What Causes a Box Office Film to Succeed?

Introduction

As time has gone on, movie business has not lost any stride as movies still make billions every year and consumers love the films that are put out. Although, movies are getting very expensive to make and considering social media platforms like YouTube and Twitter, advertisement is beginning to cost more as well. What most don’t realize is when most companies make a movie, they either stay safe and invest less, or largely invest and hope for a blockbuster. What most do not realize is that the film industry can be considered a large gamble.

Production budget and availability is one of the most important factors for a film, but higher budget does not always equal higher ticket sales. According to data scientist Randy Olson, there is a weak correlation between production budget and ticket sales, but there is still a correlation1. Analyzing the sales data of over 11,000 films, Olsen makes the claim that “the more that’s spent on film production, the less likely the film will end up making that investment back.”1 When companies invest more into a movie, the outcome can be gamble as movies can either be a complete hit or an utter flop.

When researching what causes a film to draw in consumers, what does the consumer research to determine the quality of the film? Critical reviews are readily available before a movie is released to inform consumers of the movie’s quality, but does this have an effect on ticket sales? Alec Kennedy’s article, “Predicting Box Office Success: Do Critical Reviews Really Matter?” acknowledges this query using data from over 200 films released in 2007 comparing box office gross and Metacritic score2. Using his research, Kennedy then concluded that when a film has a great critical score, it shows in ticket sales, although it may not be because of the review2. Kennedy makes the conclusion that critical reviews are not nearly as important as marketing and advertisements, as a consumer will have heard of a movie from marketing and then find a review for the film to decide whether to attend or not2. Great reviews for the newest blockbuster movie seem to assist its box office gross some, but it is not as significant as an advertisement to decide for the consumer whether they want to purchase a ticket or not.

What determines a film’s box office success? The goal of this experiment is to answer that question by using real data from films made in the past few years and running it through a neural network to find what factors determine success or failure. Although to simplify the process of data-collecting and input, the output will determine the success of the opening domestic box office weekend.

The neural network is made in python, using multiple libraries. The data was collected from multiple box office websites including the-numbers.com and boxofficereport.com which contains precise numbers for variables like production budget, theater count, and trailer view counts from YouTube. Collecting data for reviews became very simple using Rotten Tomatoes percentages, which is an average percentage of reviews created from numerous websites and critics which can be collected from rottentomatoes.com. After data was collected, the neural network’s accuracy was determined using certain methods and compared to other accuracies.

Procedures

INITIALIZATION:

The neural network was built and initialized in Python. Python 3.7.3 was the latest version used for the experiment. The libraries installed and imported into python include numpy, sci-kit, pandas, seaborn, and matplotlib. The experiment will function properly on any IDE.

DATA COLLECTION:

The input variables for this experiment included the following: Production budget, Opening theater count, YouTube trailer view count, and Rotten Tomatoes score. The source of these data came from multiple sources.

* Production budget and Opening theater count – the-numbers.com4
* YouTube trailer view count – boxofficereport.com5
* Rotten Tomatoes score – rottentomatoes.com6

The output of the neural network is whether the movie was a success or failure, how this was determined and where it was collected is researching if the domestic opening weekend box office exceeded or fell below predictions. These predictions came from boxofficereport.com while some came from various news sites who report whether movies met expectations or not. The difficulty in this method is deciding if a movie’s opening weekend is really a success or failure just based on predictions.

Data & Results

After 90 data points were collected, the data was normalized to simplify the numbers for the neural network using an equation for the data points to be between 0 and 1. All of the data was normalized. The data was then transferred to a CSV file and imported into the neural network using pandas.

