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

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]
```

Replacing Blanks with 0 and changing type to float

```
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.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
```

```
# to check if there are null values in dataset

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()
```

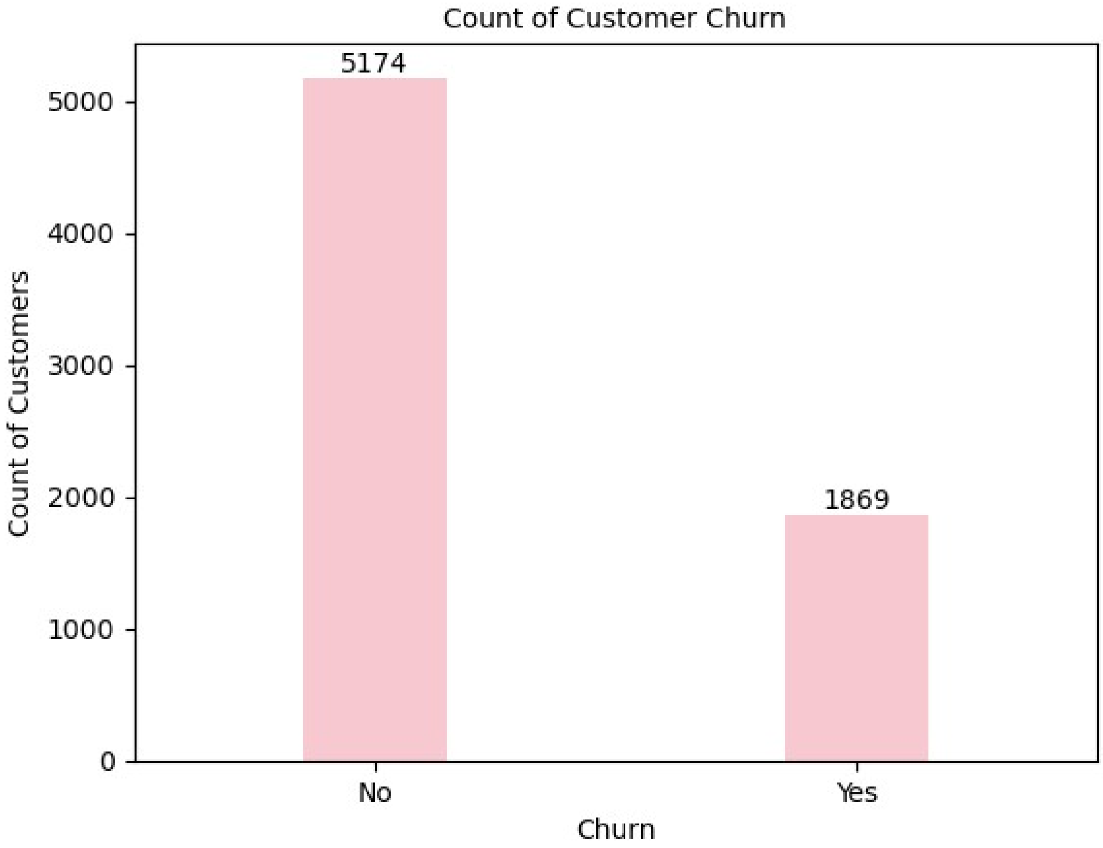
```
0
```

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

df["SeniorCitizen"]= df["SeniorCitizen"].apply(convert_type)

df
```

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

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

     DeviceProtection TechSupport StreamingTV StreamingMovies
Contract  \
0                  No          No          No              No  Month-
to-month
1                 Yes          No          No              No  
One year
2                  No          No          No              No  Month-
to-month
3                 Yes         Yes          No              No  
One year
4                  No          No          No              No  Month-
to-month
...               ...         ...         ...             ...  
...
7038              Yes         Yes         Yes             Yes  
One year
7039              Yes          No         Yes             Yes  
One year
7040               No          No          No              No  Month-
to-month
7041               No          No          No              No  Month-
to-month
7042              Yes         Yes         Yes             Yes  
Two year

