

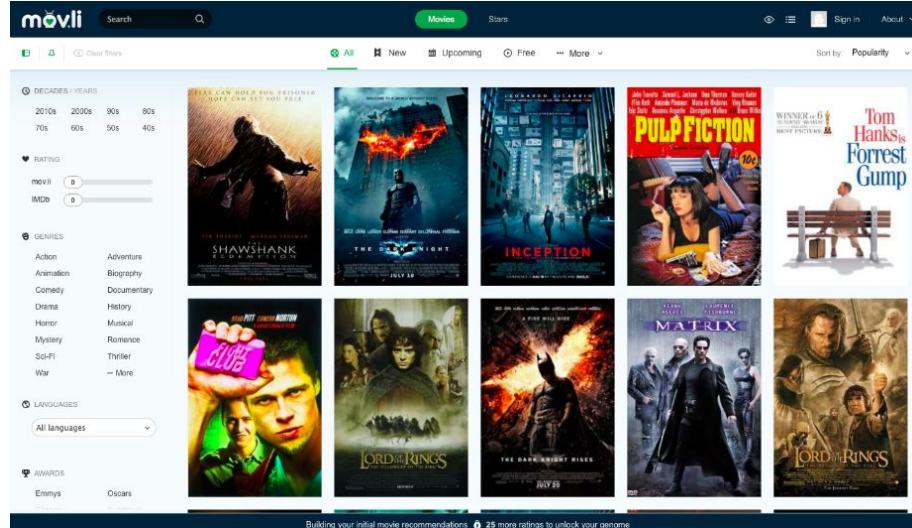
'Bad Movies'

Recommendation

Xinyu Zang & Lin Jiang



Tired of these recommended movies?



Famous & Highly rate &
Seen before

We will recommend some “bad” ones you might haven’t seen!

Intuition --- Only Bad for Movie Critics

- Some movies were surprisingly panned by critics, but had been very popular by general audiences
- These movies might not get the headlines in magazine or newspaper written by professional critics and therefore less exposure are very likely to be ignored by us.

Part of the Collection: Marvel Cinematic Universe

BLACK PANTHER
2018

TOMATOMETER **97%**
Average Rating: 8.2/10
Reviews Counted: 364
Fresh: 353
Rotten: 11

AUDIENCE SCORE **79%**
Critics Consensus: *Black Panther* elevates superhero cinema to thrilling new heights while telling one of the MCU's most absorbing stories – and introducing some of its most fully realized characters.

All News Videos Images Shopping More Settings Tools

About 9,810,000 results (0.19 seconds)

'Black Panther' To Be Highest Grossing Superhero Of All Time
Forbes - 7 hours ago
After it reaches \$1.2 billion this week, **Black Panther** still has enough gas left in its tank to add an additional \$60-80+ million in global revenue to its cume, depending on what happens over the next few weekends as some potentially major contenders arrive, including Pacific Rim Uprising, Ready Player One ...

Black Panther Breaks Twitter Record
Voice of America - 22 hours ago

'Black Panther' breaks another record, becoming the most tweeted ...
Los Angeles Times - 19 hours ago

DEATH WISH
2018

TOMATOMETER **18%**
Average Rating: 3.9/10
Reviews Counted: 108
Fresh: 19
Rotten: 89

AUDIENCE SCORE **81%**
Critics Consensus: *Death Wish* is little more than a rote retelling that lacks the grit and conviction of the original – and also suffers from spectacularly bad timing.

Add Your Rating

All Videos News Images Shopping More Settings Tools

About 2,520,000 results (0.37 seconds)

Death Wish: Better than Original?
Winchester News Gazette - Mar 20, 2018
Bruce Willis stars in a remake of the 1974 Charles Bronson movie **Death Wish**. Willis plays Dr. Paul Kersey whose wife and daughter are attacked in a brutal home invasion. After his wife dies, the Chicago doctor purchases some guns and begins hunting down the killers. The vigilante is soon dubbed "The ...

Mission Statement

Based on user's taste such as favorite movie or favorite genres, recommend similar movie that users are very likely to ignore because of low ratings from critics



BIG DADDY

TOMATOMETER **40%**
All Critics | Top Critics

Critics Consensus: Adam Sandler acquits himself admirably, but his charm isn't enough to make up for *Big Daddy's* jarringly shifts between crude humor and mawkish sentimentality.

