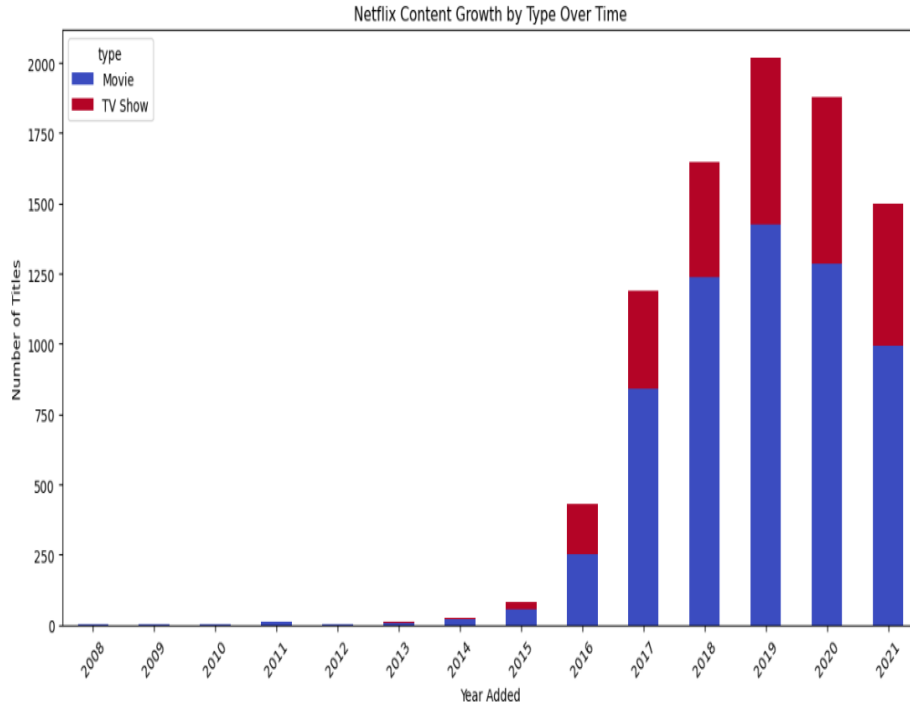


Data Visualization Pitch

Student Individual Assignment



Exploring Netflix Content Trends

Student name: Samira Doshmankosh

E-mail address: s.doshmankosh@campus.unimib.it

Githublink: https://github.com/SamiraDoshmankosh/DataVisualization_Netflix.git

Research questions

How has Netflix's content catalog evolved over time?

- How does the number of Movies vs TV Shows compare?
- How has Netflix's content grown over the years?
- Which countries produce the most Netflix content?

About Data

Exploring Netflix Content Trends. Data link [here](#)

- During the data search process, a key challenge was finding a dataset that was easy to understand while still allowing meaningful analysis and visualization. This dataset was chosen because it is well-structured, publicly available, and suitable for exploratory data analysis.
- The dataset used in this project is **"Netflix Movies and TV Shows"**, obtained from Kaggle in CSV format. It contains information about 8,807 titles, including content type, country, release year, rating, duration, and genre
- The dataset is shared on Kaggle under a Creative Commons Attribution (CC-BY) license, which allows the data to be used, analyzed, and visualized for educational purposes, provided that proper attribution to the original source is given
- Some data quality issues were identified, including missing values in columns such as *director*, *cast*, and *country*, as well as multi-valued categorical fields(e.g., genres and countries).
- These issues required basic data cleaning, such as handling missing values and formatting date variables, which were performed using Python.

Methodology

- **Data Collection:** The dataset used for this project was obtained from Kaggle, a public data-sharing platform. The data was manually downloaded in CSV format and contains information about movies and TV shows available on Netflix, including metadata such as content type, country of production, release year, rating, and content categories stored in the `listed_in` column. No web scraping or API access was required.
- **Data Processing & Cleaning:** Data processing and cleaning were performed using Python, primarily with the pandas library. The dataset was inspected to understand its structure, data types, and the presence of missing values. Missing values in categorical columns such as *director*, *cast*, and *country* were handled by replacing them with a placeholder value ("Unknown"). Minor formatting issues, such as inconsistent date formats in the `date_added` column, were corrected to ensure the data was suitable for analysis. While exploratory charts were generated using Matplotlib and Seaborn in Python, the final interactive visualizations were created using Datawrapper.

