Strategic Content Insights for Netflix: Genre popularity, regional trends and Audience Segmentation

A Proposal report for the BDM capstone Project

Submitted by

Name: Aryan

Email: 23f3001987@ds.study.iitm.ac.in

Roll number: 23f3001987

IITM Online BS Degree Program,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

Table of Contents

Declaration Statement

I hereby declare that the project titled **"Strategic Content Insights for Netflix: Genre popularity, regional trends and Audience Segmentation"** has been conducted using **secondary data**, specifically obtained from **publicly available sources on Kaggle**.

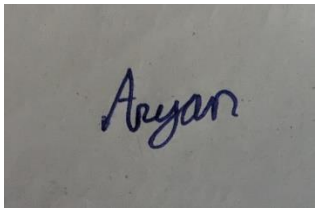
While the dataset used was externally sourced, **I affirm that all aspects of data preprocessing, analytical methodology, visualization, interpretation, and business recommendations presented in this report are entirely my original work.** No portion of the analysis has been copied or derived from any pre-existing research, academic submission, or published material.

The insights and conclusions have been developed independently, applying techniques and concepts learned during the **Business Data Management (BDM)** course. All procedures and analytical approaches used throughout this project have been clearly documented and justified in the report to ensure transparency and academic rigor.

I understand the importance of academic honesty and integrity and confirm that this work is the result of my individual effort. I am fully aware that plagiarism or misrepresentation of work may lead to disciplinary actions as per the institution's policies.

Additionally, I acknowledge that all findings and recommendations provided in this report are specific to the scope of this project and are not endorsed by **IIT Madras** for any broader use. This project is intended solely for academic purposes.

Signature of Candidate:



Name: Aryan

Date: 13/10/2025

1. Executive Summary

A comprehensive analysis was performed on Netflix's 8,800+ title dataset to identify factors influencing **content diversity, regional representation, maturity ratings and audience engagement**, while developing predictive models for age group, maturity rating and the primary country of a show. Deep exploratory data analysis and visualization techniques revealed clear disparities in content availability among different regions, highlighting underrepresented regions with limited local productions.

The analysis revealed that movies dominated most of the content on the platform, with 69% of the catalog indicating a strong focus on globally appealing content. Moreover, it was observed that most of the content came from the USA and India (more than 50%). However, regional content preferences and variations in maturity rating was observed. By analyzing the correlation between the genre, maturity ratings, origin country, and popularity patterns in evolving content trends were established.

These findings and predictions will provide actionable insights for Netflix to invest in content creation for some of the underrepresented regions or for different maturity and age ratings, making the content more diversified and engaging. This can improve customer engagement and help in market penetration. This report outlines key recommendations and a data-driven roadmap to support strategic content planning and future expansion initiatives for Netflix.

2. Proof of Originality and Data Collection

This research project has been conducted using secondary data that was obtained from publicly available secondary sources. The dataset titled "Netflix_titles.csv" was obtained from the public website Kaggle which contained the detailed information about the TV shows and the Movies on Netflix. No primary data was collected from any source and all the analysis and the insights presented in this report are based solely on this publicly available dataset.

To ensure transparency, reproducibility, and proper documentation of the analytical process the used files have been uploaded in the google drive folder:

1) Netflix_titles.csv – The main dataset containing all records of Netflix movies and TV shows used for analysis
2) Netflix_data_analysis.ipynb – The Colab notebook that documents the complete data processing, exploratory data analysis, and predictive modeling workflow.

3. Meta Data and Descriptive Statistics

a. Meta Data

The dataset used for the analysis comprises of 8807 rows and 12 columns, providing a comprehensive overview of the Netflix Movies and TV Shows data.

Table 1: Detailed Description of each variable outlining the column names, Data types and corresponding definitions.

Netflix Meta data

Column Name	Data Type	Description
show_id	object	Unique identifier for each show
type	object	Indicates whether the title is a **Movie** or **TV Show**
title	object	Name of the title
director	object	Director of the movie/show (if available)
cast	object	Cast members featured in the title
country	object	Country where the title was produced
date_added	object	Date when the title was added to Netflix
release_year	int64	Year in which the title was originally released
rating	object	Audience maturity rating (e.g., TV-MA, PG-13)
duration	object	Duration (in minutes or number of seasons)
listed_in	object	Genre or category of the title
description	object	Brief summary or plot description

b. Descriptive Statistics

As there was only one single numerical column (release_year), we calculated the Mean, Median, Standard Deviation, Min and Max and tabulated it as a part of Descriptive Statistics of our data. These metrics provide an overview of the temporal distribution of Netflix's content production and have been tabulated below.

Table 2: Summary of the descriptive statistics of the release_year column.

Descriptive Statistics for release_year

Statistic	Value
Mean	2014.18
Median	2017.0
Standard Deviation	8.82
Minimum	1925.0
Maximum	2021.0

Key Statistical Insights

The mean release year was calculated to be **2014**, with a **median of 2017**, which shows that a large portion of Netflix's content was released after the year 2015. The relatively **high Standard Deviation** indicates that the Netflix's catalog contains content from a wide timeline hence comprising of both classic titles from as early as **1925** and modern releases up to **2021**. This shows that Netflix includes both old and new titles in its library, offering a mix of classic and recent content to attract different types of viewers.

