

The film industry is one of the most lucrative in terms of profit to be made. The industry has the potential for enormous financial gains, while also posing great risks if a film underperforms. Therefore, it has become incredibly important to determine the predictors of success for well-performing films, as there is a lot of money to be gained and lost. As the industry has grown and matured since its inception, so have the techniques for analyzing the success of a given film. We are interested in looking at the 100 top grossing domestic films of all time and analyzing a few key variables that may have impacted their success.

Due to the lucrative nature of the film industry, there has been a plethora of research conducted on what makes films successful. Doing a thorough review of the literature is beyond the scope of this study, but it is still important to touch on a few other research projects to show some motivation behind the choices we made. One such study is “A Predictor for Movie Success” by Ericson and Grodman from Stanford University. We structured our data collection similarly to theirs’, where we took data from a wide variety of aggregate movie data sites (Ericson and Grodman). Similarly, we structured our questions to follow the style of Siminoff and Sparrow’s “Predicting Movie Grosses: Winners and Losers, Blockbusters and Sleepers”. Siminoff and Sparrow focused their research on overall trends with successful films, with a focus on first weekend release revenue (Siminoff and Sparrow).

We also wished to differentiate our study by creating new questions, with one question in particular having a specific focus on how movies perform on an annual basis. For instance, the year 2008 was a hard year for the United States as this was the year the housing bubble popped. It was very hard for people to make ends meet and people did not have as much disposable income to do things that they wanted to, such as travel with their family and have fun experiences that would cost large amounts of money. We wanted to see if this economic downturn had any noticeable effect on the average of the top grossing films around that time. We leaned partially on the work of Cieply and Barnes and their article “In Downturn, Americans Flock to the Movies” to help answer this question.

By looking at a collection of some of the top earning movies in the United States, we are attempting to find which aspects serve as good predictors of how well the movie does when compared to other top grossing films. The main questions that we want to answer are: In which release quarter do movies best perform? Which distributors are most successful? Are opening weekend sales and opening weekend profits good predictors for the overall domestic gross? And what is the general trend for domestic gross in terms of year, specifically in regards to events such as economic downturns?

To answer these questions, we chose the following variables to examine:

- Domestic Gross
- Release Year
- Distributor

- Release Quarter
- Opening Weekend Sales
- Production Budget
- Profit from Opening Weekend

It's important to note that the “domestic gross” variable has not been adjusted for inflation, which means the data set of the 100 top grossing domestic films will be naturally skewed towards more modern film releases. The “release quarter” variable is also of note due to its lack of presence in other similar datasets. We have defined “release quarter” as the fiscal period of a particular film’s release date. Films releasing between January and March are 1st quarter films, those releasing between April and June are 2nd quarter films, and so on.

The data itself was collected from a number of aggregate film data websites, including: Box Office Mojo, The Numbers, and The Movie Database. The data was scraped through Python scripts and imported into R Studio for analysis. Variables that appeared from multiple sources were also compared against each other to check for discrepancies to ensure that the data was reliable.

We did a statistical analysis to address every question we initially posed. The results of our data collection and visualization are as follows:

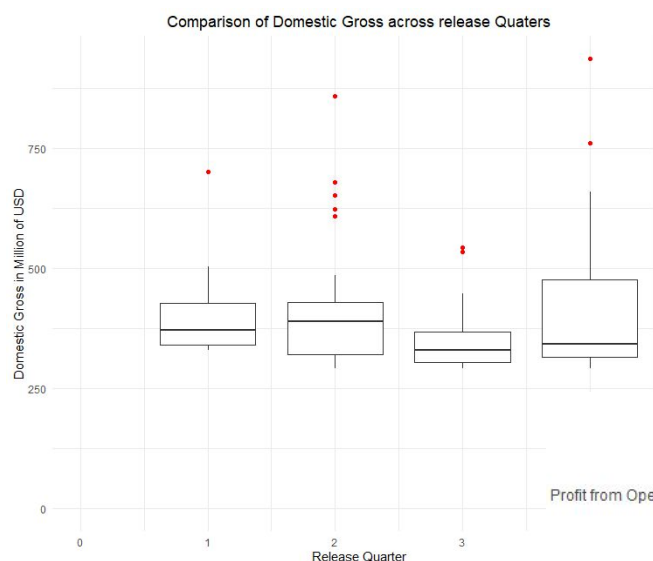


Figure 1: The boxplot shows the relationship between domestic gross and release quarter in addition to showing outliers within each quarter. Smaller boxes and whiskers mean more similarly earning movies within the quarter, while larger ones indicate more variation.

