```python
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.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]
```

```
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 blanks with 0 as tenure is 0 and no of total charge are recorded

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

```
0
```

```
# display statistcal summary
```

```
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.duplicated().sum()
```

```
0
```

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

```
0
```

```
df_cleaned = df.drop_duplicates()
```

```
print("Original rows:",len(df))
print("Rows after removing duplicates:", len(df_cleaned))
print("No. of Duplicates remove:", len(df) - len(df_cleaned))
```

```
Original rows: 7043
Rows after removing duplicates: 7043
No. of Duplicates remove: 0
```

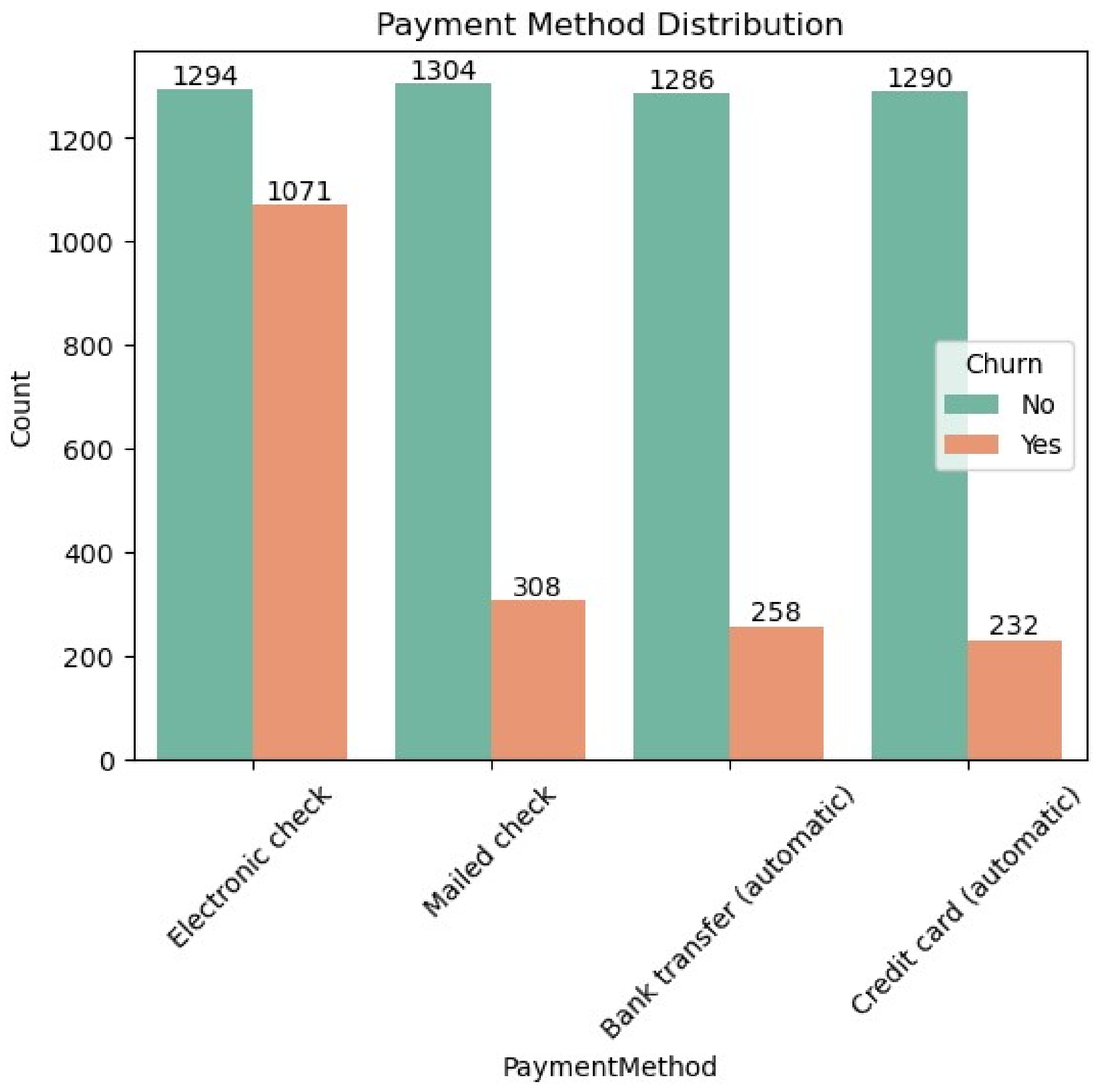
```
# Converted 0 and 1 values of senior citizen to yes/no to make it
easier to understand

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

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

df.head(20)
```

```
    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

       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  ...

