

Analysis of Movies and TV Shows on Netflix

By: Akin Pham Vu, Syed Fawad, Syed Sajid
5-7-2021

Executive Summary

Netflix is the largest streaming service with over 204 million subscribers. With so many users, Netflix must figure out which movies and TV shows are most popular to create and cater better content. In addition, in recent years several competitors have launched their own streaming services such as Disney+ and HBO MAX. The goal of our project is to analyze Netflix's data to find insights and trends to help them better understand their consumers.

Our team used data from a third-party Netflix search engine called Flixable. Before we could conduct any useful analysis of the data, we had to import and clean the data. There were several columns with missing values. To solve this issue, we either replaced the values or dropped them. After cleaning the data, we created several charts to make comparisons between several variables.

We found that more than two-thirds of the content on Netflix are movies. However, recently Netflix has been adding significantly more TV shows to their catalog. Other insights include the most popular genres, countries that create the most content, content trends over the years, and more.

The greatest insights we gained from this analysis are the fact that Netflix adds more movies each year compared to tv shows, however the gap has been shrinking for the last few years. This indicates a shift in the content that Netflix should plan to license or create. In addition, international movies, dramas, and comedies, are the most popular genres available. This shows a trend of what users typically watch and can once again help Netflix better cater to their consumers.

Although we gained valuable data, our analysis was limited due to having minimal quantitative data categories. With our categorical data types, we could potentially recommend movies to users based off the description and tf-idf. Another way to improve our data analysis in the future is to import additional data from sources such as IMDb or TMDb to help analyze popular content and make recommendations. This way we can find greater insights and further improve Netflix.

Libraries Used

We imported a few libraries to aid in our analysis. These libraries include:

NumPy – To help perform algebraic functions.

Pandas – For data manipulation and analysis. We used Pandas to import the data from Flixable, cleanse the data of null or unneeded values, and create tables to aid with making charts.

Matplotlib – To aid in creating graphs and charts. This was used to create various charts and customize various aspects such as the labels and values.

Seaborn – Used to further customize charts. This library adds additional functionality on top of Matplotlib. We used Seaborn to further customize our charts and change the theme.

Data

Our team used data provided by Flixable, a third-party Netflix search engine. This data included numerous categories including show id, type, cast, title, director, country, rating, release year, date added, duration, description, and listed in. After importing the data, we noticed that the categories 'director', 'cast', and 'country' were missing numerous values. Therefore, we replaced those null values with string inputs. The categories 'date added', and 'rating' were only missing a few values, so we decided to drop missing rows entirely.

```
Total of Columns with Missing Values:
show_id      0
type         0
title        0
director    2389
cast        718
country     507
date_added   10
release_year  0
rating       7
duration     0
listed_in    0
description  0
dtype: int64
```

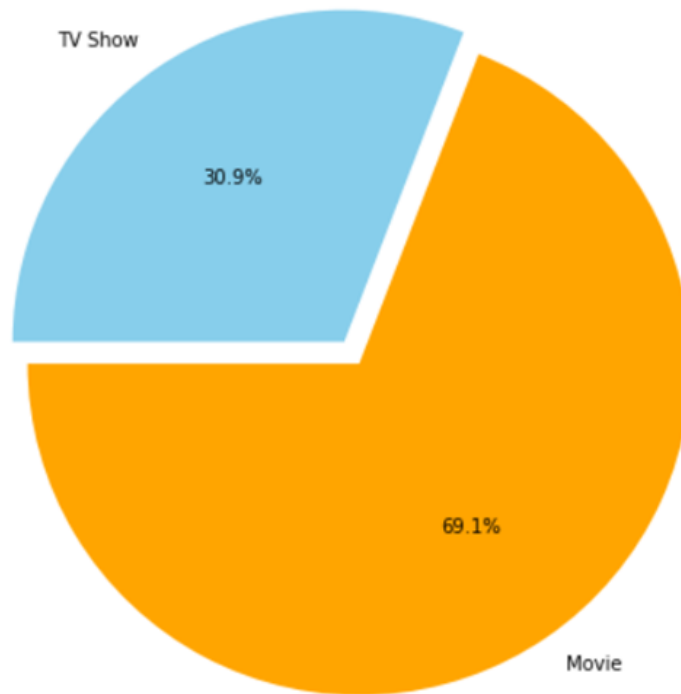
```
After Data Cleansing
show_id      0
type         0
title        0
director     0
cast         0
country      0
date_added   0
release_year  0
rating       0
duration     0
listed_in    0
description  0
dtype: int64
```

