Streaming the Past: Analyzing Decades of Global Film and TV Trends Using 2021 Netflix Catalog Data

Vivian Ning
TU Delft
Delft, the Netherlands
V.Y.Ning@student.tudelft.nl

Elena Mihalache
TU Delft
Delft, the Netherlands
e.mihalache@student.tudelft.nl

Madalena B. G. P. Plácido
TU Delft
Delft, the Netherlands
mbarrosogomesp@tudelft.nl

Abstract

In this study, we explore global trends in film and television production using the Netflix dataset[1], focusing on content released over the past several decades. By analysing metadata such as release year, country of origin, genre, content type, and rating, we aim to uncover how media production has evolved across time and regions. We apply time series analysis to identify long-term patterns and shifts in global media output, and we investigate trends in content genres and maturity ratings. The same trends are explored across the content's country of production. Furthermore, we employ unsupervised machine learning techniques to cluster content and reveal latent structures in the data. These clusters highlight temporal and geographic patterns in media production and distribution. While the Netflix dataset provides a rich source of information, we acknowledge its limitations, particularly the fact that it only includes content available on the platform during a limited time frame in 2021. Despite this constraint, our findings offer meaningful insights into the dynamics of global content creation as reflected through a major streaming platform.

1 Introduction

Let's be honest, who hasn't used Netflix at some point? Whether it's watching a new series, rewatching a comfort movie or exploring something from a different country. Netflix has become a central hub for global entertainment, but it isn't just a streaming service. It's also a goldmine of data, with content from dozens of countries spanning many genres and decades. So Netflix offers a unique opportunity to study global media trends. We can gain insights into cultural shifts, industry priorities and the globalization of entertainment, by analysing what gets produced, when and where. In a world where media consumption shapes perception and identity, understanding Netflix's catalogue means understanding modern storytelling itself.

In this paper, we dig into a 2021 Netflix dataset of movies and TV series to explore global trends in media production. Global trends in this context refer to patterns and changes in film and television content across multiple countries. By analysing metadata such as release year, country of origin, genre, content type (movies vs. TV shows), and maturity rating, we aim to uncover how content creation and distribution have changed over time and space at an international scale. We look at long-term trends and apply machine learning to discover hidden structures in the data.

Our research is guided by the following **research question**:

> How have global film and television release trends evolved over time, and what **temporal or regional patterns** can be uncovered in the Netflix dataset?

To explore this further, we consider three key **sub-questions**:

S1. How has the distribution of genres, content types, and maturity ratings changed over time globally, and what patterns can be observed in their release timing?
S2. How do content type, genre, and maturity rating distributions differ across countries, and how might these patterns reflect cultural norms or commercial strategies?
S3. What latent temporal or regional patterns can be uncovered through unsupervised learning techniques applied to Netflix metadata, and what do these patterns reveal about global content productions?

Through our analysis, we hope to better understand the dynamics of modern content production and how cultural signals are embedded in what we choose to watch, and when. The implementation of our project can be found **in this repository**.

2 Background

2.1 Context

The rise of Netflix as a global entertainment powerhouse has transformed both content production and consumption. Since its international expansion in 2010, Netflix now operates in over 190 countries [2]. This expansion coincides with a strategic shift from licensed programming to a prolific portfolio of original productions spanning multiple languages and regions. By late 2022, Netflix Originals constituted over 83% of the global Top-10 content, underscoring their dominance in attracting viewers [4].

One prominent theme in media studies is **glocalization**, how Netflix blends global distribution with local production. Neira et al. [10] highlight Netflix's strategic efforts to place non-English-language series into the global mainstream, using data-driven promotion and recommendation algorithms to reach audiences across more than 50 countries. Meanwhile, scholars like Lobato argue that Netflix's global libraries are rather shaped by regional licensing deals, regulatory frameworks, and cultural consumption patterns [16].

Given Netflix's expansive reach and influential role in shaping viewing habits, its catalogue provides a rich lens through which to study global media trends. The Netflix dataset captures not only what content is made and where, but also how platform strategies

[1] **Netflix dataset used:** https://www.kaggle.com/datasets/ariyoomotade/netflix-data-cleaning-analysis-and-visualization

DSAIT4055-Q4-25, TU Delft

evolve over time to cater to diverse regional markets. By analysing this dataset, we gain valuable insights into the intersection of technology, culture, and commerce. This should reveal how global media is produced, curated, and consumed in the streaming era.

2.2 Existing work

On the audience side, recent analyses reveal that viewing habits continue to reflect deep-rooted historical, linguistic, and cultural affinities. For instance, network-based studies of Netflix viewing across 71 countries show strong clustering along continental and language lines, even as mega-hits like Squid Game extend South Korean influence globally [5]. These studies usually focus on content preferences, and since Netflix emphasizes producing content that reflects diverse regions and cultures to engage unique audiences [5], it is also interesting to look the other way. Namely, whether these shifts in content preferences are reflected in the features available in the Netflix catalogue, in terms of country or year of production.

2.3 Supporting work

Hidalgo-Marí et al. [1] examine Netflix's original productions using quantitative methods, including genre and temporal trends, and validate the need for data-driven analysis of production patterns. They specifically point out Netflix's lack of transparency in the absence of official totalling data on the company's offer as a reason for the need for more comprehensive data and analysis to shed "encyclopedic light" on Netflix's original content. Their study is limited to fictional series available on the Spanish version of Netflix from 2013 to 2019, and they suggest repeating similar descriptive and exploratory studies is essential in providing "very relevant information on the production standards of the global television market in streaming" [1, p. 11].

Thesni et al. [14] explored Netflix's global catalogue through Tableau-based visualizations, offering a descriptive overview of content availability and genre popularity. Our work goes deeper into the analysis by incorporating temporal and regional exploration, and using unsupervised clustering techniques to uncover latent patterns within the dataset.

3 Methodology

Our approach follows the structure of the three sub-questions laid out in Section 1. Each question zooms in on a different angle of the Netflix dataset: how content has changed over time (S1), how it varies across countries (S2), and what patterns we can uncover using unsupervised learning (S3). Instead of treating the dataset as one giant block, we break it down to better understand the who, what, when, and where of Netflix content. Each part of the analysis is designed to shed light on specific pieces of the bigger puzzle.