       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
```

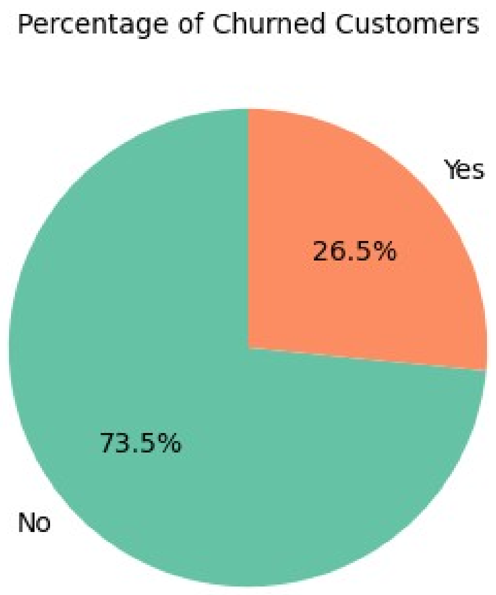
```
12                  Yes                  No                  Yes
13                  Yes                  No                  Yes
14                  Yes                 Yes                  Yes
15                  Yes                 Yes                  Yes
16  No internet service  No internet service  No internet service
17                  Yes                  No                  Yes
18                  Yes                 Yes                   No
19                  Yes                  No                   No

        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

                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 rows x 21 columns]

df['PaymentMethod'].value_counts()

PaymentMethod
Electronic check             2365
Mailed check                 1612
Bank transfer (automatic)    1544
Credit card (automatic)      1522
Name: count, dtype: int64

# checking for missing value

missing_values = df.isnull().sum()
print(missing_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

# Bar Chart

# Create the countplot
ax = sns.countplot(x='PaymentMethod', data=df, hue="Churn", 
palette="Set2")

# Add title, labels, and bar labels
plt.title("Payment Method Distribution")
plt.xlabel("PaymentMethod")
plt.ylabel("Count")

