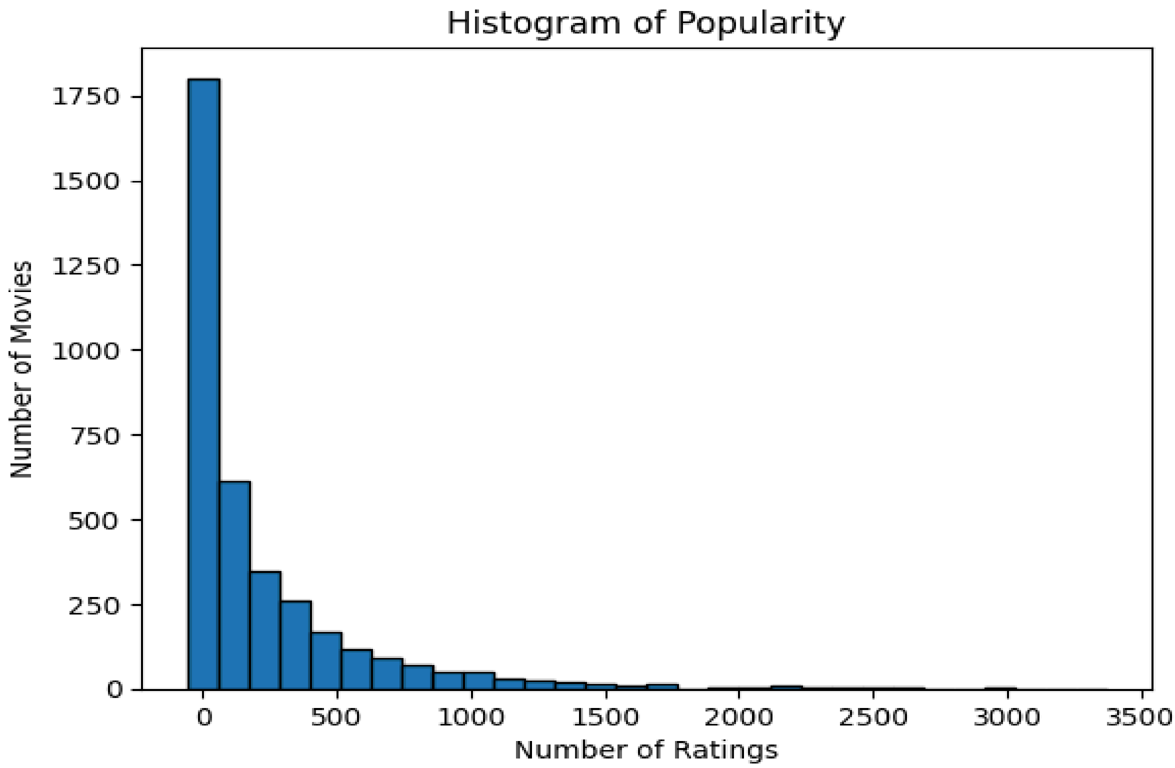
Case 2: Recommendation System Report

Team 56: Boqing Zheng, Yiqiao Wang, Rajat Bajaj, Genna Barge, Kainat Nazir

Exercise: Plot the histogram of popularity (x-axis: # of given ratings, y-axis: # of movies with the given # of ratings). Include the plot in your report.

