

Report

Features Created and Their Purpose

Six new features were engineered from the original dataset to deepen understanding of user and movie behaviours:

1. **Release Year:** Extracted from the movie title (e.g., "Toy Story (1995)") to analyse trends by movie age.
2. **Movie Age:** Calculated as the difference between the current year and release year. Helps explore how a movie's age affects its rating.
3. **Genre Count:** The number of genres per movie (e.g., "Action|Comedy" = 2). Indicates diversity of a movie's themes.
4. **Main Genre:** The first listed genre (e.g., "Action|Comedy" → "Action"). Simplifies comparison of ratings by genre.
5. **Rating Year:** Extracted from timestamps to analyse how ratings evolve over time.
6. **Average Movie Rating:** The mean rating for each movie, useful for identifying overall popularity and quality.

Key Insights from Exploratory Analysis

- **Overall Ratings:** Ratings are mostly high, with 4.0 being the most common value, suggesting a positive bias among users.
- **Movie Age Trends:** Older movies (pre-1920s) received higher ratings, indicating strong appreciation for classics.
- **Top and Bottom Movies:** "Gena the Crocodile (1969)" and "The Act of Killing (2012)" were among the highest-rated, while "Tooth Fairy 2 (2012)" and "Anaconda: The Offspring (2008)" had the lowest ratings.
- **Genre Patterns:** Movies with more genres (7–10) tend to receive higher ratings, possibly because they appeal to broader audiences.
- **Rating Trends Over Time:** Ratings peaked between 2011–2013 and dipped slightly around 2017, showing mild fluctuations in audience behaviour.
- **Genre Differences:** "Film-Noir" movies are rated highest, while "Horror" films receive lower ratings, suggesting some genres are consistently favoured.

Implications for a Movie Recommendation System

The engineered features and insights can strongly support the development of a personalized recommendation system:

- **Release Year and Movie Age:** Can help recommend older or newer films based on a user's preference trends.
- **Main Genre and Genre Count:** Allow clustering users and movies by dominant genres or diversity of interests.
- **Average Movie Rating:** Serves as a popularity metric to recommend highly rated movies to new users (cold-start problem).

Report

- **Rating Year:** Enables time-based filtering, so the system prioritizes more recent, relevant movies.

Together, these features improve personalization, helping the recommendation engine understand not only what users like, but also *why* they like it — leading to smarter, context-aware suggestions.

A few visualizations/tables that support your findings

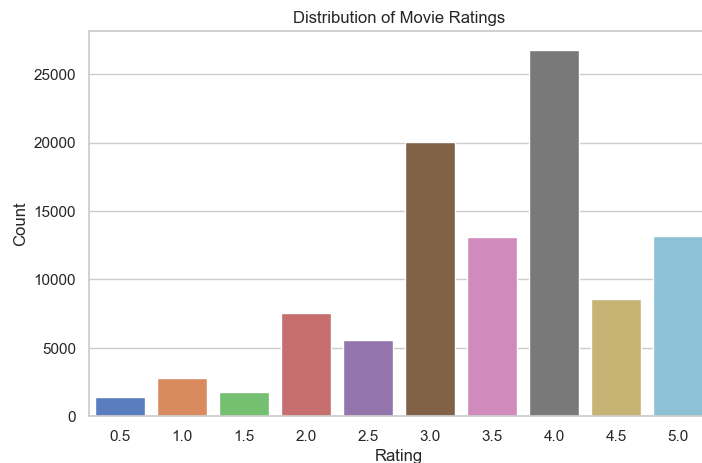
1. Do people tend to rate movies high or low?

My Observation:

From the bar chart, I can see that the rating value of **4.0** occurs most frequently among users, followed by 3.0 and 3.5.

My Conclusion:

This suggests that people generally **tend to rate movies quite high**, showing a positive bias toward the films they watch.



2. Do newer movies get higher ratings?

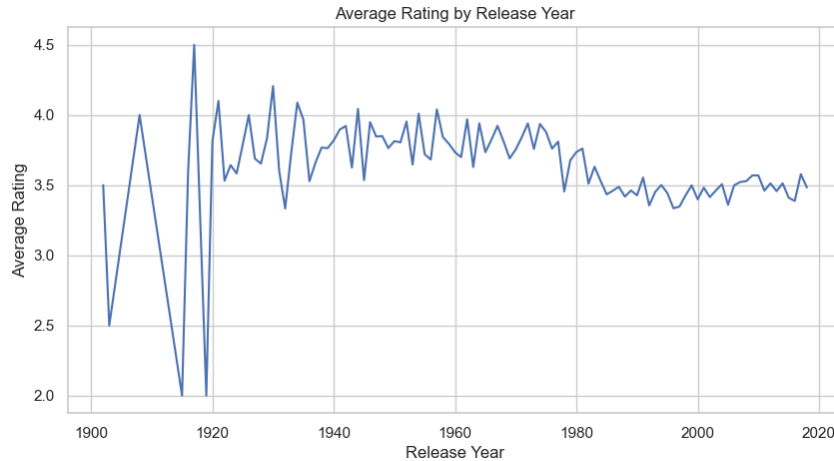
My Observation:

The line chart of average ratings against release year shows that **older movies**, especially those from the **late 1910s and before the 1920s**, have **very high ratings** compared to newer ones.

