```
# Importing libraries

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

print('Libraries loaded ...')
```

```
Libraries loaded ...
```

```
# load dataset
df= pd.read_csv(r'D:\Learnig Enviroment\Python\Work\Customer
Churn.csv')

# read_csv() for reading csv data
print('Data loaded...')
```

```
Data loaded...
```

```
df.shape # return size of data
```

```
(7043, 21)
```

```
df.head() # returns first 5 rows
```

```
   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]
```

```
df.tail() # returns last 5 rows of data
```

```
      customerID  gender SeniorCitizen Partner Dependents  tenure  \
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  ...  \
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
1990.50
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
7038     No
7039     No
7040     No
7041    Yes
7042     No

[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
```

replacing the value below

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

Changing the datatype below from string to float

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

Checking Null

```
df.isnull() # returns True/false cell wise
```

```
      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]
```

```
df.isnull().sum() # summing column wise total null values
```

```
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
```

checking the information

```
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
```

Basic Statistics

```
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
```

Check Duplicate in data

```
df.duplicated() # Returnds True or False
```

```
0       False
1       False
2       False
3       False
4       False
        ...
7038    False
7039    False
7040    False
7041    False
7042    False
Length: 7043, dtype: bool
```

```
# Summing Total Duplicate value
df.duplicated().sum()
```

```
0
```

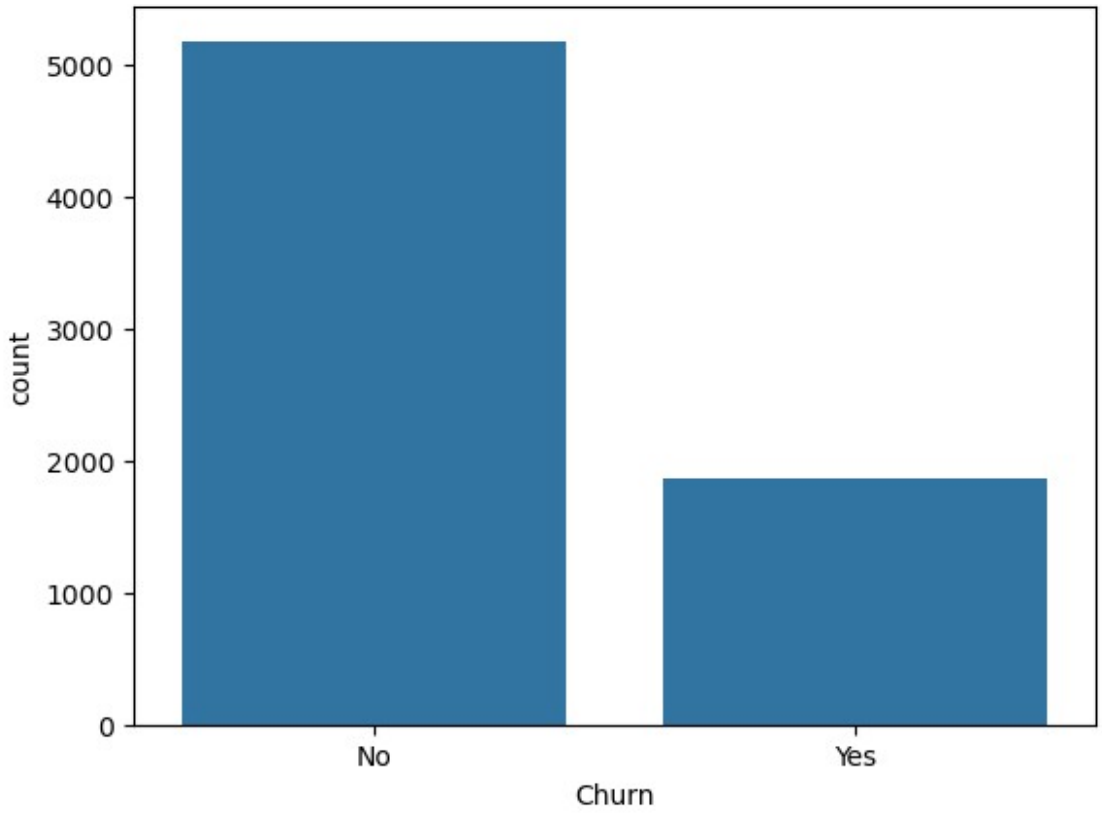
```
# converting the Senior Citizen Yes No

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


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

df.head(10)
```

```
   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
```

