# Content producing strategy for Netflix to increase user engagement

**Final Report**

Kalyan Dhar, Jugal Shah, and Prasad Yacham

D590 – Data Visualization Summer 2022

# Contents

# 1. Abstract

Netflix, an online movie subscription rental service, allows you to watch or rent movies for a fixed monthly fee. Netflix's digital library contains millions of web shows and movies that are released across multiple countries at a very rapid pace. This report analyzes the content on Netflix platforms and its trends and sentiments across the geolocations. Based on this analysis, we are trying to identify the reason behind Netflix's success as well as provide business insight to the digital streaming giant in order to improve customer experience. We have studied several existing visualizations that were generated using a similar dataset. We find several limitations with existing visualization. The article discusses the existing visualization and its technique. Also, in this article, we have provided more in-depth visualization, which can provide better insights into Netflix content. Netflix has recently started its own production house as well and this will help the company to decide what kind of content it should publish

# 2. Introduction

## 2.1. Motivation

Netflix has been the pioneer in digital content creation and providing entertainment at minimal subscription cost which has changed the trend in the entertainment industry. As of June 2022, Netflix has a total of 222 million subscribers (*Figure 2-2*) across the globe.

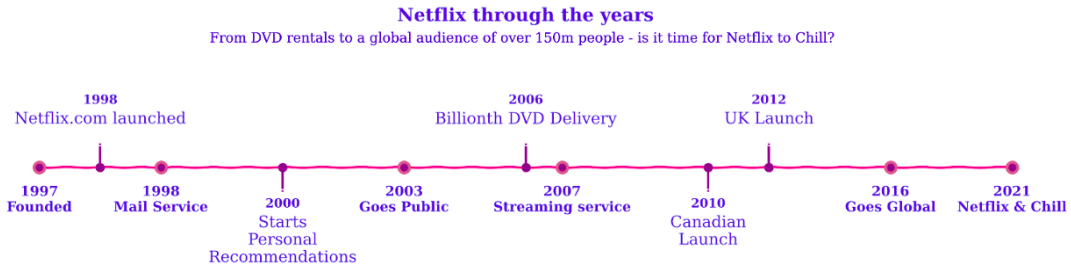



*Figure 2-1 : Journey of Netflix Inc (Reference 3)*

Netflix started its business as a mail-based DVD rental company and quickly identified opportunities in the online video streaming market. Within a few years, Netflix became a streaming giant on OTT platforms. In recent times, there were more streaming companies came into the

market and provided stiff competition to Netflix. Amazon Prime, next best to Netflix, has approx., 200 million subscribers as of June 2022

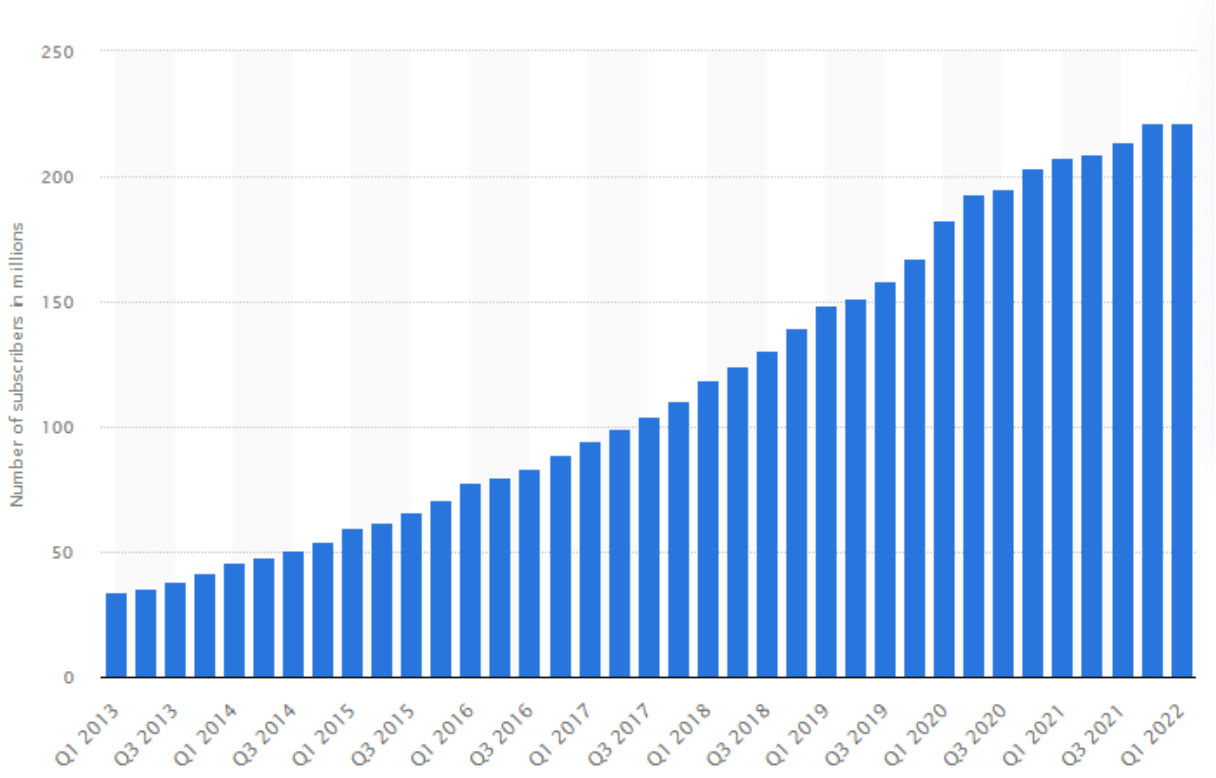



*Figure 2-2 : Netflix Subscription Growth over the time (Reference 7)*

Even with such a large customer base and constant growth, we have observed that recently Netflix stock price has been dropping dramatically (*Figure 2-3*) over the last couple of quarters. At the same time, the company also announced that they are losing subscribers at a faster pace than they had anticipated. It could very well possibly be due to an increase in competition in the streaming market or a lack of good content availability. With so many providers to choose from, nowadays customers are very choosy and hence a better content strategy is needed to address the current market need. Given the current trend, people constantly binge-watch movies or shows, and hence to keep user engagement high, Netflix must constantly produce new TV shows and movies. It must also go back to its own success and analyze what types of content kept users engaged in a better way. This paper is trying to address these questions and hopes that it will give Netflix clear and concise insights on the type of content that it should produce or focus on.

