Exploring Content Trends in Netflix and Disney+ Using Age Categorization and Rotten Tomatoes Scores

Nixon Mundattuchundayil Paulson

December 03, 2024

Table of Contents

1. Introduction

In recent years, movie streaming platforms like Netflix, Hulu, Amazon Prime, and Disney+ have become central to how people consume media, offering a wide range of movies and TV shows for diverse audiences (Song, 2021) While Netflix has often been viewed as catering to a broad demographic, with content ranging from children's shows to adult-oriented films, Disney+ was traditionally associated with family-friendly content. However, Disney+ has increasingly begun offering more adult-oriented content, raising questions about the nature of its evolving audience and the age appropriateness of its catalogue (Adamou & Petridis, 2023; Badr et al., 2024).

This shift has sparked growing interest in understanding whether the content on these streaming platforms aligns with their perceived target audiences. Specifically, two key questions arise:

1. Is the age restriction for movies on Disney+ lower than those on Netflix?
2. Are the movies on Netflix generally rated higher on Rotten Tomatoes compared to those on Disney+?

To address these questions, we will perform both descriptive and inferential statistical analyses on a dataset containing movie details such as title, year of release, age appropriateness, RT scores, and availability on different streaming platforms. Descriptive statistics will be used to compare the distribution of age ratings and RT scores for movies on both streaming platforms. Hypothesis testing will then be employed to determine whether significant differences exist between the streaming platforms in terms of age restriction and movie ratings.

This study investigates the distribution of age restriction and RT scores across Netflix and Disney+. Findings reveal that Netflix dominates the total number of movies across all age restriction categories, particularly for 13+, 16+, and 18+ ratings, while Disney+ has a stronger presence in All Ages and 7+ categories. Additionally, Disney+ movies have significantly higher RT scores than Netflix movies, though the effect size is small.

The structure of the report is as follows: Section 2 provides a detailed description of the data and the problem. Section 3 outlines the methods used, including both descriptive analysis and hypothesis testing. Section 4 presents the results of these analyses, and Section 5 concludes with a summary of findings and a discussion on their implications.

2. Description of the Problem

This report aims to explore two key aspects of the movies available on Netflix and Disney+: the comparison of age restriction and the evaluation of RT scores. Specifically, we seek to examine the following:

- How does the age restriction for movies on Disney+ compare to those on Netflix?
- What is the difference in Rotten Tomatoes ratings between movies on Netflix and Disney+?

Data Overview

The dataset for this analysis was sourced from Kaggle and includes data on movies available on Netflix, Hulu, Prime Video, and Disney+. The key variables in the dataset are shown in **Appendix 1**. The dataset's collection method is not explicitly documented but is assumed to be aggregated from public movie databases or Application Programming Interface (APIs), as is common with Kaggle datasets (*Find Open Datasets and Machine Learning Projects,* 2010).

Missing Data

The dataset contains 9,515 entries, but significant missing data was identified:

- 4,177 entries (43%) were missing the age restriction information.
- 7 entries (0.15%) were missing the RT scores.

Given the potential impact of missing data on analysis, two separate data frames were created for the two research questions. In the first data frame (used for age restriction analysis), entries with missing age restriction values were excluded, leaving 5,338 observations. The 7 missing RT entries were retained, as they are irrelevant to this analysis. In the second data frame (used for RT scores analysis), the 7 entries with missing RT scores were removed, leaving 9,508 observations. Using separate datasets for each research question maximizes data retention, leading to greater statistical power and more representative results (Little & Rubin, 2014). This approach also ensures that each research question focuses on its specific subset of data, without being constrained by the missing data in the other. Finally, both the data frames were filtered to include only Netflix and Disney+ films. Three movies appeared on both streaming platforms, and they were retained in both for consistency.

Data Cleaning

The Age Restriction column, initially presented in a categorical format, was transformed into numerical values to facilitate analysis. The categories were mapped to numerical values as follows —"All ages" = 0, "7+" = 1, "13+" = 2, "16+" = 3, and "18+" = 4. This transformation ensured that the data was converted to a categorical ordinal variable facilitating further analysis (Field, 2017).

3. Methods

Sample Size

The analysis focused on two datasets: one for age restriction and another for RT scores, which were filtered to include only Netflix and Disney+ films. After cleaning the datasets by removing rows with missing values, the final sample included 2,623 observations for age restriction and 4,610 observations for RT scores.