# Add bar labels
for container in ax.containers:
    ax.bar_label(container, fmt="%.0f")

# Rotate x-axis labels
plt.xticks(rotation=45)

# Show the plot
plt.show()
```

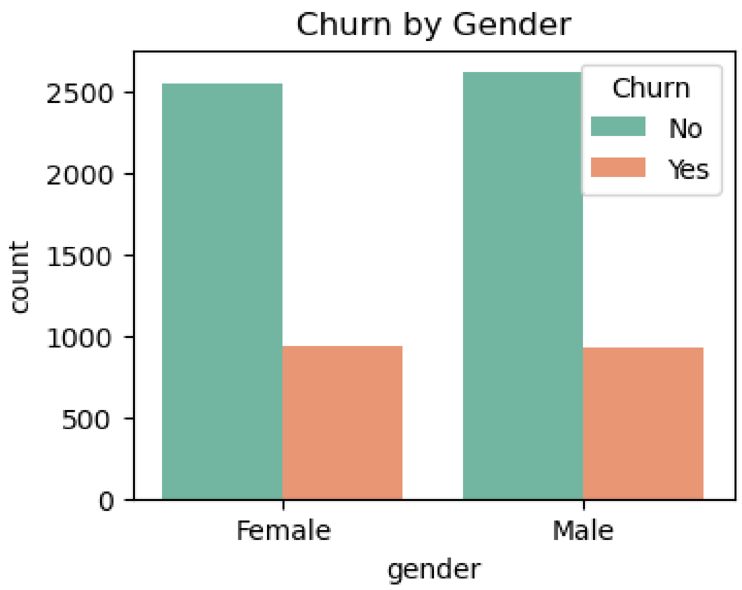


```
# The customers is likely to churn when they is using electrcity check
as a payemnt
```

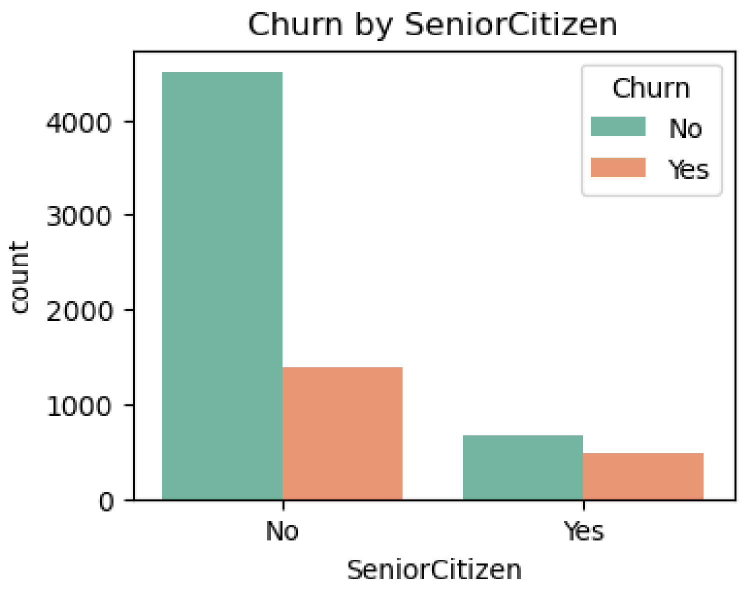
```
gb = df.groupby("Churn").agg({'Churn': "count"})
plt.figure(figsize=(5, 4))
colors = sns.color_palette("Set2")
plt.pie(gb['Churn'], labels=gb.index, autopct="%1.1f%%", startangle=90
,colors=colors)
plt.title("Percentage of Churned Customers", fontsize=10)

plt.show()
```



```
#from the given pie chart we can conclude that 26 54% of our customers
have churned out
#not let's explore the reas behind it


plt.figure(figsize =(4,3))
sns.countplot(x="gender", data= df, hue="Churn", palette="Set2")
plt.title("Churn by Gender")
plt.show()
```

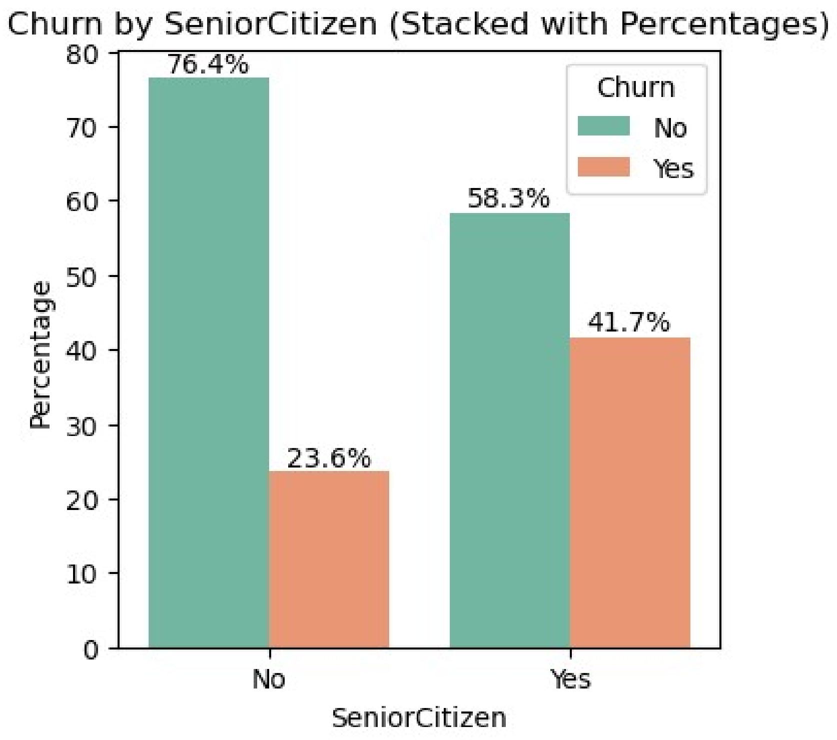