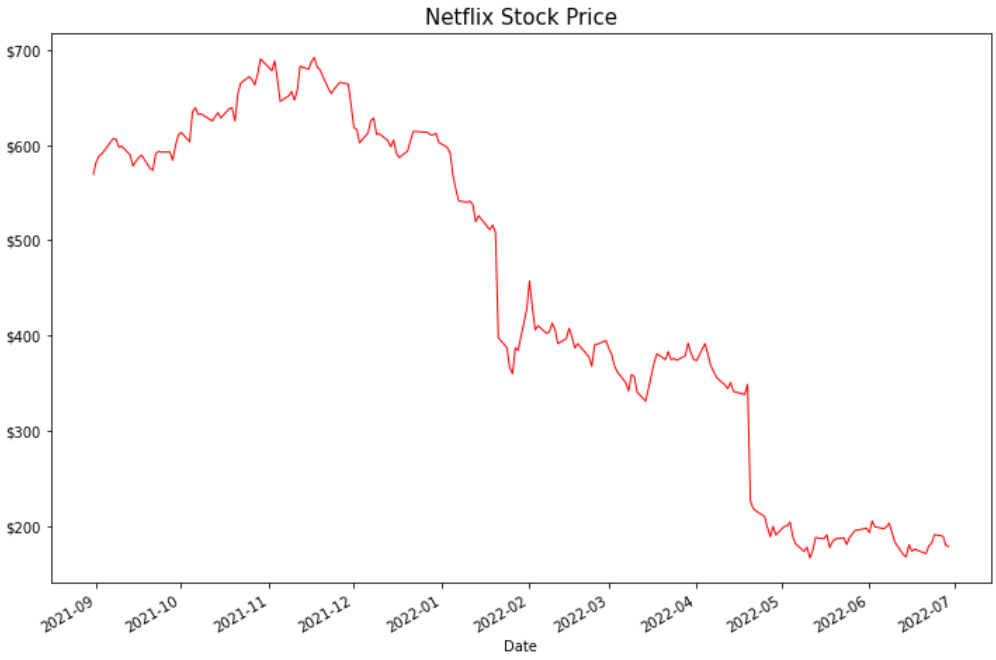



*Figure 2-3 : Netflix Stock Price*

## 2.2. Analysis and Critique on Existing work

On analyzing further and reviewing the existing work, we found several visualizations that tried to provide insights on Netflix's business but with serious misrepresentations or without providing complete pictures.

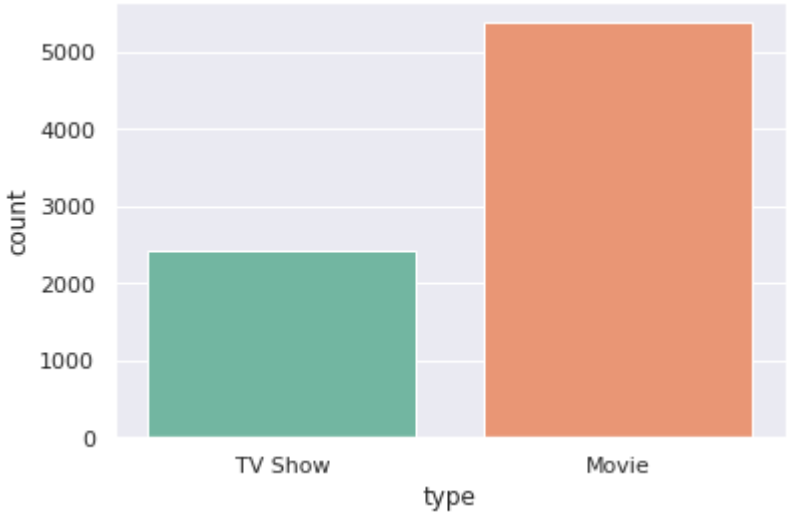



*Figure 2-4 : Distribution of movies and TV-Show on Netflix (Reference 2)*

The visualization shown in *Figure 2-4* represents the distribution of Movies and TV shows. However, it will be more interesting to see the percentage distribution of movies and TV shows rather than plain counts. Also, it unnecessarily colors both categories. As it is a simple bar chart, both charts could be the same color. It is very important to avoid unnecessary colors in visualization. Another example is shown in *Figure 2-5*. It is a time series plot that can be better represented by a line chart. It uses unnecessary colors which **reduces the data-ink ratio**. Also, a better way to show time series is a line chart which can help to identify trends effectively.

