Factors Affecting Movie Ratings

University of Maryland

INST627: Data Analytics for Information Professionals

FALL SEMESTER (2022)

Professor Christopher Antoun

December 12, 2022

Sarah Engleman: senglema@umd.edu
Surbhi Garg: sgarg31@umd.edu
Tanya Gupta: tgupta21@umd.edu

INTRODUCTION

Viewers rate movies based on how they perceive its content. This may seem like a simple concept and one that we are all familiar with, however, we were intrigued to find out if this simplicity is merely on the surface or if movie characteristics influence the thumbs up or thumbs down and the number from 1-10 that we see when clicking into a movie title. Scott Wallsten of Technology Poly Institute analyzed IMDb ratings for a set of about 4,000 movies and concluded that "an additional point on IMDB (on a scale of 1 to 10) is associated with an additional $38 million in worldwide revenues." What affects this impactful number?

Zeroing in on the relationships between genre of movie and rating of movie, our research question is as follows: Do the genres of comedy and horror affect movie ratings? These two genres were chosen for comparison, as we found that these genres did not overlap for movies in our dataset. Analyzing genres with such polarizing characteristics would lead to more definitive comparison results than other genres. If there are factors affecting movie ratings, audiences and production companies alike should be aware of these instances.

METHODS

The dataset used in this research is sourced from Internet Movie Database (IMDb), last updated in 2022, and found on Kaggle.com. There are a total of 340 observations observed, with 259 comedy observations and 81 horror. The dataset lists movies available on Netflix and their genres. Comedy genre and horror genre were the dependent variables for our research question, and release year and duration are dependent variables in an additional analysis. All movies were described with more than one genre, such as "romance, comedy". We found zero titles with comedy and horror genres overlapping, and therefore our analysis compares movies containing one of the two genres. The independent variable is movie ratings, which are derived from IMDb. Viewers create IMDb accounts and may give one rating to each movie. Ratings are on a scale of 1-10, with 1 being the lowest possible rating and 10 the highest.

Four key methods were used in this analysis. Descriptive statistics were found first to better understand the dataset. A two-sample t-test was then conducted to understand how significant any differences in variable means are. Our analysis went two steps further, as we tested additional dependent variables against movie ratings with correlation tests and a multiple linear regression model. Our analysis was conducted in the data analytics platform, R Studio.

ANALYSIS

Descriptive statistics for each variable are shown in Table 1. Figures 1B-5B in Appendix B show histogram distributions and boxplots for each variable. Most notably, the mean rating for comedy movies is 6.19 and the mean horror rating is 5.25.

Table 1: Descriptive Statistics

	All Genres (rating)	**Comedy (Rating)**	**Horror (Rating)**	**Release Year**	**Duration (Mins)**
Mean	5.97	6.19	5.25	2012	99
Max	8.2	8.2	7.6	2022	174
Min	2	2.8	2	2005	50
Standard Deviation	1.39	1.01	1.24	10	18.82

H*o*: The mean movie ratings are the same for both comedy and horror movies.

H*a*: The mean movie ratings are different for comedy and horror movies.

The two-sample t-test provided a t-value of 6.2 and a p-value of less than our significance level of 0.05. Because the *p*-value is less than 0.05, the results are statistically significant. The effect size is 0.503, which indicates that there is a medium size effect between the difference of the mean between the two groups. Hence, we reject the null hypothesis and conclude that ratings do get influenced by comedy and horror genres.

Correlation tests were conducted to find the relationship between movie ratings (all genres) and additional dependent variables. Figure 1 shows a scatterplot of release year and ratings. The correlation value between these two variables is -0.15, indicating a moderate negative relationship. Figure 2 shows a scatterplot of duration and ratings. The correlation value between these two variables is 0.30, which indicates a moderate positive relationship between movie duration and ratings. We conclude that as movie release dates increase, ratings decrease, and as movie durations increase, ratings also increase.