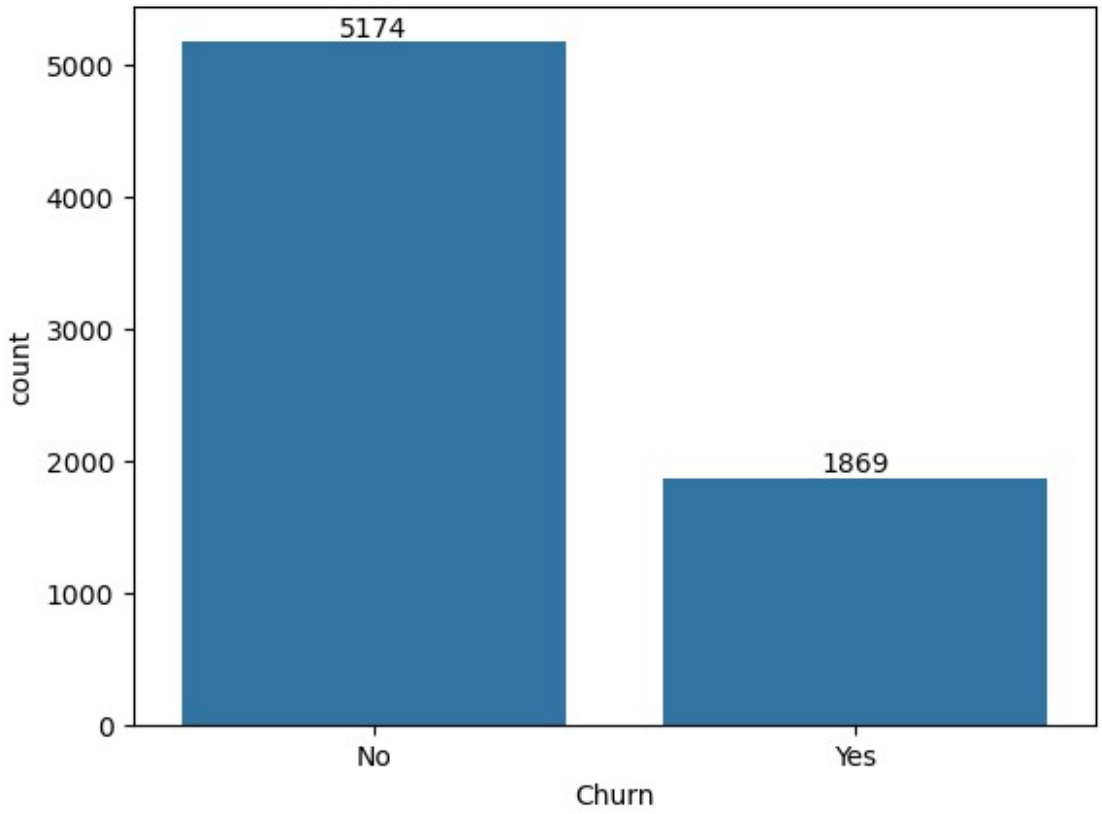
```
2  3668-QPYBK    Male            no      No         No       2
Yes
3  7795-CFOCW    Male            no      No         No      45
No
4  9237-HQITU  Female            no      No         No       2
Yes
5  9305-CDSKC  Female            no      No         No       8
Yes
6  1452-KIOVK    Male            no      No        Yes      22
Yes
7  6713-OKOMC  Female            no      No         No      10
No
8  7892-POOKP  Female            no     Yes         No      28
Yes
9  6388-TABGU    Male            no      No        Yes      62
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
5               Yes     Fiber optic             No  ...
Yes
6               Yes     Fiber optic             No  ...
No
7  No phone service             DSL            Yes  ...
No
8               Yes     Fiber optic             No  ...
Yes
9                No             DSL            Yes  ...
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
5          No          Yes              Yes  Month-to-month
Yes
6          No          Yes               No  Month-to-month
Yes
7          No           No               No  Month-to-month
No
8         Yes          Yes              Yes  Month-to-month
Yes
9          No           No               No        One year
No

                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            Electronic check          99.65        820.50    Yes
6    Credit card (automatic)          89.10       1949.40     No
7                Mailed check          29.75        301.90     No
8            Electronic check         104.80       3046.05    Yes
9  Bank transfer (automatic)          56.15       3487.95     No

[10 rows x 21 columns]
```

