In [2]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
data=pd.read_csv("C:/Users/hp/Documents/WA_Fn-UseC_-Telco-Customer-Churn (1).csv")
print(data.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]
```

Showing top 5 rows of each column of the dataset

In [18]:
```
data["TotalCharges"]=data["TotalCharges"].replace(" ","0")
data["TotalCharges"]=data["TotalCharges"].astype("float")
data["tenure"]=data["tenure"].astype("float")
print(data.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   float64
 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(3), int64(1), object(17)
memory usage: 1.1+ MB
None
```

In [20]:
```
print(data.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
```

Checking for null values in the dataset

In [31]:
```
print(data["customerID"].duplicated().sum())
```

```
0
```

Checking for duplicated data vased on customer Id

In [47]:
```
data["SeniorCitizen"]=data["SeniorCitizen"].astype("object")
data["SeniorCitizen"]=data["SeniorCitizen"].replace(1,"YES")
data["SeniorCitizen"]=data["SeniorCitizen"].replace(0,"NO")
data["SeniorCitizen"]=data["SeniorCitizen"].replace(1,"YES")
print(data.info())
```

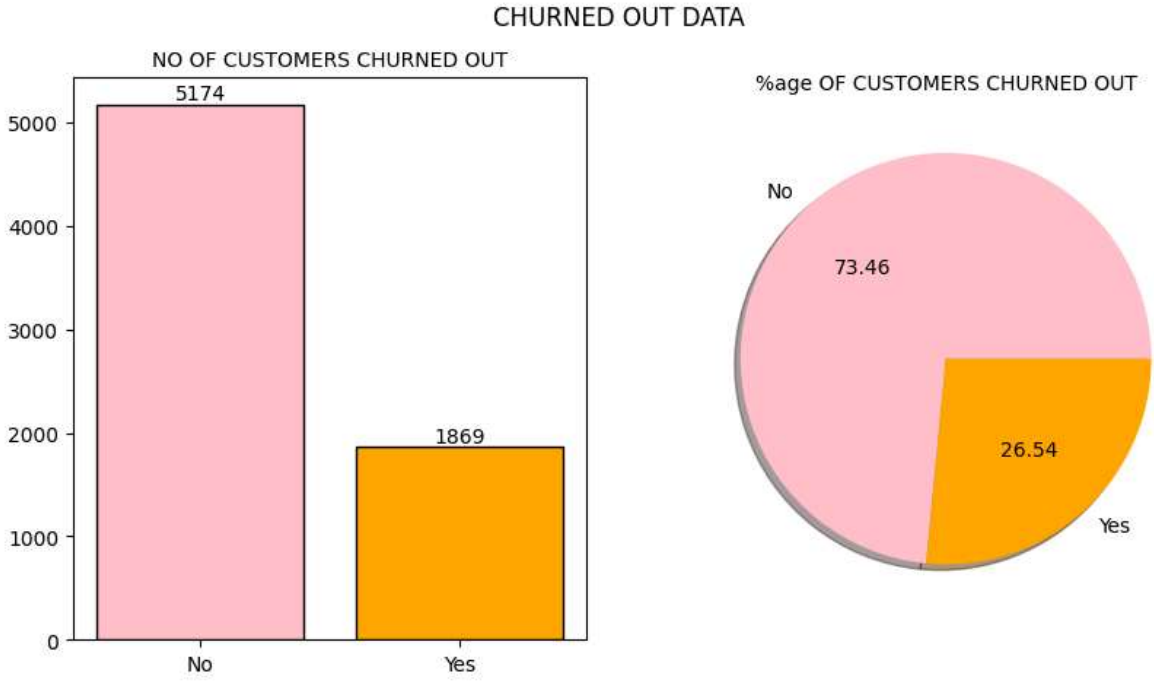
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 21 columns):
 #   Column            Non-Null Count  Dtype  
---  ------            --------------  -----  
 0   customerID        7043 non-null   object 
 1   gender            7043 non-null   object 
 2   SeniorCitizen     7043 non-null   object 
 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(1), object(19)
memory usage: 1.1+ MB
None
```

