Evolving Landscapes of Television: An Analytical Dive into TV Show Trends on Netflix

Project Outline

We decided to tackle several key questions:

Genre Production Patterns

- What can be inferred about Netflix's content strategy based on the genres and types of shows produced?
- How has the diversity of production regions and languages changed over time on Netflix?
- How does Netflix's content portfolio compare with its initial years versus the more recent years?

IMDb Score, Votes, and Genre Relationship

- Are certain genres consistently receiving higher ratings and votes?
- Are older shows maintaining popularity, or is there a trend towards newer releases?
- What factors contribute to higher viewer engagement and popularity for TV shows on Netflix?

Maturity Rating Evolution

- How do different age certifications (like TV-MA, TV-14) perform in terms of quantity and popularity?
- Does this reflect any broader market trends or shifts in audience demographics?

Pandemic and Reality TV Popularity

- Was there a significant rise in the popularity of reality TV shows during the pandemic, and has this increase been sustained post-pandemic?
- How do scripted TV shows perform in comparison to non-scripted ones in terms of IMDb scores, viewer engagement, and production rates?

Data Clean-up and Preparation

The primary goal is to prepare a clean, well-structured dataset focusing exclusively on TV shows from Netflix for detailed analysis. Adding language data through the OMDB API enriches the dataset, offering more dimensions for analysis, particularly concerning linguistic diversity. The cleaning and preparation steps ensure data quality and usability for subsequent analyses.

Process:

1. Import Dependencies: The notebook begins by importing necessary Python libraries (`requests`, `pandas`) and an API key from a configuration file.

2. Data Loading: It loads the dataset `imdb_movies_shows.csv` from Kaggle into a data frame named `netflix_data`.
3. Data Filtering: The dataset is filtered to include only entries where the type is 'SHOW', resulting in a new data frame `netflix_shows`.
4. Data Analysis and Clean-up: The notebook performs an analysis of the Data Frame's structure using `info()`, removes any rows with missing values (`dropna()`), and resets the index (`reset_index`).
5. Unique IMDB IDs: Ensures that all IMDB IDs in the DataFrame are unique.
6. API Data Integration: The notebook uses the OMDB API to add language information to each show in `netflix_shows` using the `imdb_id` field. It iterates through the data frame, makes API requests, and updates the language column.

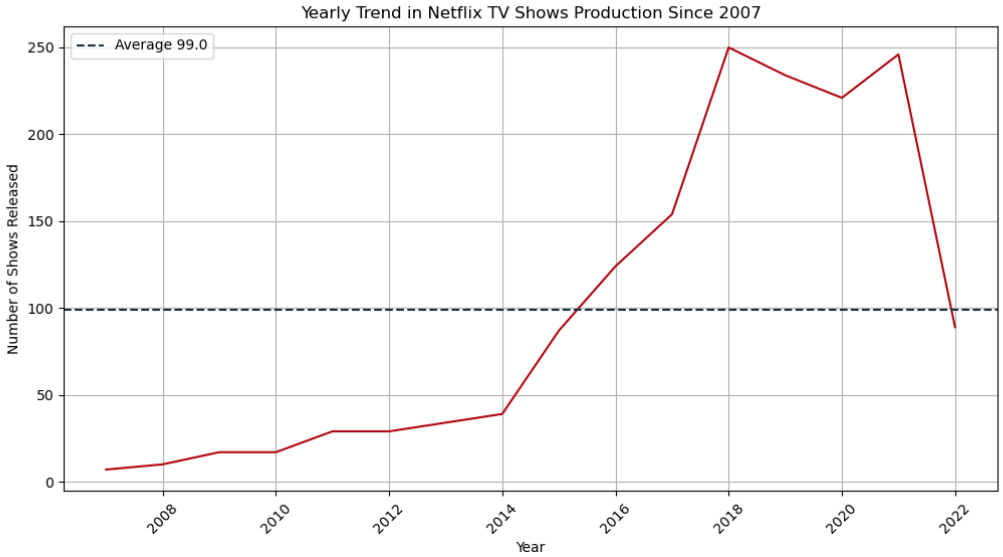
```
#Using OMDB to add a column for languages using the imdb_id

netflix_shows["language"] = ""

# Iterate through the netflix_shows DataFrame
for index, row in netflix_shows.iterrows():
    # Get imdb_id from the DataFrame
    imdb_id = row["imdb_id"]

    # Set base URL for OMDB API
    base_url = "http://www.omdbapi.com/"

    # Make an API request using the imdb_id
    response = requests.get(base_url, params={"i": imdb_id, "apikey": api_key})

    # Convert the API response to JSON format
    show_data = response.json()

    # Grab the language from the results and store it in the netflix_shows DataFrame as a list
    try:
        netflix_shows.loc[index, "language"] = show_data["Language"]
    except:
        # If title is not found, print the message and set as N/A
        print(f"{imdb_id} - not found")
        netflix_shows.loc[index, "language"] = "N/A"
```

7. Further Clean-up: Removes any additional N/A entries and resets the index again.
8. Data Frame Refinement: Creates a new Data Frame `shows_df` by selecting specific columns relevant for analysis.
9. Data Export: The cleaned and refined Data Frame is saved as a new CSV file `netflix_shows.csv` in the Resources directory. The dataset contains 1,670 data points, each representing a different show on Netflix.
The resultant `netflix_shows.csv` file is now a cleansed and enriched resource ready for deeper analysis in subsequent notebooks. The inclusion of language data and the focus on key attributes pave the way for multifaceted analysis, including genre trends, popularity metrics, and demographic shifts. The rigorous data clean-up process sets a strong foundation for accurate and meaningful insights in the later stages of the project.

Genre and Language Trends Analysis

The focus on data post-2007 allows the analysis to be relevant to Netflix's streaming era, providing insights into content strategy changes after the introduction of streaming. Visualizing trends with line graphs and heatmaps enables easy identification of patterns and

anomalies over time, such as shifts in genre production or language inclusion. The analysis of non-English languages provides insights into Netflix's global content strategy and diversification.

