



Data is the new script. We let the numbers tell us what audiences love

-
- Title

Analysis to Launch a Successful Movie Studio

- Subtitle

Uncovering What Makes a Movie Successful

Presented By:

- 1. Joackim Kisienya
- 2. Eric Metobo
- 3. Hezron Rumenya
- 4. Joy Sila
- 5. Lynn Kyalo
- 6. Newton Njeri

Problem statement

A company is about to invest in a new movie studio at a time when entertainment giants are pouring billions into original films and series. While the potential rewards are enormous, production costs are equally high, and every misstep carries significant financial risk. Today's challenge is to replace educated guesswork with rigorous data science: we need to uncover precisely which kinds of films by genre, by release timing, by runtime, and by audience appeal consistently deliver strong box office returns. Our goal is to empower your leadership team with clear, evidence-based guidance so you can greenlight projects with confidence rather than uncertainty.

Project Objectives

This analysis has three central objectives.

- Explore and quantify the relationship between film attributes such as genre, runtime, release season, and audience ratings and domestic box office revenue.
- Identify clear patterns and trends in this high-stakes data, calling out which combinations of characteristics tend to win big at the box office.
- Translate these insights into three strategic, actionable recommendations.

By the end of this presentation, we will have a data-driven roadmap for selecting the right films to produce for maximum commercial success.

Importance of the project

1

Reduce Financial Risk

Invest in the genres, release windows, and runtimes most likely to deliver strong returns, rather than testing ideas at random.

2

Maximize Marketing Efficiency

Align promotional spend with audience behavior patterns—knowing when viewers are most receptive to theatrical releases.

3

Build Competitive Advantage

Join the ranks of industry leaders who already leverage data analytics to produce blockbusters, ensuring the new studio makes informed, strategic decisions from day one.

Data sources and scope

To build our analysis, we assembled five datasets covering U.S. theatrical releases from 2010 through 2019:

- **IMDB (im.db):** Contains comprehensive title, genre, runtime, and user rating information for every film.
- **Box Office Mojo (bom.movie_gross.csv.gz):** Provides the official domestic gross revenue figures for each title.
- **Rotten Tomatoes (rt.movie_info.tsv and rt.reviews.tsv):** Offers critic scores and review details to add qualitative sentiment.
- **TheMovieDB (tmdb.movies.csv):** Supplies supplemental metadata such as production companies and alternate titles.
- **The Numbers (tn.movie_budgets.csv):** Includes production budget data to examine cost versus revenue relationships.

Although all five sources enrich our perspective, this presentation will focus primarily on IMDB's title, Box Office Mojo and The Numbers.

Building the Master Dataset



We started by merging IMDB's movie_basics and movie_ratings tables on the shared film identifier, ensuring that each record includes both descriptive metadata (title, genre, runtime) and audience sentiment (average rating, vote count).



Next, we enriched this combined IMDB dataset by joining it with **IMDB (im.db)** metadata, matching on cleaned title text and release year to append supplemental fields such as production company, original language, and alternate titles.



Finally, we integrated Box Office Mojo's domestic gross figures by performing a join between our enriched IMDB+ **tn.movie_budgets.csv** dataframe and the BOM revenue data, again matching on title and release year.



The result is a single, comprehensive master dataframe that contains every film's key attributes, user ratings, and box office performance providing the complete foundation for our subsequent analysis and visualization.



Data analysis

In this phase, we translated our cleaned dataset into actionable insights by performing the following analyses:

Genre-Level Revenue and Rating Averages

- Computed the average domestic gross and average IMDB user rating for each individual genre, allowing us to compare which categories deliver the strongest financial and audience appeal.

Seasonality Impact Assessment

- Grouped films by release quarter and month, then calculated mean revenues to identify when audience turnout peaks, revealing the optimal windows for launching major titles.

Runtime Performance Evaluation

- Segmented films into runtime brackets (<90 min, 90–120 min, >120 min) and compared their average gross and rating metrics, pinpointing the runtime sweet spot that maximizes both box office returns and viewer satisfaction.

Correlation Analysis

Measured the statistical relationship between IMDB user ratings and domestic gross, confirming that higher audience approval generally corresponds to stronger financial performance.

