

Exploring Movie Data to Guide Investment Decisions

Phase 2 Data Science Project

By: Stephen Munene Gachingu

July 2025

Project Goal

- ▶ Analyze Box Office Mojo & IMDb data to understand what drives movie success.
- ▶ Business Goal: Help stakeholders decide which films to invest in for high Return On Investment.

Data Sets Used

- ▶ IMDb: Contains movie metadata like genres, ratings, runtimes, and more.
- ▶ Box Office Mojo: Provides data on gross earnings, budgets, and release years.
- ▶ Combined using SQL queries and pandas for analysis.

Data Cleaning Steps Used

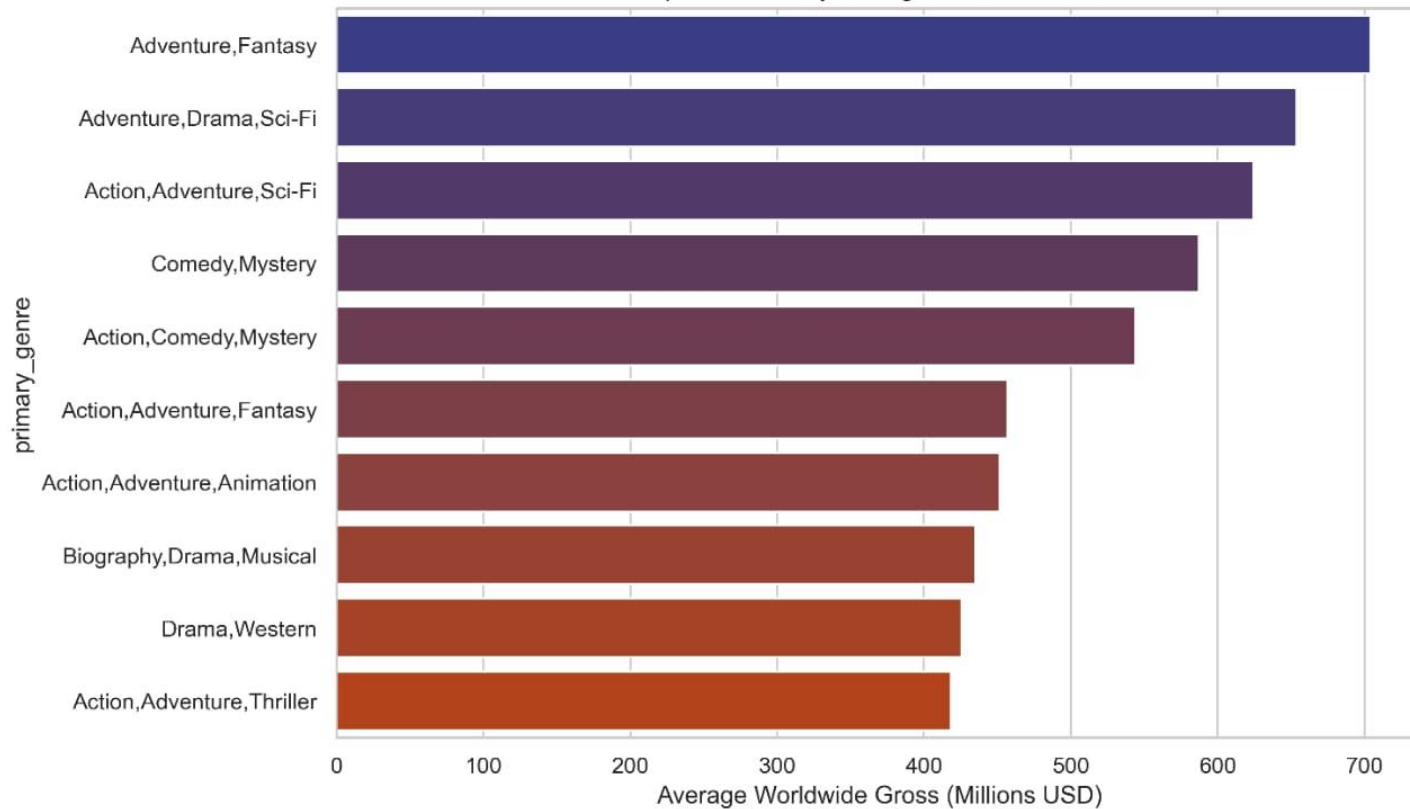
- ▶ Removed nulls and duplicates
- ▶ Filtered unrealistic or missing budgets/gross values
- ▶ Converted data types (e.g., strings to integers, dates)
- ▶ Merged datasets using IMDb IDs

Key insights I got from my EDA

- ▶ Action, Adventure, and Animation films tend to earn higher gross.
- ▶ Budgets are positively correlated with gross earnings.
- ▶ Runtime shows a sweet spot around 100-120 minutes.
- ▶ Movies with IMDb rating above 7 often perform better at the box office.