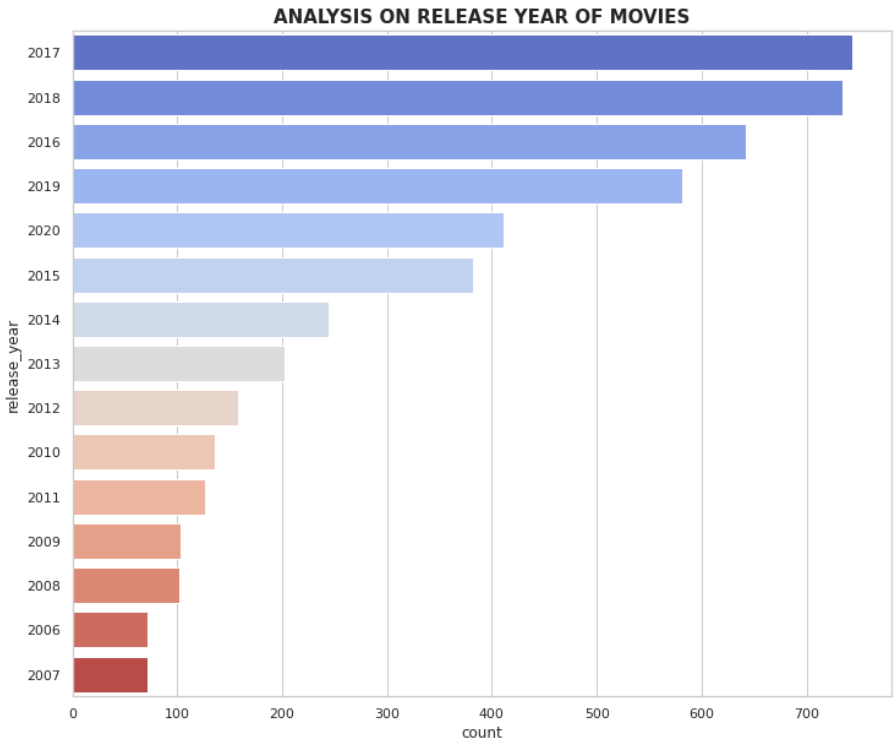



*Figure 2-5 : Number of Movies released each year (Reference 5)*

There are some visualizations that answer the potential untapped market that Netflix should target. For example, current Netflix's digital catalog library has very less content for Kids. The visualization in *Figure 2-6* shows it beautifully. It has separate backgrounds for each age group and compared different ratings among each age group. It is visually appealing and brings out the underlying point that the kid's section seriously lacks content availability. This is where Netflix can have opportunities to gain more engaged users.

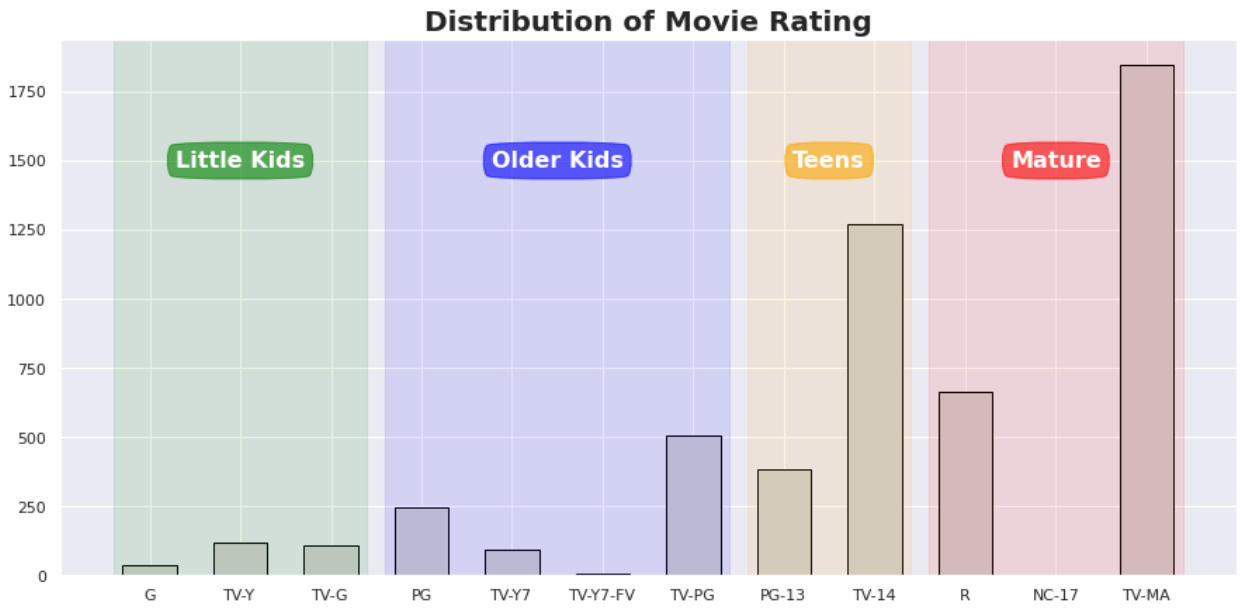



*Figure 2-6 : Movies based on Age Group (Reference 7)*

## **2.3.** Contribution to existing work

After understanding the data behind Netflix content, there are several elements along with proper data visualization techniques that can be used to provide insight into increasing the engagement of the user. Geographic data can be used to answer region-specific questions and it is possible to provide recommendations to produce content based on the culture of the region. For example, content that may be successful in Mexico may not be successful in India. Also, none of the existing visualizations has found a correlation between Netflix subscriber counts and country of content origin.

Apart from that, there are two major kinds of content, Movies and TV Shows, available on Netflix. Existing visualization does not provide a detailed analysis of market trends between movies and TV shows. If correctly analyzed, this could provide insight to Netflix as to which type of content, they should be investing in.

Correlation between different genres is also another important factor. For example, it is very likely that a user who is a founder of dramas, also be interested in independent movies or documentaries. This will further give Netflix insight into what kind of shows they need to be produced.

Further, there is a very popular dataset that contains IMDB ratings of movies and TV shows. None of the visualizations has currently connected those two datasets and tried to find more insight. It

would be interesting to see the average rating of Netflix content and see if they are always providing high-rating content to their subscribers.

# 3. Dataset and Visualization Methodology

## 3.1. About the dataset

The dataset analyzed in this paper is obtained from Kaggle (*Reference 6*)The dataset contains information regarding current TV shows and movies on Netflix's digital catalog. It also provides data related to each TV show and movie in detail with a timestamp of content addition, genres of each content, and origin country. There are a total of more than 8000 movies or TV shows streaming on Netflix as of mid-2021. Each content has details such as

- Title
- Type of content – Movies or TV Shows
- Cast and Director
- Country of origin
- Content release and added date
- Genres
- Rating
- Genres and duration

## 3.2. Ideas behind the visualization

The dataset contains several numerical and categorical columns. Genres, Ratings (PG13, R, etc.), and country of origin are a few of the categorical columns. While content release and added date are date-time functions. These columns can be utilized to provide time series visualization. Time series visualization will help us envision the trend of the current content addition strategy. In addition, we can also recommend the right time to add digital content for each location.

