Analysing the Impact of Movie Run time on Ratings and Box Office Collections

Anudeep Nayak, Bhuvvaan Punukolu, Kanishka Ghodke, Nikhil Rowland, Siddharth Kalyanasundaram, Trishala Thakur

University of Colorado Boulder

Abstract

The duration of a movie, commonly referred to as its run time, holds substantial significance in shaping a film's success, where success is defined as achieving higher ratings and greater revenue. Our objective in this study is to examine the potential influence of a movie's run time on both its ratings and box office earnings. The hypothesis will be evaluated using a z-test methodology. For the analysis, data was acquired by web-scraping a random selection of 1000 movies from the IMDb database.

1. Introduction

This report aims to test whether a movie's duration has an impact on its commercial success through hypothesis testing. Our research relies on IMDb, an extensive online database encompassing diverse details about released movies, as the primary source. Specifically, our sample dataset consists of a random selection of 1000 movies collected as of December 13th, 2023. To conduct our analysis, a z-test methodology is being utilized. Initially, a sample of 1000 movies is being examined to draw inferences applicable to the entire IMDb collection. In this context, the entire IMDb collection of movies represents the total population for our study.

2. Data

The data were randomly scraped from the IMDb website, comprising a total of 1000 movies from various countries, languages and genres. It consists of eight columns, with four columns being categorical and four columns being quantitative.

Movie Name	Year	Runtime(m)	Genre	Rating	Votes	Revenue($)	Country
chr	*int*	*int*	*chr*	*float*	*int*	*int*	*chr*
Avatar	2009	162	Action	7.9	1,367,686	760,507,625	United States
The Avengers	2012	143	Action	8	1,442,411	623,279,547	United States
Minions	2015	91	Animation	6.4	254,006	336,045,770	United States
Inception	2010	148	Action	8.8	2,495,317	292,576,195	United States
Toy Story 3	2010	103	Animation	8.3	877,676	415,004,880	United States

Table 1: Top grossing movies from the dataset

The data consists of columns such as movie name, release year, genre, and country of origin, which are categorical, as well as user rating, user votes, gross revenue in dollars, and run time in minutes, which are quantitative. Among these columns, the focus is on the run time, ratings,

and revenue for our analysis. IMDb users assign scores to movies on a scale of 1 to 10, and the aggregated score is presented as the movie's rating. The reported revenue figures represent the domestic box office collections of these movies in the United States.

2.1. Biases in the data

The data inherently has the bias of being sourced solely from the IMDb database. While the IMDb database is not exhaustive, it is believed to be comprehensive enough to generalize our findings.

While movies from all countries released over a 70-year period are included in the base data, the analysis was limited to movies exclusively from the US and released in the 21st century. This decision aimed to mitigate two potential biases in the data:

1. Culturally, people might show varying preferences for longer movies in different countries.
2. Including movies as old as the 1950s could skew perceptions of their box office collections due to differences in ticket prices and inflation.

3. Methods

3.1. Data Cleaning and Preparation

Prior to conducting a thorough analysis of the data, it was necessary to perform data cleaning. Within the data set, certain movies exhibited missing values in the revenue column. As the number of such movies were less such movies were simply removed from the sample. Additionally, the revenue data contained commas, resulting in non-integer values within the column. To address this, the commas were removed, transforming the data within the column from string to integer. Furthermore, some movies had non-integer data within the release year column. This non-integer data was removed, leaving only the release year represented as integers.

The data set was divided into two groups: movies with a run time exceeding two hours and those with a run time below two hours. This benchmark was selected as movies surpassing the two-hour mark are typically classified as long, and the 120-minute threshold is sufficiently close to the 107-minute mean run time of the sample, ensuring the significance of our findings.

3.2. Analysis Conducted

Two hypothesis tests were performed on the data set. Hypothesis tests were chosen because of their effectiveness at finding statistically significant differences between two samples. The sample with shorter movies had 311 movies and the sample with longer movies had 65 movies. For this hypothesis test, the mean has been chosen as the test statistic. The initial hypothesis test assessed the mean ratings of both samples, while the second test examined the mean revenue of each sample.

