

Capstone Project

NETFLIX MOVIES /TV SHOWS

CLUSTERING

By – RAHUL SINGH WALDIA

Netflix, Inc. is an American technology and media services provider and production company headquartered in Los Gatos, California. Netflix was founded in 1997 by Reed Hastings and Marc Randolph in Scotts Valley, California. The company's primary business is its subscription-based streaming service, which offers online streaming of a library of films and television series, including those produced in-house.

Netflix is a popular entertainment service used by people around the world. This EDA will explore the Netflix dataset through visualizations and graphs using python libraries, matplotlib, and seaborn.

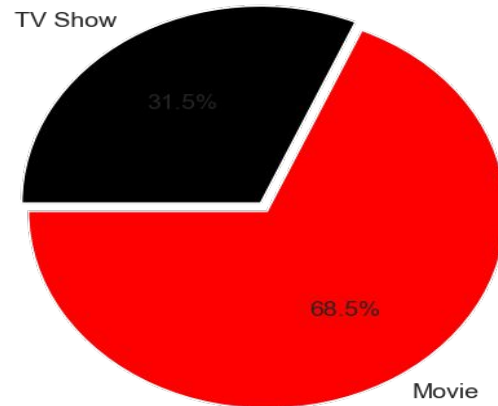


NETFLIX

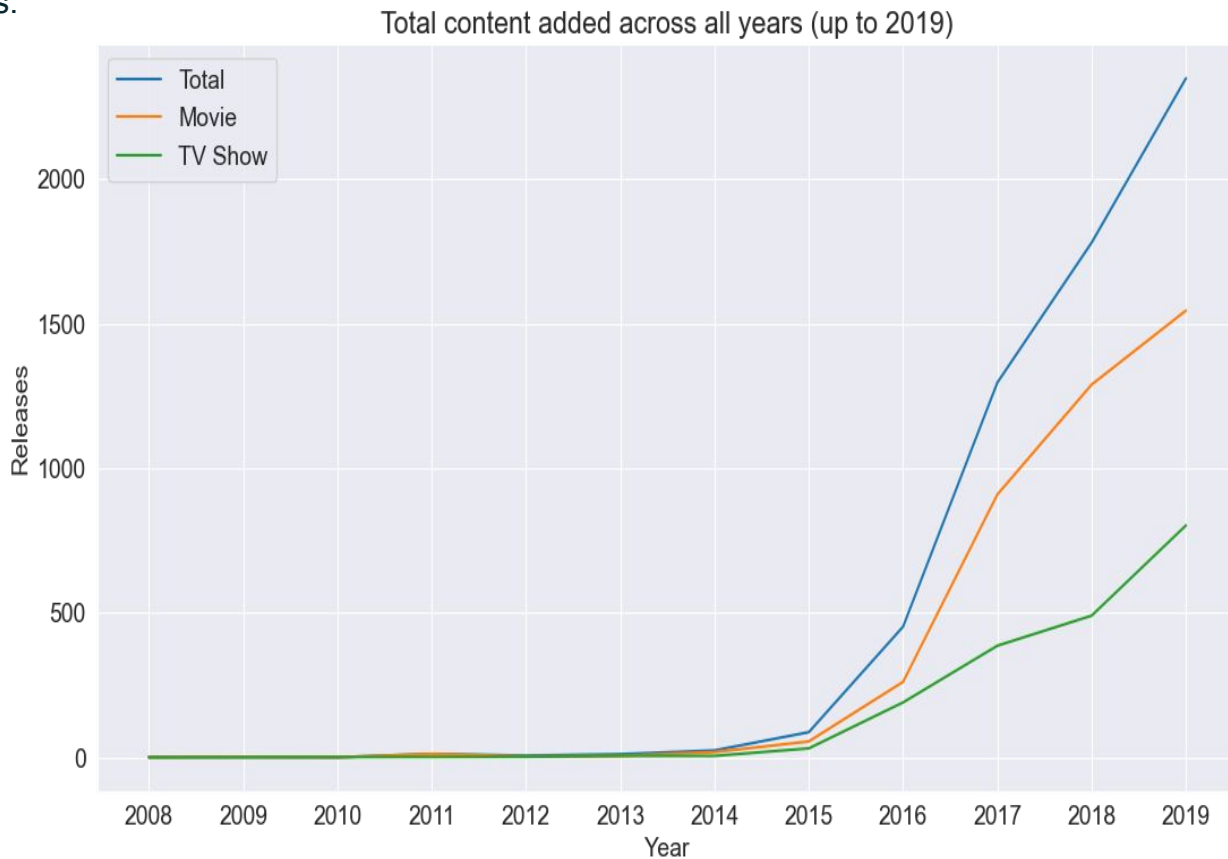
Let's check the first 5 data . The dataset contains over 6234 titles, 12 descriptions. After a quick view of the data frames, it looks like a typical movie/TVshows data frame without ratings. We can also see that there are NaN values in some columns.