Variables

The variables used in the analysis are age restriction (all ages, 7+,13+,16+, and 18+), RT scores (0-100), and streaming platform (Netflix and Disney+).

Descriptive analysis

Descriptive analysis was employed to visualize the distribution of age restriction and RT scores across the two streaming platforms. The specific methods used for these visualizations are detailed below:

Bar Graph (Count Plot): A bar graph (or countplot) was employed to visualize the distribution of age restriction for Netflix and Disney+. This method allows for comparing the frequency of different categories of age restriction between the two streaming platforms. A bar graph is constructed by counting the occurrences for each category of a variable (in this case, age restriction) and representing these counts as bars. The height of each bar corresponds to the frequency of the respective category. By summarizing the data visually, bar graphs highlight disparities or similarities in age restriction distributions across the streaming platforms (Cleveland, 1994; Russell, 2021).

Heatmap: A heatmap was used to represent the frequency of age restriction across streaming platforms. A heatmap is a data visualization tool that uses colour to represent values in a matrix or grid format (in this case, age restriction and streaming platform). Each cell in the heatmap corresponds to a value in the dataset, and the colour intensity indicates the magnitude of that value. By examining the colour gradient, one can quickly discern areas with higher or lower frequencies of age restriction, providing insights into the distribution and enforcement of age-related content across various streaming platforms (Cleveland, 1993; Friendly, 2002; Wilkinson & Friendly, 2009).

Boxplot: A boxplot was used to visualize the distribution of RT scores for each streaming platform, highlighting the median, interquartile range (IQR), and potential outliers. “A Boxplot displays batches of data” (Tukey, 1977).This method is based on key quartiles: the first quartile (Q1), representing the 25th percentile; the third quartile (Q3), representing the 75th percentile; and the median (Q2), which is the 50th percentile. The IQR, calculated as $Q3 - Q1$, measures the spread of the middle 50% of the data. The boxplot is constructed by plotting a box from Q1 to Q3, with a line indicating the median. Whiskers extend to $Q1 - 1.5 \times \text{IQR}$ and $Q3 + 1.5 \times \text{IQR}$, capturing the range of the data excluding outliers, which are points lying beyond the whiskers (McGill et al., 1978).

Histogram: A histogram was used to represent the distribution of RT scores for each streaming platform. “A histogram conveys visual information of both the frequency and relative frequencies of observations; that is, the essence of a density function. The classical frequency histogram is formed by constructing a complete set of nonoverlapping intervals, called bins, and counting the number of points in each bin” (Scott, 2015).The histogram provided a straightforward representation of the data, facilitating a comparative analysis of the two streaming platforms. Kernel Density Estimation (KDE) was overlaid on the histograms to produce a smooth curve, offering insights into the continuous distribution of scores. Bin sizes were selected to ensure meaningful granularity while maintaining visual clarity, as per best practices for histogram analysis (Cleveland, 1994; Scott, 2015).

Hypothesis Testing

Mann-Whitney U Test for Age restriction

To compare the distributions of age restriction between Netflix and Disney+, a Mann-Whitney U test was employed as the variable age restriction is ordinal.

Null Hypothesis (H_0): The distribution of age restriction for movies on Disney+ is the same as or higher than that for Netflix. $\text{Age restriction (Disney+)} \geq \text{Age restriction (Netflix)}$

Alternate Hypothesis (H_1): The distribution of age restriction for movies on Disney+ is lower than that for Netflix. $\text{Age restriction (Disney +)} < \text{Age restriction (Netflix)}$

The Mann-Whitney U test is a non-parametric test used to determine whether there is a significant difference between two independent samples. Let the age restriction for Netflix and Disney+ be represented by two independent samples $X = \{x_1, x_2, ..., x_n\}\ and\ Y = \{y_1, y_2, ..., y_m\}$, repectively. The U statistic is defined as:

$$U = mn + \frac{m(m+1)}{2} - T_1$$

Where, T_1 is the sum of ranks for sample X (Netflix), m and n are the sample sizes for the two groups, and $\frac{m(m+1)}{2}$ is the expected rank sum under the null hypothesis that the distributions of both samples are equal. Similarly, U for the second group can be calculated. The test statistic used is the smaller U. (Mann & Whitney, 1947).

