

# Analysing **NETFLIX's** Content Strategy For Business Growth

# NETFLIX

**Harjinder Singh**

DSML April 2023 Batch

# Table of Content

1. Problem Statement and Analysing basic metrics

2. Data Preprocessing

3. DATA FORM EXPLORATION – SHAPE AND TYPE ANALYSIS

4. Data Visualization

5. Box Plot and Outlier Checks

6. Heatmaps

7. Top Actors and Directors For Movies In Top 5 Most Popular Countries

8. Top Actors and Directors For TV Shows In Top 5 Most Popular Countries

9. Seasonal Content Strategy

10. Business Strategy Recommendation

## 1. Problem Statement and Analysing basic metrics

The goal is to optimize Netflix's content library by recommending which type of movies and TV shows the streaming platform should focus on and improving user satisfaction in different countries.

### About the dataset

The dataset is available in a single csv file and consists of following columns:

<b>Show ID</b>	The ID of the show
<b>Type</b>	Identifier - A Movie or TV Show
<b>Title</b>	Title of the Movie / Tv Show
<b>Director</b>	Director of the Movie
<b>Cast</b>	Actors involved in the movie/show
<b>Country</b>	Country where the movie/show was produced
<b>Date_Added</b>	Date it was added on Netflix
<b>Release_Year</b>	Actual Release year of the movie/show
<b>Rating</b>	TV Rating of the movie/show
<b>Duration</b>	Total Duration - in minutes or number of seasons
<b>Listed_in</b>	Genre
<b>Description</b>	The summary description

### Data Exploration Before Pre-processing - Shape And Type Analysis

#### Shape of Dataset

<b>Rows</b>	8807
<b>Columns</b>	12

### Attribute Analysis

Column Name	Non-null Count	Data type
Show ID	8807	object
Type	8807	object
Title	8807	object
Director	6173	object
Cast	7982	object
Country	7976	object
Date_Added	8797	object
Release_Year	8807	int64
Rating	8803	object
Duration	8804	object
Listed_in	8807	object
Description	8807	object

## 2. Data Preprocessing

### a. Identifying Nested Data

- i. In our dataset nested data are columns that have concatenated information using ','
- ii. The check function will be used to identify columns that contain ',' in them
- iii. Check Function Identified these columns that contain ',':
  - title
  - director
  - cast
  - country
  - date\_added
  - listed\_in
  - description

## b. Identifying candidates that require un-nesting

- i. The un-nesting needs to be performed on columns that have multiple data concatenated using ','.
- ii. Let's identify columns that require un-nesting from remaining columns:
  - **Title:** Doesn't require un-nesting as ',' is for grammatical purposes.
  - **Director:** Requires un-nesting as ',' is used to separate multiple data.
  - **Cast:** Requires un-nesting as ',' is used to separate multiple data.
  - **Country:** Requires un-nesting as ',' is used to separate multiple data.
  - **Date Added:** Doesn't require un-nesting as ',' is to create a date.
  - **Listed In:** Requires un-nesting as ',' is used to separate multiple data.
  - **Description:** Doesn't require un-nesting as ',' is for grammatical purpose.
- iii. The identified candidates for un-nesting are:
  - Director
  - Cast
  - Country
  - Listed\_in

## c. Un-nesting identified columns

Let's identify columns that require un-nesting from remaining columns:

- i. **Step 1:** Create a copy of the original dataframe.
- ii. **Step 2:** Prepare nested columns for split.
- iii. **Step 3:** Split Values:
  - **First check:** check if the values are already in the form of a list. If yes then ignore. This is necessary in order to maintain data integrity.
  - **Second check:** check if the function contains ','. If yes it will split the data into a list.
  - **Third check:** if the column contains null value it will create a list with np.nan() to maintain the data integrity.
- iv. **Step 4:** Requires un-nesting as ',' is used to separate multiple data.
- v. **Step 5:** Doesn't require un-nesting as ',' is to create a date.

## d. Removing leading and trailing spaces from each column

### e. Handling Missing Values

In our raw dataset, following columns have missing values:

- a. **Director:** 2634
- b. **Cast:** 825
- c. **Country:** 831
- d. **Date\_added:** 831
- e. **Rating:** 4
- f. **Duration:** 3

### f. Handling Missing Values

- i. **Director:** Replacing nan with Unknown Director
- ii. **Cast:** Replacing nan with Unknown Actor
- iii. **Country:** Replacing nan with Unknown Country
- iv. **Rating:** Replacing nan with Unknown Rating
- v. **Duration:** Replacing nan with 0.

### g. Converting date from string to date datatype in dd/mm/yyyy format

## 3. DATA FORM EXPLORATION – SHAPE AND TYPE ANALYSIS

### Shape of Dataframes

Dataframe	Number of Rows	Number of Columns
og_df	8807	8
dir_df	9612	2
cast_df	64951	2
country_df	10850	2
listed_in_df	19323	2

## Dataframe Attribute Analysis

Dataframe	Column Name	Non-null Count	DType
og_df	show_id	8807	object
og_df	type	8807	object
og_df	title	8807	object
og_df	date_added	8807	object
og_df	release_year	8807	int64
og_df	rating	8807	object
og_df	duration	8807	object
og_df	description	8807	object
dir_df	show_id	9612	object
dir_df	director	9612	object
cast_df	show_id	64951	object
cast_df	cast	64951	object
country_df	show_id	10850	object
country_df	country	10850	object
listed_in_df	show_id	19323	object
listed_in_df	listed_in	19323	object

## Dataframe Attribute Analysis

### a. TV Shows vs. Movies

- There are 6131 movies and 2676 TV shows in the dataset.
- With 6131 movies and 2676 TV shows, the dataset showcases a diverse collection of content.

### b. TV Shows vs. Movies

- There are 6131 movies and 2676 TV shows in the dataset.
- With 6131 movies and 2676 TV shows, the dataset showcases a diverse collection of content.

c. TV Shows vs. Movies

- i. There are 6131 movies and 2676 TV shows in the dataset.
- ii. With 6131 movies and 2676 TV shows, the dataset showcases a diverse collection of content.

d. Directors of TV Shows vs. Movies

- i. Total number of directors: 5040
- ii. Number of directors directing the movies: 4742
- iii. Number of directors directing TV Shows: 298

e. Cast of TV Shows vs. Movies

- i. Total number of actors: 40809
- ii. Number of casts acting in the movies: 25947
- iii. Number of casts acting in TV Shows: 14862

f. Country wise TV Shows vs. Movies

- i. Total number of countries: 124
- ii. Number of Countries with Movies: 119
- iii. Number of Countries with TV Shows: 67

g. Genre Movies

Movie Genre	Count
International Movies	2752
Dramas	2427
Comedies	1674
Documentaries	869
Action & Adventure	859
Independent Movies	756
Children & Family Movies	641
Romantic Movies	616
Thrillers	577
Music & Musicals	375
Horror Movies	357
Stand-Up Comedy	343



Sci-Fi & Fantasy	243
Sports Movies	219
Classic Movies	116
LGBTQ Movies	102
Cult Movies	71
Anime Features	71
Faith & Spirituality	65
Movies	57

#### h. Genre Movies

TV Show Genre	Count
International TV Shows	1351
TV Dramas	763
TV Comedies	581
Crime TV Shows	470
Kids' TV	451
Docuseries	395
Romantic TV Shows	370
Reality TV	255
British TV Shows	253
Anime Series	176
Spanish-Language TV Shows	174
TV Action & Adventure	168
Korean TV Shows	151
TV Mysteries	98
Science & Nature TV	92
TV Sci-Fi & Fantasy	84
Teen TV Shows	75
TV Thrillers	69

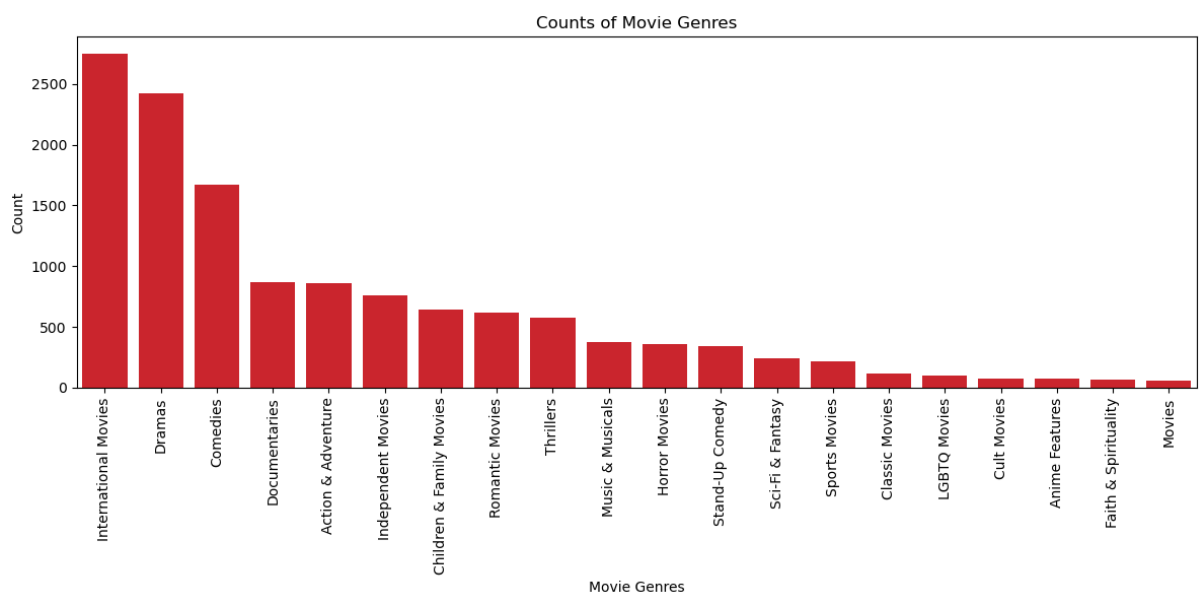
Stand-Up Comedy & Talk Shows	56
Classic & Cult TV	28
TV Shows	16

