```
import pandas as pd

# Load dataset
df = pd.read_csv('WA_Fn-UseC_-Telco-Customer-Churn.csv')

# Tampilkan 5 baris pertama
df.head()
```

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...
0	7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	...
1	5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	...
2	3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	...
3	7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes	...
4	9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	...

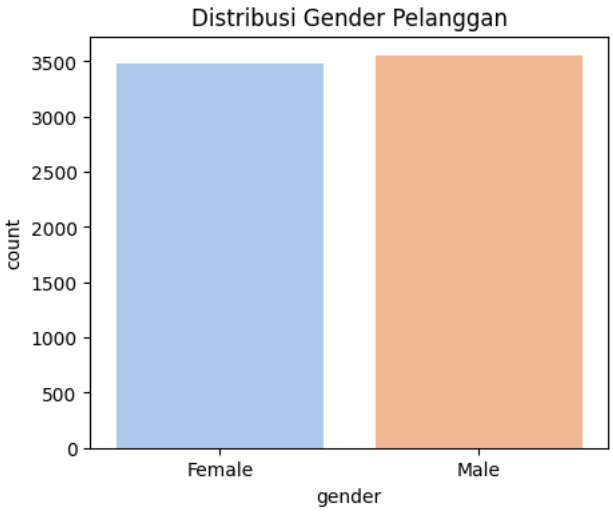
5 rows × 21 columns

```
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
```

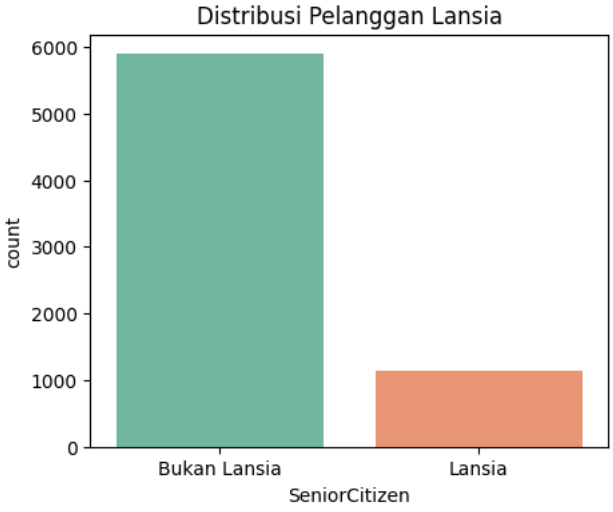
```
import seaborn as sns
import matplotlib.pyplot as plt

# Angka jumlah pelanggan per gender
print("Jumlah pelanggan berdasarkan gender:")
print(df['gender'].value_counts())

# Visualisasi
plt.figure(figsize=(5, 4))
sns.countplot(data=df, x='gender', palette='pastel')
plt.title('Distribusi Gender Pelanggan')
plt.show()
```

```
Jumlah pelanggan berdasarkan gender:
gender
Male      3549
Female    3483
Name: count, dtype: int64
C:\Users\lenovo ideapad 330\AppData\Local\Temp\ipykernel_2144\1831591583.py:10: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `le

  sns.countplot(data=df, x='gender', palette='pastel')
```



