Dprep Final Report - The Nostalgia Effect: How Release Year and Genre Shape IMDb Movie Ratings Over Time

Team 4

2025-03-22

Contents

1 The Nostalgia Effect: How Release Year and Genre Shape IMDb Movie Ratings Over Time

Link to our GitHub page to have a broader overview of the whole project: https://github.com/course-dprep/team-project-spring-2025-team-4

1.1 Introduction

This project aims to explore how nostalgia influences IMDb movie ratings by analyzing the relationship between a film's release year and audience perception, with genre as a moderating factor. As nostalgia increasingly shapes consumer preferences, understanding its impact on movie reception can reveal biases in ratings and cultural memory effects. This study provides insights into how audiences evaluate films over time, helping industry professionals leverage nostalgia in content creation, marketing, and distribution strategies.

1.2 Motivation

In today's entertainment industry, maximizing audience engagement and profitability is crucial, and nostalgia has been identified as a key factor influencing movie reception (Ulker-Demirel et al., 2018). While studies

have explored whether nostalgia motivates audiences to watch remakes and reunions (Tetik & Türkeli, 2023) or influences viewer engagement (Andrade, 2023), little research has examined how nostalgia manifests in IMDb user ratings and whether it directly impacts the perceived quality of a movie. From an academic perspective, this study contributes to film studies, behavioral psychology, and consumer research by empirically analyzing rating patterns over time and assessing whether nostalgia bias or selective survival bias systematically influence IMDb ratings. Additionally, it examines how factors such as genre interact with this relationship, offering a deeper understanding of audience preferences and cultural memory in film evaluation.

From a real-world and managerial standpoint, this research has direct implications for film studios, streaming platforms, and marketers in shaping their content strategies. If older films receive higher ratings due to nostalgia rather than objective quality, this insight could influence decisions on which films to remake, re-release, or highlight in marketing campaigns. Streaming platforms can refine their recommendation algorithms by accounting for nostalgia-driven biases, improving user engagement and retention. Additionally, movie studios could leverage these findings to predict the potential success of sequels and franchise revivals, ensuring they appeal to audiences who value nostalgic elements. Understanding how genre moderates this effect also allows industry professionals to tailor their approach depending on whether certain film types (e.g., sci-fi, drama, or animation) are more influenced by nostalgia than others. Ultimately, this study provides a data-driven framework for maximizing audience appeal, optimizing content distribution, and making strategic investment decisions in the evolving film industry.

1.3 Research Question

To what extent is (i) IMDb movie user rating influenced by the (ii) movie released year? And to what extent does (iii) movie genre influence this relationship?

1.4 Context

Why IMDb? The Internet Movie Database (IMDb) is widely regarded as a valuable resource for both general users and researchers interested in film and television.
The platform offers extensive information on millions of films and television programs including details about cast and crew, plot summaries, and production notes. Moreover, it allows users to rate and review titles, fostering an interactive community where individuals can share opinions and discover new content based on collective feedback.
In addition, the website IMDb provides a variety of datasets that are valuable for research purposes. These datasets include information on movies, television shows, and video games, encompassing details such as cast and crew, ratings, reviews, release dates, and genres. Researchers can utilize these datasets to conduct in-depth analyses of entertainment content and audience interactions, which can inform business decisions, drive innovation, and advance scholarly research in the field.
Indeed, for the purpose of our research we will work on these datasets to answer our Research Question and, more specifically, we will focus on two of them: title.ratings.tsv.gz and title.basics.tsv.gz, that provide us all the necessary information for our analysis.

For this study, the amount of observations on which we will base our study is 51870 in the sample.