3.3. Z-Test: Assumptions and Use Cases

In this study, a z-test has been employed to test and validate the hypotheses. The z-test can only be used when certain conditions are met, such as the existence of an independent and identically distributed (IID) sample of random variables. In our case, we randomly sampled from an IMDb list of independent movies, satisfying this criterion.

Additionally, the z-test is applicable when the underlying distribution is approximately normal or when the sample size is large enough to assume the applicability of the Central Limit

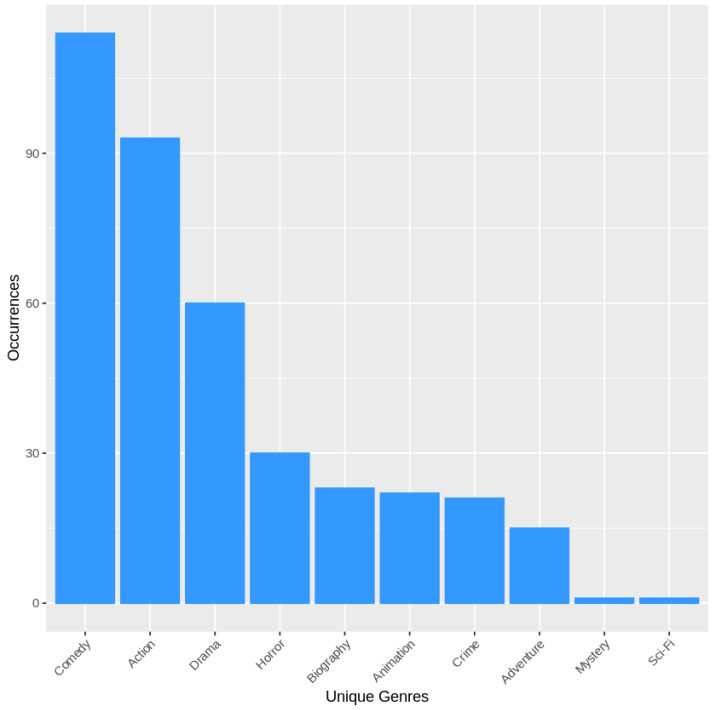

Figure 1: Genre distribution within the sampled movies

Theorem. In our case, as illustrated in Figure 3, the underlying distribution closely resembles normality, and the sample sizes of 311 and 65 are large enough to apply CLT and justify the use of a two-sample z-test.

3.4. Hypothesis Tests

The mean ratings and revenue of movies in both the buckets are given in Table 2. The means of both the features are clearly smaller for shorter movies. So, a single tailed test was used instead of a two tailed test to determine if people have a statistically significant predilection in that direction.

Sample	Mean Rating	Mean Revenue ($)
Movies under 2 hrs	6.2	52,676,476
Movies over 2 hrs	7.0	101,171,824

Table 2: Means of samples

In the terminology of hypothesis testing both the tests can be stated as follows. Here μ_a is the mean of the sample with shorter movies and μ_b is the mean of the sample with longer movies.

Test 1:

Sample	Standard Deviation of Rating	Standard Deviation of Revenue ($)
Movies under 2 hrs	0.91	60,635,535
Movies over 2 hrs	0.83	74,771,318

Table 3: Standard deviation of samples

- H_0: Movies that have a run time of over two hours have the same rating when compared to movies that are under two hours i.e., $\mu_a - \mu_b = 0$.
- H_a: Movies that have a run time of over two hours have a higher rating i.e., $\mu_a < \mu_b$.

Test 2:

- H_0: Movies that have a run time of over two hours earn the same revenue when compared to movies that are under two hours i.e., $\mu_a - \mu_b = 0$.
- H_a: Movies that have a run time of over two hours earn more revenue i.e., $\mu_a < \mu_b$.

