Project Title: Customer Segmentation Visualization

Project Overview:

The project aims to analyze customer churn in a telecommunications company and develop predictive models to identify at-risk customers. The ultimate goal is to provide actionable insights and recommendations to reduce churn and improve customer retention.

Load the dataset using pandas:

```
# Load the dataset using pandas.
import pandas as pd
df = pd.read_csv("C:/Users/pinma/Downloads/Telco_Customer_Churn_Dataset  (3).csv")
```

Display the first 10 rows:

```
# Display the first 10 rows.
df.head(10)
```

Identify the data types of each column:

```
# Identify the data types of each column.
df.info()
```

```
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)
```

Standardize column names:

```python
# Standardize column names (convert to capitalize and replace spaces with underscores).
df.columns = [col.capitalize() for col in df.columns]
```

```
# checking the column name
df.columns
```

```
['Customerid', 'Gender', 'Seniorcitizen', 'Partner', 'Dependents',
 'Tenure', 'Phoneservice', 'Multiplelines', 'Internetservice',
 'Onlinesecurity', 'Onlinebackup', 'Deviceprotection', 'Techsupport',
 'Streamingtv', 'Streamingmovies', 'Contract', 'Paperlessbilling',
 'Paymentmethod', 'Monthlycharges', 'Totalcharges', 'Churn'],
dtype='object')
```

Handling blanks with 0 and changing datatype for Total charges column:

```
# Replacing blanks with 0
df['Totalcharges'] = df['Totalcharges'].replace(" ","0")
df['Totalcharges'] = pd.to_numeric(df['Totalcharges'],errors='coerce')
```

Checking for null values:

```
# checking for any null values
df['Totalcharges'].isnull().sum()
```

```
np.int64(0)
```

Checking for duplicated values:

```
# checking for duplicate values
df.duplicated().sum()
```

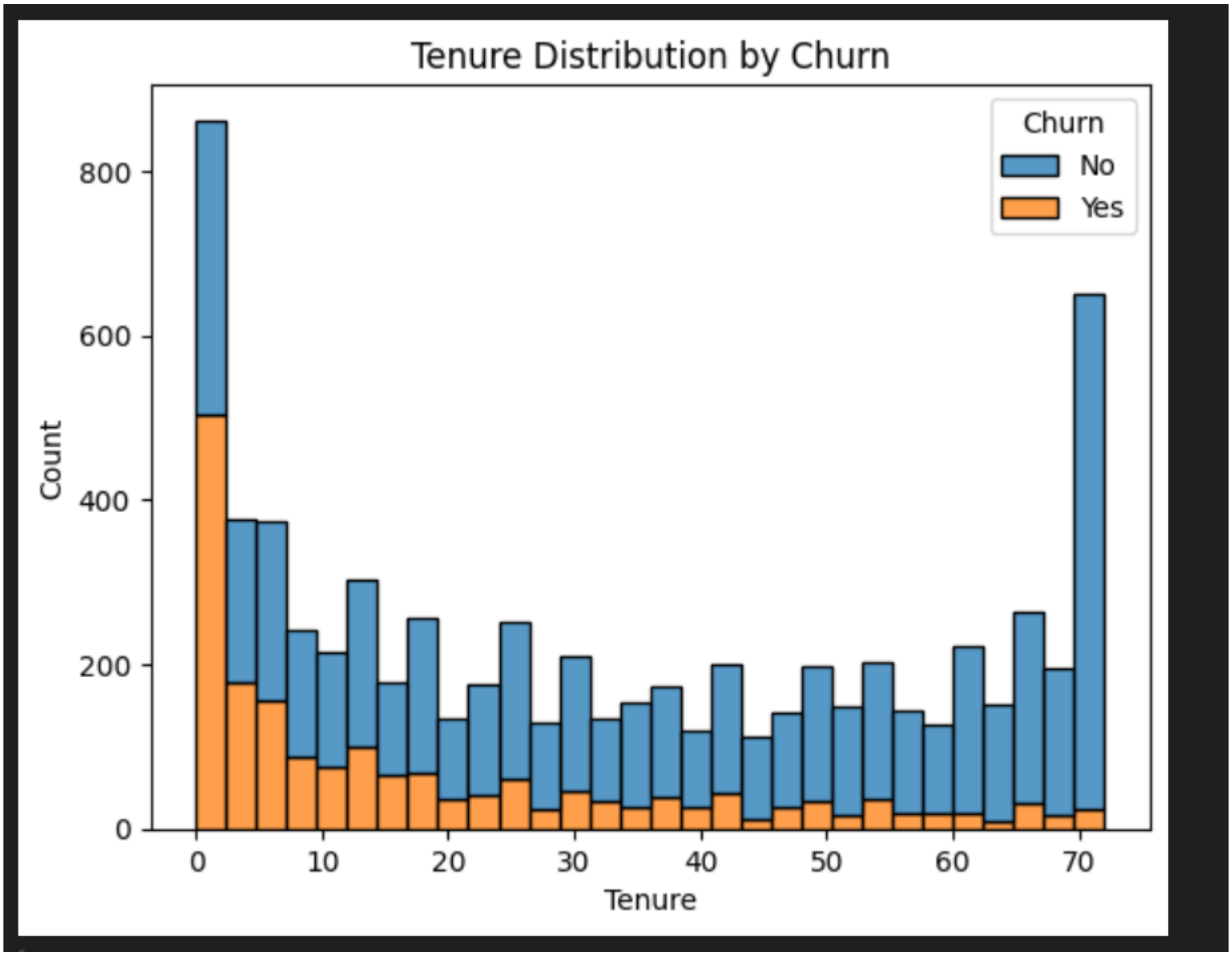
Importing Visualize library:

```
# Visualize libray
import seaborn as sns
import matplotlib.pyplot as plt
```

Understand trends and distributions in the data:

Distribution between Tenure vs Churn (Hist plot):

```python
# Churn Vs Tenure
sns.histplot(data=df,x='Tenure',hue='Churn',legend=True,bins=30,multiple='stack')
plt.xlabel('Tenure')
plt.ylabel('Count')
plt.title('Tenure Distribution by Churn')
plt.show()
```



High churn at very short tenure: At 0 months, churn is extremely high. Many new customers leave almost immediately.

