```
!pip install pandas
!pip install matplotlib
!pip install seaborn



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

```
Requirement already satisfied: pandas in e:\anaconda\lib\site-packages
(2.2.2)
Requirement already satisfied: numpy>=1.26.0 in e:\anaconda\lib\site-
packages (from pandas) (1.26.4)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\
dell\appdata\roaming\python\python312\site-packages (from pandas)
(2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in e:\anaconda\lib\site-
packages (from pandas) (2024.1)
Requirement already satisfied: tzdata>=2022.7 in e:\anaconda\lib\site-
packages (from pandas) (2023.3)
Requirement already satisfied: six>=1.5 in c:\users\dell\appdata\
roaming\python\python312\site-packages (from python-dateutil>=2.8.2-
>pandas) (1.17.0)
Requirement already satisfied: matplotlib in e:\anaconda\lib\site-
packages (3.8.4)
Requirement already satisfied: contourpy>=1.0.1 in e:\anaconda\lib\
site-packages (from matplotlib) (1.2.0)
Requirement already satisfied: cycler>=0.10 in e:\anaconda\lib\site-
packages (from matplotlib) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in e:\anaconda\lib\
site-packages (from matplotlib) (4.51.0)
Requirement already satisfied: kiwisolver>=1.3.1 in e:\anaconda\lib\
site-packages (from matplotlib) (1.4.4)
Requirement already satisfied: numpy>=1.21 in e:\anaconda\lib\site-
packages (from matplotlib) (1.26.4)
Requirement already satisfied: packaging>=20.0 in c:\users\dell\
appdata\roaming\python\python312\site-packages (from matplotlib)
(24.2)
Requirement already satisfied: pillow>=8 in e:\anaconda\lib\site-
packages (from matplotlib) (10.3.0)
Requirement already satisfied: pyparsing>=2.3.1 in e:\anaconda\lib\
site-packages (from matplotlib) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\dell\
appdata\roaming\python\python312\site-packages (from matplotlib)
(2.9.0.post0)
Requirement already satisfied: six>=1.5 in c:\users\dell\appdata\
roaming\python\python312\site-packages (from python-dateutil>=2.7-
>matplotlib) (1.17.0)
```

```
Requirement already satisfied: seaborn in e:\anaconda\lib\site-
packages (0.13.2)
Requirement already satisfied: numpy!=1.24.0,>=1.20 in e:\anaconda\
lib\site-packages (from seaborn) (1.26.4)
Requirement already satisfied: pandas>=1.2 in e:\anaconda\lib\site-
packages (from seaborn) (2.2.2)
Requirement already satisfied: matplotlib!=3.6.1,>=3.4 in e:\anaconda\
lib\site-packages (from seaborn) (3.8.4)
Requirement already satisfied: contourpy>=1.0.1 in e:\anaconda\lib\
site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (1.2.0)
Requirement already satisfied: cycler>=0.10 in e:\anaconda\lib\site-
packages (from matplotlib!=3.6.1,>=3.4->seaborn) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in e:\anaconda\lib\
site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (4.51.0)
Requirement already satisfied: kiwisolver>=1.3.1 in e:\anaconda\lib\
site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (1.4.4)
Requirement already satisfied: packaging>=20.0 in c:\users\dell\
appdata\roaming\python\python312\site-packages (from matplotlib!
=3.6.1,>=3.4->seaborn) (24.2)
Requirement already satisfied: pillow>=8 in e:\anaconda\lib\site-
packages (from matplotlib!=3.6.1,>=3.4->seaborn) (10.3.0)
Requirement already satisfied: pyparsing>=2.3.1 in e:\anaconda\lib\
site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\dell\
appdata\roaming\python\python312\site-packages (from matplotlib!
=3.6.1,>=3.4->seaborn) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in e:\anaconda\lib\site-
packages (from pandas>=1.2->seaborn) (2024.1)
Requirement already satisfied: tzdata>=2022.7 in e:\anaconda\lib\site-
packages (from pandas>=1.2->seaborn) (2023.3)
Requirement already satisfied: six>=1.5 in c:\users\dell\appdata\
roaming\python\python312\site-packages (from python-dateutil>=2.7-
>matplotlib!=3.6.1,>=3.4->seaborn) (1.17.0)
```