## 4. Data Visualization

### Dataframe Attribute Analysis

#### a. Countplot (Count of Movies in Each Genre)

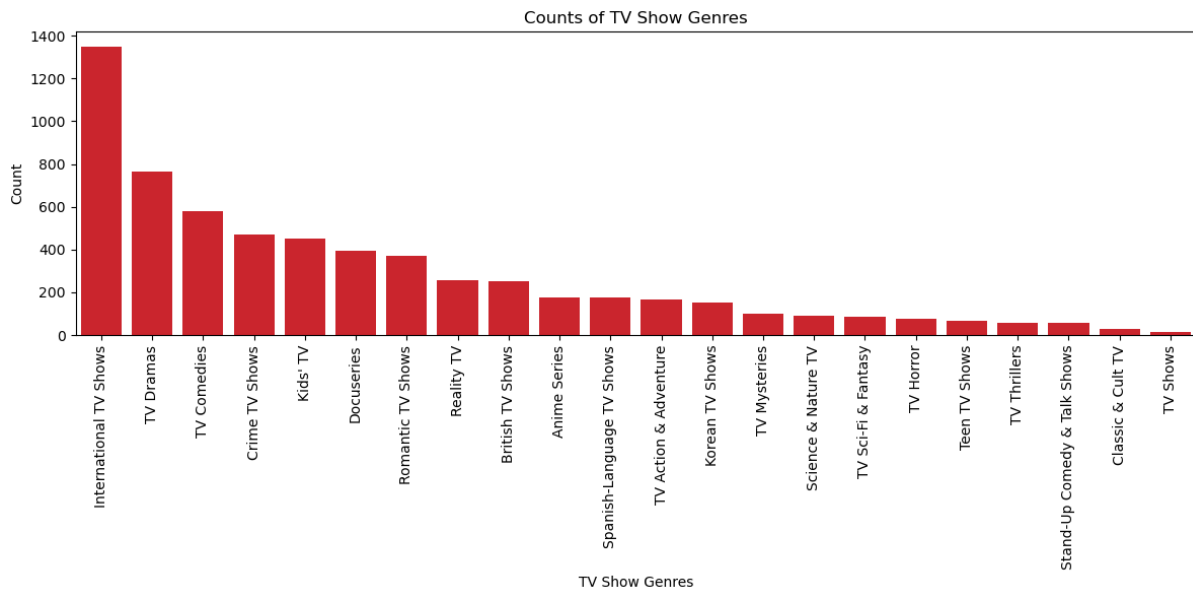
**Insight:** In movies, International movies, Dramas, and Comedies are top 3 genres in the entire dataset.



#### b. Countplot (Count of TV Shows in Each Genre)

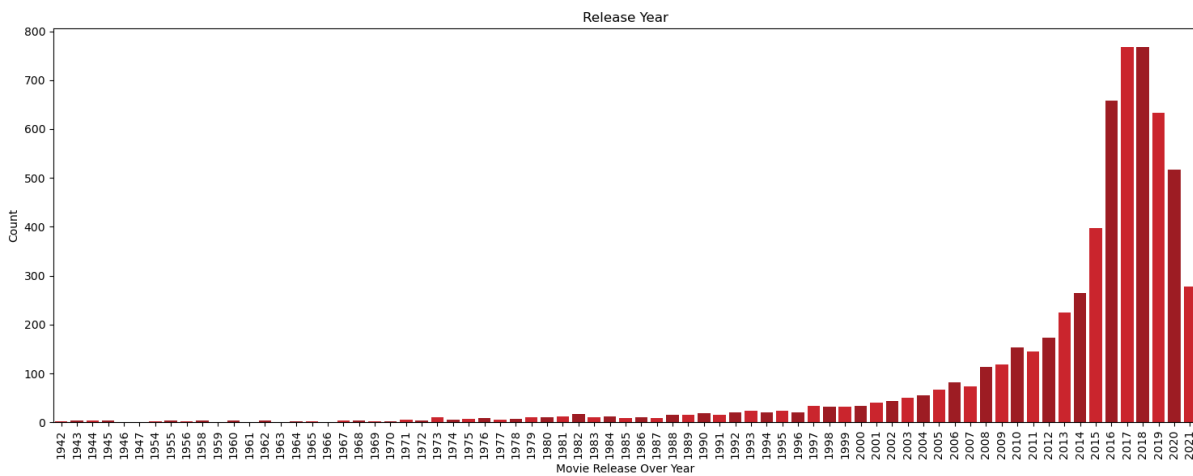
**Insight:**

- In TV Shows, International TV Shows, TV Dramas, and TV Comedies are top 3 genres in the entire dataset.
- It is interesting to see that both TV Shows and Movies have the same top 3 genres.



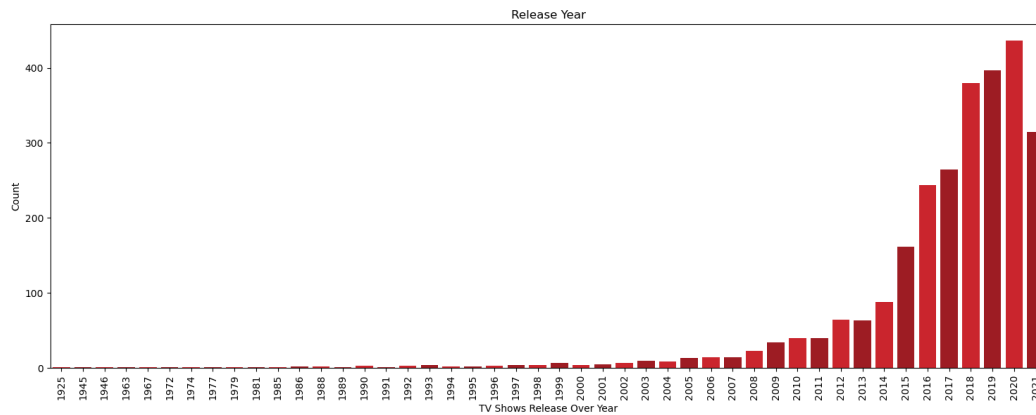
### c. Barplot (Count of Movies Over Release Year)

**Insight:** The Netflix platform predominantly features a higher number of movies that were released between 2016 and 2019. In contrast, there is a relatively smaller selection of movies released in the years before 2016.



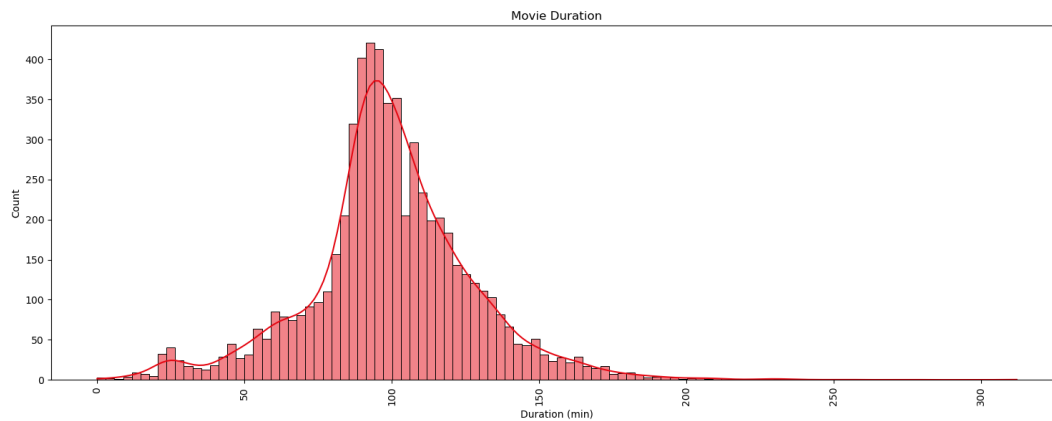
#### d. Barplot (Count of TV Shows Over Release Year)

**Insight:** The Netflix platform predominantly features a higher number of TV Shows that were released after 2018.



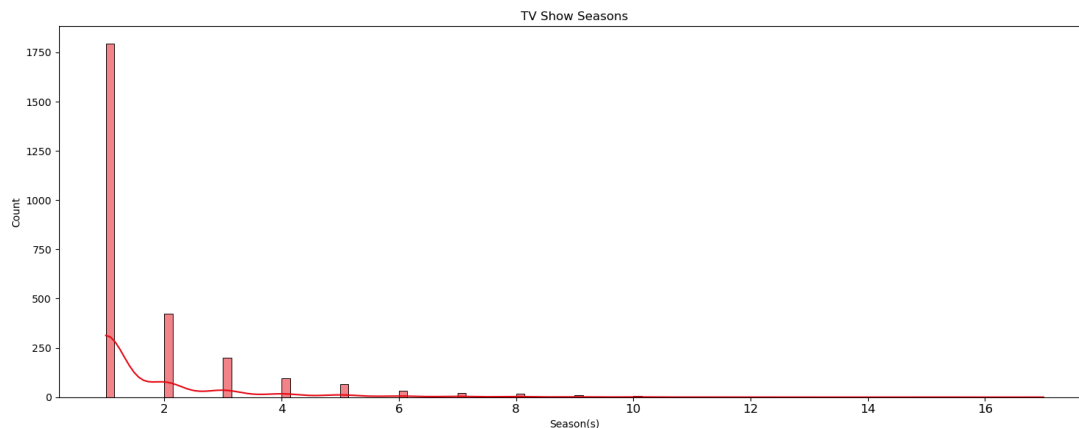
#### e. Histplot (Distribution of Movie Duration)

**Insight:** Most movies listed on the Netflix platform are having duration between 90 mins to 120 mins.



#### f. Histplot (Distribution of TV Show Seasons)

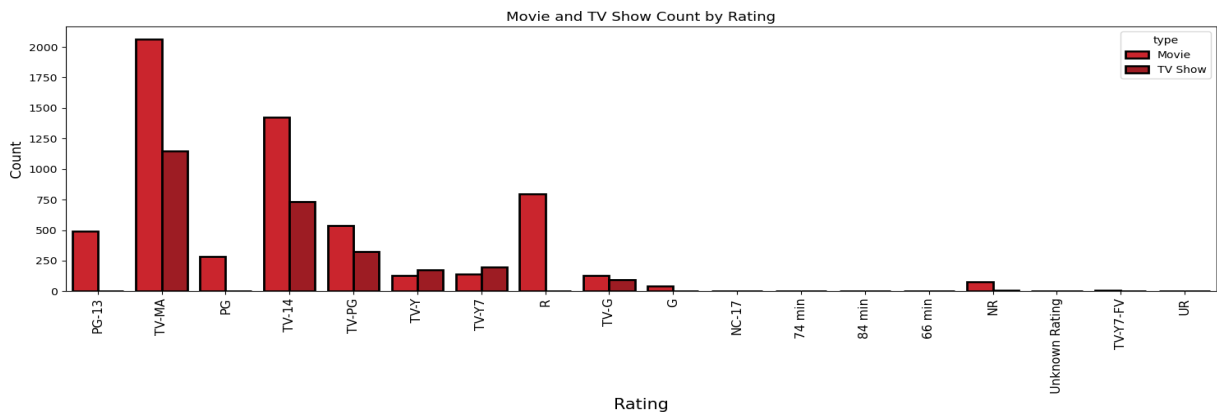
**Insight:** Most TV Shows listed on the Netflix platform have only 1 season.



