Streaming Platform Analysis

Kimberly B Valentine

Western Governors University

Table of Contents

[A. Proposal Overview 4](#_Toc2137083979)

[A.1 Research Question or Organizational Need 4](#_Toc1815560889)

[A.2 Context and Background 4](#_Toc1384268585)

[A.3 and A3A Summary of Published Works and Their Relation to the Project 4](#_Toc2079458267)

[Review of Work 1 5](#_Toc694593955)

[Review of Work 2 5](#_Toc475195288)

[Review of Work 3 5](#_Toc1156669356)

[A.4 Summary of Data Analytics Solution 5](#_Toc374341712)

[A.5 Benefits and Support of Decision-Making Process 6](#_Toc1116633891)

[B. Data Analytics Project Plan 6](#_Toc1561262715)

[B.1 Goals, Objectives, and Deliverables 6](#_Toc1217623361)

[B.2 Scope of Project 7](#_Toc427690838)

[B.2.A Included in Project Scope 7](#_Toc1243851001)

[B.2.B Not included in Project Scope 7](#_Toc2036717847)

[B.3 Standard Methodology 7](#_Toc204762054)

[B.4 Timeline and Milestones 8](#_Toc1049840195)

[B.5 Resources and Costs 8](#_Toc313275103)

[B.6 Criteria for Success 9](#_Toc1337049307)

[C. Design of Data Analytics Solution 9](#_Toc764513894)

[C.1 Hypothesis 9](#_Toc1169566000)

[C.2 and C.2.A Analytical Method 9](#_Toc1622461191)

[C.3 Tools and Environments 10](#_Toc53928856)

[C.4 and C.4.A Methods and Metrics to Evaluate Statistical Significance 10](#_Toc1386867513)

[C.5 Practical Significance 11](#_Toc157862960)

[C.6 Visual Communication 11](#_Toc1263386613)

[D. Description of Dataset 12](#_Toc34323025)

[D.1 Source of Data 12](#_Toc1867631323)

[D.2 Appropriateness of Dataset 12](#_Toc418252330)

[D.3 Data Collection Methods 12](#_Toc1665164903)

[D.4 Observations on Quality and Completeness of Data 12](#_Toc57902398)

[D.5 and D.5.A Data Governance, Privacy, Security, Ethical, Legal, and Regulatory Compliances 13](#_Toc1649406101)

[References 14](#_Toc1322159537)

[Appendix A 14](#_Toc1046046101)

[Title of Appendix 15](#_Toc804972029)

[Appendix B 15](#_Toc128761729)

[Title of Appendix 16](#_Toc1200922991)

[Appendix C 16](#_Toc1897962072)

[Title of Appendix 17](#_Toc1862588384)

[Appendix D 17](#_Toc1292381938)

[Title of Appendix 18](#_Toc390418831)

# Proposal Overview

## A.1 Research Question or Organizational Need

Consumers are increasingly challenged when it comes to choosing the right streaming platform from the multitude of options available. This research aims to answer the pivotal question: which streaming platform, or combination of platforms, offers the best value for consumers based on high-rated content (IMDb ≥ 7.5), exclusive high-rated content, availability of popular content, and high-rated content across specific genres? The analysis will focus on platforms that excel in genres such as Action, Drama, or Documentary, while also evaluating overall content quality and genre-specific offerings to cater to diverse consumer preferences.

## A.2 Context and Background

The streaming industry has become increasingly competitive, with major platforms like HBO, Netflix, Hulu, Amazon, and Apple vying for consumer attention. Each offers a unique mix of exclusive titles and extensive licensed content libraries. However, consumers often struggle to assess the actual value of these platforms when prioritizing high-quality content. This project addresses this challenge by evaluating the significant platforms using objective metrics such as content quality (IMDb ratings), exclusivity, and popularity. In addition to evaluating high-rated content, the project will focus on the availability of high-rated content within specific genres, ensuring that the analysis reflects diverse viewing preferences and providing actionable insights for consumers seeking premium content.

