User Behavior Behind Movie Ratings

Project 1 : Storytelling through EDA

Harsh Srivastava

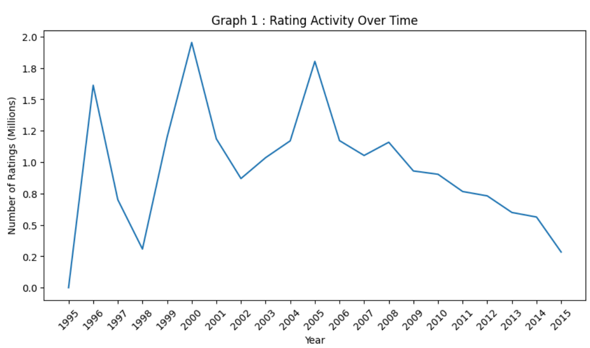
Overview

MovieLens is a web-based recommender system and virtual community that recommends movies for its users to watch, based on their film preferences. The datasets describe ratings and free-text tagging activities from MovieLens, a movie recommendation service. It contains 20000263 ratings and 465564 tag applications across 27278 movies. These data were created by 138493 users between January 09, 1995 and March 31, 2015. This dataset was generated on October 17, 2016.

Problem:

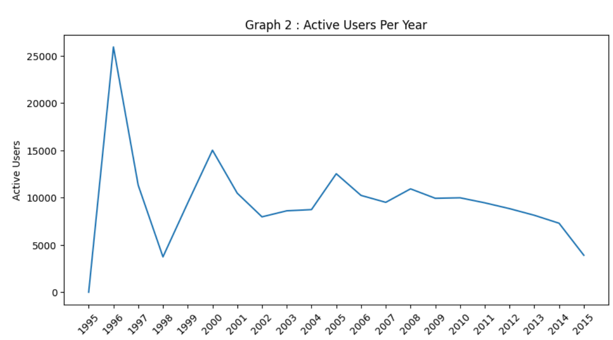
"**How has user engagement on MovieLens evolved over time?**"

Early Growth and the Expansion Phase



Graph 1 shows the number of ratings per year from 1995 to 2015, revealing an explosive rise in activity during the late 1990s. This period corresponds to the early adoption phase of MovieLens, when **online movie communities were novel and participation barriers were low**. Ratings increased sharply through the mid-to-late 1990s, peaking around the year 2000.

This growth is mirrored in Graph 2, which shows the number of active users per year. The late 1990s exhibited a surge in new users, culminating in a peak around 1996–1997. The alignment between active users and total ratings suggests that **early growth was driven primarily by scale**: more users naturally led to more ratings, rather than existing users becoming significantly more active.



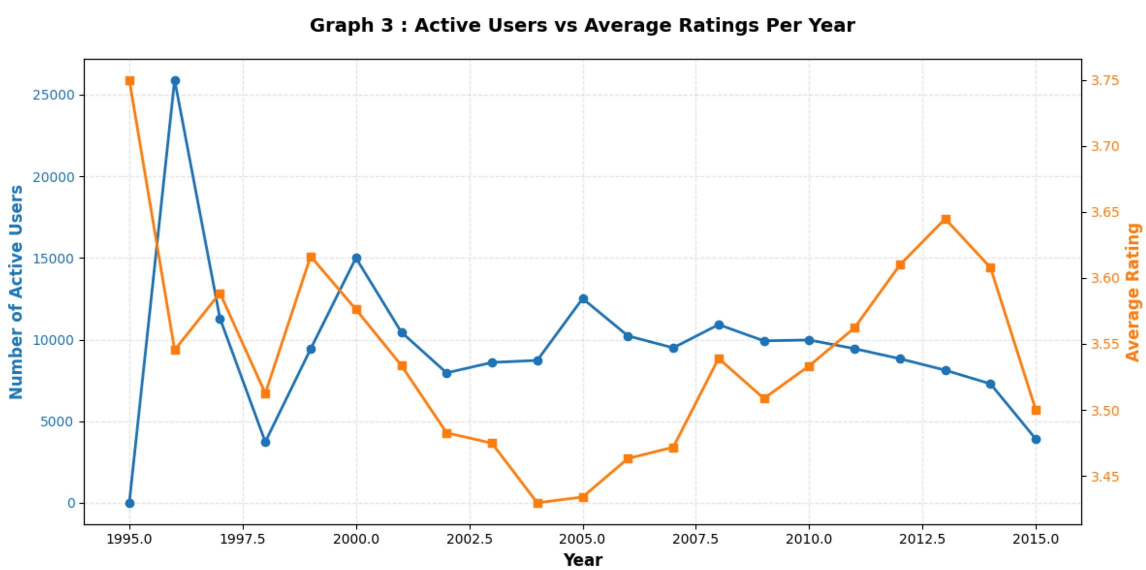
Interestingly, both graphs reveal noticeable dips around 1997 and 1998. At first glance, these downturns appear puzzling, especially given the overall upward trend of internet adoption during this period. However, later genre-level analysis show no corresponding disruption in content preferences, suggesting that **these dips are unlikely to be content-driven**. Prior research on the MovieLens dataset shows that **rating behavior and selection bias evolve over time as items age**, indicating that long-term engagement and preference patterns are not static but shift as the platform matures and as users' interaction habits change. This temporal evolution aligns with the observed dip in ratings and tag activity in later years.[1]

