Movie Dataset Analysis & Visualization Project

This project is designed to help students practice data analysis and visualization in Python using a movie dataset. The dataset contains various features such as budget, revenue, ratings, votes, genre, and more. Students are required to explore the dataset, analyze patterns, and visualize findings.

1. Basic Exploration

- Load the dataset and check its shape, column names, and data types.
- Find the number of unique movies, genres, directors, and lead actors.
- What are the earliest and latest release years in the dataset?
- Identify missing values, if any, and suggest ways to handle them.
- Show summary statistics (mean, median, min, max, std) for numerical columns like Budget, Box Office, IMDb rating, etc.

2. Univariate Analysis

- Plot the distribution of movie release years (Histogram).
- Which are the top 5 most common genres? (Bar chart).
- Plot the distribution of IMDb ratings (Histogram/Density plot).
- Show the distribution of Rotten Tomatoes scores across movies (Histogram).
- What's the average budget of movies by country? (Bar chart).

3. Bivariate Analysis

- Is there a correlation between Budget and Global Box Office revenue? (Scatter plot).
- Compare the average IMDb rating across genres (Bar/Box plot).
- Do higher budgets lead to better Rotten Tomatoes scores? (Scatter/Line plot).
- Compare US vs Global Box Office revenue — are some genres more dependent on international sales?
- Relationship between Opening Day Sales and First Week Sales (Scatter plot).

4. Time-Based Analysis

- How has the number of movies released per year changed over time? (Line chart).
- Has the average movie budget increased over the decades? (Line/Bar chart).
- Do IMDb ratings show any trend across release years? (Line chart).
- Which genre has grown most in popularity since 2000? (Line/Bar chart).
- Compare average Global Box Office revenue of movies before 2000 vs after 2000 (Bar chart).

5. Categorical Insights

- Which director has the highest average IMDb rating? (Top 10 directors bar chart).
- Which lead actors have appeared in the most movies? (Bar chart).
- Which country produces the most movies? (Bar chart).
- Which genre tends to have the highest US vs Global box office ratio?

- Compare IMDb and Rotten Tomatoes ratings for each genre — are they aligned? (Side-by-side bar chart).

6. Advanced / Critical Thinking

- Do higher IMDb ratings always mean higher box office revenue? (Correlation test).
- What percentage of movies had a budget above $100M but failed to make a profit (Global revenue < Budget)?
- Which genres are most 'profitable' (average Global Box Office / Budget ratio)? (Bar chart).
- Does a higher Rotten Tomatoes score guarantee higher first-week sales? (Scatter plot).
- Predictive thinking: Based on budget, genre, and director, what factors seem most important for a movie’s success?

7. Visualization Ideas

- Heatmap of correlations among numeric variables (Budget, Revenue, Ratings, Votes).
- Bubble chart of Budget vs Global Revenue, with bubble size = IMDb votes, color = Genre.
- Trend line of IMDb rating vs Release Year.
- Pie chart or Treemap of genre distribution.
- Box plot of IMDb rating by Genre.