```
plt.figure(figsize =(4,3))
sns.countplot(x="SeniorCitizen", data= df, hue="Churn",
palette="Set2")
plt.title("Churn by SeniorCitizen")
plt.show()
```



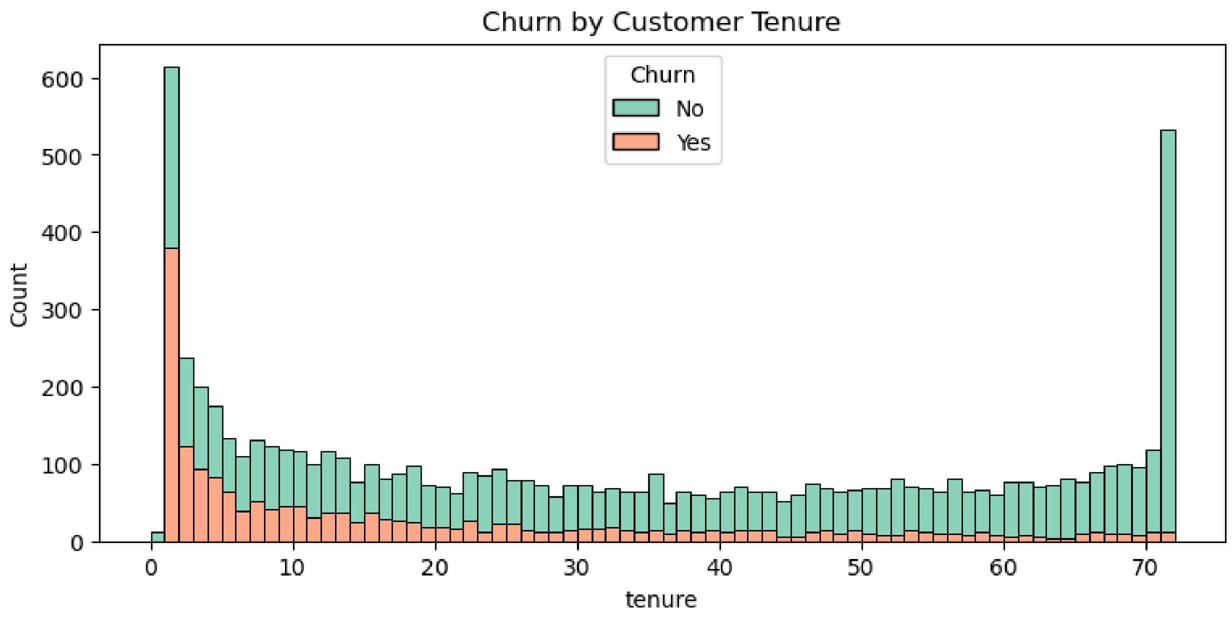
```
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd

# Group the data and calculate the count of SeniorCitizen and Churn
```

```
df_grouped = df.groupby(['SeniorCitizen',
'Churn']).size().reset_index(name='Count')

# Calculate the percentage for each SeniorCitizen group
df_grouped['Percentage'] = df_grouped.groupby('SeniorCitizen')
['Count'].transform(lambda x: 100 * x / x.sum())

# Create a figure
plt.figure(figsize=(4, 4))

# Plot a stacked bar chart with percentages
ax = sns.barplot(x='SeniorCitizen', y='Percentage', hue='Churn',
data=df_grouped, palette="Set2")

