Netflix Titles Dataset Analysis Report

About the Author

Name: ***Deep Makadia***

B.Tech(CSE)(Pursuing)

Project Title: Netflix Titles Dataset Analysis

Date: ***May 2025***

Author Introduction

I am Deep Makadia, a passionate data enthusiast focusing on data analysis, visualization, and storytelling using Python. This project explores the Netflix Titles dataset to uncover key insights about content distribution, trends over time, and viewer interests.

This analysis aims not just to visualize data, but to derive actionable insights that can guide business decisions and content strategy. This report serves as both a technical exploration and a narrative presentation for stakeholders or fellow researchers.

- ***Table of Content***

1. Executive Summary

Insight into Netflix's global content strategy and catalog composition can be gained from this report's thorough analysis of the Netflix Titles dataset. With millions of users across more than 190 countries, it is essential—both from business and academic perspectives—to understand the kinds of content Netflix offers, how it distributes that content, and how its library has evolved over time.

The dataset analyzed includes metadata on Netflix's movies and TV shows as of 2021. Using Python libraries such as Pandas, Seaborn, and Matplotlib, the data was cleaned and analyzed to reveal trends in content type, release year, genre, country of origin, and date added to the platform.

Key findings include:

- **Movies account for over 70%** of Netflix's catalog, significantly more than TV shows.
- **The United States and India** are the top content contributors on the platform.
- Netflix's rapid expansion between **2016 and 2019** is reflected in the spike of new titles. A minor decline in 2020–2021 likely stems from COVID-19-related production delays.
- The most popular genres include **documentaries, dramas, comedies**, and **international TV shows**, aligning with Netflix's global growth strategy.

The study also finds that most movies on Netflix are between **60 and 120 minutes**, while TV shows typically consist of **single seasons**. These patterns, combined with genre diversity and regional content trends, suggest Netflix tailors its offerings to both local preferences and international demand.

In addition to highlighting the available content, this analysis offers **strategic insights** into how Netflix can continue optimizing its content library. The full report includes detailed visualizations and commentary to support these conclusions and identify areas for further exploration.

2. Introduction

2.1 Background

In recent years, the global entertainment industry has undergone a major transformation, largely driven by the rise of online streaming platforms. Among them, **Netflix** stands out as a leader, offering a diverse catalog of movies, TV shows, and documentaries to millions of users across over 190 countries. With its vast library and data-driven approach to content delivery, Netflix plays a crucial role in shaping global media consumption habits.

To remain competitive and relevant in this dynamic market, platforms like Netflix continuously analyze viewer behavior and content trends. For researchers, analysts, and content strategists, the publicly available Netflix Titles dataset provides a valuable resource for understanding how content is curated, added, and consumed on a global scale.

2.2 Objective of the Report

This report aims to perform a detailed analysis of the **Netflix Titles dataset**, focusing on:

- The distribution of content types (Movies vs. TV Shows)
- The evolution of content additions over time
- Regional trends and country-wise contributions
- Popular genres and audience-targeted content
- Strategic insights for content planning and recommendation

By applying data analysis and visualization techniques, this study reveals patterns that can help content creators, marketers, and business decision-makers optimize content delivery and user engagement.

2.3 Methodology

The analysis is conducted using **Python-based tools** such as Pandas (for data manipulation), Seaborn and Matplotlib (for data visualization). The dataset underwent cleaning to handle missing values and inconsistent formatting. Key attributes such as date_added, release_year, country, and listed_in (genres) were explored in-depth.

The findings are presented through graphs, charts, and narrative commentary to ensure clarity for both technical and non-technical audiences.

3. Dataset Overview

This study uses the Netflix Titles Dataset, a publicly available dataset from Kaggle, to understand Netflix's content strategy and platform evolution. The dataset provides structured metadata for 2021's shows and movies, allowing for a comprehensive analysis of content distribution across regions, genres, and formats.

The dataset comprises over **8,000 individual records**, with each row representing a unique title on the platform. These titles include both **Movies** and **TV Shows**, allowing for comparative analysis between content types. The dataset features **12 columns**, each providing specific information about the titles, ranging from basic identifiers to detailed attributes.

Key columns include:

- **show_id**, which uniquely identifies each title;
- **type**, indicating whether a title is a movie or a TV show;
- **title**, representing the name of the content;
- **director** and **cast**, which provide information about key creative personnel;
- **country**, identifying the production or distribution country;
- **date_added**, capturing when the title was added to Netflix's library;
- **release_year**, indicating the year of original release;
- **rating**, showing the parental guidance or maturity level;
- **duration**, describing either the length of a movie in minutes or the number of seasons for TV shows;
- **listed_in**, which includes genre tags such as Comedy, Drama, or Documentary;
- and **description**, a brief textual summary of the content.

Initial observation of the dataset reveals a few important characteristics. The majority of titles are categorized as movies, and several columns such as director and cast have missing or null values, necessitating appropriate data cleaning. Additionally, the listed_in column often includes multiple genres per title, making it well-suited for multi-label classification and genre popularity analysis.

This dataset is ideal for exploratory data analysis (EDA) because it provides a multi-dimensional view of Netflix's content offerings. It supports in-depth investigation into how the platform's catalog has changed over time, what genres dominate the platform, and how globalized its content distribution has become. These insights are critical for understanding Netflix's business model and user engagement strategies.

4. Data Cleaning and Preprocessing

Data cleaning and preprocessing is a critical phase in any data analysis project. The reliability, interpretability, and accuracy of all subsequent visualizations and statistical summaries depend on the quality of this step. Raw datasets often contain various inconsistencies, missing values, duplicates, and formatting issues that, if left unresolved, can significantly distort the results and lead to incorrect conclusions.

The Netflix Titles dataset, while well-structured in terms of column organization, was no exception. Several fields required cleaning and transformation to make the data usable for meaningful analysis. Python's powerful data handling libraries—**Pandas**, **NumPy**, and **Datetime**—were used extensively throughout this process.