### g. Barplot (Movie and TV Show Count by Rating)

#### Insight:

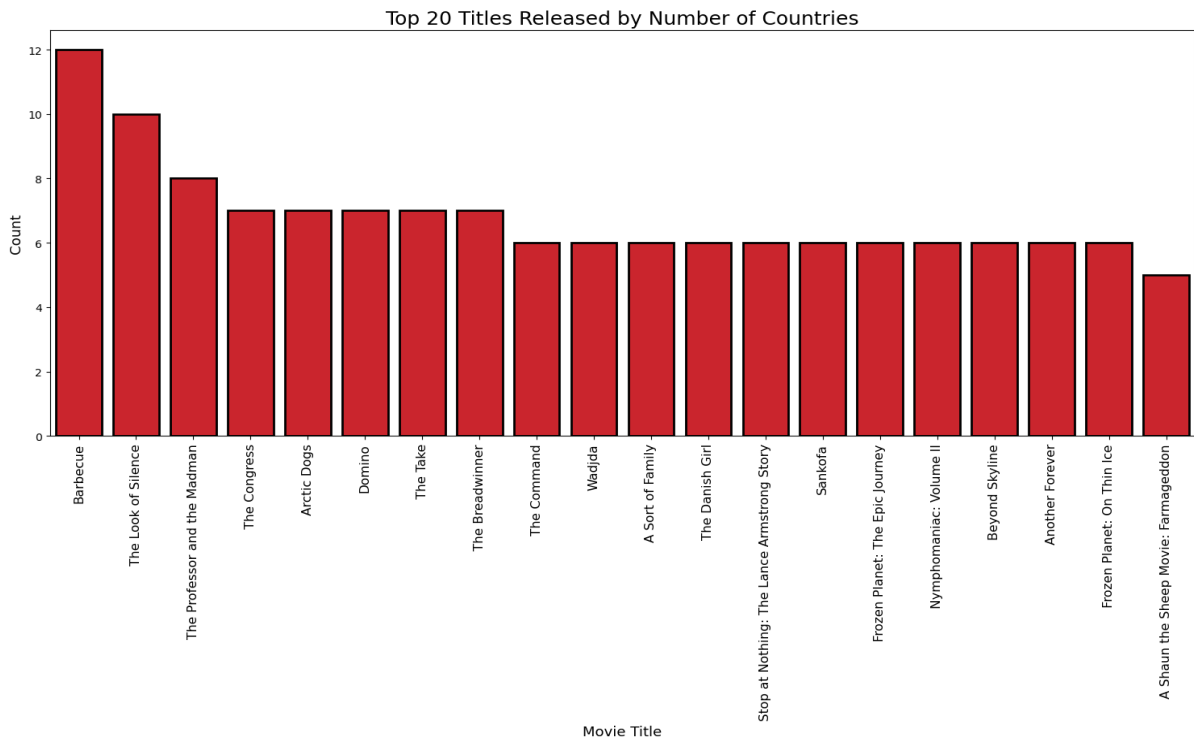
- Most movies listed on the Netflix platform have a TV-MA rating followed by TV-14 rating.
- Most TV shows listed on the Netflix platform have a TV-MA rating followed by TV-14 rating.
- R-rated movies are quite prevalent as compared to TV shows.



### h. Barplot (Multi Country Movie Releases)

#### Insight:

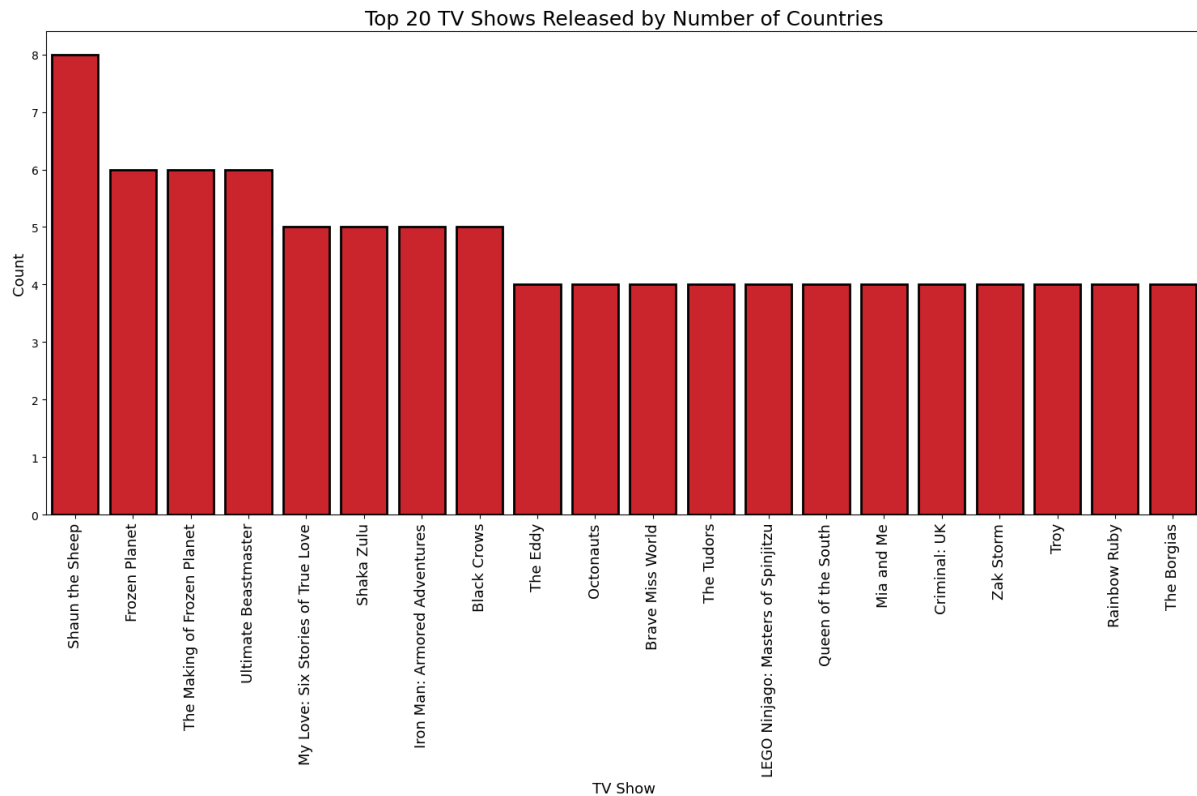
- A total of 19 movies were produced in more than 5 countries.
- Barbecue was released in 12 countries, which is the highest for any movie.



#### i. Barplot (Multi Country TV Show Releases)

##### Insight:

- 8 TV Shows were released in 5 or more countries.
- Shaun The Sheep was released in 8 countries, which is highest for any TV Shows in our dataset.

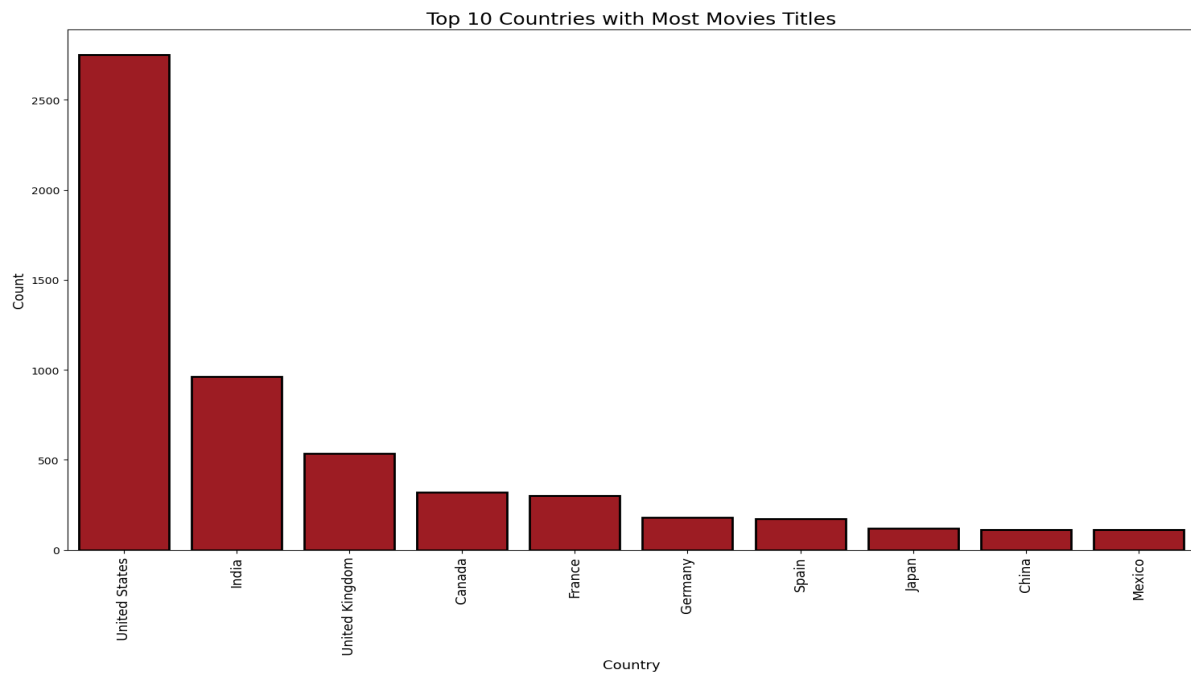


#### j. Barplot (Top 10 Countries with Most Movies Titles)

##### Insight:

- The number of movies released in the United States is significantly higher than any other country in our dataset.
- It is interesting to see that only Netflix has released more than 500 movies in only 3 countries.

