Streaming Platform Analysis

Kimberly B Valentine

Western Governors University

Table of Contents

[A. Project Highlights 4](#_Toc1868160802)

[B. Project Execution 5](#_Toc2011082291)

[C. Data Collection Process 5](#_Toc1191971028)

[C.1 Advantages and Limitations of Data Set 6](#_Toc10980602)

[D. Data Extraction and Preparation 6](#_Toc512824111)

[E. Data Analysis Process 6](#_Toc1571068809)

[E.1 Data Analysis Methods 6](#_Toc1409685802)

[E.2 Advantages and Limitations of Tools and Techniques 7](#_Toc737998485)

[E.3 Application of Analytical Methods 7](#_Toc825242289)

[F Data Analysis Results 7](#_Toc1939531306)

[F.1 Statistical Significance 7](#_Toc109772385)

[F.2 Practical Significance 8](#_Toc1419440648)

[F.3 Overall Success 8](#_Toc1210557849)

[G. Conclusion 9](#_Toc1495356842)

[G.1 Summary of Conclusions 9](#_Toc613065263)

[G.2 Effective Storytelling 9](#_Toc16322062)

[G.3 Recommended Courses of Action 9](#_Toc660897025)

[H Panopto Presentation 9](#_Toc2048387211)

[References 12](#_Toc1410715514)

[Appendix A 12](#_Toc1695981861)

[Title of Appendix 13](#_Toc1655579738)

[Appendix B 13](#_Toc1686431257)

[Title of Appendix 14](#_Toc334098601)

[Appendix C 14](#_Toc1371871733)

[Title of Appendix 15](#_Toc38061623)

[Appendix D 15](#_Toc724305577)

[Title of Appendix 16](#_Toc1812292947)

# A. Project Highlights

**Research Question:**

With so many streaming platforms to choose from, it can be overwhelming for consumers to pick the one that offers the best value. I set out to answer a key question: Which streaming platform, or combination of platforms, provides the best experience for viewers? To find the answer, I focused on factors like high-rated content (IMDb ≥ 7.5), exclusive shows and movies, popular titles, and top-performing genres such as Action, Drama, and Documentary. My goal was to help viewers make more informed decisions by prioritizing content quality and variety.

**Scope of Project:**

I analyzed five major streaming platforms: HBO Max, Hulu, Netflix, Amazon Prime, and Apple TV+. The evaluation covered:

* **Content Quality**: Focusing on titles rated 7.5 or higher on IMDb.
* **Exclusive Content**: Highlighting the shows and movies that are unique to each platform.
* **Popular Titles**: Identifying the top 25% of titles based on IMDb ratings and vote counts.
* **Genre Availability**: Analyzing content across specific genres like Action, Drama, and Documentary.

**What Wasn’t Included:**

* Subscription costs or pricing plans.
* Content availability outside of the U.S.
* Real-time updates to each platform’s library.
* Usability factors such as platform interface or ease of use.

**Overview of Solution:**

* I took a structured approach using the Waterfall Methodology, breaking the project down into clear, step-by-step phases. After gathering data from sources like Kaggle and the IMDb API, I applied statistical methods such as ANOVA tests, Kruskal-Wallis tests, and descriptive statistics to analyze the data.
* For the analysis, I used Python (in Visual Studio Code) along with libraries like Pandas for data management, Scipy for statistical tests, and Matplotlib/Seaborn for visualizations such as bar charts, heatmaps, and pie charts.
* The result was a detailed set of reports and visualizations that clearly show each platform's strengths and weaknesses. My aim was to make it easier for viewers to decide which streaming platform best suits their preferences, particularly when it comes to content quality and variety.

# B. Project Execution

**Project Plan (Goal 1: Data Preparation and Goal 2: Platform Evaluation)**  
The main objectives from Task 2 were mostly on track, but I ran into a bit of a delay with the data cleaning process. There were some issues with missing data and formatting inconsistencies between datasets, which meant the cleaning and merging took longer than I initially planned. This extra time spent on data cleaning extended that phase by a couple of days, which in turn delayed the start of the statistical analysis.  
**Difference:** The data cleaning phase took an additional two days, which meant I had to start the analysis later than originally planned. However, once the data was cleaned, the rest of the project progressed smoothly.