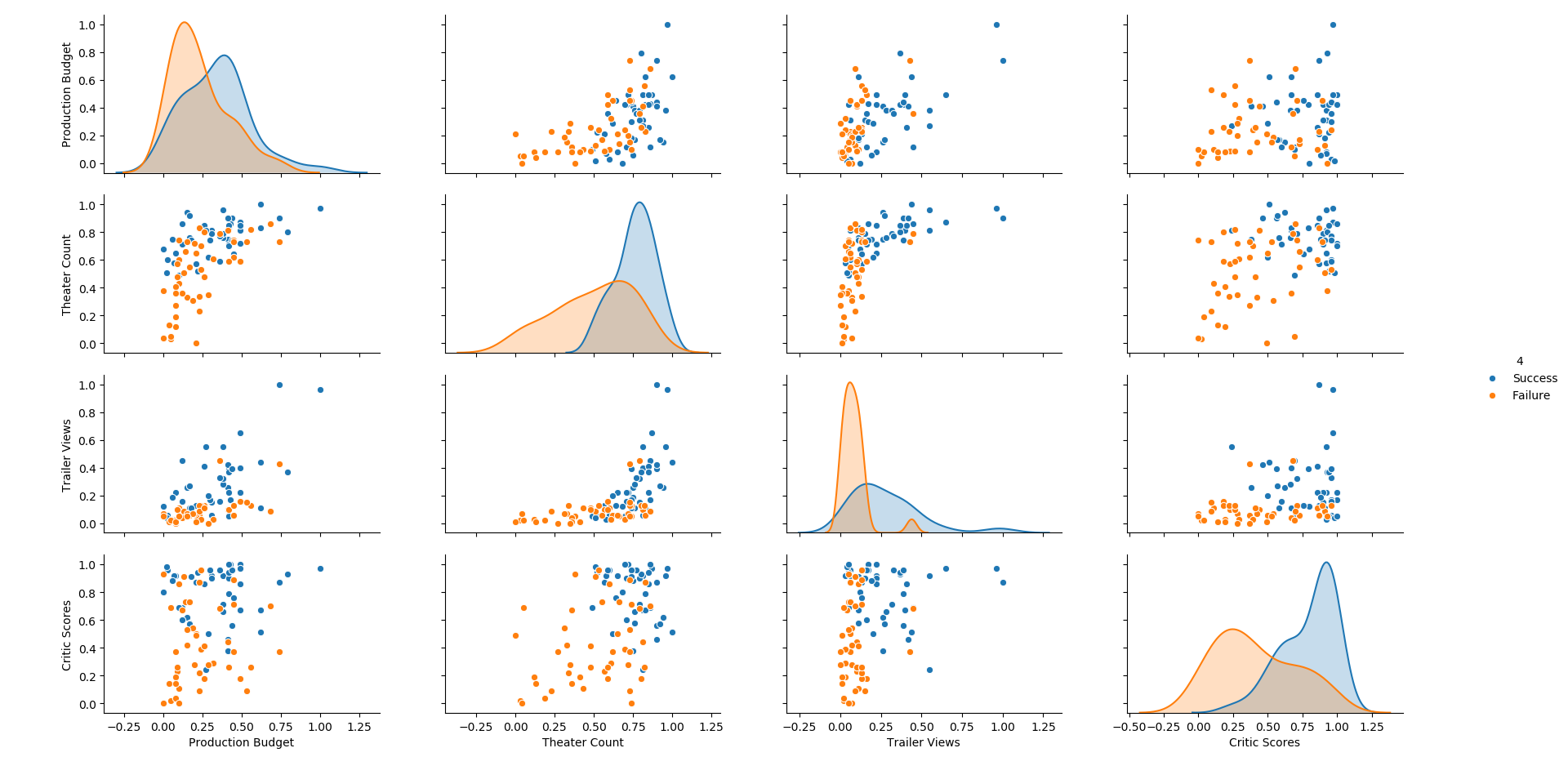
Figure 1The results of the input data represented in scatter plots for two input variables compared to one another, along with histograms of the input variables themselves.