**Recommendation:** Netflix can look to release more movies in countries other than the US, India, and UK.

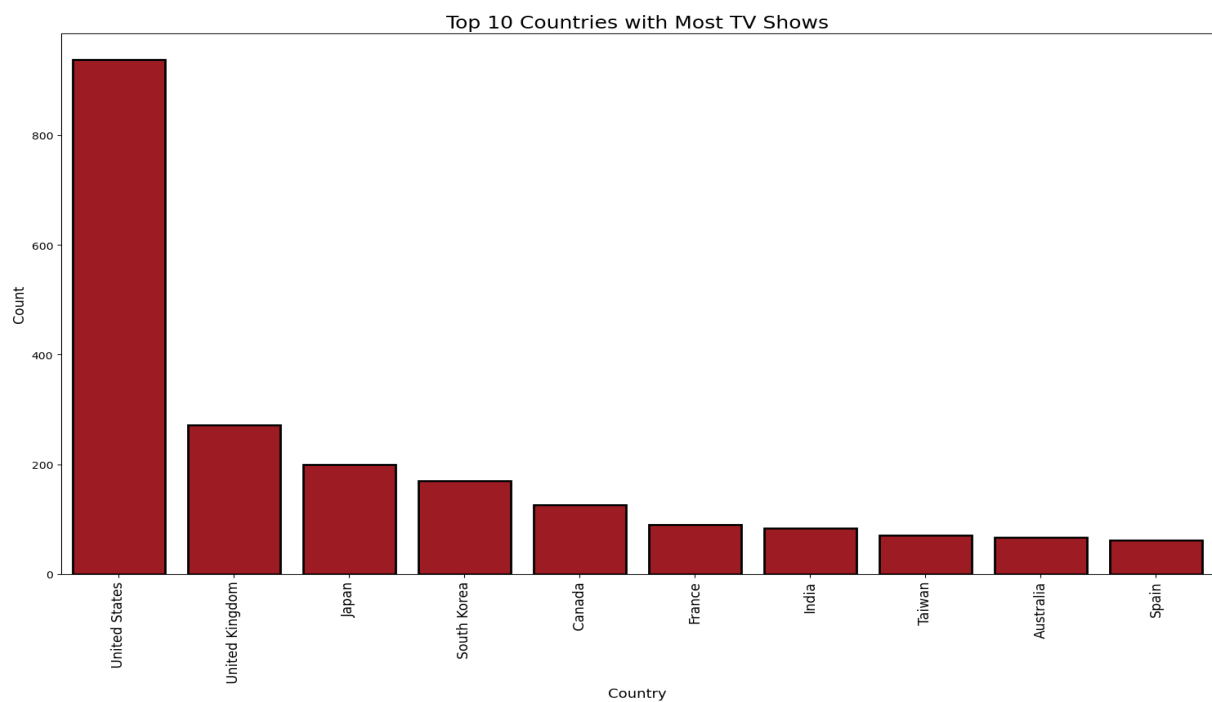


#### k. Barplot (Top 10 Countries with Most TV Shows)

##### Insight:

- Just like for movies, United States has a significantly high number of TV shows as compared to other countries.
- Interestingly only 3 countries have 200 or more TV shows.

**Recommendation:** Netflix can focus on releasing more TV shows in countries other than the US.



## 5. Box Plot and Outlier Checks

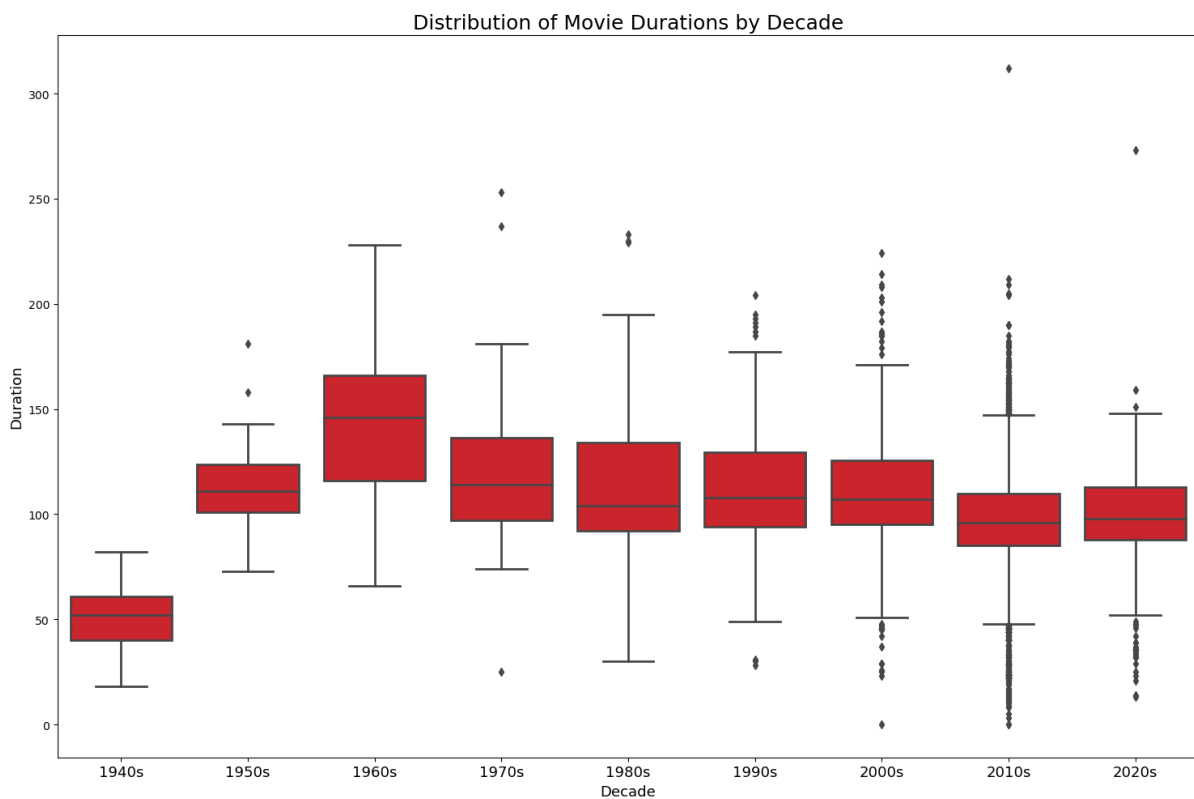
### a. Distribution of Movie Durations by Decade

#### Insight:

- Median duration of movies in the 1950s became more than the maximum duration in the 1940s.
- Median duration of movies was highest in the 1960s but seems to have gone down slightly since the 1980s onwards.
- Median duration was almost constant between the 1980s and 2000s.

#### Outlier Detection:

- Decades 2000s, 2010s, 2020s had the most number of outliers as compared to prior decades.
- The max duration of movies was more than 300 mins in 2010s which is exceptionally higher than all the movies that are on Netflix's platform.
- Number of outliers have increased with each decade suggesting more experimentation with movie duration by directors.





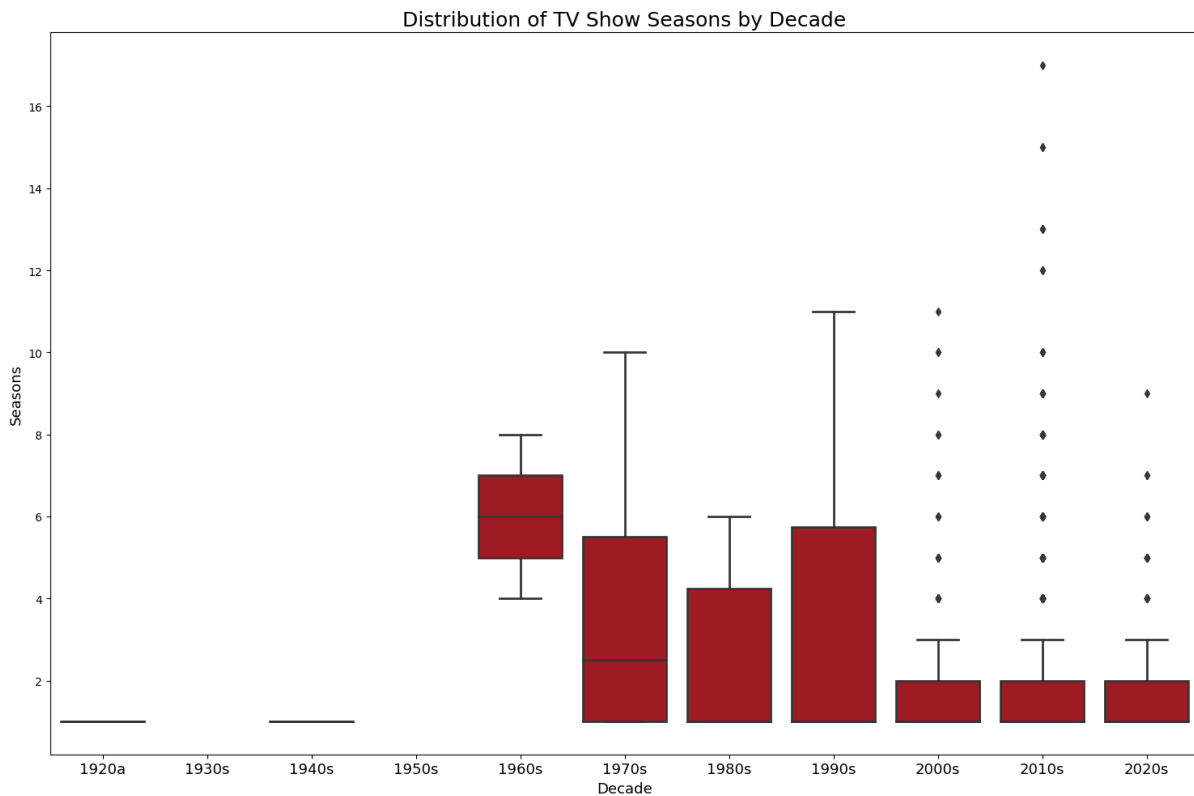
## b. Distribution of TV Show Seasons by Decade

### Insight:

- There are no TV shows on Netflix from the 1930s and 1950s.
- From 2000 onwards the median number of seasons in the TV show were 2 but still there were multiple TV shows having the number of seasons between 4 to 17.
- In the 1960s the median number of seasons for TV shows was highest which i.e. 6 which reduced to 3 in the next decade.
- Median seasons were constant after the 1980s.

