Analyzing the Timing of Theatrical Film Releases to Additional Distribution Channels

Danielle Randazza

July 26, 2024

Introduction

Box office revenues and home entertainment revenues combined totaled $36.8 billion in the United States in 2021. Out of this $36.8 billion, $4.5 billion consisted of box office revenue and $32.3 billion consisted of home entertainment revenue, which includes revenue from streaming services, digital rentals and purchases, and disc sales. Interestingly, 80% of this home entertainment revenue comes from digital streaming only (Motion Picture Association, 2021). Therefore, it makes sense that film companies would want to utilize this digital distribution space for their feature films and reach the home entertainment market.

In recent years, it has become more common for films to be released in the box office and then, shortly after, be released to a streaming platform for at-home viewing. While some film companies still wait for the movie to leave the box office before releasing it to streaming, films are also being released to streaming while they are still currently showing in theaters. A similar phenomena occurs with releasing theatrical films to home entertainment markets in general. This behavior leads to a natural inclination to understand what the basis for this decision is and how the film companies are moving to streaming and home entertainment channels without being overly concerned for their box office revenue.

The purpose of this paper is to explore if there is a proportion of total box office revenue at which a movie that is still currently showing in theaters is additionally released to streaming and to home entertainment. The underlying business importance to determining if there is a standard proportion of box office revenue is to use the process that is already being used by film companies in the market currently. By looking at the current players in the market, a new film company can try to employ similar tactics to create a similar level of success.

This approach, however, does not address the innate differences in the characteristics of theatrical movies that can influence the decision to release a movie to streaming or home entertainment. In order to address this gap, this paper will additionally look into using movie characteristics to understand how these certain movie features impact the release of a theatrical film to home entertainment channels, which will exclusively include digital purchases, digital rentals, television

rentals, and disc sales for the purpose of this analysis.

This paper focuses on three different analyses regarding theatrical movies and their release to additional distribution channels, including subscription-based streaming platforms and home entertainment channels. This analysis discovers that theatrical movies are released to streaming when they reach around a standard proportion equal to 0.9917607 of the total box office revenue. Theatrical movies that were released to home entertainment channels did not show to be released at a standard proportion of total box office revenue. However, the implementation of the proportional hazards model highlights that other movie characteristics such as MPAA rating, running time, year, month, day of the week, genre, and whether the film was released before the Oscars each year are important factors that impact the timing of home entertainment release. The results of this analysis and interpretation can provide new studios and distributors with guidance and understanding of when to release theatrical movies to streaming and to home entertainment.

Institutional Information

Before continuing with the analysis process, it is important to note a few key distinctions in the film industry. The film industry has a general process of idea development, financing, production, marketing and distribution, and exhibition. Film production companies are only responsible for the financing aspect and filmmaking process. After a final product is created, film distributors will market and sell the film to the theaters and streaming services for viewing. However, movie studios are responsible for both the production and distribution process (Greenwald and Landry, 2022). There are five major movie studios in the United States that make up 80% of domestic box office revenue which consist of Universal Pictures, Warner Bros., Paramount Pictures, Walt Disney Studios, and Sony Pictures. Each of these major movie studios have a streaming service (Peacock, Max, Paramount+, Disney+ respectively) except for Sony Pictures.

Literature Review

In related literature, the analysis of other distribution methods for films and their intersection with the typical theatrical release window is limited to DVD sales, DVD rentals, and at-home digital purchases. While these distribution options vary considerably from subscription-based streaming services, other researcher's analysis on these distribution methods and how they relate to the theatrical release window for films can still provide insight into this topic in addition to the knowledge gained regarding the home entertainment channels.

An article titled "Optimal Inter-Release Timing for Sequentially Released Products", which was published in the journal *Consumer Needs and Solutions*, focused on theatrical releases and the corresponding DVD releases that occurred after the movie had already left the box office for a sample of movies. This research took a consumer utility-based approach to their analysis due to the availability of both box office revenues and DVD sales data. The findings of this research show movies that consistently perform well in the box office are released to DVD later than the movies that do not consistently generate substantial revenue in the box office (Luan and Sudhir, 2022). This implies that box office revenues could be acceptable to use as a metric for the timing of film release to different distribution channels, which will be explored here. This paper also provides information regarding the movie characteristics that impact the timing from theatrical release to DVD release, which is a starting point understanding timing to all home entertainment channels.

Another article published in the *Journal of the Academy of Marketing Science*, titled "Exclusivity Strategies for Digital Products Across Digital and Physical Markets", focused on digital purchases, digital rentals, and DVD purchases to determine if the sequential, staggered release of digital versions before the release of physical DVD versions of films provides different revenue results than the typical method of release to digital purchasing/rentals and DVDs simultaneously. The researchers found that having a film in digital purchasing or digital rental form exclusively for a period of time before releasing the film on DVD had a positive impact on overall revenue (Seifert et al., 2023). Exclusively showing a film through one distribution channel for a period of time providing more total revenue inspires the current investigation into how box office revenues

can be used as an indicator for the release of a film to streaming and home entertainment.

Overall, the literature lacks a robust analysis of the timing of a film's release to streaming services and home entertainment when the additional release overlaps with the theatrical showing. This shows that additional research and analysis is needed on this topic, especially as the popularity of streaming continues to grow in the future and the theatrical exclusivity window is shortening.

Economic Theory

For the particular question being investigated, typical economic models are difficult to apply directly. Therefore, the basis of this research will be exploratory and require thinking about the topic using a loose application of a two-armed bandit model to frame this question and understand the analysis process.

Multi-armed bandit models, often referred to as k-armed bandit models, deal with there being multiple choices to choose from at a given moment, but the result of the choice is not known. The decision maker must learn from each decision they make and the outcomes of those decisions to try to inform themselves on the possible outcomes (Lattimore and Szepesvári, 2020). Therefore, for a two-armed bandit situation, there are two different options to choose between for each moment and a choice must be made by the decision maker without having full knowledge of what to expect in the outcome.

Applying this concept to the topic at hand, the two choices that film distributors face each day that the film is showing in the box office are whether to keep the film exclusively in theaters or to show the film in theaters and through streaming/home entertainment at the same time. The film distributors and movie studios do not know for certain what the outcome of this decision will be on the movie's revenue, so each day they must decide if they are willing to go to streaming or home entertainment yet. However, unlike in the typical examples of multi-armed bandit models, the distributors and studios do not have the ability to continue making the choice between theater-exclusive showing and simultaneous multi-channel showing after they choose to move to

streaming/home entertainment, at least not on the daily frequency they had prior to the decision to additionally release to additional channels and with the same ease.

Therefore, the theory behind this exploration is to use the daily box office revenues for a set of movies and see if there is a standard proportion of this total box office revenue where movie studios and film distributors are deciding to make the jump and switch from choosing theater-exclusive showing to simultaneous theater and streaming/home entertainment showing.

It is important to note that this concept uses the total box office revenue earned for each movie, which is only acquired after the film in no longer in the box office. This means that if there is a standard proportion of box office revenue that distributors release films to streaming at, then the distributor must determine a reliable way to forecast their total box office revenue from the current day onwards. This is not what the scope of this particular project will be focusing on, but it is important to understand how the new film distribution firms and studios can actually implement this approach.

Statistical Model

To further investigate the timing of theatrical movie release to home entertainment, a Cox proportional hazard model with LASSO (least absolute shrinkage and selection operator) regression was implemented. The proportional hazard model is a model within the overall topic of survival analysis, which focuses on the timing between the start of observation and a subsequent event occurring.

The proportional hazard model adds another layer by not only allowing to analyze the time until an event of interest occurs, but also factors, commonly referred to as covariates, influencing the timing as well (Fox and Weisberg, 2023). The Cox proportional hazard model in particular is applicable to this analysis because it is able to handle right-censored data, which occurs in the data because not all movies have been release to home entertainment at the time of this analysis (Cox and Oakes, 1984). By performing analysis using this model, the extent of influence the covariates

have on the timing for home entertainment release can be determined.