Figure 1: Year vs. Rating Scatter Plot

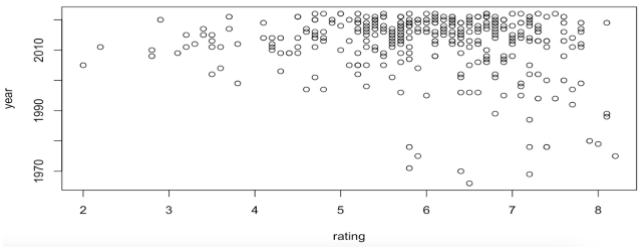
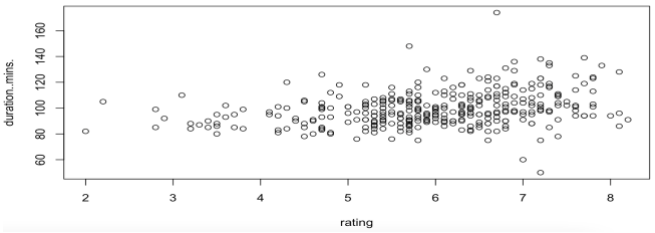


Figure 2: Duration Vs Rating Scatter Plot



Wrapping up our analysis, we conducted a multiple linear regression using the top-down approach to learn more about the influence our dependent variables have on the independent variable of movie ratings. From the parameters in Table 2, we found that the regression coefficient of all three variables is significant, as *p*-values are less than 0.05. Hence, rating is influenced by genre, year, and duration. The regression equation is as follows:

Rating = (-0.85)*genre + -0.016*year+0.023*duration(mins). Note: R^2: 0.225 and adjusted R^2: 0.215.

Table 2: Summary Table of Multiple Linear Regression

	Estimate	**Std. Error**	***t*-Value**	***p*-Value**
Intercept	37.043	10.91	3.395	0.00076
Genre (Horror)	-0.85	0.129	-6.589	<0.05
Release Year	-0.016	0.005	-3.039	0.002
Duration (Mins)	0.023	0.004	6.007	<0.05

RESULTS

Do the genres of comedy and horror affect movie ratings? Our analysis predicts that ratings of comedy and horror movies are not the same, with comedy movies having higher ratings than horror. Upon conducting a regression analysis, we found that our initial analysis of comedy movies having higher ratings than horror does hold strong (as seen from the equation). The regression analysis also showed that movie duration and release year also influence the ratings of a movie.

CONCLUSION

As a group including a quantitative researcher, we are interested to hear first-hand from viewers after watching comedy and horror films, why they give these ratings. Could the serotonin from a comedy movie increase the chances that viewers rate comedy movies higher in general? Are viewers more critical of horror movies than comedy movies?

Certain limitations exist in this research, starting with the dataset. Millions of movies are available worldwide, and therefore our dataset is considered small. This dataset only includes movies available on Netflix. Expanding to analyze movies available on other platforms may provide a more holistic representation of the movie population. We also acknowledge the vast number of variables that may impact movie ratings that were not included as part of this analysis. Additional variables include more genres and production budgets.

This data is also imperative to the everyday movie connoisseur. As a person is scrolling through Netflix and deciding between watching a comedy or horror movie, comparing ratings may not be the best indication of what movie to ultimately watch, as the comedy movies are rated higher than horror. Next steps in this research include investigating how viewers find movie ratings. Do a significant number of people rely on IMDb ratings in comparison to other rating tools and sites such as Rotten Tomatoes and

the movie platform itself (i.e., Netflix)? In addition, we are interested to see if our research findings would be similar for television shows, and if viewers compare ratings between movies and television shows when deciding what to watch. What factors affect television ratings?

Awareness surrounding factors that affect movie ratings would be of utmost importance to production companies, as Technology Poly Institute's research shows how impactful just one additional rating point on IMDb is associated with an $38 million in global revenue. Knowing this information before investing in a movie could decrease the risk of the company losing money and serve as a powerful tool to production companies.

