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

df = pd.read_csv(r"C:\Users\NISHA\Desktop\python\import_files\Customer
Churn.csv")
df.head()
```

```
   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]
```

```
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
```

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

```
df[df["TotalCharges"] == 0]
```

```
      customerID  gender  SeniorCitizen Partner Dependents  tenure  \
488   4472-LVYGI  Female              0     Yes        Yes       0
753   3115-CZMZD    Male              0      No        Yes       0
936   5709-LVOEQ  Female              0     Yes        Yes       0
1082  4367-NUYAO    Male              0     Yes        Yes       0
1340  1371-DWPAZ  Female              0     Yes        Yes       0
3331  7644-OMVMY    Male              0     Yes        Yes       0
```

```
3826  3213-VVOLG    Male              0     Yes         Yes       0
4380  2520-SGTTA  Female              0     Yes         Yes       0
5218  2923-ARZLG    Male              0     Yes         Yes       0
6670  4075-WKNIU  Female              0     Yes         Yes       0
6754  2775-SEFEE    Male              0      No         Yes       0

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

         DeviceProtection          TechSupport          StreamingTV  \
488                   Yes                  Yes                  Yes
753   No internet service  No internet service  No internet service
936                   Yes                   No                  Yes
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                   No                  Yes                   No

          StreamingMovies  Contract PaperlessBilling  \
488                    No  Two year              Yes
753   No internet service  Two year               No
936                   Yes  Two year               No
1082  No internet service  Two year               No
1340                   No  Two year               No
```

```
3331  No internet service  Two year               No
3826  No internet service  Two year               No
4380  No internet service  Two year               No
5218  No internet service  One year              Yes
6670                   No  Two year               No
6754                   No  Two year              Yes

                   PaymentMethod MonthlyCharges  TotalCharges  Churn
488   Bank transfer (automatic)          52.55           0.0     No
753                Mailed check          20.25           0.0     No
936                Mailed check          80.85           0.0     No
1082               Mailed check          25.75           0.0     No
1340    Credit card (automatic)          56.05           0.0     No
3331               Mailed check          19.85           0.0     No
3826               Mailed check          25.35           0.0     No
4380               Mailed check          20.00           0.0     No
5218               Mailed check          19.70           0.0     No
6670               Mailed check          73.35           0.0     No
6754  Bank transfer (automatic)          61.90           0.0     No

[11 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   float64
 20  Churn             7043 non-null   object
```

```
dtypes: float64(2), int64(2), object(17)
memory usage: 1.1+ MB
```

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

```
0
```

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

```
0
```

```
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 it easier to understand

```
df[df["SeniorCitizen"] == "Yes"]
```

```
      customerID  gender SeniorCitizen Partner Dependents  tenure  \
20    8779-QRDMV    Male           Yes      No         No       1
30    3841-NFECX  Female           Yes     Yes         No      71
31    4929-XIHVW    Male           Yes     Yes         No       2
34    3413-BMNZE    Male           Yes      No         No       1
50    8012-SOUDQ  Female           Yes      No         No      43
...          ...     ...           ...     ...        ...     ...
7023  1035-IPQPU  Female           Yes     Yes         No      63
7029  2235-DWLJU  Female           Yes      No         No       6
7031  3605-JISKB    Male           Yes     Yes         No      55
7032  6894-LFHLY    Male           Yes      No         No       1
7041  8361-LTMKD    Male           Yes     Yes         No       4

     PhoneService     MultipleLines InternetService
OnlineSecurity  ...  \
20             No  No phone service             DSL
```

```
No  ...
30           Yes               Yes     Fiber optic
Yes  ...
31           Yes                No     Fiber optic
No  ...
34           Yes                No            DSL
No  ...
50           Yes               Yes     Fiber optic
No  ...
...          ...               ...            ...             ...  ..
.
7023         Yes               Yes     Fiber optic
No  ...
7029          No  No phone service            DSL
No  ...
7031         Yes               Yes            DSL
Yes  ...
7032         Yes               Yes     Fiber optic
No  ...
7041         Yes               Yes     Fiber optic
No  ...

     DeviceProtection TechSupport StreamingTV StreamingMovies
Contract  \
20                Yes          No          No              Yes  Month-
to-month
30                Yes         Yes          No               No
Two year
31                Yes          No         Yes              Yes  Month-
to-month
34                 No          No          No               No  Month-
to-month
50                 No          No         Yes               No  Month-
to-month
...               ...         ...         ...              ...
...
7023              Yes          No         Yes              Yes  Month-
to-month
7029               No          No         Yes              Yes  Month-
to-month
7031               No          No          No               No
One year
7032               No          No          No               No  Month-
to-month
7041               No          No          No               No  Month-
to-month

     PaperlessBilling              PaymentMethod MonthlyCharges
TotalCharges  \
```

```
20                 Yes           Electronic check          39.65
39.65
30                 Yes    Credit card (automatic)          96.35
6766.95
31                 Yes    Credit card (automatic)          95.50
181.65
34                  No  Bank transfer (automatic)          45.25
45.25
50                 Yes           Electronic check          90.25
3838.75
...                ...                        ...            ...
...
7023               Yes           Electronic check         103.50
6479.40
7029               Yes           Electronic check          44.40
263.05
7031                No    Credit card (automatic)          60.00
3316.10
7032               Yes           Electronic check          75.75
75.75
7041               Yes               Mailed check          74.40
306.60

      Churn
20      Yes
30       No
31       No
34       No
50       No
...     ...
7023     No
7029     No
7031     No
7032    Yes
7041    Yes

[1142 rows x 21 columns]
```

