Your Paper

Dallin Loftus

February 25, 2025

Abstract

Your abstract. I need to fill in this abstract. I plan on doing it after I finish the study.

1 Introduction

Investigate how different movie attributes (genre, plot, acting, pacing, cinematography, soundtrack) influence review scores and box office performance.

Significance: Helps studios and marketers predict success and optimize movie production.

Research Questions:

How do specific movie attributes (genre, acting, cinematography, soundtrack, runtime, director, writer, star power) correlate with review scores and box office revenue? Which attributes are most important for each types of movie genre?

How do certain actors, writers, directors, run times, and genres impact review score and box office performance?

Do higher review scores lead to higher box office revenue?

Predicting Box Office Revenue Using Review Sentiment and Movie Attributeschoose your plan.

2 Literature Review

2.1 Do stars drive movie success

Elberse, A. (2007). The Power of Stars: Do Star Actors Drive the Success of Movies? Journal of Marketing, 71(4), 102-120. This study examines how stars influence theatrical revenues, revealing that stars do boost expected revenues by an average of $3 million. The impact is especially significant when a new star with past box office established cast. However, the study also suggests that stars may not increase the overall value of film companies, as they tend to

2.2 Do ratings impact box office

Moon, S., Bergey, P. K., Iacobucci, D. (2010). Dynamic Effects among Movie Ratings, Movie Revenues, and Viewer Satisfaction. Journal of Marketing, 74(1), 108-121. This study explores how movie ratings from critics, communities, and viewers influence movie revenues and viewer satisfaction. The research reveals that early box office success boosts later ratings, and strong advertising maximizes revenues when combined with high ratings. Sequels tend to earn more but receive lower ratings than originals. These findings provide valuable insights for my project, which will examine how different attributes (e.g., plot, pacing, and soundtrack) influence review scores and sales. By incorporating cross-genre comparisons, my project will expand this analysis to understand how these dynamics vary across genres and how they impact sales.

2.3 Other sources to be cited later

Cherradi, Mohamed, and Anass El Haddadi. "Comparative Analysis of Machine Learning Algorithms for Sentiment Analysis in Film Reviews." Acadlore Transactions on AI and Machine Learning 3 (July 23, 2024): 137-47. https://doi.org/10.56578/ataiml030301.

Pei, J., Zhang, ZL. Liu, WA. Sentiment classification of movie reviews: a powerful method based on ensemble of classifiers and features. Int. J. Mach. Learn. Cyber. 15, 6027–6048 (2024). https://doi.org/10.1007/s13042-024-02299-w
Deliverable 6 begins here

3 Methodology

I've run multiple regression models to analyze factors influencing movie ratings and box office performance. One model examines which aspects of Rotten Tomatoes data contribute to critic and audience scores. Another explores the key drivers behind IMDb ratings. A third model evaluates whether review data can be used to predict box office revenue.

4 Results

4.1 What influences TomatoMeter

Genres significantly impact tomatoMeter scores. Documentaries, Animation, and Drama tend to receive higher ratings, while Kids & Family, Action, and Crime movies generally score lower. This suggests that critics favor certain genres over others, possibly due to storytelling depth or artistic merit.

Newer movies tend to have lower scores, which could be attributed to harsher critic standards or shifting audience preferences. As film evaluation criteria evolve, recent releases may be judged more critically compared to older films. Additionally, longer movies tend to receive higher scores, indicating that extended runtimes may allow for more in-depth storytelling and character development.

Ratings also influence scores. Films rated PG, PG-13, and TV-Y7 are generally associated with lower scores, while TV-MA may have a positive effect, though this is not statistically strong. This trend could reflect differences in content maturity and target audience expectations.

Despite these trends, the model explains only about 20% of the variance in tomatoMeter scores, indicating that other factors such as cast, director, budget, marketing, and audience sentiment likely play a significant role in determining a movie's critical reception.

4.2 What influences Audience Score:

The model explains only about 19.26% of the variance in audience scores, indicating that other factors such as cast, marketing, and cultural relevance likely play a significant role. Additionally, the residual standard error of 19.44 suggests a large average deviation between actual and predicted audience scores.

