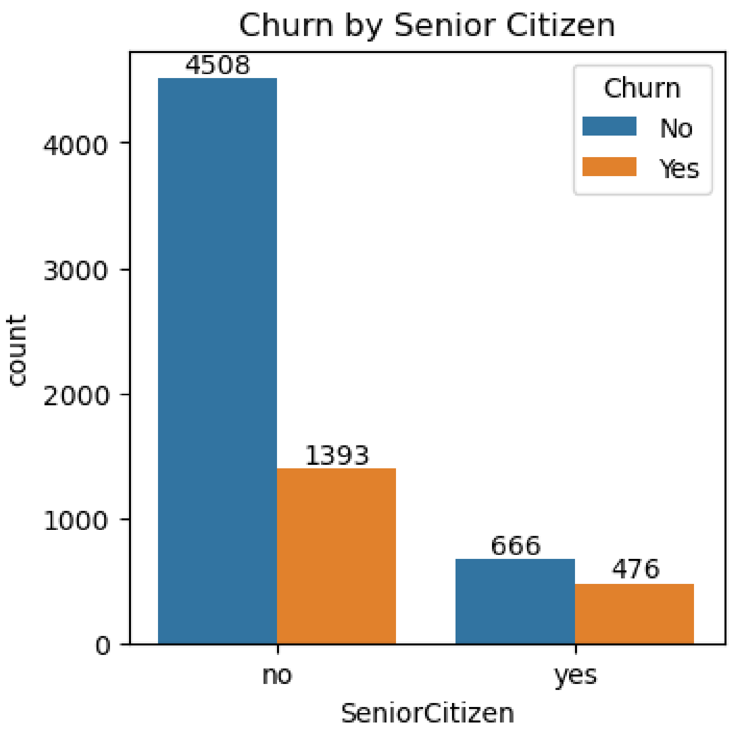
[34]:

```
plt.figure(figsize=(4, 4))
ax = sns.countplot(x='gender', data=df, hue='Churn')
for container in ax.containers:
    ax.bar_label(container)  # Adds the labels for all containers
plt.title("Churn by Gender")
plt.show()
```



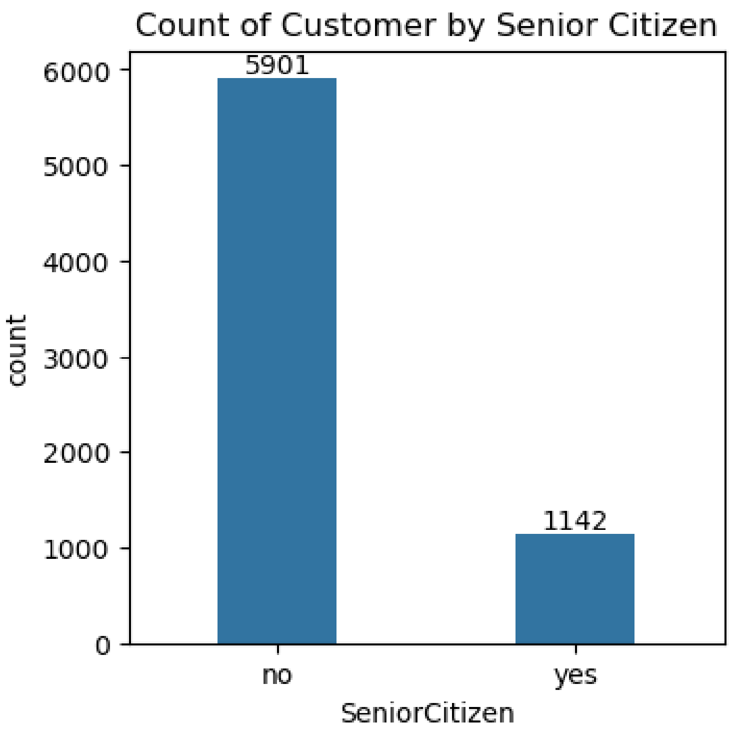
[36]:

```
plt.figure(figsize=(4, 4))
ax = sns.countplot(x='SeniorCitizen', data=df, hue='Churn')
for container in ax.containers:
    ax.bar_label(container)  # Adds the labels for all containers
plt.title("Churn by Senior Citizen")
plt.show()
```