Key Insight 1: Genre & Revenue

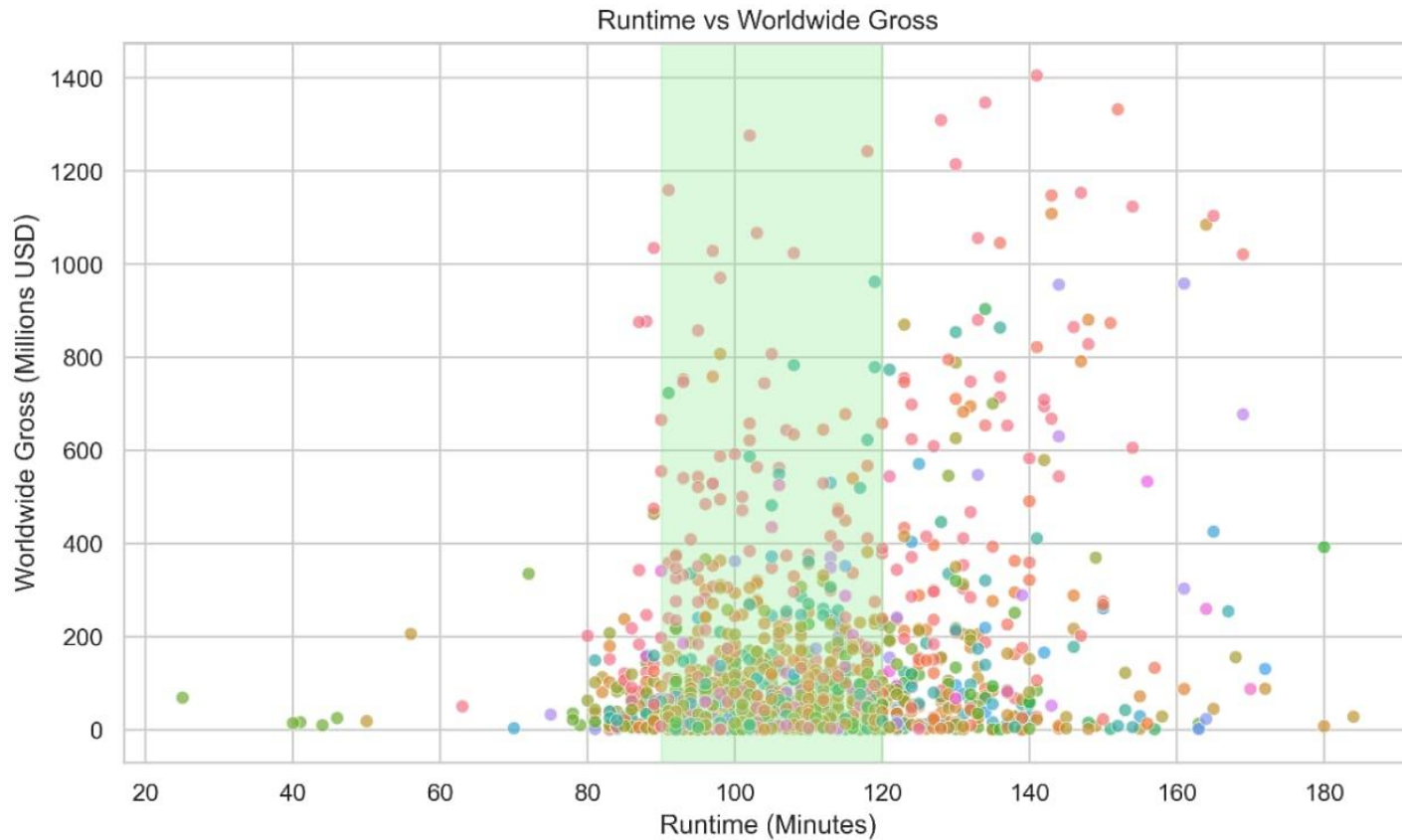
Top 10 Genres by Average Worldwide Gross



Genre & Revenue *(Cont.)*

- ▶ Top genres: **Action, Adventure, Animation and Sci-Fi**
- ▶ The above have higher revenues
- ▶ Implication: **Prioritize these genres for higher box office potential and higher revenue margins.**