Using Choropleth map visualization based on country of origin, we can identify which country is the top digital content producer. Also, we can further produce country-specific analyses based on the type of content and its ratings. This will give us an idea about variations across different countries.

*Figure 3-1* shows visualization sketches that we are planning to cover in this paper.

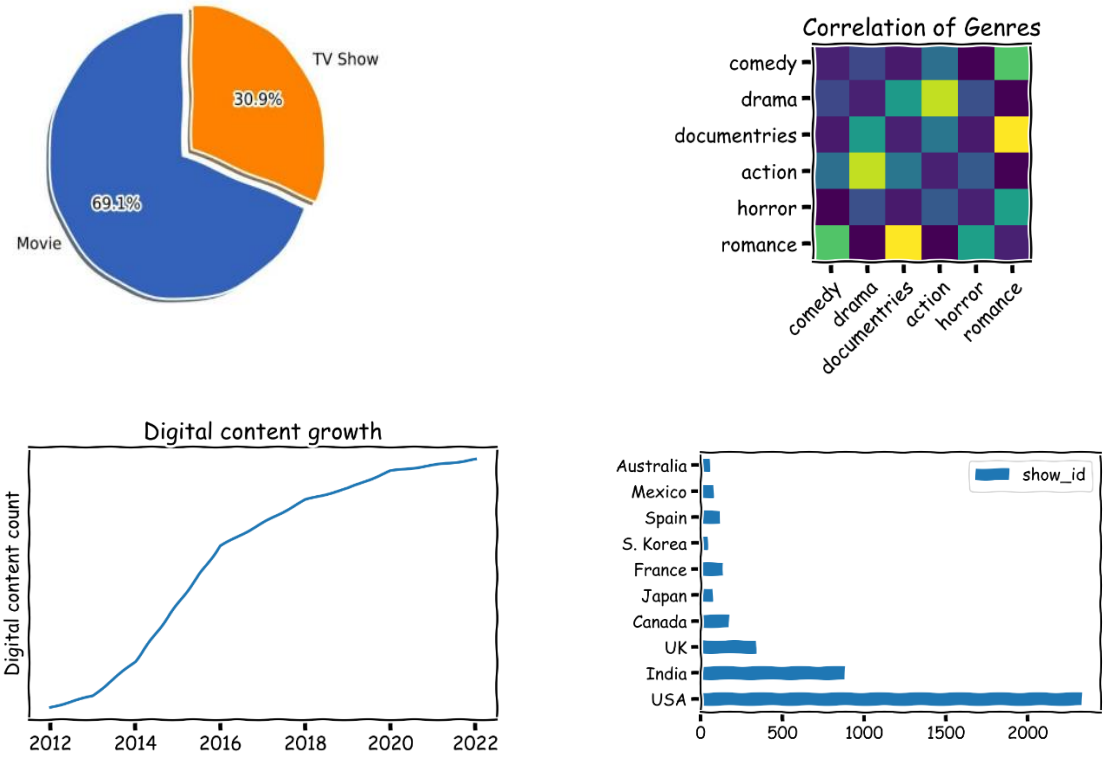
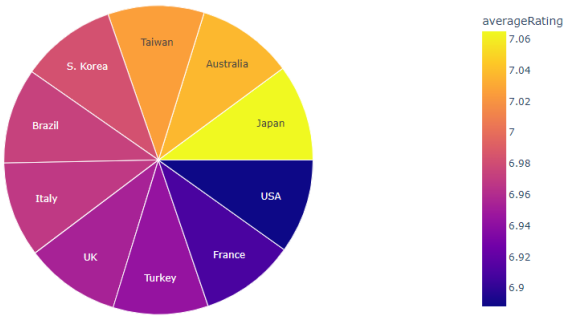



*Figure 3-1 : Example of visualizations*

## 3.3. Visualization Process and Methodology

The original dataset contains several categorical columns. However, we could not find features related to IMDB rating. IMDB rating shows content popularity and quality index. It is a very popular gauge viewer that is often looked at by customers before deciding to watch the content. None of the earlier existing visualizations we have searched contains a combination of Netflix Content and IMDB rating. Based on the content title, both the tables can be joined, and we can see the top content-producing countries with their average ratings.

Chart, pie chart Description automatically generated




It would be interesting to see the average movie rating for each country and how that rate on the IMDB scale. We have filtered countries that have at least 100+ digital content available on Netflix. Before the filter was applied, we were also getting countries like Iran which do not have enough content for analysis. Hence, we intentionally decided to filter countries of origin with 100+ titles. It is seen that the average rating for Japan is slightly above 7. The USA, producing the largest number of digital content ranks at the bottom 10. Netflix has potential improvement opportunities in this area, and they should specifically focus on how to make high-rated movies and TV shows available on their platforms which can increase user engagement levels.

Apart from the above, several visualization iterations were performed for map visualization (*Figure 3-2*). We were interested in seeing how different regions contribute to the digital content library of Netflix. Initially, we used map visualization and ran into several difficulties with map visualization as discussed in the example.

For example, there are several contents that were jointly produced by production houses across different countries. So, we have created an additional feature called “first country” which contains the name of the first country.

