**Date:** April 23, 2019

**To:** Shawn Janzen, Founder, Janzen Consulting Group

**From:** Andrew Carroll, Janzen Consulting Group

**RE:** The Relationship Between a Movies Budget and IMDb Rating

For a movie studio to be successful, it is important that they pick and choose movies that they feel will be popular with audiences and do well at the box office. Often, when studios feel that a project will mesh well with audiences, they tend to give that movie a bigger budget. However, is a bigger budget necessarily indicate that a movie will be better received by audience? It is vital that movie studios understand if a high budget indicates a good movie because it is a direct determinant of the profit a studio makes.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table 1: Summary Statistics of Budget (in Millions of USD) and IMDb Score** | | | | | | |
|  | Freq | Min | Max | Med. | Mean | SD |
| Budget | 3509 | $0.0002 | $300.00 | $20.00 | $35.76 | $42.69 |
| IMDb Score | 3509 | 1.60 | 9.30 | 6.50 | 6.36 | 1.11 |

For our analysis, the target population is all movies that have been made. I hypothesized that a bigger budget does not mean a movie will have a better IMDB rating. During my testing, I found that this may indeed be the case.

In order to find an answer for my hypothesis, I used a sample the IMDB movie data set. In this sample, there were 3,509 independent observations of movies that were used in my model. For this test my independent variable was the Budget variable. This variable describes the amount of money that the studio spent on the film in USD. This variable had to be recoded because some of the values were not in USD. Since converting all the different currencies would take far too long to do, in this analysis I only included movies that were made in the USA. My dependent variable was the IMDb Score variable. This variable rates movies on a score from one to ten. While IMDb has not officially explained how their score is calculated, we do know that it is the weighted average review score of all user submitted reviews.

Before beginning my test, I checked the assumptions and for an OLS Regression. Originally, there were many outliers in my data. Upon further investigation, some of the movie’s budgets appeared to be incorrectly entered into the dataset and had budgets of over $1 billion. Therefore, I capped the budget at $400 million since that is how much *Pirates of the Caribbean: On Stranger Tides,* the most expensive movie ever made, cost to make. After this recode, there were no more outliers in my data. Finally, the only assumption and diagnostic my model did not pass was having normally distributed errors. However, my QQ plot was close enough to being normally distributed that I felt comfortable enough to not do anymore recoding to my variables.

A close up of a map

Description automatically generatedIn order to test my hypothesis, I conducted a Simple OLS Regression test on the Budget and IMDb score variables at an alpha level of .05. My model shows there was a nonsignificant correlation between the two variables, r(3507) = .08, p < .001. Despite a nonsignificant correlation, we reject the hypothesis that claims there is no correlation between Budget and IMDb score. Budget was also a statistically significant predictor in the model (t = 4.63, p < .001) I am confident that my results can be generalized to my target population since my dataset is so large and accounts for over 100 years of movies.

The outputted regression model, = 6.283 + 0.00022(Budget) tells us that for each $10,000 in the budget, the IMDb score should increase by 0.00002 points (p < .001). The y-intercept coefficient is not meaningful because the budget of a movie will never be $0.00, and neither is slope coefficient because such a small number will have little effect on the score. Additionally, budget insignificantly explained the proportion of variance in IMDb score, R2 =.006, F(1, 3507) = 21.45 , p < .001.

Despite my results, my study still had its’ limitations. One such is that I was limited to only using movies made in the United States. While most movies are made in the United States, foreign films were unrepresented in this model. Had foreign films been included, we may have gotten significantly different results than those that were outputted.

A close up of a map

Description automatically generatedA close up of a map

Description automatically generatedA close up of a map

Description automatically generatedA close up of a map

Description automatically generated