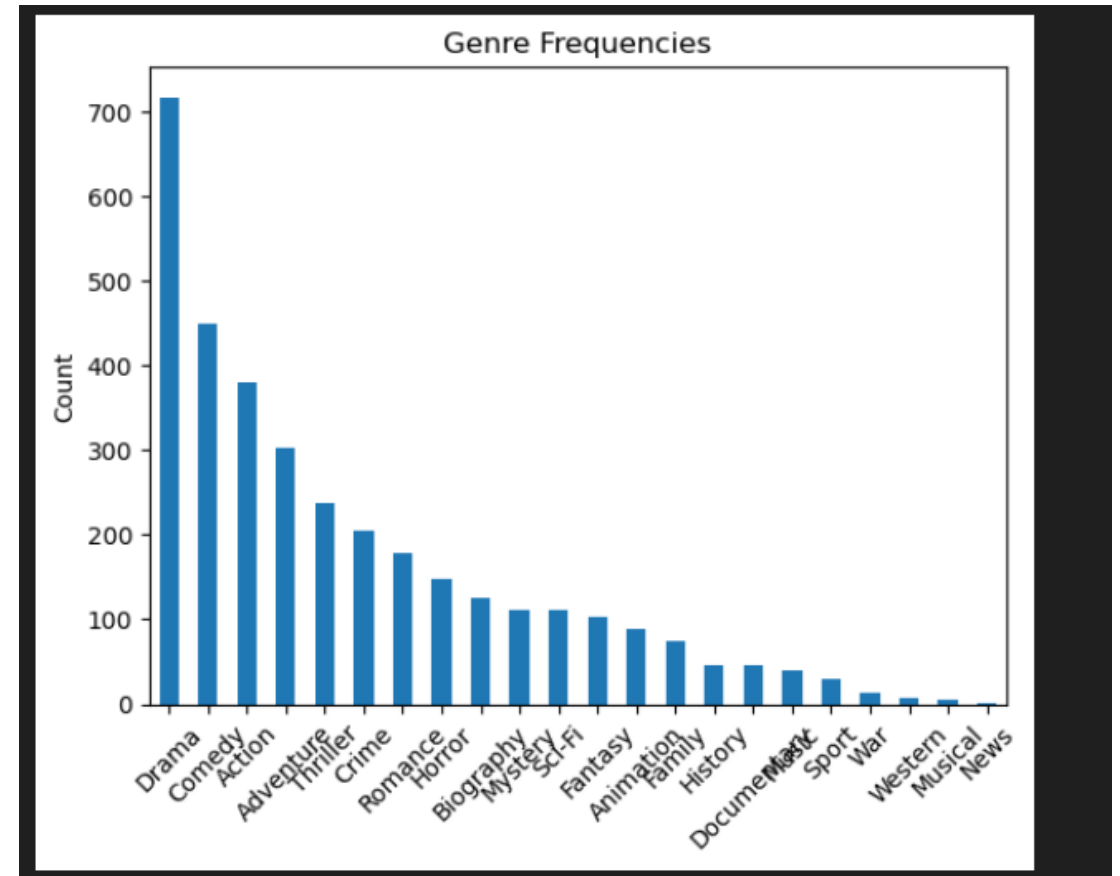
ROI and Profitability Ranking

- For each title with budget data available, calculated return on investment (ROI) as $(\text{domestic_gross} - \text{production_budget}) / \text{production_budget}$. Ranked films by ROI to spotlight the most cost-effective wins.

Data Visualization

- In this section, we translate our analytical findings into clear, impactful visuals that bring the data to life and make the insights immediately accessible.
- By walking through these visualizations, you will see how some factors influence box office outcomes, enabling your team to make data-driven decisions with confidence.

The bar chart beside displays the count of films in our cleaned master dataset for each genre. The horizontal axis lists individual genres ; Drama, Comedy, Action, Adventure, Thriller, and so forth while the vertical axis shows how many titles fall into each category

