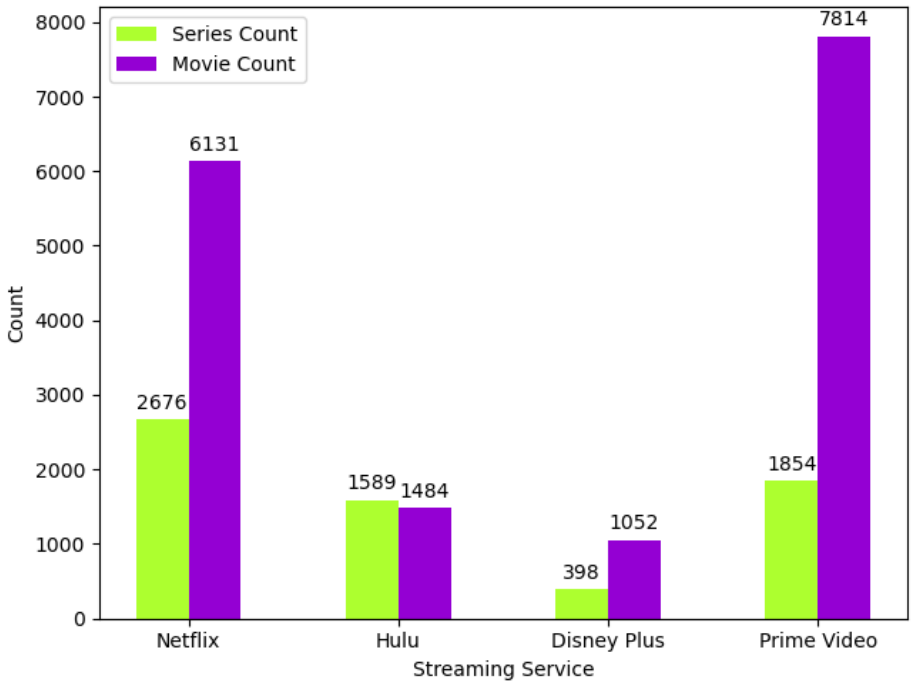
Introduction

At the beginning, when we were brainstorming potential project ideas, we sought to identify a common passion/interest that could serve as a launchpoint for any research. We identified entertainment- TV, movies, and music, amongst others- as an area that not only us (as the people who were doing the initial research) but also our peers would be able to relate to. We decided to pursue an analysis of SVOD (Streaming Video on Demand) services as we found that that category had the most data freely available and there were many angles we could explore as our research evolved. In addition, we acknowledged the recent influx of new SVOD platforms into the market, which has resulted in an increasingly saturated market; this combined with the recent inflation means that consumers who are subscribed to several platforms have or likely will begin to cut their consumption of certain brands. In the course of our initial research, we found that many datasets confirmed that the top 3 platforms are Netflix, Prime Video, and Hulu (we added Disney+ because Hulu and Disney+ have the same parent company) (Stoll, 2023). We were curious if this is because those three brands have established themselves as industry giants through effective promotion and a keen awareness of what direction ever-changing audience interests are moving in. Therefore, we determined to establish the following null hypothesis: "There is no connection between the content on a streaming service and its brand identity, and its wider appeal to its target audience." The following analysis will first assess the content offered on each platform and determine whether it may offer some insight into potential brand identities that each platform leverages during its promotional activities. Then, it will move into discussing how well these offerings reflect and/or stem from existing subscribers' interests, and potentially opportunities for each platform to gain an advantage in the market. Finally, we will close by providing an overview of how well each platform is meeting its audience's interests at this current moment in time, and revisit our null hypothesis to determine whether we disproved it.

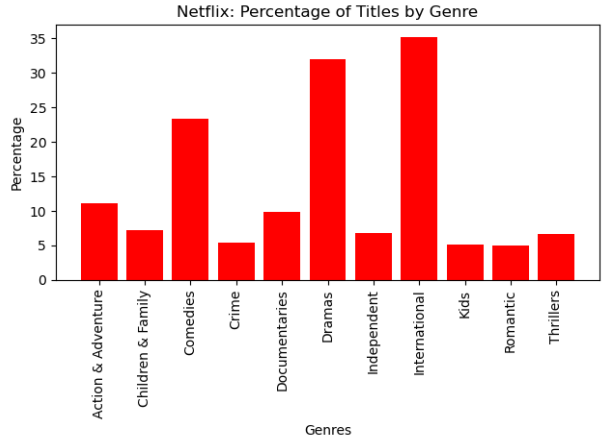
What types of content and genres are streaming services adding to their libraries?

