



Movie Recommender System

Sheetal Munjewar

Bellevue University



Introduction & Problem Statement

- The rise in content across entertainment and e-commerce platforms has made it difficult for users to find relevant products or movies.
- A recommender system helps sort content based on user preferences, preventing users from feeling lost in a sea of options.

Problem Statement Continue ...

The volume of content increasing on entertainment/media platforms.

The number of choices available is increasing rapidly on e-commerce sites.

Recommender system sorts a large list of products/contents as per user's taste/earlier choices.

Without recommendations users might get lost in the volume of content availability.

Good recommendations keep a user engaged for the long term.

Good recommendations attract users to products/contents of his/her choice.

Types of Recommender System

Content-Based
Recommender System

Collaborative filtering-based
Recommender System

Popularity-Based
Recommender System

Hybrid Recommender
System



Major Steps in Building a Recommender

Data Understanding

Data Cleaning and Preparation

Building a Simple Model with Weighted Ratings

Content-Based Model

Hybrid Model (Weighted Ratings + Content-Based)

Collaborative Filtering-Based Model

Hybrid Model (Content-Based + Weighted Rating + Collaborative Filtering)

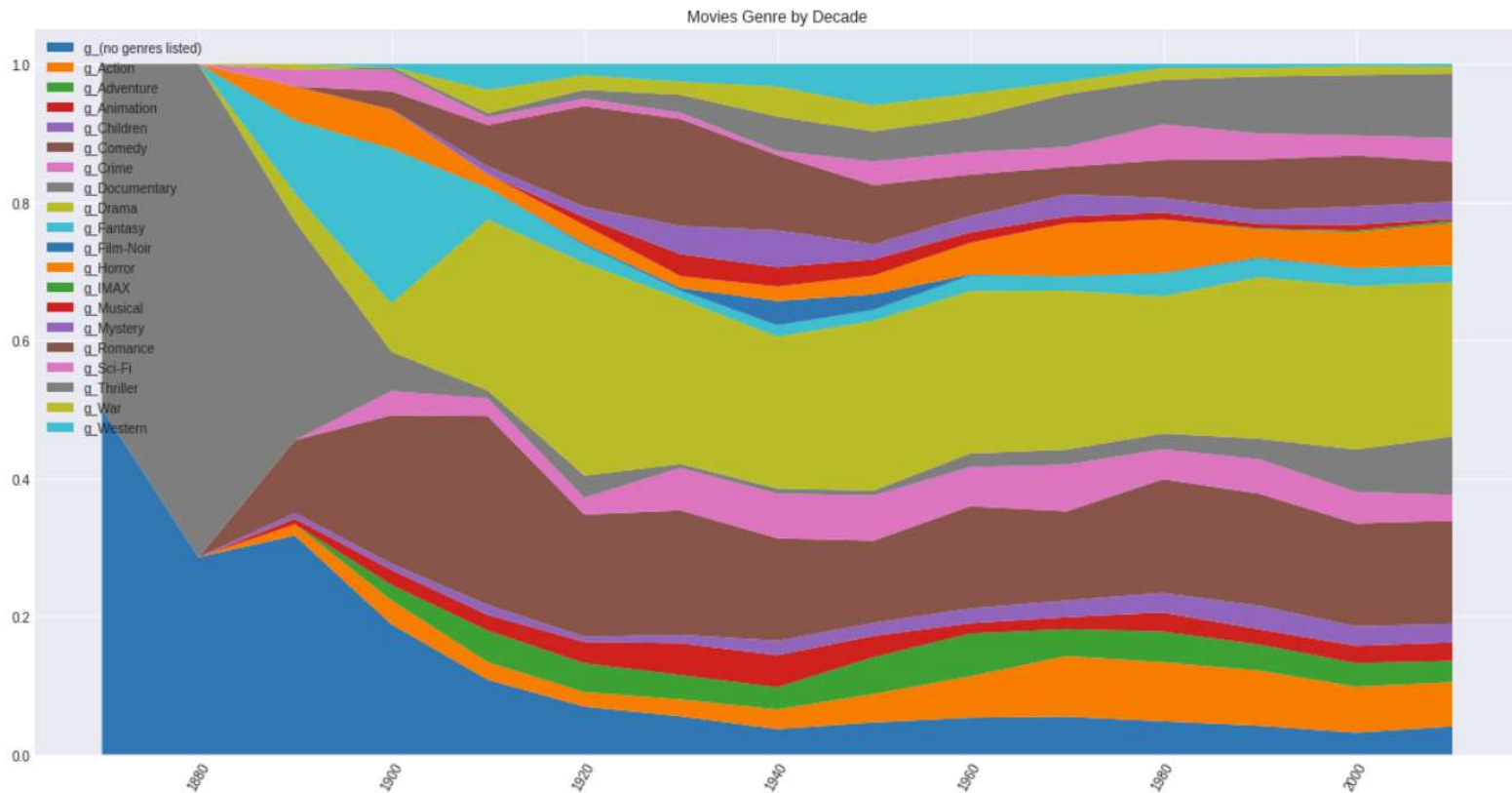
Dataset

Movie rating dataset taken from grouplens.org

Ratings dataset contains 25 million movie ratings given by 162K users for 62K movies

The movie metadata dataset taken from [Kaggle.com](https://www.kaggle.com).

The metadata dataset contains metadata about 45K movies.