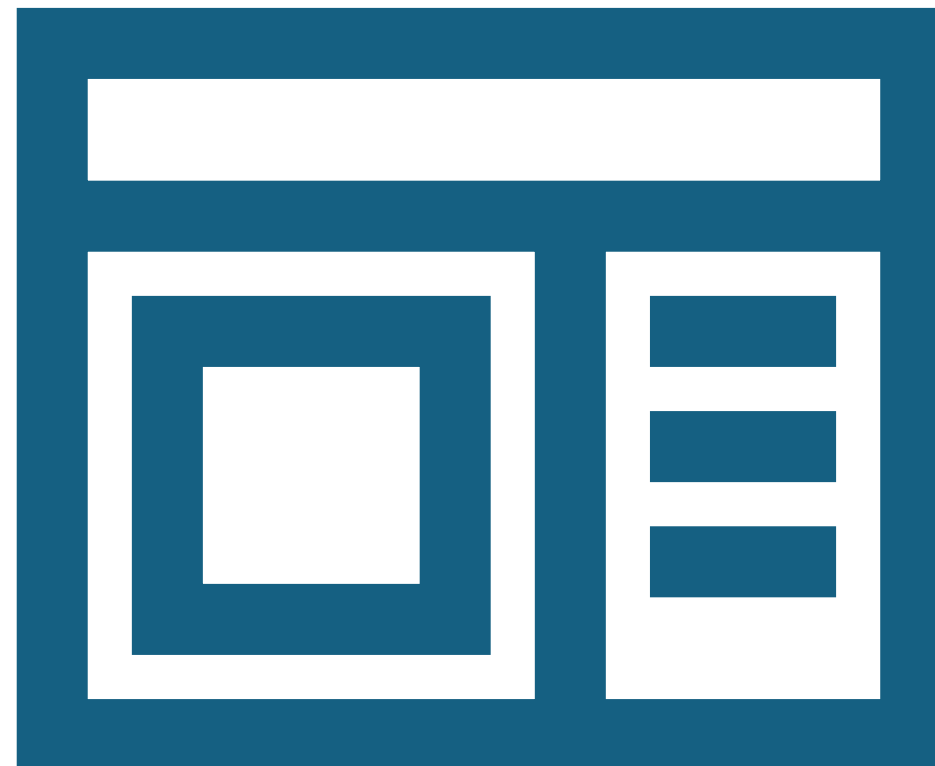


From the above visualization, we observe;

- **Drama Leads the Pack:** With over 700 titles, Drama is the most frequently produced genre in our 2010–2019 sample, representing nearly 25 % of all films.
- **Popular Blockbusters:** Comedy , Action , and Adventure follow closely behind, indicating studios consistently invest in these crowd-pleasing categories.
- **Long Tail of Niche Genres:** Beyond the top six, genre counts drop off sharply, War, Western, Musical, and News each have fewer than 15 films, suggesting lower volume (and potentially higher risk) in those areas.

Key Takeaways

- **Prioritize High-Volume Genres:** Because Drama, Comedy, Action, and Adventure dominate industry output, these categories benefit from established production pipelines, marketing channels, and audience familiarity factors that can reduce risk and accelerate time to market.
- **Consider Underserved Niches Strategically:** While niche genres are less common, they may offer opportunities for differentiation if your studio can target the right audience segment with specialized marketing.