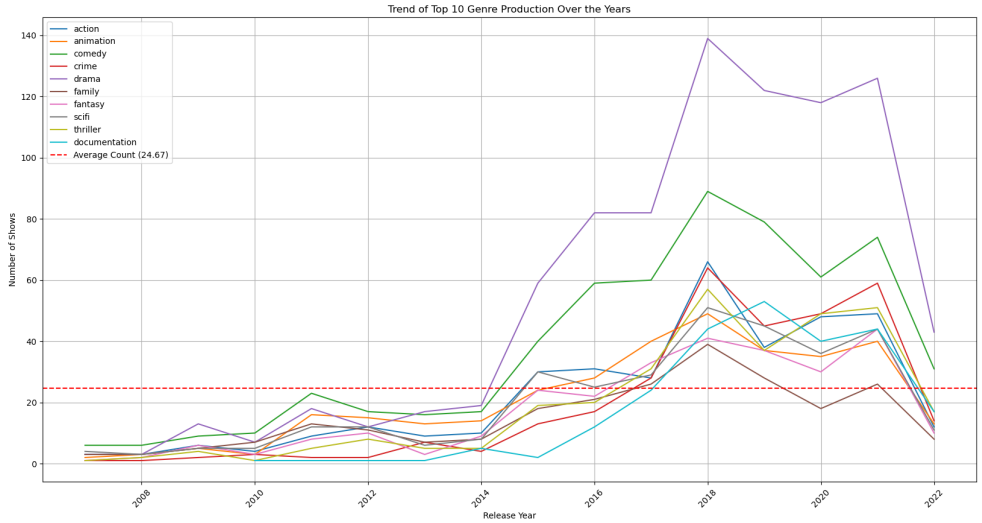


The graph demonstrates a significant investment in original content by Netflix over the years, particularly from 2015 leading up to the peak. This trend aligns with Netflix's strategy to expand its library of original content as a differentiator in the competitive streaming market. The sharp decrease in the most recent year could be indicative of various factors such as strategic realignment, budget constraints, changes in consumer preferences, or external factors such as the global COVID-19 pandemic.

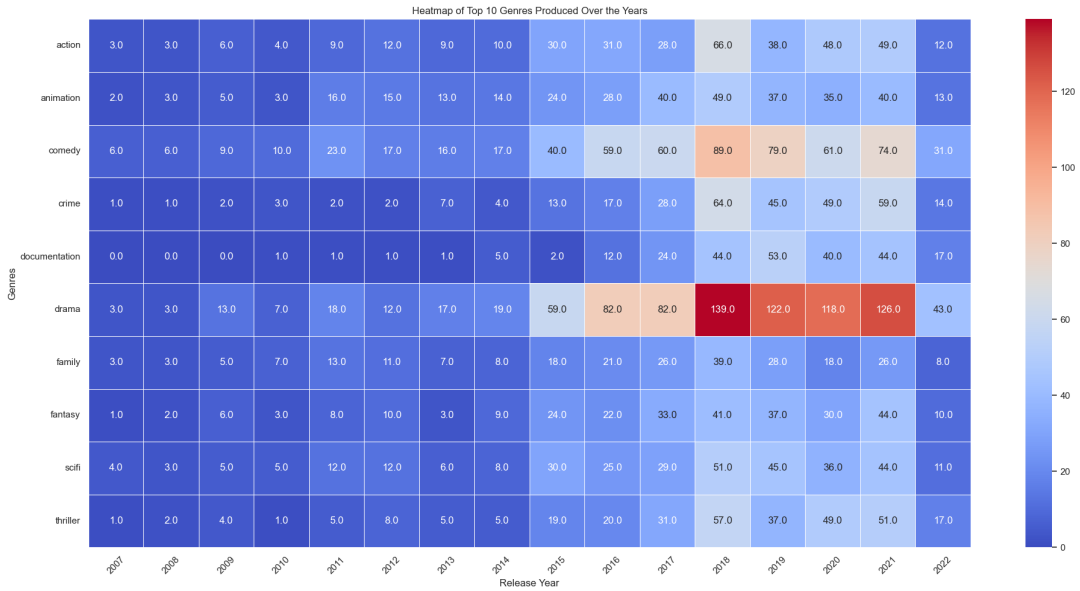
The placement of the average line suggests that while the number of shows has fluctuated, the overall trend has been towards an increase in production, with recent years far exceeding the average until the decline in the latest year. This visualization provides a clear narrative of aggressive growth followed by a recent pullback in content production, which may signal a new phase in Netflix's content strategy.

Content Strategy Based on Genres and Types of Shows:

Netflix has significantly diversified its genre offerings, with a focus on drama, comedy, and action. The increase in genre variety over the years suggests a strategy to cater to a wide range of viewer interests and demographics. The sustained growth in content production until a peak in 2018-2019, followed by a decline, might imply a phase of aggressive expansion followed by a strategic consolidation or response to external factors.



We identified growth patterns of each genre, with drama and comedy seeing substantial increases. There is a general upward trend for all top genres, indicating an increase in diversity of content.



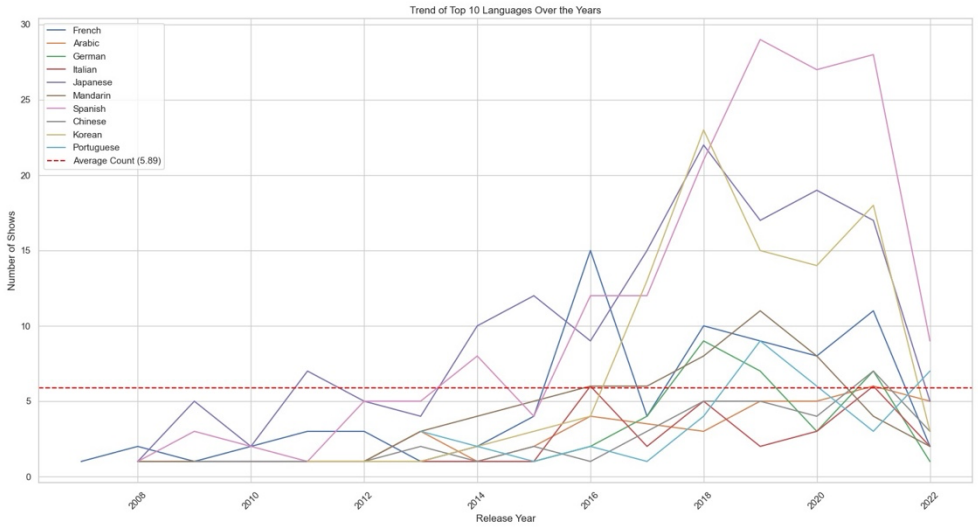
The heatmap visualizes the production counts of the top ten genres on Netflix from 2007 onwards. The heatmap provides a colour-coded representation of the number of shows produced in the top genres over time. Darker colours represent a higher production count. Drama, comedy, and action genres dominate the production, with the drama genre experiencing the highest production numbers, especially in recent years.

This heatmap is valuable for understanding how Netflix has varied its content production over time, focusing on certain genres that align with its strategic goals or viewer demand trends. It also illustrates the platform's response to the evolving entertainment landscape.