Several factors negatively impact audience scores. Newer movies tend to receive lower ratings, as indicated by the negative coefficient for release year. R-rated and PG-rated movies also tend to score lower compared to the presumed G-rated baseline. Certain genres, including Horror, Action, Crime, Mystery/Thriller, Romance, History, and Kids/Family, are associated with lower audience scores. Additionally, TV-14 and TV-Y7 ratings show lower audience engagement.

On the other hand, some factors positively influence audience scores. Longer movies tend to receive higher ratings, with each additional minute slightly boosting the score. Documentary, Sports, Anime, Nature, Faith/Spirituality, and Biography genres perform particularly well, with Faith/Spirituality showing the highest positive effect. Other genres such as Animation, Comedy, Fantasy, War, and Musical also contribute positively to audience ratings.

Production-related variables also play a role. Movies with an unknown director tend to perform significantly worse, with a strong negative impact on audience scores. Additionally, factors such as sound mix and distributor choices show minor but varying effects on audience ratings.

4.3 Ratings influencing IMDB Dataset Box Office

A correlation of 0.325 between IMDb ratings and box office revenue indicates a weak to moderate positive relationship. Higher IMDb ratings tend to be associated with higher box office revenue, but the connection is not strong. Other factors, such as marketing, budget, genre, and competition, likely

have a greater impact. Since the correlation is far from 1.0, highly rated movies don't always earn the most, and lower-rated films can still perform well financially.

4.4 Everything looking to predict Box Office

The regression results show that approximately 57.45

Among the key findings, movies rated PG, PG-13, and R tend to generate significantly higher box office revenue, with PG-13 films earning around $72 million more on average. In contrast, films with a "Not Rated" certificate te$ million in revenue. Runtime also plays a role, as longer movies are associated with higher earnings, with an estimated increase of $766K per additional minute.$

Audience perception, as measured by IMDb ratings, is a strong predictor of financial success. A one-point increase in IMDb rating corresponds to an additional $15 million in revenue. However, genre alone does not show a signif$

Directors have a substantial influence on revenue, with some directors contributing significantly to higher earnings. For instance, certain directors are associated with revenue increases ranging from $268 million to $334 million, indicating the strong impact of well-known or highly skilled filmmakers.

Deliverable 6 ends here

4.5 How do movie attributes correlate with review scores and box office revenue?

xyz...

4.6 How do certain actors, writers, directors, run times, and genres impact review score and box office performance?

jfnvjdnvjdfnvjdnv

4.7 Do higher review scores lead to higher box office revenue?

4.8 Predicting Box Office Revenue Using Review Sentiment and Movie Attributes

jfnvjdnvjdfnvjdnv

4.9 Why This Study Matters

Better Prediction Models for Studios and Marketers

Studios can use my findings to optimize production and marketing budgets, maximizing return on investment.

Understanding What Truly Drives Box Office Success

My study helps differentiate between factors that drive critical acclaim vs. box office revenue.If review scores don't always correlate with financial success, studios can make more informed decisions about balancing artistic quality with commercial appeal.

Advancing Machine Learning Applications in Entertainment

Using AI for sentiment analysis and predictive modeling isn't just about movies; it has applications in predicting product success across industries (e.g., books, music, video games).My approach, especially with NLP techniques, could lead to more refined models in entertainment analytics.

Reassessing the Role of Critics and Audience Reviews

If top critics' opinions don't align with audience sentiment or revenue outcomes, it raises questions about their influence.My study could provide insights into whether studios should focus more on audience feedback rather than critic scores.

Contributing to Film Studies and Storytelling Science

Identifying which attributes resonate most within each genre can help filmmakers refine storytelling, cinematography, and character development strategies.It can help film schools and aspiring filmmakers understand how different elements contribute to both artistic and financial success.
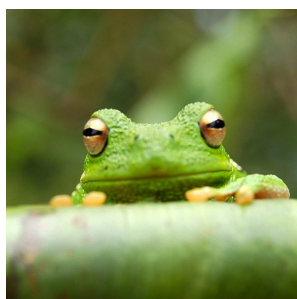
Has This All Been Done Before?