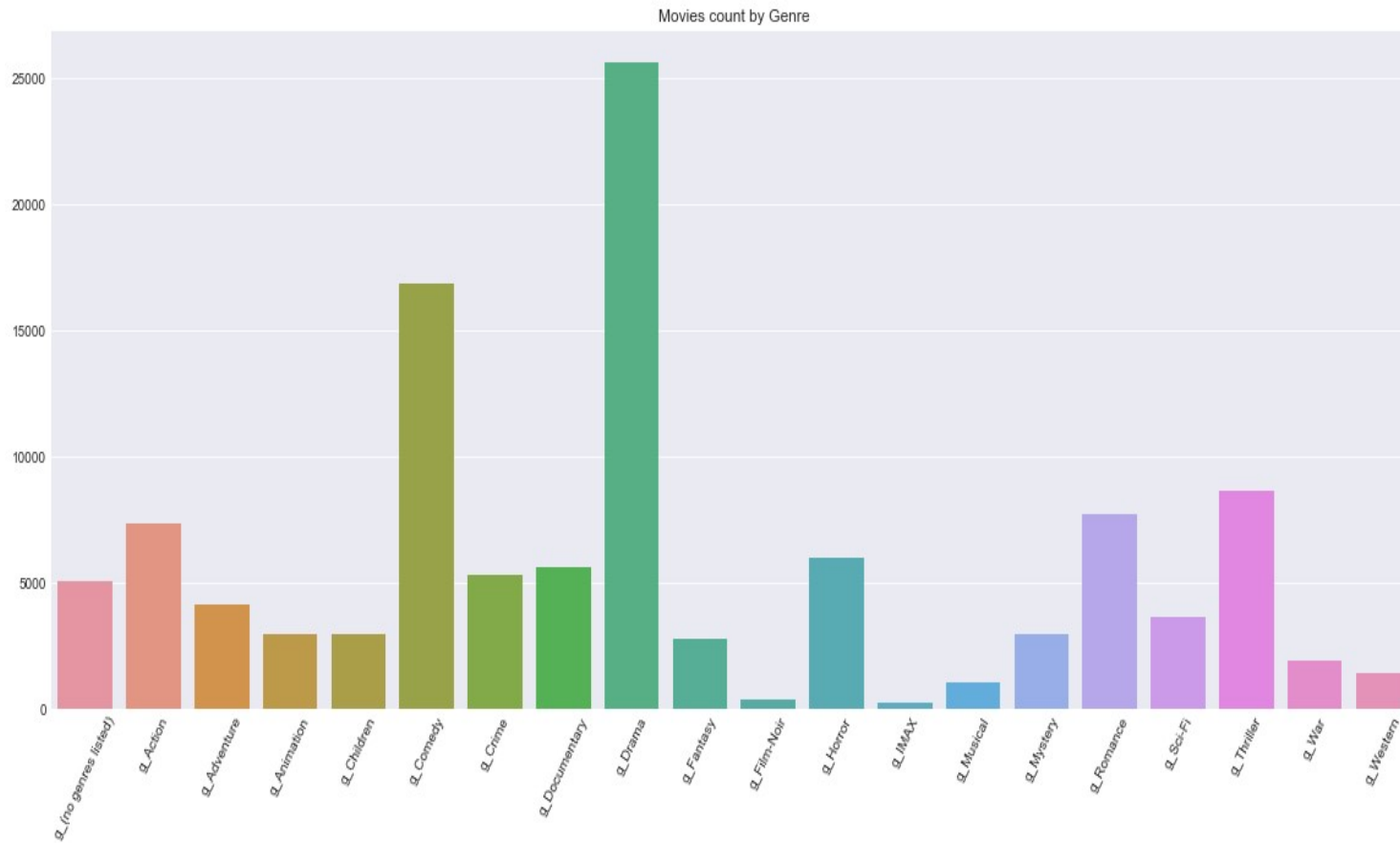


Exploratory Data Analysis

DRAMA AND ROMANCE MOVIES ARE MAJOR PART THROUGH MOVIE HISTORY

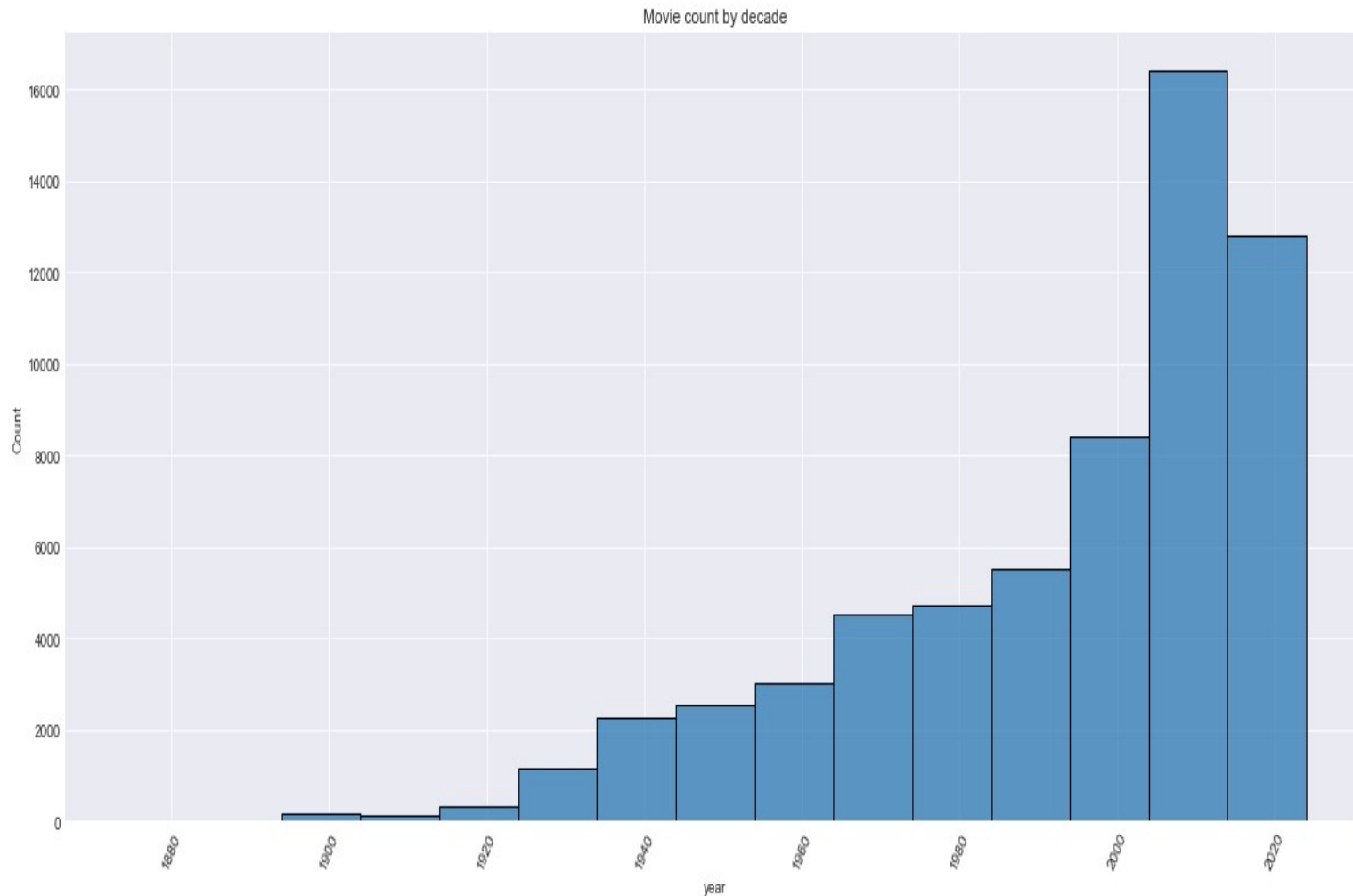
Exploratory Data Analysis

➤ Movies are from different genres



Exploratory Data Analysis

➤ Movies dataset contains movies from 1885-2020



Data cleaning and preparation

- Ratings column converted to int8
- Converted “movieId” column from ratings dataset and “id” column from metadata dataset to int for linking
- Create year column from release date
- Filled null values with empty strings

Weighted-Rating based recommender

➤ Problem with Rating based model

➤
$$\text{Weighted Rating (WR)} = \left(\frac{v}{v+m} \cdot R\right) + \left(\frac{m}{v+m} \cdot C\right)$$

- v is the number of rating for the movie
- m is the minimum number of rating required to be listed in the chart
