

# Telecom Customer Churn Analysis

## Using Python

A Data-Driven Approach to Understand Customer Behavior

### **Introduction:**

Customer churn refers to the loss of customers, and analyzing churn helps businesses improve their retention strategies. Customer churn is a critical business metric that measures the percentage of customers who stop using a service or product. High churn rates can indicate issues with customer satisfaction, service quality, or pricing. This project aims to analyze customer churn using Python to uncover key factors contributing to churn and develop strategies to retain customers.

### **Objective:**

The purpose of this analysis is to identify factors contributing to customer churn, predict churn likelihood, and provide actionable insights for improving customer retention.

### **Project Objective:**

The goal is to predict customer churn and identify the key factors that contribute the most to churn, helping the business reduce churn rates and improve customer retention.

### **Key Question to Address**

1. How does the choice of payment method relate to customer churn?
2. What percentage of our customers have churned, and what could be the reasons?
3. Is there a difference in churn rate based on gender?
4. Does age play a role in customer churn?
5. Do senior citizens have a lower churn rate?
6. Does the length of time a customer has been using the service affect churn?
7. How does the contract type influence churn?

- 8.Which services are associated with higher churn rates?
- 9.How does the type of internet service affect churn?
- 10.Does the amount of monthly charges affect the churn rate?
- 11.How does tenure distribution relate to customer churn?

## Dataset Description :

The dataset for the **Customer Churn Analysis** project contains key customer data, including demographics, contract details, service usage, and monthly charges.The dataset consists of **Total Columns: 21 xTotal Rows: 7043** , capturing various customer behaviors and demographics to help in predicting churn. Major features include:

- **Demographics:** Gender, SeniorCitizen, Partner, Dependents
- **Service Information:** PhoneService, InternetService, Streaming, etc.
- **Contract & Payment:** Contract type, PaymentMethod, MonthlyCharges
- **Outcome:** Churn (Yes/No)

## Tools and Technologies Used

For this project, we utilized the following tools:

- **Python:** Main programming language.
  - **Pandas:** Data manipulation.
  - **NumPy:** Numerical operations.
- **Visualization:**
  - **Matplotlib & Seaborn:** For data visualization.

## Data Loading and Overview

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

df = pd.read_csv('Customer Churn.csv')
```

I have used Python's pandas library to load the customer data for analysis. Below is a snapshot of the data to provide an overview of the dataset.

## Study Data:

### Display the number of Rows and Columns

```
rows, columns = df.shape
print(f'Total Rows: {rows}')
print(f'Total Columns: {columns}')
```

Total Rows: 7043  
Total Columns: 21

```
df.head()
```

## Dataset Preview

```
df.head()
```

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

```
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
```

## Data Cleaning

```
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 1
```

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

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