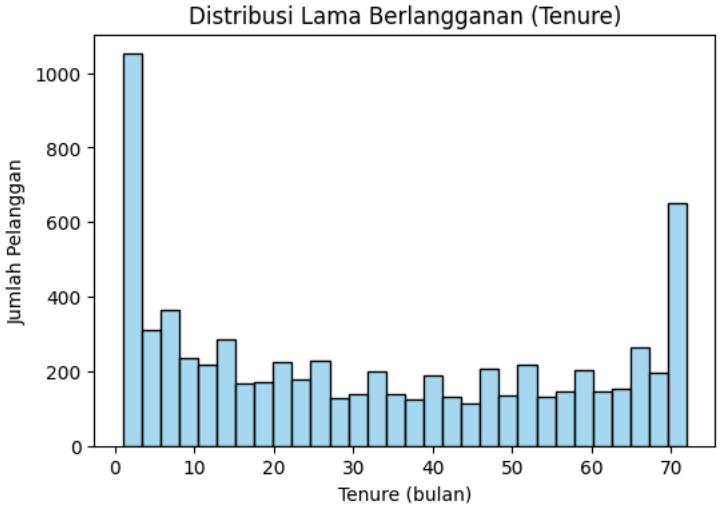

```
# Angka pelanggan lansia vs bukan lansia
print("Jumlah pelanggan lansia vs bukan:")
print(df['SeniorCitizen'].value_counts())
print("Persentase:")
print(df['SeniorCitizen'].value_counts(normalize=True) * 100)

# Visualisasi
plt.figure(figsize=(5, 4))
sns.countplot(data=df, x='SeniorCitizen', palette='Set2')
plt.title('Distribusi Pelanggan Lansia')
plt.xticks([0, 1], ['Bukan Lansia', 'Lansia'])
plt.show()
```

```
Jumlah pelanggan lansia vs bukan:
SeniorCitizen
0    5890
1    1142
Name: count, dtype: int64
Persentase:
SeniorCitizen
0    83.759954
1    16.240046
Name: proportion, dtype: float64
C:\Users\lenovo ideapad 330\AppData\Local\Temp\ipykernel_2144\436423924.py:9: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `le

  sns.countplot(data=df, x='SeniorCitizen', palette='Set2')
```



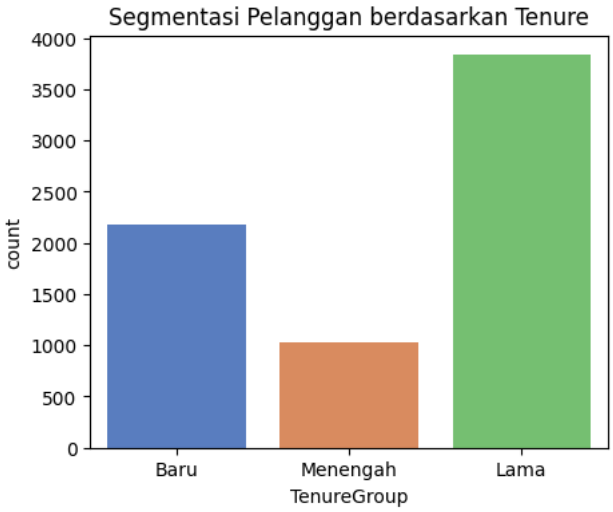
```
# Statistik deskriptif tenure
print("Statistik Tenure (lama berlangganan dalam bulan):")
print(df['tenure'].describe())

# Visualisasi
plt.figure(figsize=(6, 4))
sns.histplot(data=df, x='tenure', bins=30, color='skyblue')
plt.title('Distribusi Lama Berlangganan (Tenure)')
plt.xlabel('Tenure (bulan)')
plt.ylabel('Jumlah Pelanggan')
plt.show()
```

```
Statistik Tenure (lama berlangganan dalam bulan):
count    7032.000000
mean       32.421786
std        24.545260
min         1.000000
25%         9.000000
50%        29.000000
75%        55.000000
max        72.000000
Name: tenure, dtype: float64
```