```
df = pd.read_csv('Customer Churn.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
```

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

```
SeniorCitizen     7044.530758
tenure            7264.930630
MonthlyCharges    7470.551740
dtype: float64
```

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

```
0
```

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

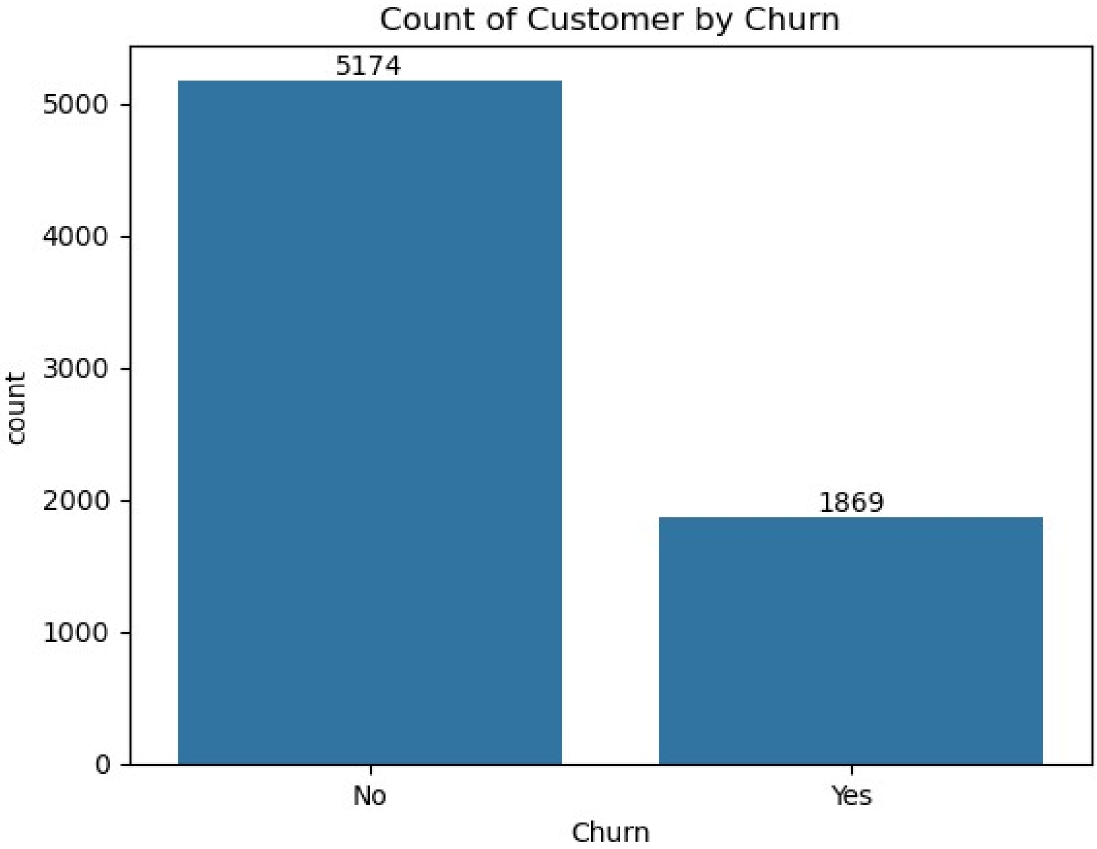
```
0
```

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

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

df['SeniorCitizen'] = df['SeniorCitizen'].astype("object")


df.info()
df
```