# Add labels to the bars
for container in ax.containers:
    ax.bar_label(container, fmt="%.1f%%")

# Add title and show the plot
plt.title("Churn by SeniorCitizen (Stacked with Percentages)")
plt.ylabel("Percentage")
plt.show()
```

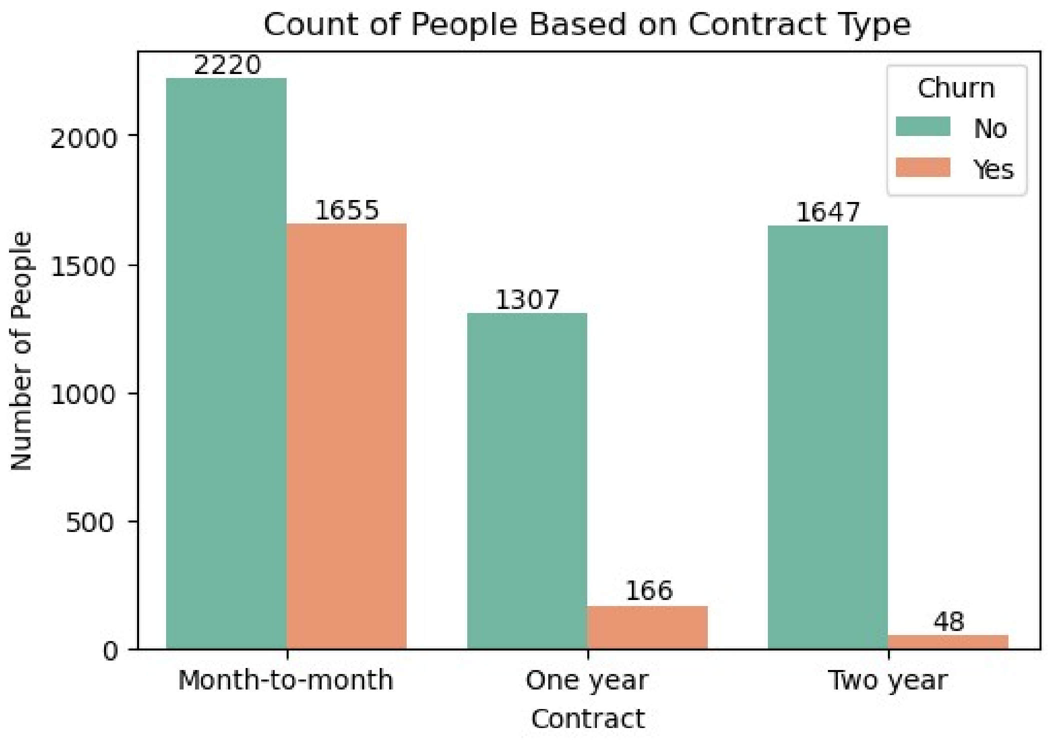


```
# comparative a greater percentage of people in senior citizen
category have chured
```

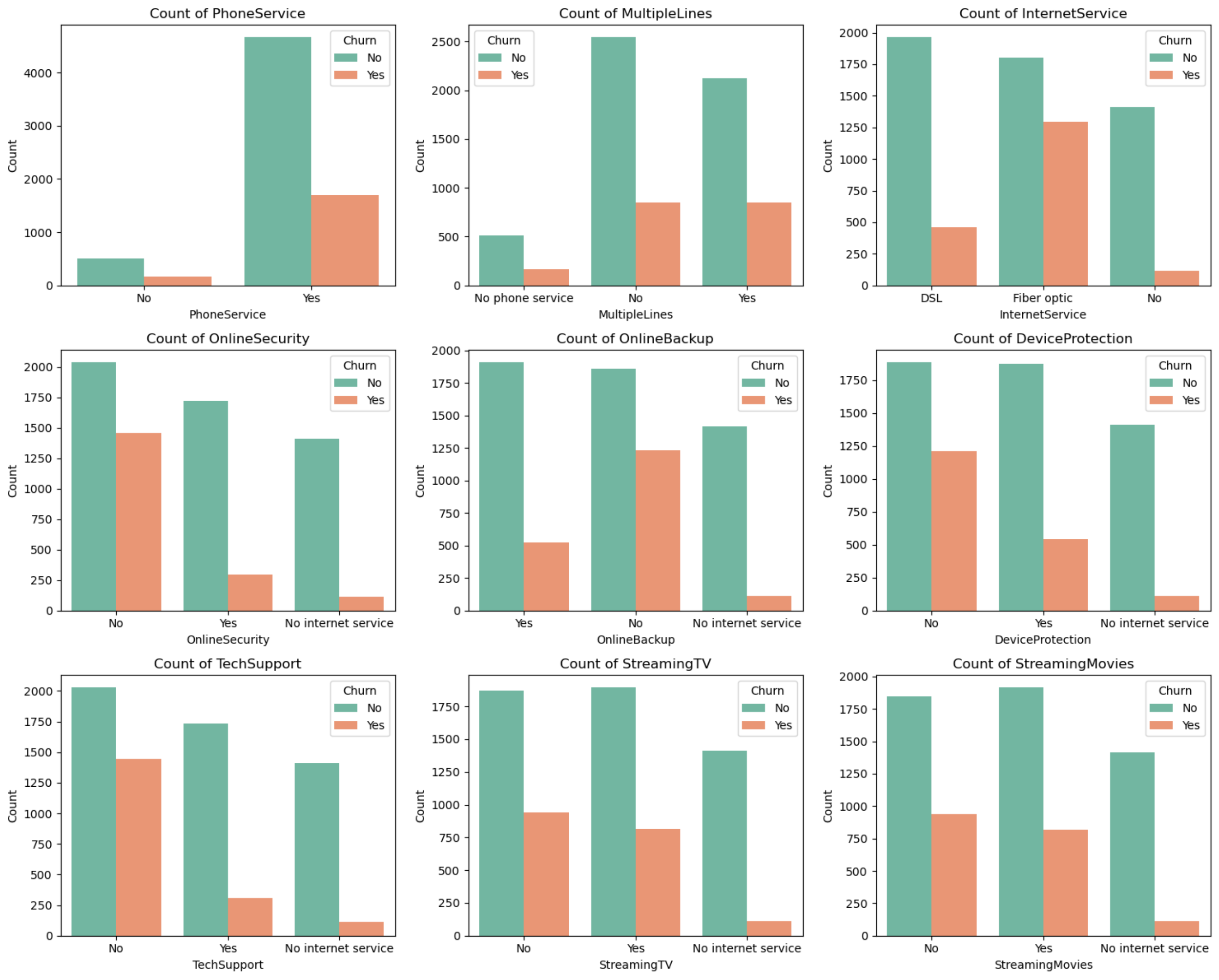
```python
plt.figure(figsize=(9,4))
sns.histplot(data=df, x='tenure', hue='Churn', multiple='stack',
palette="Set2", bins=72)
plt.title("Churn by Customer Tenure")
plt.show()
```