```
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
```

## Exploratory Data Analysis (EDA)

### Jupyter Notebook:

Used to write, run, and document Python code. It's an interactive environment where code, visualizations, and markdowns can be presented together, making it ideal for analysis. Seaborn and Matplotlib were used during EDA to create visualizations and explore data patterns.

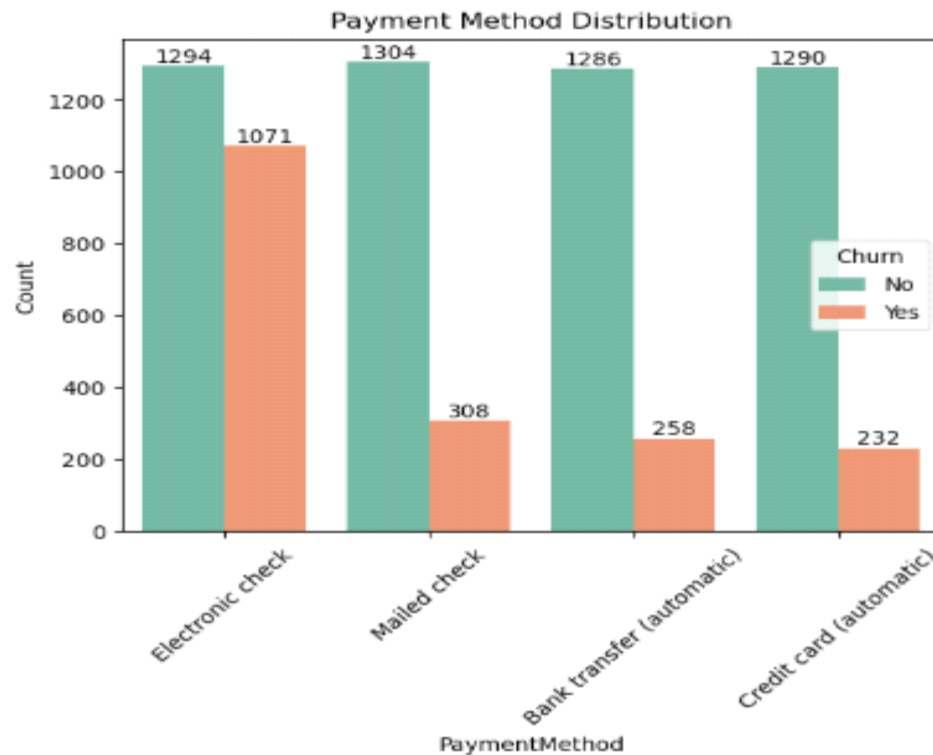
### 1. Customer Payment Method and Churn

**Question:** How does the choice of payment method relate to customer churn?

#### Data Load & Overview:

```
# Create the countplot
ax = sns.countplot(x='PaymentMethod', data=df, hue="Churn", palette="Set2")
plt.title("Payment Method Distribution")
plt.xlabel("PaymentMethod")
plt.ylabel("Count")
for container in ax.containers:
    ax.bar_label(container, fmt="%.0f")

plt.xticks(rotation=45)
plt.show()
```



#### Details:

- Customers who use electricity checks as a payment method tend to churn more often. This could be linked to dissatisfaction with the service or other external factors influencing their decision to leave.
- Visualization:** sns.countplot to show the relationship between payment method and churn.

## 2.Churn Rate of Customers

**Question:** What percentage of our customers have churned, and what could be the reasons?

#### Load Data & Overview

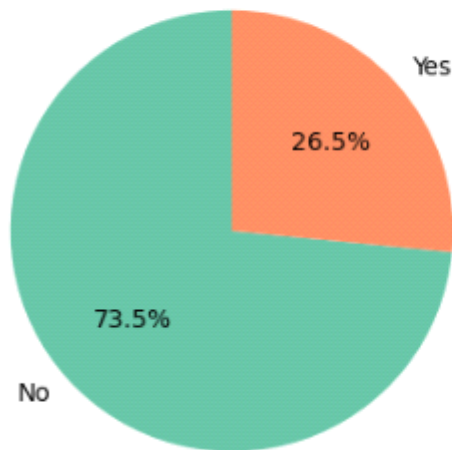
```