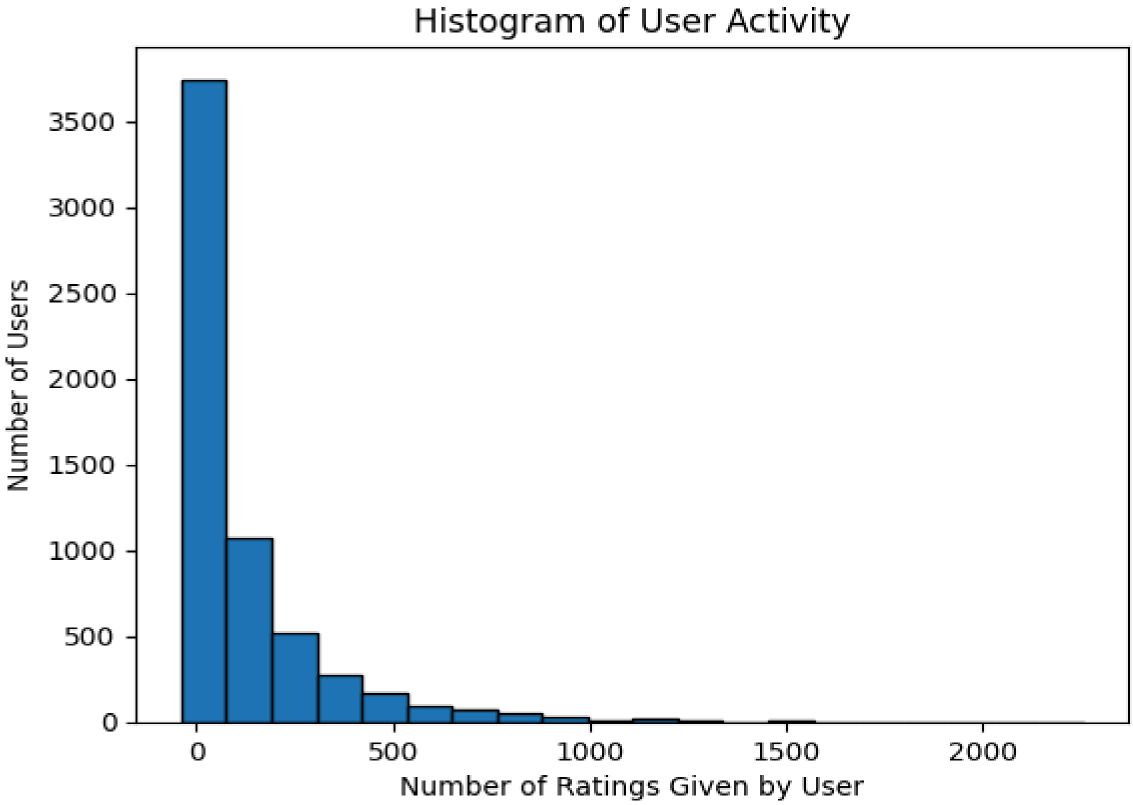


The histogram shows the distribution of movies based on the number of ratings they received. The x-axis represents the number of ratings given to movies, while the y-axis represents the number of movies that fall into each range of ratings.

Insights:

1. **Highly Skewed Distribution**:
   - Most movies have a small number of ratings, with a sharp decline in popularity as the number of ratings increases. This indicates that the majority of movies are niche, appealing to smaller audiences.
   - A few movies received a large number of ratings, likely blockbuster or popular films.
2. **Long-Tail Phenomenon**:
   - The "long tail" of movies with fewer ratings represents an opportunity for personalized recommendations. Catering to this niche audience can improve user satisfaction and uncover hidden gems in the catalog.
3. **Business Implications**:
   - Popular movies drive engagement and could be leveraged for promotions and advertisements.
   - The lesser-rated movies represent a potential for better recommendation algorithms to increase their visibility and improve the diversity of movies viewed by users.

Exercise: Plot the histogram of user activity (x-axis: # of given ratings, y-axis: # of users with the given # of ratings). Include the plot in your report.



This histogram shows the distribution of user activity in terms of the number of ratings given. The x-axis represents the number of ratings provided by each user, and the y-axis shows the count of users corresponding to those numbers.

Insights:

1. **Significant Impact of Highly Active Users**
   A small proportion of highly active users contribute a substantial number of ratings. Understanding their behavior is crucial for improving recommendation system performance.
2. **Majority of Users Are Less Active**
   The dataset contains a large number of passive users who give only a handful of ratings. These users provide limited data, making it more challenging to model their preferences effectively.
3. **Highly Active Users as Key Outliers**
   Highly active users can be considered outliers, contributing disproportionately to the overall dataset. They play a critical role in shaping the dataset and influencing model outcomes.
4. **Wide Range of User Activity**
   There is a significant variation in user activity levels, from casual users to highly engaged ones. This indicates differing levels of interest and engagement across the user base.

Exercise: Play with the hyperparameters (embedding_dim, n_iter, learning_rate, L2 regularization, etc) to beat the benchmarks presented in Surprise. List the hyperparameters and the RMSE (square root of the mean sequared error) and MAE on test data in the report. Save the model state to model_cf.pt and submit it to Canvas.

```
Best Hyperparameters: {'embedding_dim': 64, 'n_iter': 30, 'learning_rate': 0.0005, 'l2': 1e-05}
Best RMSE: 0.8536, Best MAE: 0.6692
```

	embedding_dim	n_iter	learning_rate	l2	RMSE	MAE
0	32	30	0.0010	0.000010	0.8725	0.6821
1	32	30	0.0010	0.000001	0.8929	0.6977
2	32	30	0.0005	0.000010	0.8656	0.6807
3	32	30	0.0005	0.000001	0.8614	0.6750
4	32	50	0.0010	0.000010	0.8983	0.7003
5	32	50	0.0010	0.000001	0.9428	0.7312
6	64	30	0.0010	0.000010	0.9055	0.7079
7	64	30	0.0010	0.000001	0.9539	0.7436
8	64	30	0.0005	0.000010	0.8536	0.6692
9	64	30	0.0005	0.000001	0.8538	0.6679
10	64	50	0.0010	0.000010	0.9499	0.7400
11	64	50	0.0010	0.000001	1.0375	0.8056