Methodology(continues)

- **Data Transformation:** Several data transformations were applied to support analysis and visualization. The *date_added* column was converted from a text format to a datetime format to enable time-based analysis. Aggregations were performed to count the number of titles by year, content type (Movies vs. TV Shows), country, content categories stored in the *listed_in* column. These transformations allowed the identification of trends and comparisons across different categories.
- **Use of AI Tools:** AI-powered tools were used to support the project. ChatGPT and Gemini was used to assist with code generation and data cleaning strategies.
- **Analytical Techniques:** The analysis relied on descriptive statistics and exploratory data analysis techniques. Comparative analysis was used to examine differences between Movies and TV Shows, while trend analysis was applied to study changes in Netflix content over time. Frequency counts were used to identify the most common countries and genres in the dataset.
- **Reproducibility:** The analysis is fully reproducible. The dataset is publicly available on Kaggle, and all data processing, transformation, and visualization steps were performed using Python code. By running the provided code on the same dataset, another user could replicate the analysis and obtain the same results.

Insights from the Data

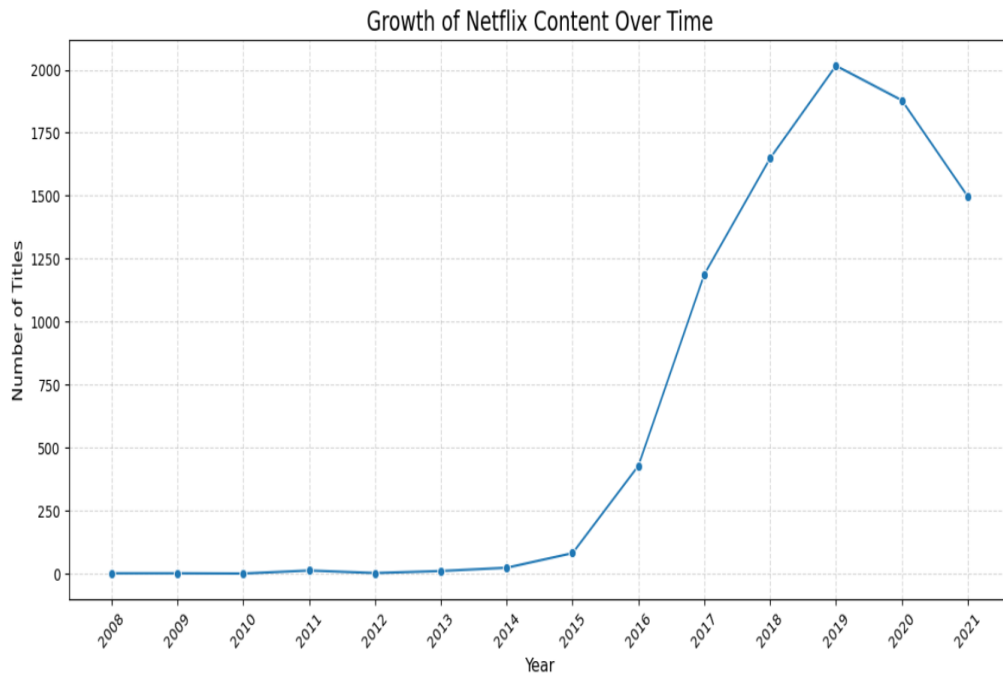
- The dataset contains 8,807 Netflix titles, with Movies representing the majority compared to TV Shows. Netflix's content catalog shows a strong growth trend over time, especially after 2015, indicating a rapid expansion of the platform. The United States is the largest producer of Netflix content, followed by a smaller number of dominant countries, while many countries contribute only a limited number of titles. appear most frequently, suggesting a focus on broadly appealing content. Some metadata fields, such as *director* and *country*, contain missing values, highlighting common limitations in real-world media datasets.
- Descriptive statistics were used to understand the size and structure of the dataset. Frequency counts and aggregation were applied to compare content types and countries. Trend analysis was used to examine how the number of Netflix titles changed over time.

Growth of Netflix Content Over Time

This visualization shows the number of movies and TV shows added to Netflix over time, based on the year each title was added to the platform. The chart highlights how Netflix's content library has evolved across different years.

The visualization reveals a strong upward trend in Netflix content, particularly after 2015. This period marks a rapid expansion of the platform, with a significant increase in the number of titles added each year.

The sharp growth in content suggests a strategic shift by Netflix toward expanding its catalog to attract a broader global audience. This trend also reflects the increasing competition in the streaming industry and the platform's investment in content production and acquisition. The data was aggregated by year using the `date_added` variable. A trend analysis was then performed to count the number of titles added per year, and the results were visualized using a line chart.