Results:

Once we cleaned the data, we were able to pull accurate information and create charts and graphs. In total we created nine graphs that tell us more about the dataset.

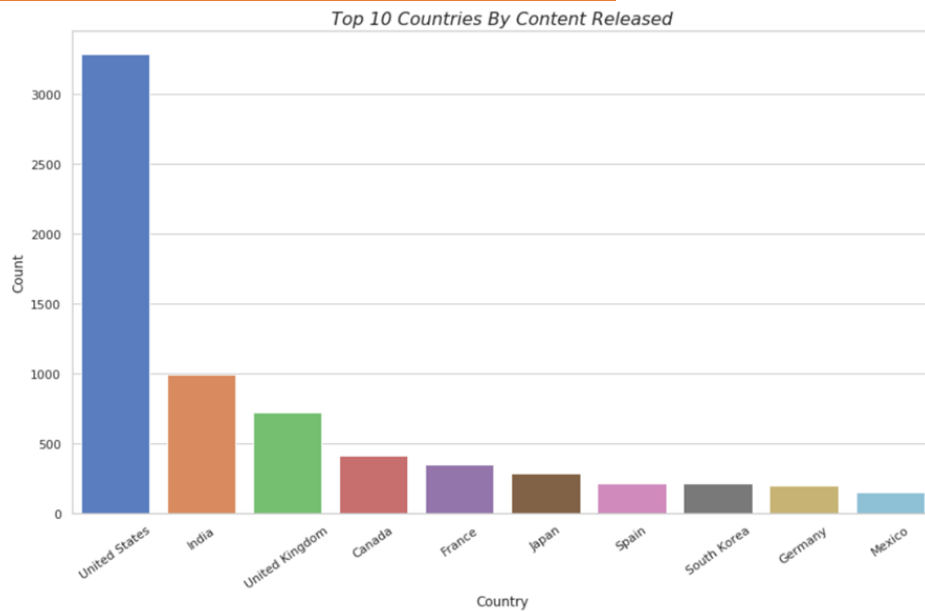
Percentage of TV Shows Compared to Movies

Percentage of Movies vs. TV Shows



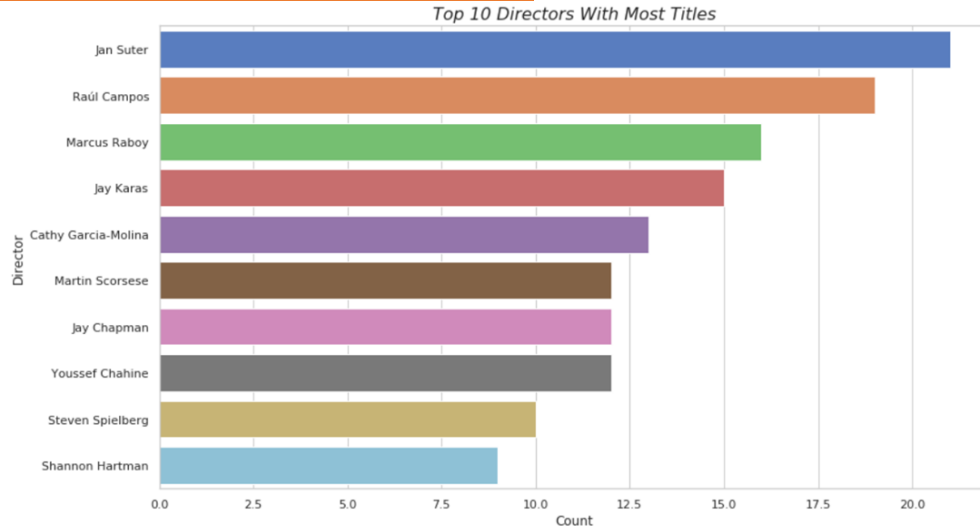
First, we compared the amount of movies vs tv shows. We found that Netflix consist of 69.1% movies and 30.9% tv shows.