Figure 1 is a collection of scatter plots and histograms of input variable compared to each other to find possible trends in the data for the program to train and test with. The 4th column is the possible outputs of the opening weekend, which is either a success or failure. The correlation in the graphs in Figure 1 is hard to determine but there are a few that can be observed. There are multiple observable things that can cause a failure for a movie’s opening weekend as you can see in the histograms but there’s not a clear correlation for certain success. Availability seems to have an effect, the graph in figure 1 which compares Theater count and critic scores shows that most successes have higher theater count and higher critic scores, while most of the failures have lower, with a few outliers.

Table 1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Movie Title | Production Budget (USD) | Opening Theater Count | YouTube Trailer count (m) | Rotten Tomatoes Critic scores (%) | Box Office Success or Failure |
| *Avengers: Endgame (2019)* | $400,000,000 | 4,662 | 233.4 | 94 | Success |
| *Joker (2019)* | $55,000,000 | 4,374 | 112.9 | 69 | Success |
| *Dark Phoenix (2019)* | $200,000,000 | 3,721 | 41.6 | 23 | Failure |
| *Maleficent: Mistress of Evil (2019)* | $185,000,000 | 3,790 | 18.8 | 40 | Failure |
| *Zombieland: Double Tap (2019)* | $48,000,000 | 3,468 | 15.5 | 69 | Success |

Table 1 shows a few sample data that show the varying factors for the failure or success of a domestic opening weekend, this experiment is purely based on data available before a film has made its debut, so the actual gross made on the opening weekend is not tested, as this is a experiment is to test which factors of a movie affects its opening weekend the most.

Figure 2

Figure 2 shows the various tests for the grid search with different amounts of nodes for one hidden layer.

From the grid search which was used on the neural network, the number of recommended hidden layers and nodes were shown through running multiple tests. The number of hidden layers recommended was 1 with 3 nodes within that layer. Although when the specific node count was tested in Figure 2, the results are very odd and relatively the same throughout, which can mean that node count isn’t as significant as one would think for this neural network.

The best train and test score were then gained from cross validating the data. The train score resulted in a 84% accuracy from training with 60% of the data, while the test score resulted in a 90% accuracy using the remaining 40% of the data.

Conclusion

The neural network was built to predict domestic box office weekend results. The input variables are from the data which is available before a movie’s opening weekend has begun which include: production budget, opening theaters, amount of YouTube trailer views, and Rotten Tomatoes critic scores. The neural network was then made using Python and various libraries imported into the program. Using gathered data, the neural network was trained and tested. Then the best accuracy was determined using grid search.

Some of the few difficulties during the process were first, the decision on what describes success and failure for a film, where to collect the data online, and what correlations are made based on the graphical data shown from the neural network. All of these were solved however they were an obstacle in the process. The difference between success and failure could be improved and more precise for another experiment which could help the experiment for relating it to real data. There was no strong correlation between any of the factors and domestic box office success, although there were a few weak ones that could be explored and experimented more to find if they have a real effect on a movie’s success, as the opening weekend to a film is its most important3.

1. By. Does a bigger film production budget result in more ticket sales? Dr. Randal S. Olson. 2014 Dec 29 [accessed 2019 Dec 18]. <http://www.randalolson.com/2014/12/29/does-a-bigger-film-production-budget-result-in-more-ticket-sales/>
2. Predicting Box Office Success: Do Critical Reviews Really ... [accessed 2019 Dec 18]. <https://www.stat.berkeley.edu/~aldous/157/Old_Projects/kennedy.pdf>
3. Kumar R, Sinha, Sitabhra. The statistical laws of popularity: universal properties of the box-office dynamics of motion pictures. NASA/ADS. [accessed 2019 Dec 18]. <http://adsabs.harvard.edu/abs/2010NJPh...12k5004P>
4. Where Data and the Movie Business Meet. The Numbers. [accessed 2019 Dec 18]. <https://www.the-numbers.com/>
5. Box Office Report. Box Office Report. [accessed 2019 Dec 18]. <http://boxofficereport.com/>
6. Movies: TV Shows: Movie Trailers: Reviews. Rotten Tomatoes. [accessed 2019 Dec 18]. <https://www.rottentomatoes.com/>