

Exploratory Data Analysis (EDA) on Netflix Movies and TV Shows Dataset



Exploratory Data Analysis (EDA) on Netflix Movies and TV Shows Dataset

Prepared by: Atharva Kolhe

Date: 12/01/2025

Platforms: [LinkedIn](#) | [GitHub](#)

Introduction

This report presents an exploratory data analysis (EDA) of a Netflix dataset, providing insights into the platform's content library. This project leverages Python's data analysis libraries (Pandas, Matplotlib, Seaborn) and Jupyter Notebook as the development environment to explore and analyze the dataset effectively. The dataset contains information about movies and TV shows available on Netflix as of mid-2021. The analysis explores various aspects such as content type distribution, temporal trends, genre popularity, geographic origins, ratings, and key contributors (cast and directors).

Dataset Overview

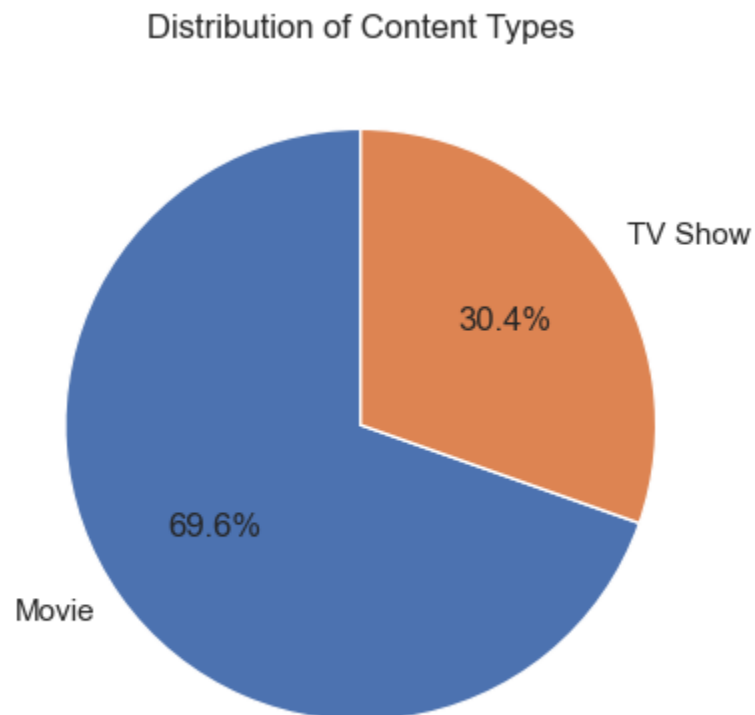
- **Source:** The dataset was obtained from <https://www.kaggle.com/datasets/shivamb/netflix-shows/data>
 - **Scope:** Limited to Netflix content available as of mid-2021.
 - **Features:** The dataset includes the following columns:
 - show_id: Unique identifier for each title.
 - type: Indicates whether the title is a "Movie" or "TV Show."
 - title: The name of the content.
 - director: Director(s) of the content.
 - cast: Cast members featured in the content.
 - country: Country of origin.
 - date_added: The date the content was added to Netflix.
 - release_year: The year the content was released.
 - rating: Content rating (e.g., PG, R).
 - duration: Duration of movies (in minutes) or TV shows (number of seasons).
 - listed_in: Categories or genres.
 - description: Brief description of the content.
-

Key Insights from the EDA

1. Distribution of Content Types

The dataset contains the following distribution of content types:

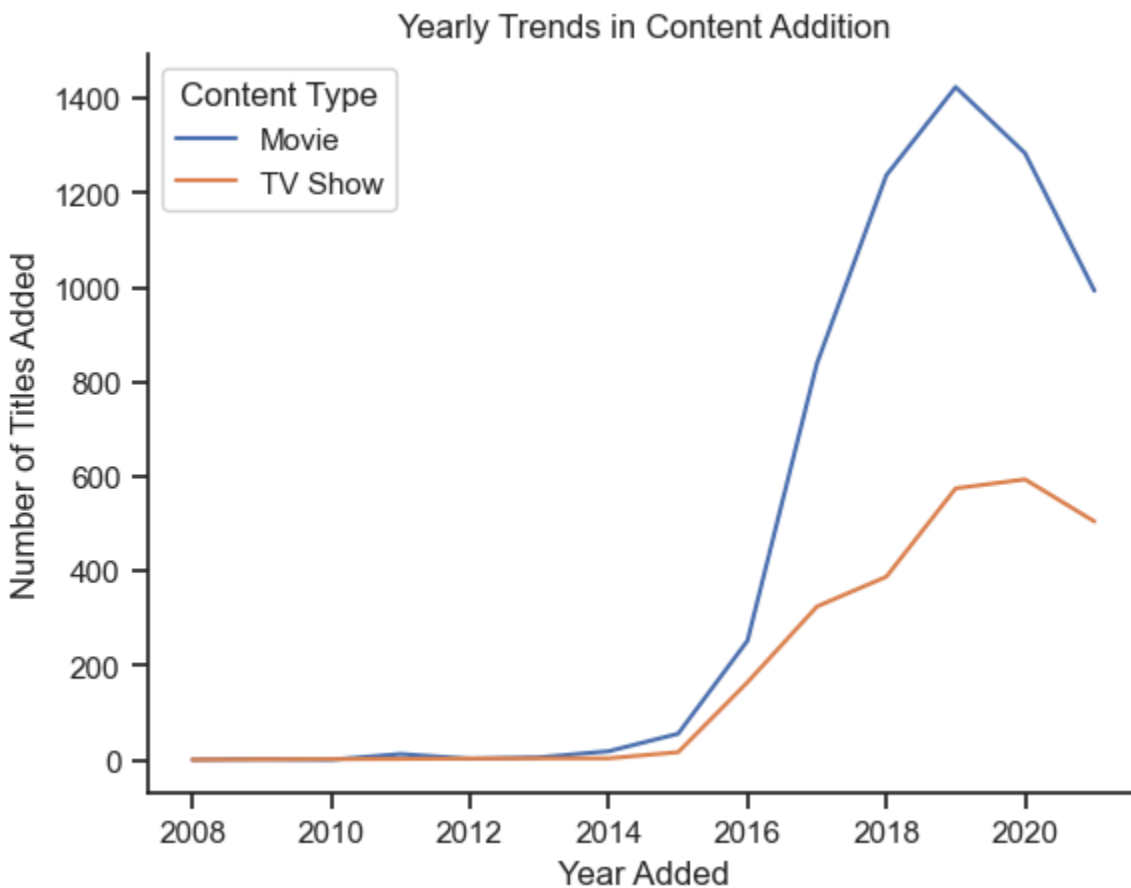
- Movies: 69.6 %
- TV Shows: 30.4 %



2. Yearly Trends in Content Addition

A line chart was generated to analyze how Movies and TV Shows were added to Netflix over the years. This revealed key trends, such as:

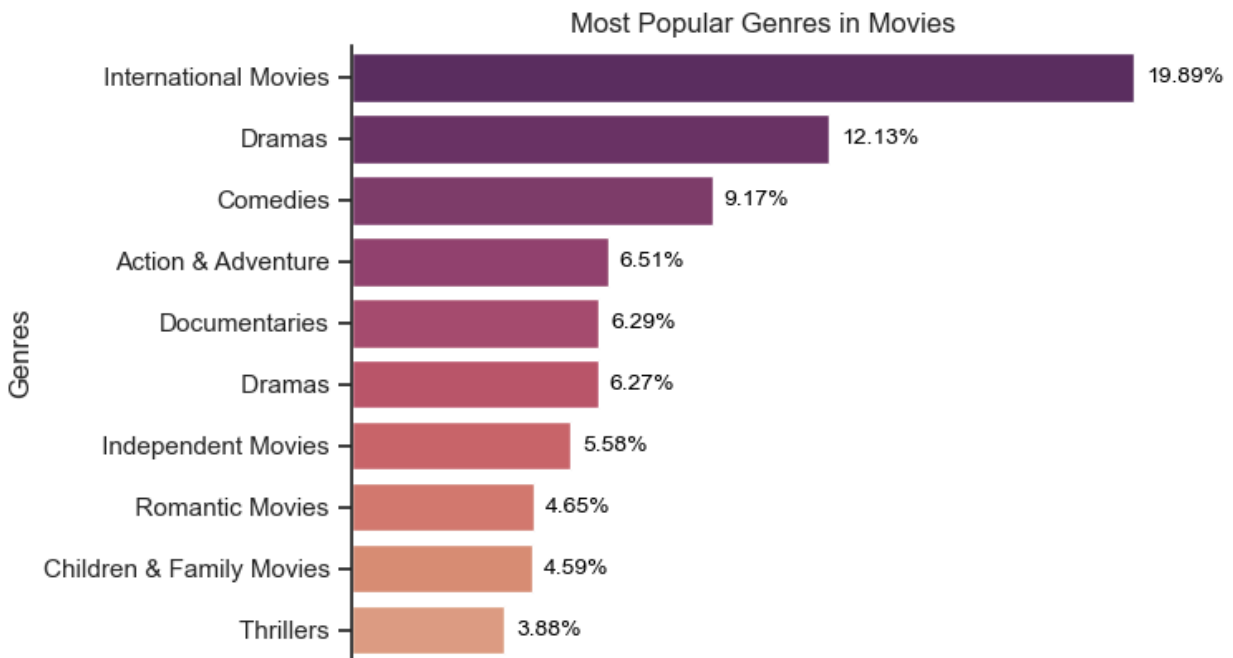
- A significant increase in content addition from 2016.
- The ratio of Movies to TV Shows added annually.