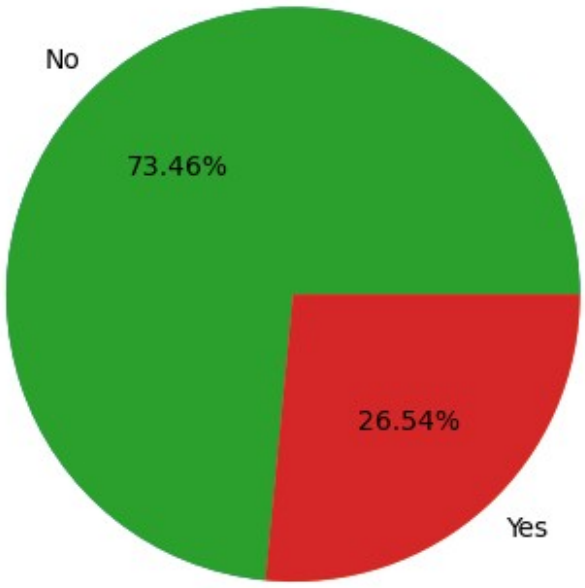
```
sns.countplot(x='Churn', data=df)
plt.show()
```




```
ax=sns.countplot(x='Churn', data=df)
ax.bar_label(ax.containers[0])
plt.show()
```

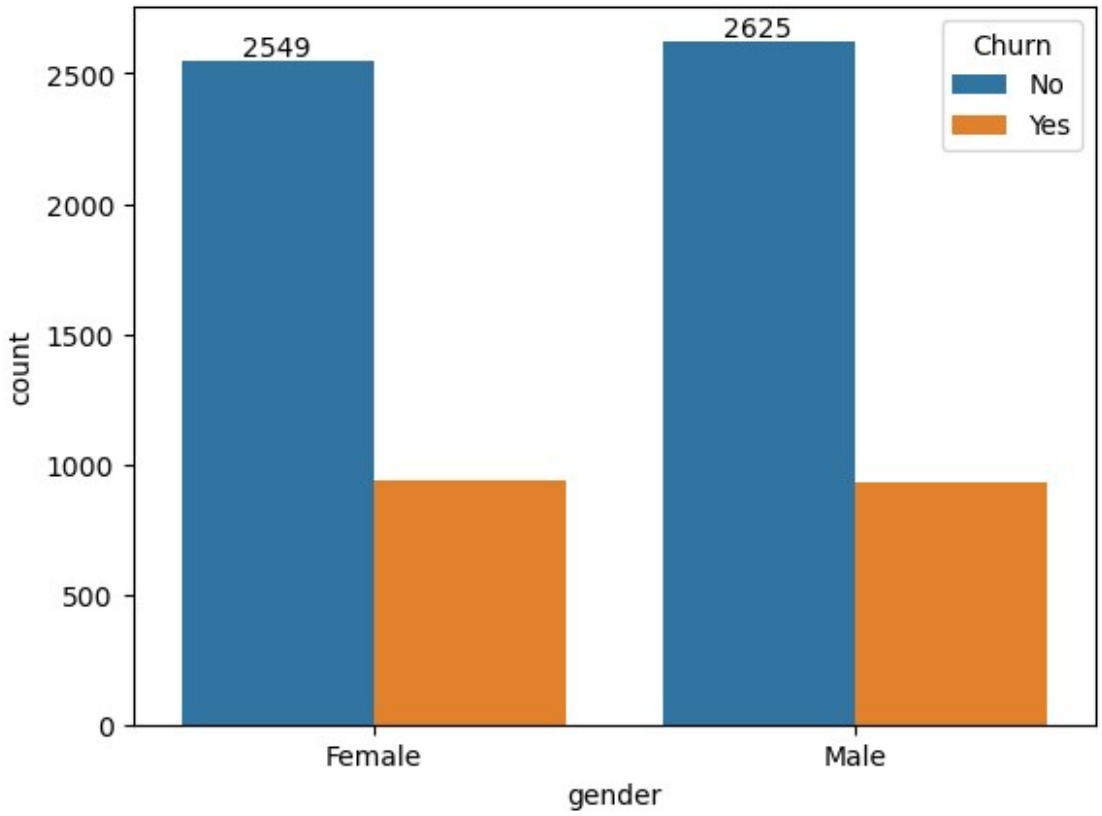



```
gb = df.groupby('Churn').agg({'Churn':'count'})
plt.pie(gb['Churn'],labels=gb.index, autopct='%1.2f%%')
plt.pie(gb['Churn'])

plt.show()
```