Average Rating: 4.9/10
Reviews Counted: 93
Fresh: 37
Rotten: 56

AUDIENCE SCORE **74% liked it**
All Critics | Top Critics

Average Rating: 3.5/5
User Ratings: 1,052,024

ADD YOUR RATING

- NOT INTERESTED + WANT TO SEE

Add a Review (Optional)



HOTEL TRANSYLVANIA

TOMATOMETER **44%**
All Critics | Top Critics

Critics Consensus: Hotel Transylvania's bouncy, giddy tone may please children, but it might be a little too loud and thinly-scripted for older audiences.

Average Rating: 5.3/10
Reviews Counted: 142
Fresh: 63
Rotten: 79

AUDIENCE SCORE **72% liked it**
All Critics | Top Critics

Average Rating: 3.8/5
User Ratings: 150,781

ADD YOUR RATING

- NOT INTERESTED + WANT TO SEE

Add a Review (Optional)

Post

Data

MovieLens Latest Datasets

Collected by GroupLens, a research group at University of Minnesota.

Datasets include 100,000 ratings and 1,300 tag applications applied to 9,000 movies.

The Open Movie Database

Dataset contains movie ratings from both Rotten Tomatoes critics and the general audience. It also contains urls of movie posters.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
id	title	imdbID	spanishTitle	imdbPicture	year	rtID	rtAllCriticsRating	rtAllCriticsNumReviews	rtAllCriticsNumFresh	rtAllCriticsNumRt	rtAllCriticsScore	rtTopCriticsRating	rtTopCriticsRt	rtTopCriticsrt	rtTopCriticsrt	rtAudience	rtAudienceScore	rtPictureURL							
1	Toy story	114709	Toy story	http://ia.r	1995	toy_story	9	73	73	0	100	8.5	17	17	0	100	3.7	102338	81	http://content7.flixster.com/movie/10/93/63/10936393_det.jpg					
2	Jumanji	113497	Jumanji	http://ia.r	1995	1068044_jt	5.6	28	13	15	46	5.8	5	2	3	40	3.2	44587	61	http://content8.flixster.com/movie/56/79/73/5679734_det.jpg					
3	Grumpy O	107050	Dos viejos	http://ia.r	1993	grumpy_o	5.9	36	24	12	66	7	6	5	1	83	3.2	10489	66	http://content6.flixster.com/movie/25/60/256020_det.jpg					
4	Waiting tc	114885	Esperando	http://ia.r	1995	waiting_tc	5.6	25	14	11	56	5.5	11	5	6	45	3.3	5666	79	http://content9.flixster.com/movie/10/94/17/10941715_det.jpg					
5	Father of	113041	Vuelve el	http://ia.r	1995	father_of	5.3	19	9	10	47	5.4	5	1	4	20	3	13761	64	http://content8.flixster.com/movie/25/54/255426_det.jpg					

Data Preprocessing

MovieLens Latest Datasets

- Clean movie title to build dictionary

The Open Movie Database

- Check missing values
- Select movies: Critics Score<50 AND Audience Score>50 AND Number of Audience Ratings>200 as potential recommendations

A	B
moviedb_id	title
1	1 Toy Story (1995)
2	2 Jumanji (1995)
3	3 Grumpier Old Men (1995)
4	4 Waiting to Exhale (1995)
5	5 Father of the Bride Part II (1995)
6	6 Heat (1995)
7	7 Sabrina (1995)
8	8 Tom and Huck (1995)
9	9 Sudden Death (1995)
10	10 GoldenEye (1995)
11	11 American President, The (1995)
12	



A	B
moviedb_id	title
1	1 Toy Story
2	2 Jumanji
3	3 Grumpier Old Men
4	4 Waiting to Exhale
5	5 Father of the Bride Part II
6	6 Heat
7	7 Sabrina
8	8 Tom and Huck
9	9 Sudden Death
10	10 GoldenEye
11	11 The American President
12	