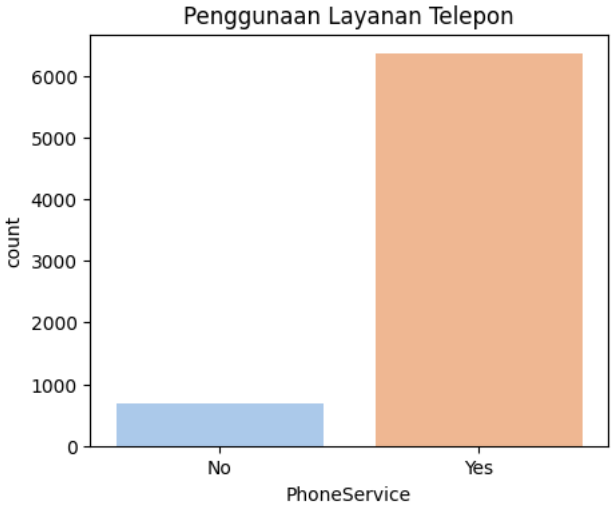

```
# Buat kolom TenureGroup
def segment_tenure(t):
    if t <= 12:
        return 'Baru'
    elif t <= 24:
        return 'Menengah'
    else:
        return 'Lama'

df['TenureGroup'] = df['tenure'].apply(segment_tenure)

# Lihat jumlah dan persentase
print("Jumlah pelanggan per segmentasi Tenure:")
print(df['TenureGroup'].value_counts())
print("Persentase:")
print(df['TenureGroup'].value_counts(normalize=True) * 100)

# Visualisasi
plt.figure(figsize=(5, 4))
sns.countplot(data=df, x='TenureGroup', order=['Baru', 'Menengah', 'Lama'], palette='muted')
plt.title('Segmentasi Pelanggan berdasarkan Tenure')
plt.show()
```

```
Jumlah pelanggan per segmentasi Tenure:
TenureGroup
Lama        3833
Baru        2175
Menengah    1024
Name: count, dtype: int64
Persentase:
TenureGroup
Lama        54.507964
Baru        30.930034
Menengah    14.562002
Name: proportion, dtype: float64
C:\Users\lenovo ideapad 330\AppData\Local\Temp\ipykernel_2144\4184869882.py:20: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `le

  sns.countplot(data=df, x='TenureGroup', order=['Baru', 'Menengah', 'Lama'], palette='muted')
```



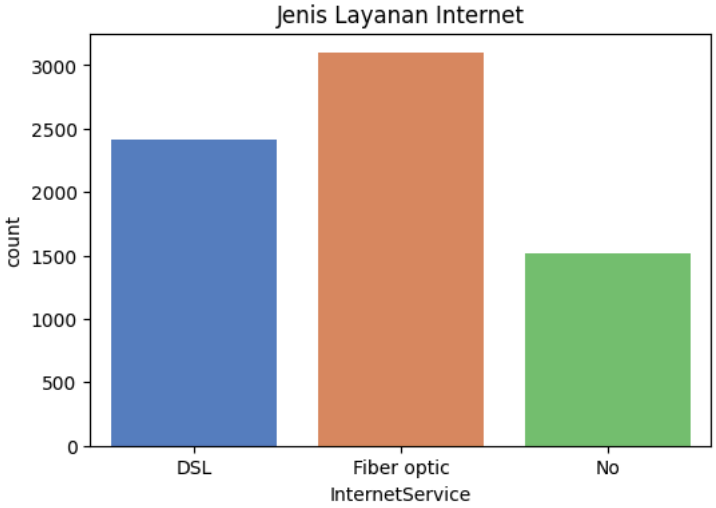

```
#2: Analisis Produk & Layanan yang Digunakan
print("Distribusi PhoneService:")
print(df['PhoneService'].value_counts())
print("Persentase:")
print(df['PhoneService'].value_counts(normalize=True) * 100)

plt.figure(figsize=(5, 4))
sns.countplot(data=df, x='PhoneService', palette='pastel')
plt.title('Penggunaan Layanan Telepon')
plt.show()
```

```
Distribusi PhoneService:
PhoneService
Yes    6352
No      680
Name: count, dtype: int64
Persentase:
PhoneService
Yes    90.32992
No      9.67008
Name: proportion, dtype: float64
C:\Users\lenovo ideapad 330\AppData\Local\Temp\ipykernel_2144\1595151209.py:8: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `le

  sns.countplot(data=df, x='PhoneService', palette='pastel')
```