To answer the main research question, we mix straightforward techniques like visualizations and statistical summaries with more exploratory ones like clustering and dimensionality reduction. While we go into detail in each sub-section, the core idea is to use the right tools for each question: timeline plots and genre breakdowns for S1, regional comparisons for S2, and pattern discovery for S3.

By combining the results of these three parts, we build up a layered view of global media trends on Netflix. S1 and S2 give us clear signals about how and where content is produced and released, while S3 helps us find subtler groupings and relationships that are not immediately obvious. Together, these perspectives allow us to answer the research question not just with surface-level patterns, but with insights into how global entertainment is evolving, both in what we watch and how that content is shaped behind the scenes.

3.1 Preprocessing

Regarding the preprocessing for genres, the original dataset had multiple genres associated per title. These were expanded, which resulted in 42 distinct genres. Since this number is too high for effective analysis, we grouped them into 19 broader categories, following Netflix codes[2]. Notably, the category "TV Show" was deemed too broad, so its subcategories are also analysed separately, which can be seen in Appendix A Table 2 and in Figure 2e. This approach mirrors prior work on systematic genre evolution, such as the thematic analysis of Bollywood genres by Mohanty et al. [8]. Here, data-driven grouping of themes over time enabled clearer temporal insights.

Similarly, for maturity ratings we conducted an analysis with both detailed and generalized ratings (Appendix A Table 3). This is because some ratings only differ in terminology depending on whether the content type is a movie or TV show. Additionally, we mapped the 14 maturity ratings to 7 age-appropriateness categories, for a more meaningful analysis. Moreover, the dataset initially included 85 countries. However, over 70% of these countries had less than 50 shows each. Considering this limited data may not provide representative insights, those countries were excluded from the analysis, resulting in a dataset of 24 countries.

Lastly, to ensure fair comparisons and improve interpretability, we normalized values across content type, maturity rating, and genre to account for differences in total yearly output. These preprocessing steps enabled consistent trend analysis across time, while preserving enough granularity to reveal meaningful patterns.

3.2 Temporal Trends in Global Content (S1)

To begin, we explore how Netflix content has changed over time (S1). This involves a series of temporal analyses aimed at understanding the evolution of content in terms of volume, type, and thematic trends. Using the metadata, we first compare the total number of releases per year to observe overall growth or decline in Netflix's content output. We then break this down further by analyzing the number of movies and shows released each year, and how these distributions vary by genre. This helps identify which genres are gaining or losing prominence over time, and whether different content types (e.g., movies vs. series) follow distinct trends.

We also examine maturity ratings, such as PG, PG-13, R, etc., by year to explore how the tone and target audience may have shifted. For instance, a rise in mature-rated content over time could indicate a trend toward edgier shows and movies.

3.3 Regional Variation in Content Characteristics (S2)

After analysing the temporal evolution of content releases, the next natural step is to examine how these releases vary across countries.

[2] https://www.netflix-codes.com/

This involved exploring metadata from the Netflix dataset, namely content type, maturity rating and genre, to understand regional differences.

We begin by visualizing the number of releases by country, distinguishing between movies and TV shows. Since the amount of releases varies significantly per country, the data used in the analysis is normalized to ensure fair comparisons. Accordingly, we visualize the proportion of content types by country. Next, normalized stacked bar charts are used to illustrate the relative distribution of maturity ratings and age appropriateness categories within each country. Given the wide range of genres, heatmaps are used to visualize genre distributions across countries. This provides a clear overview of genre popularity and regional variation. Additionally, more detailed visualizations are provided for grouped genres and sub-genres within the “TV-show” category.

3.4 Uncovering Patterns with Unsupervised Learning (S3)

To identify latent temporal and regional patterns in the Netflix dataset, we employ unsupervised machine learning techniques, specifically clustering analysis combined with dimensionality reduction. Our approach focuses on two distinct clustering analyses: **temporal clustering of years**, and **regional clustering of countries**. Both are based on genre distributions.

For the **temporal analysis**, we first construct a year-genre matrix by aggregating content releases for each year and normalizing genre proportions to account for varying annual content volumes. This normalization ensures that our analysis captures relative genre preferences rather than absolute counts to allow for meaningful comparison across years. We then apply **Principal Component Analysis (PCA)** [3, 11] to reduce the high-dimensional genre space to two components in order to facilitate visualization and preserve the most significant variance in the data. **K-means clustering** [6, 7] is then performed on the PCA-transformed data to group years with similar genre distribution patterns. The regional analysis follows a parallel methodology by constructing a country-genre matrix. Countries are clustered based on their normalized genre distributions, again using PCA for dimensionality reduction followed by K-means clustering. This approach allows us to identify groups of countries that exhibit similar content preferences.

To determine the optimal number of clusters for both analyses, we employ a systematic evaluation using the **elbow method** and **silhouette analysis** across a range of k values (2 to 15 clusters). For both the temporal and regional comparisons, we chose 4 clusters. The choices were made to optimize for both a **high silhouette value** [12] and to be close to the **elbow point** [15]. The respective graphs can be found in Appendix C.

For each resulting cluster configuration, we conduct **one-way Analysis of Variance (ANOVA)** tests [9] with a 1% significance level to statistically validate which genres significantly differentiate between clusters.

The clustering results are visualized through scatter plots of the PCA-transformed data with years or countries labeled and color-coded by cluster membership. Additionally, we generate cluster profile visualizations showing the average genre proportions for each cluster to assess the dominant content characteristics that define each group. This enables us to uncover both obvious and subtle patterns in global content trends that might not be apparent through traditional descriptive analysis alone.

4 Results

4.1 Temporal Trends in Global Content (S1)

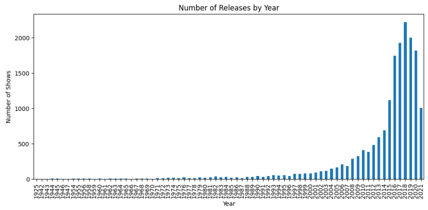
The analysis of temporal trends in Netflix’s content reveals several key insights into the platform’s evolving strategy. Figure 1a demonstrates a significant increase in the total number of releases over time, with a sharp rise beginning in the early 2010s, peaking around 2020, and subsequently stabilizing. This growth aligns with Netflix’s expansion as a global streaming service [1]. The full table containing the actual number of release per year can be found in Table 1 in Appendix A.

