

Analysis of Business Factors Influencing Film Revenue

Hanyi Zhang^{1,a,*}

¹King's College London, London, WC2R 2LS, The United Kingdom

a. K21035756@kcl.ac.uk

**corresponding author*

Abstract: The level of film income directly determines the success of the film's commercial awareness. It not only directly reflects the audience's acceptance and market influence of the film, but also has a profound impact on the development of the film industry. In order to explore the various factors affecting film revenue, this paper selects all the feature films that have been released worldwide from 2022 to 2023, and uses statistical methods to conduct linear regression analysis on six influencing factors, including film budget, audience rating, number of votes, popularity value, release date and film duration, with the help of R data analysis software, so as to evaluate the impact of each factor on the film revenue. In addition, by establishing a linear regression model between the influencing factors and the total income of the film, and based on the analysis of variance based on this model, this paper analyzes these influencing factors one by one. The results showed that except for the release date, which had no significant negative correlation effect on revenue, the other influencing factors were all significantly positive. When considering the significant interaction factors, the combination of number of votes and audience rating has a strong positive impact on revenue, while the combination of audience rating and popularity value and the number of votes and popularity value have negative impacts.

Keywords: Film Revenue, Film Industry, Linear Regression, Empirical Research, Analysis of Variance

1. Introduction

Since the birth of film in the 19th century, film has been influencing people's lives as a representative form of entertainment. By the early 20th century, film companies pooled their funds to build production bases, purchase expensive production equipment, and raise funds for the production of feature films, relying on large audiences to collectively buy tickets to watch films to make a profit. This formed the rudiments of a commercial film production and sales model [1]. Today, a film's box office performance and its surrounding revenues are generally accepted as an important measure of their success. Therefore, it is crucial for investors to analyze and understand the factors that influence film revenue, as it can help them gauge how to make choices to maximize their returns.

With the development of the times, more and more film technology has begun to advance. More and more viewers are able to learn a lot of information before the film is released, such as the approximate investment of the film, the release date, the type of film, and the overall rating of the film in advance, and this information will also be promoted in advance to help the audience choose.

This paper selects all the feature films that have been released worldwide from 2022 to 2023, and uses statistical methods to conduct linear regression analysis on six factors: film budget, audience rating, number of votes, popularity value, release date and film duration, and evaluates the impact of each factor on the box office. The six main influencing factors of film budget (budget), audience rating (vote_average), number of votes (vote_count), popularity value (popularity), release date (release_date) and film duration (runtime) have an impact on the total revenue of the film.

2. Methodology

2.1. Scope of Study

The data selected in this paper are 899 feature films (According to the Academy of Motion Picture Arts and Sciences, the American Film Institute and the British Film Institute a feature film runs for more than 40 minutes) [2] released worldwide between January 2022 and December 2023. The data is from the "Full TMDB Films Dataset 2024 (1M Films)" dataset on the Kaggle website, and all the data is from the official TMDB website.

2.2. Variable Selection and Definition

The dependent variable of the model discussed in this article is the total revenue generated by the film, including the box office and its derivative income. Seven main influencing factors, including film budget, audience rating, number of votes, popularity value, release date and film duration, were selected as the independent variables of the model.

- Film Duration (Duration of the film in minutes): This article selects data for feature-length films with a duration of more than 40 minutes and less than 300 minutes.
- Budget (Budget allocated for the film): The data selected in this article are feature-length films with a budget greater than or equal to 0.
- Release date (Date when the film was released): Select the specific date of the film's first release worldwide. In consideration of the release, the date was converted to a quarter for analysis. This article divides the release quarters into four categories, with one quarter containing the number of films released in a three-month period.
- Audience rating (Average vote or rating given by viewers): Film rating is a comprehensive evaluation of all aspects of the film by netizens, reflecting the quality of the film and people's recognition of the film, affecting the word-of-mouth communication of the film, and is an important guarantee for box office revenue. When people have doubts about the promotional information of a film, they usually turn to professional film word-of-mouth websites for real information. As a result, the relationship between film ratings and box office has become more and more close [3]. The audience rating data selected in this article comes from the average score of the film displayed by TMDB.
- Number of votes (Total count of votes received for the film): The number of votes is the number of votes for all the people who have rated the film on TMDB.
- Popularity (Popularity score of the film): The popularity value selected in this article is from the definition on the TMDB official website, which is defined as: Popularity is a fairly important metric here on TMDB. You can think of popularity as being a "lifetime" popularity score that is impacted by the attributes below.

Number of votes for the day, Number of views for the day, Number of users who marked it as a "favorite" for the day, Number of users who added it to their "watchlist" for the day, Release date, Number of total votes, and Previous days score"[4].