Top 10 Countries by Content Released



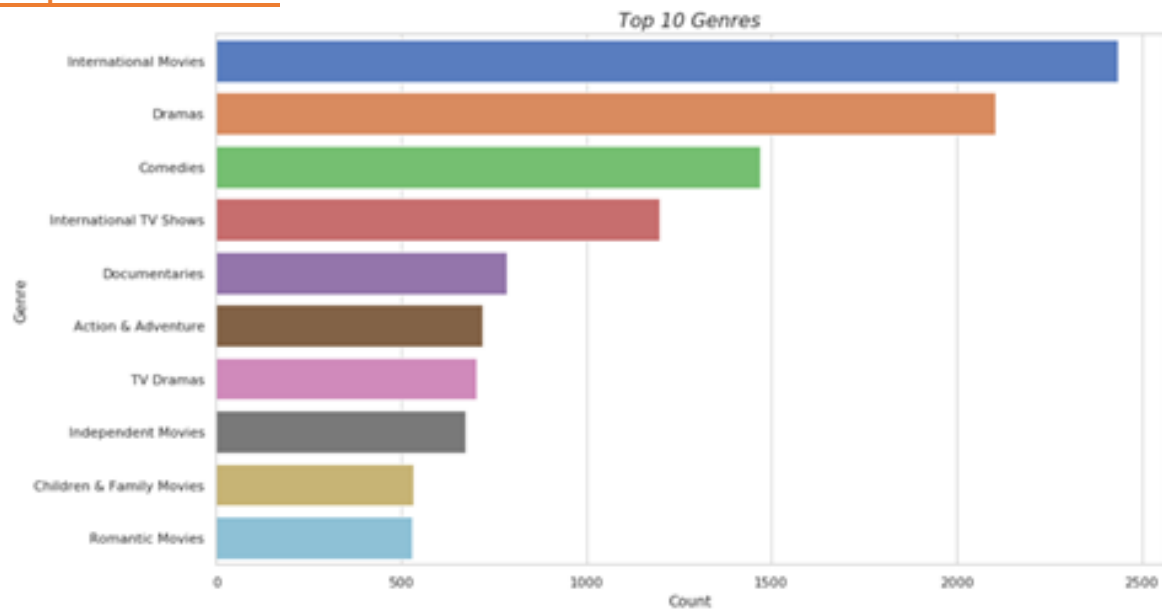
Second, we sorted the top 10 Countries by amount of content released. The United States is at the top of the list having more than 3x the amount content by the next leader India.