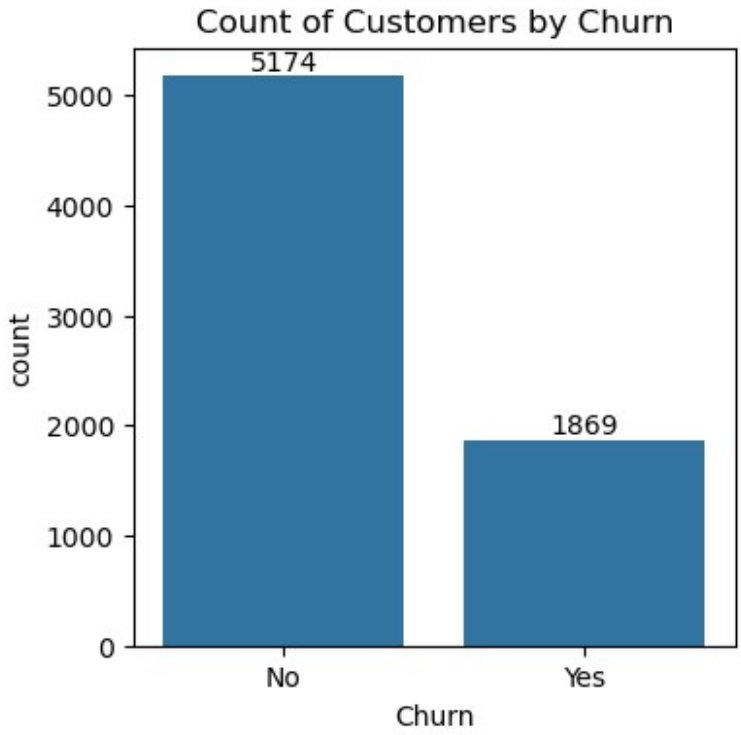
```
df[df["SeniorCitizen"] == "No"]
```

```
      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
...          ...     ...           ...     ...        ...     ...
7037  2569-WGERO  Female            No      No         No      72
7038  6840-RESVB    Male            No     Yes        Yes      24
7039  2234-XADUH  Female            No     Yes        Yes      72
```