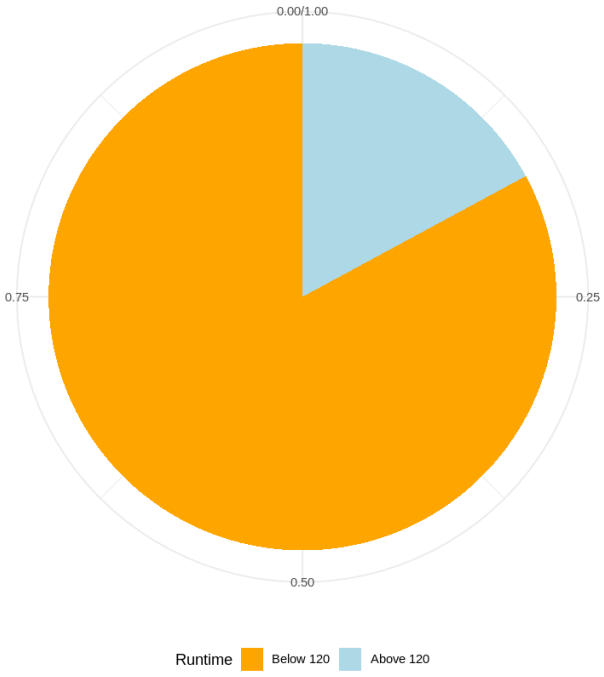

Figure 2: Proportion of movies that are over/under 2 hr run time

4. Results

The results of the z-test implied that, with a high level of statistical significance, movies with longer run time than two hours are popular among movie goers and do well at the box office than movies that are under two hours. The p-values of the z-test are given in Table 4.

Test	p-value
Hypothesis Test 1	3.085e-12
Hypothesis Test 2	0.001596

Table 4: p-values

A threshold of 5% is set to the significance level (α) to determine the statistical significance of the test. In both tests conducted, the obtained p-values are markedly lower than this predefined significance level. Given that the p-values are well below 5%, we find substantial evidence to reject the null hypothesis in favor of the alternate hypothesis in both the tests.

Figure 3 shows that the distribution of the underlying sample is close to normal, which ensures accurate results for the z-test. All three figures, namely Figure 3, Figure 4, and Figure 5, support the obtained result by showing higher ratings and greater revenue for longer movies.