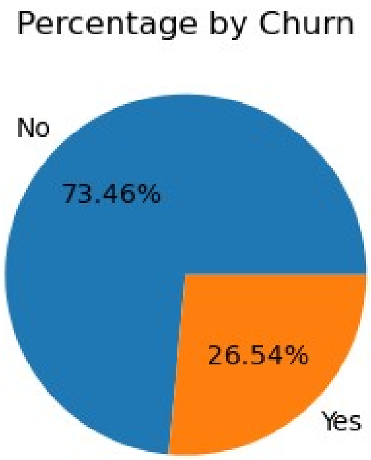
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 21 columns):
 #   Column            Non-Null Count  Dtype 
---  ------            --------------  ----- 
 0   customerID        7043 non-null   object
 1   gender            7043 non-null   object
 2   SeniorCitizen     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

      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
1889.50
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
```

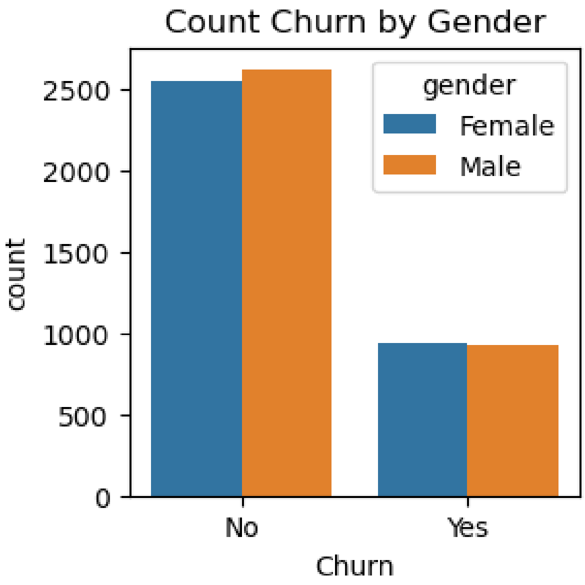
```
6844.50

      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]

ax = sns.countplot(x = 'Churn', data = df)
ax.bar_label(ax.containers[0])
plt.title("Count of Customer by Churn")
plt.show()
```



```
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 Churn")
plt.show()
```

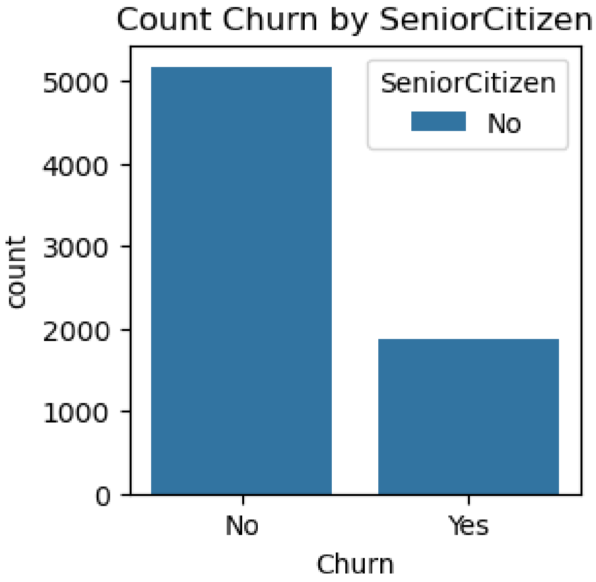