     PaperlessBilling              PaymentMethod MonthlyCharges 
TotalCharges  \
0                 Yes           Electronic check          29.85  
29.85
1                  No               Mailed check          56.95  
1889.50
2                 Yes               Mailed check          53.85  
108.15
3                  No  Bank transfer (automatic)          42.30  
1840.75
4                 Yes           Electronic check          70.70  
151.65
...               ...                        ...            ...  
...
7038              Yes               Mailed check          84.80  
1990.50
7039              Yes    Credit card (automatic)         103.20  
7362.90
7040              Yes           Electronic check          29.60  
346.45
7041              Yes               Mailed check          74.40  
306.60
7042              Yes  Bank transfer (automatic)         105.65  
6844.50

      Churn
```

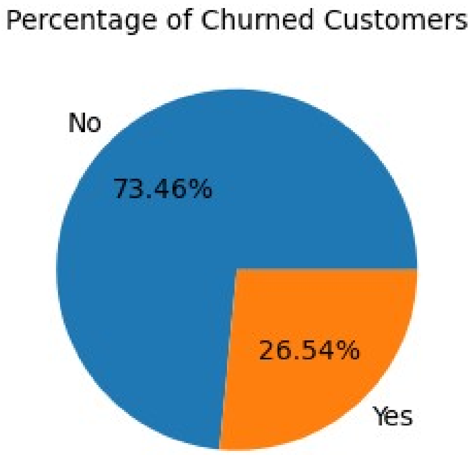
```
0         No
1         No
2        Yes
3         No
4        Yes
...      ...
7038      No
7039      No
7040      No
7041     Yes
7042      No

[7043 rows x 21 columns]
```

```
ax=sns.countplot(x="Churn",data=df,color='pink',width=0.3)
ax.bar_label(ax.containers[0])
plt.ylabel("Count of Customers")
plt.title("Count of Customer Churn",fontsize=10,color='black')
plt.show()
```

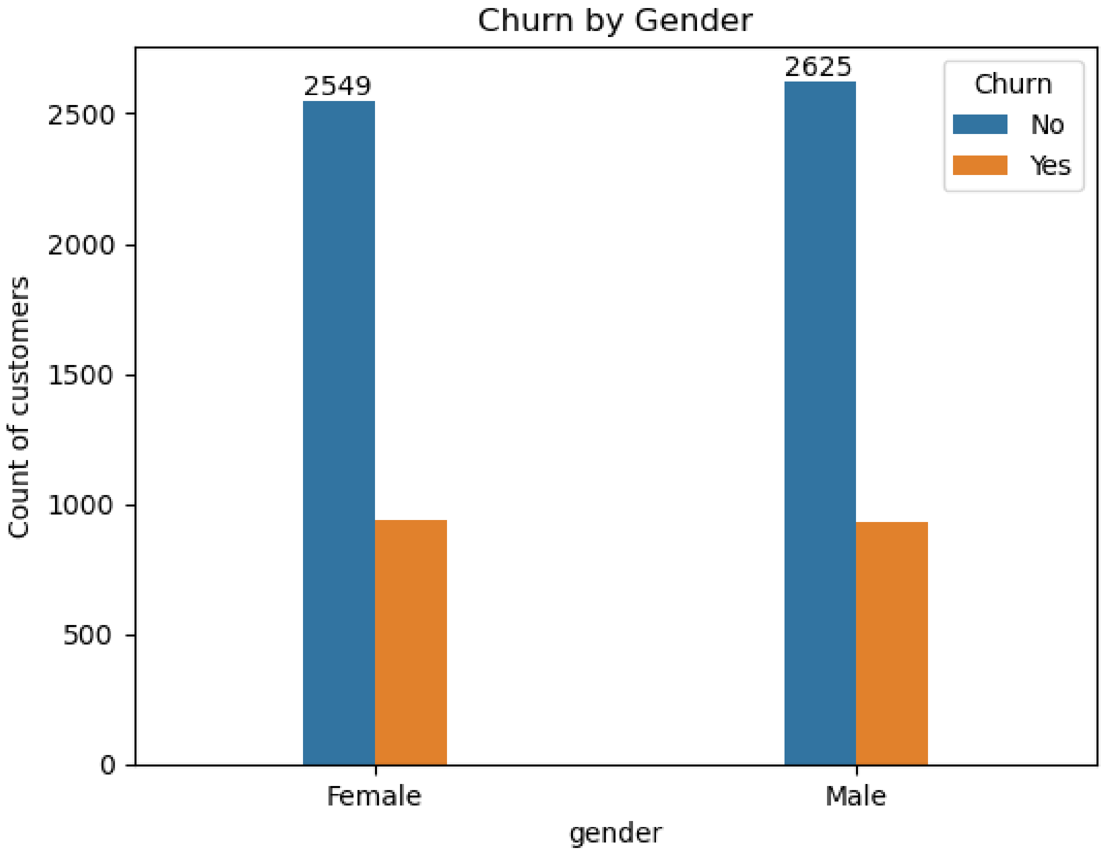