**Project Planning Methodology (Waterfall Approach)**  
The Waterfall methodology worked as intended. I followed a linear, step-by-step approach, with each phase building on the one before it. There weren’t any significant surprises or changes in how I executed the project.  
**No Changes:** I completed the five phases—Requirements, Design, Implementation, Verification, and Maintenance—in order. However, I did overlap some tasks from the Implementation phase with Verification to make up for the time lost during the data cleaning process.

**Project Timeline and Milestones**  
As mentioned, the delay in data cleaning had a small impact on the overall timeline, but everything else ran pretty close to plan. Here's an overview of how things went:

* **Milestone 1: Data Collection and Cleaning** – Took 6 days (12/17/24 to 12/23/24), instead of the planned 4 days.
* **Milestone 2: Statistical Analysis** – Completed on time (12/23/24 to 12/25/24), though I spent more time reviewing the data due to the extra work done during the cleaning phase.
* **Milestone 3: Visualization Creation** – Completed within the planned 2-day window (12/23/24 to 12/25/24).
* **Milestone 4: Report Creation** – Started on time, but I extended it by 1 day to ensure thoroughness.
* **Milestone 5: Review and Finalization** – Completed by the end of the day on 12/31/24, as planned.

Despite the data cleaning delay, I adjusted the schedule to ensure that everything was still completed on time.

**Conclusion on Execution vs. Plan**  
Overall, the project stayed largely on track with a few minor adjustments. The data cleaning phase took longer than expected, but I was able to adapt without impacting the final delivery. The project is set to be completed by 12/31/24, with all objectives met and deliverables on schedule.

# C. Data Collection Process

**Data Selection and Collection Process**  
The data collection process went mostly as planned, with a few adjustments along the way to keep things manageable and meaningful for the analysis. Here’s an overview of how I approached it:

* **Streaming Platform Data from Kaggle:**  
  I started by collecting datasets for the five streaming platforms—Hulu, Apple TV+, Amazon Prime, Netflix, and HBO Max—from Kaggle. Each platform's dataset was downloaded as a .csv file, following my original plan.
* **IMDb Data:**  
  Initially, I had planned to use the entire IMDb catalog from Kaggle, but I ran into a couple of issues. The dataset was too large, which ended up crashing my system, and it was more data than I could handle effectively. So, I switched to two more manageable datasets from IMDb’s developer section: title.basics.tsv.gz and title.ratings.tsv.gz. These files contained key information, like titles, genres, ratings, and votes, in a more accessible format (gzipped, tab-separated values with UTF-8 encoding).
* **IMDb Top-Rated Titles Dataset (Kaggle):**  
  To focus on the most popular titles, I used the "IMDb Top Rated Titles: Movies and TV Series" dataset from Kaggle. This dataset includes over 6,000 titles (both movies and TV series) with a minimum IMDb rating of 7 and over 10,000 votes, making it perfect for identifying top-rated content.

**Handling Obstacles During Data Collection**  
While the collection process was mostly smooth, there were a few obstacles I had to address. Here's how I dealt with them:

* **Missing IMDb ID and Rating Information:**  
  Some entries in the IMDb data had missing IMDb IDs and ratings. For the missing IDs, I dropped the affected rows. For the missing ratings, I tried to fill in the gaps with additional info from the IMDb dataset. However, in cases where this wasn’t possible, I chose to drop those rows entirely.
* **Additional Metadata:**  
  I originally planned to stick with just the IMDb data, but I realized adding a column with country availability (for where content is available) would enrich the analysis. I narrowed this down to just the U.S. to keep things focused.
* **Merging Platform Data:**  
  I merged the streaming platform datasets and ensured that I kept track of where each title was available. I used one-hot encoding to indicate whether a title was available on a platform (creating columns for each platform with a 1 if the title was available).
* **Cleaning Outliers:**  
  During the cleaning process, I identified and removed outliers and any data anomalies that might skew the analysis. This helped ensure that only valid data remained.
* **Handling Duplicate Columns and Mislabeling:**  
  I also cleaned up any duplicate columns and renamed them for clarity. Additionally, I corrected any mislabeling in the content types (like distinguishing between movies and TV shows).
* **Genre One-Hot Encoding:**  
  The genre column in the IMDb dataset had comma-separated values, so I applied one-hot encoding to break it down into individual genre columns. This made it easier to analyze genre-specific content availability.
* **Handling Missing Data:**  
  For missing data, if I couldn’t fill the gaps with supplemental information, I chose to drop the problematic columns. Since my focus was on popular content, missing data for less popular titles didn’t have much impact on the analysis.