Model Selection

- Memory-Based (User-User & Item-Item) v.s. Model-Based (SVD)
- Use Cross Validation and Root Mean Squared Error (RMSE) to evaluate the model
 - In each iteration of CV, use ratings from the training folds to estimate the rating in the test fold and compute RMSE
 - Average RMSE in each iteration to get the CV-RMSE
- Model-Based Collaborative Filtering has relatively low RMSE.

```
#Print out the RMSE
print('User-based CF cross-validation RMSE: ' + str(np.mean(user_rmse)))
print('Item-based CF cross-validation RMSE: ' + str(np.mean(item_rmse)))
```

```
User-based CF cross-validation RMSE: 3.09088051696
Item-based CF cross-validation RMSE: 3.43873881381
```

```
user_ratings_mean = np.mean(train_data_matrix, axis = 1)
normal_train_data_matrix = train_data_matrix - user_ratings_mean.reshape(-1, 1)

u, s, vt = svds(normal_train_data_matrix, k = 20)
s_diag_matrix=np.diag(s)
X_pred = np.dot(np.dot(u, s_diag_matrix), vt) + user_ratings_mean.reshape(-1, 1)
print('User-based CF MSE with K=20: ' + str(rmse(X_pred, test_data_matrix)))
```

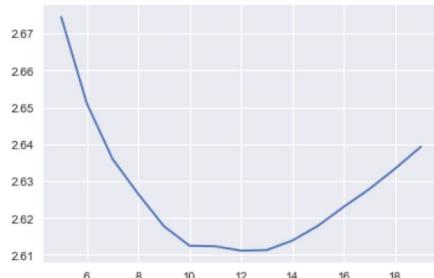
```
User-based CF MSE with K=20: 2.6461986099138306
```

Model-Based Collaborative Filtering

- Apply SVD to get a low- dimensional hidden feature matrix from the sparse rating matrix
- Use 5-folds Cross Validation to choose the optimal K-value which achieves minimal RMSE.

```
plt.plot(k_range, mean_rmse_value)
k_optima=k_range[np.argmin(mean_rmse_value)]
print('The optimal K value is ' + str(k_optima) + ' with RMSE=' + str(np.min(mean_rmse_value)))
```

The optimal K value is 12 with RMSE=2.61112085349



Model-Based Collaborative Filtering

- Measure cosine similarity of the hidden feature matrix. For input movie, find 10 similar underrated movies Ids

```
def top_cosine_similarity(low_dim_data, movie_name, top_n):  
    movie_name = movie_name  
    low_case_movie_name = movie_name.lower()  
    if low_case_movie_name in lower_case_movie_dict:  
        index = Rating_df.columns.get_loc(lower_case_movie_dict[low_case_movie_name])  
        movie_row = low_dim_data[:, index]  
        magnitude = np.sqrt(np.einsum('ij, ij -> i', low_dim_data.T, low_dim_data.T))  
        similarity = np.dot(movie_row, low_dim_data) / (magnitude[index] * magnitude)  
        pre_sort_indexes = np.argsort(-similarity)  
        #Delete the index of input movie itself  
        sort_indexes = np.delete(pre_sort_indexes, 0)  
        top_movie_id = Rating_df.columns[sort_indexes]  
        frozen_selected_id = frozenset(selected_id)  
        selected_top_movie_id = [x for x in top_movie_id if x in frozen_selected_id][:top_n]  
        return selected_top_movie_id  
    else:  
        return("Sorry, no matching found in the dataset")
```

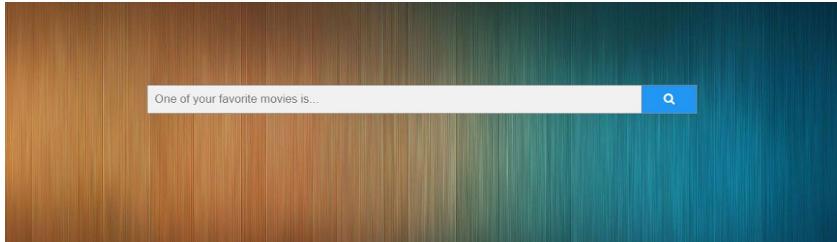
- Retrieve the relative information from the recommended Ids

```
def get_movie_infor(movie_df, movie_id):  
    indexes = []  
    for id in movie_id:  
        index = movie_df[movie_df['id']==id].index.astype(int)[0]  
        indexes.append(index)  
    get_content = movie_df.iloc[indexes]  
    return get_content[['title', 'rtAllCriticsScore', 'rtAudienceScore', 'rtPictureURL']]
```

- For recommendation based on genres, get top 10 highest audience scores movies from each genre.

