Netflix Data : Cleaning, Analysis and Visualization

Data Science Project

BY : Siddharth Rawat

INTRODUCTION

Netflix is a popular streaming service that offers a vast catalog of movies, TV shows, and original contents. The data consist of contents added to Netflix from 2008 to 2021. The oldest content is as old as 1925 and the newest as 2021. This datasets about Netflix Movies & TV Shows. Datasets have 12 columns with some null values. To analysis of dataset are used Pandas, plotly.express and Datetime libraries. Analysis process I divided into several parts for step wise analysis and to find out trending questions on social media for Bollywood actors and actress. The purpose of this dataset is to test my data cleaning and visualization skills.

PROBLEM STATEMENT

- **Content production :** Netflix can use data analysis to decide what content to produce and upload.

- **Release timing :** Netflix can use data analysis to determine the optimal timing for releasing content.

- **Subscriber engagement :** Netflix can use data analysis to create detailed profiles of its subscribers to better engage with them. For example, Netflix can analyze whether users binge-watch entire seasons of TV shows.

- **Viewer preferences :** Netflix can use data analysis to understand viewer preferences, such as whether there's a correlation between the number of seasons a TV show has and its viewer ratings.
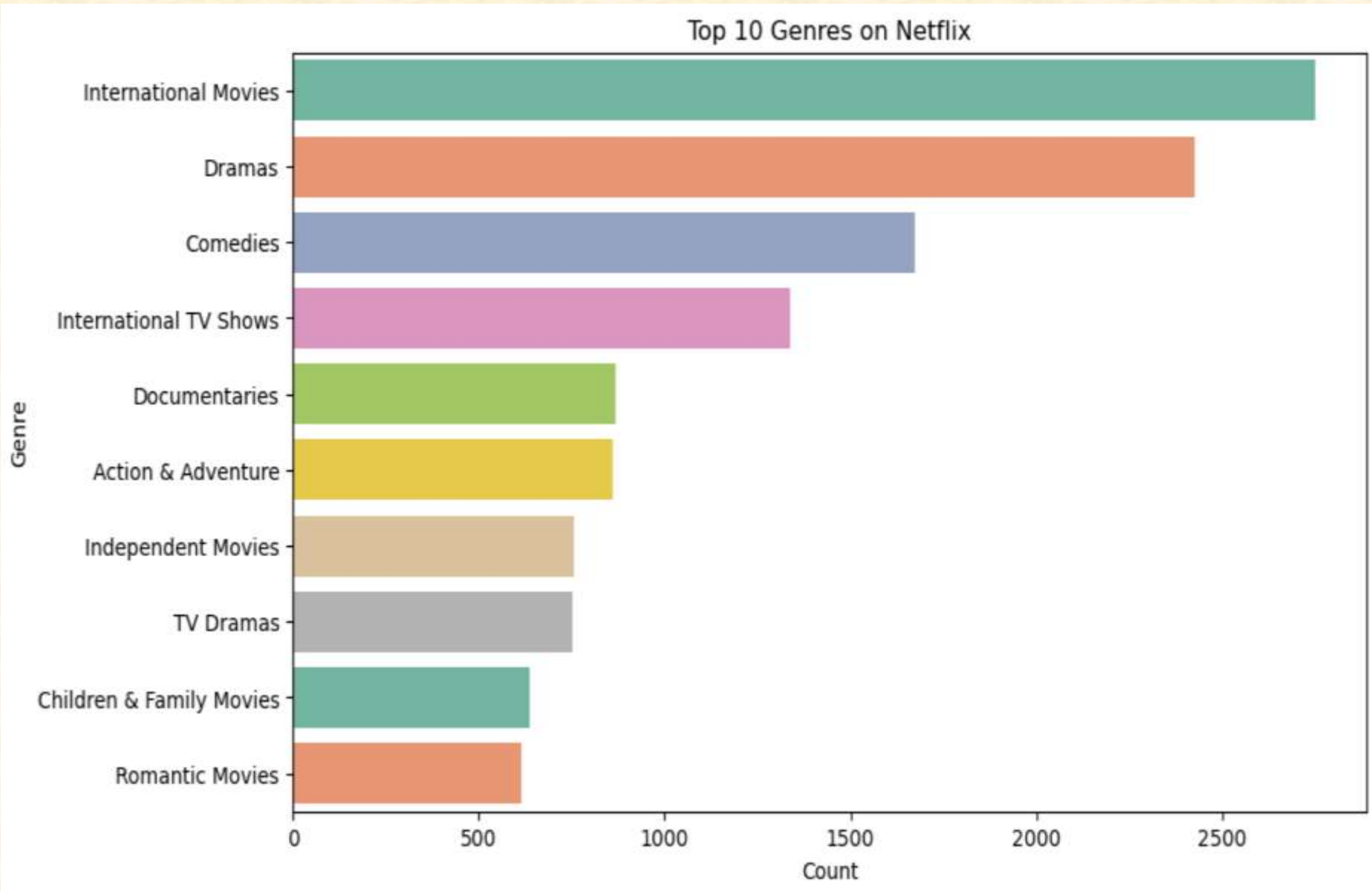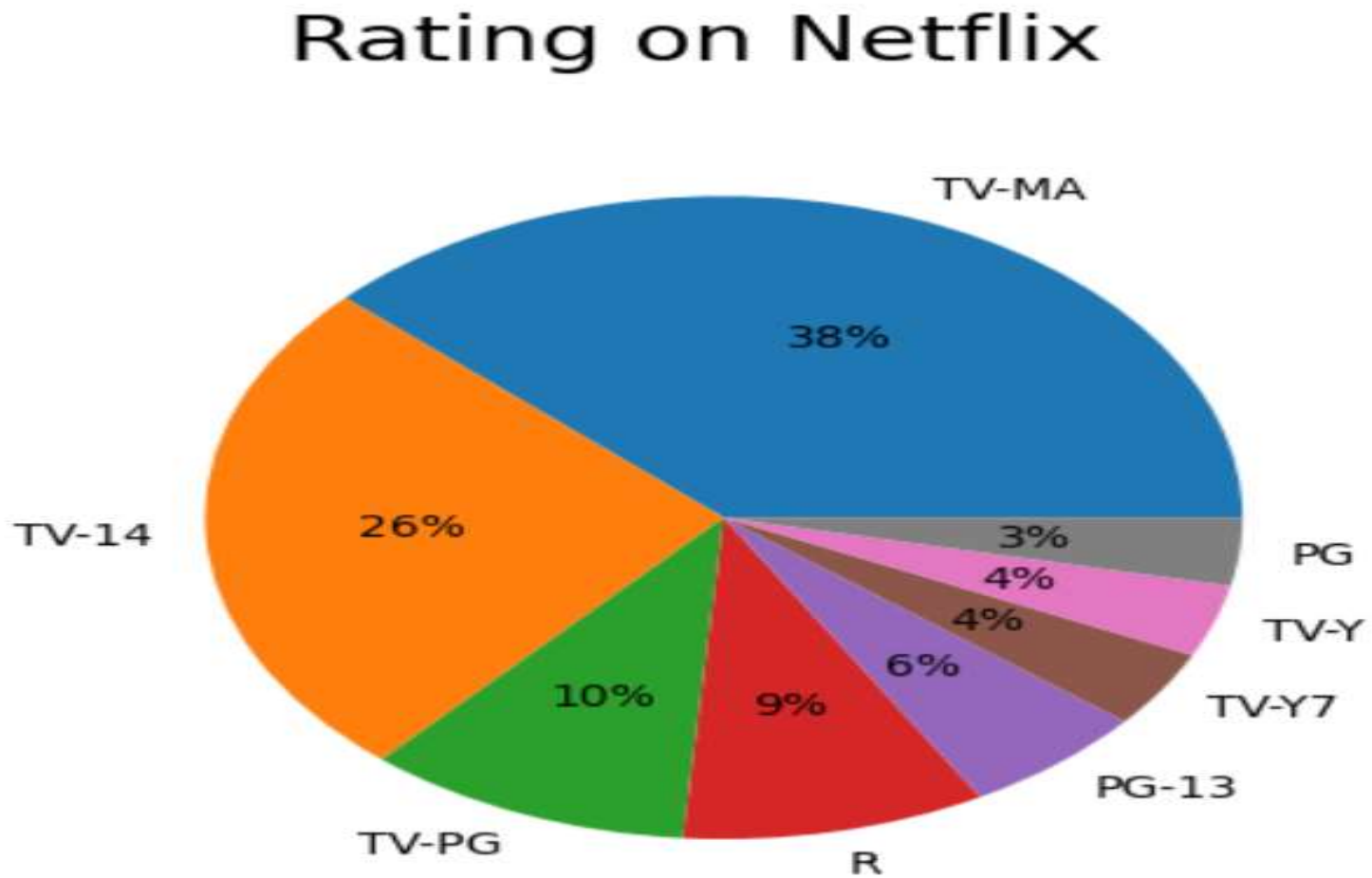
TOTAL CONTENT ON NETFLIX