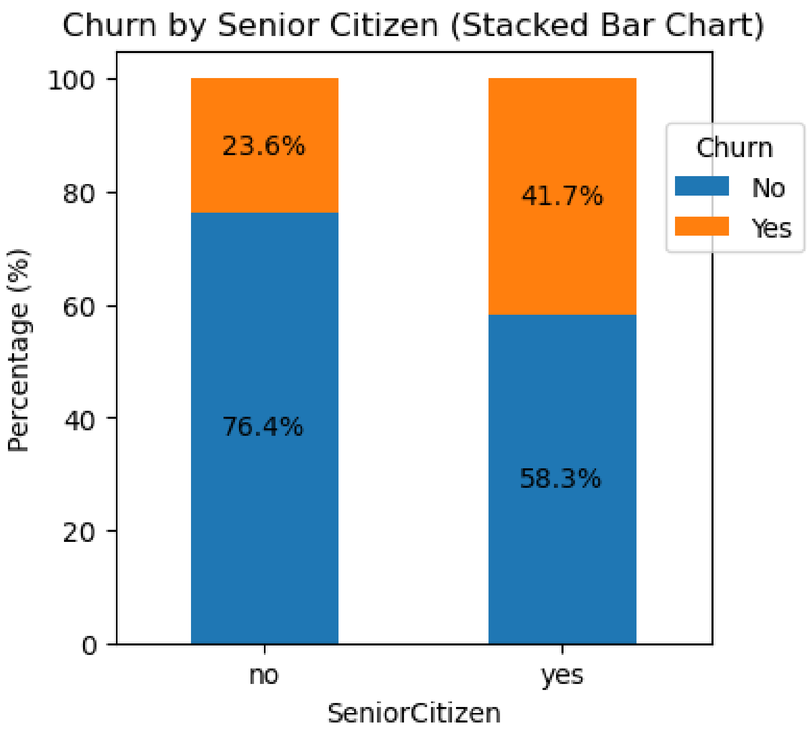


[38]:

```
plt.figure(figsize=(4,4))
ax = sns.countplot(x='SeniorCitizen', data=df, width=0.4)
ax.bar_label(ax.containers[0])
plt.title("Count of Customer by Senior Citizen")
plt.show()
```