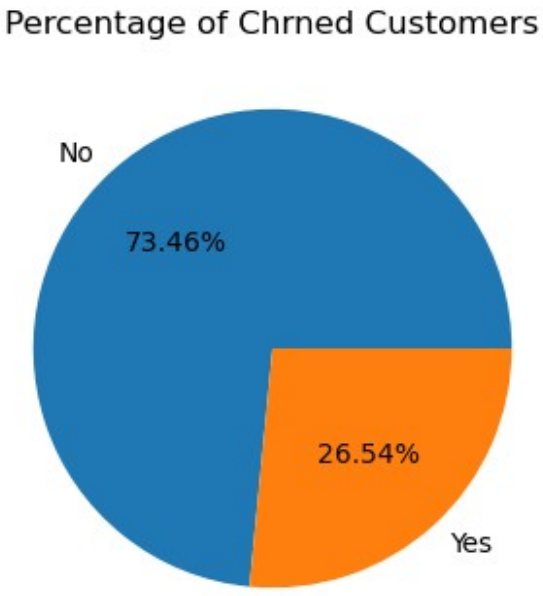
```
7040  4801-JZAZL  Female            No     Yes        Yes      11
7042  3186-AJIEK    Male            No      No         No      66

     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  ...
...           ...               ...             ...                   .
..  ...
7037          Yes                No              No  No internet
service  ...
7038          Yes               Yes             DSL
Yes  ...
7039          Yes               Yes     Fiber optic
No  ...
7040           No  No phone service             DSL
Yes  ...
7042          Yes                No     Fiber optic
Yes  ...

         DeviceProtection          TechSupport          StreamingTV  \
0                      No                   No                   No
1                     Yes                   No                   No
2                      No                   No                   No
3                     Yes                  Yes                   No
4                      No                   No                   No
...                   ...                  ...                  ...
7037  No internet service  No internet service  No internet service
7038                  Yes                  Yes                  Yes
7039                  Yes                   No                  Yes
7040                   No                   No                   No
7042                  Yes                  Yes                  Yes

          StreamingMovies        Contract PaperlessBilling  \
0                      No  Month-to-month              Yes
1                      No        One year               No
2                      No  Month-to-month              Yes
3                      No        One year               No
4                      No  Month-to-month              Yes
...                   ...             ...              ...
7037  No internet service        Two year              Yes
7038                  Yes        One year              Yes
```

```
7039                  Yes        One year              Yes
7040                   No  Month-to-month              Yes
7042                  Yes        Two year              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
...                           ...            ...           ...    ...
7037  Bank transfer (automatic)          21.15       1419.40     No
7038                Mailed check          84.80       1990.50     No
7039    Credit card (automatic)         103.20       7362.90     No
7040            Electronic check          29.60        346.45     No
7042  Bank transfer (automatic)         105.65       6844.50     No

[5901 rows x 21 columns]
```

```
plt.figure(figsize = (4,4))
ax = sns.countplot(x = df["Churn"])
ax.bar_label(ax.containers[0])
plt.title("Count of Customers by Churn")
plt.show()
```

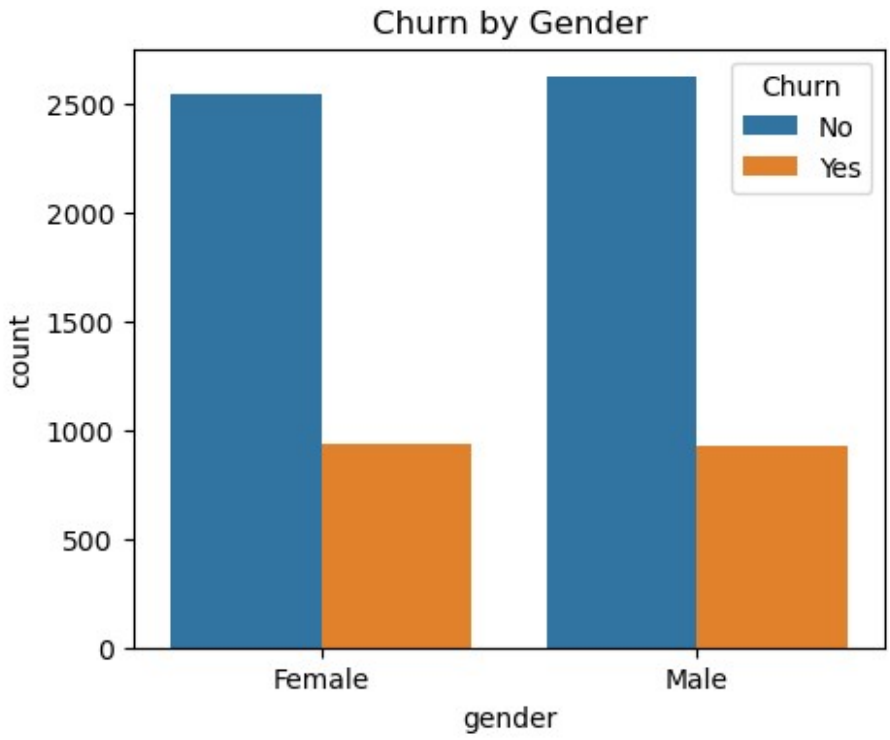


```
plt.figure(figsize = (4,4))
gb = df.groupby("Churn").agg({"Churn":"count"})
plt.pie(gb["Churn"],labels = gb.index, autopct= "%1.2f%%")
plt.title("Percentage of Chrned Customers")
plt.show()
```