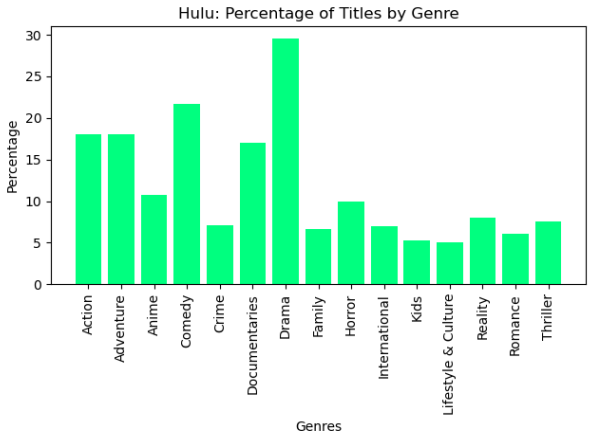
Using datasets from Kaggle, it was possible to look at the content that streaming services are adding to their catalogs. The specific streaming services that were chosen were Netflix, Hulu, Disney+, and Prime Video. These datasets separated the type of content being added into two categories: movies and series (TV shows).

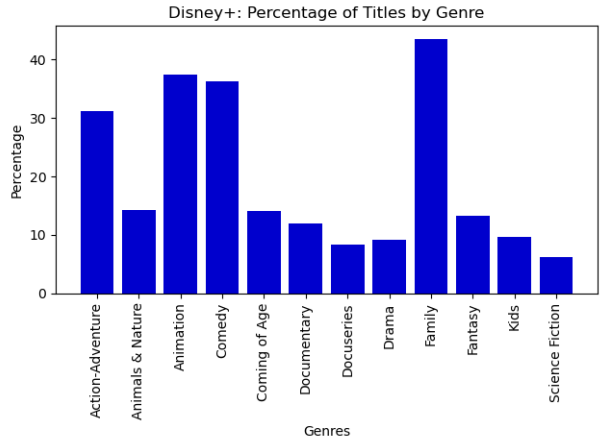


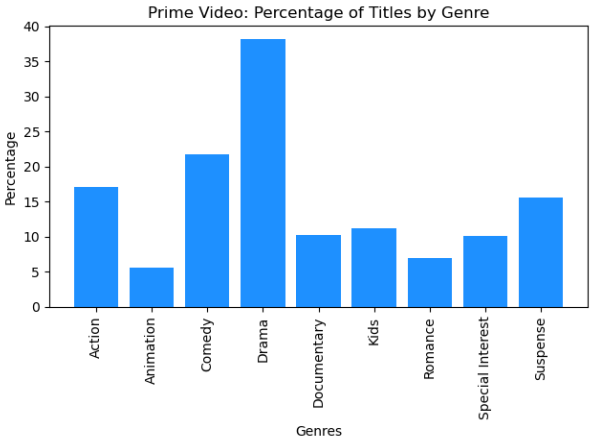
An analysis of the number of movies and series that each streaming service added to their catalog revealed that Netflix, Amazon Prime, and Disney+ had all added more movies to their respective libraries than series. Prime Video had the visually greater difference between movies and series, with a difference of 5,960 titles. Prime Video also had the highest number of movies added at 7,814 movies. Hulu was the only streaming service that added a greater number of series than movies, although the difference was only 105 titles. One limitation of these datasets is that they did not provide definitive reasoning for the difference in types of content. Some possible reasons are that movies might be more accessible to streaming services or they might better fit the brands that these streaming services are trying to create. When looking at the overall number of titles added, it was clear that Hulu and Disney+ have the least amount of content added in general. One possible reason for this is that Hulu and Disney+ are owned by the same company, so they might try to spread content out between the two streaming services. Another reason might be that Disney+ made it a point during their launch that content would not have as much turnover.

Another way to look at the content added to the streaming services is to separate the titles by genres. By taking each genre used by a streaming service, the data can be used to reveal how much of the content on a streaming service was included in each genre. In the dataset, each title is associated with 1-3 genres that the streaming services have applied. Because of this, a title could appear in multiple categories. This allowed for a better picture of how each streaming service truly categorizes their content. In the analysis, genres were only considered if they accounted for at least 5% of the total amount of titles added to a streaming service. This allowed the analysis to reveal the genres most often associated with content within a streaming service. Several conclusions could be drawn by comparing and contrasting the individual streaming services and what genres best represent their content.









Looking at every streaming service, drama was either the largest or second largest category on every streaming service, except Disney+. This means that three out of four streaming service were describing a large amount of their content as dramas. It also means that these streaming services must find value in obtaining shows of this genre. Looking at unique genres on different streaming services, the analysis showed that over 35% of titles added to Netflix were assigned as international. This was the largest genre on Netflix. This could be an attempt by Netflix to individualize themselves and create a brand as platform for international content. Other streaming services had significantly less international titles. Disney+ did not apply the genre to any of its titles. Looking at Disney+, over 40% of the content on the streaming service fell under the family genre. The next largest genre was animation. Disney+ clearly wants to create a brand as a family friendly streaming service. Comparing Disney+ and Hulu, the analysis also showed how they are being curated to create their own individual brands, despite being owned by the same company.