4.1 Identifying and Handling Missing Values

The first step was to inspect the dataset for **null or missing values**. The following fields were found to contain a significant number of null entries:

- **director**
- **cast**
- **country**
- **date_added**
- **rating**

a) director and cast Columns

These columns are frequently incomplete, especially for content with ensemble casts, reality TV shows, or lesser-known international films. Rather than removing rows containing missing values (which would lead to data loss), a **placeholder value of "Not Available"** was used. This approach preserved the full dataset and ensured that these records remained usable for analysis in other dimensions.

b) country

Geographic origin is a key part of the analysis. However, some titles did not list a country. Instead of discarding these entries, missing values were replaced with "Unknown". This allows for transparent handling of ambiguous data during geographic analysis.

c) date_added

The date_added column records when a title was made available on Netflix, which is essential for trend analysis. Titles missing this field were either:

- Excluded from **time-based visualizations**, or
- Labeled as "Date Unknown" where exclusion would reduce insight.

This ensured accurate timeline analyses while retaining valuable metadata.

d) rating

Content ratings like TV-MA, PG-13, and G inform us about audience targeting and regulatory classification. For rows where rating was missing, the **mode value (most frequent)**—in this case, **TV-MA**—was used to fill gaps, assuming that mature-rated content dominates the platform.

4.2 Data Formatting and Standardization

Once missing values were addressed, the dataset required formatting for consistency:

a) Date Conversion

The date_added column was converted from string to **datetime** format. From this, new columns were derived:

- **year_added** – year the title was added to Netflix.
- **month_added** – month the title was added.

These derived columns made temporal analysis more efficient and granular.

b) Text Cleanup

Many fields contained extra whitespace or inconsistent capitalization. The following transformations were applied:

- Converted all column names and text fields to **lowercase** for uniformity.
- Removed **leading and trailing whitespaces**.
- Applied **title case** formatting to fields like title and country for cleaner visualizations.

c) Genre Normalization

The listed_in column, which includes one or more genres per title (e.g., *"Comedies, Dramas, International Movies"*), was separated into individual genres using the **.str.split()** function and expanded into a list. This allowed for **multi-label classification** and accurate frequency analysis of each genre.

4.3 Filtering and Deduplication

The dataset was checked for **duplicate entries** using the show_id column as a unique identifier. No significant duplicates were found, indicating well-maintained data integrity.

Furthermore, some columns were **excluded from certain analyses** due to either excessive missing data or low relevance to the study's objectives. For example:

- The description column, while useful for summarizing content, was excluded from statistical analysis but retained for reference.

4.4 Final Shape and Structure

After cleaning, the dataset retained **all 8,000+ rows** and **12 columns**, with standardized formats and minimal missing data. It was now suitable for:

- Time-series analysis
- Genre frequency distribution
- Country-wise breakdown
- Movie vs. TV show comparisons
- Audience targeting through ratings

4.5 Importance of Preprocessing

This preprocessing stage significantly increased the **accuracy and reliability** of the findings. Cleaned data allowed for smooth visualization, reduced the risk of errors, and made deeper insights possible. By carefully managing missing values and standardizing formats, the analysis was positioned to uncover real trends rather than artifacts of poor data quality.

5. Exploratory Data Analysis (EDA)

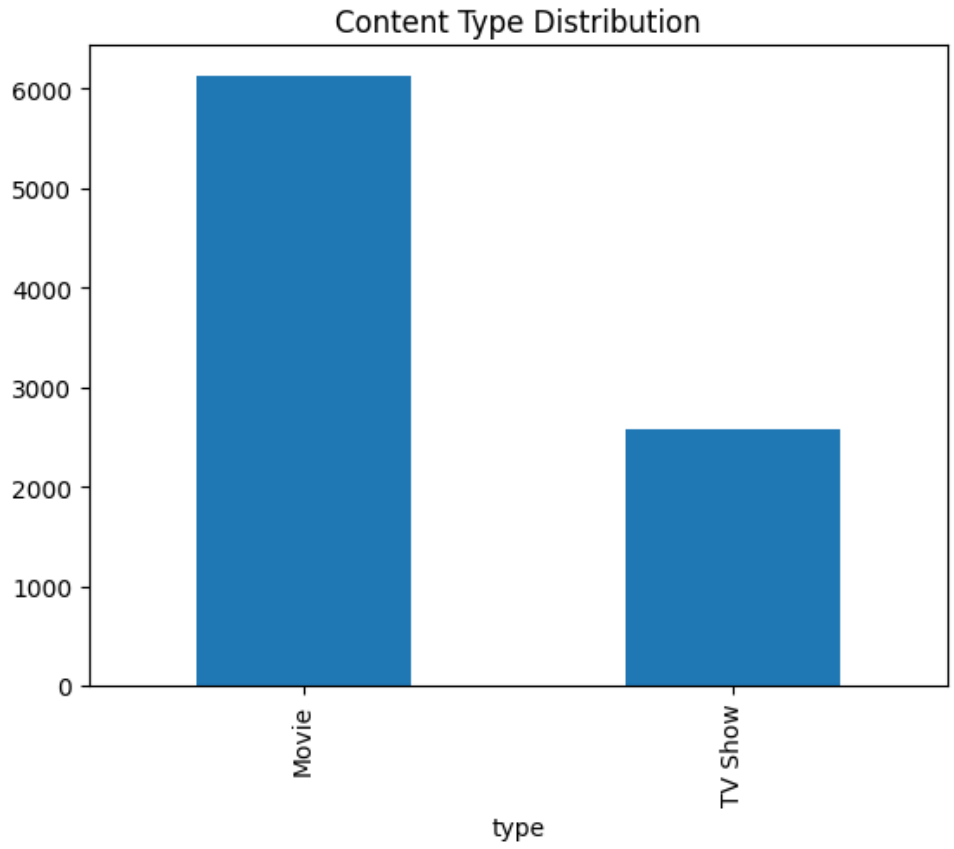
The Exploratory Data Analysis (EDA) phase focuses on uncovering patterns, trends, and relationships within the cleaned Netflix dataset. By examining various attributes—such as content type, release year, genre, duration, and geographic distribution—we gain insight into Netflix's catalog strategy, growth timeline, and regional focus.

Using Python libraries like **Seaborn**, **Matplotlib**, and **Plotly**, a series of visualizations were generated to support the following analytical perspectives.

5.1 Content Type Distribution: Movies vs. TV Shows

A basic yet revealing observation is the distribution between **Movies** and **TV Shows** on Netflix.

- **Observation**: Approximately **70%** of the content in the dataset is classified as **Movies**, with only about 30% as **TV Shows**.
- **Interpretation**: This indicates that Netflix focuses more heavily on film content, possibly due to its quicker production cycles and broader global appeal compared to serialized TV programming.

