**THE ANALYSIS REPORT**

**THE NETFLIX TITLES**

BY,

Piyush Upadhyay

Ponnoju karthika Bharani

Pradeep P

Pradeep S

Pramitha P

Pritesh Parmar

Ragavander S

Ravi Malapati

**INTRODUCTION**

Netflix.INC is an American [technology and media services provider](https://en.wikipedia.org/wiki/Over-the-top_media_service) and [production company](https://en.wikipedia.org/wiki/Production_company) headquartered in [Los Gatos, California](https://en.wikipedia.org/wiki/Los_Gatos,_California). Netflix was founded in 1997 by [Reed Hastings](https://en.wikipedia.org/wiki/Reed_Hastings) and [Marc Randolph](https://en.wikipedia.org/wiki/Marc_Randolph) in [Scotts Valley, California](https://en.wikipedia.org/wiki/Scotts_Valley,_California). The company’s primary business is its subscription-based [streaming service](https://en.wikipedia.org/wiki/Streaming_service_provider), which offers online streaming of a library of films and television series, including those produced in-house. Netflix is the world’s leading internet entertainment service with 158 million paid memberships in over 190 countries enjoying TV series, documentaries, and feature films across a wide variety of genres and languages.

Netflix is one of the leading streaming services worldwide, with a vast library of movies and TV shows available to subscribers. The Netflix Titles dataset provides a comprehensive list of titles available on Netflix across different countries, along with various attributes such as genre, rating, and release year. In this analysis report, we will explore the Netflix Titles dataset to gain insights into the content library, identify similar content, analyze the network of actors and directors, and understand the trends in content production by Netflix.

**METHODOLOGY**

We performed four tasks to gain insights into the Netflix Titles dataset. In Task 1, we analyzed the content library across different countries. In Task 2, we identified similar content by clustering titles based on text-based features. In Task 3, we analyzed the network of actors and directors. In Task 4, we analyzed the trends in content production by Netflix.

**DATA DESCRIPTION**

The data set consisted of 8807 Rows and 12 Columns. The columns and their descriptions were as listed below:

* **SHOW-ID** - Unique id of each show (not much of a use for us in this notebook)
* **TYPE** - Show category. Could be either a Movie or a TV Show
* **TITLE** - Name of the show
* **DIRECTOR** - Name of the director(s) of the show
* **CAST** - Names of Actors/ Actresses in the show
* **COUNTRY** - Countries where the show is available to watch on Netflix
* **DATE ADDED** - Date when the show was added to Netflix
* **RATING** - Show rating on Netflix
* **RELEASE YEAR** – The release year of the show
* **DURATION** - Time duration of the show
* **LISTED IN** - Genre of the show
* **DESCRIPTION - Brief insight into what the show is about**

**DATA CLEANING**

The data cleaning process involves identifying incorrect, incomplete, inaccurate, irrelevant, or missing pieces of data and then modifying, replacing, or deleting them as needed.

In the dataset, the following column contains the null values director,” “cast,” “country,” “date\_added,” listed in” and “rating.” To handle the null values without losing information, input the blank cells with “not mention”, and there are no duplicate values. we split the listedin into columns with comma values generate two more columns and take only one column for analysis