Breaking down the content by type, Figure 1b illustrates the normalized proportion of movies versus TV shows per year. While movies dominated the early years, TV shows have gained prominence. Particularly after 2010 which reflects a strategic shift toward serialized content. This trend is further supported by the genre-specific analysis in 1c, which highlights the rising popularity of genres such as TV shows, dramas, and comedies, while traditional genres like classic movies have seen a relative decline. Another genre becoming more popular is foreign movies, this suggests a diversification strategy targeting varied audiences. These shifts point to an intentional reallocation of production resources toward more popular or commercially viable genres.

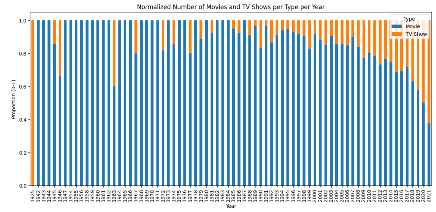
Figure 1e presents the normalized distribution of maturity ratings for movies and TV shows. Each bar represents the proportion of titles released in a given year, categorized by maturity rating. The data reveals several key trends:

- **Early Years (1925-1950s):** Content in the earliest years was predominantly labelled as suitable for *All Ages* or *Young Children (PG Suggested)*. A consistent absence of explicit ratings is observed, likely due to the lack of standardized classification systems during that period.
- **1960s-1980s:** During this period, there is a noticeable increase in the proportion of titles marked *Not Rated*, as well as a gradual emergence of content labelled *Adult (17+)*. However, the dataset for these years appears sparse, likely reflecting limited availability of titles from this period in the Netflix catalogue.
- **1990s-2000s:** A more diverse rating distribution becomes evident, with a steady presence of *Adult (17+)* content dominating the proportions. Ratings such as *Older Children (7+)* and *Teens (PG Strongly Cautioned)* also begin to appear more frequently, suggesting a broader content strategy targeting various age groups.
- **2010s-2021:** From 2010 onwards, the dominance of *Adult (17+)* content becomes even more pronounced, comprising approximately 40–60% of titles annually. The proportions of titles rated for *Young Children* or *All Ages* decline significantly, indicating a shift toward more mature programming in Netflix’s catalogue.

Overall, the data suggests a long-term trend toward an increased focus on adult-oriented content, particularly in the last two decades. While content for children and teens remains present, it forms a



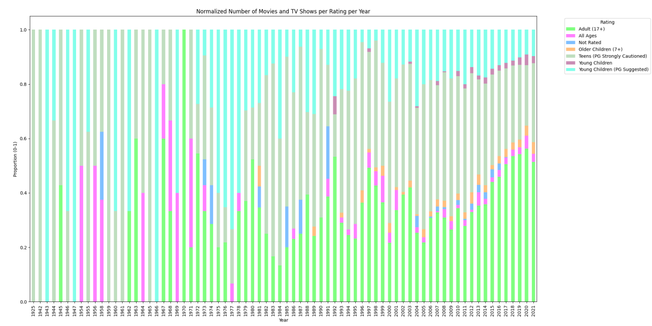
(a) Releases by Year



(b) Type of Releases by Year

(c) Releases per Genre by Year

(d) Releases per Genre by Year (Repeated)



(e) Maturity Ratings by Year

Figure 1: Overview of Netflix Dataset Trends: Release counts, types, genres, and ratings over time.

relatively small proportion of the overall offering. The diversity in maturity ratings in more recent years reflects the platform's expansion into a global and varied audience.

4.2 Regional Variation in Content Characteristics (S2)

The number of releases per country is presented in Figure 2a. The majority of the dataset's content was produced by the United States of America (US), India, and the United Kingdom (UK). This dataset represents the Netflix US catalogue, so a large proportion of American content was expected. Nevertheless, the dominance of the United States and India in the film industry naturally places them among the top contributors. The United Kingdom (UK) also ranks highly, which can be attributed to the shared language and strong media ties with the US. Beyond these, many other countries are represented with smaller amount but meaningful content. These contributions highlight Netflix's efforts to offer a more globally inclusive catalogue and cater to diverse audience interests [10].

The proportions of content type, Movie or TV Show, are visualized for each country's releases in Figure 2b. South Asian countries, namely Indonesia, Philippines, Hong Kong and India, show a strong preference for producing movies. This can be attributed to long-established film industries, such as Bollywood. On the contrary, Pakistan and East Asian countries, such as Japan, South Korea and Taiwan, have a higher proportion of TV shows in their production distribution, reflecting the global popularity of K-Dramas and Anime. On the other hand, European and American countries generally appear to have a more balanced mix of content types, due to more diversified media industries that produce both formats at scale.

Figure 2c depicts the proportions of age appropriateness categories, derived from the maturity ratings, across each country's releases. Content from European and American countries tends to have a higher proportion of content rated for adults (17+), which reflect more liberal cultural attitudes towards sensitive or complex topics in media as well as a stronger emphasis on artistic freedom. In contrast, many Asian and African countries produce a bigger portion of teen-friendly content, possibly due to stricter cultural norms or censorship laws that constrain the prevalence of adult-rated content. The percentage of *Not Rated* content is minimal across all regions.

Next, we look into the correlation between countries and the genres of the content they release. Figure 2d shows a heatmap of the broader genre categories we derived, broken down by country of release. The United States of America dominates in the genre group *TV Show*. *Comedies*, *Dramas* and *Documentaries* are also highly represented in the American content. Meanwhile, genres such as *Children & Family Movies*, *Independent Movies* and *Action & Adventure* appear moderately in the Netflix dataset, reflecting the country's diverse entertainment industry. In the case of India, a substantial proportion of its content falls under *Foreign Movies* and *Dramas*, with a smaller part in *Comedies*. Both the UK and Pakistan contribute to the *TV Show* category. South Korea, while slightly less dominant in this category, still remains a notable presence, which is consistent with its global influence in K-Dramas.