Converted 0 and 1 to yes/no making data for senior citizen easy to understand

In [104…
```
gp=data.groupby("Churn")["Churn"].count()
print(gp)
plt.figure(figsize=(10,5))
plt.subplot(1,2,1)
gr1=plt.bar(gp.index,gp.values,edgecolor="black",color=['pink','orange'])# bra grap
plt.title("NO OF CUSTOMERS CHURNED OUT",fontsize=10)
for bar in gr1:
    h=bar.get_height()
    plt.text(bar.get_x() + bar.get_width()/2,h,f'{h}',ha="center",va="bottom")
plt.subplot(1,2,2)
plt.pie(gp.values,labels=gp.index,autopct="%1.2f",colors=["pink","orange"],shadow=T
plt.title("%age OF CUSTOMERS CHURNED OUT",fontsize=10)
plt.suptitle("CHURNED OUT DATA")
```

```
Churn
No     5174
Yes    1869
Name: Churn, dtype: int64
```

Out[104… Text(0.5, 0.98, 'CHURNED OUT DATA')

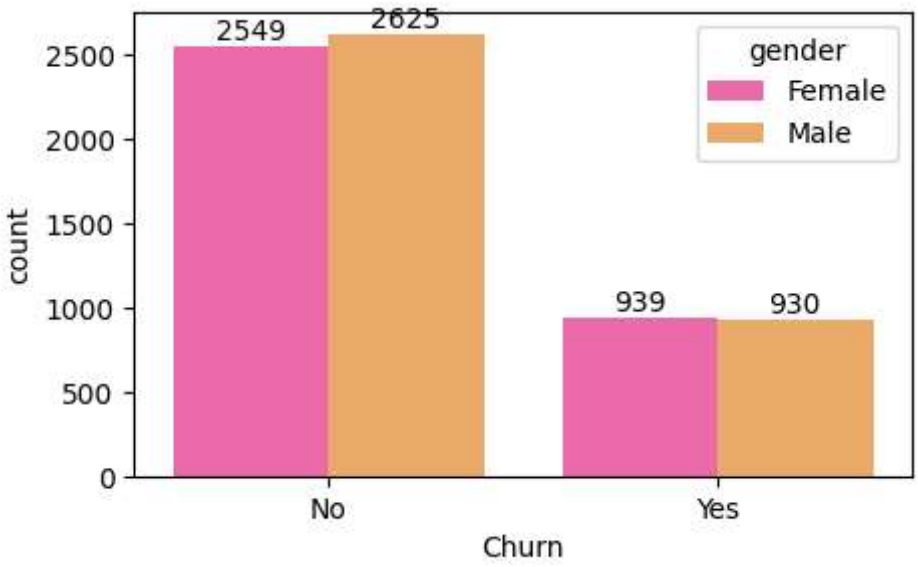


Graphical represntation of how many curstomers have churned out from the piechart we can conclude that 26.54% of customers have churned out

In [104…

```
gp=data.groupby("Churn")["Churn"].count()
print(gp)
group=data.groupby(["Churn","gender"])["Churn"].count()# churn and gender are group
print(group)
plt.figure(figsize=(5,3))
ax1=sns.countplot(data=data,x="Churn",palette="spring",hue="gender")
ax1.bar_label(ax1.containers[0])
ax1.bar_label(ax1.containers[1])
plt.show()
```

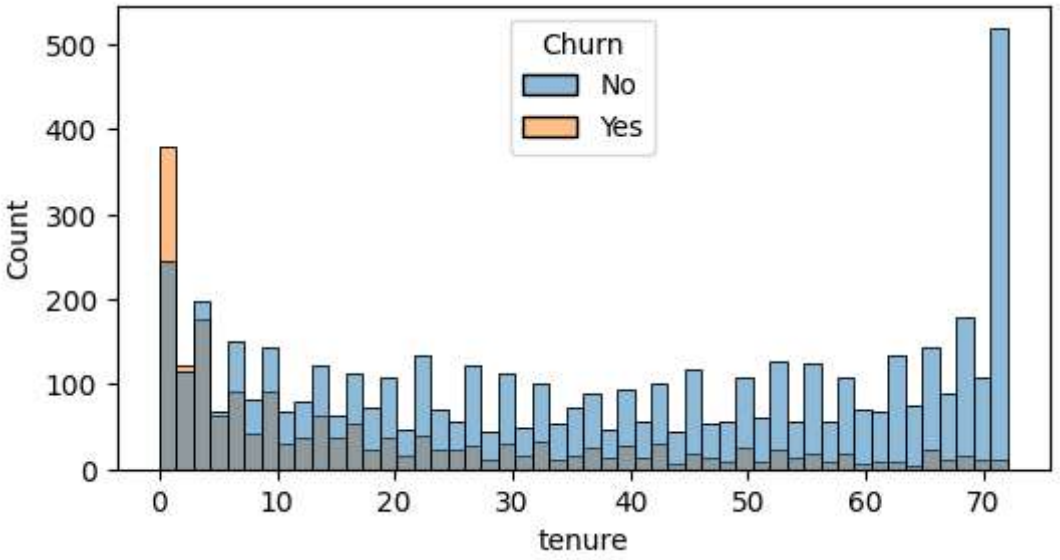
```
Churn
No     5174
Yes    1869
Name: Churn, dtype: int64
Churn  gender
No     Female    2549
       Male      2625
Yes    Female     939
       Male       930
Name: Churn, dtype: int64
```