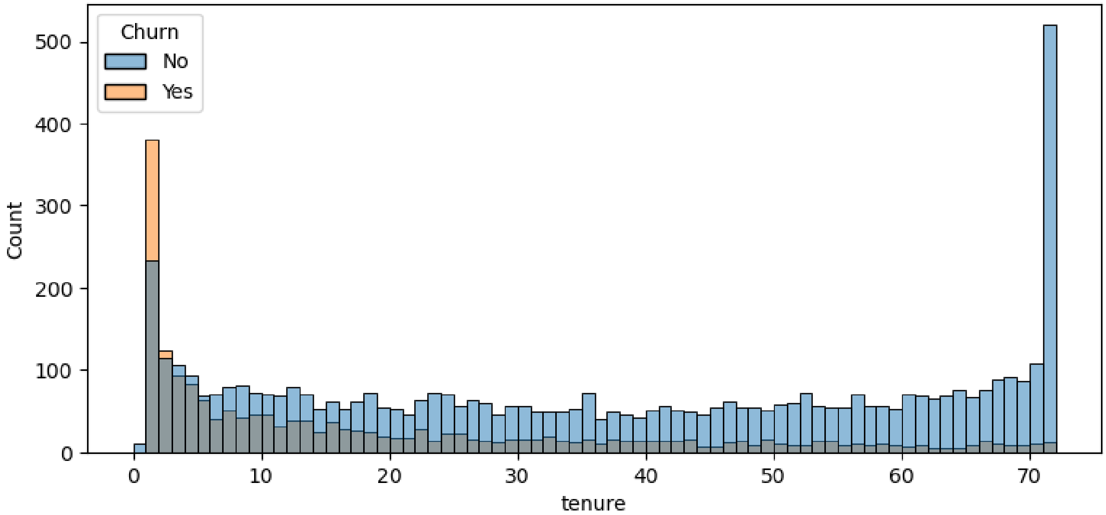


[40]:
```python
total_count = df.groupby('SeniorCitizen')['Churn'].value_counts(normalize=True).unstack() * 100

fig, ax = plt.subplots(figsize=(4,4))

total_count.plot(kind='bar', stacked= True, ax=ax, color=['#1f77b4', '#ff7f0e'])

for p in ax.patches:
    width, height = p.get_width(), p.get_height()
    x, y = p.get_xy()
    ax.text(x + width / 2, y + height / 2, f'{height:.1f}%', ha = 'center', va='center')


plt.title('Churn by Senior Citizen (Stacked Bar Chart)')
plt.xlabel('SeniorCitizen')
plt.ylabel('Percentage (%)')
plt.xticks(rotation=0)
plt.legend(title='Churn', bbox_to_anchor = (0.9,0.9))

plt.show()
```



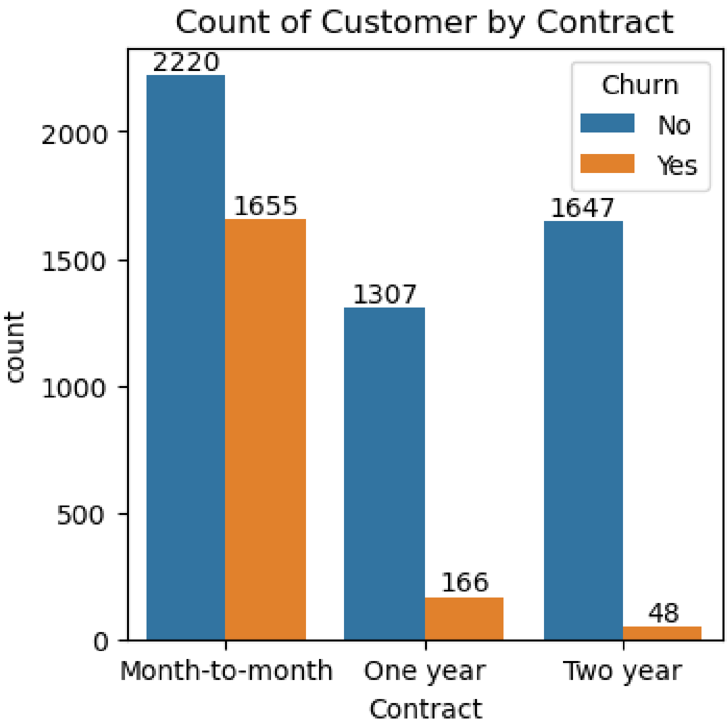
#comparative a greater percentage of people in senior citizen category have churned.

[43]:
```
plt.figure(figsize=(9,4))
sns.histplot(x= 'tenure', data=df, bins=72, hue='Churn')
plt.show()
```



#people who have used our services for a long time have stayed and people who have used services for 1 or 2 months have churned.

[46]:
```
plt.figure(figsize=(4,4))
ax = sns.countplot(x='Contract', data=df, hue='Churn' )
for container in ax.containers:
    ax.bar_label(container)
plt.title("Count of Customer by Contract")
plt.show()
```