Analysis entire Netflix dataset consisting of both movies and shows. Let's compare the total number of movies and shows in this dataset to know which one is the majority. So there are about 4,000++ movies and almost 2,000 TV shows, with movies being the majority. There are far more movie titles (68,5%) that TV shows titles (31,5%) in terms of title.

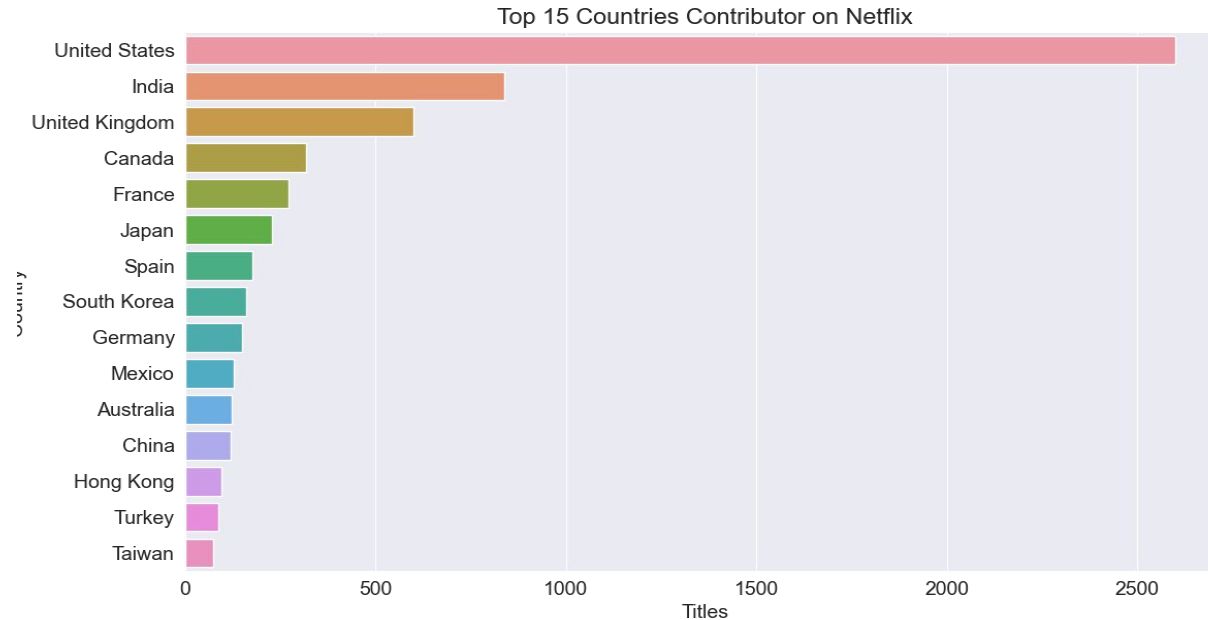
Percentation of Netflix Titles that are either Movies or TV Shows



Next, we will explore the amount of content Netflix has added throughout the previous years. Since we are interested in when Netflix added the title onto their platform, we will add a “year_added” column to show the date from the “date_added” columns.

