Exploring Netflix: Unveiling Trend, Contributors and Insights

Introduction:

Netflix is one of the most popular media and video streaming platforms. They have over 10000 movies or tv shows available on their platform, as of mid-2021, they have over 222M Subscribers globally.

In this analysis, we delve into a dataset comprising 8807 movies and TV shows to uncover trends and insights crucial for content curation and business expansion.

By leveraging data-driven insights, Netflix can make informed decisions about the types of shows and movies to produce, as well as strategize for growth across different markets.

Summary statistics of the dataset:

Jupiter Notebook Link: https://drive.google.com/file/d/1tbRryqkd-is_1AILv_qN5ZkZJ6xlMw9F/view?usp=sharing

Shape: (8807, 12) -> 8807 unique contents, each with 12 columns providing various details about it.

Columns:

- Show_id: Unique ID for every Movie / Tv Show
- Type: Identifier - A Movie or TV Show
- Title: Title of the Movie / Tv Show
- Director: Director of the Movie
- Cast: Actors involved in the movie/show
- Country: Country where the movie/show was produced
- Date_added: Date it was added on Netflix
- Release_year: Actual Release year of the movie/show
- Rating: TV Rating of the movie/show
- Duration: Total Duration - in minutes or number of seasons
- Listed_in: Genre
- Description: The summary description

Datatype and No. of Nulls in each column:

#	Column	Non Null Count	Dtype
0	show_id	8807 non null	Object
1	type	8807 non null	Object
2	title	8807 non null	Object
3	director	6173 non null	Object
4	cast	7982 non null	Object
5	country	7976 non null	Object
6	date_added	8797 non null	object
7	release_year	8807 non null	int64
8	rating	8803 non null	Object
9	duration	8804 non null	Object
10	listed_in	8807 non null	Object
11	description	8807 non null	object

No. of unique elements:

Column	Count of unique values
show_id	8807
type	2
title	8807
director	4528
cast	7692
country	748
date_added	1767
release_year	74
rating	17
duration	220
listed_in	514
description	8775

Type of shows: [Movie, TV Show]

Shows available in the range of years:

- Based on released: (1925 – 2021)
- Based on addition (year added to Netflix): (2008 – 2021)

Handling Missing Data in the Analysis:

From the given data,

- **Categorical Column**: We cannot impute specific values based on prediction. Therefore, we updated all categorical columns with 'unknown_[column_name]'.
- **Numerical column**: Imputing 0 to numerical rows would affect calculations such as mean or median. Hence, we retained all missing values as null.

Handling Columns with multiple values:

There are multiple comma-separated values within certain columns for each row.

These columns include:

- Director
- Country
- Cast
- Genre (under the column name 'listed_in')

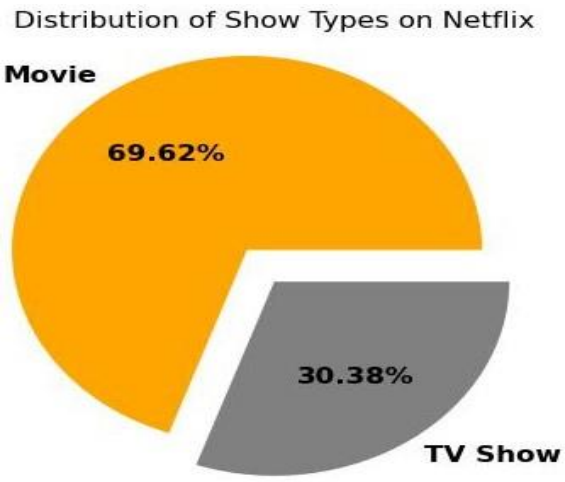
Unnesting all columns in the same table can lead to discrepancies during further analysis. Therefore, we maintain separate tables for the nested columns by unnesting and including additional necessary columns from the main dataset.