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()

```

Percentage of Churned Customers



#### Details:

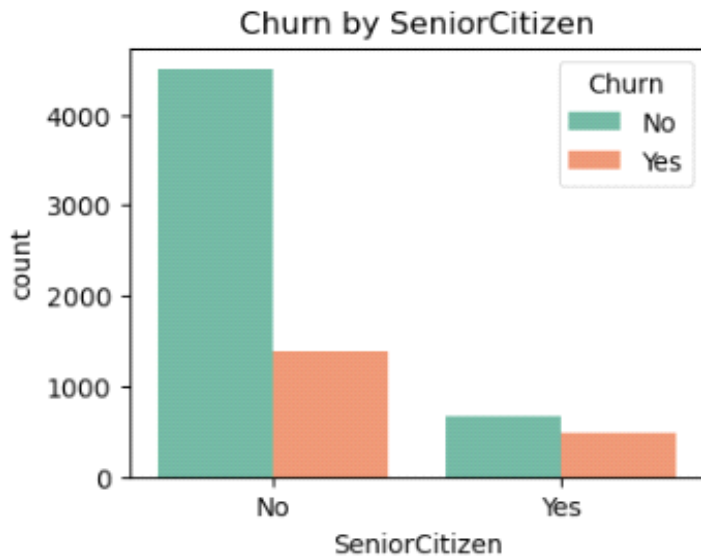
- The pie chart shows that 26.54% of customers have churned. Analyzing the reasons behind this high churn rate can help identify areas for improvement.
- **Visualization:** plt.pie to show churn percentage.

### 3. Gender and Churn Rate

**Question:** Is there a difference in churn rate based on gender?

#### Load Data & Overview

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



#### Details:

- Female customers seem to have a higher churn rate compared to males, as indicated by the larger proportion of "Yes" (churned) in the female category.
- **Visualization:** sns.countplot to show gender-based churn comparison.

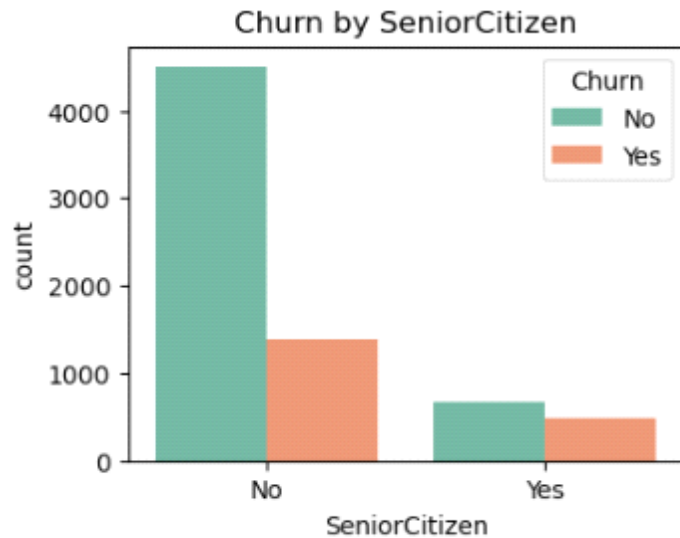
## 4. Age Group and Churn Rate

**Question:** Does age play a role in customer churn?

#### Load Data & Overview



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



#### Details:

- Non-senior citizens exhibit a higher churn rate compared to senior citizens, with the majority of churn coming from younger age groups.
- **Visualization:** sns.countplot to visualize churn by age groups.

## 5. Senior Citizens and Churn

**Question:** Do senior citizens have a lower churn rate?

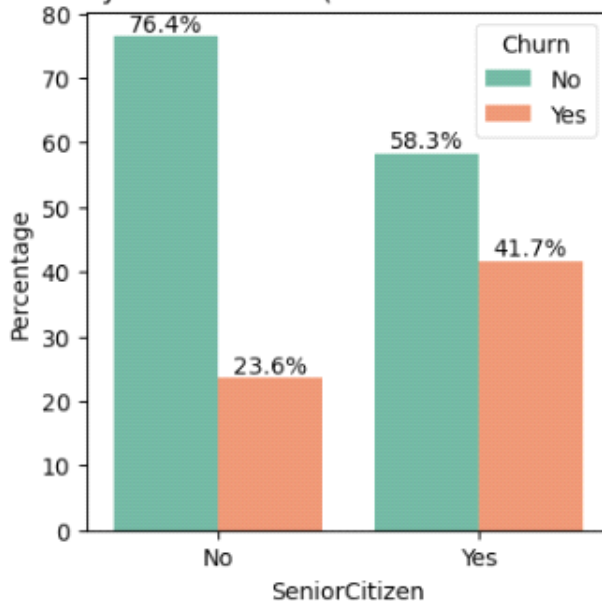
### Load Data & Overview

