



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΜΕ ΕΦΑΡΜΟΓΕΣ
ΣΤΗ ΒΙΟΙΑΤΡΙΚΗ**

**COLLECTION AND ANALYSIS OF DATA
RELATING TO TELEVISION SHOWS**

AGATHANGELOS PAPADOPOULOS

BSc THESIS

**SUPERVISOR
GERASIMOS RAZIS**

Lamia, 2025

Με ατομική μου ευθύνη και γνωρίζοντας τις κυρώσεις ⁽¹⁾, που προβλέπονται από της διατάξεις της παρ. 6 του άρθρου 22 του Ν. 1599/1986, δηλώνω ότι:

1. Δεν παραθέτω κομμάτια βιβλίων ή άρθρων ή εργασιών άλλων αυτολεξεί **χωρίς να τα περικλείω σε εισαγωγικά** και χωρίς να αναφέρω το συγγραφέα, τη χρονολογία, τη σελίδα. Η αυτολεξεί παράθεση χωρίς εισαγωγικά χωρίς αναφορά στην πηγή, είναι λογοκλοπή. Πέραν της αυτολεξεί παράθεσης, λογοκλοπή θεωρείται και η παράφραση εδαφίων από έργα άλλων, συμπεριλαμβανομένων και έργων συμφοιτητών μου, καθώς και η παράθεση στοιχείων που άλλοι συνέλεξαν ή επεξεργάστηκαν, χωρίς αναφορά στην πηγή. Αναφέρω πάντοτε με πληρότητα την πηγή κάτω από τον πίνακα ή σχέδιο, όπως στα παραθέματα.

2. Δέχομαι ότι η αυτολεξεί **παράθεση χωρίς εισαγωγικά**, ακόμα κι αν συνοδεύεται από αναφορά στην πηγή σε κάποιο άλλο σημείο του κειμένου ή στο τέλος του, είναι αντιγραφή. Η αναφορά στην πηγή στο τέλος π.χ. μιας παραγράφου ή μιας σελίδας, δεν δικαιολογεί συρραφή εδαφίων έργου άλλου συγγραφέα, έστω και παραφρασμένων, και παρουσίασή τους ως δική μου εργασία.

3. Δέχομαι ότι υπάρχει επίσης περιορισμός στο μέγεθος και στη συχνότητα των παραθεμάτων που μπορώ να εντάξω στην εργασία μου εντός εισαγωγικών. Κάθε μεγάλο παράθεμα (π.χ. σε πίνακα ή πλαίσιο, κλπ), προϋποθέτει ειδικές ρυθμίσεις, και όταν δημοσιεύεται προϋποθέτει την άδεια του συγγραφέα ή του εκδότη. Το ίδιο και οι πίνακες και τα σχέδια

4. Δέχομαι όλες τις συνέπειες σε περίπτωση λογοκλοπής ή αντιγραφής.

Ημερομηνία: 24 / 02 / 2025

Ο – Η Δηλ.

(Παπαδόπουλος Αγαθάγγελος)

(1) «Όποιος εν γνώσει του δηλώνει ψευδή γεγονότα ή αρνείται ή αποκρύπτει τα αληθινά με έγγραφη υπεύθυνη δήλωση του άρθρου 8 παρ. 4 Ν. 1599/1986 τιμωρείται με φυλάκιση τουλάχιστον τριών μηνών. Εάν ο υπαίτιος αυτών των πράξεων σκόπευε να προσπορίσει στον εαυτόν του ή σε άλλον περιουσιακό όφελος βλάπτοντας τρίτον ή σκόπευε να βλάψει άλλον, τιμωρείται με κάθειρξη μέχρι 10 ετών.



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΜΕ ΕΦΑΡΜΟΓΕΣ
ΣΤΗ ΒΙΟΙΑΤΡΙΚΗ

ΣΥΛΛΟΓΗ ΚΑΙ ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ ΣΧΕΤΙΖΟΜΕΝΩΝ
ΜΕ ΤΗΛΕΟΠΤΙΚΕΣ ΣΕΙΡΕΣ

ΑΓΑΘΑΓΓΕΛΟΣ ΠΑΠΑΔΟΠΟΥΛΟΣ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΕΠΙΒΛΕΠΩΝ
ΓΕΡΑΣΙΜΟΣ ΡΑΖΗΣ
ΜΕΤΑΔΙΔΑΚΤΟΡΙΚΟΣ ΕΡΕΥΝΗΤΗΣ

Τριμελής Επιτροπή:

Ραζής Γεράσιμος, Μεταδιδακτορικός Ερευνητής (επιβλέπων)

Αναγνωστόπουλος Ιωάννης, Καθηγητής

Κακαρούντας Αθανάσιος, Αναπληρωτής Καθηγητής

ΛΑΜΙΑ, 2025

Abstract

TV shows have been in our lives for many years, providing a wide range of entertainment - from laughter and life lessons to heart-pounding excitement. They have also evolved from being a weekly event to a non-stop binge-watch that keeps people at home for consecutive days. Given the changes that have undoubtedly been made on the TV scene, the focus of this thesis was given on the analysis of TV shows' evolution to better understand these changes throughout the years, why they were realized and to take a closer look at shifts in the recent years, especially after the rise of the streaming services (Netflix, Amazon Prime, Disney+, etc.) worldwide. This analysis utilized metadata that were extracted from IMDb and stored in a local database for this purpose. Relying on the visualization of the extracted data, this work highlights the trends in audience preferences, the increasing diversity in production, and changes in the way TV series are structured. The findings show a growing preference for binge-watching and shorter storytelling formats, with US-produced shows continuing to dominate average ratings, though non-US productions are steadily gaining recognition for their quality. These findings will make the shifts in this field clearer and show that the future of TV is in the hands of the audience, which plays a crucial role in shaping the future of the industry.

Table of Contents

Abstract	4
1) Introduction	7
1.1) The Evolution of Television and Media Consumption	7
1.2) Audience Influence in the Evolution of Television	7
1.3) Television Production in the Age of Streaming and Big Data	8
1.4) Research Objectives and Contribution	9
2) Related Work and methods.....	10
3) Bayesian Average Ranking	12
3.1) Empirical Analysis	13
4) Collection and Storage of Data.....	15
4.1) Limitations of the Data	17
5) Metrics and Results	17
5.1) Quantity and quality	17
5.2) Diversity	19
5.3) Structure Changes	22
5.4) Audience Engagement.....	26
5.5) Genre Preferences.....	26
6) Conclusions and Research Contribution.....	31
Bibliography	33

Table of Figures

Figure 1 – Top 20 TV Shows (Highest to lowest IMDb Ratings)	13
Figure 2 – Top 20 in the Custom Ranking List (Highest to lowest Bayesian Average Rating – 25 th percentile).....	13
Figure 3 - Top 20 in the Custom Ranking List (Highest to lowest Bayesian Average Rating – 75 th percentile).....	14
Figure 4 - Top 20 in the Custom Ranking List (Highest to lowest Bayesian Average Rating – 98 th percentile – Final Form).....	14
Figure 5 - MySQL Database Schema for the Table of Episodes.....	16
Figure 6 – MySQL Database Schema for the Table of TV Shows.....	16
Figure 7 – Quantity of TV Shows Over Time: Pre- Streaming vs. Streaming Era.....	18
Figure 8 - Quality Trends Across Time (Including TV shows that haven't ended yet every year, up until their end year - Using IMDb Ratings as a metric).....	18
Figure 9 – Quality Trends Across Time (Using Bayesian Average as a metric).....	19
Figure 10 – Diversity in Content over time (Countries and Languages).....	20
Figure 11 – Comparison of Total TV Shows vs. Non-USA TV Shows Per Year (Number of new releases per year).....	20
Figure 12 – Average IMDb Ratings of Non-US TV Shows (1990-2023).....	21
Figure 13 – Average IMDb Ratings of Non-US TV shows (2006-2023).....	21
Figure 14 – Average IMDb Ratings of US, Non-US, and All shows (1990-2023).....	22
Figure 15 – Average Number of Seasons Per Year	23
Figure 16 – Average Number of Episodes Per Season Every Year.....	23
Figure 17 – Average TV Show Rating Per Show's Total Number of Seasons	24
Figure 18 – Total Number of Shows with Same-Day Release of First Season Episodes Per Year	25
Figure 19 – Average IMDb Ratings by Release Type (Same Day vs. Different Dates) ..	25
Figure 20 – Audience engagement over time (showing total IMDb Rating votes per year)	26
Figure 21 - Top 5 Genres based on number of appearances (split into four groups according to the corresponding decade)	27
Figure 22 - Top 5 Genres of Non-US TV Shows based on number of appearances (split into four groups according to the corresponding decade).....	27
Figure 23 - Top 5 Genres based on Average IMDb Rating (split into four groups according to the corresponding decade)	28
Figure 24 – Top 5 Genres based on Bayesian Average (split into four groups according to the corresponding decade)	29
Figure 25 - Top 20 Genre Combinations (Including Solo Genres)	30
Figure 26 - Top 20 Genre Combinations (Excluding Solo Genres)	30

1) Introduction

1.1) The Evolution of Television and Media Consumption

For many years, TV Shows have been an integral part of people's lives and their daily routine. During these years, there have been different mediums for new TV series, starting with Cable television, which first appeared in the USA in 1948 [1]. For many decades it was the only platform on which people could watch a series on its first appearance. That was until the creation of streaming platforms which effectively managed to dethrone Cable TV [2]. Television has become an influencer and still holds power in shaping societal norms, leisure activities, and interpersonal relationships. Research [3] has shown that the implications of owning a TV are significant in the way people allocate their time. For instance, television set owners tend to spend less time on activities like household care, social visits to friends and relatives and, more importantly, sleep. So, the time that was previously spent on other activities was substituted for time in front of the screen. This phenomenon also shows the impact of television—while some activities are displaced, there may also be an increase in family engagement within the household. We should also keep in mind that this research was published in 2002 where TV was the only device available for someone to watch their favorite shows. Now, with the way technology has advanced there are countless options and devices to choose from ranging from computers to smartphones, tablets and even cars [4], meaning that media consumption is no longer confined to a single screen or device. With the increase in on-demand content, personalized recommendations and the flexibility people have that gives them the option to watch TV shows anytime and anywhere, the influence of the media on daily life has become even stronger and heavily integrated into our routines. This highlights how much more significant the role of screen time has become in shaping modern lifestyles.

1.2) Audience Influence in the Evolution of Television

From the beginning of television content, opinions and audience preferences have played a big role in shaping the industry. Broadcasters and networks have always tried to understand what viewers want to watch, when they are watching, and why. This information has been essential for adjusting programming, attracting advertisers and justifying financing models like subscription fees or licenses. Audience data has traditionally been collected through methods like surveys, interviews and focus groups, to discover how television fits into people's cultural lives and meets their needs for information, entertainment and identity.

Looking at history [3], broadcasters relied on metrics such as ratings and audience shares to measure the success of their programs. These insights were used to aid programming decisions and make the effect television had on social interactions and cultural behavior clearer. Audience research initially focused on understanding television as a shared cultural experience, with particular attention given to the social and psychological needs it met for

different audiences. Over time, with the continuous evolution of the industry, audience understanding became more and more complex, paving the way for more comprehensive methods of data collection and analysis.

1.3) Television Production in the Age of Streaming and Big Data

Television, broadcast or streamed, has always been closely linked to its audience, with preferences of the viewer playing a crucial role in shaping its evolution. The advent of on-demand services such as catch-up and streaming platforms has changed the way people consume content, raising new questions about the role of television in society. Television has tried to represent a common cultural identity in the past but has also embraced regional and ethnic diversity. In contexts like British television, public services and social representation were central. Over time, the study of specific programs and the flow of programming has emphasized how the audience engage with television as a unique medium. This involvement has made the analysis of audience's reactions increasingly important, both for television producers who want to create interesting content and for researchers who study the impact of television on society [5].