Declining churn with longer tenure: As tenure increases, the proportion of churned customers (orange) decreases steadily.

Near-zero churn at long tenure: At around 72 months, almost all customers are non-churned (blue), showing strong loyalty among long-term customers.

Business Interpretation:

Early lifecycle risk: Customers in their first months are the most vulnerable. If they don't see value quickly, they churn.

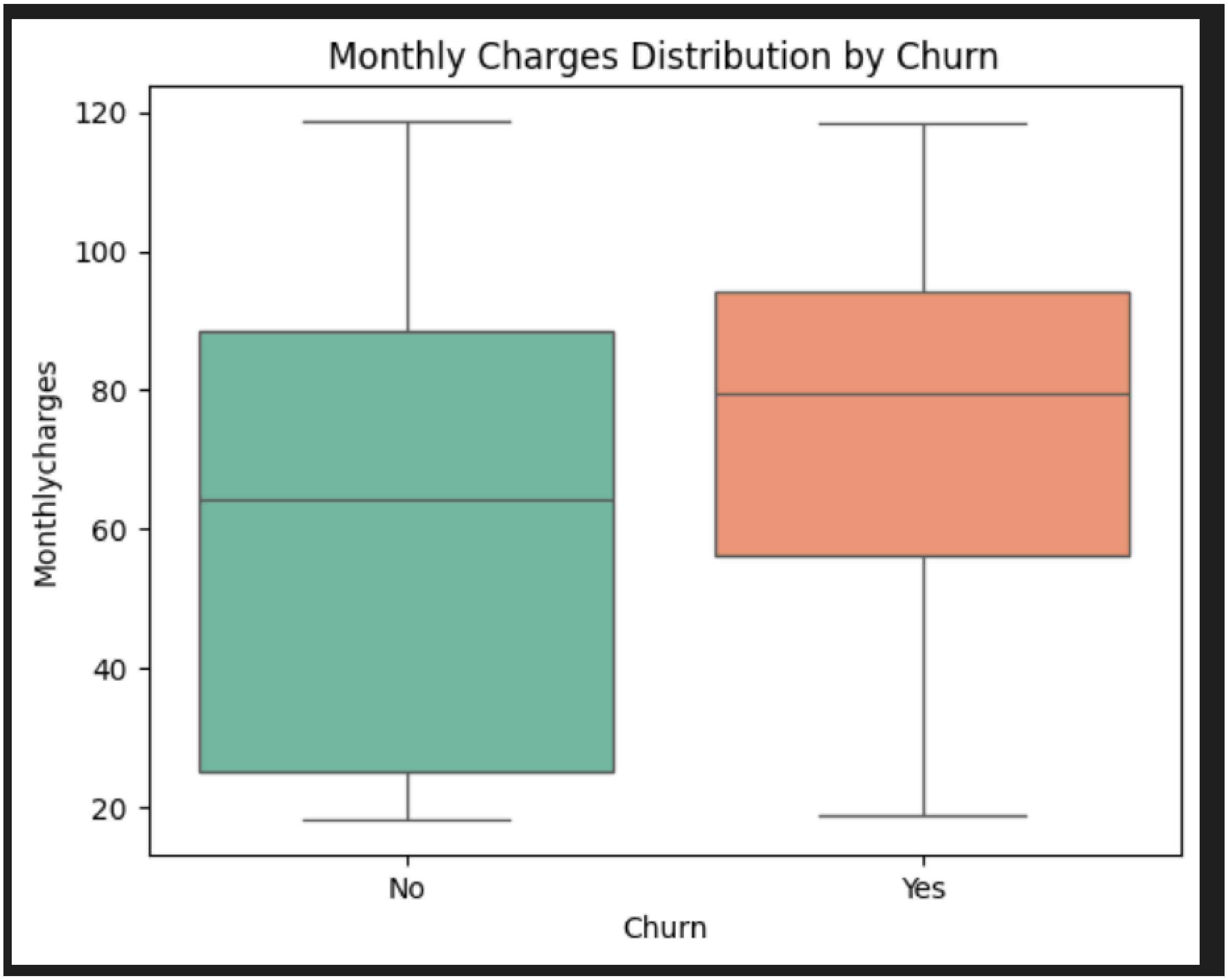
Retention stabilizes over time: Once customers pass the early stage, they are much more likely to stay.

Loyalty segment: Long-tenure customers represent a highly stable base, ideal for upselling or advocacy.

Generate summary statistics (mean, median, mode):

Distribution between Monthly charges vs Churn (Boxplot):

```
# Boxplot : monthly vs Churn
sns.boxplot(data=df,x='Churn', y='Monthlycharges',hue='Churn',palette='Set2',legend=False)
plt.title('Monthly Charges Distribution by Churn')
plt.show()
```



Key Insights from the Figure:
Higher Monthly Charges Linked to Churn
Customers who churned generally had higher monthly charges compared to those who stayed. The median monthly charge for churned customers is around 80, while for non-churned customers it's closer to 65.

Spread of Charges:
For churned customers, the distribution is wider, meaning there's more variability in what they were paying.
Non-churned customers show a tighter distribution, suggesting their charges are more consistent and generally lower.