Total Content on Netflix
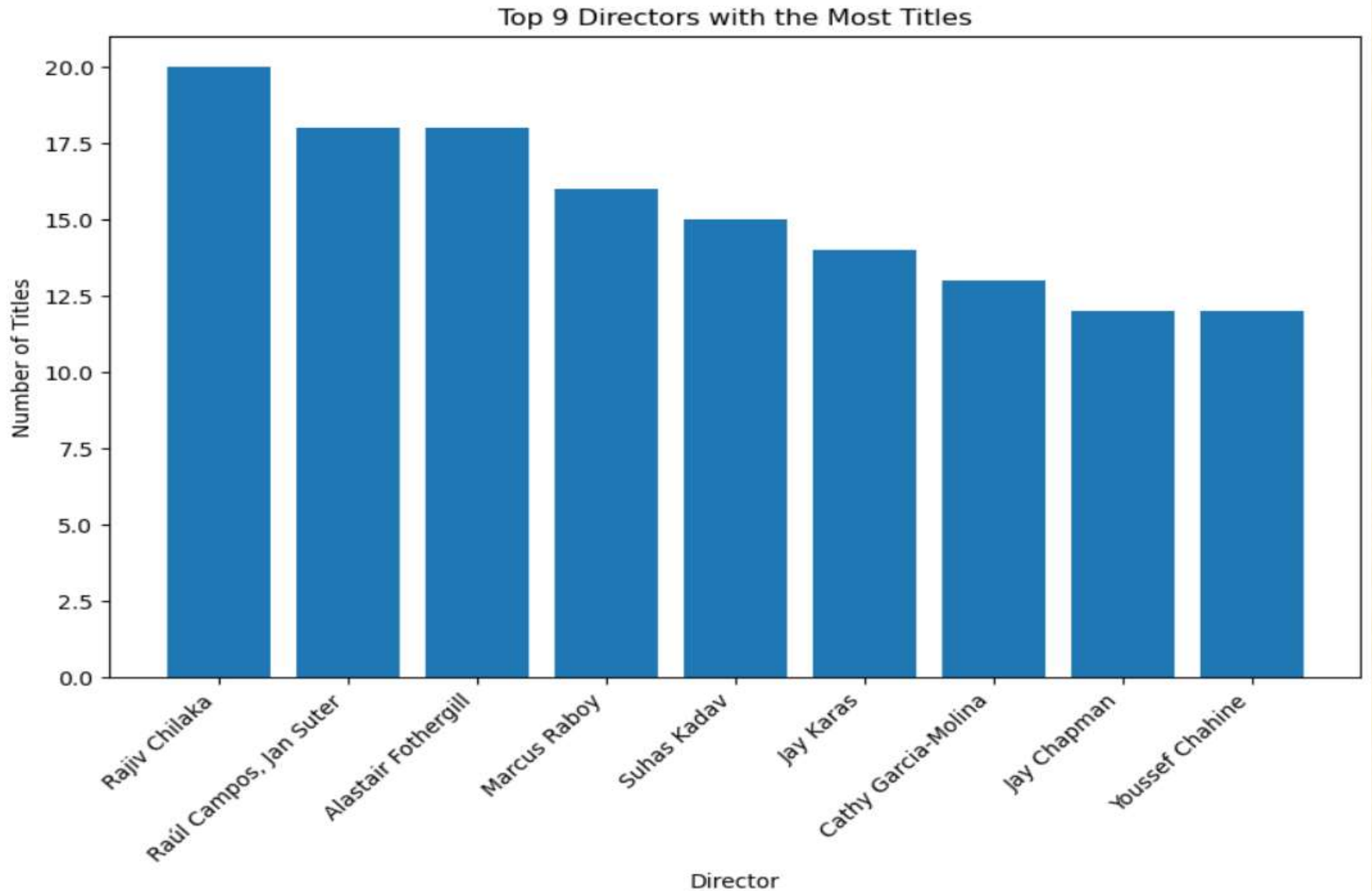
TOP 10 COUNTRIES WITH MOST CONTENT ON NETFLIX
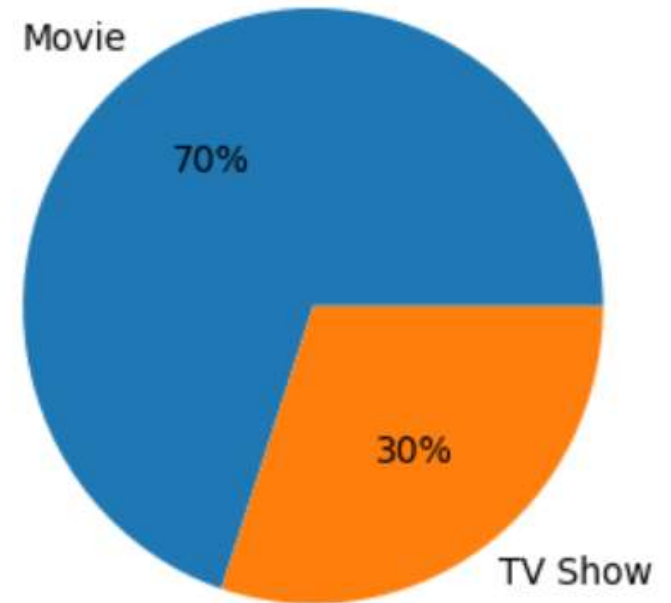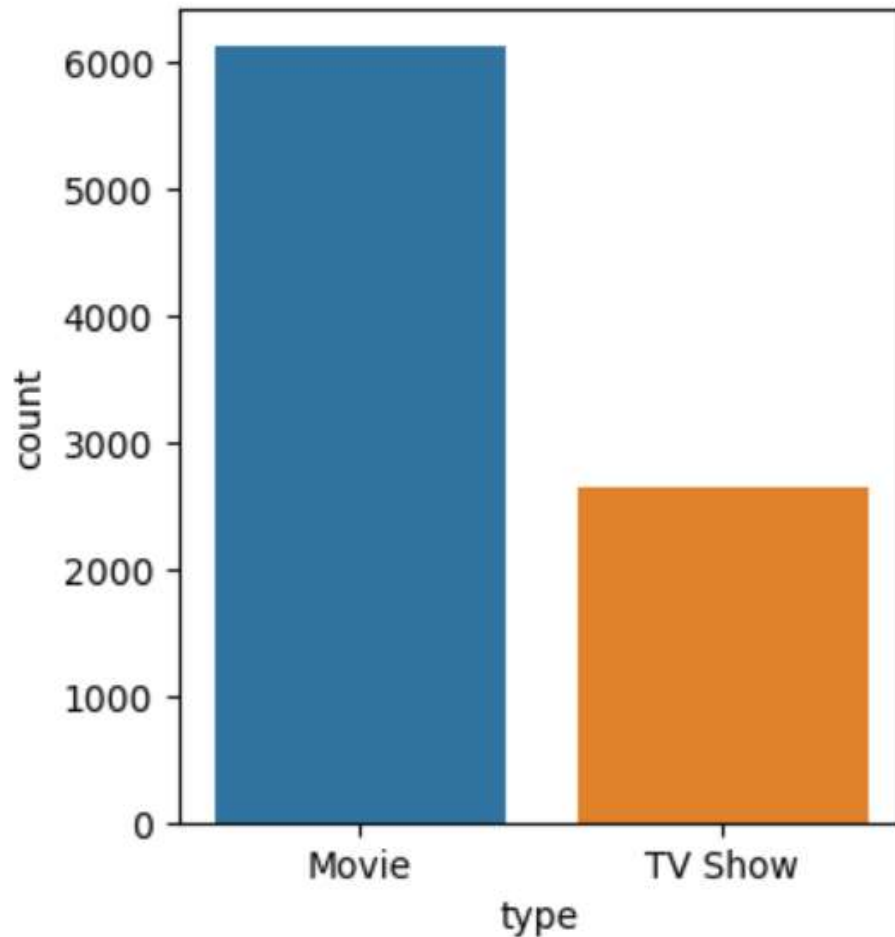
Top 10 Countries with most Content on Netflix