```

df_grouped = df.groupby(['SeniorCitizen', 'Churn']).size().reset_index(name='Count')
df_grouped['Percentage'] = df_grouped.groupby('SeniorCitizen')['Count'].transform(lambda x: 100 * x / x.sum())
plt.figure(figsize=(4, 4))
ax = sns.barplot(x='SeniorCitizen', y='Percentage', hue='Churn', data=df_grouped, palette="Set2")
for container in ax.containers:
    ax.bar_label(container, fmt="%.1f%%")
plt.title("Churn by SeniorCitizen (Stacked with Percentages)")
plt.ylabel("Percentage")
plt.show()

```

Churn by SeniorCitizen (Stacked with Percentages)



#### Details:

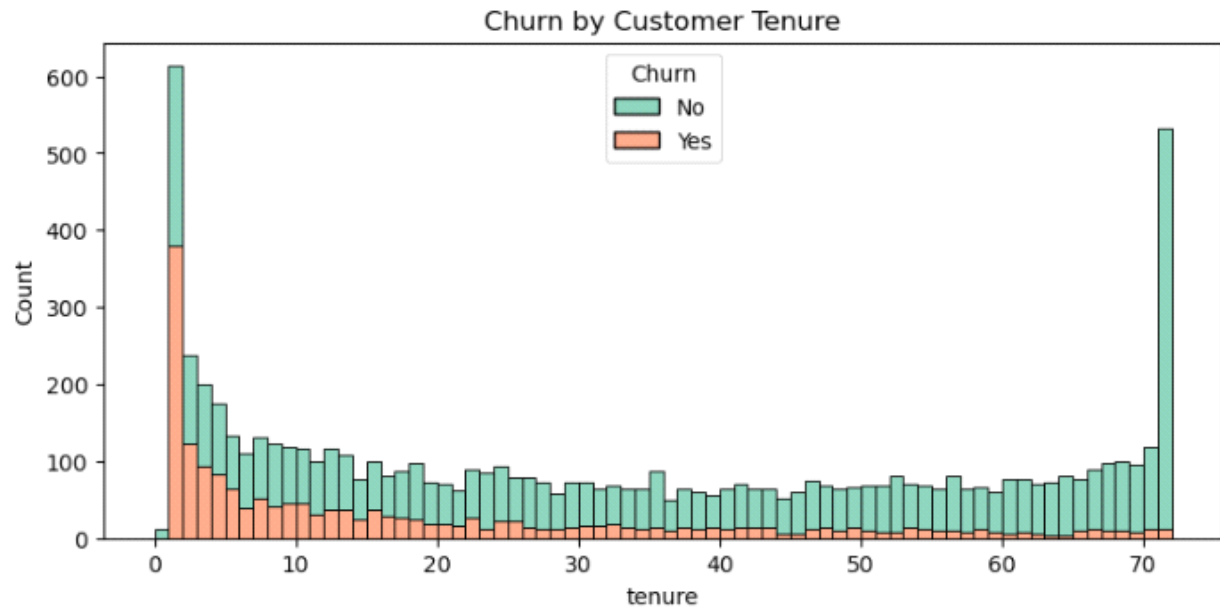
- A comparative bar chart shows a greater percentage of senior citizens churned compared to younger customers.
- **Visualization:** sns.barplot to compare churn rate in senior citizens versus non-senior citizens.

## 6. Tenure and Customer Retention

**Question:** Does the length of time a customer has been using the service affect churn?

### Load Data & Overview