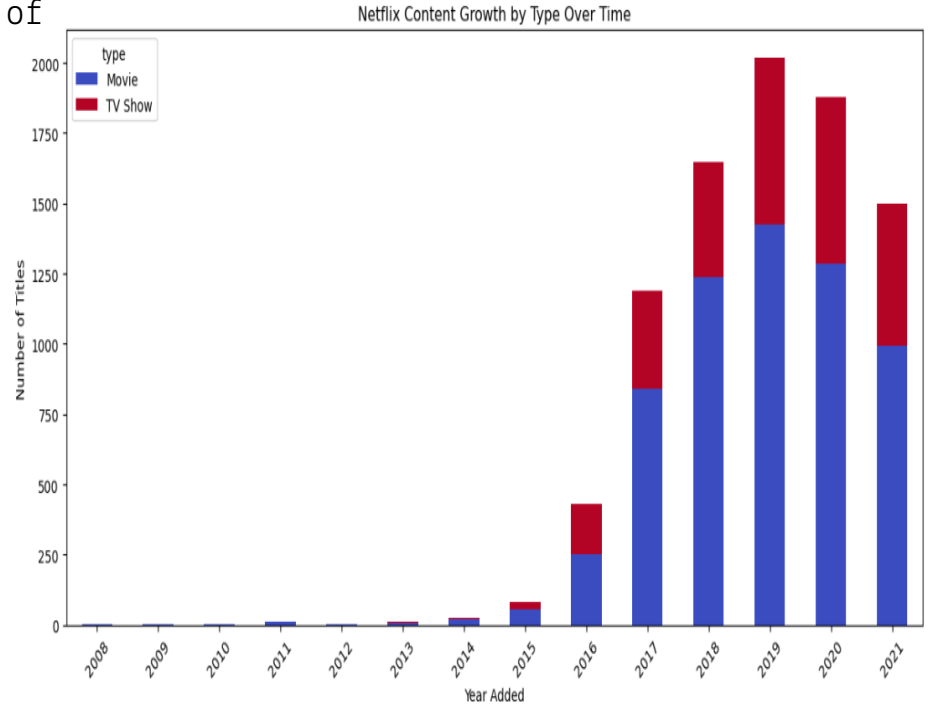


Made by python (pandas + matplotlib/seaborn)

Netflix Content Growth Over Time by Content Type

This visualization shows the number of Netflix titles added each year, separated by content type: Movies and TV Shows. The chart highlights how Netflix's content library has grown over time and how the contributions of different types of content have evolved.

Netflix's content catalog has grown steadily, with a sharp increase after 2015, reflecting the platform's rapid expansion. Movies consistently dominate the catalog, contributing the largest share of titles added each year. TV Shows show steady growth, indicating Netflix's ongoing investment in series alongside films. The overall trend demonstrates Netflix's strategy to expand both its volume and diversity of content to attract and retain a global audience.



Made by python (pandas + matplotlib)

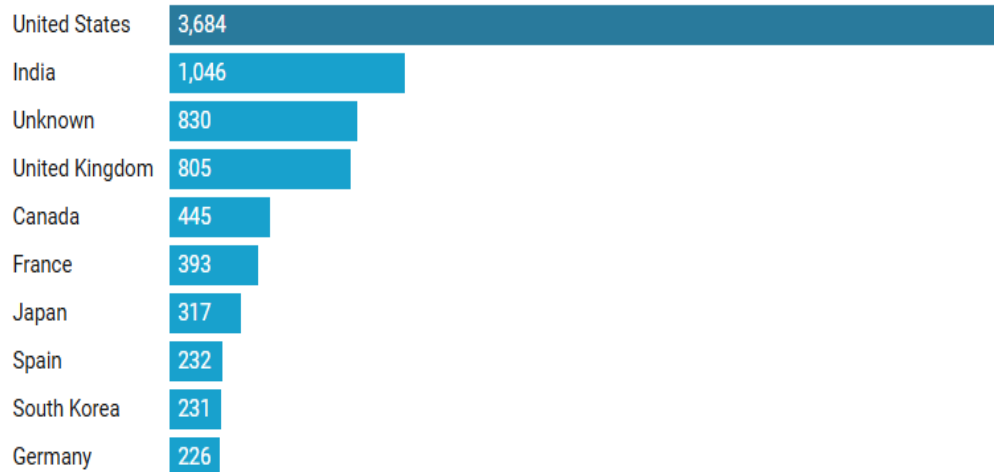
Top 10 Countries Producing Netflix Content

This chart displays the top 10 countries that produce the most content available on Netflix. It shows how Netflix's catalog is distributed globally in terms of production.

The United States is the largest contributor by a significant margin, followed by countries like India, the United Kingdom, and Canada. This highlights the dominance of certain countries in Netflix's content production, while many other countries contribute only a few titles.

Understanding the geographic distribution of Netflix content provides insight into the platform's global strategy and investment in different regions. It also reflects cultural diversity and market focus in Netflix's catalog.

Top 10 Countries Producing Netflix



Made with datawrapper
Interactive version available on GitHub

LICENCE

Slides are shared with the following license:

“Exploring Netflix Content Trends” by **Samira Doshmankosh** is licensed under Creative Commons Attribution (CC-BY).