4. Detailed Explanation of Analysis Process/Method

The analytical process for this project was divided into four major steps, each addressing a specific research objective:

1. Understanding the key characteristics and the content catalog of Netflix dataset.

2.Exploring trends across regions, and identifying patterns related to age rating, maturity genres, and time periods.

3.Performing Descriptive analysis on the dataset.

4.Giving data-driven actionable insights for data-driven decision-making.

4.1 Data Cleaning and Preprocessing

Before doing any type of analysis the raw data should be properly cleaned and preprocessed to ensure accuracy, consistency, and readiness for exploration. Proper preprocessing is a critical stage that determines the reliability of later analysis and interpretation. Errors, duplicates, and inconsistencies were handled to maintain data integrity. The steps are described below:

1.Handling Missing Values

The raw data had some missing values in the columns in some of the columns. The rows which had missing values for critical columns such as **date_added, rating, duration and description were dropped from our dat**a. The rows with less important features missing such as director, cast, and country were handled by filling 'Unknown' at the null places. This preserved record completeness while maintaining neutrality. This approach ensured that no observation was discarded unnecessarily and that statistical summaries remained unbiased.

2. Duplicate Record Removal

Then the data was checked for any duplicates. Although there were no duplicate records in our data, we still ensured the duplicates were removed by the function drop_duplicates(). This ensured that each movie or show title appeared only once in the dataset keeping the data unbiased. The removal of duplicate records was necessary to prevent inflated counts or misrepresentation of patterns in visualization.

3. Data Type Conversion

Some of the columns were not the right data type for the analysis. We changed the **release_year** column of the dataset from **string to int** so that we can perform descriptive statistics on that column and know the mean, median, min, max, and standard deviation of

that feature. Also, the **date_added column was converted into a datetime format** to allow chronological sorting and temporal analysis, such as understanding content addition patterns over time.

4. Categorical Standardization

The categorical columns like **type, rating, and country were standardized** to ensure uniformity. The inconsistencies in text format such as spacing, capitalization, and punctuation—were corrected using string methods. This ensured accuracy in group by operations and visualizations.

5. Column Renaming and Organization

The dataset columns were kept the same with descriptive naming for visualization and clarity of data. Irrelevant columns were retained although they did not contribute to our main motive but this was done to maintain the completeness of the data. These columns were not included in the visualization plots.

4.2 Identification of Key Trends and Patterns

In this step we identified the major trends within the data, specifically those related to content type, genre distribution, release patterns, and regional representation. The aim was to know **how the content library has evolved over time and what are the major factors contributing to its diversity**. The steps are described below:

1. Content Type Distribution

The variable 'type' was analyzed to determine the ratio of Movies to TV shows available on Netflix. Count plots were used to visualize this distribution. This simple analysis showed that Netflix focused more on full-length movies rather than serialized shows as movies contributed more than 65% of the Netflix catalog.

2. Country-wise Content Availability

Then the column country was analyzed to identify the top countries contributing to Netflix's content. A bar chart of the **top 10 countries** was plotted, showing how United States dominated the content count followed by countries like India, the United Kingdom, and Japan. This analysis was essential for the global expansion strategies of Netflix.

3. Genre Analysis

Then the listed_in column was analyzed which had the information of genre. This was done using bar graph to identify popular categories such as Drama, Comedy, and Action. This showed that the Dramas and Comedies dominated the catalog with Action and adventure and documentaries just below them. Horror movies and Standup Comedy were the least popular categories.

4. Temporal Analysis of Content Additions

The release_year and date_added columns were analyzed to understand Netflix's timeline of releases. The number of titles released per year was visualized using line and bar plots. It was observed that there was a significant increase in the number of shows after 2014, which displayed the aggressive content expansion of Netflix during the streaming boom. This analysis also shows the company's focus on acquiring more recent content.

4.3 Descriptive Statistical Analysis

The only numerical column in the processed dataset was the release_year column. Statistical measures such as Mean, Median, Standard Deviation, Minimum, and Maximum were computed to provide a summary of content distribution across years.

The minimum release year was calculated to be **1925** and the maximum was **2021** which showed the wide variety of Netflix's content. The **mean was calculated as 2014** and the **median was 2017** signifying that most of the movies were released after 2015. The **standard deviation** of **8.82** reflects a wide spread in the release years, suggesting that Netflix's catalog consists of both old and new titles.

This descriptive statistical analysis reveals that Netflix maintains a balance between the old titles and the new ones delivering a good user experience to a variety of different users with varied preferences ranging from vintage film enthusiasts to those seeking the latest releases. This statistical insight is crucial for understanding the temporal dynamics of Netflix's content strategy.

4.4 Visualization and Insight Extraction

This step focused on converting the quantitative findings to visualizations. Visualization was performed using Python libraries such as Matplotlib and Seaborn, enabling clear and interpretable graphical representations. The following visual analyses were conducted:

1. Content by Country

A horizontal bar plot was created that showed the **10 most popular countries** which produced the highest number of Netflix Titles. The visualization revealed that USA was at the top, India being second and counties like Mexico and Spain were at the bottom. This reflected Netflix's global reach and investment in regional productions.

2. Movies vs TV Shows

A count plot demonstrated that the Netflix titles were dominated by movies contributing more than 65% to the Netflix catalog. This highlights the company's traditional emphasis on films while still expanding in the TV shows domain.