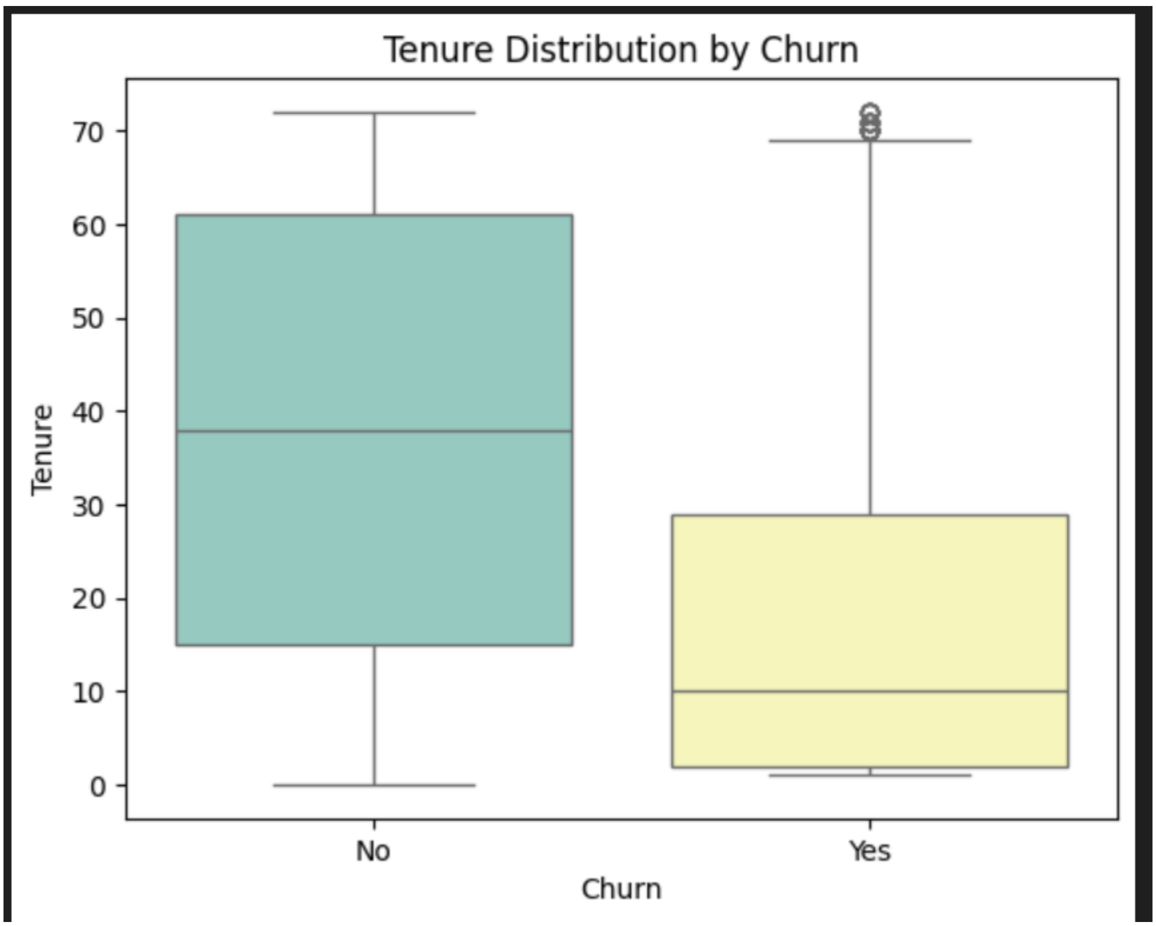
Business Implication:
This pattern suggests that pricing pressure could be a major driver of churn.

Customers paying higher fees may feel less satisfied or perceive less value, making them more likely to leave.

Retention Strategy Angle:
Offering tiered pricing or discounts to high-charge customers could reduce churn. Alternatively, improving the value proposition (better service, added benefits) for those paying more might help retain them.

Distribution between Tenure vs Churn (Boxplot):

```
# Boxplot: Tenure vs Churn
sns.boxplot(x='Churn', y='Tenure', data=df,  hue='Churn', palette='Set3',legend=False)
plt.title("Tenure Distribution by Churn")
plt.show()
```



Key Insights:

Tenure Strongly Correlates with Churn
Customers who churned generally had shorter tenures. The median tenure for churned customers is around 10 months, while for non-churned customers it’s closer to 38 months.

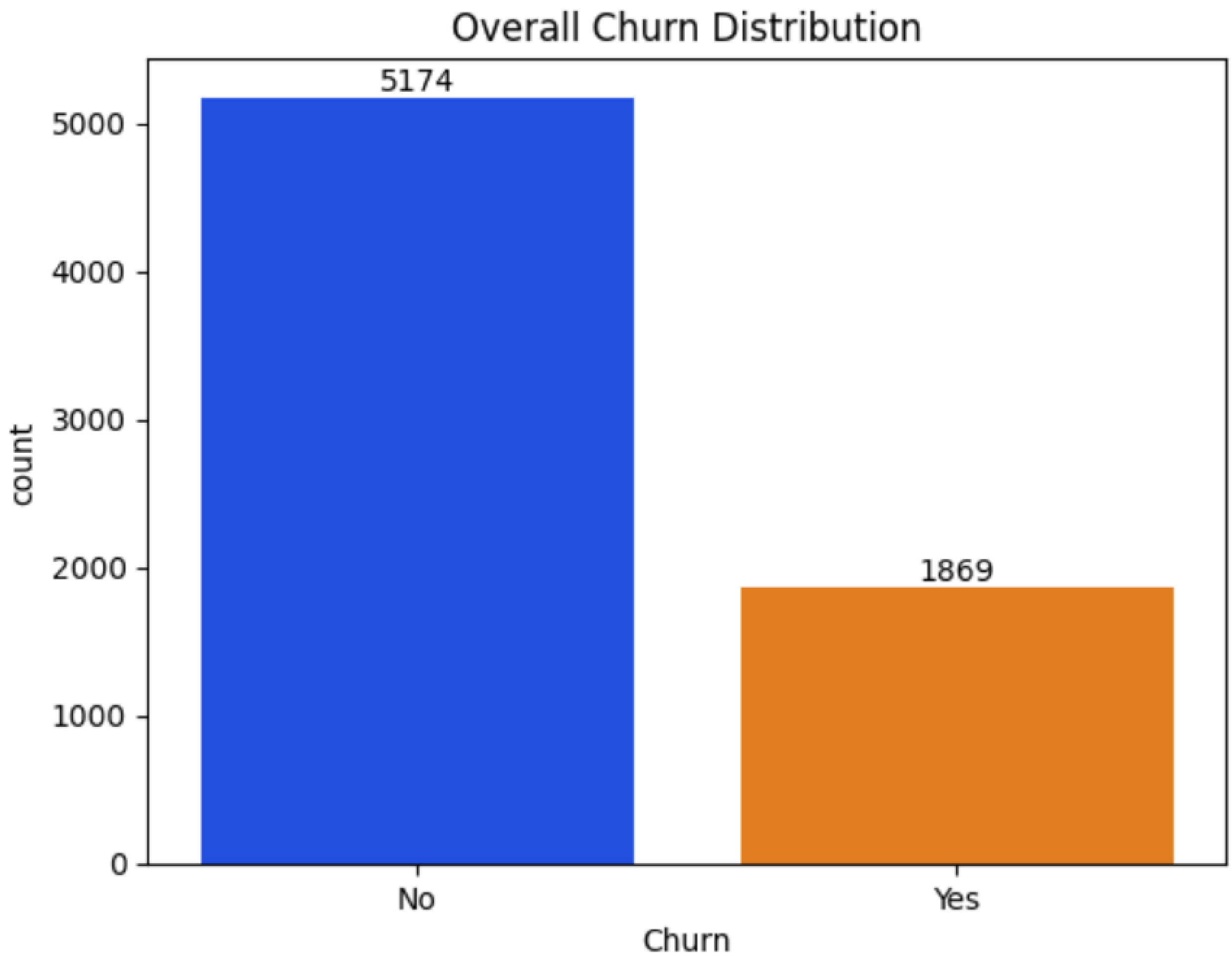
Distribution Differences:
Non-churned customers: Tenure values are spread widely, with many long-term customers (up to 70 months).
Churned customers: Most left early, within the first 30 months. A few outliers stayed longer but still churned.