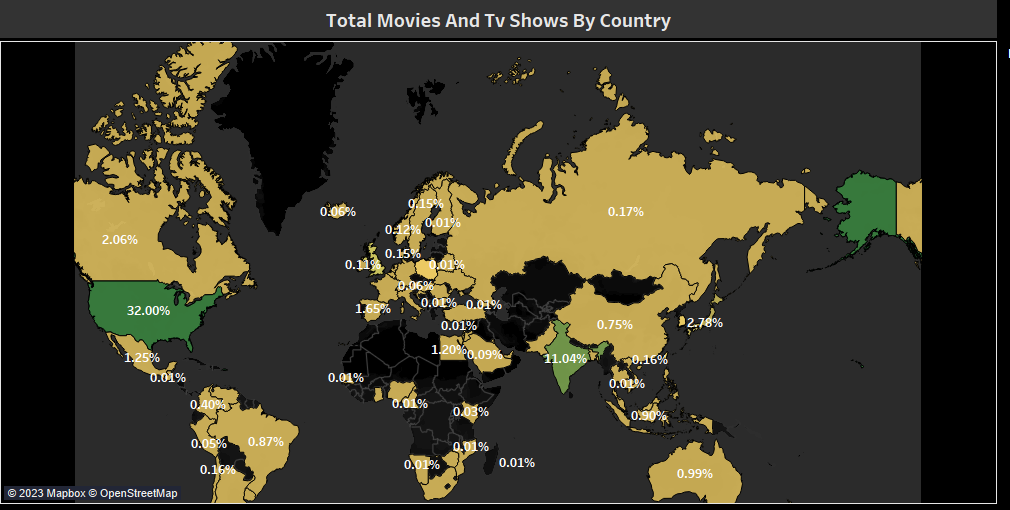
**DATA ANALYSIS**

**1-Understanding what content is available in different countries**

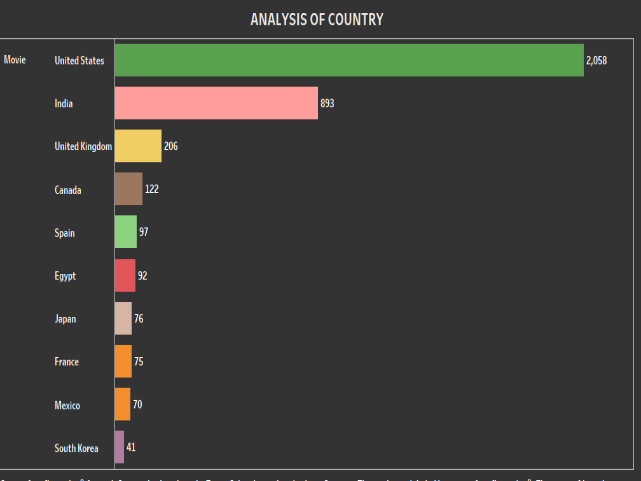
We can gain insights into the distribution of content across different countries and regions. We may find that certain countries, such as the United States, have a much larger library of titles than other countries. We can also look at the distribution of content by region, such as North America, Europe, Asia, and so on.

This can help us explore the content available in a particular country or compare the content across different regions. By understanding the distribution of content across different countries, we can gain insights into the cultural and regional preferences of Netflix users, as well as the strategies used by Netflix to expand its global reach.

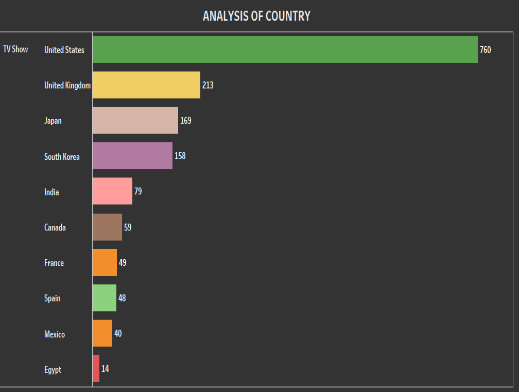
by analyzing the content available in different countries, we can gain insights into the cultural and regional preferences of Netflix users. This can be useful for content creators and distributors who want to understand what types of content are most popular in different regions, and how to tailor their offerings to different audiences.



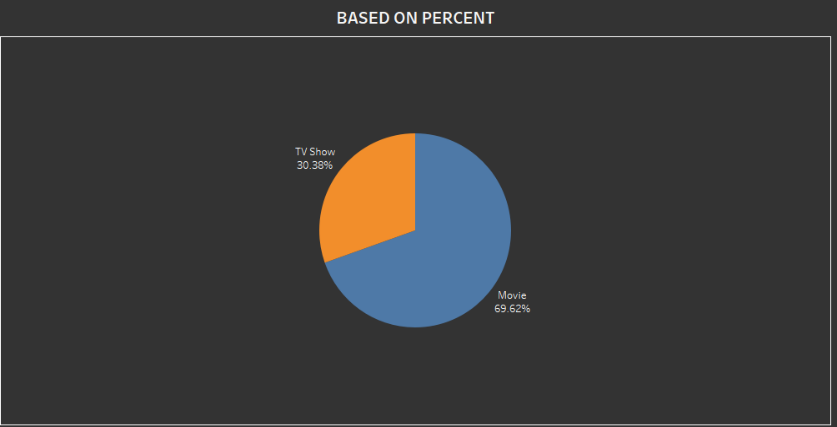
The above map chart are showing how much content is produced by each country. From the graph, we can see that the USA produces 32% of content out of the total which makes the country majority of content producers, and 2nd in India which is giving 11.04% of the content.