Gender wise Churn Analysis

```
ax= sns.countplot(x='gender', data=df, hue='Churn')
ax.bar_label(ax.containers[0])
plt.show()
```

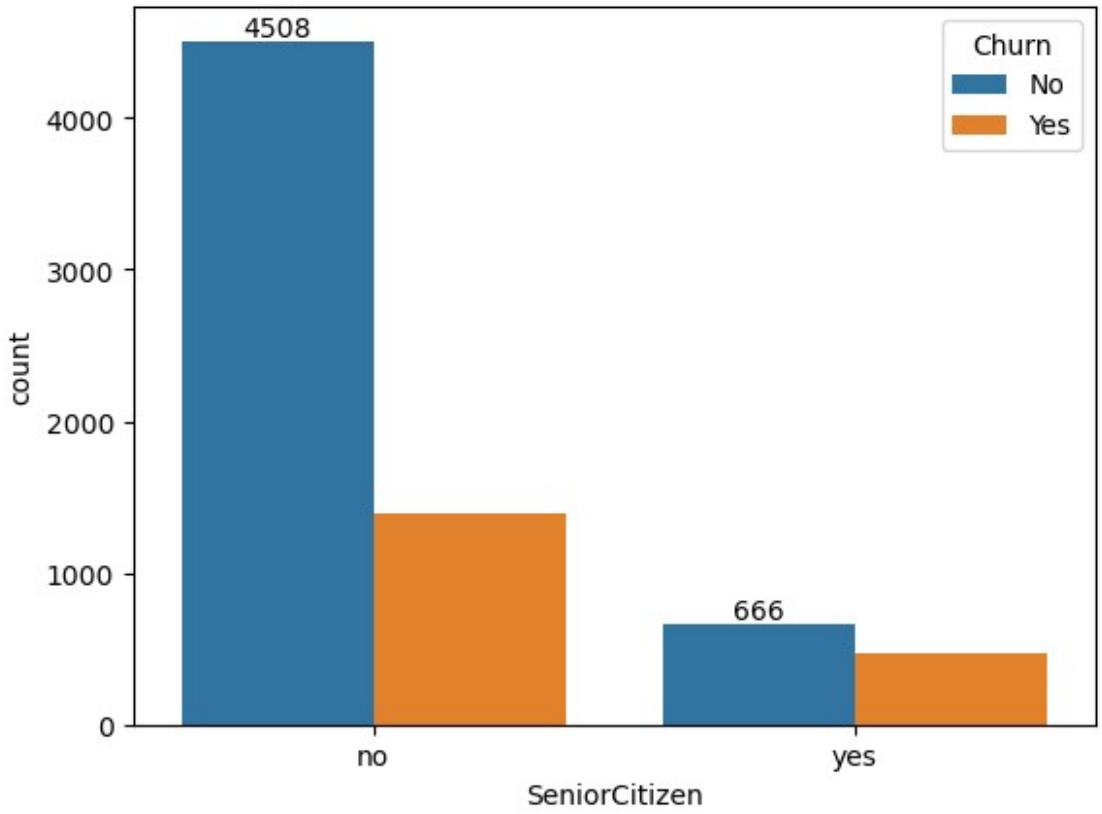



```
df.columns
```

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

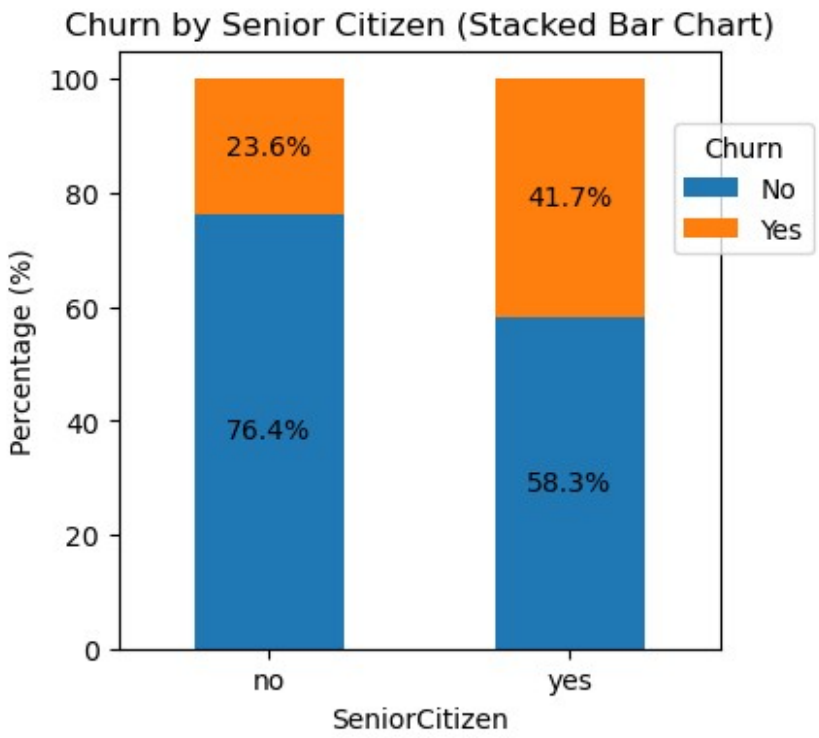
Senior Citizen Churn Analysis

```