The LASSO regression, or L1 regularization, performed along with the proportional hazards model helps to create a more significant model. The function of LASSO is to increase the predictive accuracy of the model and to make the model output easier to interpret (Tibshirani, 1996). This is enacted by shrinking down some of the coefficients and setting some coefficients to zero in order to keep only the most significant covariates.

Empirical Model

Box Office Revenue Proportions at Additional Release

The empirical specification for determining if there is a standard proportion of box office revenue at which distributors are releasing films to streaming/home entertainment looked at a sample of movies for a total of N movies that are each in the box office for a different number of total days, which will be notated as T_{max}. The analysis process begins with obtaining the revenue generated each day that the movie is in the box office.

Box office revenue for movie *n* on day *t* of theatrical showing:

$$R_t^n$$

$$n = 1, 2, ..., N \text{ movies}$$

$$t = 1, 2, ..., T_{max}^n \text{ days in box office for movie } n$$

The revenue for each day was used to calculate a total sum of the box office revenue for all days that the film was showing in theaters. This involved taking the sum of daily revenues up to the maximum number of days that the film was in theaters as follows.

Total box office revenue for movie *n* for all days of theatrical showing, T_{max}, for movie *n*:

$$\sum_{t=1}^{T_{max}^n} R_t^n$$

Once the total box office revenue for movie *n* was found, the daily box office revenue was used in order to find the proportion of total box office revenue that was earned for each day, which will be referred to as P_t^n.

Proportion of total box office revenue earned on day *t* of theatrical showing for movie *n*:

$$P_n^t = \frac{R_t^n}{\sum_{t=1}^{T_{max}^n} R_t^n}$$

To get a final function for the accumulation of box office revenues for each movie over time, the proportion of total box office revenue for each day is summed.

Proportion of total revenue movie *n* at time *t*:

$$F_t^n = \sum_{t=1}^{t} P_t^n$$

$$0 \leq F_t^n \leq 1$$

Once the function F_t^n was found, then $F_{\rho_n}^n$ indicated the proportion of total box office revenue when ρ_n = time of film release to streaming (or home entertainment release) for movie *n*. For simplicity of symbolization, let $F_{\rho_n}^n = q_n$. This gives each movie a time of release (ρ_n) and a proportion of revenue (q_n) that was plotted and the distribution of values were observed and further analyzed.

If there is a relatively wide range of values for ρ_n, but the values for q_n fall in a small range, this would indicate that there is a similar situation to a control limit policy taking place and film distributors are waiting until box office revenues reach a certain proportion of a forecasted total box office revenue, assuming that the box office revenue is appropriately forecasted by the firm. Once this threshold of total box office revenue is reached, then the distributor is deciding to switch from theater-exclusive distribution to theater and streaming distribution.

If there is not a small range of values for q_n, then this indicates that there are additional variables that impact the time to additional release. This indicates a need for further research on what is

impacting the timing to additional streaming/home entertainment release.

Proportional Hazards Model

For the proportional hazards model, a matrix of covariates that could impact the the timing of theatrical movies being released to home entertainment channels is input. This is used in conjunction with the baseline hazard function determines the coefficients for the covariates and the hazard function as shown below.

$$h(t|x) = h_0(t)\exp(x\beta)$$

$$h(t|x) = \text{Hazard function at time } t \text{ given predictors } x$$
$$h_0(t) = \text{Baseline hazard function}$$
$$\exp(x\beta) = \text{Exponential function of covariates and coefficients } x\beta$$

The baseline hazard function does not need to be determined when using the Cox proportional hazards model in order to determine the coefficients of the covariates. The coefficients that are obtained as a result of this model indicate how much the respective covariate is impacting the hazard function. If the coefficient is close to zero, this indicates that the covariate has a small impact on the hazard.

Empirical Results

Streaming Data - Box Office Revenue Proportions

The sample used to determine if there is a proportion of box office revenue that theatrical movies release to streaming at contained 125 movies that have gone to a streaming platform. Out of these 125 movies, 17 movies were released to a streaming platform while they were still showing in theaters. The remaining movies were released to streaming some point after the movie had already been removed from theaters.

	Movie	Day of Release	Proportion of Revenue
1	Paw Patrol : The Mighty Movie	46	0.99057590
2	Dungeons & Dragons: Honor Among Thieves	46	0.98943583
3	Ant-Man and the Wasp: Quantumania	89	0.99858898
4	Theater Camp	62	0.97935707
5	Haunted Mansion	68	0.99358585
6	The First Omen	53	0.99998024
7	Barbie	147	0.99995542
8	Saltburn	35	0.94569471
9	The Holdovers	62	0.88150329
10	The Super Mario Brothers. 2023	120	0.99864288
11	Elemental	89	0.99860276
12	The Meg 2: The Trench	56	0.99800756
13	Champions 2023	49	0.99631389
14	Black Panther: Wakanda Forever	82	0.99851755
15	Guardians of the Galaxy Volume 3	89	0.99990138
16	Asteroid City	56	0.98894098
17	Transformers:Rise of the Beasts	46	0.99207085

Table 1: Total box office proportion (q_n) at time of streaming release (ρ_n) for theatrical movies

The day of release following theatrical opening (ρ_n) varied considerably, with the shortest release window being only 35 days and the longest being 147 days. As shown in Table 1, the majority proportion of total box office revenue (q_n) at day of streaming release fall within 0.98-0.99+. However, there were two movies, *Saltburn* and *The Holdovers*, that were released at a proportion of 0.88 and 0.94 respectively. The mean proportion that movies were released to streaming at was 0.985275, which is taking into account the two movies that released at lower proportions. Without incorporating these two movies, the mean proportion of box office revenue that movies are released to streaming at is 0.9917607. The distribution of ρ_n and q_n values for the streaming sample can be found in Figure 1.

For simplicity, only four plots of the cumulative box office revenue have been presented here. The remaining 13 movies that do not have plots shown look similar to the *Black Panther: Wakanda Forever* plot that is shown in Figure 2. This is representative of the movies that were released to streaming at the 0.98-0.99+ proportions.

The remaining three plots have been chosen because they represent the atypical cases. Figure

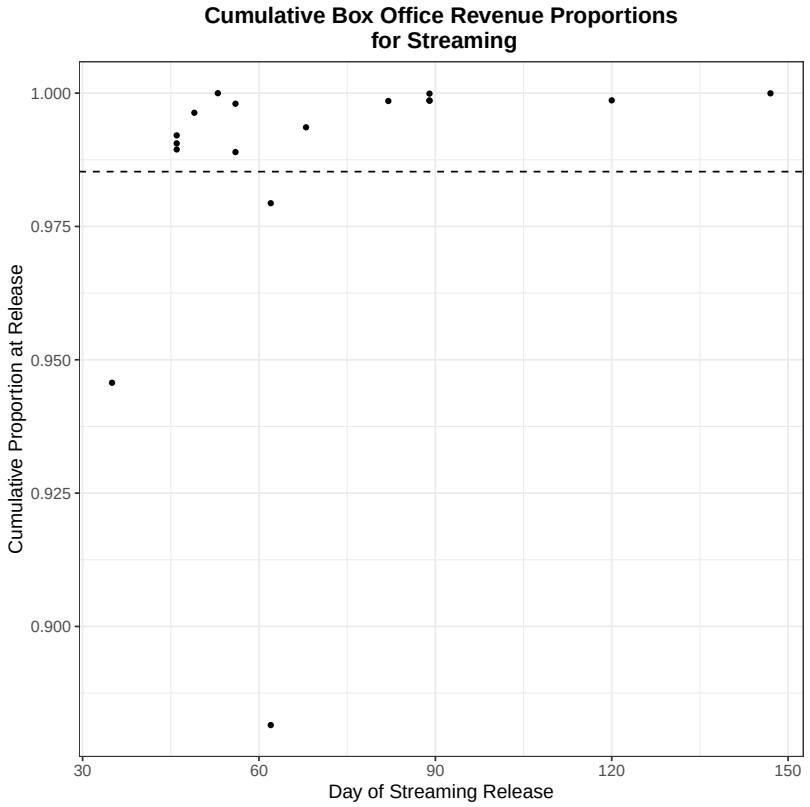

Figure 1: Proportions of total box office revenue (q_n) at day of streaming release (ρ_n). The mean proportion is represented by the dotted line.