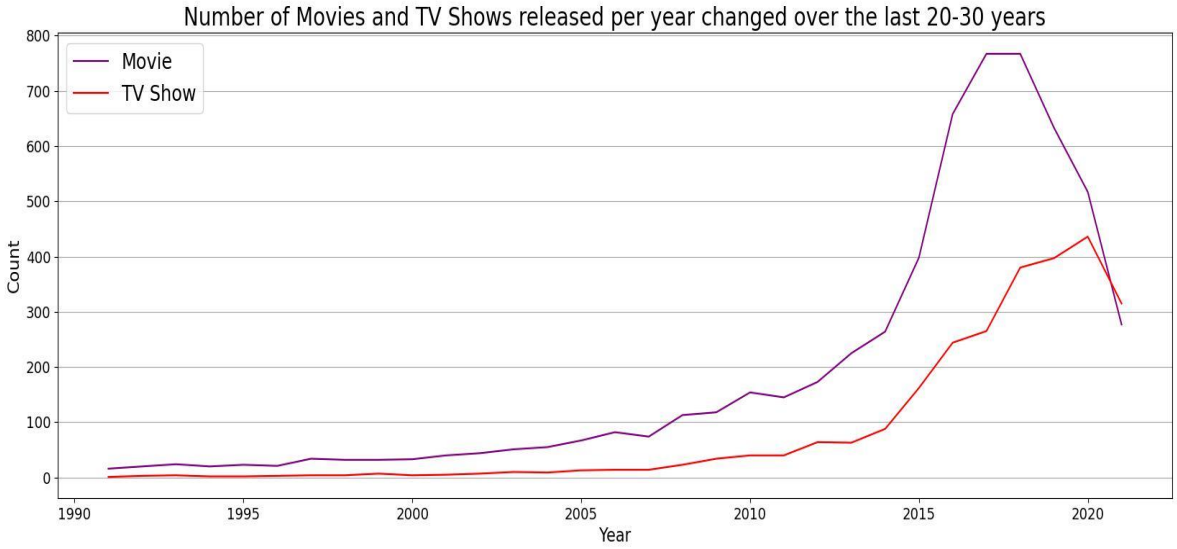
After unnesting those columns, the unique counts are,

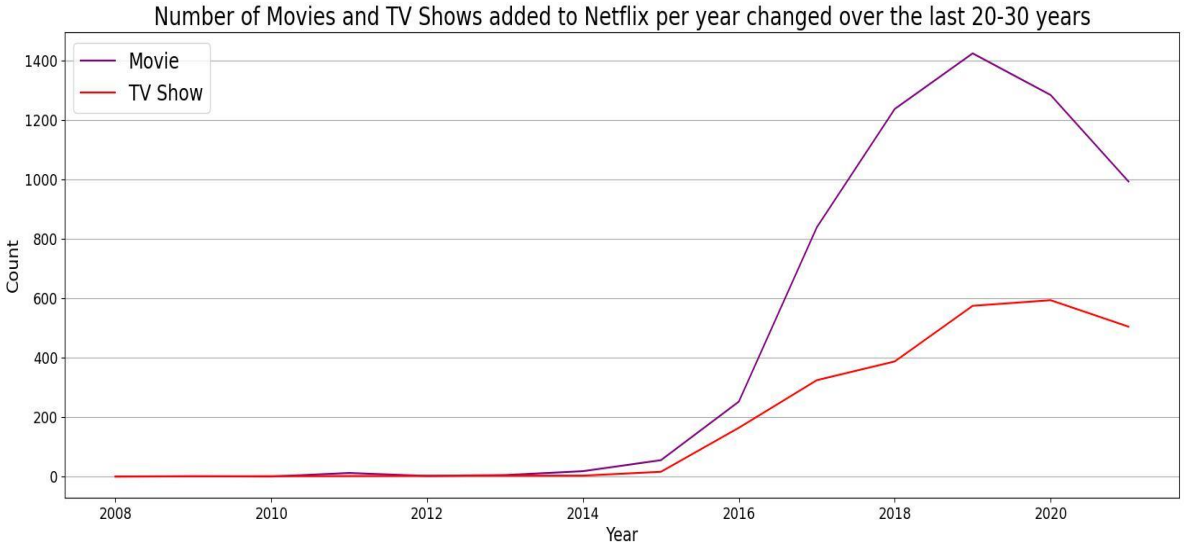
- Director : 4989 + 1(unknown_column)
- Country : 123 + 1(unknown_column)
- Cast : 36427 + 1(unknown_column)
- Genre : 4

Univariate and Bivariate Analysis:

1.Analysis based on Show type:



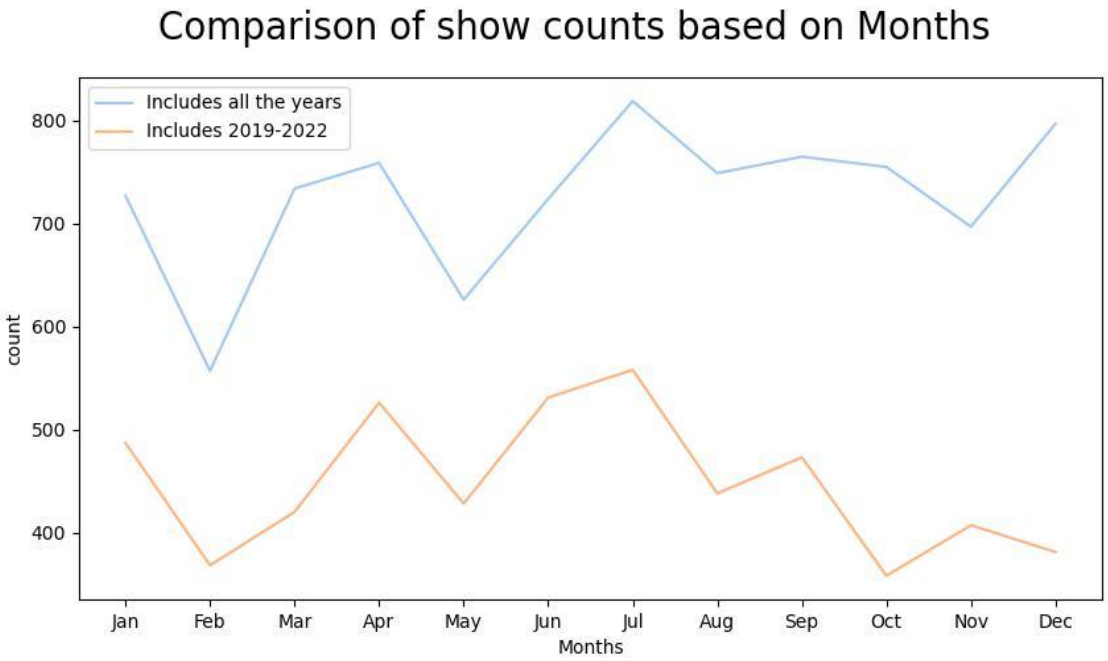




Insights:

- ✓ From 2015 to 2019, Netflix significantly increased the addition of both movies and TV shows. However, there is a noticeable decline in show additions after 2019. This drop could be attributed to the inclusion of older shows in the 2015-2019 range.
- ✓ Prior to 2015, the trend line suggests minimal show additions, indicating Netflix was in the developmental stages during that period."
- ✓ In comparison to movies, the decline in TV show additions is relatively less pronounced.

In the following plot, the blue line represents the total counts of shows added each month, while the red line illustrates the count of shows added over the period from 2019 until the latest year in the given data. The red line serves to indicate the recent trend in the addition of shows to Netflix.

