How to earn the highest profit from a single movie

Charles Bui, Christopher Turner, Jack Jeffries, Sarah Brittle

## **Background**

The purpose of this research project is to find the secret to making the most profit from a single movie as possible. This profit is only calculated by movie sales, and does not include big money makers such as merchandise. An exploration of the movie data between the years 2005-2015 is expected to return the answers to our three questions:

1. Does budget affect overall profit?
2. Do ratings need to be high to be profitable?
3. Do any genres stand out as big money makers?

By answering these questions, we hope to propose a fool proof money making movie.

## **Data** **Collection**

The original intention of this research project was to utilize the API of OMDb. However, OMDb only allows for searches with a movie title or an IMDb ID. Since this project was predicated on searching for movies by years in order to find the titles of the movies, a new dataset had to be introduced before OMDb could be utilized. A complete movie dataset was found on Kaggle, which included 45,000 movies and the years we were needing for our own research.

The dataset was cleaned to include our stipulations in our research, including deleting duplicates, cutting down the years to the time frame needed, and dropping any movies which did not have budget or revenue information. These changes whittled our dataset down to 2,316 movies across the years 2005-2015. The dataset was then further reduced to cut out any movie which had a budget or revenue of less than $5,000. This was done to cut out any amateur or very low budget movies from our research.

With this new dataset, OMDb could be utilized to search for the titles through an API call. OMDb API provided data on ratings, number of votes, genre, and box office sales. The Kaggle dataset and the OMDb dataset were merged to create the final dataset for our research.

## **Budget** **and** **Revenue**

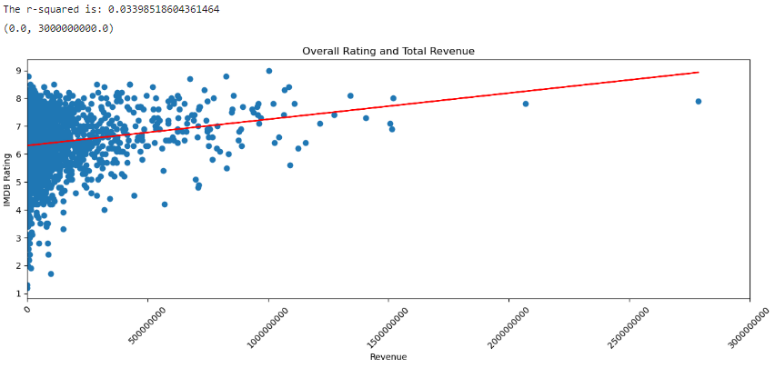
Budget and revenue of the movies were plotted in a scatter graph (fig. 1). The graph shows a relative correspondence between amount spent and amount earned. Worth noting however, is that of the top ten movies by budget (fig. 2), only *Avengers: Age of Ultron* appears in the top ten movies by revenue (fig. 3). So, while there appears to a general r value between money spent and money earned, it is not necessarily a rule written in stone.

This becomes even more apparent when revenue is subtracted from budget to get the return on investment (ROI). The top ten movies by ROI all generally show share a commonality in having a small budget and a large revenue. In other words, even though the top ten movies by revenue earned millions of dollars, these movies also spent millions to be made. So it seems that one of the key factors to high earnings is to have a relatively moderate budget and a popular movie, as can be seen with the example of *Paranormal Activity* (fig. 4). The r-value between budget and revenue however is .6. This is a low to moderate correlation between the two variables. However, as discussed above, a high budget does not always equate to ROI, as movies that make more also are the movies that spend more.

Another factor to consider is the total budget, revenue, and ROI over the ten-year period (fig. 5). The data shows that movies are becoming cheaper to make, which in turn means profits grow higher. This could be due to a new technology available to film makers within this ten-year period which reduces the overall cost of the production.