In the streaming age, the collection and analysis of big data changed the way producers adapt content to the audience's preferences. Streaming platforms collect huge amounts of demographic and behavioral data, not only through direct interactions such as viewing habits, but also through tracking mechanisms embedded in devices like cookies and device identifiers [6]. These data enable platforms to identify patterns in viewer preferences like specific genres, themes or even the time of day that content is consumed. This information enables producers to make more informed decisions and to make content that matches what is most likely to benefit the viewers. [7] In addition to the data collected by streaming companies, there are also other metrics that are of significance to producers and that is data collected from websites like IMDb [8] which acts as a public opinion hub, where user reviews and ratings provide a form of the audience's opinion on a product. These opinions often affect not only the future of current shows, but also wider decisions on genre trends or marketing strategies. Together, these tools show how data-driven insights are important to the creative choices for TV shows, making audience preferences an integral part of the decision-making process in modern television.

The importance of understanding audience preferences and viewing trends is highlighted by these changes. By analyzing metadata on TV programs and episodes across decades, my goal is to discover the changing patterns that have shaped the television landscape. Understanding these trends is crucial, not only to appreciate the cultural impact of television, but also to identify how changes in audience behavior, preferences and technology have influenced content creation, storytelling, and viewer engagement over time. Also, it is important to see how television has adapted to social and technological transformations, and what this could mean for the future.

1.4) Research Objectives and Contribution

There are many areas to analyze in television history and every TV show. However, this research is done to find out more about how the numbers have evolved, what changed in people's preferences and how producers have adapted to the streaming models by altering the new series' format. The following are the primary areas of focus in this research:

Quantity of TV Shows

The increasing numbers of TV shows over the years and especially the plethora of content over the last few years has been the first thought of this research. So, the goal was to track the number of new TV shows released annually, showcasing the significant growth of the medium, especially after the rise of streaming platforms. This is further explored by examining trends in the United States versus globally, highlighting the diversification of TV production worldwide.

Quality of TV Shows

Quality is examined using mainly IMDb ratings as a metric. The research explores trends in average ratings over time, identifying shifts in audience preferences and the factors contributing to changes in perceived quality. Also, attention is given to the evolution of non-US shows in recent years in quality, to see if they can challenge the consistent dominance of US-produced content.

Content Diversity

Another interesting objective for this analysis is the diversification of television content in terms of countries of origin, languages, and cultural representation. The aim is to analyze the number of non-US productions to find trends, specifically in the last decade where streaming platforms rule. This is further examined by analyzing the languages spoken and countries represented in TV shows as well as the trend in average ratings during this era.

Structural Changes in TV Shows

The research also examines how the structure of TV shows has evolved. This is done by analyzing the length of shows and their seasons every year to find patterns that show the changes and what they could mean. A connection between these changes and the rise of binge-watching culture is made as I also compare the different release methods of TV shows, especially the last few years where 'same-day' episode releases are more common, showing audience's preferences for more immediate viewing experiences.

Audience Engagement

Audience engagement is analyzed through IMDb rating votes, a metric that captures the active participation of viewers in evaluating TV content. By calculating total votes per year, the research aims to identify trends in audience involvement and how much the audience participates in a TV show's evaluating process.

Comparative Analysis of US and Non-US Shows

Throughout the research, a comparative lens is applied to study the differences between US and non-US shows. These comparisons include quality, quantity and preferences in genres.

Genre Preferences Over Time

The analysis also dives into the genres that dominate each decade, both globally and within specific regions.

This document provides an analysis of the evolution of television. Following this introduction, Section [2](#) contains a review of related work which situates the thesis within the broader academic context. Section [3](#) outlines the use of Bayesian averages to rank TV shows and interpret metadata effectively, followed by a description of the data collection and storage processes that are required for the analysis in Section [4](#). Section [5](#) includes the findings that explore the aspects of television's transformation, including trends in quantity and quality, shifts in content diversity, changes in TV shows' structures, audience engagement through IMDb rating votes, and the evolution of genre preferences. Finally, Section [6](#) summarizes the results and discusses the thesis' contribution to understanding television's role in the modern media landscape.

2) Related Work and methods

Some significant studies and metrics have been documented in the field of data analysis for TV Shows, whether it be for ratings predictions or for showing how people from all over the world react to "foreign" productions.

One of the methods used in this thesis is the implementation of a more normalized ranking table for TV shows and episodes based on Bayesian models (Bayesian Average Ranking). These models have already been utilized in some form in a different study [\[9\]](#). This study uses Bayesian hierarchical models to analyze television viewership patterns, taking both time-to-event and count data into account to predict viewer behavior. Furthermore, Moral et al. focus on how important it is to overcome systematic biases in empirical data and high dimensionality of the dataset.

Another method that deals with TV series analysis is presented [\[10\]](#) with the use of Machine Learning (ML) algorithms to predict TV show ratings. This is achieved by using a combination of data from TV shows including genre, audience sentiment, and viewer behaviors like word from the viewers on Twitter.

Another research [\[11\]](#) applies ML and statistical methods to analyze drama TV shows, using metadata such as shows' plots, season counts, and production information to make predictions on their ratings and understand what the audience prefers. This study supports the idea that metadata - like genres, year of release, and episode characteristics - is important to better understand trends and pattern changes in TV series ratings.

In an article [12], it is mentioned how important it is to combine statistical methods with analyses of traditional humanities when studying television. A sitcom called "Happy Days" is used as a case study to illustrate how descriptive statistics can pinpoint significant trends in editing styles, showing that different metrics like average shot length and cutting rates can also provide insight into the way TV productions and techniques change over the course of time.

Conducted research [13], focuses on predictive analytics of television ratings by using a linear regression approach for an examination on how variables like the number of episodes and IMDb ratings connect, revealing a negative correlation where an increase in episodes leads to lower ratings. This finding highlights the importance of content quality over quantity, which appears to be a significant factor for the way a season's length affects audience reception. Finally, emphasis is given to the need for a more complete understanding of various aspects that can influence the success of TV shows, suggesting that future research should include additional variables such as genre and other production elements.

Another study [14] examines how linguistic, geographic, and cultural factors can shape global TV show consumption. An analysis is made on Netflix's daily Top 10 shows across 80 different countries which reveals that smaller linguistic distance increases shared viewing preferences, geographic distance affects content accessibility, and cultural similarities play an important role in shaping audience behavior. The study gives an idea of how global content consumption remains connected to regional and cultural contexts, even in the streaming era.

These methods have significantly advanced the field of media analytics. This thesis is inspired by these approaches as one of its aspects lies in the application of Bayesian averages to TV show ratings. This is because a tool was needed to offer more reliable insights into TV show quality and popularity over time. The integration of Bayesian analysis with genre- and year-specific trends gives this thesis a complementary role in the contribution to the broader discussion on media analytics, offering additional perspectives on the evolution of television content and audience dynamics.

3) Bayesian Average Ranking

In this thesis, Bayesian averaging was utilized to establish a more accurate representation of a TV show's and an episode's popularity, specifically for constructing fair and reliable ranking tables. While not used directly for analyzing the quality of content, this method was employed to create a clearer image of the popularity landscape by addressing the limitations of traditional averaging techniques.

Every website that gives the option to a user or a customer to rate a product should have an algorithm that takes these products and ranks them according to the best or the worst ratings. But what happens in case a product, a movie or a TV show with 5 rating votes has a rating of 9.5/10 and another entry has a rating of 9/10 with -let's say- 500 votes? Ranking them from best to worst using the rating as a metric puts the latter into a clear disadvantage. This could be considered unfair because having a big number of votes like this and still being at 9/10 is a great achievement so the main thought was that there should be a way of changing this metric into something else that takes note on the number of rating votes too. Different companies that deal with ratings like Amazon have this kind of algorithm that can rank their products in fair terms and give the most popular products (that still have a good rating of course) an advantage. An interesting finding in Algolia's Documentation [15] was examined, talking about a custom ranking list based on the Bayesian Average metric. What they do is combine individual ratings with the overall distribution of ratings in the dataset, also considering the number of votes each item has received. This way, TV shows like the example above with high ratings and extremely low rating votes go further down on the list as more popular titles surpass them. The formula they propose for this method is:

$$B = \frac{Rv + Cm}{v + m}$$

Equation 1 – Bayesian Average Calculation

Where:

- B : Bayesian average rating.
- R : IMDb rating of the show or episode.
- v : Number of IMDb votes for the show or episode.
- C : Average IMDb rating across the dataset.
- m : Minimum number of votes required to assign "full trust" to an item's rating.

3.1) Empirical Analysis

When sorting from highest to lowest rating, as seen in fig. 1, we can see the problem described in the previous section:

Title	Year	Year End	IMDbRating	IMDbRatingVotes
DVD Talk with Cody Clarke	2010	2010	10.00	7
Guitar Basics with Cody Clarke	2009	2009	10.00	6
Live at the Laugh Zone	2008	2008	10.00	6
Welcome to Italy	2015	NULL	10.00	5
The Chaplain with Teachers in Cars...	2020	2020	10.00	11
Isev	2020	NULL	10.00	7
Robbie Tucker's the Boy in the Bat...	2019	2019	10.00	6
Rang Tara Ishq No	2020	2020	10.00	6
Livros Animados	2014	2014	10.00	7
Jeff Seid Youtube	2018	NULL	10.00	10
Flashpoint	1992	2015	10.00	7
Garden Gray the with Voices	2015	2015	10.00	13
The Message of Fatima	2017	2017	10.00	8
Red Bull Conquest	2018	NULL	10.00	6
K-15	1995	NULL	9.90	157
Bikram malati show	2019	NULL	9.90	10
Saga 2	2020	NULL	9.90	8
Dewlover420 Weekly	2017	NULL	9.90	10
The Edge of Adventure	2017	NULL	9.90	9
Octopus Monday Live	2019	NULL	9.90	18

Figure 1 – Top 20 TV Shows (Highest to lowest *IMDb Ratings*)

None of these shows should be in the top-20 of the last three decades when they have so few rating votes. So the testing on how this list could change was a necessity to have a better view of the popularity of shows in the database.

For the application of Eq. 1 to the complete dataset of TV Shows, the most logical value to assign to the variable m had to be located. Of course, the TV show list was sorted from highest to lowest regarding the number of the rating votes to pinpoint the perfect spot to set my threshold. Starting from low trust levels, I selected the value located at the 25th percentile of the sorted list. After the process, the top 20 will look as presented in fig. 2 (the value of the Bayesian average is assigned to 'AV'):

Title	Year	AV	IMDbRating	IMDbRatingVotes
Christina Grimmie Zeldaxlove 64	2009	9.83...	9.90	618
Friday Five Sharp	2015	9.79...	9.80	4124
Londra Merkez	2018	9.77...	9.80	1822
Upanishad Ganga	2012	9.75...	9.80	1006
Jerma985	2011	9.72...	9.80	570
BnftV	2018	9.72...	9.80	561
Secrets 2	2024	9.71...	9.80	489
Steven Reeds Weber Cooks	2012	9.69...	9.90	194
Hiç Komik Degil	2011	9.69...	9.80	364
Choufli Hal	2005	9.68...	9.70	2769
Call Me Kevin	2016	9.67...	9.70	1866
Kothao Keu Nei	1990	9.67...	9.70	1327
OdBita Pot	2022	9.65...	9.70	874
Amrutham	2001	9.65...	9.70	861
Marmadesam	1997	9.65...	9.70	859
K-15	1995	9.65...	9.90	157
The Simonetta Lein Show	2020	9.64...	9.80	257
The Why Files	2020	9.64...	9.70	692
Raja Shivchhatrapati	2008	9.64...	9.70	675
Prime Time	2016	9.61...	9.90	135

Figure 2 – Top 20 in the Custom Ranking List (Highest to lowest *Bayesian Average Rating – 25th percentile*)

The improvement is clear but it's still not what the final table should look like, because of the low threshold. This threshold still gives TV shows with fewer rating votes significant influence on the rankings, so the thought was that we needed higher trust levels for the variable m to better balance this influence and give more representation to highly rated shows with a larger number of votes. Testing it with the value contained at the 75th percentile, we get the result that is displayed on fig. 3.

