

# **Netflix Dataset Analysis Report**

**Prepared by : R.R.Pragadeeshwari**

**Date : July 17, 2025**

# Introduction

The Netflix dataset contains information about movies and TV shows available on the Netflix streaming platform, including details such as titles, directors, genres, release years, ratings, and countries of production. The dataset provides an opportunity to analyze Netflix's content library and understand patterns in content production and release.

The purpose of this analysis is to explore trends in Netflix content, such as the distribution of movies versus TV shows, popular genres, and the number of releases over time. We aim to answer key questions such as:

- How has Netflix's content library grown over the years?
- What types of content (movies or TV shows) dominate Netflix?
- Which genres are most common on Netflix?
- Who are the top directors contributing to Netflix content?

This analysis will help in understanding Netflix's content strategy and viewer preferences.

## Data Overview

The Netflix dataset contains 8,807 rows and 12 columns, providing detailed information about the streaming content available on the platform.

Key columns in the dataset include:

- `show_id`: Unique identifier for each show or movie
- `type`: Indicates whether the content is a Movie or TV Show
- `title`: Name of the movie or show
- `director`: Name(s) of the director(s)
- `cast`: List of main actors
- `country`: Country or countries where the content was produced
- `date_added`: Date the content was added to Netflix
- `release_year`: Year the content was originally released
- `rating`: Content rating (e.g., PG, TV-MA)
- `duration`: Length of the movie or number of seasons for TV shows
- `listed_in`: Genre categories

During the data cleaning process, some issues were identified:

- Missing values in columns such as *director* and *country*
- Some entries contained multiple comma-separated values that required splitting for analysis
- Occasional inconsistencies in formatting and spelling

These data quality considerations were addressed through appropriate cleaning steps before analysis.

## Data Cleaning and Preparation

Before performing the analysis, the dataset underwent several cleaning and preparation steps to ensure accuracy and usability:

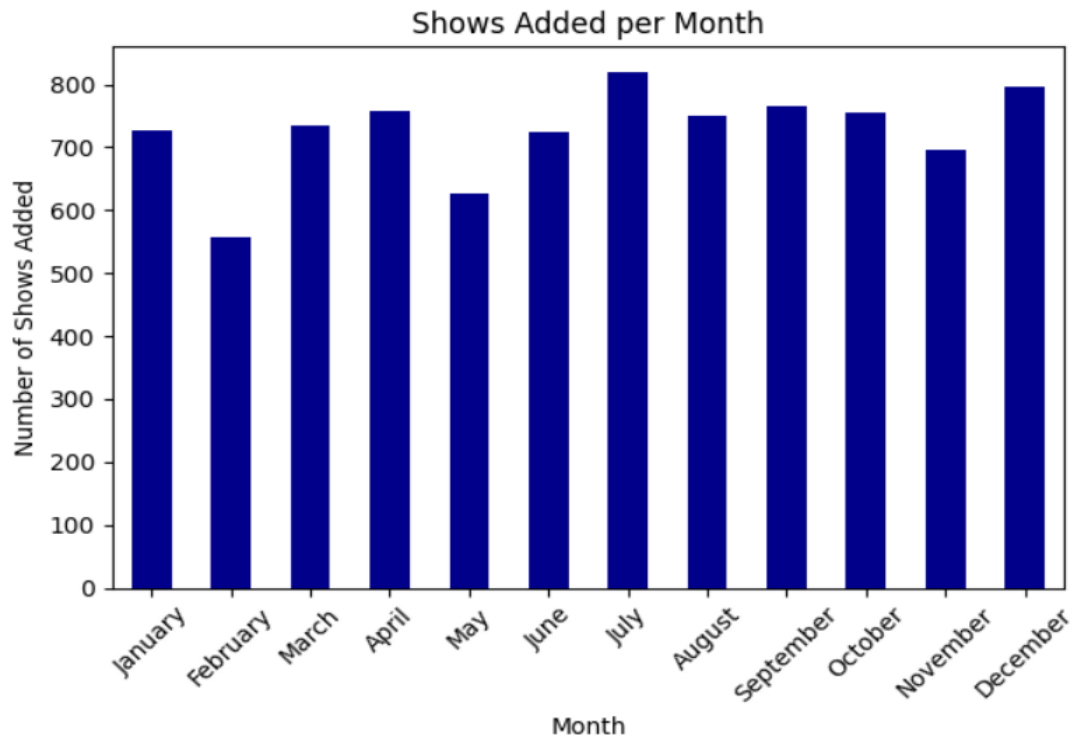
- Handling Missing Values: Missing data in columns like *director* and *country* were carefully reviewed. Rows with missing critical information were either filled with appropriate placeholders or excluded depending on the analysis requirement.
- Splitting Comma-Separated Values: Columns such as *country*, *cast*, and *listed\_in* contained multiple values separated by commas. These were split into individual entries using string operations, and the data was expanded using the `explode` method to facilitate granular analysis.
- Stripping Whitespace: After splitting, extra spaces were removed from the values to maintain consistency.
- Converting Data Types: The *date\_added* column was converted from string to datetime format to allow time-based analysis.
- Creating New Columns: No additional columns were created in this analysis, but the dataset was reorganized to improve clarity and ease of analysis.