```
plt.figure(figsize=(3,4))

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

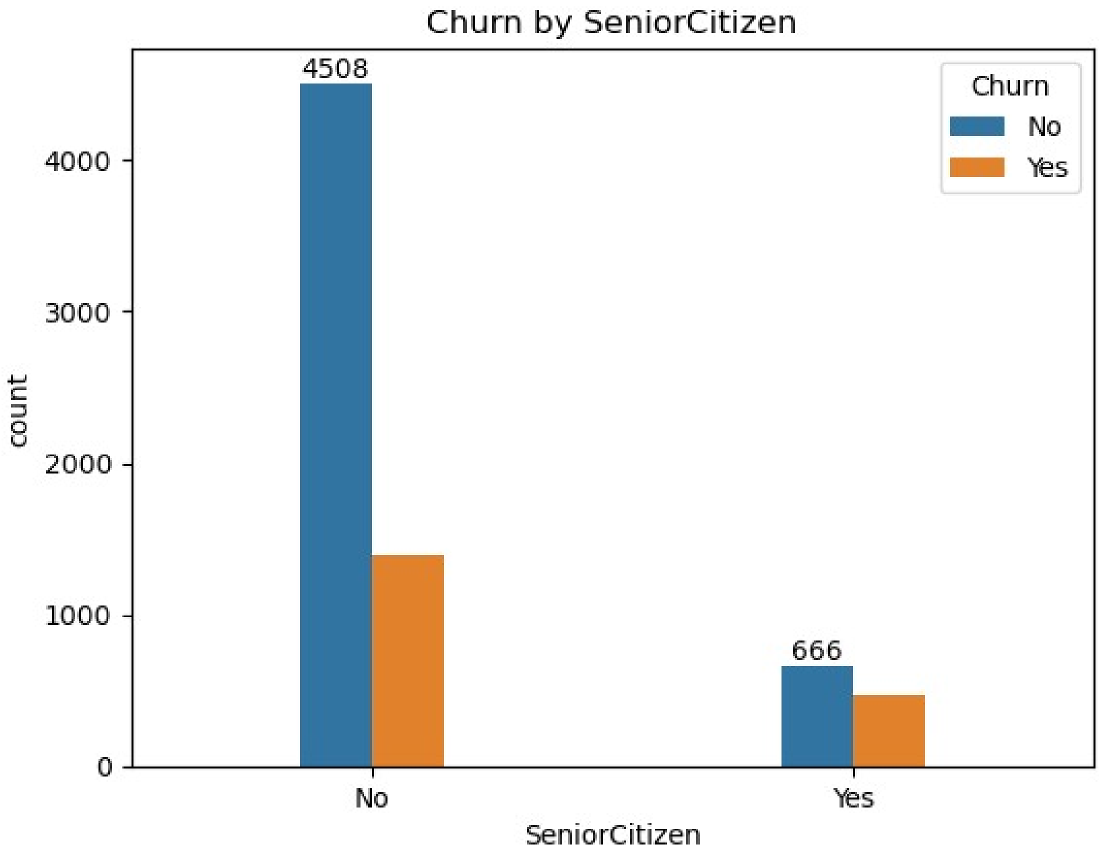
```python
gb

plt.pie(gb["Churn"],labels=gb.index,autopct="%1.2f%%")
plt.title("Percentage of Churned Customers",fontsize=10)
plt.show()
```