**Data Governance, Privacy, and Security Issues**  
Regarding data governance and ethical considerations, the datasets I used followed standard practices for privacy and security:

* **Privacy and Security:**  
  The datasets from both Kaggle and IMDb are non-commercial, meaning they don’t contain any personal or sensitive information. This addressed any privacy concerns upfront. I stored the datasets securely on my encrypted system to prevent unauthorized access.
* **Ethical Considerations:**  
  I made sure to handle the data ethically, maintaining transparency in how it was processed and ensuring that no biases were introduced during the analysis. I was careful to communicate the results clearly and avoid any misrepresentation.
* **Compliance with Regulations:**  
  I adhered to the terms of use for both Kaggle and IMDb. Since the datasets didn’t contain personal information, there weren’t any specific legal or regulatory issues to worry about.

**C.1 Advantages and Limitations of the Dataset**

* **Advantage:**  
  One of the biggest advantages of the datasets I used was the structure of the Kaggle platform data. Each streaming platform dataset was consistently formatted, which made it easy to merge them and combine the necessary information without needing extensive data cleaning or transformation. This saved time and ensured consistency across the datasets.
* **Disadvantage:**  
  A limitation of the data was that titles were only uniquely identified by their IMDb ID. This created difficulties when trying to supplement the data with additional information from other sources, like tracking whether TV shows had been canceled or were still ongoing. This limited the scope for additional analysis, such as studying the lifecycle of content or its long-term popularity..

# D. Data Extraction and Preparation

The data collection for this project closely followed the plan outlined in Task 2, and I gathered two primary datasets: one for the streaming platform data and another for IMDb data.

**Streaming Platform Data from Kaggle:**  
To assess the content available on major streaming platforms, I downloaded the respective .csv files for five platforms from Kaggle:

* Hulu
* Apple TV+
* Amazon Prime
* Netflix
* HBO Max

Each dataset provided detailed information about the platform’s catalog, including title names, genres, release years, and other key metadata. This made them ideal for analyzing content quality and availability across different platforms. This method of data collection aligned well with the plan I initially set out.

**IMDb Supplemental Data:**  
For supplemental information, I sourced the IMDb dataset from their official Non-Commercial Datasets page, ensuring the data's credibility and reliability. The dataset includes ratings, vote counts, and other metadata for titles available across various platforms. These metrics were essential for evaluating content quality and identifying the most popular titles.

**IMDb Top-Rated Dataset:**  
Rather than using the full IMDb catalog, which was too large and unwieldy, I opted for a more focused dataset: the "IMDb Top-Rated Titles: Movies and TV Series" from Kaggle. This dataset includes over 6,000 titles (movies and TV shows) with a minimum IMDb rating of 7 and over 10,000 votes, making it perfect for identifying top-rated content.

To ensure I was working with the most popular titles, I filtered this dataset further by calculating the 75th percentile for both average ratings and number of votes. I then kept only the titles that exceeded the 75th percentile in both categories, focusing on content that was both highly rated and widely viewed.

From this filtered dataset, I extracted a list of titles along with their corresponding IMDb IDs. These were crucial for identifying the "most popular" content in my analysis.

# E. Data Analysis Process

## E.1 Data Analysis Methods

**ANOVA (Analysis of Variance)**

**Description:**  
ANOVA is a statistical test that helps determine whether there are significant differences between the means of three or more groups. In this project, I used one-way ANOVA to compare the percentage of high-rated content (IMDb ≥ 7.5) and exclusive content across different streaming platforms.

**Why It Was Appropriate:**  
ANOVA is ideal for comparing multiple groups, in this case, streaming platforms, to see if there are differences in content quality. It allows for the comparison of means across all platforms simultaneously, ensuring that no pairwise comparisons are overlooked. This method was particularly useful because it can test for differences in continuous variables, such as the percentage of high-rated content, across multiple categories (i.e., the streaming platforms).

**Kruskal-Wallis Test**

**Description:**  
The Kruskal-Wallis test is a non-parametric statistical test used to compare the distributions of a variable across multiple groups. Unlike ANOVA, it doesn't assume a normal distribution. I used it to test for significant differences in the availability of popular content across the different platforms.