These steps ensured the data was clean, consistent, and ready for exploratory data analysis.

## Exploratory Data Analysis (EDA)

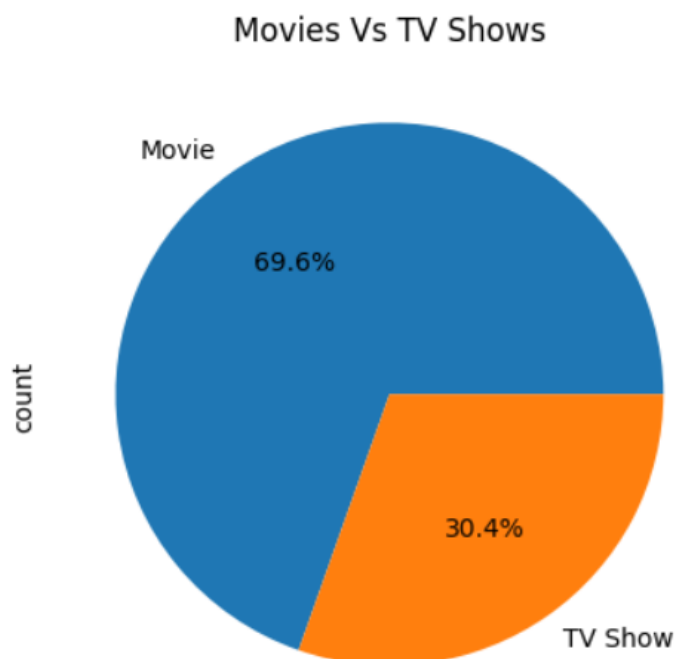
### Bar Chart of Shows Added Per Month

This bar chart displays the number of shows added to Netflix each month. It reveals seasonal trends and highlights months with the highest content additions, indicating Netflix's content release strategy over the calendar year.



