EDA of Customer Churn using python

Importing libraries

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

Data Preprocessing

Reading CSV file 'Customer-Churn.csv'

```
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]
```

Displaying few rows of datasets

```
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]
```

Getting summary of Data frame structure like column name, no. of not null column,datatype nd memory usages

```
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
```

Column 'TotalCharges' contain object Dtype, it has numerical value & has to be float Dtype, so identifying that due to null value it datatype is object so replacing it with 0 and then converting it into float

```
df['TotalCharges']=df['TotalCharges'].replace(" ","0")
df['TotalCharges']=df['TotalCharges'].astype("float64")
```

```
 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       1889.50     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]
```

After conversion 'TotalCharges' from object to 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
```

To check if the column conatin null value or not

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

```
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
```

Getting the Summary of statistics of numerical column in datasets

```
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
```

Function to convert SeniorCitizencolumn value

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

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

Checking that SeniorCitizen value had converted

```
df.head(30)
```

```
    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
10  9763-GRSKD    Male            No     Yes        Yes      13
Yes
11  7469-LKBCI    Male            No      No         No      16
Yes
12  8091-TTVAX    Male            No     Yes         No      58
Yes
13  0280-XJGEX    Male            No      No         No      49
Yes
14  5129-JLPIS    Male            No      No         No      25
Yes
15  3655-SNQYZ  Female            No     Yes        Yes      69
Yes
16  8191-XWSZG  Female            No      No         No      52
Yes
```

```
17  9959-WOFKT    Male            No      No         Yes      71
Yes
18  4190-MFLUW  Female            No     Yes         Yes      10
Yes
19  4183-MYFRB  Female            No      No          No      21
Yes
20  8779-QRDMV    Male           Yes      No          No       1
No
21  1680-VDCWW    Male            No     Yes          No      12
Yes
22  1066-JKSGK    Male            No      No          No       1
Yes
23  3638-WEABW  Female            No     Yes          No      58
Yes
24  6322-HRPFA    Male            No     Yes         Yes      49
Yes
25  6865-JZNKO  Female            No      No          No      30
Yes
26  6467-CHFZW    Male            No     Yes         Yes      47
Yes
27  8665-UTDHZ    Male            No     Yes         Yes       1
No
28  5248-YGIJN    Male            No     Yes          No      72
Yes
29  8773-HHUOZ  Female            No      No         Yes      17
Yes

       MultipleLines InternetService       OnlineSecurity  ...  \
0   No phone service             DSL                   No  ...
1                 No             DSL                  Yes  ...
2                 No             DSL                  Yes  ...
3   No phone service             DSL                  Yes  ...
4                 No     Fiber optic                   No  ...
5                Yes     Fiber optic                   No  ...
6                Yes     Fiber optic                   No  ...
7   No phone service             DSL                  Yes  ...
8                Yes     Fiber optic                   No  ...
9                 No             DSL                  Yes  ...
10                No             DSL                  Yes  ...
11                No              No  No internet service  ...
12               Yes     Fiber optic                   No  ...
13               Yes     Fiber optic                   No  ...
14                No     Fiber optic                  Yes  ...
15               Yes     Fiber optic                  Yes  ...
16                No              No  No internet service  ...
17               Yes     Fiber optic                  Yes  ...
18                No             DSL                   No  ...
19                No     Fiber optic                   No  ...
20  No phone service             DSL                   No  ...
```

```
21                No             No  No internet service  ...
22                No             No  No internet service  ...
23               Yes            DSL                   No  ...
24                No            DSL                  Yes  ...
25                No            DSL                  Yes  ...
26               Yes    Fiber optic                   No  ...
27  No phone service            DSL                   No  ...
28               Yes            DSL                  Yes  ...
29                No            DSL                   No  ...