```python
# People who have used our services for a long time have stayed and
people who haved used our services 1 or 2 month have churned

# Group by 'Contract' and 'Churn' and get the count
df_grouped = df.groupby(['Contract',
'Churn']).size().reset_index(name='Count')

# Create the bar plot
plt.figure(figsize=(6, 4))
ax = sns.barplot(x='Contract', y='Count', data=df_grouped,
hue='Churn', palette="Set2")

# Add labels to the bars
for container in ax.containers:
    ax.bar_label(container, fmt="%.0f")

# Set the title and show the plot
plt.title("Count of People Based on Contract Type")
plt.ylabel("Number of People")
plt.show()
```

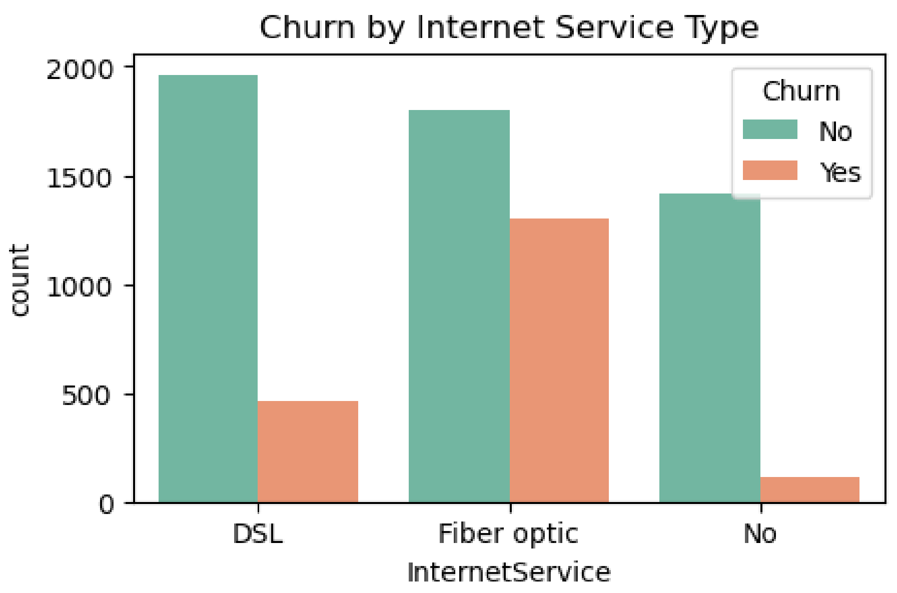