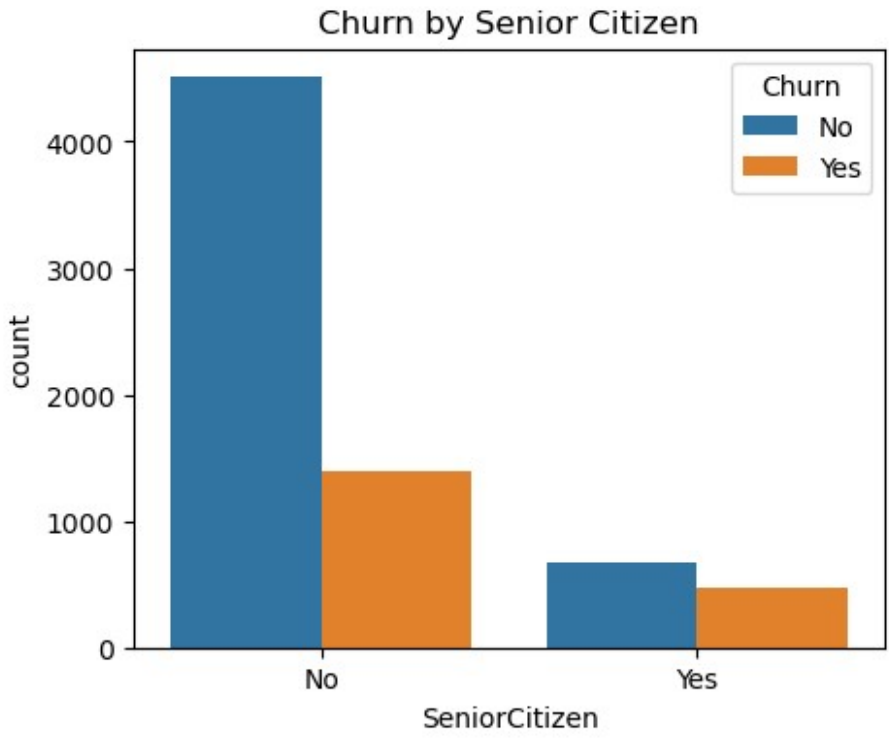


from the given pie chart we can conclude that 26.54% of our customershave churned out. #not let's explore the reason behind it

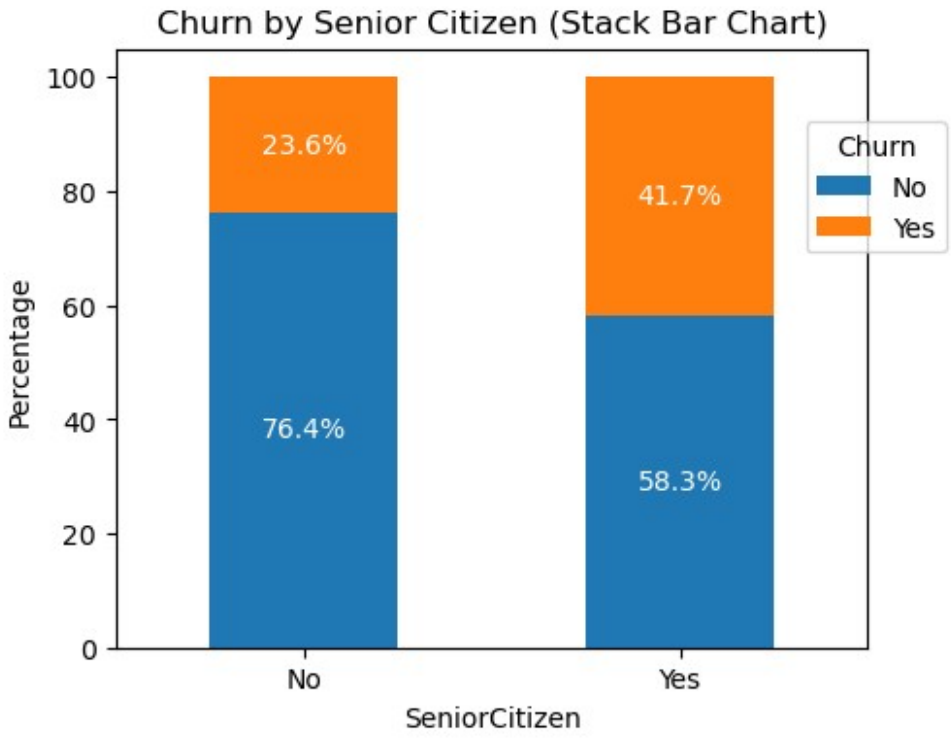
```
plt.figure(figsize = (5,4))
sns.countplot(x = "gender", data = df , hue = "Churn")
plt.title("Churn by Gender")
plt.show()
```