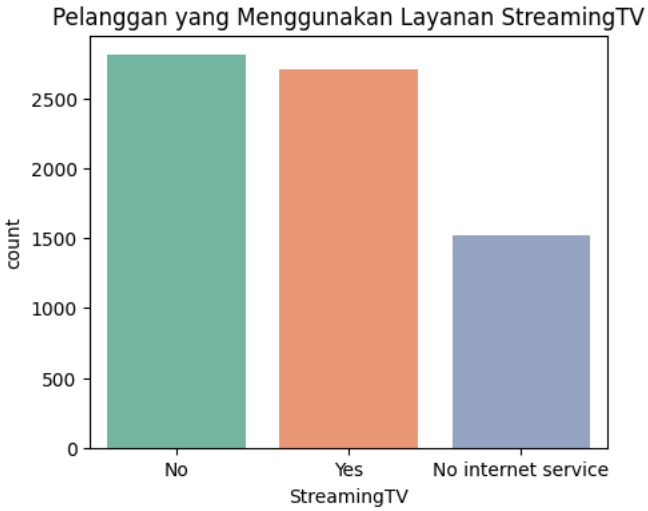

```
print("Distribusi InternetService:")
print(df['InternetService'].value_counts())
print("Persentase:")
print(df['InternetService'].value_counts(normalize=True) * 100)

plt.figure(figsize=(6, 4))
sns.countplot(data=df, x='InternetService', palette='muted')
plt.title('Jenis Layanan Internet')
plt.show()
```

```
Distribusi InternetService:
InternetService
Fiber optic    3096
DSL            2416
No             1520
Name: count, dtype: int64
Persentase:
InternetService
Fiber optic    44.027304
DSL            34.357224
No             21.615472
Name: proportion, dtype: float64
C:\Users\lenovo ideapad 330\AppData\Local\Temp\ipykernel_2144\4204503194.py:7: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `le

  sns.countplot(data=df, x='InternetService', palette='muted')
```



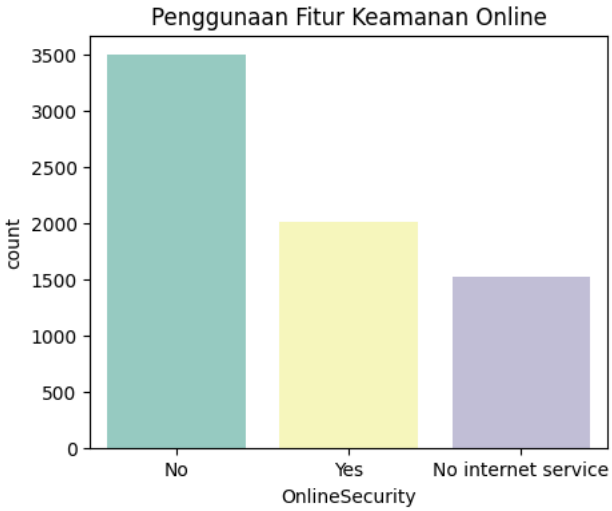

```
print("Distribusi StreamingTV:")
print(df['StreamingTV'].value_counts())
print("Persentase:")
print(df['StreamingTV'].value_counts(normalize=True) * 100)

plt.figure(figsize=(5, 4))
sns.countplot(data=df, x='StreamingTV', palette='Set2')
plt.title('Pelanggan yang Menggunakan Layanan StreamingTV')
plt.show()
```

```
Distribusi StreamingTV:
StreamingTV
No                     2809
Yes                    2703
No internet service    1520
Name: count, dtype: int64
Persentase:
StreamingTV
No                     39.945961
Yes                    38.438567
No internet service    21.615472
Name: proportion, dtype: float64
C:\Users\lenovo ideapad 330\AppData\Local\Temp\ipykernel_2144\4073750741.py:7: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `le

  sns.countplot(data=df, x='StreamingTV', palette='Set2')
```