Top 10 Directors with Most Titles



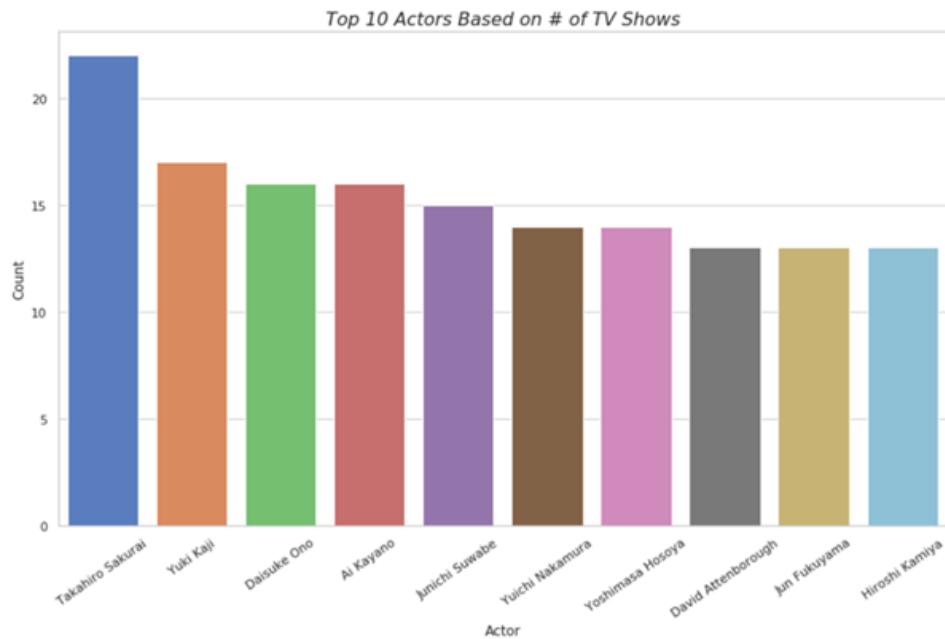
Next, we have the top 10 directors with most titles directed. At the top of the list there is Jan Suter. Among the top 10 are some famous directors like Martin Scorsese and Steven Spielberg.

Top 10 Genres



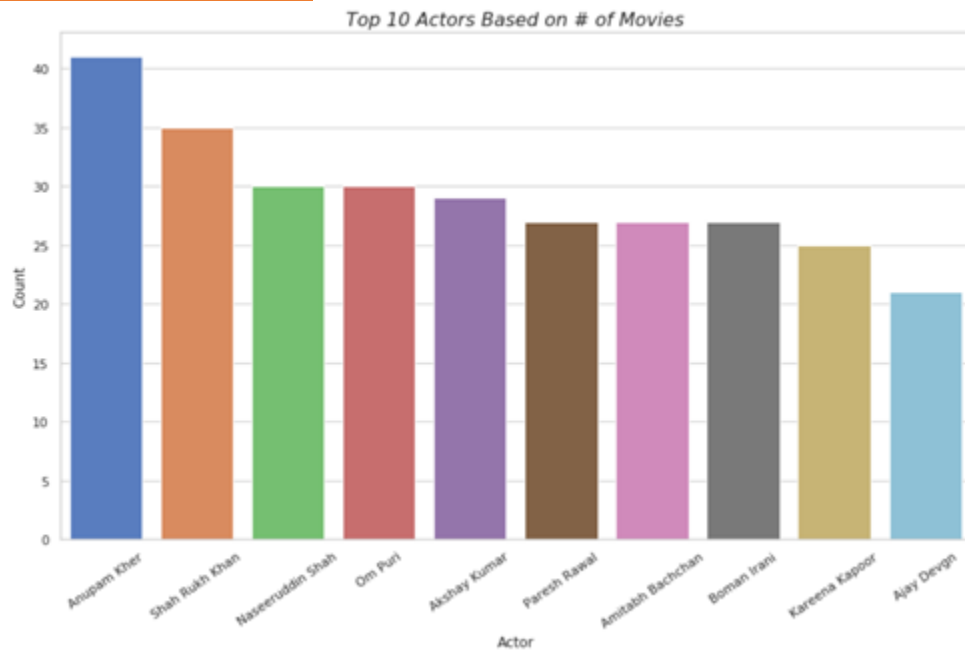
Next, we have the top ten genres. The first is international movies followed by dramas and comedies.