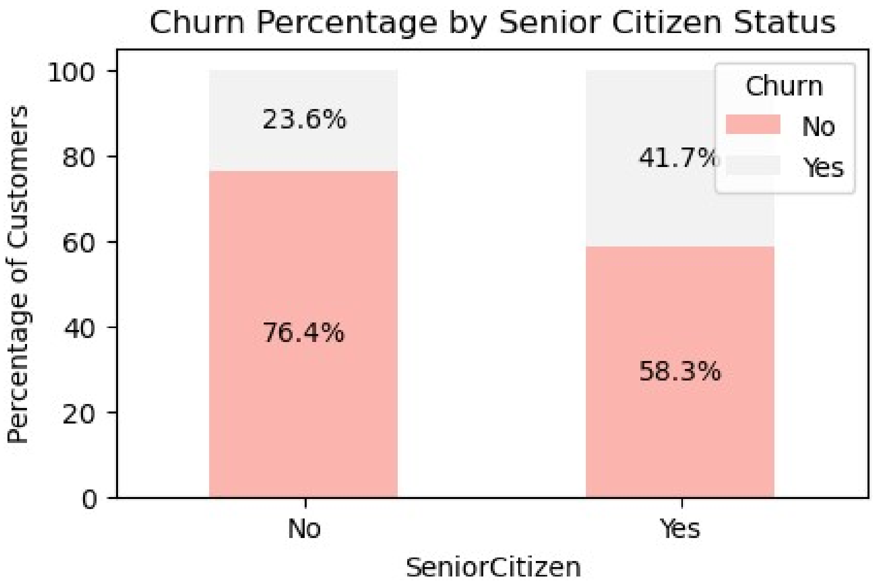
```python
a=sns.countplot(x='gender',data=df,width=0.3,hue="Churn")
a.bar_label(a.containers[0])
plt.ylabel("Count of customers")
plt.title("Churn by Gender")
plt.show()
```



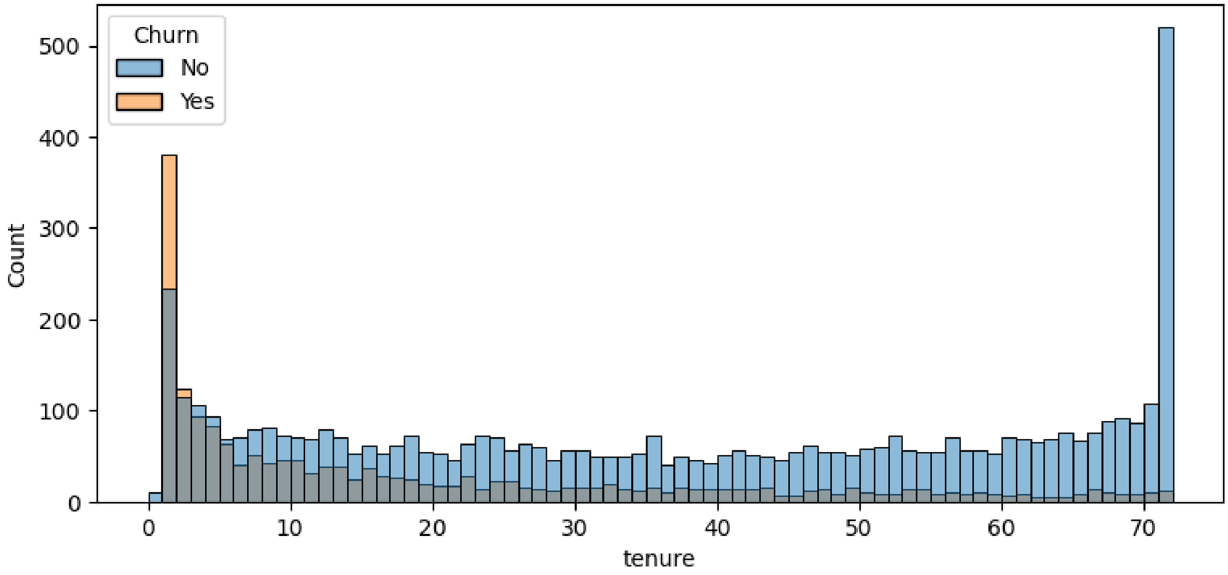
```
aq=sns.countplot(x='SeniorCitizen',data=df,width=0.3,hue="Churn")
aq.bar_label(aq.containers[0])
plt.title("Churn by SeniorCitizen")
plt.show()
```