```
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()
```



#### Details:

- Customers who have been using the service for a longer period tend to stay, while those who have used the service for only 1 or 2 months are more likely to churn.
- **Visualization:** sns.histplot to visualize churn based on customer tenure.

## 7. Contract Type and Churn

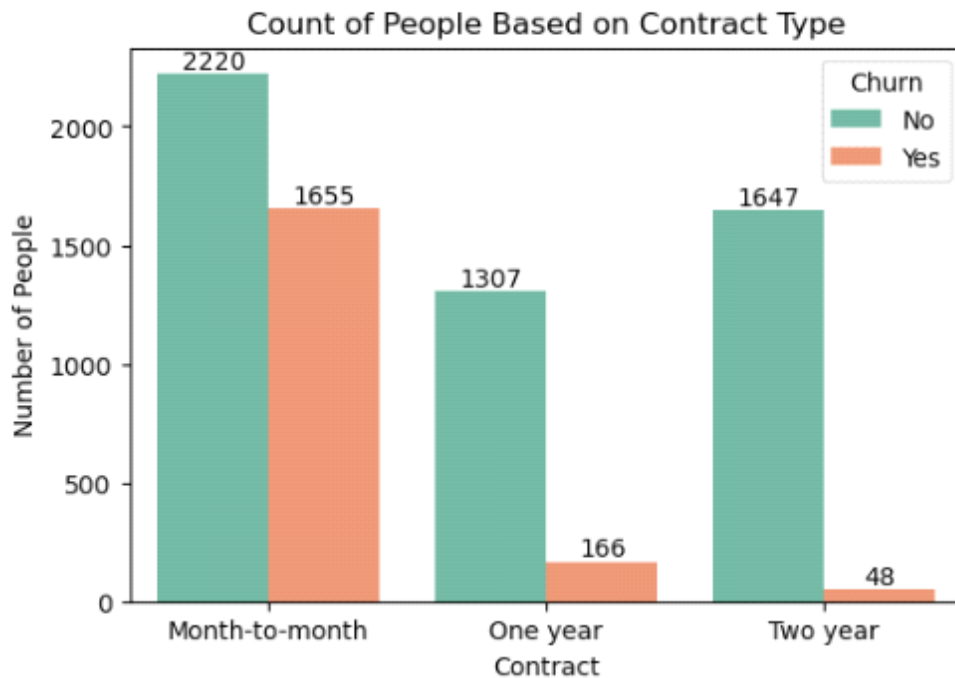
**Question:** How does the contract type influence churn?

### Load Data & Overview

