Abstract

Why do some films resonate with audiences while others fall flat? This study examines what drives IMDb ratings for moderately popular films - those receiving between 100 and 1000 votes. This range captures movies that are visible enough to reflect genuine audience sentiment, yet not so prominent that ratings are dominated by franchise hype, marketing campaigns, or extreme fan behavior. Using a dataset of 71,331 films from IMDb, we analyze how genre, runtime, release period, director experience, actor experience, and repeated director–actor collaboration relate to audience ratings.

Our analysis combines descriptive visualization with a predictive regression model ($R^2 = 0.26$) to understand both broad patterns and joint effects. The results show that genre overwhelmingly shapes audience ratings: documentaries and biographies receive much higher scores than thrillers and horror films, with gaps exceeding three rating points. Release period also matters, with older films rated more highly than recent ones, consistent with survivorship bias. Runtime has a positive effect only when it aligns with genre expectations, helping some genres while hurting others. In contrast, director and actor experience have minimal influence once genre and time are accounted for, while repeated collaborations offer only a modest boost.

Overall, the findings suggest that IMDb ratings reflect cultural expectations and narrative fit more than professional credentials. For filmmakers, this implies that understanding genre conventions and aligning creative choices accordingly matters more than relying on experience or star power.

Introduction

IMDb ratings have become one of the most widely used indicators of film quality. Unlike box office revenue or awards, they reflect how everyday viewers respond to a movie after watching it. Because millions of users contribute ratings, IMDb provides a unique window into audience judgment at scale. Yet not all ratings are equally informative. Extremely popular films often attract intense fan activity, review bombing, or marketing-driven attention, while obscure films may lack enough votes to form stable opinions.

To address this, this study focuses on moderately popular films - those with 100 to 1000 IMDb votes. These films occupy a middle ground: they are visible enough to generate meaningful audience feedback but not so dominant that ratings are distorted by hype or polarized fan communities. Studying this group allows us to ask a clearer question: what actually drives audience appreciation when noise from extremes is reduced?

Prior research points to several possible drivers. Genre acts as a cognitive framework that shapes expectations; audiences judge documentaries, comedies, and horror films by very different standards (Litman, 1983). Runtime has been linked to perceptions of ambition and seriousness, especially in genres associated with complex narratives (Simonton, 2004). At the same time, reputation-based theories suggest that audiences use experience and star power as signals of quality, rewarding films made by well-known directors and actors (Elberse, 2007). Repeated collaborations between directors and actors may further enhance outcomes by fostering creative chemistry.

Based on this literature, we examine five hypotheses: that genre, runtime, director experience, actor experience, and director–actor collaboration influence IMDb ratings. Rather than treating these factors in isolation, we combine visual exploration with a regression model to understand their relative importance and interactions.

The goal is not to predict ratings perfectly, but to tell a clear, evidence-based story about how audiences evaluate films that sit outside the spotlight.