Since the derived genre category *TV Show* is quite broad, we provide a detailed visualization of its sub-genres, i.e. the original dataset' genre categories that were aggregated into this category. Figure 2e shows regional patterns in these sub-genres. The US shows a very strong presence in the *TV Comedies* genre, with *Dramas* and *Kids' TV* also highly represented in its content, reflecting the central role of the US in global TV content production. Genres such as *Crime TV Shows* and *Reality TV*, on the other hand, appear moderately in the Netflix dataset. Some trivial relationships are confirmed, such as the strong association between the UK and *British TV Shows*, and between South Korea with *Korean TV Shows*. Both countries are also well represented in the *International TV Shows* category, while Pakistan and Japan have a slightly bigger presence in said category, suggesting the growing international contributions from these regions.

4.3 Uncovering Patterns with Unsupervised Learning (S3)

4.3.1 Temporal Clustering. Applying PCA-guided K-means clustering to the normalized year-genre matrix (with k = 4) reveals four distinct eras in Netflix's catalogue (Figure 4b). **Cluster 0** represents the modern streaming era (roughly the last 60 years between 1960 and 2021 but primarily the twenty-first century) and is characterized by a balanced mix of *Dramas* (17%), *International Movies* (16%), *Comedies* (12%), *Action & Adventure* (9%), and *Romantic Movies* (4%). **Cluster 1** (middle of the century, 1942-1947 content) is the early catalogue period and is predominantly *Classic Movies* (42%) and *Documentaries* (50%). **Cluster 2** is a transitional period (second half of the century, 1950s-1990s) dominated by a strong blend of 29% *Classic Movies* and 20% *Dramas*, but with slightly elevated *Action & Adventure* and *International Movies* shares (11%). **Cluster 3** (the year 1925) is 100% *TV shows*, but that is because it is an outlier with a single production. **One-way ANOVA** confirms that 13 genres, including *Documentaries* ($p \approx 1 \times 10^{-36}$), *Classic Movies* ($p \approx 7 \times 10^{-25}$), and *TV Shows* ($p \approx 4 \times 10^{-37}$) differ significantly across these temporal clusters, validating the meaningfulness of the splits.

4.3.2 Regional Clustering. Applying PCA-guided K-means clustering to the normalized country-genre matrix (with k = 4) uncovers four clearly separated clusters of production profiles, visualized in the world choropleth (Figure 3) and the two-dimensional PCA scatter plot (Figure 4a). Importantly, "country" here denotes the country of production not of audience preference, so the clusters therefore capture supply patterns in global film and TV making rather than viewing habits. **Cluster 0** comprises countries across Africa (Namibia, Mozambique), the Middle East (Iran, Syria) and Latin America (Guatemala, Jamaica), where a lot of International (34%) and Independent (7%) Movies are made. The modest share of Independent and Sports films hints at festival-oriented or niche productions. **Cluster 1** is made of mostly TV series producers (more than 50%) around East Asia and Eastern Europe, the most significant being South Korea, Japan, or Russia. The map reveals that these countries are neither geographically contiguous nor linguistically homogeneous, suggesting that the common denominator is exportable series. In **Cluster 2** we find mostly countries from the Global South (India, Nigeria, Indonesia, Kenya, the Philippines, Peru, etc.) and Europe which are characterized by *International Movies* (33%), *Dramas* (14%) and *Comedies* (11%). **Cluster 3** includes all big Anglophones (US, UK, Canada, Australia, New Zealand), most of Western Europe (France, Germany, Spain, Italy) and the larger Latin American countries (Brazil, Argentina). No single genre exceeds 18%, creating the most balanced profile of all clusters. Their scatter plot centroid sits near the origin and underscores how volume rather than extreme specialization distinguishes these markets. A **one-way ANOVA** confirms that 10 genres differ significantly across clusters at the 1% level, with the sharpest separations for *Dramas* ($p \approx 1.4 \times 10^{-20}$), *International Movies* ($p \approx 8.8 \times 10^{-14}$) and *International TV Shows* ($p \approx 8.6 \times 10^{-18}$), validating the distinctiveness of the four groups.

5 Discussion

5.1 Findings and Implications for S1

The observed trends in Netflix's release patterns over time point to a clear evolution in content strategy that mirrors broader industry and audience shifts. The sharp rise in content volume (as can be seen in Figure 1a) after 2010 aligns with Netflix's transformation from a regional DVD rental company to a global digital streaming powerhouse. This growth phase not only involved scaling production but also diversifying the type, genre, and maturity of content offered to serve an expanding and segmented user base.

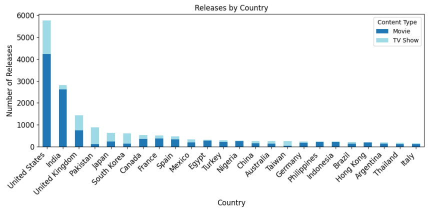
The increasing share of TV shows relative to movies (Figure 1b) indicates a deliberate pivot toward serialized storytelling. This shift likely reflects changing consumer preferences for bingeable, long-form content, which enhances user engagement and subscription retention. Moreover, serialized formats allow for deeper character development and long-term viewer investment, making them ideal for original productions that build platform loyalty [13].

Genre diversification (Figure 1c), particularly the growth in foreign movies and TV-centric genres such as Drama and Comedy, signals Netflix's strategic commitment to catering to international audiences. As Netflix has expanded into new markets, localized content that reflects regional tastes has become critical [16]. This move boosts global user acquisition and supports cultural relevance, which is increasingly important in a crowded content ecosystem.

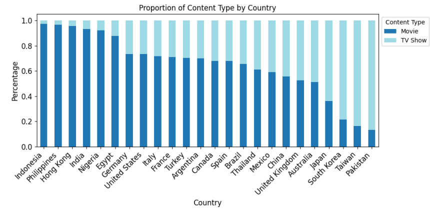
In terms of maturity ratings (Figure 1e), the pronounced shift toward adult-oriented content may reflect both demand and creative freedom associated with streaming platforms. Compared to traditional broadcasters constrained by network standards, Netflix can host edgier, more mature content that appeals to adult demographics. While this has resulted in a decline in the proportion of children-focused programming, the continued presence of a range of ratings suggests that Netflix is still maintaining a family-inclusive catalogue. Finally, the reduction in "Not Rated" content in recent decades, along with the structured classification of newer titles, points to improved metadata standards and the influence of regulatory or parental control expectations in modern content platforms.