For data preparation, we conducted the study in following steps:
We started by retrieving and upload the IMDb datasets we needed (title.basics.tsv.gz and title.ratings.tsv.gz) directly from the public repository.
We automated the process using R scripts that handle downloads and extraction, ensuring reproducibility and minimizing errors from manual handling. During this stage, we also selected the essential variables to address our Research Questions.
From Title Ratings dataset we focused on:
averageRating, our primary dependent variable and
numVotes, to understand films' popularity.
While from Title Basics dataset we focused on:

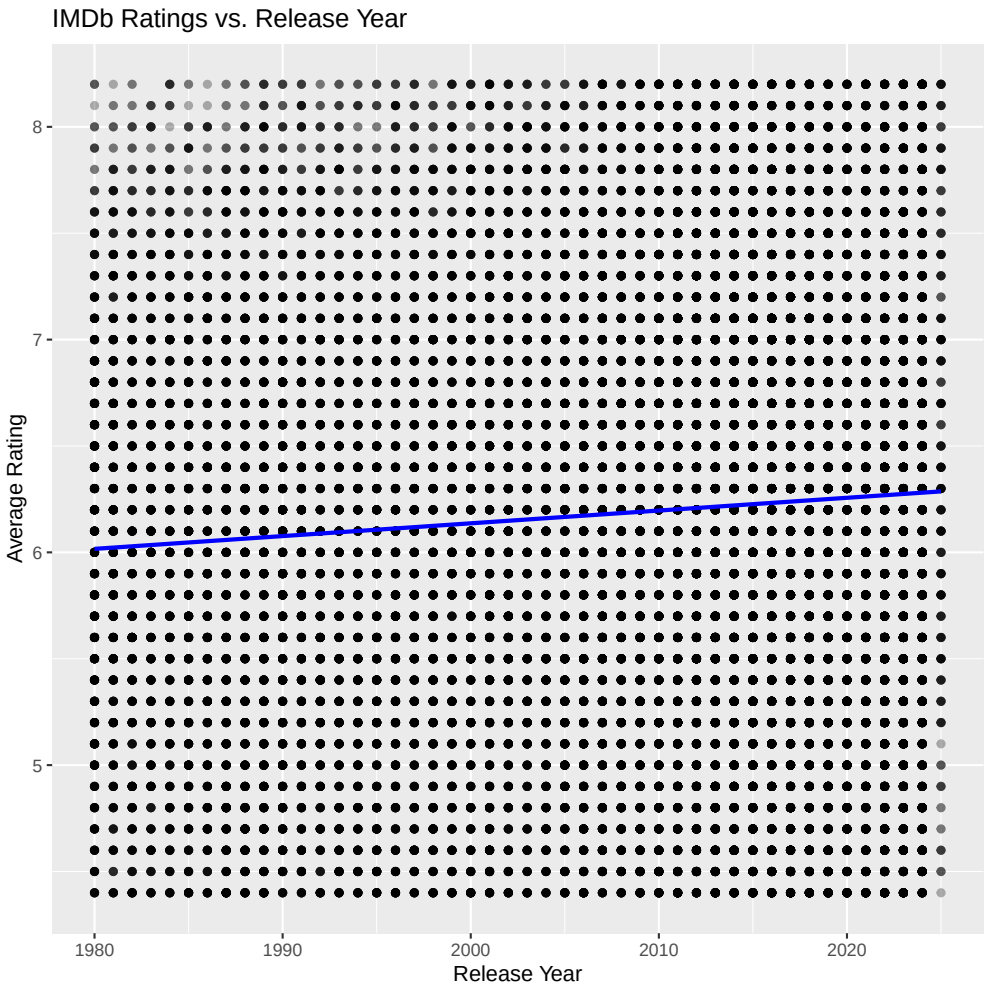
startYear, being the year of the film's release and that we hypothesized independent variable able to capture nostalgic sentiment overtime.
genres, our moderator variable, to investigate genre-specific moderating effects.
We chose these variables based on the academic premise that nostalgia may bias audience perceptions of film quality and that certain genres might be more susceptible to this phenomenon.

Next, we started the data cleaning process by focusing on: converting averageRating and numVotes into numeric data, removing from raw data possible missing values (e.g., titles with no valid release year or rating), and columns with variables we didn't need, we also filtered out future-dated entries. This step ensured that our dataset accurately represented actual releases and legitimate user feedback, thus strengthening the validity of our analysis. We then merged the cleaned tables on their unique identifier (tconst) to consolidate information on both film characteristics and user ratings in a single dataset.
To approach this, we implemented a full joyn merging, as all the variables we retained from the two datasets we relevant for our research.
Once merged, we did another cleaning on this final dataset by removing possible outliers on the averageRating variable implementing the IQR method, as we wanted to focus on the general trend.
Then, as the variable genres had multiple genre types (e.g., a film could be assigned to Horror, Thriller, Dramatic) creating a huge number of combined genres type, we kept only the first one indicated, in order to simplify and keep the research more clear.
An additional filtering on genres was made, by selecting only the most popular ones, by selecting only the ones with votes higher than the mean. Moreover, we focused our study on a subset of years, starting in 1980 until today.

