Data-Driven Insights for Box Office Success— Phase 2 GRP 2 Project

Problem Statement

❖**Context:** The company is launching a new movie studio to enter original video content, aiming to compete with successful industry players and generate box office revenue

❖**Problem:** The company has no internal experience or knowledge of movie production and lacks clarity on which types of films perform best at the box office, creating uncertainty in investment decisions

❖**Objective:** Analyse box office trends to identify top-performing film types and provide actionable guidance on which films the studio should produce

Business Understanding

❖**The Goal**: Maximizing Worldwide Gross

❖**The Challenge**: A new studio must strategically allocate its production budget to minimize risk and maximize returns in a highly competitive global market

❖**Critical Success Metric:** Total Worldwide Gross Revenue (Domestic + Foreign) in millions of USD.

Data Understanding & Acquisition

Data Source	Content	Purpose	Data Source
IMDB Database (im.db)	Movie Metadata (Runtime, Genres, Title), Audience Ratings (Average Rating, Votes).	The 'What': Provides film characteristics and audience perception.	IMDB Database (im.db)
Box Office Mojo (bom.movie_gross.csv)	Financial Gross Data (Domestic Gross, Foreign Gross).	The 'How Much': Provides the revenue outcome.	Box Office Mojo (bom.movie_gross.csv)

Data Preparation

🧩 **Data Integration-**Datasets were merged using movie titles and unique IDs

💰 **Revenue Processing -** Worldwide revenue was calculated and standardized to Millions (USD)

🎭 **Genre Engineering-Multi-genre films were exploded to enable individual genre analysis**

🚫 **Data Filtering-**Films with zero reported revenue were excluded

⏱️ **Missing** Values HandlingMissing runtimes were imputed using the median-

Data Analysis - Key Findings

1. Genre vs. Revenue

❖ Top 3 Genres by Average Revenue: Sci-Fi, Adventure, and Animation

❖ Statistical Validation (ANOVA):The difference in average revenue between major genres is statistically significant ($p < 0.05$). The high-ranking genres are reliably better financial bets, not just due to random chance

2. Runtime Analysis

❖ The single most profitable 10-minute segment is 180–200 minutes

❖ The overall commercial "sweet spot" is broadly between140 and 200 minutes

3. Rating Correlation

❖ Correlation (R) between IMDB Rating and Revenue is {0.17}.

❖ This represents a very weak positive correlation, meaning a better rating offers only a negligible, unreliable boost to box office gross
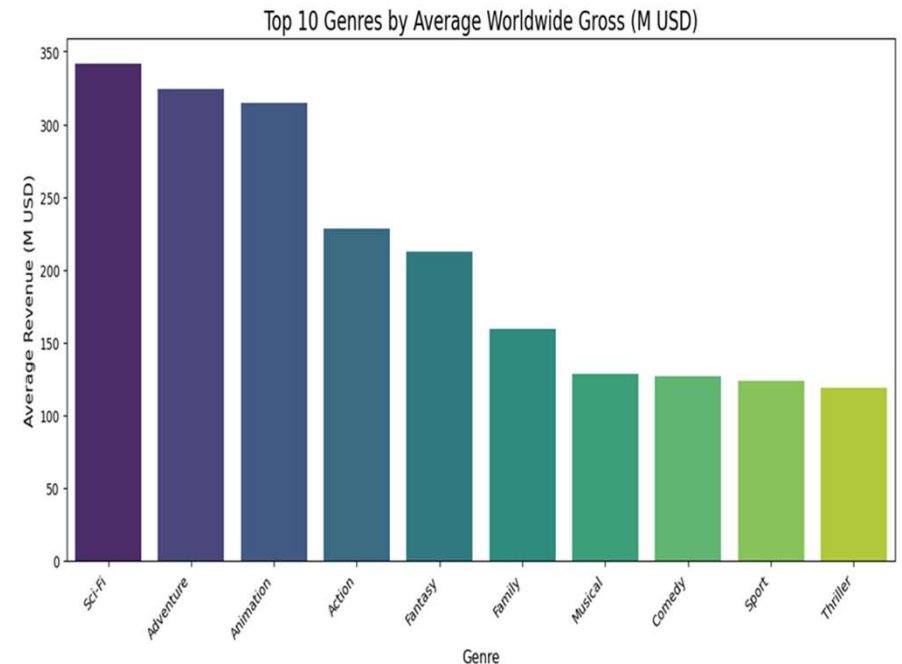
Top 10 Revenue-Generating Genres

Goal: Identify the statistically proven genres for highest average returns

Observed outcomes:

❖Sci-Fi and Adventure lead by a significant margin

❖Animation is the third-highest-performing genre

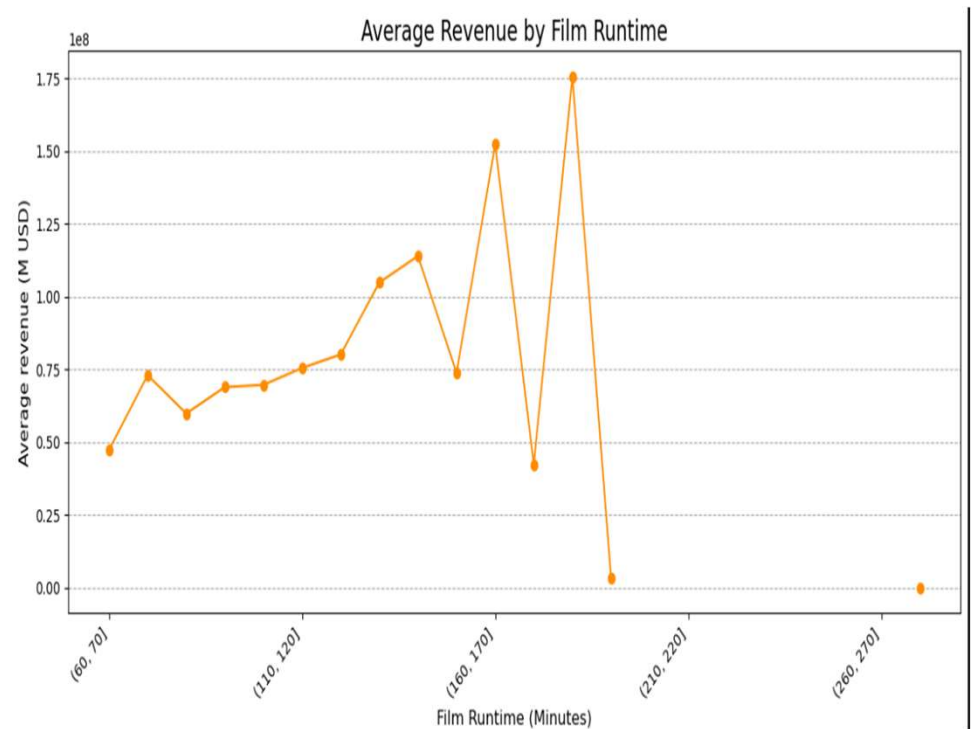❖Action and Comedy, while frequently produced, fall well outside the top 3 by average return

Top 10 Genres by Average Worldwide Gross (M USD)

Optimal Runtime for Maximum Revenue

Goal: Determine the commercially optimal length for films

Observed outcomes:

❖Revenue generally increases with runtime until the 180–200 minutes segment, where it peaks

❖Films significantly over 200 minutes see a sharp decline in average revenue, likely due to fewer daily screening slots
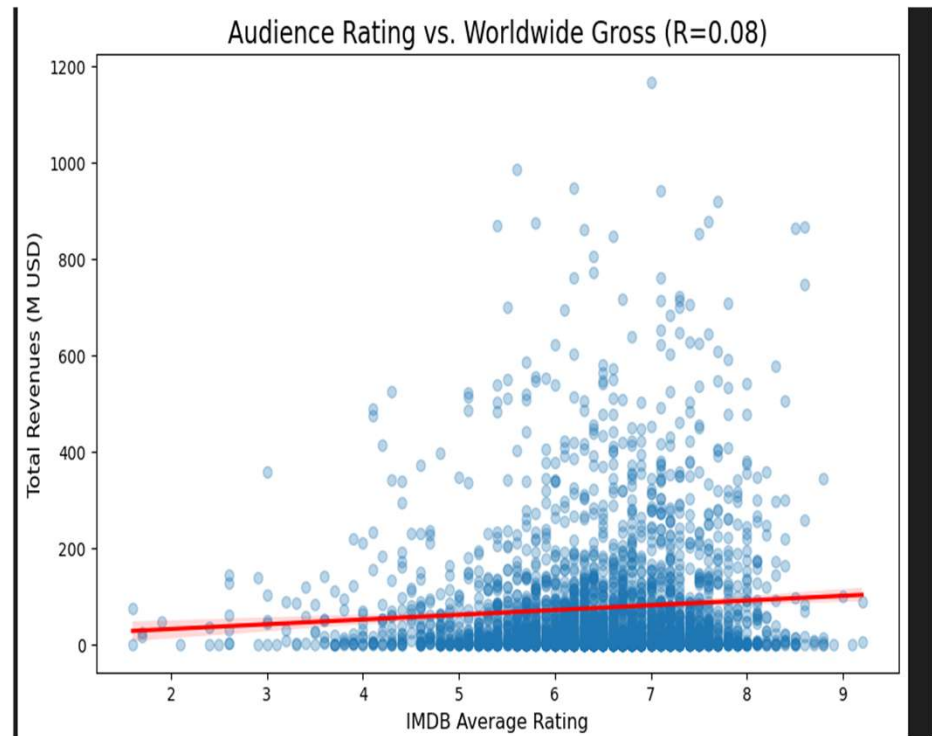
Audience Rating vs. Worldwide

Goal: Assess the predictive power of audience acclaim on financial success.