3 shows that *The Holdovers* experience slow growth in box office revenue in the beginning, while also having a sudden increase in box office revenue towards the end of the theatrical run. This could likely have been caused by this movie being an Oscar nomination during the time of the increase in revenue. In Figure 4, slow box office revenue growth was seen in the beginning, as well as a noticeably slower revenue growth rate compared to the movies represented by *Black Panther: Wakanda Forever*. This difference in the proportion for *Saltburn* could be due to this film being released to a small selection of theaters initially and the controversial nature of the movie's content. Lastly, *Theater Camp*, which had the lowest proportion, excluding *Saltburn* and *The Holdovers*, is shown in Figure 5. This again shows slower total box office revenue growth compared to the movies that were released in the 0.98-0.99+ range.

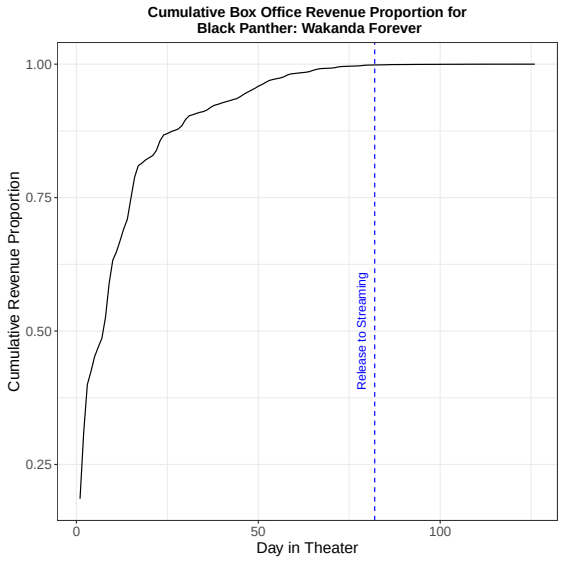

Figure 2: Box office revenue graph for Black Panther: Wakanda Forever

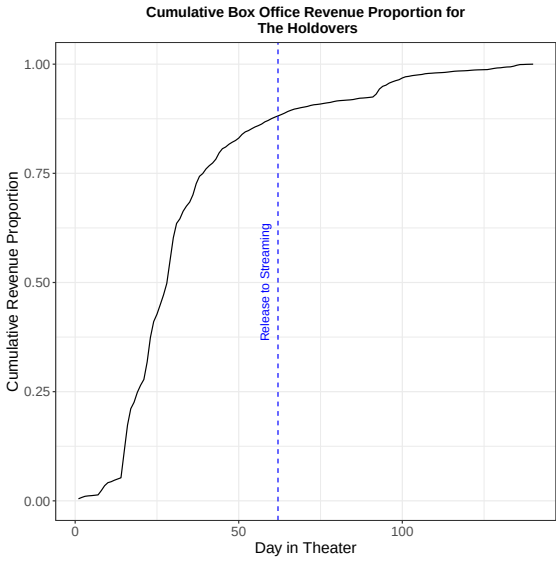

Figure 3: Box office revenue graph for The Holdovers

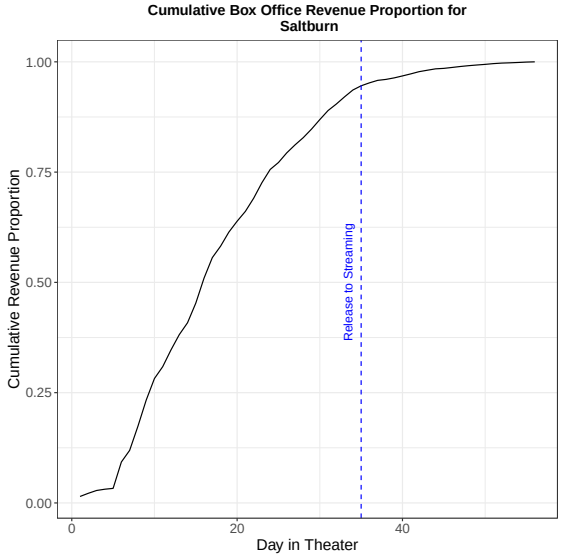

Figure 4: Box office revenue graph for Saltburn

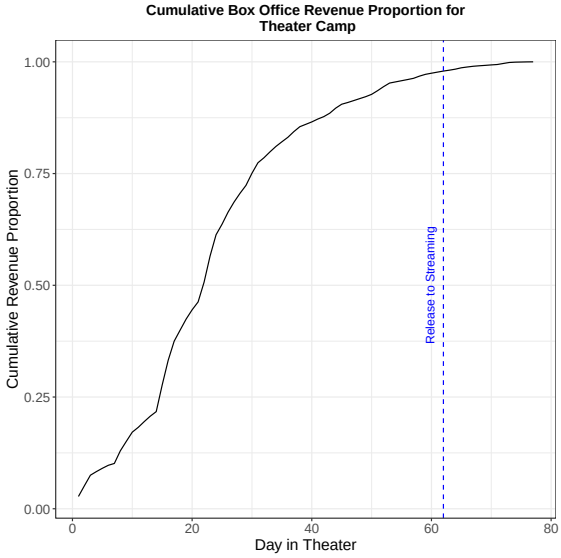

Figure 5: Box office revenue graph for Theater Camp

Figure 6: Sample of graphs of box office revenue with day of streaming release (ρ_n) marked by the dotted line. The corresponding cumulative revenue proportion (q_n) can be found on the y-axis.

Home Entertainment Data - Box Office Revenue Proportions

For determining if there is a proportion of total box office revenue that theatrical movies go to home entertainment channels, a sample of 268 movies that were in the box office between 2021 and 2024 was used. Out of the 268 total movies, 116 movies were released to home entertainment while still showing in theaters. The remaining movies were either released to home entertainment after they were removed from theaters or they were not released to home entertainment at all at the point of data collection.

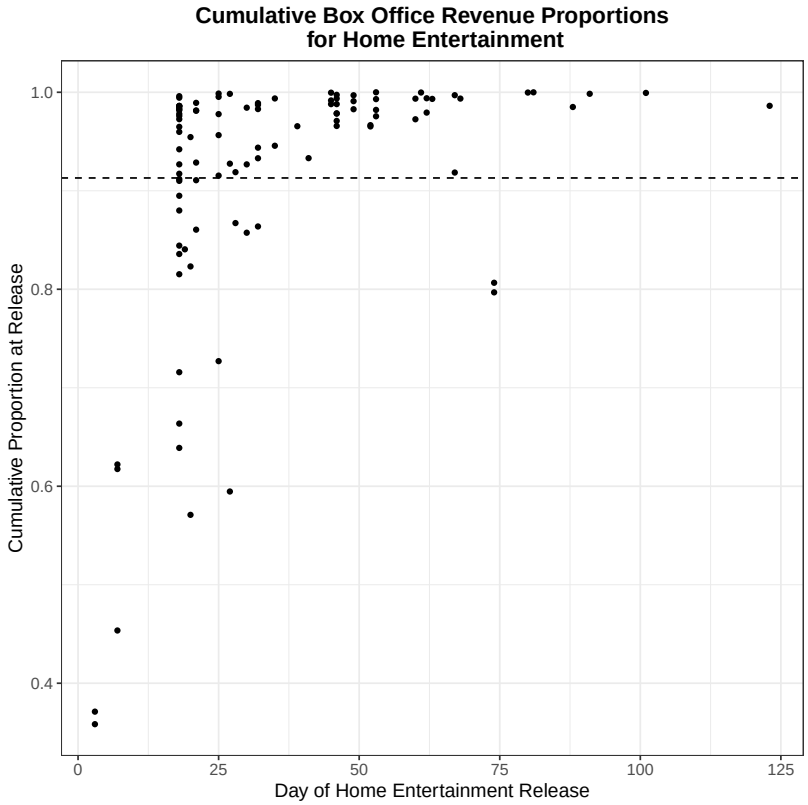

Figure 7: Proportions of total box office revenue (q_n) at day of home entertainment release (ρ_n). The mean proportion is represented by the dotted line.