- R is the average rating of the movie
- C is the mean rating across the whole report

Top 10 movies by weighted rating

```
build_top_movie_chart(metadata, percentile=0.95, no_of_movies=10)
```

	title	year	vote_count	vote_average	popularity	id	weighted_rating
314	The Shawshank Redemption	1994	8358.0	8.5	51.645403	278.0	8.357746
834	The Godfather	1972	6024.0	8.5	41.109264	238.0	8.306334
12481	The Dark Knight	2008	12269.0	8.3	123.167259	155.0	8.208376
2843	Fight Club	1999	9678.0	8.3	63.869599	550.0	8.184899
292	Pulp Fiction	1994	8670.0	8.3	140.950236	680.0	8.172155
351	Forrest Gump	1994	8147.0	8.2	48.307194	13.0	8.069421
522	Schindler's List	1993	4436.0	8.3	41.725123	424.0	8.061007
23673	Whiplash	2014	4376.0	8.3	64.29999	244786.0	8.058025
5481	Spirited Away	2001	3968.0	8.3	41.048867	129.0	8.035598
1154	The Empire Strikes Back	1980	5998.0	8.2	19.470959	1891.0	8.025793



Top 10 Crime movies

```
build_top_movie_chart(metadata, genre="Crime", percentile=0.90, no_of_movies=10)
```

	title	year	vote_count	vote_average	popularity	id	weighted_rating
314	The Shawshank Redemption	1994	8358.0	8.5	51.645403	278.0	8.389613
834	The Godfather	1972	6024.0	8.5	41.109264	238.0	8.349302
12481	The Dark Knight	2008	12269.0	8.3	123.167259	155.0	8.229593
292	Pulp Fiction	1994	8670.0	8.3	140.950236	680.0	8.201554
1178	The Godfather: Part II	1974	3418.0	8.3	36.629307	240.0	8.064967
289	Leon: The Professional	1994	4293.0	8.2	20.477329	101.0	8.016979
3030	The Green Mile	1999	4166.0	8.2	19.96678	497.0	8.011851
46	Se7en	1995	5915.0	8.1	18.45743	807.0	7.970079
1170	GoodFellas	1990	3211.0	8.2	15.424092	769.0	7.961638
586	The Silence of the Lambs	1991	4549.0	8.1	4.307222	274.0	7.933982