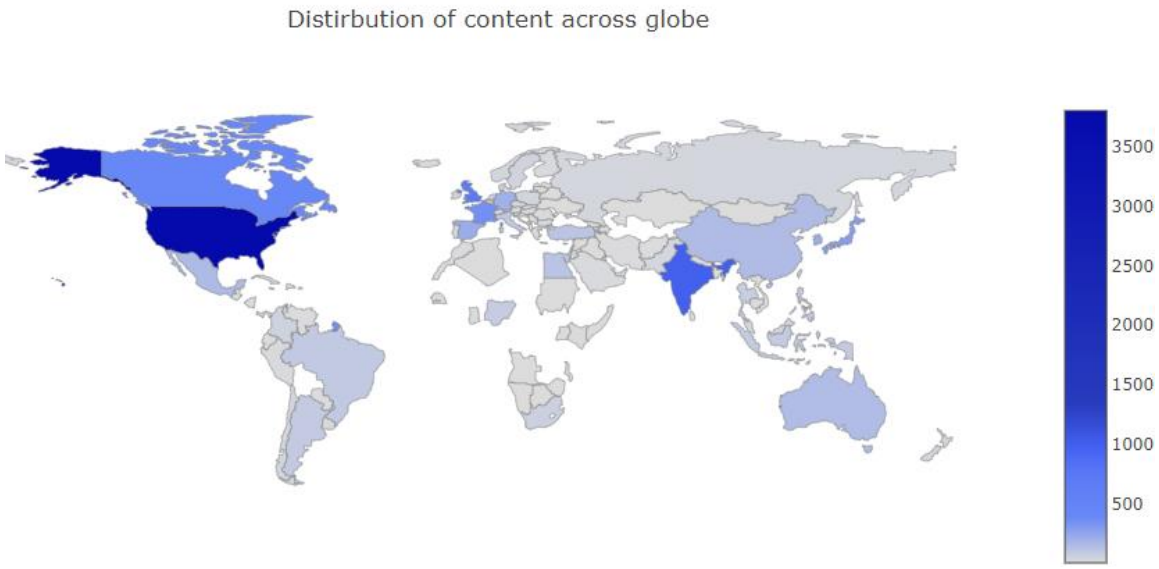



*Figure 3-2 : Distribution of digital content across globe*

The map visualization provides the overall idea that the USA, Canada, and India are top providers of digital content on Netflix. It is also noted that these countries have the highest number of Netflix Subscribers. However, map visualization does not give many details. Map visualization does not show several variables at the same time. Alternatively, we found a simple bar chart (*Figure 3-3*) to resolve some of our problems.

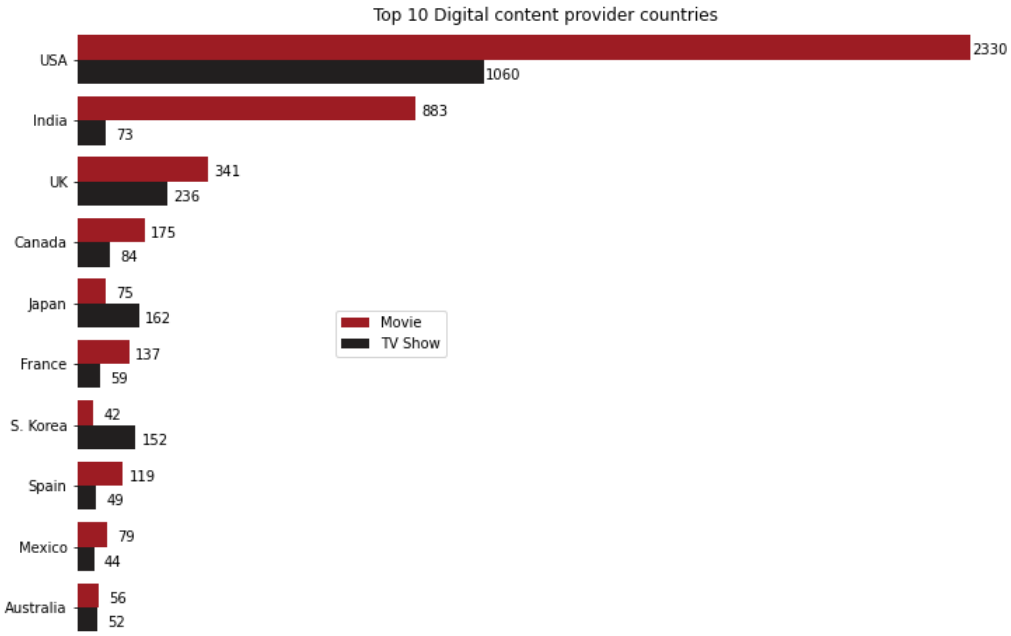



*Figure 3-3 : Top 10 countries in digital content producing*

As seen in the figure above, we can answer several questions. For example, from visualization Figure 3-3, we can identify that India and USA are the largest movie producers and people in that region tend to watch more movies than TV-Series. On the other hand, Japanese and South Korean have more TV shows available than movies. So, it can be inferred that Japanese and South Korean TV shows, they get more popular. A famous example of this is the south Korean "Squid Game" TV show released by Netflix, which has drastically increased user engagement on Netflix.

Data visualization is not a straightforward process. It requires several iterations, and the final visual should be able to tell good stories. The next section describes some of the visualization generated for this dataset and gives insights into how the user engagement level can be increased.

# 4. Results

## 4.1. Growth of Digital contents

Netflix started its digital streaming service in 2007. We have data for content addition over time.

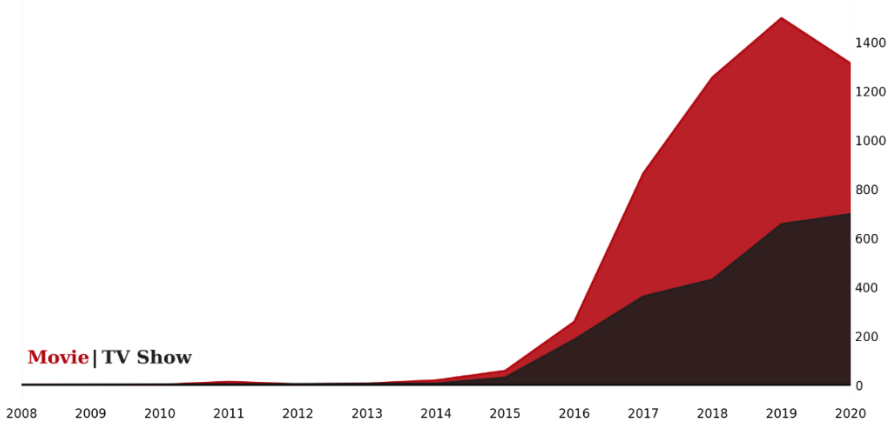



*Figure 4-1 : Movies and TV Show growth timeline*

As we can see in the above visualization, Netflix started adding more movies and TV-shows after 2014. However, this visualization does not clearly tell what the growth rate in content addition. Nowadays most people binge-watch their content on Netflix and other online streaming platforms. So, it becomes imperative to constantly add new content in order to keep users engaged.