3. Content Over the Years

A time series visualization of **release_year** was plotted that showed the sudden increase in the number of titles after 2014 marking Netflix as a dominant global streaming platform.

4. Genre Distribution

Visualization of top genres emphasized Drama, Comedy, and Documentaries as the most popular categories. Whereas the least popular genres were Horror and Standup Comedy. This aligns with Netflix's strategy to include both mainstream and niche genres to retain diverse viewer groups.

5. Rating Distribution

A horizontal bar chart was created that showed the number of titles in each maturity rating. This revealed that most of the shows belonged to the TV-MA and TV-14, indicating that Netflix primarily caters to mature audiences while still maintaining a balanced mix of family-friendly options.

6. Other

Several other visualizations were done in the notebook to analyze other features of the Netflix catalog. This included the distribution of movies based on its duration that revealed most of the movies were around 100-120 min long. The number of seasons of the TV shows was also analyzed which showed that most of the TV shows were around 2-3 seasons long. Moreover, cast Size distribution, Country vs Rating, Rating vs Release year were also visualized.

4.5 Actionable Insights and Recommendations

This is the final stage of analysis that focused on translating statistical and visual findings into actionable insights that can guide decisions of Netflix towards content acquisition and production. These insights can help Netflix enhance user engagement and optimize its catalog diversity.

1. Content Strategy Optimization

The analysis revealed that most of the content on Netflix was released after 2015. Netflix can maintain user engagement by continuing to invest in newer content while preserving a curated set of content for the audience who prefer older cinema and shows.

2. Regional focus

Given the majority of the content comes from USA and India, Netflix should try to expand its international collaborations and licensing agreements to include the underrepresented countries like Mexico and Spain making the content more diverse.

3. Genre Diversification

Since Drama and Comedy dominate the catalog, Netflix could explore increasing its offerings in genres such as Sci-Fi, Mystery, and Animated content to attract more varied audiences and differentiate itself from competitors.

4. Age rating balance

Most of the content on Netflix falls under mature-rated content (TV-MA). This suggests there is a potential for expanding family and children’s content increasing the platform’s user base and making the platform more appealing to viewers of all age groups.

5. Data-Driven curation

The insights derived from analysis can be used to recommend personalized collections for users by the use of machine learning algorithms for better content discovery and user experience.

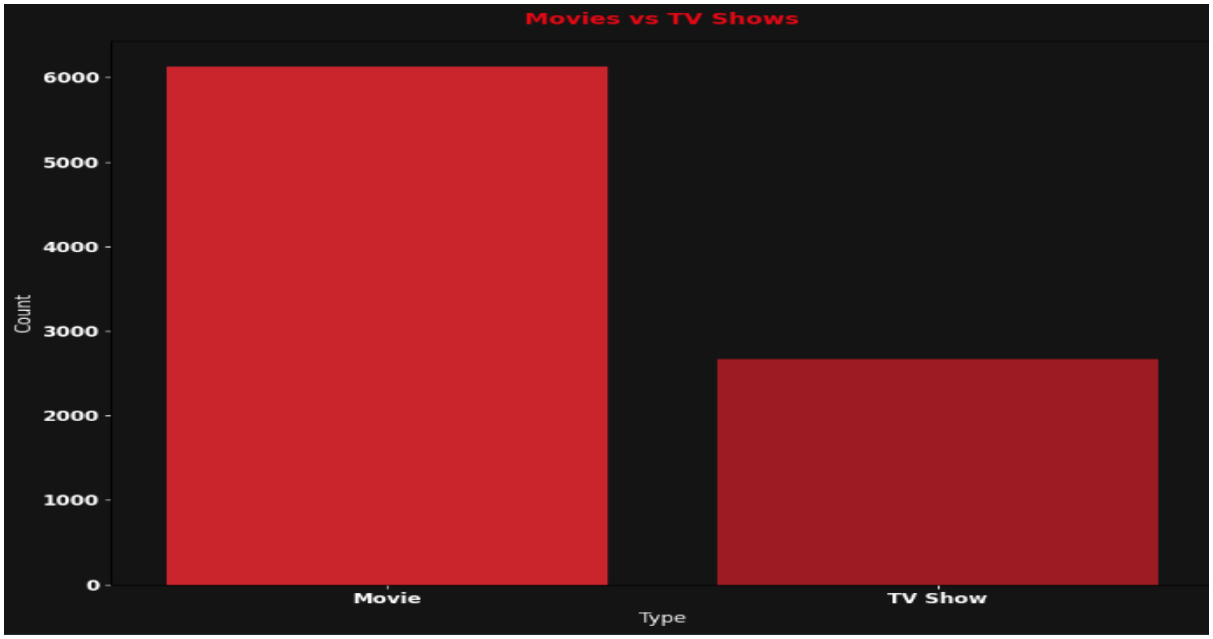
This was the overall approach – from preprocessing to visualization – which provided a comprehensive understanding of Netflix's content library. We gathered actionable insights and key findings by converting the raw data into useful data and analyzing the descriptive statistics and the visualizations related to age rating, maturity rating, and country of origin. These findings emphasize how the evolving catalog has contributed in Netflix being one of the largest streaming services in the world.