### Outlier Detection:

- The boxplot does not exhibit any outliers on the minimum side.
- Outliers can be found 2000s onwards
- Decade of the 2020s is having TV show with maximum numbers of seasons i.e. 17.



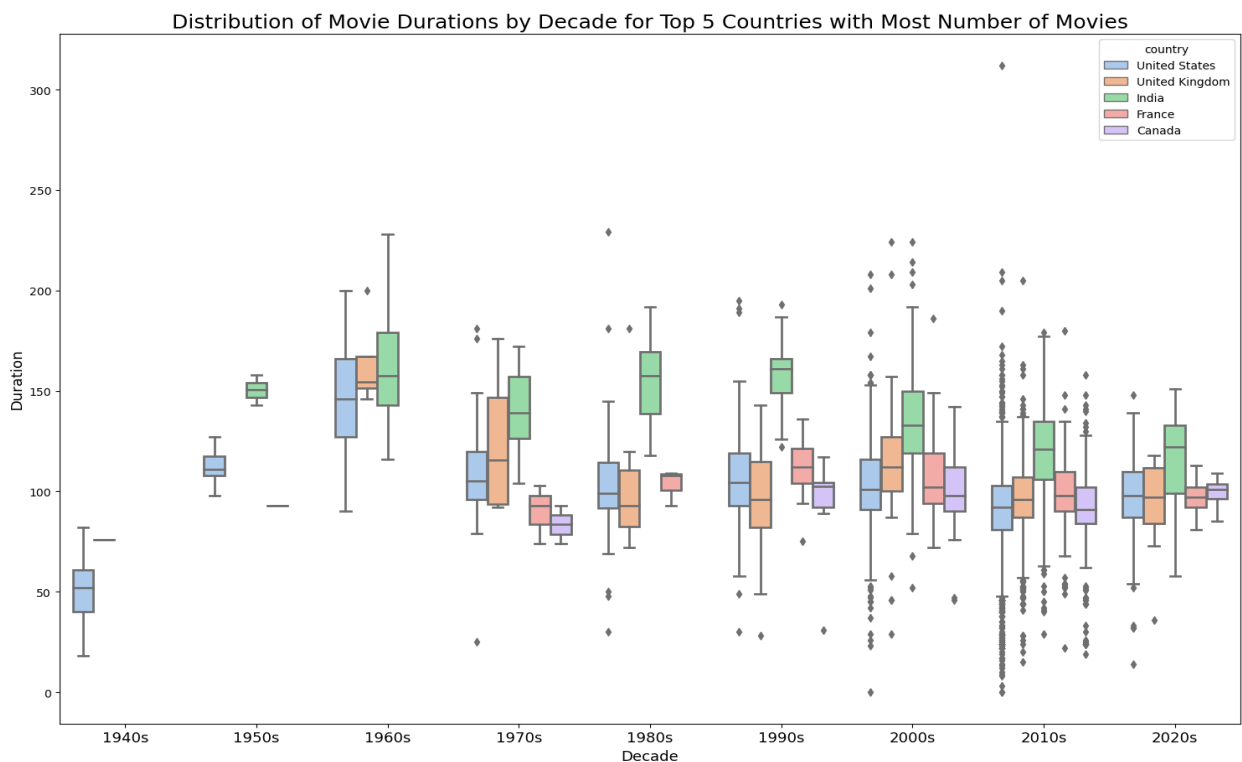
### c. Distribution of Movie Durations by Decade for Top 5 Countries with Most Number of Movies

#### Insight:

- The oldest movies available on the Netflix platform date to the 1940s and originate from the United States and some of them from the United Kingdom.
- Netflix offers movies from the United Kingdom and India which were released in the 1960s.
- Canadian Netflix content includes movies from the 1990s, making them the oldest from Canada.
- The median duration of Indian movies on Netflix has been gradually decreasing since the 1990s.
- French movie durations on Netflix experienced an increase from the 1970s to the 1990s, followed by a decline from that decade onwards.
- For movies from the United States, the median duration remained relatively stable from the 1970s to the 2000s, with a slight decline thereafter.
- Canadian movie duration has remained nearly constant since the 1990s.

#### Outlier Detection:

- The most number of outliers for all countries was between 1990s till 2010s for all countries.
- The United States seems to have most number of outliers in 2000s.



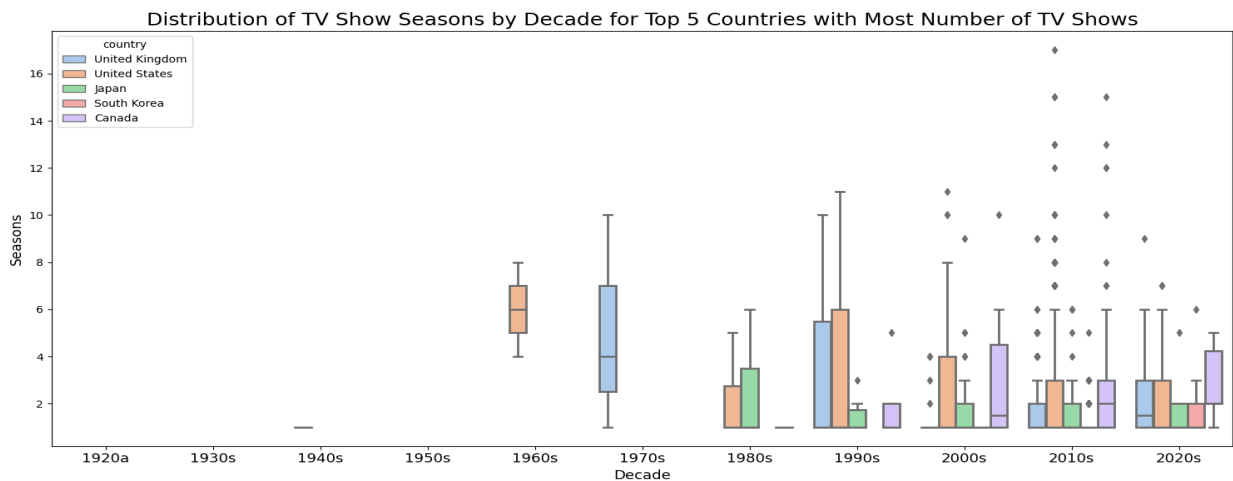
#### d. Distribution of TV Show Seasons by Decade for Top 5 Countries with Most Number of TV Shows

##### Insight:

- Median number of seasons for TV show have fallen down from 1970s onward
- The number of TV Shows from Canada has increased since the 2000s while the number of TV shows from the United States has reduced a lot.

##### Outlier Detection:

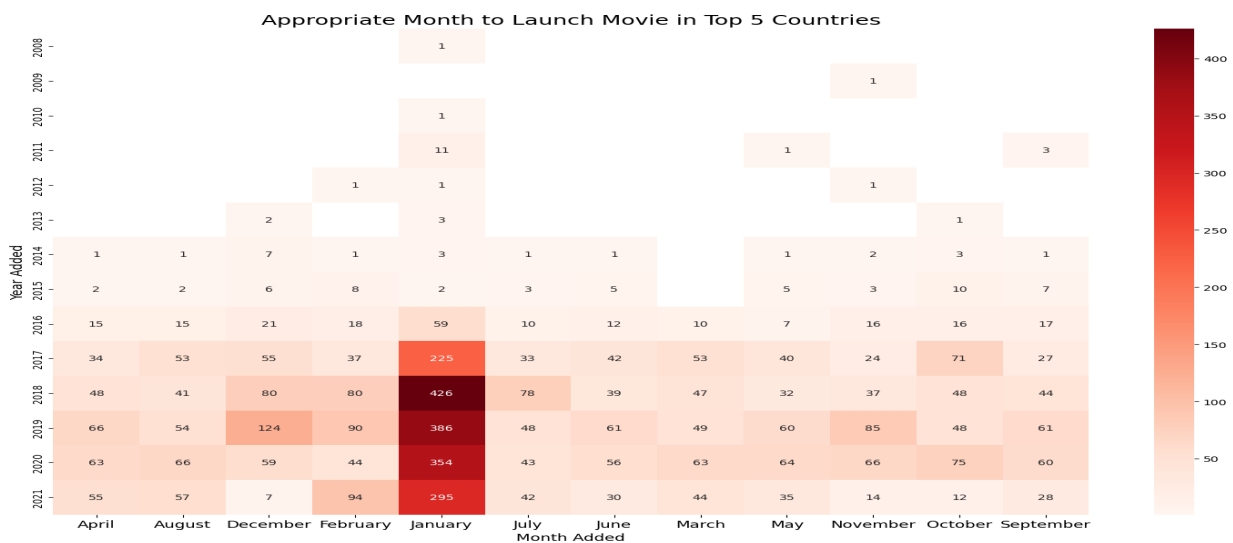
- No outliers are observed on the minimum side of the box plot.
- Outliers have significantly increased since the 2000s.
- The United States has shown a significant increase in outliers since 2010s.



