

Course Project

Netflix Dataset Analysis

Tumma Karthikeyan

M.S Analytics

Visualization, Feedback, and Dissemination – AA-5200-12

Prof. Keith Hacke

Date: 14 Oct 2023

Executive summary

This report explains why Netflix's content strategy prioritizes movies over TV shows, focusing on movies made in the U.S., India, and the U.K. The platform mostly shows mature material and gives new movies, especially dramas and foreign films, a top priority. For continued global appeal, it's best to include content from regions that aren't well-represented and to regularly review your offers based on what viewers want.

DEFINITION [D]

Objective: This study's objective is to analyze the distribution and attributes of the content accessible on the Netflix platform, utilizing the Netflix dataset. The aim is to gain a deeper understanding of the many types of content, their origins, ratings, release years, and genres to derive meaningful insights.

The dataset appears to contain information about Netflix content. Here's a summary of the columns:

- **show_id:** An identifier for the show or movie.
- **type:** The type of content (e.g., Movie or TV Show).
- **title:** The title of the content.
- **director:** The director of the content.
- **country:** The country where the content was produced or is most relevant.
- **date_added:** The date the content was added to Netflix.
- **release_year:** The year the content was released.
- **rating:** The content rating (e.g., PG-13, TV-MA).
- **duration:** The duration of the content (e.g., in minutes for movies or the number of seasons for TV shows).
- **listed_in:** The genre(s) or category/categories the content falls under.

PREPARATION [P]

Dataset: <https://www.kaggle.com/datasets/ariyoomotade/netflix-data-cleaning-analysis-and-visualization>

Data cleaning:

This process of data cleaning involves identifying and rectifying errors, inconsistencies, and inaccuracies in a dataset to ensure its quality and reliability.

The missing values in the 'country' column were addressed by handling items marked as "Not Given". The dataset was examined for duplicate entries by comparing the values in the 'show_id' and 'title' columns to ascertain their uniqueness. Textual data like 'country' has been normalized and 'listed_in' to maintain consistency.

Data Transformation:

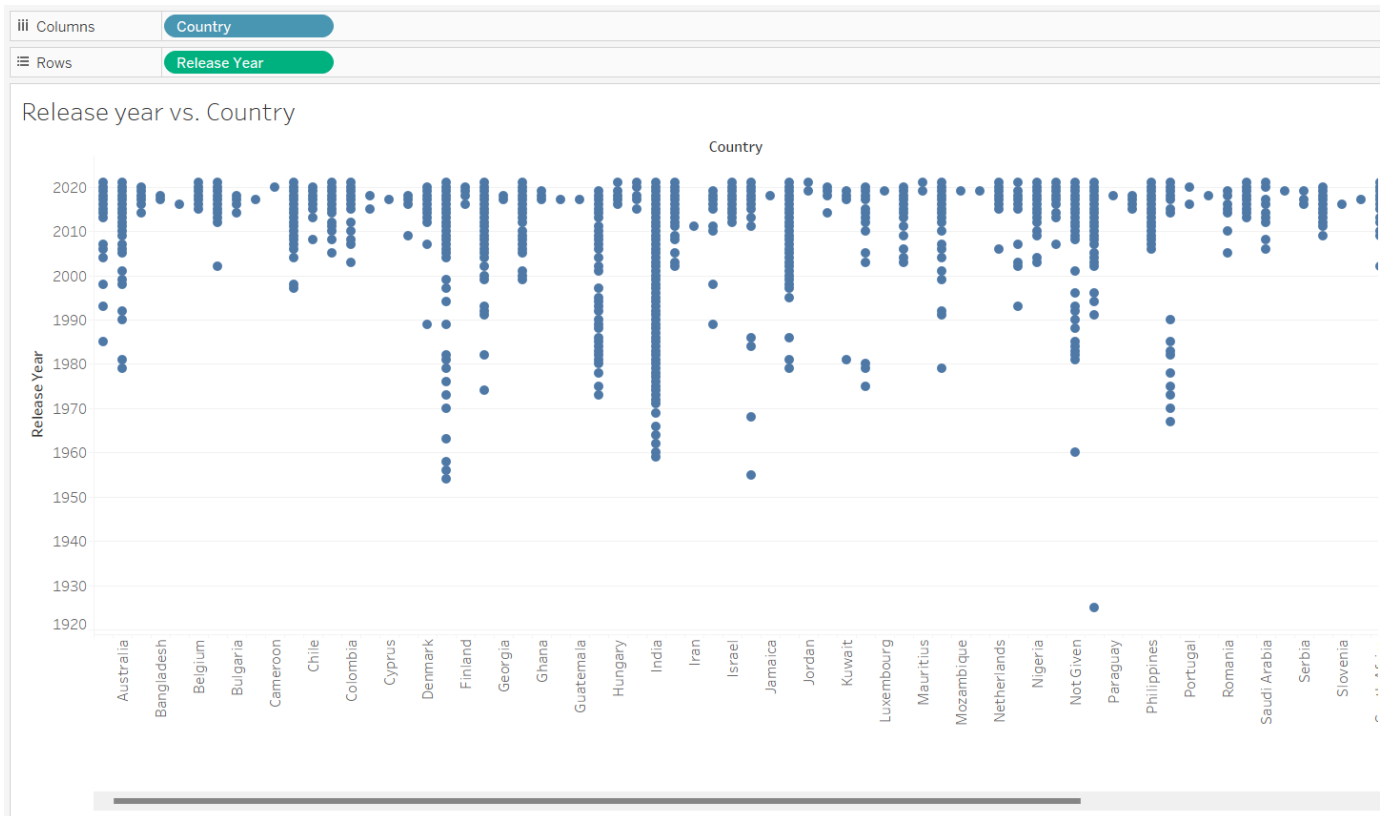
The 'listed_in' column was analyzed to find various genres or categories from the comma-separated values. The 'date_added' variable was transformed from a string format to a date and time format to facilitate time