Approximately same proportion of males and females have churned from the company and the %age of men churning out is less

In [157…
```
plt.figure(figsize=(6,3))
sns.histplot(data=data,x="tenure",hue="Churn",color="pink",bins=50)
```

Out[157…
```
<Axes: xlabel='tenure', ylabel='Count'>
```



From the above plot it is clear that customers that have joined recently(1 or 2 months) # are churning off but we can see that most of them are with us for long time

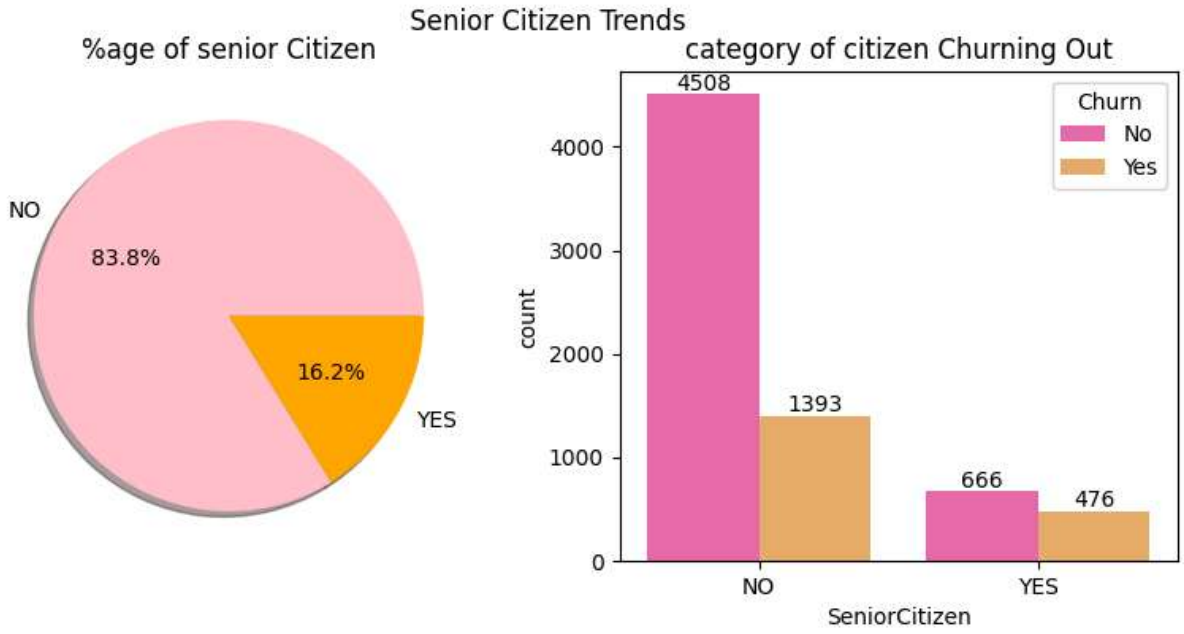
In [48]:
```
print(data.head())
```

```
   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        1889.5    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]
```

In [102…

```
plt.figure(figsize=(10,4))
gp2=data.groupby("SeniorCitizen")["Churn"].count()
print(gp2)
gp3=data.groupby(["SeniorCitizen","Churn"])["Churn"].count()
print(gp3)
plt.subplot(1,2,1)
plt.pie(gp2.values,labels=gp2.index,colors=['pink','orange'],autopct="%1.1f%%",shad
plt.title("%age of senior Citizen")
plt.subplot(1,2,2)
ax2=sns.countplot(data=data,x="SeniorCitizen",hue="Churn",palette="spring")
ax2.bar_label(ax2.containers[0])
ax2.bar_label(ax2.containers[1])
plt.title("category of citizen Churning Out")
plt.suptitle(" Senior Citizen Trends")
plt.show()
```