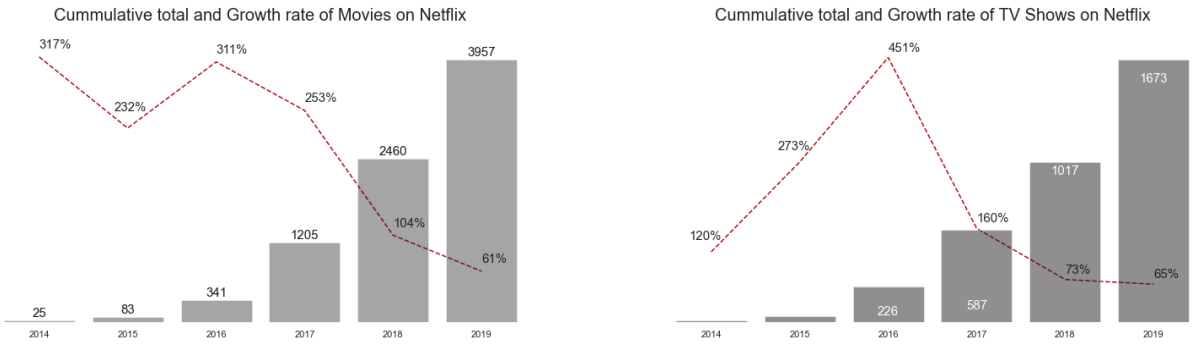



*Figure 4-2 : Growth rate of content addition*

As we can see in *Figure 4-2* visualization, Netflix content addition growth dropped significantly after the year 2018. It is getting lower and lower and in order to keep users engaged, **Netflix must keep adding new content for their users**.

## 4.2. Seasonal Analysis

Netflix adds new content throughout the year. Again, nowadays people binge-watch. They spend lots of hours watching and completing one TV show or movie over a day or weekend. Major subscribers to Netflix are based in the USA. Typically, the USA has holidays in the months of December and January. Hence it is a good strategy to release content during those months. As shown in *Figure 4-3* and *Figure 4-5*, Month of December and January are ideal to release new content. Also, as shown in *Figure 4-5*, more TV shows are released in the month of December whereas more movies are released in January on Netflix. This is in line with our assumptions that more people nowadays binge-watch. Since TV series are of far longer duration than movies, it is ideal to release TV shows in the month of December.

As we see in *Figure 4-4* and *Figure 4-6*, majority of content is released on Friday. As most people across the globe observe weekends on Saturday and Sunday, it is better to release new content on Friday. Also, from *Figure 4-6*, we can see that more than 30% movies and TV-shows are released on Friday. Also, newer TV-shows should be released on Friday than any other day of the week.

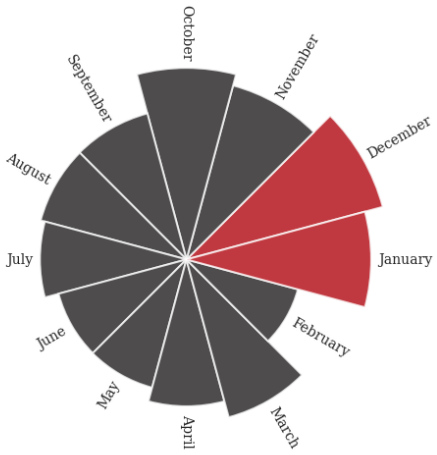



*Figure 4-3 Monthly content addition aggregation*

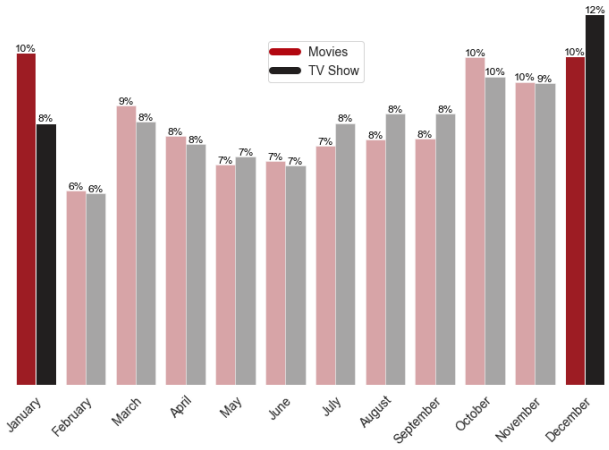



*Figure 4-5 Monthly Distribution of Movies and TV-shows*

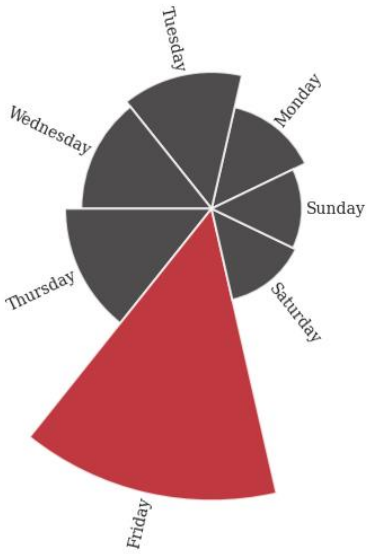



*Figure 4-4 Weekday Content addition aggregation*

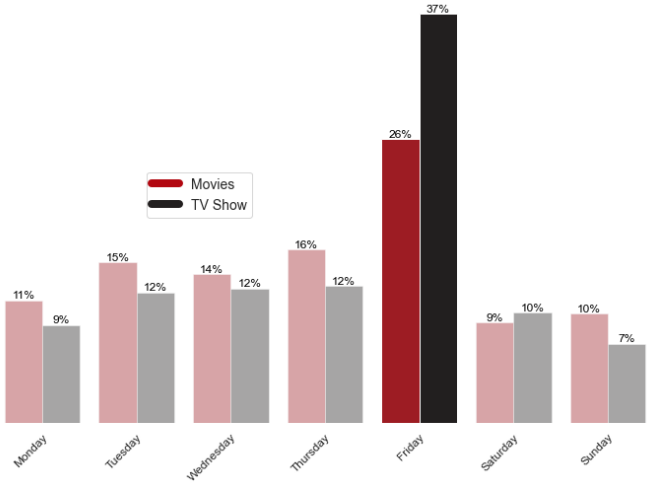



*Figure 4-6 Weekly Distribution of Movies and TV-shows*

## 4.3. Correlation Analysis

Another set of features we wanted to analyze is the correlation between the different genres. It is very important to understand the relationship between one genre and another. As shown in *Figure 4-7*, we can see that certain genres are correlated with one another. For example, independent movies are more likely to be drama movies.