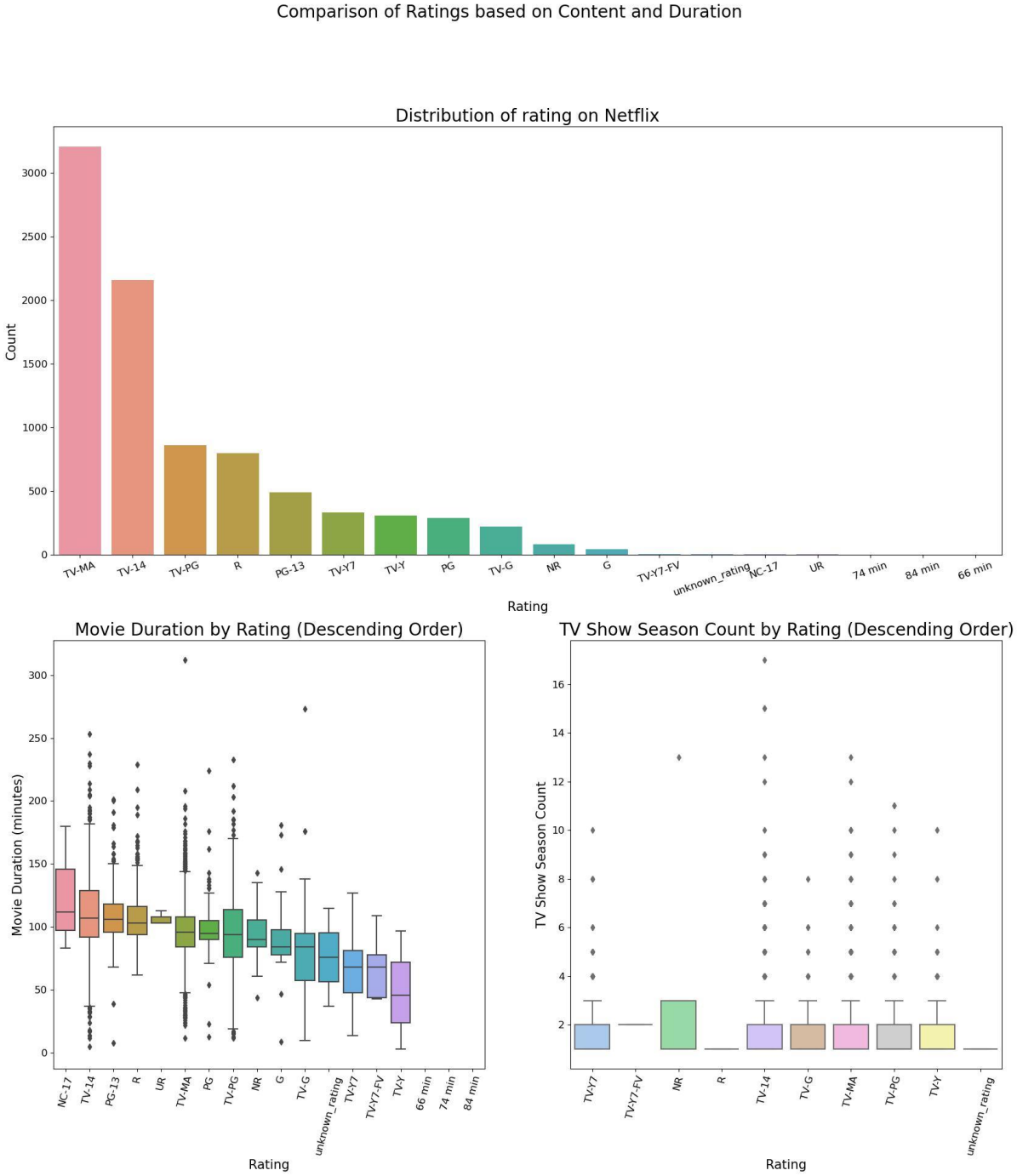


Insights:

Based on the recent trend line,

- ✓ January, April, June, and July were the peaks in adding content.
- ✓ However, February, October, and December show a decline in content addition.

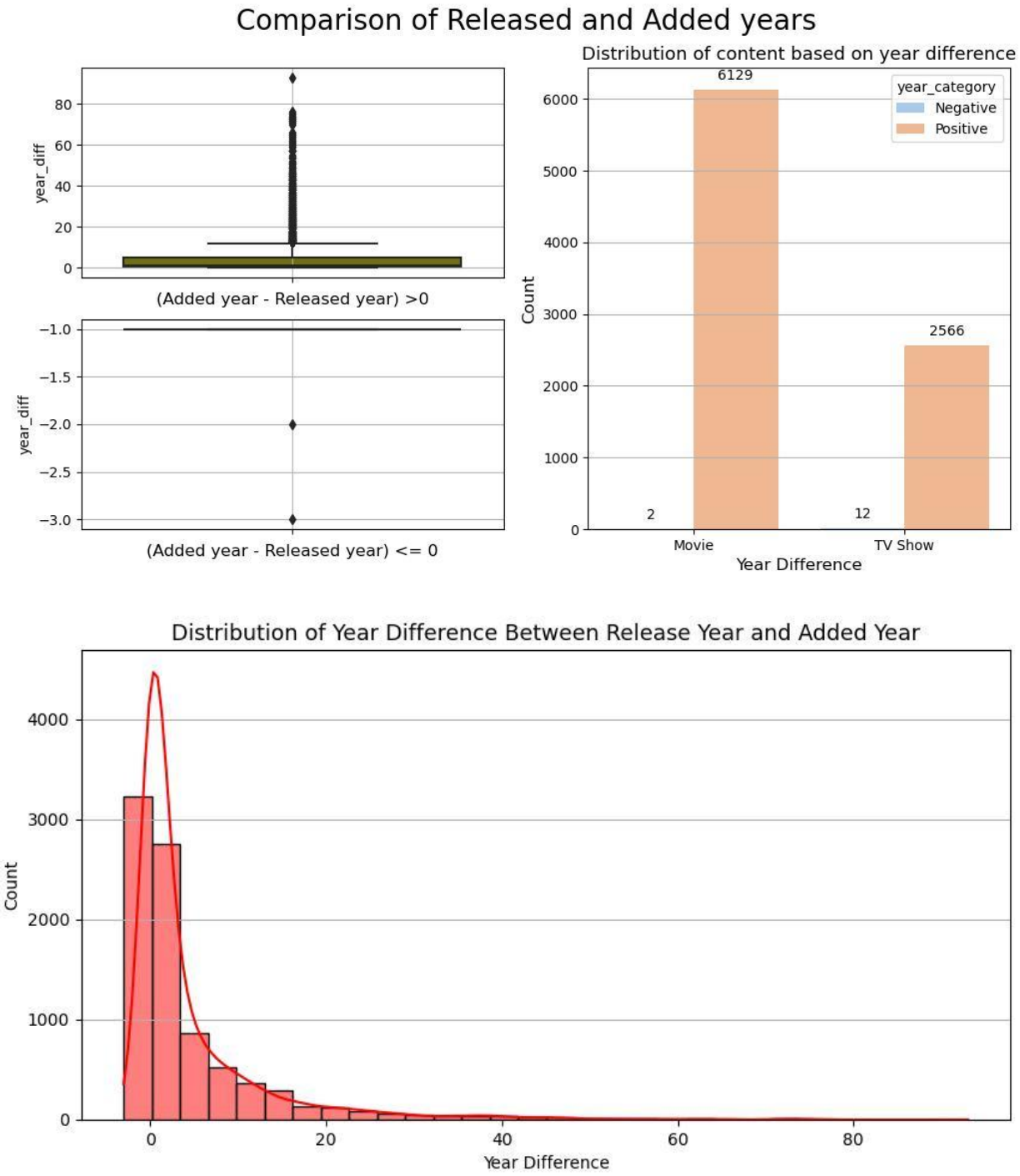
2.Analysis based on Rating:



Insights:

- ✓ In Netflix, TV-MA and TV-14 ratings each have over 3000 and 2000 shows respectively, accounting for almost 90% of the total shows.
- ✓ TV-PG and R ratings have approximately 700 shows each, showing a similar distribution.
- ✓ There is only one movie each with a rating of 66 minutes, 74 minutes, and 84 minutes respectively.
- ✓ When considering the duration of shows, NC-17 and TV-14 have the highest average duration. However, TV-MA shows exhibit a wide range of durations, with numerous outliers.
- ✓ There isn't a significant difference in the count of TV show seasons. However, TV-14 has the highest number of shows with multiple seasons.

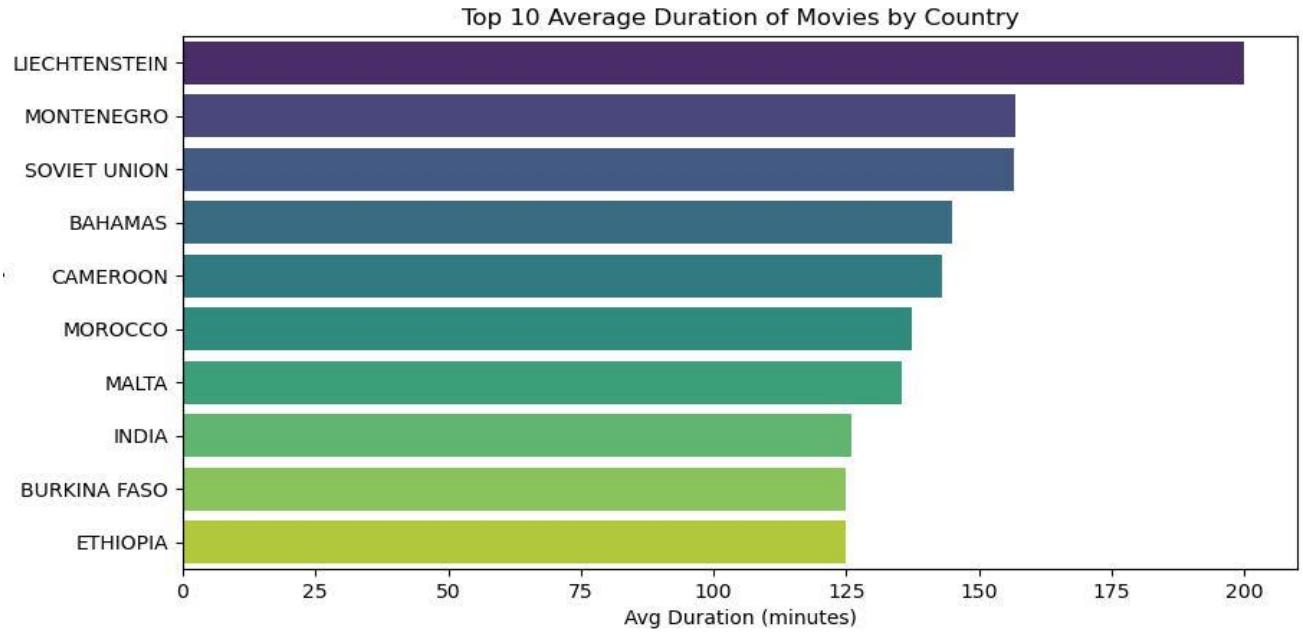
3.Analysis Based on Shows Added and Released (in Years):



Insights:

- ✓ The majority of shows on Netflix have a year difference of either 0 or 1, as indicated by the histogram.
- ✓ Interestingly, the data suggests that 12 TV shows and 2 movies were added to Netflix before their release dates.

