Project Report

Netflix Movie Ratings

Reno Chen, Jimin Huh, Wei Ching Lim, Yuqing Luo

1. Introduction

The popularity of Netflix has soared in recent years, making it one of the most famous movie streaming sites. The evaluation of a movie can be influenced by a person's personality, the genre he or she pursues, the environment in which he or she has grown, and various factors. Ratings, specifically, present an essential standard for the audiences when evaluating different movies, and also they can help the audiences compare and choose the best movies.

Our project utilizes the data of Netflix's movie ratings from IMDb to analyze the relationship among the ratings, released year, durations, and genres.

1.1 Research Questions

- Is there a relationship between star ratings and durations (in minutes)?
- How does each feature (country, genre, age rating, duration, or year released) impact the star ratings of movies?
- Is there any difference in average star ratings between the recent movies (released before 2018) and the new movies (released after 2018)?

2. Data

We gathered two pieces of data: a pre-existing dataset from Kaggle about Netflix movies and TV shows and a dataset scraped from IMDb.

2.1 Netflix Movies

The Netflix Movies and TV Shows1 dataset obtained from Kaggle contains 8,807 listings of Netflix movies and TV shows along with 12 features related to each title (e.g., genre, duration, country). The director, cast, and description were removed from this dataset as it does not relate to our research questions.

[1] https://www.kaggle.com/shivamb/netflix-shows

2.2 IMDb Netflix Movie Ratings

The IMDb Netflix2 data is a collection of 20,619 movies and TV shows. The information is reduced to the top 10,000 movies and TV shows, and each movie or TV show contains title, year released, age rating, duration in minutes, genres, star rating out of 10, description and casts. These irrelevant features were removed from the data, such as descriptions and casts. We specifically scrape data from IMDb other than from Kaggle because we wanted to scrape the star ratings scored by users and integrate the two data. We utilized the rvest package to scrape the titles accurately, release year, genres, duration, and star ratings from IMDb.

[2] https://www.imdb.com/search/title/?companies=co0144901&start=

2.3 Integrating Netflix Movies and IMDb Netflix Movie Ratings

After initial data cleaning, we merged the Kaggle and IMDb data. Since some features are slightly different for data from Kaggle and IMDb (e.g., duration is 98 minutes on Kaggle data and 99 minutes on IMDb data for the same movie), we assumed that the data from Kaggle was correct. The primary focus of our analysis is on film only, we removed records that were categorized as TV shows after the integration, and the number of rows was reduced from 8,807 to 1,727 and 11 columns. Table 1 below shows the name, type, and description of the features we used in the project.

Table 1 Description of the data

Column	Type	Description
show_id	text	Unique ID for each movie
title	text	Title of the movie
country	text	Country of production
date_added	date	Date the movie was added on Netflix
release_year	numeric	Actual year the movie is released
age_rating	text	Age rating of the movie
duration	text	Length of time for each movie
genre	factor	Genre of the movie
star_rating	numeric	Star rating of the movie
movie_timeFrame	text	Recent movies (before 2018) vs. New movies (after 2018)

3. Analysis

This project aims to examine the relationship between movie ratings and genres to understand which features (genre, duration, year released, and age rating) impact the star ratings of movies more or less. Then, we see the difference in average star ratings between the recent film (released before 2018) and the new film (released after 2018).

3.1 relationship between star ratings and durations

We wanted to figure out if there is a relationship between the ratings and the durations, and Figure 1 below displays the result.

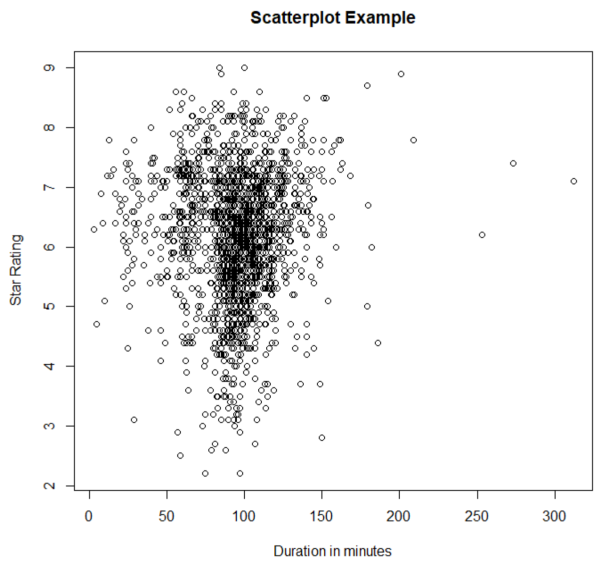

Figure 1 Scatterplot of relationships between star ratings and durations

The figure above indicates no special relationship between the star ratings and the durations. We further analyzed this relationship by carrying out a correlation test. Our correlation value is 0.02850527 showing a weak and positive correlation between the duration of movies and star ratings.