TOP 10 GENRES ON NETFLIX
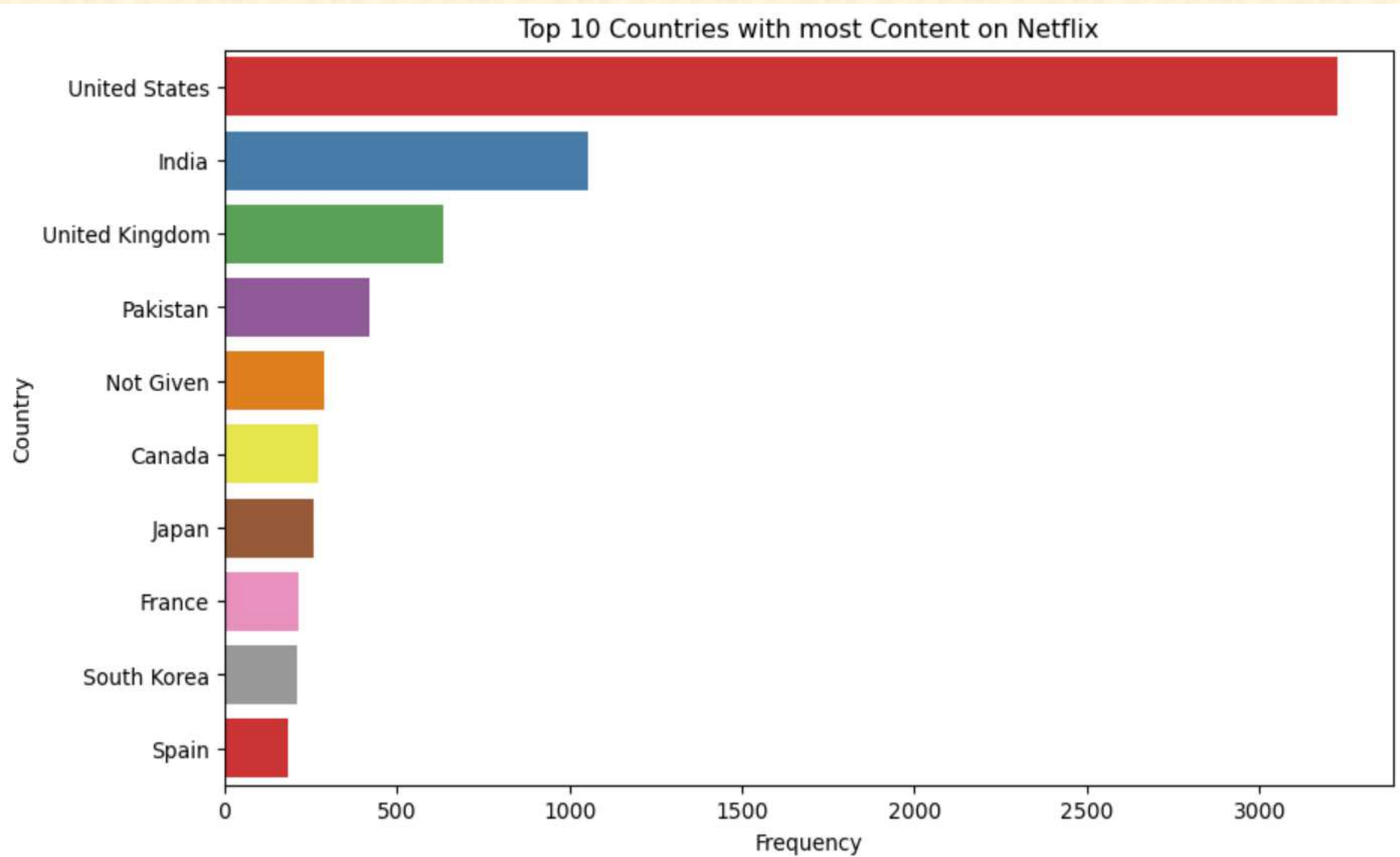
Top 10 Genres on Netflix

