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.

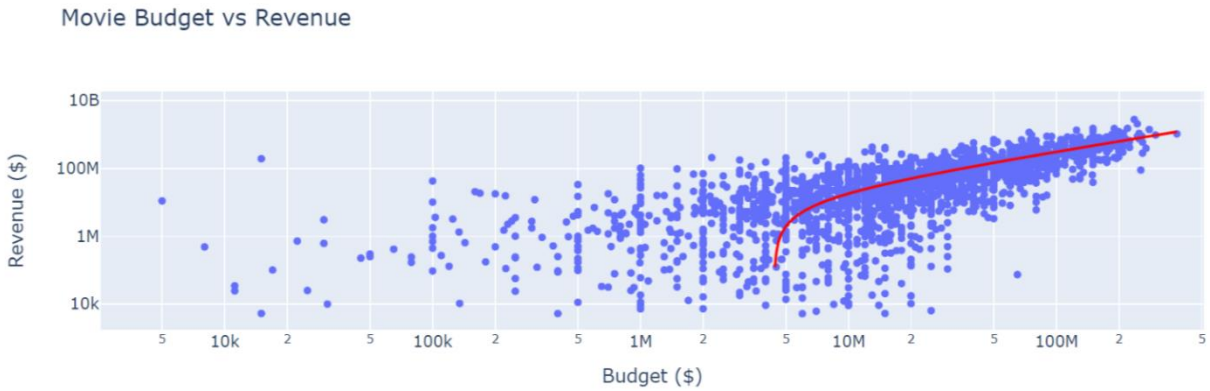

Figure 1 Scatter plot showing the relationship between money spent and money earned

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, only *Avengers: Age of Ultron* appears in the top ten movies by revenue (fig. 2). So, while there appears to a general r value between money spent and money earned, it is not necessarily a rule written in stone.

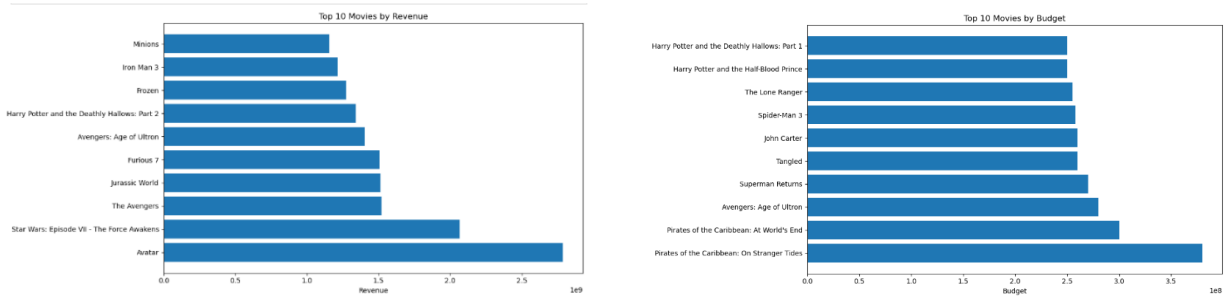

Figure 2 Top ten movies by budget(in millions) and by revenue (in billions). Only one movie exists in both of these graphs.

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 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. 3). 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.

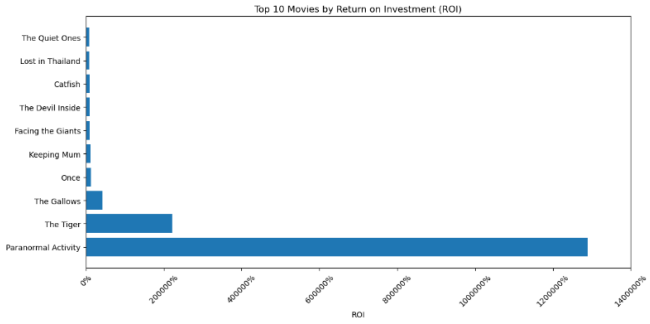

Figure 3 The highest return on investment (in billions). "Paranormal Activity" stands out as a clear outlier.