1.5 Analysis

This analysis explores the evolution of IMDb ratings over time and across genres using a cleaned dataset of 51,870 films and TV movies. After filtering to include only the five most engaged genres—*Action*, *Adventure*, *Biography*, *Crime*, and *Horror*—based on having above-average vote counts (mean 23,934), we conducted a series of regression models to test various hypotheses. We began with a simple linear regression to assess whether ratings have increased over the years; results showed a statistically significant positive effect of release year on average rating (= 0.006, p < 2e-16), although the model's explanatory power was limited (R^2 0.0057).
To account for genre-specific trends, a second model included interaction terms between release year and genre. This revealed that the relationship between release year and rating varies significantly across genres—for example, *Adventure* and *Biography* showed steeper declines over time (interaction coefficients -0.006, p < 1e-8).
The model fit improved substantially (R^2 0.148), but also introduced high multicollinearity, as shown by VIF values around 168 for interaction terms—expected given the model structure.
An ANOVA test further confirmed that ratings differ significantly between genres ($F(4, 51865) = 2169$, p < 2.2e-16). To assess whether the apparent trend over time was driven by popularity bias, a third regression model controlled for the number of votes. The results showed that both startYear (= 0.0059) and numVotes (= 2.2e-6) were significant predictors, yet the year's effect remained stable, suggesting that the observed trend is not merely a reflection of title popularity.
Throughout the analysis, we checked key regression assumptions: Levene's test repeatedly rejected homoscedasticity (p < 2.2e-16), indicating non-constant variance in residuals, while Lilliefors (Kolmogorov–Smirnov) tests consistently rejected normality—both globally and within each genre. These violations suggest caution in interpreting p-values and support the consideration of robust or non-parametric models in future work.

IMDb Ratings vs. Release Year



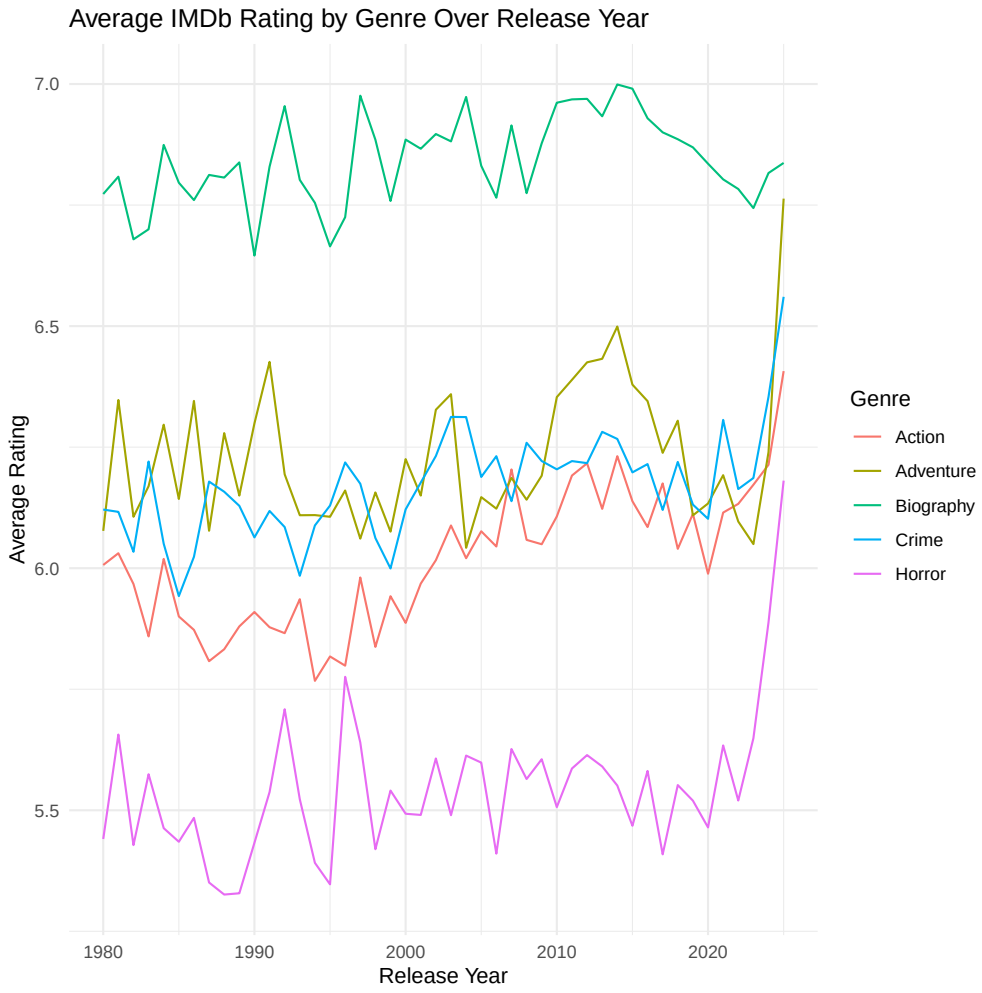
A final scatter plot with a linear fit visually confirmed the gradual upward trend in IMDb ratings over time. The findings indicate a small but statistically significant increase in ratings across decades, varying by genre, with minimal distortion due to popularity.

