1. Project Title: Customer Churn Analysis & Insights

A complete, client-ready exploratory data analysis (EDA) of a telecom churn dataset.

2. Business Problem Statement

A telecom company wants to understand **why customers are leaving (churning)** and what factors drive churn. The goal is to: - Identify high-risk customer groups - Extract actionable business insights - Provide recommendations to reduce churn

3. Business Objectives

1. Perform complete data cleaning and validation
2. Conduct exploratory analysis to identify churn behaviour patterns
3. Visualize key trends clearly and professionally
4. Produce insights a business manager can act upon

4. Import Libraries

```
import pandas as pd
import numpy as np
import seaborn as sns
sns.set_theme(style="whitegrid", palette="deep")
import matplotlib
import matplotlib.pyplot as plt
import matplotlib_inline
```

5. Load the Dataset

```
df = pd.read_csv('/content/drive/MyDrive/data-analytics-portfolio-anurag/P1-Customer_Churn/data/raw_telco_custom
```

6. Dataset Overview

```
df.head()
df.info()
df.describe(include='all')
```

Show hidden output

Key Notes: - Categorical & numerical mix - TotalCharges often loaded as object - Missing values generally present in tenure-related fields

7. Data Cleaning Summary

Steps performed to clean the dataset:

1. Converted TotalCharges to numeric
2. Replaced blank values with NaN
3. Imputed or corrected invalid values
4. Validated tenure and charges consistency
5. Removed duplicates
6. Encoded categorical variables for later analysis

8. Clean TotalCharges Column

```
# Replace empty strings with NaN
df['TotalCharges'] = df['TotalCharges'].replace(" ", np.nan)

# Convert to numeric
df['TotalCharges'] = pd.to_numeric(df['TotalCharges'])
```

```
# Impute values where tenure > 0 but TotalCharges was missing
mask = df['TotalCharges'].isna() & (df['tenure'] > 0)
df.loc[mask, 'TotalCharges'] = df.loc[mask, 'tenure'] * df.loc[mask,
'MonthlyCharges']
```