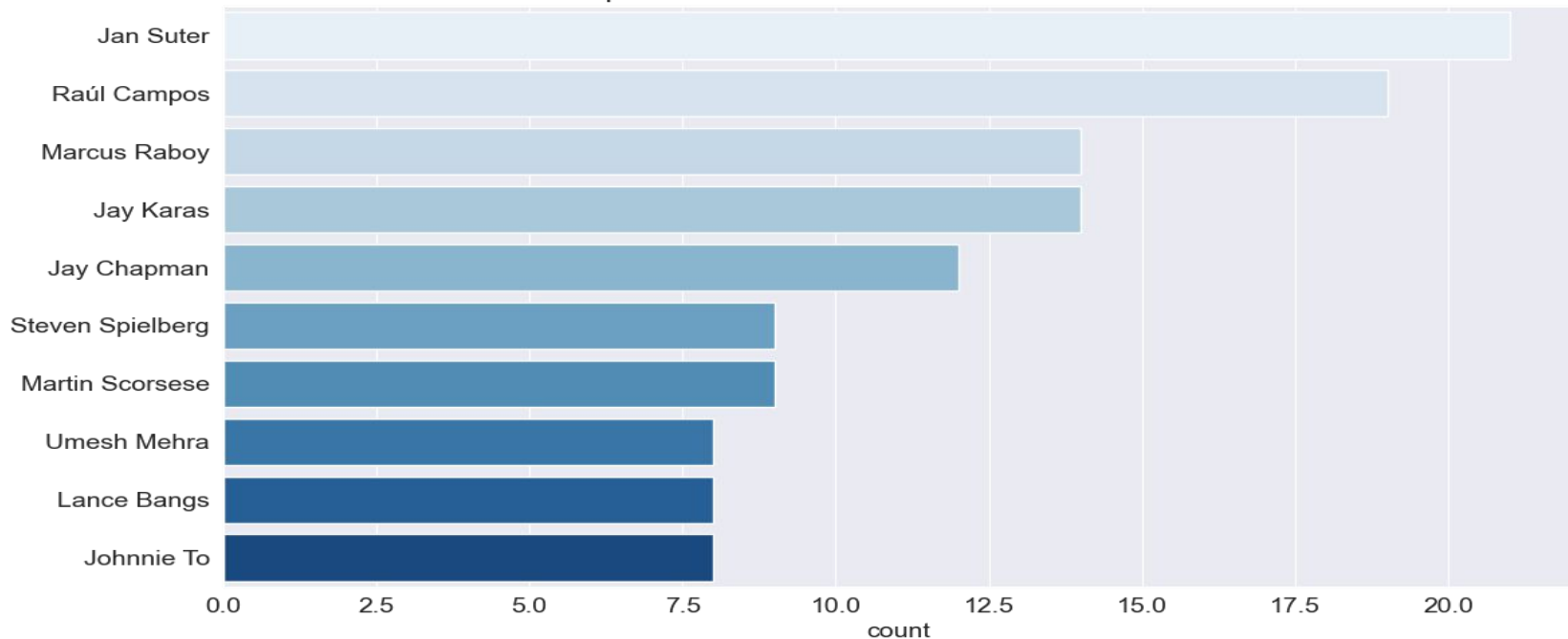


Next is exploring the countries by the amount of the produces content of Netflix. We need to separate all countries within a film before analyzing it, then removing titles with no countries available. From the images above, we can see the top 15 countries contributor to Netflix. The country by the amount of the produces content is the United States.