```
plt.figure(figsize = (3,3))
sns.countplot(x = 'Churn' , data = df, hue = 'gender')
plt.title("Count Churn by Gender")
plt.show()
```



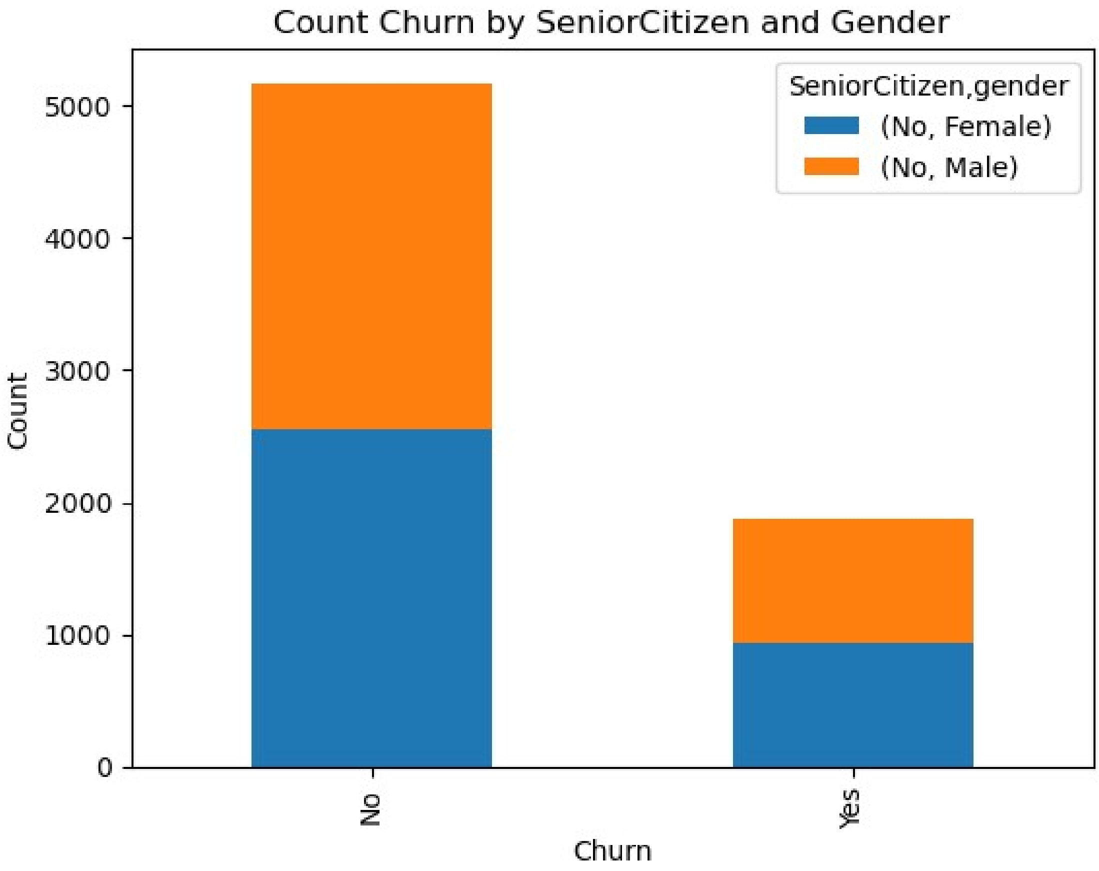
```
def conv(value):
    if value == 1:
```

```python
        return "Yes"
    else:
        return "No"
df['SeniorCitizen'] =df['SeniorCitizen'].apply(conv)
plt.figure(figsize = (3,3))
sns.countplot(x = 'Churn' , data = df, hue = 'SeniorCitizen')
plt.title("Count Churn by SeniorCitizen")
plt.show()
```



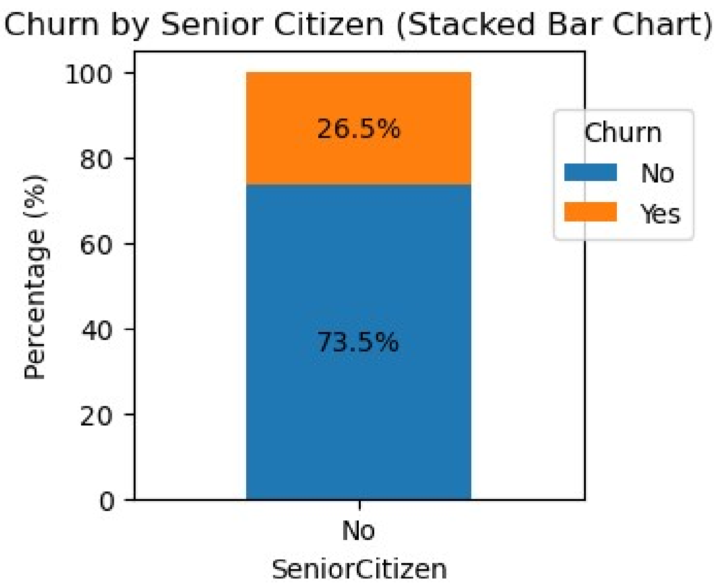
```python
# Create a cross-tabulation
cross_tab = pd.crosstab(index=df['Churn'],
                        columns=[df['SeniorCitizen'], df['gender']])