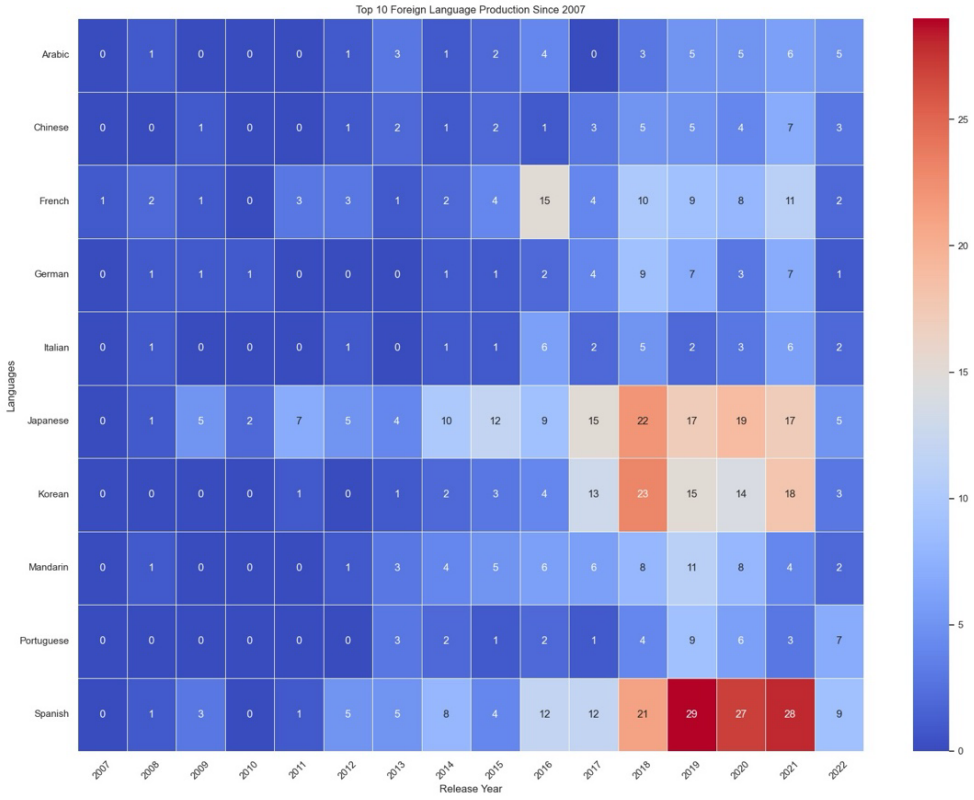
Diversity of Languages

The increase in productions across various languages indicates Netflix's strategic shift towards global markets and localization of content.

Peaks in productions of certain languages like Japanese and Korean reflect a successful penetration into these markets, capitalizing on the global popularity of K-dramas and anime.



A line chart focuses on the production trend of content in the top 10 foreign languages on Netflix, highlighting the growth in international content.



The heatmap shows the trend in the top foreign languages, with a noticeable increase in production in Japanese, Korean, and Spanish, indicating a strategic focus on these markets.

Content Portfolio Comparison: Initial Years vs. Recent Years

In the initial years, Netflix focused on a narrower range of genres and primarily English-language content. Over time, Netflix significantly increased its genre and language offerings, reflecting a pivot towards a more global audience.

The data shows a content portfolio that has evolved from a primarily English-speaking, less diverse base to a broad, multi-lingual, and multi-genre platform.

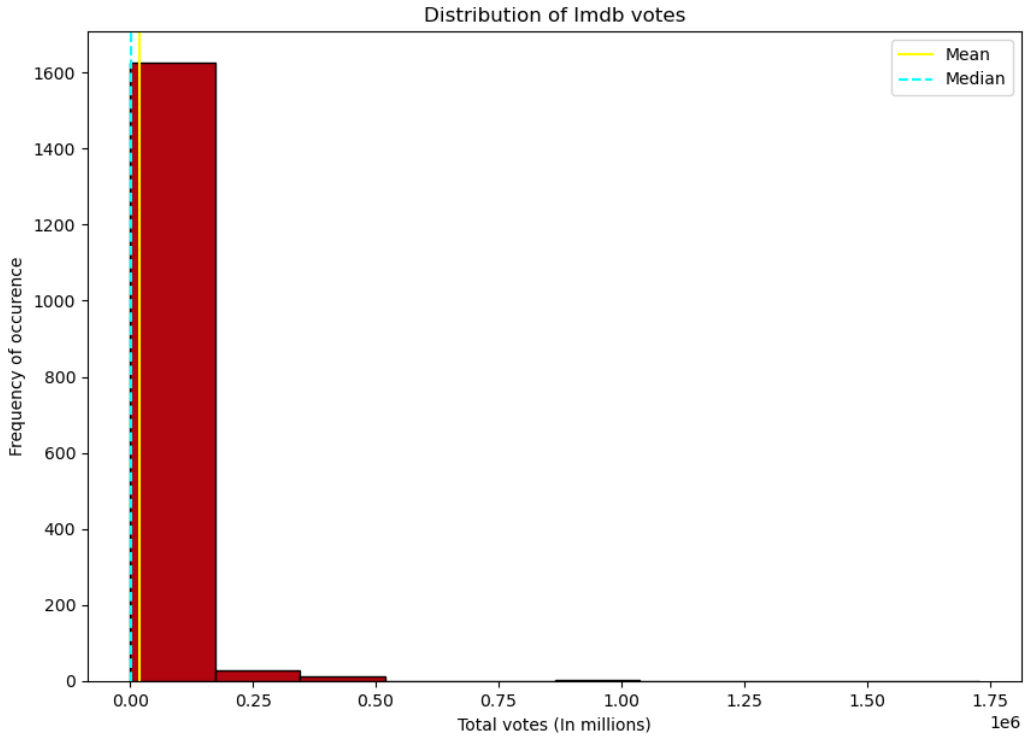
Overall, the data from these visualizations indicates that Netflix has been adapting its content strategy to become a more diverse and global entertainment provider, which aligns with the trends in viewer consumption and the increasing demand for varied content across different regions and languages. The recent dip in production suggests a re-evaluation of content strategy or a response to market saturation, increased competition, or other operational challenges.

IMDb Score, Votes, and Genre Relationship

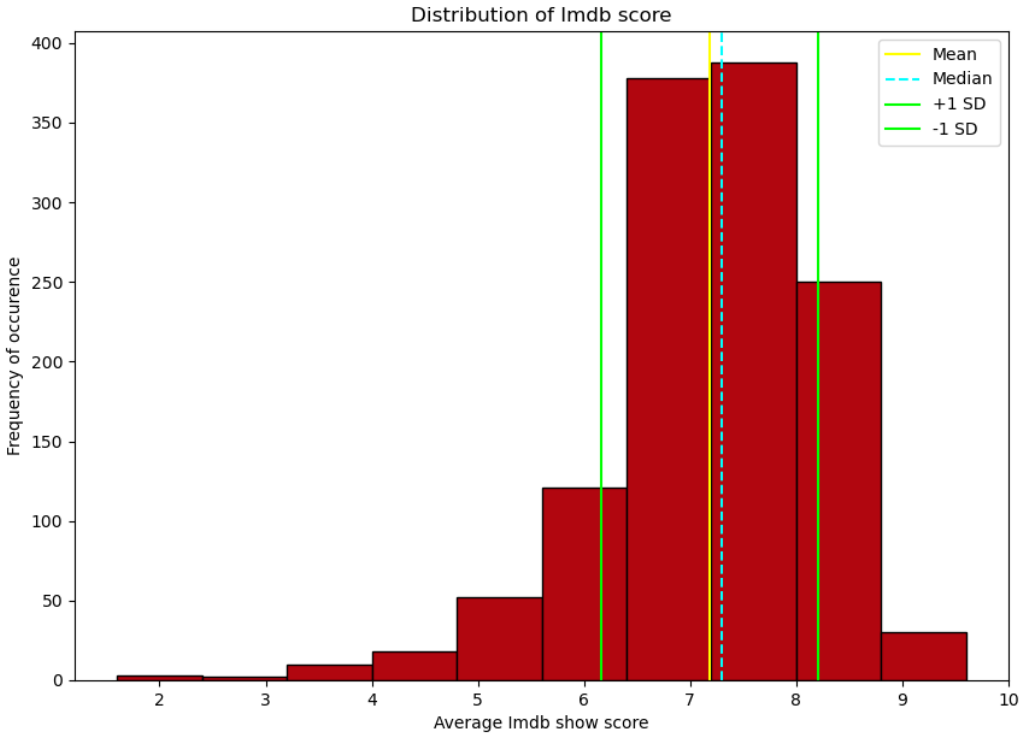
An initial exploration involved examining descriptive statistics of IMDb votes, which include measures like mean, median, mode, standard deviation, minimum, and maximum values. Histograms and box plots were used to visualize the distribution of IMDb votes and scores. These plots reveal that the data is not normally distributed, with votes being right-skewed and scores being left-skewed. Therefore, the dataset is filtered to remove entries with an IMDb vote count below a certain threshold to ensure a better representation of the overall population's opinion.