**Why It Was Appropriate:**  
Since we were comparing categorical groups (platforms) and only had one value (percentage) for each group, the Kruskal-Wallis test was a good fit. It is robust against violations of normality and works well when comparing rankings or distributions, making it particularly useful for understanding the availability of popular titles across platforms.

**Tukey HSD (Honest Significant Difference) Test**

**Description:**  
The Tukey HSD test is a follow-up analysis used after ANOVA if significant differences are found. It helps identify exactly which pairs of groups (platforms) differ in terms of high-rated and exclusive content availability.

**Why It Was Appropriate:**  
Once ANOVA identified significant differences, the Tukey HSD test helped pinpoint which specific platforms differed in terms of their content availability. This test is valuable because it controls for the risk of Type I errors when making multiple comparisons, ensuring reliable and precise conclusions.

## E.2 Advantages and Limitations of Tools and Techniques

**ANOVA (Analysis of Variance)**

* **Advantage**: ANOVA is a powerful method for comparing the means of multiple groups. It’s particularly useful for examining differences between multiple streaming platforms in terms of content availability. The method can process multiple groups simultaneously, reducing the need for pairwise comparisons and offering a clear insight into whether significant differences exist across the platforms.
* **Limitation**: One limitation of ANOVA is that it assumes the data follows a normal distribution, which may not always be the case with real-world data. If this assumption is violated, the results may not be as reliable. Additionally, ANOVA is sensitive to outliers, which could distort the analysis if they aren't properly managed.

**Kruskal-Wallis Test**

* **Advantage**: The Kruskal-Wallis test is non-parametric, meaning it doesn’t rely on the assumption of normal distribution, which makes it well-suited for datasets with skewed distributions. It is ideal for comparing groups like streaming platforms where the data is not always normally distributed, ensuring more flexibility in analysis.
* **Limitation**: A limitation of the Kruskal-Wallis test is that it only tells you whether there are differences between the groups but doesn’t pinpoint which specific groups differ. While it's useful for detecting differences, follow-up tests like the Dunn's test are needed for more detailed comparisons, adding to the complexity.

**Tukey HSD (Honest Significant Difference) Test**

* **Advantage**: The Tukey HSD test is a robust post-hoc test that helps identify exactly which pairs of platforms differ after finding significant results with ANOVA. It controls for the risk of Type I errors when making multiple comparisons, which ensures that the conclusions are reliable even when testing many groups.
* **Limitation**: A limitation of the Tukey HSD test is that it assumes equal sample sizes across groups. If the sizes are unequal, the test may not be as accurate, potentially leading to misleading conclusions if not accounted for properly.

## E.3 Application of Analytical Methods

**1. ANOVA to Compare High-Rated Content Across Platforms**

**Implementation Steps:**

1. **Reshape the Data**: To compare platforms easily, I first converted the data from a wide format (where each platform had its own columns) into a long format, where each row represented a single platform-title pair. This made it easier to analyze the platforms side by side.
2. **Calculate High-Rated Content Percentages**: I then calculated the percentage of high-rated content (IMDb ratings of 7.5 or higher) for each platform. This was done by filtering out titles that met the rating requirement and then calculating the ratio of those titles to the total number of titles available on each platform.
3. **Run ANOVA**: I performed a One-Way ANOVA test to compare the percentage of high-rated content across platforms. This test helped determine if one platform stood out in terms of offering high-rated content compared to the others.
4. **Post-hoc Analysis (Tukey's HSD Test)**: If the ANOVA test indicated significant differences, I followed up with Tukey’s HSD test. This allowed me to identify exactly which platforms showed differences in their content availability.

**Requirements and Verification:**

* **ANOVA Assumptions**: I ensured the data met the assumptions of the ANOVA test by visually inspecting the distribution of the data. Since the data were categorical (platforms) and continuous (content percentages), ANOVA was an appropriate choice.
* **Tukey's HSD Verification**: After the ANOVA test showed significant differences, I applied Tukey's HSD test to pinpoint which platform pairs had the differences, ensuring a more specific understanding of the results.