Business Implication:
This suggests that early-stage customers are at the highest risk of churn. If they don’t find value quickly, they leave before becoming long-term loyal customers.

Retention Strategy Angle:
Focus on onboarding and early engagement: The first year is critical.
Provide welcome offers, personalized support, and quick wins to ensure new customers see value fast.
Monitor customers with tenure under 12 months closely, as they’re the most vulnerable.

Churn distribution:

```
# Visualize churn distribution
ax = sns.countplot(data=df,x='Churn',hue='Churn',palette='bright', legend=False)
ax.bar_label(ax.containers[0])
ax.bar_label(ax.containers[1])
plt.title('Overall Churn Distribution')
plt.show()
```



Non-churned customers ("No"): 5,174.
Churned customers ("Yes"): 1,869.
That means about 26–27% of customers churned, while the majority (around 73–74%) stayed.

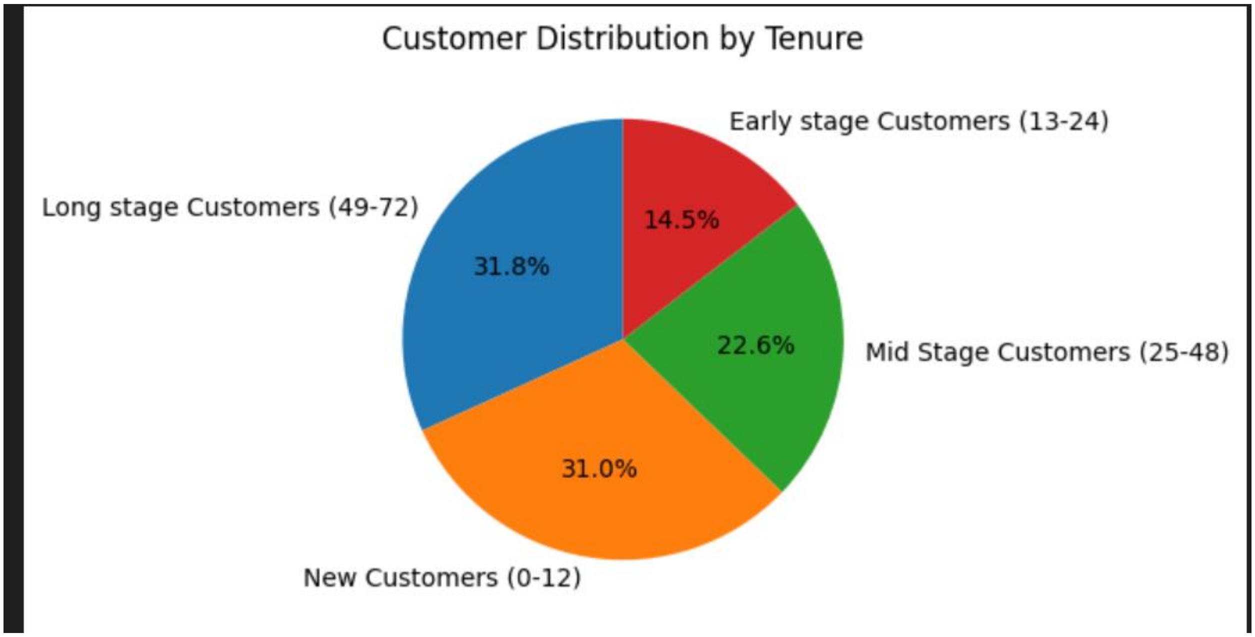
Customer Segmentation by Tenure:

```
# Tenure Segmentation
df['Tenure Group'] = pd.cut(df['Tenure'],bins=[0,12,24,48,72],labels=['New Customers (0-12)',
                                                                      'Early stage Customers (13-24)',
                                                                      'Mid Stage Customers (25-48)',
                                                                      'Long stage Customers (49-72)'], include_lowest=True)
```

Visualization:

```
# Count customers in each tenure group
tenure_counts = df['Tenure Group'].value_counts()

# Pie chart
plt.figure(figsize=(4,4))
plt.pie(tenure_counts, labels=tenure_counts.index, autopct='%1.1f%%', startangle=90)
plt.title("Customer Distribution by Tenure")
plt.show()
```



Long Stage Customers **(49–72 months): 31.8%**

New Customers **(0–12 months): 31.0%**

Mid-Stage Customers **(25–48 months): 22.6%**

Early-Stage Customers **(13–24 months): 14.5%**

Insights:
The distribution is bimodal: large proportions at the extremes (new and long-term customers), with fewer in the middle stages.

This suggests:
Retention strategies are working well for customers who pass the early hurdles and stay long term.