9. Handle Missing Values

```
df.isna().sum()
# If any remain
df.fillna(method='ffill', inplace=True)
```

Show hidden output

10. Data Validation

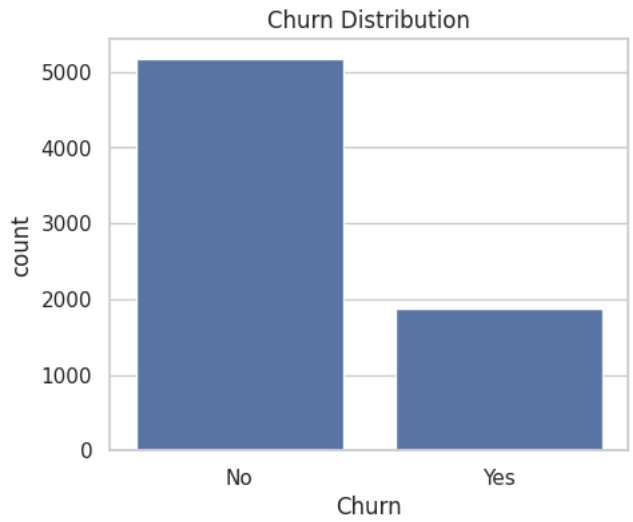
```
df.duplicated().sum()
```

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

11. Exploratory Data Analysis (EDA)

11.1 Churn Distribution

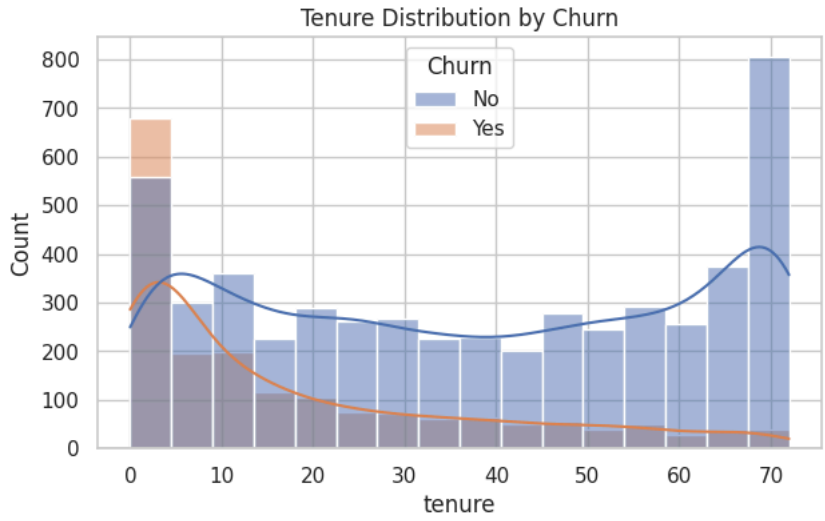
```
plt.figure(figsize=(5,4))
sns.countplot(x='Churn', data=df)
plt.title('Churn Distribution')
plt.show()
```



Insight: Most customers stay, but churn rate is significant (~26-28%).

11.2 Tenure vs Churn

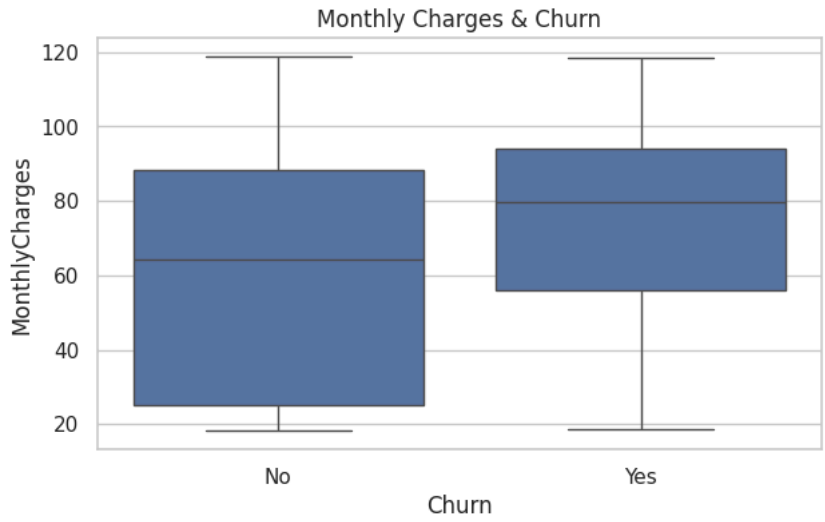
```
plt.figure(figsize=(7,4))
sns.histplot(data=df, x='tenure', hue='Churn', kde=True)
plt.title('Tenure Distribution by Churn')
plt.show()
```



Insight: New customers (low tenure) churn at a much higher rate.

11.3 Monthly Charges vs Churn

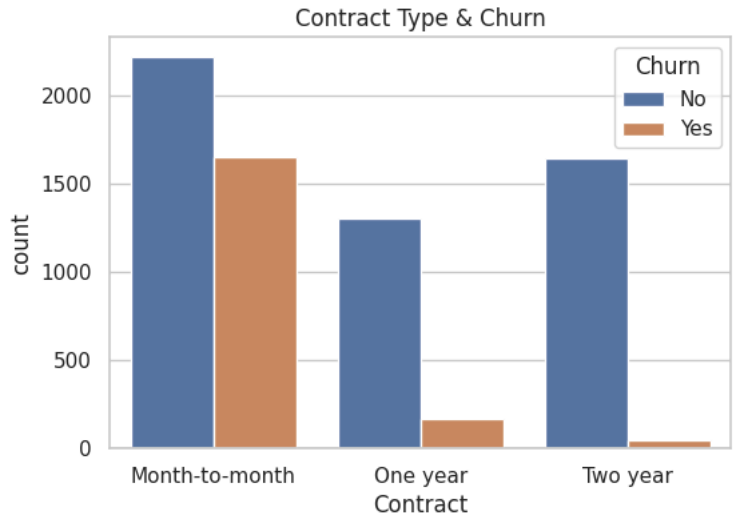
```
plt.figure(figsize=(7,4))
sns.boxplot(x='Churn', y='MonthlyCharges', data=df)
plt.title('Monthly Charges & Churn')
plt.show()
```



Insight: Customers paying higher monthly charges are more likely to churn.