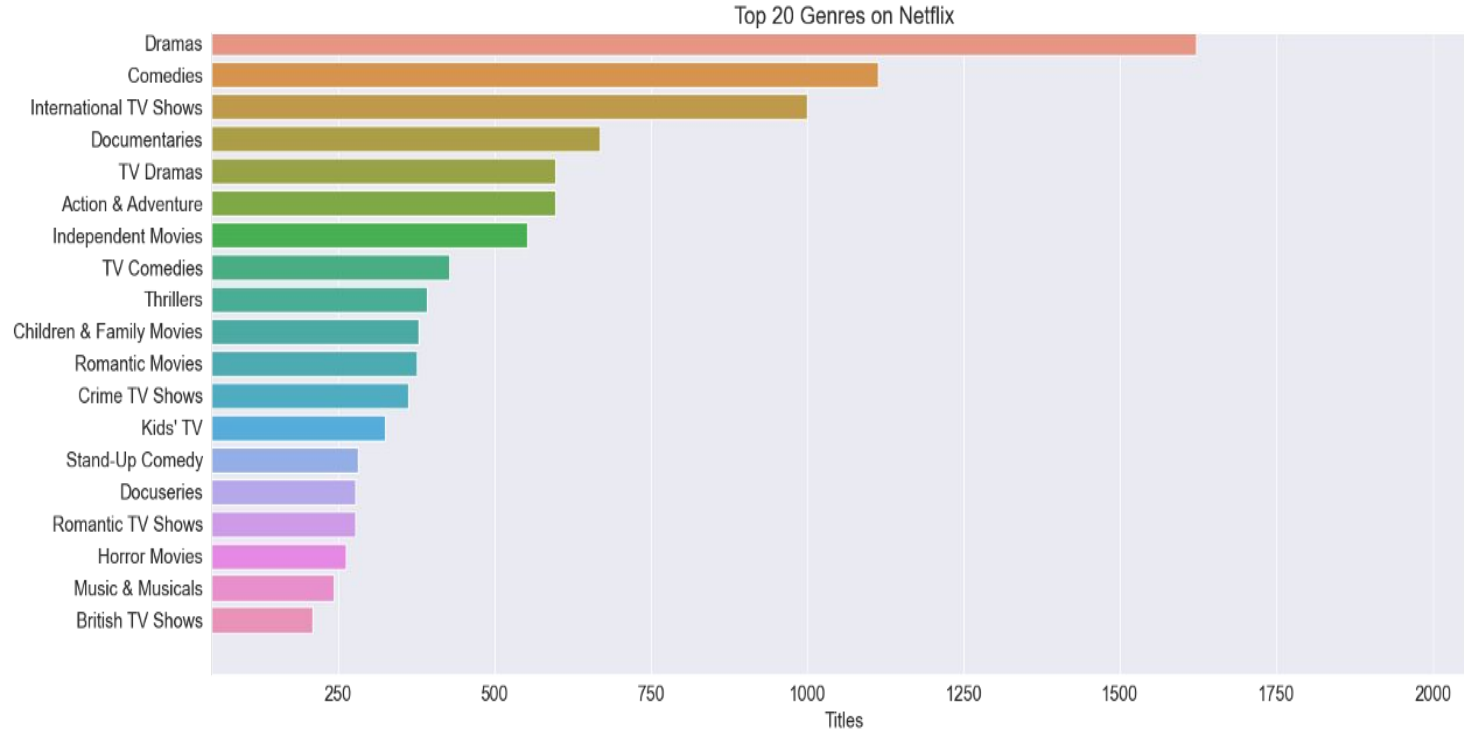


To know the most popular director, we can visualize it. The most popular director on Netflix, with the most titles, is mainly international.

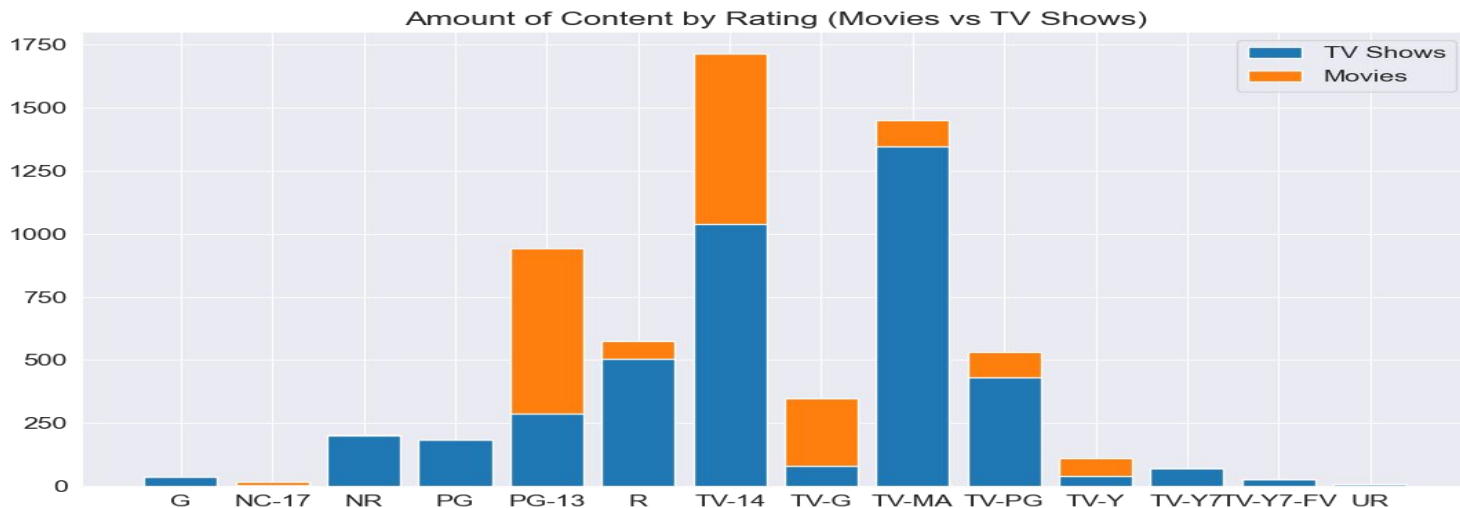
Top 10 Director Based on The Number of Titles