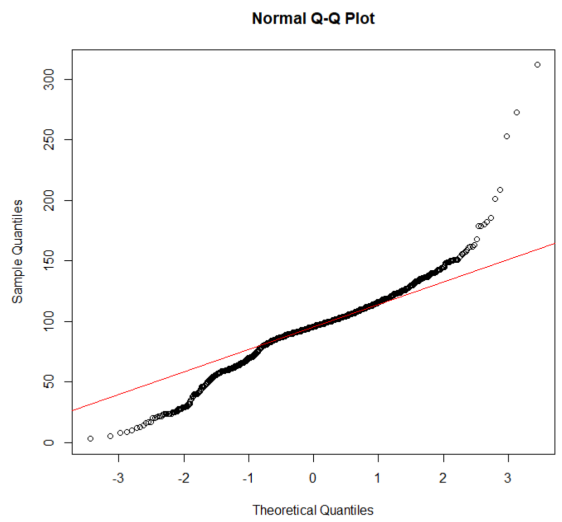

Figure 2 qqplot of normal distribution between star ratings and durations

The figure above also indicates that star ratings and durations do not have a normal distribution. To be more precise, we tried to use the Shapiro test and found out the p-value is 2.2e-16, which is very low. Because the p-value is less than 0.05, we reject the null hypothesis. Therefore, this did not follow the normal distribution.

3.2 Effect of each feature affecting star ratings

To determine whether the features, country, genre, age rating, duration, and released year impacted the star ratings of movies, we carried out two statistical tests on these features. First, the ANOVA test was used on the three categorical parts, country, genre, and age rating of movies, while the linear regression test was used on all the features mentioned previously.

3.2.1 Country on Star Rating

From the ANOVA tests, we reject the null hypothesis that the means for star ratings of movies is equal for all countries because there is a significant difference ($p = 7.52e-09$). Following this result, we tested the normality and identical variance assumptions of the ANOVA. Certain countries violated the normality assumption as the p-values were lower than 0.05. The results for the equal variance test were not violated as the p-value is 0.3366. We fail to reject the null hypothesis that star ratings based on countries have equal variance at the confidence level of 0.05. Therefore, one of the ANOVA assumptions is violated. Then, we conducted a follow-up test using Kruskal-Wallis, a non-parametric test. The p-value is 8.322e-09; thus, we reject the null hypothesis that the median star rating is equal for all countries. Our tests infer that some countries may impact the star ratings of Netflix movies.

Following the ANOVA test, our linear regression model results indicate three countries with p-values less than 0.05, whereby the estimate is not equal to zero. Colombia, Egypt, and the United Arab Emirates, with p-values of 0.03448, 0.0361, and 0.00103 respectively, had caused the star rating of movies to decrease by 0.9357504, 0.710664, and 3.2790686, as it changes from Argentina, the intercept value.

3.2.2 Genre on Star Rating

From the ANOVA test, we reject the null hypothesis that the means of movie star ratings are equal for all genres as there is a significant difference ($p < 2e-16$). Next, we tested the normality and identical variance assumptions of the ANOVA. Some genres violated the normality assumption as the p-values were less than 0.05. The results for the equal variance test were also violated as the p-value is 0.001854, which is lower than 0.05. We reject the null hypothesis that star ratings based on genres have equal variance. Therefore, both assumptions of ANOVA are violated. We conducted a follow-up test using Kruskal-Wallis, a non-parametric test. The p-value was <2e-16; thus, we reject the null hypothesis that the median star rating is equal for all genres. Our tests infer that specific genres may impact the star ratings of Netflix movies.

In our linear regression model that follows the ANOVA test, results indicate that certain genres impact the star ratings of movies both positively and negatively. Five genres increase the star ratings when they change from Action & Adventure, the intercept value. They are Comedies ($p = 0.04261$), Documentaries ($p = 6.09E-06$), Dramas ($p = < 2e-16$), Music & Musicals ($p = 1.51E-07$) and Stand-Up Comedy ($p = 2e-16$). The star rating of Comedies, Documentaries, Dramas, Music & Musicals, and Stand-Up Comedy is estimated to increase the star ratings by 0.2000357, 1.3812277, 0.4085715, 1.9639778, and 1.3408016, respectively.