The above list chart is showing top ten countries according to the number of movies they produce. Here USA is producing 2058 movies which is more than double that of the 2nd country who is producing 893 and India is 4 times more movies than the 3rd UK.



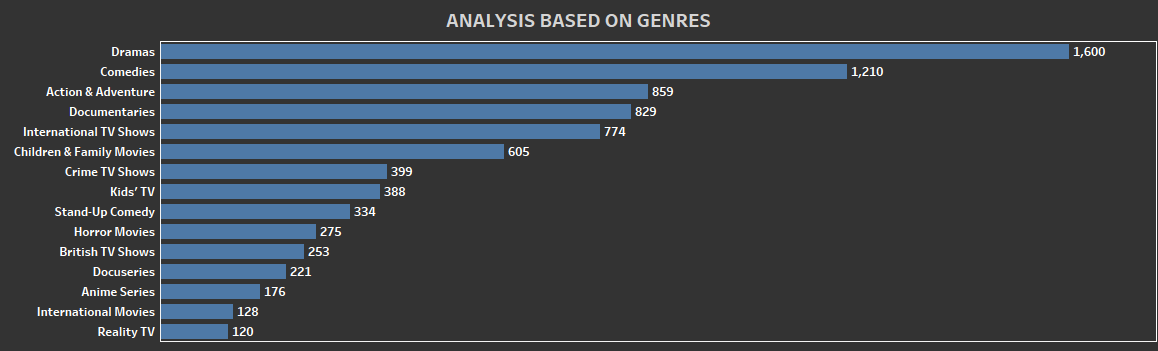
The above list chart is showing top ten countries according to the number of TV shows they produce. Here USA is again 1st by producing 760 shows. 2nd is the UK producing 213 shows and Japan is 3rd producing 169 shows.



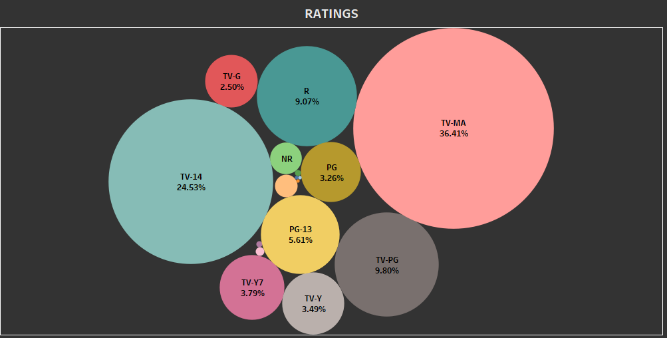
There are total 8807 content data we have In which 2676(30.38%) are TV shows and 6131(69.62%) are movies

**2-Identifying similar content by matching text-based features**

we can create a visualization that shows the frequency of different terms in the title and description of the titles in the dataset. This can help us identify common themes and topics that appear in multiple titles. by analyzing text-based features in the Netflix Titles dataset, we can identify similar content, group titles based on common themes or topics, and create recommendations for users. These insights can be valuable for content creators, distributors, and streaming platforms who want to understand audience preferences and tailor their offerings to different viewers.