Figure 1: This frog was uploaded via the file-tree menu.

Item	Quantity
Widgets	42
Gadgets	13

Table 1: An example table.

While there have been past studies on box office prediction, my work expands on existing research by integrating a broader range of variables, using advanced sentiment analysis, and testing multiple predictive models.The industry constantly evolves (e.g., streaming vs. theatrical releases, changing audience preferences), making continuous research necessary. Additionally many of these studies were done before AI and so I have a chance to broaden this field as a whole.

Limitations

While there are many limitations I believe that the biggest is that my data doesn't have the budget for these different movies. I am unfortunately assuming that a bigger box office is better. However, a movie may have a bigger box office than another but because of a more expensive budget it could be less successful. This is a huge drawback to my data.

4.10 How to add + every other piece of code I still need

You can simply upload a `.bib` file containing your BibTeX entries, created with a tool such as JabRef. You can then cite entries from it, like this: [Gre93]. Just remember to specify a bibliography style, as well as the filename of the `.bib`. You can find a video tutorial here to learn more about BibTeX.

If you have an upgraded account, you can also import your Mendeley or Zotero library directly as a `.bib` file, via the upload menu in the file-tree.

4.11 How to add Tables

Use the table and tabular environments for basic tables — see Table 1, for example. For more information, please see this help article on tables.

4.12 How to add Comments and Track Changes

Comments can be added to your project by highlighting some text and clicking "Add comment" in the top right of the editor pane. To view existing comments, click on the Review menu in the toolbar above. To reply to a comment, click on the Reply button in the lower right corner of the comment. You can close the Review pane by clicking its name on the toolbar when you're done reviewing for the time being.

Track changes are available on all our premium plans, and can be toggled on or off using the option at the top of the Review pane. Track changes allow you to keep track of every change made to the document, along with the person making the change.

4.13 How to add Lists

You can make lists with automatic numbering . . .

1. Like this,

2. and like this.

. . . or bullet points . . .

- Like this,

- and like this.

4.14   How to write Mathematics

LATEX is great at typesetting mathematics. Let $X_1, X_2, \ldots, X_n$ be a sequence of independent and identically distributed random variables with $\mathrm{E}[X_i] = \mu$ and $\mathrm{Var}[X_i] = \sigma^2 < \infty$, and let

$$S_n = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{1}{n}\sum_i^n X_i$$

denote their mean. Then as n approaches infinity, the random variables $\sqrt{n}(S_n - \mu)$ converge in distribution to a normal $\mathcal{N}(0, \sigma^2)$.

4.15   How to change the margins and paper size

Usually the template you're using will have the page margins and paper size set correctly for that use-case. For example, if you're using a journal article template provided by the journal publisher, that template will be formatted according to their requirements. In these cases, it's best not to alter the margins directly.

If however you're using a more general template, such as this one, and would like to alter the margins, a common way to do so is via the geometry package. You can find the geometry package loaded in the preamble at the top of this example file, and if you'd like to learn more about how to adjust the settings, please visit this help article on page size and margins.

4.16   How to change the document language and spell check settings

Overleaf supports many different languages, including multiple different languages within one document.

To configure the document language, simply edit the option provided to the babel package in the preamble at the top of this example project. To learn more about the different options, please visit this help article on international language support.

To change the spell check language, simply open the Overleaf menu at the top left of the editor window, scroll down to the spell check setting, and adjust accordingly.

4.17   Good luck!

We hope you find Overleaf useful, and do take a look at our help library for more tutorials and user guides! Please also let us know if you have any feedback using the Contact Us link at the bottom of the Overleaf menu — or use the contact form at https://www.overleaf.com/contact.

References

[Gre93] George D. Greenwade. The Comprehensive Tex Archive Network (CTAN). *TUGBoat*, 14(3):342–351, 1993.