Observed outcomes:
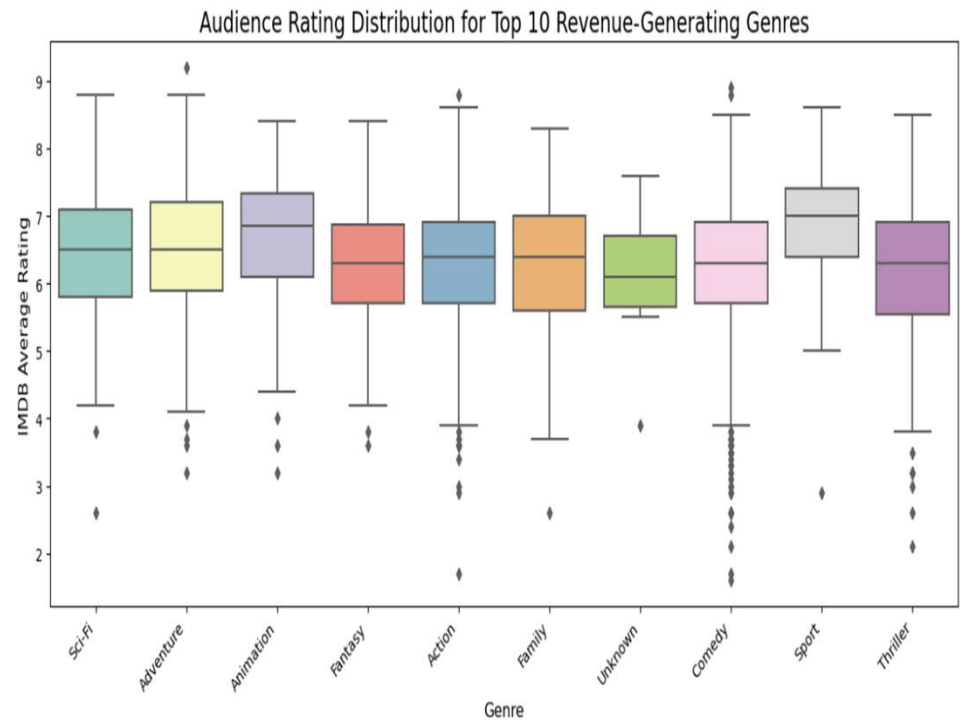
❖ The scatter plot shows a wide, dispersed cloud of data points, confirming the weak correlation (R = 0.17).

❖ Many of the highest-grossing films (top of the Y-axis) have only average ratings (e.g., 6.0 - 7.0), not necessarily top ratings.

Rating Consistency

Goal: Assess the predictive power of audience acclaim on financial success

❖ High profitability does not always correspond with consistent audience acclaim. Fantasy and Adventure genres show wide variability in ratings, indicating higher critical risk despite strong revenues

❖ Animation films achieve strong profitability with more consistent audience ratings. This suggests that audience acclaim alone is a weak predictor of financial success
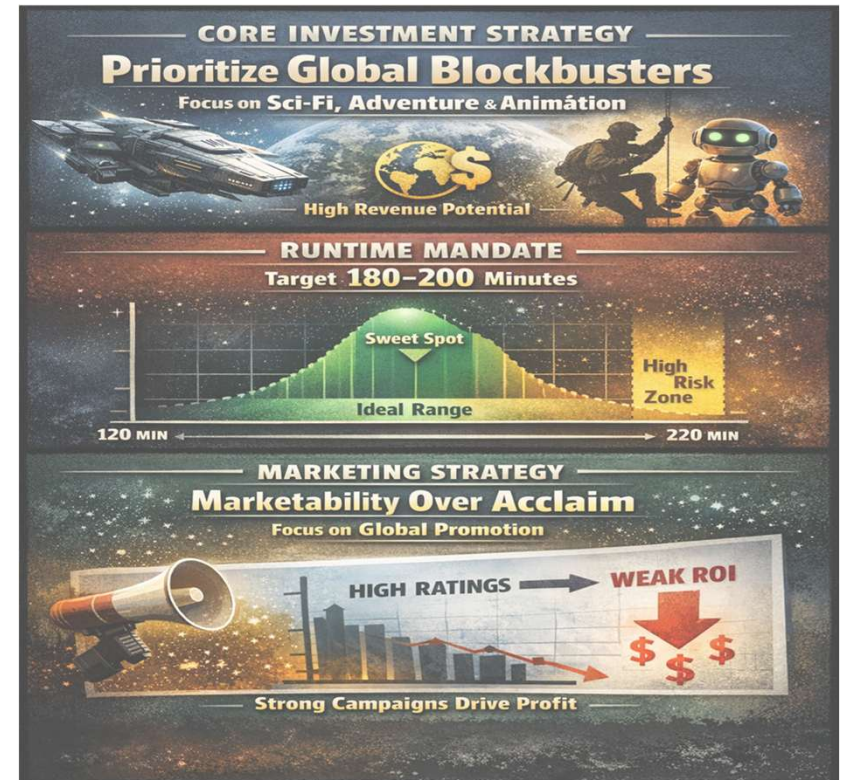