Two genres that negatively impact the star ratings are Horror Movies and Thrillers when they change from the intercept value, Action & Adventure. Horror Movies has a p-value of 6.99E-05 and decreases the star ratings by 0.5181144. The genre Thrillers has a p-value of 0.02009 and reduces the star ratings by 0.5906233.

3.2.3 Age Rating on Star Rating

The results of the ANOVA test indicate that there is no significant difference between the means of each category of Netflix movies' age rating. The p-value is 0.0731, higher than 0.05; therefore, we cannot reject the null hypothesis. We can infer that the star rating is equal for all age ratings and that it does not impact the star rating of movies.

For our linear regression test, we will still include age ratings with the other four features in our model in order to parse out its effects.

3.2.4 Duration on Star Rating

Our linear regression model rejects the null hypothesis that the estimate is equal to zero as there is a significant difference ($p = 1.35e-14$). The estimate shows that if the duration of a movie increases by one unit, the star rating increases by 0.0095436. We can infer from further analysis that films with longer time get slightly higher star ratings from the viewers.

3.2.5 Released Year on Star Rating

Our linear regression model rejects the null hypothesis that the estimate is equal to zero as it shows a significant difference (p-value = 2.14e-05). The estimate indicates that if the released year increases by one unit, the star rating decreases by 0.0213684. We can infer that in recent years, star ratings of movies have dropped as viewers become critical of their movie consumption.

The assumptions of our linear regression model were tested. First, the linearity assumption of duration on star ratings and the released year on star rating are violated. The second assumption, normality, is also violated as the p-value of our linear regression model is 1.502e-14; therefore, our model does not follow a normal distribution. Third, our model is not suffering from homoscedasticity as the plot in Figure 3 does not show any visible trend. Lastly, one feature suffers from multicollinearity while four other features do not. The genre has a VIF value of 14.819237; thus, this regression coefficient was poorly estimated. Country, age rating, duration, and released year have VIF values of 1.838962, 4.847813, 1.656278, and 6.055044, respectively. These coefficients do not suffer from multicollinearity.

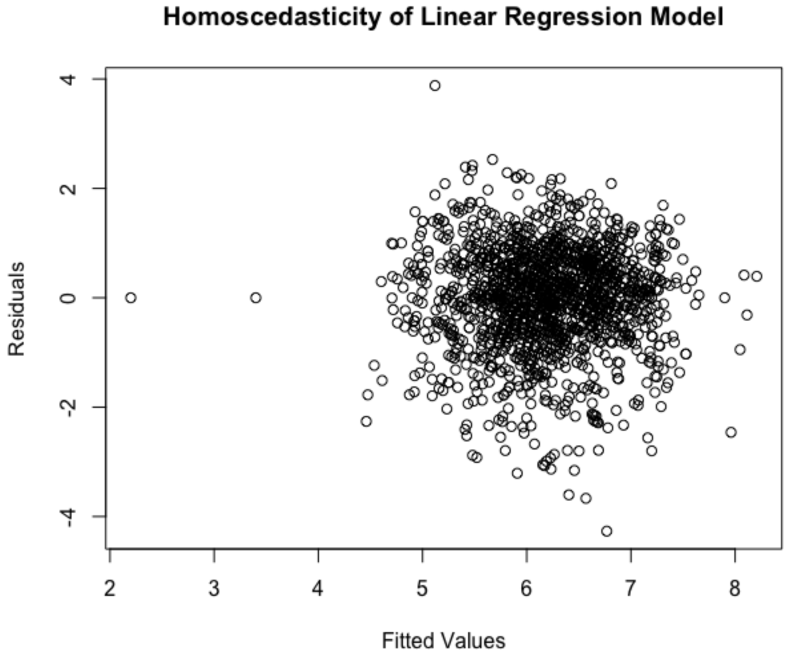

Figure 3 Homoscedasticity plot for the linear regression model of movie features on star ratings

3.3 Recent Movies vs. New Movies

We created dplyr summary tables to see the different star ratings between movies released before and after 2018. Recent movies are the movies released before 2018, and new movies are the movies released after 2018. The table below displays the resulting summary table.