RATING ON NETFLIX

Rating on Netflix

- TV-MA 38%
- TV-14 26%
- TV-PG 10%
- R 9%
- PG-13 6%
- TV-Y7 4%
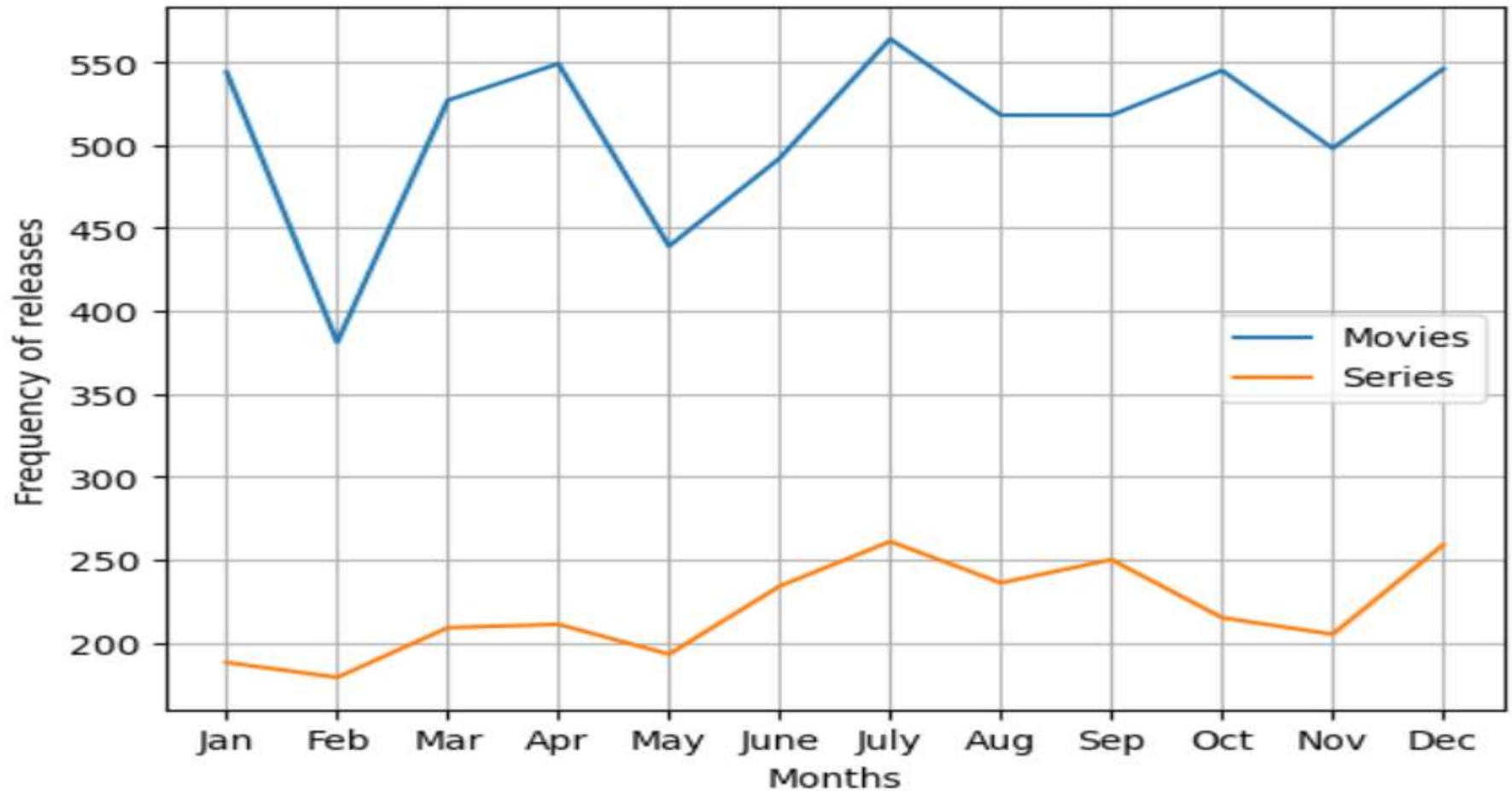- TV-Y 4%
- PG 3%