The day of home entertainment release, ρ_n, ranged from 0, meaning the movie was released simultaneously to theaters and home entertainment, to 123 days. The proportion of total box office revenue, q_n, for this sample ranged from 0.35 to 0.99+, excluding the movies that were simultaneously released to theaters and home entertainment. Due to the complete table of day of release and proportion of total box office revenue being quite long, refer to the Additional Tables

section to see this table. However, the distribution plot of these values represents the data well, which is shown in Figure 7.

The distribution of ρ_n and q_n shows that both of these values are dispersed across a relatively wide range of values, which is a key difference compared to the distribution of these values for the streaming movies sample. The mean proportion that theatrical movies were released at was 0.91229847, but this is not a good estimation for determining if there is a standard proportion considering the distribution of ρ_n and q_n.

For the same reasons stated prior, only four plots of individual movies' cumulative box office revenues and day of home entertainment release are shown here. These plots have been chosen because they are representative of a variety of ρ_n and q_n values.

Figure 8 illustrates a similar plot to those in the streaming movies sample that were released at a proportion of 0.99+ for *Peter Rabbit 2: The Runaway* being released to home entertainment. However, *Mrs. Harris Goes to Paris*, shown in Figure 9, has a similar trend in cumulative revenue, but this occurs over a greater period of time. Interestingly, this movie was released to home entertainment at a much lower proportion of revenue.

Looking at the plot for *Crisis* in Figure 10 and the plot for *Blithe Spirit* in Figure 11, it is clear that these movies do not follow a similar trend in their cumulative revenue growth over time and their proportion values, q_n, are quite different.

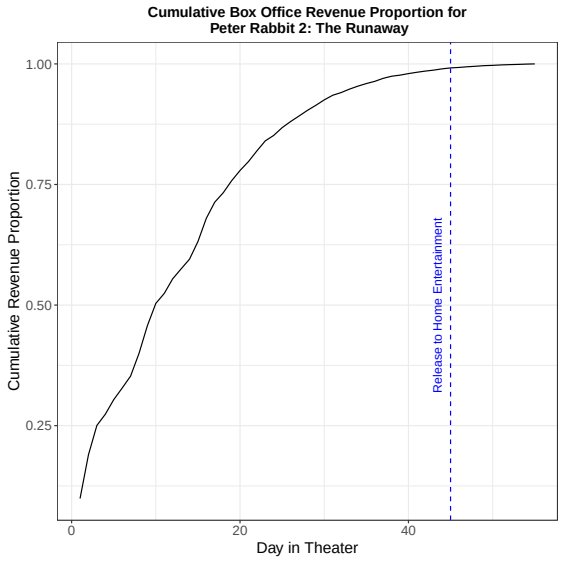

Figure 8: Box office revenue graph for Peter Rabit 2: The Runaway

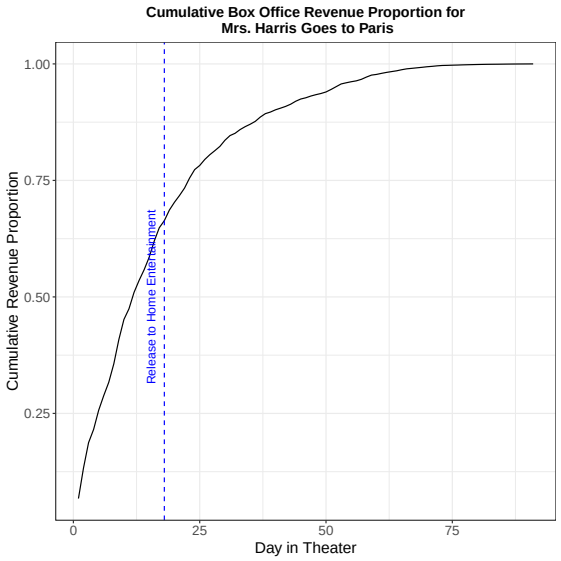

Figure 9: Box office revenue graph for Mrs. Harris Goes to Paris

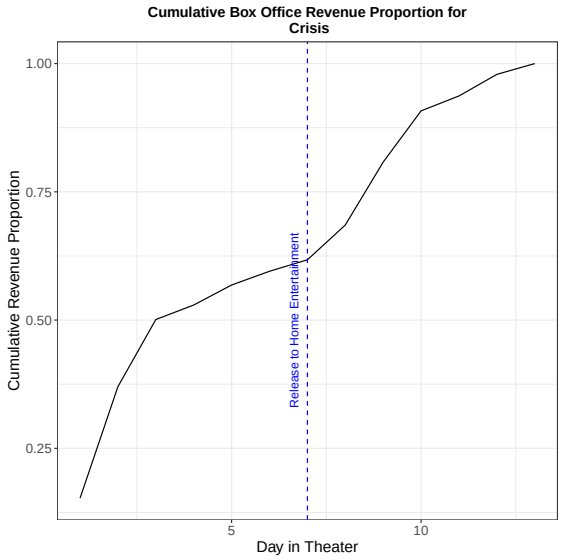

Figure 10: Box office revenue graph for Crisis

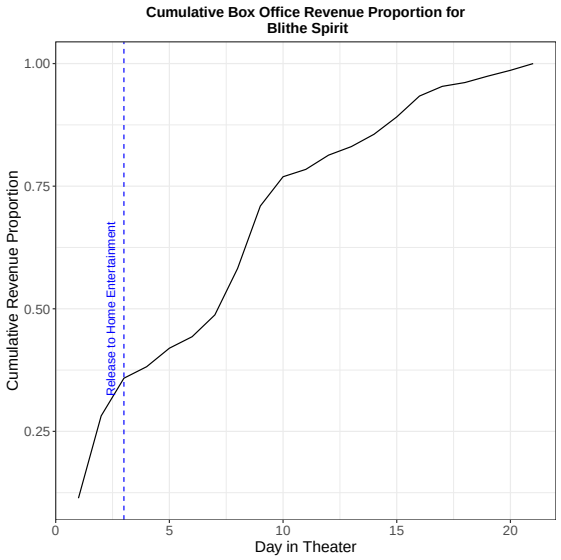

Figure 11: Box office revenue graph for Blithe Spirit

Figure 12: Sample of graphs of box office revenue with day of home entertainement release (ρ_n) marked by the dotted line. The corresponding cumulative revenue proportion (q_n) can be found on the y-axis.

Home Entertainment Data - Proportional Hazards Model

The sample for the proportional hazards model consisted of 233 movies from the larger home entertainment sample explained above. The data was split into training data, which the coefficient results presented here are based off of, and testing data in order to implement model on new data and see the predicted hazard ratios for films. The covariates originally implemented included: distributor, opening theater count, MPAA rating, running time, year of theatrical release, month of theatrical release, day of the week, season, proportion of opening box office revenue out of total box office revenue, genre. Season encompassed popular movie-going time of year including the holiday season, summer blockbusters, and the months surrounding the Oscars.

The coefficients of the Cox proportional hazards model were estimated using the *survival* library and the *glmnet* library in R. The *survival* library provided the ability to input the number of days until home entertainment release, or the number of days since theatrical release for movies that did not go to home entertainment yet, and a binary indicator of the event occurrence to create a survivor object.

This survivor object is used by the *glmnet* library using the cv.glmnet function, which automatically performs cross-validation to allow for selection of the optimal regularization parameter. The regularization parameter is important because it determines how strongly the coefficients are going to be shrunken towards zero. For this sample, the regularization parameter set to 0.06143375. The glmnet function was able to fit the Cox proportional hazards model while implementing LASSO regression to determine which covariates impacted the hazard function.