       DeviceProtection          TechSupport          StreamingTV  \
0                    No                   No                   No
1                   Yes                   No                   No
2                    No                   No                   No
3                   Yes                  Yes                   No
4                    No                   No                   No
5                   Yes                   No                  Yes
6                    No                   No                  Yes
7                    No                   No                   No
8                   Yes                  Yes                  Yes
9                    No                   No                   No
10                   No                   No                   No
11  No internet service  No internet service  No internet service
12                  Yes                   No                  Yes
13                  Yes                   No                  Yes
14                  Yes                  Yes                  Yes
15                  Yes                  Yes                  Yes
16  No internet service  No internet service  No internet service
17                  Yes                   No                  Yes
18                  Yes                  Yes                   No
19                  Yes                   No                   No
20                  Yes                   No                   No
21  No internet service  No internet service  No internet service
22  No internet service  No internet service  No internet service
23                   No                  Yes                   No
24                   No                  Yes                   No
25                   No                   No                   No
26                   No                   No                  Yes
27                   No                   No                   No
28                  Yes                  Yes                  Yes
29                   No                   No                  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
5                   Yes  Month-to-month              Yes
```

6	No	Month-to-month	Yes	
7	No	Month-to-month	No	
8	Yes	Month-to-month	Yes	
9	No	One year	No	
10	No	Month-to-month	Yes	
11	No internet service	Two year	No	
12	Yes	One year	No	
13	Yes	Month-to-month	Yes	
14	Yes	Month-to-month	Yes	
15	Yes	Two year	No	
16	No internet service	One year	No	
17	Yes	Two year	No	
18	No	Month-to-month	No	
19	Yes	Month-to-month	Yes	
20	Yes	Month-to-month	Yes	
21	No internet service	One year	No	
22	No internet service	Month-to-month	No	
23	No	Two year	Yes	
24	No	Month-to-month	No	
25	No	Month-to-month	Yes	
26	Yes	Month-to-month	Yes	
27	No	Month-to-month	No	
28	Yes	Two year	Yes	
29	Yes	Month-to-month	Yes	

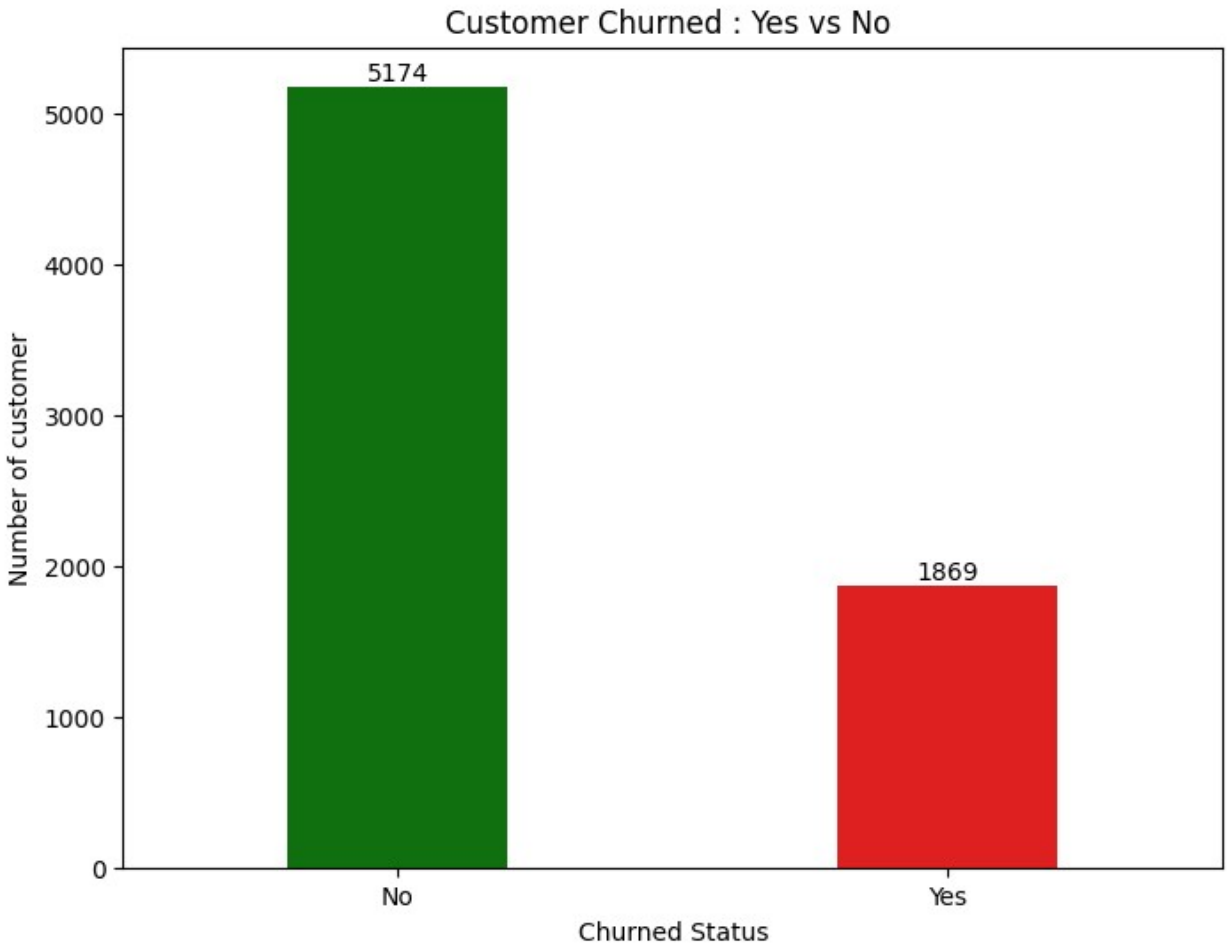
	PaymentMethod	MonthlyCharges	TotalCharges	Churn
0	Electronic check	29.85	29.85	No
1	Mailed check	56.95	1889.50	No
2	Mailed check	53.85	108.15	Yes
3	Bank transfer (automatic)	42.30	1840.75	No
4	Electronic check	70.70	151.65	Yes
5	Electronic check	99.65	820.50	Yes
6	Credit card (automatic)	89.10	1949.40	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	Mailed check	49.95	587.45	No
11	Credit card (automatic)	18.95	326.80	No
12	Credit card (automatic)	100.35	5681.10	No
13	Bank transfer (automatic)	103.70	5036.30	Yes
14	Electronic check	105.50	2686.05	No
15	Credit card (automatic)	113.25	7895.15	No
16	Mailed check	20.65	1022.95	No
17	Bank transfer (automatic)	106.70	7382.25	No
18	Credit card (automatic)	55.20	528.35	Yes
19	Electronic check	90.05	1862.90	No
20	Electronic check	39.65	39.65	Yes
21	Bank transfer (automatic)	19.80	202.25	No
22	Mailed check	20.15	20.15	Yes