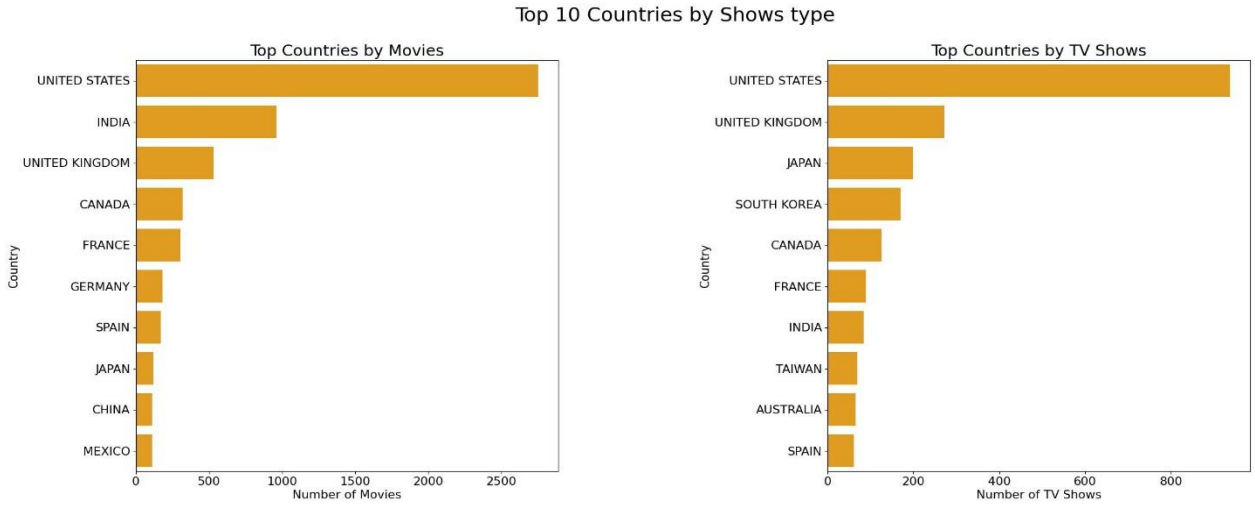
4. Top 10 Average Duration of Movies by Country



Insights:

- ✓ Liechtenstein, a European country, has the highest average duration, nearly 200 minutes.
- ✓ Montenegro and the Soviet Union hold the second place with an average time of over 150 minutes.
- ✓ More African countries secured places in the top 10 based on average duration.

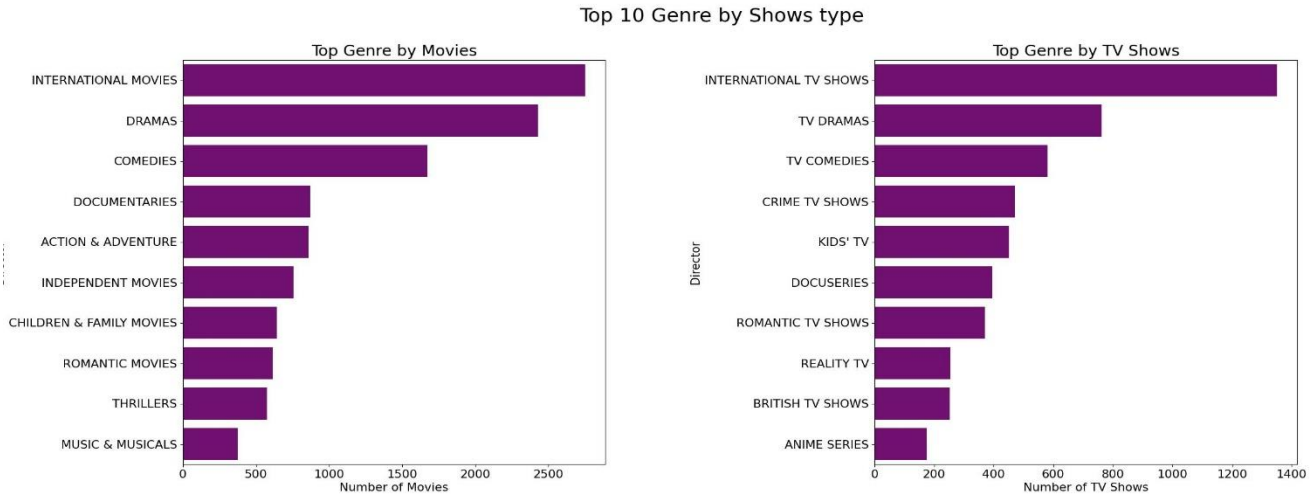
5. Top 10 Countries by Shows type



Insights:

- ✓ The United States boasts over 2500 movies and 900 TV shows, constituting almost 40% of the overall content.
- ✓ India exhibits a substantial number of movies but a lower count of TV shows.
- ✓ Meanwhile, the United Kingdom presents a reasonable balance between movies and TV shows.