Title	Year	AV	IMDbRating	IMDbRatingVotes
Friday Five Sharp	2015	9.69...	9.80	4124
Londra Merkez	2018	9.56...	9.80	1822
Choufli Hal	2005	9.54...	9.70	2769
Arkadas Canlisi	2022	9.51...	9.60	4778
Breaking Bad	2008	9.49...	9.50	2123785
Planet Earth II	2016	9.49...	9.50	156963
Call Me Kevin	2016	9.48...	9.70	1866
Run BTS!	2015	9.45...	9.60	2733
Upanishad Ganga	2012	9.40...	9.80	1006
Minecraft Epic	2019	9.40...	9.50	4009
Kothao Keu Nei	1990	9.40...	9.70	1327
Band of Brothers	2001	9.39...	9.40	523156
Planet Earth	2006	9.39...	9.40	219309
Bluey	2018	9.38...	9.40	28598
BB Ki Vines	2015	9.36...	9.40	10506
Bi' Kahve 2 Sohbet	2016	9.31...	9.60	1348
Tree of Life	2014	9.30...	9.40	3853
Coke Studio	2008	9.30...	9.60	1276
Chernobyl	2019	9.29...	9.30	868346
The Wire	2002	9.29...	9.30	375007

Figure 3 - Top 20 in the Custom Ranking List (Highest to lowest **Bayesian Average Rating** – 75th percentile)

After this, we can finally see popular shows like Breaking Bad, Band of Brothers and The Wire in the top 20 but, still not what the list should look like based on knowledge of the TV scene. Some other testing was done with the value of rating votes at the 85th and 95th percentile but I finally realized that the optimal spot was at the 98th percentile. A threshold of 11,060 rating votes makes the top 20 spots on the table look as presented in fig. 4.

Title	Year	AV	IMDbRating	IMDbRatingVotes
Breaking Bad	2008	9.48...	9.50	2123785
Band of Brothers	2001	9.34...	9.40	523156
Planet Earth II	2016	9.32...	9.50	156963
Planet Earth	2006	9.27...	9.40	219309
Chernobyl	2019	9.26...	9.30	868346
The Wire	2002	9.23...	9.30	375007
Avatar: The Last Airbender	2005	9.22...	9.30	368587
Game of Thrones	2011	9.18...	9.20	2271615
The Sopranos	1999	9.14...	9.20	466689
Scam 1992: The Harshad Mehta Story	2020	9.13...	9.30	156639
Aspirants	2021	9.12...	9.20	313589
Cosmos: A Spacetime Odyssey	2014	9.10...	9.30	128307
Sherlock	2010	9.07...	9.10	995604
Rick and Morty	2013	9.05...	9.10	600566
Attack on Titan	2013	9.05...	9.10	504353
Fullmetal Alchemist: Brotherhood	2009	8.98...	9.10	196089
Sapne Vs Everyone	2023	8.97...	9.30	71630
The Office	2005	8.96...	9.00	706191
Better Call Saul	2015	8.96...	9.00	647614
The Last Dance	2020	8.94...	9.10	151767

Figure 4 - Top 20 in the Custom Ranking List (Highest to lowest **Bayesian Average Rating** – 98th percentile – Final Form)

4) Collection and Storage of Data

IMDb [8] is one of the most popular media databases that contain details about millions of titles, ranging from movies and TV shows to video games and podcasts. It also gives its users the option to rate movies or TV Shows they've watched, creating an average and, in this way, giving each title a profile and a quality metric. It can certainly influence the audience's opinion and give them a chance to show their preferences to producers. That's why I chose to extract my information from this website as thousands of people give their opinions daily for many TV Shows.

To achieve this, I used an API I found on RapidAPI called imdb8 by ApiDojo [16] which takes the data from IMDb and provides them after a request. This API was utilized to harvest all the unique IDs of TV shows from 1990 to 2023. This timeframe was selected to set a balance between collecting a substantial amount of data and maintaining relevance to modern viewing habits. Using the Python Requests library [17], I sent requests to imdb8 and stored the necessary data in text files to be used later to extract the metadata to store in my database.

```
url = "https://imdb8.p.rapidapi.com/v2/search-advance"
after = ""

payload = {
    "first": 100,
    "after": after,
    "includeReleaseDates": False,
    "sort": {
        "sortBy": "USER_RATING",
        "sortOrder": "ASC"
    },
    "releaseDateRange": {
        "end": "1990-12-31",
        "start": "1990-01-01"
    },
    "anyTitleTypeIds": ["tvSeries", "tvMiniSeries"]
}

headers = {
    "content-type": "application/json",
    "X-RapidAPI-Key": "XXXXXXXXXXXXXXXXXXXXXXXXXXXX",
    "X-RapidAPI-Host": "imdb8.p.rapidapi.com"
}

response = requests.post(url, json=payload, headers=headers)
```

Picture 1 – POST Request (Python)

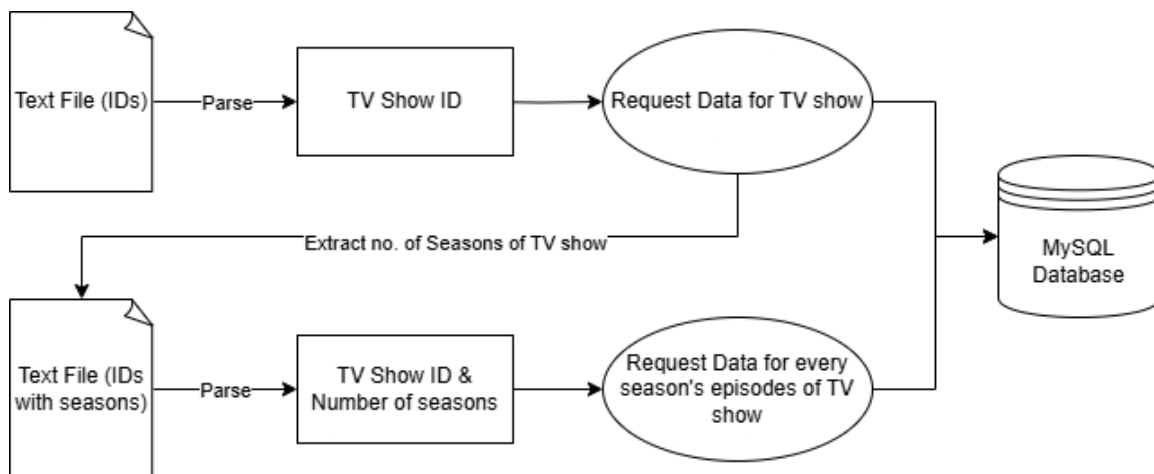
```
{
  "data": {
    "advancedTitleSearch": {
      "edges": [
        {
          "node": {
            "title": {
              "__typename": "Title",
              "id": "tt1824295"
            }
          }
        },
        {
          "node": {
            "title": {
              "__typename": "Title",
              "id": "tt9175350"
            }
          }
        },
        {
          "node": {
            "title": {
              "__typename": "Title",
              "id": "tt0350420"
            }
          }
        }
      ]
    }
  }
}
```

Picture 2 – Response (JSON format)

An example of a request is displayed in pic. 1. In this example the first 100 IDs of TV shows and miniseries that have been released in 1990 are requested and a snippet of the API response is presented in pic. 2 where the first three nodes can be seen with their unique IDs. These correspond to the unique IDs that are given by IMDb to each title.

After some research I found the API that fit the most for the purposes of this thesis from TV-API.com [18]. So, I created an automated program with Python via Visual Studio Code to read the harvested IDs of TV Shows from the text files, one by one, and send a signal to TV-API requesting metadata belonging to each entry using the URL “GET/

API/Title/{apiKey}/{id}”. The data were provided in JSON format and what my program did was navigate this dictionary and extract the values of the needed data, while nulling the missing values. For example, some TV Shows lacked genre, country, ratings, or language information and these empty values were nullled before insertion in the database. After that, new text files were created containing every ID that was in the previous files but also containing the number of seasons for each ID’s TV Show. These files were to be used later in the extraction of metadata from the episodes. Using another automated Python program, requests were sent using the URL “GET/API/SeasonEpisodes/{apiKey}/{id}/{seasonNumber}” -one show, one season at a time- to take the available metadata for each episode of a TV Show. This process is visualized in pic. 3 for better understanding.



Picture 3 – Flow Chart of the data extraction process from TV-API

Both TV Shows’ and episodes’ metadata were stored in a locally created MySQL Database for the purpose of the analysis. Two interconnected tables were created for these purposes:

- **`tv\_shows`**: Containing metadata for TV shows (table schema in fig. 5)
- **`tv\_episodes`**: Containing metadata for episodes (table schema in fig. 6)

Field	Type	Null	Key
ID	varchar(12)	NO	PRI
Title	varchar(255)	YES	
Year	varchar(255)	YES	
Year End	varchar(255)	YES	
IMDbRating	decimal(16,2)	YES	
IMDbRatingVotes	int	YES	
Genres	varchar(255)	YES	
Content Rating	varchar(45)	YES	
Date Released	date	YES	
Seasons	int	YES	
Plot	longtext	YES	
Language	varchar(255)	YES	
Country	varchar(255)	YES	
MetacriticRating	int	YES	

Figure 6 – MySQL Database Schema for the Table of TV Shows

Field	Type	Null	Key
SeriesID	varchar(12)	NO	MUL
ID	varchar(12)	NO	PRI
SeasonNum	int	YES	
EpisodeNum	int	YES	
Title	varchar(255)	YES	
IMDbRating	decimal(16,2)	YES	
IMDbRatingVotes	int	YES	
Date Released	date	YES	
Plot	longtext	YES	

Figure 5 - MySQL Database Schema for the Table of Episodes

The process of extracting and storing the data took several months because of limitations in requests (5000) per day, set by TV-API. In total, this database includes **255,303 TV Shows** and **2,184,579 Episodes** from all over the world, ranging from 1990 to 2023 (years of release).

4.1) Limitations of the Data

While the data provided by TV-API are mostly rich, there appear to be some limitations that should be acknowledged:

- **Regional Bias:** Some regions may have less data due to varying levels of audience participation on IMDb.
- **Incomplete Metadata:** Certain shows and episodes may lack complete information, particularly for older or less popular titles. For instance, there are titles that don't have a rating, or don't provide information regarding their genres.
- **Exclusivity of IMDb Ratings:** The analysis relies on IMDb as the primary audience metric, which may not fully capture viewership dynamics across all platforms.

5) Metrics and Results

After the creation of my database, everything was ready to begin the analysis of the extracted metadata to examine the Television scene and find different patterns that can potentially explain the changes made to certain aspects like the structure or the quantity of TV shows and how they adapt to the audiences' preferences. My thesis was visualized with the aid of graphs that were created using Visual Studio Code and Python tools and libraries. The list of libraries used contains **Pandas** for data manipulation, **Matplotlib** in combination with **Seaborn** for visualizations and **NumPy** for calculations and trendlines.