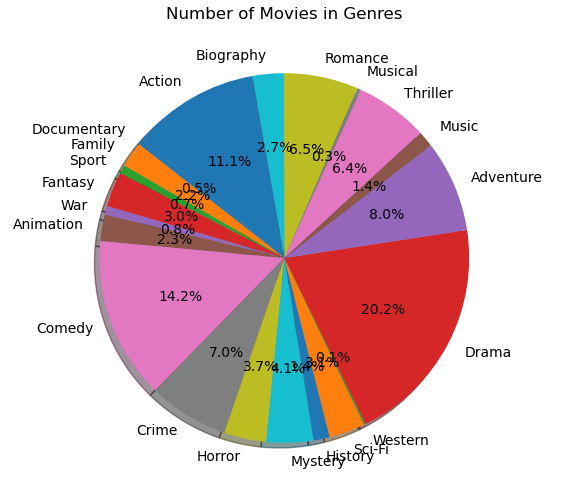
## **Ratings**

While budget, revenue, and ROI show interesting positive correlations, ratings, perhaps surprisingly, show almost no correlation to a movie’s success. In fact, over the ten-year period, the median rating for all movies peaked at 6.6 in 2007 and 2015, and slumped to 6.4 in 2008, 2010, 2011, and 2014 (table 1). This means the decade average variable is just .2. In fact, to be in the middle 80% of all movies, the rating must only be higher than 5.2 and lower than 7.6.

Not only did ratings not change with each year, the r value shows that there almost no correlation between rating and revenue (fig. 6). Additionally, the ratings remain relatively similar across different genres as well (fig. 7). Interestingly, the number of votes per genre is highly varied. The drama genre received the most votes, followed by action and adventure. Most other categories are far behind these genres in terms of votes. This could suggest that more people see these movies, or that more people are critical of these genres. Despite these higher numbers of votes though, the average ratings for these genres still remain comfortably within the range of all the genres, meaning while its possible people have higher expectations for these genres, its also possible that people are more engaged and more excited to leave a review for the movies.

## **Genre**

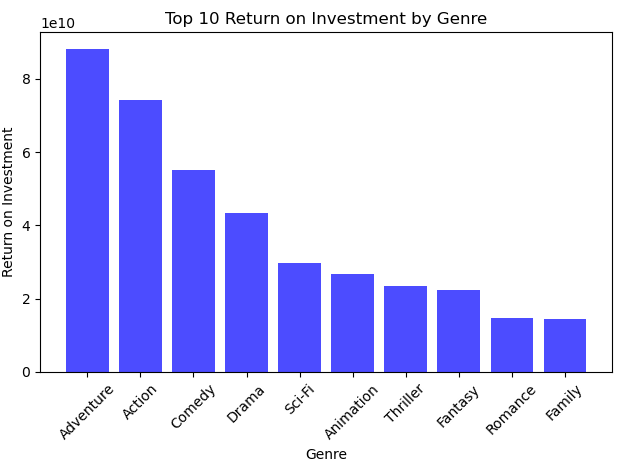
 Of the 2,316 movies, there were 21 genres recorded. Out of all of these genres, drama was the highest reported, accounting for 20% of all movies in our dataset (fig. 8). Comedy comes in second at 14.2%. Important to note, the movies in this dataset have multiple genres assigned to them. Each movie in our dataset may be counted twice or three times for each genre assigned to it. This means that our genre data is only a rough look at how each genre performs against ratings and revenue.

 Of the genres, the highest return in revenue came from adventure movies, followed by action (fig. 9). This high ROI matches with the findings of high review interaction discussed above. Adventure also only takes up 8.09% of the total movies made in this time period. The market is saturated with dramas, however action movies are few enough to have too much competition but also frequent enough to not be considered too obscure, such as documentary or western movies.

## **Conclusion**

This research set out to find the highest profit in a single movie from movie sales alone. Through our research, we conclude that being highly rated is not an important factor in making profit. As long as our movie is within the 6-7 range, we will be clustered with majority of all movies.

Our genre should be adventure, with highest return on investment. To maximize our profits, we could make an action-adventure movie to utilize the second highest earning genre. While *Paranormal Activity* has the highest ROI of our dataset, this is an outlier. In fact, in general, horror should be avoided, unless we believe we have the next groundbreaking, genre defining story in our hands.

 While there is some correlation between budget and revenue, our top ten in each of these categories suggests that a higher budget does not guarantee a higher ROI. Therefore, we suggest a medium budget of \_\_\_\_\_\_\_\_\_\_\_\_\_ in order to maximize profits.

The next step in research would be to narrow down into more distinct categories, such as the actor, director, setting, etc. While our research provides a good first stepping stone, there are still many more factors that may be considered. It is also important to note the limitations of our data, such as how many movies we needed to cut out of our dataset in order to assure good data, and how common IMDb was for leaving a review in 2005 vs. 2015. Despite this, we believe our data justifies our conclusion to create an average, middle budget action-adventure movie to maximize our total profit.