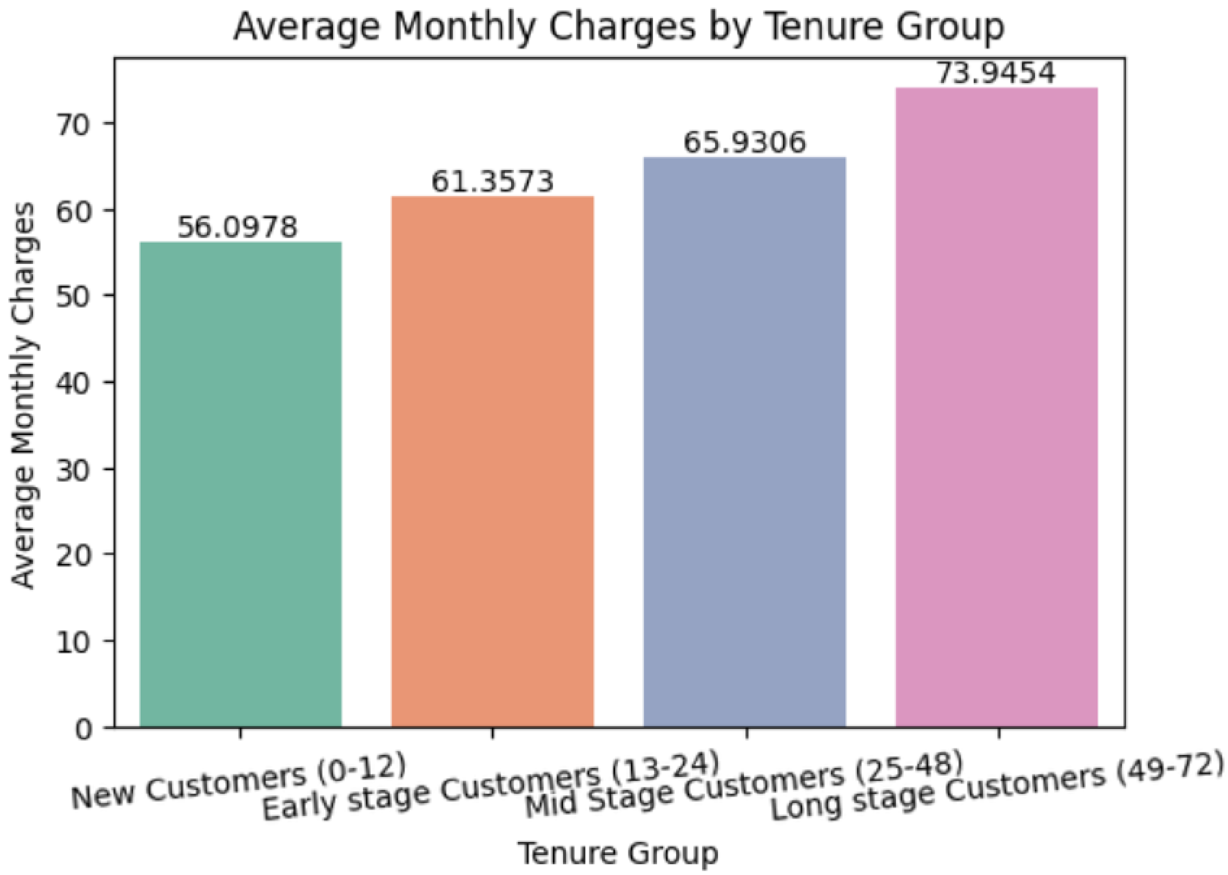
Risk of churn is highest in the first 1–2 years, especially between 13–24 months where the drop is most visible.
Businesses should focus on early engagement and loyalty programs to move more customers from the “new” and “early” stages into the “mid” and eventually “long” stage.

Calculate average monthly charges by grouping tenure groups:

```
# Group by tenure group and calculate average monthly charges
avg_charges = df.groupby('Tenure Group', observed=False)['Monthlycharges'].mean().reset_index()
```

Plotting into Bar plot:

```
plt.figure(figsize=(8,6))
ax =sns.barplot(data=avg_charges, x='Tenure Group', y='Monthlycharges',
            hue='Tenure Group', palette='Set2', legend=False)
ax.bar_label(ax.containers[0])
ax.bar_label(ax.containers[1])
ax.bar_label(ax.containers[2])
ax.bar_label(ax.containers[3])
plt.title("Average Monthly Charges by Tenure Group")
plt.xlabel("Tenure Group")
plt.ylabel("Average Monthly Charges")
plt.xticks(rotation=6)
plt.show()
```



New Customers **(0–12 months): 56.1**

Early-Stage Customers **(13–24 months): 61.4**

Mid Stage Customers **(25–48 months): 65.9**

Long Stage Customers **(49–72 months): 73.9**

Key Insights:
Clear upward trend: The longer a customer stays, the higher their average monthly charges.

Possible reasons:
Long-term customers may adopt more services or premium plans over time.
Discounts or introductory offers for new customers could explain the lower initial charges.
Cross-selling and upselling strategies likely succeed with customers who remain longer.

Business implications:

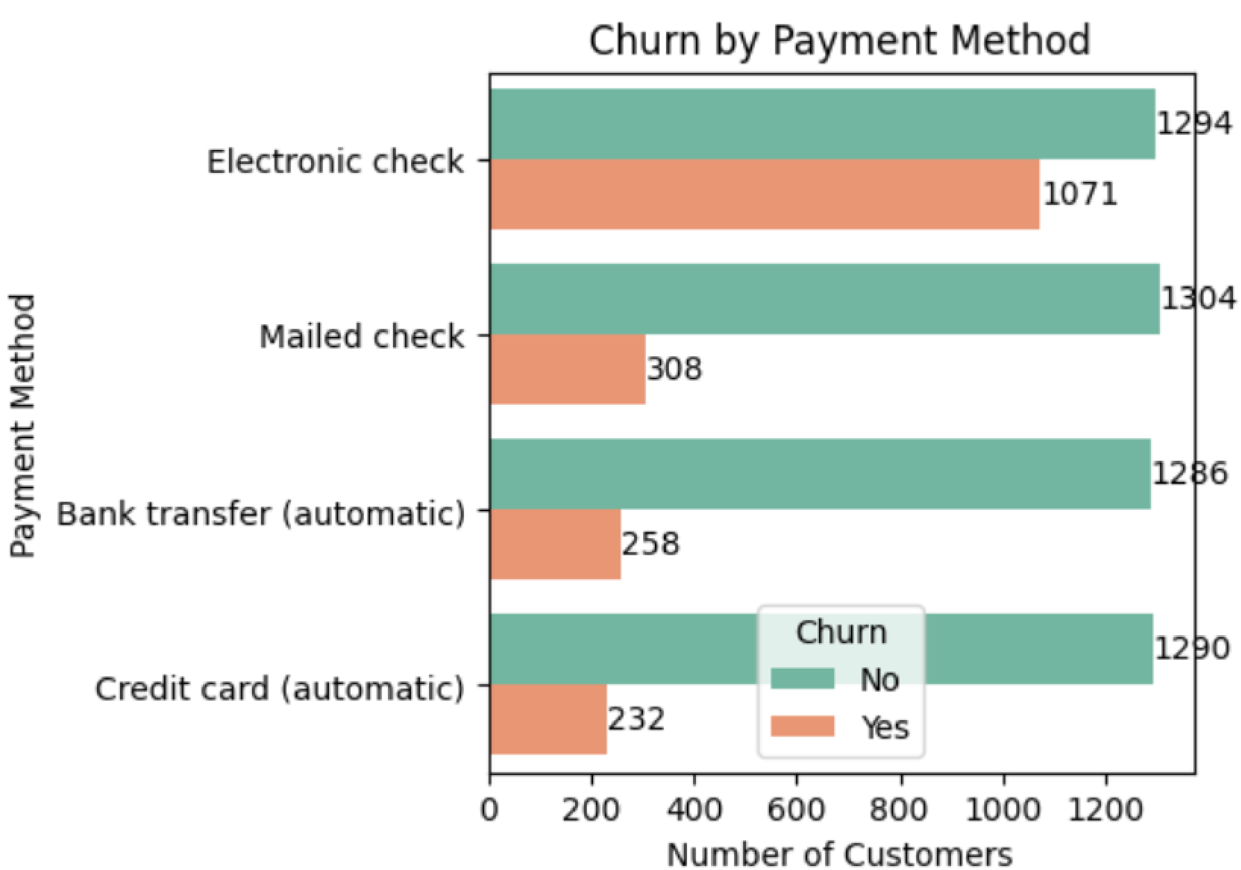
Retaining customers is financially valuable since their monthly contribution grows steadily.