3. Genre Analysis

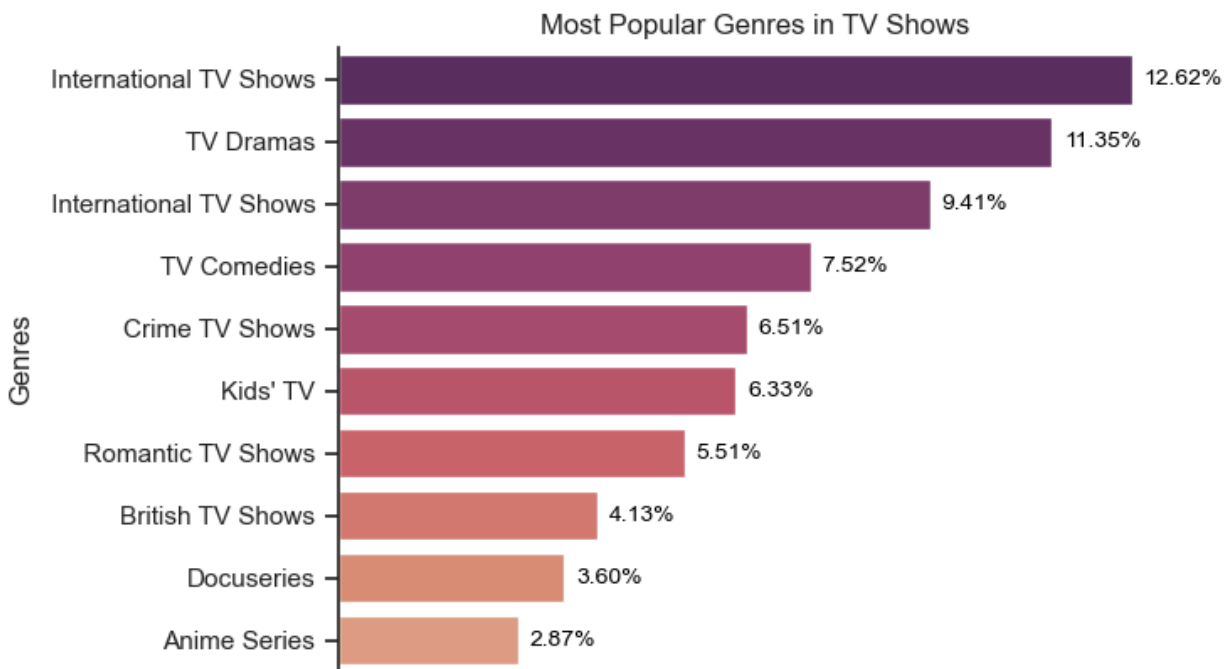
The most popular genres for Movies and TV Shows were identified and visualized.

Movies:



- International movies are the most popular genre, accounting for 19.89% of movies.
- Dramas come in second with 12.13%, followed by Comedies at 9.17%. Action & Adventure, Documentaries, and Independent Movies are also well-represented.
- Romantic Movies, Children & Family Movies, and Thrillers are the least popular genres shown, each comprising less than 5% of movies.

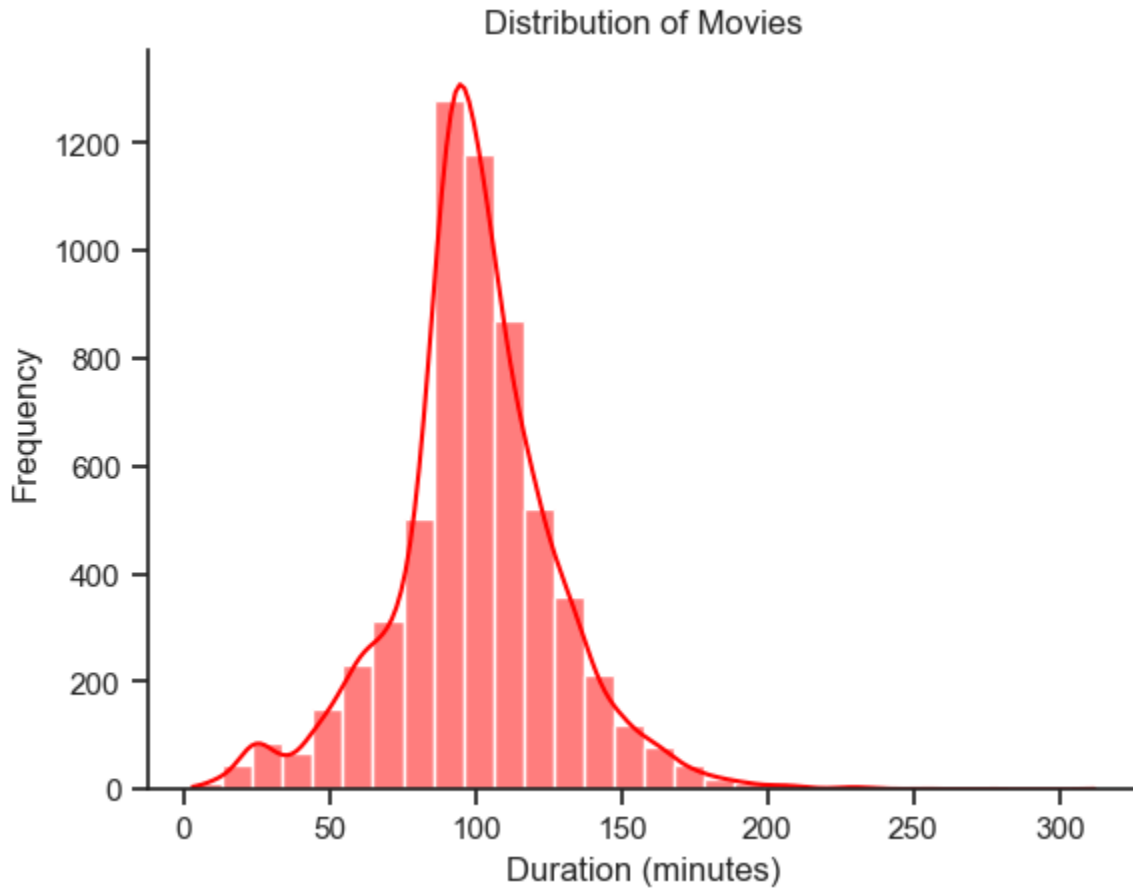
TV Shows:



- International TV Shows are the most popular genre, accounting for 12.62% of TV shows.
- TV Dramas come in second with 11.35%, followed by International TV Shows at 9.41%. TV Comedies, Crime TV Shows, and Kids' TV are also well-represented.
- Romantic TV Shows, British TV Shows, Docuseries, and Anime Series are the least popular genres shown, each comprising less than 5% of TV shows.

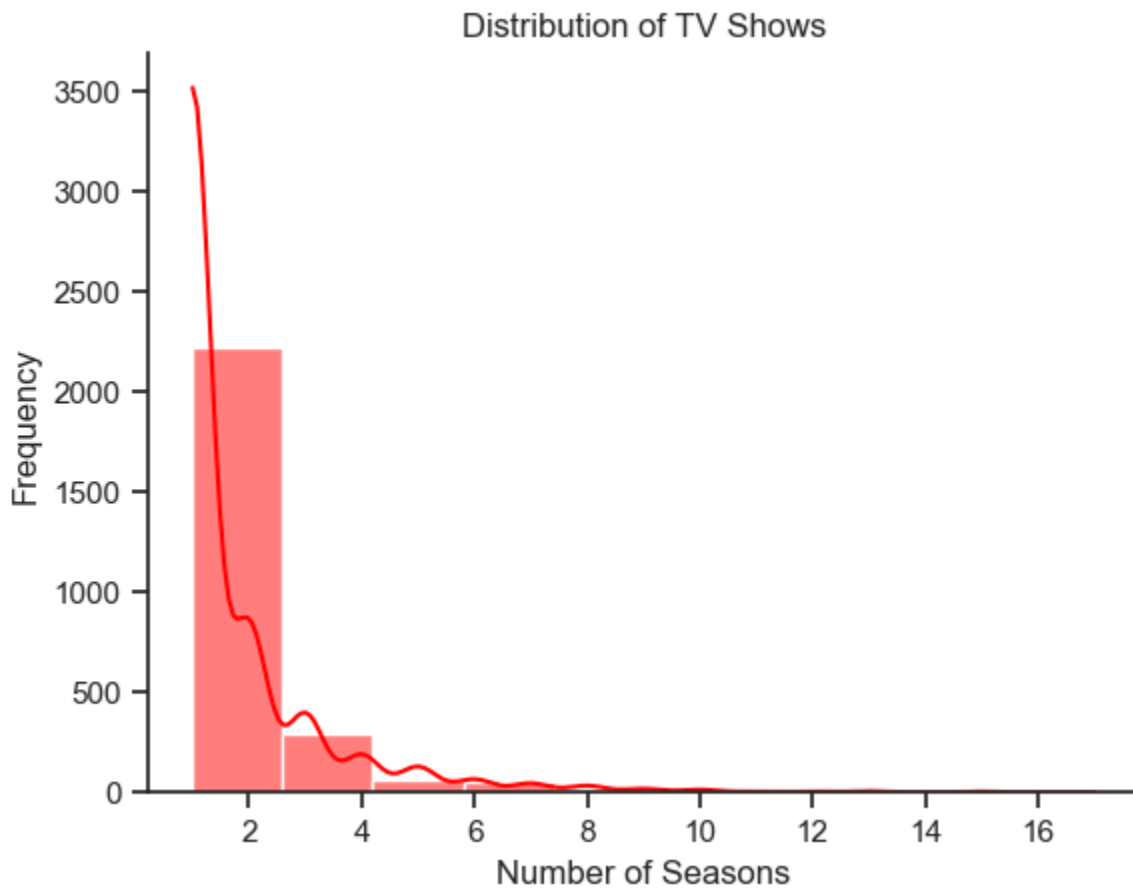
4. Duration Analysis

Movies:



- **Shape:** The distribution is right-skewed. This means there are more movies with shorter durations and fewer movies with longer durations.
- **Peak:** The peak of the distribution is around 100 minutes, indicating that the most common movie duration is around 2 hours.
- **Range:** The duration of movies ranges from approximately 0 minutes to 300 minutes.
- **Spread:** The spread of the distribution is quite wide, suggesting that there is a significant variation in movie lengths.

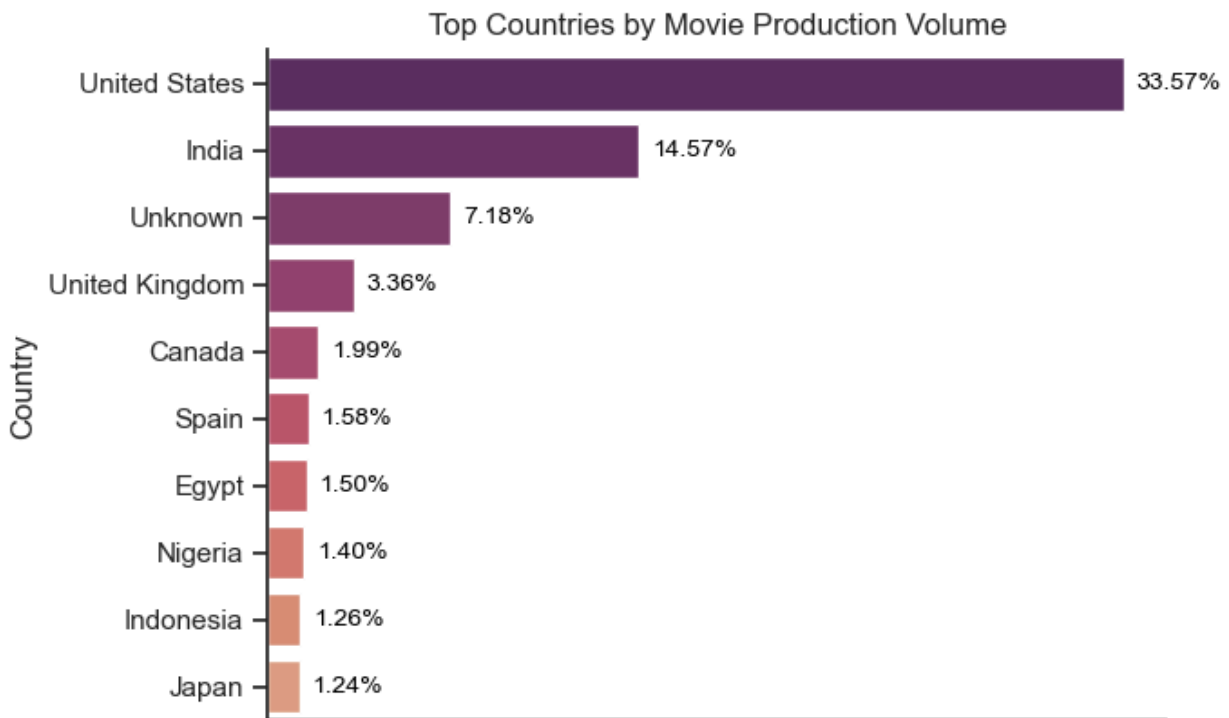
TV Shows:



- **Shape:** The distribution is right-skewed. This means there are more TV shows with fewer seasons and fewer TV shows with more seasons.
- **Peak:** The peak of the distribution is at 1 season, indicating that the most common number of seasons for a TV show is 1.
- **Range:** The number of seasons for TV shows ranges from 1 to approximately 16.
- **Spread:** The spread of the distribution is quite wide, suggesting that there is a significant variation in the number of seasons for TV shows.

