



**College of Professional Studies
Northeastern University San Jose**

MPS Analytics

**Course: ALY6070: Communication and Visualization for Data
Analytics**

Assignment:

Assignment 2 — Tableau Application

Submitted on:

March 14, 2023

Submitted to:

Prof: VENKATA DUVVURI

Submitted by:

NIKSHITA RANGANATHAN

ABSTRACT

Tableau is a powerful data visualization and business intelligence software that allows users to easily connect, visualize, and share data in a user-friendly interface.

Tableau offers a range of features, which include:

- **Interactive visualizations:** Tableau makes it easy to create interactive visualizations with just a few clicks. This allows users to quickly identify trends, outliers, and patterns in their data.
- **Data blending:** Tableau allows users to combine data from multiple sources, making it easy to analyze and visualize large datasets.
- **Real-time collaboration:** Tableau makes it easy for teams to collaborate in real-time on data analysis and visualization projects. This allows for more efficient workflows and better decision-making.
- **Mobile support:** Tableau is available on mobile devices, making it easy to access and analyze data from anywhere.

Tableau is used by organizations of all sizes and industries to gain insights from their data and for a variety of purposes, including:

- **Data analysis and visualization:** Tableau is primarily used to create powerful and interactive data visualizations that can be used to analyze and understand complex data sets.
- **Business intelligence:** Tableau allows businesses to access real-time data and create reports, dashboards, and visualizations to help them make informed decisions.
- **Data discovery:** Tableau's data blending and drag-and-drop functionality make it easy to explore data and discover insights that may not be apparent using traditional methods.
- **Data preparation:** Tableau also has features for cleaning and preparing data, allowing users to spend less time on data cleansing and more time on analysis and visualization.
- **Storytelling:** Tableau's storytelling feature allows users to create compelling narratives and present their data in a way that is engaging and easy to understand.

INTRODUCTION

Netflix has always understood the value of data analytics and has built its business model around it. Netflix started its video streaming services in 2007 but now it has transformed much more than what it offered to the customer. They are massive now that they use analytics to know their customers: What do their customer watches? What do they watch? From which geographic location do they watch? How long do they spend watching? When do they watch the content? How many times do they pause the content? Which content do they pause? Which content do they fast-forward and watch? Which content do they repeatedly watch? And many more to add.

The purpose of this project is to provide an exploratory analysis of the movies and tv shows that are available on the Netflix platform. The project focuses on the overview of the insight including the growth of Netflix content by year and percentage of the content compared to TV Shows and Movies which aims to provide movie fans to discover the Netflix contents which are presented in a variety of data visualizations.

About this Dataset: Netflix is one of the most popular media and video streaming platforms. This dataset contains information about movies and television series available for streaming on Netflix as of 2020. This dataset comprises records of all the content available on Netflix, along with details such as - genres, release dates, ratings, reviews, and other related information, etc. It also includes information about the cast and director of each title and the country of origin for each title.

The dataset is useful for analysis and research purposes, as it provides insights into the types of content that is available on Netflix and how it is distributed across different countries. It can be used to understand how Netflix's content library has changed over time, to identify popular directors and actors, and to compare the availability of titles across different countries. This dataset is useful for those who are doing research on the Netflix streaming platform, such as for a marketing or consumer behavior study. Furthermore, the data can be used to develop strategies for targeting particular audiences.

Research Questions :

1. How has the proportion of TV shows and movies available on Netflix changed over time?
2. Which countries produce the most content on Netflix?
3. Which genres are preferred on Netflix, and how do genre preferences vary by country?
4. How do the ratings of Netflix content vary for different countries?
5. What is the distribution of Movies and TV shows in countries with the most content?
6. What are the most common durations of movies and TV shows on Netflix, and how do they differ between genres?
7. Which directors are most popular on Netflix over the years and ratings?
8. Which words appear most frequently in titles?

ANALYSIS

● Why did you choose the types of visualizations that you did?

According to me, by carefully selecting the appropriate type of visualization for the data, I was able to effectively communicate the key insights and trends in the dataset in a clear and concise way.

Visualization is an important tool for making complex data more accessible and understandable, and the right type of visualization can be instrumental in helping viewers quickly grasp the message at hand. The right visualization can make complex data more understandable and engaging, while a poor visualization can lead to confusion and misinterpretation.