5.1) Quantity and quality

In fig. 7 we see the total number of TV Shows per year (globally). In blue we have the sums of TV shows that were released before the era of streaming and in green the TV shows during this era. Looking at it, the rise in the number of titles increases throughout the years, even in the period of cable television. But, through observation, the year- after-year increase is much more significant in the years of streaming. Based on this graph, we observe that as more streaming services are being launched, there appears to be significant proliferation in the rate at which the annual number of TV shows increases. For example, 2011 sees a ~23% increase in the number of TV shows, with both Netflix and Amazon Prime Video in the game of streaming that still goes on a year later (2012) with Hulu's first original TV show. Also, we observe a ~10% rise after the launch of Disney+ and Apple TV+ in 2020. It's clear that the more companies invest in their streaming platforms and engage in this competition between brands, the more content will be created. This raises the question of whether quality remains despite this rise.

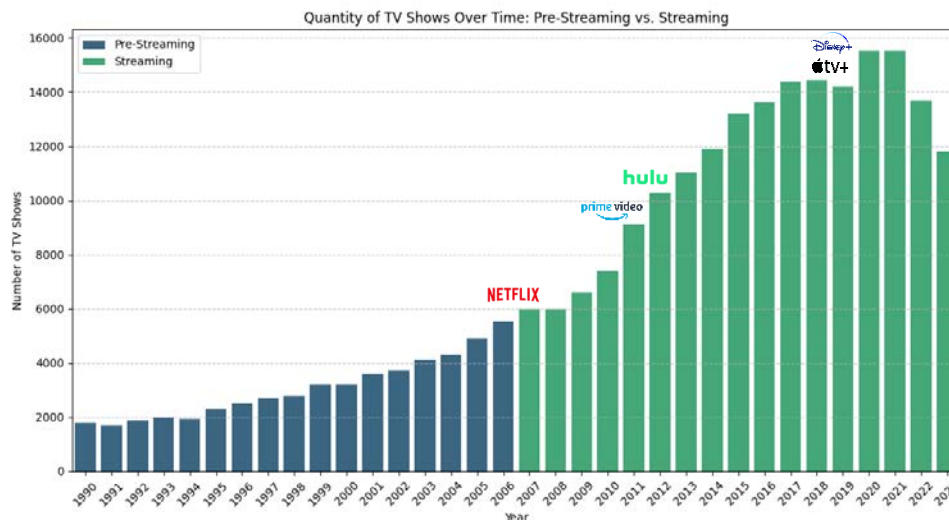


Figure 7 – Quantity of TV Shows Over Time: Pre- Streaming vs. Streaming Era

For the quality examination, the metric used was the ratings of every TV show (available) globally. However, calculating the average ratings per year for shows that were released the same year was not enough. Attention should be given to TV shows that have been releasing multiple seasons throughout the years to ensure fairer scores each year. The process of this examination required the IMDb rating scores of each show that ran for more than 1 year to be added to the sum of all the years, up to its final year. For example, the score of 'Breaking Bad' contributed to the calculation of the average scores in the years 2008-2013. The results show a constant drop in quality up to 2006 and then a rise from there with the pace picking up after 2008. Based on this graph (fig. 8), it looks like TV Shows were aided by the launch of streaming services and increased in quality despite the growing numbers and plethora of content. However, a small drop in 2021 could be because of the record-breaking number of TV shows that was displayed in fig. 7. This drop causes a momentary pause in the rise in quality but continues in the next years, still unable to achieve the scores that 1991 has.

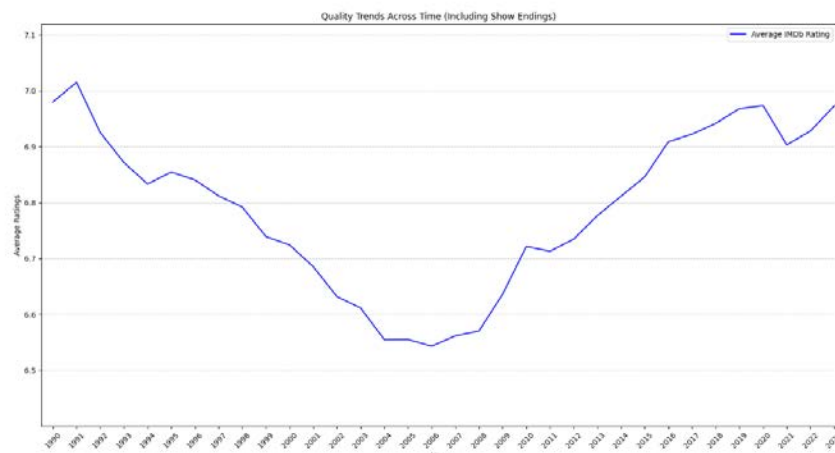


Figure 8 - Quality Trends Across Time (Including TV shows that haven't ended yet every year, up until their end year - Using IMDb Ratings as a metric)

Another thought was to apply the same analysis, but on the Bayesian average ranked table as seen in fig. 9. Obviously, this is a lot different if we note that the range of values is being compressed because of the more balanced data that this table contains. So, the differences between years are smaller and trends are less obvious. Apart from the fact that there are some small fluctuations (could be because of variations in shows' quality/popularity) nothing significant can be taken away. The difference between graphs is that IMDb ratings (fig. 8) can be influenced by outliers or highly rated shows, but Bayesian averages smooth out these extremes, making it harder to spot year-by-year changes.

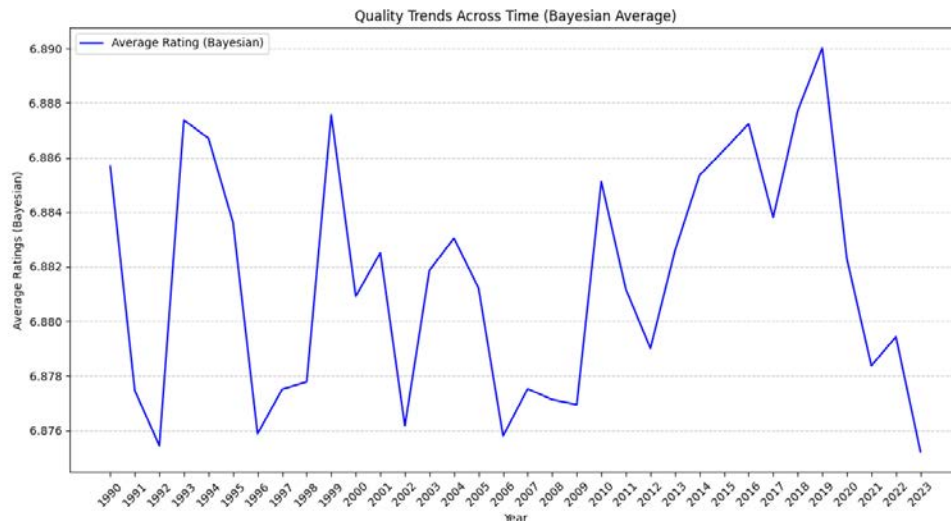


Figure 9 – Quality Trends Across Time (Using Bayesian Average as a metric)

5.2) Diversity

One way of looking at how the diversity of content has evolved globally is by choosing as a metric the different languages used in TV Shows and the number of unique countries that participate in productions every year. Fig. 10 shows that starting from ~60 different countries (blue line) and languages (green line) that were actively participating in TV shows in 1990, a point was reached at which there are more than 130 unique cultures presented with a peak of 141 countries producing shows in 2021. Apart from this upwards trend in cultural representation there also appears to be a change in 2013 onward. So, from the year of the first original Netflix series (House of Cards – 2013) and after, there seems to be less distance between the number of countries and languages. This could mean that more effort was made from non-USA countries of production to produce shows in their own language as opposed to the previous years where the distance in numbers could mean that they were co-producing a show in another language (e.g., English).

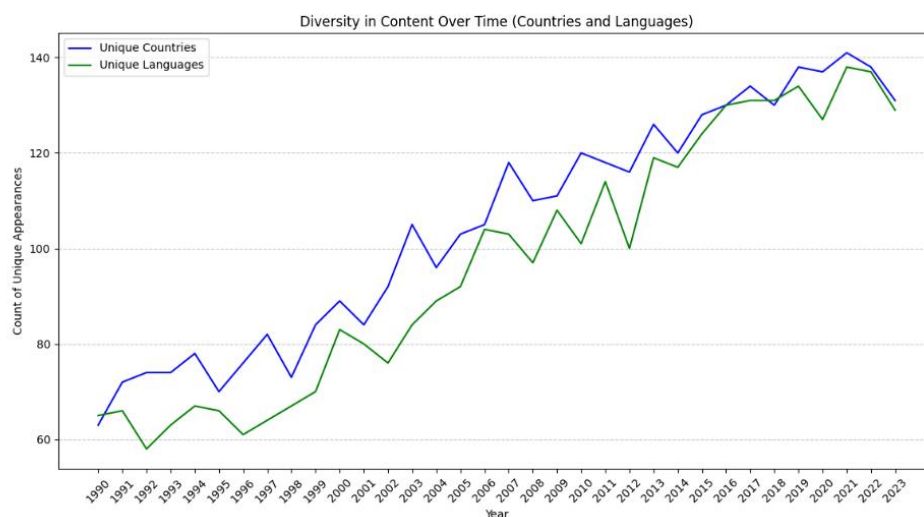


Figure 10 – Diversity in Content over time (Countries and Languages)

A different way to highlight the evolution of different cultures in TV is to measure the total number of non-USA productions per year and compare them to the global numbers in fig. 11. The increased content from countries outside the U.S is presented in this graph. Here, we can observe two things:

- The number of non-US TV Shows is higher every year and peaks in 2021
- The percentage of “foreign” TV shows goes down throughout the years, especially after 2010 with it being 80% in 1990 and 74% in 2023.

Although TV shows produced in the USA have a higher increase rate after 2010, we can see that the non-USA shows compete fiercely with these numbers and do not let the Hollywood productions move further away.

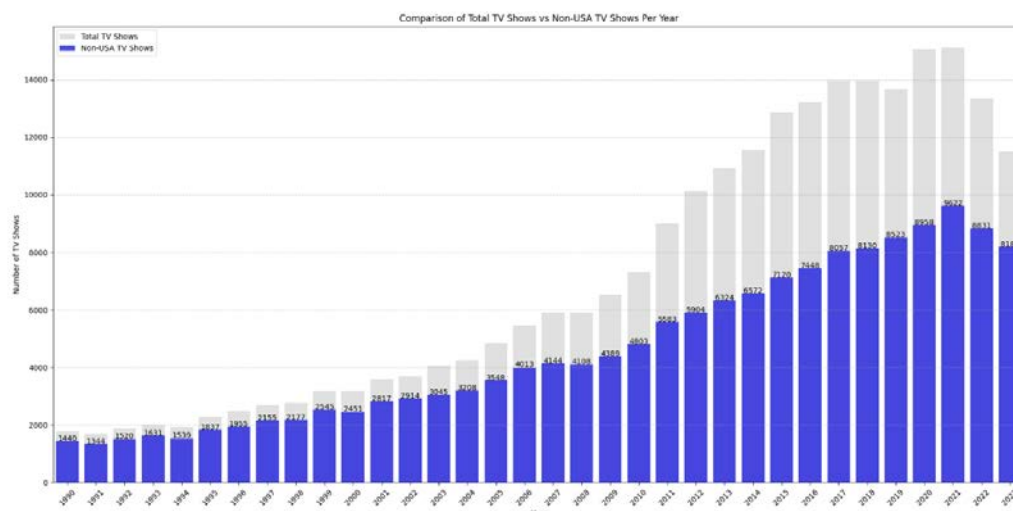


Figure 11 – Comparison of Total TV Shows vs. Non-USA TV Shows Per Year (Number of new releases per year)

Fig. 12 contains the same calculations that were made with the average ratings per year in fig. 8, this time applied to non-USA TV shows. Looking at the average ratings of these

shows per year, their pattern seems to line up with the graph in fig. 9 that showed the ratings for all TV shows. Once again, we notice an increase in quality after 2006 which could point to the fact that with the even more increasing content of streaming services, the “foreign” TV shows managed to win the audiences’ hearts. This is highlighted by the graph in fig. 13.