Results

The analysis uses a dataset constructed from IMDb's *title.basics*, *title.crew*, *title.principals*, and *title.ratings* tables. After filtering to films with 100–1000 votes, the final dataset includes titles from a wide range of genres and release periods. Several features were created to capture different dimensions of film production and reception. These include genre categories, runtime in minutes, release decade, measures of director and lead actor experience based on prior credits, and an indicator for whether a lead actor and director had collaborated previously. These features were chosen because they are observable at scale and align with common theories about audience evaluation.
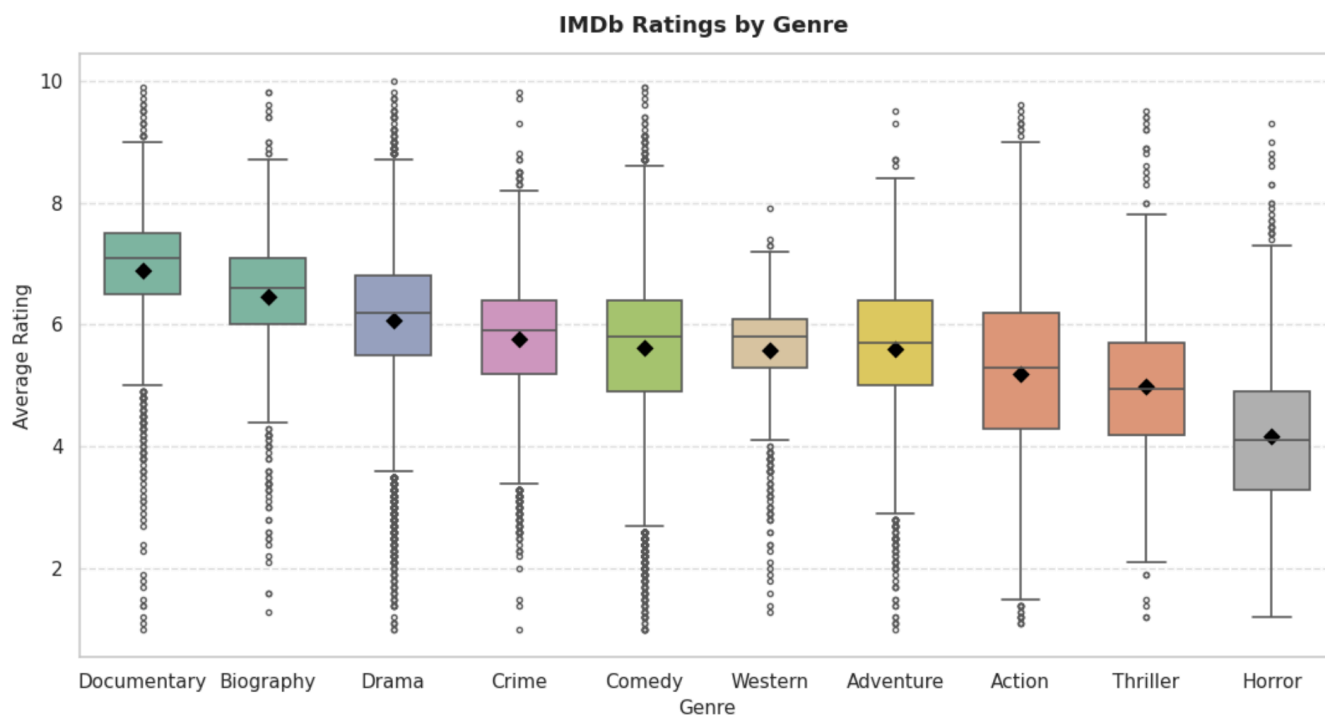

Figure 1: IMDb ratings by genre (top 10 genres)

This boxplot shows the distribution of IMDb ratings for the ten most common genres among moderately popular films. Each box represents the interquartile range, with the median marked inside and outliers shown as points. The figure reveals large and systematic differences across genres. Documentaries have the highest median rating (around 7.0), followed by biographies and dramas. In contrast, horror and thriller films cluster at much lower ratings, often below 5.0. The gap between the highest- and lowest-rated genres approaches four rating points, indicating that genre alone sets a powerful baseline for audience evaluation.