The reasoning behind the analysis is to understand viewer engagement and preferences as reflected by IMDb votes and scores. By focusing on the distribution of votes and scores, the analysis aims to gain insights into the popularity and quality of the shows as perceived by viewers. The extraction of the dominant genre is based on the assumption that the first listed genre is the most significant, which simplifies the analysis of genre-specific trends. Filtering out shows with low vote counts aims to reduce noise in the data and focus on shows that have garnered significant viewer attention.

The findings suggest that certain genres, particularly drama and sci-fi, garner more votes, indicating higher viewer engagement with these genres. The weak positive correlation between votes and scores suggests that while more popular shows tend to have higher scores, the relationship is not strong, and there are popular shows with lower scores and vice versa. The trend analysis indicates that the most recent years have seen a significant increase in both the number of shows produced and the total number of votes, pointing to a growing engagement with Netflix content.

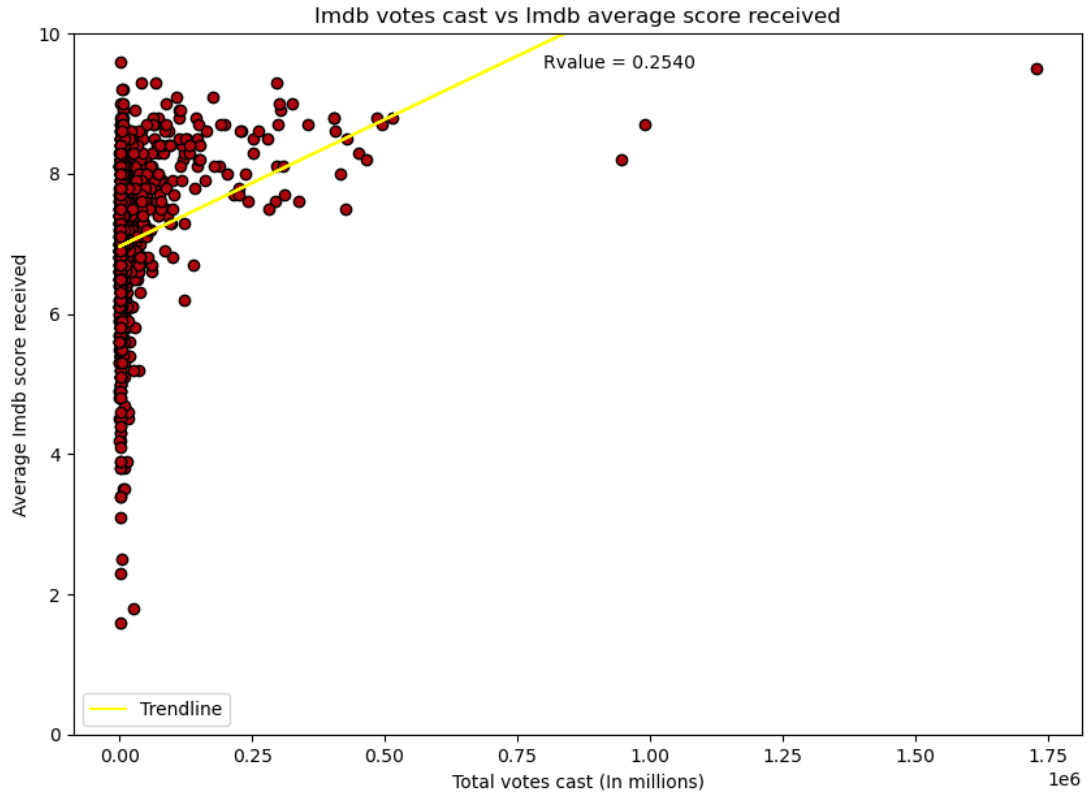


The histogram shows a highly right-skewed distribution with the majority of shows having a low number of votes, which is common for entertainment data where only a few items tend to be highly popular (a Pareto principle or "long-tail" phenomenon). The mean and median are marked, with the mean likely higher than the median due to high-vote outliers pulling the average up.