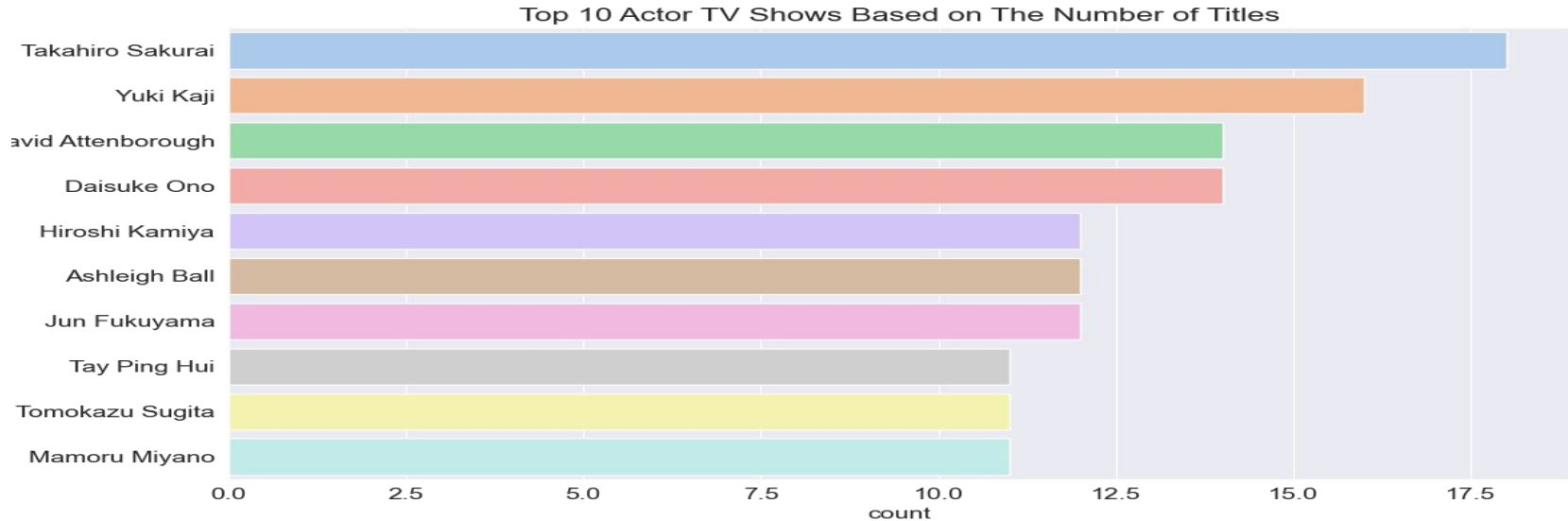
From the graph, we know that International Movies take the first place, followed by dramas and comedies.