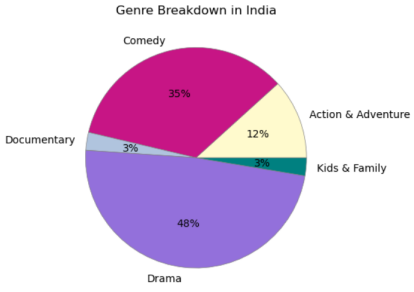
Is there a pattern in genre offerings in other countries and across different streaming services?

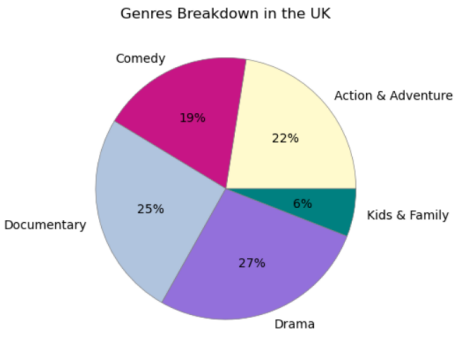
One area of interest was whether or not any pattern would emerge in the distribution of genres on streaming services in other countries. Everyone has noticed that US Netflix has different available titles in comparison to Aussie Netflix, but that is down to different individual titles, not necessarily a different catalog.

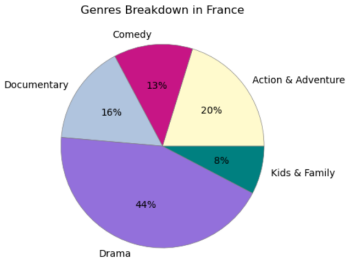
In general, I assumed the data would show a strong similarity between both what services offered in terms of genre ratios and in what each streaming service had available in different countries. For example, if Prime video had a ratio of 50% kids & family, 30% dramas, 20% action movies and tv on their platform in the US, I thought I would find a very similar if not identical ratio in Prime's UK catalog.

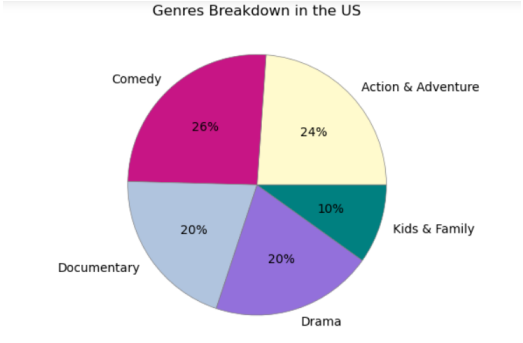
Given that Disney and Hulu are technically under the same corporate umbrella, I narrowed my focus to just Disney Plus, Prime Video, and Netflix and on the largest markets and genres by title count.

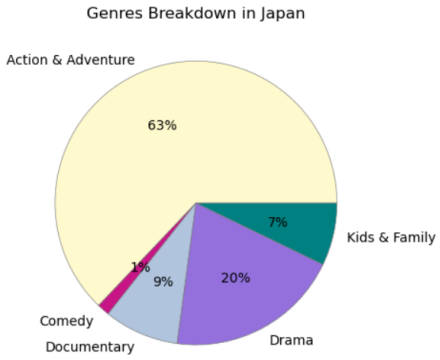
I found that Netflix wass fairly balanced across the 5 countries with the leading amount of titles, which just a few blips, such as Japan being a bit more skewed toward Action & Adventure—most anime would definitely fall under this category broadly. My pie charts represent all the 4 major streaming service's offerings by country.