## 6. Heatmaps

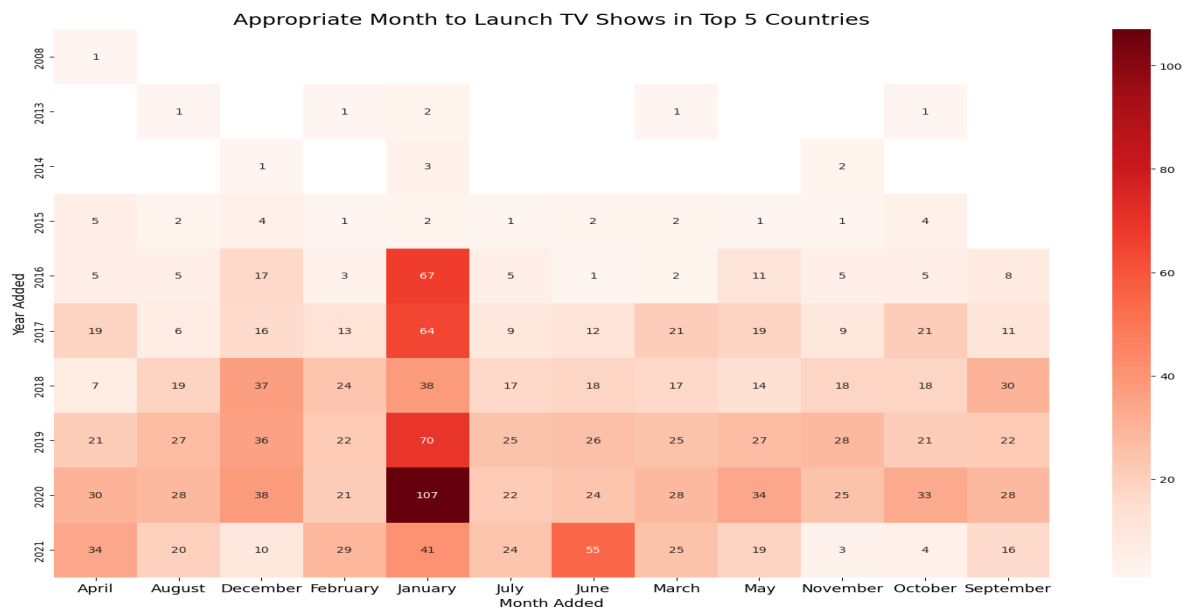
#### a. Appropriate Month to Launch Movie in Top 5 Countries

**Insight:** Most TV Shows listed on the Netflix platform have only 1 season.



## b. Appropriate Month to Launch TV Show in Top 5 Countries

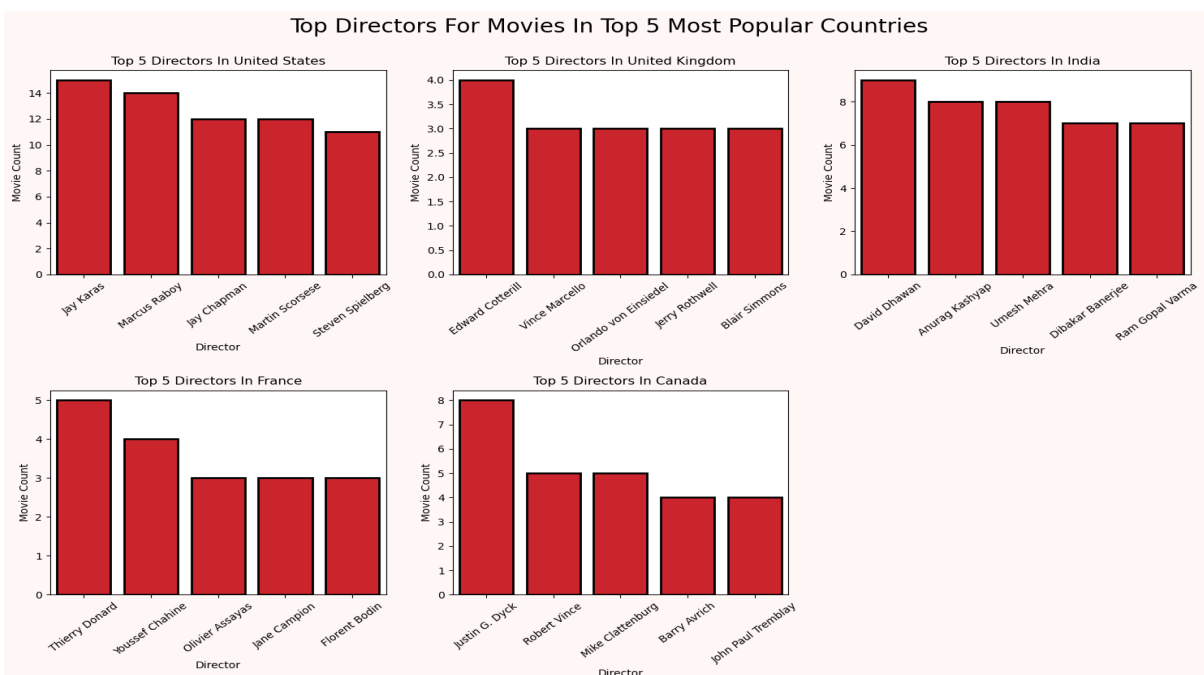
**Insight:** January seems to be the month with the most number of TV Show releases.



## 7. Top Actors and Directors For Movies In Top 5 Most Popular Countries

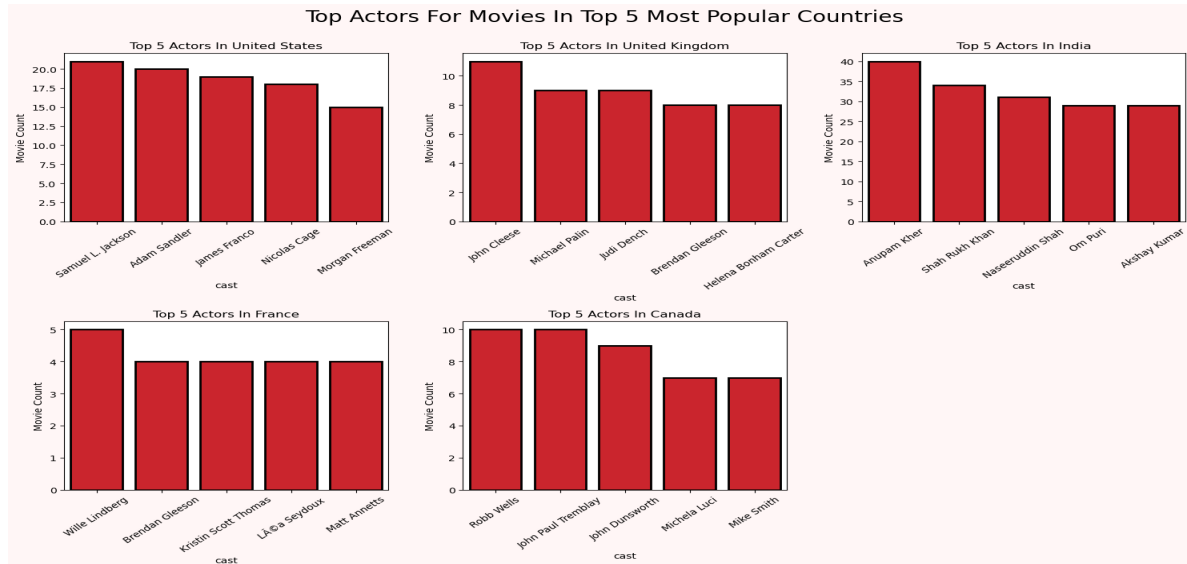
### a. Top 5 Movie Directors

**Insight:** Top 5 directors for the movie are identified for the top 5 countries. Details of the same will be available in the recommendations section.



## b. Top 5 Movie Actors

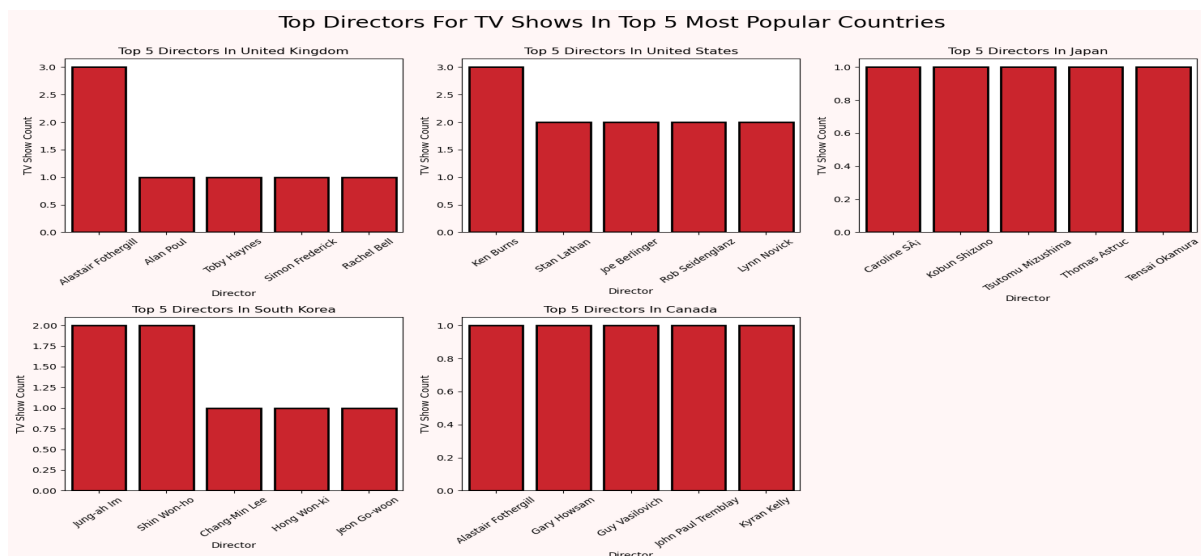
**Insight:** Top 5 actors for the movie are identified for the top 5 countries. Details of the same will be available in the recommendations section.



## 8. Top Actors and Directors For TV Shows In Top 5 Most Popular Countries

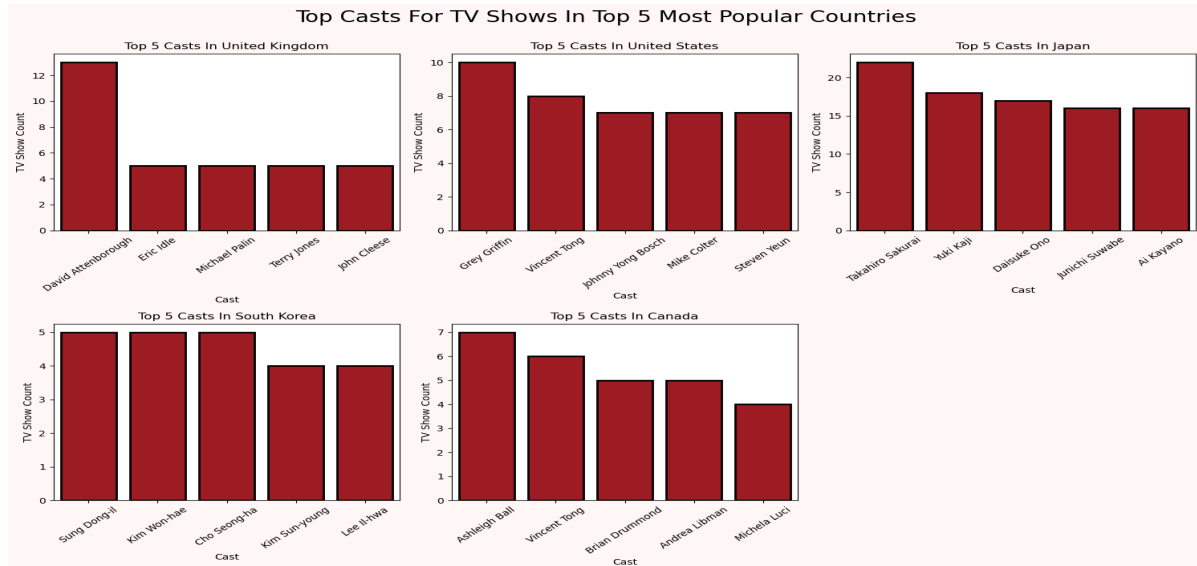
### a. Top 5 TV Show Directors

**Insight:** Top 5 directors for the TV Shows are identified for the top 5 countries. Details of the same will be available in the recommendations section.