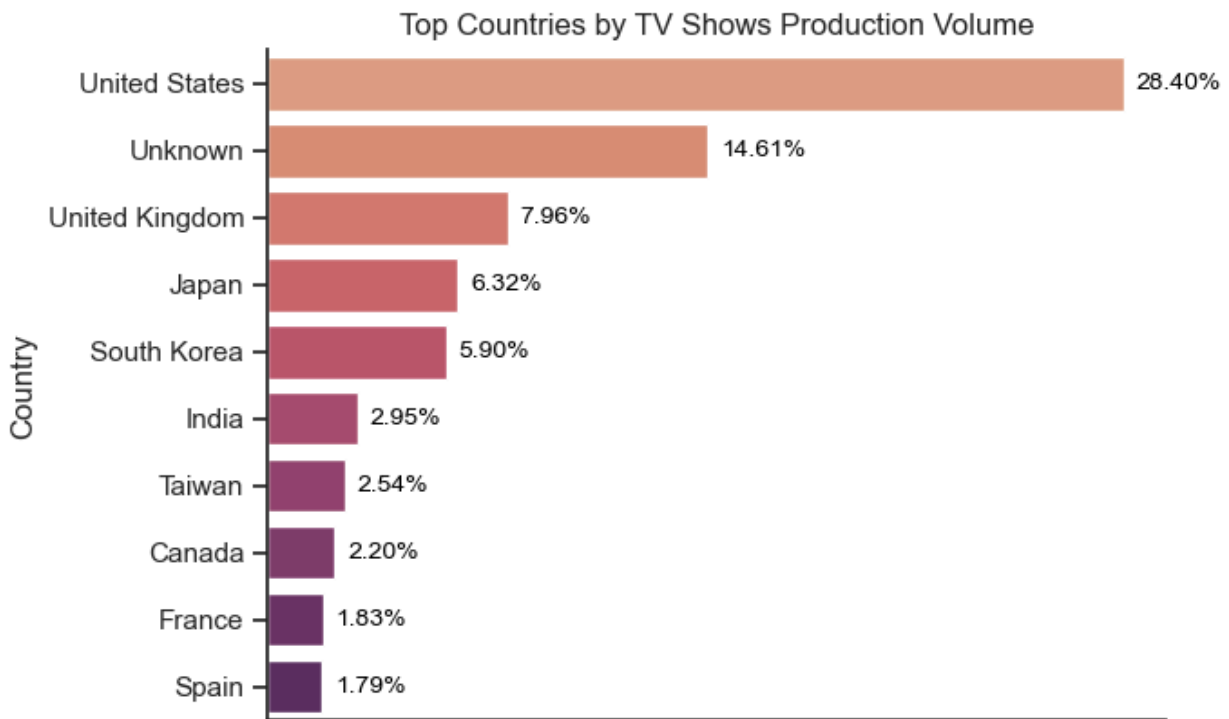
5. Top Countries in Content Production

- **Movies:**



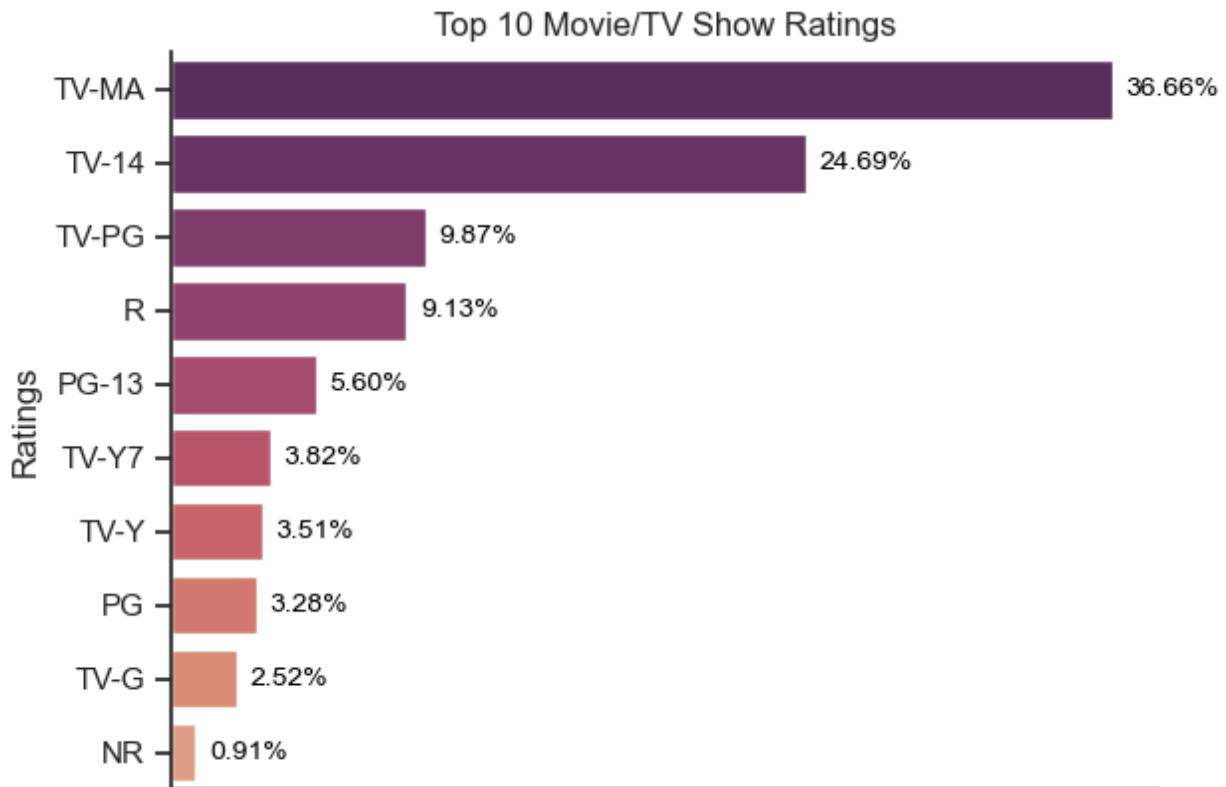
- **Country Production Volume:** The chart shows the relative movie production volume of different countries. The length of each bar corresponds to the percentage of movies produced by that country.
- **Top Producer:** The United States is the top movie producer, accounting for 33.57% of movies.
- **Other Significant Producers:** India comes in second with 14.57%, there are 7.18% Movies whose country of origin is Unknown. The United Kingdom, Canada, and Spain also have a notable share of movie production.
- **Smaller Producers:** Egypt, Nigeria, Indonesia, and Japan have smaller production volumes, each contributing less than 2% of movies.