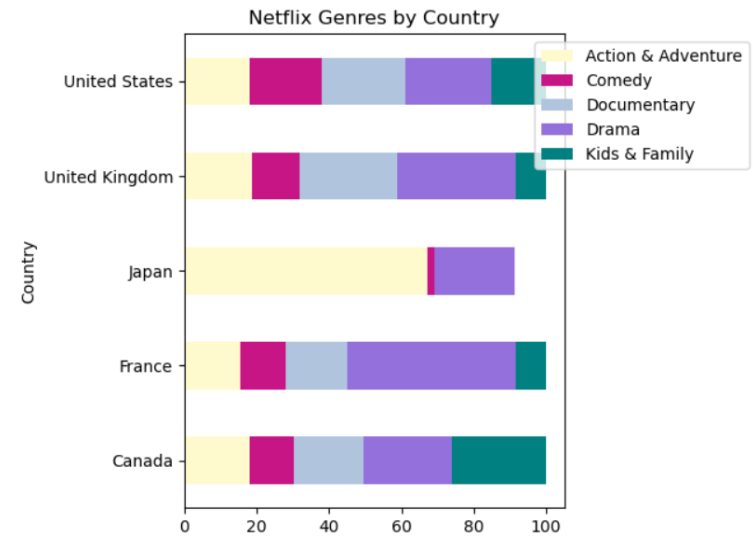


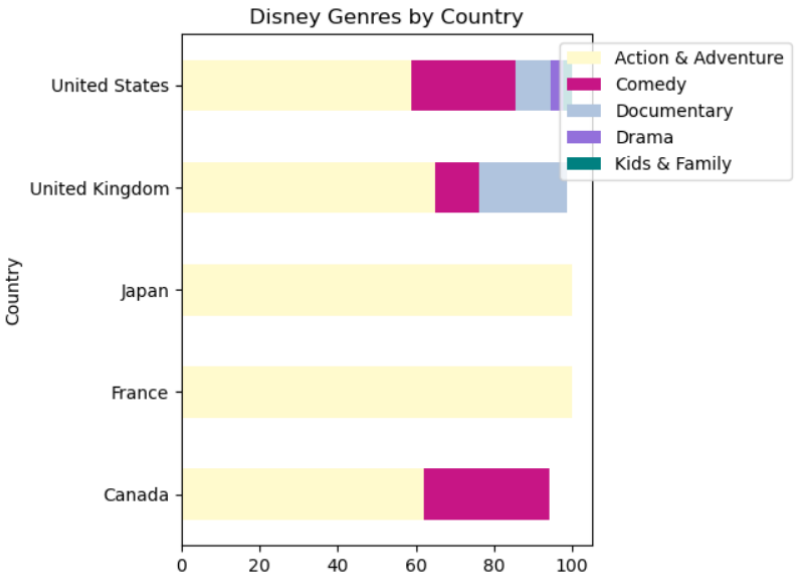


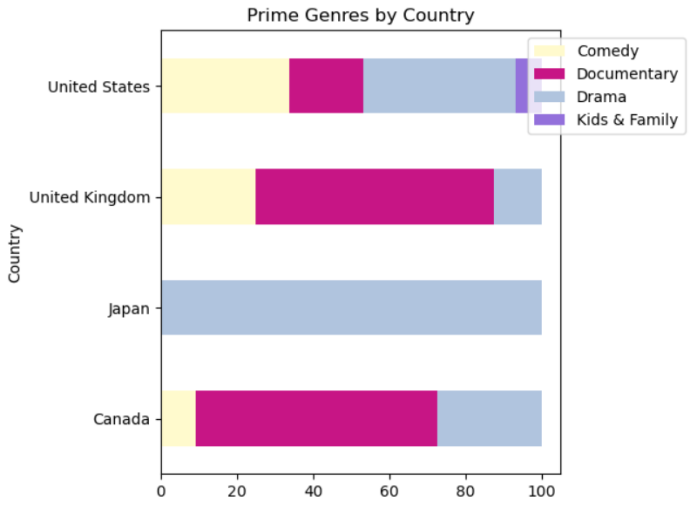




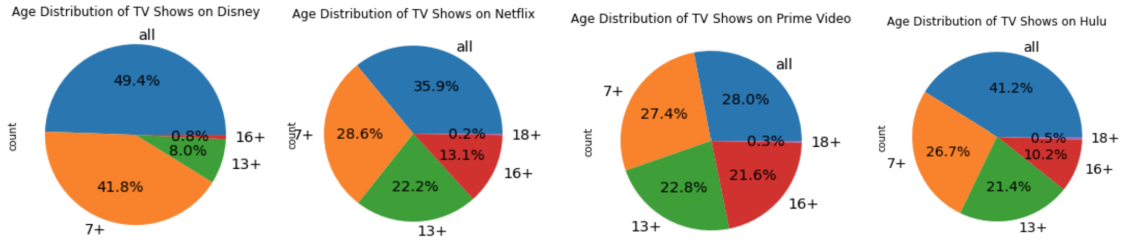



Netflix Genres by Country
United States
United Kingdom
Japan
France
Canada
Country
0
20
40
60
80
100
Action & Adventure
Comedy
Documentary
Drama
Kids & Family


Disney Genres by Country
United States
United Kingdom
Japan
France
Canada
Country
0
20
40
60
80
100
Action & Adventure
Comedy
Documentary
Drama
Kids & Family


Prime Genres by Country
United States
United Kingdom
Japan
Canada
Country
0
20
40
60
80
100
Comedy
Documentary
Drama
Kids & Family

Does the data reflect a connection between the amount of content for a specific age and the streaming platform's overall identity?



Looking at our analysis data, we can compare the charts between Netflix, Hulu, Prime Video, and Disney+. When we look at the overall distribution of TV shows on Netflix, Hulu, and Prime Video, we can see a fairly even spread of age groups with similar division between the 3 platforms, but when we look at Disney+, we see an overwhelming amount of content geared towards ages 0-13 years old. This trend continues as we take a look into Movies on each platform with Netflix, Hulu, and Prime Video having a non-insignificant percent of their content geared towards each age group meanwhile an overwhelming number of Disney+ movies are for children under 13. This is a clear correlation showing that the content on a streaming platform does contribute to its overall brand Identity at least according to the age demographic.

Is the content available on each platform reflective of audience preferences/enjoyment?