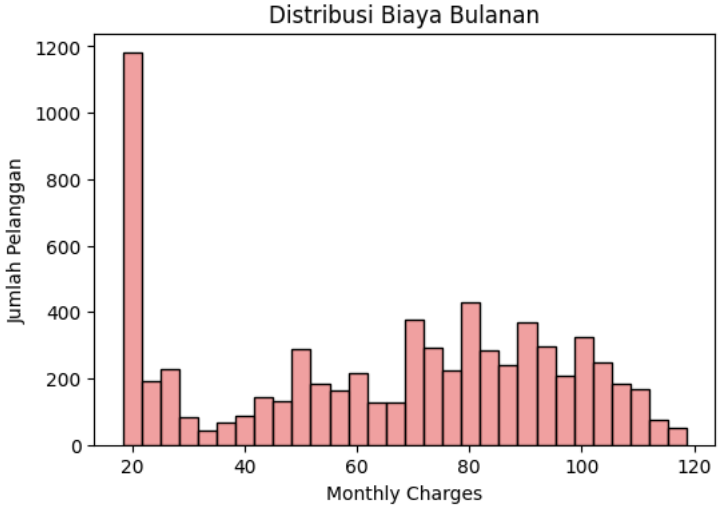

```
print("Distribusi OnlineSecurity:")
print(df['OnlineSecurity'].value_counts())
print("Persentase:")
print(df['OnlineSecurity'].value_counts(normalize=True) * 100)

plt.figure(figsize=(5, 4))
sns.countplot(data=df, x='OnlineSecurity', palette='Set3')
plt.title('Penggunaan Fitur Keamanan Online')
plt.show()
```

```
Distribusi OnlineSecurity:
OnlineSecurity
No                     3497
Yes                    2015
No internet service    1520
Name: count, dtype: int64
Persentase:
OnlineSecurity
No                     49.729807
Yes                    28.654721
No internet service    21.615472
Name: proportion, dtype: float64
C:\Users\lenovo ideapad 330\AppData\Local\Temp\ipykernel_2144\3714144382.py:7: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `le

  sns.countplot(data=df, x='OnlineSecurity', palette='Set3')
```



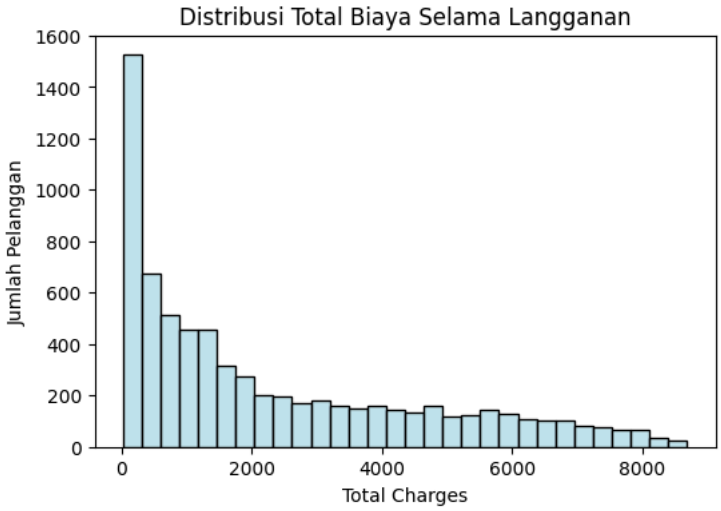

```
#3 Analisis Perilaku Pembayaran (MonthlyCharges & TotalCharges)
print("Statistik MonthlyCharges:")
print(df['MonthlyCharges'].describe())

plt.figure(figsize=(6, 4))
sns.histplot(data=df, x='MonthlyCharges', bins=30, color='lightcoral')
plt.title('Distribusi Biaya Bulanan')
plt.xlabel('Monthly Charges')
plt.ylabel('Jumlah Pelanggan')
plt.show()
```

```
Statistik MonthlyCharges:
count    7032.000000
mean       64.798208
std        30.085974
min        18.250000
25%        35.587500
50%        70.350000
75%        89.862500
max       118.750000
Name: MonthlyCharges, dtype: float64
```