Choosing the appropriate visualization involves considering factors such as the type of data being presented, the message that needs to be conveyed, the audience, and the context in which the visualization will be presented.

For example: In this assignment, I have used a donut chart to compare the distribution of TV shows and movies in the dataset from the years 1925 to 2020. A donut chart is a simple yet effective way to visualize the distribution of different categories, as it allows viewers to quickly see the proportion of each category and compare them to each other. Adding a filter helps us know the change in the distribution of content over the years.

Donut Chart

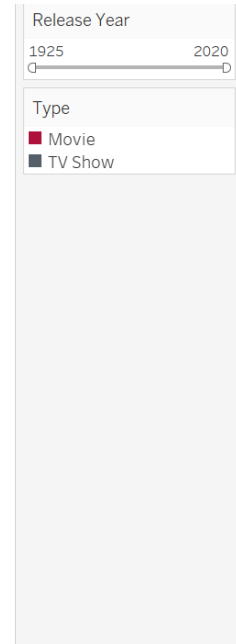
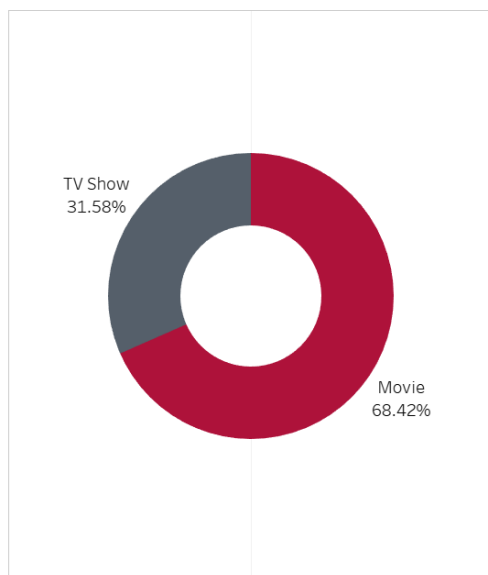


Figure 1 – Donut chart

I also wanted to understand the spread of the duration of all content (TV shows/movies) in minutes for different genres. So I used a histogram, which is a useful visualization that shows the distribution of a continuous variable. It works by dividing the data into a set of intervals (or "bins") and then counting the number of data points that fall within each bin. I have added

a filter for the genre so filtering one by one can help the audience understand the shape of the distribution.

By looking at the histogram, viewers can see the most common duration of content and how it is distributed across different ranges.

Histograms are commonly used in data analysis and statistics to identify patterns or outliers in the data and for communicating findings to a non-technical audience.