Another factor to consider is the total budget, revenue, and ROI over the ten-year period (fig. 4). 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.

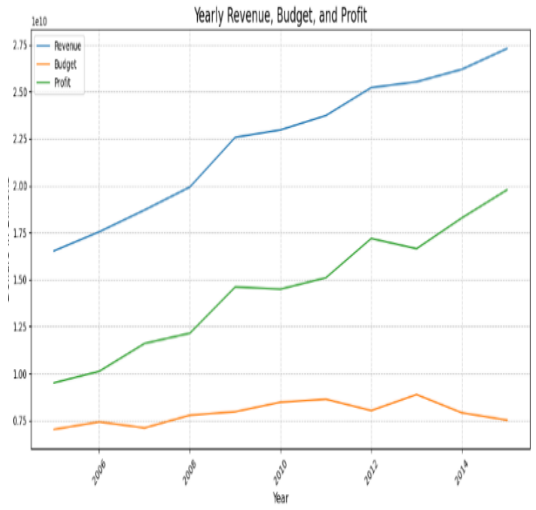

Figure 4 Line graph showing revenue, profit, and budget (in billions). 2013 shows a combined decrease in budget and increase in profit.

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.

	IMDB Rating				
	mean	median	var	std	sem
Year					
2005	6.354450	6.50	1.025020	1.012432	0.073257
2006	6.386321	6.50	1.058817	1.028988	0.070671
2007	6.506771	6.60	1.033933	1.016825	0.073383
2008	6.301932	6.40	1.204948	1.097701	0.076296
2009	6.417972	6.50	0.891296	0.944085	0.064089
2010	6.368996	6.40	0.887061	0.941839	0.062238
2011	6.327039	6.40	0.999654	0.999827	0.065501
2012	6.402913	6.45	1.011894	1.005929	0.070086
2013	6.512500	6.50	0.876345	0.936133	0.062548
2014	6.375962	6.40	1.111690	1.054367	0.073107
2015	6.491878	6.60	0.842179	0.917703	0.065384

Table 1 Across all years, average ratings remain stable.

Not only did ratings not change with each year, the r value shows that there almost no correlation between rating and revenue (fig. 5). Additionally, the ratings remain relatively similar across different genres as well (table 2). 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.

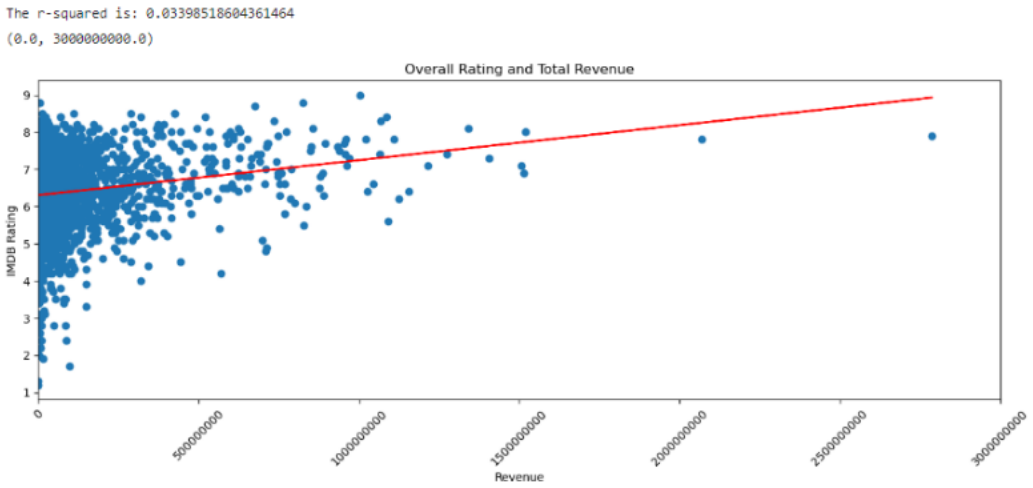

Figure 5 There is almost no correlation between rating and revenue (billions).