TOP 9 DIRECTORS WITH MOST TITLES
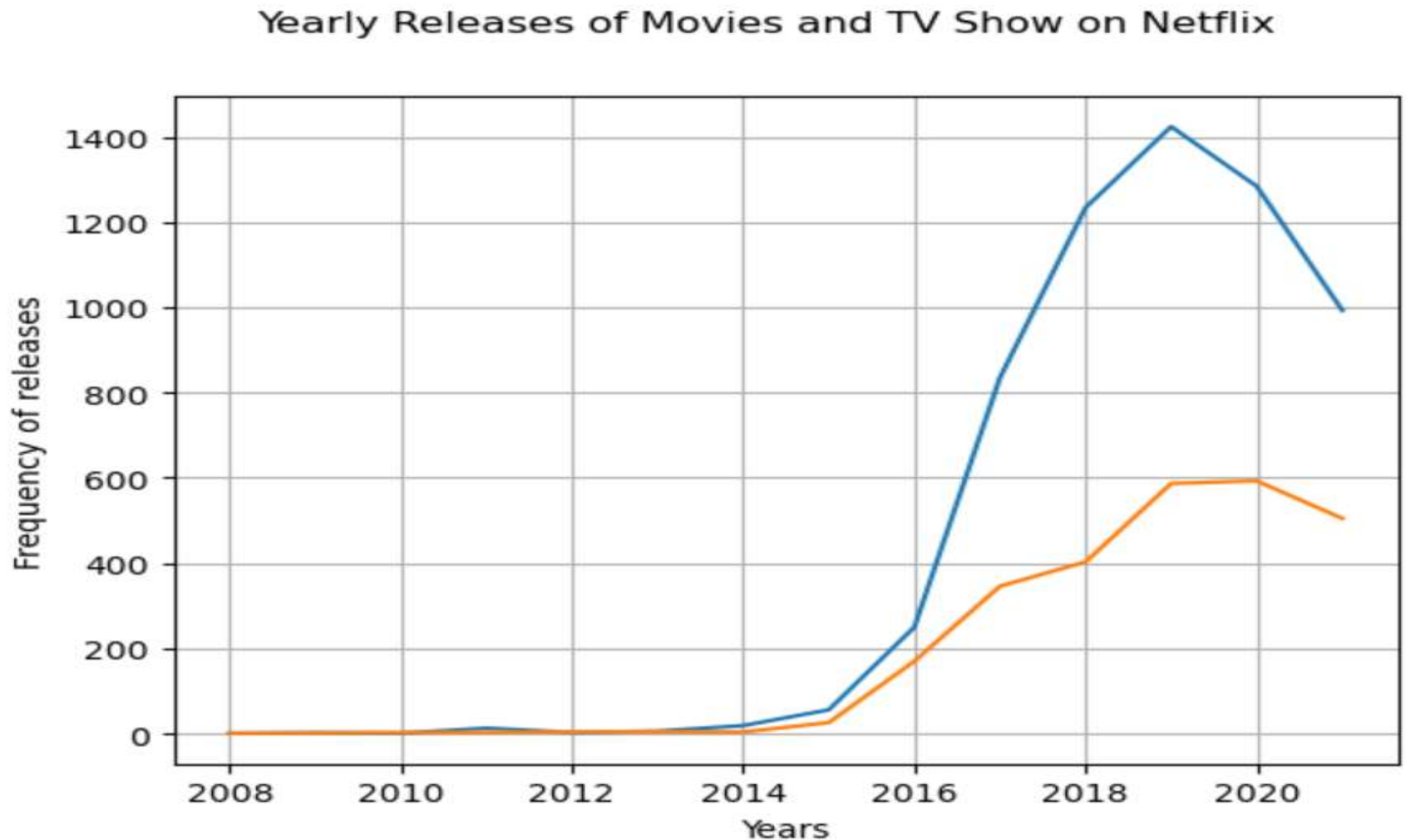

Top 9 Directors with the Most Titles

MONTHLY RELEASES OF MOVIES AND TV SHOWS ON NETFLIX
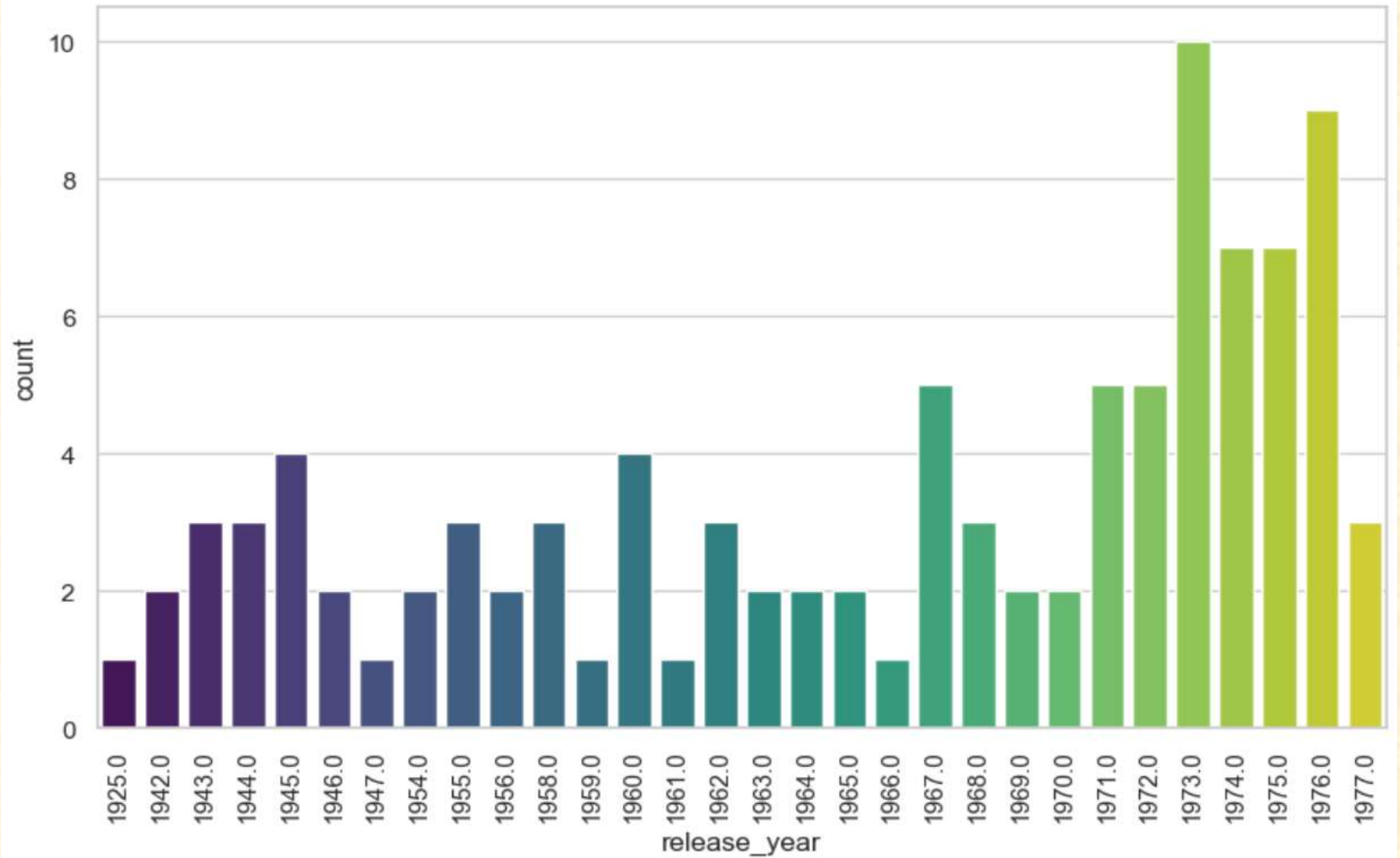