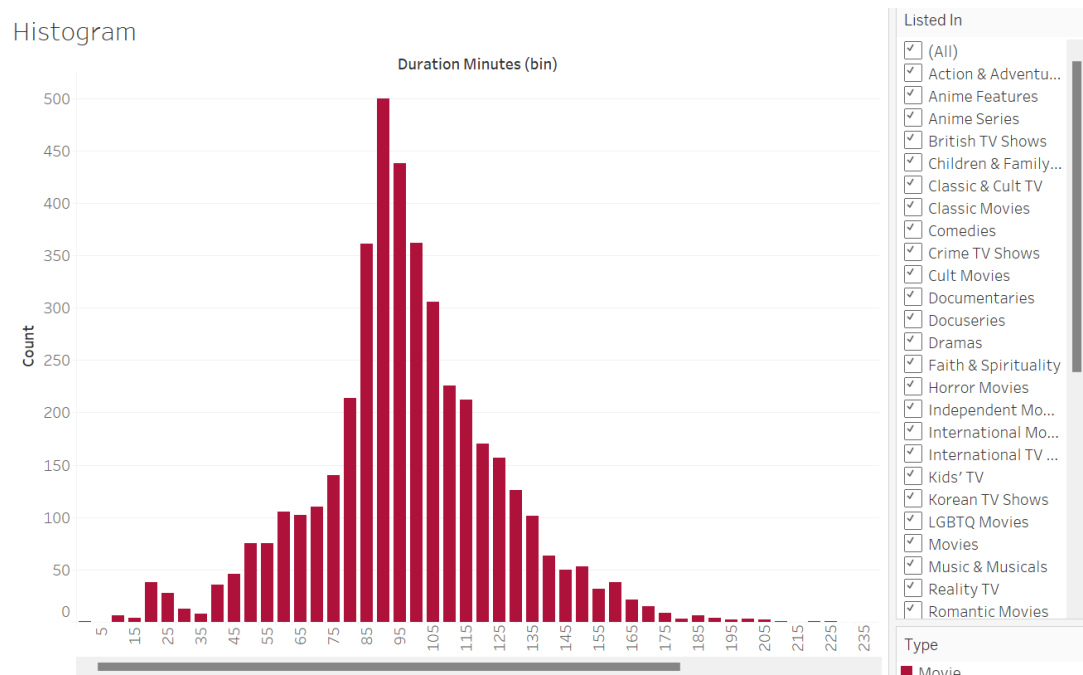


Figure 2 – Histogram

- **How are the visualizations effective and address the gestalt and design principles discussed in the course?**

Visualizations are an important tool to communicate data and findings to the general public. In order to make visualizations effective, it is important to take into account principles of visual design, such as gestalt principles and principles of design.

1. **Proximity:** The proximity principle is effectively used in the visualizations to group related information together. For example, in the stacked bar chart showing the distribution of movies and TV shows in the top 10 countries, the bars are grouped together by country, making it easy to see the comparison between countries.

Top 10 Countries-Movies vs TV shows

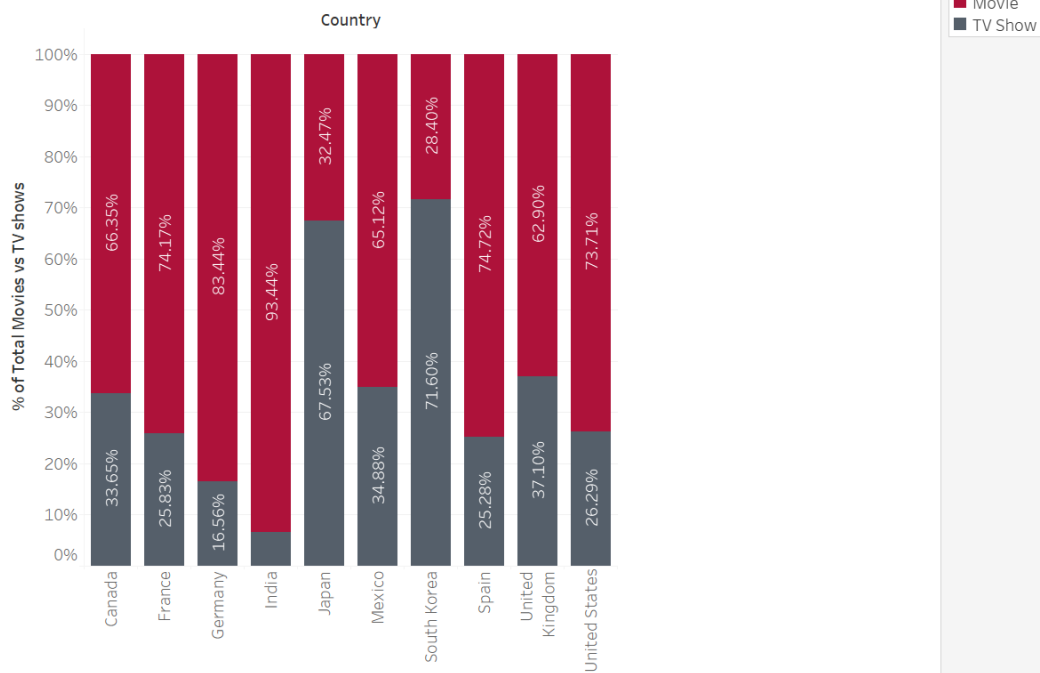


Figure 3 – 100% Stacked Bar chart

2. **Similarity:** The visualizations use similarity to create a clear visual hierarchy and to group related information together. For example, in the rating graph, the color and size of the circles are used to show the relative amount of content in each country.