11.4 Contract Type vs Churn

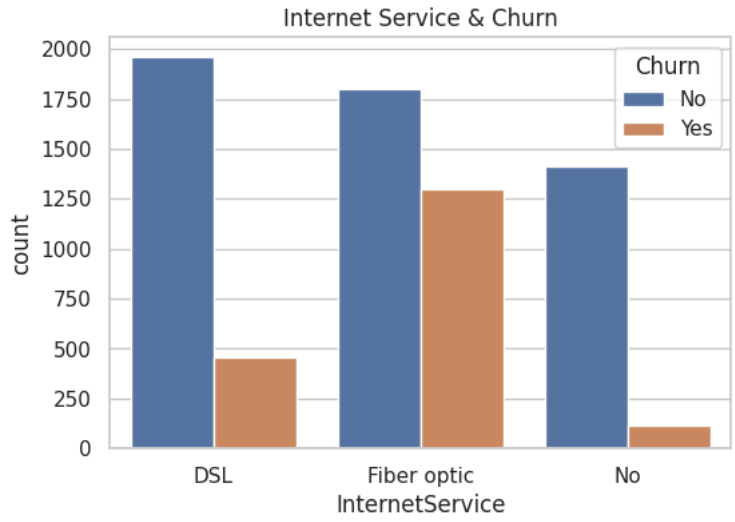
```
plt.figure(figsize=(6,4))
sns.countplot(x='Contract', hue='Churn', data=df)
plt.title('Contract Type & Churn')
plt.show()
```



Insight: Month-to-month contract customers churn heavily. Longer contracts retain customers.

11.5 Internet Service vs Churn

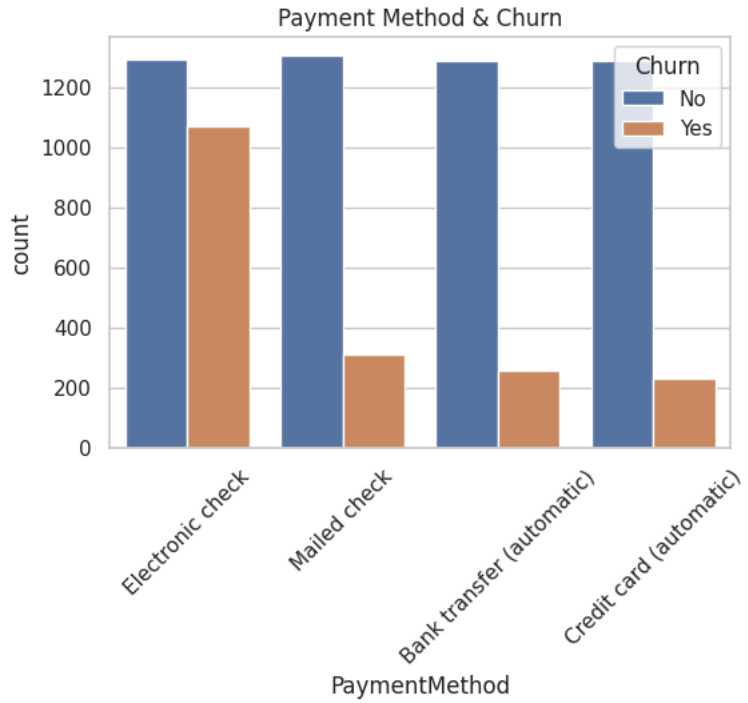
```
plt.figure(figsize=(6,4))
sns.countplot(x='InternetService', hue='Churn', data=df)
plt.title('Internet Service & Churn')
plt.show()
```



Insight: Fiber optic users churn the most — service reliability/price may be issues.