The one-sided version of the test was used because we hypothesize that Disney+ will have lower age restriction than Netflix. This is based on the ordinal nature of the data and the directional nature of the hypothesis (i.e., we expect one streaming platform to have lower age restriction) (Nachar, 2008).

Shapiro-Wilk Test for Normality for RT score

The Shapiro-Wilk test was used to assess the normality of the RT scores. The null hypothesis of the test is that the data follows a normal distribution. The test statistic W is defined as:

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2}$$

Where, $x_{(i)}$ are the ordered data points, a_i are constants dependent on the data size, and $\overline{x}$ is the sample mean. If the p-value from the Shapiro-Wilk test is less than 0.05, the null hypothesis of normality is rejected, indicating that the data significantly deviates from a normal distribution. In such cases, non-parametric methods like the Mann-Whitney U test are preferred (Shapiro & Wilk, 1965).

Mann-Whitney U Test for RT score

Since the Shapiro-Wilk test indicated that RT scores do not follow a normal distribution, the Mann-Whitney U test (two-tailed) was used to compare the distributions of RT scores between Netflix and Disney+. It is described and defined in the above paragraph. And further one-tailed Mann-Whitney U test was used to check the direction of significance (Nachar, 2008).

Null Hypothesis (H_0): There is no difference in the distribution of RT scores between movies on Netflix and Disney+. $\text{RT scores (Netflix)} = \text{RT scores (Disney +)}$

Alternate Hypothesis (H_1): There is a significant difference in the distribution of RT scores between movies on Netflix and Disney+. $\text{RT scores (Netflix)} \neq \text{RT scores (Disney+)}$

Statistical Software and Libraries

All analyses were conducted using Python, employing several libraries to streamline data manipulation, visualization, and statistical testing. Data manipulation was performed using *Pandas* (McKinney, 2010)for efficient data handling and *NumPy* (Harris et al., 2020) for numerical computations. Visualizations were created using *Matplotlib* (Hunter, 2007) and enhanced with *Seaborn* (Waskom, 2021) for advanced plotting techniques, including histograms with KDE. Statistical tests, such as the Mann-Whitney U test and Shapiro-Wilk test, were conducted using the *SciPy* library (Virtanen et al., 2020). Summary statistics, including mean and standard deviation, were computed using *NumPy*, a foundational library for numerical and statistical computations in Python. The *rcParams* module from *Matplotlib* was employed for customizing visualizations, such as setting font styles and other graphical parameters.

4. Evaluation

Descriptive Statistics

Table 1, Fig. 1, and **Fig. 2** describes the distribution of 2,623 movies across Disney+ and Netflix by age restriction categories. Of these, Disney+ hosts 725 movies (27.64%), while Netflix offers 1,898 movies (72.36%).

In the all ages category, comprising 514 movies, Disney+ accounts for 370 movies (71.98%), and Netflix for 144 movies (28.02%). For the 7+ category, which includes 600 movies, Disney+ has 278 movies (46.33%) while Netflix has 322 movies (53.67%). In the 13+ category, out of 474 movies, Disney+ contributes 70 movies (14.77%) and Netflix contributes 404 movies (85.23%). In the 16+ category, Disney+ provides 4 movies (2.58%) of the 155 total, while Netflix hosts 151 movies (97.42%). Lastly, in the 18+ category, which includes 880 movies, Disney+ has only 3 movies (0.34%), while Netflix accounts for 877 movies (99.66%). Netflix leads in the total number of movies across all categories, particularly dominating the 13+, 16+, and 18+ categories. Disney+ has a higher share in the all ages category and a moderate presence in the 7+ category.

Table 1: Summary statistics of age restriction across streaming platforms

Streaming platform	**Disney+ (%)**	**Netflix (%)**	**Total**
Age Restriction			
All ages	370 (71.98)	144 (28.02)	**514**

7+	278 (46.33)	322 (53.67)	**600**
13+	70 (14.77)	404 (85.23)	**474**
16+	4 (2.58)	151 (97.42)	**155**
18+	3 (0.34)	877 (99.66)	**880**
Total	**725 (27.64)**	**1898 (72.36)**	**2623**