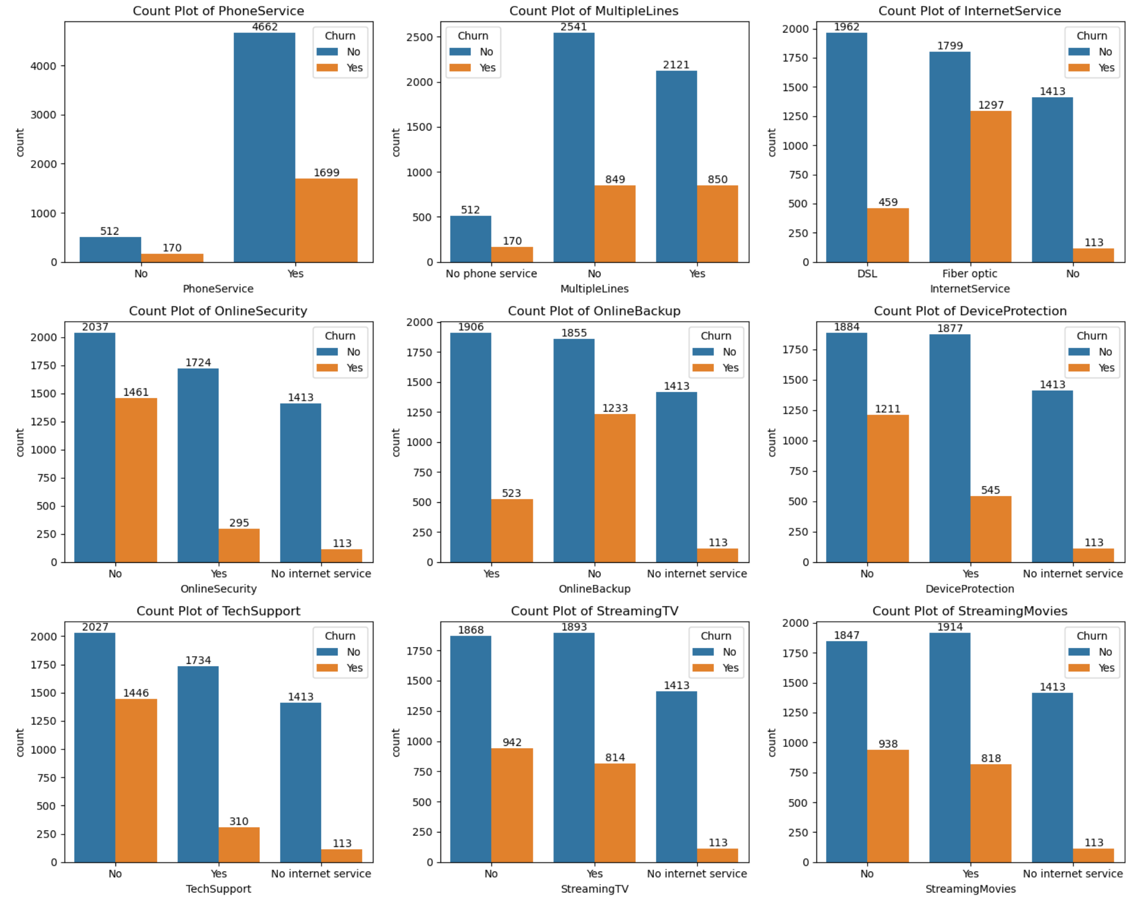


#people who have month to month contract are likely to churn then from those who have 1 or 2 years of contract.