From Peak Activity to Stabilization

Following its early peak, ratings activity entered a more volatile but generally declining phase after 2001.

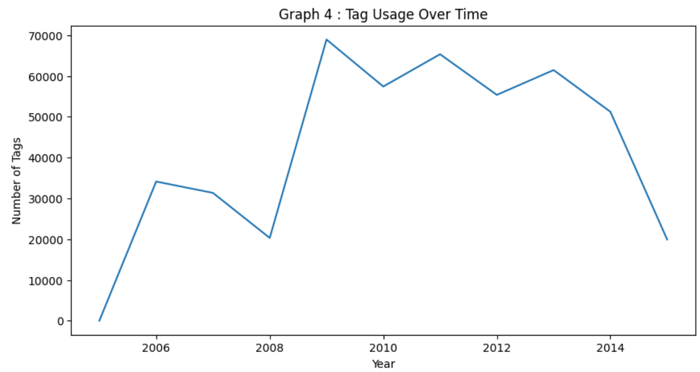
This decline, however, should not be interpreted as a simple loss of interest. The platform does not collapse but rather, it stabilises. Participation becomes less explosive but more consistent. This interpretation is reinforced by **Graph 3**.

Graph 3 shows periods of higher user counts often correspond to lower average ratings, while periods with fewer users show marginally higher averages. This suggests that **broader participation may introduce more critical or diverse opinions**, whereas **smaller communities tend to be more selective** in what they choose to rate.

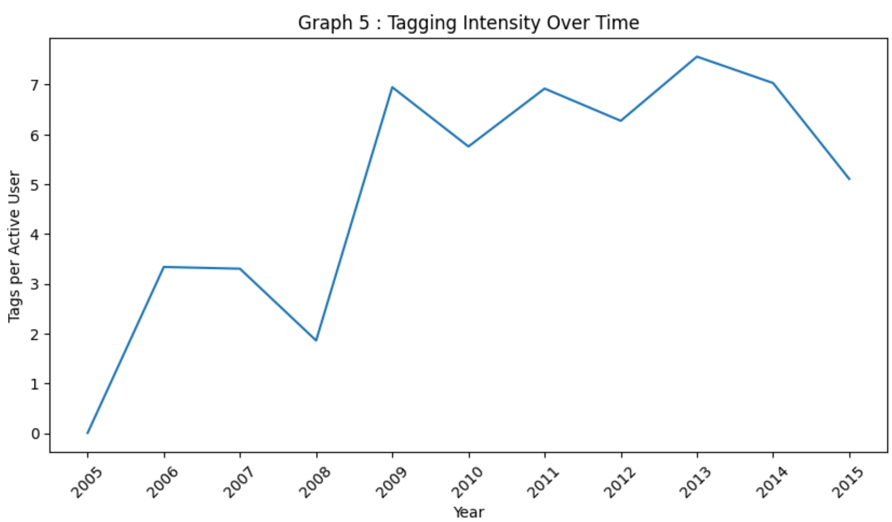


The Emergence of Tagging as a New Mode of Engagement

While ratings dominate the early years, tagging introduces a fundamentally different form of interaction.



Graph 4 shows tag usage over time, revealing that meaningful tagging activity begins only after 2005. This delay confirms that tagging was introduced as a later feature rather than a core component of the original platform design. Once introduced, tags grow rapidly peaking between 2009 and 2013 before declining again.