```
# people who have Month to Month contract are likely to churn then for
those who have 1 or 2 year of contract

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)

# List of the columns you want to create count plots for
columns = ['PhoneService', 'MultipleLines', 'InternetService',
           'OnlineSecurity', 'OnlineBackup', 'DeviceProtection',
           'TechSupport', 'StreamingTV', 'StreamingMovies']

# Create subplots
fig, axes = plt.subplots(3, 3, figsize=(15, 12))  # Adjust the size
and number of rows/columns as needed
axes = axes.flatten()  # Flatten the 2D array of axes to 1D for easier
indexing

# Loop through the columns and create count plots
for i, column in enumerate(columns):
    sns.countplot(x=column, data=df, hue='Churn', palette="Set2",
ax=axes[i])
    axes[i].set_title(f'Count of {column}')
    axes[i].set_xlabel(column)
    axes[i].set_ylabel('Count')

# Adjust layout to avoid overlapping
plt.tight_layout()
plt.show()
```

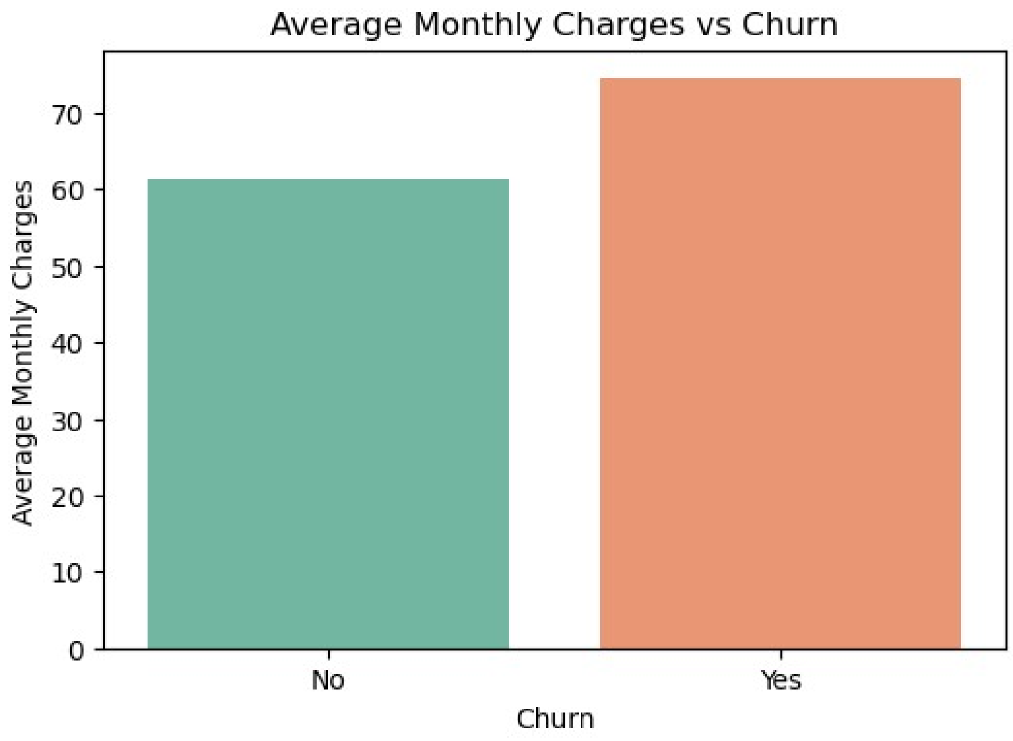


```
# The image shows multiple bar charts comparing customer churn across various services such as
# PhoneService, MultipleLines, InternetService, OnlineSecurity, OnlineBackup, DeviceProtection, TechSupport, StreamingTV, and StreamingMovies.
# It visualizes the counts of customers who churned versus those who didn’t for each service type,
# categorized by whether customers used the service or not.
```

