Introduction

Most common people do not understand how box office or ticket sales work, one may assume that the more money you pour into the movie becomes profit, which is incorrect. When most companies make a movie, they either stay safe and invest less, or largely invest and hope for a blockbuster. What most people do not realize is that the film industry can be considered a 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 affect 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 near 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.

[MAY ADD PAR. ABT TRAILERS AND ADS]

What determines a film’s box office success? Through the use of a neural network, the answer may be found by comparing the correlations between inputs and what truly affects the outcome. 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.

1. <http://www.randalolson.com/2014/12/29/does-a-bigger-film-production-budget-result-in-more-ticket-sales/>
2. <https://www.stat.berkeley.edu/~aldous/157/Old_Projects/kennedy.pdf>
3. <https://www.researchgate.net/publication/322083844_Mining_Trailers_Data_from_YouTube_for_Predicting_Gross_Income_of_Movies>

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.com
* YouTube trailer view count – boxofficereport.com
* Rotten Tomatoes score – rottentomatoes.com

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.

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.

Data & Results

Idk what to put here please help me Dr. Waddell

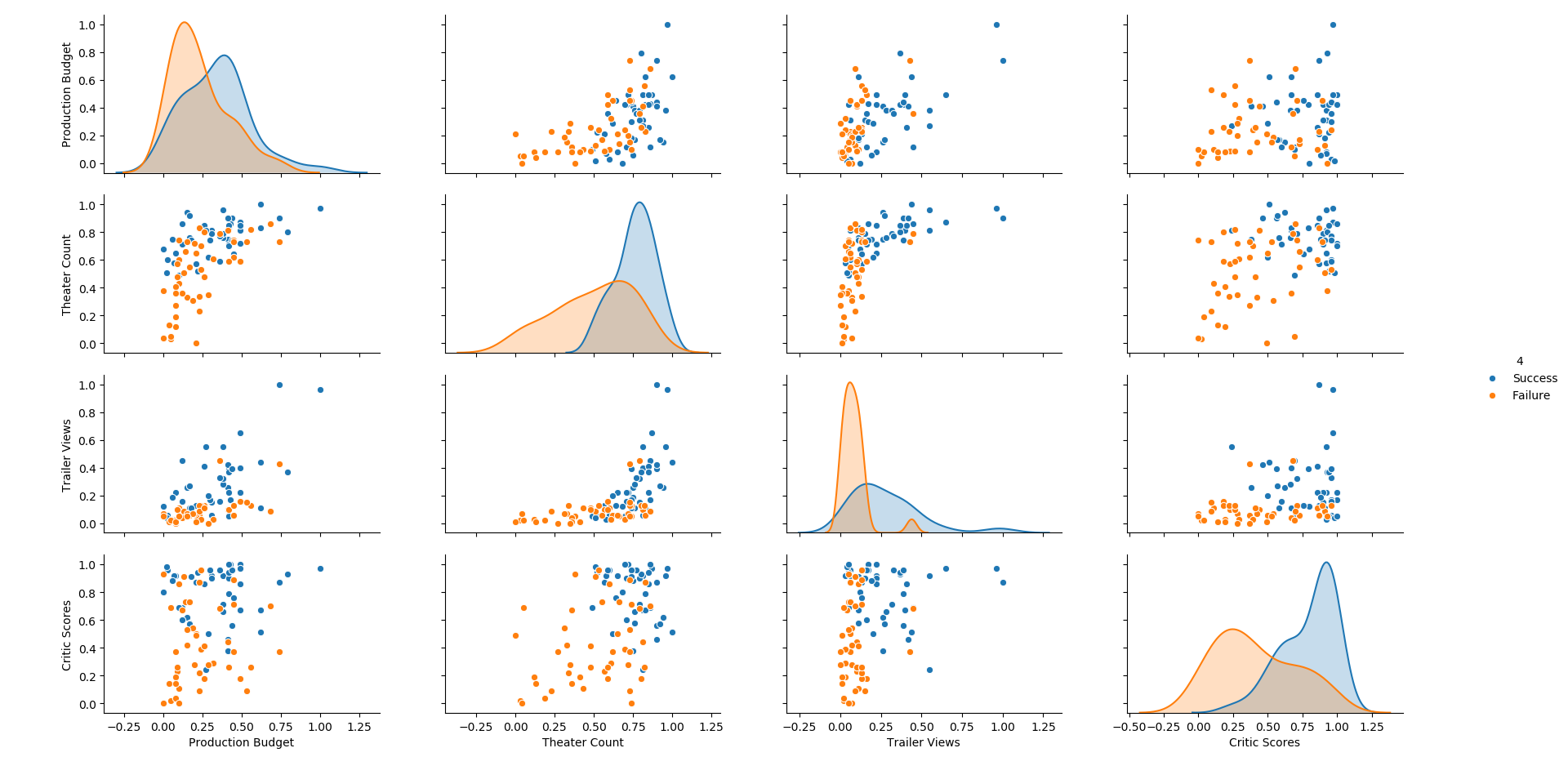


Figure 1

The 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. [STUFF HERE]

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)* | $200,0000,000 | 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

[caption]

\*Explain ranges I used for grid searches

From the grid search which was used on the program, 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.

\*Wording

\*Put some of this in procedures

\*Change this

\*\*Table maybe

Using a 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.

\*Lead to conclusion, don’t go all the way though.

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.

[CORRELATION BETWEEN DATA]. The accuracy of the neural network could be increased by adding data from films before 2016. The correlation between [SOMETHING] and a movie’s domestic opening weekend box office outcome proves that [SOMETHING] is important for a movie in becoming successful for it’s opening weekend, which is the peak time period for a movie’s income.

[ERROR STUFF]