ax= sns.countplot(x='SeniorCitizen', data=df, hue='Churn')
ax.bar_label(ax.containers[0])
plt.show()
```



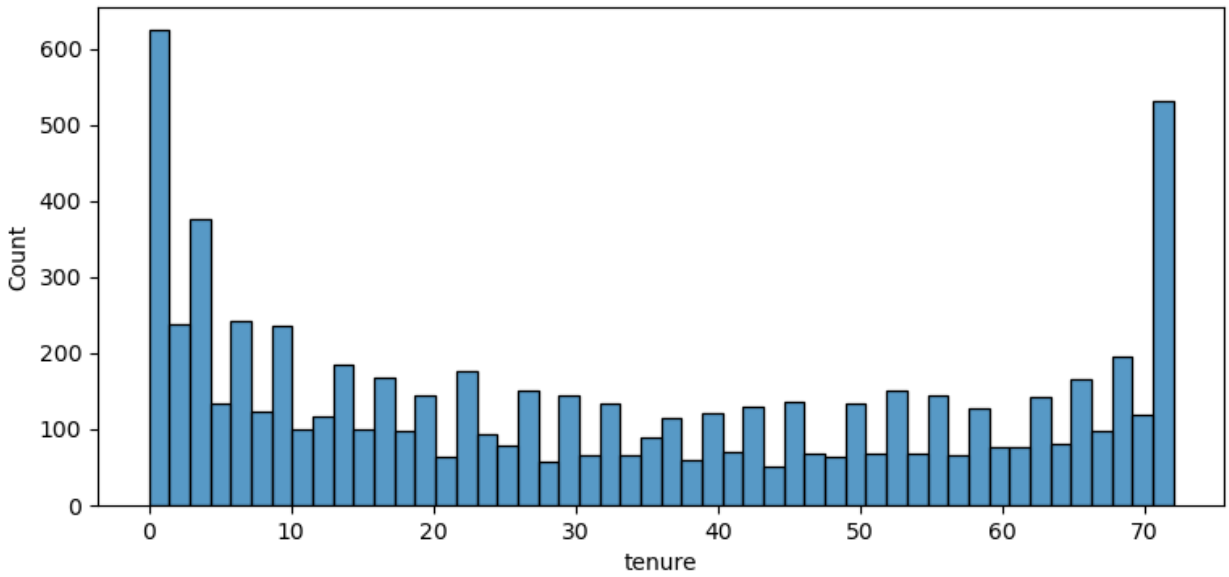

```
total_counts = df.groupby('SeniorCitizen')
['Churn'].value_counts(normalize=True).unstack() * 100

# Plot
fig, ax = plt.subplots(figsize=(4, 4))  # 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()
```

