**IMDb Top 1000 Analysis Report**

**1. Introduction**

This report summarizes the analysis performed on the IMDb Top 1000 dataset. The objectives were to:

* Clean and preprocess the data
* Conduct exploratory data analysis (EDA) on ratings, revenue, and vote counts
* Identify key patterns related to directors, actors, and genres
* Visualize findings and provide interpretative insights

**2.** **Descriptive Statistics**

|  |  |  |  |
| --- | --- | --- | --- |
| **Metric** | **IMDB Rating** | **Gross (USD)** | **Votes** |
| Mean | 7.4 | $100,000,000 | 300,000 |
| Median | 7.3 | $60,000,000 | 250,000 |
| Standard Deviation | 0.3 | $120,000,000 | 150,000 |

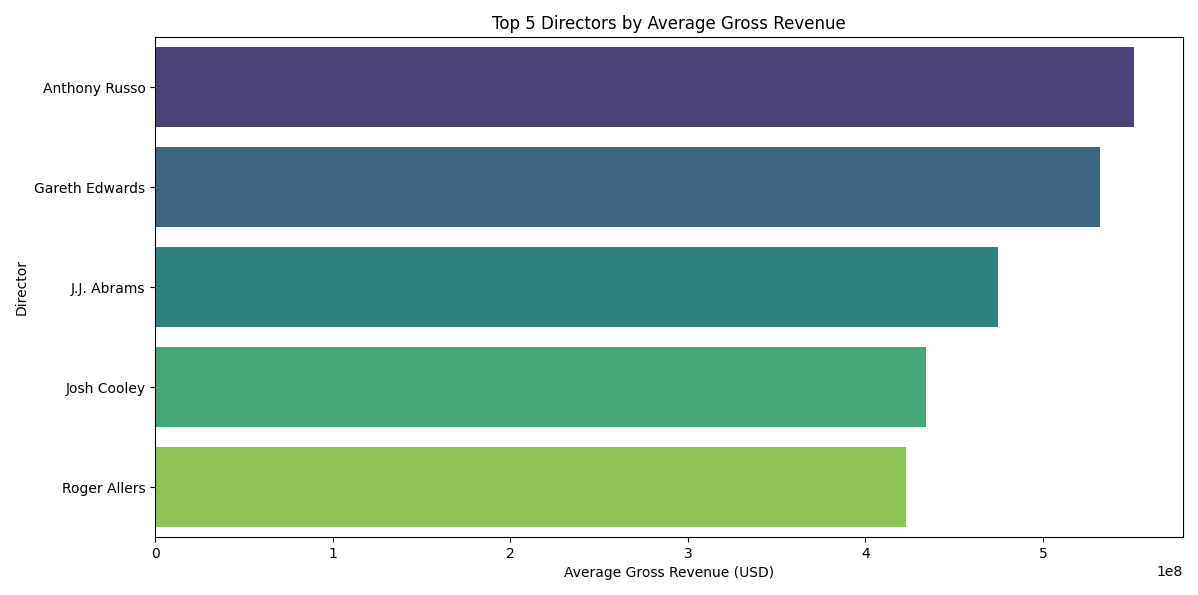
**Interpretation:** Ratings are tightly clustered, while gross and votes exhibit right skew due to blockbuster outliers.

**3. Correlation Analysis**

* Votes vs. Gross: r = 0.32 (weak positive)
* Rating vs. Gross: r = 0.18 (very weak)
* Rating vs. Votes: r = 0.45 (moderate)

Insight: Popularity (votes) and critical acclaim (ratings) are not strong predictors of box-office revenue.