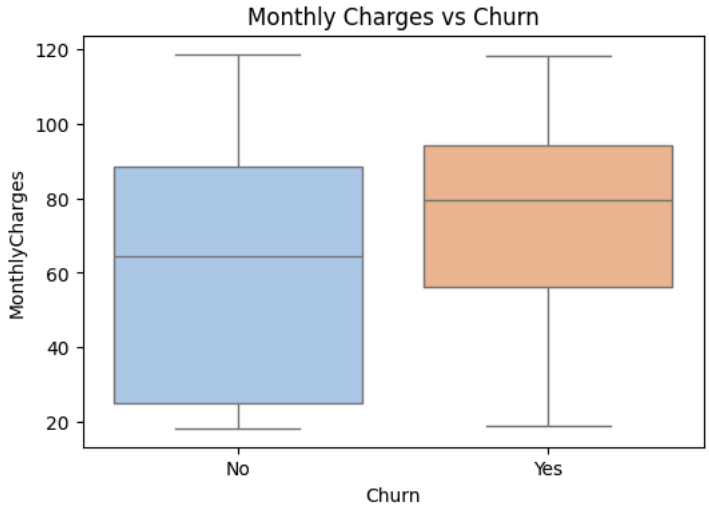
```
print("Statistik TotalCharges:")
print(df['TotalCharges'].describe())

plt.figure(figsize=(6, 4))
sns.histplot(data=df, x='TotalCharges', bins=30, color='lightblue')
plt.title('Distribusi Total Biaya Selama Langganan')
plt.xlabel('Total Charges')
plt.ylabel('Jumlah Pelanggan')
plt.show()
```

```
Statistik TotalCharges:
count    7032.000000
mean     2283.300441
std      2266.771362
min        18.800000
25%       401.450000
50%      1397.475000
75%      3794.737500
max      8684.800000
Name: TotalCharges, dtype: float64
```




```
plt.figure(figsize=(6, 4))
sns.boxplot(x='Churn', y='MonthlyCharges', data=df, palette='pastel')
plt.title('Monthly Charges vs Churn')
plt.show()
```

```
C:\Users\lenovo ideapad 330\AppData\Local\Temp\ipykernel_2144\749981182.py:2: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `le

  sns.boxplot(x='Churn', y='MonthlyCharges', data=df, palette='pastel')
```

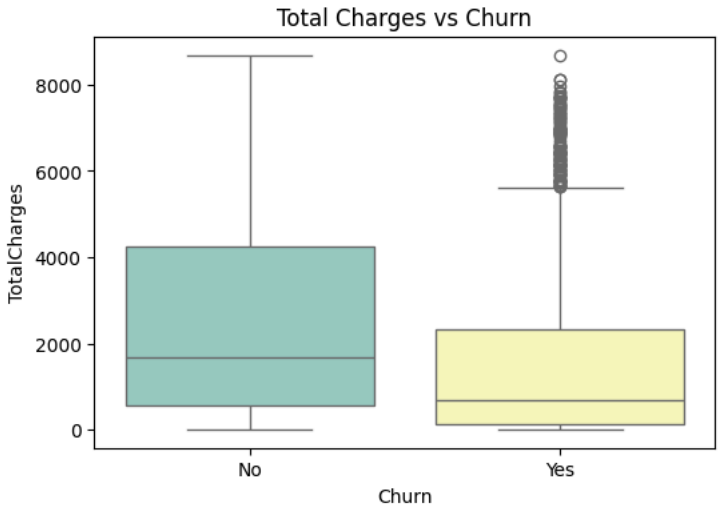



```
print("Rata-rata MonthlyCharges berdasarkan Churn:")
print(df.groupby('Churn')['MonthlyCharges'].mean().round(2))

print("\nMedian MonthlyCharges berdasarkan Churn:")
print(df.groupby('Churn')['MonthlyCharges'].median())
```

```
Rata-rata MonthlyCharges berdasarkan Churn:
Churn
No     61.31
Yes    74.44
Name: MonthlyCharges, dtype: float64

Median MonthlyCharges berdasarkan Churn:
Churn
No     64.45
Yes    79.65
Name: MonthlyCharges, dtype: float64
```

```
plt.figure(figsize=(6, 4))
sns.boxplot(x='Churn', y='TotalCharges', data=df, palette='Set3')
plt.title('Total Charges vs Churn')
plt.show()
```

```
C:\Users\lenovo ideapad 330\AppData\Local\Temp\ipykernel_2144\2754891282.py:2: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `le

  sns.boxplot(x='Churn', y='TotalCharges', data=df, palette='Set3')
```

