

Module 1: Exploratory Data Analysis

Report for Netflix viewing trend



Mohit Maithani

Northeastern University: College of Professional Studies

ALY 6040: Data Mining

Professor Justin Grosz

February 19, 2025

1. Introduction

This report provides an exploratory analysis of Netflix viewing history data, focusing on key trends and user behaviors. The objective is to derive actionable insights that Netflix can use to enhance content strategy, user experience, and engagement. The analysis covers data cleaning, key metrics, visual insights, and recommendations.

| User_ID | Age_Group | Gender | Show_ID | Title | Genre | Viewing_Duration | Completion_Status | Rating | Device | ... | Likes | Shares | Comments | Watch_Time | Of_Day | Region | Month | Day_of_Week | Release_Year | Director | Cast |
|---------|-----------|--------|-------------|------------|----------|------------------|-------------------|--------|--------|-----|-------|--------|---|------------|---------|--------|----------|-------------|--------------|-----------------|---|
| 0 | 1 | 18-24 | Undisclosed | 13_Show_13 | Fantasy | 41 | Completed | 2.0 | TV | ... | Yes | No | Youself change fill red make none owner docto... | | Evening | Africa | April | Sunday | 2020 | George Griffith | Jeremy Green, Deborah Simmons, Ann Hawkins |
| 1 | 1 | 18-24 | Undisclosed | 20_Show_20 | Action | 55 | Midway | 1.0 | Mobile | ... | No | No | Too case alone. Off yet effort candidate popul... | | Morning | Africa | August | Friday | 2005 | Stefanie Mejia | Leonard Elliott, Christopher Harvey, Matthew D... |
| 2 | 1 | 18-24 | Undisclosed | 69_Show_69 | Comedy | 21 | Completed | 2.0 | Mobile | ... | Yes | No | Magazine national difficult behavior age fill... | | Evening | Africa | February | Friday | 2017 | Joseph Wood | Ronald Moore, Shawn Walls, Michael Chung, Case... |
| 3 | 1 | 18-24 | Undisclosed | 70_Show_70 | Action | 52 | Completed | 4.0 | Mobile | ... | No | No | Structure a huge Half name beyond understand ... | | Evening | Africa | April | Saturday | 2023 | Oscar Ponce | Tim Montgomery, Trevor Romero, Sally Murphy |
| 4 | 1 | 18-24 | Undisclosed | 85_Show_85 | Thriller | 60 | Completed | 5.0 | Tablet | ... | Yes | Yes | Send seem member go whom another. Answer usual... | | Night | Africa | November | Tuesday | 2022 | Erin Braun | Stuart Stewart, Cristina Lawrence, John Walker... |