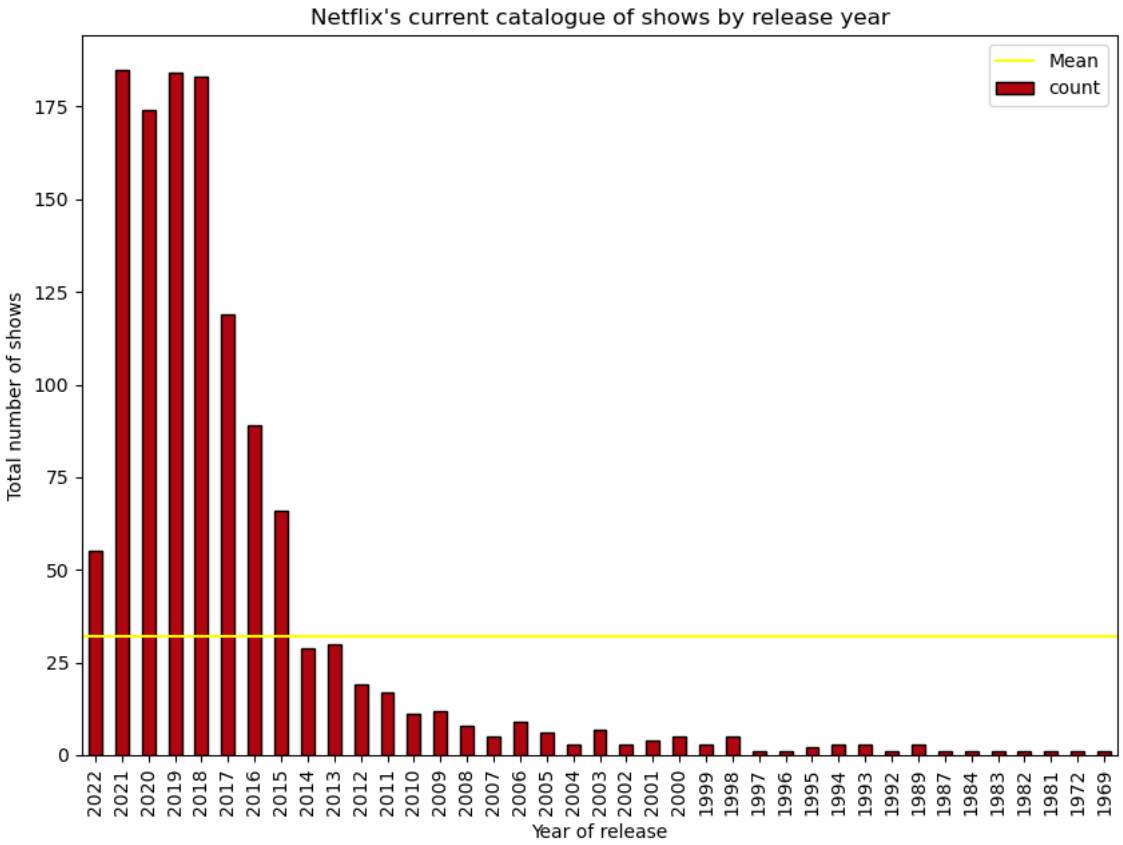


This histogram illustrates the distribution of IMDb scores for shows. The distribution appears to be roughly normal with a slight left skew, indicating that most shows have a moderate to

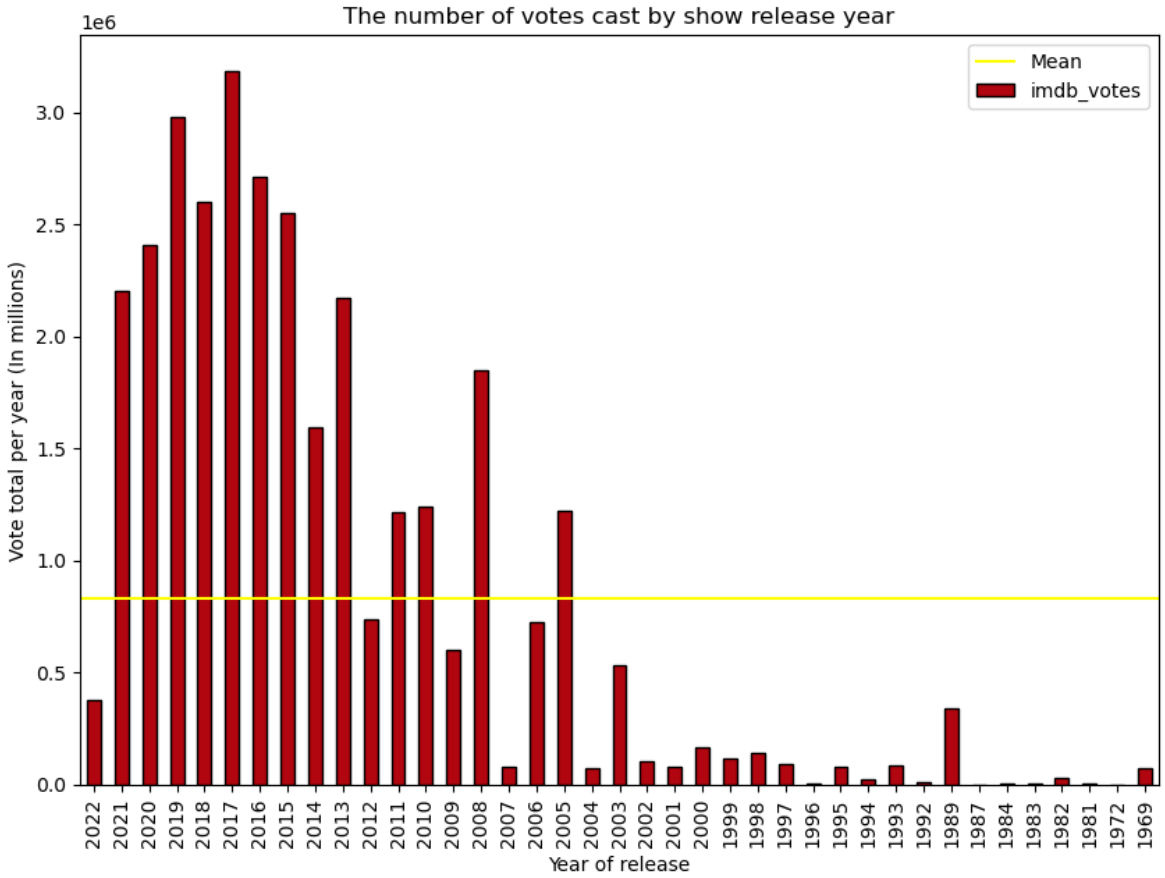
high score, with fewer shows having very low scores. The mean, median, and standard deviations are indicated, suggesting that the average score is quite consistent, but there is variability as indicated by the spread of the standard deviation.



The scatter plot presents the relationship between the total number of votes a show receives and its IMDb score. There is a positive correlation, as indicated by the trendline and the R-value, meaning shows with higher scores tend to have more votes, but the correlation is not strong (R-value = 0.2540). The distribution of points suggests that while high scores can correlate with a higher number of votes, there are many high-scoring shows with a smaller number of votes, indicating that quality (as judged by score) does not always translate to popularity (as judged by vote count).



The bar graph shows the number of shows released per year. There has been a clear increase in production in recent years, with a noticeable peak in the number of shows around 2018-2019, followed by a sharp decline. The mean line across the bars indicates that recent years have seen production numbers well above the historical average, suggesting a period of significant content expansion.



This bar graph indicates the total number of IMDb votes received by all shows released in each year. Similar to the production numbers, there's a peak in votes around the same period, aligning with the increased number of shows produced. The presence of the mean line again shows the average number of votes has increased in recent years, although there is significant year-to-year variability.

The skewed distributions of votes and scores imply that while a few shows are extremely popular, most receive moderate attention. This suggests that Netflix has both standout hits and a broad selection of content with varying popularity.

Content Engagement and Quality

There are a large number of shows with a few votes, but a small number of shows have garnered a significant number of votes, suggesting that a few hits drive a significant portion of engagement. The distribution of scores indicates that most Netflix shows are rated above average, with scores clustering around the mean and median. This suggests a generally favourable reception of Netflix content in terms of quality.

The positive but weak correlation between votes and scores suggests that while higher quality (as indicated by scores) does attract more votes, it is not the sole factor determining a show's popularity.