- **TV Shows:**



- **Country Production Volume:** The chart shows the relative TV show production volume of different countries. The length of each bar corresponds to the percentage of TV shows produced by that country.
- **Top Producer:** The United States is the top TV show producer, accounting for 28.40% of TV shows.
- **Other Significant Producers:** There are 14.61% TV Shows whose country of origin is Unknown, followed by the United Kingdom at 7.96%. Japan, South Korea, India, and Taiwan also have a notable share of TV show production.
- **Smaller Producers:** Canada, France, and Spain have smaller production volumes, each contributing less than 3% of TV shows.

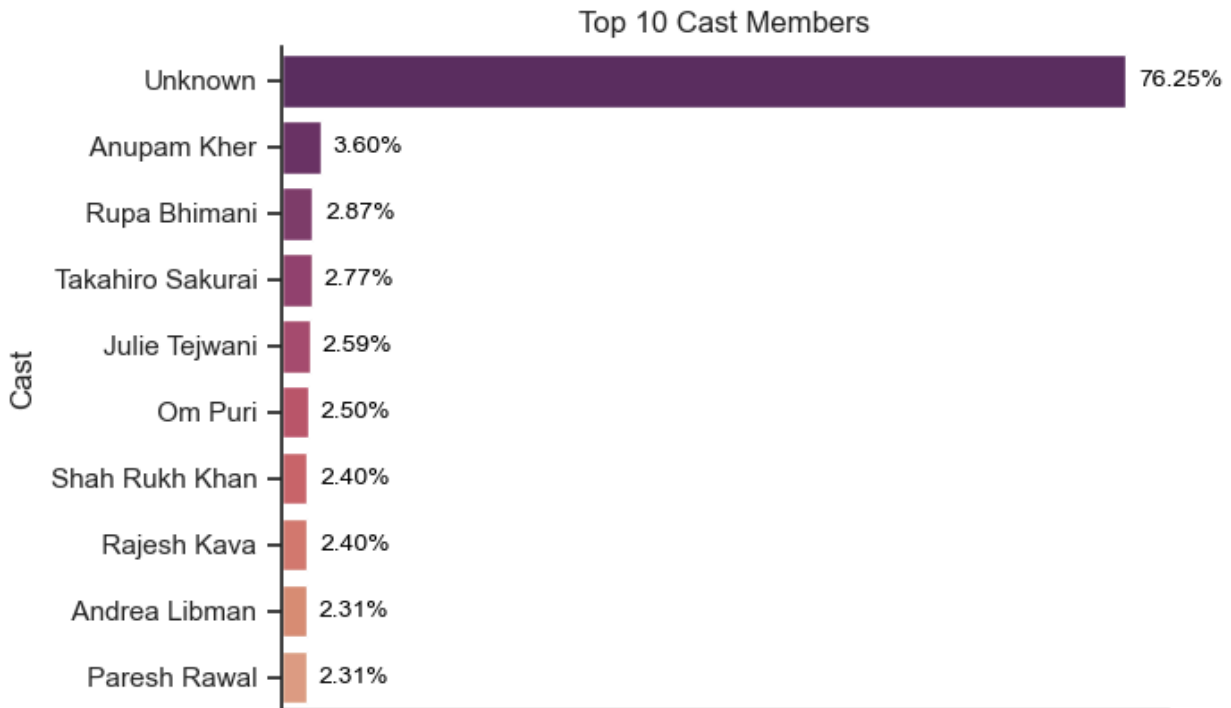
6. Content Ratings



- **Rating Distribution:** The chart shows the distribution of ratings for movies and TV shows. The length of each bar corresponds to the percentage of content that falls into that rating category.
- **Most Common Rating:** TV-MA is the most common rating, accounting for 36.66% of content.
- **Other Common Ratings:** TV-14 is the second most common rating with 24.69%, followed by TV-PG at 9.87%. R-rated content is also well-represented.
- **Less Common Ratings:** PG-13, TV-Y7, TV-Y, PG, TV-G, and NR are less common ratings, each comprising less than 5% of content.