Ratings

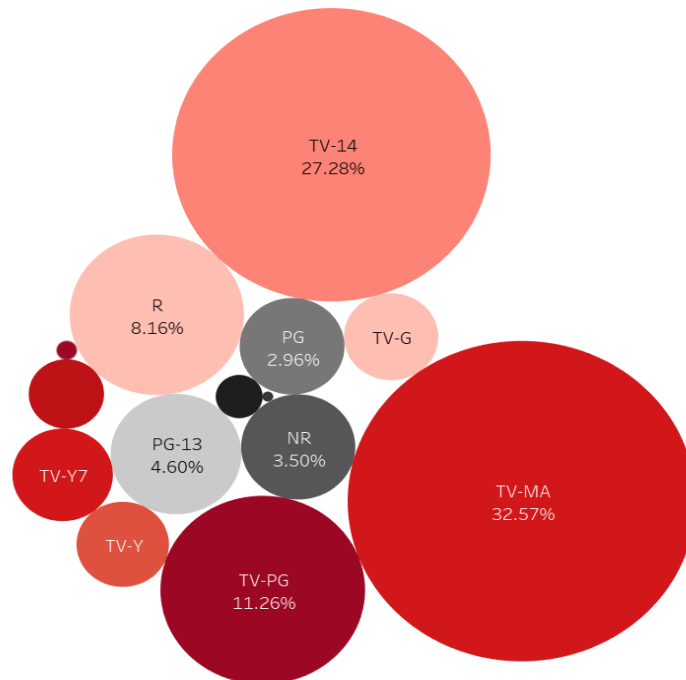


Figure 4 – Bubble chart

3. **Continuity:** The visualizations use continuity to create a sense of flow and to guide the viewer's eye through the data. For example, in the histogram showing the duration of content, the bars are arranged in a continuous sequence to show the distribution of content durations.
4. **Closure:** Closure is utilized in the visualizations to help the viewer make sense of the data and create a cohesive whole. For example, in the donut chart showing the proportion of movies and TV shows, the viewer can quickly see that the two categories add up to 100%.
5. **Enclosure:** The visualizations use enclosure to group related information together and to make it stand out from the surrounding elements. In the case of the treemap visualization showing the top 10 genres on Netflix, the enclosure principle is effectively used to group each genre into a rectangle shape based on its size.

Top 10 Genres

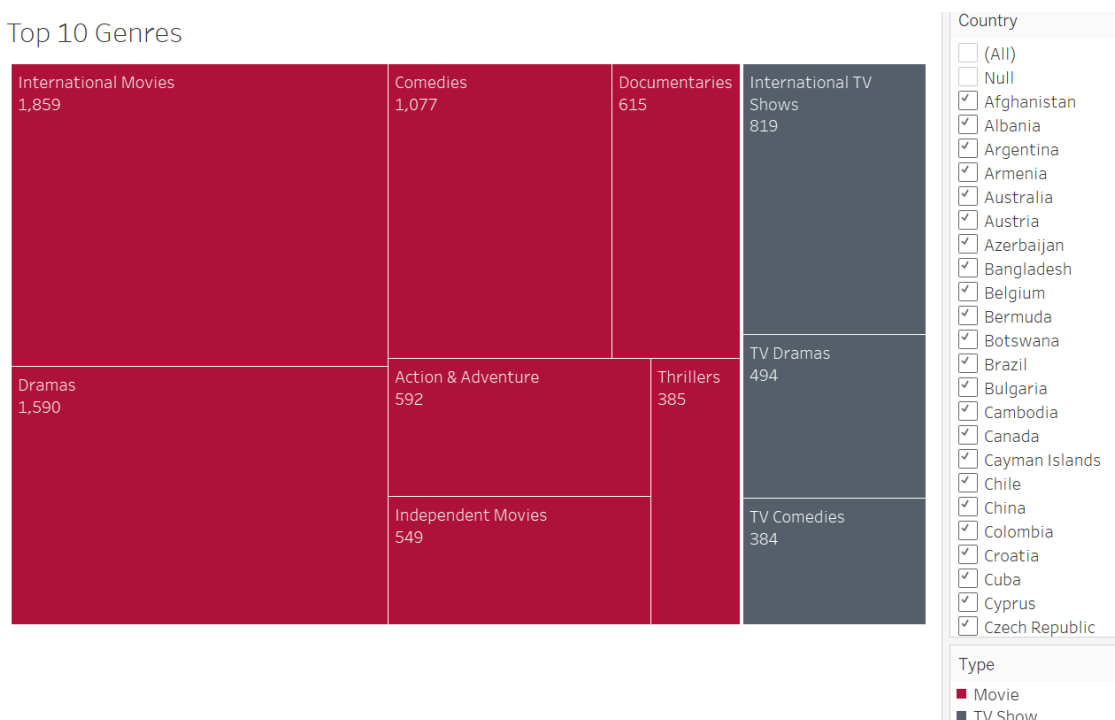


Figure 5 – Treemap

6. **Connection:** The visualizations use a connection to show the relationships between different data points.
- **How do the visualizations answer the research/business question?**
 1. **How has the proportion of TV shows and movies available on Netflix changed over time?**

The area chart shows the proportion of TV shows and movies added to Netflix. We can observe a rapid increase in content additions over the years, particularly from 2015 onwards. However, during the COVID-19 pandemic, there has been a slight decrease in the number of content additions.