"Monthly releases of Movies and TV shows on Netflix
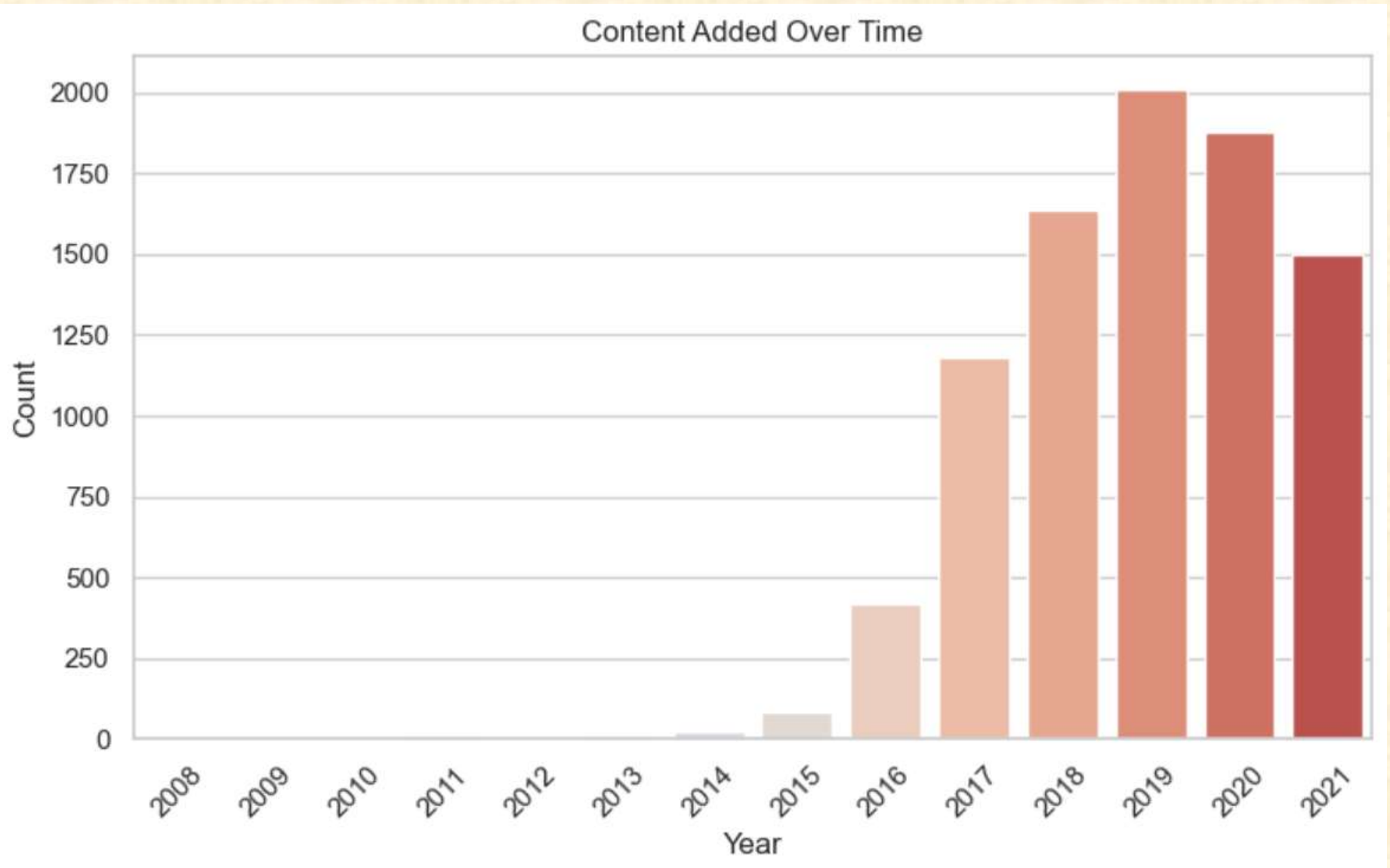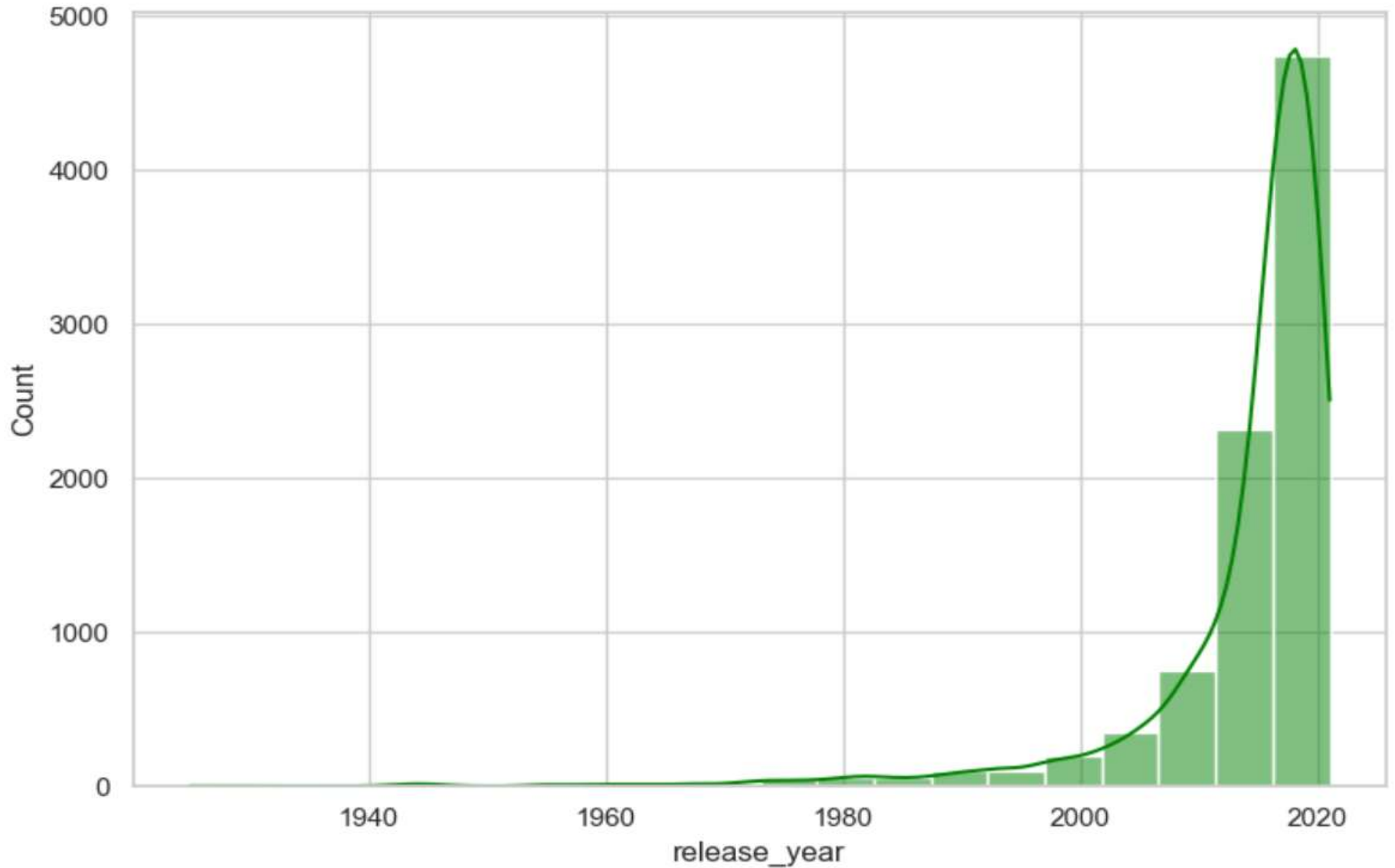
YEARLY RELEASES OF MOVIES AND TV SHOWS ON NETFLIX