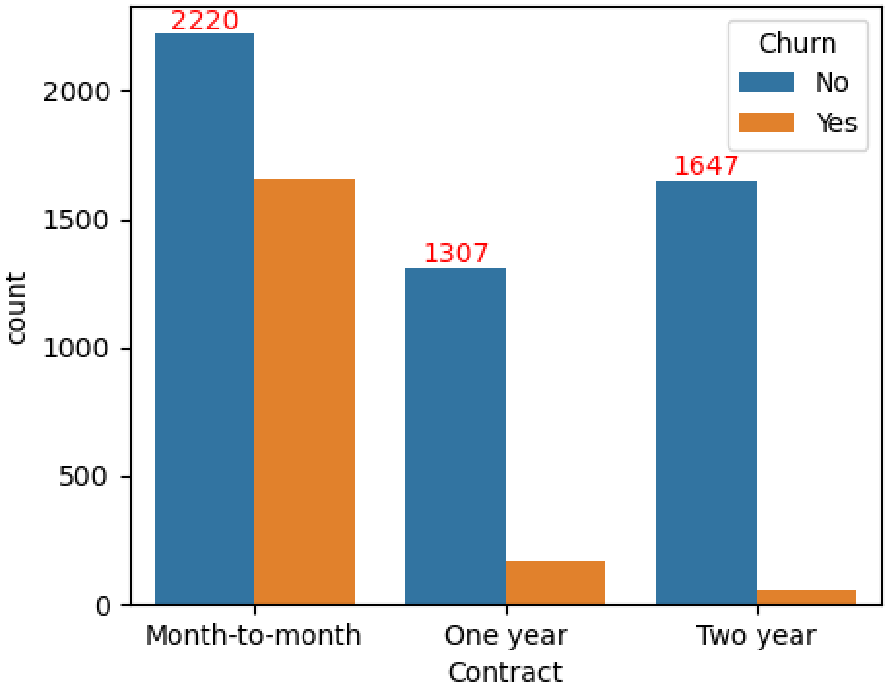
```python
senior_churn = pd.crosstab(df['SeniorCitizen'], df['Churn'],
normalize='index') * 100

# Plot the stacked bar chart
ax = senior_churn.plot(kind='bar', stacked=True, colormap='Pastel1',
figsize=(5,3))

# Add percentage labels
for c in ax.containers:
    ax.bar_label(c, fmt='%.1f%%', label_type='center')

# Formatting
plt.ylabel("Percentage of Customers")
plt.title("Churn Percentage by Senior Citizen Status")
plt.legend(title="Churn", loc="upper right")
plt.xticks(rotation=0)
plt.show()
```



```
plt.figure(figsize=(9,4))
r=sns.histplot(x='tenure',data=df,bins=72,hue="Churn")
plt.show()
```



```
plt.figure(figsize=(5,4))
c=sns.countplot(x="Contract",data=df,hue="Churn")
c.bar_label(c.containers[0],color='red')
plt.show()
```