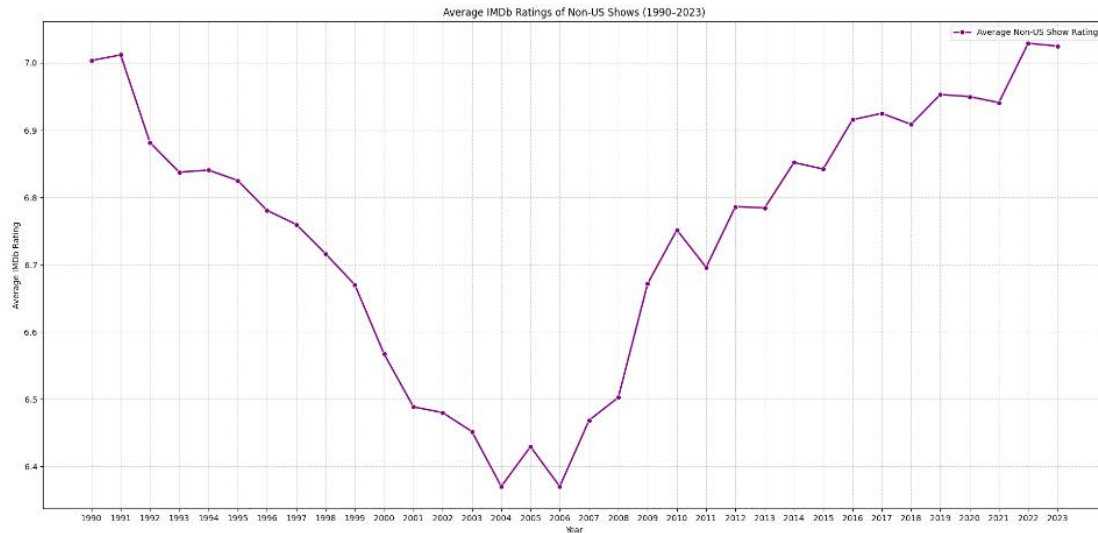


Figure 12 – Average IMDb Ratings of Non-US TV Shows (1990-2023)

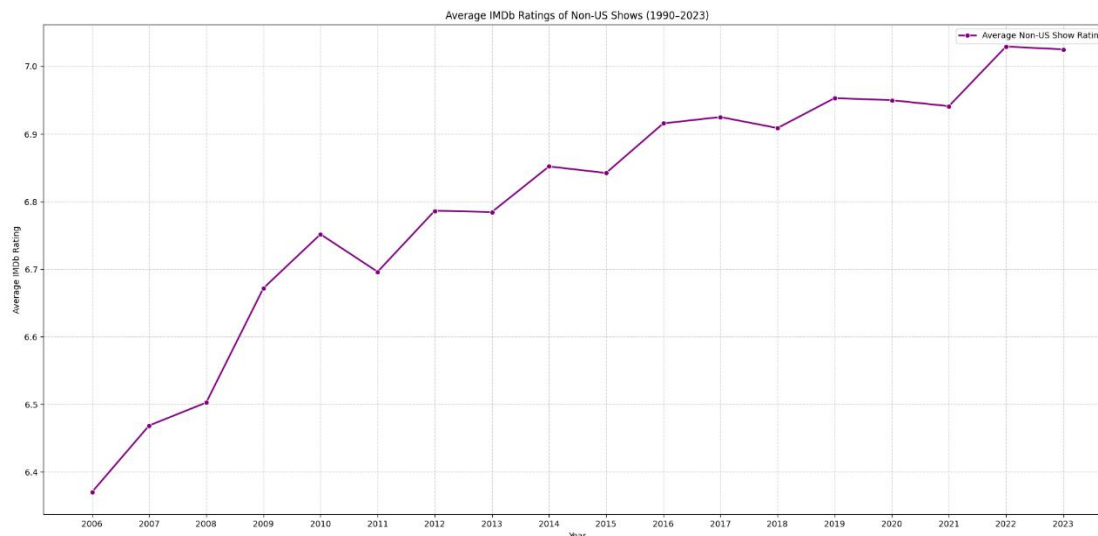


Figure 13 – Average IMDb Ratings of Non-US TV shows (2006-2023)

Another approach was made to see the overall preferences of audiences throughout the years by comparing the IMDb ratings of global (green line), USA (blue line) and non-USA (purple line) TV Shows. This comparison in fig. 14 shows that the TV series which had the

involvement of the USA in their production remain consistently higher after 1991 in IMDb ratings than the rest, lowering the average ratings each year, except for the years 2021 and 2022, where they have almost the same average ratings. These shrinking distances that are observed after 2019 between USA and “foreign” productions could mean that the TV shows outside of the USA have succeeded to reach the quality levels and audience approval that American TV shows have, over the last few years.

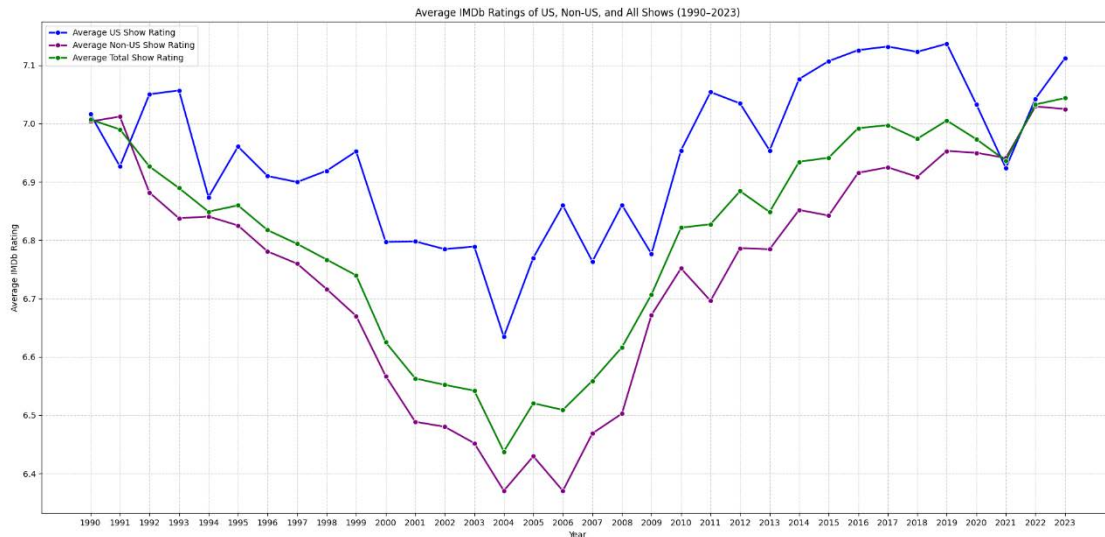


Figure 14 – Average IMDb Ratings of US, Non-US, and All shows (1990-2023)

5.3) Structure Changes

Not all TV shows have the same structure. Some are longer, some have shorter seasons, and some have only one season (limited series). One more difference which is mostly known in the last few years is the release of all episodes of a season on the same day. This method is known to encourage binge-watching as its content leaves the audience craving more, episode after episode.

To examine the changes made structure-wise I chose the average number of seasons per year (for shows released that year) as a metric in fig. 15. Of course, the data used for this graph is clear of TV Shows that have no recorded ‘End Year’, so TV shows that are still airing new episodes are removed. It’s obvious that the average number of seasons decreases more and more as the years pass, dropping lower than two seasons in average after 2018. This could only mean that an increasing number of producers prefer the method of short storytelling.

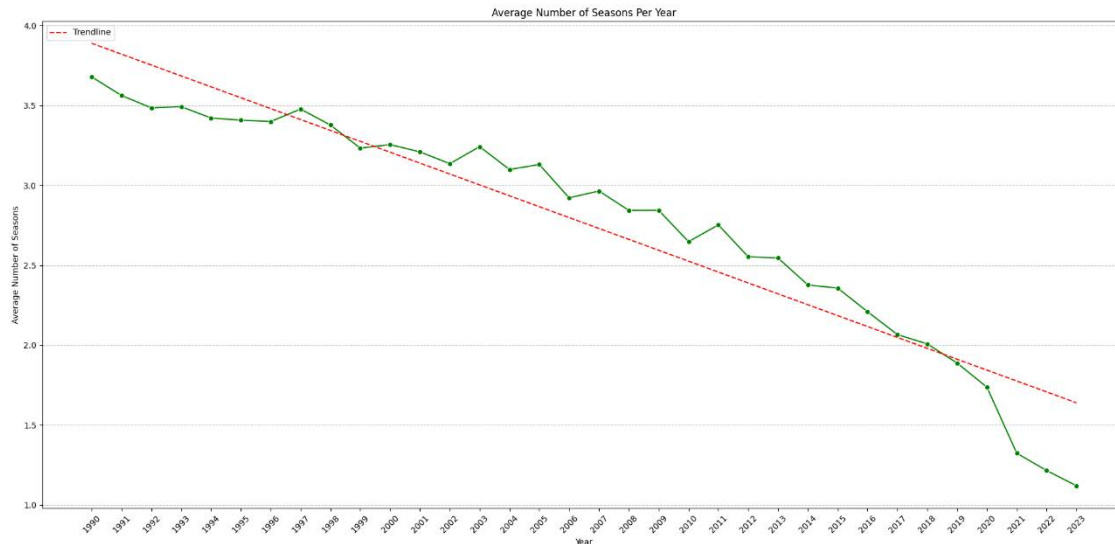


Figure 15 – Average Number of Seasons Per Year

This assumption is further supported by evidence obtained by analyzing the average number of episodes a season has per year (of a TV show's release) in fig. 16. Here we clearly see how times are changing the way TV shows are formatted with shorter seasons, especially the last 10 years when the drop in numbers is even more abrupt, as what was once a peak of ~19 episodes per season in 1994 became an average of ~10 episodes for a season of a TV show nowadays. So, we're looking at a ~50% drop in the number of episodes making it more clear that this short storytelling method is here to stay.



Figure 16 – Average Number of Episodes Per Season Every Year

Fig. 17 could show a potential reason as to why there is this need to shorten the length of a TV show. This graph contains the average IMDb Ratings per shows' total number of seasons in scatter points with the red trendline highlighting a drop in ratings the more seasons a show has. So, a reason for the decrease in length could be this slight drop of the average rating of TV shows with more than 1 season. Based on these, an assumption can be made that the longer a TV show is, the more people watching lose their interest or the show lacks in quality.