Early-stage retention is critical: moving customers past the first-year increases both loyalty and revenue.

Pricing strategies should balance attracting new customers with maximizing long-term value.

Distribution of payment vs Churn:

```
# Churn by Payment method
ax = sns.countplot(data=df,y='Paymentmethod',hue='Churn',legend=True,palette="Set2")
ax.bar_label(ax.containers[0])
ax.bar_label(ax.containers[1])
plt.xlabel('Number of Customers')
plt.ylabel('Payment Method')
plt.title('Churn by Payment Method')
plt.show()
```



Electronic check:

Did not churn: **1,294**

Churned: **1,071**

→ Very high churn rate, about **45%.**

Mailed check:
Did not churn**: 1,304**
Churned: **308**
→ Lower churn rate, about **19%.**

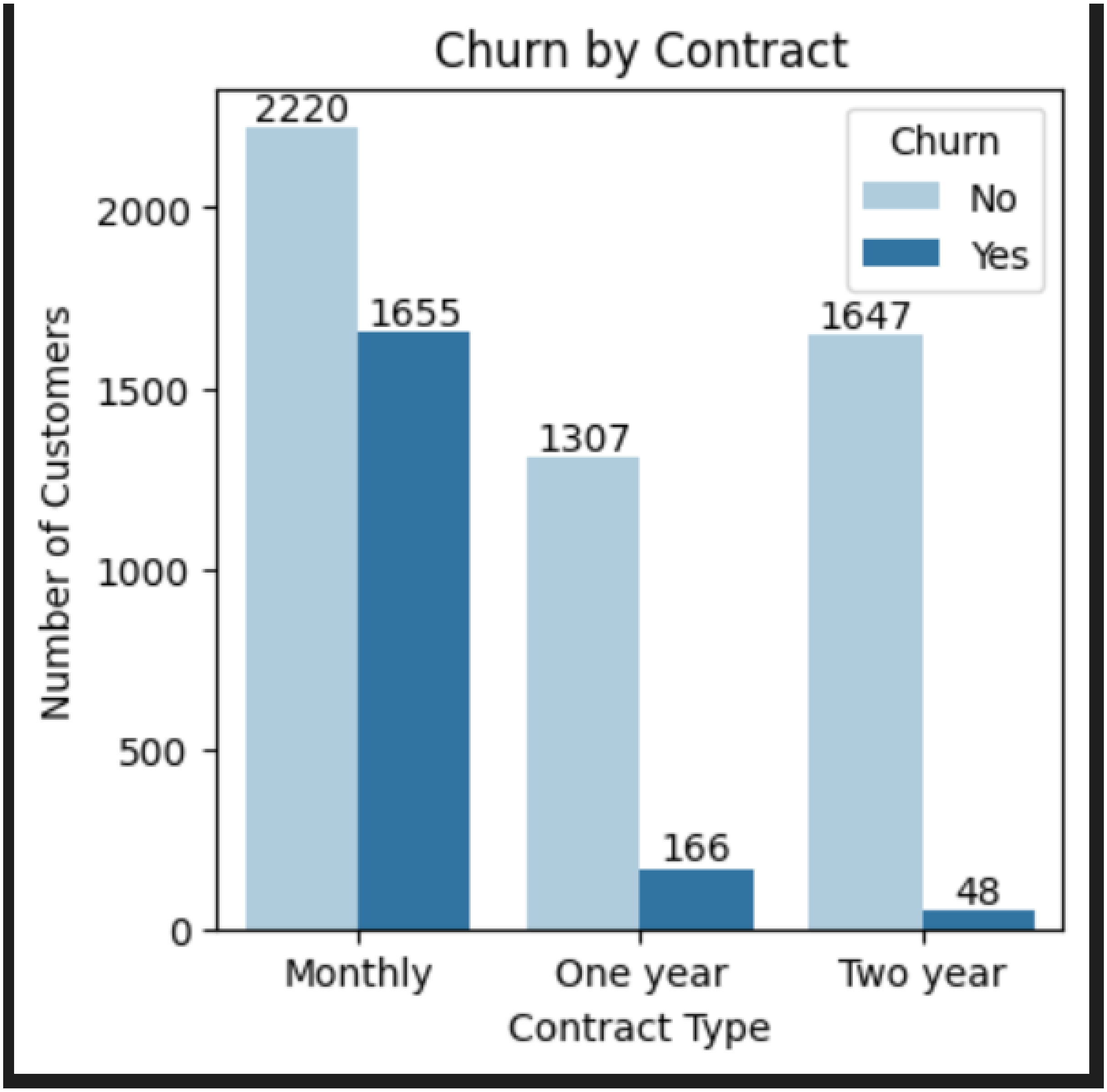
Bank transfer (automatic):
Did not churn**: 1,286**
Churned**: 258**
→ Churn rate about **17%.**

Credit card (automatic):
Did not churn: **1,290**
Churned: **232**
→ Churn rate about **15%.**

Business Interpretation:
Electronic check customers are most at risk: Nearly half of them churn, making this the most vulnerable group.
Automatic payments reduce churn: Both bank transfer and credit card automatic payments show the lowest churn rates. Customers using these methods are more stable and loyal.
Mailed check customers are in between: They churn less than electronic check users but more than automatic payment users.