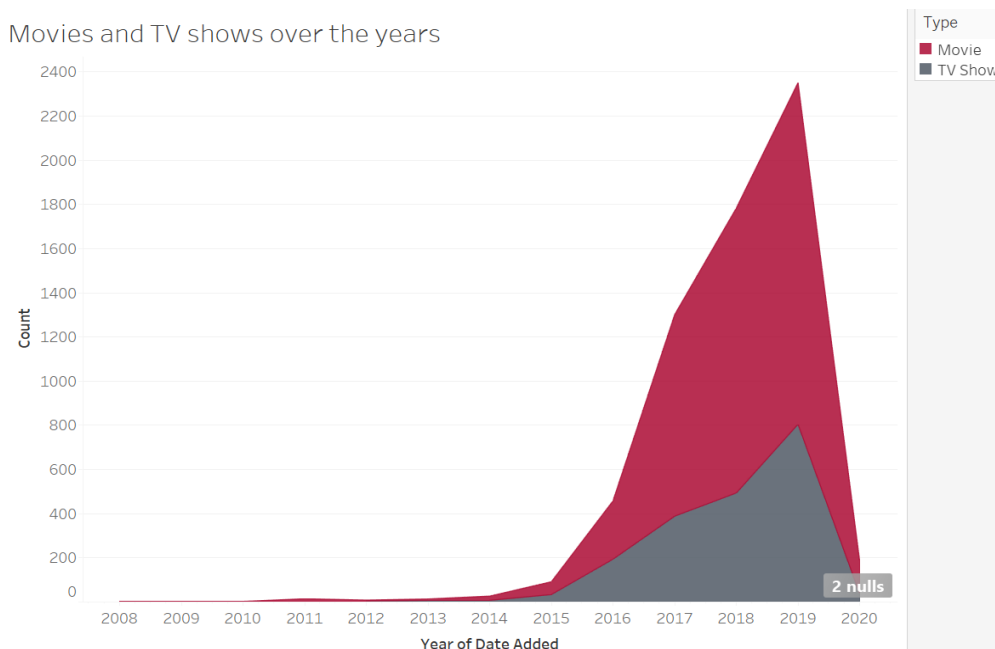


Figure 6 – Area Chart

2. Which countries produce the most content on Netflix?

The map allows us to observe how the distribution of content varies by region. The visualization shows that the United States is by far the largest producer of content, followed by India and the United Kingdom.

This information can be useful for understanding the global distribution of content on Netflix.



Figure 7 – Map

3. Which genres are preferred on Netflix, and how do genre preferences vary by country?

The treemap shows the popularity of different genres on Netflix, with the size of each rectangle representing the proportion of titles in that genre. The visualization also includes a breakdown of genre popularity by country.

The largest rectangle represents the most popular genre, which in this case is "International Movies." Smaller rectangles below represent other genres such as "Dramas," "Comedies," and "Thrillers,"

4. How do the ratings of Netflix content vary for different countries?

The most common movie ratings are TV-MA (32.57%), followed by TV-14 (27.28%), with a smaller proportion of movies rated TV-PG, PG, and G. Overall, the ratings distribution suggests that Netflix is catering more towards adult audiences than children.

5. What is the distribution of Movies and TV shows in countries with the most content?

100% stacked bar graph represents the percentage distribution of Movies vs TV shows for the top 10 countries with the most content available on Netflix. The countries include the United States, India, the United Kingdom, Canada, France, Japan, Spain, South Korea, Germany, and Australia. 100% stacked graph means the bars add up to 100% and each bar is divided into two sections, movies and TV shows.

From this graph, we can observe that Netflix has been expanding its content to cater to different interests and cultures around the world. In particular, it has focused on adding more recognizable and popular content from countries such as India, Japan, and South Korea. This includes Bollywood movies from India, anime from Japan, and K-dramas from South Korea. This is an effective way to reach out to more viewers and provide them with content that is relevant to their culture.