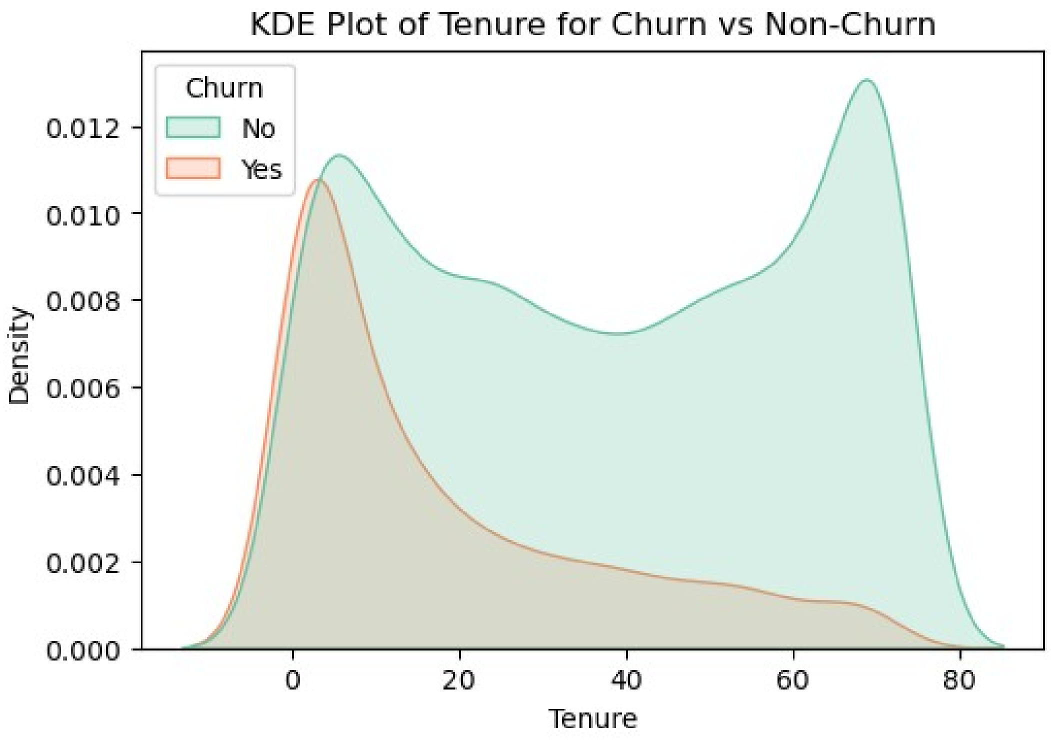
```
plt.figure(figsize=(5,3))
sns.countplot(x='InternetService', data=df, hue='Churn',
palette="Set2")
plt.title("Churn by Internet Service Type")
plt.show()
```



```
# Calculate the average monthly charges for each churn category
avg_monthly_charges = df.groupby('Churn')['MonthlyCharges'].mean()

# Create a bar plot
plt.figure(figsize=(6, 4))
sns.barplot(x=avg_monthly_charges.index, y=avg_monthly_charges.values,
palette="Set2")
plt.title('Average Monthly Charges vs Churn')
plt.xlabel('Churn')
plt.ylabel('Average Monthly Charges')
plt.show()

C:\Users\Administrator\AppData\Local\Temp\
ipykernel_13864\1257669959.py:6: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be
removed in v0.14.0. Assign the `x` variable to `hue` and set
`legend=False` for the same effect.

  sns.barplot(x=avg_monthly_charges.index,
y=avg_monthly_charges.values, palette="Set2")
```



```python
# Customers who churned have higher average monthly charges compared
to those who did not churn.

plt.figure(figsize=(6, 4))
sns.kdeplot(data=df, x='tenure', hue='Churn', fill=True,
palette="Set2")
plt.title('KDE Plot of Tenure for Churn vs Non-Churn')
plt.xlabel('Tenure')
plt.ylabel('Density')
plt.show()
```


KDE Plot of Tenure for Churn vs Non-Churn
Churn
No
Yes
Density
0.000
0.002
0.004
0.006
0.008
0.010
0.012
0
20
40
60
80
Tenure