Movies and their performance on the Box Office

TECHNICAL PROJECT REPORT

Submitted to:

Department of IT

Phoenix College of Management and IT

In partial fulfillment of the requirements for the
Extra Course Internal Evaluation

Submitted by

Student Name : Kunal Bhandari

BIT 2nd Semester

University ID: LC0003002665

2025/09/20

Table of Contents

---

1. Introduction to Dataset

- The Movie Data Science Project dataset contains information on **10 popular movies**.
- Variables include:
    - Movie (name),
    - Genre (category, e.g., Action, Sci-Fi, Romance, etc.),
    - BudgetMillion (budget in millions),
    - RevenueMillion (box office revenue in millions),
    - Rating (IMDb rating).
- Used for exploratory analysis and insights into movie success factors.

---

2. Data Overview

- **Shape:** 10 rows, 5 columns.
- **Variable Types:**
  - Numerical: BudgetMillion, Revenue Million, Rating
  - Categorical: Movie, Genre
- **Descriptive Statistics:**
  - Average budget: 144.7M144.7M144.7M
  - Average revenue: 1275.9M1275.9M1275.9M
  - Average rating: 8.288.288.28
  - No missing values detected.
- **Genre Distribution:**
  - Action, Sci-Fi, Animation, Romance, Drama.

---

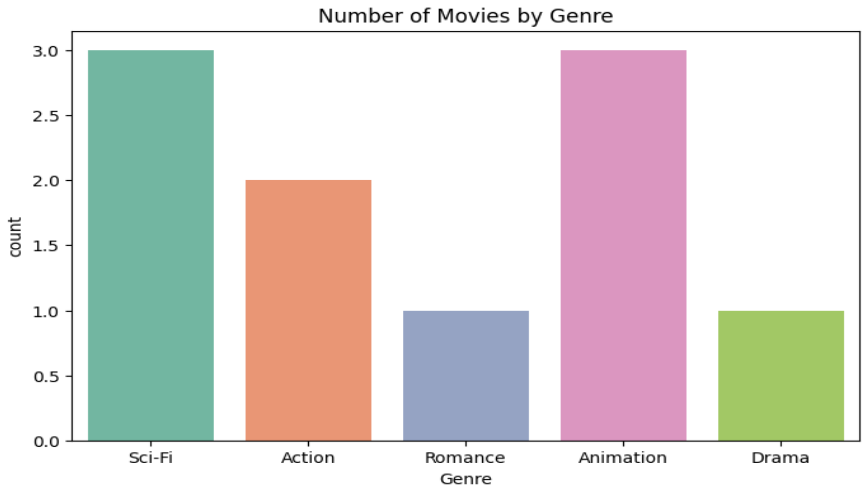
3. Data Quality

- **Completeness:** No missing values detected.
- **Duplicates:** None present.
- **Type:** Original.
- **Outliers:**
    - Revenue: Extremely high (e.g., Avatar, Titanic).
    - Ratings: All above 7, The Dark Knight exception at 9.
- No major inconsistencies, though sample size is small.
- Creation: Created by Kunal Bhandari as it the list of his favorite movies.