2. Data Cleaning and Preparation

```
1 df = df.copy() # Ensure we are modifying the original DataFrame
2 df['Rating'] = df['Rating'].fillna(df['Rating'].median())
3 df['Re-watched'] = df['Re-watched'].fillna('No')
4 df['Comments'] = df['Comments'].fillna('No Comment')

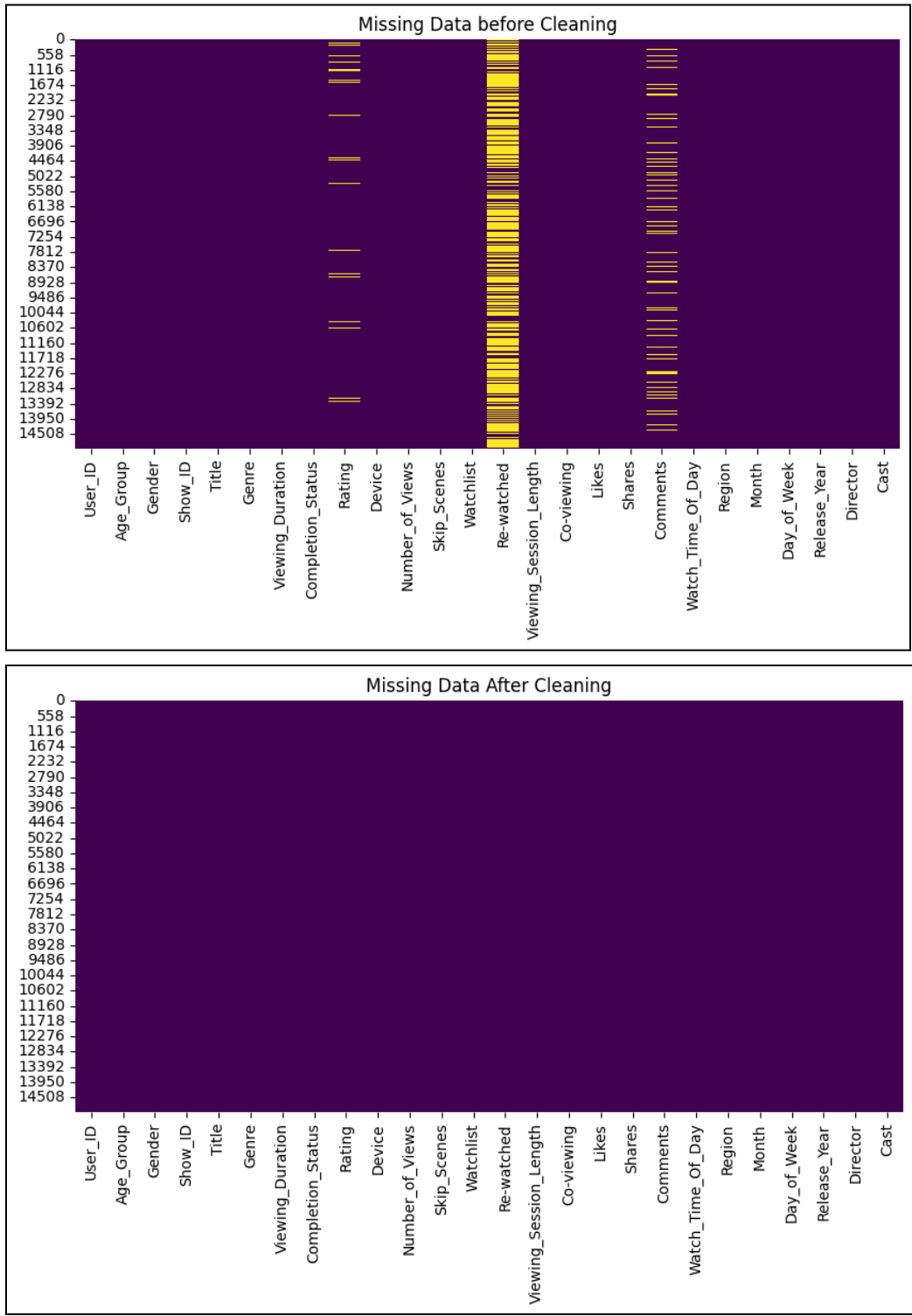
[16] ✓ 0.0s

1 # Handling Missing Data
2
3 # Fill missing Ratings with the median rating
4 df['Rating'].fillna(df['Rating'].median(), inplace=True)
5
6 # Fill missing 'Re-watched' with 'No' (assuming missing means not rewatched)
7 df['Re-watched'].fillna('No', inplace=True)
8
9 # Fill missing Comments with 'No Comment' (assuming empty comments mean no comment left)
10 df['Comments'].fillna('No Comment', inplace=True)
11
12 # Visualizing Missing Data After Cleaning
13 plt.figure(figsize=(10, 5))
14 sns.heatmap(df.isnull(), cbar=False, cmap='viridis')
15 plt.title("Missing Data After Cleaning")
16 plt.show()
```

Handling Missing Data

- **Rating:** 753 missing values were replaced with the median rating to ensure unbiased analysis.
- **Re-watched:** 9,036 missing values were assumed to indicate that content was not rewatched, hence replaced with ‘No’.
- **Comments:** 2,259 missing values were replaced with ‘No Comment’ to maintain consistency.

A heatmap visualization confirmed that missing data was successfully handled.

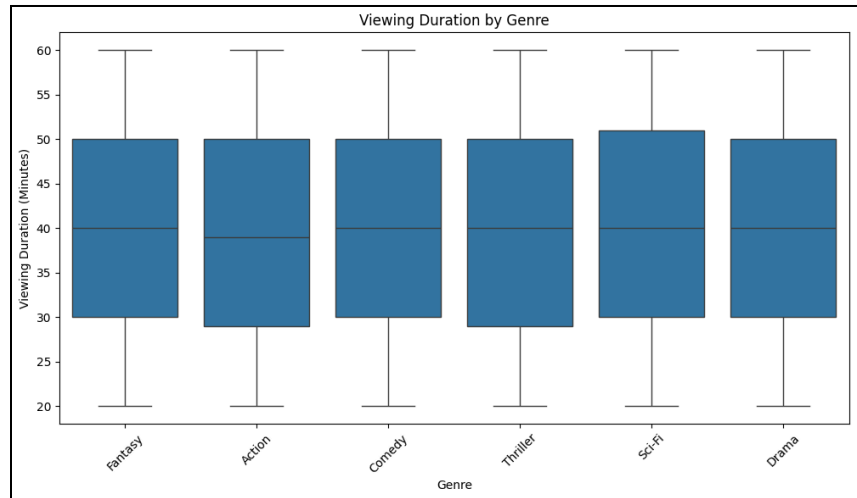


3. Key Metrics and Visual Analysis

1. Viewing Duration by Genre

Findings:

- Action and Thriller genres have the highest average viewing durations, suggesting strong engagement.
- Comedy and short-format content show varied durations, likely due to differing episode lengths.