Top Ten Television Actors



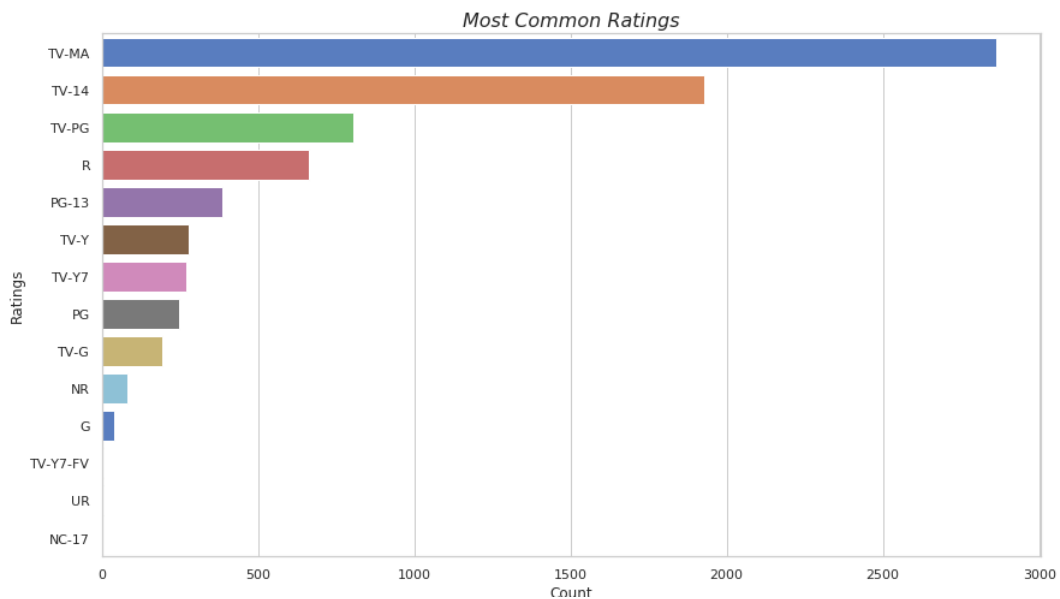
Here we have the top ten actors based on the amount of tv shows they have appeared in. Nine of the top ten are Japanese Voice Actors. The sole member who is not, has narrated numerous documentaries for the BBC.