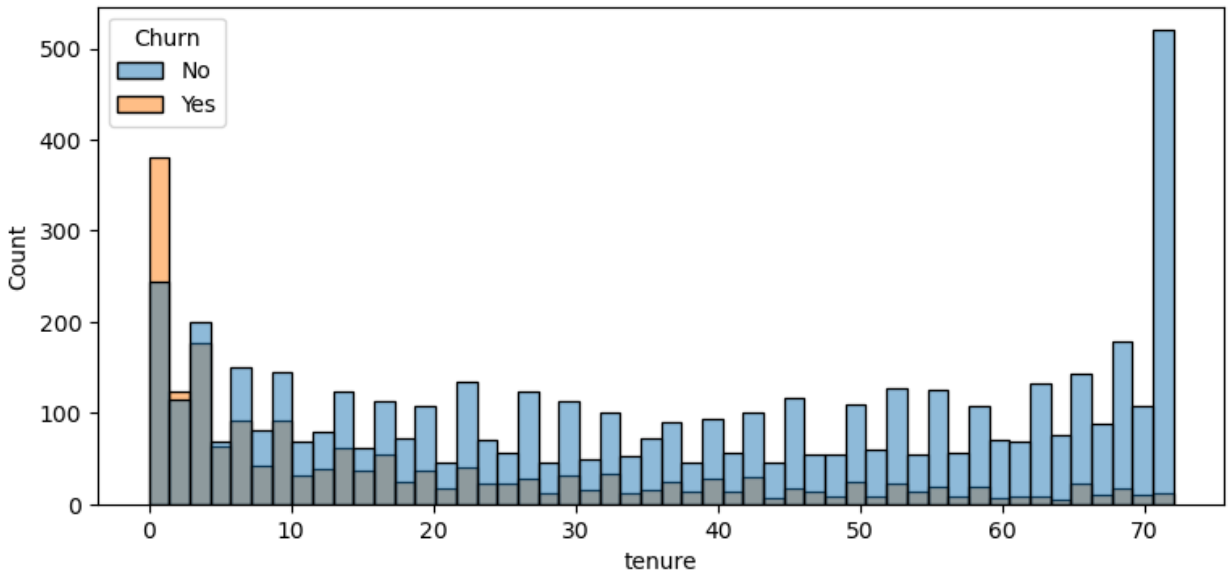


Tenure Based Churn analysis

```
plt.figure(figsize=(9,4))
sns.histplot(x='tenure', data=df, bins=50)
plt.show()
```

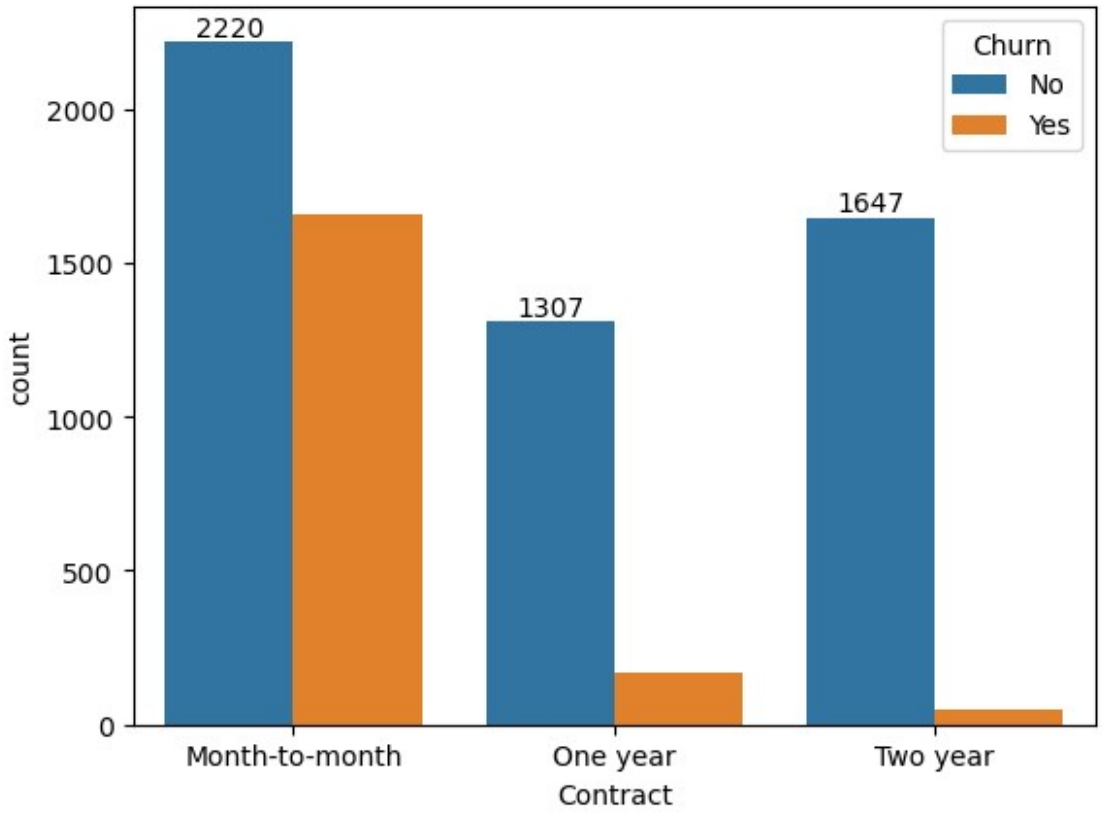



```
# Churn Status
plt.figure(figsize=(9,4))
sns.histplot(x='tenure', data=df, bins=50, hue='Churn')
plt.show()
```