We evaluate the optimal recommender system model by comparing the model performance for different combinations of hyperparameters (e.g., embedding_dim, n_iter, learning_rate, and regularization parameter L2). After training the model for each hyperparameter combination, we recorded its corresponding root mean square error (RMSE) and mean absolute error (MAE). As can be seen from the table, the optimal parameter combination of embedding_dim: 64, n_iter: 30, learning_rate: 0.0005, L2: 1e-05 corresponds to an optimal RMSE of 0.8536 and an optimal MAE of 0.6692, which suggests that the model under this hyper-parameter combination has better prediction performance.

Exercise: Compute the average ratings for every movie and find the top 20 highly rated movies. Include the list of 20 highly rated movies and their ratings in your report. Are these highly rated movies popular?

```
item_name
Lured (1947)                                                              5.000000
Bittersweet Motel (2000)                                                  5.000000
Ulysses (Ulisse) (1954)                                                   5.000000
Smashing Time (1967)                                                      5.000000
Baby, The (1973)                                                          5.000000
Gate of Heavenly Peace, The (1995)                                        5.000000
Schlafes Bruder (Brother of Sleep) (1995)                                 5.000000
Follow the Bitch (1998)                                                   5.000000
One Little Indian (1973)                                                  5.000000
Song of Freedom (1936)                                                    5.000000
I Am Cuba (Soy Cuba/Ya Kuba) (1964)                                       4.800000
Lamerica (1994)                                                           4.750000
Apple, The (Sib) (1998)                                                   4.666667
Sanjuro (1962)                                                            4.608696
Seven Samurai (The Magnificent Seven) (Shichinin no samurai) (1954)      4.560510
Shawshank Redemption, The (1994)                                          4.554558
Godfather, The (1972)                                                     4.524966
Close Shave, A (1995)                                                     4.520548
Usual Suspects, The (1995)                                                4.517106
Schindler's List (1993)                                                   4.510417
Name: ratings, dtype: float64
```

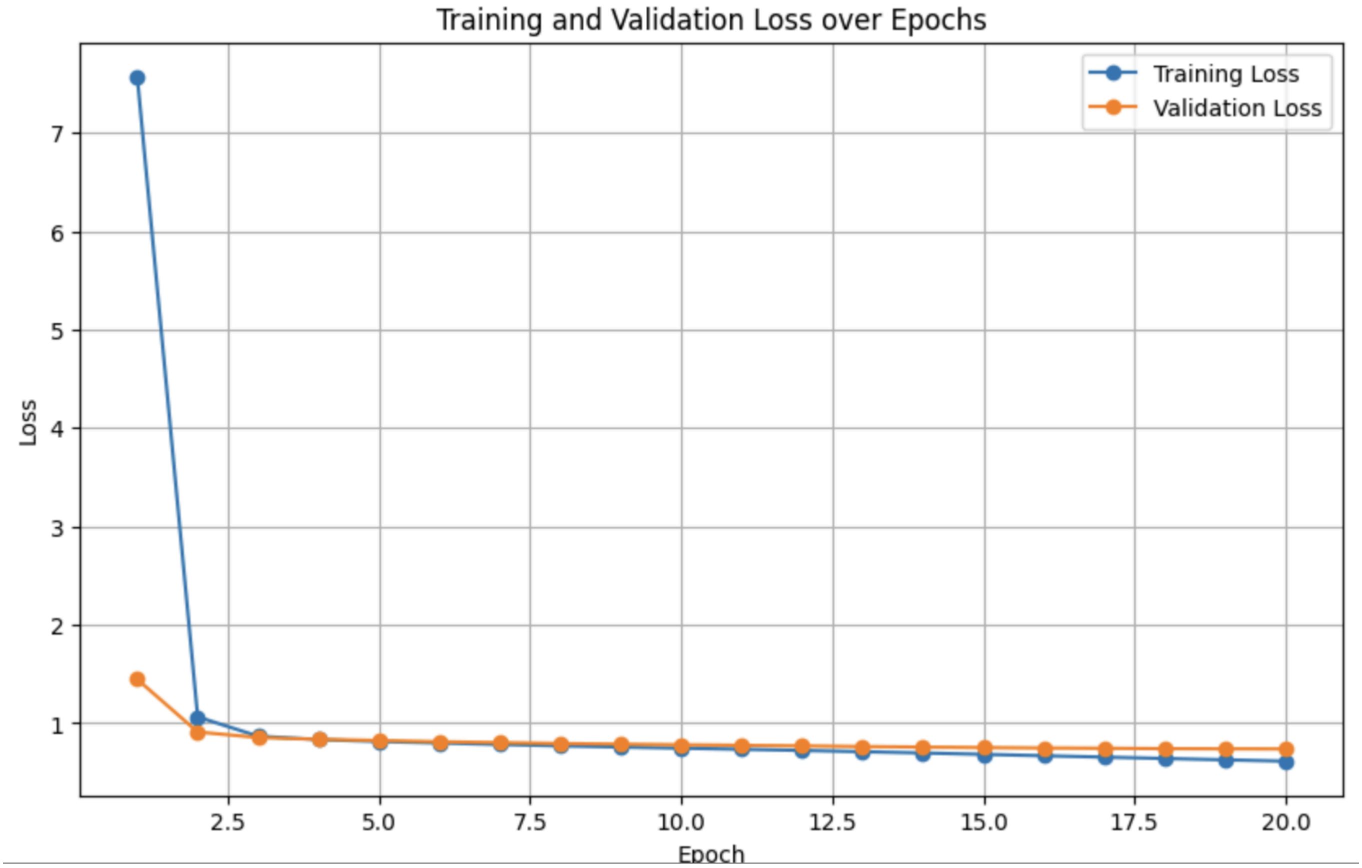
The table lists the top 20 movies ranked by their average ratings.