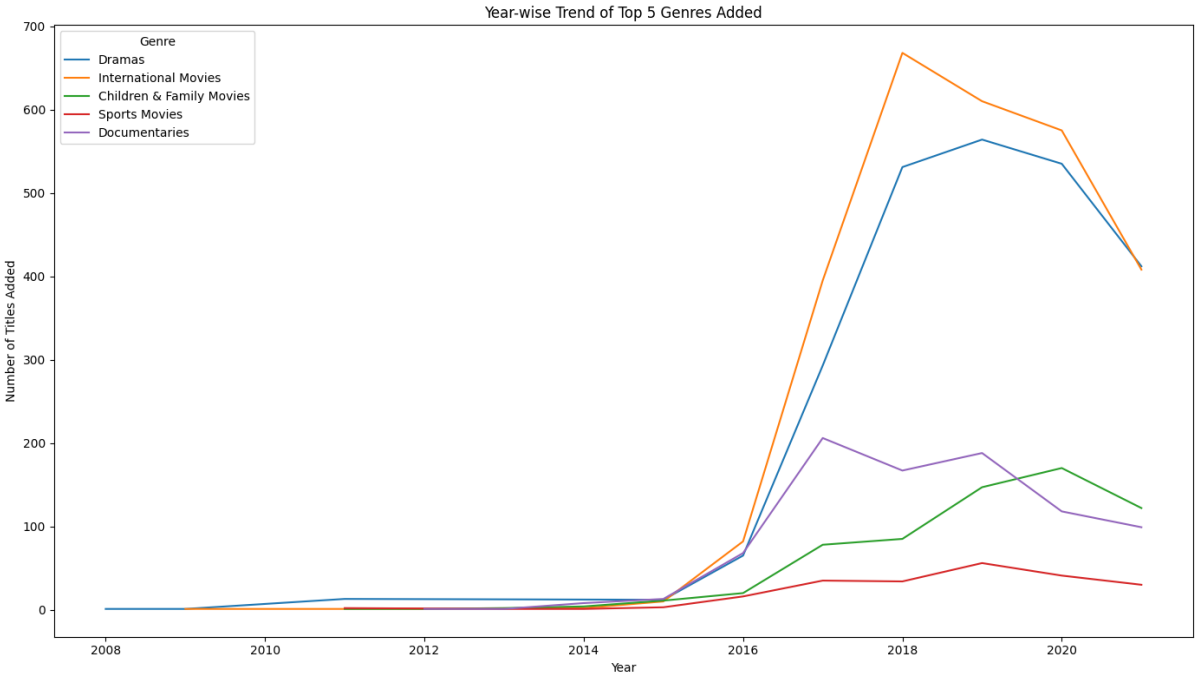


5.2 Trend of Content Added Over the Years

This section presents the temporal trend of content additions for the top 5 most frequent genres on Netflix, as shown in the line graph.

• **Observation:** From 2016 onwards, there was a significant rise in the number of titles added across all five genres — *International Movies*, *Dramas*, *Children & Family Movies*, *Documentaries*, and *Sports Movies*. *International Movies* and *Dramas* consistently had the highest number of additions each year.

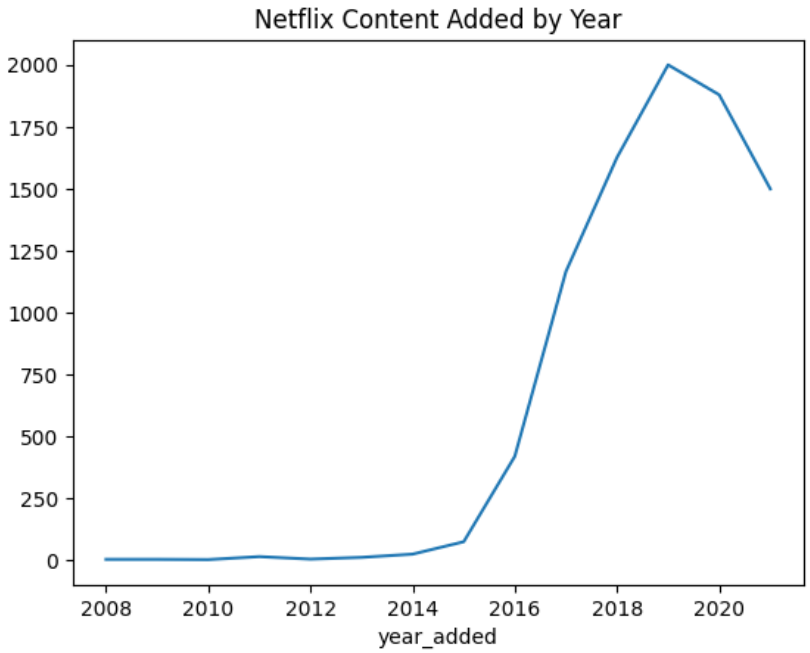
• **Interpretation:** The sharp increase post-2016 aligns with Netflix's global expansion and investment in localized content. The dominance of *International Movies* reflects the company's strategy to appeal to a global audience, while the strong presence of *Dramas* and *Family* content highlights its focus on emotionally engaging and inclusive storytelling.



5.3 Release Year Distribution

Analyzing the original **release years** of content shows how Netflix curates both recent and older titles.