**2. Exclusivity Analysis Across Platforms**

**Implementation Steps:**

1. **Label Exclusive Content**: I added a new column to the dataset to mark whether each title was exclusive to a specific platform. This was done by checking if the title was available on only one platform and marking it as exclusive.
2. **Calculate Exclusivity Percentages**: For each platform, I calculated the percentage of exclusive titles by dividing the number of exclusive titles by the total number of titles available on that platform.
3. **Run ANOVA**: Like with high-rated content, I used ANOVA to compare the percentage of exclusive content across the platforms. This helped identify if any platform offered a notably higher percentage of exclusive titles.
4. **Post-hoc Analysis (Tukey's HSD Test)**: After finding significant differences with ANOVA, I applied Tukey's HSD test to pinpoint which specific platforms differed from each other in terms of exclusivity.

**Requirements and Verification:**

* **ANOVA Assumptions**: To ensure the reliability of the ANOVA results, I visually checked the data to confirm that it met the assumptions of the test, focusing on the variability across platforms.

**3. Kruskal-Wallis Test for Popular Content Availability**

**Implementation Steps:**

1. **Identify Popular Titles**: I identified the most popular titles on each platform by using IMDb’s rating and vote counts. I focused on titles with ratings and votes above the 75th percentile.
2. **Create a Binary Matrix**: I created a binary matrix where each title was marked as either popular or not, depending on its rating and vote count. This matrix helped me track whether each platform offered the popular titles.
3. **Run the Kruskal-Wallis Test**: Given that the data didn’t meet the assumptions for normality, I used the Kruskal-Wallis test, which works well with non-normally distributed data. This test helped determine if there were significant differences in the availability of popular content across platforms.

**Requirements and Verification:**

* **Kruskal-Wallis Assumptions**: The Kruskal-Wallis test doesn’t require the data to be normally distributed, but it does assume that the groups are independent. I made sure that each platform’s data was independent and that each title was only considered once.

**4. Genre-Specific Analysis for High-Rated Content**

**Implementation Steps:**

1. **Filter High-Rated Content by Genre**: I filtered out the high-rated content (IMDb ≥ 7.5) for each genre, so I could analyze how each platform performed in specific genres like Action, Drama, and Documentary.
2. **Calculate Genre Percentages**: For each genre, I calculated the percentage of high-rated content available on each platform. This gave me insight into how platforms compared in terms of high-rated content within specific genres.
3. **Compare Across Platforms**: I repeated this process for several genres, comparing the availability of high-rated content across the platforms in each genre.

**Requirements and Verification:**

* **Data Preparation**: To ensure accurate results, I carefully cleaned the genre data, ensuring consistency and resolving any issues with missing or misnamed genres before proceeding with the analysis.

# 

# F Data Analysis Results

## F.1 Statistical Significance

**1. High-Rated Content Analysis (IMDb ≥ 7.5)**

* **Null Hypothesis (H₀):** There are no significant differences in the high-rated content percentages across platforms.
* **Statistical Test:** One-Way ANOVA
* **Metrics Generated:**
  + **F-statistic:** 5.059
  + **p-value:** 0.00045
* **Alpha Value:** 0.05 (5% significance level)
* **Conclusion:** There is sufficient evidence to reject the null hypothesis and support the claim that there are significant differences in high-rated content percentages across platforms.

**2. Exclusive High-Rated Content Analysis**

* **Null Hypothesis (H₀):** There are no significant differences in the percentage of exclusive high-rated content across platforms.
* **Statistical Test:** One-Way ANOVA
* **Metrics Generated:**
  + **F-statistic:** 242.778
  + **p-value:** 1.06e-206
* **Alpha Value:** 0.05 (5% significance level)
* **Conclusion:** There is sufficient evidence to reject the null hypothesis and support the claim that there are significant differences in the percentage of exclusive high-rated content across platforms.

**3. Popular Content Availability Analysis**

* **Null Hypothesis (H₀):** There are no significant differences in the percentage of popular content across platforms.
* **Statistical Test:** Kruskal-Wallis H-test
* **Metrics Generated:**
  + **H-statistic:** 4.00
  + **p-value:** 0.4060
* **Alpha Value:** 0.05 (5% significance level)
* **Conclusion:** There is insufficient evidence to reject the null hypothesis, suggesting no significant difference in the percentage of popular content across platforms

## F.2 Practical Significance The results of the statistical tests provide valuable insights into the availability of high-rated, exclusive, and popular content across different streaming platforms. In practical terms, these findings can guide platform selection based on content quality and exclusivity, which are crucial for users who prioritize high-quality and unique content. The following analysis details the practical significance of the results:

#### High-Rated Content Analysis (IMDb ≥ 7.5)

## The significant difference in the percentage of high-rated content across platforms (with HBO leading at 32.16% and Amazon trailing at 11.24%) demonstrates that some platforms offer far more top-rated content than others. This is practically significant because it means that users who prioritize watching high-quality content, such as critically acclaimed movies or TV shows, should gravitate towards platforms like HBO.