Figure 1 makes one result immediately clear: genre dominates everything else. Audiences do not judge all films on a single scale; instead, they apply genre-specific standards. This raises a natural question. If genre sets the baseline, can other choices - such as making a film longer, help overcome a genre disadvantage?
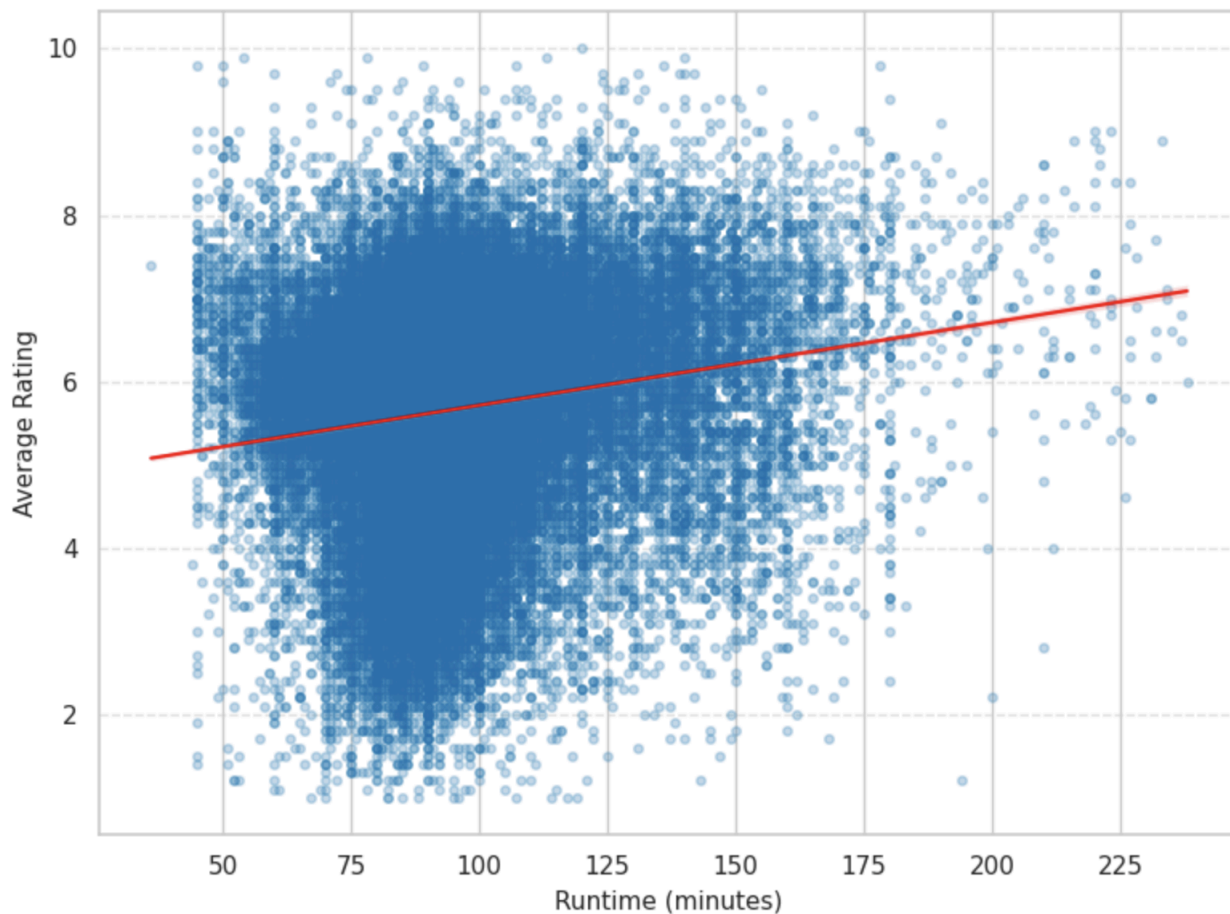
Figure 2: IMDb ratings by runtime category.

This figure groups films into runtime bins (<75, 75–90, 90–105, 105–120, 120–150, 150–180+ minutes) and shows the distribution of ratings within each bin. Median ratings increase modestly with runtime up to about 120–150 minutes, after which gains flatten. However, variability also increases for very long films. On its own, longer runtime appears to be associated with slightly higher ratings, suggesting that audiences may reward films that allow more time for narrative development.

However, this pattern turns out to be misleading when genre is taken into account. In the regression model, runtime interacts strongly with genre. Longer thrillers benefit substantially from extra time, gaining about 0.014 rating points per additional minute, while longer westerns lose about 0.008 points per minute. In other words, runtime helps only when it aligns with genre expectations. Length is not inherently good or bad; it matters whether the audience believes the genre "deserves" the extra time.