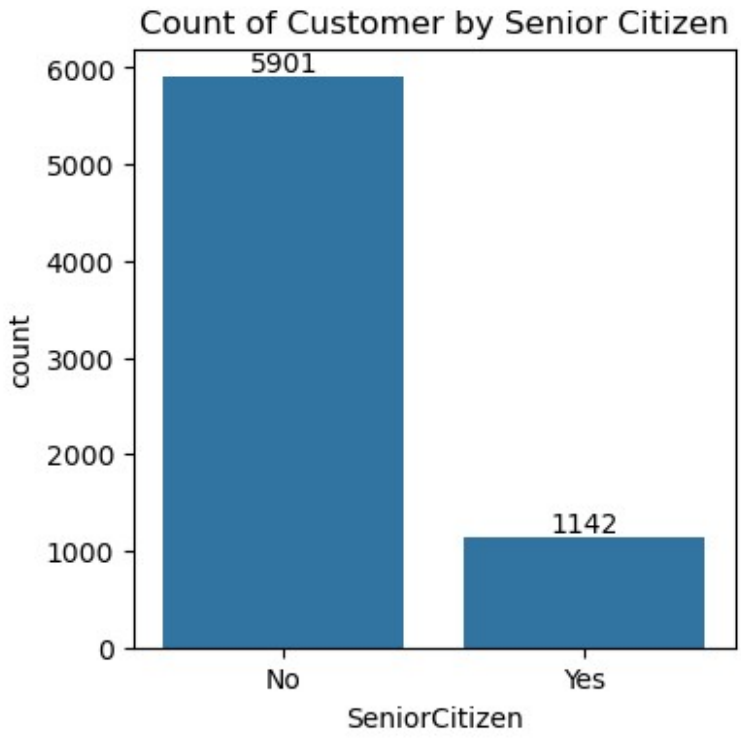

```
plt.figure(figsize = (5,4))
sns.countplot(x = "SeniorCitizen", data = df , hue = "Churn")
plt.title("Churn by Senior Citizen")
plt.show()
```




```
# Calculate counts for each combination of SeniorCitizen and Churn
count_data = df.groupby(['SeniorCitizen',
'Churn']).size().unstack(fill_value=0)

# Calculate percentages
percent_data = count_data.div(count_data.sum(axis=1), axis=0) * 100

# Create a stacked bar chart
percent_data.plot(kind='bar', stacked=True, color=['#1f77b4',
'#ff7f0e'], figsize=(5, 4))

# Adding percentage labels
for i in range(len(percent_data)):
    for j in range(len(percent_data.columns)):
        plt.text(i, percent_data.iloc[i].cumsum().iloc[j] -
percent_data.iloc[i].iloc[j]/2,
                 f"{percent_data.iloc[i].iloc[j]:.1f}%",
                 ha='center', va='center', color='white')

# Customizing the plot
plt.title("Churn by Senior Citizen (Stack Bar Chart)")
# plt.xlabel("SeniorCitizen")
plt.ylabel("Percentage")
plt.xticks(rotation=0)
#plt.legend(title='Churn', loc='upper right')
plt.legend(title='Churn', bbox_to_anchor = (0.9,0.9))
plt.show()
```