contract Based Churn Analysis

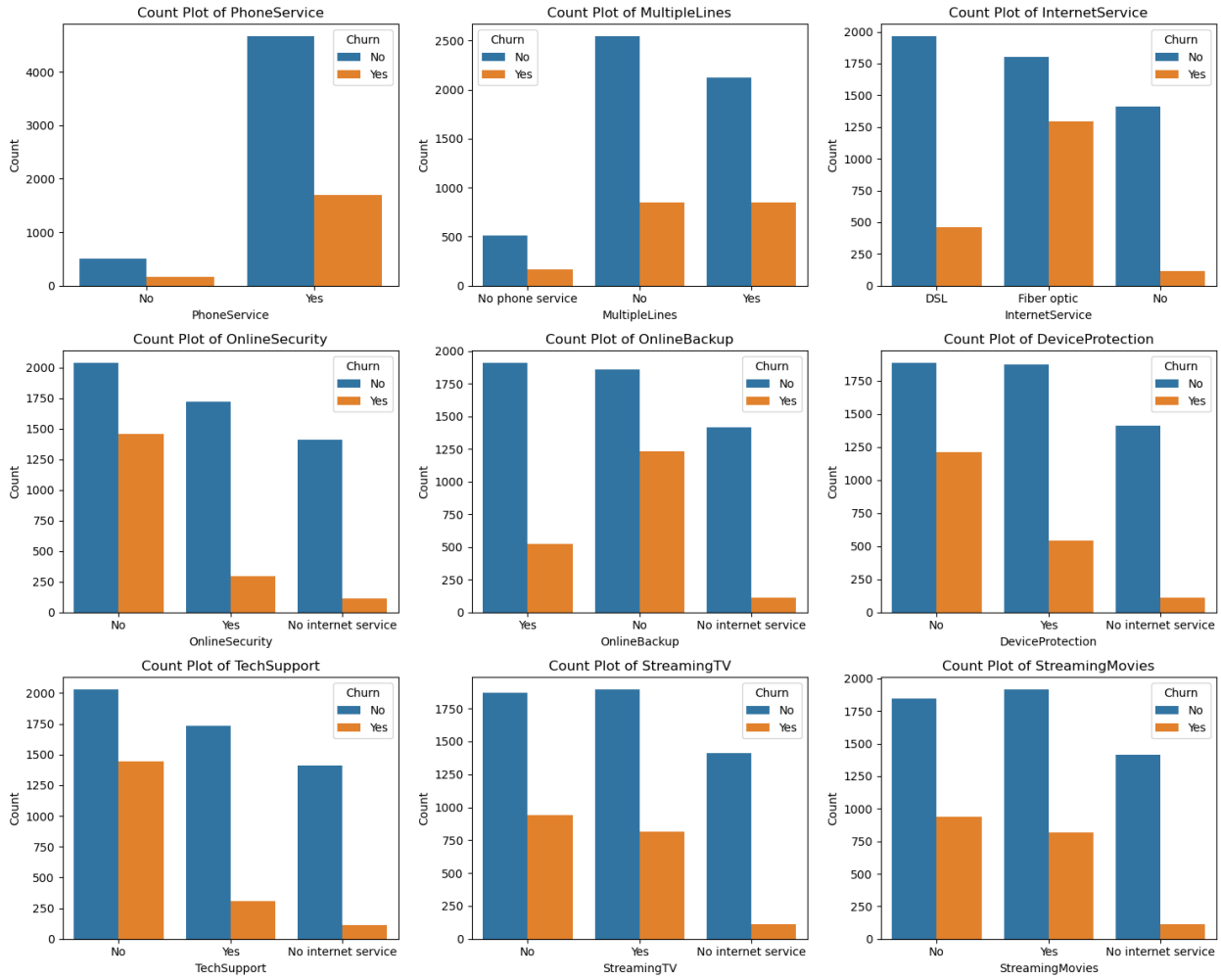
```
ax=sns.countplot(x='Contract', data=df, hue ='Churn')
ax.bar_label(ax.containers[0])
plt.show()
```



Above we can see that month to month contract customer has high churn number.