Yearly Releases of Movies and TV Show on Netflix

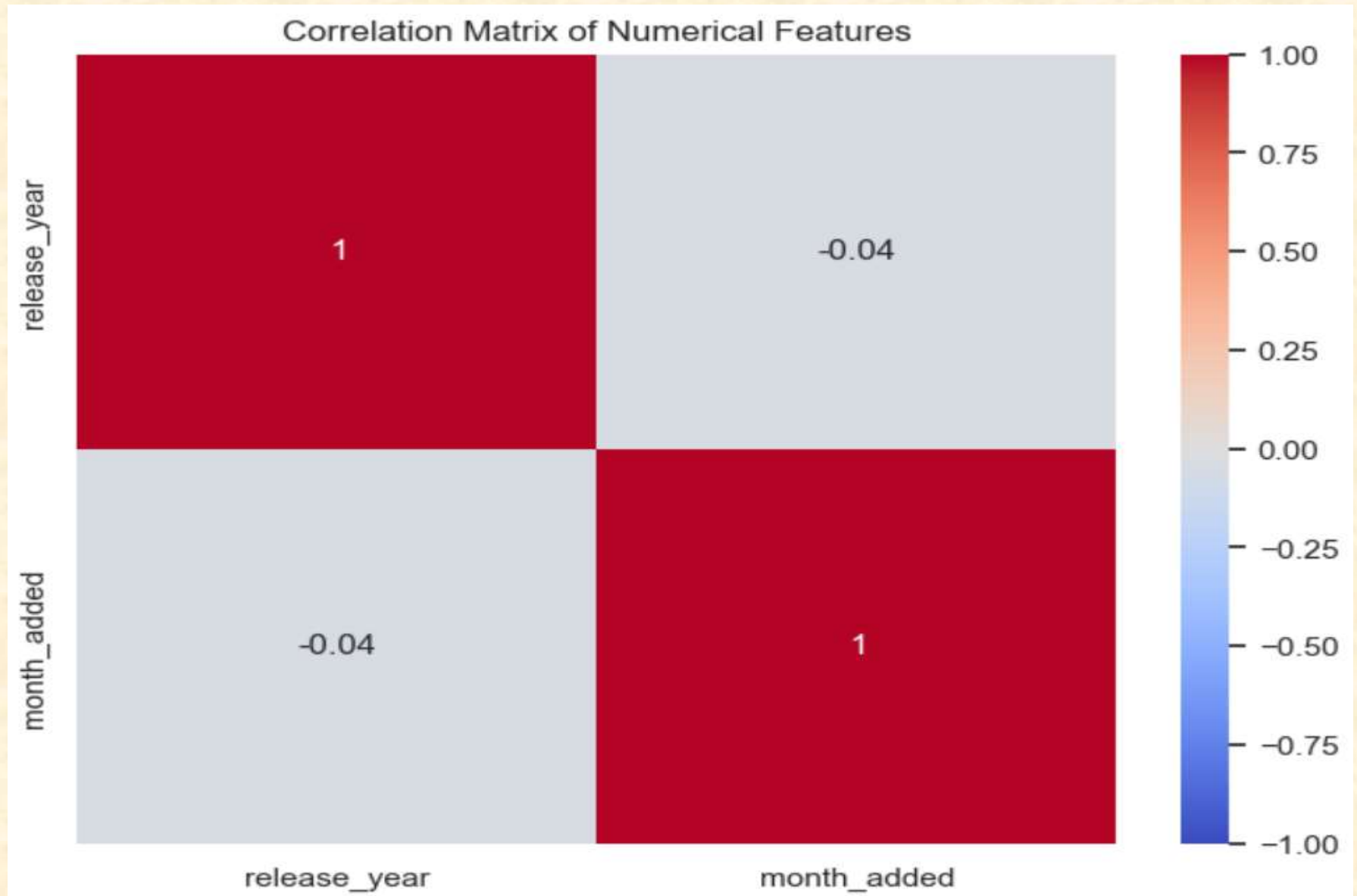# DISTRIBUTION OF YEARS IN THE FIRST 100 RELEASES