The final model only included select values for distributor, MPAA rating, year, month, day of week, season, and genre. More specifically, the distributors that impacted the hazard function in the final model were: Atlas Distribution Company, Bleecker Street Media, Crunchyroll, FUNimation Entertainment, GKIDS, Open Road Films (II), Sony Pictures Classics, Toho International, United Artists Releasing, Universal Pictures, Warner Bros., Walt Disney Studios Motion Pictures. MPAA ratings that were significant in the model were none and R. The year of releases that impact the model results are 2021, 2022, and 2024 and the months included were March, July, and December.

The two days of theatrical release that were kept in the model were Saturday and Tuesday. The time leading up to the Oscars was also included, along with the history, action, biography, and documentary genres. The resulting coefficients of the significant covariates are presented in Table 2.

The covariates that were included in the original model but not in the final model were determined by the LASSO regression to not be significant or needed in the optimal proportional hazards model.

	Coefficient
distributorAtlas Distribution Company	-0.78752680
distributorBleecker Street Media	0.84140422
distributorCrunchyroll	-1.34483493
distributorFUNimation Entertainment	-0.93243362
distributorGKIDS	-0.52450620
distributorOpen Road Films (II)	-1.19728314
distributorSony Pictures Classics	-0.39321143
distributorToho International	-0.77021514
distributorUnited Artists Releasing	-0.32967615
distributorUniversal Pictures	0.46853499
distributorWalt Disney Studios Motion Pictures	-0.05660818
distributorWarner Bros.	0.10652211
mpaa_ratingNone	-4.17985242
mpaa_ratingR	0.37739950
running_time	-0.00771794
year2021	-0.28693075
year2022	-0.31260725
year2024	0.12331083
month3	0.05061439
month7	-0.19394819
month12	-0.31538706
day_of_week_nameSaturday	-0.13511701
day_of_week_nameTuesday	-0.10116971
seasonOscar Season	0.00827121
History	-0.24923803
Action	0.18342400
Biography	-0.03173186
Documentary	-0.34499214

Table 2: Model covariates that impact the time to home entertainment release with respective coefficients

The survival curve that is utilizing the training data from the model can be found in Figure 13. This plot illustrates the probability of a movie staying exclusively in the theaters and not being released to home entertainment after the theatrical run is over as time progresses after the initial theatrical release.

Additionally, the results of predicting the hazard ratios for the testing data movies can be found in the Additional Tables section at the end of this paper. It has been omitted due to the length of the table and this additional analysis not being the primary focus of the proportional hazard model.

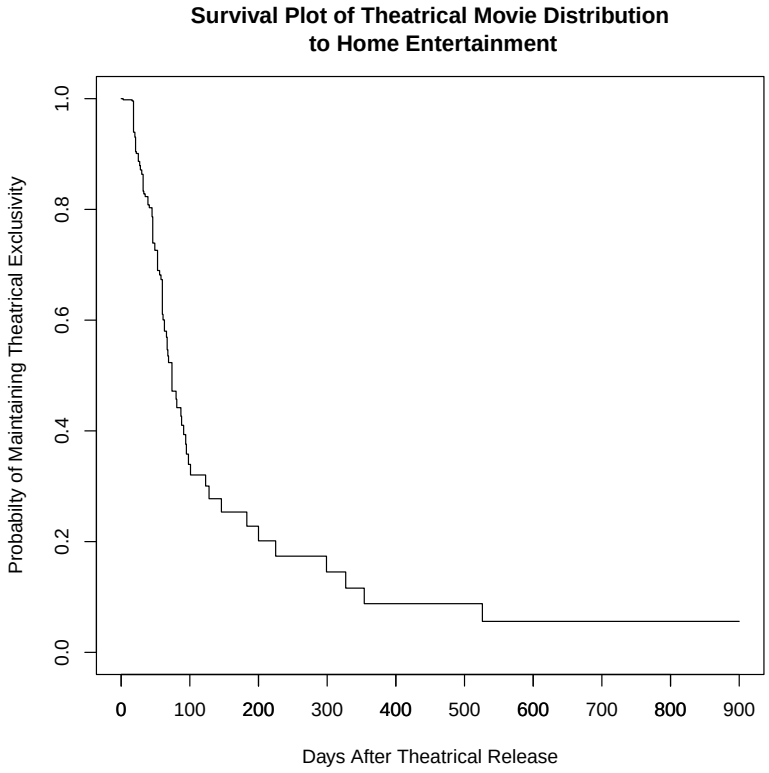

Figure 13: Survival curve representative of the aggregate final model data

Interpretation and Discussion

Streaming Data - Box Office Revenue Proportions

In order to interpret the results of the streaming data analysis, the removal of the two movies with lower proportions at streaming, *Saltburn* and *The Holdovers*, is necessary due to their unique circumstances explained prior. With that being said, there does appear to be a standard proportion of total box office revenue at which theatrical movies are being released to subscription-based streaming platforms. This proportion is 0.9917607 of total box office revenue.

This means that film distribution firms could indeed be enacting a similar policy to a control limit policy using box office revenues when determining the decision to switch to theatrical and streaming distribution channels. Since the films are being released close to 100% of their total box office revenue, this indicates that firms are potentially forecasting their predicted total box office revenue, or forecasting the future revenue at continuous points in the theatrical showing window, in order to release to streaming when the box office revenue is nearly maximized.

Using these results and interpretation, new movie studios and film distributors should aim to first reliably forecast their total box office. While this is not the focus of this project, this step is important in being able to determine when the proportion of total box office revenue reaches around 0.9917607. Once the cumulative box office revenue for the movie approaches this proportion, distributors should additionally release theatrical films to a streaming platform, if streaming is determined to be an avenue the producer and distributor are open to. Implementing this control limit-like policy ensures that new movie studios and film distributors are following the industry standard set by the large studios and distributors that have already successfully distributed movies.

However, it is important to note that these movies were sampled based off availability of the streaming release dates, which proved to be challenging to gather. For this reason, additional collection of streaming release dates and analysis may be required to provide a more robust analysis of the timing of theatrical movie release to subscription-based streaming platforms.

Home Entertainment Data - Box Office Revenue Proportions

In comparison to the streaming data, the home entertainment data sample does not indicate that there is a standard proportion of total box office revenue that movies are released to home entertainment channels at. While the mean proportion at home entertainment release is 0.91229847, this is not representative of a standard proportion since there is a wide range of proportions, q_n, that films were released to home entertainment at. Also, there is more variability in the cumulative revenue proportion trends within this sample.

This shows that there is not a control limit-like policy taking place when it comes to the current movie studios and film distributors determing when to release a theatrical film to home entertainment channels. This means that while new firms can forecast their box office revenue to use for streaming release timing, the box office revenue alone is not going to be a significant indicator to use for home entertainment release timing if these new firms want to follow the example set by the current players in the market.

Based off these results and interpretation, studios and distributors should aim to use other variables and metrics in order to determine when to release a movie currently in theaters to additional home entertainment channels. The results of this analysis inspired the inclusion of the proportional hazards model using the home entertainment data, which is one way for new studios and firms to determine which factors influence a film being released to home entertainment.

Home Entertainment Data - Proportional Hazards Model

While the covariates that were relevant to the final model were listed prior, an explanation of the interpretation of the coefficients found in Table 2 is still important for understanding the results presented here. A positive coefficient can be interpreted as that covariate, or an increase in the value of that covariate for numeric covariates, having an increased hazard of being released to home entertainment; hence, the presence, or increase in, these covariates in the data means the particular movie will be potentially released to home entertainment sooner. A negative coefficient indicates that the covariate, or an increase in the covariate for numeric covariates, will lead to a

decreased hazard of home entertainment release; therefore, the movie will be released to home entertainment later, if at all. Coefficients that have an absolute value closer to zero impact the hazard less than coefficients that have an absolute value greater than zero.

The covariates that increase the hazard were: Bleecker Street Media (distributor), Universal Pictures (distributor), Warner Bros. (distributor), R (MPAA rating), 2024 (year), March (month), Oscar season, and action (genre). The impact on the hazard by all of these covariate values are small to moderate due to their coefficients being close to zero, with the exception of Bleecker Street Media (distributor). However, these covariates are still important for understand the hazard of release to home entertainment.