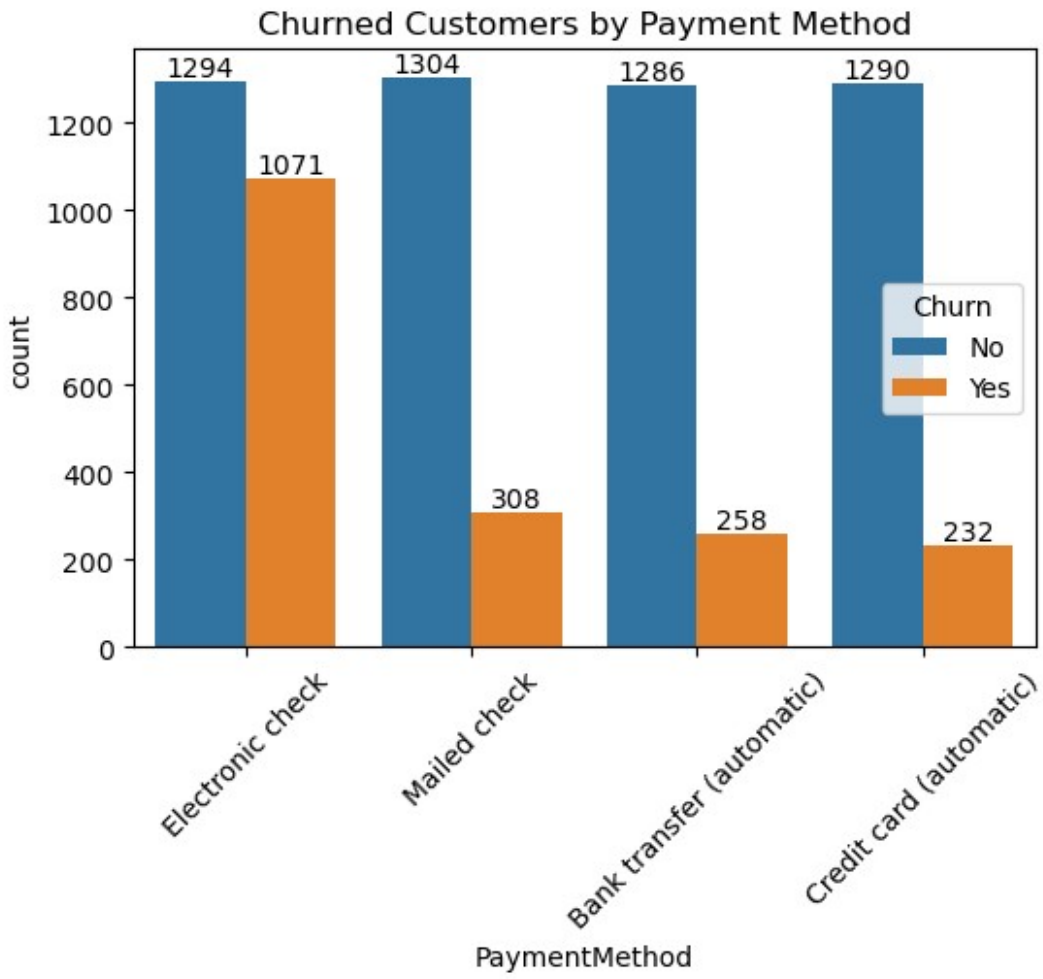
```
df.columns.values

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()
```



The majority of customers who do not churn tend to have services like PhoneService, 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("Churned Customers by Payment Method")
plt.xticks(rotation = 45)
plt.show()
```



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

```
Follow me for more Updates ...
linkedIn >>> https://www.linkedin.com/in/vivekvishwas/
GitHUb >>> https://github.com/Official-Vivek-Singh
```