Insights from the Table:

1. **High Ratings with Niche Appeal**:
    - Many of the movies in the top positions, such as *Lured (1947)* and *Bittersweet Motel (2000)*, have a perfect average rating (5.0).
    - These ratings often come from very few users, indicating a niche group of viewers who strongly appreciated these films.
    - These movies may not have broad recognition or appeal but cater to very specific audiences.
2. **Combination of Quality and Popularity**:
    - Films like *The Shawshank Redemption* and *Seven Samurai* show a good balance.
    - These movies have strong average ratings (around 4.5–4.6) and are recognized classics.
    - They have been rated by a larger number of users, reflecting a combination of critical acclaim and general popularity.
3. **Highly Rated but Questionable Reliability**:
    - Movies like *Ulysses (1954)* and *Lured (1947)* may have high average ratings, but these ratings could come from very few users.
    - The small sample size makes these ratings less reliable, as they might reflect isolated opinions rather than a consensus.
4. **Popular Classics with Consistent High Ratings**:
    - *The Godfather* and *Schindler's List* stand out as universally loved classics with high ratings.
    - These movies are highly rated by a large number of users, making them strong candidates for broader recommendations.

Netflix can promote niche movies like *Lured (1947)* as "hidden gems" for specific audiences while highlighting popular classics like *The Godfather* in general recommendations. For lesser-known films, gathering more ratings could confirm their appeal, allowing Netflix to balance mainstream and personalized recommendations effectively.

Exercise: Train the model and plot the train losses and valid losses over epoches. At which epoch, would you stop the training?



The plot illustrates the progression of training and validation losses across 20 epochs. Both losses represent the model's performance during training and validation stages.

1. **Convergence of Losses:**
   The training and validation losses decrease sharply during the initial epochs, indicating that the model is learning effectively during this stage. After about **epoch 2**, the losses stabilize and converge to a nearly constant value. This suggests that the model has reached an optimal balance between underfitting and overfitting.
2. **Low Validation Loss:**
   The validation loss is slightly higher than the training loss, indicating a good generalization capacity. However, the closeness of these two losses suggests minimal overfitting.
3. **Epoch Selection for Stopping Training:**
   Based on the validation loss curve, training could have been stopped at epoch 2. At this point, the validation loss reaches its lowest stable value, reducing computational costs without sacrificing performance.

Exercise: Now, let's compute the RMSE and mean absolute error on the test data.
Test RMSE: 0.8592322161927548

Test MAE: 0.6738846758190467

The test results show a RMSE of 0.8592 and a MAE of 0.6739, indicating the model predicts user ratings with good accuracy. The low error values suggest that the model is effective for practical recommendation tasks, with small deviations between predicted and actual ratings. The slightly higher RMSE compared to MAE indicates a few outliers but overall controlled error distribution. These results demonstrate the model's suitability for generating personalized recommendations, improving user satisfaction, and driving engagement. Further tuning of hyperparameters or adding more diverse training data could enhance performance further.

Here is comparison between the best benckmarks in Surprise and your model after hyperparameters tunning.