The distributor covariates that decrease the hazard were: Atlas Distribution Company, Crunchyroll, FUNimation Entertainment, GKIDS, Open Road Films II, Sony Pictures Classics, Toho International, United Artists Releasing, Walt Disney Studios Motion Pictures. Additional covariates that decrease the hazard were: running time, none (MPAA rating), 2024 (year), July (month), December (month), Saturday (day of week for theatrical release), Tuesday (day of week for theatrical release), history (genre), biography (genre), documentary (genre). The majority of these covariates had a moderate-to-strong impact on the hazard.

The three covariates that impact the hazard the most are FUNimation Entertainment (distributor), Open Road Films II (distributor), and having no MPAA rating. These covariates all reduce the hazard, which indicates that movies that have these covariate values will be released to home entertainment later or not at all.

Based off these results and interpretation, new movie studios and film distributors do need to take into account more than just box office revenues when determining the timing of home entertainment release for theatrical films. MPAA rating, running time, year of release, month of release, day of the week for the release, genre, and whether the film was released before the Oscars are all important when it comes to determining the timing of home entertainment release.

These results and model can be used to determine the hazard ratio relative to the baseline hazard, where a hazard ratio greater than one means an increased hazard compared to the baseline

hazard and a hazard ratio less than one means a decreased hazard compared to the baseline. An example of hazard ratios that can be obtained can be found in Additional Tables. However, it is important to note that the baseline hazard is not known at this point. If a new distributor or movie studio wants a more usable metric, the baseline hazard must be found prior.

Different distributor covariate values being incorportated into the final proportional hazard model indicate that these distributors may have a unique way of determining the time to home entertainment release. Additional analysis should be done to exclude distributors as a covariate, as well as including additional covariates such as star power, which is a measure of celebrity presence in a film, and critic reviews of the film. These additional variables may provide additional insight into the factors that impact the hazard.

Conclusion

Throughout this project, many skills were obtained and enhanced in order to conduct this analysis. These skills includes webscraping multiple websites for data, data cleaning, coding in R, coding in Python, data visualization techniques, and automation. While many of these necessary skills were obtained prior to the start of this analysis, with the exception of webscraping which was a new skill gained, working through the analysis required proficiency in each of these areas.

Overall, this analysis revealed theatrical movies can be considered to be released to streaming at a standard proportion of total revenue equal to 0.9917607. This knowledge can be used by new, developing film distribution companies and movie studios, if they forecast their total box office revenue, to follow the streaming release patterns of the already successful competitors.

For theatrical movies released to home entertainment channels, which include digital purchases, digital rentals, television rentals, and disc sales for this analysis, there was not found to be a standard proportion of total box office revenue that movies are released at. These results indicate that additional movie characteristics are requried to understand this timing decision if a new film distributor is aiming to follow the key studios and distributors in the market.

To begin further investigating the timing of home entertainment release, the Cox proportional hazard model with LASSO regression was implemented. The model found that MPAA rating, running time, year of release, month of release, day of the week for the release, genre, and whether the film was released before the Oscars impacted the hazard compared to the baseline hazard. This shows that additional characteristics, besides just box office revenues are impacting the timing of home entertainment release for theatrical movies.

Acknowledgements

Special thanks to Dr. Harry J. Paarsch for being the primary advisor on this project and throughout the business analytics program. His feedback throughout this process and expertise in the field set me up for success throughout the capstone courses that this project was produced during.

I would like to thank the other professors that were part of the business analytics curriculum for my cohort. Without them, I would not have been ready to take on such a task. My fellow classmates also deserve appreciation for their guidance and constructive criticism throughout the semester.

Lastly, I would like to thank my parents and close friends. Their constant support throughout this project truly made a huge difference.

Data Appendix

Streaming Movie Data

The streaming movie data consists of theatrical movies that were released to a streaming service at some point after theatrical release. The sample was obtained by manually recording the streaming dates, listed as SVOD, from whentostream.com (When to Stream, 2024). This was necessary due to the limited availability of accurate streaming release dates from other sources. The box office revenue and box office dates for each movie were gathering by web scraping Box Office Mojo by

IMDb Pro (IMDb, 2024). Movies that did not have daily frequency data were omitted from the final sample of streaming movies.

Release date is the date that the movie was release the theaters, while stream date is the date that movies were released to a streaming platform. Box office window is a binary variable that indicates whether the movie was released to streaming while it was still showing in the box office. This variable was filtered in order to only look at movies that were released in the box office window, and hence, had a value of one. The days between variable was found by calculating the number of days from theatrical release to streaming release, or ρ_n for this analysis.

The cumulative proportion is the proportion of total box office revenue for each movie on the day of streaming release, or q_n in this analysis. The values for this variable were calculated using the cumsum() function in the pandas package in Python. This was used in conjunction with the total revenue for each movie that was found using sum() in base Python to find the daily proportion of total box office revenue. This method adheres to the formulae found in the Empirical Model section of this paper. The revenues and dates did not require further transformation.

Data courtesy of IMDb.

release_date	streaming_date	boxoffice_window	days_between	cumulative_proportion
Min. :2022-11-11	Length:17	Min. :1	Min. : 35.00	Min. :0.8815
1st Qu.:2023-04-05	Class :character	1st Qu.:1	1st Qu.: 49.00	1st Qu.:0.9894
Median :2023-06-16	Mode :character	Median :1	Median : 62.00	Median :0.9963
Mean :2023-06-28		Mean :1	Mean : 70.29	Mean :0.9853
3rd Qu.:2023-08-04		3rd Qu.:1	3rd Qu.: 89.00	3rd Qu.:0.9986
Max. :2024-04-05		Max. :1	Max. :147.00	Max. :1.0000

Table 3: Summary table of variables used to find proportion of total box office revenue for streaming sample

Home Entertainment Movie Data

The home entertainment movie data consists of theatrical movies that were released to a home entertainment distribution channel on the same day of, or following, theatrical release. The sample was obtained by gathering a list of movies that were in the box office between the years 2021-2024

from web scraping Box Office Mojo by IMDb. This list was randomly sampled to obtain the final sample of movies that would be used for the analysis. The box office revenue and box office dates for each movie were obtained through the same web scrape methods as used for the streaming data. Movies were omitted if they did not meet the daily frequency requirement for their revenues.

Following the same method as was used for the streaming data, release date is when the movie was released in theaters and home date is the date the movie was released to home entertainment. Box office window was set equal to one only, indicating that the movie was released to home entertainment while in the theaters still. The days between variable was found by calculating the number of days from theatrical release to home entertainment release. The value is referred to as ρ_n for this analysis.

The cumulative proportion, or q_n in this analysis, is the proportion of total box office revenue for each movie on the day of streaming release. The values for this variable were, again, calculated using the cumsum() function in the pandas package in Python along with the total revenue for each movie that was found using sum() in base Python. This gives the daily proportion of total box office revenue for each movie. This method adheres to the formulae found in the Empirical Model section of this paper. The revenues and dates did not require further transformation.

Data courtesy of IMDb.

release_date	home_date	boxoffice_window	days_between	cumulative_proportion
Min. :2020-12-04	Min. :2020-12-23	Min. :1	Min. : 3.00	Min. :0.3584
1st Qu.:2022-02-18	1st Qu.:2022-04-05	1st Qu.:1	1st Qu.: 18.00	1st Qu.:0.9107
Median :2023-04-21	Median :2023-05-16	Median :1	Median : 28.00	Median :0.9727
Mean :2022-12-23	Mean :2023-01-28	Mean :1	Mean : 35.85	Mean :0.9130
3rd Qu.:2023-09-01	3rd Qu.:2023-10-24	3rd Qu.:1	3rd Qu.: 49.00	3rd Qu.:0.9890
Max. :2024-06-14	Max. :2024-07-09	Max. :1	Max. :123.00	Max. :1.0000

Table 4: Summary table of variables used to find proportion of total box office revenue for home entertainment sample

Proportional Hazards Model Data

The data that was used for the proportional hazards model consisted of a subsample of the home entertainment movies. In addition to the box office revenue and box office dates, the date of home entertainment release was gather by web scraping Rotten Tomatoes (Rotten Tomatoes, 2024). Movies were excluded from the model data if they were re-releases to the theater.