The densely packed black dots highlight the large dataset, with many movies released each year. Ratings are predominantly clustered between 5 and 8, indicating that most movies fall within this range. Highly rated movies (>8.5) are relatively rare, while very few films have ratings below 5—possibly because lower-rated movies tend to be less popular or less likely to remain relevant over time.

The findings indicate a small but statistically significant increase in ratings across decades, varying by genre, with minimal distortion due to popularity. However, model assumptions were not fully satisfied, suggesting caution in interpretation.

1.6 Results

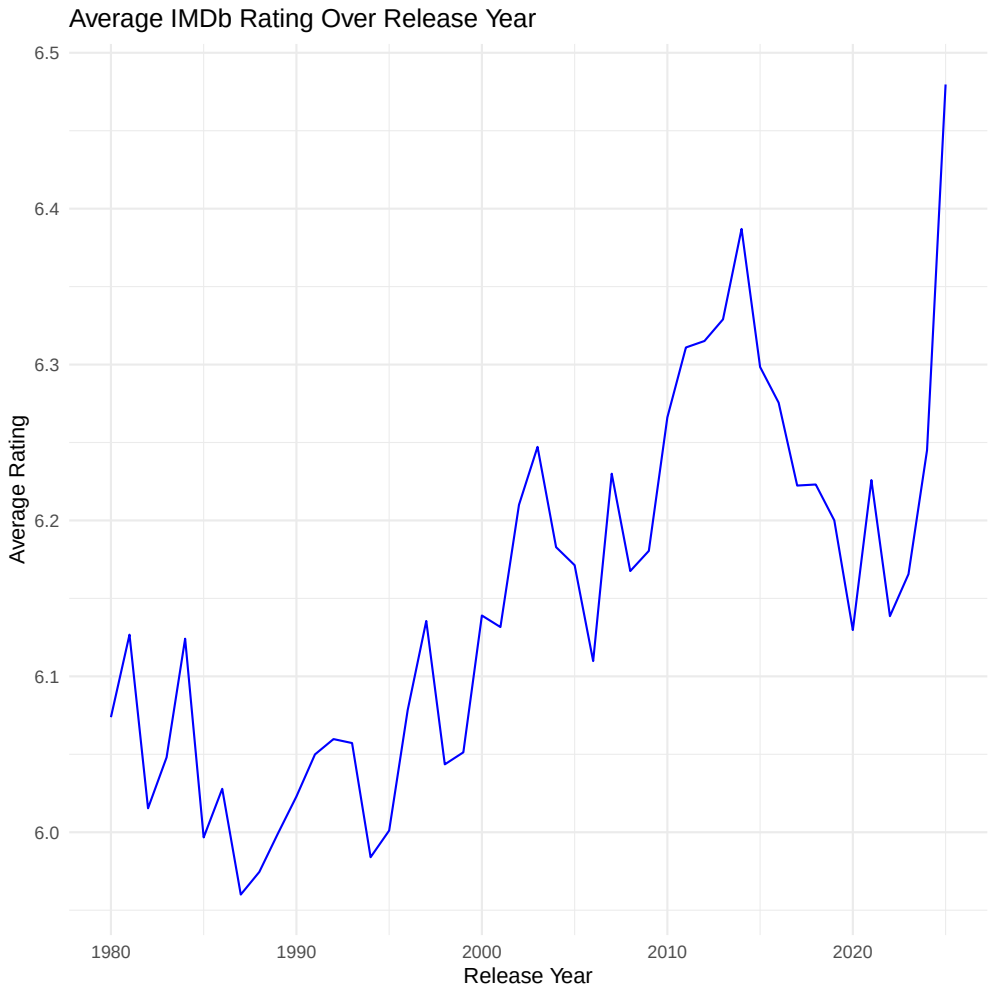
IMDb Ratings by Genre Over Release Year



The analysis of IMDb ratings by genre over time reveals that *Biography* films consistently receive the highest ratings, averaging close to 7.0, likely due to their strong narratives and character development, and critical acclaim, which result in more favorable audience reception. Biography films, in particular, often center around real-life figures or historical events, which could lead to higher engagement from audiences who appreciate realism and storytelling depth.

Crime and *Adventure* movies maintain an average IMDb rating between 5.8 and 6.5, showing greater fluctuation over time. The variability could be attributed to the impact of blockbuster hits or shifting audience preferences, to changing industry standards, evolving audience expectations, and the influence of major film franchises. Notably, action films have shown a gradual increase in ratings, reaching almost 6.5 in recent years, possibly reflecting advancements in special effects and improved storytelling. In contrast, horror films tend to receive the lowest ratings, often below 5.5, suggesting that audiences may be more critical of this genre or that the quality of horror films varies significantly. A general increase in IMDb ratings across all genres since 2020 suggests that streaming platforms, nostalgia-driven re-evaluations, and evolving user rating behaviors may contribute to this trend. Overall, the findings indicate that genre plays a significant role in shaping IMDb ratings, with biography films being the most well-received and horror films facing the most critical reception. The upward trend in ratings over time further supports the potential influence of nostalgia and audience perception shifts in film evaluation.

Average IMDb Rating Over Release Year



The trend in IMDb ratings over time reveals distinct patterns across different periods. Between 1980 and 2000, ratings fluctuated from 6.0 to 6.5, suggesting that while there is a trend, it does not dramatically reshape the perception of movies over time and it's likely due to shifts in audience preferences and changes in IMDb's rating system. Some films from this era have maintained high ratings due to cult status, while others have experienced a decline in reception.