# Plot stacked bars
plt.figure(figsize=(3,3))
cross_tab.plot(kind='bar', stacked=True)
plt.title("Count Churn by SeniorCitizen and Gender")
plt.ylabel("Count")
plt.show()
```

```
<Figure size 300x300 with 0 Axes>
```

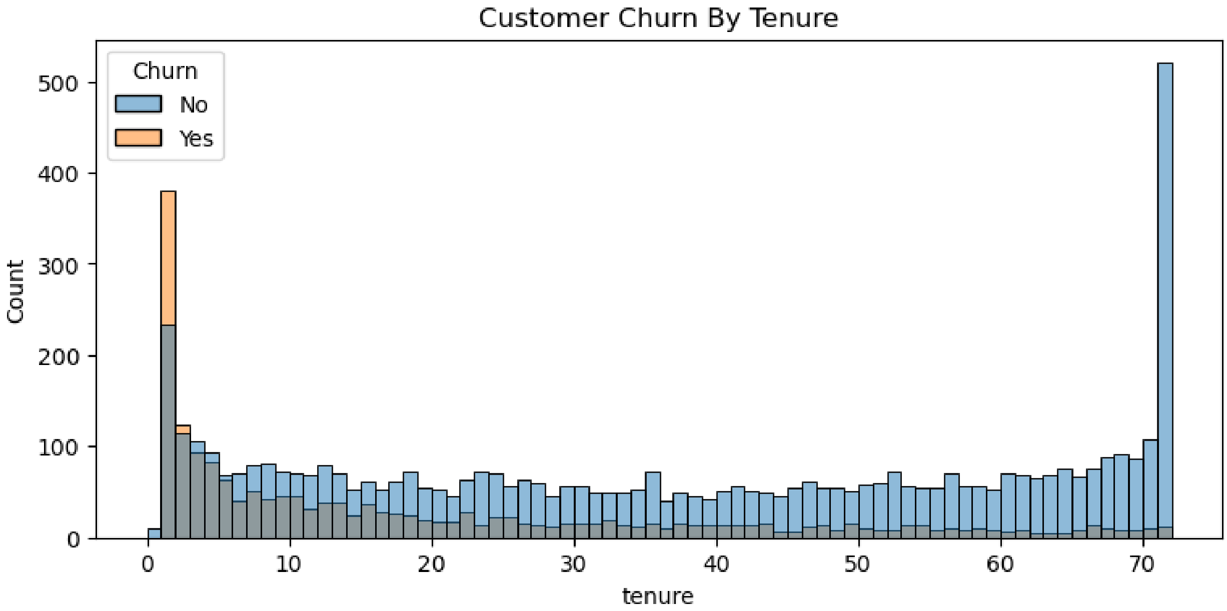


```
total_counts = df.groupby('SeniorCitizen')
['Churn'].value_counts(normalize=True).unstack() * 100

# Plot
fig, ax = plt.subplots(figsize=(3, 3))  # Adjust figsize for better
visualization

# Plot the bars
total_counts.plot(kind='bar', stacked=True, ax=ax, color=['#1f77b4',
'#ff7f0e'])  # Customize colors if desired