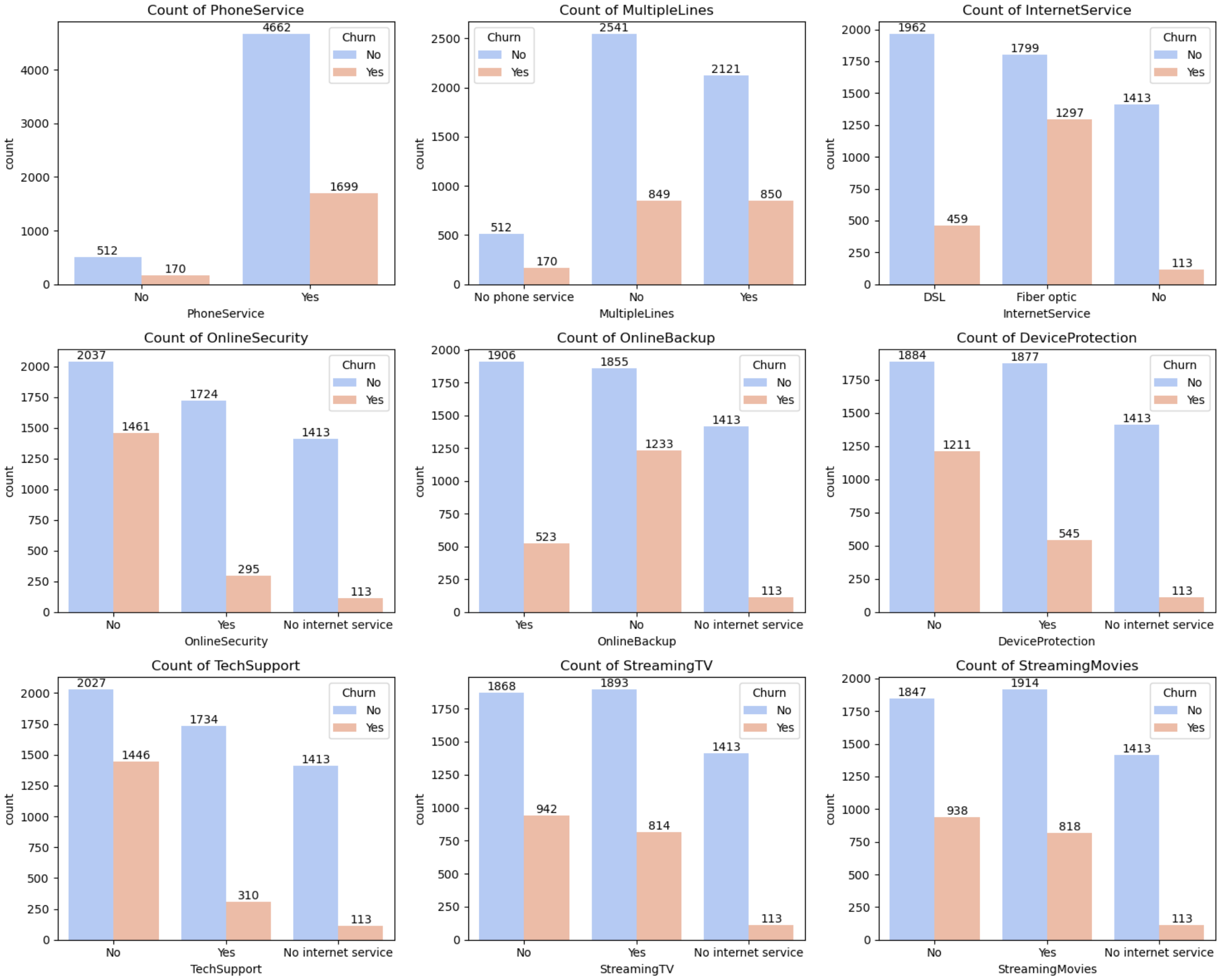


People who have month-to-month contract are more likely to churn than of the one year or two year plans

```
df.columns.values
```