```

: # Group by 'Contract' and 'Churn' and get the count
df_grouped = df.groupby(['Contract', 'Churn']).size().reset_index(name='Count')
plt.figure(figsize=(6, 4))
ax = sns.barplot(x='Contract', y='Count', data=df_grouped, hue='Churn', palette="Set2")
for container in ax.containers:
    ax.bar_label(container, fmt="%.0f")
plt.title("Count of People Based on Contract Type")
plt.ylabel("Number of People")
plt.show()

```



#### Details:

- People with a Month-to-Month contract are more likely to churn compared to those who have a 1 or 2-year contract.
- **Visualization:** sns.barplot to show churn by contract type.

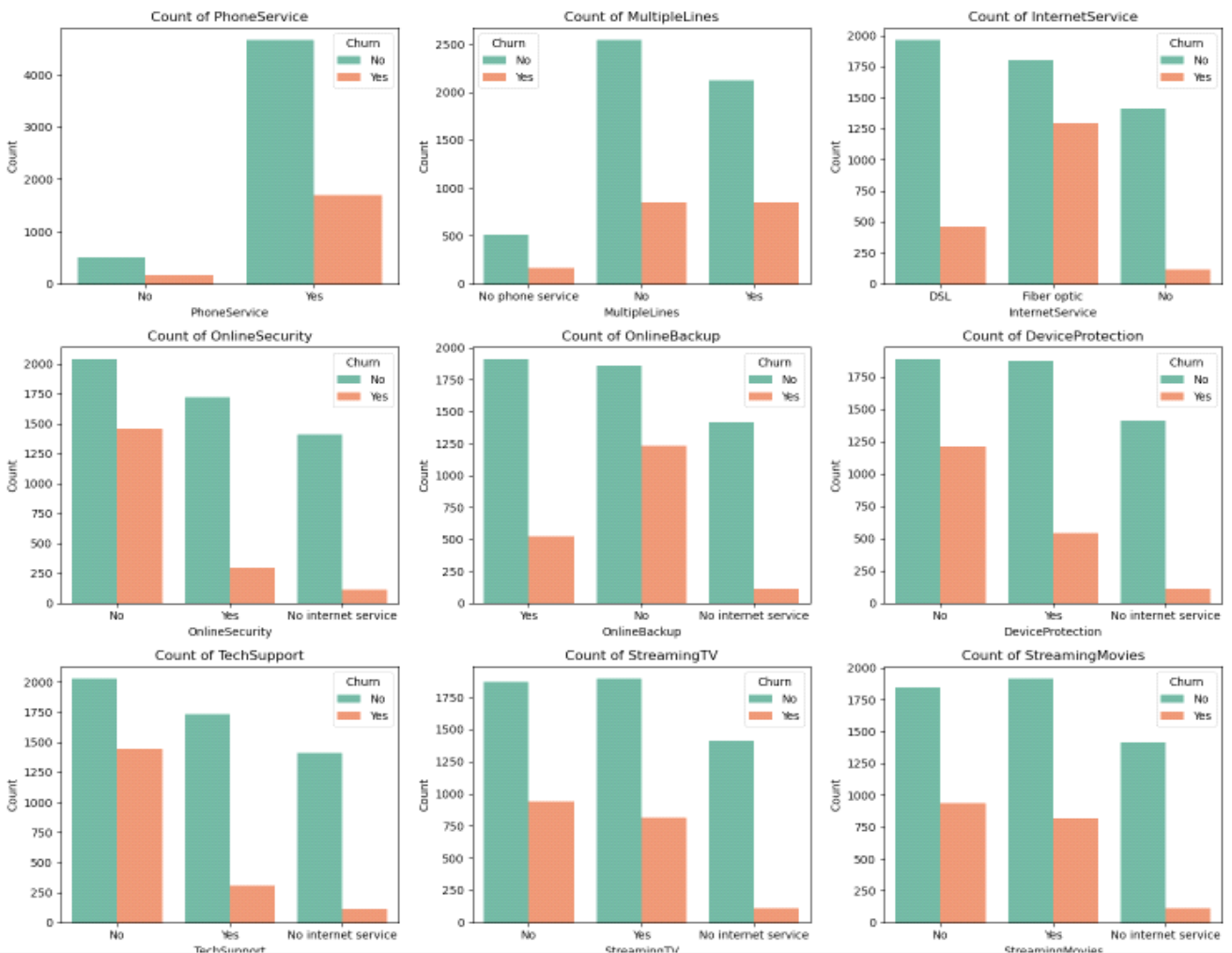
## 8. Churn Across Different Services

**Question:** Which services are associated with higher churn rates?

#### Load Data & Overview

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

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
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')
plt.tight_layout()
plt.show()
```



## Details:

- 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.

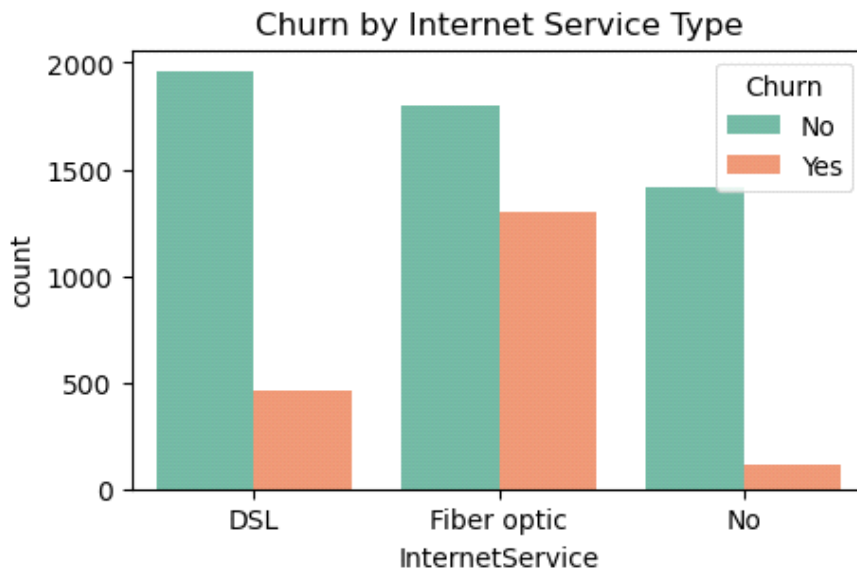
- **Visualization:** plt.tight\_layout for multiple bar charts comparing churn across services.