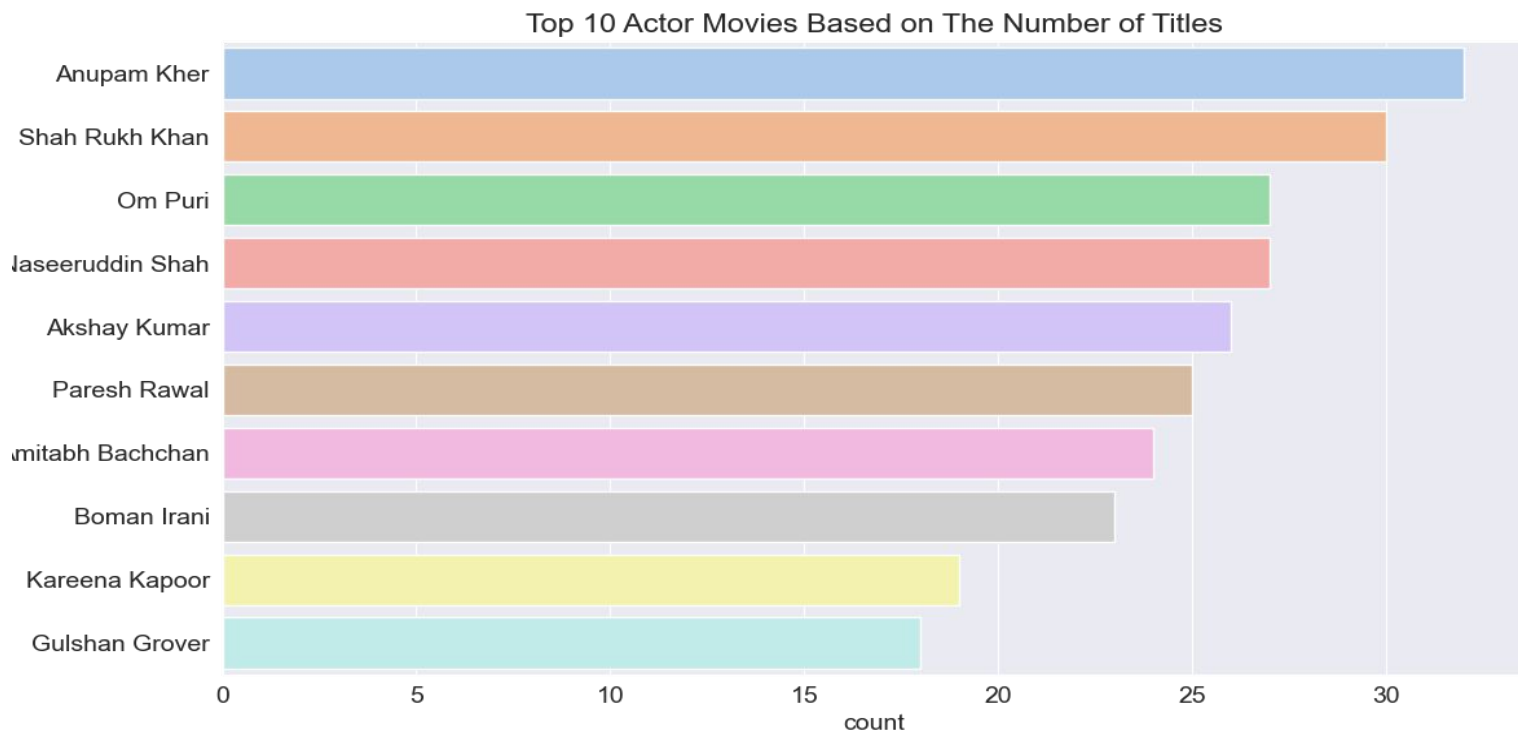
The largest count of Netflix content is made with a “TV-14” rating. “TV-14” contains material that parents or adult guardians may find unsuitable for children under the age of 14. But the largest count of TV shows is made with a “TV-MA” rating. “TV-MA” is a rating assigned by the TV Parental Guidelines to a television program designed for mature audiences only.



The top actor on Netflix TV Show, based on the number of titles, is Takahiro Sakurai.



The top actor on Netflix Movies, based on the number of titles, is Anupam Kher.



Conclusions

We have drawn many interesting inferences from the dataset Netflix titles; here's a summary of the few of them:

The most content type on Netflix is movies,

The popular streaming platform started gaining traction after 2014. Since then, the amount of content added has been increasing significantly,

The country by the amount of the produces content is the United States,

The most popular director on Netflix , with the most titles, is Jan Suter.

International Movies is a genre that is mostly in Netflix,

The largest count of Netflix content is made with a “TV-14” rating,

The most popular actor on Netflix TV Shows based on the number of titles is Takahiro Sakurai,

The most popular actor on Netflix movie, based on the number of titles, is Anupam Kher.