5.2 Findings and Implications for S2

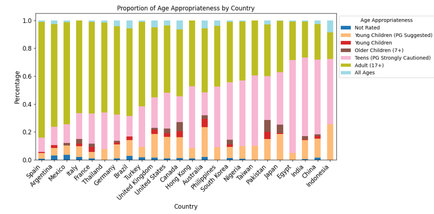
Content distributions of the Netflix dataset were analysed in terms of how metadata, such as type, maturity ratings and genres vary by country. Content type of production preferences reveal regional trends: South Asia favours movies, East Asia leans towards TV shows, and Western countries maintain a balance. In terms of maturity ratings and age appropriateness analysis, Western countries produce more adult-rated media, given the more liberal ideals, while Asian and African countries have a bigger focus on teen-friendly content, with more conservative views on the depiction of sensitive



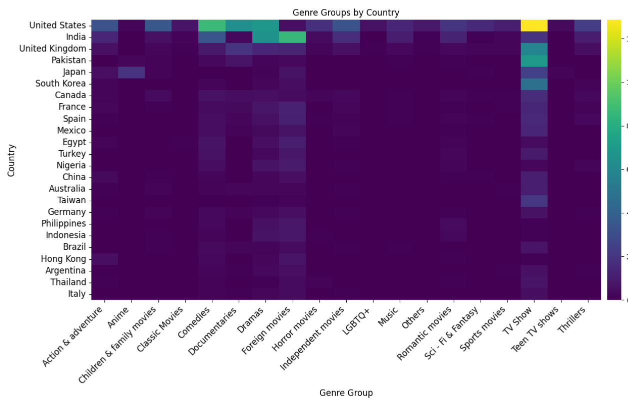
(a) Releases by Country



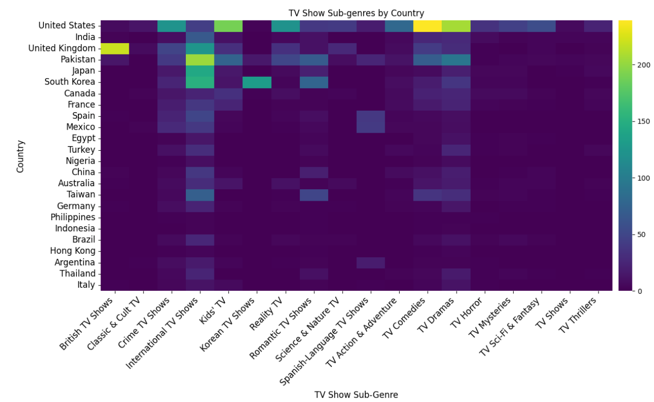
(b) Type of Releases by Country



(c) Age Appropriateness by Country



(d) Grouped Genres by Country



(e) "TV Show" Sub-genres by Country

Figure 2: Overview of Netflix Dataset Trends: Release counts, types, genres, and ratings across Countries.

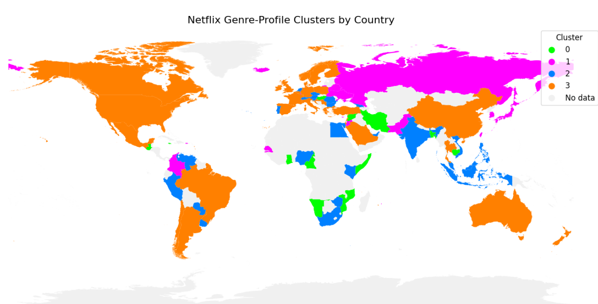


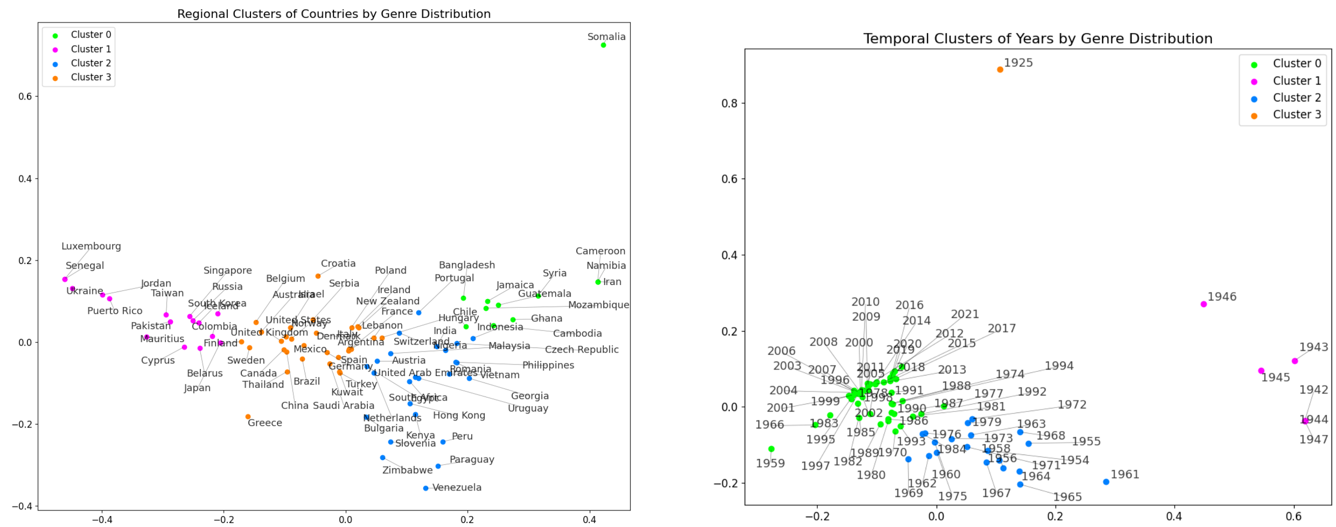
Figure 3: World choropleth.

topics. Genre visualizations across countries highlights national strengths: US leads with a dominant and diverse movie and TV industry, India with contributions in dramas and foreign movies, while the UK, Pakistan, Japan, and South Korea are significantly represented in terms of TV shows.