## Example: A client focusing on providing high-quality content for a user base that values prestige and top-tier entertainment can confidently recommend HBO as the best platform for its high-rated content, helping them attract discerning viewers.

1. **Exclusive High-Rated Content Analysis**  
   The **significant differences in exclusive high-rated content percentages** (with HBO at 29.31%) also highlight the importance of exclusive titles in a platform's catalog. Exclusive content often drives user loyalty, as it offers something that cannot be found elsewhere. HBO's higher percentage of exclusive high-rated content means it stands out in terms of offering unique and quality content that cannot be easily accessed on competing platforms.  
     
   **Example:** A client working in content marketing or platform partnership could use this insight to target potential subscribers who are looking for exclusive access to top-rated shows or movies. HBO's dominance in this area would make it the ideal recommendation for users seeking exclusive entertainment.
2. **Popular Content Availability Analysis**  
   While the Kruskal-Wallis H-test showed **no significant differences** in the percentage of popular content across platforms, the **actual percentages** still provide a meaningful comparison. HBO, with 17.97% of its catalog consisting of popular content, ranks higher than the others in terms of widely recognized titles. While the differences weren't statistically significant, the practical application of this information could guide a platform's marketing strategy in appealing to users who are interested in trending or widely recognized content.

**Example:** A client targeting casual viewers or those interested in mainstream titles can use this information to recommend platforms like HBO, which offers a higher percentage of popular content compared to others. This could be especially useful for recommending platforms to clients who prioritize watching well-known titles or following current trends.

#### Conclusion

The practical significance of these results is clear:

* **HBO** stands out as the best option for users prioritizing high-rated and exclusive content.
* The insights into popular content availability, though not statistically significant, still provide value for clients seeking to target users with mainstream tastes.

## F.3 Overall Success

The project successfully addressed the core objectives of analyzing streaming platforms based on their availability of high-rated, exclusive, and popular content. The statistical tests conducted—ANOVA for high-rated and exclusive content, and Kruskal-Wallis H-test for popular content—provided clear insights into the strengths and weaknesses of each platform, supporting the hypothesis that there are significant differences in content availability across platforms.

# 

# G. Conclusion

## G.1 Summary of Conclusions

Based on the comprehensive analysis of streaming platforms, the following conclusions have been drawn:

**Percentage of High-Rated Content (IMDb ≥ 7.5) per Platform**

* **HBO**: **32.16%** - The highest percentage of high-rated titles, showcasing its focus on quality content.
* **Hulu**: **26.54%** - Second in high-rated content, with a strong showing across certain genres.
* **Netflix**: **19.59%** - A moderate share of high-rated content but bolstered by a high volume of exclusives.
* **Apple**: **16.92%** - Offers fewer high-rated titles, focused on niche genres like Documentary and News.
* **Amazon**: **11.24%** - The lowest percentage of high-rated titles among platforms analyzed.

**2. Percentage of Exclusive Content That is High Rated**

* **HBO**: **31.71%** - Maintains its leadership in delivering high-rated exclusive content.
* **Hulu**: **24.99%** - A strong contender for high-rated exclusives, especially in Talk-Show and War genres.
* **Apple**: **16.59%** - Solid exclusivity for its high-rated niche content but falls behind HBO and Hulu.
* **Amazon**: **10.68%** - Struggles with high-rated exclusives, as most of its content is less critically acclaimed.

**3. Availability of Popular Titles (IMDb ≥ 7.5 and >10k votes)**

* **HBO**: **94 titles (17.97%)** - Leads in popular content, with a diverse range of top-performing titles.
* **Netflix**: **87 titles (16.63%)** - A close second, driven by its broad catalog of exclusives.
* **Apple**: **73 titles (13.96%)** - Focused on niche genres but slightly ahead of Amazon in popular content.
* **Hulu**: **73 titles (13.96%)** - Matches Apple in popular titles, excelling in specific genres.
* **Amazon**: **71 titles (13.58%)** - The least amount of popular titles, consistent with its overall lower ratings.