Figure 1: Distribution of age restriction by streaming platforms

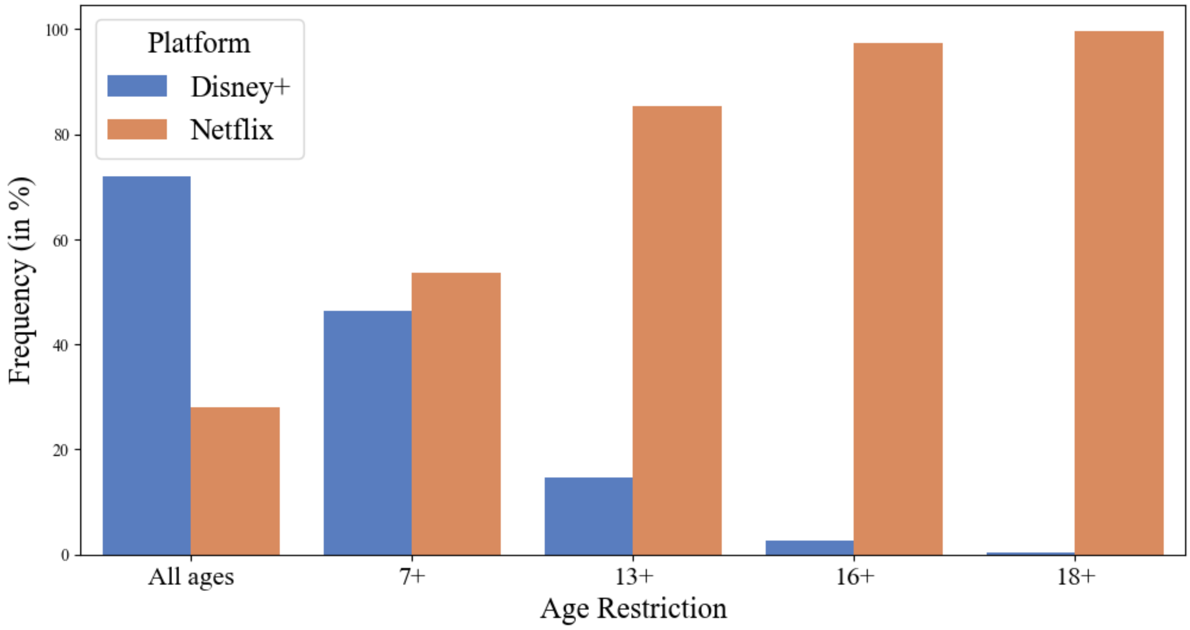


Figure 2: Distribution of age restriction across streaming platforms

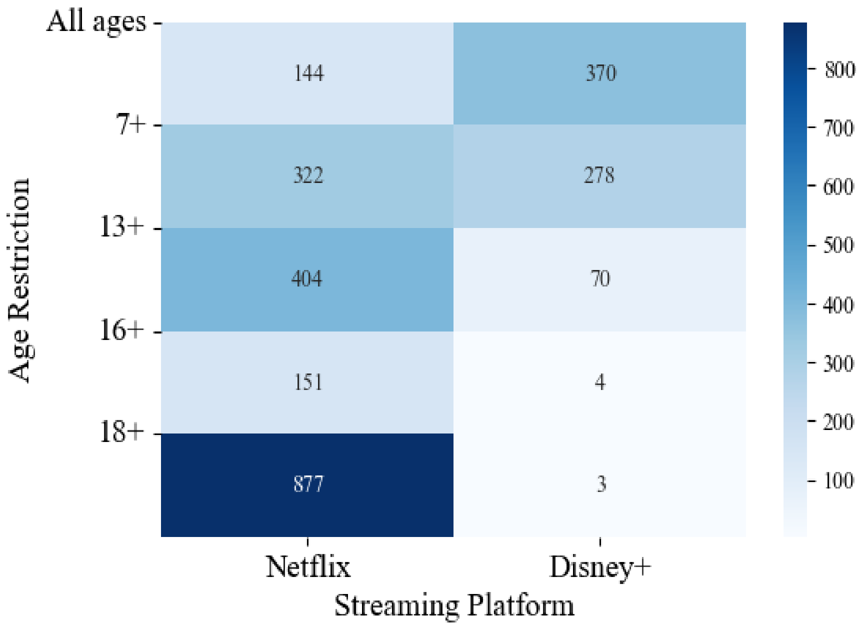
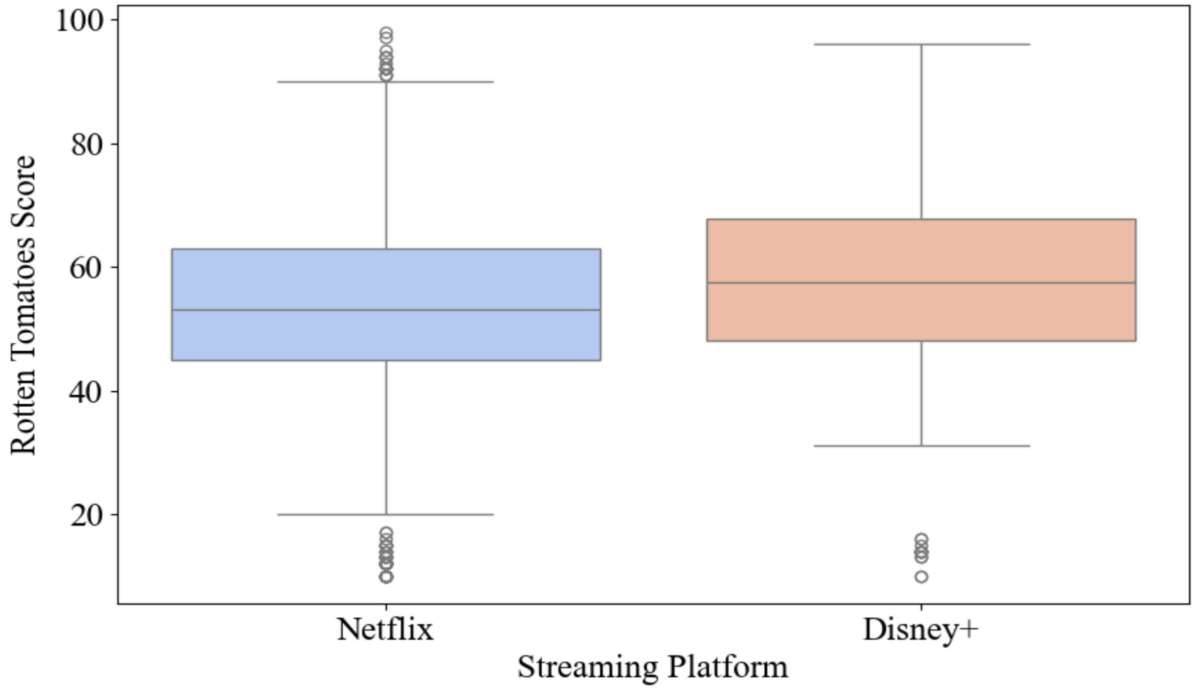


Table 2 shows the summary statistics for RT scores across Netflix and Disney+. There are 3688 movies on Netflix and 922 on Disney+. Netflix has a mean score of 54.45 (s.d. = ±13.85), while Disney+ has a slightly higher mean of 58.31 (s.d. = ±13.95). The median score for Netflix is 53, while Disney+ has a higher median of 57.50. The inter-quartile range for Netflix is 18 (ranging from 45 to 63), while Disney+ has an IQR of 19.8 (ranging from 48 to 67.8). The mode for Netflix is 46, and for Disney+, it is 48. Overall, Disney+ tends to have slightly higher RT scores, with both streaming platforms exhibiting similar variability in their ratings. **Fig. 3** shows this distribution as a box plot.

Table 2: Summary statistics of RT scores across streaming platforms

Streaming platform	Mean	Standard Deviation	Median	Inter-quartile Range	Mode	Total (n)
Netflix	54.45	13.85	53.00	18 (63-45)	46	**3688**
Disney+	58.31	13.95	57.50	19.8 (67.8-48)	48	**922**