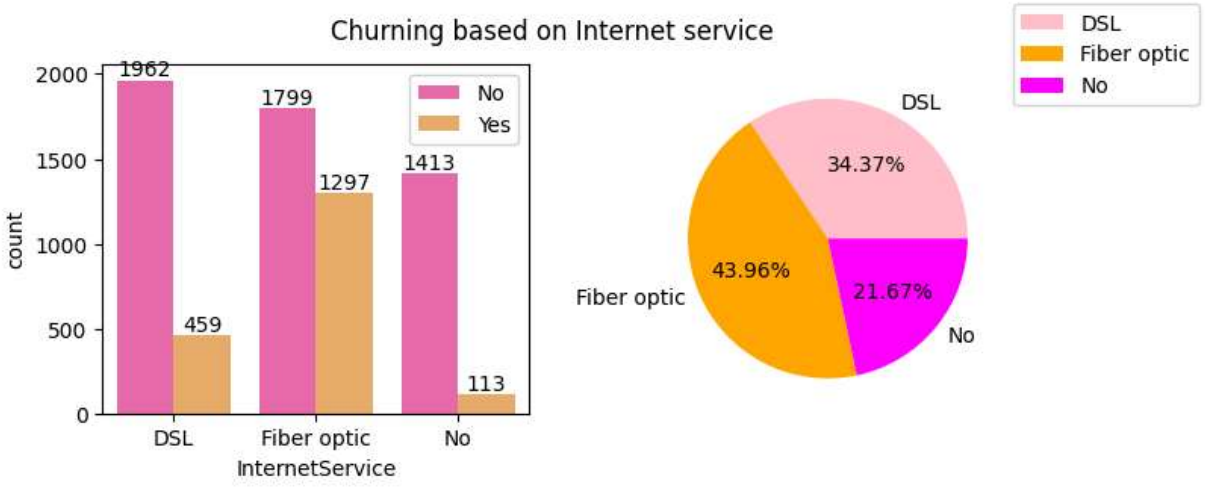
```
SeniorCitizen
NO     5901
YES    1142
Name: Churn, dtype: int64
SeniorCitizen  Churn
NO             No       4508
               Yes      1393
YES            No        666
               Yes       476
Name: Churn, dtype: int64
```



From the above pie chart it is seen that 83.8 % of customers are senior citizen and 16.2% are non # Most of the Non senior citizens are not churning out

In [159…
```
plt.figure(figsize=(8,3))
plt.subplot(1,2,1)
ax3=sns.countplot(data=data,x="InternetService",hue="Churn",palette="spring")
plt.legend()
ax3.bar_label(ax3.containers[0])
ax3.bar_label(ax3.containers[1])
plt.subplot(1,2,2)
gp4=data.groupby("InternetService")["Churn"].count()
plt.pie(gp4.values,labels=gp4.index,colors=["pink","orange","magenta"],autopct="%1.
plt.legend(bbox_to_anchor=(1,1.2))
plt.suptitle("Churning based on Internet service ")
```