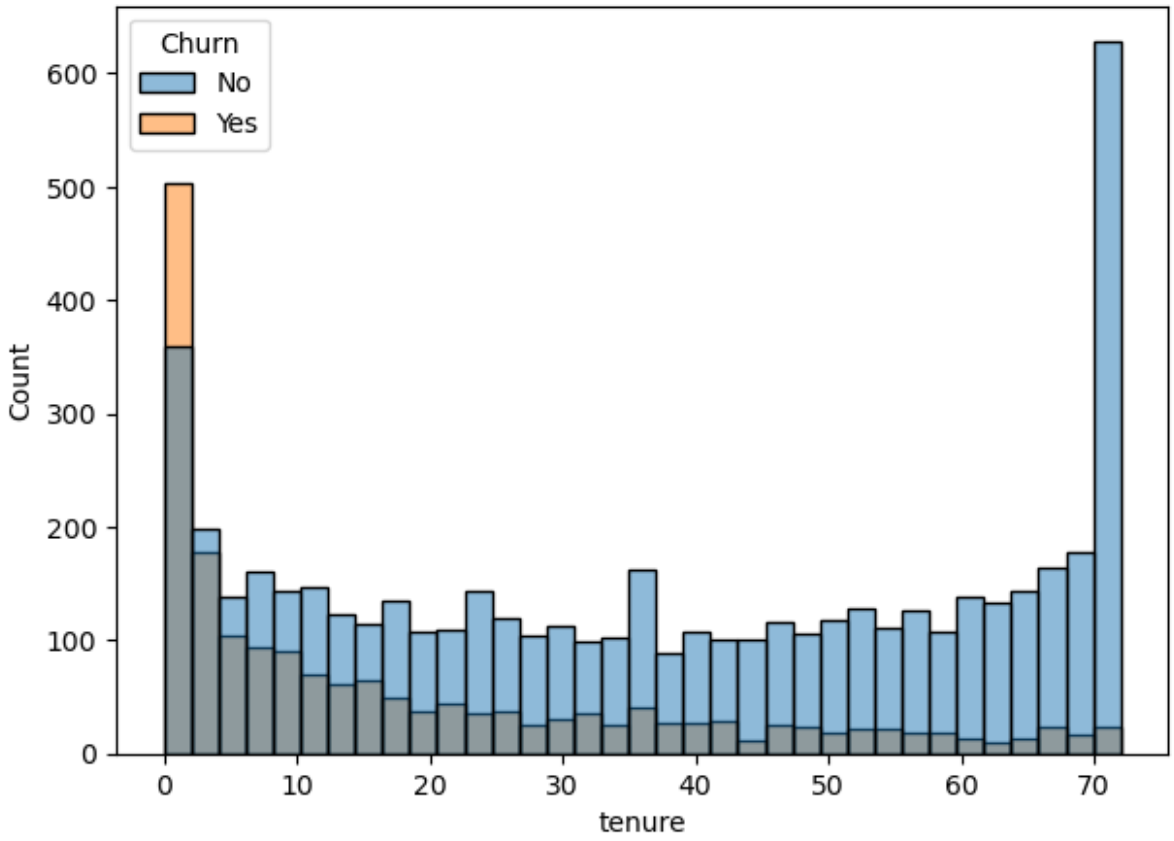



```
plt.figure(figsize = (4,4))
ax = sns.countplot(x = "SeniorCitizen", data = df)
ax.bar_label(ax.containers[0])
plt.title("Count of Customer by Senior Citizen")
plt.show()
```



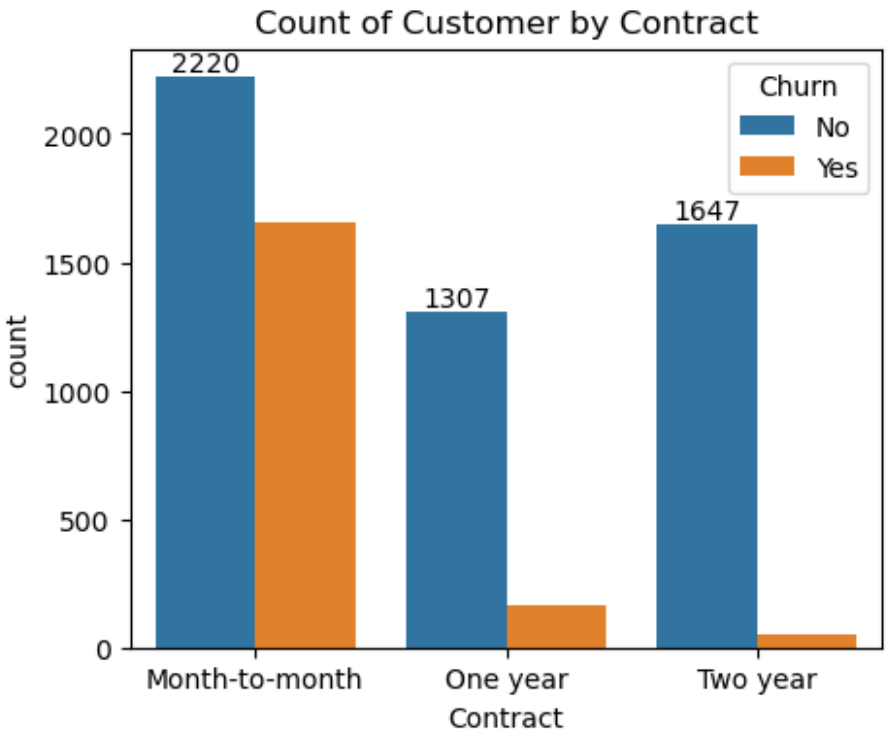
comparative a greater percentage of people in senior citizen category have churned

```
plt.figure(figsize = (7,5))
sns.histplot(x = "tenure" ,data = df, bins= 35, hue = "Churn")
plt.show()
```