Additional data for each movie was gathered from Box Office Mojo by IMDb including: opening revenue, opening theater count, distributor, MPAA rating, running time, genre, season. Opening revenue is the box office revenue earned by the movie during the opening weekend, or in the first few days for movies not released near the weekend. Opening revenue was divided by the total revenue in order to get the proportion of total revenue that opening revenue accounted for, which is the proportion opening revenue in this analysis. Opening theater count is the number of theaters that the movie debuted in. MPAA (Motion Picture Association of America) rating is the age and content suitability score and can be: no rating, G, PG, PG-13, R, NC-17. Running time is the length of the movie and was originally presented in hours and minutes. This variable was transformed to be numeric and only consist of total minutes by multiplying the number of hours by 60 and adding the additional minutes listed. The genre data obtained from Box Office Mojo contained multiple genres for each movie; therefore, each genre was turned into a binary variable and made to be a separate covariate. The genre columns include: action, adventure, animation, biography, comedy, crime, documentary, drama, fantasy, horror, music, musical, mystery, romance, sci-fi, sport, thriller, war, and western. Season accounts for the time leading up to the Oscars (January and February), spring (March, April, and May), summer blockbusters (June, July, and August), early fall (September and October), and the holiday season (November and December).

For event, this variable is a binary indicator of the movie being released to home entertainment at the time of data collection. The days between variable is the number of days since theatrical release to home entertainment release for applicable movies. For movies that have not been released to home entertainment yet, is the number of days since theatrical release up to the date of analysis.

Data courtesy of IMDb.

distributor	opening_theater_count	mpaa_rating	days_between
Universal Pictures : 20	Min. : 1	G : 0	Min. : 2.0
Focus Features : 18	1st Qu.: 298	No Rating:12	1st Qu.: 27.0
Sony Pictures Releasing : 17	Median :1407	None :68	Median : 58.0
Warner Bros. : 14	Mean :1758	PG :20	Mean : 206.1
Lionsgate Films : 13	3rd Qu.:3204	PG-13 :52	3rd Qu.: 200.0
Walt Disney Studios Motion Pictures: 10	Max. :4450	R :81	Max. :1192.0
(Other) :141			

event	running_time	year	month	day_of_week_name	Crime
Min. :0.0000	Min. : 73.0	2020: 2	4 :32	Friday :205	Min. :0.0000
1st Qu.:0.0000	1st Qu.: 98.0	2021:56	3 :26	Monday : 2	1st Qu.:0.0000
Median :1.0000	Median :111.0	2022:61	5 :25	Saturday : 2	Median :0.0000
Mean :0.7082	Mean :114.9	2023:75	12 :24	Thursday : 7	Mean :0.1245
3rd Qu.:1.0000	3rd Qu.:125.0	2024:39	7 :20	Tuesday : 1	3rd Qu.:0.0000
Max. :1.0000	Max. :206.0		8 :20	Wednesday: 16	Max. :1.0000
			(Other):86		

season	proportion_opening_revenue	Drama	History
Early Fall :30	Min. :0.0000	Min. :0.0000	Min. :0.00000
Holiday Season :41	1st Qu.:0.2361	1st Qu.:0.0000	1st Qu.:0.00000
Oscar Season :25	Median :0.3574	Median :0.0000	Median :0.00000
Spring :83	Mean :0.3687	Mean :0.4936	Mean :0.09013
Summer Blockbusters:54	3rd Qu.:0.4591	3rd Qu.:1.0000	3rd Qu.:0.00000
	Max. :2.2429	Max. :1.0000	Max. :1.00000

Mystery	Thriller	Action	Adventure	Animation
Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.00000
1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.00000
Median :0.0000	Median :0.0000	Median :0.0000	Median :0.0000	Median :0.00000
Mean :0.1502	Mean :0.2918	Mean :0.2618	Mean :0.2275	Mean :0.09013
3rd Qu.:0.0000	3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:0.0000	3rd Qu.:0.00000
Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.00000

Comedy	Family	Fantasy	Horror	Sport
Min. :0.0000	Min. :0.00000	Min. :0.0000	Min. :0.0000	Min. :0.00000
1st Qu.:0.0000	1st Qu.:0.00000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.00000
Median :0.0000	Median :0.00000	Median :0.0000	Median :0.0000	Median :0.00000
Mean :0.3519	Mean :0.09013	Mean :0.1717	Mean :0.1502	Mean :0.03433
3rd Qu.:1.0000	3rd Qu.:0.00000	3rd Qu.:0.0000	3rd Qu.:0.0000	3rd Qu.:0.00000
Max. :1.0000	Max. :1.00000	Max. :1.0000	Max. :1.0000	Max. :1.00000

SciFi	Biography	Documentary	Music
Min. :0.0000	Min. :0.00000	Min. :0.00000	Min. :0.0000
1st Qu.:0.0000	1st Qu.:0.00000	1st Qu.:0.00000	1st Qu.:0.0000
Median :0.0000	Median :0.00000	Median :0.00000	Median :0.0000
Mean :0.1416	Mean :0.09871	Mean :0.06009	Mean :0.0515
3rd Qu.:0.0000	3rd Qu.:0.00000	3rd Qu.:0.00000	3rd Qu.:0.0000
Max. :1.0000	Max. :1.00000	Max. :1.00000	Max. :1.0000

Romance	War	Musical	Western
Min. :0.000	Min. :0.00000	Min. :0.00000	Min. :0.00000
1st Qu.:0.000	1st Qu.:0.00000	1st Qu.:0.00000	1st Qu.:0.00000
Median :0.000	Median :0.00000	Median :0.00000	Median :0.00000
Mean :0.133	Mean :0.01288	Mean :0.03433	Mean :0.01288
3rd Qu.:0.000	3rd Qu.:0.00000	3rd Qu.:0.00000	3rd Qu.:0.00000
Max. :1.000	Max. :1.00000	Max. :1.00000	Max. :1.00000

Table 5: Summary table for variables used in proportional hazards model

References

Cox, D. and D. Oakes (1984). *Analysis of Survival Data*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.

Fox, J. and S. Weisberg (2023). *An R Companion to Applied Regression* (3rd ed.)., Chapter Cox Proportional-Hazards Regression for Survival Data in R.

Friedman, J., T. Hastie, R. Tibshirani, N. Simon, B. Narasimhan, and J. Qian (2021). *glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models*. R package version 4.1-1.

Greenwald, S. and P. Landry (2022). *The Business of Film: A Practical Introduction*. ISSN. Taylor & Francis.

IMDb (2024). Box office mojo. `https://www.boxofficemojo.com/?ref_=bo_nb_bn_mojologo`.

Lattimore, T. and C. Szepesvári (2020). *Bandit Algorithms*. Cambridge University Press.

Luan, J. and K. Sudhir (2022, 06). Optimal inter-release time between sequentially released products. *Customer Needs and Solutions 9*.

Motion Picture Association (2021, March). 2021 THEME report.

Rotten Tomatoes (2024). Rotten tomatoes. `https://www.rottentomatoes.com/browse/movies_at_home/`.

Seifert, R., C. Otten, M. Clement, S. Albers, and O. Kleinen (2023). Exclusivity strategies for digital products across digital and physical markets. *51*(2), 245–265.

Therneau, T. M. and E. A. Grambsch (2024). *survival: Survival Analysis*. R package version 3.5-0.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological) 58*(1), 267–288.

When to Stream (2024). When to stream. `https://www.whentostream.com/`.