Content Production and Reception Over Time

There has been a substantial increase in the number of shows produced in recent years, particularly around 2018-2019, which also corresponds to a peak in viewer engagement as evidenced by the number of votes. The decline in the number of shows in the most recent year(s) shown in the data suggests a possible strategic shift or reaction to external factors such as market saturation or changes in consumer behaviour. The number of votes per year following a similar trend to the number of shows suggests that the increase in content production did successfully translate to higher overall engagement, even if not uniformly across all shows.

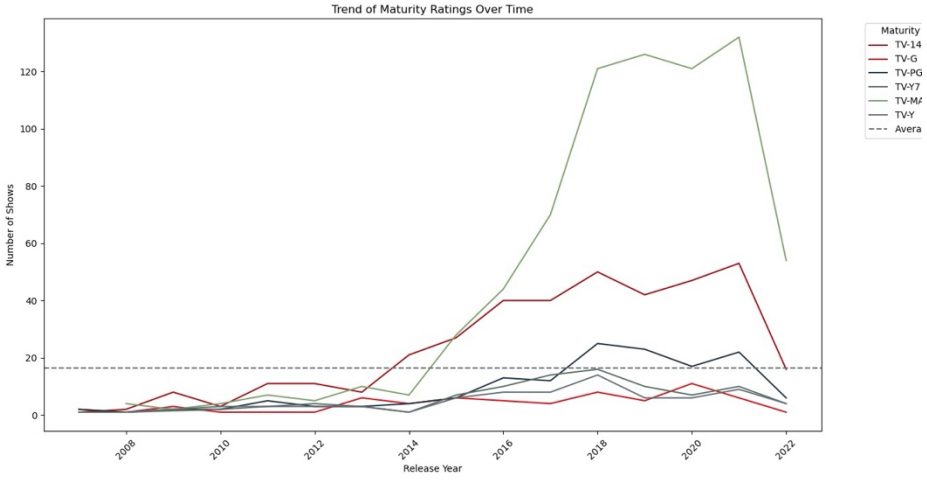
These insights could inform content strategy, indicating the importance of maintaining a portfolio of high-quality content that appeals to diverse viewer interests, while also recognizing that only some productions will become major hits. The recent decrease in both production and votes might warrant an investigation into changing market dynamics or internal strategy shifts.

In conclusion, while the data provided offers valuable insights into overall trends, answering these questions in depth would require more granular data, such as IMDb scores and vote counts by genre, as well as time-series data on votes to assess the enduring popularity of individual shows.

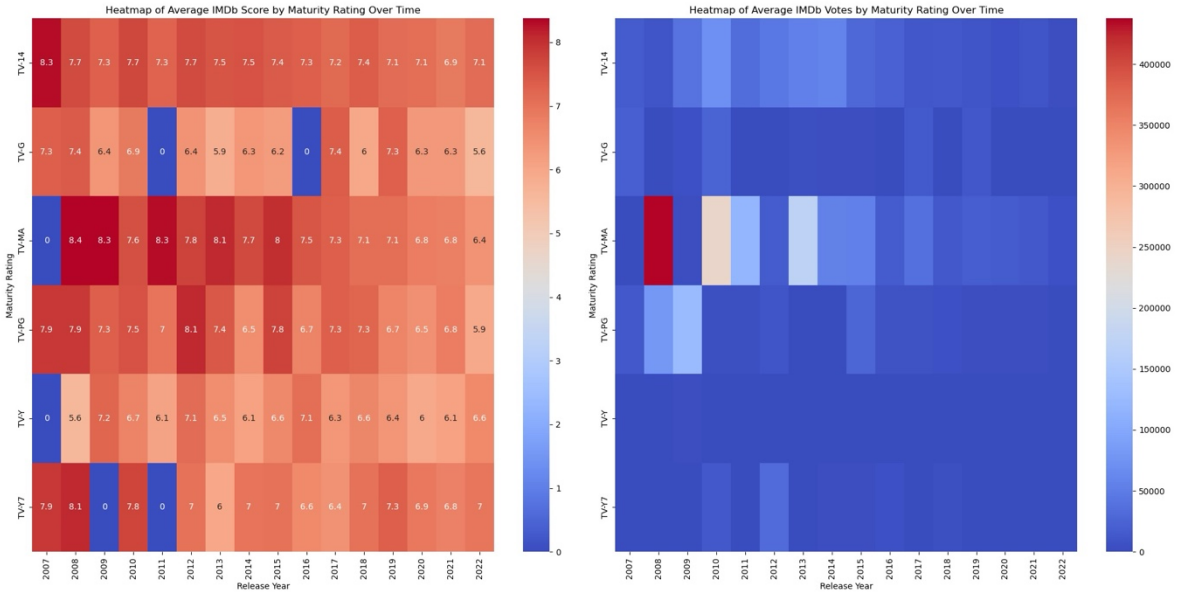
Maturity Rating Evolution

The analysis of maturity ratings over time helps to understand if Netflix is diversifying or specializing in certain audience segments. By correlating maturity ratings with IMDb scores and votes, the analysis seeks to uncover if certain types of content are more critically acclaimed or popular among viewers. The heatmap of correlations provides a quick and clear visual summary of the relationships between different variables.

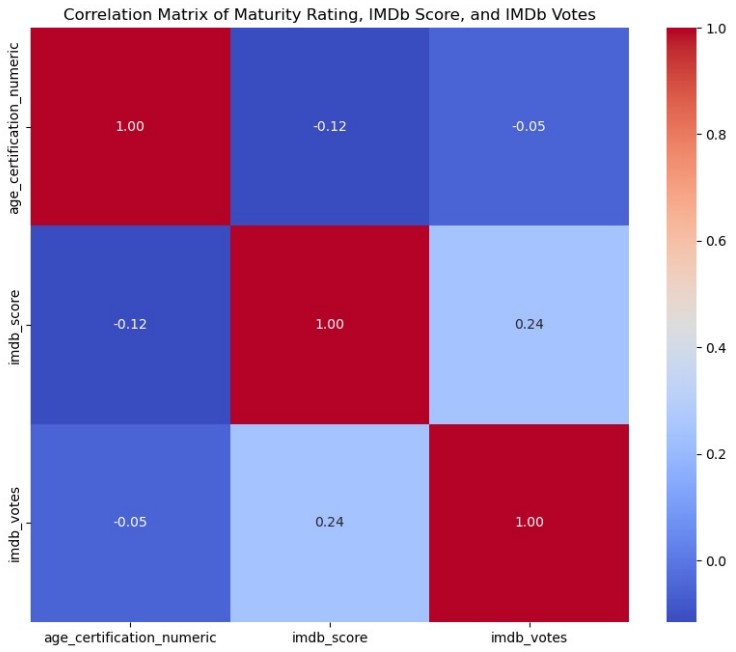
An examination into how maturity ratings have trended over time was conducted by initially counting the number of shows per rating each year. A line plot is used to visualize these trends against the overall average.



TV-MA (Mature Audience) shows have seen a significant increase, peaking around 2019, which suggests a strong trend toward content aimed at adult audiences. TV-14 (suitable for viewers over 14) also shows an increasing trend but less pronounced than TV-MA.
The average line indicates that both TV-MA and TV-14 productions are above the average number of shows produced, which shows a clear inclination towards these categories.