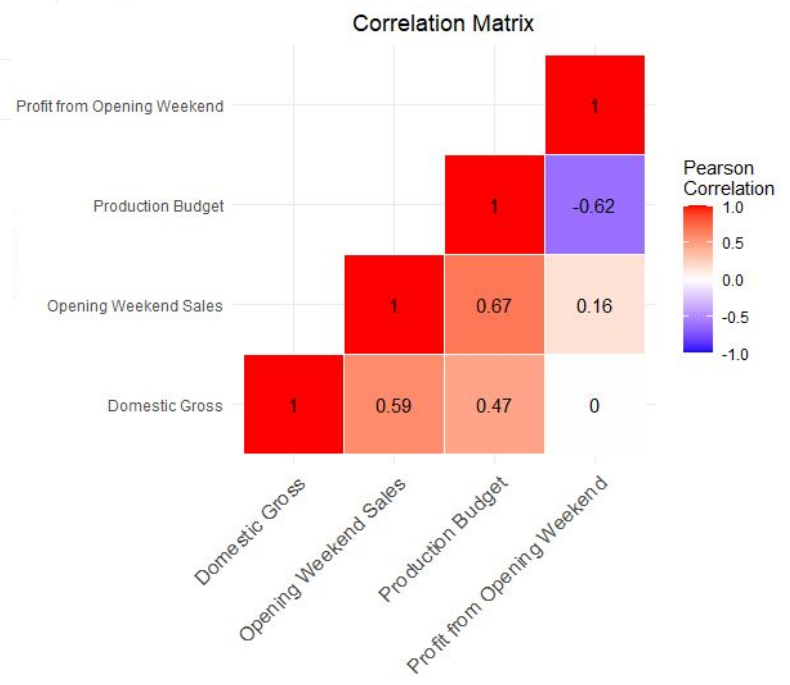


Figure 2: The correlation matrix shows the relationship amongst the continuous variables in the film dataset. Positive values indicate a positive correlation while negative values indicate a negative correlation.

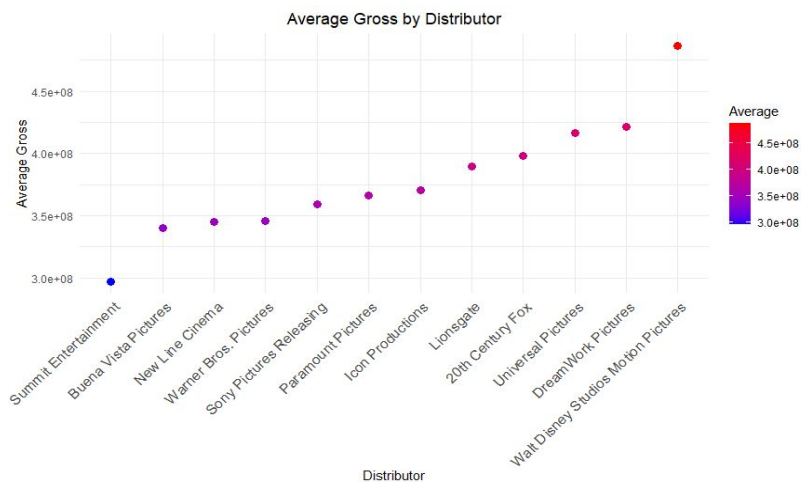
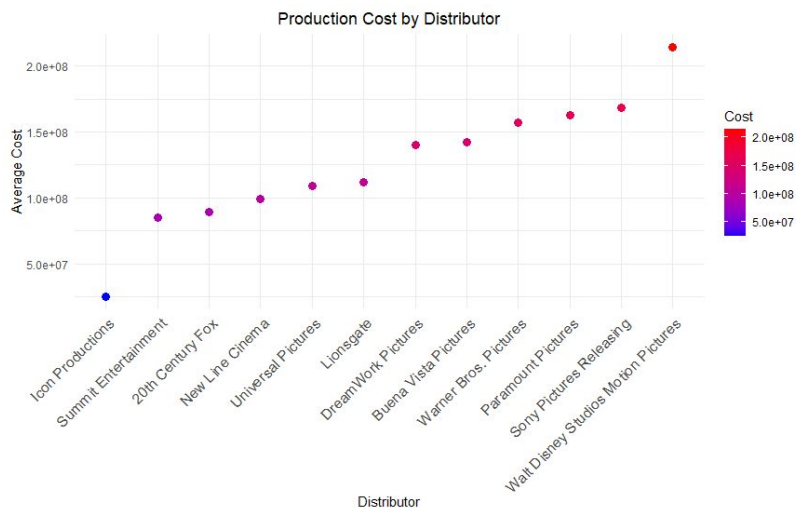


Figure 3: The average gross by distributor is shown in an increasing order.