CONTENT ADDED OVER THE YEARS

Content Added Over Time

DISTRIBUTION OF ISSUANCE RELEASE YEAR

CORRELATION MATRIX
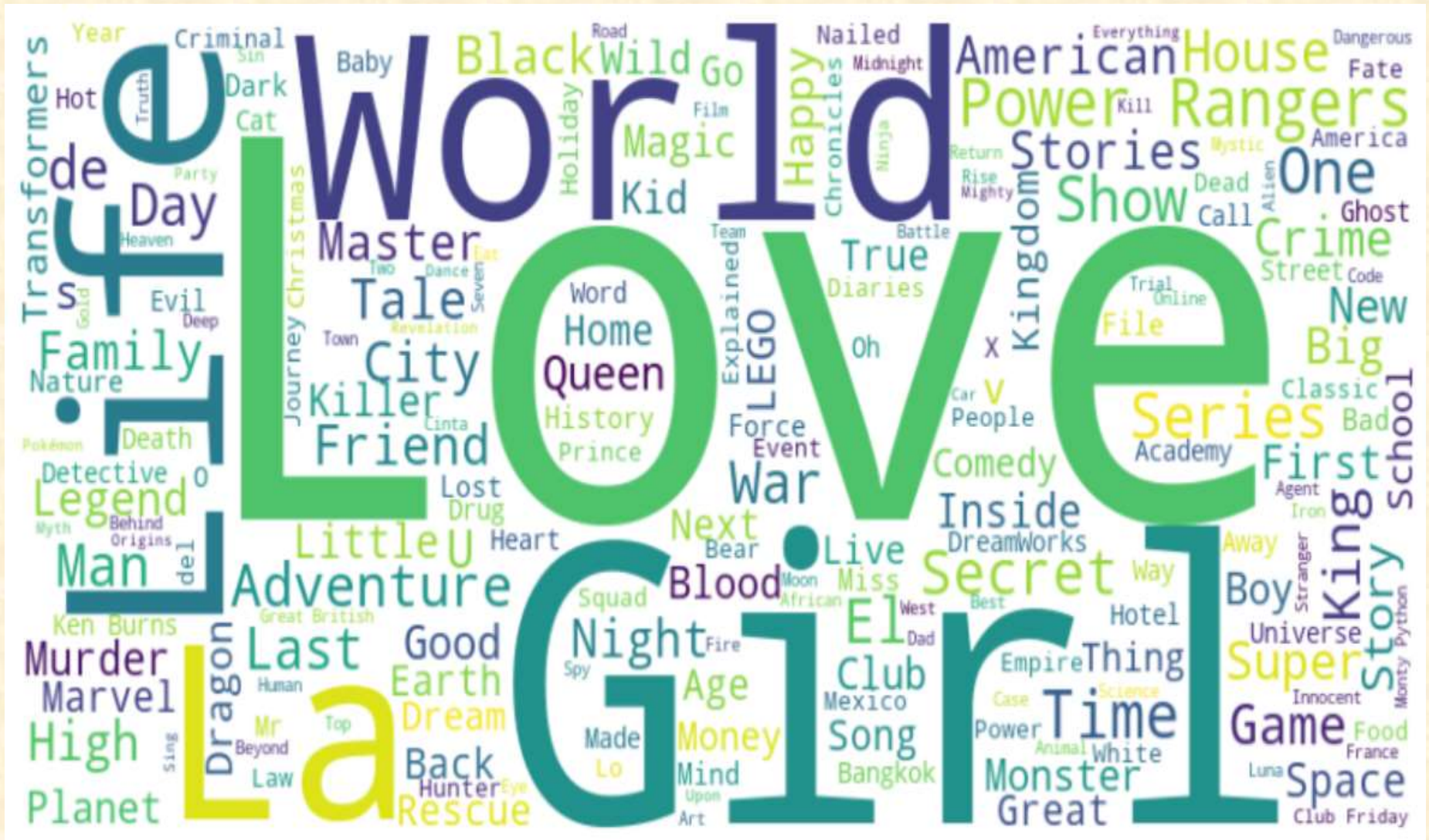

Correlation Matrix of Numerical Features

WORD CLOUD ON THE BASIS OF MOVIE TITLES

WORD CLOUD ON THE BASIS OF TV SHOW

CONCLUSION

- **Content type**: Netflix has more movies than TV shows, and most titles are dramas.

- **Content distribution**: The US is the biggest producer of Netflix content, followed by India and the UK.

- **Content popularity**: The most popular genre is International Movies, and the most popular rating is TV-MA.

- **Content trends**: Netflix's peak period was 2018, when it had over 1,000 titles.

- **Content retention**: TV shows with many seasons can help Netflix understand content retention strategies.

- **Regional tastes**: Regional tastes affect Netflix's content strategy.

- **Audience preferences**: Analyzing Netflix data can provide insights into audience preferences, which can help inform content strategies and marketing efforts.

THANK YOU!

For Reading