At the broadest level, to determine both a unique brand identity that a SVOD service might leverage and evaluate to what extent its content library reflects its audience's interests, it was pertinent to assess the alignment (or lack thereof) between the genres available and those that are amongst the highest-rated titles on each platform. Due to time and compilation constraints, the following analysis was only undertaken for Prime Video, Netflix, and Disney+, separated by movies and shows. Additionally, there is one limitation to the following analysis that ought to be noted before proceeding. The datasets from which the majority of the graphics referencing high and/or low ratings were constructed were compiled from the score information available on the TMDB API. Therefore, it is restricted by the titles which appear on that API, which seems to be quite robust, and the opinions of the API's users.

With regards to movies, certain genres- Drama, Comedy, and Action & Adventure- were common to both Prime Video and Netflix in both respects- most content in those categories total and most content in those categories amongst highly-rated titles (TMDB Score of 7+). Similarly, Disney+ appeared to have the monopoly on “Family” and “Animation”. This shifted a bit when observing the same data for shows, however, this might be simply due to the fact that the show genres across all platforms are significantly more niche than compared to those of the movies. Netflix and Disney+’s could still largely fit into the broader genres mentioned above, but an interesting find for Prime Video was that they actually had enough content for children that it not only warranted its own category but also came second in the genre rankings amongst highly-rated titles.

The significance of foregoing is that one can assume that each service is largely aware as to what pieces of content- by genre and likely by MPAA rating- tend to do well on its platform and are acquiring/producing pieces of content that reflect this evolving understanding. These findings were largely in line with our pre-existing, basic notions of what each brand had to offer i.e. Netflix and Prime Video likely have content that appeals to a more mature audience when compared to Disney+ as its brand is largely rooted targeting a younger, more family-oriented demographic.

Does providing “International” content serve as a significant competitive advantage in the U.S. market?

There was one under genre that dominated, specifically when examining the Netflix data, which has yet to be mentioned- “International”. It was the first ranked for both movies and tv, and necessitated its own discussion as it is a potential competitive advantage for the service. As we noted above, Prime Video and Netflix have similar content libraries in terms of genres offered, however, Netflix offers a significantly higher number of titles that are either produced outside of the United States and/or produced in a language that is not English. This led to us examining the ratio of popular movies (TMDB score of 7+) to all movies produced in each country in the Netflix library to determine a sort of return on investment. Shockingly, for both movies and shows, the U.S. had the lowest return; in other words, of all the highly-rated content on Netflix, the smallest amount was produced in the U.S. The countries with the highest returns for movies were France, Canada, and Hong Kong; similarly, for shows, the U.K. and France were tied followed by Spain and South Korea. This shows not only the demand for international content on Netflix specifically, but also reflects the growing phenomenon of globalization and the influence of all cultures on one another rather than simply Western/U.S. culture on the rest of the world. This prompted another question- have the other SVOD services identified this valuable niche and are they also working towards expanding their libraries. Disney+’s international content was negligible, but Prime Video may be catching on and may soon render Netflix’s competitive advantage in this area null. Prime Video, particularly its shows (we couldn’t source enough data

for its movies), certainly currently has less content categorized as “International”, however, assuming it tracks similar metrics, it’s worth noting that it also experiences the lowest return for content produced in the U.S. The countries with the highest returns were Spain, India, and Canada. Once again, this confirms that there is a demand, potentially growing, for content from abroad, and it is a lucrative avenue that SVOD services can pursue as they seek to differentiate themselves in an increasingly saturated market.