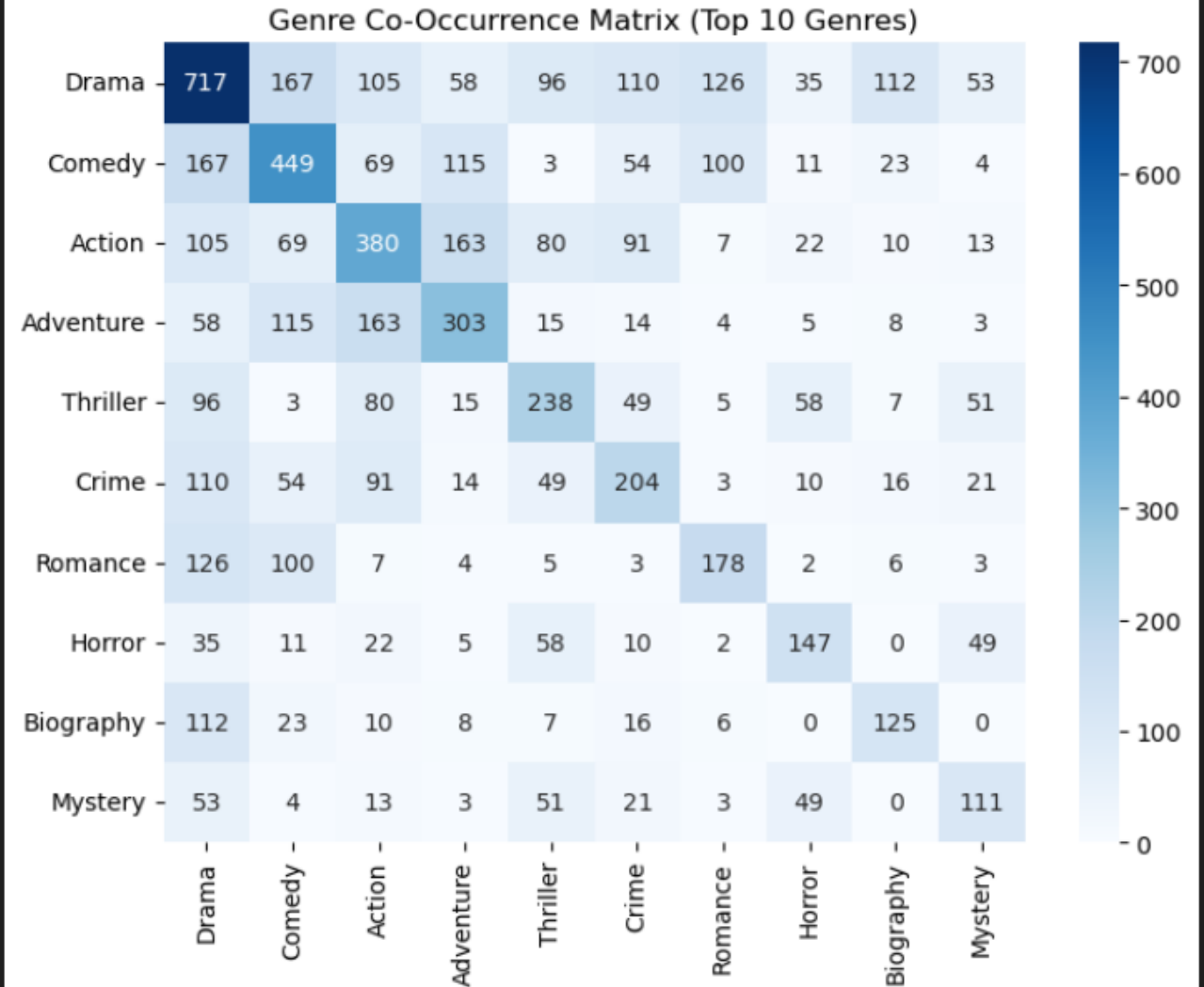


Visualization 2

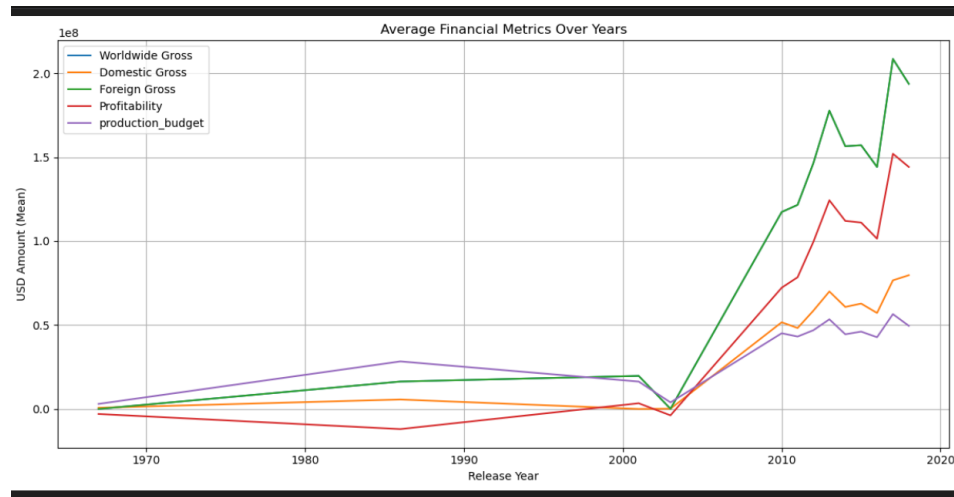
This heatmap displays a **genre co-occurrence matrix** for the **top 10 movie genres**. Each cell shows how often two genres appeared together in the same movie.

Key Insights

- **Drama** is the most common genre with **717 occurrences**, followed by **Comedy (449)** and **Action (380)**.
- Popular genre pairings include:
 - **Action & Adventure** – 163 movies
 - **Drama & Comedy** – 167 movies
 - **Thriller & Crime** – 49 movies
 - Romance & Drama (126) and Mystery & Thriller (51) also show strong relationships.
 - Rare or unlikely combinations include Romance & Horror only 2 movies Biography & Horror 0 movies
- Indicates strong patterns in genre mixing useful for content strategy, market positioning, and storytelling trends.