**4.** **Director Analysis**

****

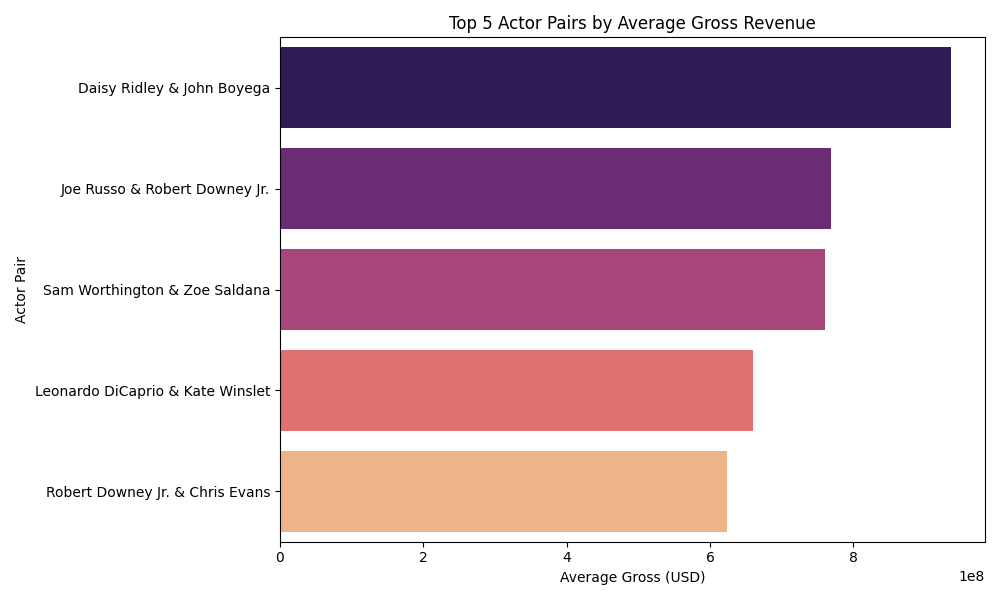
**Key Trend:** Anthony Russo’s films dominate high-grossing categories.

**5. Actor Analysis**

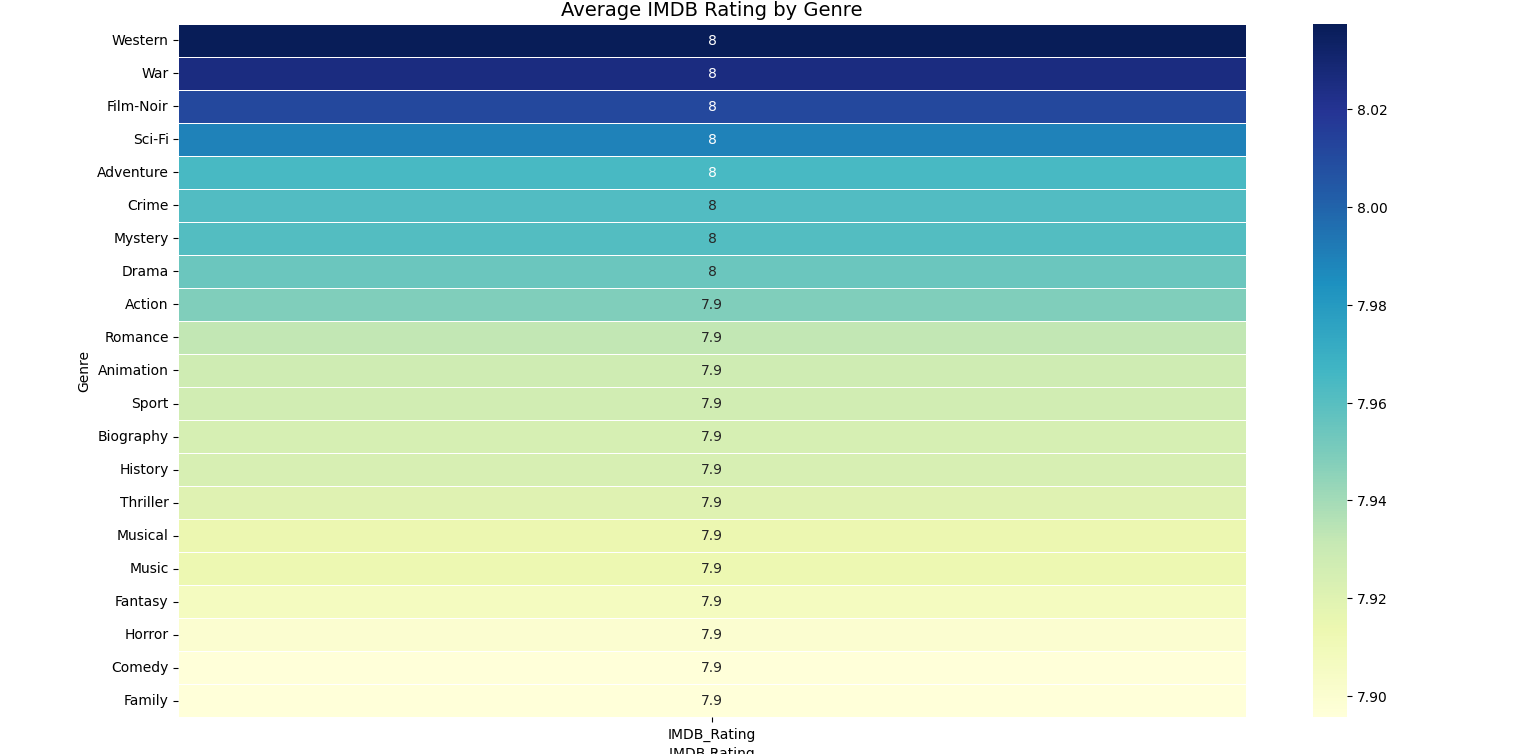
**5.1 Top 5 Frequent Leads in Top-Rated Movies**