series analysis. Additional variables such as 'month_added' and 'year_added' were derived from the existing feature 'date_added' to provide a more detailed analysis.

ANALYSIS [A]

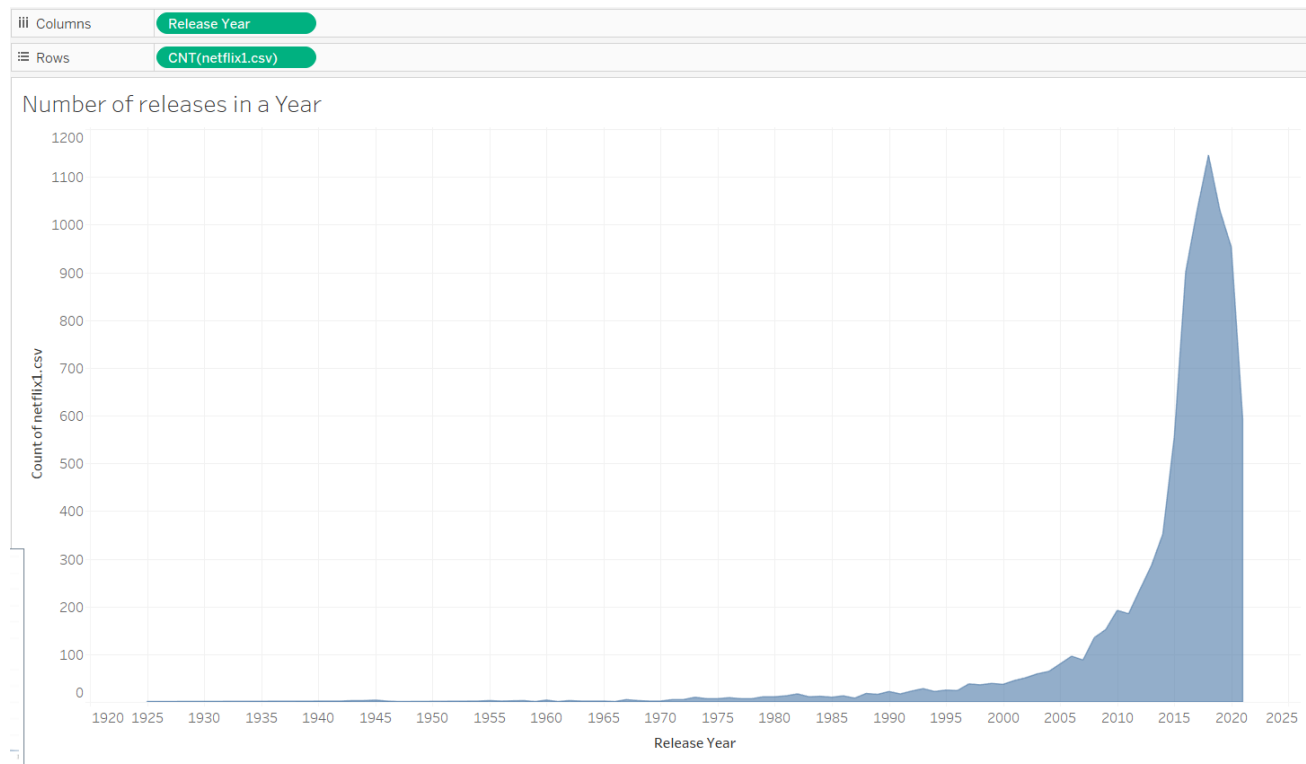
1. Release year vs. Country:

- This scatter plot shows the distribution of content releases by year and country. Each dot represents the release of a particular content in a specific year and country.

- A widespread distribution of releases for most countries around the 2010s and 2020s is observed. This suggests an increase in content production or availability in recent years.
- Some countries, like India and the United States, appear to have a more consistent distribution across the years, indicating a stable content production or release rate.

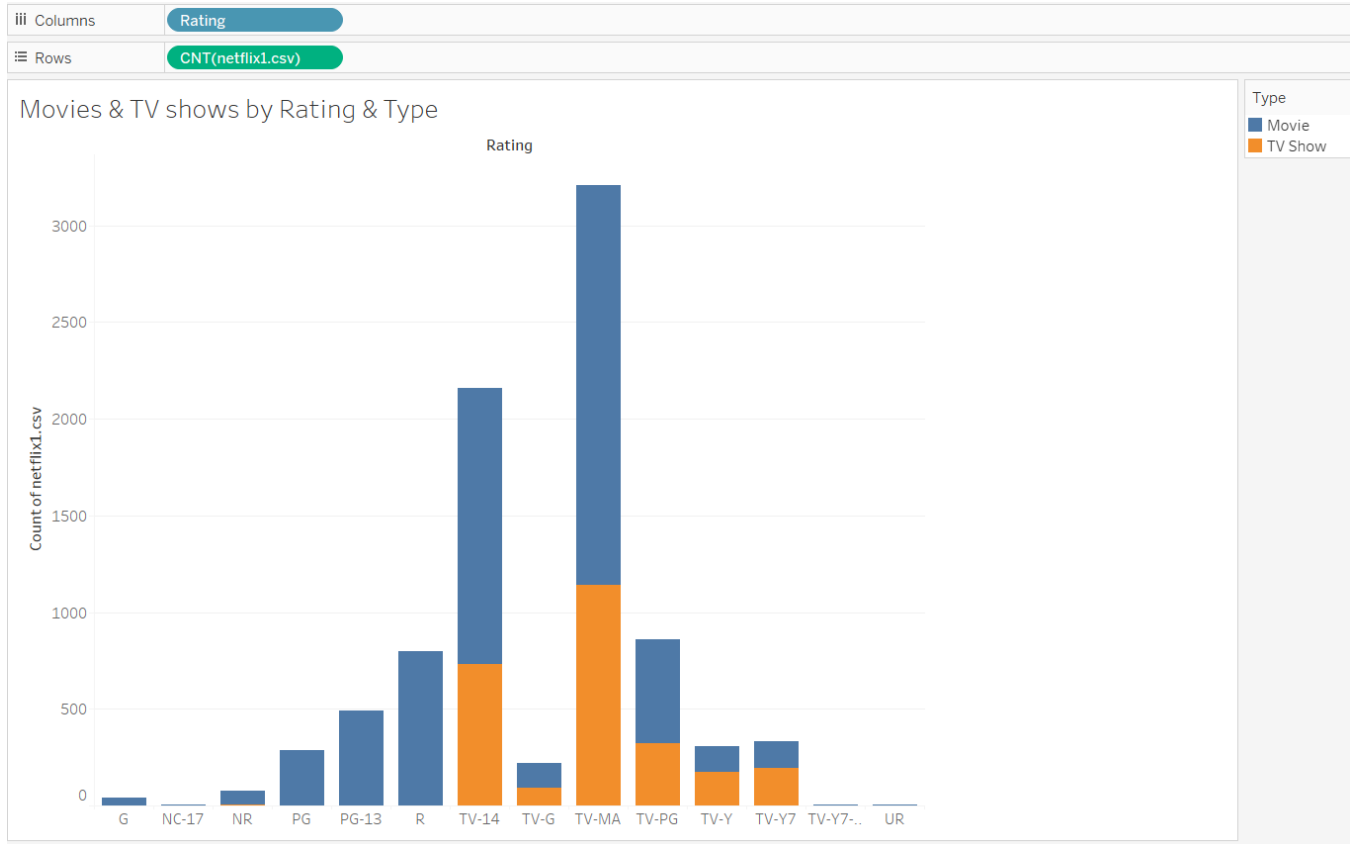


2. Number of releases in a Year:



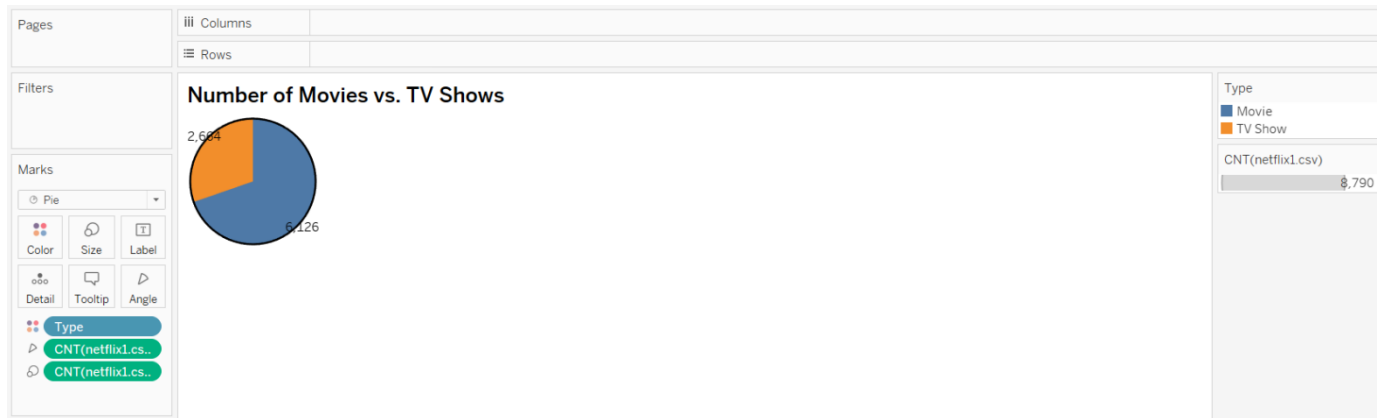
- This area chart illustrates the total number of content releases over the years.
- There's an increase in releases post the mid-2000s, with a sharp rise after 2010. This could be due to the rise of streaming platforms, increased global production of content, or better data collection methods.
- The chart peaks around the 2020s, indicating the highest number of releases.

3. Bar Graph (Movies & TV shows by Rating & Type):



- This graph showcases the distribution of Movies and TV Shows based on their ratings on Netflix.
- On the X-axis, rating categories such as G, NC-17, NR, PG, PG-13, R, TV-14, TV-G, TV-MA, TV-PG, TV-Y, TV-Y7, TV-Y7-FV, and UR are observed.
- The Y-axis represents the count of movies or TV shows within each rating category.
- Blue bars represent Movies, and orange bars represent TV Shows.
- The most prominent observation is that the category "TV-MA" has the highest count for both Movies and TV Shows. It dominates the other categories, with Movies in this category noticeably surpassing TV Shows.