5. Results and Findings

5.1 Key Insights from Netflix Content Analysis

1. Distribution of Content Types

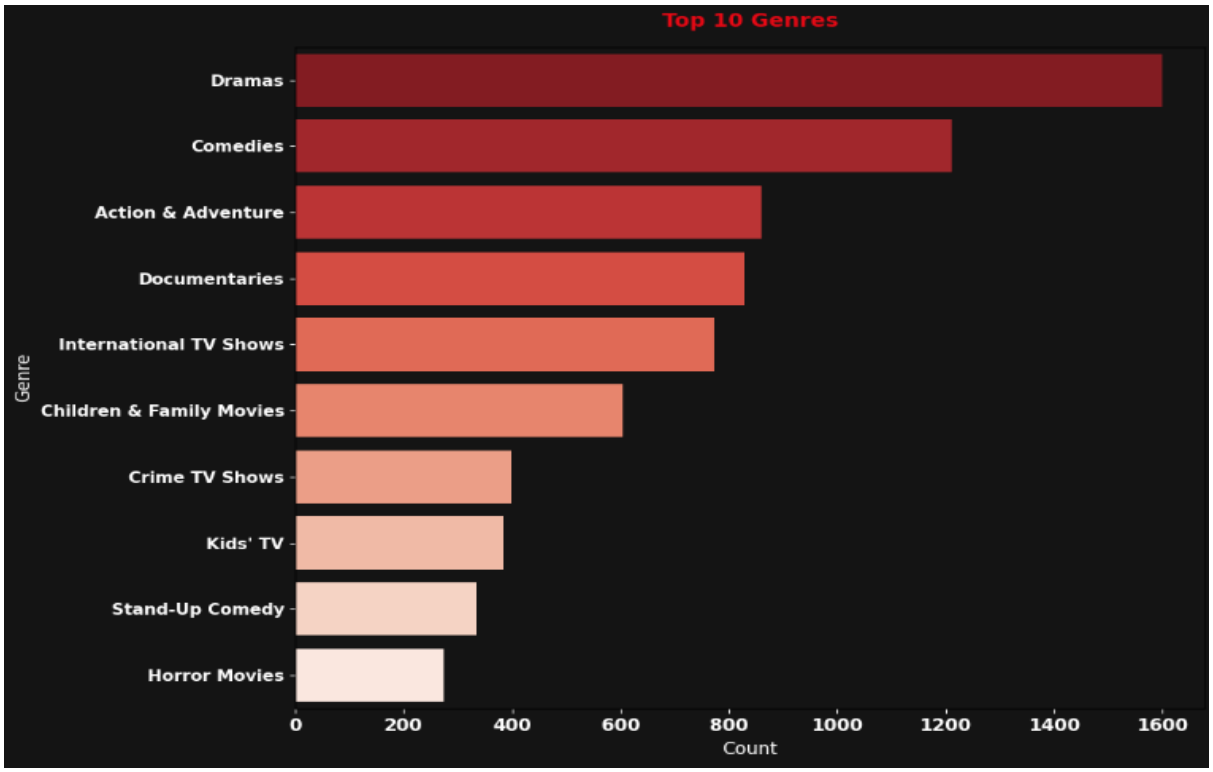
The dataset revealed that the majority of Netflix titles belong to the Movie category. Movies contributed to 69% of the Netflix's catalog and TV shows make up the remaining 31**%**. This indicates that the primary focus of Netflix is on movies, yet it maintains a good amount of serialized content.



The visualization shows the clear **dominance of Movies** over TV shows in the Netflix catalog.

2. Genre Distribution

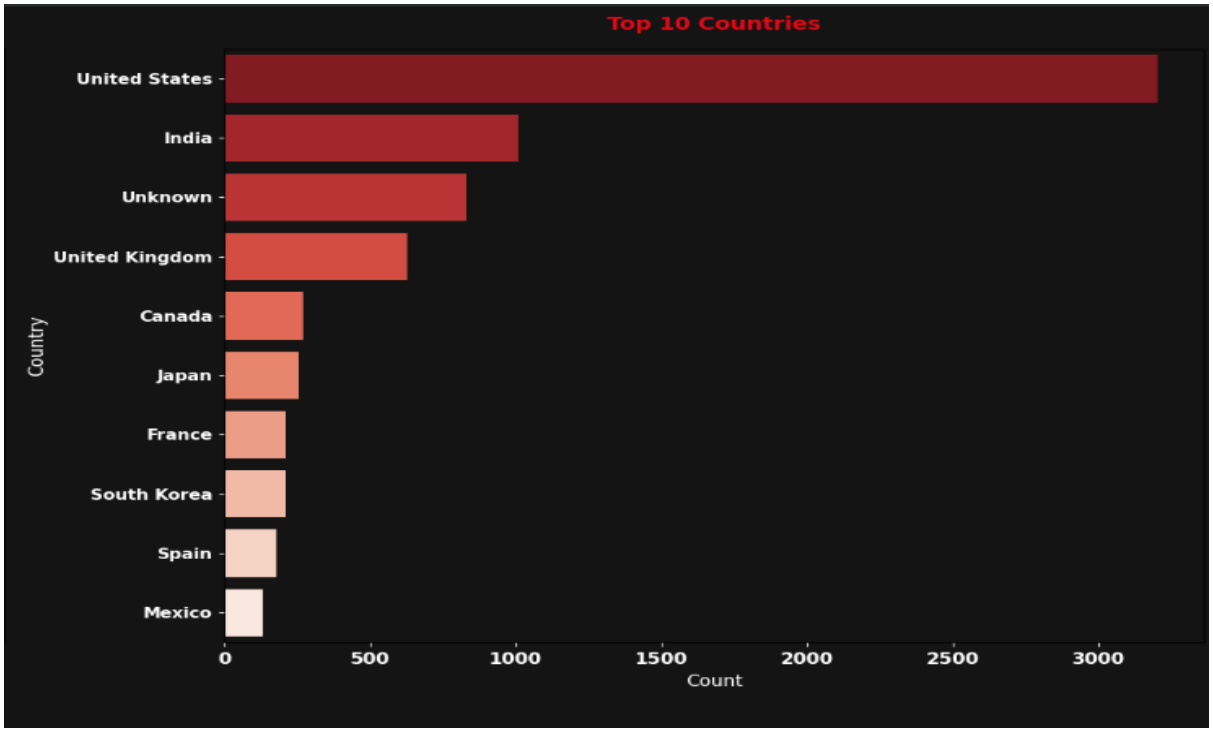
Analysis of the listed_in column showed that the most popular genres were **Dramas, and Comedies** and the least popular were **Horror and Standup Comedy**. A large proportion of shows fall under multi-genre classifications, highlighting Netflix's strategy of cross-genre appeal.