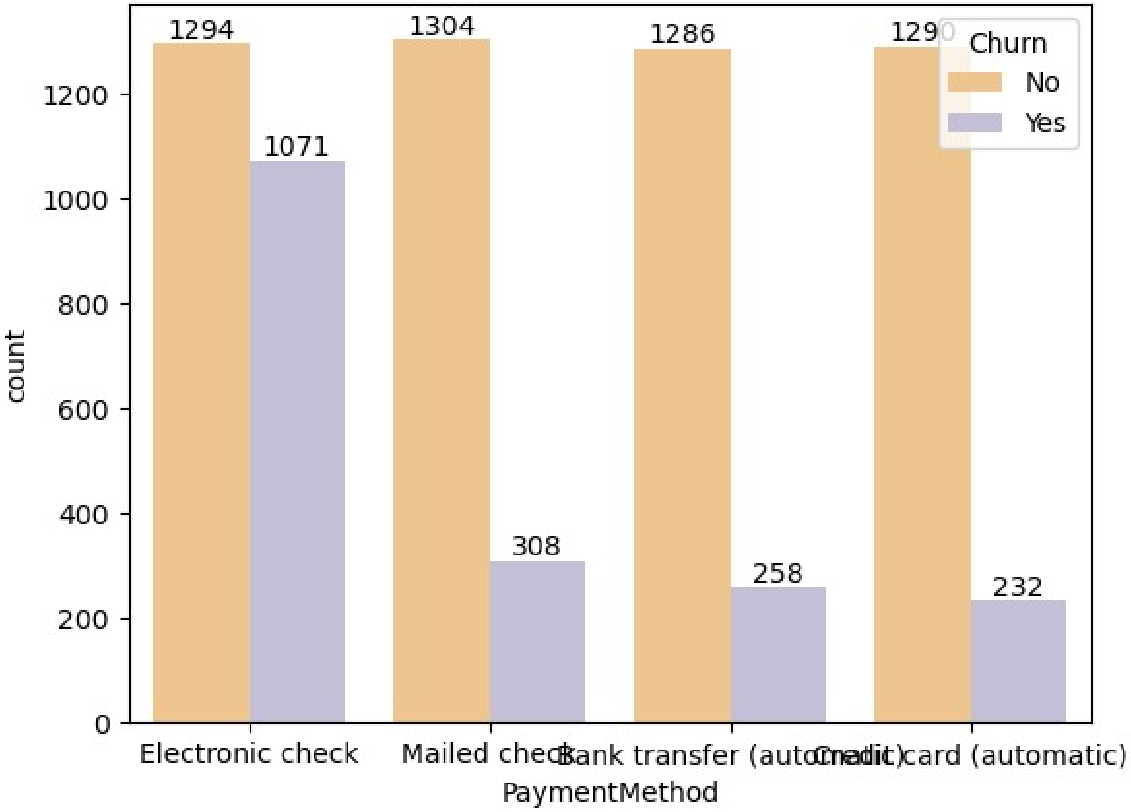
```
array(['customerID', 'gender', 'SeniorCitizen', 'Partner',
'Dependents',
       'tenure', 'PhoneService', 'MultipleLines', 'InternetService',
       'OnlineSecurity', 'OnlineBackup', 'DeviceProtection',
       'TechSupport', 'StreamingTV', 'StreamingMovies', 'Contract',
       'PaperlessBilling',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

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

# Create subplots (3 rows, 3 columns)
fig, axes = plt.subplots(3, 3, figsize=(15, 12))  # Adjust the layout

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

# Generate countplots
for i, col in enumerate(features):
    ax = axes[i]
    sns.countplot(x=df[col], data=df, ax=ax,
palette='coolwarm',hue="Churn")

    # Add count labels
    for c in ax.containers:
        ax.bar_label(c, fmt='%d', label_type='edge', fontsize=10)

    ax.set_title(f"Count of {col}")

# Adjust layout
plt.tight_layout()
plt.show()
```



Higher churn is observed among customers who lack OnlineSecurity, TechSupport, or Streaming services. Fiber optic internet users also have a higher churn rate compared to DSL users. Additionally, customers without PhoneService or MultipleLines show relatively lower churn rates.

```
pm=sns.countplot(x="PaymentMethod",data=df,palette='PuOr',hue="Churn")
pm.bar_label(pm.containers[0])
pm.bar_label(pm.containers[1])
plt.show()
```



Here most churned customers are from the electronic check payment method

```
```