Input

- One of your favourite movies



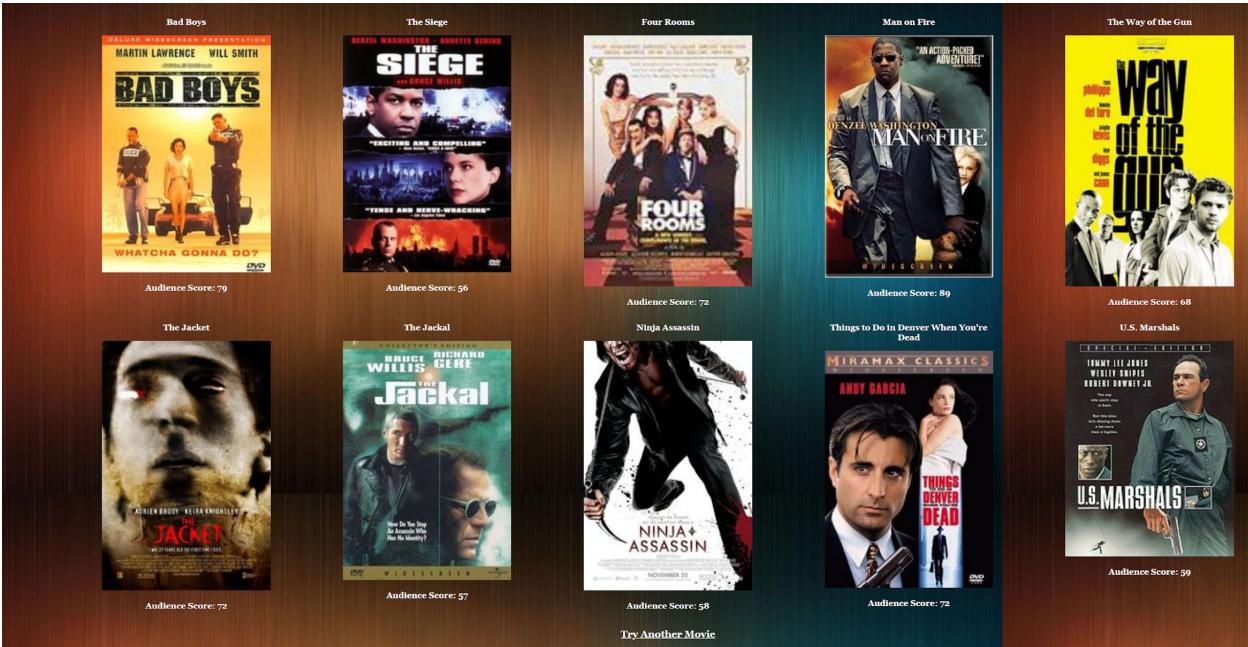
- One of you favorite film genres

Select a Movie Genre

Action	Animation	Documentary
Adventure	Children	Drama
Sci-Fi	Comedy	Musical
War	Fantasy	Romance

Output

- 10 similar ‘underrated’ movies you may like based on one of your favorite movies with audience scores



Output

- Top 10 ‘underrated’ movies by the critics from your favorite genre with audience scores

Home Recommendation Based on Movies Recommendation Based on Genres ▾

Top Audience Rated Action Movies

Movie	Audience Score
The Boondock Saints	93
Man on Fire	89
Facing the Giants	85
Extreme Days	85
The Last Dragon	84
Die Hard: With a Vengeance	83
The Guardian	83
State Property 2	82
Shooter	82
Underworld	82

No Satisfied? Try Recommendation Based on Movies

The image shows a screenshot of a movie recommendation interface. At the top, there are three navigation links: 'Home', 'Recommendation Based on Movies', and 'Recommendation Based on Genres'. Below this is a heading 'Top Audience Rated Action Movies' followed by a table listing ten movies with their audience scores. Each movie entry includes a small thumbnail image of the movie poster. The movies listed are: The Boondock Saints (93), Man on Fire (89), Facing the Giants (85), Extreme Days (85), The Last Dragon (84), Die Hard: With a Vengeance (83), The Guardian (83), State Property 2 (82), Shooter (82), and Underworld (82). At the bottom of the table, there is a link 'No Satisfied? Try Recommendation Based on Movies'.

Demo

- Let's try it with one of my favorite movies 'Heat', an American crime movie starring Robert De Niro, Al Pacino.
- Most recommendations are action/crime movies. Some with high IMDb scores.



Tears of the Sun

R 2003 · Thriller/Drama · 2h 22m

[Play trailer on YouTube](#)

6.6/10
IMDb

33%
Rotten Tomatoes

48%
Metacritic



Dead Presidents

R 1995 · Drama/Crime film · 1h 59m

[Play trailer on YouTube](#)

6.9/10
IMDb

45%
Rotten Tomatoes