## 9. Internet Service Type and Churn

**Question:** How does the type of internet service affect churn?

**Load Data & Overview:**

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



**Details:**

- Customers with fiber optic internet are more likely to churn compared to those using DSL or no internet service at all.
- **Visualization:** sns.countplot to compare churn rate by internet service type

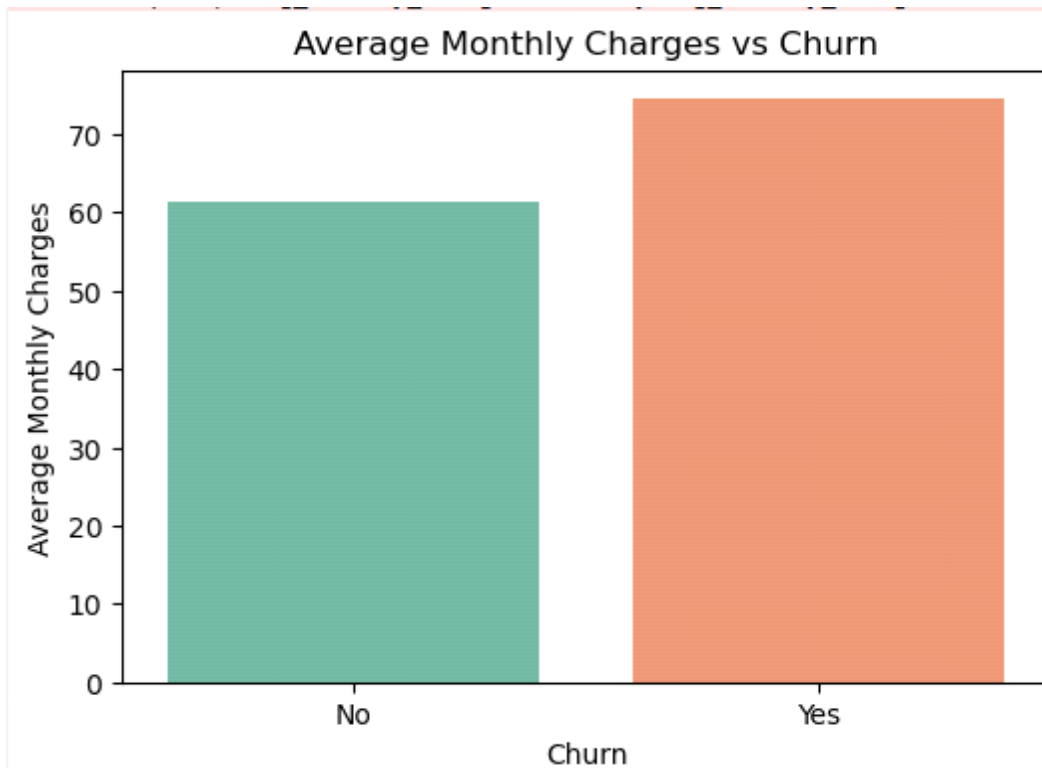
## 10. Monthly Charges and Churn

**Question:** Does the amount of monthly charges affect the churn rate?

**Load Data & Overview:**

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

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()
```



**Details:**

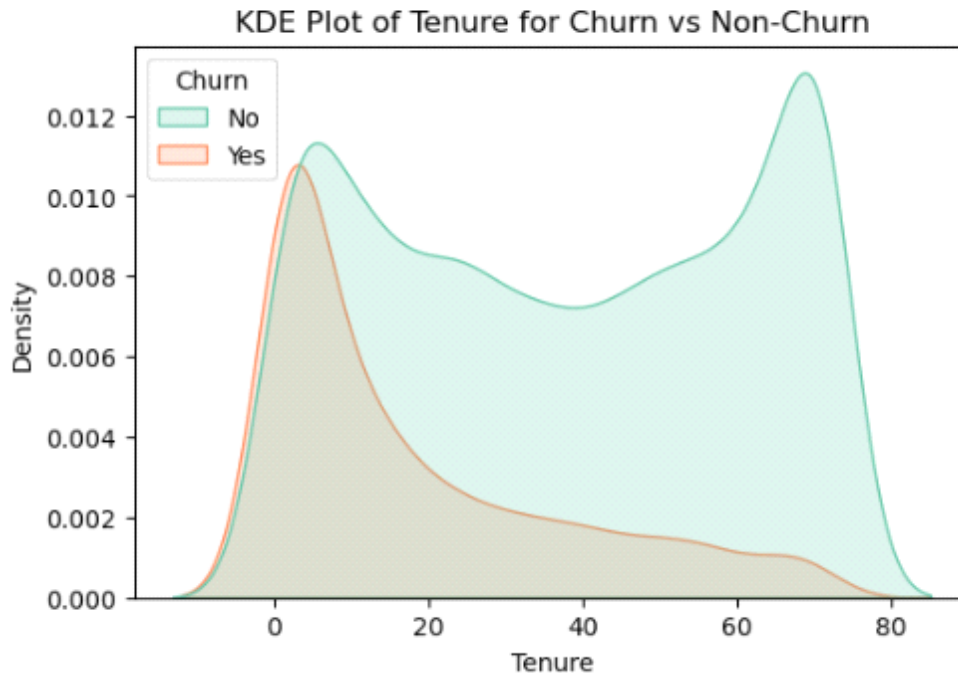
- Customers who churn tend to have higher average monthly charges compared to those who don't.
- **Visualization:** sns.barplot to show the correlation between monthly charges and churn.

## 11. Tenure Distribution and Churn

**Question:** How does tenure distribution relate to customer churn?

**Load Data & Overview:**

```
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()
```



#### Details:

- A Kernel Density Estimate (KDE) plot shows that customers with shorter tenures are more likely to churn, while those with longer tenures tend to stay.
- **Visualization:** sns.kdeplot to show churn patterns based on tenure.

