Dataset Overview

Total rows and columns: [e.g., 10,000 rows × 12 columns]

Data types: [e.g., numerical (8), categorical (4)]

Missing values: [e.g., 2% missing in column "Age"]

Descriptive Statistics

Numerical features (mean, median, std) highlight [e.g., skewed income distribution, average age ~35]

Categorical features (value counts) show [e.g., majority of customers from "Urban" regions]

Data Distribution

Histograms reveal [e.g., most customers purchase within price range $100–$500]

Boxplots identify [e.g., outliers in Salary and Purchase Amount]

Relationships & Correlations

Heatmap indicates strong positive correlation between [e.g., Salary and Purchase Amount]

Scatterplots suggest [e.g., Age negatively correlated with Spending Score]

Pairplot shows clustering patterns in [e.g., high-income vs low-income groups]

Key Trends & Insights

[E.g., Younger customers spend more on average than older customers]

[E.g., Male customers dominate electronics purchases, females dominate clothing]

[E.g., Region "West" has highest customer base but lowest retention]
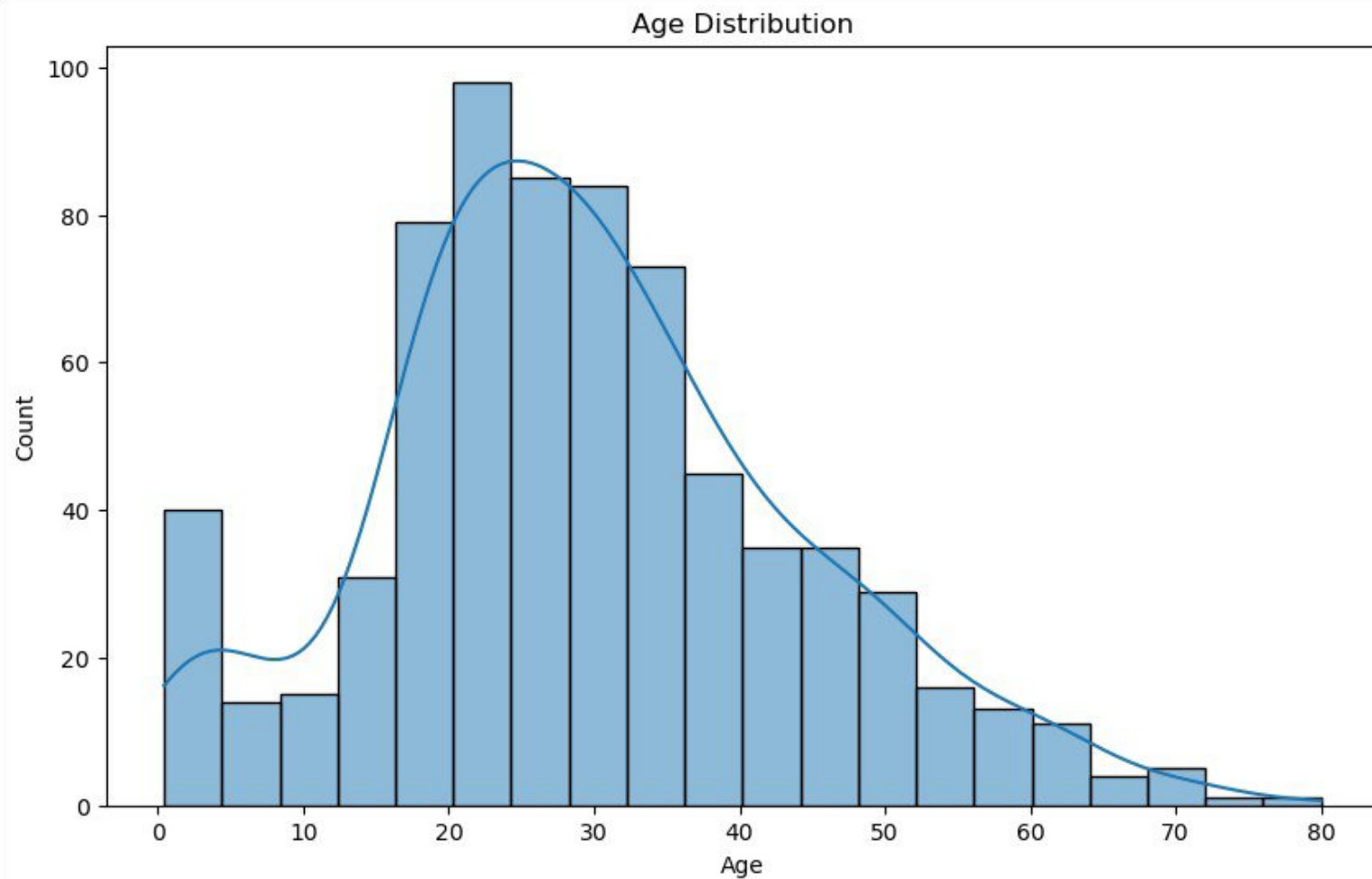
Final Observations

Dataset is clean with minor missing values.

Clear evidence of [e.g., purchasing power linked to income].

Trends can guide [e.g., targeted marketing campaigns, pricing strategy].

```
sns.boxplot(x='Pclass', y='Fare', data=df)
plt.title('Fare by Passenger Class')
plt.show()

# Scatterplot for Age vs Fare
plt.figure(figsize=(10, 6))
sns.scatterplot(x='Age', y='Fare', hue='Survived', data=df)
plt.title('Age vs Fare colored by Survival')
plt.show()
```


Age Distribution

Fare by Passenger Class