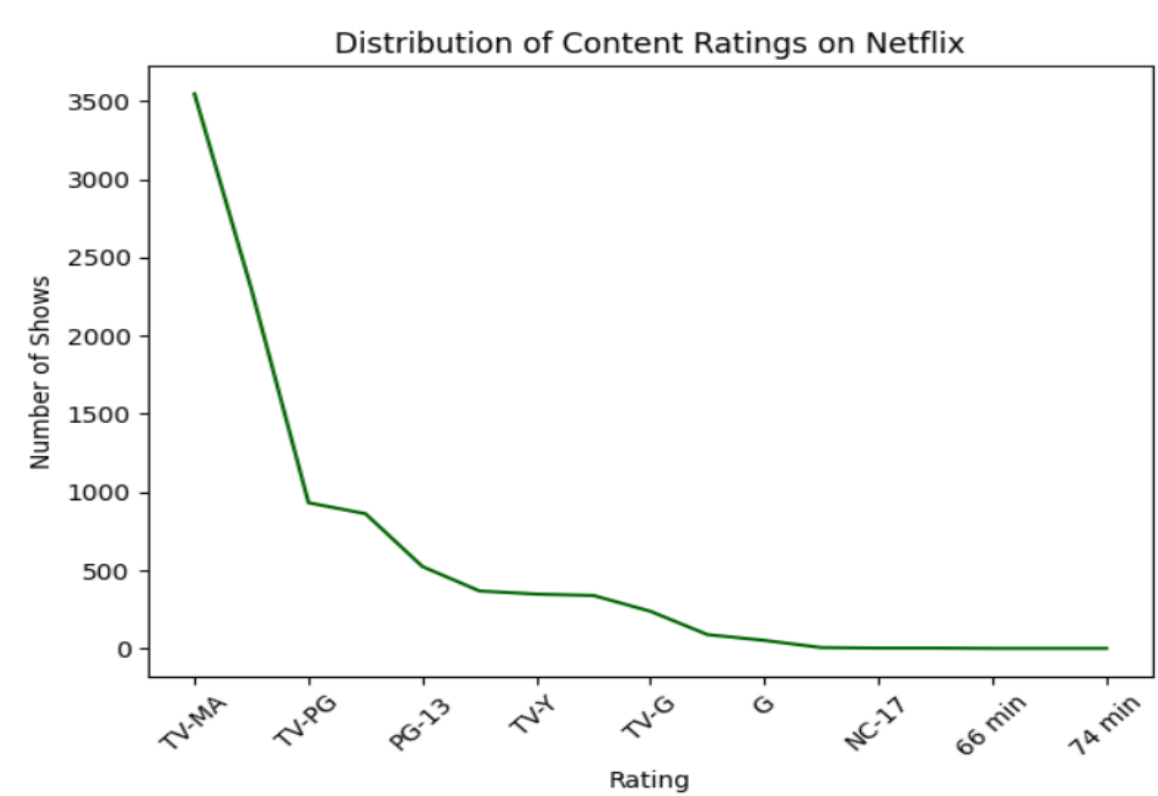
## Pie Chart of Movies vs. TV Shows

This pie chart compares the proportion of Movies and TV Shows in the dataset. It shows that movies make up a larger share of Netflix's content, although TV Shows also form a significant part of the library.



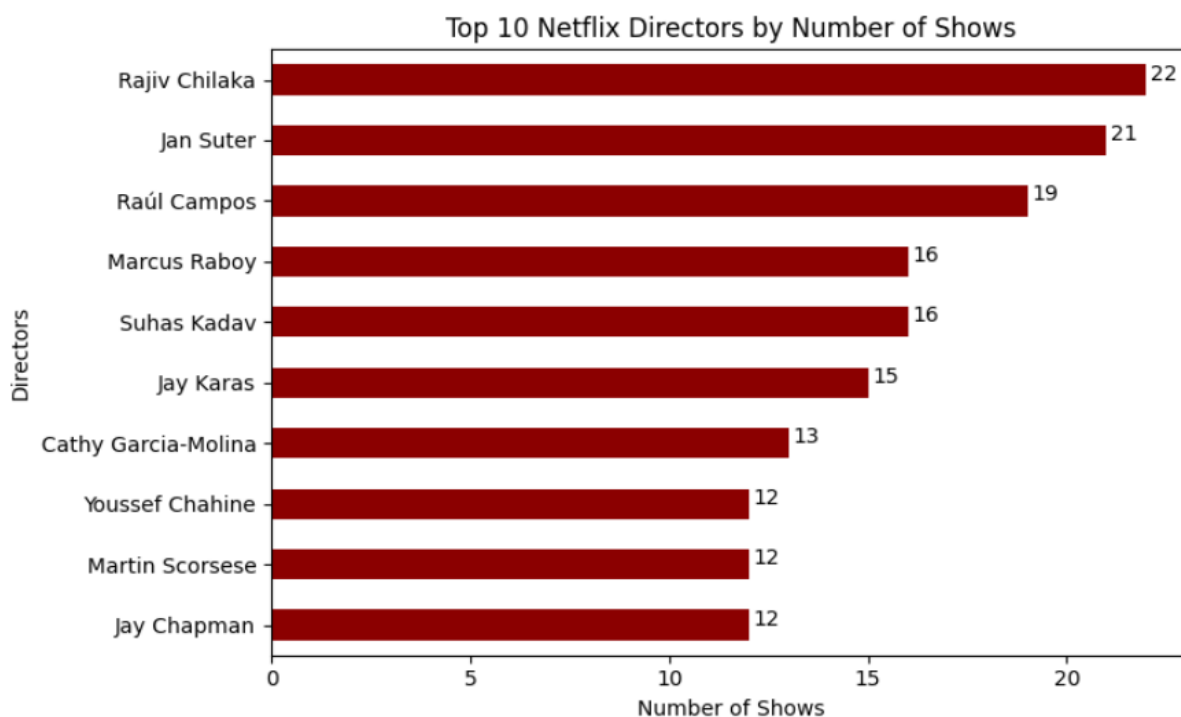
## Line Graph of Number of Releases by Rating Over Time

The line graph illustrates how the number of Netflix titles, categorized by content rating (e.g., PG, TV-MA), has changed over the years. This highlights growth trends and shifts in the types of content Netflix offers.