```
23    Credit card (automatic)           59.90       3505.10     No
24    Credit card (automatic)           59.60       2970.30     No
25  Bank transfer (automatic)           55.30       1530.60     No
26            Electronic check          99.35       4749.15    Yes
27            Electronic check          30.20         30.20    Yes
28    Credit card (automatic)           90.25       6369.45     No
29                Mailed check          64.70       1093.10    Yes

[30 rows x 21 columns]
```

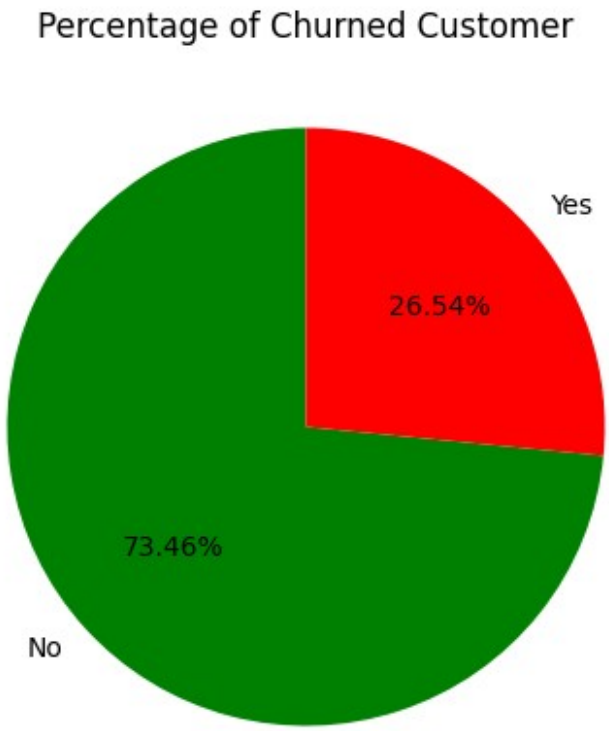
Visualization

Finding Reason why Customer has Churned

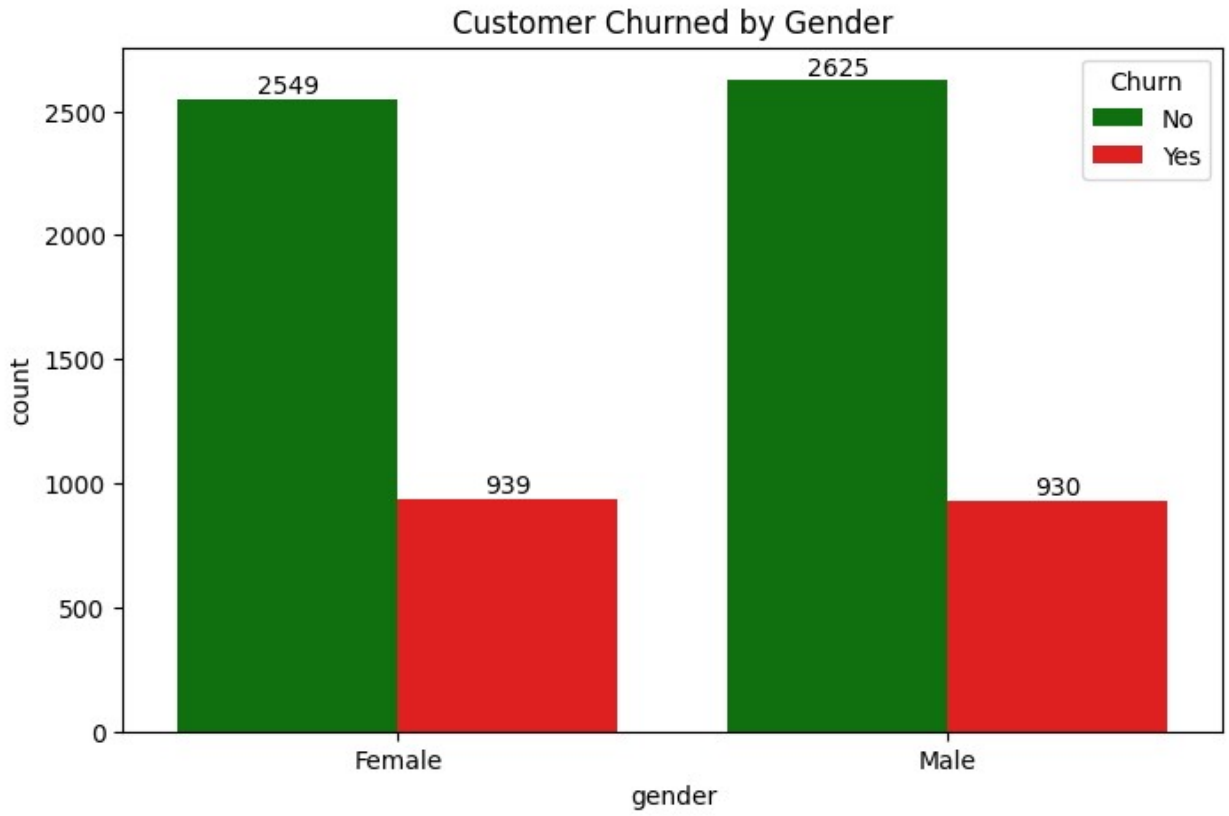
```
plt.figure(figsize=(8,6))
ax=sns.countplot(x='Churn', data=df,
hue='Churn',palette=['g','r'],width=0.4)
ax.bar_label(ax.containers[0])
ax.bar_label(ax.containers[1])
plt.title('Customer Churned : Yes vs No ')
plt.xlabel('Churned Status')
plt.ylabel('Number of customer')
plt.show()
```



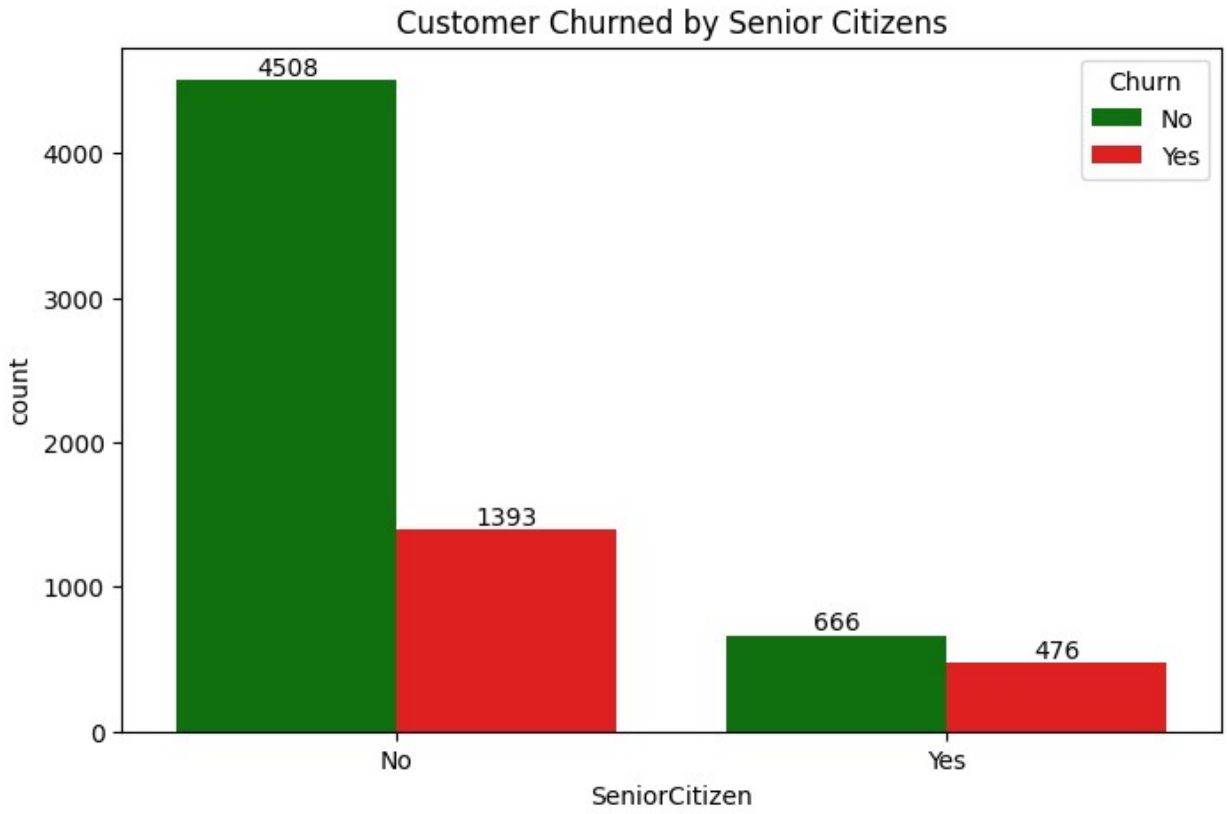