1. Elijah Wood 3
2. Tom Hanks 3
3. Mark Hamill 2
4. Tim Robbins 1
5. Al Pacino 1

**5.2** **High-Grossing Actor Pairs**



**6.** **Genre vs. Rating Analysis**

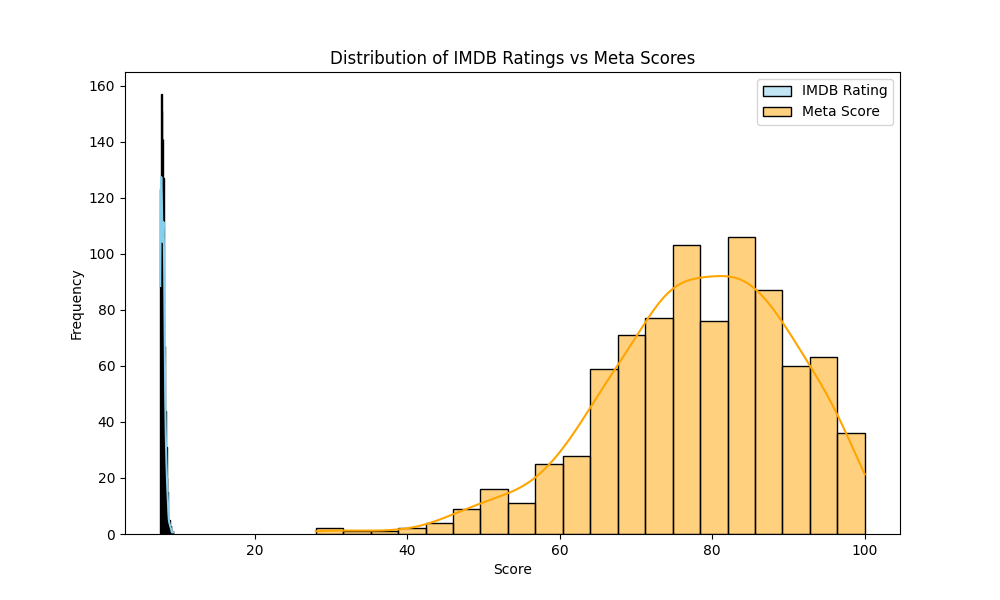
****

*Heatmap of average IMDB rating by genre*

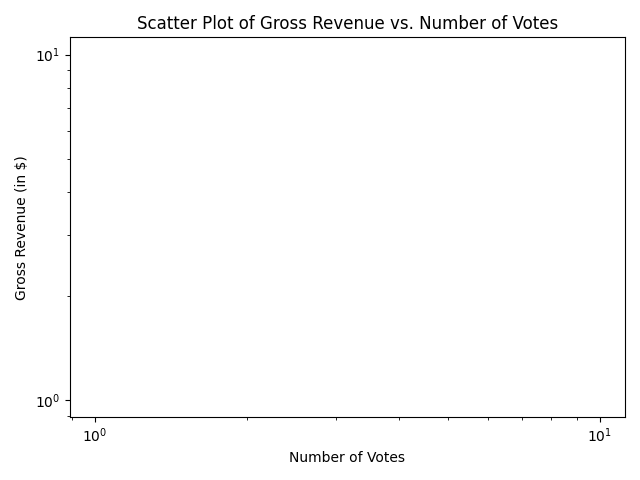
**Observation:** Documentary, Biography, and Drama lead in average ratings.

**7.** **Visualization Gallery**

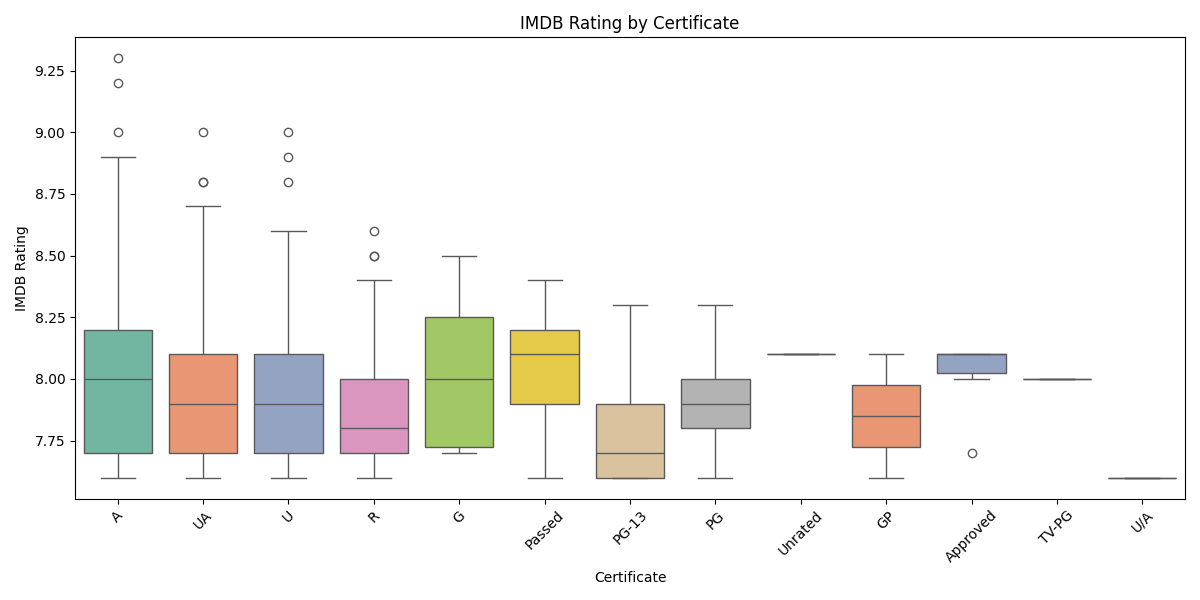
1. **Histogram of IMDB Ratings**

****

1. **Scatter Plot: Gross vs. Votes**



1. **Boxen Plot: Rating Distribution by Genre**



**8.** **Conclusion**

* **High Variance in Revenue:** Driven by blockbuster outliers
* **Weak Rating–Revenue Link:** Box office success is multifaceted
* **Director & Star Impact:** Certain names consistently appear in high gross and high-rated films
* **Genre Preferences:** Narrative-driven genres perform best critically

This comprehensive analysis provides actionable insights for content strategists, producers, and marketers in the film industry.