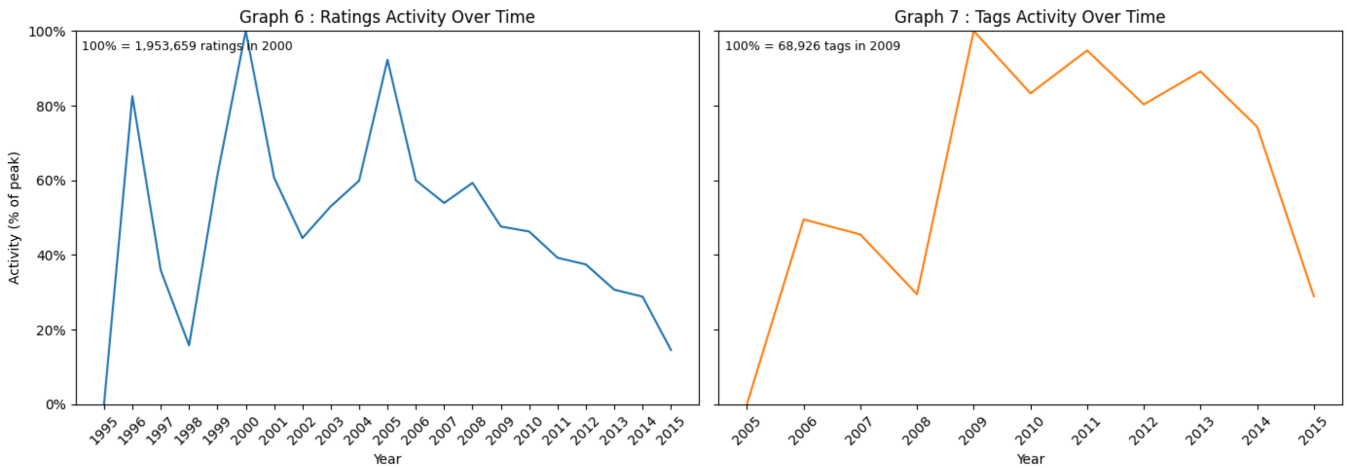
At face value, the decline in total tag counts after 2013 mirrors the decline in ratings and users. However, Graph 5 provides a deeper insight by normalising tag counts by the number of active users. This graph shows that tags per active user actually increase during the later years of the dataset.

This indicates a qualitative shift in user behaviour. **Tagging becomes a tool for expression and interpretation rather than mass participation as users provide richer metadata than simply assigning numerical ratings**.

Problem:

"**Has tagging completely overtaken ratings**?"

Ratings and Tags: Complementary, Not Competing



Graph 6 and 7 places ratings and tags on the same normalised scale, highlighting how the two interaction modes evolve relative to one another. Ratings dominate the early years almost entirely, while tags gain prominence in the mid-to-late 2000s.

Importantly, tagging does not replace rating activity but instead rises alongside a gradual decline in rating volume. This suggests that as users become more selective in how often they rate movies, they compensate by engaging more thoughtfully when they do interact. The platform evolves from a **high-throughput evaluation system into a more nuanced annotation environment**.

This transition reflects a broader pattern seen in mature online communities, where early participation is driven by novelty and scale, while **later participation emphasises depth, expertise, and expressiveness**.

A Long-Term Evolution of Engagement

When viewed holistically, the MovieLens dataset reveals three distinct phases:

1. **Expansion (1995–2000)**
   Rapid user growth, high ratings volume, exploratory participation.

2. **Stabilisation (2001–2007)**
   Slower growth, consistent rating norms, emergence of tagging.

3. **Selective Engagement (2008–2015)**
   Declining scale but increasing depth, higher tagging intensity per user.