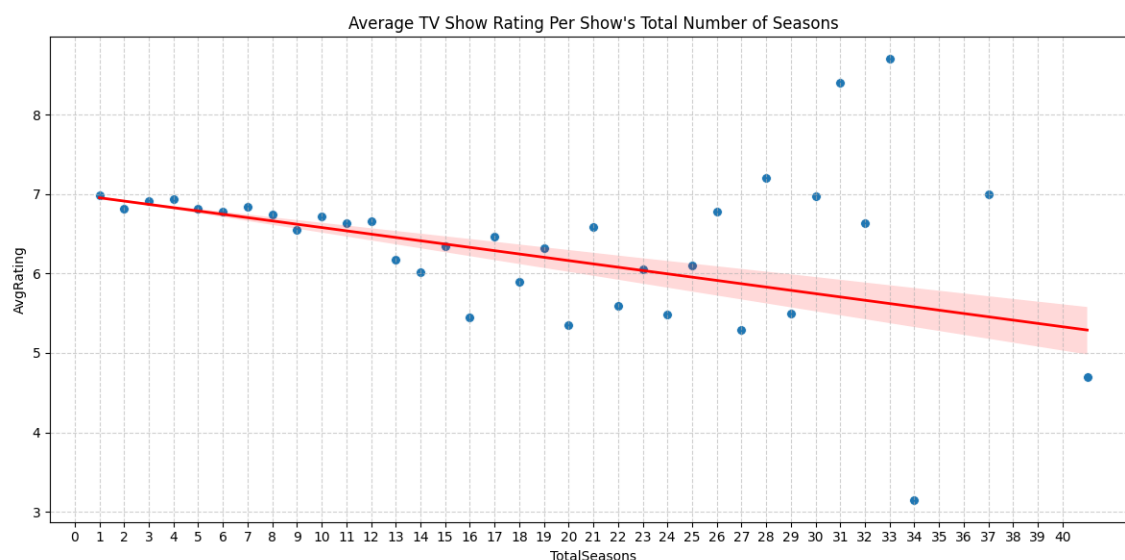


Figure 17 – Average TV Show Rating Per Show's Total Number of Seasons

As mentioned earlier, extra analysis was performed to assess the way new shows release their episodes. For this purpose, episodes from the first seasons of all TV shows were examined and were split into two groups:

- Shows that had all their 1st season episodes released on the same day.
- Shows that had a different release date for each episode.

Fig. 18 shows a steady increase in the number of shows with the 'Same Day' format after 2008 and then a sudden rise in 2021. Based on fig. 19 we could assume that these surprisingly high numbers could be the outcome of the years of higher audience scores these shows had over the ones that belong in the 'Different Days' category concerning the release of episodes. Another reason for this change could also be the state of life at the time, with the outburst of Covid-19 and the global quarantine leading streaming platform executives to release as much content as possible in a short time for people staying at home to subscribe to their services and consume it.

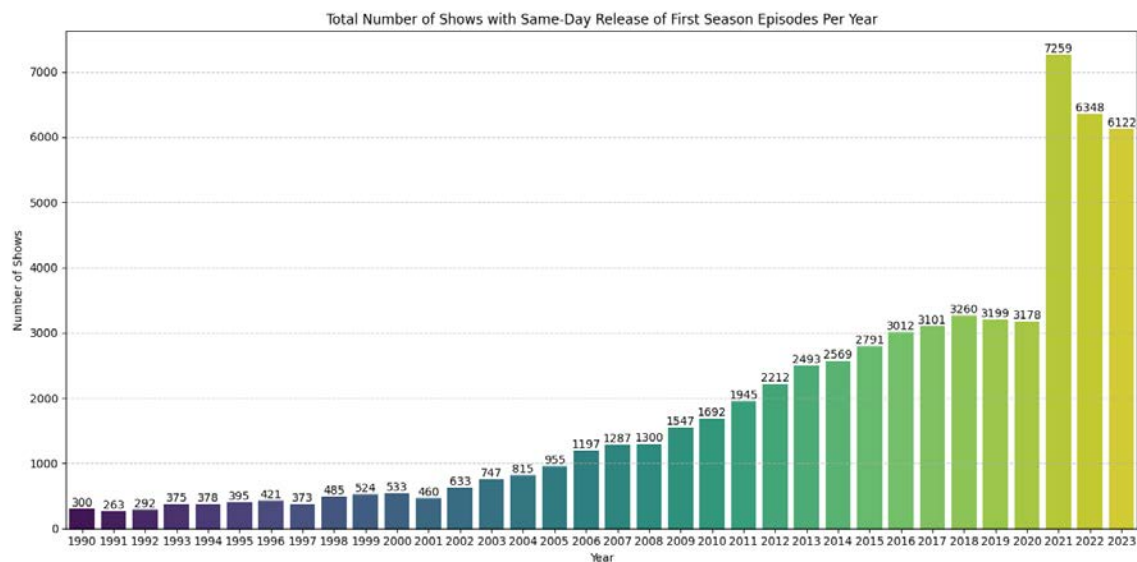


Figure 18 – Total Number of Shows with Same-Day Release of First Season Episodes Per Year

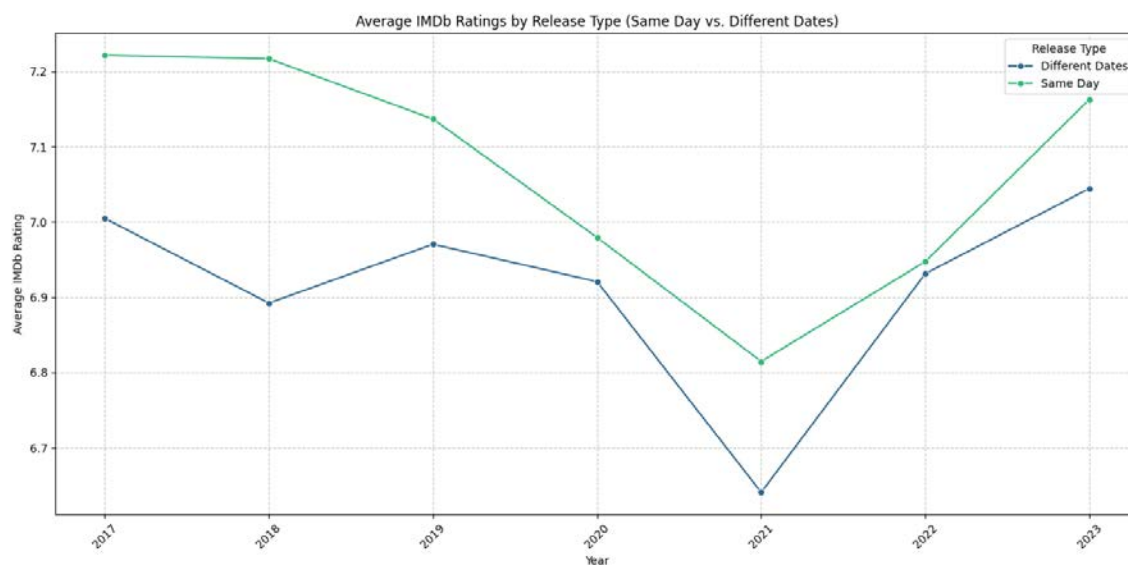


Figure 19 – Average IMDb Ratings by Release Type (Same Day vs. Different Dates)

5.4) Audience Engagement

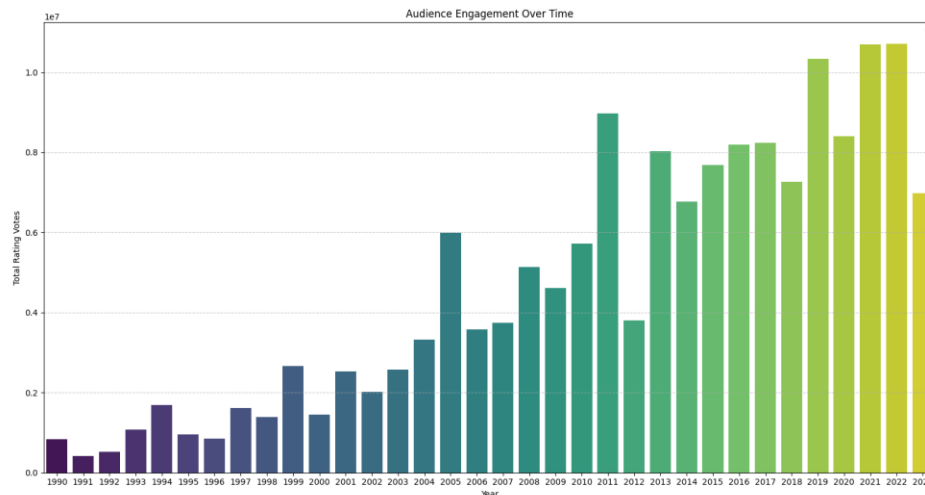


Figure 20 – Audience engagement over time (showing total IMDb Rating votes per year)

Audience engagement could only be calculated and analyzed based on the metadata that were available. This data was the total number of rating votes on IMDb for TV series that are released each year. Something that stands out in fig. 20 is the way TV shows from 2013 and up, steadily stay above 600,000 rating votes (~50%), which could mean that people are more likely to go rate a new TV show that they liked/disliked compared to the older ones. This might be because people understand their opinion's impact on how new shows are created and show it actively the last few years (over a million total votes in 2021 and 2022), so that the producers know what road to follow when making decisions on new shows.

5.5) Genre Preferences

In this section, we look at how genres have evolved throughout the years, not only using their number of appearances as a metric but also based on the average IMDb ratings and the Bayesian average scores. Genres are split into four groups according to the decade when the shows were released and the top 5 genres of each decade (using the metrics mentioned above) are analyzed.

First, an analysis on the top 5 genres based on total number of appearances (for each genre) was made. TV Shows were split into four groups according to the decade in which they were released. Though the list does not seem to change dramatically (except in the 90s where 'Family' had the place that was later for 'Reality TV'), we can see that some of the genres change places. For example, 'Documentary' surpasses Drama in the 2000s only to find itself third again instead of 'Drama' in the 2010s. The most notable change in the hierarchy is the domination of the 'Drama' category in the 2020s, but of course a lot could change since there's only data for three years. Overall, 'Comedy' outnumbers every other genre in terms of appearances throughout the years.

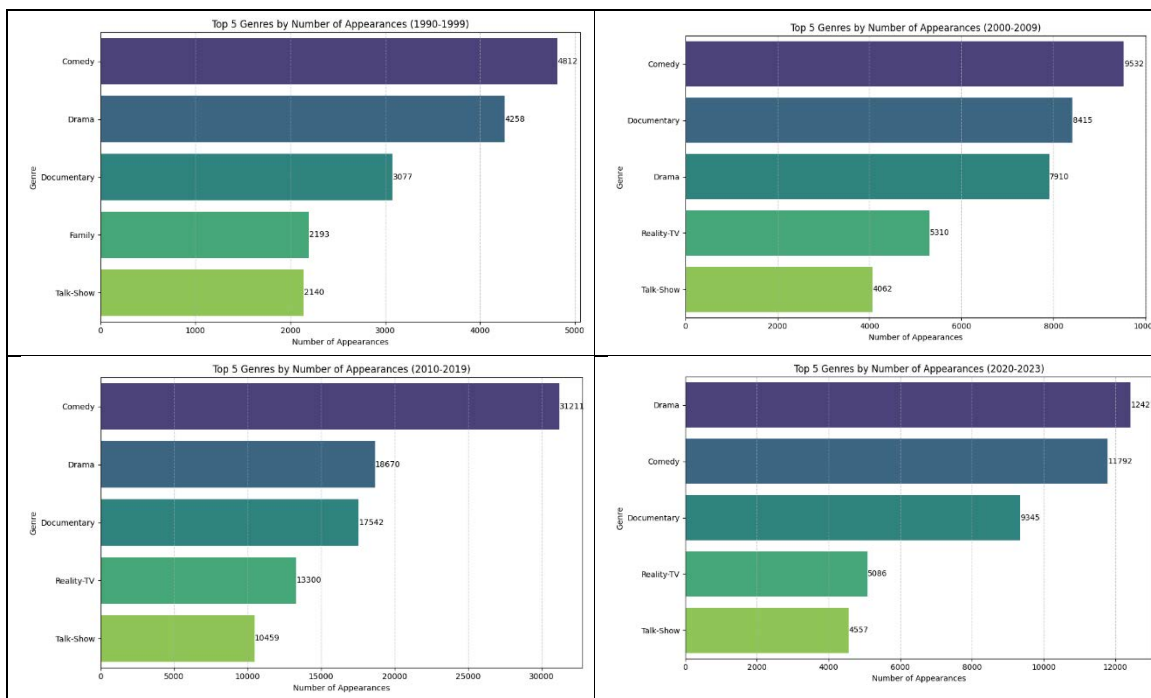


Figure 21 - Top 5 Genres based on number of appearances (split into four groups according to the corresponding decade)

The same principles were applied to the genres of non-USA TV Shows in fig. 22. Comparing the previous graphs (fig. 21) with the graphs in fig. 22 it appears that TV shows outside of the US have 'Drama' on top twice and the other genres doesn't differ significantly.