5.3 Findings and Implications for S3

For temporal clustering, the analysis reveals four distinct temporal clusters that represent different eras in global media production. The lack of early television content on Netflix is reflected in the presence of an outlier from 1925. However, there is some historical content in the form of Classic Movies and Documentaries preserved in the catalogue from around World Word II, which does render Netflix a relevant digital archive for historical media from that time period. We can also clearly see when the bridge between classic Hollywood cinema and model global entertainment was made between the 1950s and the 1990s, as this period introduced Action & Adventure and International Movies. Finally, the modern streaming era shows how Netflix tries to cater to a global audience by having varied content shown in the balanced genre distribution.

Regional clustering uncovered four distinct production profiles that reveal global patterns in content creation. Smaller countries in Africa, the Middle East and Latin America are emerging markets usually with limited production budgets which lead to a great share of Independent Movies. Then, there are countries, especially in East Asia, such as South Korea and Japan, which seem to specialize in TV Series, reflecting the global success of K-Dramas and Anime. The analysis also shows some convergence between developing and developed markets by clustering together countries from the Global South with smaller European nations, which concentrate on Dramas and Comedies. Lastly, the cluster of major English-speaking nations (the US, UK, Canada, or Australia) and other large Western (France, Germany, or Spain) and Latin American markets (Brazil and Argentina) represents mature and diversified media industries, a nature which is reflected in the balanced genre profile. Their uniform distribution suggests these markets have the resources and infrastructure to produce across all genres.



(a) Regional clusters of countries by genre distribution.

(b) Temporal clusters of years by genre distribution.

Figure 4: Clustering results: (a) regional country-genre clusters, (b) temporal year-genre clusters.

Overall, unsupervised learning uncovers that global media production is temporally phased with eras of classic, transitional, modern content, and regionally specialized with series, indie films, diversified hubs. These patterns show how Netflix is both an archive and a curator, as it amplifies certain histories (modern dramas, Korean TV), but also marginalizes others (pre-1950s cinema, low-volume regions). The clustering also reflects underlying industry structures where mature markets show diversification, and emerging markets show specialization.

5.4 Limitations

While the analysis provides meaningful insights into Netflix's content trends, several limitations must be acknowledged. The dataset is restricted to titles available on Netflix as of 2021, which introduces a selection bias toward more recent content up to 2021. As a result, the dataset underrepresents older releases, especially from before the 1990s, leading to sparsity in early-year data and skewing temporal analyses toward the 21st century. Furthermore, since the dataset only includes content that was available on the platform in 2021, it excludes titles that may have been removed prior to that year, limiting the historical completeness and potentially overlooking past strategic shifts.

It is also possible that the PCA dimensionality reduction may have missed subtle genre interactions that could provide additional insights into content strategy evolution. K-means also assumes spherical clusters, which may not capture all content production patterns. Future analyses might benefit from alternative clustering methods that preserve the original feature space complexity.

5.5 Future Work

Future research could focus on longitudinal studies tracking how these clusters evolve over time. It would also be interesting to integrate with viewing data to understand patterns on the demand side. Cross-platform analysis could further help to understand platform specific curation, whether our uncovered patterns for Netflix also hold for other streaming platforms. Furthermore, interdisciplinary research can investigate the relationship between production clusters and cultural and economic factors.

6 Conclusion

In conclusion, the global film and television release trends reflected through the Netflix dataset has evolved significantly. Not only in volume and diversity, but also through distinct temporal and regional dynamics that reflect broader shifts in media production and consumption. On the temporal level (S1) we see a transition toward serialized TV content over movies, diversified genres, and increased focus on adult-oriented productions. These patterns align with global streaming trends that prioritize engagement, localization, and flexible ratings standards. Regionally (S2), distinct preferences emerged: South Asia leans toward movies, East Asia toward TV shows, and Western countries maintain a genre-diverse balance. Maturity ratings and genre breakdowns further reflect cultural and ideological distinctions across global markets.

From a clustering perspective (S3), temporal analysis identified four main eras in global content. From classic cinema to the modern streaming age. This demonstrates how Netflix serves both as a cultural archive and trendsetter. Regional clustering exposed production archetypes: independent-focused emerging markets, TV-centric East Asia, and diversified mature industries in the West and Latin America. Together these findings show that global media production is neither uniform nor random, but structured by time, place, and platform strategy. While constrained by Netflix's 2021 catalogue, the results provide a compelling look at how data-driven curation mirrors and shapes cultural globalization in the digital media landscape.

References

[1] Tatiana Hidalgo-Marí, Jesús Segarra-Saavedra, and Patricia Palomares-Sánchez. 2021. In-depth study of Netflix's original content of fictional series. Forms, styles and trends in the new streaming scene. *Communication & Society* 34, 3 (2021), 1–13.

[2] Tatiana Hidalgo-Marí, Jesús Segarra-Saavedra, and Pilar Palomares-Sánchez. 2021. In-depth study of Netflix's original content of fictional series. Forms, styles and trends in the new streaming scene. *Communication & Society* 34, 3 (May 2021), 1–13. doi:10.15581/003.34.3.1-13

[3] Harold Hotelling. 1933. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology* 24, 6 (1933), 417.

[4] Daphne R Idiz, Julia Noordegraaf, and Rens Vliegenthart. 2025. Culture as window dressing? A threefold methodological framework for researching the locality of Netflix series. *Critical Studies in Television* 20, 1 (2025), 93–117. doi:10.1177/17496020241235579 arXiv:https://doi.org/10.1177/17496020241235579

[5] Nahyeon Lee, Jongsoo Lim, Mina Choi, and Hyeong-Chai Jeong. 2025. Global Streams, Local Currents: A Data Analysis on Global VOD Content Consumption. arXiv:2502.19043 [physics.soc-ph] https://arxiv.org/abs/2502.19043

[6] Stuart Lloyd. 1982. Least squares quantization in PCM. *IEEE transactions on information theory* 28, 2 (1982), 129–137.

[7] James MacQueen. 1967. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, Vol. 5. University of California press, 281–298.

[8] A. Mohanty, A. Mudgal, and S. Ganguli. 2023. Mapping movie genre evolution (1994–2019) using the role of cultural and temporal shifts: a thematic analysis. *F1000Research* 12 (Sept. 2023), 662. doi:10.12688/f1000research.127008.2

[9] Douglas C Montgomery. 2017. *Design and analysis of experiments*. John wiley & sons.