Out[159…
```
Text(0.5, 0.98, 'Churning based on Internet service ')
```



The Customers who have been using Fibre Optics are mostly Churning out then DSL

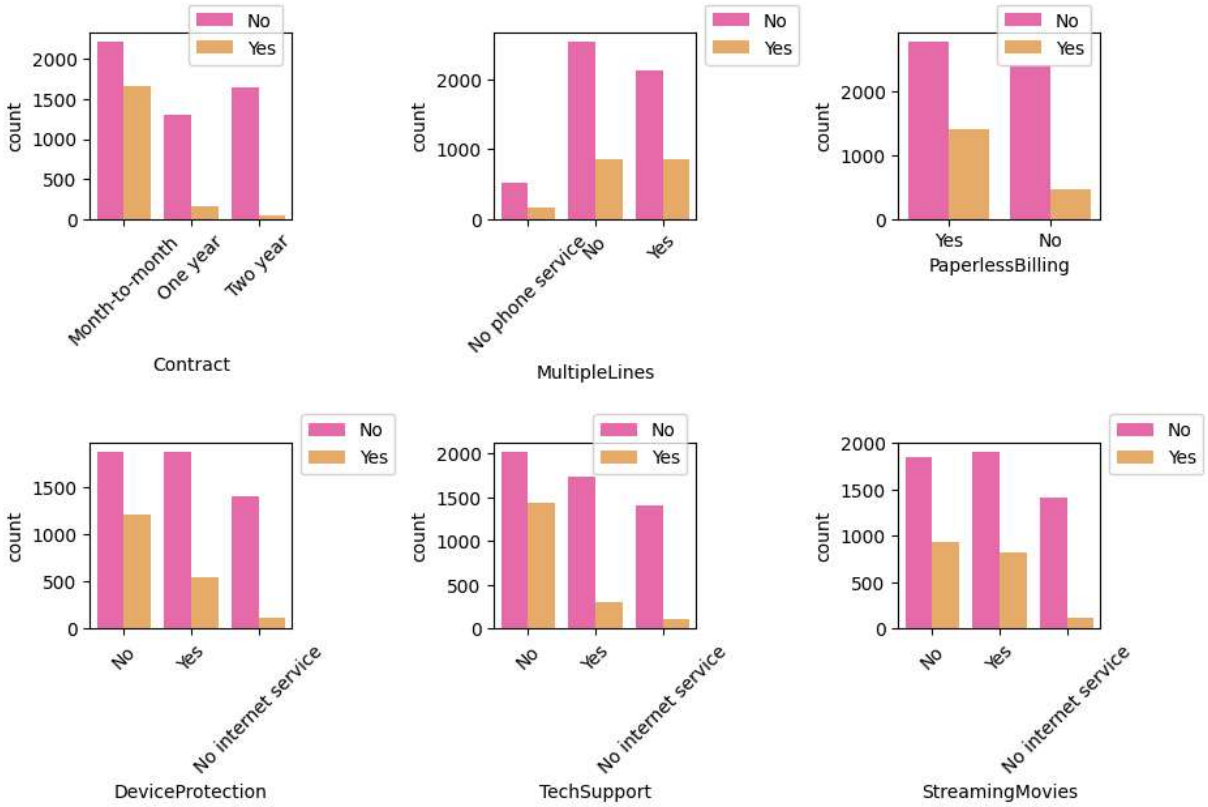
users and lastly those who ar not using any intenet service

In [195…

```
plt.figure(figsize=(10,10))
plt.subplot(3,3,1)
sns.countplot(data=data,x="Contract",hue="Churn",palette="spring")
plt.legend(bbox_to_anchor=(1,1.2))
plt.xticks(rotation=45)
plt.subplot(3,3,2)
plt.subplots_adjust(wspace=1)
plt.subplots_adjust(hspace=1.2)
sns.countplot(data=data,x="MultipleLines",hue="Churn",palette="spring")
plt.legend(bbox_to_anchor=(1,1.2))
plt.xticks(rotation=45)
plt.subplot(3,3,3)
sns.countplot(data=data,x="PaperlessBilling",hue="Churn",palette="spring")
plt.legend(bbox_to_anchor=(1,1.2))
plt.subplot(3,3,4)
sns.countplot(data=data,x="DeviceProtection",hue="Churn",palette="spring")
plt.legend(bbox_to_anchor=(1,1.2))
plt.xticks(rotation=45)
plt.subplot(3,3,5)
sns.countplot(data=data,x="TechSupport",hue="Churn",palette="spring")
plt.legend(bbox_to_anchor=(1,1.2))
plt.xticks(rotation=45)
plt.subplot(3,3,6)
sns.countplot(data=data,x="StreamingMovies",hue="Churn",palette="spring")
plt.legend(bbox_to_anchor=(1,1.2))
plt.xticks(rotation=45)
```

Out[195…

```
([0, 1, 2],
 [Text(0, 0, 'No'), Text(1, 0, 'Yes'), Text(2, 0, 'No internet service')])
```