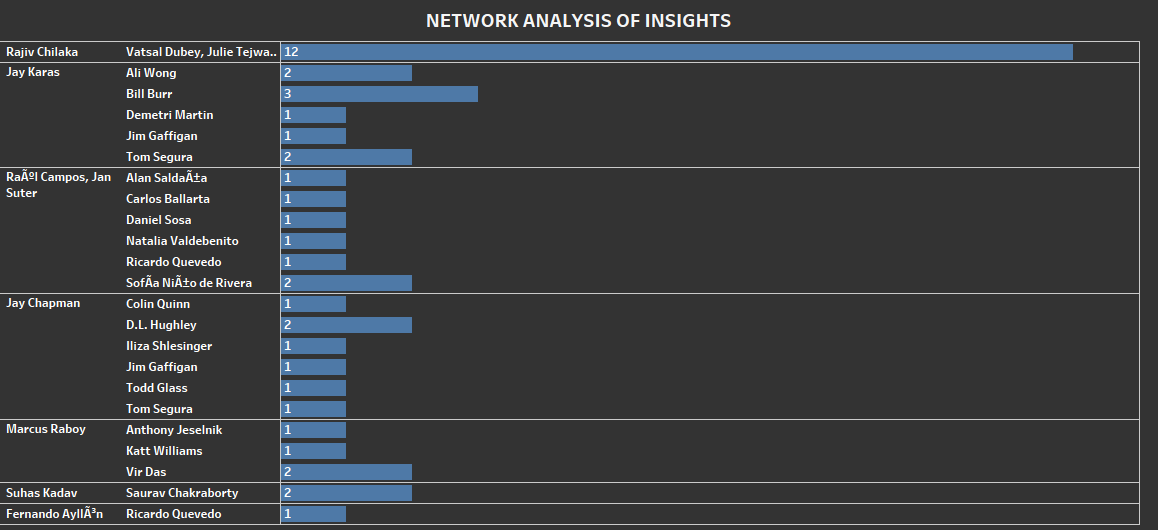
Here above bar chart is showing us the top 15 genres that Netflix is producing. Netflix is making the majority of its shows Dramas, 2nd is a comedy, and 3rd is Action & Adventure.



Here above is a bubble chart of ratings of Netflix shows. Out of 17 rating types, Netflix is making TV-MA rating shows 36.41%, 2nd is TV-14 which is 24.53%, and 3rd is TV-PG with 9.8%.

**3-Network analysis of Actors / Directors and find interesting insights**

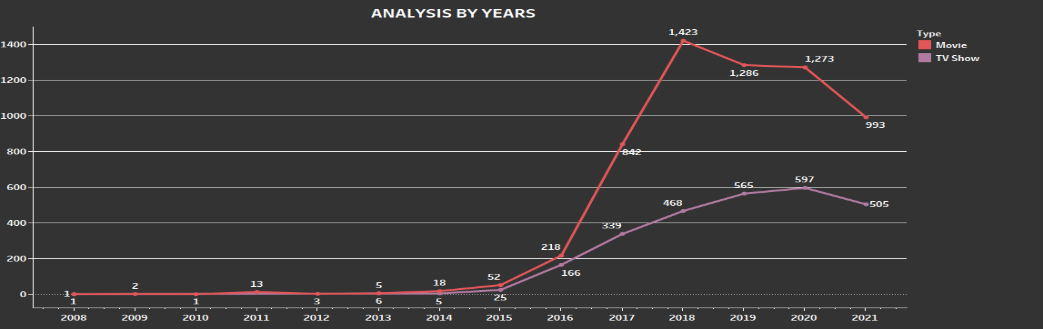
By using network analysis techniques, we can explore the relationships between different actors and directors and gain insights into the structure of the film and television industry. This visualization can help us identify which actors and directors have worked together frequently and which ones tend to work in different circles. We can also use identify clusters of actors and directors who tend to work together frequently or who have similar styles or genres. by using network analysis techniques to explore the relationships between actors and directors we can gain valuable insights into the structure of the film and television industry, identify clusters of actors and directors who tend to work together, and create recommendations for users based on their preferences.



Here the above chart is showing director and actor the number of times they work together. From the chart, we observe that Rajiv chilaka and vatsal dubey and his team worked together 12 times. And most of the other directors and cast are working together 1 to 3 times.