Key Insight 2: Runtime & Revenue



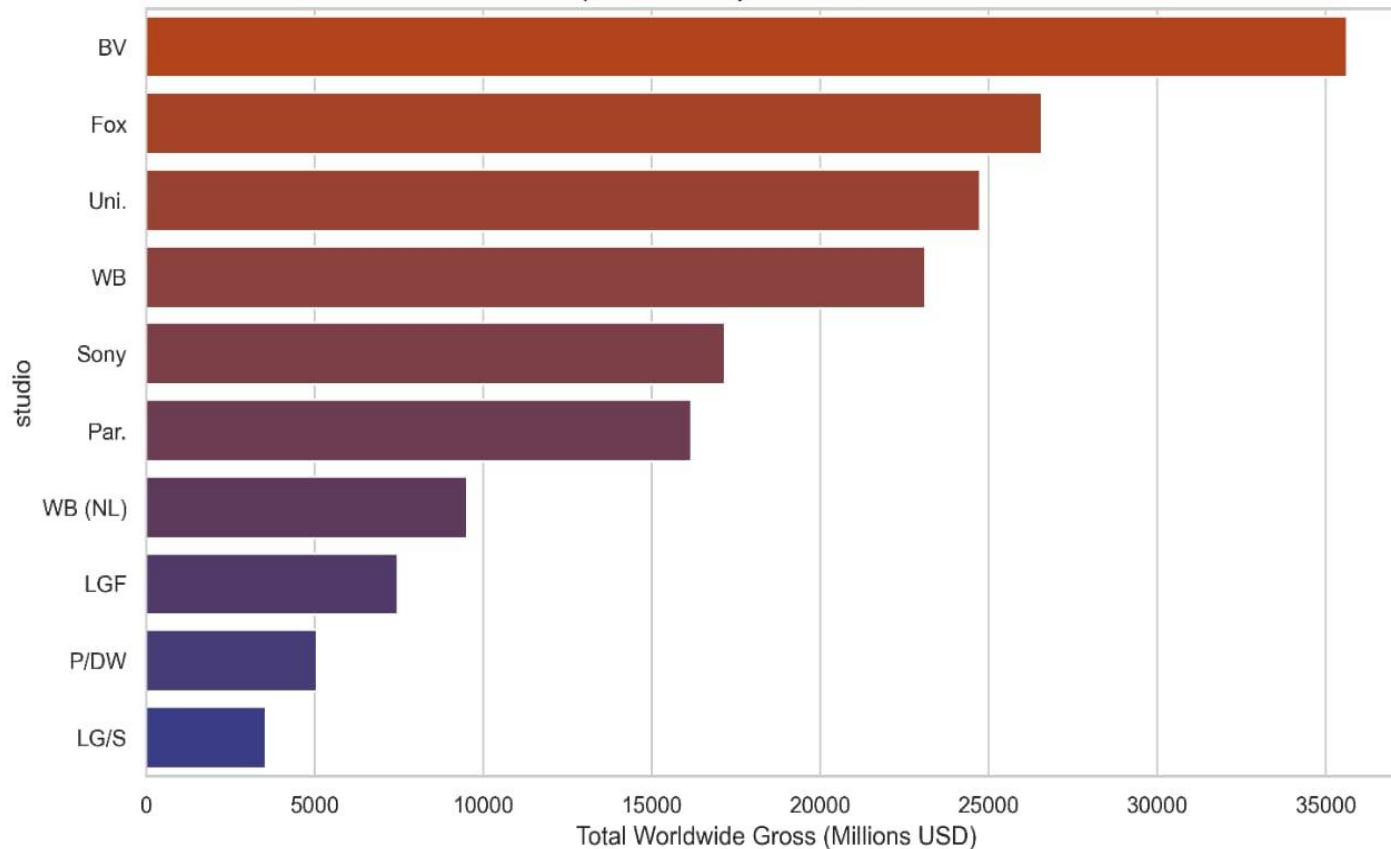
Runtime & Revenue *(Cont)*

- ▶ Sweet spot: 90-120 minutes
- ▶ Implication: Avoid overly long or short movies.