[10] E. Neira, J. C. Gavilán, and J. Sánchez-Navarro. 2023. Standing up to hollywood: the netflix glocal strategy for popularising non-english-language series worldwide. *El Profesional De La Información* (2023). doi:10.3145/epi.2023.jul.09

[11] Karl Pearson. 1901. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science* 2, 11 (1901), 559–572.

[12] Peter J Rousseeuw. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics* 20 (1987), 53–65.

[13] Bridget Rubenking and Cheryl Campanella Bracken. 2021. Binge watching and serial viewing: Comparing new media viewing habits in 2015 and 2020. *Addictive Behaviors Reports* 14 (2021), 100356. doi:10.1016/j.abrep.2021.100356

[14] Faheena Thesni, Goutham Raj, and Alwin Poulose. 2024. Visualizing the Global Streaming Landscape: An In-Depth Analysis of Netflix Content Distribution Using Tableau Data Visualization Tool. *2024 5th International Conference on Innovative Trends in Information Technology, ICITIIT 2024* (2024). doi:10.1109/ICITIIT61487.2024.10580047

[15] Robert L Thorndike. 1953. Who belongs in the family? *Psychometrika* 18, 4 (1953), 267–276.

[16] Michael Wayne. 2020. Netflix Nations: The Geography of Digital Distribution. *Journal of Communication* 70 (05 2020). doi:10.1093/joc/jqz050

A Temporal Trends in Global Content (S1) Appendix

Table 1: Number of Movies/Series per Year

Year	Count
1925	1
1942	4
1943	4
1944	6
1945	7
1946	3
1947	2
1954	6
1955	8
1956	6
1958	8
1959	2
1960	9
1961	2
1962	9
1963	5
1964	5
1965	5
1966	3
1967	10
1968	9
1969	5
1970	3
1971	15
1972	11
1973	21
1974	21
1975	15
1976	23
1977	15
1978	15
1979	27
1980	21
1981	26
1982	36
1983	24
1984	30
1985	20
1986	26
1987	16

Continued on next page

Table 1 – continued from previous page

Year	Count
1988	33
1989	29
1990	42
1991	31
1992	45
1993	55
1994	49
1995	56
1996	44
1997	73
1998	75
1999	82
2000	83
2001	95
2002	109
2003	119
2004	146
2005	165
2006	211
2007	186
2008	288
2009	324
2010	410
2011	384
2012	480
2013	592
2014	693
2015	1119
2016	1743
2017	1930
2018	2223
2019	2000
2020	1821
2021	1005

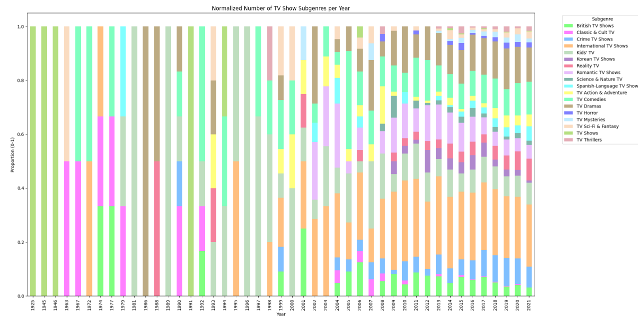


Figure 5: Normalized Stacked Bar Chart of the Subgenre TV Shows by Year.

Genre group	Detailed genres
Action & adventure	Action & Adventure
Anime	Anime Features, Anime Series
Children & family movies	Children & Family Movies
Classic Movies	Classic Movies
Comedies	Comedies, Stand-Up Comedy, Stand-Up Comedy & Talk Shows
Documentaries	Documentaries, Docuseries
Dramas	Dramas
Foreign movies	International Movies
Horror movies	Horror Movies
Independent movies	Independent Movies
LGBTQ+	LGBTQ Movies
Music	Music & Musicals
Romantic movies	Romantic Movies
Sci – Fi & Fantasy	Sci-Fi & Fantasy
Sports movies	Sports Movies
TV Show	Crime TV Shows, TV Action & Adventure, TV Dramas, TV Horror, TV Mysteries, British TV Shows, Reality TV, Kids' TV, TV Comedies, Korean TV Shows, Science & Nature TV, TV Shows, International TV Shows, Spanish-Language TV Shows, TV Thrillers, Romantic TV Shows, TV Sci-Fi & Fantasy, Classic & Cult TV
Thrillers	Thrillers
Teen TV shows	Teen TV Shows
Others	Faith & Spirituality, Cult Movies

Table 2: Broad genre groups and their constituent detailed genres.

Age appropriateness	Original ratings
Adult (17+)	TV–MA, R, NC–17
Teens (PG Strongly Cautioned)	TV–14, PG–13
Young Children (PG Suggested)	TV–PG, PG
Older Children (7+)	TV–Y7, TV–Y7–FV
Young Children	TV–Y
All Ages	TV–G, G
Not Rated	NR, UR
Other	Any rating not listed above

Table 3: Age appropriateness mappings and their constituent maturity ratings.

B Regional Variation in Content Characteristics (S2) Appendix

The proportion of maturity ratings for each country's releases is illustrated in Figure 6.

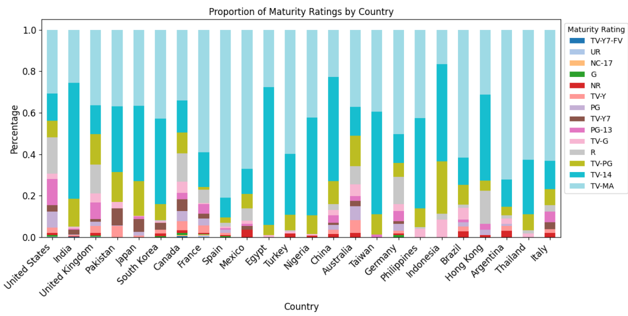


Figure 6: Normalized Stacked Bar Chart of Maturity Ratings by Country.

Figure 7 shows a heatmap of the original dataset's genres by releasing country.