Visualization 3: Trends in Average Financial Movie Metrics (1960–2019)



Key Observations:

- Steady Growth Pre-2000: Gradual increase in financial metrics with minor fluctuations.
- Significant Spike After 2000: Sharp upward trend post-2010.
- Foreign Gross Dominates: Outpaces Domestic Gross and Budgets from 2005 onward.
- Profitability Trends: Low and unstable before 2000, rising sharply post-2010.

The sharp rise in financial performance metrics after 2010, particularly driven by **Foreign Gross**, highlights the increasing importance of global markets in movie profitability. Studios appear to have controlled budget growth relative to revenue growth, enhancing overall profitability.

Conclusions

Genres

1. Overall, mainstream and emotionally resonant genres like Drama, Comedy, and Action are popular in the movie industry and on demand among the audience. Specialized or genre-specific categories such as Western, Musical, and News are comparatively rare.
2. Genre combinations having a common occurrence included Drama & Romance, Action & Adventure, and Thriller & Crime. Some genres like Horror, Mystery, and Biography were most frequently used in focused or specialized narratives.

Financial performance

1. Average performance metrics across top ten major movie genres indicate that:
 - a). Adventure and Action dominate as the most expensive genres to produce, with average production costs with relatively high returns.
 - b). Horror and Mystery emerge as the most financially efficient genres, posting the highest average return on investment (ROI) at 11.71 and 11.24, respectively. These genres are comparatively inexpensive to produce but yield disproportionately high returns, making them highly attractive to studios with limited budgets or those seeking high-margin ventures.

Markets

1. Foreign revenue consistently outperforms domestic revenue, highlighting the growing importance of international markets for film profitability. International audiences are a crucial revenue source regardless of genre, and studios might benefit from tailoring content with global appeal.
 - a). Adventure and Action genres generate the highest revenues in both foreign and domestic markets
 - b). Adventure presents a significant margin it's foreign revenue i.e More than double its domestic market
 - c). Comedy, Thriller, and Crime genres also exhibit strong international performance

Average Rating

1. Overall average rating patterns show a long-term decline followed by partial recovery and consistency, which may reflect the industry's adaptation to audience preferences, critical standards, and evolving storytelling approaches
2. Voting behavior has evolved, first with steady growth, then with digital-era acceleration, and subsequently a decline possibly due to a larger volume of films being released (diluting votes), or newer releases having had less time to accumulate votes.
3. Biography has the highest-rated genre, followed closely by Drama and Adventure, indicating these genres tend to deliver strong storytelling and quality content appreciated by viewers

Other performance measures

1. Average Runtime The industry has settled on a more standardized length for mainstream films. This could be influenced by streaming trends, theatrical scheduling preferences, or evolving attention spans of audiences.
2. Standout production houses that balance both financial performance and critical acclaim based on ROI and average rating by genres were
3. a). UTV was financially efficient and creatively successful in the production of Action and Biography genres.
4. GrtIndia was best placed in the production of high-rated Action and Drama genres

Recommendations

Investment Portfolio

1. **High budget:** The data shows Action and Adventure titles cost the most but also pull in the widest audiences and sells best overseas. They are best launched in the spring or summer, when blockbuster demand is highest and foreign box-office potential is strongest. Success strategy: pack theaters worldwide and hit the release window that maximizes revenue.
2. **Mid budget:** Conclusions point out that Drama mixed with Romance or Crime paired with Thriller delivers a healthy balance of good ratings, decent ROI, and festival attention without runaway budgets. Success strategy - Target festival buzz and good reviews for free marketing, helping these films earn steadily year-round and boosting the studio's reputation.
3. **Low budget:** Horror and Mystery/Thriller are inexpensive to make yet top the ROI tables, according to the study. By capping budgets below \$10 million and timing releases near Halloween or other proven windows, can limit risk while giving a chance at the next breakout hit.

Marketing strategy

More than two-thirds of a movie's earnings now come from foreign audiences. Every project should plan with the foreign market needs in mind. That means budgeting early for translated trailers, posters and social media content aimed at the three biggest foreign markets for each film.

THE END

Questions session