If structure matters so much, what about reputation? Do experienced filmmakers or established partnerships meaningfully raise ratings?
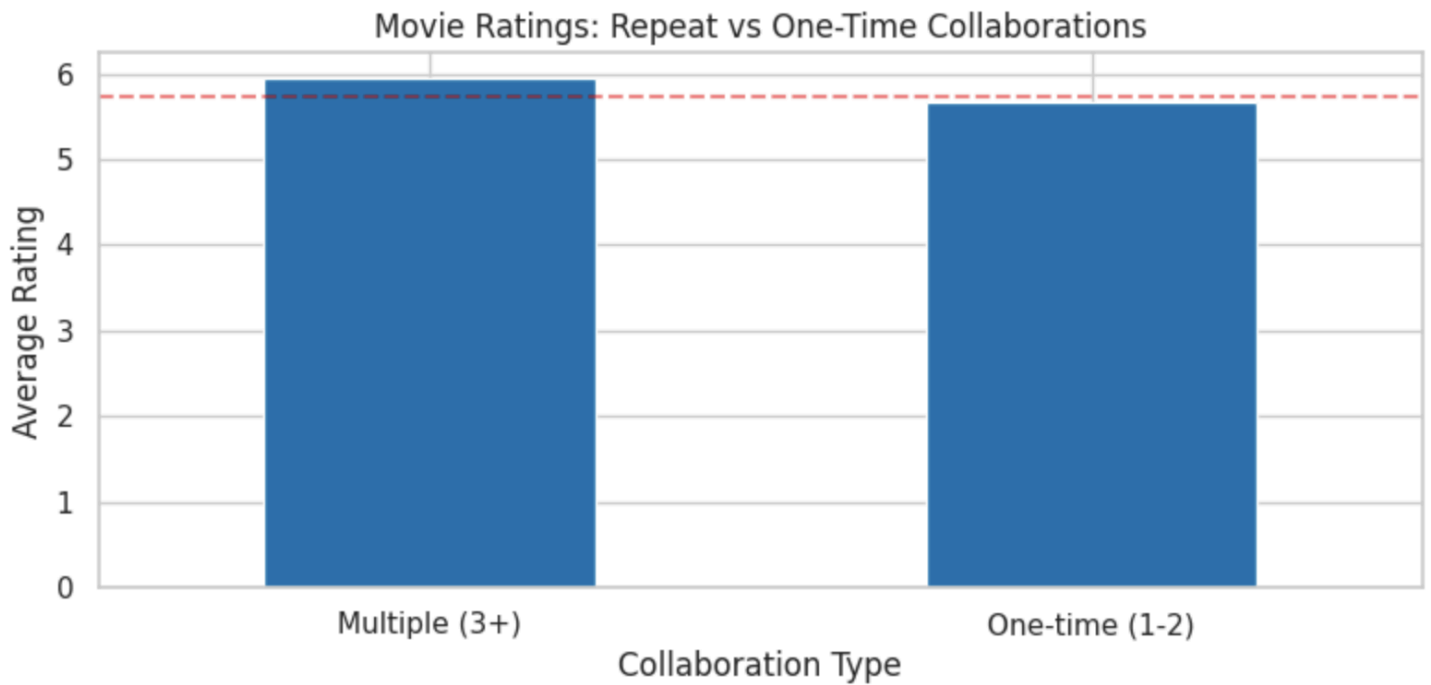
Figure 3: Average IMDb rating by director–actor collaboration type.
This bar chart compares average ratings for films where the director and lead actor had collaborated one or two times versus films with three or more prior collaborations. Films made by repeated collaborators score about 0.29 rating points higher on average. While the difference is consistent, it is small compared to genre effects. Collaboration appears to help, but only modestly.

Interestingly, director and actor experience by themselves contribute even less. In the full regression model, the coefficients on log director experience (-0.057) and log actor experience (-0.046) are close to zero and statistically insignificant once genre and decade are included. Experience does not meaningfully raise ratings; at best, it reduces the likelihood of very poor outcomes. This suggests that audiences do not reward credentials directly, they care more about the kind of film being made.

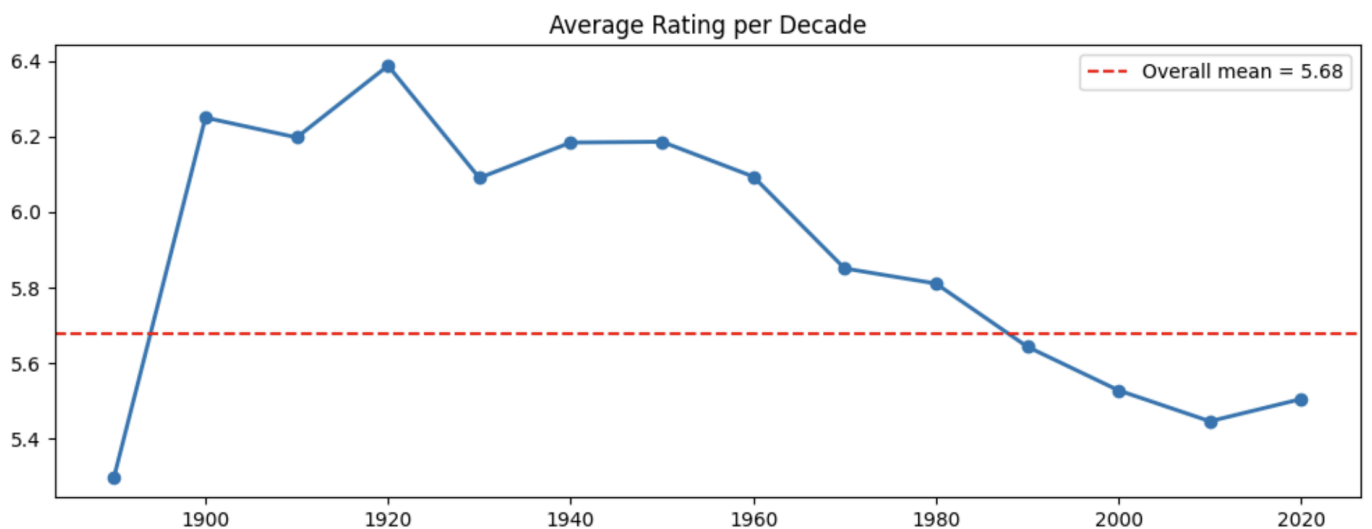Another structural factor shaping ratings is time.