Key insight 3:

Studio & Revenue

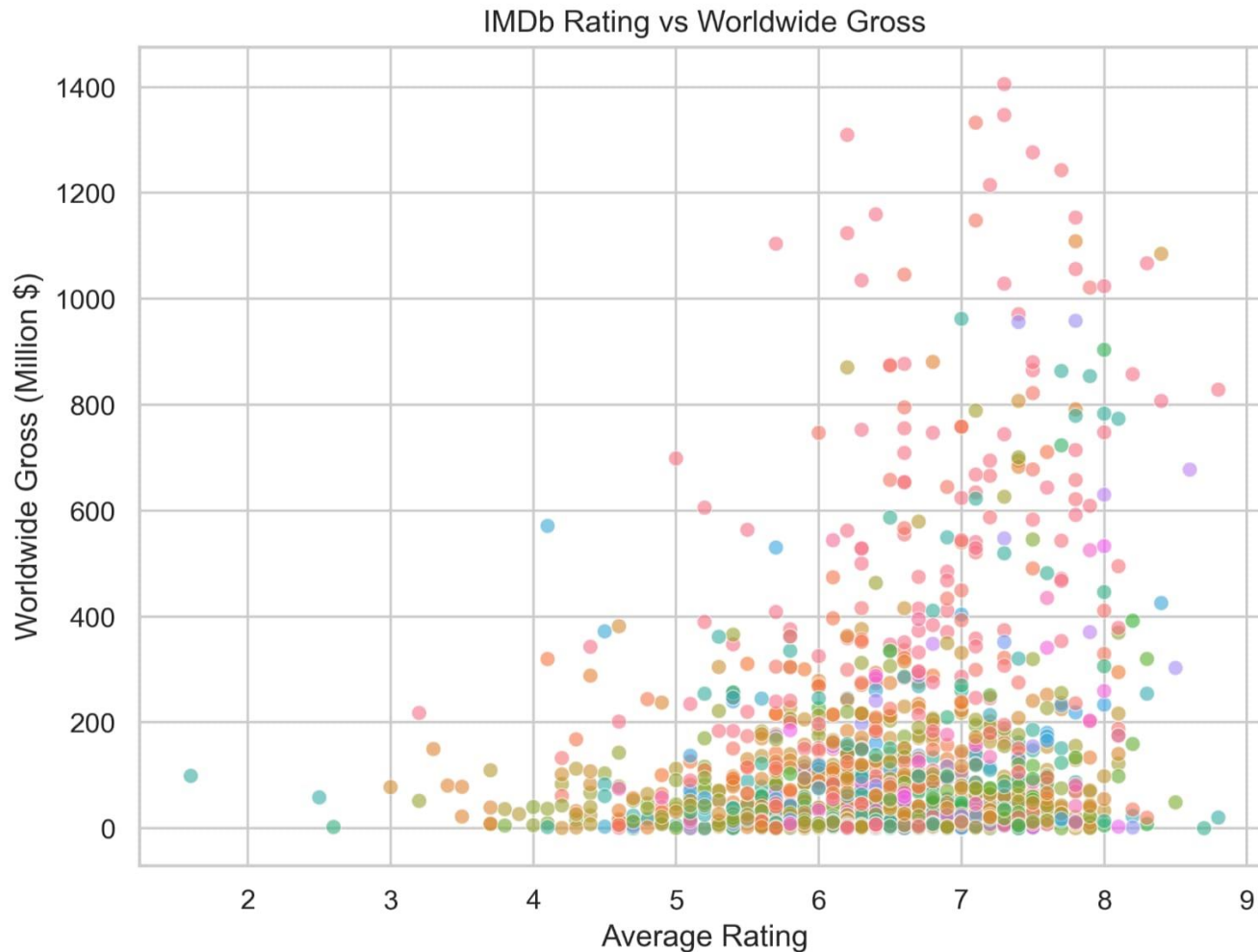
Top 10 Studios by Total Worldwide Gross



Studio & Revenue (Cont.)

- ▶ Top studios you should use to make your movies are **BV, FOX, WB and P/WD**
- ▶ Implications: Avoid unpopular studios that are not well at the Top to make your movies.

Key Insight 4: IMDb Rating & Revenue



Key Insight 4

IMDb Rating & Revenue

- ▶ Loose positive trend
- ▶ Implication: Audience perception matters; Ratings affect turnout.

Business Recommendations

- ▶ Invest in high-budget Action or Adventure films with strong scripts.
- ▶ Prioritize films with runtimes between 100-120 minutes.
- ▶ Use IMDb rating and genre combinations to predict success.
- ▶ Avoid overspending on genres like Drama or Documentary unless niche.

Other Insights...

- ▶ Incorporate more granular audience demographic data
- ▶ Use machine learning to predict movie success
- ▶ Expand to streaming performance data
- ▶ Continue refining feature engineering and visualizations

Thank You!

- ▶ Questions & Answer

- ▶ GitHub:

- <https://github.com/GACHINGU/movie-box-office-analysis>