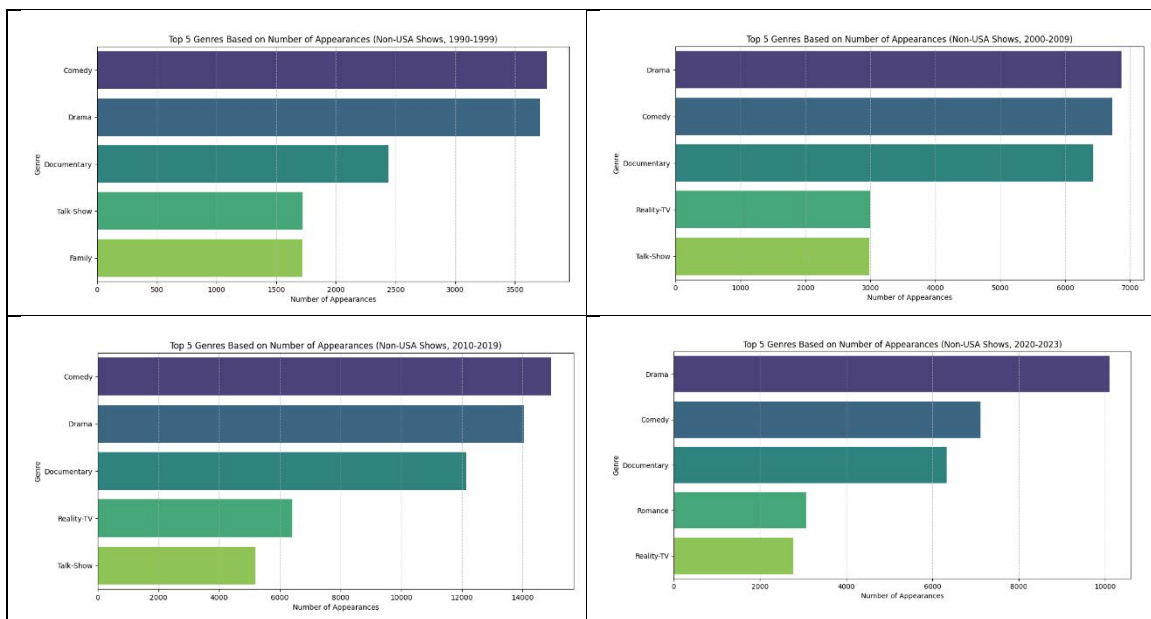


Figure 22 - Top 5 Genres of Non-US TV Shows based on number of appearances (split into four groups according to the corresponding decade)

Fig. 23 shows the analysis of the top 5 genres based on average IMDb ratings across four decades appearing a lot more different in contrast to the most frequent genres. Interestingly, none of the genres dominating in fig. 21 are present on this top 5 list. In the 90s and 2000s, ‘History’ reigns supreme, but its dominance wanes in the 2010s and 2020s, where it is eventually replaced by ‘Western’ as the most highly rated category.

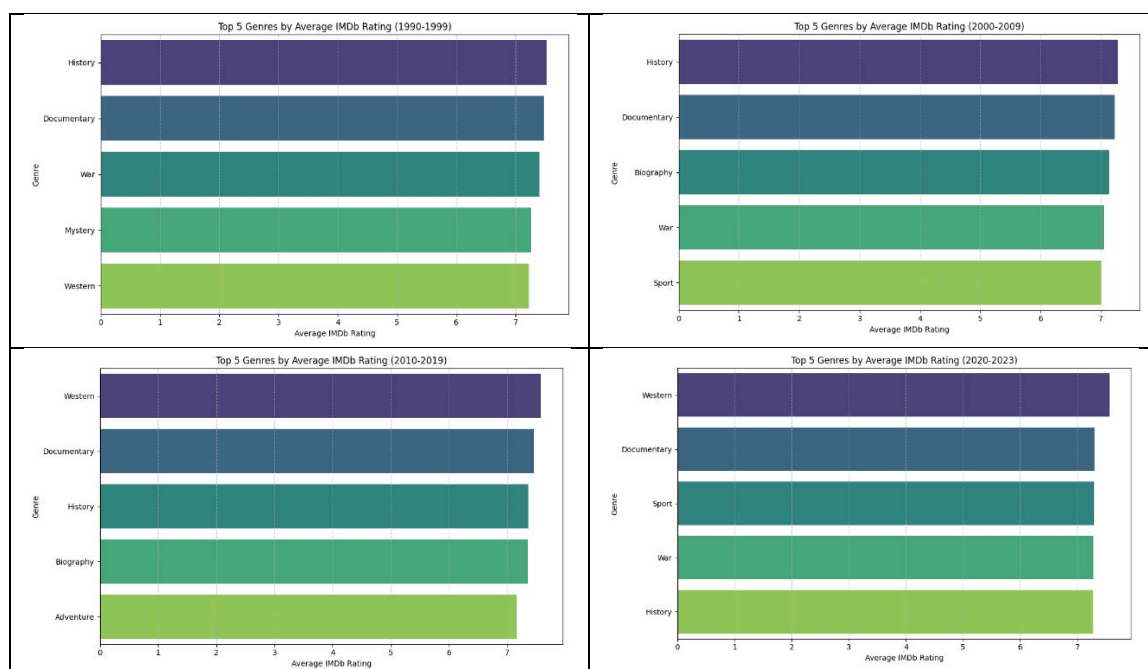


Figure 23 - Top 5 Genres based on Average IMDb Rating (split into four groups according to the corresponding decade)

Another comparison was made between decades, this time based on Bayesian average ratings for each. Fig. 24 reveals a subtle yet noteworthy shift in preferences. While ‘War’ and ‘Action’ maintain a presence across decades, genres like ‘Western’ and ‘Biography’ emerge strongly in the 2010s and 2020s. This evolution suggests a growing appreciation from audiences for these categories in more recent years. Another thing to point out is that the ratings for these genres remain remarkably close, highlighting the effectiveness of Bayesian averaging in minimizing extreme outliers and ensuring a fair comparison.

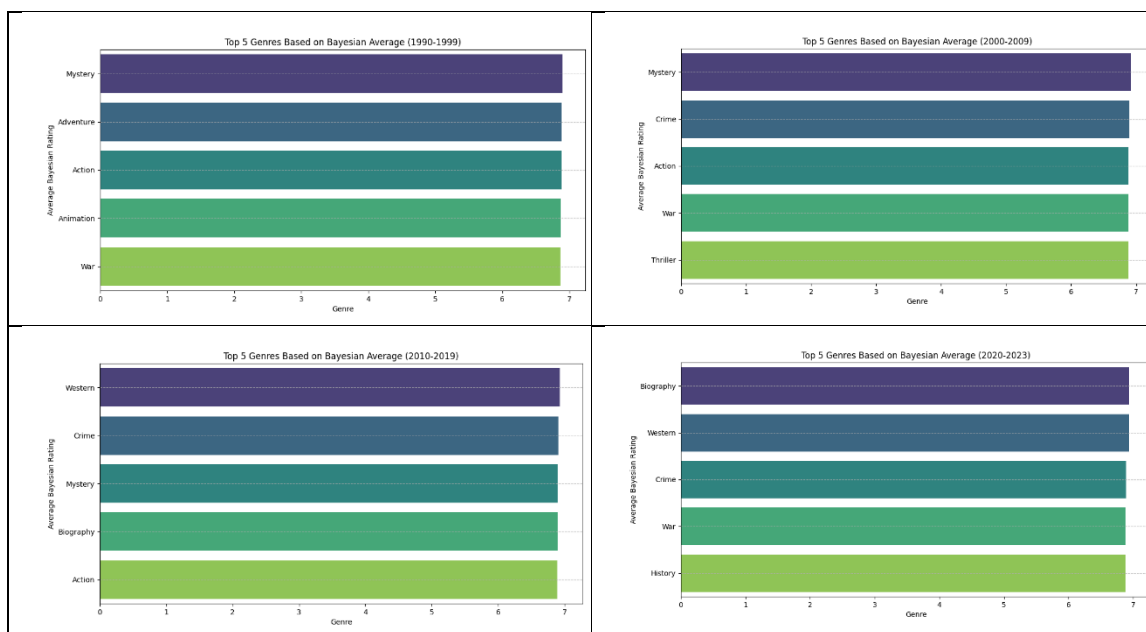


Figure 24 – Top 5 Genres based on Bayesian Average (split into four groups according to the corresponding decade)

There are some notable things in this genre analysis. ‘Comedy’ consistently dominates in terms of total appearances, with ‘Drama’ and ‘Documentary’ occasionally shifting positions. ‘Drama’ emerges as the leading genre in the 2020s, although the data is limited to three years. Outside the USA, ‘Drama’ remains the top genre, with minimal variation in others. Interestingly, when analyzing average IMDb ratings, high-ranking genres like ‘History’ and ‘Western’ stand out, differing significantly from the most frequent genres. Bayesian average ratings further reveal small changes, with ‘War’ and ‘Action’ maintaining a steady presence, while ‘Western’ and ‘Biography’ genres gain prominence in the 2010s and 2020s.

Looking at all the different genres and combinations of categories it seemed interesting to have knowledge of the top genre combinations in numbers. That was the thought behind figures 25 and 26. Of course, the top 5 were already examined in fig. 21 and genres like ‘Comedy’, ‘Documentary’ and ‘Drama’ do not seem so strange dominating the list but combinations like ‘Drama, Romance’ and ‘Comedy, Drama’ stand out when excluding single genres. If we also think about other combinations like ‘Action, Adventure, Animation’ and ‘Comedy, Drama, Romance’ we can assume that most TV shows, and extensively what the audiences are drawn to, have different themes and narratives mixed. One important thing to note is the consistent presence of ‘Comedy’ (9 appearances in fig. 26) which highlights its appeal across formats and genre pairings.