Figure 3: Distribution of RT scores across streaming platforms

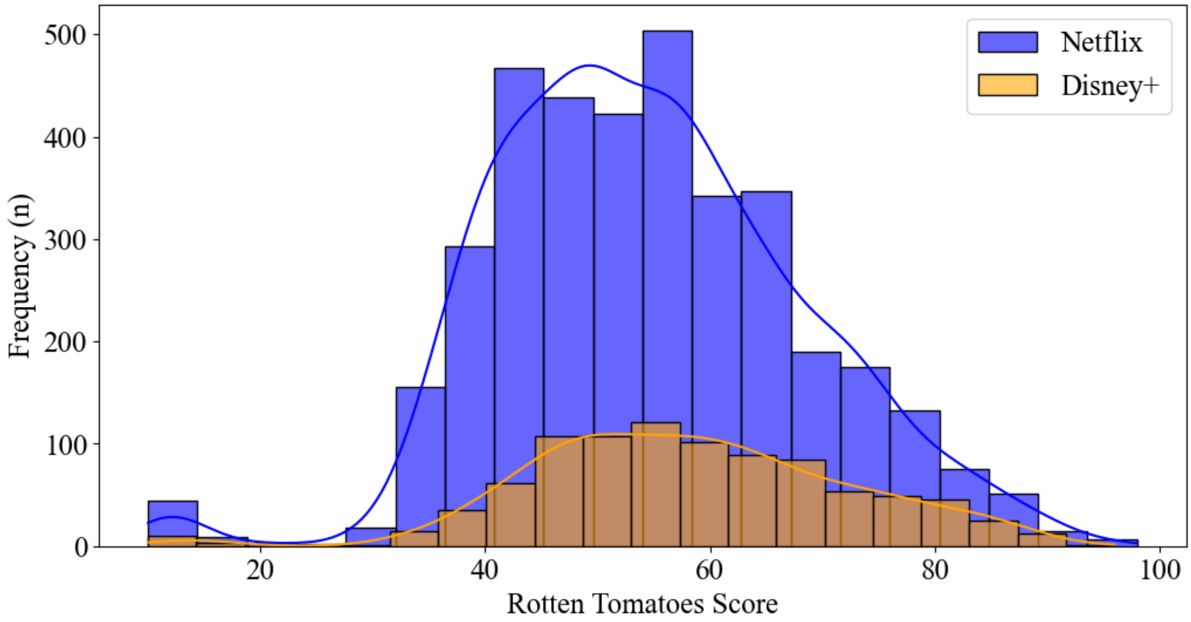


Hypothesis Testing

From the one-tailed Mann-Whitney U test to evaluate whether Disney+ has lower age restriction ratings than Netflix, it was found that age restriction for Disney+ were significantly lower than those for Netflix (U= 166350, p<0.01). The rank-biserial correlation, as the measure of effect size, was found to be -0.379, indicating a medium to large effect size **(Table 3)**.

To investigate the difference in RT scores between Netflix and Disney+, a Shapiro-Wilk normality test revealed that the data were non-normally distributed (W = 0.984, p<0.01), which can be observed in the KDE plot **(Fig. 4)**. Hence, we used Mann-Whitney U Test, a non-parametric test, to investigate the association. The test results were statistically significant (U = 1415471, p<0.01), suggesting a difference in RT scores between the two streaming platforms. The rank-biserial correlation, used as a measure of effect size, was calculated to be -0.167, indicating small practical significance **(Table 3)**.

Figure 4: KDE of RT scores across streaming platforms



To understand the directionality of association, a one-tailed Mann-Whitney U test was performed to evaluate whether Disney+ has higher RT scores than Netflix. The results indicate that RT scores for Disney+ were significantly higher than those for Netflix (U=1984865, p<0.01). The rank-biserial correlation, used as a measure of effect size, was calculated to be 0.167, indicating small practical significance despite the statistical significance of the results **(Table 3)**.

Table 3: Mann-Whitney U Test Results

Group Comparison	U Statistic	p-value	Effect Size
Age restriction (One tailed test)	166350	<0.01	-0.379
RT scores (Two tailed test)	1415471	<0.01	-0.167
RT scores (One tailed test)	1984865	<0.01	0.167

5. Summary of findings

This study aimed to explore the distribution of movies by age restriction and RT scores on Netflix and Disney+, addressing key questions about streaming platform content diversity and quality. Descriptive analysis revealed that Netflix offers a significantly larger number of movies (72.36%) compared to Disney+ (27.64%), with a notable dominance in the 13+, 16+, and 18+ categories. Conversely, Disney+ leads in the all ages category and maintains a moderate presence in the 7+ category. Statistical testing confirmed that age restriction for Disney+ were significantly lower than those for Netflix, with a medium to large effect size.

Regarding quality, Disney+ exhibited slightly higher RT scores on average than Netflix. The median score for Disney+ was 57.50, compared to 53.00 for Netflix. These differences were confirmed through the Mann-Whitney U test, with statistically significant results and a small effect size. Additional analysis highlighted a higher median and narrower interquartile range for Disney+, indicating a slight edge in content quality perception.