## A.3 and A3A Summary of Published Works and Their Relation to the Project

### Review of Work 1: "Ditch Your Cable"

This product review by Patrick Lucas Austin evaluates popular television streaming services, including Netflix, HBO GO, Hulu, Sling, and Amazon Prime Instant Video. The review outlines the strengths of each platform, such as Netflix's extensive library and Ultra HD streaming, HBO GO's exclusive content offerings, Hulu's access to current network TV shows, Sling's live television package, and Amazon Prime's bundled services. The article emphasizes these streaming options' flexibility, convenience, and affordability as viable alternatives to traditional cable television (Austin, 2015).

**Relation to Project:**

This review supports the project's focus on analyzing streaming platforms by detailing key features influencing consumer adoption, such as content quality, flexibility, and technical innovations. Discussing platform-specific attributes provides valuable insights for comparing platforms and identifying the best overall service. It aligns with the project's methodology for evaluating popular titles and consumer preferences.

### Review of Work 2: "The Show Must Go On(line): The Impact of Content and System Quality on the Usage of Television Streaming Content Libraries"

This study looks at how content quality, system quality, price, and user habits affect the use and perception of streaming services. The authors surveyed 1,038 users in Germany and found that content quality (especially its relevance, range, and freshness) and system quality (usability and reliability) are key factors for driving usage and positive word-of-mouth. Habit also plays an important role in encouraging continued use of the platform.

**Relation to Project**

The study supports the project's focus on comparing streaming platforms based on their content libraries. It highlights the importance of having a wide range of relevant, up-to-date content, which ties into the project's goal of analyzing popular titles across different platforms. Understanding these factors helps explain why some platforms might attract more users.

### Review of Work 3: “Netflix, HBO Get Ready to Rumble.”

In the article "Netflix, HBO Get Ready to Rumble," Schneider (2019) discusses the ongoing competition between Netflix and HBO, particularly in the context of Emmy nominations. Netflix, which started small, rapidly grew and surpassed HBO in 2018 with many nominations. The article explores how Netflix's strategy of submitting many shows for nominations (41 shows in 2019) contrasts HBO's more selective approach (submitting 20 shows). The article notes that while Netflix’s sheer volume of content gives it an advantage, HBO remains competitive with its high-quality contenders like Game of Thrones and Veep. Despite Netflix’s dominance in nominations, major program wins have remained elusive, with HBO continuing to lead in key categories.

**Relation to Project**

This work directly relates to the project's focus on comparing streaming platforms based on their content libraries. The discussion of Netflix's strategy of offering a high volume of content and the importance of content quality aligns with the project's goal of analyzing popular titles across platforms. The competition between Netflix and HBO also highlights the significance of content diversity and the ability to appeal to a broad audience, which is crucial in understanding how streaming services attract users.

## A.4 Summary of Data Analytics Solution

The proposed data analytics solution will evaluate catalog data from significant streaming platforms alongside IMDb data to determine the best value for consumers based on high-rated content, exclusive titles, genre dominance, and widespread content availability. We will use statistical methods like ANOVA and Kruskal-Wallis tests to compare differences across platforms, with descriptive analytics to highlight trends in exclusivity and content quality.

The data is sourced from reputable platforms such as Kaggle.com and the IMDb API, which provide comprehensive metrics such as IMDb ratings, votes, genres, release dates, and content type. Our rigorous data cleaning steps, which address missing values, standardize formats, remove duplicates, and merge data from multiple sources, ensure the reliability and consistency of our dataset.

ANOVA will compare the percentages of high-rated and exclusive content across platforms, while the Kruskal-Wallis test will evaluate differences in popular titles. Descriptive statistics will summarize factors like exclusivity, genre-specific performance, U.S. availability, and content distribution.