APPENDIX A

IMDb.com. (n.d.). *Ratings, reviews, and where to watch the best movies & TV shows*. IMDb. Retrieved December 2022, from https://www.imdb.com/

Netflix Popular Movies Dataset. Kaggle. (2022, September 24). Retrieved December 2022, from https://www.kaggle.com/datasets/narayan63/netflix-popular-movies-dataset

Wallsten, S. (2021, June 28). *Rotten tomatoes and IMDB reviews strongly correlated with movie revenues - publications*. The Technology Policy Institute. Retrieved from https://techpolicyinstitute.org/publications/miscellaneous/rotten-tomatoes-and-imdb-reviews-strongly-correlated-with-movie-revenues/

APPENDIX B

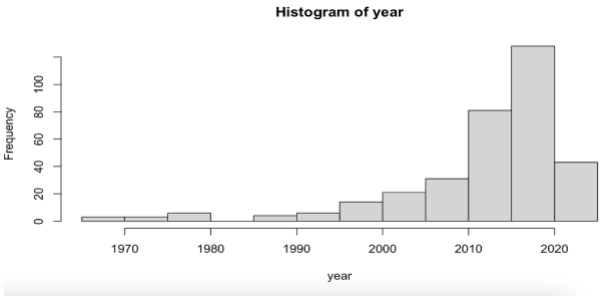


Figure 1B: Histogram of Year

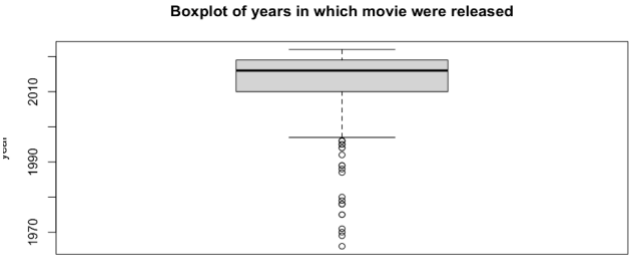


Figure 2B: Boxplot of Year

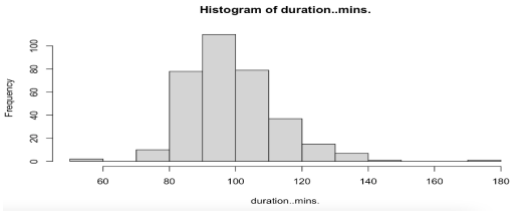


Figure 3B: Duration (Mins) Histogram

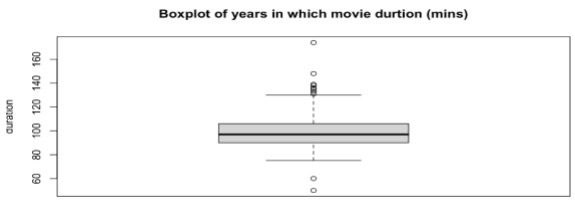


Figure 4B: Boxplot of Duration (mins)

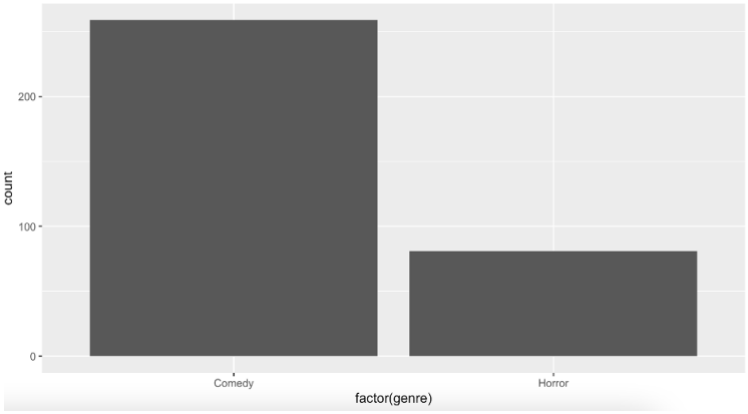


Figure 5B: Frequency Plot of Genre