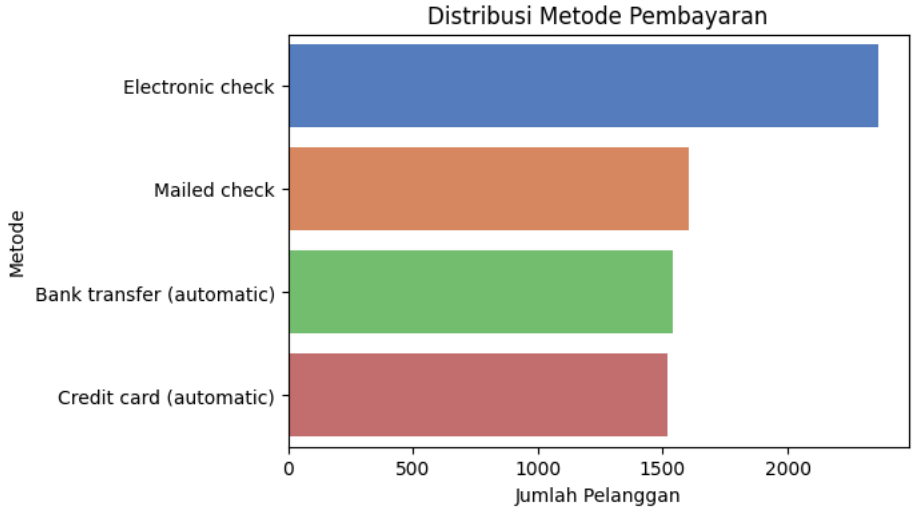



```
print("Rata-rata TotalCharges berdasarkan Churn:")
print(df.groupby('Churn')['TotalCharges'].mean().round(2))

print("\nMedian TotalCharges berdasarkan Churn:")
print(df.groupby('Churn')['TotalCharges'].median())
```

```
Rata-rata TotalCharges berdasarkan Churn:
Churn
No     2555.34
Yes    1531.80
Name: TotalCharges, dtype: float64

Median TotalCharges berdasarkan Churn:
Churn
No     1683.60
Yes     703.55
Name: TotalCharges, dtype: float64
```

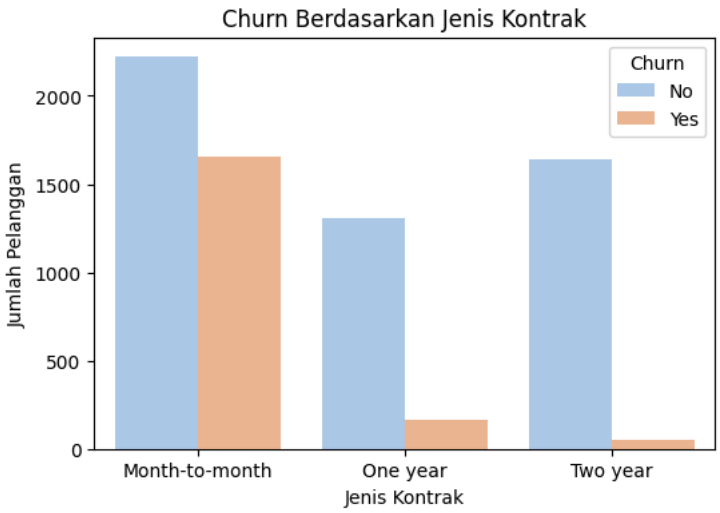
```
print("Metode Pembayaran Pelanggan:")
print(df['PaymentMethod'].value_counts())

plt.figure(figsize=(6, 4))
sns.countplot(y='PaymentMethod', data=df, order=df['PaymentMethod'].value_counts().index, palette='muted')
plt.title('Distribusi Metode Pembayaran')
plt.xlabel('Jumlah Pelanggan')
plt.ylabel('Metode')
plt.show()
```

```
Metode Pembayaran Pelanggan:
PaymentMethod
Electronic check             2365
Mailed check                 1604
Bank transfer (automatic)    1542
Credit card (automatic)      1521
Name: count, dtype: int64
C:\Users\lenovo ideapad 330\AppData\Local\Temp\ipykernel_2144\3293595780.py:5: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable to `hue` and set `le

  sns.countplot(y='PaymentMethod', data=df, order=df['PaymentMethod'].value_counts().index, palette='muted')
```




```
#Bagian 4: Analisis Churn
# Analisis Churn berdasarkan Jenis Kontrak
# 🔍 Cek jumlah pelanggan churn berdasarkan jenis kontrak (bulanan, 1 tahun, 2 tahun)
print(df.groupby(['Contract', 'Churn']).size())
# Hitung persentase churn untuk setiap jenis kontrak
print("\nPersentase churn per jenis kontrak:")
print(pd.crosstab(df['Contract'], df['Churn'], normalize='index') * 100)
# 📊 Visualisasi churn berdasarkan jenis kontrak
plt.figure(figsize=(6, 4))
sns.countplot(data=df, x='Contract', hue='Churn', palette='pastel')
plt.title('Churn Berdasarkan Jenis Kontrak')
plt.xlabel('Jenis Kontrak')
plt.ylabel('Jumlah Pelanggan')
plt.show()
```