This visualization will depict how Netflix maintains variety across its catalog, emphasizing on the production and acquisition of content of all genres.

3. Country-Wise Content Production

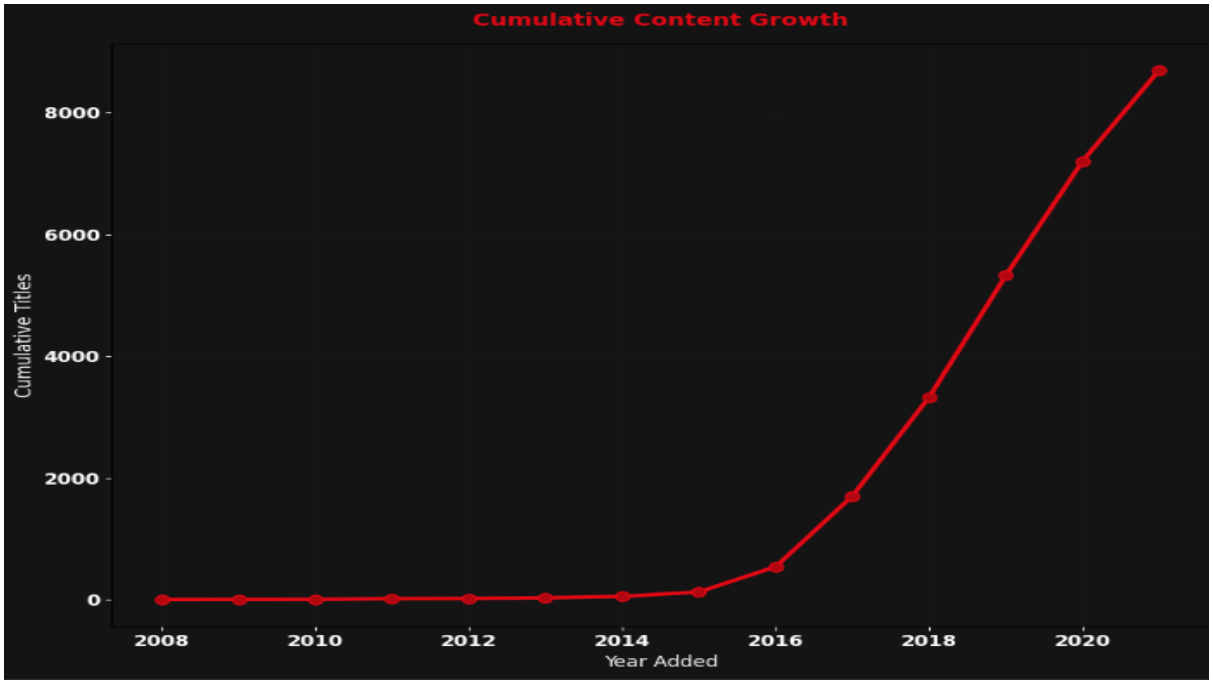
The analysis identified USA as the major content creator in the Netflix's catalog followed by India and the United Kingdom. The lowest in the charts were **Mexico and Spain**. This demonstrates Netflix's reliance on the American entertainment industry, but also highlights its growing investments in regional content from Asia and Europe.



This chart should highlight the imbalance in regional representation, showing that certain regions—especially Africa and parts of South America—remain underrepresented, suggesting future opportunities for content expansion.