The heatmaps provide a visual representation of the average IMDb scores and votes across different maturity ratings over time. The heatmap for scores shows moderate to high scores across all maturity ratings without a clear pattern favouring one rating over another. The heatmap for votes shows more pronounced hotspots, especially for TV-MA content in recent years, indicating higher engagement.



The matrix displays the correlation between age certification numeric values, IMDb scores, and votes. There is a very weak negative correlation between maturity rating and IMDb score, and a weak positive correlation between IMDb scores and votes. This suggests that higher-rated content does not necessarily score better, but shows that higher scores do tend to receive more votes.

Performance of Age Certifications
TV-MA content is increasingly dominant in terms of quantity, which implies that Netflix is investing more in adult-oriented content. This content is not only more abundant but also highly engaged, as indicated by the number of IMDb votes.
TV-14 shows are also prevalent, indicating that content suitable for early teenagers and older is a significant portion of Netflix's strategy, likely reflecting its appeal to a wide audience demographic that includes young adults.

Reflection of Market Trends and Audience Demographics
The surge in TV-MA content corresponds to broader market trends where streaming platforms are creating more bold and mature content, possibly to differentiate themselves from traditional broadcasters with more restrictions on content maturity. The high engagement with TV-MA shows reflects a possible shift in audience demographics, with platforms like Netflix increasingly targeting adult viewers who seek complex narratives and mature themes. The weaker correlation between maturity ratings and scores suggests that while maturity ratings influence production decisions, they are not a determinant of perceived quality.

Overall, the analysis indicates that Netflix is producing more mature content, which aligns with a shift in strategy to target adult audiences, reflecting broader market trends of catering to a demand for diverse and sophisticated storytelling. However, the quality of content, as judged by viewers, seems to be distributed evenly across maturity ratings, suggesting that Netflix is maintaining a balance between quantity and quality across its content spectrum.

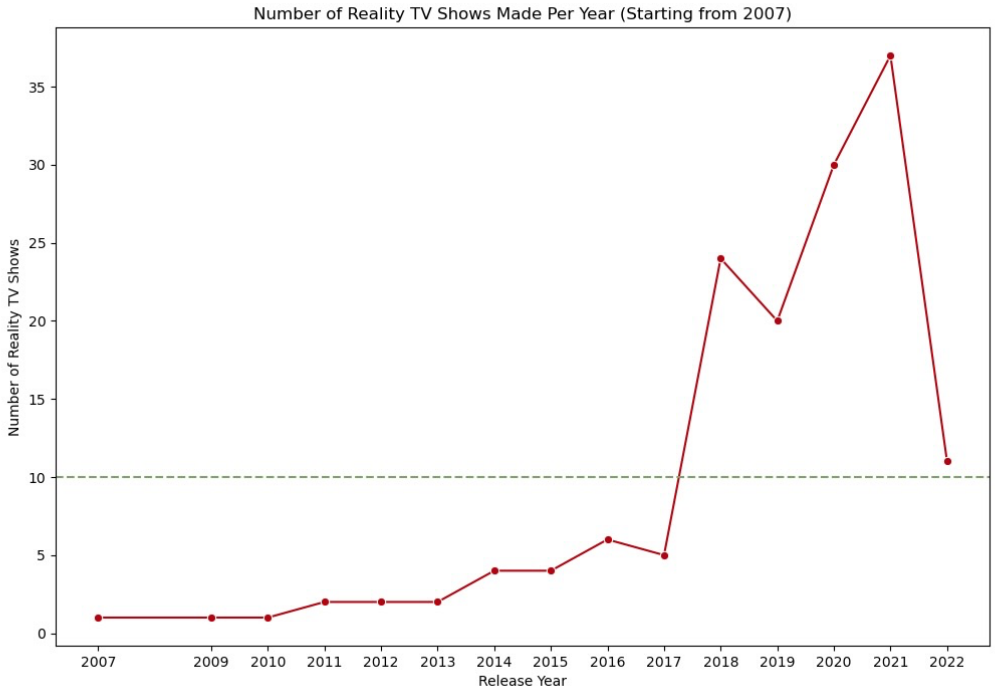
Pandemic and Reality TV Popularity

The analysis of reality TV shows can shed light on the popularity and success of different content types on Netflix. Understanding the trends in reality TV show production and viewer engagement can inform Netflix's content strategy and investment decisions. By comparing reality TV shows with other genres in terms of IMDb scores and votes, the analysis seeks to quantify the relative appeal and success of reality content.

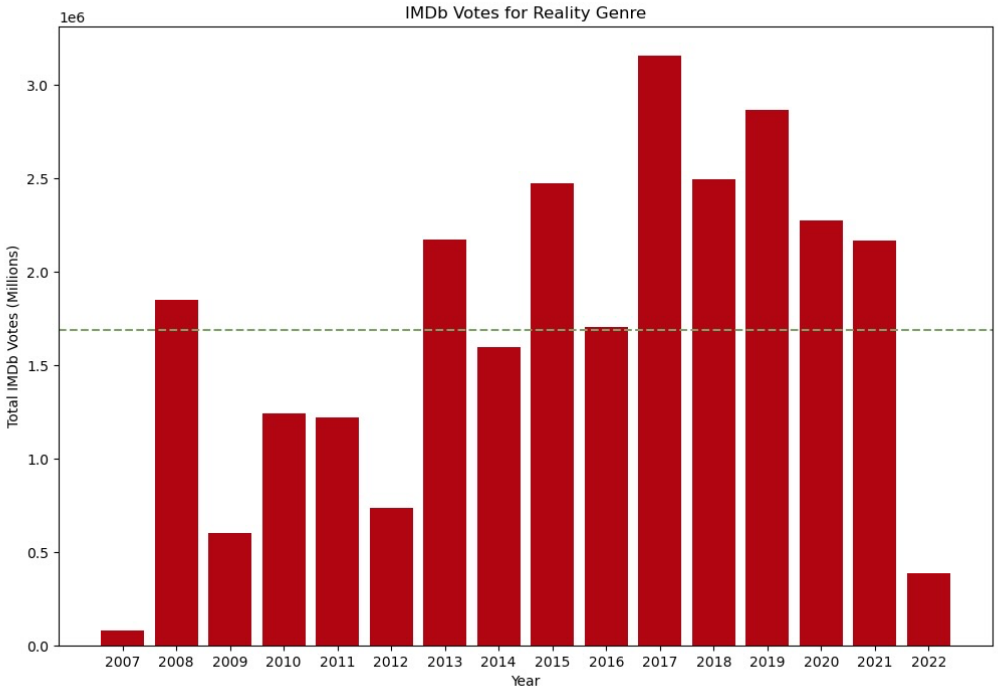
The production trends of reality TV shows over time may reflect changing audience preferences, possibly influenced by cultural shifts or significant global events like the pandemic. A significant decline in the production or viewer engagement with reality shows could suggest a shift in strategy or audience fatigue with the genre.