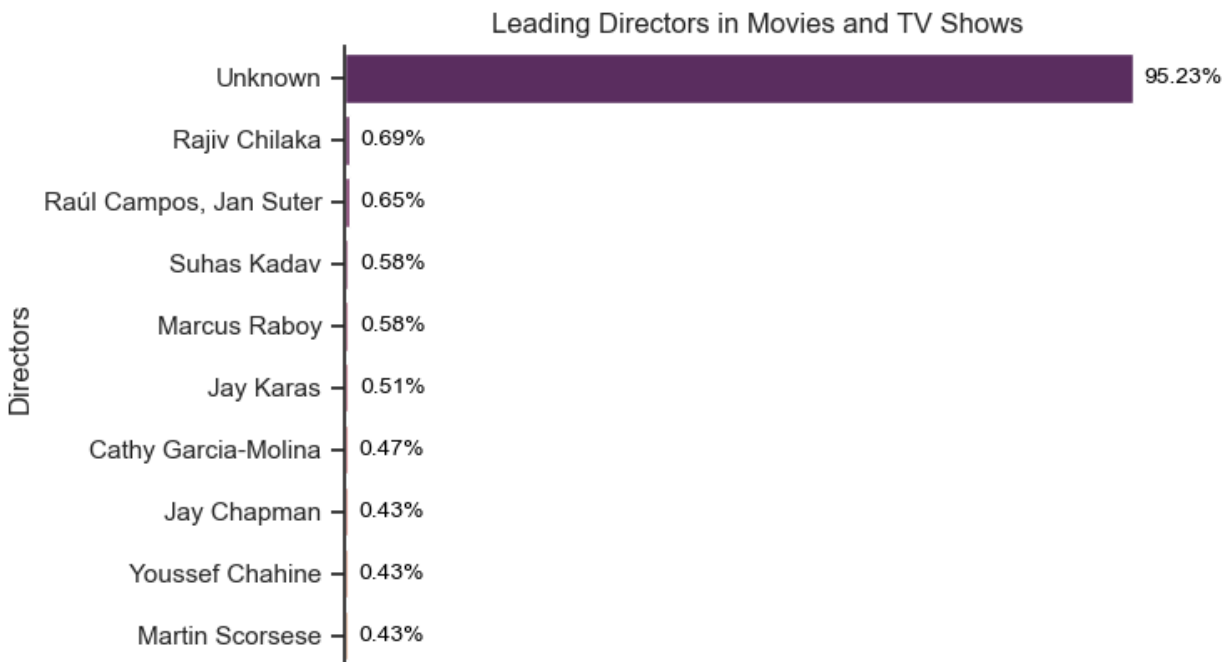
7. Frequent Cast Members and Directors

Cast Members:



- **Cast Member Frequency:** The chart shows the relative frequency of cast members appearing in movies and TV shows. The length of each bar corresponds to the percentage of content featuring that cast member.
- **Most Frequent Cast Member:** "Unknown" is the most frequent cast member, appearing in 76.25% of content. This likely indicates that cast information is not always available.
- **Other Frequent Cast Members:** Anupam Kher, Rupa Bhimani, Takahiro Sakurai, and Julie Tejjwani are also relatively frequent, each appearing in more than 2.5% of content.
- **Less Frequent Cast Members:** Om Puri, Shah Rukh Khan, Rajesh Kava, Andrea Libman, and Paresh Rawal are less frequent, each appearing in less than 2.5% of content.

Directors:



- **Director Frequency:** The chart shows the relative frequency of directors appearing in movies and TV shows. The length of each bar corresponds to the percentage of content featuring that director.
- **Most Frequent Director:** "Unknown" is the most frequent director, appearing in 95.23% of content. This likely indicates that director information is not always available.
- **Other Frequent Directors:** Rajiv Chilaka, Raúl Campos, Jan Suter, Suhas Kadav, and Marcus Raboy are also relatively frequent, each appearing in more than 0.5% of content.
- **Less Frequent Directors:** Jay Karas, Cathy Garcia-Molina, Jay Chapman, Youssef Chahine, and Martin Scorsese are less frequent, each appearing in less than 0.5% of content.

Conclusion

This analysis provides insights into the content trends on Netflix, helping understand its catalog's composition and audience preferences. The findings can be useful for content creators, analysts, and Netflix enthusiasts.

For the complete code and visualizations, visit the [GitHub Repository](#).