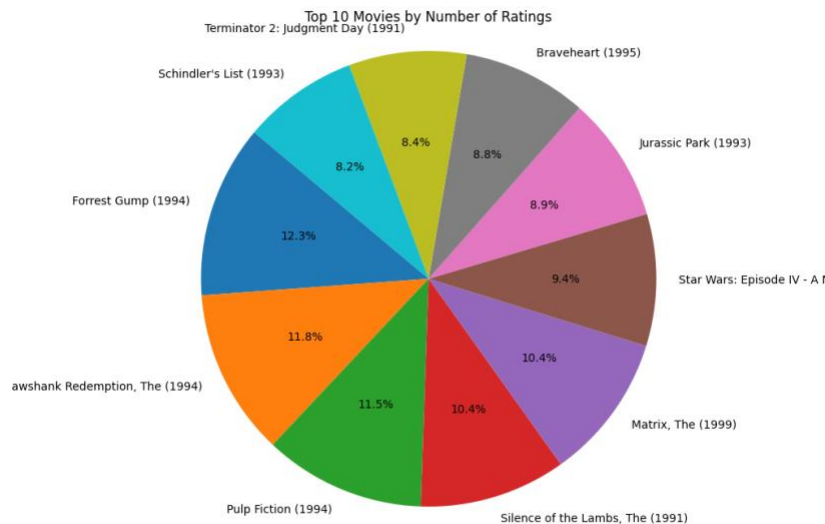


# Data Analysis Report

To identify factors influencing user movie rentals, the analysis can focus on the following aspects:

- **Movie Popularity and Ratings:** Analyzing the frequency of ratings and average rating scores for each movie can reveal popular and highly rated movies.
- **Genre Preferences:** Understanding which genres are most popular among users.
- **User Behavior:** Investigating patterns in how different users rate movies.
- **Tag Analysis:** Examining the tags to understand user perceptions and preferences about movies.

## Analyzing movie popularity and ratings:

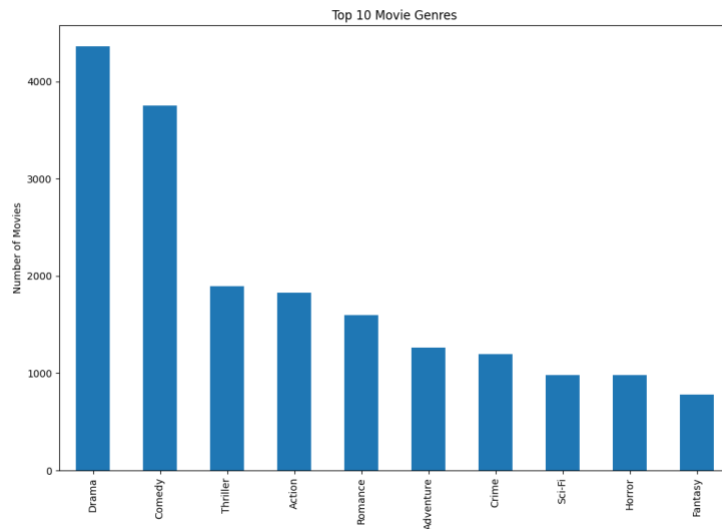


The top 10 movies in the MovieLens dataset, based on the number of ratings they received, are as follows:

- **Forrest Gump (1994)** - Average Rating: 4.16, Number of Ratings: 329
- **Shawshank Redemption, The (1994)** - Average Rating: 4.43, Number of Ratings: 317
- **Pulp Fiction (1994)** - Average Rating: 4.20, Number of Ratings: 307
- **Silence of the Lambs, The (1991)** - Average Rating: 4.16, Number of Ratings: 279
- **Matrix, The (1999)** - Average Rating: 4.19, Number of Ratings: 278
- **Star Wars: Episode IV - A New Hope (1977)** - Average Rating: 4.23, Number of Ratings: 251
- **Jurassic Park (1993)** - Average Rating: 3.75, Number of Ratings: 238
- **Braveheart (1995)** - Average Rating: 4.03, Number of Ratings: 237
- **Terminator 2: Judgment Day (1991)** - Average Rating: 3.97, Number of Ratings: 224
- **Schindler's List (1993)** - Average Rating: 4.23, Number of Ratings: 220

These movies are not only popular in terms of the number of ratings they have received, but also generally have high average ratings, indicating that they are well-received by the audience. This suggests that the popularity and reception of a movie (as indicated by the number of ratings and the average rating) could be significant factors influencing user movie rentals.

Next, let's analyze genre preferences.