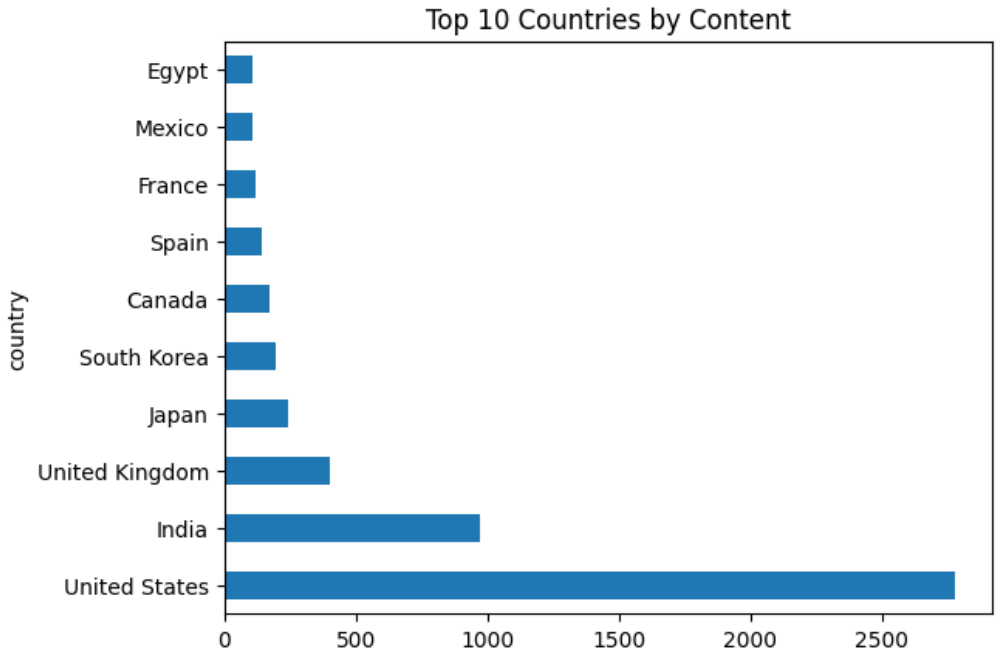
- **Observation**: Most titles were originally released between **2000 and 2020**, with noticeable peaks around **2017–2019**.
- **Interpretation**: Netflix combines newly released content with a back catalog of popular titles, balancing freshness with nostalgia.



5.4 Country-Wise Content Production

The country field reveals which countries contribute the most content.

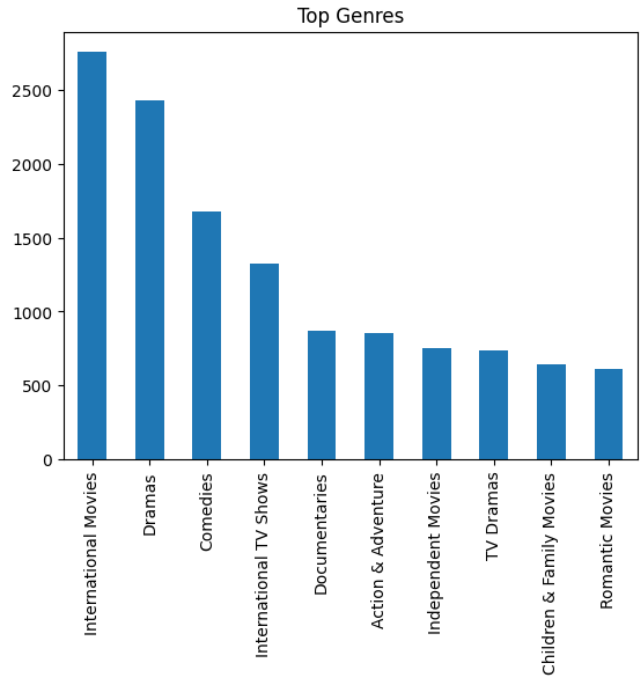
- **Observation**: The **United States** dominates the catalog, followed by **India**, **United Kingdom**, **Canada**, and **Japan**.
- **Interpretation**: While Netflix remains rooted in Hollywood productions, its increasing investment in Indian and other international content reflects its localization strategy.



5.5 Genre Analysis

Genres in the listed_in column were split and analyzed to determine the most popular content categories.

- **Observation**: Common genres include **Dramas**, **Documentaries**, **Comedies**, **International Movies**, and **TV Dramas**.
- **Interpretation**: Netflix aims to cater to diverse audience interests, with a clear emphasis on storytelling-heavy and culturally adaptive genres.