# Add percentage labels on the bars
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))  # Customize
```

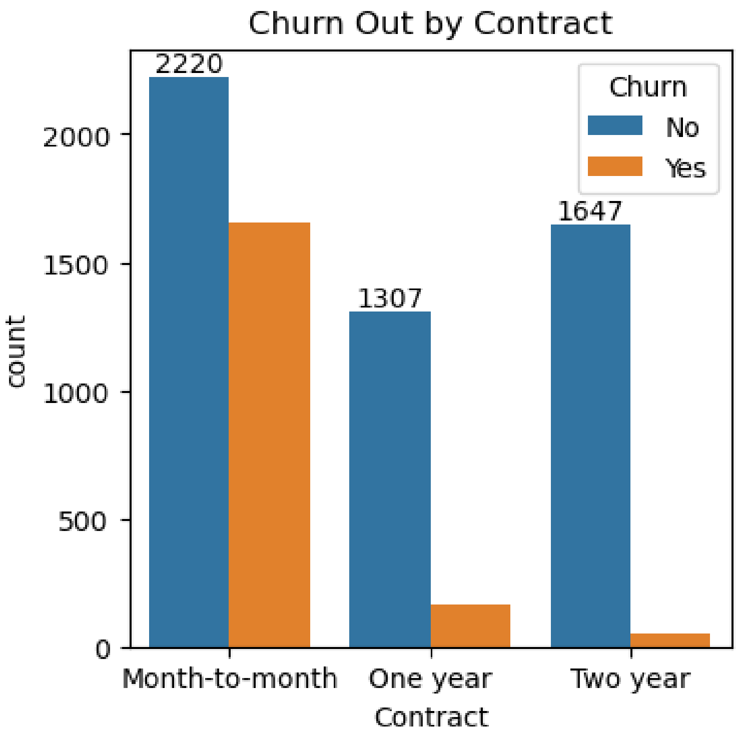
```
legend location