Additional Tables

The complete output for the home entertainment movies that went to streaming while in theaters is as follows:

Movie	Day of Release	Proportion of Revenue
Firebrand	18	0.984138972809668
The Marksman	67	0.918499202079802
Peter Rabbit 2: The Runaway	45	0.991690139182998
Five Nights at Freddy's	0	0
Out of Darkness	18	0.996037500525225
Spider-Man: Across the Spider-Verse	67	0.997031923933
Everything Everywhere All at Once	74	0.796833532843129
Oppenheimer	123	0.986233159424529
Immaculate	25	0.977823673406671
Jesus Revolution	46	0.99366903859082
Gran Turismo	32	0.943752119414015
The Machine	25	0.995328604535466
Belfast	20	0.570956029000516
A Thousand and One	18	0.978044246798548
Operation Fortune: Ruse de Guerre	18	0.98610971925571
Haunted Mansion	68	0.9935858548898
You Hurt My Feelings	27	0.927506459089809
Blithe Spirit	3	0.358419472078096
Strays	18	0.894906745353108
Spider-Man: No Way Home	88	0.985044221475468
Mrs. Harris Goes to Paris	18	0.663613801722556
Firestarter	0	0
DC League of Super-Pets	25	0.726998515232964
Dune: Part Two	46	0.965788140977109
Blue Beetle	39	0.965519913400373
Barbie	53	0.975615586094886
Bros	18	0.942094044933143
Tár	74	0.80658409543131
Creed III	28	0.918873444094208
Orphan: First Kill	0	0
Space Jam: A New Legacy	0	0
The Flash	32	0.988966415927717
The First Omen	53	0.999980241680578
Raya and the Last Dragon	0	0
Renfield	18	0.959777876564083
American Skin	0	0
Titane	18	0.910033208869374
Polite Society	18	0.986138069358178

I.S.S.	18	0.972678963868533
Together	25	0.998679975745137
The Watchers	21	0.981281965710823
Cocaine Bear	25	0.915394765308282
The Eyes of Tammy Faye	0	0
Bottoms	28	0.867190132062199
Perfect Days	27	0.594650661411661
The French Dispatch	52	0.965303089267026
Together Together	3	0.371154510647547
Insidious: The Red Door	25	0.956486682163345
Sisu	18	0.926819892364343
Amsterdam	35	0.993725435207954
Ghostbusters: Frozen Empire	46	0.970928889806433
Copshop	27	0.99838008702069
Cabrini	80	0.999716471789712
Redeeming Love	45	0.999588127657635
A Haunting in Venice	46	0.987968189049142
Godzilla x Kong: The New Empire	46	0.9783412444438
Joy Ride	21	0.989252731611596
Benedetta	18	0.986033855266118
Plane	20	0.823158419492575
Spencer	18	0.87994804866402
Cruella	91	0.998444797090897
Wrath of Man	53	0.993028195194515
Killers of the Flower Moon	46	0.978588911466068
Supernova	18	0.917270087951499
Pig	18	0.844295622981629
Where the Crawdads Sing	60	0.972586454996057
Honk for Jesus. Save Your Soul.	0	0
The Little Mermaid	60	0.99350222141541
The Super Mario Bros. Movie	41	0.93310385553077
The Long Game	18	0.911908994653665
Violent Night	18	0.715769601315278
The Last Voyage of the Demeter	18	0.975955806112407
The Mauritanian	18	0.638908546673425
The Exorcist: Believer	18	0.835797333720695
Alice	18	0.981744731840565
No Way Up	0	0
The Northman	21	0.86050770730442
Retribution	21	0.981476415224882
The Marsh King’s Daughter	18	0.994470426486624
Priscilla	49	0.990953172469993
Fantastic Beasts: The Secrets of Dumbledore	45	0.987949714871577
Trolls Band Together	32	0.863745061902347

Judas and the Black Messiah	0	0
Teenage Mutant Ninja Turtles: Mutant Mayhem	30	0.857397617166721
About My Father	21	0.92862485794363
Last Night in Soho	20	0.954483899236
Thor: Love and Thunder	62	0.993921927205352
Godzilla vs. Kong	0	0
Madame Web	30	0.984344150889381
Queen Bees	0	0
Ghostbusters: Afterlife	52	0.966597852552835
Summer of Soul	7	0.453547261994382
Dog	81	0.999936967923034
Guardians of the Galaxy Vol. 3	63	0.993263985542561
Saltburn	35	0.945694706265952
Nefarious	49	0.982788847202433
Thanksgiving	32	0.98726733908964
The Contractor	0	0
Ordinary Angels	32	0.982941701342638
The Equalizer 3	32	0.933000761222742
Crisis	7	0.617476213005309
Fast X	21	0.910707001461901
Monster Hunter	0	0
Strange World	30	0.926789651853278
The Unholy	53	0.982118536242373
Evil Dead Rise	18	0.815255573545715
Sound of Freedom	101	0.999343400511652
Back to Black	18	0.98263963603323
Werewolves Within	7	0.622122732133077
Infinity Pool	18	0.964898990843772
Half Brothers	19	0.840536044032507
PAW Patrol: The Movie	0	0
The Garfield Movie	46	0.997392035239354
Beast	49	0.996965132464981
Theater Camp	62	0.97935707347607
Taylor Swift: The Eras Tour	61	0.999753215791224

The complete output for the hazard ratio predictions:

Movie	**Hazard Ratio**
Firebrand	0.505403322805746
The Outfit	0.499085868820029
Spider-Man: Across the Spider-Verse	0.407755443047486
Everything Everywhere All at Once	0.461181732969938
Salaar	0.00347397749739338

Sight	0.00590202740551578
Kandahar	0.211211648407856
You Hurt My Feelings	0.711507746558826
Blithe Spirit	0.352489071876668
Lyle, Lyle, Crocodile	0.00493915325984753
Dad, I'm Sorry	0.0042762468307232
2000 Mules	0.00398846628259506
Glass Onion	0.25022344211996
DC League of Super-Pets	0.35808206353794
Barbie	0.380118850926384
Poor Things	0.358360009076056
Creed III	0.371235416200485
Official Competition	0.0046077167210128
Nope	0.514812139071488
Together	0.859807466521948
Wonka	0.331493836448563
Mack and Rita	0.00541844676182738
Deep in the Heart: A Texas Wildlife Story	0.00363566296899955
Raging Fire	0.340967788801654
Dragonkeeper	0.530971278965899
The French Dispatch	0.479337234905126
Cabrini	0.385291840192622
Terrifier 2	0.252162123478675
Asphalt City	0.00693849927927759
A Haunting in Venice	0.42675036594443
Here Today	0.304242621170635
Godzilla x Kong: The New Empire	0.65462558185013
Benedetta	0.192993283402034
Scream	0.00468198233649625
Resident Evil: Welcome to Raccoon City	0.00604104172187465
Breaking	0.0117253358379533
The Little Mermaid	0.333360710790772
Scoob!	0.0062321772480436
Laal Singh Chaddha	0.00328098295433635
The Long Game	0.371451658763864
The Mauritanian	0.307941864382608
The Northman	0.347017648678518
Priscilla	0.590692532146911
Fantastic Beasts: The Secrets of Dumbledore	0.271978379812924
Treasure	0.0170461864555811
Big George Foreman	0.00547699537836713
Anita	0.0030121206286023
Spy x Family Code: White	0.0023182887247062
Last Night in Soho	0.447171828060377

Thor: Love and Thunder	0.275145121260261
Ghostbusters: Afterlife	0.288241852889196
Summer of Soul	0.166646928152831
Saltburn	0.590297515500778
The Beast	0.36659099860828
One Piece Film: Red	0.00144241892478943
Crisis	0.443979438687585
Separation	0.144766376368029
The Roundup	0.387792470247053
Strange World	0.377939417910362
American Fiction	0.431284803114077
Christmas with the Chosen: The Messengers	0.00324228509365794
Evil Dead Rise	0.773367684873589
Werewolves Within	0.51779727019298
The Black Phone	0.0080758871193253
Infinity Pool	0.596111028997
Half Brothers	0.347736815763145
Beast	0.998983285794096
Theater Camp	0.404947144519363