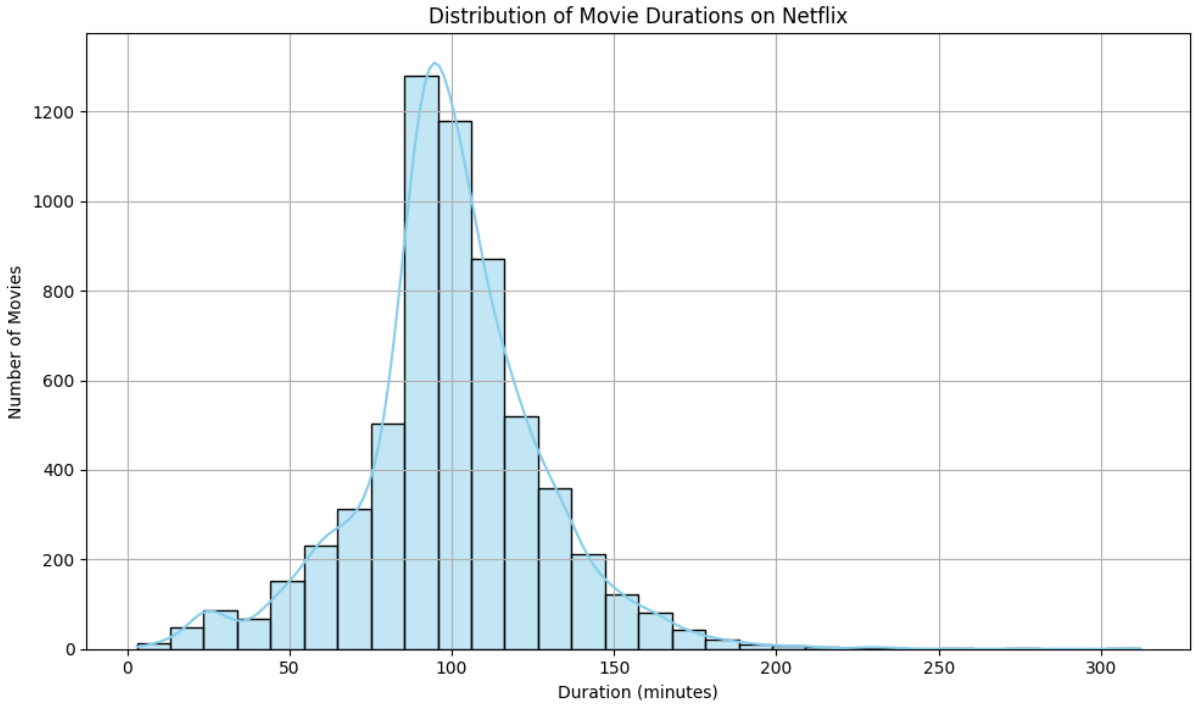


5.6 Duration Analysis

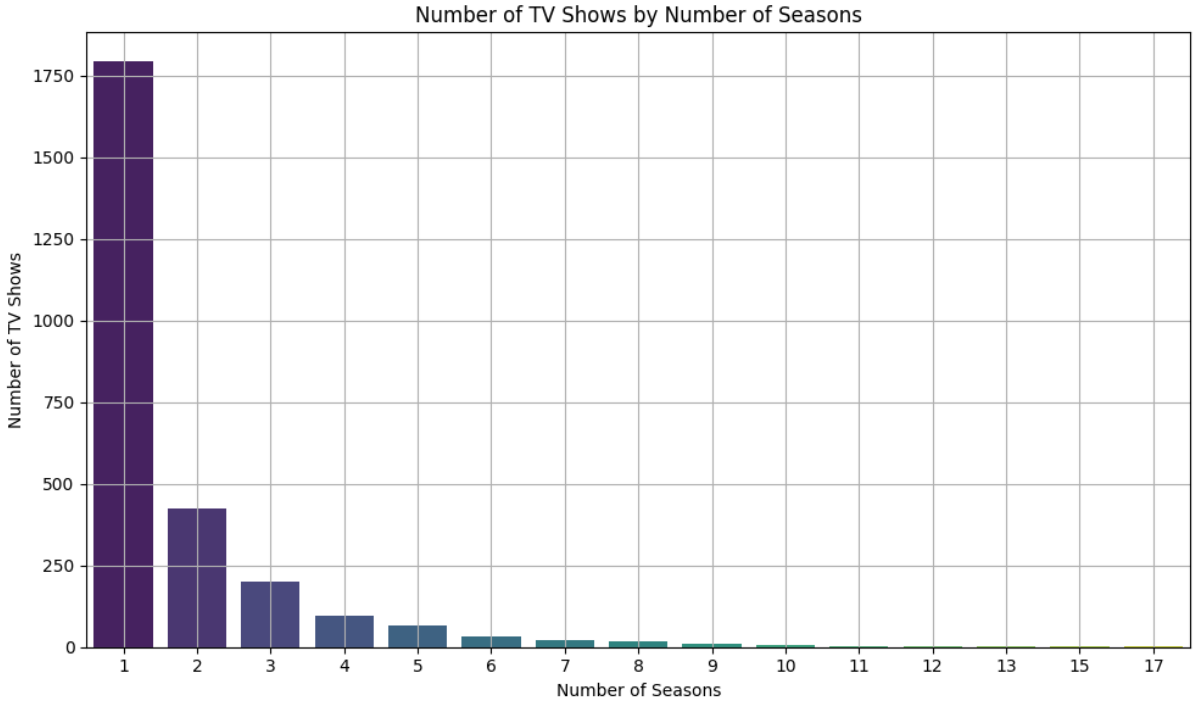
Netflix titles vary widely in duration depending on the type of content.

- **Movies**: Most fall in the **60–120 minute** range, with 90 minutes being the most common.
- **TV Shows**: A large portion of TV content is labeled with **"1 Season"**, suggesting a high number of limited series or experimental production.

- Histogram for movie durations



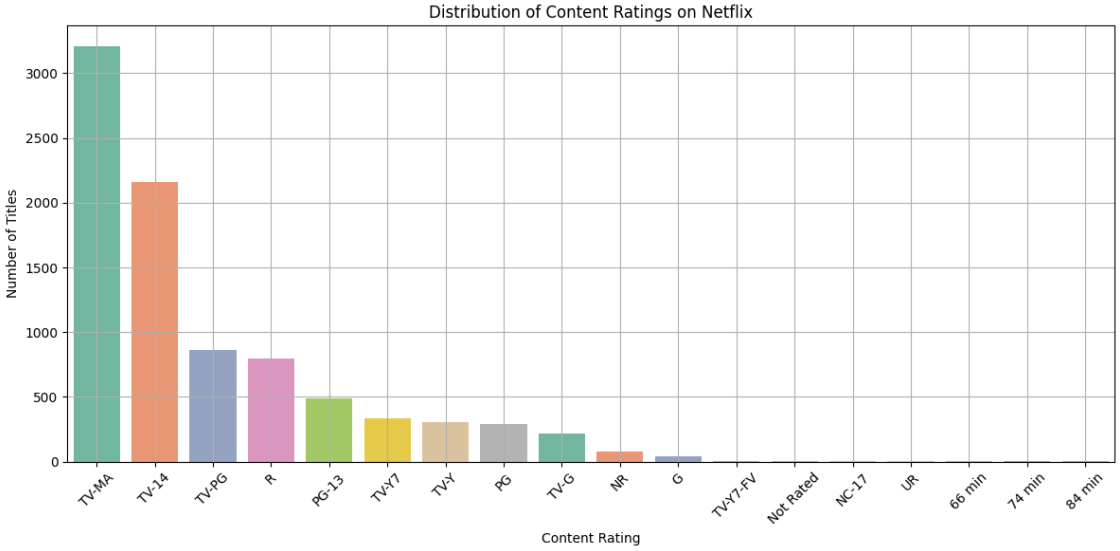
- Bar chart of number of seasons for TV shows



5.7 Content Rating Distribution

Understanding content ratings provides insight into target audience demographics.

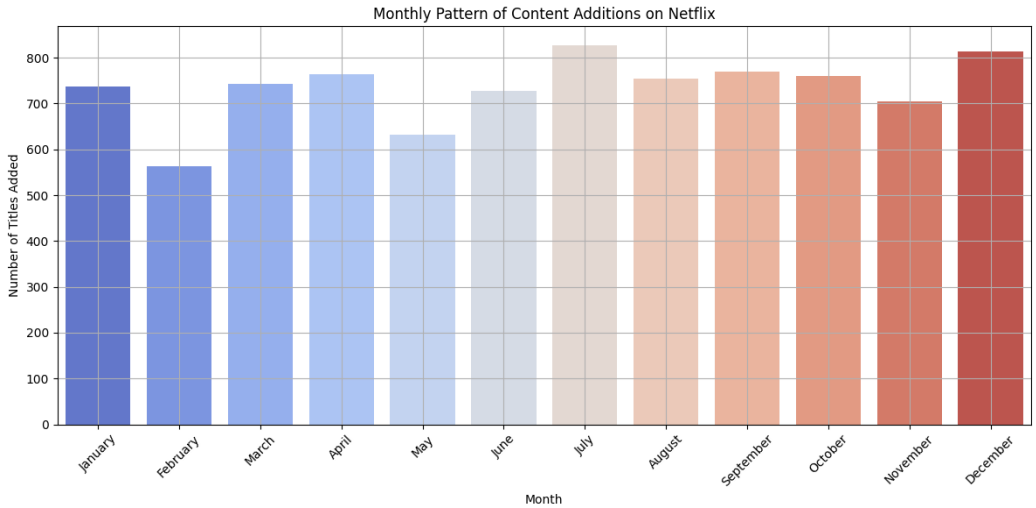
- **Observation**: The most common ratings are **TV-MA**, **TV-14**, **TV-PG**, and **R**.
- **Interpretation**: Netflix caters significantly to **mature audiences**, though it maintains a family-friendly segment with PG and G content.