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

```
plt.figure(figsize = (5,4))
ax = sns.countplot(x = "Contract", data = df , hue = "Churn")
ax.bar_label(ax.containers[0])
plt.title("Count of Customer by Contract")
plt.show()
```



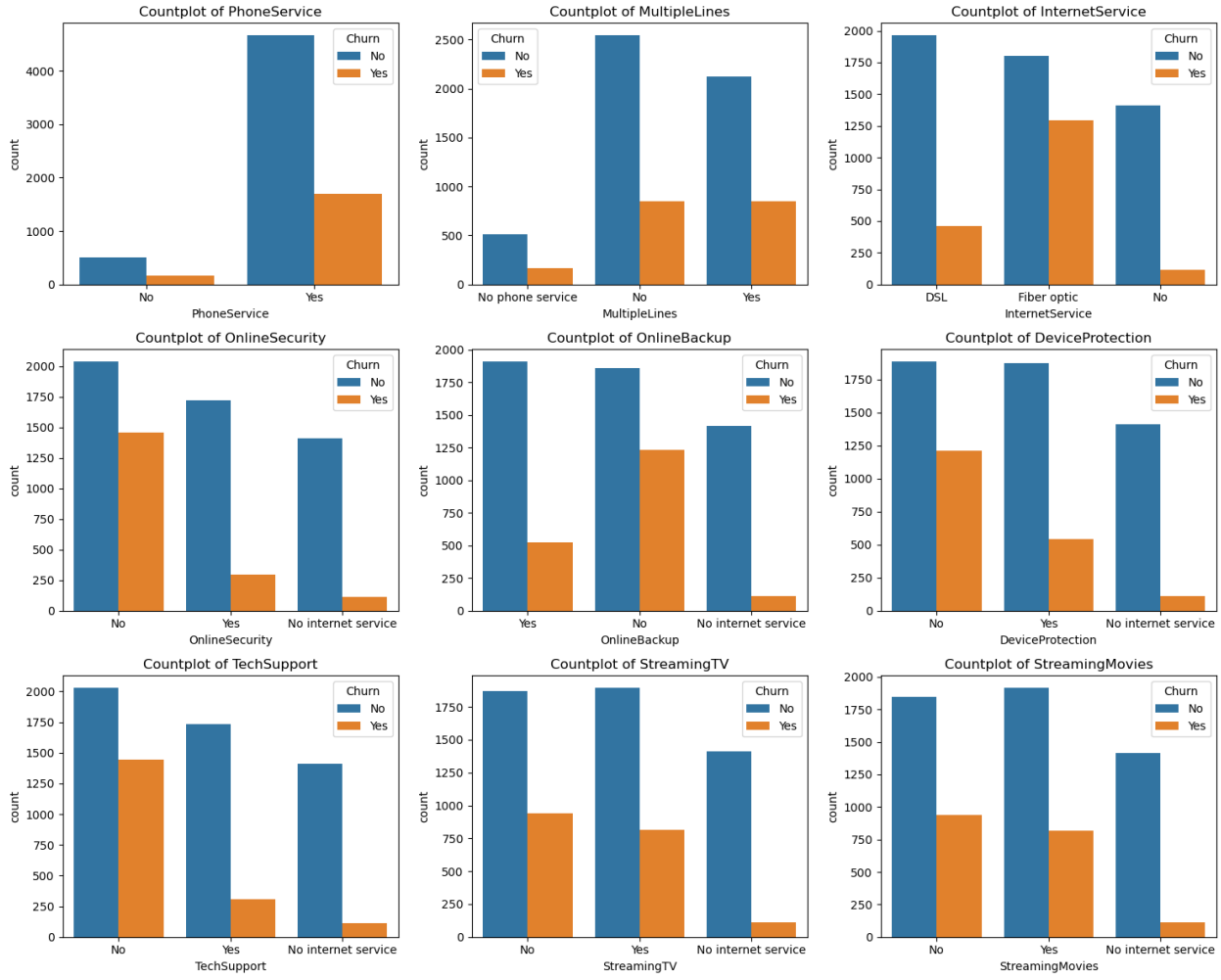
People who have month to month contract are likely to churn thrn from those who have 1 or 2 years of contract.

```
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)
```