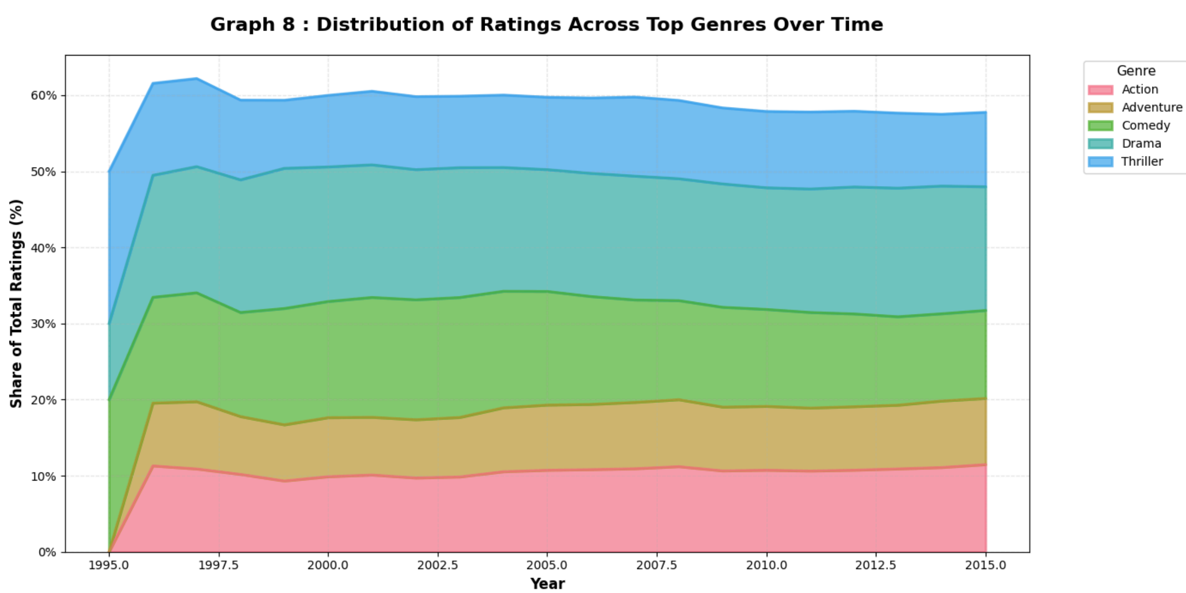
Rather than depicting a platform in decline, the data portrays a system maturing alongside its users.

<u>**Problem**</u>:

“**Has the user’s taste changed over time?**”

Temporal-Shifts in Genre-Level Rating Composition

A key observation from Graph 8 is the stability in genre composition over time, particularly for Drama and Comedy, which consistently account for a substantial proportion of total ratings. Despite fluctuations in overall user activity and rating volume observed in earlier analyses, the relative prominence of these genres remains largely intact. This suggests that while the number of users and ratings may rise or fall,the **foundational content preferences of the audience exhibit long-term persistence**.



Importantly, these patterns highlight that **platform-level volatility does not necessarily translate into proportional changes in content preference**. The genre ecosystem appears to absorb changes in user behaviour while maintaining a relatively stable internal structure.

The Taste Tribes: Understanding Who Our Users Really Are

Raw averages hide behaviour. A 3.5-star rating can mean very different things depending on *who* is doing the rating and *how often*. To move beyond surface-level trends, we segmented users based on two fundamental behavioural signals:

- **How frequently they rate (engagement intensity)**
- **How generously they rate (sentiment bias)**

This revealed three distinct **Taste Tribes** which are essentially *behavioural personas* that consistently interact with movies in different ways.

The Taste Tribes Landscape:

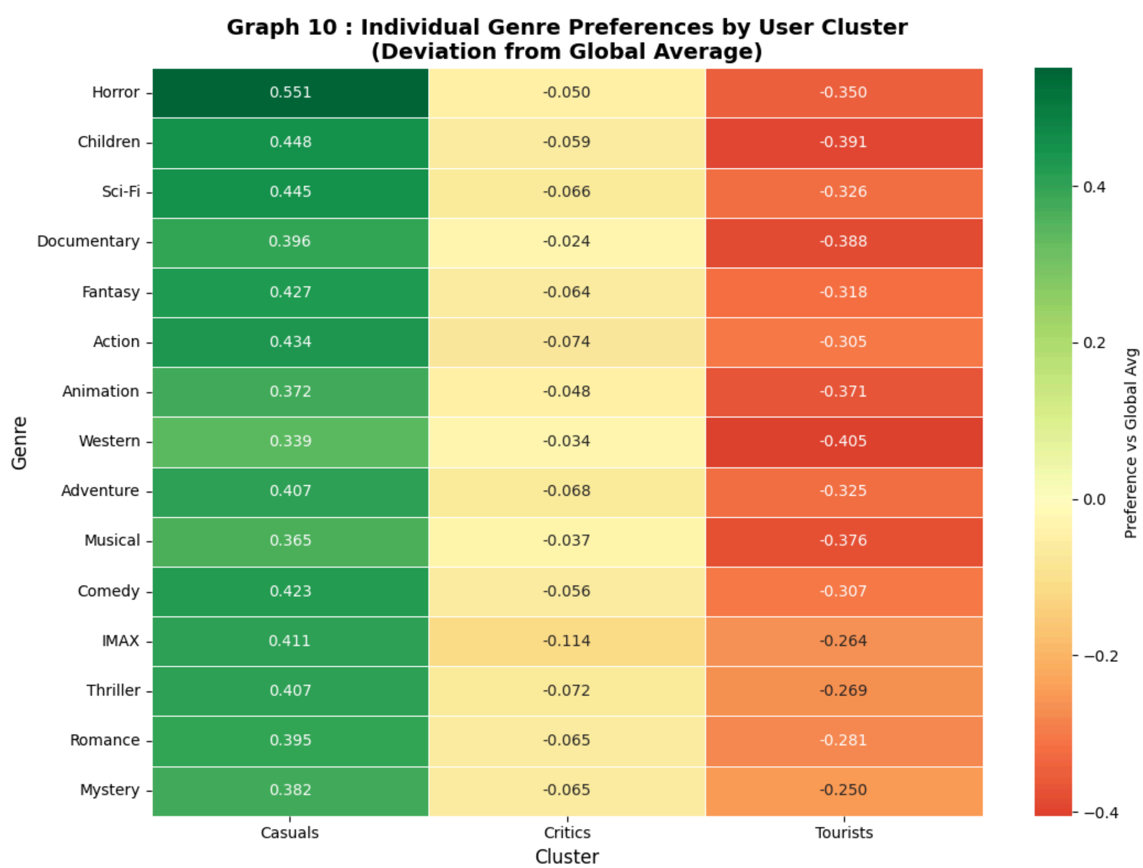
We can classify users in three different tribes according to the Average Rating they give and how frequently they rate. These tribes are:

- **Casuals** : Users who have low-moderate activity and give consistently high ratings. They engage when they like something, not to critique it.
- **Critics** : Users who have very high activity and give moderate to lower ratings. They watch broadly, rate frequently and apply stricter standards.
- **Tourists** : Users who have low activity and give low average ratings. They dip in and out of the platform and mostly rate when they have to show unmet expectations.

<u>Problem</u>:

"**Do different tribes systematically prefer different genres**?"

Genre Preferences by Tribe (Deviation from Global Average)



From Graph 10,we can see how different tribes rate different genres

Casuals :

- Consistently rate all major genres above the global average,showing broad-based enjoyment rather than selective taste.
- Show strongest positive preference for **Horror, Sci-Fi, Action, Fantasy, and Adventure**.
- Signal **high satisfaction when expectations are met**.

Critics **:**

- Ratings remain **close to the global average** across genres.
- Exhibit **mild negative deviation** instead of strong likes or dislikes.
- Reflect a more **evaluative and comparative** viewing mindset.

***Tourists*:**

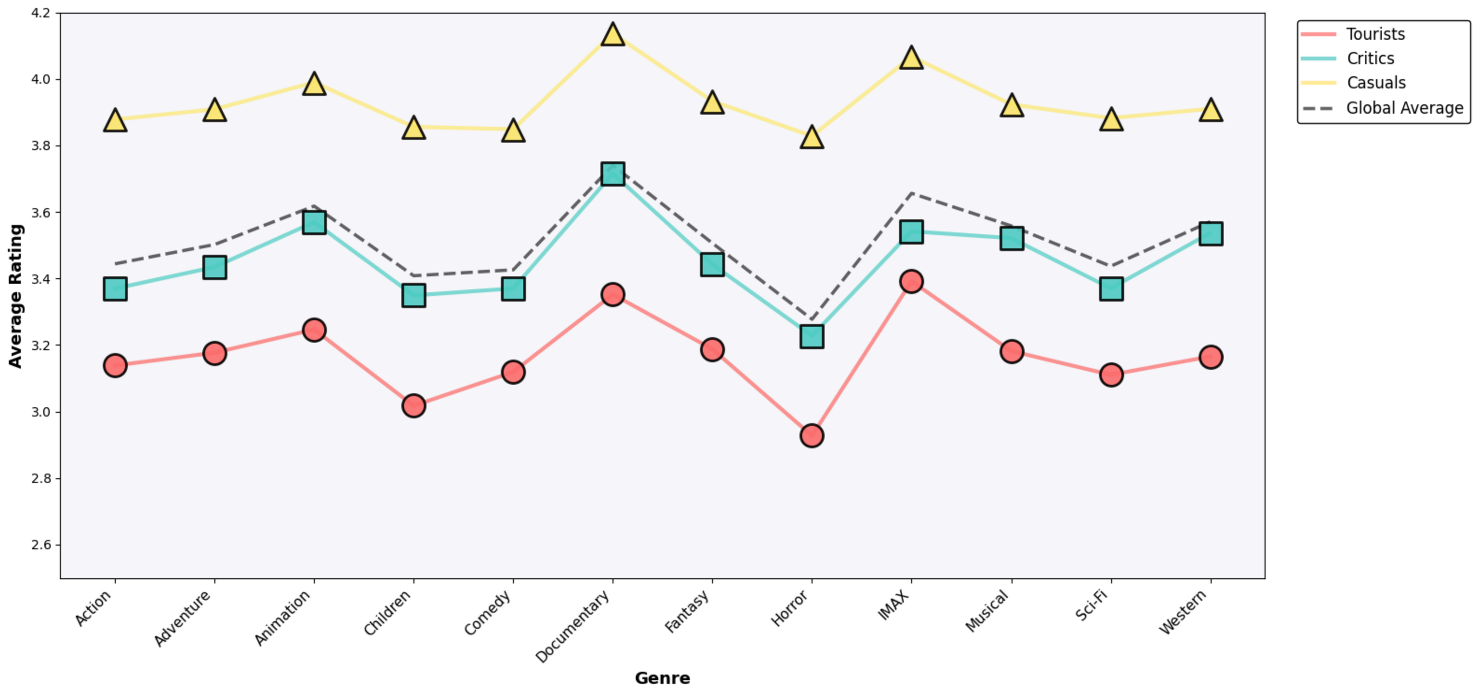
- Rate almost every genre **below the global average**.
- Strongest negative deviation in **Children, Documentary, Western, and Horror**.
- Lack of clear genre affinity suggests **disengagement** rather than taste-driven choice,indicating low commitment rather than genre-specific aversion.