Table 2 Summary table for the IMDb rating of new movies and recent movies

	total	average	median	min	max
New Movie (>2018)	1,036	6.12	6.2	2.2	9
Recent Movie (<2018)	691	6.32	6.4	2.2	9

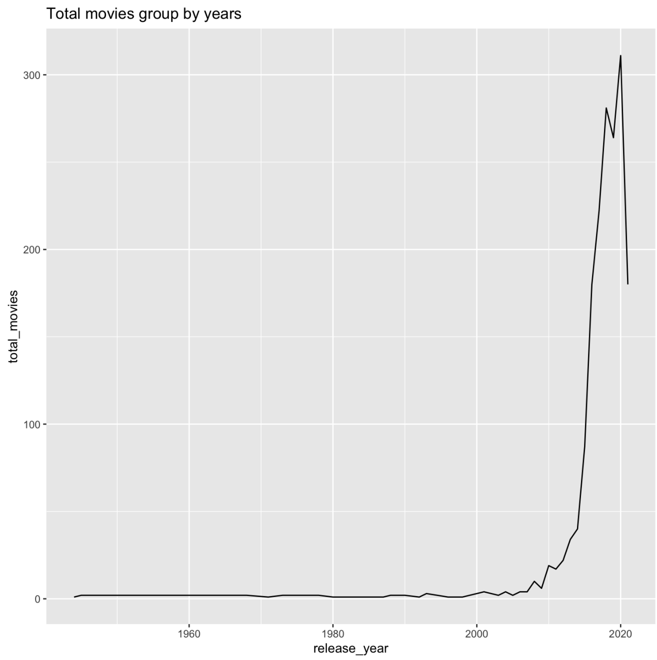

Figure 4 Total movies by years

Table 2 and Figure 4 above indicate that the movie market is becoming more active since about 350 movies were released in 3 years after 2018, which is half of the films released before 2018. The ratings of new movies were slightly lower than that of recent films, and this supports the analysis we have made above that the star ratings decrease as time goes by. We ran a t-test which also proved that the average ratings for new movies and recent movies have a significant difference (p-value = 0.0001737).

All code for analysis and visualization is included in the R script "19_Project_Analysis.R"

4. Conclusion

This project integrates Netflix movie data from Kaggle and IMDb to investigate the effect of features affecting Netflix movie ratings. We, in order to evaluate, used three following analyses: relationships between star ratings and durations, features that impact the star ratings the most or the least (specifically analyzed country, genre, age rating, durations, and released year on star ratings), and the comparison between recent movies (released before 2018) and new film (released after 2018). Using the summary table, visualization, ANOVA test, and linear regression test, we found out which features impact the star ratings for Netflix movies. Some countries possibly affect the star rating: Colombia, Egypt, and the United Arab Emirates had lower p-value; it is reasonable to indicate the star ratings of those countries had caused the star rating of movies to decrease. Several genres positively impacted the star ratings: Comedies, Documentaries, Dramas, Music & Musicals, and Stand-Up Comedy, while Horror Movies and Thrillers had a negative effect on it. The rating and duration, on the other hand, seem not to impact the ratings (although a slight increase is expected as duration gets longer), but we could predict that in recent years, star ratings of movies have dropped as viewers become critical of their movie consumption.

Although we used this dataset to analyze and test the hypothesis, we still encountered some limitations. After we merged the two datasets (one from Kaggle, another from Web Scraping), we had 1,727 instances for the Netflix Movies that had the IMDb rating. However, most movies in the dataset were released between 2010 to 2021. In other words, we had very few classic films that were released before 2000. Therefore, we did not have a complete dataset to contain both old movies and new movies with an IMDb rating. Second, some films have several genres, but we chose only to keep the first genre, which is the primary genre. And this might affect the overall ratings of specific genres. Third, some movies' production should also have multiple countries, but we chose to clarify and change those countries as international. This might affect the overall ratings of countries' movie production. Future work is utilizing web scraping and collecting more old movies (before 2000) with IMDb ratings, which helps us dig into more details on the ratings of different films. Also, we can add more features into the dataset to do further and broader analysis, such as the primary age range of people who like to watch this movie. In addition, we only focus on the film on Netflix and not generalizable data for the movie in other streaming services, such as Hulu, Disney Plus, and HBO Max, etc. Finally, the data only aim at the movie but not other types of videos, like TV shows, and we can expand more on different types of videos.