Weighted-Rating based recommender

Top 10 Drama movies

```
build_top_movie_chart(metadata, genre="Drama", percentile=0.90, no_of_movies=10)
```

	title	year	vote_count	vote_average	popularity	id	weighted_rating
10309	Dilwale Dulhania Le Jayenge	1995	661.0	9.1	34.457024	19404.0	8.493149
314	The Shawshank Redemption	1994	8358.0	8.5	51.645403	278.0	8.452756
834	The Godfather	1972	6024.0	8.5	41.109264	238.0	8.434910
12481	The Dark Knight	2008	12269.0	8.3	123.167259	155.0	8.270123
2843	Fight Club	1999	9678.0	8.3	63.869599	550.0	8.262251
522	Schindler's List	1993	4436.0	8.3	41.725123	424.0	8.219148
23673	Whiplash	2014	4376.0	8.3	64.29999	244786.0	8.218078

Weighted-Rating based recommender

Content-based recommender

Problem with weighted-rating based recommender

Features considered:

- genres
- spoken_languages
- Tagline
- Overview

Used matrix of TF-IDF features using n-gram range of 1-2

Calculated cosine similarity score using linear kernel

Content-based recommender

- Recommendations using content-based recommender for movie “The Apartment”

```
get_recommendations('The Apartment', 10)
```

17658	Miss Nobody
8943	LadyBugs
32824	Unlocking the Cage
2298	Nothing in Common
29887	Company Limited
14094	Birdsong
7443	Shadow of the Thin Man
8697	The Holy Mountain
8352	The Shopworn Angel
21390	Black Nativity

Name: title, dtype: object

Hybrid recommender Content-based + Weighted rating based

Similar movies to "The Family" movie by cosine similarity and weighted rating

```
cosine_sim_plus_weighted_rating('The Family', 10, 0.80)
```

	title	year	vote_count	vote_average	popularity	id	weighted_rating
1186	Harold and Maude	1971	266.0	7.7	10.878112	343.0	7.625912
11131	All the Invisible Children	2005	14.0	7.3	2.271382	24553.0	6.614655
5358	The Last Kiss	2001	84.0	6.5	5.50323	12308.0	6.410125
42714	The Olive Tree	2016	38.0	6.5	2.541633	366505.0	6.323994
20489	Those Daring Young Men in Their Jaunty Jalopies	1969	10.0	6.8	4.336514	38792.0	6.227586
34492	Violent Summer	1959	10.0	6.7	0.906731	88395.0	6.177586
16581	Monte Carlo	2011	611.0	6.0	8.34835	59860.0	5.994447
9230	Remember Me, My Love	2003	47.0	5.8	4.996071	38970.0	5.774592
32177	Felix and Meira	2015	12.0	5.6	0.776885	285848.0	5.625078
5839	Pinocchio	2002	144.0	5.6	6.463613	10599.0	5.603583

Hybrid recommender Content-based + Weighted rating based

Similar movies to "The Apartment" movie by cosine similarity and weighted rating

```
cosine_sim_plus_weighted_rating(['The Apartment', 10])
```

	title	year	vote_count	vote_average	popularity	id	weighted_rating
8697	The Holy Mountain	1973	143.0	7.7	4.190266	8327.0	7.447292
7642	Safety Last!	1923	76.0	7.7	4.394823	22596.0	7.276613
1203	The Day the Earth Stood Still	1951	323.0	7.3	9.360003	828.0	7.205488
80	Things to Do in Denver When You're Dead	1995	87.0	6.7	4.486179	400.0	6.529401
4908	48 Hrs.	1982	364.0	6.5	15.297121	150.0	6.463430
10068	Baby Face	1933	30.0	6.9	3.363332	27449.0	6.459398
2337	Running Scared	1986	58.0	6.4	4.819467	15698.0	6.254187
3888	The House of Mirth	2000	25.0	6.4	2.492547	25520.0	6.154357
29506	Chloe and Theo	2015	24.0	6.4	2.775514	300686.0	6.149153
2298	Nothing in Common	1986	30.0	6.1	8.750517	29968.0	6.008271

Collaborative filtering-based recommendations

Rating based

Predict estimated rating for a movie using already given ratings by users

Used Surprise Library: SurPRISE - Simple Python Recommendation System Engine

Used Grid Search to compare and hyperparameter tuning

Compared SVD and KNN

Collaborative filtering-based recommendations

➤ Grid Search Results

```
[Parallel(n_jobs=10)]: Using backend LokyBackend with 10 concurrent workers.  
[Parallel(n_jobs=10)]: Done 18 out of 24 | elapsed: 16.3s remaining: 5.4s  
[Parallel(n_jobs=10)]: Done 24 out of 24 | elapsed: 20.9s finished
```

