Streaming Services: Which One Is For You?

**Author:**

Zach Mason

Nicolas Mallett

Mauricio Gil

Introduction

As the years have gone by the number of available streaming services has continued to increase as have their individual costs. Some providers offer bundles and lower prices, which make them more affordable overall but leave you with a service you don’t necessarily use. Allthewhile, companies have continued to add shows and movies relentlessly to become more competitive, but in a world of plenty sometimes it’s hard to actually decide what you want to watch. In search of answers to some of these modern day problems we decided to look at three major companies (Netflix, HBO, Amazon Prime) to discern which company offers the most value in terms of your time and money based on varying factors.

Methodology

We began by merging and cleaning three datasets that we found on Kaggle. Once cleaned, our first goal was to identify the top 20 shows and movies on each streaming service based on IMDb scores. To identify the top shows we sorted the IMDb ratings on our dataframes in a descending order from greatest to least (10-0.0) for each of the three platforms (Image 1).

Next, we wanted to analyze which platforms had a broader range of cinema based on the release year. Using our initial merged dataframe, we sorted films and tv shows into bins beginning at the year 1963, creating a new bin ever ten years until 2023. This gave us a new column we named ‘Time Period’ and later injected in our initial dataframe. The next step was to count how many shows were present in each time period using the .value\_counts() function. Finally , we utilized matplotlib to create a bar graph comparing the Time Period counts for the three platforms (Image 2). The graph shows that Amazon Prime does a good job of providing younger and older viewers options. Out competing its competitors in older cinema (~1200 titles) while still having many newly produced programs.

With individuals and families purchasing many different services we wanted to uncover who offers the widest variety of family friendly content and who does not. To make the data wrangling easier we separated the various rating categories (TV-MA ,R, TV-14, PG-13, PG, TV-PG, G, TV-Y7, TV-Y, TV-G, NC-17, TV-Y7-FV) into 4 distinct groups; Mature, Teen, Family, and Children. After performing a count of each rating in the datasets we depicted our findings in three unique pie charts (Images 3, 4, 5). The pie charts tell us that there seems to be more Children programs on Netflix (9.6%) so if you have a family with young kids Netflix is the best platform providing content for all ages. HBO seems to have the most mature rated content (47.8%) and is presumably the best choice, with the most variety for those without kids.

Recently, shows have began producing shows called limited series, but how well are they being received by critics. We wanted to know if the length of a series was at all related to its IMDb score. To effectively plot our data we first had to clean the data set of any blank values in the Seasons and IMDb Score columns we wanted to analyze. Then to plot our data we used matplotlib to import a scatter plot design where we illustrated a line of best fit. The results show no significant results indicating that there is no correlation between number of seasons of a show and its IMDb score (Image 6).

With costs for streaming services increasing every year it makes sense to purchase one that you’ll get your money’s worth. To find the average ratings for each platform we first created an aggregate summary (mean, median, mode, standard variation and variance) focusing on media type (show or movie) and compiling IMDb Score of each (Image 7). The aggregate summary allowed us to create a box and whisker plot to better visualize how each streaming service compared in movies and shows (Image 8).From the boxplot we can see that HBO has the higher average score for movies (6.8) and a higher median(7.0) with a larger majority of its shows scoring within 7.1 - 2.3 out of 10 ( image 9). When looking at the aggregate summary of shows imposed on a boxplot there is no clear winner.

Lastly, we observed the number of genres on each streaming platform. Many shows and movies are considered to be multiple genres such as rom-com or mystery-thriller, etc. For us to categorize each of these we had to separate the genre rating by comma using a string split function. After counting each genre we used matplotlib to apply our counted genres to a bar graph. Once this was completed for each platform we compiled them into a dataframe and graphed the data as a bar (Image 10). Seeing the genre counts all laid out on the graph may help point those seeking a new streaming platform in the direction of which one best tailors to their interests.