4. Year of Release Trends

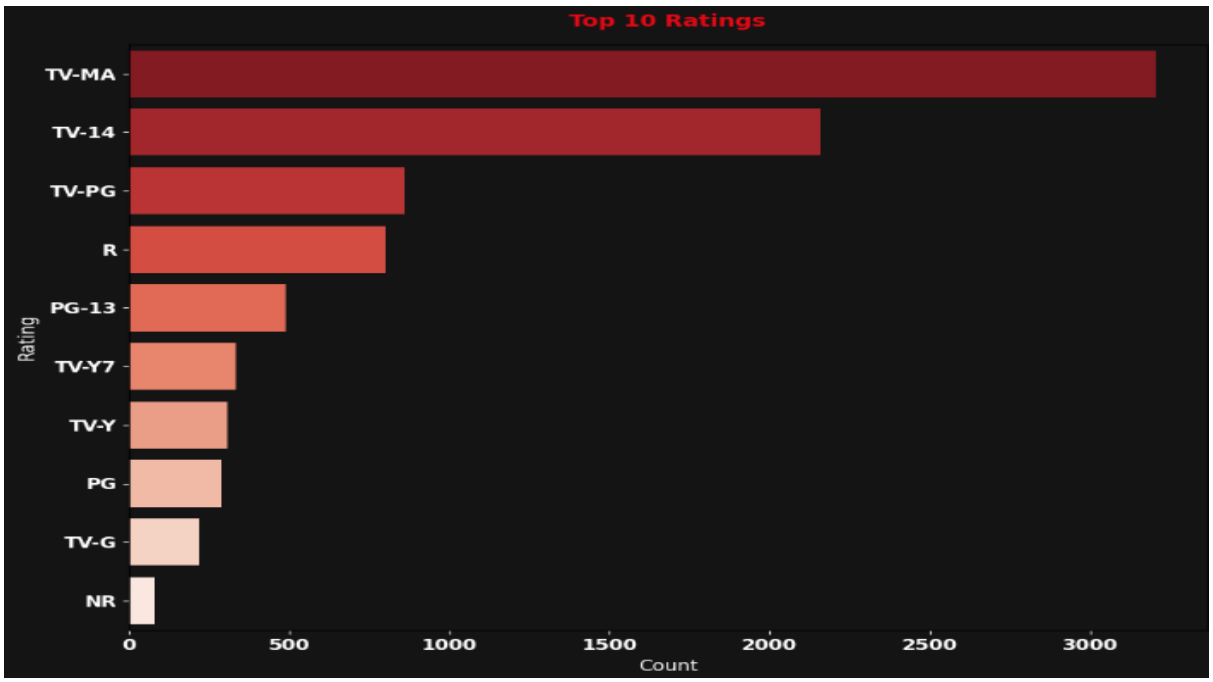
This analysis showed the upward trend in Netflix's content production and acquisition after 2014. There was a spike in between 2017 and 2020 resulting in Netflix's global expansion.



5.2 Analysis of Ratings and Audience Maturity

1. Maturity Rating Distribution

The maturity ratings showed that TV-MA (Mature Audience) titles dominate Netflix's library, comprising nearly **40%** of all content, followed by **TV-14 and R-rated categories**. This shows that Netflix mainly focuses on a more mature and adult audience.

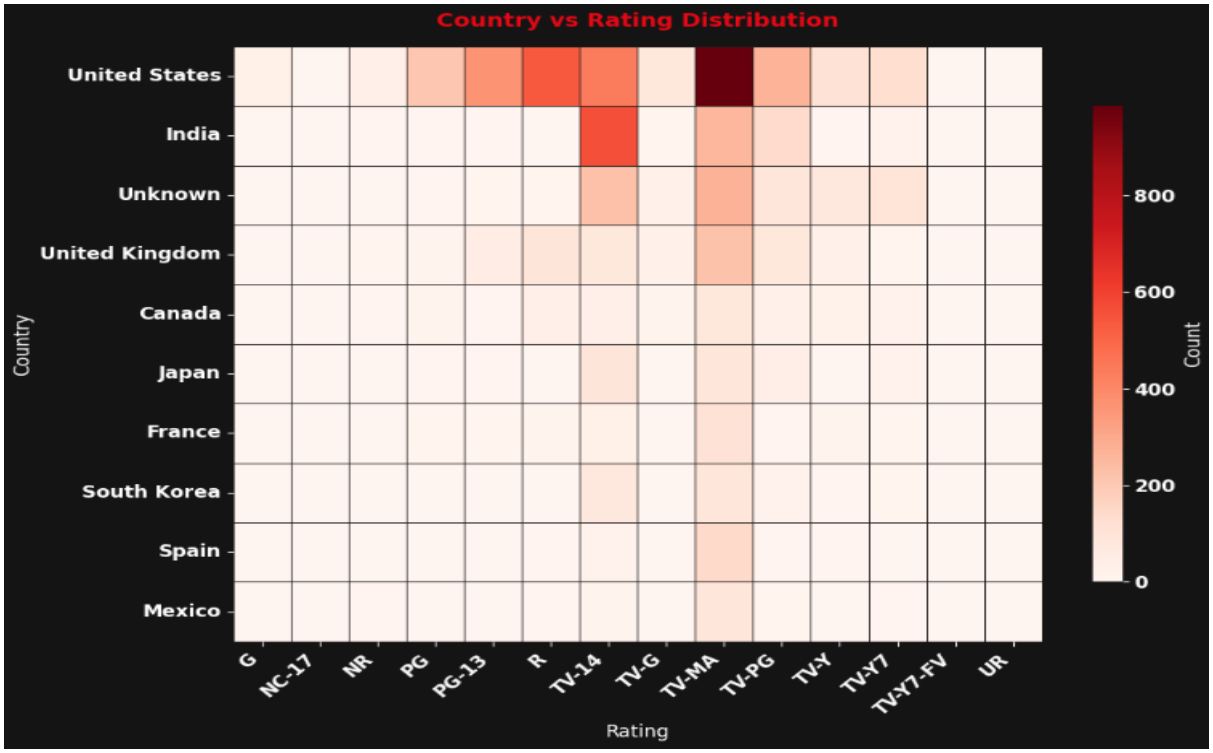


2. Scope for Family and Children's Content

Titles rated **TV-G, TV-Y, and PG** had very little contribution towards the content. This limited no of shows for family and children suggests that there is an opportunity to expand Netflix's content towards these ratings enriching its catalog while diversifying customers. Expanding in this segment could help attract younger audiences and families, thereby broadening Netflix's subscriber base.