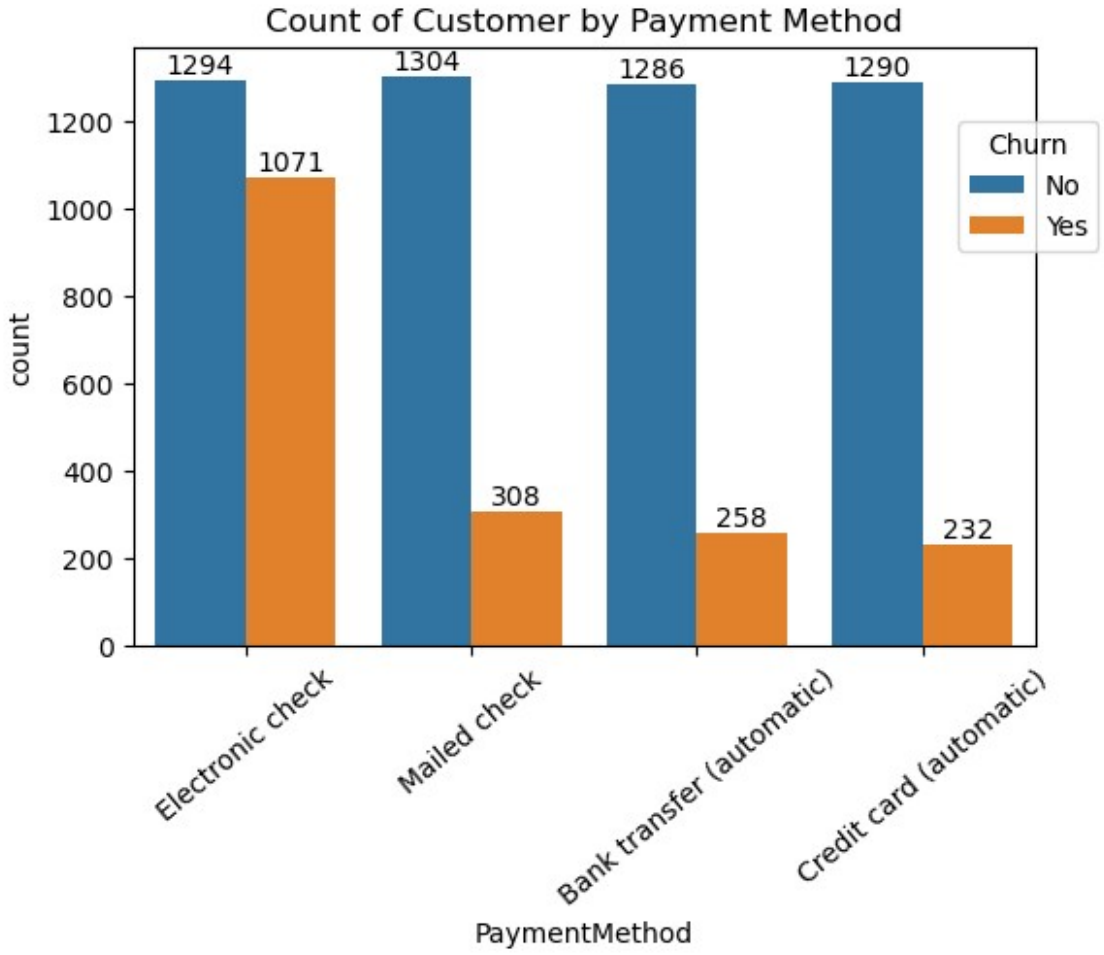
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Assume you have a DataFrame 'df' with the columns specified
# Example: df = pd.read_csv('your_data.csv')

# Set up the matplotlib figure
plt.figure(figsize=(15, 12))

# List of columns to create subplots for
columns = [
    'PhoneService', 'MultipleLines', 'InternetService',
    'OnlineSecurity', 'OnlineBackup', 'DeviceProtection',
```

```
    'TechSupport', 'StreamingTV', 'StreamingMovies'
]

# Create a count plot for each column
for i, column in enumerate(columns, 1):
    plt.subplot(3, 3, i)  # 3 rows, 3 columns
    sns.countplot(data=df, x=column, hue=df["Churn"])  # Create a
countplot for the current column
    plt.title(f'Countplot of {column}')  # Set the title using an f-
string

plt.tight_layout()  # Adjust layout to prevent overlap
plt.show()
```



The majority of customers who do not churn tend to have services like PhoneSevice, InternetService (particularly DSL), and OnlineSecurity enabled. For services like OnlineBackup, TechSupport and StreamingTV. churn rates are noticeably higher when these services are not used or are unavailable.

```
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 Customer by Payment Method")
plt.xticks(rotation = 40)
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 payment method