My Conclusion:

This means that **classic or older movies are often rated higher**, possibly because they are viewed as iconic, nostalgic, or of high cultural value.

Report



3. Which movies have the highest and lowest average ratings?

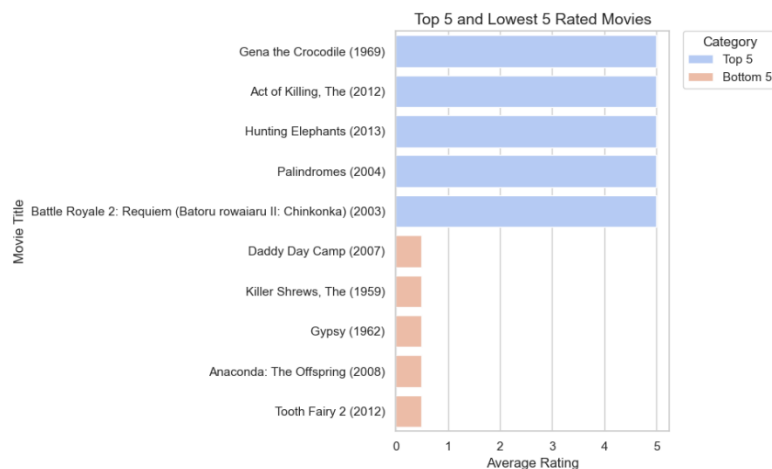
My Observation:

The movie **Gena the Crocodile (1969)** has the highest average rating, followed by **The Act of Killing (2012)**.

Meanwhile, **Tooth Fairy 2 (2012)** and **Anaconda: The Offspring (2008)** have the lowest average ratings.

My Conclusion:

This suggests that **critically acclaimed or classic films** receive higher ratings, while **sequels** tend to receive lower ratings from users.



4. Do movies with more genres tend to have higher or lower ratings?

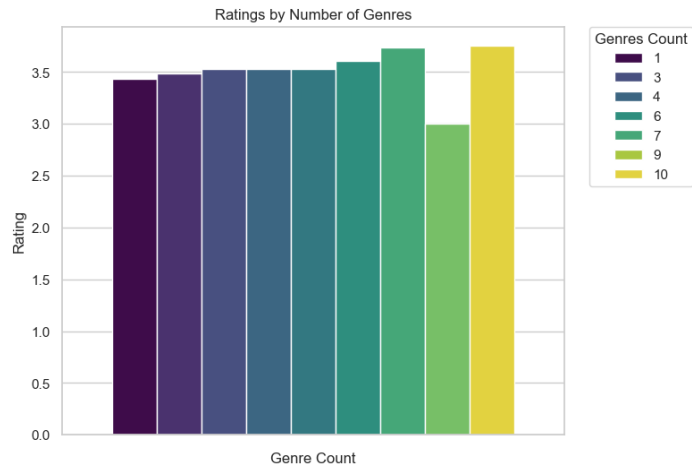
My Observation:

The bar chart shows that movies with **7 or 10 genres** have relatively higher ratings, while those with **9 genres** have the lowest ratings overall.

Report

My Conclusion:

Generally, **movies with more genres tend to attract higher ratings**, possibly because they appeal to a broader audience by blending multiple themes — though this trend is not always consistent.



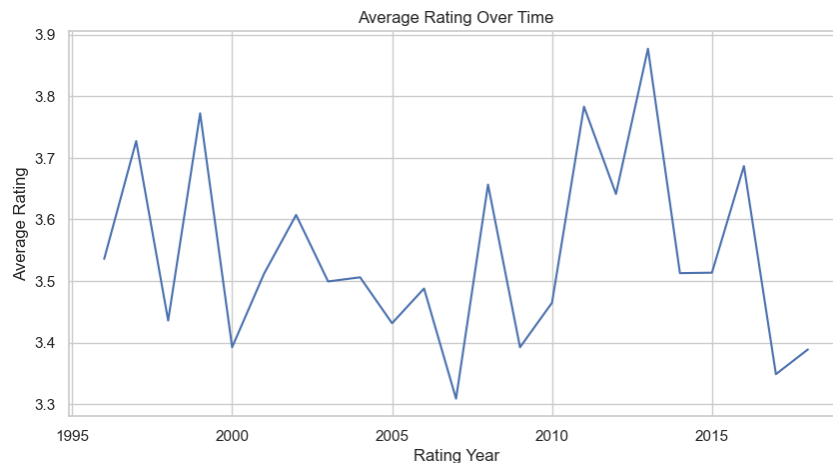
5. Does rating behaviour change over time?

My Observation:

The line chart shows that ratings **peaked around 2011–2013**, and then remained relatively stable over the years, with a **slight drop around 2017**.

My Conclusion:

This indicates that **audience rating behaviour changes slightly over time**, possibly influenced by movie quality trends or shifts in viewer expectations during different periods.



Report

6. Do certain genres tend to have higher or lower ratings?

My Observation:

The bar chart reveals that **Film-Noir** movies have the **highest average ratings**, while **Horror** movies receive the **lowest**.

My Conclusion:

This shows that **more artistic or story-driven genres like Film-Noir** tend to receive higher ratings, while **genres like Horror**, which are often **less popular or predictable**, tend to be rated lower on average.