3. Regional Trends in Ratings

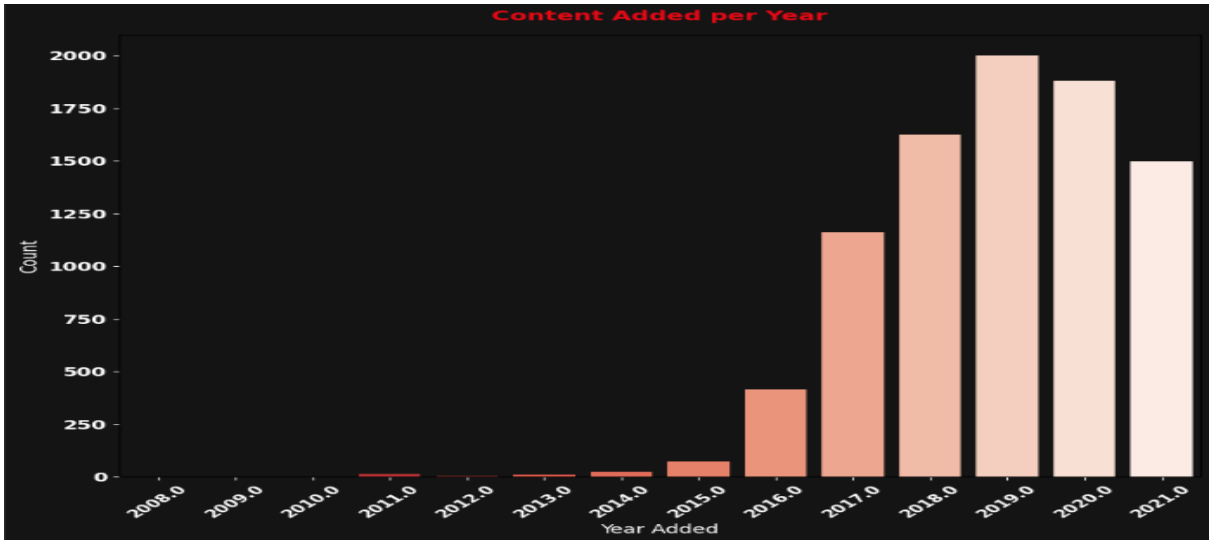
The country vs Rating analysis revealed that countries like USA, India and United Kingdom had more movies in the mature ratings whereas countries **like Japan and South Korea focused more on shows with ratings TV-14 or PG-13**. This indicates that cultural factors and certification norms play some role in determining the content rating.



5.3 Descriptive and Statistical Insights

1. Descriptive Statistics – Release Year

release_year was the only numerical column in our dataset. Upon performing Descriptive analysis, we found out the mean to be 2014 and the median to be 2017 with a standard deviation of 8.82. This showed that Netflix focused more on recent content with some old titles.

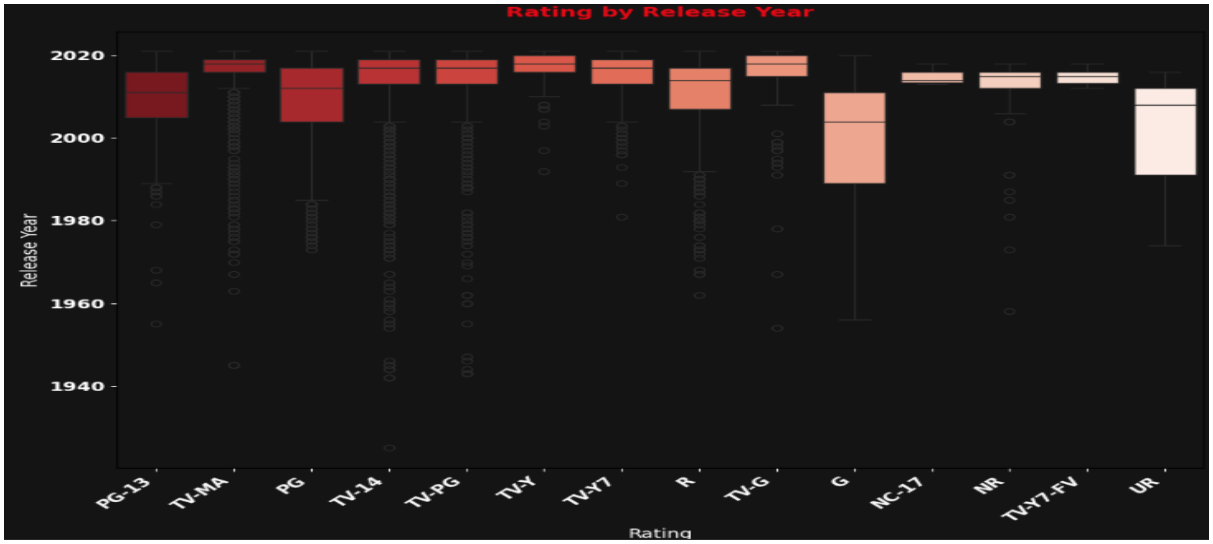


2. Content Longevity and Trends

By the above bar graph, we can also see that **the data is skewed towards recent years**, showing Netflix's attempt at keeping the catalog fresh while keeping some of the old titles. This is consistent with Netflix's shift toward producing original shows and films, which helps differentiate it from competitors such as Disney+ or Amazon Prime Video.