Visual representations like bar charts, pie charts, and heatmaps will present these comparisons. This solution aligns with the research question using objective metrics and statistical methods to assess platform strengths. The approach is feasible with readily accessible data and tools, providing actionable insights for consumers to make informed decisions based on their content preferences and quality standards.

## A.5 Benefits and Support of Decision-Making Process

The proposed solution significantly benefits consumers by offering a data-driven evaluation of streaming platforms based on content quality, exclusivity, and popularity. The solution eliminates guesswork by objectively ranking platforms like HBO, Netflix, and others, empowering consumers to make informed subscription decisions. This analysis directly supports decision-making by identifying platforms that align with specific preferences, such as access to high-rated content, exclusive titles, and high-quality content in preferred genres. By focusing on genre-specific content, the solution enables consumers to choose platforms that best suit their tastes, enhancing the personalization of their viewing experience. These insights allow consumers to make more informed choices, ensuring they invest in platforms that offer the most value based on their content priorities.

# Data Analytics Project Plan

## B.1 Goals, Objectives, and Deliverables

Goal 1: Ensure the data is clean, reliable, and ready for analysis.

* Objective 1.1: Clean and prepare the data for analysis.
  + Deliverable 1.1.1: Clean the datasets by handling missing values, standardizing formats, and removing duplicates.
  + Deliverable 1.1.2: Merge datasets from multiple sources (IMDb, Kaggle) to ensure consistency and readiness for statistical analysis.
  + Objective 1.2: Ensure the data is correctly formatted for statistical analysis and visualization.
    - Deliverable 1.2.1: Standardize the dataset columns (e.g., content type, IMDb ratings, genre) for straightforward interpretation and analysis.
    - Deliverable 1.2.2: Ensure the data can be efficiently used in visualizations such as heatmaps, pie charts, and bar graphs

Goal 2: Evaluate the best streaming platform(s) for consumers based on high-rated content, exclusivity, access to popular content, and genre-specific availability.

* Objective 2.1: Analyze the availability of high-rated content (IMDb ≥ 7.5) across platforms.
  + Deliverable 2.1.1: Conduct an ANOVA test to compare the percentages of high-rated content on each platform.
  + Deliverable 2.1.2: Create a bar chart to represent the distribution of high-rated content across streaming platforms visually.
* Objective 2.2: Assess the availability of exclusive high-rated content across streaming platforms.
  + Deliverable 2.2.1: Use a Kruskal-Wallis test to evaluate significant differences in exclusive high-rated content across platforms.
  + Deliverable 2.2.2: Design a heatmap to show the concentration of exclusive high-rated content on each platform.
* Objective 2.3: Investigate the availability of popular content across platforms.
  + Deliverable 2.3.1: Compile a list of popular content (based on IMDb ratings, votes, or other metrics) available on each platform.
  + Deliverable 2.3.2: Create a pie chart showing the proportion of popular content available on each platform.
* Objective 2.4: Analyze high-rated content for genres.
  + Deliverable 2.4.1: Perform genre-specific analysis using descriptive statistics to highlight the number of high-rated content in various genres available on each platform.
  + Deliverable 2.4.2: Create a series of pie charts or bar graphs to compare genre-specific high-rated content availability across platforms.

Goal 3: Provide actionable insights for consumers to make informed decisions based on platform content offerings.

* Objective 3.1: Create a comparative report summarizing each platform’s performance.
  + Deliverable 3.1.1: Prepare a detailed report comparing the platforms based on high-rated content availability, exclusivity, and genre-specific offerings.
  + Deliverable 3.1.2: Summarize the report into an executive summary highlighting key findings for quick decision-making.
* Objective 3.2: Develop a recommendation system or scorecard for consumers based on content preferences.
  + Deliverable 3.2.1: Create a recommendation scorecard that ranks platforms based on the data analysis and consumer priorities (e.g., high-rated content, exclusivity, preferred genres).
  + Deliverable 3.2.2: Provide consumer guidelines to help identify the best platform or combination of platforms based on content preferences.

## B.2 Scope of Project

### B.2.A Included in Project Scope

* Evaluation of Streaming platforms. These include HBO Max, Hulu, Netflix, Amazon Prime and Apple TV+.
  + The evaluation will be based on:
    - Content Quality
    - Quality of exclusive content
    - Availability of high-quality content for specific genres
    - Access to popular titles. This will be defined by the top 25% of titles on the internet movie database regarding high vote counts and high ratings.
* Analytical methods
  + Anova tests with Tukey HSD to supplement findings
  + Chi-Square tests to help compare key metrics.
  + Descriptive statistics to summarize trends and visualize findings through bar-charts, heatmaps, and pie charts.
* Data Sources
  + Data will be sourced from Kaggle.com. The data for each platform is individually collected.
    - This includes data for Netflix, Amazon Prime, Apple TV+, Hulu and HBO-Max
  + Additional IMDb information for each title will be acquired directly from the IMDb API.

### B.2.B Not included in Project Scope

* Pricing or Cost Analysis
  + The project does not include an analysis of subscription costs or pricing strategies for each streaming platform. Its focus is solely on content quality, exclusivity, and availability.
* International Content Availability
  + This project will only focus on streaming content available in the United States, and will not evaluate platform quality in other geographic areas.
* Real Time Content Updates
  + The analysis will be based on static datasets, which may not reflect real-time updates or changes in content availability. Continuous updates and tracking over time will not be included.
* Platform Usability or Interface
  + The project does not include evaluating the streaming platforms' usability, user interface, or other non-content related features.

## B.3 Standard Methodology

I will be using the Waterfall Methodology to implement this project. The approach is linear and structured, with each phase building upon the previous one. The methodology will be broken down into five distinct phases, as described below:

**Phase 1: Requirements**

* Goal: Gather and define all project requirements.
* Steps:
  + Data Collection: Gather data from platforms (e.g., Netflix, HBO, etc.) through Kaggle and the IMDb API.
  + Define Metrics: Define metrics for evaluating high-rated content (IMDb ≥ 7.5), exclusivity, and genre-specific availability.
  + Establish Research Questions: Refine the research question to ensure clarity and focus on genre-specific content and exclusivity.

**Phase 2: Design**

* Goal: Design the analytics solution and determine the tools, methods, and techniques for the analysis.
* Steps:
  + Data Preparation Plan: Outline the steps for cleaning and merging datasets, including handling missing values, standardizing formats, and removing duplicates.
  + Analytical Approach: Decide on the statistical methods (ANOVA, Chi-Square, descriptive statistics) and ensure alignment with the project's goals.
  + Visualization Strategy: Determine the types of visualizations (bar charts, pie charts, heatmaps) to represent the findings comprehensibly.

**Phase 3: Implementation**

* Goal: Execute the data analysis as per the design specifications.
* Steps:
  + Data Cleaning: Clean and prepare the data by handling missing values, removing duplicates, and standardizing formats.
  + Statistical Analysis: Apply ANOVA and Kruskal-Wallis test to assess the availability and exclusivity of high-rated content. Use descriptive statistics for genre-specific analysis.
  + Create Visualizations: Develop visual representations of the data such as bar charts, heatmaps, and pie charts.
  + Generate Report: Prepare a detailed report based on the analysis, summarizing key insights and conclusions.

**Phase 4: Verification**

* Goal: Ensure that the solution meets the project’s objectives and goals.
* Steps:
  + Review Results: Evaluate the effectiveness of the statistical tests and visualizations.
  + Test the Report: Check if the report effectively answers the research questions and provides actionable insights.

**Phase 5: Maintenance**

* Goal: Provide long-term support and ensure the solution remains relevant.
* Steps:
  + Documentation: Create detailed documentation on the data sources, cleaning processes, and analysis methods for future reference.
  + Update Strategy: Develop a plan for periodically updating the datasets to reflect changes in platform content (this could be manual or automated, depending on available resources).

## B.4 Timeline and Milestones

|  |  |  |  |
| --- | --- | --- | --- |
| **Milestone/Deliverable** | **Duration** | **Projected Start Date** | **Projected End Date** |
| Milestone 1: Data Collection and Cleaning | 4 days | 12/17/24 | 12/21/24 |
| Milestone 2: Statistical Analysis | 2 days | 12/21/24 | 12/23/24 |
| Milestone 3: Visualization Creation | 2 days | 12/23/24 | 12/25/24 |
| Milestone 4: Report Creation | 4 days | 12/25/24 | 12/29/24 |
| Milestone 5: Review and Finalization | 2 days | 12/29/24 | 12/31/24 |

## B.5 Resources and Costs

|  |  |
| --- | --- |
| **Resource** | **Cost** |
| Hardware: Laptop for data processing and analysis | $800 (one-time) |
| Software: Python and Jupyter Notebook | Free |
| Data Sources: Kaggle dataset and IMDb API access | Free |
| Work Hours: 112 hours for project tasks ( data cleaning, analysis, report creation) | $5,600 (112 hours @ $50/hour) |
| Total Estimated Project Cost | $6,400 |

## B.6 Criteria for Success

For this project to be considered successful, the following criteria must be met:

* The completeness and accuracy of the cleaned datasets will measure success.
  + Metric: 100% of missing data should be handled and all datasets merged correctly without duplicates.
* Statistical analysis should provide statistically significant findings (or appropriately inconclusive results\_ based on the data
  + Metric: Perform and interpret ANOVA and Kruskal-Wallis tests with a significance level 0.05.
* Visualizations should accurately represent the distribution of high-rated content, exclusive titles, and genre-specific content across platforms. These should be clear and help illustrate trends.
  + Metric: At least three visualizations created (e.g., bar charts, pie charts, heatmaps) with appropriate labels, titles, and legends for clarity
* The final report should summarize the analysis in a clear, concise, and actionable manner. It should provide insights on the best streaming platform(s) based on content quality, exclusivity, and availability.
  + Metric: The report should include a detailed comparison and actionable insights for consumers, plus an executive summary.

# Design of Data Analytics Solution

## C.1 Hypothesis

Null Hypothesis (H₀):

* When considering high-rated content, exclusivity, and access to popular titles, the overall value of streaming platforms is not significantly different.

Alternative Hypothesis (H₁):

* At least one streaming platform provides a significantly better overall value based on high-rated content, exclusivity, and access to popular titles.

## C.2 and C.2.A Analytical Method

ANOVA:

* Relevance to Hypothesis:
* ANOVA helps determine if there are statistically significant differences in key metrics (high-rated content percentage and exclusivity) across platforms, directly addressing whether platforms differ in value.
* Specificity of Metrics:
* The F-statistic and associated p-value provide clear criteria for evaluating differences, ensuring objectivity and adherence to statistical precision.

## C.3 Tools and Environments

1. IDE: Visual Studio Code
2. Programming Language: Python
3. Libraries:

* Pandas: Data manipulation.
* Scipy: ANOVA and Kruskal-Wallis test tests.
* Matplotlib/Seaborn: Visualizations of results.

1. Environment: Local with Python 3.9 and libraries installed.

## C.4 and C.4.A Methods and Metrics to Evaluate Statistical Significance

**ANOVA (Analysis of Variance)**

* **Null Hypothesis (H₀):**

There are no significant differences in the mean percentage of high-rated content and exclusivity across streaming platforms.

* **Planned Statistical Test:**

One-way ANOVA will compare group means of high-rated content percentage and exclusivity across multiple platforms.

* **Metrics Generated:**
  + **F-statistic:**

Represents the variance ratio between group means to the variance within groups.