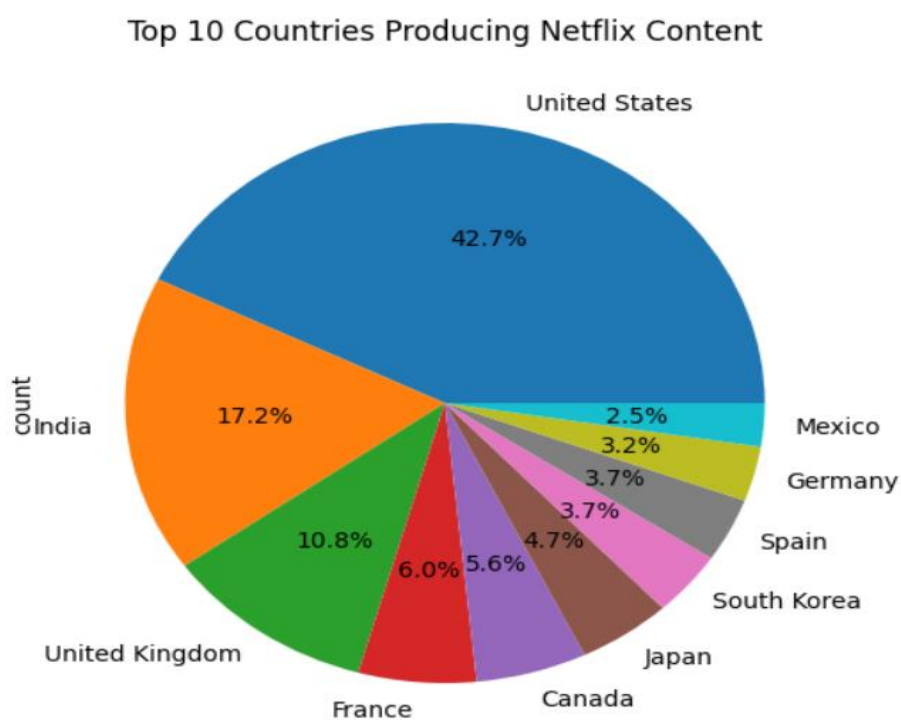
## Horizontal Bar Chart of Top 10 Directors by Number of Shows

The horizontal bar chart ranks the top 10 directors based on how many shows they have on Netflix. This shows which directors are most prolific on the platform and may indicate popular or frequently featured creators.