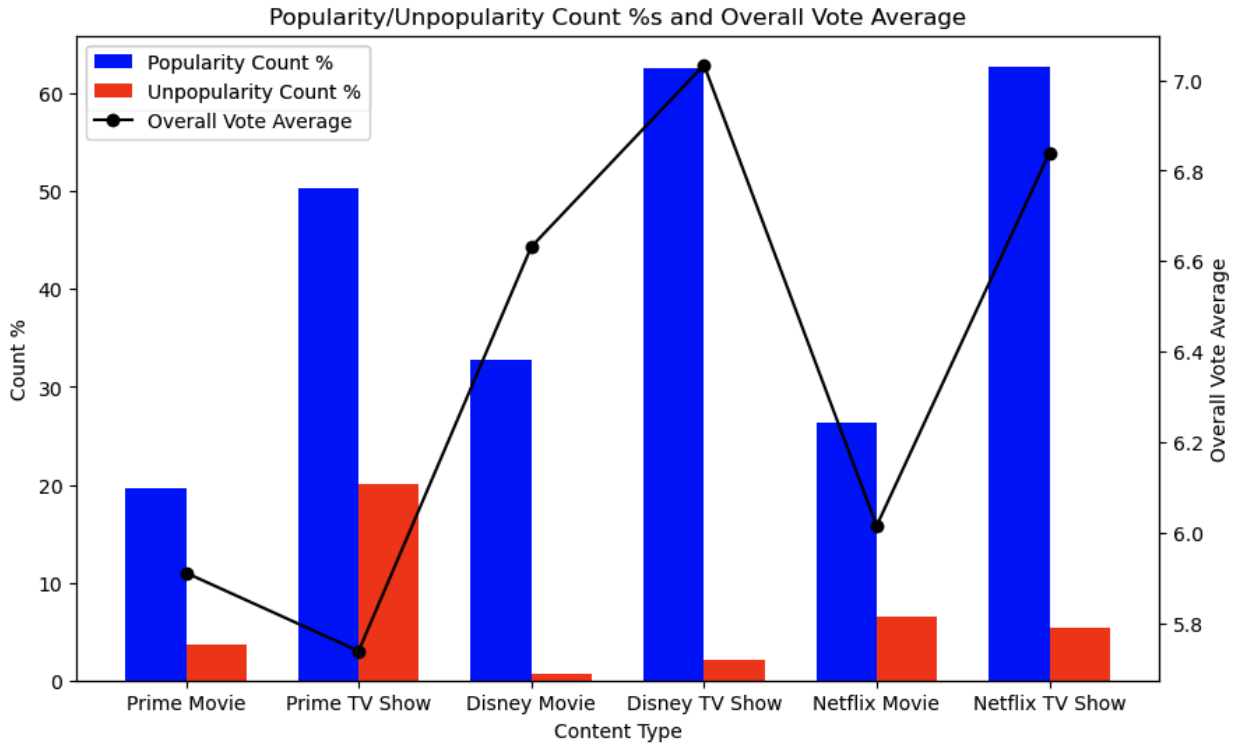
Are SVOD Services Effective in Identifying and Prolonging the Life of Successful Shows ?

To further tie together this concept of whether SVOD services are accurate in interpreting the ever-evolving trends of audience preferences in real-time, we elected to explore how effective platforms were at extending the runtime of popular shows. In addition to the limitations mentioned above, another consideration here is that as the runtime of shows grows longer, the individuals who are returning to rate each consecutive season might be those who continue to be fans until they experience significant upset at the turn of the series.The chart detailing the runtimes of highly-rated shows reveals that all SVOD platforms, albeit to varying extents, are not quite as successful as identifying or transitioning a first season’s popularity. Of the three, it is clear that Disney+ actually appears to have the “highest” success rate, and we believe this relates back to the platform generally experiencing lower turnover in its content compared to its competitors. In other words, one could say that Disney+ is stays true to its brand and is perhaps slightly more effective at attracting its preferred demographic segment. There is a general consensus, although no confirmation has ever been given by Netflix or any other platform, that SVOD services typically measure the popularity of a show within a first week before determining whether or not extend the show; if that is true, this chart perhaps emphasizes that they should be tracking audience sentiment towards the show for a longer time before making a decision for better or for worse. There are many reasons why a show might perform well in the first week when it may not deserve such high praise and equally as many reasons why a show might perform poorly when it deserves significantly higher consideration. This information, tracking a show’s reception over longer, more closely spaced time points, might prove to be an extremely valuable resource for all the platforms. The overarching significance of this is that this may play an unexpected role in brand identity as consumers may eventually begin to associate particular platforms with not being great at assessing their interests.

What is the overall quality of content on the streaming service based on viewer ratings?

The overall quality of content on various streaming services can be inferred from the “Overall Vote Average” obtained from The Movie Database dataset. The Movie Database’s vote rating scale of 1 to 10 (10 being the highest and 1 the lowest). For the analysis, we consider vote ratings <= 3 as an unpopular rating, while a score >=7 a popular one.

When using a combined chart (bars and line) to visualize the data, we noted that vote averages range from a high of 7.034 for Disney TV Shows and a low of 5.738 - again where a higher number indicates better viewer ratings.