```
plt.figure(figsize=(8,5))
gb=df.groupby('Churn').agg({'Churn':'count'})
plt.pie(gb['Churn'],labels=gb.index,autopct='%1.2f%
%',colors=['g','r'],startangle=90)
plt.title('Percentage of Churned Customer')
plt.show()
```



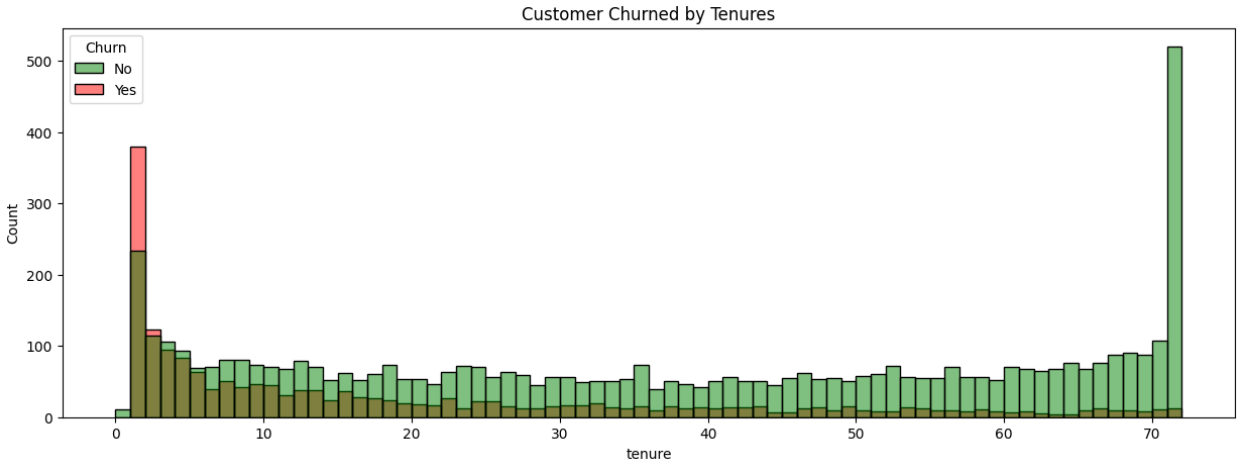

```
## 26.54% of the Customer Churned

plt.figure(figsize=(8,5))
ax=sns.countplot(x='gender', data=df, hue='Churn', palette=['g','r'])
ax.bar_label(ax.containers[0])
ax.bar_label(ax.containers[1])

plt.title('Customer Churned by Gender')
plt.show()
```



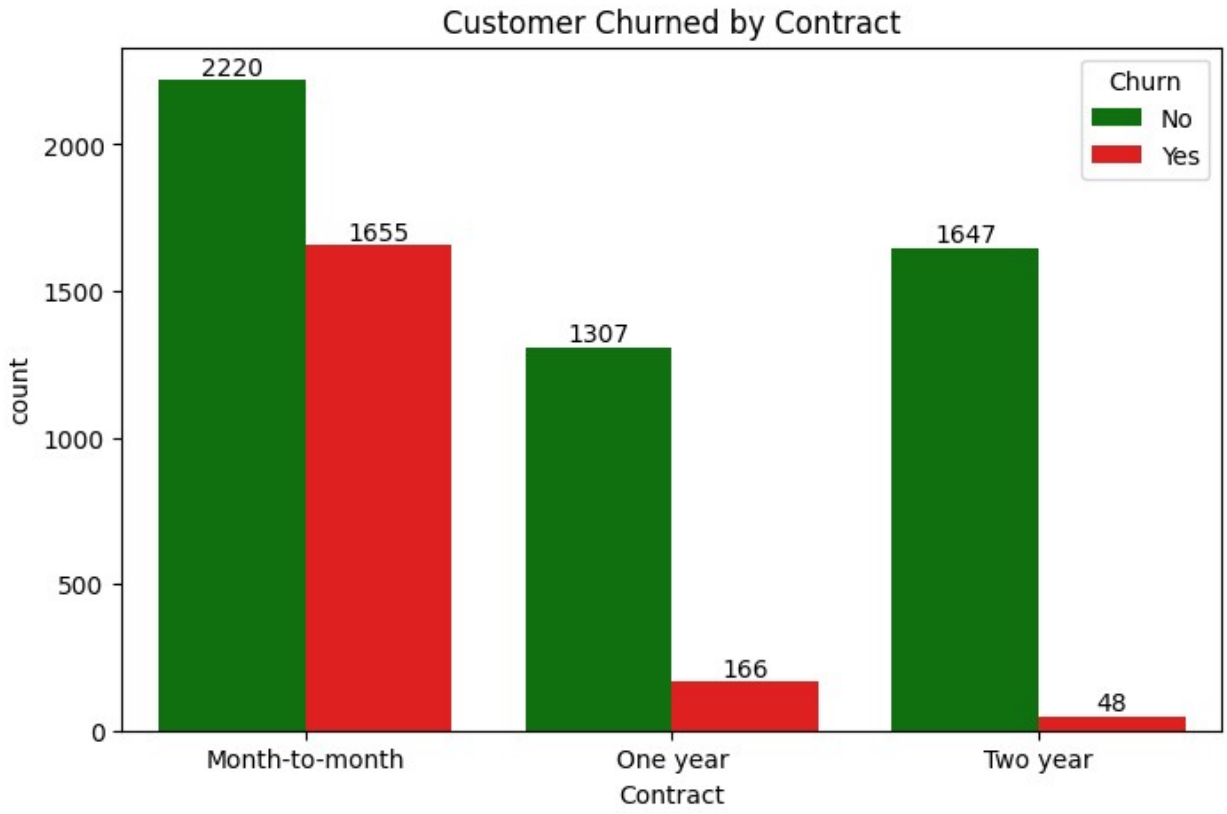