Model Name SVD

0.935093115053419

`{'n_epochs': 10, 'lr_all': 0.005, 'reg_all': 0.4}`

```
[Parallel(n_jobs=10)]: Using backend LokyBackend with 10 concurrent workers.  
[Parallel(n_jobs=10)]: Done 21 tasks | elapsed: 17.6s
```

Model Name KNN

0.9494451125698792

`{'n_epochs': 5, 'lr_all': 0.002, 'reg_all': 0.4, 'k': 50}`

```
[Parallel(n_jobs=10)]: Done 72 out of 72 | elapsed: 56.8s finished
```

Collaborative filtering-based recommendations

➤ Performance of SVD on Test dataset (25%)

```
svd = SVD(verbose=True, n_epochs=10, lr_all=0.005, reg_all=0.4)
```

```
RMSE: 0.9275
```

```
0.9275297885294524
```

```
svd.predict(uid=5, iid=100)
```

```
Prediction(uid=5, iid=100, r_ui=None, est=3.433044024923659, details={'was_impossible': False})
```

**Hybrid
Recommendation:
Content-Based +
Weighted Rating +
Collaborative
Filtering**

Find	Find similar movies using content-based recommendations
Apply	Apply weighted rating filter based on parameter
Predict	Predict estimating rating for each movie
Sort	Sort based on ratings and return top results

Hybrid Recommendation: Content-Based + Weighted Rating + Collaborative Filtering

Recommended movies for user 5, who watched movie Jumanji and filter_on_weighted_rate=False

```
hybrid_model_cosine_weighted_rate_svd(5, 'Jumanji', filter_on_weighted_rate=False)
```

	title	vote_count	vote_average	year	id	popularity	est
21577	Guardians of the Galaxy	10014.0	7.9	2014	118340.0	53.291601	3.887143
9592	The Phone Box	23.0	6.7	1972	52943.0	1.487562	3.846220
9358	Word Wars	7.0	6.1	2004	28105.0	0.549944	3.804895
14863	Le Pont du Nord	6.0	7.7	1982	59484.0	0.019669	3.786413
14115	The Spousals of God	4.0	8.8	1999	101342.0	0.503167	3.737727
18563	Wreck-It Ralph	4656.0	7.1	2012	82690.0	13.697597	3.669460
38987	Snowed Under	0.0	0.0	1936	217672.0	0.001754	3.662513
9019	Nirvana	57.0	6.6	1997	8765.0	3.376578	3.603538
16401	The Dark Angel	1.0	7.0	1935	120977.0	0.11704	3.594286
8048	Masques	6.0	5.9	1987	60760.0	0.807687	3.591593



Recommended movies for user 5, who watched movie Jumanji and filter_on_weighted_rate=True

```
hybrid_model_cosine_weighted_rate_svd(5, 'Jumanji', filter_on_weighted_rate=True)
```

	title	year	vote_count	vote_average	popularity	id	weighted_rating	est
21577	Guardians of the Galaxy	2014	10014.0	7.9	53.291601	118340.0	7.898630	3.887143
9592	The Phone Box	1972	23.0	6.7	1.487562	52943.0	6.475066	3.846220
9358	Word Wars	2004	7.0	6.1	0.549944	28105.0	5.875148	3.804895
14863	Le Pont du Nord	1982	6.0	7.7	0.019669	59484.0	6.656410	3.786413
18563	Wreck-It Ralph	2012	4656.0	7.1	13.697597	82690.0	7.098086	3.669460
9019	Nirvana	1997	57.0	6.6	3.376578	8765.0	6.505983	3.603538
8048	Masques	1987	6.0	5.9	0.807687	60760.0	5.756410	3.591593
42636	Quiz	2012	11.0	5.5	1.031617	96823.0	5.539819	3.570156
41766	Liar Game: Reborn	2012	6.0	5.8	0.605467	133977.0	5.706410	3.565825
14744	The Wild Hunt	2009	20.0	6.2	1.847803	44680.0	6.064497	3.564569

Hybrid Recommendation: Content-Based + Weighted Rating + Collaborative Filtering

Thank You!