2.3. Data Analysis of Model Construction Set

In this paper, multiple linear regression will be carried out on all data, and the generalized linear model will be based on the constructed:

$\text{lm}(\text{revenue} \sim \text{budget} + \text{month_group} + \text{vote_average} + \text{vote_count} + \text{vote_average} * \text{vote_count} + \text{popularity} + \text{vote_count} * \text{popularity} + \text{vote_average} * \text{popularity} + \text{vote_average} * \text{popularity} * \text{vote_count} + \text{runtime})$

The analysis of variance will be performed based on the constructed model and the ANOVA table to analyze the influencing factors.

3. Results

3.1. Multiple Linear Regression

After basic processing of all the data by Python, data integration using R and analysis after building the model, the following conclusions about the probability p-value can be drawn: (p-value < 0.05 proves that the independent variable has a significant effect on the dependent variable) (independent effect)

- Insignificant independent variables: film duration (0.873) > audience rating (0.824) > release date (0.489)
- Significant independent variables: popularity value (0.049) > number of votes (0.003) > film budget (<2e-16)

Note: The above p-value comes from the constructed general linear model, which has an explanatory degree of 75.43%. The results show the degree of impact based on the combination of all influencing factors. Based on the above results, and considering the correlation between the variables, the independent influence of a particular independent variable on the dependent variable will be studied under the condition of controlling for other variables, and the relationship between the independent influence of the above-mentioned independent variable and the dependent variable will be explained in the order of increasing impact significance according to the p-value.

3.2. Analysis of Variance

3.2.1. Film Duration

Under the independent influence of film duration, film duration has a significant impact on total film revenue (sig. = 1.66e-14 < 0.05).

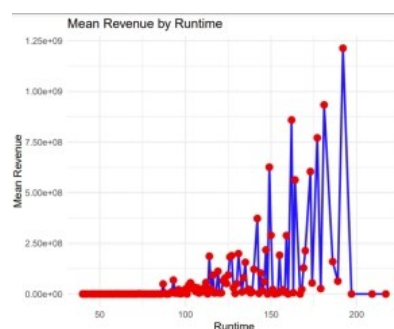


Figure 1: Mean Revenue by Runtime Line Chart

There is a positive correlation between film duration and revenue. After the analysis of variance through R, it was found that there was a significant impact between the film duration and the film

box office, and the mean line plot between the film duration and the film box office was made (Figure 1). Combined with the data in the graph and the results of ANOVA, the longer the film, the more likely it is to increase its revenue. This may be due to the fact that longer films tend to have more complex plots and a richer viewing experience for viewers, all of which contribute to higher box office revenues. However, it can be noted that very long durations also come with high risks. The following is a discussion of the standard deviation of the film duration segment:

Less than 60 minutes (6.392) < 60-90 minutes (7.832) < 90-120 minutes (8.300) < more than 120 minutes (19.736)

After the film duration exceeds 120 minutes, the income shows a clear upward trend, but the extremely high standard deviation compared with other segments indicates that the income is extremely unstable.

3.2.2. Release Date

According to the data analysis, the release date had no significant impact on the film revenue under the independent influence of the release date ($\text{sig.} = 0.19 < 0.05$). The reason for this may be due to the rise of streaming platforms and the impact of the pandemic.

Since the global spread of the new crown epidemic in 2020, the global economic and social development has been significantly affected. As a cultural and creative industry, the film industry has been particularly hard-hit. The impact of the epidemic on the film industry is not something that can be eliminated in a short period of time, but a long-term, continuous and structural change in the entire industry chain of the global film and television industry - the production, marketing, and distribution methods of films have been reformed to varying degrees, the strong entry of streaming media has changed the form of film distribution, the distribution of theaters has shrunk sharply, a large number of theaters have gone bankrupt and closed, and audiences' film-watching habits have changed to home platforms [5], thus greatly weakening the impact of release dates on film revenue. As shown in Figure 2, the number of films released has generally shown a downward trend over time. Overall, the first quarter had the highest number of film releases compared to other quarters, which may be related to the holiday schedule. The fourth quarter has the least number of films released, only about half of the number in the first quarter.