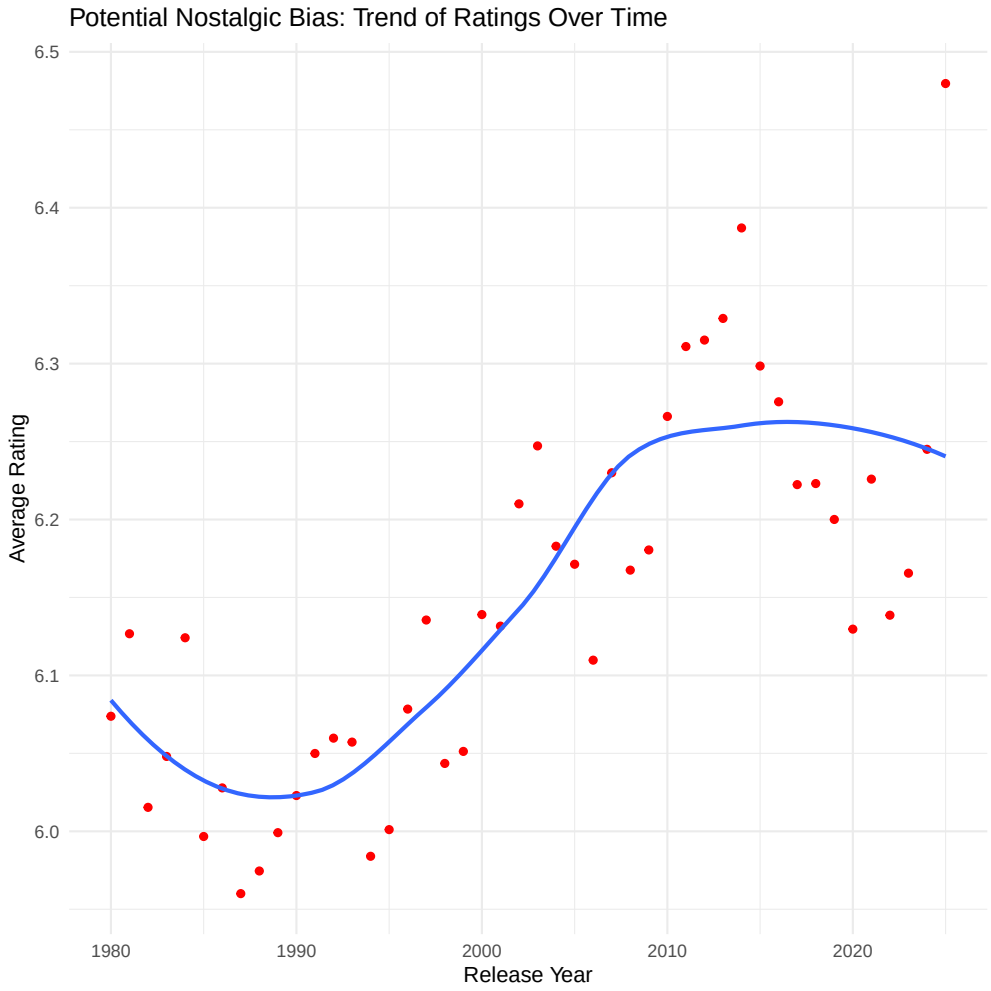
Some sharp declines (e.g., early 1990s and mid-2020s) suggest that certain periods may have had a larger share of lower-rated films or that audience expectations changed. There is a spike in rating after 2000, a more consistent upward trend emerges, potentially influenced by improvements in filmmaking techniques and the growing impact of streaming services.

The increased accessibility of films has contributed to a broader and more engaged voting audience, leading to more stable and possibly higher ratings. Moreover, cyclical trends where films of certain eras gain renewed appreciation due to nostalgia, leading to a reevaluation over time. Older movies might experience rating inflation due to nostalgic viewers revisiting and rating them more favorably years later, whereas newer movies are rated by audiences with fresh, unfiltered perspectives.

Since 2020, IMDb ratings have surged across various film categories, with potential contributing factors including the nostalgia effect, where older films are reassessed more positively, as well as the role of streaming platforms in making films more accessible and widely reviewed. Additionally, social media influence has played a role in shaping audience perception and generating hype around certain films.

Overall, the analysis suggests that newer movies tend to receive higher IMDb ratings, but this trend may be influenced by changes in audience behavior, nostalgia bias, and streaming accessibility. Future research could explore whether this pattern continues or stabilizes as more films are added to the IMDb database.

Trend of IMDb Ratings Over Time



This analysis explores how IMDb ratings have evolved over the decades, with a particular focus on the role of nostalgia bias in shaping audience perception. The graph presents a nonlinear trend in average IMDb ratings over time, deviating from a simple upward trajectory. Instead, we observe a U-shaped pattern, suggesting that certain eras of film are rated more favorably than others, potentially due to nostalgia-driven biases. While one might expect a steady increase in ratings due to improvements in filmmaking techniques, the data implies that audience memory, selective rewatching, and cultural context play significant roles in rating behavior. The average IMDb rating has shown a steady increase from 1980 to 2025, with notable fluctuations over time. A slight decline is visible in the late 1980s, followed by a consistent rise starting in the 1990s. This drop may reflect shifting audience expectations, changes in industry trends, or a lack of strong nostalgic attachment to films from this period.