[49]:
```
df.columns.values
```

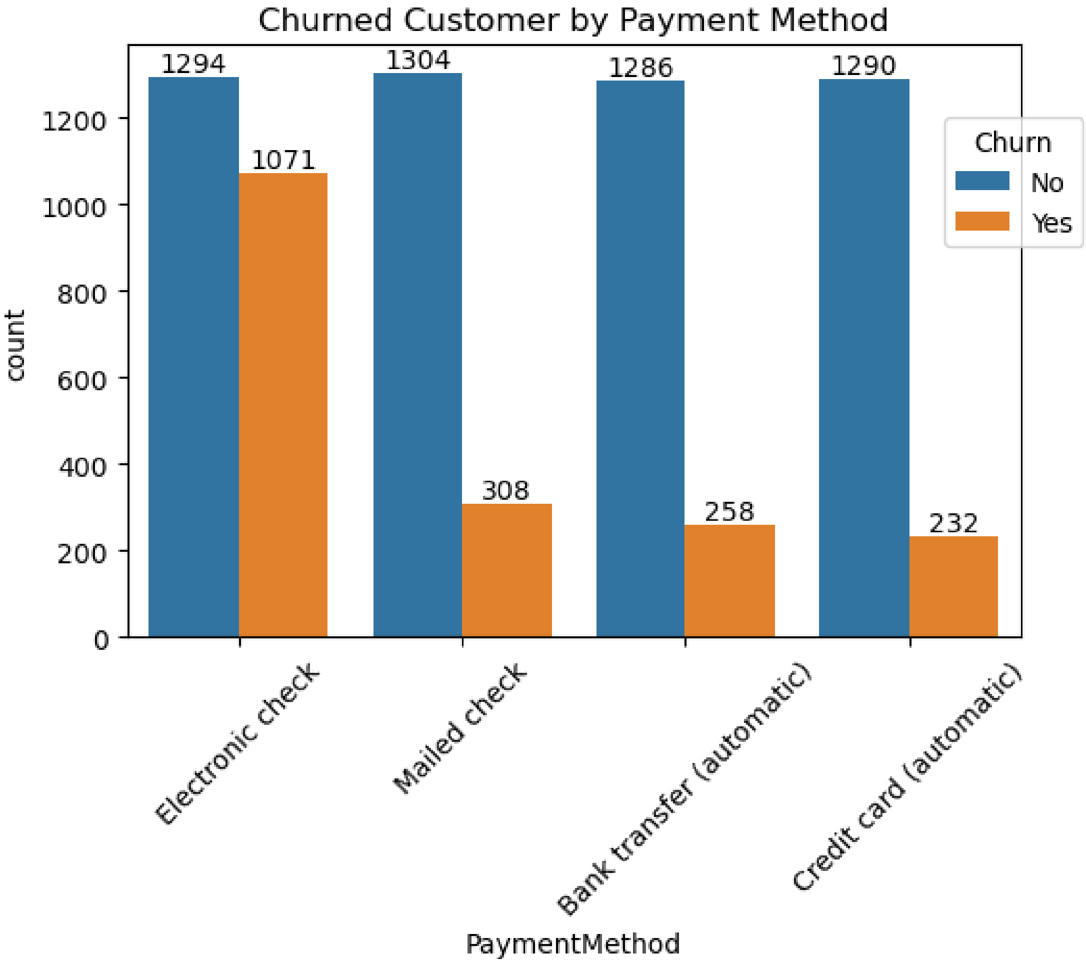
[49]:
```
array(['customerID', 'gender', 'SeniorCitizen', 'Partner', 'Dependents',
       'tenure', 'PhoneService', 'MultipleLines', 'InternetService',
       'OnlineSecurity', 'OnlineBackup', 'DeviceProtection',
       'TechSupport', 'StreamingTV', 'StreamingMovies', 'Contract',
       'PaperlessBilling', 'PaymentMethod', 'MonthlyCharges',
       'TotalCharges', 'Churn'], dtype=object)
```

[51]:
```
# List of columns to plot
columns = [
    'PhoneService', 'MultipleLines', 'InternetService',
    'OnlineSecurity', 'OnlineBackup', 'DeviceProtection',
    'TechSupport', 'StreamingTV', 'StreamingMovies'
]

# Create a grid of subplots
num_cols = 3  # Number of columns in the subplot grid
num_rows = -(-len(columns) // num_cols)  # Calculate rows needed using ceiling␣
 ↪division
fig, axes = plt.subplots(num_rows, num_cols, figsize=(15, num_rows * 4))

# Flatten axes array for easier iteration
axes = axes.flatten()

# Plot countplots for each column
for i, col in enumerate(columns):
    ax = sns.countplot(x=col, data=df, ax=axes[i], hue= df["Churn"])
    for container in ax.containers:
       ax.bar_label(container)
    axes[i].set_title(f'Count Plot of {col}')


# Remove any unused subplots
for j in range(len(columns), len(axes)):
    fig.delaxes(axes[j])

# Adjust layout for better spacing
plt.tight_layout()
plt.show()
```



#Services like "InternetService" and "TechSupport" display clear differences in churn rates based on customer preferences or availability, while categories like "OnlineSecurity" and "DeviceProtection" highlight significant variation between "Yes" and "No internet service" responses.

[54]:
```
plt.figure(figsize=(6,4))
ax = sns.countplot(x = 'PaymentMethod', data=df, hue='Churn')
for container in ax.containers:
    ax.bar_label(container)
plt.title("Churned Customer by Payment Method")
plt.xticks(rotation = 45)
plt.legend(title='Churn', bbox_to_anchor = (0.9,0.9))
plt.show()
```



#Customer is likely to churn when he is using electronic check as a payment method.

```
[ ]:
```