plt.show()
```



```
plt.figure(figsize = (9,4))
sns.histplot(x = "tenure" , data = df, bins = 72, hue = "Churn")
plt.title("Customer Churn By Tenure")
plt.show()
```



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



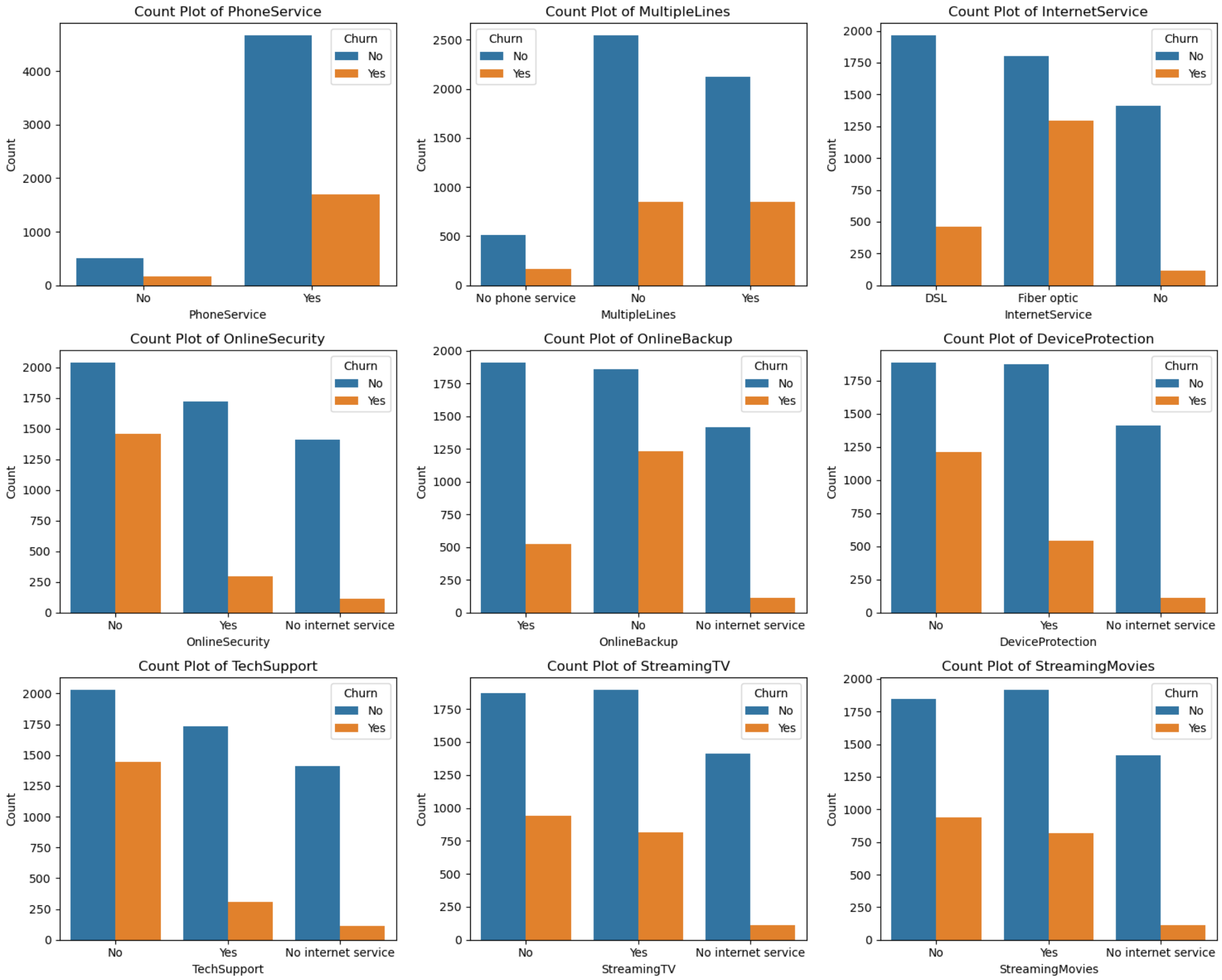
#people who have month to month contract has churn than people who have one year or two year contract

```
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)
```

```
columns = ['PhoneService', 'MultipleLines', 'InternetService',
'OnlineSecurity',
           'OnlineBackup', 'DeviceProtection', 'TechSupport',
'StreamingTV', 'StreamingMovies']