11.6 Payment Method vs Churn

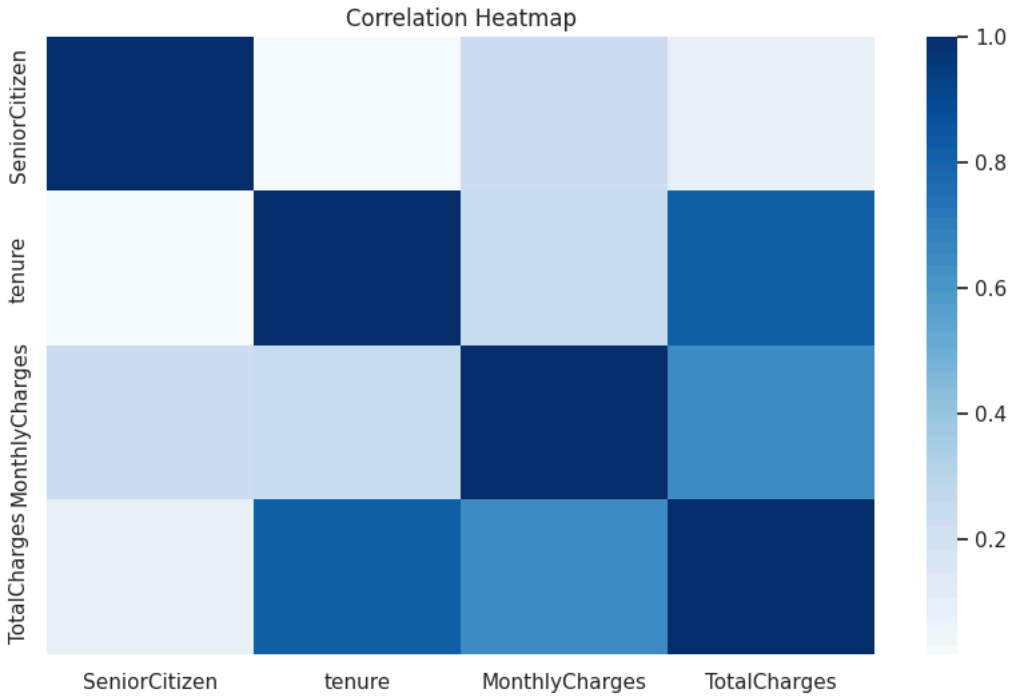
```
plt.figure(figsize=(6,4))
sns.countplot(x='PaymentMethod', hue='Churn', data=df)
plt.xticks(rotation=45)
plt.title('Payment Method & Churn')
plt.show()
```



Insight: Electronic check customers have the highest churn.

12. Correlation Heatmap

```
plt.figure(figsize=(10,6))
corr = df.corr(numeric_only=True)
sns.heatmap(corr, annot=False, cmap='Blues')
plt.title('Correlation Heatmap')
plt.show()
```



Insight: Monthly charges & tenure have the strongest relationship to churn.

13. Key Insights Summary

1. **Month-to-month customers churn the most** → lock-in offers can help.
2. **Fiber optic customers churn at much higher rates** → indicates dissatisfaction.

3. **Electronic check users churn the most** ⟶ risky customer segment.
4. **Low tenure = high churn** ⟶ early onboarding improvements needed.
5. **High monthly charges drive churn** ⟶ pricing strategy may need review.
6. **Customers lacking tech support/security churn more** ⟶ bundle offers recommended.

14. Business Recommendations

1. Improve first-month onboarding

Reduce early churn by targeted onboarding for new customers.

2. Incentivize longer contracts

Offer discounts or benefits for 1-year or 2-year contracts.