```
plt.figure(figsize=(8,5))
ax=sns.countplot(x='SeniorCitizen',data = df,
hue='Churn',palette=('g','r'))
ax.bar_label(ax.containers[0])
ax.bar_label(ax.containers[1])

plt.title('Customer Churned by Senior Citizens')
plt.show()
```



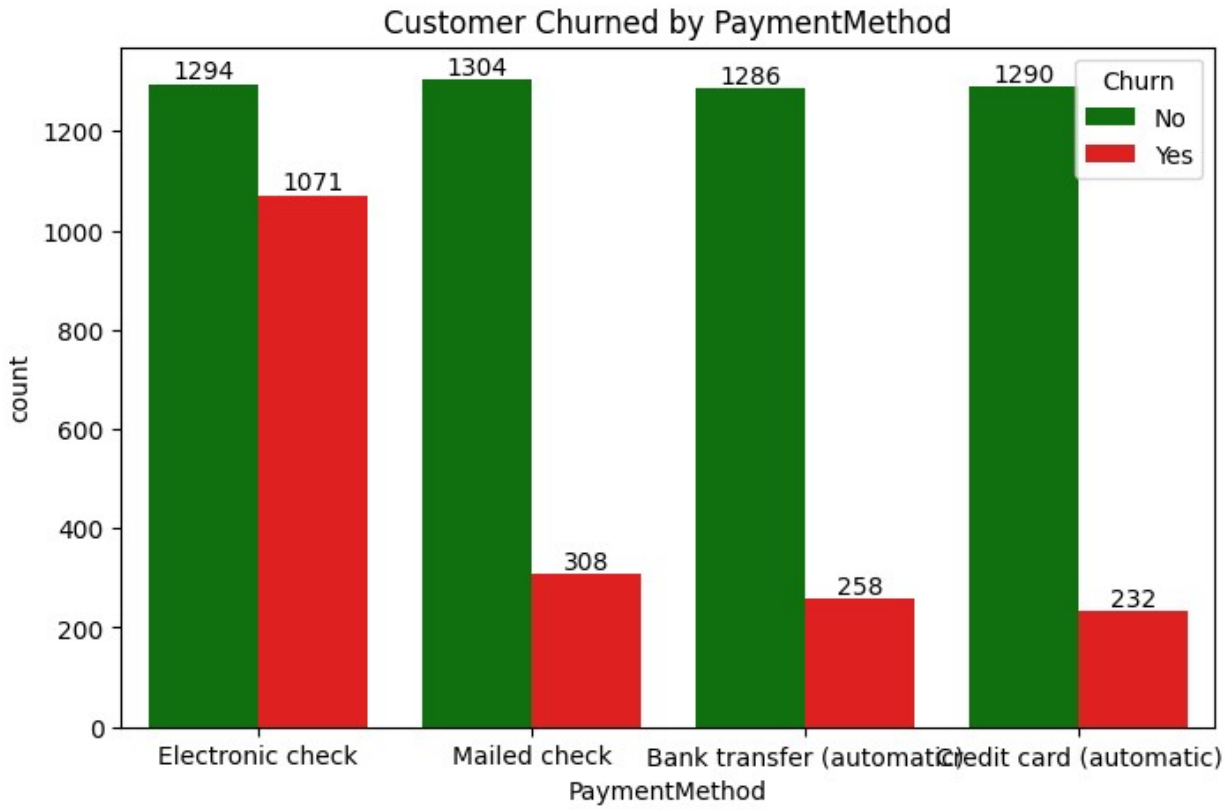

```
## Senior Citizens have higer rate of churned rate as compared to non-Senior Citizens

plt.figure(figsize=(15,5))
ax=sns.histplot(x='tenure', data=df,hue='Churn',
palette=['g','r'],bins=72)


plt.title('Customer Churned by Tenures')
plt.show()
```



```
plt.figure(figsize=(8,5))
ax=sns.countplot(x='Contract', data=df, hue='Churn',
palette=['g','r'])
ax.bar_label(ax.containers[0])
ax.bar_label(ax.containers[1])