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. 6). 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. 7). 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.

	count	Genre Rating	Genre Votes
Romance	395.0	6.400506	37600416.0
Sci-Fi	191.0	6.354974	61118846.0
Action	675.0	6.359407	142350453.0
Thriller	387.0	6.323773	62828996.0
Western	9.0	6.844444	2252130.0
Biography	165.0	7.022424	24281916.0
Drama	1225.0	6.688735	155484897.0
Sport	45.0	6.548889	4658356.0
Documentary	33.0	6.960606	897367.0
Musical	20.0	6.745000	1723656.0
Fantasy	185.0	6.082162	32832156.0
Music	85.0	6.548235	6819152.0
Comedy	859.0	6.210477	98746160.0
Animation	141.0	6.612766	24866380.0
Crime	427.0	6.524356	60101355.0
History	86.0	6.839535	7486626.0
Adventure	486.0	6.365638	112160872.0
Mystery	249.0	6.260643	37134978.0
War	46.0	6.863043	6089730.0
Family	136.0	6.015441	15538751.0
Horror	224.0	5.670982	23211040.0

Table 2 No one genre rates noticeably higher or lower than any other genre.

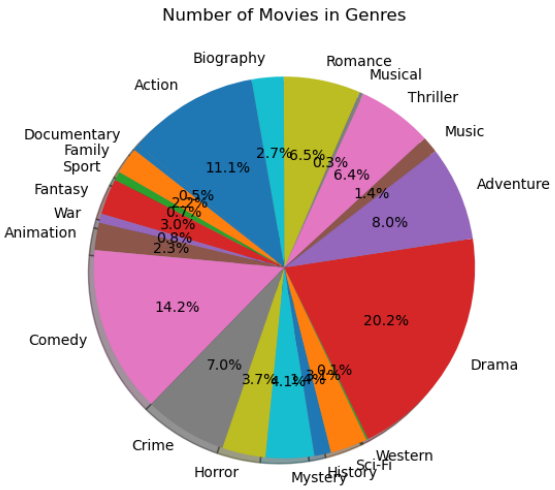

Figure 6 Drama dominates, with 1/5 of the entire market.

Our genre should be adventure, which has the 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 $37,441,333 in order to maximize profits.

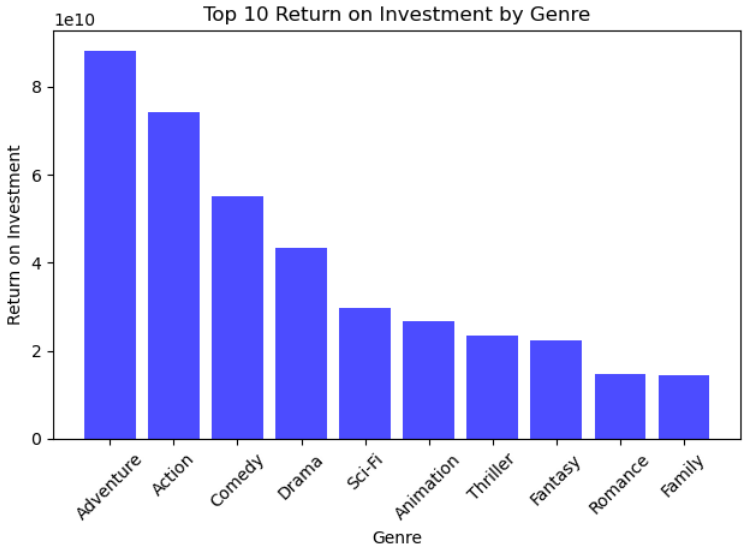

Figure 7 Adventure and action movies deliver the highest return on investment (in billions).

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.