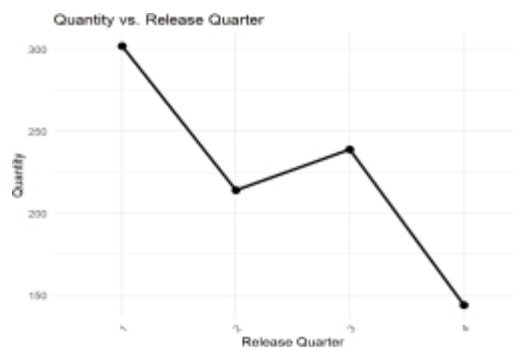


Figure 2: Quantity vs Release Quarter

3.2.3. Popularity Value, Audience Rating and Number of Votes

Because these three factors are strongly correlated, the three of them are discussed together. All three factors were positive and significant in the independent discussion. However, if a multivariate ANOVA is performed on these three independent variables, the results show that although the combined variables are all significant factors, some of the combined influencing factors become negative effects. The order is number of votes and audience rating (strong positive correlation) >

audience rating and popularity value (strong negative correlation) > number of vote and popularity value (strong negative correlation).

Result description: (1) The interaction effect of the strong positive correlation between the number of votes and the audience rating is significant: the analysis of the score and the number of votes together can reflect the popularity of the film (similar to the popularity value to some extent), for example, a higher-quality film will attract a large number of viewers to vote for the rating, and then improve the film score. This combination has an important impact on revenue. (2) The interaction effect of the strong and negative correlation between audience rating and popularity value is significant: popularity value usually reflects marketing intensity and audience attention. Ratings may have a negative correlation with high-profile blockbusters, such as Marvel, DC and other "fan-oriented" films, which may already have a large audience base, so low ratings may still harvest a part of the market. And some niche, high-quality independent films may have limited revenue due to a small audience. (3) The negative correlation between the number of votes and the popularity value is significant: Films with high popularity (e.g., early release) may have fewer votes, and vice versa. This interaction may reflect a complex process of changes in box office revenues over time.

The interaction effects in the above data indicate that a single variable may not be sufficient to explain income, but the combination of variables has an important impact on income. In particular, audience ratings have a less obvious impact on income than other factors, but they can have a significant impact on income through the interaction with the number of votes and popularity.

It is true that with the continuous development of the film industry, the output of films will continue to grow, in the face of huge and rapidly expanding film resources, how to choose the right film for the audience to watch needs to be solved urgently: even if it is selected by tags and indexes, such as distribution region, film genre or release time, the search results are ultimately only a preliminary narrowing of the scope, and the question of how to choose a film that is worthy and suitable for their own viewing is still lingering around the audience. The emergence of online film ratings has alleviated the above problems to a certain extent: at present, audiences can not only search for their satisfactory results according to tags and indexes, such as ratings and rating rankings on websites, but also further filter on the basis of traditional film selection tags. This is where the value of online film ratings comes in as a label and index of Internet integration resources [6].

3.2.4. Film Budget

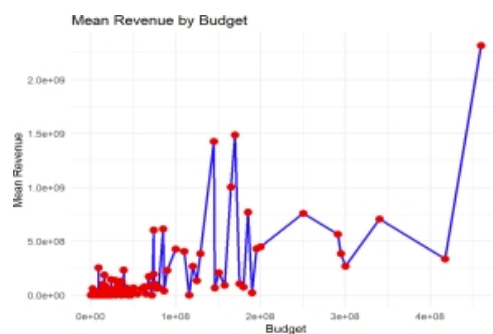


Figure 3: Mean Revenue by Budget Line Chart

Film budget is one of the key factors that affect film revenue. According to the results of regression analysis, there is a significant positive correlation between film budget and revenue under the independent influence of film duration ($\text{sig.} = 2\text{e-}16 < 0.05$). As can be seen from Figure 3, film budgets and film revenues regress roughly in a positive linear direction.

This shows that the higher the budget, the higher the revenue. This result is in line with industry belief, as higher budgets often mean better production quality, a stronger cast, and a wider market

outreach. This may be due to the improvement of production scale and quality, the blessing of publicity and distribution resources, and the market appeal of actors and directors. A higher budget allows studios to spend more on things like visual effects, costume design, and shooting equipment, improving the overall quality of the film and thus attracting more audiences. High-budget films tend to be accompanied by large-scale publicity and global distribution, which helps to appeal to a wider audience. For example, events such as trailers, posters, social media promotions, and celebrity meet-and-greets all require a lot of financial support. High budgets often allow for the hiring of well-known directors and actors, who often have their own box office appeal and can directly improve the film's appeal and market performance. All in all, the knock-on effect of a high-budget film can capture the attention of a larger audience. However, it's important to note that a high budget doesn't necessarily guarantee a high income. For example, in Figure 3, some films that are over-budgeted but under-quality or have a poor market response may face losses (e.g., "box office bombs"). In the eyes of many investors and producers, a high budget is the same as a high box office. However, as the number of high-budget film productions climbs and consumers gradually return to rationality, the film market is bound to gradually mature. The film industry itself is an industry of both opportunities and risks. Due to the economic nature of the film itself, a high-budget investment also means a high risk [7]. In addition, the analysis shows that the marginal impact of the budget on income may be diminishing, i.e., when the budget reaches a certain level, the increased budget becomes limited in its effect on the increase in revenue. This may be because audiences are more focused on the story and quality of the film than just the cost of production. According to the analysis results of this paper, film budget has a significant positive impact on income, especially in large-budget films. However, the budget is only one of the many factors that affect revenue, and its effect is also moderated by the promotional strategy, the quality of the film's content, and audience preferences. Therefore, while increasing budgets, studios should focus on content quality and market positioning to maximize box office revenue.