* + **p-value:**
  + The probability of observing the results if the null hypothesis is true.
* **Alpha Value (α):**

Set at 0.05. If p < 0.05, the null hypothesis will be rejected, indicating significant differences in the metrics across platforms.

* **Justification for ANOVA**:

ANOVA is the appropriate test for comparing means across more than two groups (in this case, streaming platforms) when analyzing continuous numerical variables like high-rated content percentage and exclusivity.

**Kruskal-Wallis Test**

* **Null Hypothesis (H₀):**
* There are no significant differences in the median availability of popular content across streaming platforms.
* **Planned Statistical Test:**
* The Kruskal-Wallis test will compare the distributions of popular content availability across multiple streaming platforms.
* **Metrics Generated:**
  + **Kruskal-Wallis H Statistic (H):** Measures the difference in the ranks of content popularity across groups.
  + **p-value:** The probability of observing the data if the null hypothesis is true.
  + **Alpha Value (α):** Set at 0.05. If p < 0.05, the null hypothesis will be rejected, indicating significant differences in popular content availability across platforms.
* **Justification for Kruskal-Wallis Test:**

The Kruskal-Wallis test is a non-parametric alternative to ANOVA. It is ideal for comparing medians across groups when data assumptions for parametric tests, such as normality, are violated. This makes it suitable for analyzing popular content availability, which may not follow a normal distribution and is better represented through rankings.

## C.5 Practical Significance

Practical significance will be based on whether the insights provided by the analysis meaningfully inform consumer decisions regarding streaming platform subscriptions.

**High-Rated or Genre-Specific Content:**

* If a platform offers more highly rated or genre-specific content than others, this will be practically significant because it directly supports informed decision-making.

**Exclusive High-Quality Content:**

* If there is a significant difference between the availability of exclusive high-quality content, this is practically significant as it helps identify unique value choices for each platform.

**Customized Consumer Preferences:**

* Insights into platform strengths across specific genres or high-quality exclusives will be significant for customers with niche interests, enabling them to make personalized subscription choices.

## C.6 Visual Communication

We will use at least two graphic visualizations to communicate our findings in the project.

1. Bar Chart for High-rated content comparison across platforms
   1. This bar chart will display the percentage of high-rated content available on each platform.
   2. Each bar represents a streaming platform, and the height indicates the proportion of high-rated content in its library.
   3. I will be using matplotlib and seaborn to create this.
2. Heatmap for overlap of popular content across platforms.
   1. This heatmap will display the overlap percentage between the platforms. It will be filtered to titles labeled as popular. We will define this by titles within the top 75% of the IMDB catalog when considering rating and number of votes.
   2. The darker the squares between platforms, the higher the content overlap. Each cross section will also include a numeric value representing the percentage of the overlap.
   3. Again, I plan to use matplotlib and seaborn to create this heatmap.

# Description of Dataset

## D.1 Source of Data

**Streaming Datasets (Kaggle):**

* Hulu Dataset: [Full Hulu Dataset on Kaggle](https://www.kaggle.com/datasets/octopusteam/full-hulu-dataset)
* Apple TV+ Dataset: [Full Apple TV+ Dataset on Kaggle](https://www.kaggle.com/datasets/octopusteam/full-apple-tv-dataset)
* Amazon Prime Dataset: [Full Amazon Prime Dataset on Kaggle](https://www.kaggle.com/datasets/octopusteam/full-amazon-prime-dataset)
* Netflix Dataset: [Full Netflix Dataset on Kaggle](https://www.kaggle.com/datasets/octopusteam/full-netflix-dataset/data)
* HBO Max Dataset: [Full HBO Max Dataset on Kaggle](https://www.kaggle.com/datasets/octopusteam/full-hbo-max-dataset)

**IMDb Dataset:**

* Non-Commercial Datasets from IMDb: [IMDb Developer Page](https://developer.imdb.com/non-commercial-datasets/)

## D.2 Appropriateness of Dataset

**Streaming Datasets (Kaggle):**

* These datasets provide detailed information about the catalogs of major streaming platforms (Hulu, Apple TV+, Amazon Prime, Netflix, and HBO Max).
* Data points include title names, genres, release years, and other metadata essential for analyzing content quality and availability.
* They directly align to compare platforms based on high-rated content, exclusive offerings, and genre-specific availability.

**IMDb Dataset:**

* This dataset includes IMDb ratings, vote counts, and other key metrics for identifying high-rated and popular content.
* It is a widely recognized and credible source for assessing content quality, ensuring the analysis is based on reliable and standardized metrics.
* This dataset complements the streaming platform catalogs by providing objective quality indicators for each title.

## D.3 Data Collection Methods

S**treaming Datasets (Kaggle):**

* The data for each streaming platform was collected by downloading the respective .csv files from the Kaggle website:
  + Hulu: [Download Link](https://www.kaggle.com/datasets/octopusteam/full-hulu-dataset)
  + Apple TV+: [Download Link](https://www.kaggle.com/datasets/octopusteam/full-apple-tv-dataset)
  + Amazon Prime: [Download Link](https://www.kaggle.com/datasets/octopusteam/full-amazon-prime-dataset)
  + Netflix: [Download Link](https://www.kaggle.com/datasets/octopusteam/full-netflix-dataset/data)
  + HBO Max: [Download Link](https://www.kaggle.com/datasets/octopusteam/full-hbo-max-dataset)

**IMDb Dataset:**

* The IMDb dataset was downloaded from the official IMDb developer page, specifically the section dedicated to non-commercial datasets: [IMDb Non-Commercial Datasets](https://developer.imdb.com/non-commercial-datasets/).
* This ensures that all ratings, votes, and metadata for the analysis are sourced directly from IMDb, maintaining data reliability and credibility.

## D.4 Observations on Quality and Completeness of Data

1. Quality of Data:

* The Kaggle streaming platform and IMDb datasets are of high quality, as they originate from reputable sources and are widely used in data analysis projects.
* The data contains essential attributes such as title, genre, release year, ratings, and vote counts, all of which are formatted consistently and provide a comprehensive foundation for analysis.
* Preliminary checks indicate that data fields are accurately populated, with minimal inconsistencies in formatting or missing values.

1. Completeness of Data:

* The datasets are comprehensive, covering a wide range of content across major streaming platforms and detailed IMDb metrics.
* However, data gaps exist, such as missing ratings or vote counts for lesser-known titles. These gaps do not significantly impact the overall analysis, as the focus is on popular and high-rated content, which typically has complete data.
* These entries will either be excluded from specific analyses or appropriately imputed based on reasonable assumptions (e.g., average ratings within genres) to accommodate missing data.

**D.5 and D.5.A Data Governance, Privacy, Security, Ethical, Legal, and Regulatory Compliances**

The data is used in accordance with the terms of use published by Kaggle and IMDb. It does not contain personal or sensitive information, eliminating direct privacy concerns. The datasets are stored on secure, encrypted systems to prevent unauthorized access. Finally, results will be communicated transparently to avoid bias or misuse.

# References

Austin, P. L. (2015). Ditch Your Cable. Black Enterprise, 45(6), 38.

Zabel C, Kunz RE, Telkmann V, O’Brien D. The show must go on(line): the impact of content and system quality on the usage of television streaming content libraries. *Journal of Media Economics*. 2023;35(3/4):63-86. doi:10.1080/08997764.2024.2322825

Schneider M. Netflix, HBO Get Ready to Rumble. Variety. 2019;344(15):19. Accessed December 24, 2024. https://search.ebscohost.com/login.aspx?direct=true&db=brb&AN=137368483&site=eds-live&scope=site

# 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.