Top Ten Movie Actors



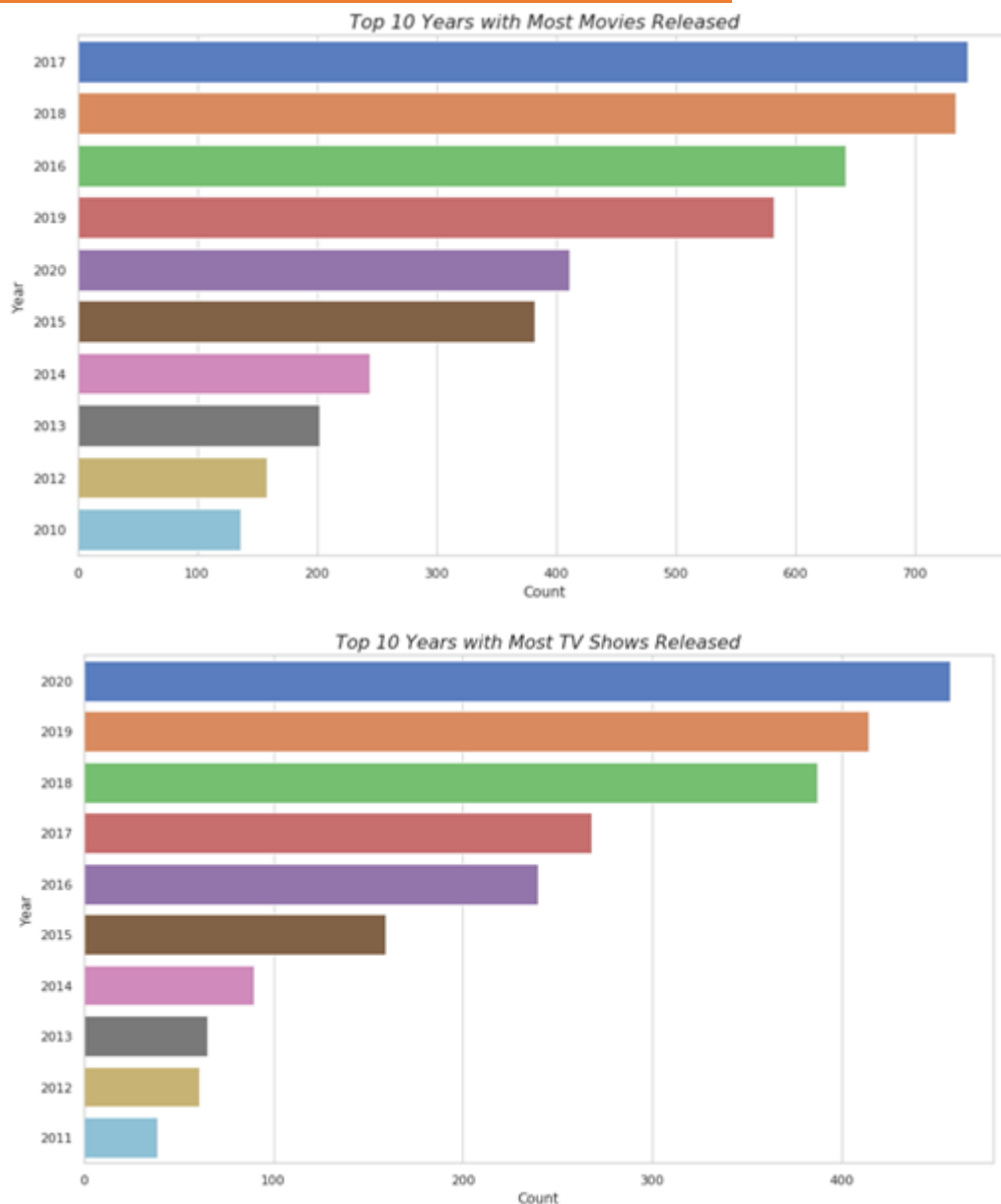
Here we have the top ten movie actors, based on the number of movies they have appeared in. All ten are Bollywood actors

Common Ratings



Next, we have a graph that shows the most common ratings on Netflix. The most common rating is TV-MA followed by TV-14 and TV-PG

Top Ten Years with Most Movies/TV Shows



Lastly, we have the two graphs here. One shows the top ten years with most movies and released and the other shows the most tv shows. When looking at the top years for movies, 2017 was the peak and the number of movies has steadily declined in the years after. Whereas when you look at the top ten years for tv shows, there has been a year over year increase since 2011. If you compare the two charts Netflix is still adding more movies than tv shows but that number is decreasing. If this trend continues soon TV shows will be added more than movies. This will change the initial content spread we found of 69.1 % Movies and 30.1% TV Shows, to a more favorable spread for tv shows.

Conclusion

We analyzed the data set provided by Flixable that consist of what content Netflix offers. We used several libraries to analyze the code, NumPy, Matplotlib, Seaborn, and Pandas. After importing the data, we first had to clean it. This reduces errors and missing information in the data and will provide more accurate reports. We came away with these key findings

- Movies make up 69% percent of content, TV Shows 31%
- The United States produces the most content
- Jan Suter has directed the most titles
- International movies are the top genre followed by dramas and comedies
- 9/10 of the top tv actors based on quantity are Japanese Voice Actors
- All the top movie actors based on quantity are Bollywood Actors
- The most common rating is TV-MA followed by TV-14 and TV-PG
- More movies are being added than TV shows, but the gap is shrinking

If these trends continue, soon TV Shows will make up for a much larger percentage of content on Netflix in a few years, eventually surpassing the number of movies.