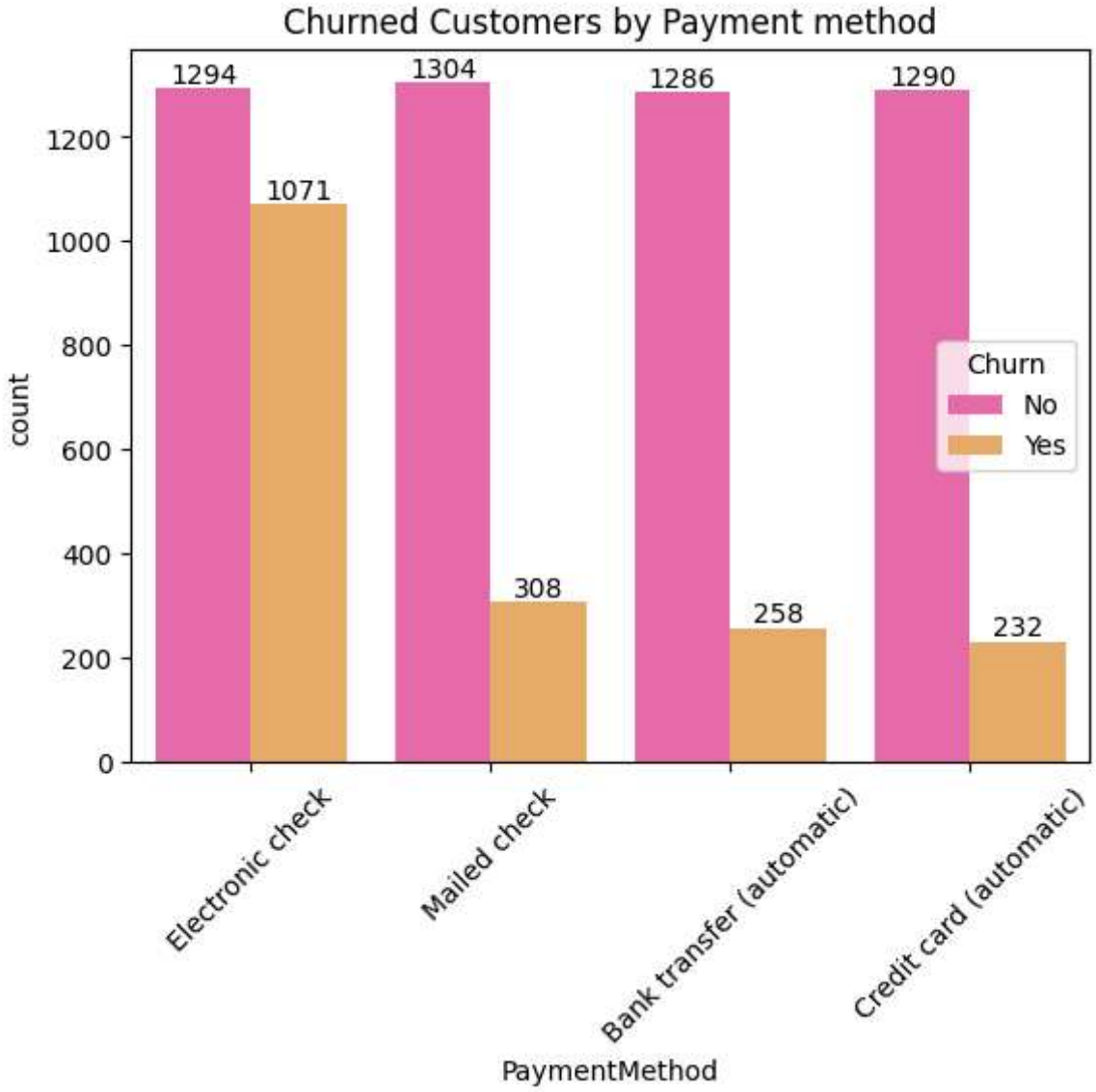
Data related to Churning of customers on the basis of Contract ,Multiple lines,paperlessbilling,device protection,tech support,streaming movies

In [199…

```
ax4=sns.countplot(data=data,x="PaymentMethod",hue="Churn",palette="spring")
ax4.bar_label(ax4.containers[0])
ax4.bar_label(ax4.containers[1])
plt.title("Churned Customers by Payment method")

plt.xticks(rotation=45)
```

Out[199…

```
([0, 1, 2, 3],
 [Text(0, 0, 'Electronic check'),
  Text(1, 0, 'Mailed check'),
  Text(2, 0, 'Bank transfer (automatic)'),
  Text(3, 0, 'Credit card (automatic)')])
```



Here most of the customers who are using Electronic chek for payment are Churned out more