References

* [https://www.kaggle.com/datasets/dgoenrique/amazon-prime-movies-and-tv-showsows](https://www.kaggle.com/datasets/dgoenrique/amazon-prime-movies-and-tv-shows)
* <https://www.kaggle.com/datasets/dgoenrique/hbo-max-movies-and-tv-shows>
* <https://www.kaggle.com/datasets/dgoenrique/netflix-movies-and-tv-shows>

Appendices

Image 1:

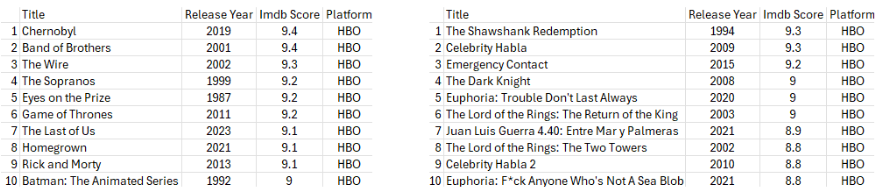


Image 2:

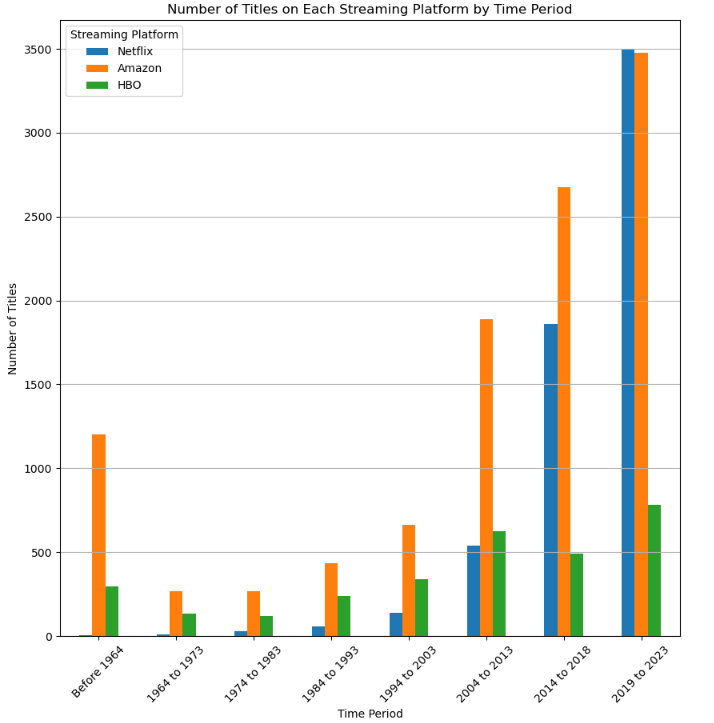


Image 3:

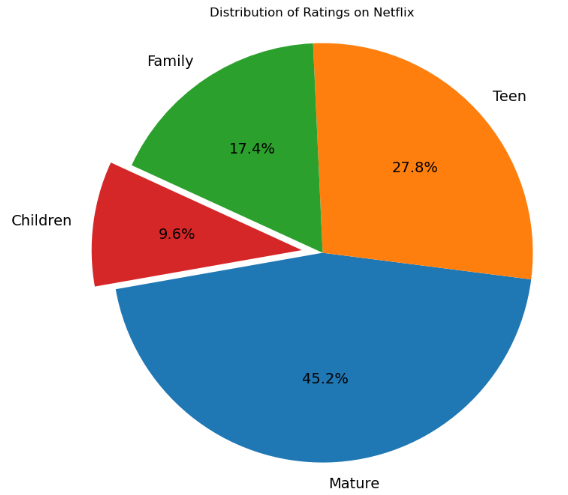


Image 4:

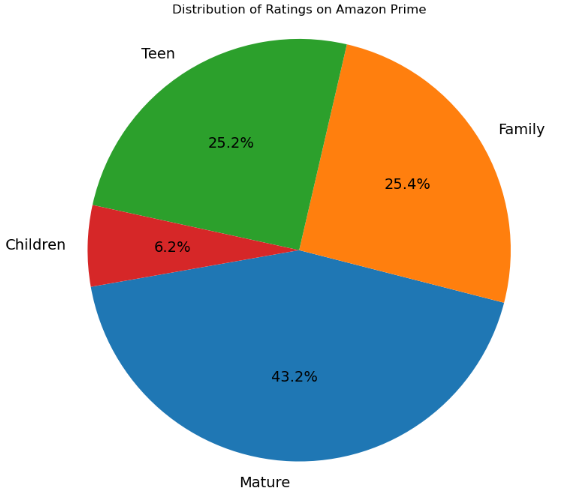


Image 5:

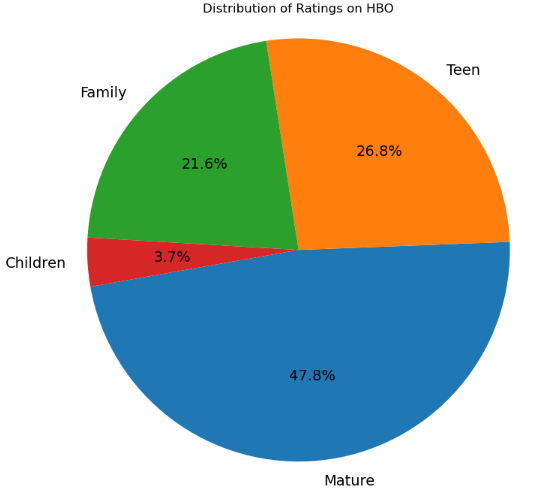


Image 6:

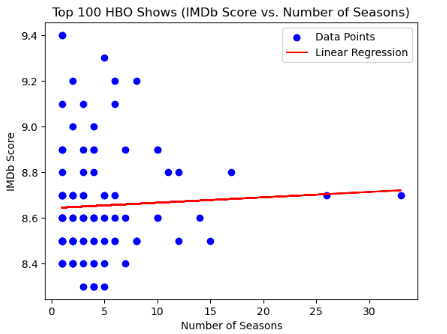


Image 7:

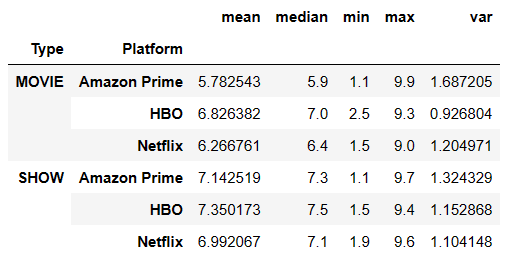


Image 8:

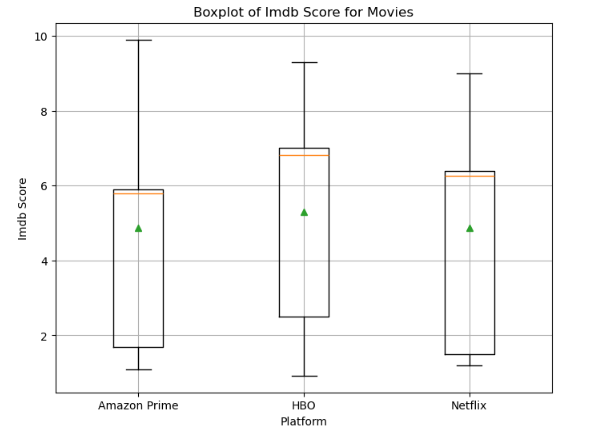


Image 9:

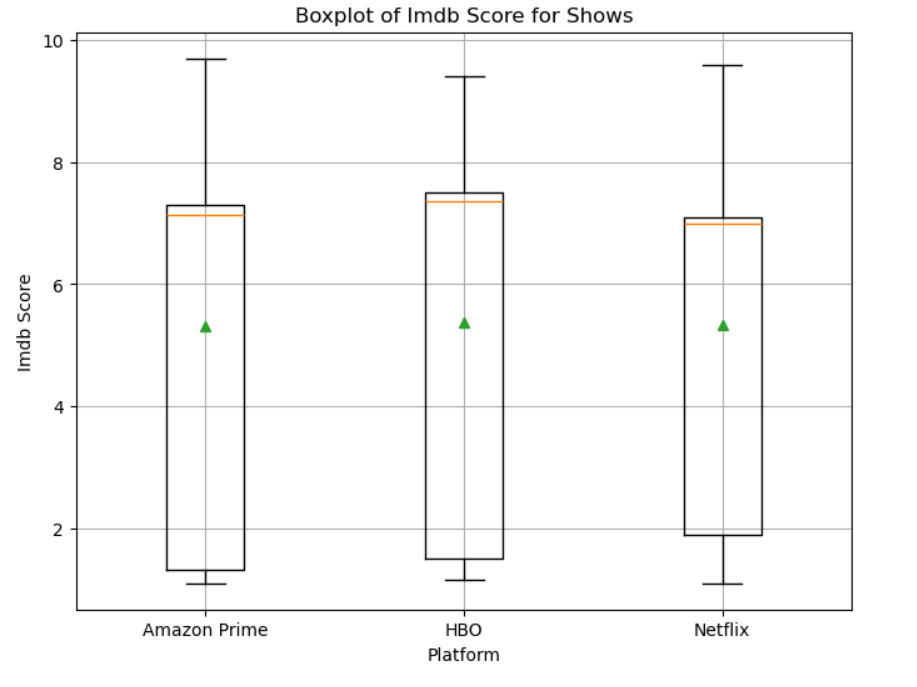


Image 10:

