**Introduction**

Streaming platforms have revolutionized content consumption by offering diverse selections tailored to varying audience preferences. Services like Disney+, Netflix, Hulu, and Prime Video dominate the industry by targeting specific demographics. These platforms significantly influence individual perspectives and personalities, often tailoring content to age-specific audiences. Disney+ positions itself as a family-friendly service, emphasizing children's programming that fosters prosocial behavior and traditional values. In contrast, Netflix caters to a more mature demographic with content that explores complex themes, including mental health, sexuality, and societal issues. This study aims to critically assess whether Disney+ primarily serves younger audiences compared to Netflix’s adult-oriented viewership. Furthermore, it will conduct a statistical analysis using Rotten Tomatoes scores to determine whether Netflix films are rated higher than Disney+ offerings. Understanding age-based content differentiation and popularity metrics can benefit both consumers and businesses, enabling more strategic engagement with these platforms.

**Detailed Description of the Problem**

**Aim of this report**   
This report's main goal is to examine and contrast the films that are accessible on Disney+ and Netflix based on their age limits and Rotten Tomatoes ratings. Two main questions are intended to be addressed by this study:

1. Are movies on Disney+ generally aimed at a younger audience compared to Netflix?
2. Are movies on Netflix rated higher on average than those on Disney+ based on their Rotten Tomatoes scores?

In order to achieve this, this report will use a combination of statistical hypothesis testing, visualizations, and descriptive statistics to extract valuable information from the data.

**Research Questions**

 **Age Restrictions Analysis:**

* Is the general perception of Disney+ being a children-centric platform supported by data?
* Are the movies available on Disney+ more likely to have lower age restrictions compared to Netflix?

 **Rotten Tomatoes Score Comparison:**

* Are the movies on Netflix generally better in quality (as assessed by Rotten Tomatoes scores) compared to Disney+?
* How significant is the difference, if any, between the two platforms in terms of movie ratings?

**Detail about the Data**

The dataset used for this analysis is available on Kaggle: [Movies on Netflix, Prime Video, Hulu, and Disney+](https://www.kaggle.com/datasets/ruchi798/movies-on-netflix-prime-video-hulu-and-disney). The dataset includes information on movies across four streaming platforms: Netflix, Prime Video, Hulu, and Disney+. For the purpose of this report, we will focus on movies from Disney+ and Netflix.

The dataset contains the following key attributes:

 **ID:** A unique identifier assigned to each movie.

 **Title:** The name of the movie.

 **Year:** The year of release for the movie

 **Age:** The age restriction for the movie (e.g., "7+", "18+"), denoting the target audience

 **Rotten Tomatoes:** The Rotten Tomatoes score for the movie, presented as a percentage (e.g., "35/100," "68/100").

 **Netflix, Hulu, Prime Video, Disney+:** Binary indicators (0 or 1) that specify the availability of a movie on these platforms. For instance, a value of "1" in the Disney+ column indicates the movie is available on Disney+.

 **Type:** A column indicating the content type (potentially distinguishing between movies and TV shows).

**Methodology**

### Data Cleaning and Analysis

To clean and analyze the dataset, MS Excel, and Python were utilized. Various data preprocessing techniques were applied to ensure consistency, accuracy, and reliability in the analysis.

#### **Data Preprocessing**

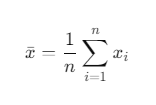
An exploratory data analysis (EDA) was conducted to understand the dataset’s structure, distribution, and relationships. String values were converted into numerical formats, enabling the handling of missing values.

#### **Imputation Techniques**

Imputation replaces missing values with estimated ones to maintain data integrity (Finucane et al., 2024). Mean imputation was applied to the "Rotten Tomatoes Score" after extracting numerical values. However, as 43.90% of the "Age" variable was missing, imputation was avoided to prevent biases.

#### **Descriptive Statistics**

Key statistical measures, including minimum, maximum, mean, and median, were used to summarize data characteristics. The mean, calculated as:



provides the average value, while the median, a robust measure against outliers, represents the dataset's midpoint. It is determined using:

A math equation with numbers and symbols

Description automatically generated

#### **Data Visualization**

**Pie Chart:** Pie charts visually represent proportional data using color-coded slices. While useful for small datasets, they are less effective for precise comparisons.

**Bar Chart:** A graphical representation of categorical data using rectangular bars, where bar height indicates value magnitude (Larsen, 1985). Bar charts effectively illustrate trends and patterns, making them essential for data analysts and decision-makers.

**Boxplot:** Boxplots summarize data distribution through key metrics—minimum, first quartile (Q1), median (Q2), third quartile (Q3), and maximum. The interquartile range (IQR) identifies variability, and whiskers highlight data spread, aiding in outlier detection (Cox, 2009).

These statistical techniques and visual tools enhance data-driven insights, ensuring a comprehensive analysis of streaming platform trends.

**Contingency tables**

A contingency table, or cross-tabulation, represents the frequency distribution of two or more categorical variables, organizing data into rows and columns. It is essential for analyzing relationships, dependencies, and associations between variables (Everitt, 1992).

**Chi-squared test for independence**

The chi-squared (χ²) test determines whether two categorical variables in a contingency table are statistically independent (S et al., 2024). By comparing observed and expected frequencies, it assesses the strength of association, rejecting the null hypothesis if significant deviation occurs.

The chi-square test can be calculated using the following formula:

|  |  |  |
| --- | --- | --- |
|  |  | ……..… (3) |

Where, are observed frequencies and (expected frequencies) is given by:

|  |  |  |
| --- | --- | --- |
|  |  | ……..… (4) |

With, is the total for row i, and is the total for column j, and N is the grand total of all observations.

Assumptions include group frequencies >5, independent observations, and no extreme outliers.

**Mann-Whitney test with Holm-Bonferroni correction**

The Mann-Whitney U test, introduced by Mann and Whitney (1947), is a non-parametric test used when normality assumptions are violated. With Holm-Bonferroni correction, it compares distributions of two independent groups, making it widely applicable in various scientific fields.

|  |  |  |
| --- | --- | --- |
|  |  | ……..… (5) |

**Evaluation**

**Data cleaning and preparation**

|  |
| --- |
|  |
| Figure 1: |

Though the dataset consists of ten variables only “Age” and “Rotten Tomatoes” had missing values. Figure 1 illustrates the percentage of missing and non-missing values for the “Age” variable, where 43.90% of the data is missing, and 56.10% is non-missing. This means that the total number of missing values is 4,177 and the total number of non-missing values is 5338.

|  |
| --- |
| A graph of missing value and missing value  Description automatically generated |
| Figure 2: |

Figure 2 is a bar diagram that represents the total percentage of missing values and non-missing values for the variable “Rotten Tomatoes” where approximately 0.07% of the data was missing for that variable and 99.93% of the data was non-missing. The total number of the missing values are for the variable “Rotten Tomatoes” are 7.

As the variable “Age” was a categorical variable which represents the age restrictions on movies across different platforms such as Disney+ and Netflix, we have extracted the numeric value from this value and created a new variable called “Numeric\_Age”. Furthermore, A variable named “Rotten\_Tomatoes\_Score” has been created from extracting the numerical part from the string variable.

So, as the variable “Age” consist of a significant number of the missing value (43.90%), imputing this value may result in some unwanted bias which will affect our analysis. On the other hand, the variable “Rotten Tomatoes” consists of 0.07% missing value, so this will not create any bias after imputing this missing value. So, two datasets have be prepared for analyzing this two variables, where mean Imputation technique has been used for the variable “Rotten Tomatoes” and for “age” variable missing values was dropped (see appendix).

**Descriptive Analysis**

**Age Restrictions**

|  |
| --- |
|  |
| Figure 3: |

Age restriction is categorized into five categories such as, “nan”, “7+”, “13+”, “16+”, “18+”. Figure 3 depicts the overall percentage of the categories in the varia ble age restriction in a donut chart. Almost 13.08% of the movies have no age restriction, where 20.42% of the movies are “7+”. Furthermore, 18.70% of the movies has restricted age for “13+”. “16+” age restriction categories had the lowest percentage among all categories which is 5.17%. Additionally, “18+” age category has the highest percentage which is 42.64%.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Table 1: | | | | | |
| Age Category | Disney+ | | Netflix | | Total |
| No | % | No | % |
| 0 | 370 | 71.98% | 144 | 28.02% | 514 |
| 7+ | 278 | 46.33% | 322 | 53.66% | 600 |
| 13+ | 70 | 14.77% | 404 | 85.23% | 474 |
| 16+ | 4 | 2.58% | 151 | 97.42% | 155 |
| 18+ | 3 | 0.34% | 877 | 99.66% | 880 |
| Total | | | | | 2623 |

The number and percentage of titles in each age category are highlighted in Table 1, which displays the distribution of age restrictions across Disney+ and Netflix. Compared to Netflix (28.02%), Disney+ has a larger percentage of titles with no age restrictions (71.98%), however Netflix leads the "13+," "16+," and "18+" categories with 85.23%, 97.42%, and 99.66% of titles, respectively. Netflix also holds a slim lead (53.66%) over Disney+ (46.33%) in the "7+" category. 2,623 titles in all were examined, indicating different content distributions between the two platforms. As the age restrictions for 16+ and 18+ are less than 5 we have to recategorize this. The age restrictions category has been recategorized into three categories such as No age restrictions, age restriction less than 16 years, and age restriction greater than 16 years.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Table 2: | | | | | |
| Age Category | Disney+ | | Netflix | | Total |
| No | % | No | % |
| No age restriction | 370 | 71.98% | 144 | 28.02% | 514 |
| <16 | 348 | 32.40% | 726 | 67.60% | 1074 |
| 16+ | 7 | 0.68% | 1028 | 99.32% | 1035 |
| Total | | | | | 2623 |

**Table 2** provides the counts and percentages for each platform's distribution of new age groups for movies on Disney+ and Netflix. Netflix has a smaller percentage (28.02%) than Disney+ (71.98%) of titles without an age restriction. Netflix has the most titles (67.60%) in the "<16" category, with Disney+ contributing 32.40%. With 99.32% of the titles in the "16+" category, Netflix has a significant lead over Disney+, which has just 0.68%. The two platforms have 2,623 titles in total, which shows notable variations in the material addressed by age group.

**Rotten tomatoes**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table 3: | | | | |
| Platform | Rotten Tomatoes Score | | | |
| Mean | Median | Mode | Std Dev |
| Netflix | 54.45 | 53 | 46 | 13.83 |
| Disney+ | 58.31 | 57.5 | 48.00 | 13.95 |

**Table 3** summarizes the Rotten Tomatoes ratings for Netflix and Disney+ films. At 54.45, Netflix's mean score is little lower than Disney+'s (58.31). The median scores for Netflix and Disney+ are 53 and 57.5, respectively, suggesting that Disney+ has a greater central tendency. For Netflix and Disney+, the mode which indicates the most frequent score, is 46 and 48, respectively. With standard deviations of 13.83 for Netflix and 13.95 for Disney+, the two platforms' scores exhibit comparable variability.

|  |
| --- |
| A graph with blue and black lines  Description automatically generated with medium confidence |
| Figure 4: |

**Figure 4** illustrates a box plot that shows how the Rotten Tomatoes ratings for films on Netflix and Disney+ are distributed. According to the orange horizontal line in each box, Disney+ has a higher median score than Netflix. Compared to Netflix, Disney+ has a somewhat smaller interquartile range (IQR), which is indicated by the height of the boxes. This suggests that there is less variation within the middle 50% of scores. There are more outliers on Netflix than on either platform, yet both have outliers below their minimum whisker values. Both platforms have a comparable overall score range, ranging from the smallest to the largest whiskers.

|  |
| --- |
| A graph with a red line  Description automatically generated |
| Figure 5: |

In **Figure 5** the scores plotted against an arbitrary index, the scatter plot displays the Rotten Tomatoes ratings for Disney+ and Netflix films. Netflix movies are shown by red dots, and Disney+ movies are shown by blue dots. According to the plot, a greater range of scores are often covered by Netflix films, with many falling into the mid- to low-range. With ratings primarily in the mid- to high range, Disney+ films, on the other hand, exhibit a closer grouping. According to the clustering pattern, Netflix's Rotten Tomatoes scores vary more than Disney+'s, whilst Disney+'s scores show a more stable distribution.

**Data Distribution**

The findings of normalcy tests for Netflix's and Disney+'s age limits are shown in Table 4. At a p-value < 0.001, the Shapiro-Wilk test statistic for Disney+ is 0.7683, whereas for Netflix, it is 0.7583. Netflix has a p-value of <0.001 and a Kolmogorov-Smirnov test statistic of 0.5000, while Disney+ has a p-value of 0.9241. These findings show that there is a substantial departure from normalcy in the age restriction data for both platforms.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table 4: | | | | |
| Test name | Disney+ | | Netflix | |
| Statistic value | p-value | Statistic value | p-value |
| Shapiro-Wilk Test | 0.7683 | <0.001 | 0.7583 | <0.001 |
| Kolmogorov-Smirnov Test | 0.9241 | <0.001 | 0.5000 | <0.001 |

Disney+ exhibits a large concentration at the lowest age restriction (0), whereas Netflix shows a higher frequency at the maximum age restriction (18+), indicating that both Netflix and Disney+ have unevenly distributed age restriction scores based on the histogram. The rejection of normalcy is supported by this lopsided distribution.

So, based on this information we can say that the age restriction variable is non-parametric.

|  |
| --- |
| A graph of age restriction scores  Description automatically generated |
| Figure 6: |

The results of the normalcy test for Netflix and Disney+'s Rotten Tomatoes ratings are shown in Table 5. Disney+ had a statistic value of 0.9844 with a p-value of less than 0.001 and Netflix had a statistic value of 0.9800 with a p-value less than 0.001 according to the Shapiro-Wilk test. For both systems, the Kolmogorov-Smirnov test yielded a p-value of less than 0.001 and a statistic value of 1.00.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table 5: | | | | |
| Test name | Disney+ | | Netflix | |
| Statistic value | p-value | Statistic value | p-value |
| Shapiro-Wilk Test | 0.9844 | <0.001 | 0.98 | <0.001 |
| Kolmogorov-Smirnov Test | 1.00 | <0.001 | 1.00 | <0.001 |

|  |
| --- |
| A graph of a number of tomatoes  Description automatically generated |
| Figure 7: |

The histogram shows the distribution of Rotten Tomatoes ratings for films that are accessible on Netflix (blue) and Disney+ (orange). A bell-shaped curve more closely resembles Netflix's distribution, with a larger frequency of scores centered in the middle (50–60). With lower frequencies in the middle and somewhat flat frequencies across other ranges, Disney+, on the other hand, has a wider and more scattered distribution. Furthermore, the density curves superimposed on the histogram indicate that Disney+ scores do not closely follow a normal pattern, displaying random variances across the range, whereas Netflix scores follow a slightly skewed bell-shaped distribution.

Based on this analysis we can say that the variable Rotten tomatoes score is non-parametric.

**Is the age restriction for movies on Disney+ lower than for movies on Netflix?**

**Table 6** represents the summary of the chi-square test of the age restriction for movies on Disney+ and Netflix. In the table we can see that p-value is less than 0.001 and the degree of freedom is 2, and the chi-square test yielded a statistic value of 895.65, demonstrating a very significant difference between Disney+ and Netflix in terms of age restriction categories.

|  |  |  |  |
| --- | --- | --- | --- |
| Table 6: | | | |
| Test | Chi-square | p-value | Degree of freedom |
| Test scores | 895.65 | <0.001 | 2 |

**Table 7** represents the findings of Mann-Whitney U test for Age Restriction. The age restriction scores between the two platforms were compared using the Mann-Whitney U test, which yielded a statistic value of 166350.5 with a p-value of less than 0.001. The statistical significance of the variations in age limitation scores was further supported by the adjusted p-value (<0.001).

|  |  |
| --- | --- |
| Table 7: | |
| Mann-Whitney U Test for Age Restriction | |
| Name | Values |
| Statistic | 166350.5 |
| p-value | <0.001 |
| Adjusted p-value | <0.001 |
|  |  |

In the **Figure 8**, the average age restriction scores for both platforms are displayed in the bar chart. Netflix has a mean score of 13.55, whereas Disney+ has a far lower mean score of 4.1. This striking disparity suggests that Netflix offers a wider variety of content with stricter age limits, while Disney+ predominantly provides content appropriate for younger children.

|  |
| --- |
|  |
| Figure 8: |

Finally observing all the test data, we can conclude that Disney+ has substantially less age restrictions for its films than Netflix is strongly supported by the statistical tests and visual data. This is in line with Disney+'s emphasis on kid-friendly and family-friendly programs, whereas Netflix provides a more varied selection aimed at an adult audience.

**Is there a difference in Rotten Tomatoes Score for movies on those two platforms?**

**Table 8** represents the summary of the chi-square test of the Rotten tomatoes’ sores for movies on Disney+ and Netflix. Here for performing chi-square test we had to recategorize the numerical values of the Rotten Tomatoes where score below 40 was categorized as low, values from 41 to 70 was recategorized as average and the value above that was recategorized as High.

With a degree of freedom of two and a p-value of less than 0.001, the chi-square test yielded a statistic value of 52.44. These findings show that the two platforms' Rotten Tomatoes scores differ significantly.

|  |  |  |  |
| --- | --- | --- | --- |
| Table 8: | | | |
| Test | Chi-square | p-value | Degree of freedom |
| Test scores | 52.44 | <0.001 | 2 |

**Table 9** represents the findings for the Mann-Whitney U test for Rotten Tomatoes Score between two platforms Disney+ and Netflix. With a p-value of less than 0.001 and an adjusted p-value of less than 0.001, the Mann-Whitney U test produced a statistic value of 1988799.0. This demonstrates that there is a statistically significant difference between Disney+ and Netflix's Rotten Tomatoes scores.

|  |  |
| --- | --- |
| Table 9: Mann-Whitney U test for Rotten Tomatoes Score | |
| Name | Values |
| Statistic | 1988799.0 |
| p-value | <0.001 |
| Adjusted p-value | <0.001 |
|  |  |

In **Figure 9** the bar chart shows the mean Rotten Tomatoes scores for the two platforms. Disney+ has a higher mean value of **58.31**, compared to Netflix’s mean value of **54.45**. This suggests that movies on Disney+ tend to receive slightly better ratings than the Netflix.

|  |
| --- |
|  |
| Figure 9: |

Together, the visual data and statistical testing show that Disney+ and Netflix have significantly different Rotten Tomatoes ratings. Disney+ films typically receive higher ratings, which may indicate greater reception or a better fit with the tastes of viewers and reviewers.

**Summary**

In order to compare age limits and Rotten Tomatoes ratings, this study examined the content that was accessible on Disney+ and Netflix. With most of its titles having lower age limits, we discovered through descriptive statistics that Disney+ generally targets younger audiences. Netflix, on the other hand, has a wider selection of content, many of which are geared toward older viewers. The average Rotten Tomatoes rankings for Disney+ films were marginally higher than those for Netflix films, suggesting that reviewers may favor Disney+ content. Significant variations in age limitations and ratings between the two platforms were validated by statistical tests. These findings demonstrate how Disney+ and Netflix have different content strategies.

Future Research

The effects of economic, geographic, and cultural factors on streaming preferences might be investigated in more detail. Furthermore, looking at measures related to user engagement, like watch durations and subscription patterns, may yield information for improving platform content strategies.

References

Cox, N. J., 2009. Speaking Stata: Creating and varying box plots. *The Stata Journal Promoting Communications on Statistics and Stata* [online], 9 (3), 478–496.Available from: https://doi.org/10.1177/1536867x0900900309.

Dave, N. A., Kolli, N. R. K., Mokkapati, N. C., Goel, N. O., Khan, N. Dr. S., and Jain, N. A., 2022. Techniques for Enhancing User Engagement through Personalized Ads on Streaming Plat-forms. *Universal Research Reports* [online], 9 (3), 196–218.Available from: https://doi.org/10.36676/urr.v9.i3.1390.

Edmond, M., Khoo, O., Perkins, C., and Trott, V., 2024. Streaming diversity: Studying screen diversity in the streaming era. *Convergence the International Journal of Research Into New Media Technologies* [online], 30 (4), 1315–1330.Available from: https://doi.org/10.1177/13548565241270887.

Everitt, B. S., 1992. *The analysis of contingency tables* [online].Available from: https://doi.org/10.1201/b15072.

Finucane, K., Brennan, L., and Gormley, I. C., 2024. Missing data imputation using a truncated infinite factor model with  application to metabolomics data. *arXiv (Cornell University)* [online].Available from: http://arxiv.org/abs/2410.10633.

Forni, D., 2020. YOUNG ADULTS AND TV SERIES.NETFLIX AND NEW FORMS OF SERIAL NARRATIVES FOR YOUNG VIEWERS. *MeTis - Mondi Educativi Temi Indagini Suggestioni* [online], 10 (1).Available from: https://doi.org/10.30557/mt00124.

Fraser, C., 2024. Association Between Categorical Variables: Contingency Analysis with Chi Square. *In*: *Business Statistics for Competitive Advantage with Excel and JMP* [online]. 47–72.Available from: https://doi.org/10.1007/978-3-031-42555-4\_3.

Garman, 2011. Family on Film : How Disney visual media influences children’s perceptions of family. Honors. [online]. Ball State University.Available from: https://studylib.net/doc/10950223/family-on--film---how-disney-visual--media-influences...-...

Gupta, S. N., 2011. Mean, median, mode: an introduction. *In*: *International Encyclopedia of Statistical Science* [online]. 788–791.Available from: https://doi.org/10.1007/978-3-642-04898-2\_353.

Kalita, J. K., Bhattacharyya, D. K., and Roy, S., 2024. Data preparation. *In*: *Elsevier eBooks* [online]. 31–46.Available from: https://doi.org/10.1016/b978-0-32-391778-0.00010-7.

Kaur, H. and Ashfaq, R., 2023. The Impact of Netflix on viewer behaviour and media Consumption: An exploration of the effects of streaming services on audience engagement and entertainment preferences. *Journal of Media Culture and Communication* [online], (34), 9–23.Available from: https://doi.org/10.55529/jmcc.34.9.23.

Lakshmanan, M., 2014. Histogram and Vertical Bar diagram: Often Misapprehended Concept. *Cardiovascular Drugs and Therapy* [online], 28 (4), 387.Available from: https://doi.org/10.1007/s10557-014-6529-6.

Larsen, R. D., 1985. Box-and-whisker plots. *Journal of Chemical Education* [online], 62 (4), 302.Available from: https://doi.org/10.1021/ed062p302.