**4. Overlap of Popular Titles Across Platforms (as a % of each platform's popular titles)**

This heatmap shows where popular titles are shared between platforms:

* **Amazon**: Shares the most popular titles with Apple (**12.50%**) and Hulu (**7.46%**). Least overlap with Netflix and HBO.
* **Hulu**: Shares 8.84% of its popular titles with Netflix and 7.35% with Apple, showing moderate cross-platform availability.
* **Netflix**: Has a low overlap with other platforms, emphasizing its focus on exclusivity.
* **Apple**: Significant overlap with Amazon (**12.50%**) and Hulu (**7.35%**), indicating less exclusive popular content.
* **HBO**: Minimal overlap with other platforms, reinforcing its reputation for unique, high-quality content.

These insights provide a clear picture of how the platforms compare in terms of quality, exclusivity, and popular content availability, with HBO consistently leading in both high-rated and popular content, while other platforms cater to specific niches or prioritize exclusivity over quality.

## G.2 Effective Storytelling

The visuals used in this project helped make the analysis more understandable and engaging. Here’s how they supported the story:

* **Bar and Pie Charts**: These helped show the percentage of high-rated content across platforms and highlight the differences between them in terms of content quality. It was a quick way to compare the platforms at a glance. A graph of different colored bars

  Description automatically generated

A graph of different colored bars

Description automatically generated

A pie chart with different colored circles

Description automatically generated

**Heatmap of Platform Overlap**: This heatmap visualized the overlap of popular titles (those rated ≥ 7.5 with over 10k votes) across platforms. It clearly showed which platforms share the most popular content, helping to identify where there are gaps or redundancies in a platform's library.A chart with numbers and a number of squares

Description automatically generated with medium confidence

## 

Using tools like **Matplotlib**, **Seaborn**, and **Power BI** made these visuals easy to understand, allowing the audience to quickly grasp the key points from the data.

## G.3 Recommended Courses of Action

Based on everything we found, here are two recommendations:

**Best Single Platform: HBO**

HBO emerges as the best single streaming platform for consumers seeking high-quality content. Here's why:

1. **Highest Percentage of High-Rated Content:** 32.16% of its library is rated 7.5 or higher on IMDb, the best among all platforms.
2. **Strong Genre Leadership:** HBO dominates in 17 out of 25 genres, including Drama, History, Action, and Music, ensuring a well-rounded selection.
3. **Top in Exclusive High-Rated Content:** With 31.71% of its exclusives being high-rated, HBO delivers a unique, quality-driven library.
4. **Most Popular Titles:** 94 titles rated 7.5+ with 10k+ votes, making HBO a clear leader in critically acclaimed content.

For a consumer seeking a single, versatile platform, HBO offers the best combination of quality, exclusivity, and genre diversity.

**Best Pair of Platforms: HBO + Netflix**

For consumers wanting both depth and breadth, a combination of HBO and Netflix provides the most comprehensive library:

1. **HBO's Quality and Genre Leadership:** HBO’s strength in high-rated and exclusive content anchors the combination. It ensures access to top-rated titles and genre dominance.
2. **Netflix’s Quantity and Exclusivity:** Netflix complements HBO with a vast library of content, boasting 94.16% exclusivity and excelling in genres like Animation, Sci-Fi, and Crime.
3. **Wide Popular Title Coverage:** Together, they provide access to **181 of the 523 popular titles**, covering the majority of high-rated, high-vote content.

This pairing offers the best of both worlds: HBO’s emphasis on quality and Netflix’s extensive catalog, catering to diverse consumer preferences.

# H Panopto Presentation

<https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=7da61749-2c25-4c64-9203-b25400311ae9>   
 

# References

No sources were cited.

# Appendix A

# Title of Appendix

Put any supporting material in these appendices. Add additional or delete superfluous appendices as needed.

# Appendix B

# Title of Appendix

Put any supporting material in these appendices. Add additional or delete superfluous appendices as needed.

# Appendix C

# Title of Appendix

Put any supporting material in these appendices. Add additional or delete superfluous appendices as needed.

# Appendix D

# Title of Appendix

Put any supporting material in these appendices. Add additional or delete superfluous appendices as needed.