As we move into the late 1990s and 2000s, ratings begin to rise sharply, peaking in the 2010s. This era aligns with the childhood and adolescence of many IMDb users today, suggesting that nostalgic recall is inflating ratings for films from this period. People tend to romanticize films from their youth, and as streaming services like Netflix and Disney+ make older films more accessible, audiences may revisit and rate these movies more favorably than they would have upon first release.

The plateau or slight decline in ratings post-2015 further supports the nostalgia hypothesis. Recent films are evaluated in real-time, without the benefit of nostalgic hindsight, leading to harsher assessments compared to films that have had decades to earn their place in cultural memory. Furthermore, modern franchise fatigue and sequel comparisons may contribute to lower ratings for contemporary films, as audiences judge them

against well-loved originals from the past.
However, the sharp rise in ratings after 2020 raises another point of consideration. The COVID-19 pandemic altered media consumption patterns, with audiences revisiting comfort films and selectively engaging with content they were predisposed to enjoy. The rise of streaming platforms has also allowed audiences to rediscover and appreciate older films, potentially contributing to higher ratings. Changes in filmmaking quality, including advancements in technology, storytelling, and cinematography, could explain the overall upward trend in ratings. Additionally, audience behavior and social media influence may shape movie ratings, as online discussions, reviews, and viral trends create shifts in public perception. Overall, the U-shaped trend in the graph suggests that nostalgia significantly shapes the perception of film quality, making older movies seem better in hindsight, while more recent films face greater scrutiny and lack the advantage of retrospective appreciation.

1.7 Conclusion

This study examined how nostalgia influences IMDb ratings over time by analyzing the relationship between a film's release year and audience perception, with genre acting as a moderating factor. Our findings indicate that while IMDb ratings have generally increased over time, this trend is not linear; instead, a U-shaped pattern emerges, suggesting that nostalgia plays a crucial role in shaping how films are remembered and evaluated.

The results indicate that movies from the late 1990s and early 2000s receive higher ratings, likely benefiting from nostalgic reassessment by IMDb users who grew up during that period, while newer films (post-2015) face immediate scrutiny and do not experience the same retrospective boost. Genres also plays a moderating role, especially for certain types, such as *Biography* and *Crime*, maintain consistently high ratings, suggesting that these genres may be less affected by shifts in audience preferences and more resilient to the influence of nostalgia, but not queally for all the genres. Indeed, *Horror* movies that consistently receive lower ratings.

Another key observation is the recent peak in IMDb ratings since 2020. This pattern could be attributed to the increased accessibility of films through streaming platforms, the re-evaluation of older movies through a nostalgic lens, and the influence of social media in shaping public perception.

Our analysis confirms that nostalgia significantly impacts IMDb ratings, influencing how films from different eras are evaluated, with genre-specific effects further shaping this trend.
These findings have practical implications for film studios, streaming platforms, and marketers looking to leverage nostalgia in content curation and promotional strategies. Future research could further explore how nostalgia interacts with generational differences in rating behavior, the impact of critic reviews versus user ratings, and how digital accessibility continues to reshape audience engagement with cinema over time.
Ultimately, IMDb ratings should not be interpreted solely as indicators of film quality but as reflections of cultural memory, emotional attachment, and the evolving ways in which audiences engage with movies across decades.

1.8 References

• Tetik, T., & Türkeli, Ö. (2023). POPULAR CINEMA AS NOSTALGIA INDUSTRY: REUNIONS, REMAKES, AND REQUELS. Sinecine: Sinema Araştırmaları Dergisi, 14(1), 7-31. https://doi.org/10.32001/sinecine.1253910
• Ulker-Demirel, E., Akyol, A. and Simsek, G.G. (2018), "Marketing gitand consumption of art products: the movie industry", Arts and the Market, Vol. 8 No. 1, pp. 80-98. https://doi.org/10.1108/AAM-06-2017-0011
• Andrade, S. (2023). The impact of nostalgia and social setting on consumer motivation to watch a movie at the cinema vs. through streaming. Handle.net. http://hdl.handle.net/10400.14/42261