4. Conclusion

As a high-risk industry, the income of film may be related to many factors. Research in Western media economics shows that factors such as the director, star, production cost, advertising cost, distributor, schedule, awards, and audience's evaluation of the quality of the film will affect the box office performance of the film [8].

From the current analysis, the film budget has the greatest impact factor. If you want to reap high profits, it is a more effective strategy to choose to invest higher production costs in the early stage of film production. Before and after the release of the film, attention should be paid to the publicity and distribution of the film. In this day and age, people are more willing to pursue popularity. In this context, the action to promote the film before and during its release is especially important. The public effect can be achieved through social networks, offline roadshows, TV shows, etc. Another way is to encourage viewers to actively post their thoughts and reviews about the film on the internet [9]. According to the analysis results, compared with only focusing on improving the quality of the film in the hope that the audience will buy it (judged by audience ratings), if a lot of publicity can be carried out in the early stage of the film's release, and the popularity of film-related topics can be increased to attract more audiences, its income will increase significantly, exceeding that of similar films in the same period. For film duration, it's best to keep the film duration to around 120 minutes when making a film; if it exceeds 120 minutes, the film's revenue will be more likely to fall short of the expected revenue. Regarding the release schedule of films, the new crown epidemic has largely determined the direction of the film industry. At a time when cinemas in public spaces are clearly being defeated by online cinemas, restrictions imposed on public spaces due to the ongoing pandemic

will rapidly reduce interest in cinemas in the coming years [10]. As many films have changed from offline cinemas to online screenings, the impact of film schedules has been weakened.

At the same time, there are few studies on how to understand the influencing factors that affect film revenue. How to balance the artistic value and commercial value of the film will be a question that should be paid attention to in follow-up research.

References

- [1] "Film." baidubaike, baidu, Accessed 19 Dec. 2024. <https://baike.baidu.com/item/film/31689>.
- [2] "FAQ." British Film Institute, 24 Nov. 2019. <https://www.bfi.org.uk/bfi-national-archive/search-bfi-archive/bfi->
- [3] Wang Zheng, Xu Min. *Analysis of the influencing factors of movie box office: A study based on the Logit model* [J]. *Exploration of economic issues*, 2013(11):96-102. DOI:10.3969/j.issn.1006-2912.2013.11.017.
- [4] "Popularity & Trending" TMDb, (n.d.), <https://developer.themoviedb.org/docs/popularity-and-trending>
- [5] Jing Yi, Guangming Hou, Jianxun Wu. (2021) *The four dimensions of building a film power in the post-epidemic era* [J]. *Contemporary cinema*, (6):111-121. DOI:10.3969/j.issn.1002-4646.2021.06.017.
- [6] Qi Wei. (2015) *Fan Movie Preference, Viewing Orientation and Box Office Balance: A Study on the Current Situation of Online Movie Ratings* [C]// *Proceedings of the 3rd China Film and Television Youth Forum*.pp.39-49.
- [7] Hong Zhang, Qinqin Wang. (2008) *Risk Control of High-budget Film Operations: Analysis of the Production and Marketing of "Transformers"* [J]. *Contemporary cinema*, (12):70-73
- [8] Xiaoli Hu, Bo Li, Zhangpeng Wu. *Analysis of the influencing factors of the movie box office* [J]. *Journal of Communication University of China (Natural Science Edition)*, 2013, 20(1):62-67,39. DOI:10.3969/j.issn.1673-4793.2013.01.011.
- [9] Hao, B. (2023). *The Analysis of the Factors that Influence the Film Revenue. Highlights in Science, Engineering and Technology*, 47, 154-159. <https://doi.org/10.54097/hset.v47i.8184>
- [10] Okumuş, M. Sami. (2022) *The effects of Covid-19 pandemic on audience practices in cinema, television, and OTT platforms*. *İstanbul Ticaret Üniversitesi Sosyal Bilimler Dergisi*, 21.43: 133-147.