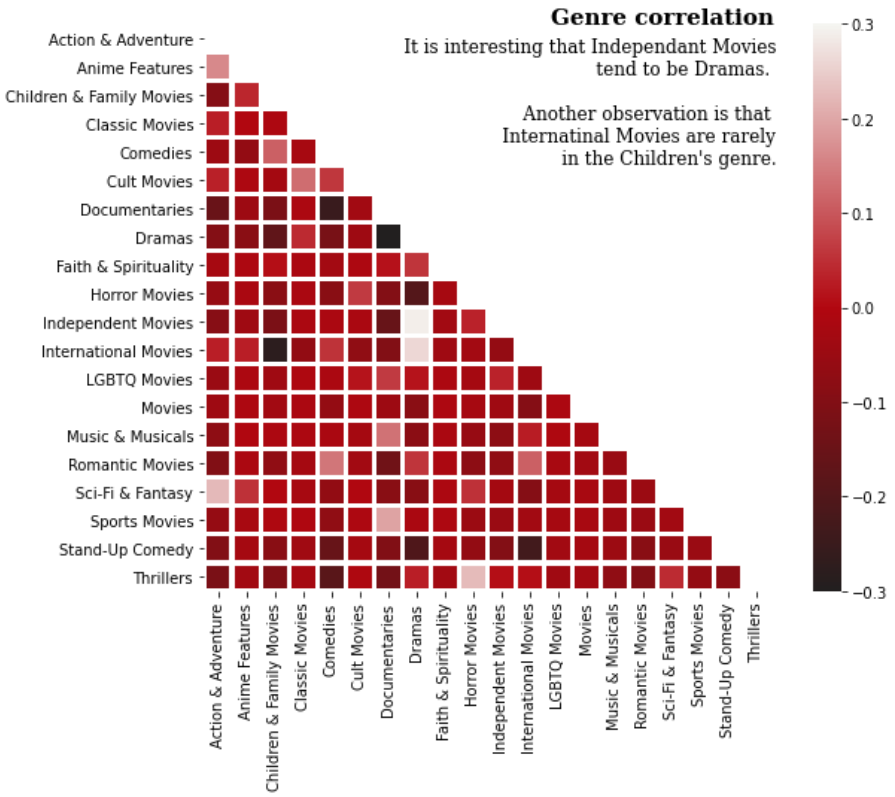



*Figure 4-7 : Weekly aggregation of content addition*

Another way to visualize correlation is to use country against the genre (*Figure 4-8)*. It is interesting to see the popularity of various genres in different countries. As we can see, typically Korean dramas are more popular in South Korea. Japan is more into anime movies and TV shows. British TV shows engage more users from England. The largest number of international movies are from India. One of the key insights this heatmap gives is that local contents are more popular in each of their countries. Hence, Netflix should focus on catering to tailor-made needs for each geographical region.

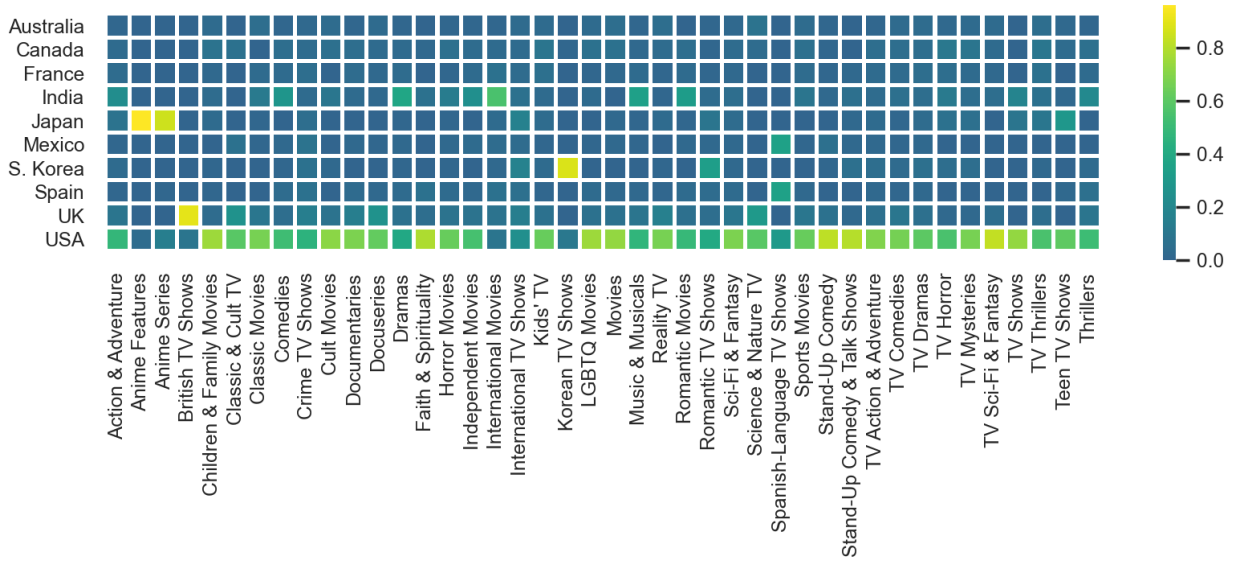



*Figure 4-8 : Countries and Genres correlation*

Further correlation between various age groups and countries also provides different insights (*Figure 4-9*). In India, the teenage audience is more engaged with Netflix. Again, this correlation shows that there is very less content available for kids. This is a potentially untapped market that Netflix should target.