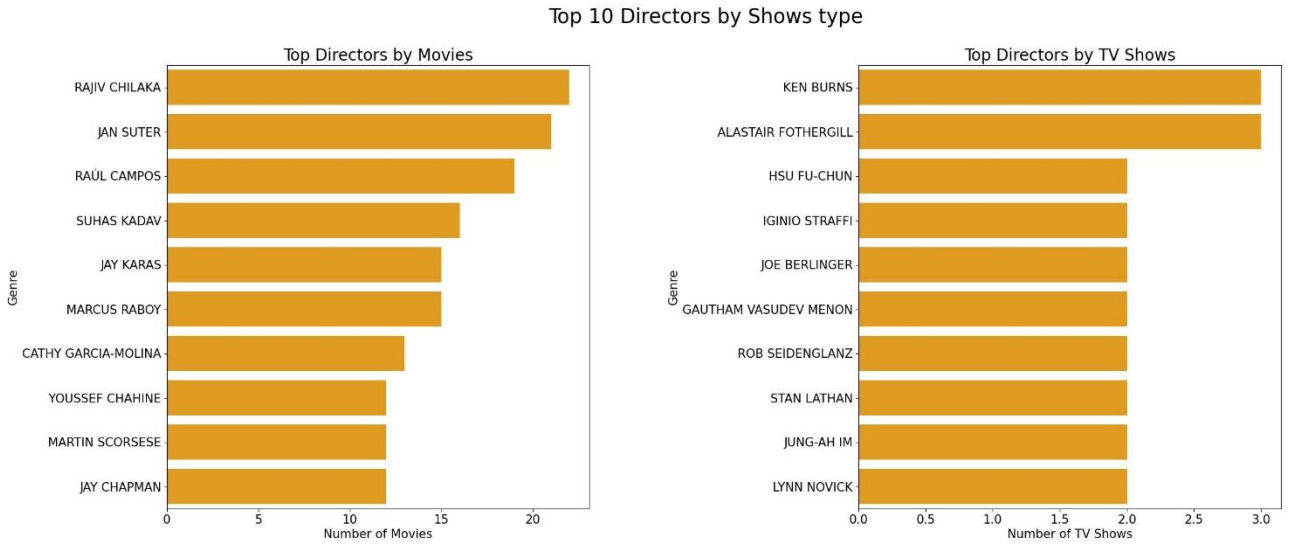
6. Top 10 Genre by Shows type



Insights:

- ✓ Nearly 45% of the shows comprised international movies and TV shows.
- ✓ The top three genres in both movies and TV shows were consistent, with international content, dramas, and comedies leading the charts.

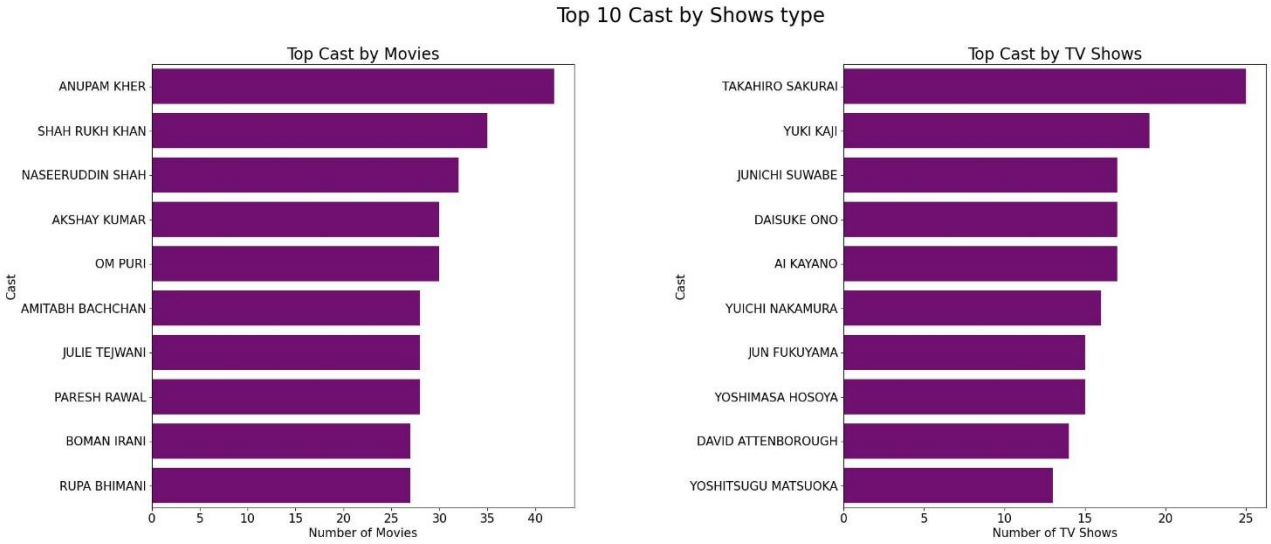
7. Top 10 Directors by Shows type



Insights:

- ✓ Rajiv Chilaka, the Indian director, stands out in movies, having directed nearly over 20 films.
- ✓ Each director in the top 10 list of movie counts has directed more than 10 movies.
- ✓ Ken Burns, an American director, and Alastair Fothergill, a British director, hold the first place in TV shows, with 3 TV shows each.
- ✓ Subsequently, the next 8 directors in the top 10 list hold the second place with 2 TV shows each.

7. Top 10 Cast by Shows type



Insights:

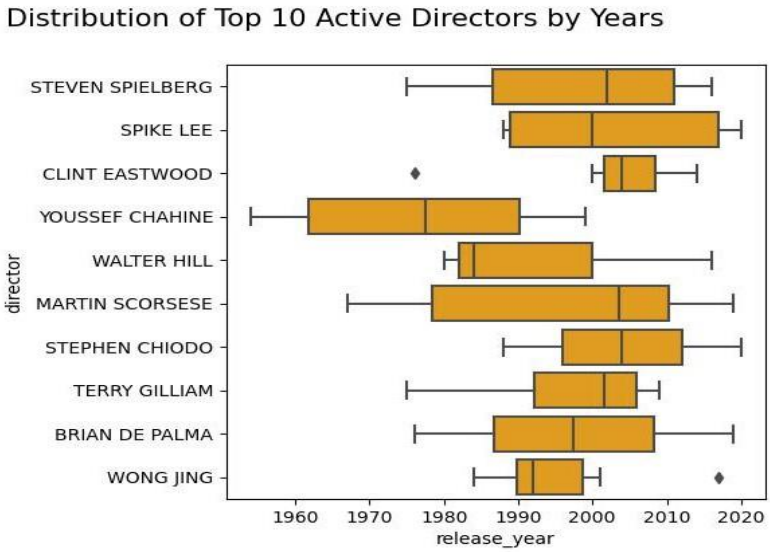
Nine out of the top 10 cast members in movies are from India.

Similarly, nine out of the top 10 cast members in TV shows are from Japan.

The plot illustrates that the majority of the top cast members are from the Asian continent.

8.Top 10 active Directors:

	Director Name	Years Active In Netflix
1	Martin Scorsese	52
2	Youssef Chahine	45
3	Brian De Palma	43
4	Steven Spielberg	41
5	Clint Eastwood	38
6	Walter Hill	36
7	Terry Gilliam	34
8	Wong Jing	33
9	Stephen Chiodo	32
10	Spike Lee	32



Insights:

- ✓ Martin Scorsese emerges as the most prolific director on Netflix, with a career spanning over 52 years, showcasing his work across different decades, from the 1960s to the present.
- ✓ Youssef Chahine, with an active career of 45 years, contributed from the 1950s until the early 2000s, although he is not currently active.
- ✓ Notable currently active directors include Stephen Chiodo, Martin Scorsese, Brian De Palma, and Spike Lee.

Recommendation:

Comprehensive Viewer Preference Analysis Metrics: Perform in-depth surveys to determine viewer preferences, such as favourite show genres, favourite show types, directors whose works are worth seeing again, languages or countries that viewers prefer to watch in addition to their native language, and actors whose performances are worth seeing again.

Type of shows preference: Analyse user viewing behaviour by country to determine whether people prefer watching TV shows or movies. This analysis will help in selecting shows that align with the preferences of users in each country, thereby enhancing content selection and improving the overall user experience.

Responsive Content Acquisition: Stay agile in content acquisition by monitoring recent trends and adapting to evolving viewer preferences. Utilize real-time data analysis to identify emerging trends and acquire content that resonates with the audience.

Regional Preferences: Analyse regional preferences to tailor content offerings based on audience interests. For example, if certain genres or languages are more popular in specific regions, prioritize acquiring content that aligns with those preferences.

Diversification: Continuously diversify the content library to cater to a broader audience. Incorporating a mix of genres, languages, and production styles can attract viewers from different demographics and regions.

Evaluation of Off-Peak Months: Investigate reasons behind the decline in content additions during months like February, October, and December. It could be due to factors such as production schedules, holidays, or viewer behaviour.

Data-Driven Decision Making: Continuously monitor and analyse viewer trends and preferences to inform future content acquisition and release strategies, allowing for more agile and responsive decision-making.