Distribution between Churn and Contract Type:

```
# Churn by Contract Type
ax = sns.countplot(data=df,x='Contract',hue='Churn',legend=True,palette='Paired')
ax.bar_label(ax.containers[0])
ax.bar_label(ax.containers[1])
plt.title('Churn by Contract')
plt.xlabel('Contract Type')
plt.ylabel('Number of Customers')
plt.show()
```



Month-to-month contracts:
Did not churn: 2,220
Churned: 1,655
→ Very high churn rate, nearly 43%.

One-year contracts:
Did not churn: 1,307
Churned: 166
→ Much lower churn rate, around 11%.

Two-year contracts:
Did not churn: 1,647
Churned: 48
→ Extremely low churn rate, about 3%

Business Interpretation:
Customers on month-to-month plans are far more likely to leave, while those on longer contracts are highly loyal.
Commitment effect: Longer contracts lock customers in, but they also suggest that these customers are more satisfied or see greater value.
Risk segment: Month-to-month customers are the most vulnerable group and should be the focus of retention strategies.

Distribution between Seniors citizen vs churn:

```
# SeniorCitizen distribution
ax = sns.countplot(data=df,x='Seniorcitizen',hue='Churn',legend=True, palette='pastel')
ax.bar_label(ax.containers[0])
ax.bar_label(ax.containers[1])
plt.title("Churn by Senior Citizen")
plt.show()
```

Non-senior citizens:
Did not churn: **4,508**
Churned: **1,393**

Senior citizens:
Did not churn: **666**
Churned: **476**

This means that while most customers are not senior citizens, the churn rate among senior citizens is much higher.
Senior citizens are more vulnerable to churn: Nearly 42% of senior citizens churned (476 out of 1,142), compared to only about 24% of non-senior citizens.

Possible reasons for churn:

Senior citizens may be more price-sensitive.

They may face usability or service challenges.

They might prefer simpler or lower-cost alternatives.

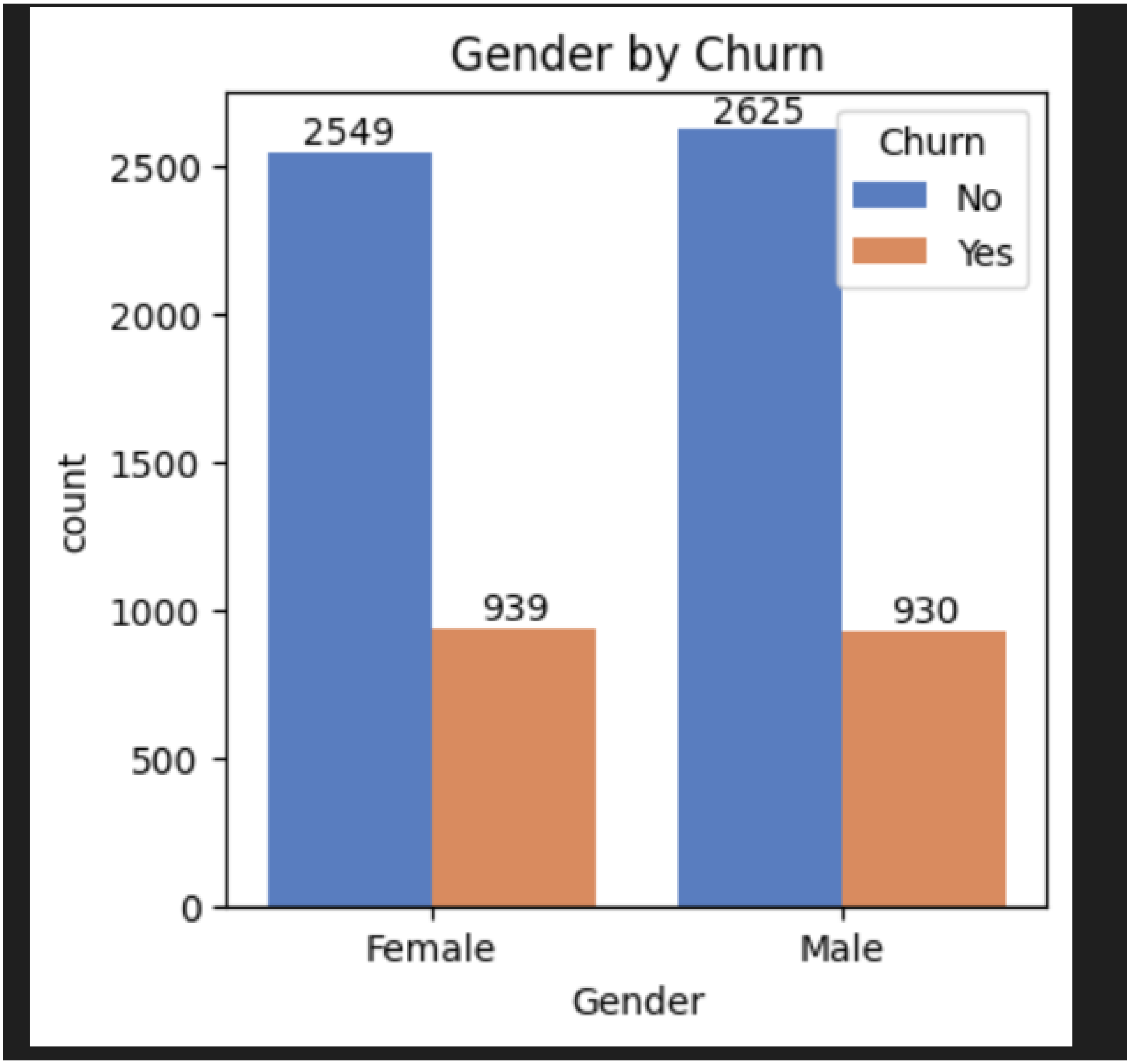
Strategic Takeaways:

Retention focus: Design special plans or support services tailored for senior citizens.