6. What are the most common durations of movies and TV shows on Netflix, and how do they differ between genres?

Overall, around 500 movies/Tv shows have a duration of 90 minutes according to the histogram. The distribution seems like a normal distribution.

7. Which directors are most popular on Netflix over the years and ratings?

Jan Suter, Raul Campos and Marcus Raboy are the most popular directors on Netflix. This indicates that Netflix has been investing in content from popular directors to attract a wider audience.

Top 10 directors

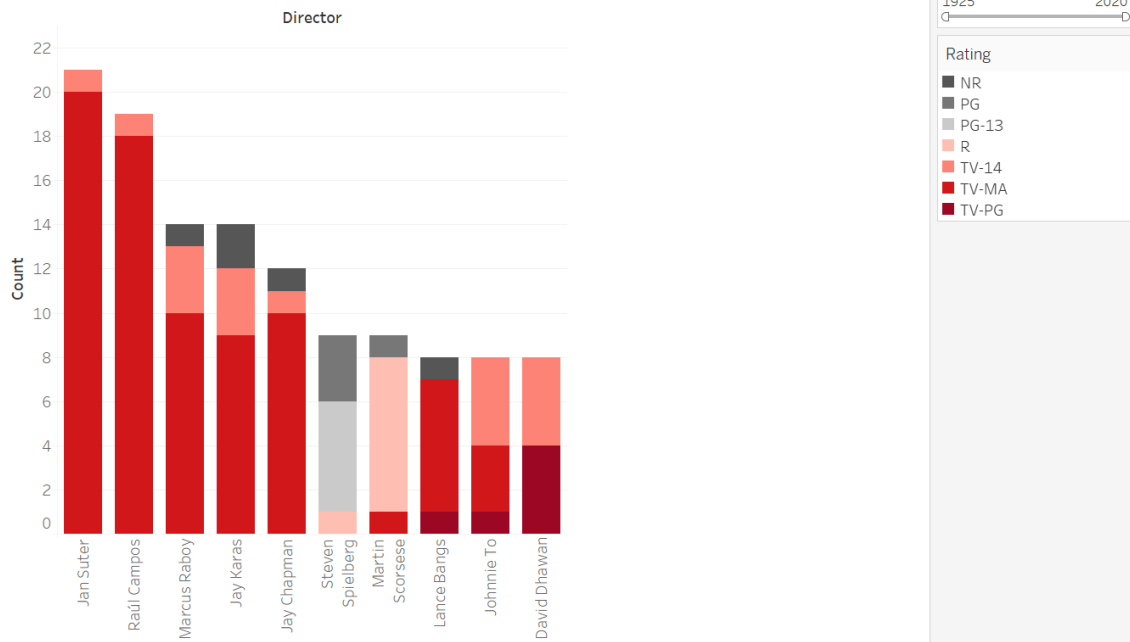


Figure 8 – Stacked Bar graph

8. Which words appear most frequently in titles?

A word cloud visualization could be used to show the most frequently occurring words in the titles of movies and TV shows. This could give an idea of the most common themes and topics of content available on Netflix.

Word cloud



Figure 9 – Word Cloud

• What story do the visualizations tell?

The visualizations in the Netflix dataset tell a story of a company that has adapted and innovated in order to become one of the most popular streaming services in the world.

Through investing in original content and catering to a global audience, Netflix has been able to grow and evolve over time, while also providing a diverse range of content that appeals to a wide range of viewers.

The visualizations have effectively conveyed this information through clear and concise graphics, utilizing design principles such as color, contrast, and layout to enhance the storytelling. The use of treemaps, bar charts, and heat maps has provided a comprehensive view of the data, allowing the viewer to easily interpret the trends and patterns.

The Netflix dataset and accompanying visualizations demonstrate the power of data in understanding business growth and consumer behavior. It is clear that Netflix's success is not just due to luck, but rather a combination of strategic decision-making and an understanding of what its users want. As Netflix continues to expand and evolve, it will be interesting to see how these trends continue to change and shape the streaming industry.

REFERENCES

P. (n.d.). *basic-dataset/netflix_titles.csv at master · prasertcbs/basic-dataset*. GitHub.

https://github.com/prasertcbs/basic-dataset/blob/master/netflix_titles.csv

Resources. (n.d.). Tableau. <https://www.tableau.com/resources>