Figure 4: The average production cost by distributor is shown in an increasing order.



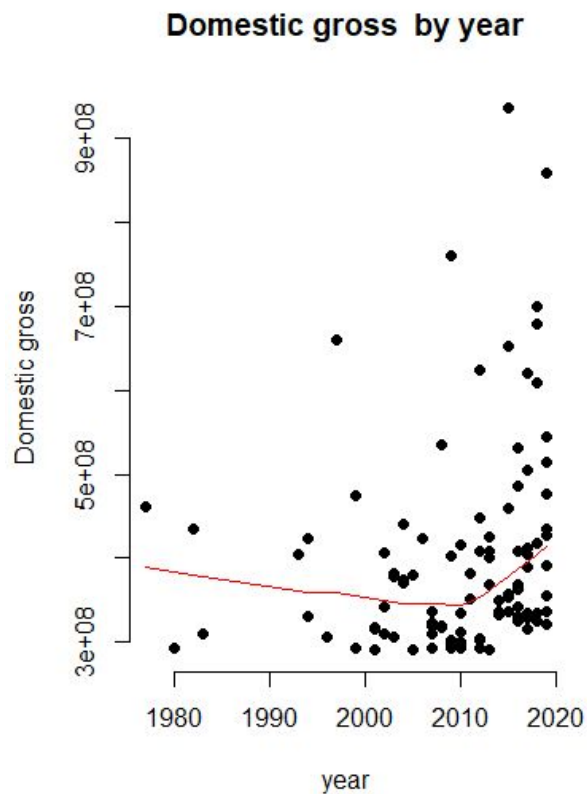
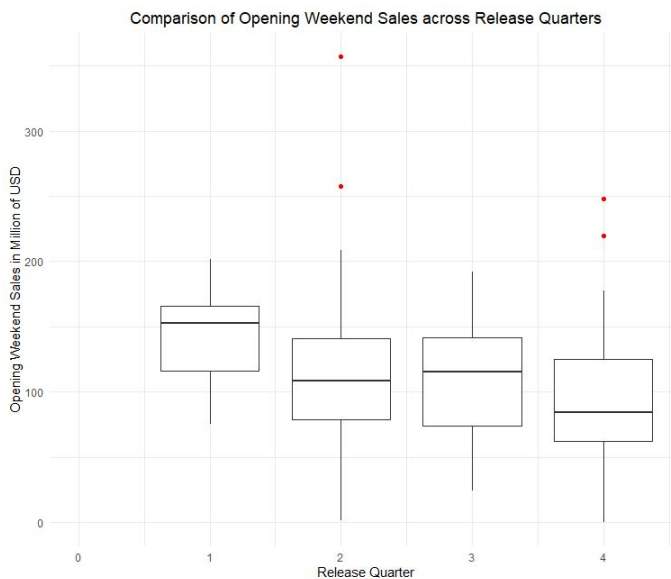


Figure 5: The scatter plot shows the relationship of the domestic gross over the years while the line helps to visualize the trend of the graph.

Figure 6: The box plot shows the relationship between opening weekend sales and release quarter.



In order to find which release quarter movies generally perform best in, we looked at the data in Figures 1 and 6 to determine the spread of domestic gross for each quarter as well as the opening weekend sales for each film. The boxplots show the median values for each respective quarter, as well as the difference between the extremes for each quarter. A tighter spread indicates a more consistent quarter while a greater spread means it can be more variable. To find which distributors are most successful we looked at the data represented in Figure 3 and Figure 4, which show average gross revenue and average production cost respectively. To find out if opening weekend sales and profits are good indicators of domestic gross we look at Figure 2. In the correlation matrix we can look at the values that show the relationship between opening weekend sales, profit, and domestic gross and get a good idea of how closely correlated the variables are. Finally, we look at the trend for average domestic gross per year to find the effects of economic downturn on the movie industry in Figure 5.