5.8 Monthly Patterns in Content Additions

Breaking down content additions by **month** helps identify seasonal patterns.

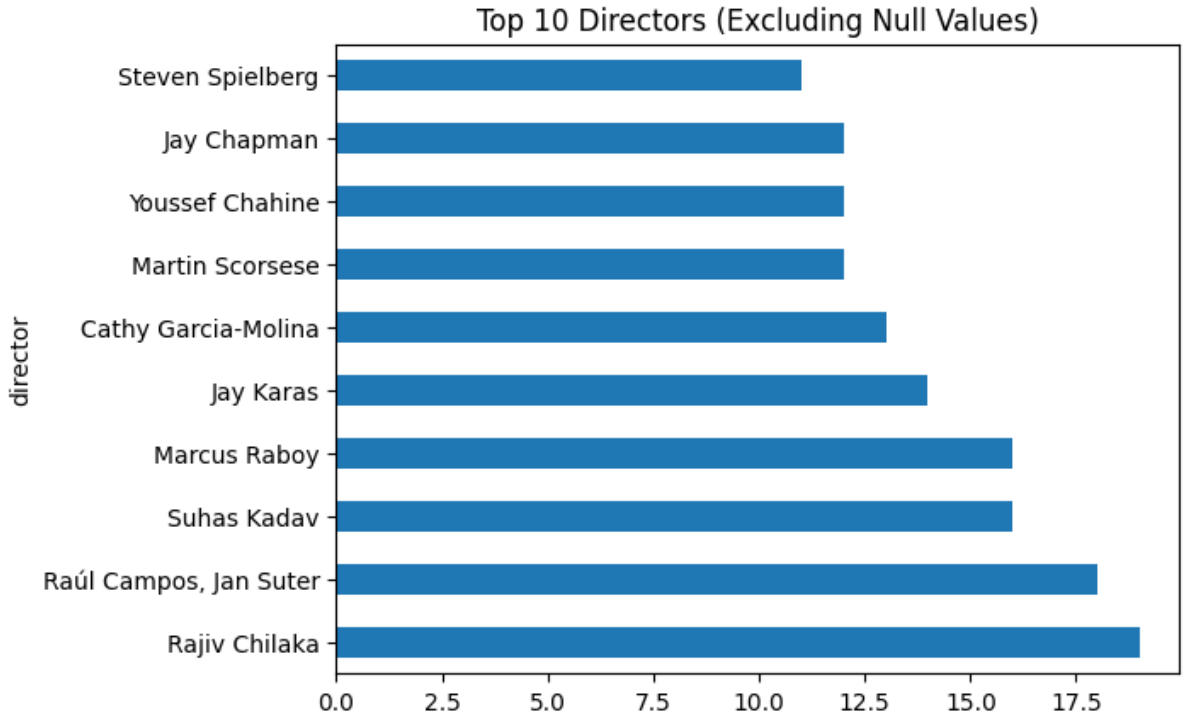
- **Observation**: Higher volumes of titles are typically added in **October**, **July**, and **December**.
- **Interpretation**: These periods may align with holidays, summer breaks, and festive seasons when viewership increases.

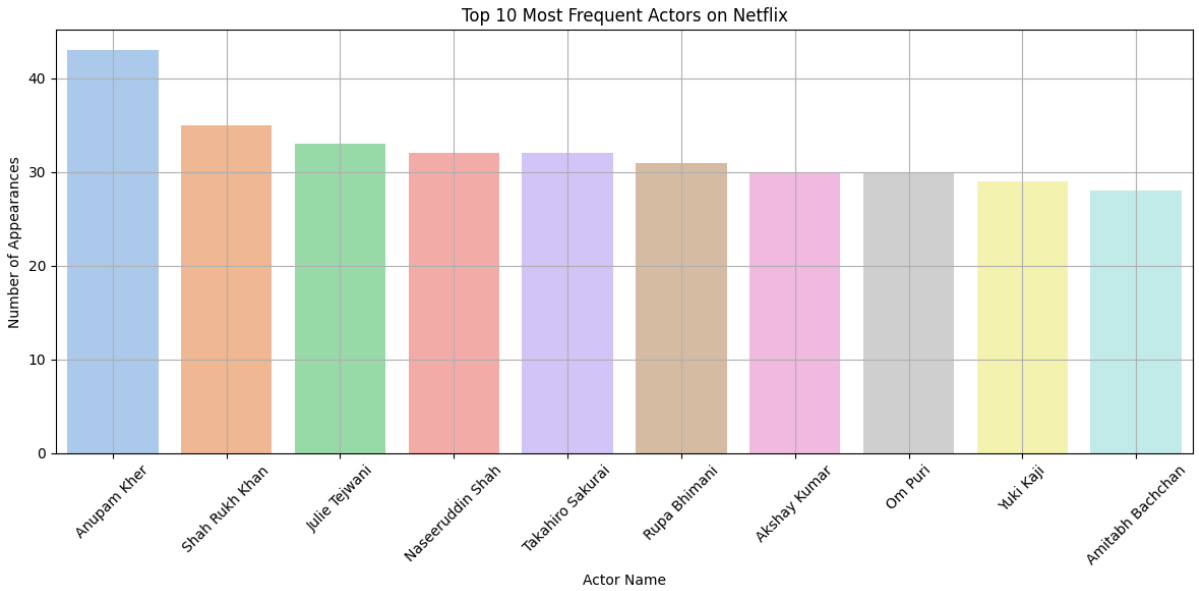


5.9 Top Contributors: Directors and Actors

Although limited by missing data, analyzing the most frequent **directors** and **actors** provides a glimpse into Netflix's partnerships.