Figure 4: Average IMDb rating by release decade.

This line plot shows the average IMDb rating of films by decade. Movies released before 1960 receive notably higher ratings than more recent films, while movies from the 2010s and 2020s are rated about 0.6 points lower on average. This pattern is consistent with survivorship bias: poorly received older films fade from view, while modern films are evaluated more comprehensively. Ratings therefore reflect not only quality, but also historical filtering and changing audience standards.

Finally, we combine all variables into a single predictive model to understand their relative importance.

```
Features by Importance

(+) = increases rating  and (-) = decreases rating

genre_Documentary                              2.4352
genre_Biography                                1.8672
genre_Animation                                1.5689
genre_Thriller                                -1.3018
genre_Horror                                  -1.0798
genre_Fantasy                                 -1.0267
genre_Adult                                    0.9331
genre_Other                                    0.9094
genre_Western                                  0.8797
genre_History                                  0.7220
genre_Romance                                  0.6544
decade_2010s_2020s                            -0.5903
genre_Sci-Fi                                  -0.5702
genre_Drama                                    0.5658
decade_2000s                                  -0.5636
```

Figure 5: Top 15 standardized regression coefficients predicting IMDb ratings.

This coefficient plot shows the magnitude and direction of the most influential predictors from the regression model. Genre indicators dominate the chart, with documentaries (+2.44) and biographies (+1.87) far outweighing all other effects, while thrillers (−1.30) and horror (−1.08) are strongly negative. Decade indicators show a clear time penalty for recent films. Runtime interactions appear only in smaller coefficients, and experience variables are near zero. The model explains 26 percent of the variance in ratings, highlighting both the strength of structural factors and the limits of prediction.

Taken together, the figures tell a consistent story. Genre sets expectations, time shapes evaluation through historical bias, runtime matters only conditionally, and reputation plays a surprisingly small role.

DISCUSSION

The central insight from this study is that IMDb ratings reflect cultural expectations more than professional credentials. Genre functions as an implicit contract between filmmakers and audiences. When a film delivers what its genre promises, it is rewarded; when it violates those expectations, it is penalized. This helps explain why documentaries and biographies rate so highly: audiences approach them with different standards, often valuing insight and authenticity over spectacle. In contrast, horror and thrillers face tougher grading, with small missteps leading to harsh judgments.

The strong effect of release period reinforces this interpretation. Older films benefit from historical filtering, while recent films are judged in real time, good and bad alike. This does not mean modern films are worse, but that ratings are shaped by which movies remain visible over time.

Perhaps the most surprising result is how little experience matters. Once genre and time are accounted for, director and actor experience nearly vanish as predictors. This challenges the idea that audiences consistently reward reputation or star power. Instead, experience seems to stabilize outcomes rather than elevate them. Collaboration offers a small benefit, suggesting that chemistry helps, but it cannot overcome structural disadvantages.

There are clear limitations. We cannot measure script quality, cinematography, performances, or marketing - factors that clearly influence audience enjoyment. Director identification in IMDb data may also obscure shared creative control. Importantly, 74 percent of rating variation remains unexplained. This is not a failure of the model, but a reminder that artistic success cannot be fully reduced to measurable inputs.

For filmmakers, the implications are practical. Choosing the right genre and matching runtime to genre expectations matters far more than maximizing experience or star casting. For audiences, the results suggest that ratings are less about who made a film and more about how well it fulfills its narrative promise.

In the end, the study highlights a reassuring conclusion: even in a data-rich world, movie quality still resists full prediction. Algorithms can explain patterns, but they cannot replace creativity, originality, or emotional impact.