#### Conclusion and Recommendation:

Customer churn analysis helps businesses identify at-risk customers, understand the reasons behind their decision to leave, and take proactive steps to retain them. By focusing on churn, businesses can improve their customer service, personalize offerings, and ultimately reduce the overall churn rate, leading to better customer satisfaction and long-term success. In summary, this Customer Churn Analysis project helped identify key factors driving customer attrition, such as month-to-month contracts, higher monthly charges, and electronic check payment methods. Predictive modeling highlighted customers at risk of churning, providing opportunities for proactive retention efforts.

#### Key Recommendations:

- **Incentivize Long-Term Contracts:** Offer discounts for longer contracts to reduce churn from month-to-month customers.



- **Improve Service for Fiber Optic Customers:** Address any service-related issues for fiber optic users, who have a higher churn rate.
- **Offer Flexible Payment Options:** Provide better payment methods to reduce dissatisfaction with electronic checks.
- **Enhance Onboarding for New Customers:** Support new customers during their initial months to reduce early churn.

### **Future Scope:**

- **Add More Data:** Incorporate customer feedback and sentiment data for deeper insights.
- **Real-Time Prediction:** Develop a real-time churn prediction system for immediate action.
- **Test Advanced Models:** Experiment with more advanced machine learning algorithms for better accuracy.
- **Personalized Strategies:** Create retention strategies tailored to specific customer profiles.

This project offers a strong starting point to reduce churn and retain more customers, with the potential for continuous improvement.