Problem:

"Are we hyper-focusing on active users while misunderstanding real audience satisfaction?"

Aggregate Ratings Reflect Activity More Than Satisfaction

Graph 11: How Each Cluster Rates Genres vs Global Average



Graph 10 shows that ratings are systematically influenced by user behavior rather than representing a uniform measure of satisfaction. **Casual users consistently rate genres above the global average**, while **Critics,who contribute most of the ratings,cluster at or below the average across genres. Tourists rate infrequently and tend to score lower, further depressing aggregate values.**

This pattern explains why overall ratings are often pulled toward the preferences of highly active users, even when they are not the most satisfied group. In the context of this analysis, the graph confirms that **aggregate ratings overweight critical voices and underrepresent quieter but more positive users**, reinforcing the need to interpret ratings through the lens of user tribes rather than as standalone indicators.

Key Takeaways:

- **User ratings are not neutral signals**: they systematically reflect differences in user behavior, not just content quality.
- **Highly active users dominate aggregate ratings**, despite being more critical on average, leading to a downward bias in perceived satisfaction.
- **Less active users consistently rate more positively**, indicating that satisfaction is often underrepresented in summary metrics.
- **Genre performance cannot be interpreted in isolation**: the same genre is experienced very differently by different user tribes.
- **Declines or plateaus in average ratings do not necessarily imply declining quality** : they may reflect shifts in audience composition or platform maturity.

Future Plans:

- **Cross-validation of clustering :** Validate tribe definitions using alternative feature sets (e.g., rating variance, genre diversity).
- **Temporal genre shifts by tribe :** Examine whether genre preferences within tribes change over time or remain stable.
- **User lifecycle analysis :** Analyze how rating behavior evolves from a user's first rating to later stages of engagement.

References :

- Harper, F. M., & Konstan, J. A. (2015). *The MovieLens Datasets: History and Context.* ACM Transactions on Interactive Intelligent Systems.
- GroupLens Research. *MovieLens 20M Dataset.* University of Minnesota.

- Nielsen, J. (2006). *Participation Inequality: Encouraging More Users to Contribute.* Nielsen Norman Group.
- Preece, J., & Shneiderman, B. (2009). *The Reader-to-Leader Framework: Motivating Technology-Mediated Social Participation.* AIS Transactions on Human-Computer Interaction.
- Golder, S. A., & Huberman, B. A. (2006). *Usage Patterns of Collaborative Tagging Systems.* Journal of Information Science.
- Sen, S., et al. (2006). *Tagging, Communities, Vocabulary, Evolution.* Proceedings of the ACM CSCW Conference.

- Adomavicius, G., & Tuzhilin, A. (2005). *Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art.* IEEE Transactions on Knowledge and Data Engineering.
- Huang, J., Oosterhuis, H., & de Rijke, M. (2022).*It Is Different When Items Are Older: Debiasing Recommendations When Selection Bias and User Preferences Are Dynamic.* Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining (WSDM ’22)