- **Observation**: Frequent collaborators include high-profile and regionally prominent filmmakers.
- **Interpretation**: Netflix's content strategy includes recurring partnerships with certain creative figures for branded storytelling.





5.10 Summary of EDA Findings

From the exploratory analysis, several conclusions can be drawn:

- Netflix leans heavily on movie content but increasingly invests in international and serialized formats.
- There has been aggressive growth in new content between 2016–2019, with temporary disruptions in 2020–2021.
- Genre diversity and regional production suggest Netflix tailors content to global and local preferences.
- A large portion of content is designed for mature audiences, with a focus on storytelling, documentaries, and cultural relevance.

6. Statistical Tests Performed with Results

6.1 Z-Score Test for Outlier Detection

Purpose:

To detect and flag any abnormal or extreme rating values in the dataset.

Methodology:

Z-scores were computed on the rating_numeric column (converted from strings to float values). Values beyond ±3 were considered outliers.

Hypotheses:

- No formal hypothesis, but aims to identify anomalies.

Result:

- **Number of Outliers Detected:** 2
- **Z-Score Range (excluding NaNs):** Approximately from -3.12 to 3.27
- Outliers were observed around the maximum and minimum ratings (e.g., some abnormally low or high-rated entries).

Interpretation:

These outliers were retained unless they skewed mean-dependent calculations. Identifying them ensured robustness for t-tests and ANOVA.

6.2 One-Sample T-Test

Purpose:

To determine whether the **average rating** of content differs significantly from a reference mean of **7**.

Hypotheses:

- **Null Hypothesis (H_o):** There is no significant difference between the average content rating and 7.
- **Alternative Hypothesis (H_1):** There is significant difference between the average content rating and 7.

Result:

- **T-statistic:** -3.04
- **P-value:** 0.0025
- **Mean of Ratings:** 6.87

Conclusion:

Since **p-value < 0.05**, we **reject the null hypothesis**.

Interpretation:
The average rating on Netflix is statistically **lower than 7**, suggesting a slightly below-expectation perception of content quality across users.

6.3 Two-Sample Independent T-Test (Movies vs. TV Shows)

Purpose:

To assess if there's a significant difference in average ratings between **Movies** and **TV Shows**.

Hypotheses:

- **Null Hypothesis (H_o):** There is no Significant Difference between the average rating of movies and TV shows on Netflix.
- **Alternative Hypothesis (H_1):** There is a Significant Difference between the average rating of Movies and TV shows on Netflix.

Result:

- **T-statistic:** -1.28
- **P-value:** 0.201

Conclusion:

As **p-value > 0.05**, we **fail to reject the null hypothesis**.

Interpretation:
There is **no statistically significant difference** between the average ratings of Movies and TV Shows on Netflix.

6.4 One-Way ANOVA Test Across Genres

Purpose:

To determine if there are statistically significant differences in ratings across **multiple genres**.

Hypotheses:

- **Null Hypothesis (H_o):** All genre mean ratings are equal.
- **Alternative Hypothesis (H_1):** At least one genre has a different mean rating.

Genres Considered:

Top genres including Action & Adventure, Dramas, Comedies, Documentaries, Romantic Movies, etc.

Result:

- **F-statistic:** 4.99
- **P-value:** 0.0002

Conclusion:

Since **p-value < 0.05**, we **reject the null hypothesis**.

Interpretation:

There is a **significant difference in mean ratings** among genres. Viewers rate some genres (e.g., Documentaries, Dramas) differently from others (e.g., Horror, Reality TV).

---

7. Key Insights and Strategic Interpretation

Based on the exploratory data analysis of Netflix's 2021 content catalog, several key insights emerge that provide a strategic understanding of the company's operational direction, market penetration, and content diversification. These findings not only reflect the composition of Netflix's global library but also hint at deliberate business strategies that align with consumer preferences and international market dynamics.

7.1 Dominance of Movie Content

One of the most significant findings is that **movies account for over 70%** of all titles on Netflix. This suggests a strategic emphasis on **short-form, standalone content** that can be consumed in a single sitting. Movies often require fewer episodes, smaller budgets (compared to TV series), and faster production timelines, allowing Netflix to maintain a high turnover of fresh content.

Strategic Implication:

By prioritizing movies, Netflix can rapidly scale content offerings, experiment with diverse genres and formats, and keep user engagement high without long production lead times. This approach also caters well to mobile users and global audiences with shorter attention spans.

7.2 International Expansion and Localized Content

The dataset reveals that after the United States, **India** is the second-largest content provider on Netflix. Other top contributors include the **United Kingdom**, **Canada**, and **Japan**. This reflects Netflix's **localization strategy**, which aims to appeal to regional tastes while maintaining a global brand presence.

Strategic Implication:

Netflix's aggressive investment in **non-English and regional content** demonstrates its understanding that future subscriber growth lies outside the saturated U.S. market. Tailored content—such as Indian dramas, Korean thrillers, and Latin American documentaries—helps boost subscriptions in emerging markets.

7.3 Growth Timeline and COVID-19 Disruption

Content additions peaked between **2016 and 2019**, aligning with Netflix's global expansion and its increased focus on original productions. A noticeable dip in 2020 and 2021 likely

reflects the **impact of the COVID-19 pandemic** on global film and television production pipelines.

Strategic Implication:

While Netflix adapted well by releasing previously completed content and acquiring independent films, the temporary production slowdown may have affected its release schedule. However, the rebound in 2022–2023 (not shown in this dataset) likely resumed the upward trend.

7.4 Genre Diversification and Audience Segmentation

Popular genres include **Drama**, **Documentary**, **Comedy**, and **International TV Shows**. These genres appeal across cultural and demographic boundaries and are often cost-effective to produce.

Strategic Implication:

Netflix strategically balances **mainstream entertainment** with **niche storytelling**, enabling it to target multiple audience segments. This genre diversity enhances personalization, a key factor in user retention, and supports Netflix's recommendation algorithm.

7.5 Preference for Short Durations and Single Seasons

The majority of movies have a duration between **60 and 120 minutes**, while TV shows are often limited to **one season**. This reflects a tendency toward **limited series formats** and one-off stories.

Strategic Implication:

Netflix leverages **limited series and concise formats** to reduce risk, test market responses, and streamline content development. Single-season shows allow flexibility in decision-making for renewals, based on viewership data.