# Number of columns for the subplot grid (you can change this)
n_cols = 3
n_rows = (len(columns) + n_cols - 1) // n_cols  # Calculate number of
```

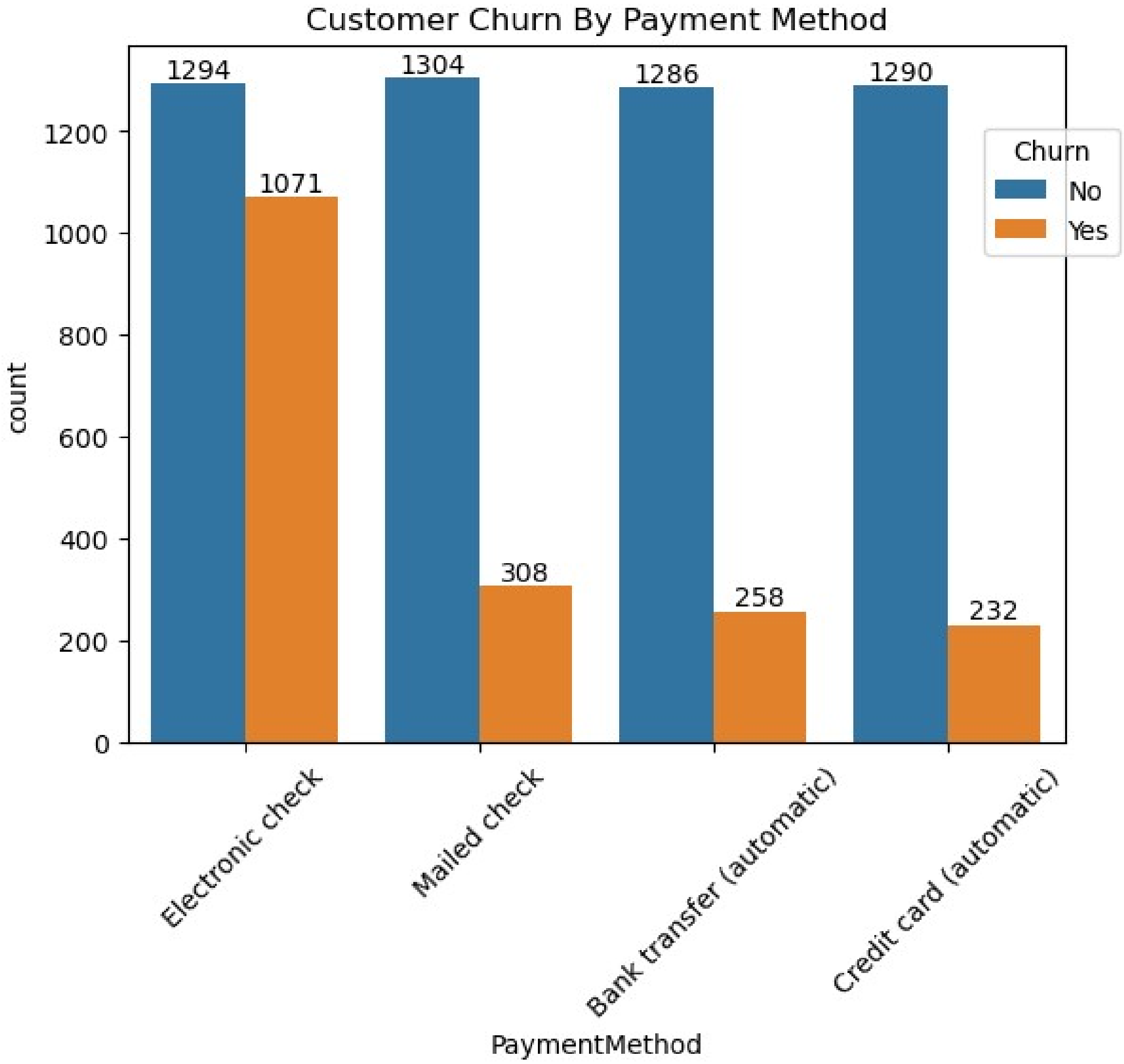
```
rows needed

# Create subplots
fig, axes = plt.subplots(n_rows, n_cols, figsize=(15, n_rows * 4))  #
Adjust figsize as needed

# Flatten the axes array for easy iteration (handles both 1D and 2D
arrays)
axes = axes.flatten()

# Iterate over columns and plot count plots
for i, col in enumerate(columns):
    sns.countplot(x=col, data=df, ax=axes[i], hue = df["Churn"])
    axes[i].set_title(f'Count Plot of {col}')
    axes[i].set_xlabel(col)
    axes[i].set_ylabel('Count')

# Remove empty subplots (if any)
for j in range(i + 1, len(axes)):
    fig.delaxes(axes[j])

plt.tight_layout()
plt.show()
```



#Customers without internet Services Churn more Frequently

#Fiber optic internet show higher churn compared to DSL users

```
ax = sns.countplot(x = "PaymentMethod", data = df, hue = "Churn")
ax.bar_label(ax.containers[0])
ax.bar_label(ax.containers[1])
plt.xticks(rotation = 45)
plt.title("Customer Churn By Payment Method")
plt.legend(title='Churn', bbox_to_anchor = (0.9,0.9))
plt.show()
```



#Customer Churn out mostly when the payment method is Electrronic Check