plt.title('Customer Churned by Contract')
plt.show()
```

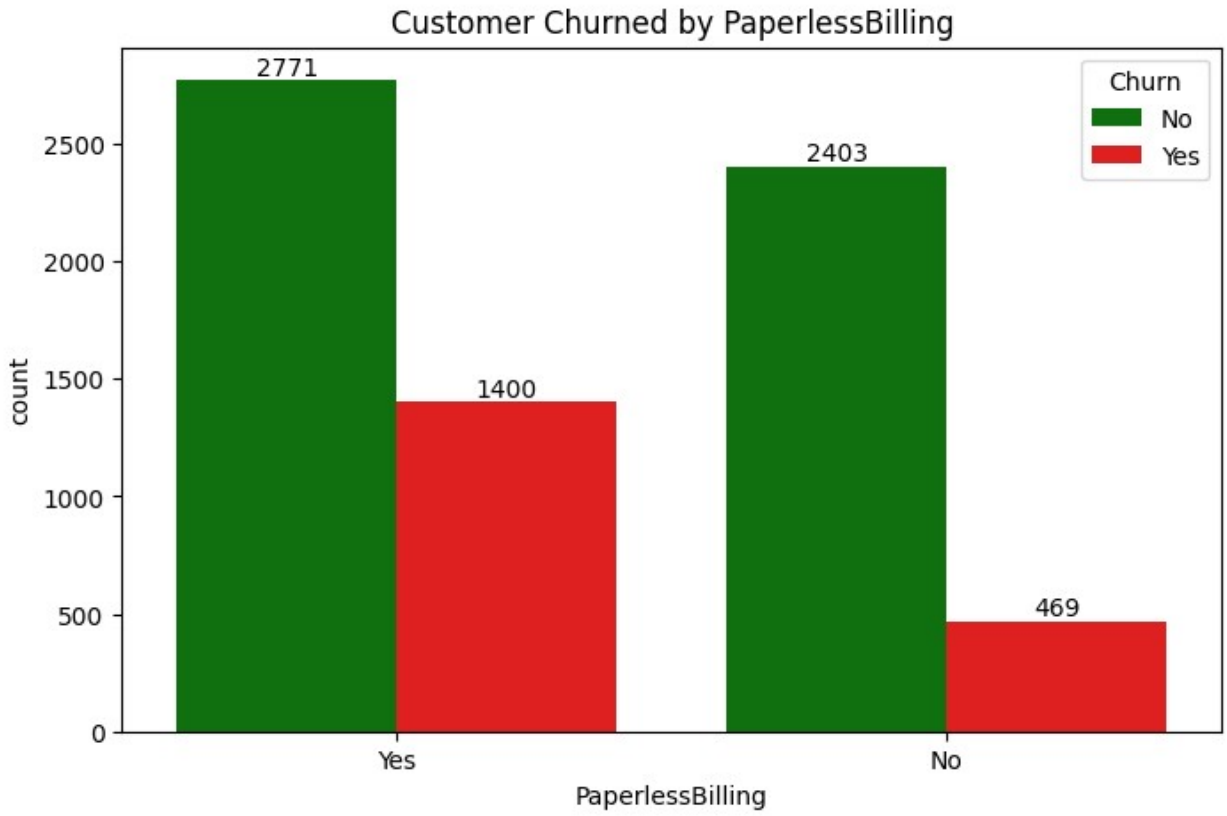



```
## month to month contract have higher rate of churning than year wise
contract

plt.figure(figsize=(8,5))
ax=sns.countplot(x='PaymentMethod', data=df, hue='Churn',
palette=['g','r'])
ax.bar_label(ax.containers[0])
ax.bar_label(ax.containers[1])