The correlation between reality TV show ratings and their popularity could indicate how well reality content resonates with the audience compared to other genres. Statistical

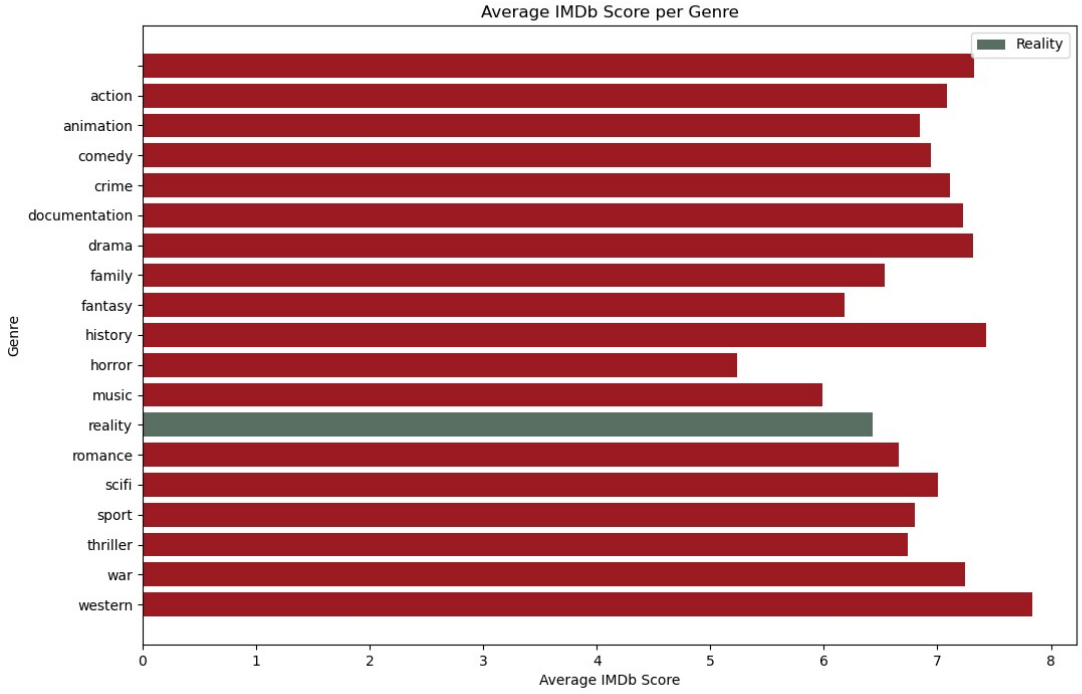
significance in the difference of IMDb scores between reality shows and other genres points to a niche appeal or a divergence in quality perception among viewers.



This line graph indicates the number of reality TV shows produced each year from 2007 onwards. There is a notable peak in 2020, which could be associated with the onset of the COVID-19 pandemic. The trend shows an increase from previous years, suggesting a rise in the production of reality TV shows during the pandemic. However, in 2022, there is a sharp decline, indicating that the increase was not sustained post-pandemic.



This bar chart presents the total IMDb votes for reality TV shows per year. There is a peak in 2020, supporting the previous chart's implication of the increased popularity of reality shows during the pandemic. The votes are significantly higher compared to other years, including the preceding and following years. This suggests that while production dropped in 2022, viewer engagement in terms of votes remained relatively high.



The horizontal bar chart compares the average IMDb scores across various genres, with a particular emphasis on reality TV, which has one of the lowest scores. This implies that while reality TV may have a solid viewer base, it does not necessarily translate into high critical acclaim or high scores on IMDb.

Popularity During Pandemic
There was a significant rise in the production of reality TV shows during the pandemic, as seen in 2020. This increase coincided with a peak in viewer engagement (IMDb votes). However, the production rates were not sustained post-pandemic, indicating a possible return to pre-pandemic levels in 2022.

Performance Comparison
Scripted TV shows like sci-fi, action, and drama tend to receive higher viewer engagement (IMDb votes) and potentially higher production rates compared to non-scripted ones like reality TV. However, the IMDb score for reality TV is notably lower than for most scripted genres, suggesting that while reality TV may engage viewers, it does not perform as well critically.

These observations suggest that reality TV experienced a surge in popularity during the pandemic, likely due to increased viewership while people were at home. However, the production of new reality shows has not sustained the pandemic peak, and they generally do not fare as well as scripted genres in terms of critical reception.

Conclusions

Diversification and Growth
Netflix has significantly diversified its content offering, with a particular increase in drama, comedy, and action genres. This reflects a content strategy designed to cater to a broad range of viewer interests and demographics.

Global Expansion
There is a clear trend toward producing content in multiple languages, with notable increases in Japanese, Korean, and Spanish productions. This indicates a strategic move to capitalize on global market trends and the popularity of certain international genres, such as K-dramas and anime.

Mature Content
There has been a growth in TV-MA-rated content, suggesting a focus on adult-oriented programming. This aligns with a broader trend in the streaming industry to produce more sophisticated and mature narratives.

Reality TV During Pandemic
Reality TV saw a surge in production and viewer engagement during the pandemic, which did not sustain post-pandemic, suggesting that this genre's growth was likely an anomaly due to the unique circumstances of the global pandemic.

Viewer Engagement vs. Critical Acclaim

While reality TV shows have had solid viewer engagement, they do not necessarily receive high critical acclaim, indicating a dichotomy between popularity and perceived quality.

Content Quality and Popularity There is a weak positive correlation between the number of IMDb votes and scores, suggesting that while higher-rated shows tend to be more popular, there are also many high-quality shows that do not receive as many votes.

Future Implications for Exploratory Data Analysis

Content Strategy Adjustments
Netflix may need to continue adjusting its content strategy to balance the creation of broadly appealing content with the production of niche shows that cater to specific audience segments.

Investment in Quality
The data suggests that investing in high-quality content can lead to better viewer engagement. Netflix might need to consider how to leverage this relationship further, perhaps by directing more resources toward shows with higher engagement potential.

Post-Pandemic Content Planning
Given the drop in reality TV production post-pandemic, Netflix could explore whether other genres might fill the void or whether there are new emergent genres that could capture audience attention in the longer term.

Understanding Audience Behaviour
The increase in international content and mature ratings points to changing viewer preferences. Netflix might benefit from deeper audience behaviour analysis to understand the evolving tastes and how they correspond to global trends.

Long-Tail Content Value
Considering the right-skewed distribution of votes, Netflix could evaluate the long-tail value of content that does not initially receive high engagement but maintains steady viewership over time.

Sustainability of High Production Rates
The recent decline in the number of shows produced may reflect market saturation, increased competition, or strategic realignment. Understanding this trend will be crucial for planning future production volumes.

In summary, the exploratory data analysis points to a dynamic and evolving content landscape for Netflix. The insights gained from this data can guide strategic decisions to enhance viewer engagement, expand global reach, and adjust content offerings to match audience preferences and emerging market trends.