Customer experience: Provide easier onboarding, simplified billing, or senior-friendly customer service.

Distribution between Gender and Churn:

```
# Gender distribution
ax = sns.countplot(data=df,x='Gender',hue='Churn',palette='muted',legend=True)
ax.bar_label(ax.containers[0])
ax.bar_label(ax.containers[1])
plt.title("Gender by Churn")
plt.show()
```



Female customers:
Did not churn: 2,549
Churned: 939

Male customers:
Did not churn: 2,625
Churned: 930

The churn counts are very similar across genders, with no major difference in churn behavior between male and female customers.
Gender is not a strong predictor of churn: Both groups show nearly identical churn rates (around 26–27%).

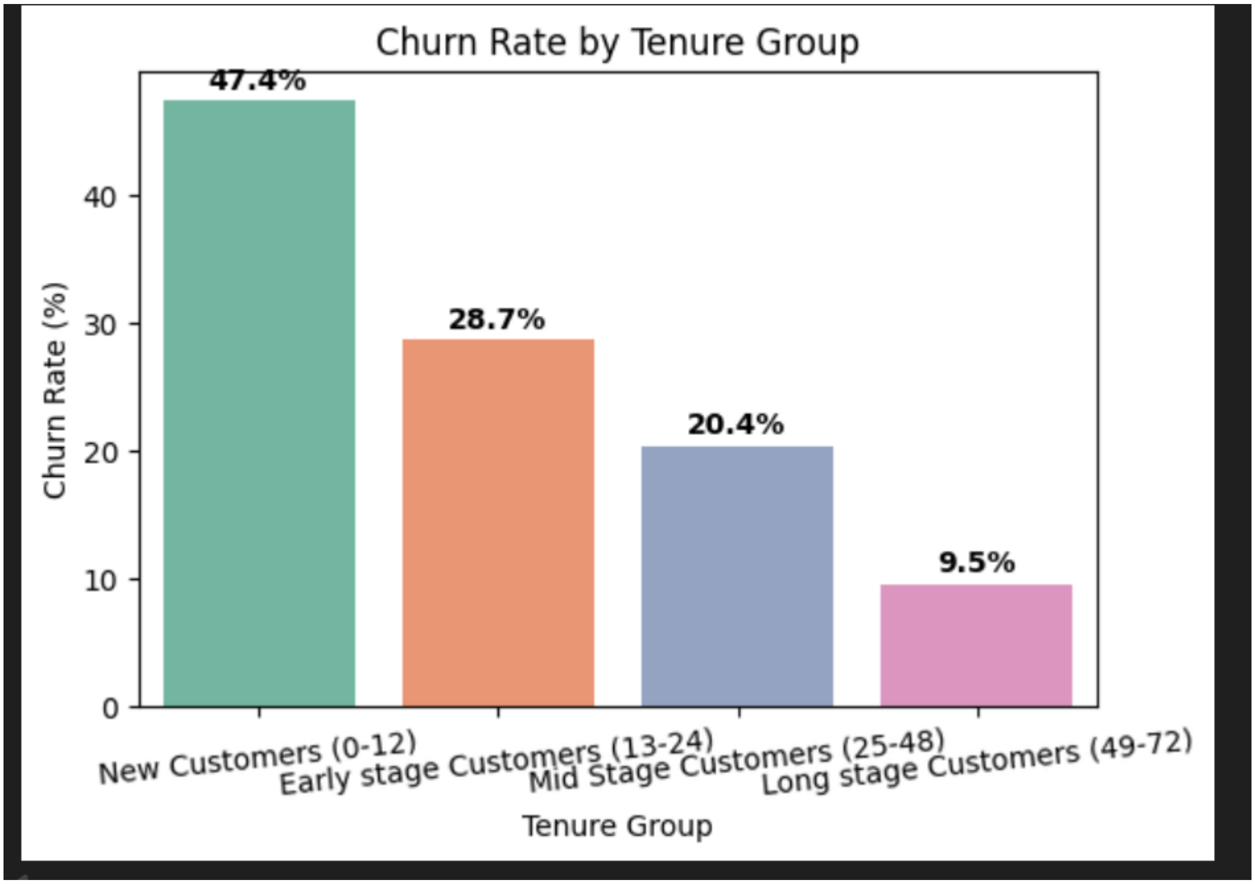
Calculate churn rate by tenure group:

```
# Churn rate by tenure group
tenure_stats = df.groupby('Tenure Group', observed=True).agg(
    Avg_MonthlyCharges=('Monthlycharges','mean'),
    Churn_Rate=('Churn', lambda x: (x=='Yes').mean()*100)
).reset_index()
```

```
# Churn Rate by Tenure Group
plt.figure(figsize=(8,6))
barplot = sns.barplot(data=tenure_stats, x='Tenure Group', y='Churn_Rate',
            hue='Tenure Group', palette='Set2', legend=False)

for index, row in tenure_stats.iterrows():
    barplot.text(index, row['Churn_Rate'] + 1, # position slightly above bar
                 f"{row['Churn_Rate']:.1f}%", # formatted label
                 ha='center', fontweight='bold', color='black')

plt.title("Churn Rate by Tenure Group")
plt.ylabel("Churn Rate (%)")
plt.xticks(rotation=6)
plt.show()
```



New Customers **(0–12 months): 47.4%**

Early-Stage Customers **(13–24 months): 28.7%**

Mid Stage Customers **(25–48 months): 20.4%**

Long Stage Customers **(49–72 months): 9.5%**

Key Insights:
Inverse relationship: The longer customers stay, the less likely they are to churn.
Critical period: The first year is the most dangerous stage—almost half of customers leave.
Retention payoff: Once customers pass the early stage, churn risk drops steadily, and long-term customers are the most stable.

Business implication:
Focus retention efforts on new and early-stage customers with onboarding, engagement, and loyalty programs.

Long-tenure customers are highly valuable—not only do they churn less, but as seen in your monthly charges chart, they also pay more.