## b. Top 5 TV Show Actors

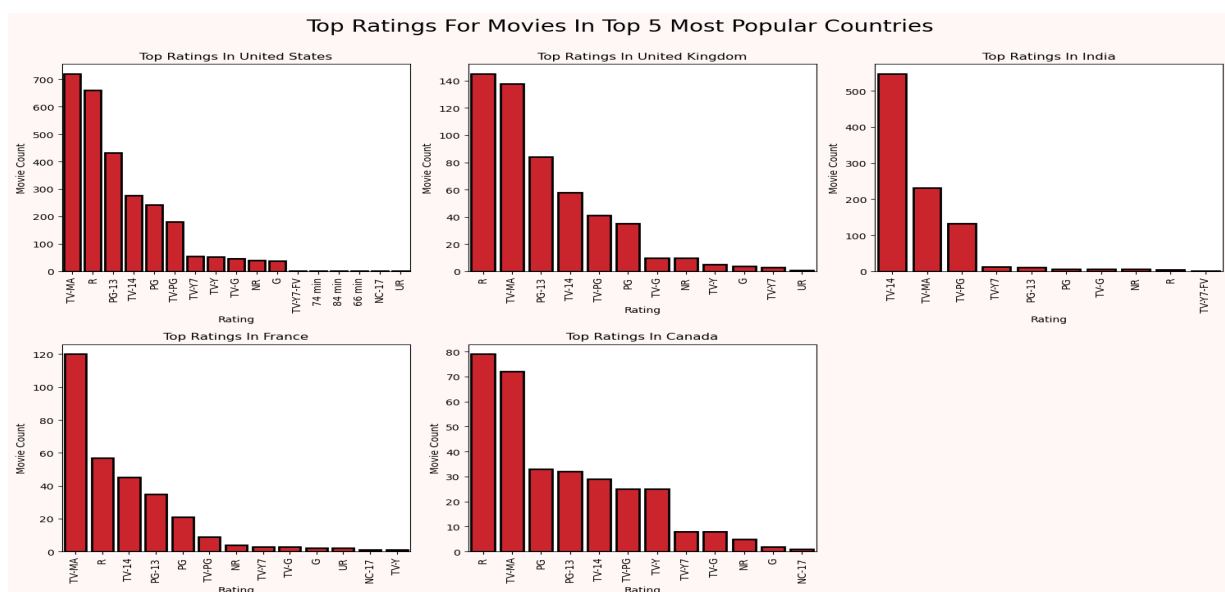
**Insight:** Top 5 actors for the TV Shows are identified for the top 5 countries. Details of the same will be available in the recommendations section.



## c. Movie Ratings in Top 5 Countries

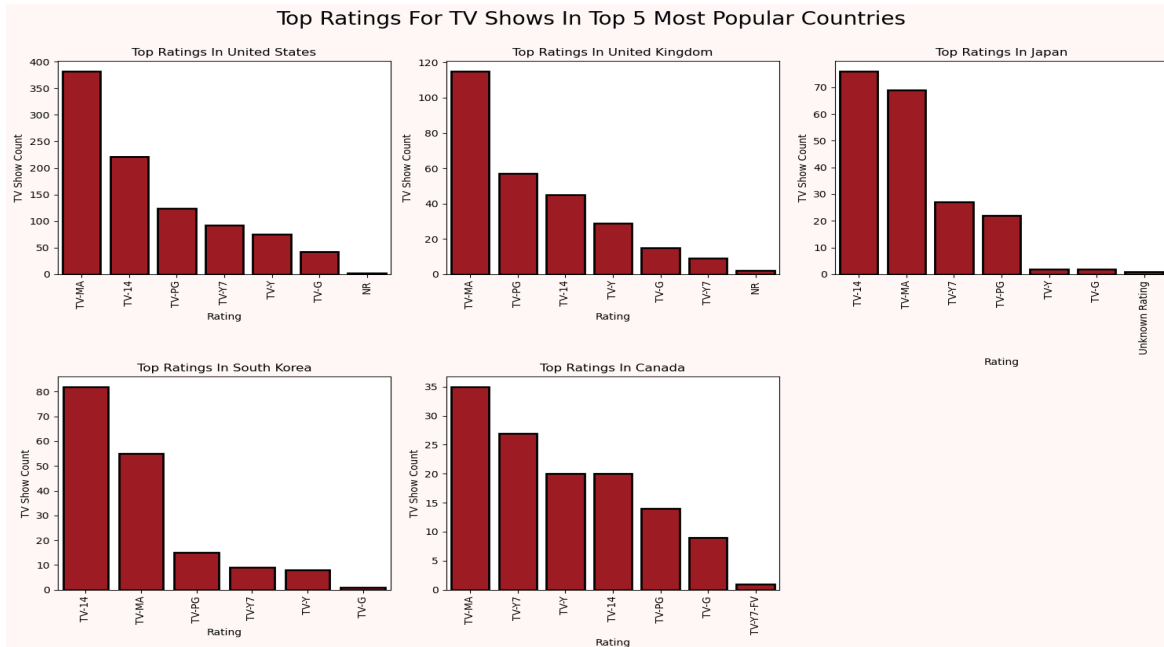
**Insight:**

- TV-MA is the most popular rating in 3 out of 5 countries.
- Except India which has very few R-rated movies, R-rated movies are popular in all top countries.
- India has the most number of TV-14 movies.



#### d. TV Show Ratings in Top 5 Countries

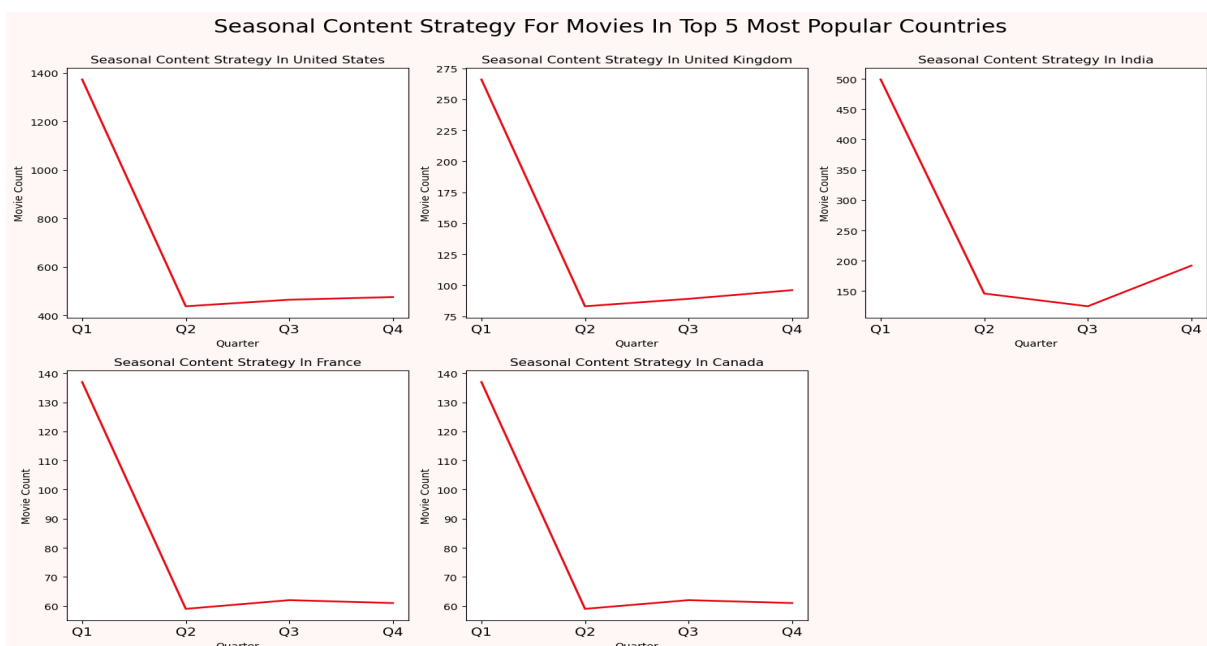
**Insight:** TV-MA and TV-14 are the most popular ratings for TV Shows on Netflix Platform.