Audience Rating Distribution for Top 10 Revenue-Generating Genres

Business Recommendation

1. Core Investment Strategy

Prioritize Global Blockbusters- Dedicate the majority of high-budget investment to projects in the Sci-Fi and Adventure genres, which exhibit the highest average revenue potential for worldwide success. Animation should be viewed as a reliable, high-yield genre for franchise building
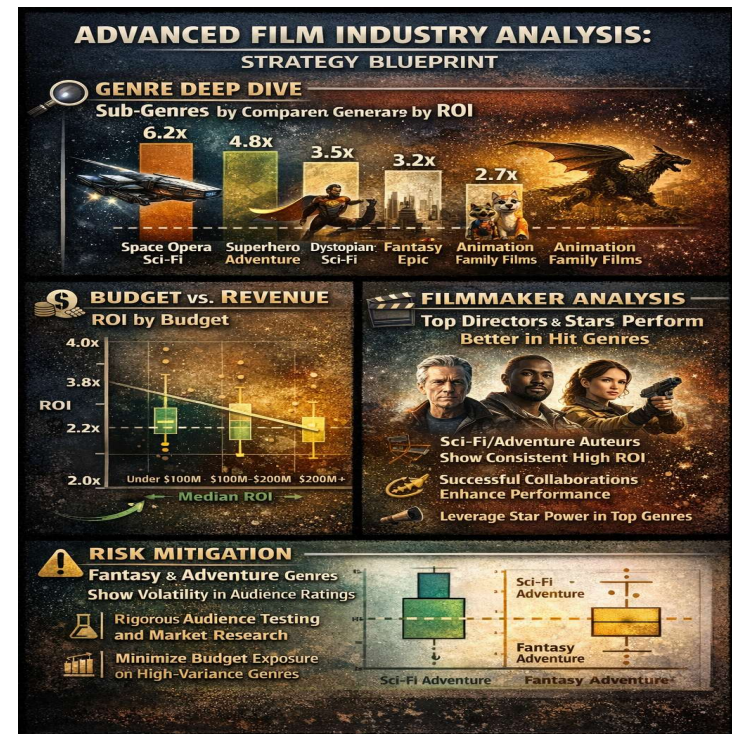
2. Runtime Mandate-Target the Sweet Spot: Implement a production mandate to keep the majority of commercial features within the 180–200 minutes runtime range, with the ideal peak being near the 180–200 minutes segment. Films significantly exceeding 200 minutes should be treated as high-risk exceptions

3.Marketing Strategy -Focus on Marketability Over Acclaim: Given the very weak positive correlation of 0.17, the studio should allocate resources more heavily toward strategic global marketing and distribution rather than relying on high critical acclaim to drive profits. The data shows that high ratings are not a reliable primary predictor of box office success

Next Steps for the Studio

A. **Genre Deep Dive**: Analyze the specific sub-genres (e.g., "Space Opera Sci-Fi" vs. "Dystopian Sci-Fi") that perform best.

B. **Budget vs. Revenue:** Integrate film budget data to calculate Return on Investment (ROI) for each genre and runtime segment

C. **Filmmaker Analysis:** Investigate the performance of key directors and stars within the top-performing genres

D. **Risk Mitigation** : The analysis found that high-profit genres like Fantasy and Adventure show high volatility in audience ratings (wide box plots in the notebook analysis)

E. **Mitigation:** When developing projects in these high-variance genres, employ rigorous audience testing or minimize production budget exposure, as the revenue potential is high but the risk of critical/audience failure is also elevated

CONTRIBUTORS

Yvonne Rajula

Lauren Njoki

Claire Njeri

Isaac Macharia

Dahir Ahmed

Abdullahi Hassan