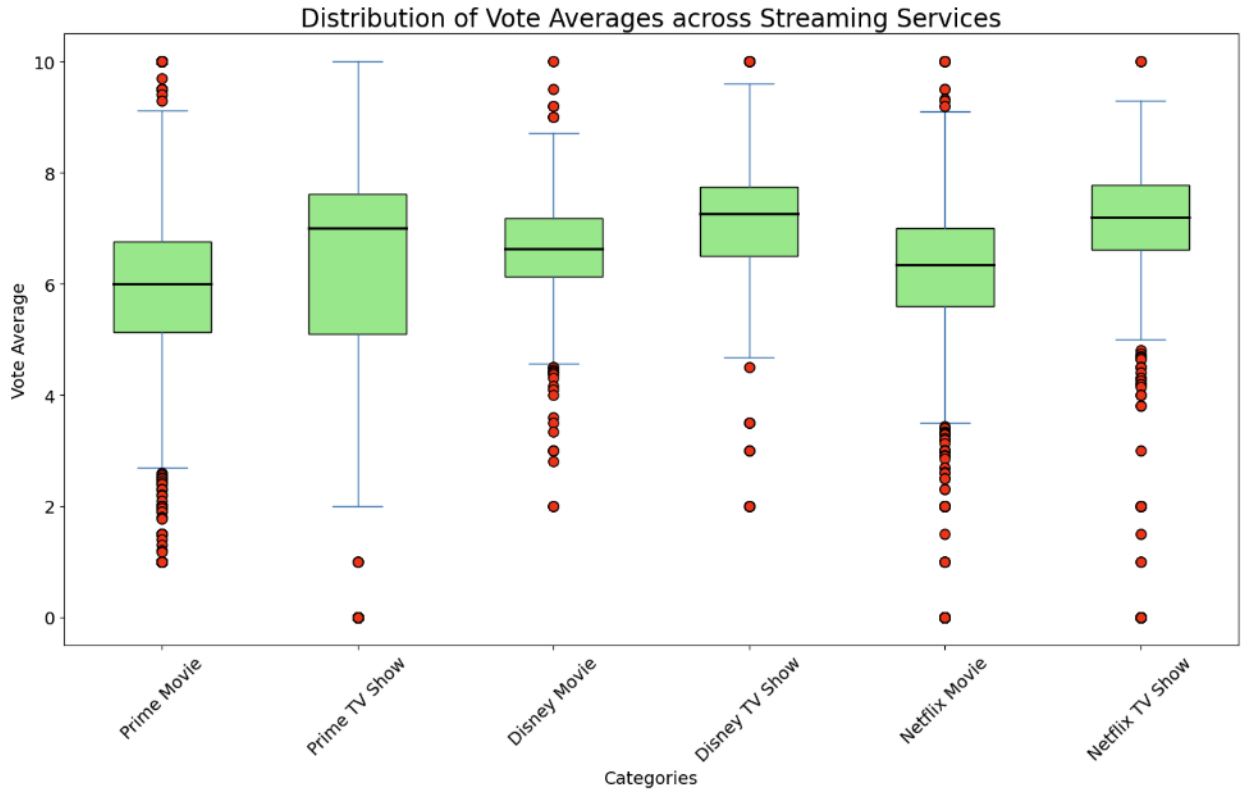
	Type	Popular Count	Unpopular Count	Total Count	Popularity Count %	Unpopularity Count %	Overall Vote Average
0	Prime Movie	1292	239	6571	19.66	3.64	5.911152
1	Prime TV Show	275	110	547	50.27	20.11	5.738009
2	Disney Movie	339	7	1033	32.82	0.68	6.631720
3	Disney TV Show	234	8	374	62.57	2.14	7.033741
4	Netflix Movie	995	244	3775	26.36	6.46	6.015155
5	Netflix TV Show	543	46	865	62.77	5.32	6.838901

We examine the “Popularity Count %”, which represents the percentage of popular content, Netflix TV Shows lead with 62.77%, closely followed by Disney TV Shows at 62.57%. This suggests that a high rate of Netflix and Disney viewers mark their respective TV show content as popular.

Conversely, the “Unpopularity Count %” shows Disney Movies with a very low percentage of 0.68%, indicating that Disney viewers mark a very small proportion of the movie content as unpopular.

In summary of the combined chart visualization, based on viewer ratings, Disney TV Shows appear to offer the highest quality content among the categories (again, followed closely by Netflix TV Shows). Then, Prime Movies and TV Shows have the lowest overall vote average which suggest a comparatively lower viewer quality rating.

When using box plots to visualize the data, it represents the three streaming services’ distribution of vote averages for movies and TV shows. From the median line, we can see the central tendency of viewer ratings across categories. For example, all three streaming services’ TV show content tend to have higher median ratings compared to their movie counterparts, which suggest viewers have a higher quality perception of TV Show content.



	Category	LowerQ	UpperQ	IQR	LowBnd	UpBnd	Min	Max	Median	Std Dev	LB Outlier Count	UB Outlier Count	Total Count	LB Outlier %	UB Outlier %
0	Prime Movie	5.134	6.7670	1.6330	2.68450	9.21650	1.0	10.0	6.0000	1.401284	167	108	6571	2.541470	1.643585
1	Prime TV Show	5.100	7.6225	2.5225	1.31625	11.40625	0.0	10.0	7.0000	2.961591	99	0	547	18.098720	0.000000
2	Disney Movie	6.134	7.1880	1.0540	4.55300	8.76900	2.0	10.0	6.6310	0.889080	21	9	1033	2.032914	0.871249
3	Disney TV Show	6.500	7.7490	1.2490	4.62650	9.62250	2.0	10.0	7.2615	1.283308	12	8	374	3.208556	2.139037
4	Netflix Movie	5.600	7.0000	1.4000	3.50000	9.10000	0.0	10.0	6.3430	1.737335	257	20	3775	6.807947	0.529801
5	Netflix TV Show	6.621	7.7810	1.1600	4.88100	9.52100	0.0	10.0	7.2000	1.738837	69	3	865	7.976879	0.346821

Next, the presence of outliers indicates content that falls outside the typical vote average range. For example, Prime and Netflix Movies have a noticeable number of low bound and upper bound outliers, indicating a significant variation about how viewers rate movie content.