plt.title('Customer Churned by PaymentMethod')
plt.show()
```



Electronic check have more churning than automatic payment method

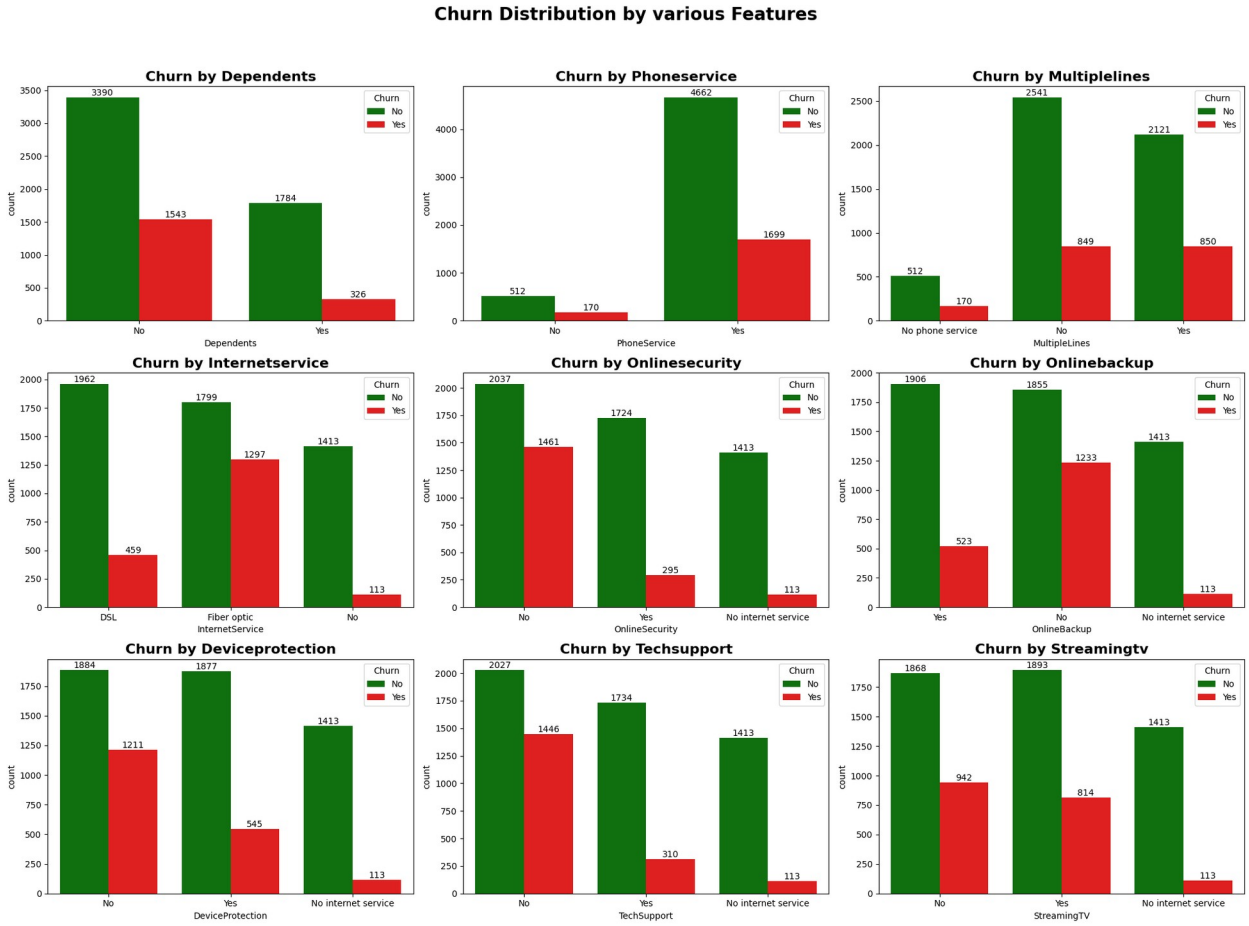
```
plt.figure(figsize=(8,5))
ax=sns.countplot(x='PaperlessBilling', data=df, hue='Churn',
palette=['g','r'])
ax.bar_label(ax.containers[0])
ax.bar_label(ax.containers[1])

plt.title('Customer Churned by PaperlessBilling')
plt.show()
```



Paperless Billing have more churning than paper billing

```
columns = ['Dependents', 'PhoneService', 'MultipleLines',
'InternetService','OnlineSecurity', 'OnlineBackup',
'DeviceProtection','TechSupport', 'StreamingTV']
nrows=3
ncols=3
fig,axes= plt.subplots(nrows, ncols, figsize=(20,15))
fig.suptitle('Churn Distribution by various Features',
fontsize=20,fontweight='bold')
axes=axes.flatten()
for i,column in enumerate(columns):
    bx=sns.countplot(data=df,x=column, ax=axes[i],hue='Churn',
palette=['g','r'])
    bx.bar_label(bx.containers[0])
    bx.bar_label(bx.containers[1])
    axes[i].set_title(f'Churn by {column.capitalize()}',
fontsize=16,fontweight='bold')
    axes[i].set_xlabel(column)
    axes[i].set_ylabel('count')
for j in range(len(columns), nrows*ncols):
    fig.delaxes(axes[j])
```

```
plt.tight_layout(rect=[0,0,1,0.95])
plt.show()
```



customers who lack supplementary services are at higher risk of churning like dependents, multiplelines, online security, backup, device protection & techsupport re more likely to churn.

INSIGHTS

Service Features:

Dependents: Customers without dependents churn more.

Multiple Lines: Higher churn for customers with multiple lines.

Internet Service: Fiber optic users churn more than DSL users.

Online Services: Higher churn without online security, backup, device protection, and tech support.

Streaming: No streaming services lead to higher churn.

Demographics:

Senior Citizens: Higher churn rate among senior citizens.

Gender: Equal churn rates between male and female.

Contract and Payment:

Tenure: High churn in month 1.

Contract Type: Month-to-month contracts have higher churn.

Payment Method: Electronic payments have more churn than other methods.

Billing: Paperless billing shows higher churn.

RECOMMENDATIONS

1. Enhance Service Add-Ons: Focus on online security, backup, tech support, and streaming services.

2. Target Senior Citizens: Develop retention strategies for senior citizens.

3. Improve Onboarding: Reduce churn in the first month through better onboarding.

4. Long-Term Contracts: Encourage longer-term contracts to reduce churn.

5. Optimize Payments: Review electronic payment methods to minimize churn.

6. Incentivize Billing: Offer retention incentives for paperless billing customers.