In the real-world context, these findings suggest strategic differences between the streaming platforms: Netflix appears to cater more broadly to older audiences, while Disney+ emphasizes family-friendly content. The higher average quality ratings on Disney+ may reflect its focus on carefully curated content. These insights are valuable for viewers making informed choices and for content producers aiming to tailor offerings to audience demographics.

Future research could explore other factors influencing streaming preferences, such as the diversity of genres, regional content availability, or subscription pricing. Expanding the analysis to include additional streaming platforms and qualitative measures of audience satisfaction would further enrich our understanding of the streaming landscape.

Bibliography

Adamou, C., & Petridis, S. (2023). *Television by Stream: Essays on Marketing, Content and Audience Worldwide*. McFarland.

Badr, N., Sharaf, S., & Mahrous, A. (2024). Streaming wars: An analysis of the growth and competitive landscape of the subscription video-on-demand services market. *EuroMed J. of Management*, *6*, 23–41. https://doi.org/10.1504/EMJM.2024.135992

Cleveland, W. S. (1993). *Visualizing data*. AT & T Bell Laboratories.

Cleveland, W. S. (1994). *The elements of graphing data* (Rev. ed.). AT&T Bell Laboratories.

Field, A. (2017). *Discovering Statistics using IBM SPSS Statistics* (5th ed.). SAGE Edge.

Find Open Datasets and Machine Learning Projects. (2010). Kaggle. https://www.kaggle.com/datasets

Friendly, M. (2002). *Visualizing Categorical Data*.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, *9*(3), 90–95. Computing in Science & Engineering. https://doi.org/10.1109/MCSE.2007.55

Little, R. J. A., & Rubin, D. B. (2014). *Statistical Analysis with Missing Data*. John Wiley & Sons.

Mann, H. B., & Whitney, D. R. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics*, *18*(1), 50–60. https://doi.org/10.1214/aoms/1177730491

McGill, R., Tukey, J. W., & Larsen, W. A. (1978). Variations of Box Plots. *The American Statistician*, *32*(1), 12–16. https://doi.org/10.2307/2683468

McKinney, W. (2010). *Data Structures for Statistical Computing in Python*. 56–61. https://doi.org/10.25080/Majora-92bf1922-00a

Nachar, N. (2008). The Mann-Whitney U: A Test for Assessing Whether Two Independent Samples Come from the Same Distribution. *Tutorials in Quantitative Methods for Psychology*, *4*(1), 13–20. https://doi.org/10.20982/tqmp.04.1.p013

Russell, J. M. (2021). 2.2 Displaying and Describing Categorical Data. In *Significant Statistics: An Introduction to Statistics*. https://pressbooks.lib.vt.edu/introstatistics/chapter/displaying-categorical-data/

Scott, D. W. (2015). *Multivariate Density Estimation*.

Shapiro, S. S., & Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, *52*(3/4), 591–611. https://doi.org/10.2307/2333709

Song, M. (2021). A Comparative Study on Over-The-Tops, Netflix & Amazon Prime Video: Based on the Success Factors of Innovation. *International Journal of Advanced Smart Convergence*, *10*(1), 62–74. https://doi.org/10.7236/IJASC.2021.10.1.62

Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley Pub. Co.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., Van Der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … Vázquez-Baeza, Y. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Waskom, M. (2021). seaborn: Statistical data visualization. *Journal of Open Source Software*, *6*(60), 3021. https://doi.org/10.21105/joss.03021

Wilkinson, L., & Friendly, M. (2009). The History of the Cluster Heat Map. *The American Statistician*, *63*, 179–184. https://doi.org/10.1198/tas.2009.0033s

Appendix

Appendix 1: Key variables in the dataset

Variable	**Description**	**Scale level**
Movie Title	A categorical variable indicating the title of the movie	Nominal
Year of Release	A numerical variable representing the release year of the movie.	Interval

Age Restriction	A categorical variable indicating the age rating of the movie (all age, 7+, 13+, 16+, and 18+)	Ordinal
RT Score	A continuous variable representing the RT scores of the movie, ranging from 0 to 100.	Interval
Streaming platform	A categorical variable indicating which streaming platform the movie is available on (Netflix, Hulu, Prime Video, Disney+).	Nominal