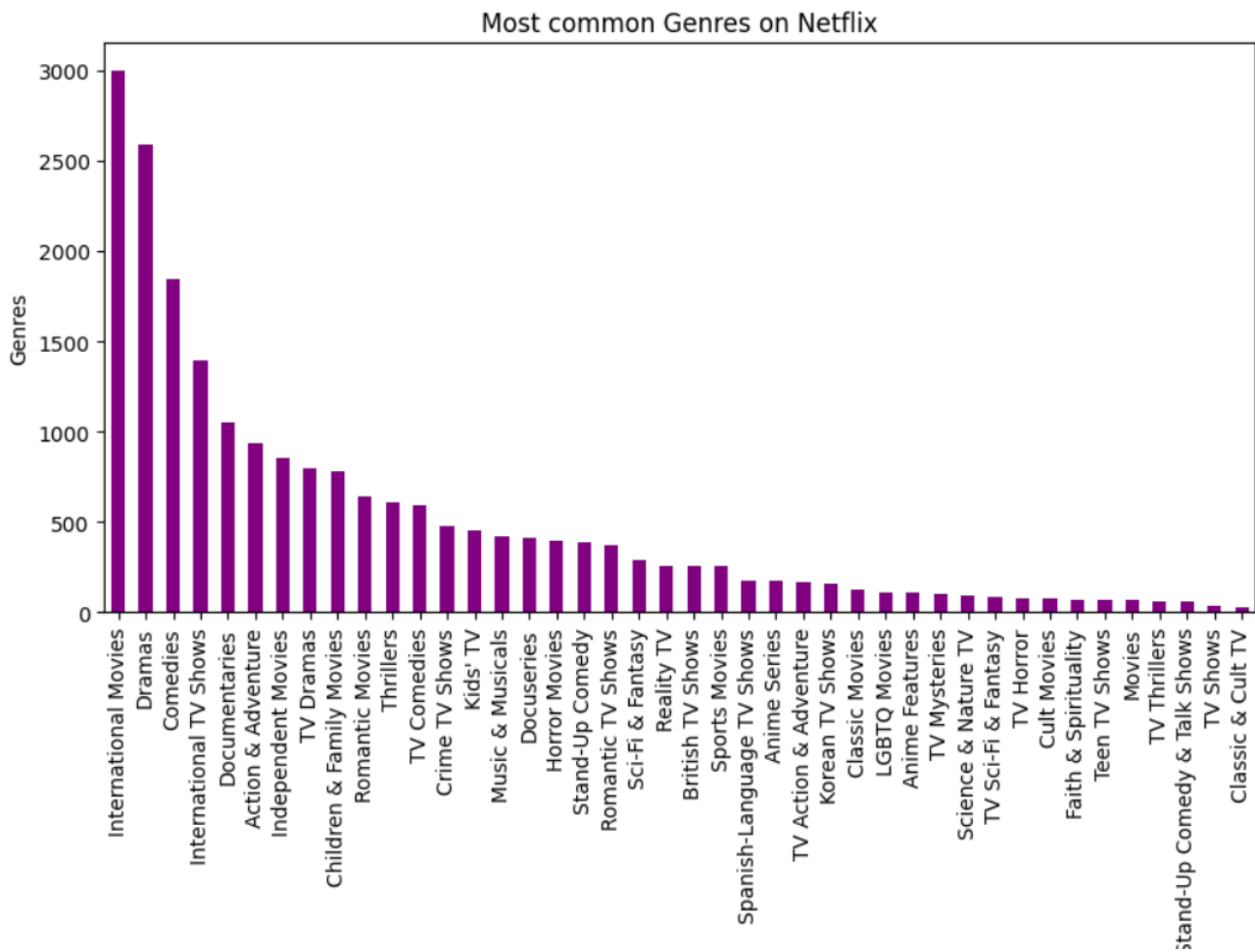
## Pie Chart of Top 10 Countries by Number of Titles

This pie chart shows the top 10 countries contributing the most titles to Netflix. The United States leads significantly, followed by India, the United Kingdom, and Canada. This reflects Netflix's strong presence in English-speaking markets and growing international reach.



# Bar Chart of Most Common Genres on Netflix

This bar chart displays the most frequent genres listed in Netflix’s content library. *International Movies* and *Dramas* top the list, indicating Netflix’s focus on diverse and emotionally engaging storytelling. Other popular genres like *Comedies* and *Documentaries* highlight the platform’s wide appeal to different audiences.



## Insights

From the analysis of the Netflix dataset, several important insights emerged:

- Content release trends over time indicate clear monthly patterns, suggesting strategic scheduling by Netflix to maximize audience engagement during peak periods.

- The distribution of content ratings shows Netflix's emphasis on broad audience accessibility, with a focus on general and family-friendly categories.
- The content type split reveals a slightly higher availability of movies compared to TV shows, highlighting Netflix's strong investment in film content.
- Analysis of directors demonstrates that a small group of prolific creators dominate Netflix's library, potentially influencing content style and audience retention.
- Genre analysis confirms that International Movies and Dramas constitute the bulk of Netflix's offerings, reflecting the platform's commitment to diverse and emotionally rich storytelling.
- Geographical distribution of content underscores the United States as the primary content producer, while the presence of countries like India and the UK signals ongoing global expansion and localization efforts.

These insights reveal Netflix's strategic approach to content diversity and growth, helping the platform stay competitive and appeal to a wide audience worldwide.

## Conclusion

The analysis of the Netflix dataset provides valuable insights into the platform's content strategy and trends. Netflix's growing content library, with a strong emphasis on movies and a significant share of TV shows, demonstrates its commitment to catering to diverse audience preferences.

The increasing number of releases over the years reflects Netflix's aggressive investment in original and licensed content. Additionally, the identification of top directors shows the key contributors shaping Netflix's offerings.

Overall, this analysis highlights how Netflix continues to evolve its content to maintain its position as a global streaming leader, focusing on both quantity and variety to attract and retain viewers worldwide.