Padilla-Walker, L. M., Coyne, S. M., Fraser, A. M., and Stockdale, L. A., 2013. Is Disney the nicest place on Earth? A content analysis of prosocial behavior in animated Disney films. *Journal of Communication* [online], 63 (2), 393–412.Available from: https://doi.org/10.1111/jcom.12022.

S, V., AS, H., and Sundar, J. S., 2024. CHI-SQUARE TESTS: a QUICK GUIDE FOR HEALTH RESEARCHERS. *International Journal of Advanced Research* [online], 12 (10), 1214–1222.Available from: https://doi.org/10.21474/ijar01/19746.

**Appendix**

**Artifact**

**(link:** [**https://colab.research.google.com/drive/1ZX8o1LHHaqXd4gSaWN4-xJO2hZPx7Txr?usp=sharing**](https://colab.research.google.com/drive/1ZX8o1LHHaqXd4gSaWN4-xJO2hZPx7Txr?usp=sharing) **)**

# Library Installation  
# (Ensure the following libraries are installed)  
import pandas as pd  
import matplotlib.pyplot as plt  
import seaborn as sns  
import numpy as np  
from scipy.stats import chi2\_contingency, shapiro, normaltest, mannwhitneyu  
  
# Section 1: Loading the Data  
from google.colab import drive  
drive.mount('/content/drive')  
  
file\_path = '/content/drive/My Drive/Colab Notebooks/MoviesOnStreamingPlatforms.csv'  
data = pd.read\_csv(file\_path)  
  
# Output: Display a sample of the data  
data.sample(10)  
  
# Output: Data Information  
data.info()  
  
# Output: Data Types  
print(data.dtypes)  
  
# Output: Unique values in each column  
for column in data.columns:  
 print(f"Unique values in '{column}':")  
 print(data[column].unique())  
 print()  
  
# Section 2: Data Cleaning  
# Checking Missing Values  
total\_rows = len(data)  
missing\_values = data.isnull().sum()  
present\_values = total\_rows - missing\_values  
  
# Preparing data for visualization  
missing\_summary = pd.DataFrame({  
 "Variable": data.columns,  
 "Present Values": present\_values.values,  
 "Missing Values": missing\_values.values  
})  
  
# Visualization of Missing Values  
plt.figure(figsize=(10, 6))  
sns.barplot(data=missing\_summary.melt(id\_vars="Variable", var\_name="Status", value\_name="Count"),  
 x="Variable", y="Count", hue="Status", palette=["blue", "red"])  
plt.title("Presence and Missing Values per Column", fontsize=14)  
plt.xlabel("Variables", fontsize=12)  
plt.ylabel("Count", fontsize=12)  
plt.legend(title="Data Status", fontsize=10)  
plt.xticks(rotation=45)  
plt.tight\_layout()  
plt.show()  
  
# Display Missing Summary as Table  
fig, ax = plt.subplots(figsize=(10, 4))  
ax.axis("tight")  
ax.axis("off")  
table = ax.table(cellText=missing\_summary.values, colLabels=missing\_summary.columns, cellLoc="center", loc="center")  
table.auto\_set\_font\_size(False)  
table.set\_fontsize(10)  
table.auto\_set\_column\_width(col=list(range(len(missing\_summary.columns))))  
plt.show()  
  
# Section 3: Preparing Data for Analysis  
# Dropping unnecessary columns  
df = data.drop(columns=data.columns[[0,1]].tolist() + ['ID', 'Type'])  
  
# Checking for duplicate values  
duplicate\_count = df.duplicated().sum()  
print(f"Number of duplicate rows: {duplicate\_count}")  
  
# Extracting Rotten Tomatoes scores  
df['Rotten\_tomatoes\_score'] = df['Rotten Tomatoes'].str.extract(r'(\d+)', expand=False).astype(float)  
  
# Extracting numeric age from 'Age' column  
df['Age'] = df['Age'].replace('all','0')  
df['Numeric\_Age'] = df['Age'].str.extract(r'(\d+)', expand=False).astype(float)  
  