7.6 Mature Audience Focus

Ratings such as **TV-MA**, **TV-14**, and **R** dominate the dataset, indicating that Netflix produces a significant amount of content for mature audiences.

Strategic Implication:

Netflix is positioning itself as a **premium entertainment provider** for adults, distinguishing its brand from family-friendly platforms like Disney+. However, it still maintains a catalog of PG and G-rated content for broader household appeal.

7.7 Content Drop Patterns and User Behavior

Content additions spike around **October, July, and December**, which may correspond to global holiday periods, summer breaks, and the year-end holiday season.

Strategic Implication:

Netflix aligns its release strategy with **peak viewing periods**, ensuring that major content drops coincide with school vacations and festive holidays when viewer engagement is highest.

7.8 Director and Actor Concentration

Although limited by missing data, a few directors and actors appear multiple times across different titles. This suggests that Netflix maintains ongoing relationships with certain creators, facilitating recurring collaborations.

Strategic Implication:

By investing in trusted creative partners, Netflix ensures consistent content quality and brand association, while enabling creators to experiment within a supportive ecosystem.

7.9 Platform Optimization Potential

The presence of older titles, low-duration content, and single-season shows also suggests there may be room for **library optimization**—such as retiring underperforming titles or licensing them to regional platforms.

Strategic Implication:

Strategic curation can help reduce content clutter, improve search accuracy, and ensure users discover high-value titles more easily.

8. Conclusion

This comprehensive analysis of Netflix's global title catalog—based on its 2021 dataset—reveals the platform's strategic focus on scale, regional expansion, and content diversification. By examining over 8,000 titles, the report has uncovered major trends in content type, genre popularity, geographic distribution, release timelines, and audience targeting.

Key conclusions include:

- **Dominance of Movies**: Over 70% of the content consists of movies, reflecting Netflix's emphasis on short-format, quickly consumable content.
- **U.S. and India as Key Content Hubs**: These two countries contribute the majority of the catalog, highlighting both Netflix's domestic strength and its investment in international markets.
- **Peak Growth from 2016–2019**: A significant increase in content additions during this period aligns with Netflix's aggressive global expansion.
- **High Popularity of Dramas, Documentaries, and Comedies**: These genres form the core of Netflix's content offerings, appealing to a wide range of viewers.
- **Focus on Mature Audiences**: Ratings such as TV-MA and TV-14 dominate the platform, indicating a deliberate effort to capture the adult viewership segment.
- **Localized and Seasonal Content Strategies**: Content drop patterns align with global holidays and audience demand cycles.

Netflix's content strategy reflects a blend of **global scalability**, **regional customization**, and **algorithm-driven personalization**, reinforcing its position as a market leader in the OTT space.

9. Strategic Recommendations

Based on the patterns and insights identified, the following strategic actions are recommended:

9.1. Strengthen Regional Content Production

With growing contributions from countries like India, Japan, and South Korea, Netflix should continue investing in **region-specific productions** that align with cultural preferences. This also includes increasing partnerships with local creators and producers to enhance authenticity and engagement.

Rationale: Regionally resonant content boosts subscriptions in international markets and supports Netflix's multilingual expansion.

9.2. Expand Successful Genre Lines

Genres such as **drama**, **documentary**, and **international TV** have high viewer engagement. Netflix should continue to build on these themes while exploring sub-genres such as **true crime**, **biographical series**, and **sci-fi dramas**.

Rationale: Deepening existing genre categories enhances user retention and maximizes the utility of Netflix's recommendation engine.

9.3. Optimize Content Duration Strategy

Continue to develop **limited series** and **single-season formats** to maintain narrative quality while controlling production costs. Netflix can also analyze completion rates of shorter versus longer content to inform future durations.

Rationale: Viewers increasingly prefer binge-worthy but concise content due to time constraints and content overload.

9.4. Improve Catalog Search and Discovery

Use the existing metadata (genre, cast, country, duration) to refine **recommendation algorithms** and **user interface filters**, especially in non-English regions.

Rationale: Enhanced content discovery tools increase user satisfaction and reduce churn by helping users find relevant content faster.

9.5. Use Data to Inform Licensing Decisions

Analyze viewership and metadata to **retire underperforming titles** or relicense them to niche platforms. Focus on content that delivers sustained engagement.

Rationale: A leaner, higher-performing library reduces operational costs and improves the overall viewer experience.

9.6. Expand Family and Educational Content

While the catalog leans mature, there is a clear opportunity to expand offerings for **children**, **families**, and **educational programming**, especially in developing markets where family viewership is common.

Rationale: Diversifying audience demographics allows for growth in under-targeted segments and supports Netflix's image as an all-ages platform.

9.7. Monitor and Adapt to Post-Pandemic Trends

Given the temporary drop in content addition during 2020–2021, Netflix should continue tracking production recovery, content demand shifts, and regional viewing patterns to guide its post-pandemic content pipeline.

Rationale: Agile adaptation will be key in maintaining market share as global media consumption patterns continue to evolve.

10. Appendix: Code Overview

This appendix outlines the key steps and methods used in the analysis of the Netflix dataset.

10.1. Data Cleaning and Preprocessing

- **Missing Values Handling:**
    - Replaced missing values in the director column with 'Unknown'.
    - Dropped rows with missing date_added to ensure time-based analysis was accurate.
- **Date Conversion:**
    - Converted date_added column to datetime format.
    - Extracted year_added and month_added for trend analysis.

10.2. Genre Normalization

- The listed_in column, which contained comma-separated genre strings, was split to allow per-genre analysis.
- Used the explode() function to flatten genres into individual rows for aggregation.

10.3. Z-Score Analysis

- Converted string-based rating to numeric format where possible.
- Applied Z-score test to detect statistical outliers based on the numeric rating data.

10.4. T-Test and ANOVA

- **T-Test:**
    - Performed to compare average ratings between Movies and TV Shows.
- **ANOVA Test:**
    - Conducted across multiple genres to determine if there were statistically significant differences in average ratings.

10.5. Visualizations

- **Content Type Distribution:**
    - Bar plot to show the proportion of Movies vs. TV Shows.
- **Top Directors:**

    - Horizontal bar chart of the top 10 directors by content count, excluding 'Unknown'.

- **Year-wise Genre Trends:**
    - Line chart showing the addition of top 5 genres across years, derived from year_added.

Source:

1. Netflix logo
2. Dataset