3. Rating over time

The boxplot of Rating by Release Year shows a clear trend toward more mature content in recent years. Earlier titles were mostly rated **G**, **PG**, or **TV-G**, reflecting older, family-friendly content. While in the recent years especially after 2015 there has been an increase in the more mature content rated TV-MA and R. This shows Netflix's growing focus on more adult and mature content to attract a wider global audience.



5.4 Overall Findings and Business Implications

1. Global Expansion and Content Diversity

Netflix's catalog demonstrates a strong dominance of US while showing inclusion of other countries like India and Japan. Though some of the countries of Africa and South America are very underrepresented showing that there is an opportunity for growth in these regions.

2. Shift Toward Originals and Modern Content

The increasing trend in content after 2015 shows Netflix's heavy investment in original productions. This gave Netflix creative control over its content and long-term profitability.

3. Audience targeting

The dominance of mature rated titles shows its focus on mature audience. Although it can be concluded that there is **room for expansion in the family friendly and kids content type**. This could help attract a more diverse subscriber base.

4. Strategic Insights

1) Netflix should focus on increasing regional diversity.

2) Producing more kids and family-oriented content will help in balancing maturity distribution.

3) Investment in culturally varied genres could further enhance global market adaptability.

4) Maintaining a mix of movies is recommended to ensure both short term and long-term engagement of users.

6. Interpretation of Results and Recommendation

6.1 Interpretation of Results

1. Content Distribution Dynamics

The Netflix catalog shows a dominance of movies over serialized content, but the serialized content has been growing from 2015 showing the shift of Netflix towards TV shows and web series which will result in a more engaging experience for the user leaning towards the platform more as compared to its competitors.

2. Regional and Genre Insights

Content distribution shows a **clear domination of USA and Asian countries** but the Netflix's catalog still contains a wide variety of regionally diversified content. This

imbalance suggests potential for regional market growth. Genre analysis further indicated that International TV Dramas, Documentaries, and Comedies dominate the content mix, reflecting Netflix's effort to appeal to diverse audiences. However, there was less content in localized genres suggesting a growth opportunity for Netflix there.

3. Maturity and Age Rating Trends

The maturity rating trends showed a clear focus of Netflix towards more mature content in the recent years. The number of kids shows and family friendly shows are very less. The mean release year of 2014 and median of 2017 confirm Netflix's heavy inclusion of recent, original content. This trend emphasizes a deliberate focus on mature themes to attract a more global and diverse adult audience.

4. Country-wise Maturity Variation

A cross-country comparison showed that the **U.S., India and U.K. account for the majority of the mature-rated titles**, while countries like **Japan and South Korea focused more on PG-13 and TV-14 content.** This demonstrates how cultural and regional certification norms influence Netflix's content rating strategy and suggest the need for region-specific rating standards and age-targeted recommendations.

6.2 Recommendations for Strategic Growth

1. Content Diversification and Regional Expansion

Netflix should actively invest in content creation from underrepresented regions to broaden its global appeal. Collaborating with local production houses and creators cab help in generating culturally rich content. **Expanding regional language content will attract new customers of that region** thus diversifying customer base and enhancing Netflix's reach.

2. Balanced Content and Audience Segmentation

While mature rated title have grown rapidly attracting a lot of new viewers, the lack of kids and family friendly content limits the viewer count. **Acquisition and making original content in the family friendly rating can help in inclusion of household audience**. Maintaining an optimal balance between Movies and TV Shows will also ensure both short-term viewer engagement and long-term retention.

3. Data-Driven Personalization and Recommendation Systems

Leveraging insights from maturity ratings, country trends, and genre preferences, Netflix can improve its recommendation system to suggest more region-specific content increasing the viewer engagement and user satisfaction, while ensuring culturally appropriate recommendations across demographics.

4. Strategic Production Planning

Netflix can use the insights from data analysis to plan its future content production more effectively. By focusing on the most popular genres, regions, and maturity ratings, the company can invest its budget in areas that are more likely to attract viewers. Creating a wider range of content will help Netflix reach new audiences and stay relevant in different markets. This will also strengthen its brand value and long-term growth.

7. Conclusion

This analysis of Netflix's catalog has provided valuable insights into the platform's content strategy, regional representation, and maturity trends. The analysis shows how Netflix is transitioning from being a movie-dominated platform to one that includes more serialized content over the years.

The dominance of content from USA and India, and the underrepresentation of several regions, shows that there is opportunity to expand in localized content production supporting Netflix's global expansion.

The recommendation proposed – increasing regional diversification, balanced content development, and personalized recommendations. By integrating these data-driven insights Netflix can strengthen its global reach and improve customer engagement while maintaining its place in the market as a leading content streaming platform.