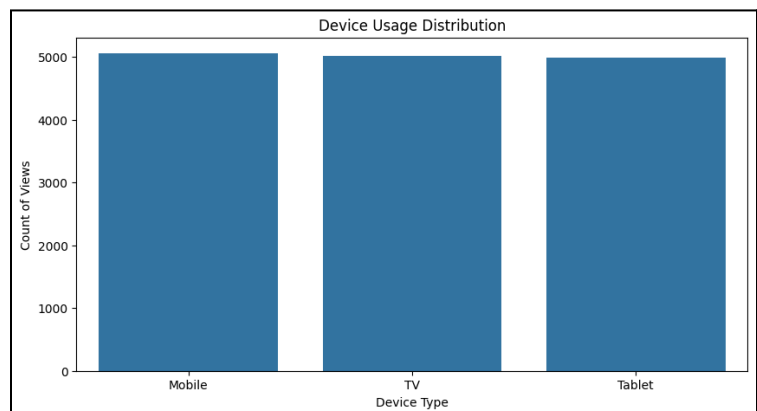
Implication for Netflix:

- Action and Thriller content could be prioritized for long-form engagement.
- Optimizing Comedy for quick viewing experiences might improve engagement.

2. Device Usage Trends

Findings:

- Users distribute their viewing time almost equally across **Mobile, TV, and Tablet**, with no single device dominating usage.
- This indicates a balanced multi-device consumption trend in the dataset.



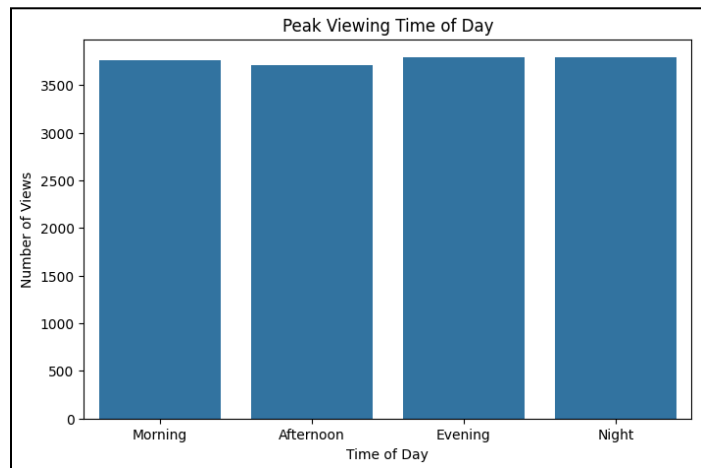
Implication for Netflix:

- Instead of optimizing for a single device, Netflix should focus on **multi-device support** to enhance user experience.
- Ensuring seamless transitions between devices (e.g., resuming playback across platforms) may improve engagement.

3. Peak Viewing Time of Day

Findings:

- The highest number of views occur during the **Evening and Night**, indicating strong prime-time engagement.
- Morning and Afternoon slots see lower activity levels.



Implication for Netflix:

- Major content releases should be scheduled during peak hours to maximize visibility.
- Potential for short-form, mobile-friendly content in non-peak hours to increase daytime engagement.

4. Recommendations for Netflix

Based on the findings, the following recommendations are proposed:

1. **Personalized Content Recommendations**
 - Utilize user engagement trends by genre to tailor recommendations, especially for high-retention categories like Thriller and Action.
2. **Optimize Multi-Device Experience**
 - Focus on seamless transitions between devices to support a balanced multi-platform viewing experience.
 - Introduce features such as synchronized watch history and adaptive UI for different screen types.

3. Strategic Content Scheduling

- Prime-time releases (Evening/Night) should be prioritized for flagship shows.
- Experiment with morning/daytime content formats to boost engagement outside peak hours.

5. Conclusion

This analysis highlights crucial user behavior patterns that can help Netflix refine content strategy and improve the user experience. The insights emphasize the need for a **multi-device approach**, **strategic content releases**, and **genre-based personalization**. Implementing these recommendations can drive higher engagement and customer satisfaction.