## 9. Seasonal Content Strategy

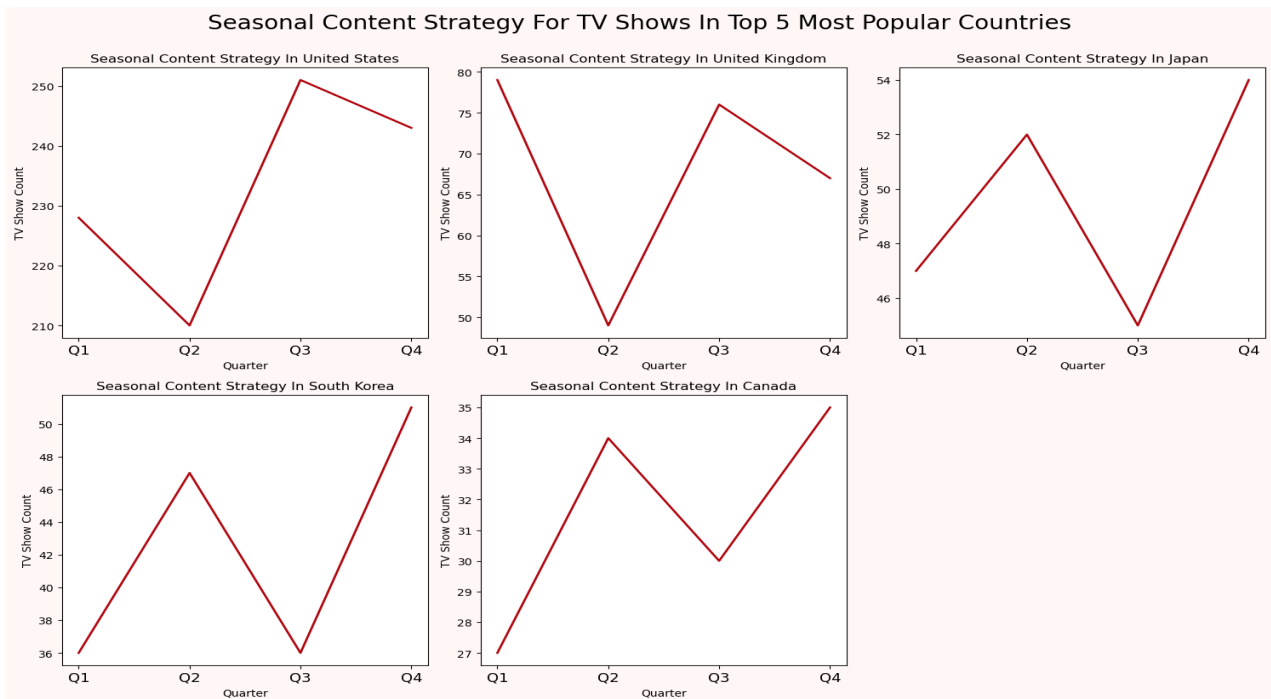
#### a. Movies

**Insight:** Q1 is the most busy month for the top 5 countries.



## b. TV Shows

**Insight:** Q3 is the most busy month for the top 5 countries.



## 10. Business Strategy Recommendation

### a. Target market

- Currently Netflix is doing well in some countries for Movies and some others for TV Shows.
- The platform should continue to grow their business in these markets as it knows the market dynamics and business conditions very well.
- Following are the top countries where Netflix should consider increasing their focus:
  - a. United States
  - b. United Kingdom
  - c. India
  - d. South Korea
  - e. Japan
  - f. France
  - g. Canada



## b. Recommendations for Movies

- For each country, the top director & cast is identified who have performed well in the trending genres in that country.
- Following table lists the country wise, director & cast recommendations:

Country	Director	Actor
United States	<ul style="list-style-type: none"><li>• Jay Karas</li><li>• Marcus Raboy</li><li>• Jay Chapman</li><li>• Martin Scorsese</li><li>• Steven Spielberg</li></ul>	<ul style="list-style-type: none"><li>• Samuel L. Jackson</li><li>• Adam Sandler</li><li>• James Franco</li><li>• Nicolas Cage</li><li>• Morgan Freeman</li></ul>
United Kingdom	<ul style="list-style-type: none"><li>• Edward Cotterill</li><li>• Vince Marcello</li><li>• Orlando von Einsiedel</li><li>• Jerry Rothwell</li><li>• Blair Simmons</li></ul>	<ul style="list-style-type: none"><li>• John Cleese</li><li>• Michael Palin</li><li>• Judi Dench</li><li>• Brendan Gleeson</li><li>• Helena Bonham Carter</li></ul>
India	<ul style="list-style-type: none"><li>• David Dhawan</li><li>• Anurag Kashyap</li><li>• Umesh Mehra</li><li>• Dibakar Banerjee</li><li>• Ram Gopal Varma</li></ul>	<ul style="list-style-type: none"><li>• Anupam Kher</li><li>• Shah Rukh Khan</li><li>• Naseeruddin Shah</li><li>• Om Puri</li><li>• Akshay Kumar</li></ul>
France	<ul style="list-style-type: none"><li>• Thierry Donard</li><li>• Youssef Chahine</li><li>• Olivier Assayas</li><li>• Jane Campion</li><li>• Florent Bodin</li></ul>	<ul style="list-style-type: none"><li>• Wille Lindberg</li><li>• Brendan Gleeson</li><li>• Kristin Scott Thomas</li><li>• L'Ã©a Seydoux</li><li>• Matt Annetts</li></ul>
Canada	<ul style="list-style-type: none"><li>• Justin G. Dyck</li><li>• Robert Vince</li><li>• Mike Clattenburg</li><li>• Barry Avrich</li><li>• John Paul Tremblay</li></ul>	<ul style="list-style-type: none"><li>• Robb Wells</li><li>• John Paul Tremblay</li><li>• John Dunsworth</li><li>• Michela Luci</li><li>• Mike Smith</li></ul>

### c. Recommendations for TV Shows

- For each country, the top director & cast is identified who have performed well in the trending genres in that country.
- Following table lists the country wise, director & cast recommendations:

Country	Director	Actor
United Kingdom	<ul style="list-style-type: none"><li>• Alastair Fothergill</li><li>• Alan Poul</li><li>• Toby Haynes</li><li>• Simon Frederick</li><li>• Rachel Bell</li></ul>	<ul style="list-style-type: none"><li>• David Attenborough</li><li>• Eric Idle</li><li>• Michael Palin</li><li>• Terry Jones</li><li>• John Cleese</li></ul>
United States	<ul style="list-style-type: none"><li>• Ken Burns</li><li>• Stan Lathan</li><li>• Joe Berlinger</li><li>• Rob Seidenglanz</li><li>• Lynn Novick</li></ul>	<ul style="list-style-type: none"><li>• Grey Griffin</li><li>• Vincent Tong</li><li>• Johnny Yong Bosch</li><li>• Mike Colter</li><li>• Steven Yeun</li></ul>
Japan	<ul style="list-style-type: none"><li>• Caroline Săi</li><li>• Kobun Shizuno</li><li>• Tsutomu Mizushima</li><li>• Thomas Astruc</li><li>• Tensai Okamura</li></ul>	<ul style="list-style-type: none"><li>• Takahiro Sakurai</li><li>• Yuki Kaji</li><li>• Daisuke Ono</li><li>• Junichi Suwabe</li><li>• Ai Kayano</li></ul>
South Korea	<ul style="list-style-type: none"><li>• Jung-ah Im</li><li>• Shin Won-ho</li><li>• Chang-Min Lee</li><li>• Hong Won-ki</li><li>• Jeon Go-woon</li></ul>	<ul style="list-style-type: none"><li>• Sung Dong-il</li><li>• Kim Won-hae</li><li>• Cho Seong-ha</li><li>• Kim Sun-young</li><li>• Lee Il-hwa</li></ul>
Canada	<ul style="list-style-type: none"><li>• Alastair Fothergill</li><li>• Gary Howsam</li><li>• Guy Vasilovich</li><li>• John Paul Tremblay</li><li>• Kyrán Kelly</li></ul>	<ul style="list-style-type: none"><li>• Ashleigh Ball</li><li>• Vincent Tong</li><li>• Brian Drummond</li><li>• Andrea Libman</li><li>• Michela Luci</li></ul>

#### d. Ratings Recommendation For Movies

This document presents the analysis about ratings of the movies that need to be considered while getting listed.

These are the ratings for the content:

Country	Rating
United States	<ul style="list-style-type: none"><li>• TV-MA</li><li>• R</li><li>• PG-13</li><li>• TV-14</li><li>• PG</li><li>• TV-PG</li><li>• TV-Y7</li><li>• TV-Y</li><li>• TV-G</li></ul>
United Kingdom	<ul style="list-style-type: none"><li>• R</li><li>• TV-MA</li><li>• PG-13</li><li>• TV-14</li><li>• TV-PG</li></ul>
India	<ul style="list-style-type: none"><li>• TV-14</li><li>• TV-MA</li><li>• TV-PG</li><li>• TV-Y7</li><li>• PG-13</li></ul>
France	<ul style="list-style-type: none"><li>• TV-MA</li><li>• R</li><li>• TV-14</li><li>• PG-13</li><li>• PG</li></ul>
Canada	<ul style="list-style-type: none"><li>• R</li><li>• TV-MA</li><li>• PG</li><li>• PG-13</li><li>• TV-14</li></ul>

### e. Ratings Recommendation For TV Shows

This document presents the analysis about ratings of the TV show that need to be considered while getting listed.

These are the ratings for the content:

Country	Rating
United Kingdom	<ul style="list-style-type: none"><li>● TV-MA</li><li>● TV-PG</li><li>● TV-14</li><li>● TV-Y</li><li>● TV-G</li></ul>
United States	<ul style="list-style-type: none"><li>● TV-MA</li><li>● TV-14</li><li>● TV-PG</li><li>● TV-Y7</li><li>● TV-Y</li></ul>
Japan	<ul style="list-style-type: none"><li>● TV-14</li><li>● TV-MA</li><li>● TV-Y7</li><li>● TV-PG</li><li>● TV-Y</li></ul>
South Korea	<ul style="list-style-type: none"><li>● TV-14</li><li>● TV-MA</li><li>● TV-PG</li><li>● TV-Y7</li><li>● TV-Y</li></ul>
Canada	<ul style="list-style-type: none"><li>● TV-MA</li><li>● TV-Y7</li><li>● TV-Y</li><li>● TV-14</li><li>● TV-PG</li></ul>

While launching any new show or the movie it's critical to be aware about the favourable ratings in that country. As the rating is one of the major driving factors for content consumption.

**f. Launch Favourable Quarters For Movies**

- Here we present the analysis about considering which quarter to be considered while adding content to the platform in top 5 countries
- Following table lists the country wise ideal Quarterly release time for movies:

Country	Quarter
United States	Q1, Q2, Q3
United Kingdom	Q1
India	Q1, Q2
France	Q1
Canada	Q1

**g. Launch Favourable Quarters For TV Shows**

- Here we present the analysis about considering which quarter to be considered while adding content to the platform in top 5 countries
- Following table lists the country wise ideal Quarterly release time for TV Shows:

Country	Quarter
United Kingdom	Q1, Q2
United States	Q1, Q2, Q3, Q4
Japan	Q2, Q4
South Korea	Q4, Q2
Canada	Q4, Q2