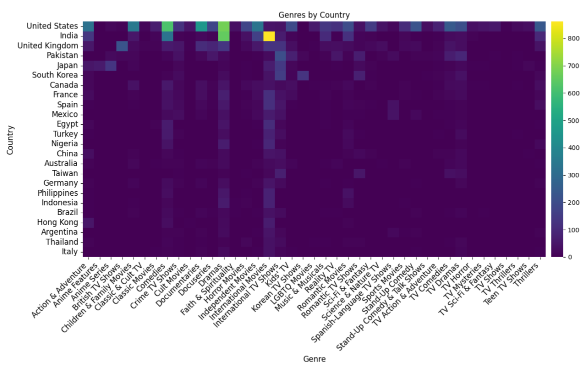


Figure 7: Heatmap of Genres by Country.

C Uncovering Patterns with Unsupervised Learning (S3) Appendix

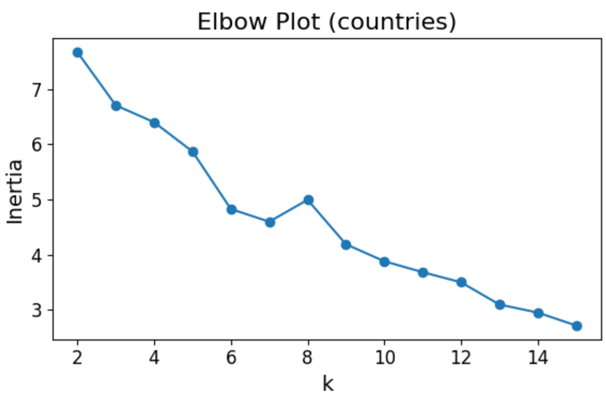


Figure 8: Elbow plot for determining the optimal number of clusters on the country feature set.

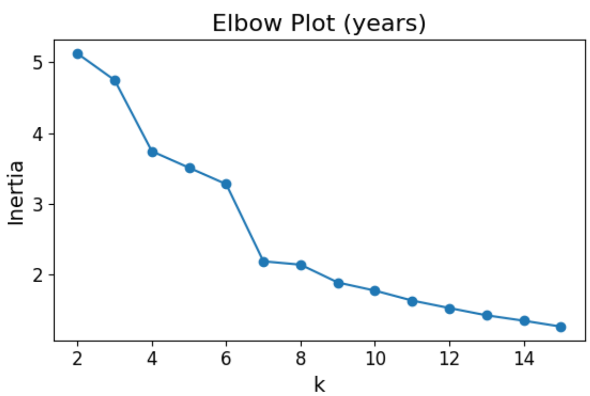


Figure 9: Elbow plot for determining the optimal number of clusters on the temporal (year) feature set.

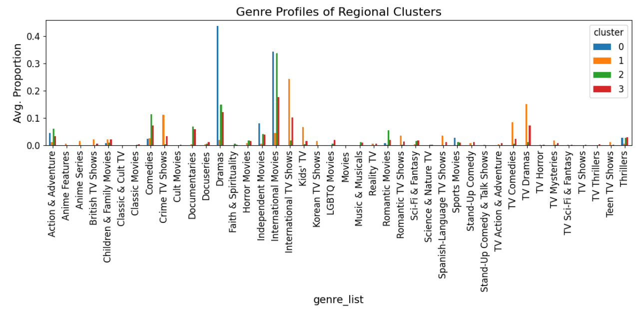


Figure 10: Mean feature profiles of each regional cluster, showing how genres and ratings vary by cluster.

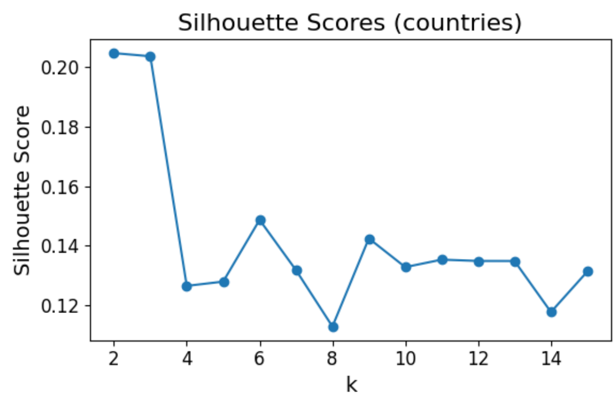


Figure 11: Silhouette scores for the country clustering, indicating cluster cohesion and separation.

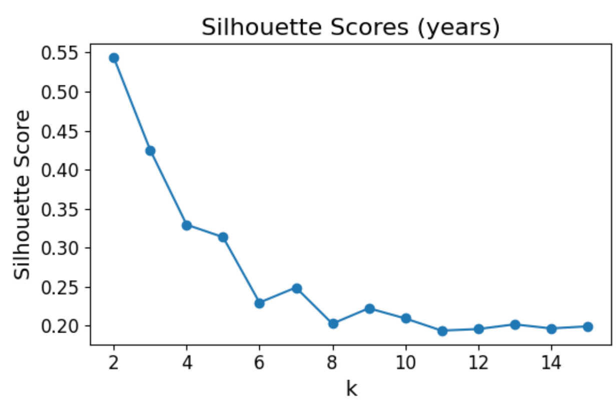


Figure 12: Silhouette scores for the temporal (year) clustering, indicating cluster cohesion and separation.

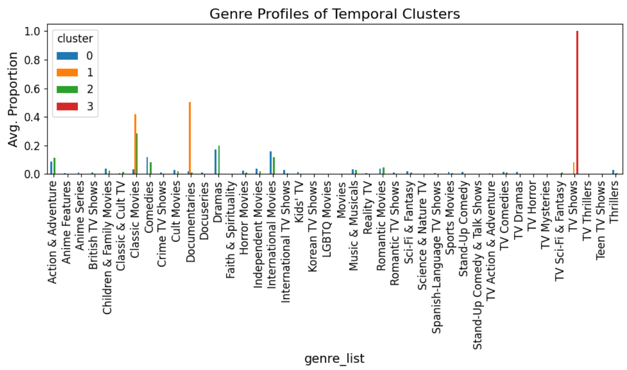


Figure 13: Mean feature profiles of each temporal cluster, showing how genre and rating distributions evolve over decades.