Referencing the statistical summary, it complements the findings by providing specific statistical results. For example, Prime TV shows have a median vote average of 7.0, which is quite high, where its standard deviation of 2.96 illustrates the ratings spread around the median. In comparison, Disney Movies' lower standard deviation of 0.89 indicates that viewer ratings are more consistently clustered around the median.

Referencing the outlier percentages, we gain insight into the proportion of content that is rated significantly lower or higher than the average. A notable example is that Prime TV Shows do not have any upper bound outliers, which suggest a consistent quality of said content. In contrast to Netflix Movies, which has a higher percentage of lower bound outliers of 6.81% (in the context of the lower bound count), indicating a more varied quality with some content rated much lower than average.

In summary of the box plot visualization, while there are variations within Prime, Disney, and Netflix movies and TV shows content, we can infer that Disney and Netflix TV Shows tend to have a more favorable viewer reception. In contrast, Netflix and Prime Movies indicate a wider range of viewer ratings, which could suggest more content quality diversity.

References

Stoll, J. (2023, October 2). *SVOD service viewership shares in the U.S. 2023*. Statista. https://www.statista.com/statistics/1412763/svod-service-viewership-share-us/#:~:text=In%20the%20first%20half%20of,playbacks%20through%20the%20Reelgood%20app.

Links to Datasets

Shivam Bansal. *Netflix Movies and TV Shows*. Kaggle. https://www.kaggle.com/datasets/shivamb/netflix-shows/

Shivam Bansal. *Hulu Movies and TV Shows*. Kaggle. https://www.kaggle.com/datasets/shivamb/hulu-movies-and-tv-shows

Shivam Bansal. *Disney+ Movies and TV Shows*. Kaggle. https://www.kaggle.com/datasets/shivamb/disney-movies-and-tv-shows

Shivam Bansal. *Amazon Prime Movies and TV Shows*. Kaggle. https://www.kaggle.com/datasets/shivamb/amazon-prime-movies-and-tv-shows

Coding Resources

In the file named **"Data_Compilation"**:

1. The "pd.to_numeric", which changes the values in the "Vote Average" column to floats was sourced from a Stack Overflow board, which can be found here: https://stackoverflow.com/questions/15891038/change-column-type-in-pandas.
    a. It was used in Cells 5, 9, 14, 19, 22, 25, 30, 35.
2. The ".apply (lambda x:str(x).split(",")).tolist()" and the "...DF.stack()", which unnest the values in the "Genre" and/or "Country" columns respectively, were sourced from Kaggle user who demonstrated an extensive dataset cleanup. It can be found here: https://www.kaggle.com/code/janiket3/netflix-data-cleaning-visualization-recommendation#Graphical-analysis.
    a. It was used in Cells 6, 10, 15, 16, 20, 21, 23, 24, 26, 27, 31, 32, 36, 37.

In the file named **"Streaming Service Analysis"**:

1. The "plt.tight_layout()", which ensured the axis labels wouldn't be cut off when the figure was saved locally, was sourced from docs.kanaries.net, which can be found here: https://docs.kanaries.net/topics/Matplotlib/matplotlib-savefig-cuts-off-labels.
    a. It was used to produce all of the charts within the first 18 cells in the file.

2. Under the “Question To Be Answered: What is the overall quality of content on the streaming service based on viewer ratings (Vote Average)?”
    a. For Prime, Disney, and Netflix sections, the “pd.concat()” method was used to stack the movies and TV shows DataFrames for each of the streaming services (i.e., append the rows of the TV Shows DataFrame under the movies DataFrame)

       Code reference:
       https://www.geeksforgeeks.org/how-to-stack-multiple-pandas-dataframes/

    b. Under the “Box Plot Charts - All Content” section: the “medianprops”, “boxprops” and “ax = df[].plot()” were used to prepare the box plots in one figure, including chart colors and styles

       Code reference:
       - medianprops: https://matplotlib.org/stable/gallery/statistics/boxplot.html
       - boxprops: https://matplotlib.org/stable/gallery/statistics/boxplot.html
       - Box Plots in same figure: https://www.tutorialspoint.com/plot-multiple-boxplots-in-one-graph-in-pandas-or-matplotlib

3. Used code from https://matplotlib.org/stable/gallery/lines_bars_and_markers/barchart.html and https://matplotlib.org/stable/api/_as_gen/matplotlib.axes.Axes.bar.html#matplotlib.axes.Axes.bar to create a grouped bar chart
    a. Used in cell 11

4. Used code (bbox_to_anchor=()) from https://matplotlib.org/stable/gallery/pie_and_polar_charts/pie_and_donut_labels.html# to adjust location of legends
    a. Used in cells 22, 32, 42, 52