# Output: Display sample rows  
df.sample(10)  
  
# Section 4: Data Analysis and Visualization  
## Compare the age restrictions of Disney+ and Netflix  
# Filtering non-missing age data  
df1 = df.dropna(subset=['Age'])  
  
# Filtering data for Disney+ and Netflix  
filtered\_df = df1[(df1['Netflix'] == 1) | (df1['Disney+'] == 1)]  
  
# Age statistics  
mean\_age\_netflix = filtered\_df[filtered\_df['Netflix'] == 1]['Numeric\_Age'].mean()  
mean\_age\_disney = filtered\_df[filtered\_df['Disney+'] == 1]['Numeric\_Age'].mean()  
  
range\_age\_netflix = filtered\_df[filtered\_df['Netflix'] == 1]['Numeric\_Age'].max() - filtered\_df[filtered\_df['Netflix'] == 1]['Numeric\_Age'].min()  
range\_age\_disney = filtered\_df[filtered\_df['Disney+'] == 1]['Numeric\_Age'].max() - filtered\_df[filtered\_df['Disney+'] == 1]['Numeric\_Age'].min()  
  
# Output statistics  
print("Descriptive Statistics for Disney+")  
print("Mean age:", mean\_age\_disney)  
print("Range of ages:", range\_age\_disney)  
  
print("\nDescriptive Statistics for Netflix")  
print("Mean age:", mean\_age\_netflix)  
print("Range of ages:", range\_age\_netflix)  
  
# Visualize age restrictions  
disney\_data = df1[df1['Disney+'] == 1]['Numeric\_Age']  
netflix\_data = df1[df1['Netflix'] == 1]['Numeric\_Age']  
sns.boxplot(data=[disney\_data, netflix\_data], palette="coolwarm")  
plt.xticks([0, 1], ["Disney+", "Netflix"])  
plt.title("Boxplot: Age Restrictions")  
plt.ylabel("Age")  
plt.show()  
  
# Section 5: Rotten Tomatoes Analysis  
# Handling Missing Values  
missing\_rt\_values = df['Rotten\_tomatoes\_score'].isnull().sum()  
mean\_rt = df['Rotten\_tomatoes\_score'].mean()  
df['Rotten\_tomatoes\_score'].fillna(mean\_rt, inplace=True)  
  
# Rotten Tomatoes score statistics  
disney\_scores = df[df['Disney+'] == 1]['Rotten\_tomatoes\_score']  
netflix\_scores = df[df['Netflix'] == 1]['Rotten\_tomatoes\_score']  
  
mean\_disney\_rt = disney\_scores.mean()  
mean\_netflix\_rt = netflix\_scores.mean()  
  
print("Mean Rotten Tomatoes Score:")  
print(f"Disney+: {mean\_disney\_rt:.2f}, Netflix: {mean\_netflix\_rt:.2f}")  
  
# Histograms  
sns.histplot(disney\_scores, color="blue", kde=True, label="Disney+")  
sns.histplot(netflix\_scores, color="red", kde=True, label="Netflix")  
plt.legend()  
plt.title("Histogram: Rotten Tomatoes Scores")  
plt.xlabel("Score")  
plt.ylabel("Frequency")  
plt.show()  
  
# Section 6: Hypothesis Testing  
# Mann-Whitney U Test  
age\_stat, age\_p = mannwhitneyu(disney\_data, netflix\_data, alternative='less')  
print(f"Mann-Whitney U Test: Statistic = {age\_stat:.2f}, p-value = {age\_p:.4f}")  
  
# Additional statistics for Rotten Tomatoes scores  
rt\_stat, rt\_p = mannwhitneyu(disney\_scores, netflix\_scores, alternative='two-sided')  
print(f"Rotten Tomatoes Mann-Whitney U Test: Statistic = {rt\_stat:.2f}, p-value = {rt\_p:.4f}")