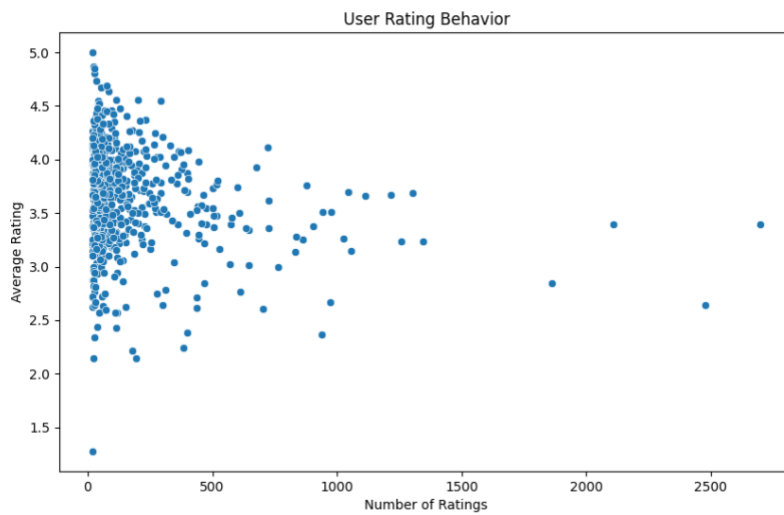


The top 10 genres in the MovieLens dataset, based on the number of movies in each genre, are as follows:

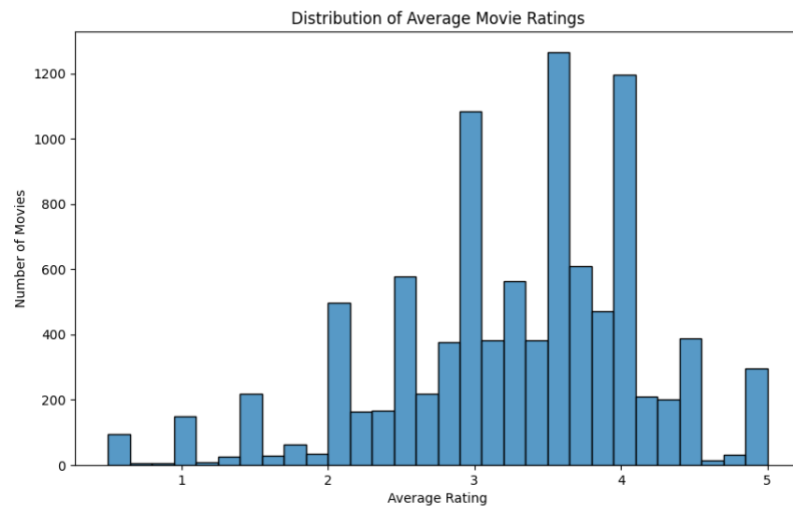
- **Drama:** 4,361 movies
- **Comedy:** 3,756 movies
- **Thriller:** 1,894 movies
- **Action:** 1,828 movies
- **Romance:** 1,596 movies
- **Adventure:** 1,263 movies
- **Crime:** 1,199 movies
- **Sci-Fi:** 980 movies
- **Horror:** 978 movies
- **Fantasy:** 779 movies

This genre distribution indicates that Drama and Comedy are the most common genres. This could suggest that these genres are popular among users, potentially influencing their rental choices.

Now, let's explore user behavior by analyzing rating patterns.



Areas with a higher density of points indicate common behavior among a larger group of users. Sparse areas suggest less common behavior.



Above distribution indicates the distribution of movie ratings.

The top 10 users in the MovieLens dataset, based on the number of ratings they have given, along with their average rating, are as follows:

- **User 414:** 2,698 ratings, Average Rating: 3.39
- **User 599:** 2,478 ratings, Average Rating: 2.64
- **User 474:** 2,108 ratings, Average Rating: 3.40
- **User 448:** 1,864 ratings, Average Rating: 2.85
- **User 274:** 1,346 ratings, Average Rating: 3.24
- **User 610:** 1,302 ratings, Average Rating: 3.69
- **User 68:** 1,260 ratings, Average Rating: 3.23
- **User 380:** 1,218 ratings, Average Rating: 3.67
- **User 606:** 1,115 ratings, Average Rating: 3.66
- **User 288:** 1,055 ratings, Average Rating: 3.15

These users are very active in terms of rating movies, indicating that user engagement (frequency of ratings) can be a significant factor in understanding user preferences and movie rental behavior.

Additionally, the variation in their average ratings highlights differing user tastes and preferences, suggesting the importance of personalized recommendations in influencing movie rentals.

Finally, let's briefly look at the tags data to understand user perceptions of movies.

The top 10 most common tags used by users in the MovieLens dataset are:

- **In Netflix queue:** 131 occurrences
- **Atmospheric:** 36 occurrences
- **Thought-provoking:** 24 occurrences
- **Superhero:** 24 occurrences
- **Funny:** 23 occurrences
- **Surreal:** 23 occurrences
- **Disney:** 23 occurrences
- **Religion:** 22 occurrences
- **Sci-fi:** 21 occurrences
- **Quirky:** 21 occurrences

These tags reflect user perceptions and interests in movies. For example, the high frequency of "In Netflix queue" suggests a direct interest in viewing these movies. Tags like "atmospheric", "thought-provoking", and "funny" indicate specific qualities that users might seek in movies. Understanding these tags can help in creating more accurate recommendations and understanding the nuanced preferences of different user groups.