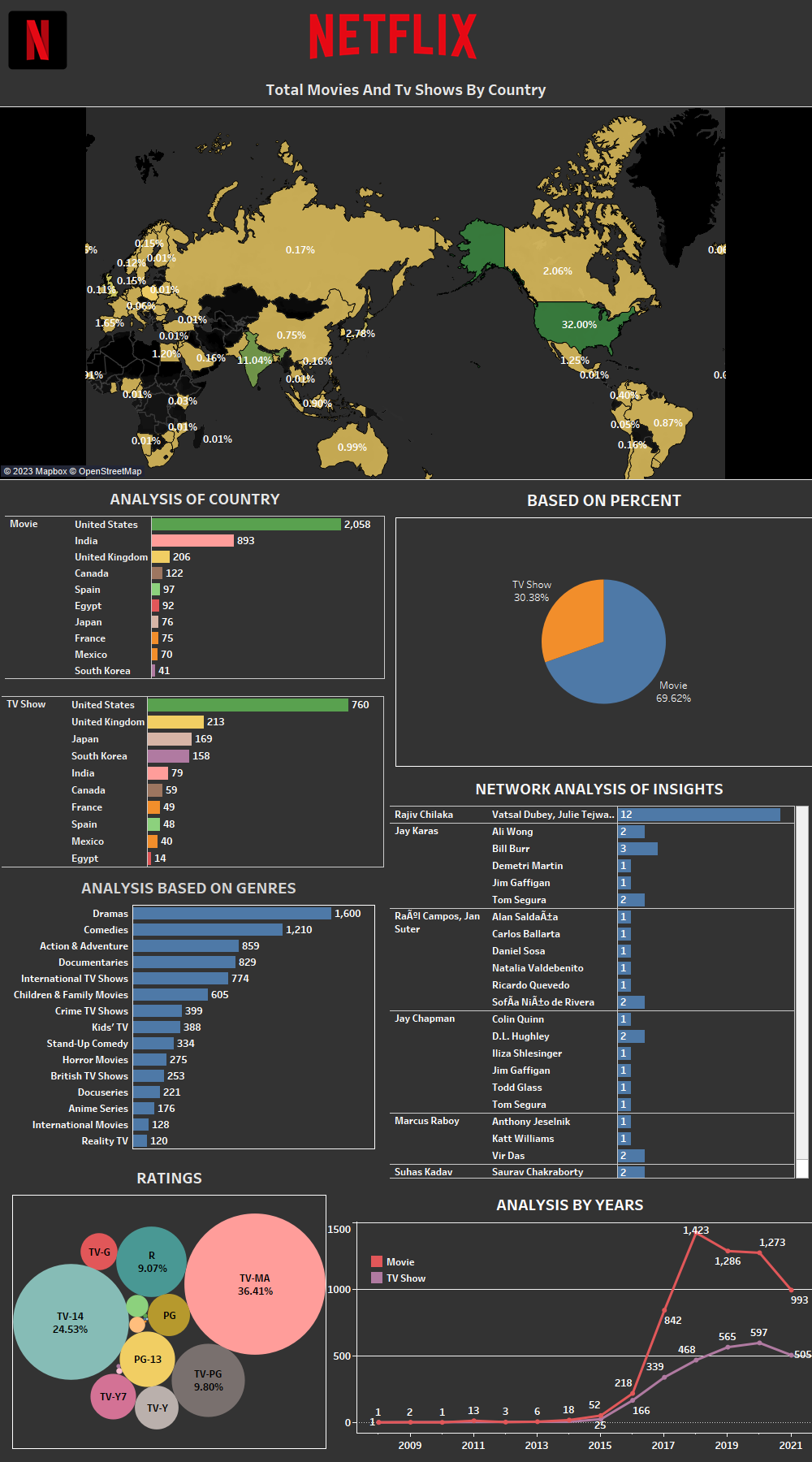
**4-Does Netflix has more focus on TV Shows than movies in recent years?**

The visualization shows the number of TV shows and movies added to the platform each year. By plotting this data over time, we can identify whether there has been a shift in focus toward TV shows in recent years. we can analyze if any patterns in the distribution of TV shows and movies across different genres. If we see a higher proportion of TV shows in certain genres, such as drama or comedy, this could also indicate a shift in focus toward TV shows. analyzing trends in the distribution of titles by type (TV shows vs. movies) over, we can gain insights into whether Netflix has more focus on TV shows than movies in recent years.



Here above we have a timeline chart that shows how many movies and TV shows are produced every year. We have data from the years 2008 to 2021 and from the graph, we can see that as the number of years increases movies and TV shows are released more and more. After 2018 Netflix decreases their movies from the previous year and continue to produce tv shows. Due to covid situation b, both categories have fallen in the number of releases.

**DATA VISUALIZATION**



**CONCLUSION**

Netflix has a major presence in the United States and India. Its focus has shifted recently towards TV shows which were previously on movies. More TV shows are released than movies. the Netflix Titles dataset provides a wealth of information for exploring various aspects of the streaming service. By applying different analytical techniques, we can gain insights into the content library, similarities between titles, the dynamics of the entertainment industry, and trends in content production.