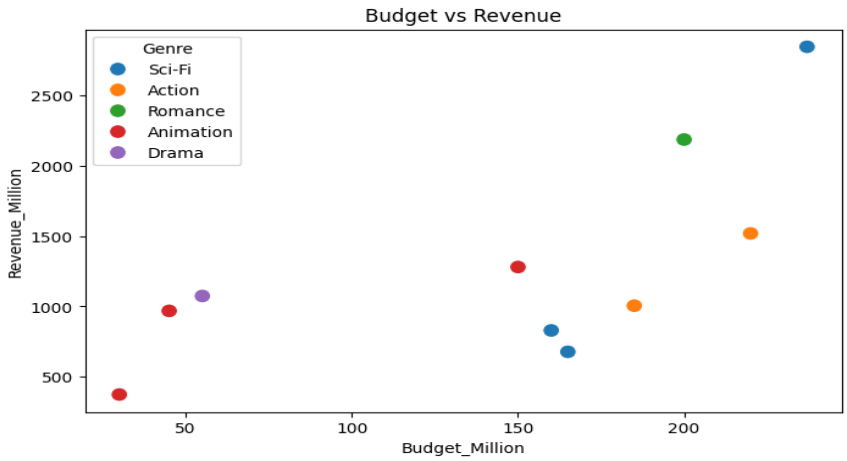
---

4. Visualization Overview

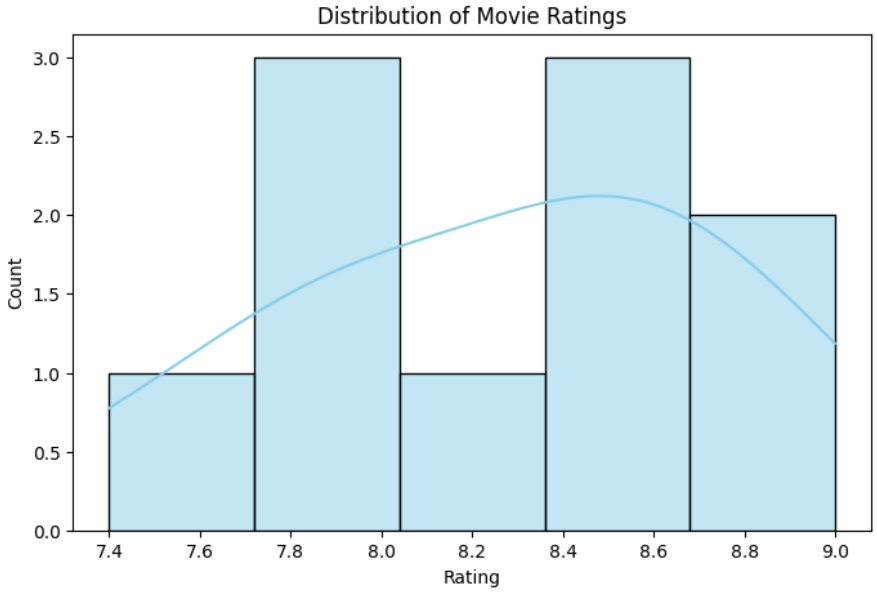
- **Genre Popularity:**
  - Countplot: Highlights Animation and Sci-Fi as frequent genres in high-grossing movies.



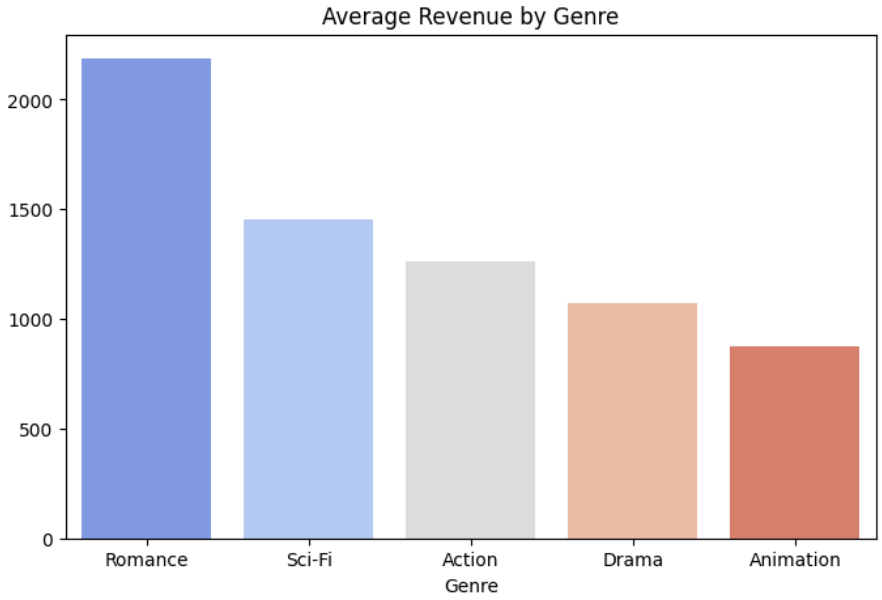
- **Budget vs Revenue:**
  - Scatterplot: Shows that higher budgets often correlate with higher revenue, but not always with higher ratings.



- **Rating Distribution:**
  - Histogram: Most movies rated above 8, indicating strong critical performance.



- **Average Revenue by Genre:**
  - Barplot: Action and Sci-Fi genres show highest average revenues.



---

5. Feature Engineering

- **Encoding:**
  - One-hot encoding suitable for Genre if used for further machine learning.
- **Binning:**
  - Rating can be grouped into categories (High, Medium, Low) for interpretability.
- **Redundancy:**
  - No constant or unused columns detected.

---

6. Key Findings and Conclusion

- **Insights:**
  - Action and Sci-Fi movies tend to generate higher revenues.
  - Animation movies also perform strongly, particularly in revenue relative to budget.
  - High budget does not guarantee higher ratings.
  - All movies analyzed are rated well (above 7), with notable outlier performances (The Dark Knight).
- **Conclusion:**
  - The dataset provides an overview of popular, high-performing movies.
  - Further analysis with larger datasets may yield deeper insights into genre trends and success factors.
  - Feature engineering and visual exploration offer actionable perspectives for understanding movie industry patterns.

GitHub Repo link: https://github.com/Anyonomous1/Python-Data-Science/blob/main/Movie_DataScience_Project.ipynb