4. Pie Chart (Number of Movies vs. TV Shows):



- The pie chart visually compares the number of Movies and TV Shows on Netflix.
- The larger blue segment represents Movies, and the smaller orange segment represents TV Shows.
- The total count, based on the filter to the right, suggests 8,790 titles. Movies make up the majority, with approximately 6,126 titles, while TV Shows account for around 2,664 titles.

DEPLOYMENT [D]

Strategic Suggestions: To reach more people, Netflix could consider adding more titles with family-friendly scores and broadcasting more kids' shows.

Because there is a lot of mature material, parental controls, and profile separation should be strong to keep kids from seeing inappropriate content.

Partnerships for Content: To get more titles in the G, TV-Y, and TV-Y7 ratings, work with companies or creators who make content that kids will enjoy.

Look into forming partnerships with studios in other countries to add more mature material from different cultures and languages to the "TV-MA" category and make it more diverse.

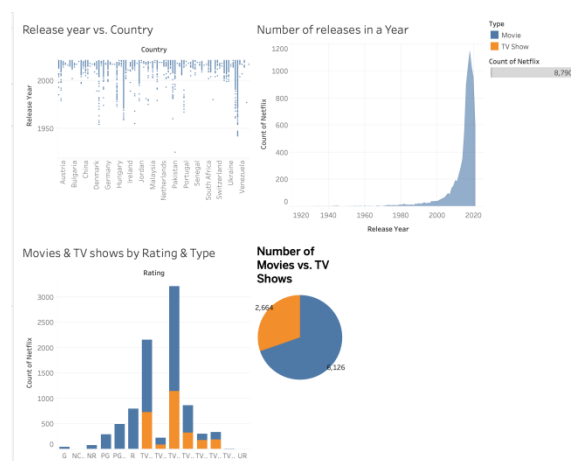
Improve user experience: Make content filtering based on ratings more precise. This will make it easier for users to find content that fits their tastes or watching environment.

Analyzing how users rate content can help you make better content suggestions. For example, if a user mostly watches "TV-14" material, the recommendation engine might give those titles more weight.

Feedback Loop: After the app is released, monitor user activity and comments to see how well the changes made based on the analysis worked. Review the research often to ensure your content strategy remains relevant as your audience's tastes change.

Marketing and promotion: Use the platform's wide range of scores in advertising to show how appealing it is to a wide range of people.

Run targeted marketing campaigns focusing on new titles that are good for kids or any new partnerships with mature material.



CONCLUSION

Distribution of Content Based on Ratings:

Netflix has a lot of content with a "TV-MA" rating, which means "Mature Audiences." Therefore, much of the content on Netflix is probably meant for adults and may contain serious themes that kids shouldn't see.

Some other grades, like PG-13, R, and TV-14, are also easy to spot. This makes it even clearer that Netflix is mainly for teens and adults.

The small number of titles with child-friendly ratings like G, TV-Y, and TV-Y7 shows that Netflix does have content for younger audiences, but it's not their focus.

Comparison of Movies and TV Shows:

There are more movies than TV shows on Netflix. This says that Netflix puts a lot of money into one-time watching content or movies, even though series or episodic content is popular and necessary.

This preference has many reasons, such as shorter production times, different tastes among audiences worldwide, or strategic partnerships with movie companies.

Reference

Ariyo, A. (2008). Netflix Data: Cleaning, Analysis and Visualization. Cleaning and Visualization with Pgsql and Tableau. Retrieved from <https://www.kaggle.com/datasets/shivamb/netflix-shows>. CC0: Public Domain.

Note: -

Grammarly and Grammarly Go have been used to edit and rewrite parts of this assignment. They have been used in the context of spell checking, proper use of grammar, and making sure the sentence is clear.