Results and Conclusion

In which release quarter do movies generally perform best? On average, quarter two has the highest domestic gross. Quarter two also has significant outliers, including the second highest grossing film of all time. It's important to note that quarter two also has the most movies released during the quarter, likely because it falls towards the beginning of the Summer blockbuster season. Production companies may also have some further insider knowledge about how successful the quarter can be. A side note is that quarter three appears to be the least successful in terms of gross revenue.

Which distributors are most successful? From Figures 3 and 4, we can see that Walt Disney Studios is on average the biggest earner and spender by a large margin. After comparing the graphs, we also notice that average production cost does not perfectly line up with average earnings suggesting that individual distributor budgets have more of an impact than the base amount of money spent.

Are opening weekend sales and opening weekend profits good predictors for the overall domestic gross? Looking at Figure 2, we can see that opening weekend sales and overall domestic gross have a correlation coefficient of 0.59, indicating that opening weekend sales are a fairly strong indicator of domestic box office performance. At the same time, opening weekend profits and overall domestic gross have a correlation coefficient of 0. This indicates that there is no linear relationship between opening weekend profits and overall domestic gross. It may, however, be a more complex curvilinear relationship that we didn't model. Further research may be required.

What is the general trend for domestic gross in terms of year, specifically in regards to events such as economic downturns? Figure 5 shows the trend line for domestic gross slowly lower from just before 1980 to approximately 2008 where it dips heavily to 2010 before skyrocketing from the 10s onwards. We were curious to see if the 2008 economic crash had any major impact on movie sales. Looking at the graph a

person may blame the dip on the economic crash of 2008. However, with the amount of data that was used to make this graph it can be a little misleading. According to Michael Cieply and Brooks Barnes, there often aren't drop offs in movie sales during economic downturns, but, in fact, just the opposite. In one example, they stated, "In 1982, theater attendance jumped 10.1 percent to about 1.18 billion (the top seller was 'E.T.: The Extra-Terrestrial') as unemployment rose sharply past 10 percent. Then admissions fell nearly 12 percent, an unusually sharp drop, in 1985". This, and the many other examples given in their article, imply that movie sales may in fact jump in times of economic hardship, likely due to the idea that movies can be a form of escape from difficult times. We ourselves don't mean to imply that the economic crash in 2008 resulted in the great boost to box office performance, but it may have had a non-trivial effect.

In conclusion, after analyzing the data we collected, we found some notable trends in the data that may indicate what may help bolster a film in the box office. We found that films distributed by Walt Disney Studios in quarter two perform the best overall, while the staying power of a movie in theaters can be determined by a number of factors including opening weekend sales, while factors like opening weekend profits are not relevant to the overall performance of a film in the domestic box office. Furthermore, it seems like recent years have been better than ever for releasing films with the hopes of wild success, as the vast majority of the best performing films were released post-2010. Finally, we determined that economic downturns are often correlated with stronger box office success, so a film releasing during one may do better than one released during a strong fiscal year. Further research may benefit from going more in-depth on any number of these questions with a much richer and refined dataset of films.

Works Cited

- Bell, Travis. "The Movie DB." *The Movie Database (TMDb)*, Fan TV, 2008, www.themoviedb.org/?language=en-US.
- Cieply, Michael, and Brooks Barnes. "In Downturn, Americans Flock to the Movies." *The New York Times*, The New York Times, 28 Feb. 2009, www.nytimes.com/2009/03/01/movies/01films.html.
- Ericson, Jeffrey, and Jesse Grodman. "A Predictor for Movie Success." *Stanford University*, cs229.stanford.edu/proj2013/EricsonGrodman-APredictorForMovieSuccess.pdf.
- IMDb. *Box Office Mojo*, Amazon, 1998, www.boxofficemojo.com/.
- Simonoff, Jeffrey S., and Ilana R. Sparrow. "Predicting Movie Grosses: Winners and Losers, Blockbusters and Sleepers." *Chance*, vol. 13, no. 3, 2000, pp. 15–24., doi:10.1080/09332480.2000.10542216.
- Nash, Bruce. "Where Data and the Movie Business Meet." *The Numbers*, Nash Information Services, LLC, 1997, www.the-numbers.com/.

<https://www.themoviedb.org/?language=en-US>