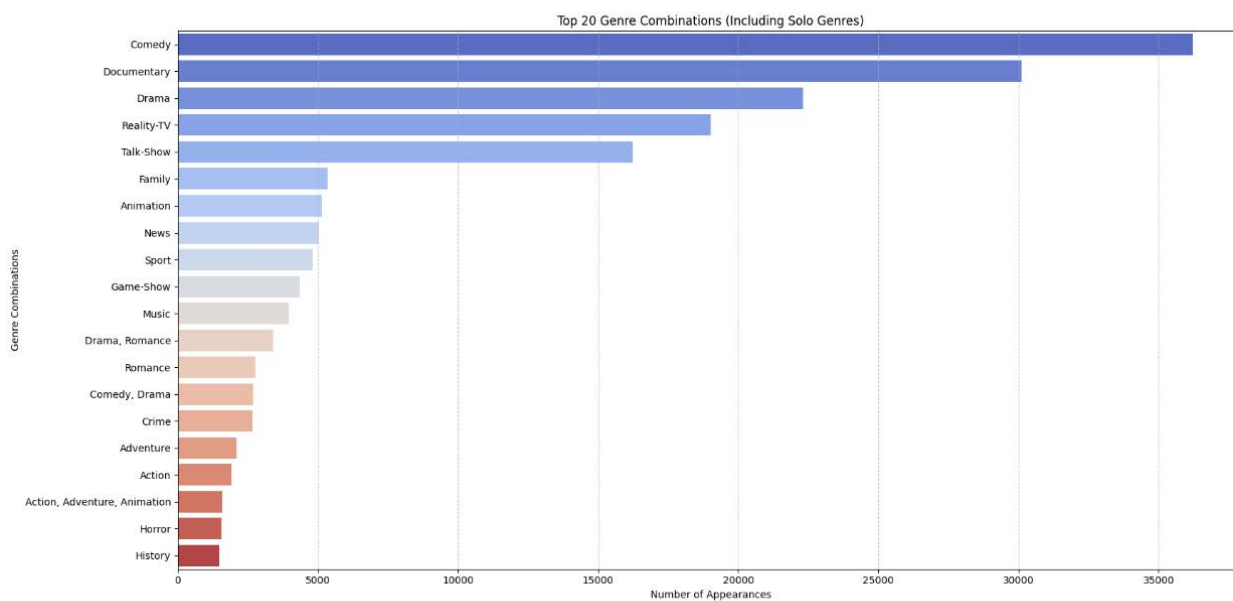


Figure 25 - Top 20 Genre Combinations (Including Solo Genres)

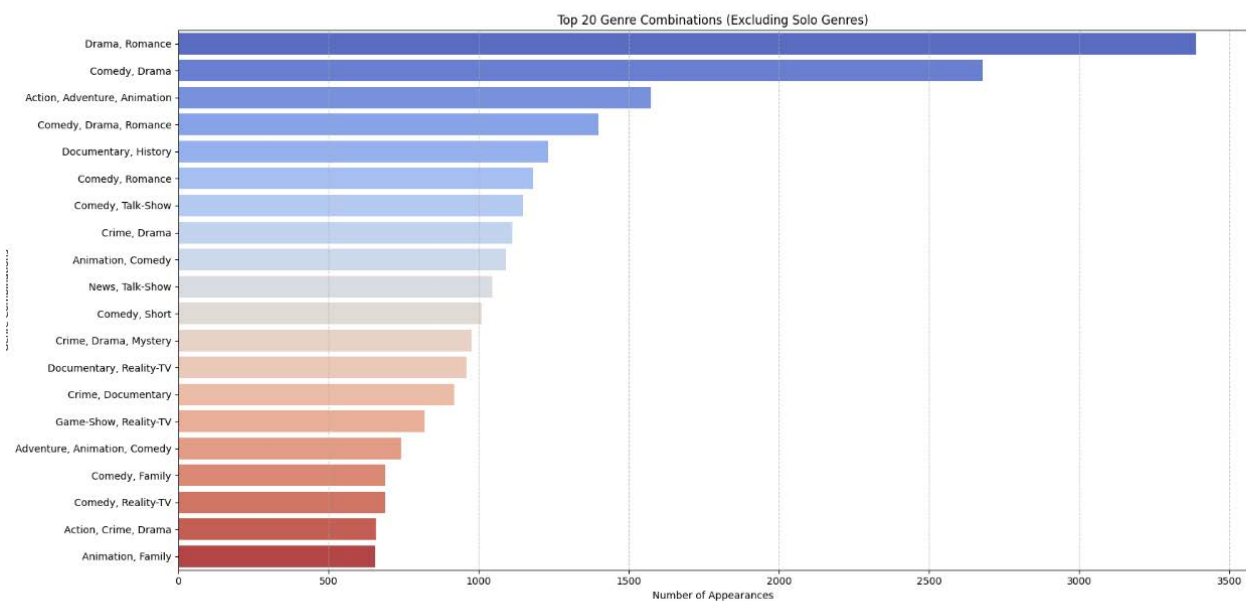


Figure 26 - Top 20 Genre Combinations (Excluding Solo Genres)

6) Conclusions and Research Contribution

The television industry has changed a lot over the course of the last three decades, going from TV to cable to being able to watch from anywhere, even from mobile phones. The way TV shows, and their numbers have evolved and grown shows that there was an impact by the rise of streaming platforms and evolving audience consumption patterns. My thesis' purpose was to shed some light on all these changes and highlight certain trends so that there's a better understanding of the industry's evolution.

After the metadata analysis, it seems that TV shows have changed during the three decades that were observed. This is based on the findings:

- While content had already been increasing year- after-year, the rate at which it went on was proliferated during the years of streaming and especially after the introduction of other platforms, competitors to Netflix. So, the need for diverse and high-quality content was growing more and more in order to attract subscribers globally. This could be assessed by further analyzing the number of TV shows that were released on streaming services vs. cable TV in the future.
- Quality (as measured by audience scores on IMDb) had already been dropping when Netflix launched its streaming service. After that year, a massive increase in TV shows' average ratings per year is observed, meaning that the increased numbers of TV series not only did not lower their quality but raised it instead, giving more options for people to watch. In the future, detailed analysis on the ratings of streaming platform originals vs cable/hybrid TV shows could be made to assess audiences' preferences.
- In combination with the increased quantity of TV Shows worldwide, the diversity of content has also grown as the unique number of countries participating in productions has been steadily growing. Also, while there have always been fewer unique languages than countries, the difference was minimized the last few years meaning that more countries produced TV shows in their own language and not just as co-producers with English-speaking countries. Though the ratings for "foreign" shows have been consistently lower than the USA productions, an increase can be observed during the years of streaming. Future research on this subject could include an assessment of whether the rise of streaming services helped in the quality of these shows and the exposure to worldwide audience.
- The modifications in the structure of TV shows can be seen in this thesis, as it appears that the series, on average, are getting smaller not only in the number of seasons, but in the number of episodes as well. Not only that but most streaming platforms emphasize on the release of entire seasons at once, something sparsely seen in the past years. This shift is something that changes entirely the way people consume content, as they get something smaller in length (we see TV shows with 6-10 episodes compared to the 20 episode per season series that were once popular) and they sit for hours in front of their TVs, laptops, and smartphones watching episode after episode until there's none left. This situation could be examined in future research to find out if it is the reason that binge-watching culture is so popular

and if producers have seen how much it's preferred by the audience (huge increase of 'same day' releases of episodes in 2021 and 2022).

- Audience engagement has also grown significantly because of this sudden explosion in content from streaming services. That is why we see rating votes on IMDb that are consistently higher than 600,000 annually for the last ten years. Platforms like IMDb give people the power to talk about their preferences and show the creators of TV shows what they did right or wrong. Further research could be conducted on how much people discuss about new TV shows online. Also to see whether this is the reason that leads the audience to leave ratings and reviews more easily than they would do in previous decades and if producers/writers choose to go the safe way in storytelling just to please the fans, this way making the TV show experience more interactive.
- Finally, an analysis in the top genres and their combinations has shown that not much has changed throughout the years. One thing that is for certain is the love people have for the 'Comedy' genre as it is the most appeared, both in combinations and on each own. However, recent trends should be considered, like how 'Western' has risen in popularity the last 13 years and how there are a lot of different styles and pairings combined in TV shows, giving more diverse options to people.

This thesis shows how important it is to use data to understand trends in the television industry, a practice utilized by both traditional media companies and modern streaming platforms through different methods like questionnaires, focus groups, and advanced data analysis techniques. The use of the Bayesian Average Ranking is also important in these kinds of analyses because it is more fitting of what is popular without biases that exist in traditional metrics. Of course, apart from future research to be conducted on the subjects already mentioned above, further work could be done using this metadata in addition to other data as well. For example, a financial analysis of TV shows throughout the years, what is spent and what is earned from streaming companies and how that affects the quality of the final product as well as the response from the audience. Also, a look at how Artificial Intelligence could change the way content is made in the future. There are lots of interesting things to analyze in this field to further contribute to an understanding of how technological advancements and changing audience behaviors can redefine the television industry.

Bibliography

- [1] NCTA, Cable's Story, Last Accessed 25 January 2025, <https://www.ncta.com/cables-story>
- [2] Valinsky, J. (2022, August 19). The cord has been cut. Streaming is more watched than cable. CNN Business. <https://www.cnn.com>
- [3] John P. Robinson, 2002, Television's Impact on Everyday Life: Some Crossnational Evidence, National Institute of Mental Health
- [4] Netflix, How to use Netflix on your Tesla display, Last Accessed 25 January 2025, <https://help.netflix.com/en/node/112323/us>
- [5] Bignell, J., & Woods, F. (2022). An Introduction to Television Studies (4th ed.). Routledge. <https://doi.org/10.4324/9781315619675>
- [6] Jet Stream, How Streaming Services Track and Use Your Personal Information, Last Accessed 25 January 2025, <https://jet-stream.com/personal-information-at-risk-how-streaming-services-track-and-use-your-data/>
- [7] Leiner, D. J., & Neuendorf, N. L. (2022). Does Streaming TV Change Our Concept of Television? *Journal of Broadcasting & Electronic Media*, 66(1), 153–175. <https://doi.org/10.1080/08838151.2021.2013221>
- [8] IMDb, Last Accessed 25 January 2025, <https://www.imdb.com>
- [9] Moral, Rafael & Chen, Zhi & Zhang, Shuai & McClean, Sally & Palma, Gabriel & Allan, Brahim & Kegel, Ian. (2022). Profiling Television Watching Behavior Using Bayesian Hierarchical Joint Models for Time-to-Event and Count Data. IEEE Access. 10.113018-113027. 10.1109/ACCESS.2022.3215682.
- [10] Cammarano, Maria & Guarino, Alfonso & Malandrino, Delfina & Zaccagnino, Rocco. (2024). TV shows popularity prediction of genre-independent TV series through machine learning-based approaches. Multimedia Tools and Applications. 83. 1-24. 10.1007/s11042-024-18518-z.
- [11] Lo, C. M., & Syu, Z. S. (2023). Analyzing drama metadata through machine learning to gain insights into social information dissemination patterns. *PloS one*, 18(11), e0288932. <https://doi.org/10.1371/journal.pone.0288932>
- [12] Butler, J. (2014). Statistical Analysis of Television Style: What Can Numbers Tell Us about TV Editing? *Cinema Journal* 54(1), 25-44. <https://dx.doi.org/10.1353/cj.2014.0066>.

- [13] Husna, Meryatul & Purba, Lampson & Rinaldy, Muhammad & Lubis, Arif. (2024). Predictive Analytics for IMDb Top TV Ratings: A Linear Regression Approach to the Data of Top 250 IMDb TV Shows. *Journal of Applied Informatics and Computing*. 8. 115-118. 10.30871/jaic.v8i1.7600.
- [14] Jang, M., Kim, D., & Baek, H. (2021). How do global audiences of TV shows take shape?: Evidence from Netflix. *Applied Economics Letters*, 30(3), 285–291. <https://doi.org/10.1080/13504851.2021.1983916>
- [15] Algolia, Using the Bayesian average in custom ranking, Last Accessed 25 January 2025, <https://www.algolia.com/doc/guides/managing-results/must-do/custom-ranking/how-to/bayesian-average/>
- [16] RapidAPI, IMDb8, Last Accessed 25 January 2025, <https://rapidapi.com/apidojo/api/imdb8>
- [17] PyPI, Requests, Last Accessed 25 January 2025, <https://pypi.org/project/requests/>
- [18] TV-API, Last Accessed 25 January 2025, <https://tv-api.com/>