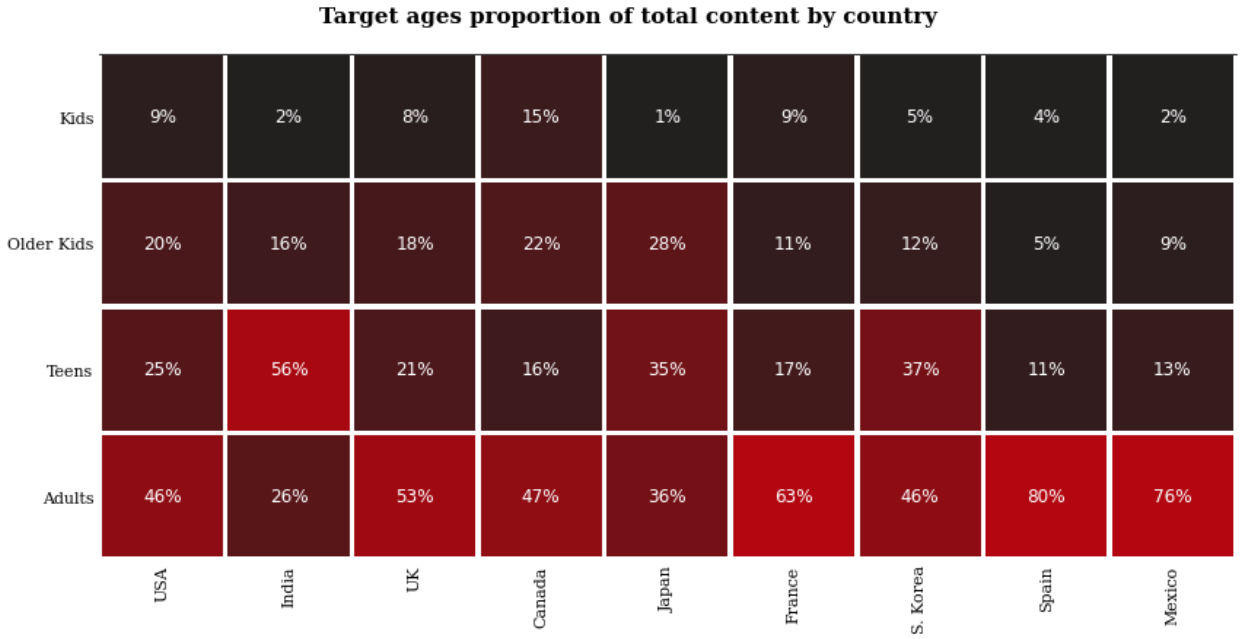



*Figure 4-9 Countries and Targeted Age Group correlation*

## 5. Discussion and Conclusion

This This paper tries to address the strategy for Netflix to increase user engagement levels. This paper solely suggests a strategy based on available content data on Netflix. Since more people are constantly binge-watching TV shows or movies, it is leading to additional pressure on Netflix to add new content. Also, the content needed for each geographical region varies largely. Hence, this paper provides insight into considering the different needs of aspects of geographic regions.

One of the prime reasons we observed is that new content addition growth has significantly declined on Netflix in recent years. It could be an effect of the global pandemic, however, there is a constant need for entertainment, and for that Netflix should constantly add new content. This paper does not consider content that was on Netflix and already removed due to the unavailability of that information.

Most users complete the TV series over a weekend. Hence, it is very important to release content at an ideal time. This article identifies the ideal time for releasing new content as well. It is noted that there is no significant difference in movie content between other months vs months with Major Holidays. However, demand for TV shows is high during the month of December and January. As there is the majority of subscribers of Netflix from the USA, and the USA does observe several holidays during this month, it is the ideal time for releasing new content. In terms of weekdays, Friday is clearly a winner. As most of the users observe weekends on Saturday and Sunday, Friday is by far the best day to release new content.

In this paper, we have also analyzed the correlation between several important features. We have analyzed correlations between different genres. This correlation provides important insights regarding related genres. Netflix can use this insight to find the likelihood of users' preferences. It is very likely that users fond of one genre may be interested in similar genres and this could play a huge role in their recommendations system. Also, the correlation between the different countries and genres is also interesting. It is observed that most countries prefer their local content. Also, some genres are more popular in one country than another. However, it is noted that most of the content is produced by the USA, and hence, there is a chance of a biased conclusion here. It is recommended to perform a thorough analysis of these features before implementing any strategy.

This paper also addresses the potential untapped market that Netflix should focus on. It is observed that there is very less content available for kids. Adding more content related to kids will increase Netflix user engagement level. We have analyzed the correlation between countries and the targeted age groups for Netflix content. For most countries, Netflix content is highly focused on adults. However, kids and teenagers can be engaged if Netflix adds more relevant content for these age groups. This is a great opportunity for Netflix.

In summary, the article addresses a few important strategies that Netflix can implement to increase user engagement levels. At the same time, it is important to note that Netflix should not rely on just content information. They should combine this information with other data such as user demographic, engagement duration as well as revenues. This will help Netflix get back on track and improve the company's financials.

## 6. References


1. https://medium.com/analytics-vidhya/netflix-movies-and-tvshows-exploratory-data-analysis-eda-and-visualization-using-python-80753fcfcf7
2. https://www.kaggle.com/code/niharika41298/netflix-visualizations-recommendation-eda
3. https://www.dataquest.io/blog/comical-data-visualization-in-python-using-matplotlib/
4. https://en.wikipedia.org/wiki/Netflix
5. https://www.kaggle.com/code/nikunjmalpani/netflix-movies-and-tv-shows-data-visualization/notebook
6. https://www.kaggle.com/datasets/shivamb/netflix-shows?select=netflix_titles.csv
7. https://www.statista.com/statistics/250934/quarterly-number-of-netflix-streaming-subscribers-worldwide/