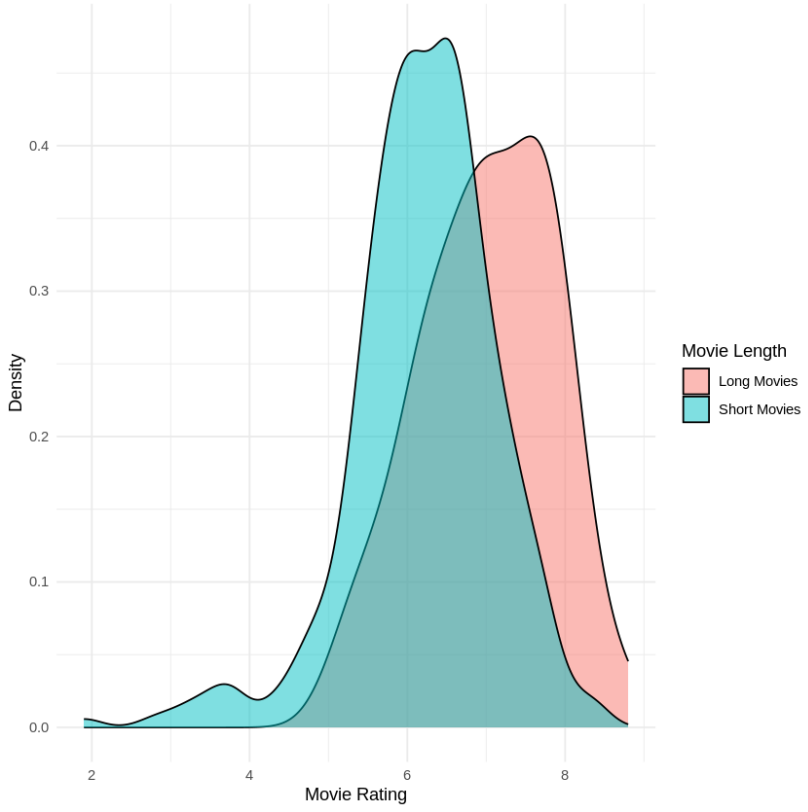

Figure 3: Density plot of movies and their ratings

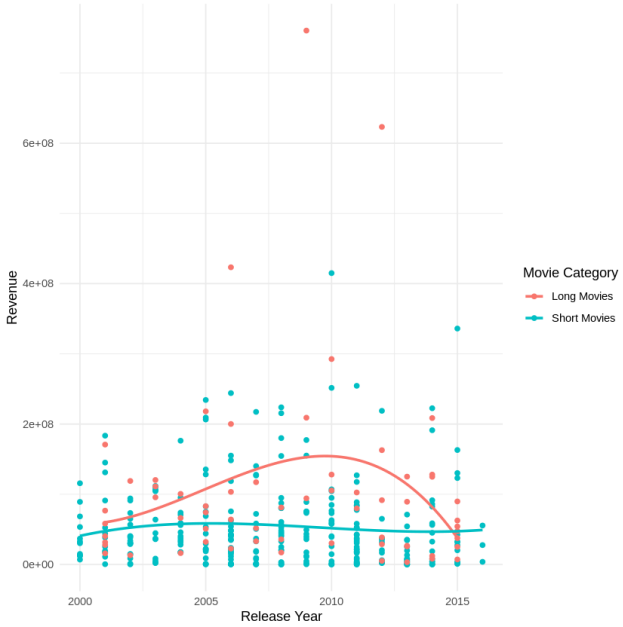

Figure 4: Revenue earned by movies over time

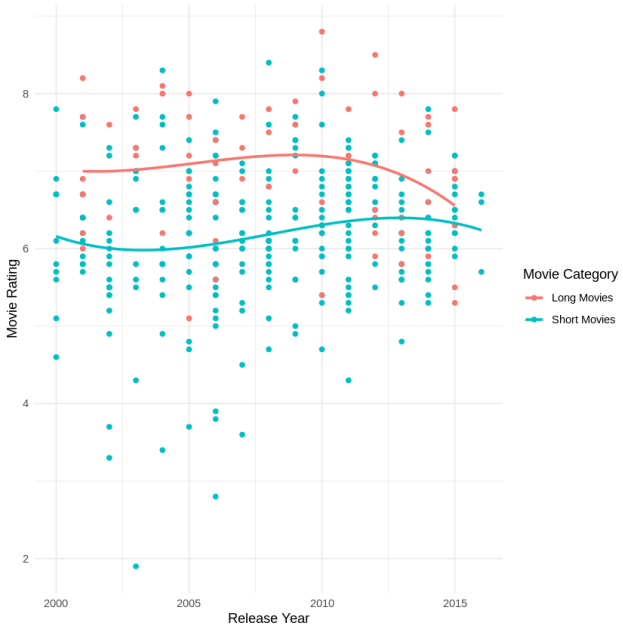

Figure 5: Ratings of movies over time

5. Conclusions

5.1. Discussion on Findings

This paper examines the effect of a movie's run time on its ratings and revenue. To accomplish this, hypothesis testing was used to compare the means of two groups of movie samples, one group with movies under two hours, and one group with movies over two hours. As can be seen in the results, movies above two hours are more likely to both have higher ratings and higher revenues.

Consumers might prefer longer movies or shorter movies for a variety of reasons. With their longer run time, longer movies can tell more complex stories with characters the audience spends more time with, allowing them to be fully immersed into the film. Another reason could be that longer films might have larger budgets, which allows the director more resources to create a fun experience. The value proposition of a longer movie might also appear more appealing to consumers, since if both a short movie and a long movie are equally priced, the long movie would give more screen time per dollar.

5.2. Further Research

Future research on this topic should investigate the correlation between run time and budget, as well as the effects of budget on ratings and revenue. Another topic that could be of interest could be performing the same analysis as the one performed in this paper, but using a sample of much older movies to test whether consumer preferences have changed overtime.

References

,. IMDb. https://www.imdb.com/list/ls063676189/?sort=list_order,ascst_dt=mode=detailpage=1. Accessed: December 17, 2023.