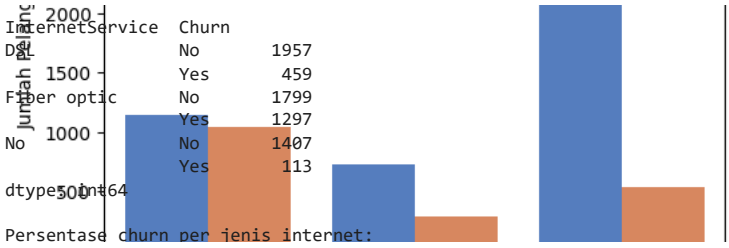
```
Contract        Churn
Month-to-month  No       2220
                Yes      1655
One year        No       1306
                Yes       166
Two year        No       1637
                Yes        48
dtype: int64

Persentase churn per jenis kontrak:
Churn                 No        Yes
Contract
Month-to-month  57.290323  42.709677
One year        88.722826  11.277174
Two year        97.151335   2.848665
```




```
#Analisis Churn berdasarkan Lama Berlangganan (Tenure Group)
# Cek jumlah pelanggan churn berdasarkan kategori lama berlangganan
#Hitung persentase churn berdasarkan kategori Tenure
print(df.groupby(['TenureGroup', 'Churn']).size())
print("\nPersentase churn per kategori tenure:")
print(pd.crosstab(df['TenureGroup'], df['Churn'], normalize='index') * 100)
# 📊 Visualisasi churn berdasarkan TenureGroup (Baru, Menengah, Lama)
plt.figure(figsize=(6, 4))
sns.countplot(data=df, x='TenureGroup', hue='Churn', order=['Baru', 'Menengah', 'Lama'], palette='muted')
plt.title('Churn Berdasarkan Lama Berlangganan')
plt.xlabel('Kategori Tenure')
plt.ylabel('Jumlah Pelanggan')
plt.show()
```

```
TenureGroup  Churn
Baru         No       1138
             Yes      1037
Lama         No       3295
```

```
# Analisis Churn berdasarkan Jenis Internet
# Cek jumlah pelanggan churn berdasarkan jenis internet (DSL, Fiber, Tidak ada)
print(df.groupby(['InternetService', 'Churn']).size())

# Hitung persentase churn per jenis layanan internet
print("\nPersentase churn per jenis internet:")
print(pd.crosstab(df['InternetService'], df['Churn'], normalize='index') * 100)

# 📊 Visualisasi churn berdasarkan jenis Internet
plt.figure(figsize=(6, 4))
sns.countplot(data=df, x='InternetService', hue='Churn', palette='Set2')
plt.title('Churn Berdasarkan Jenis Layanan Internet')
plt.xlabel('Jenis Internet')
plt.ylabel('Jumlah Pelanggan')
plt.show()
```




```
InternetService  Churn
DSL              No       1957
                 Yes       459
Fiber optic      No       1799
                 Yes      1297
No               No       1407
                 Yes       113
dtype: int64

Persentase churn per jenis internet:
```