Movielens 1M	RMSE	MAE
SVD++ (cache_ratings=False)	0.862	0.672
SVD++ (cache_ratings=True)	0.862	0.672
Yours	**0.862**	**0.675**

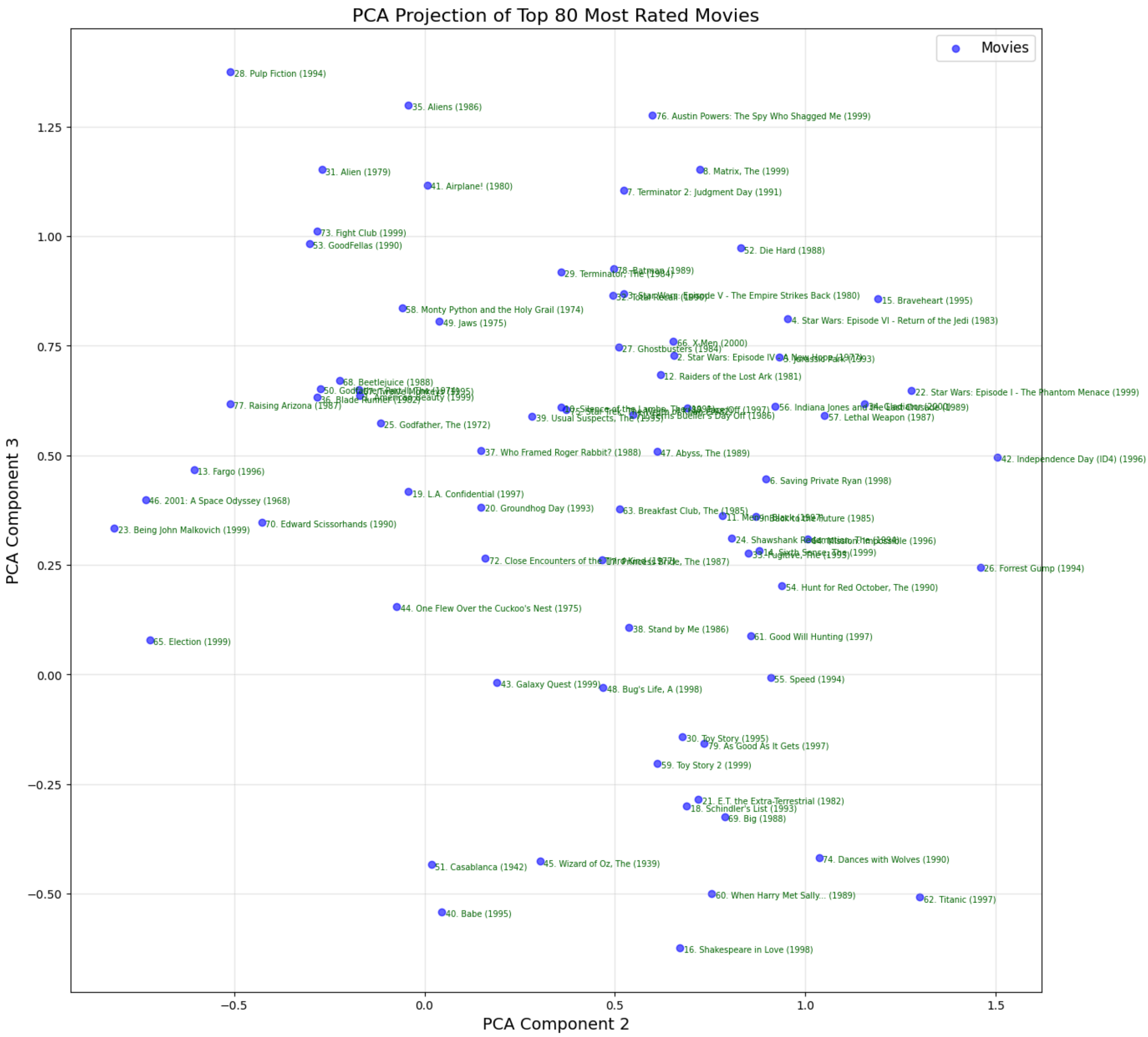
Our model achieves an RMSE of **0.862** and an MAE of **0.675**, which is comparable to the best benchmark provided by the SVD++ algorithm in the Surprise library. This result indicates that our model performs on par with industry-standard techniques for rating prediction. While there is a slight increase in MAE, the RMSE remains identical, showcasing consistent prediction reliability. This level of performance demonstrates that our hyperparameter tuning and model adjustments are effective in aligning predictions closely with actual user preferences.

Exercise: What are your top 10 movies with the largest values of movie bias? What do you think about this ranking?

```
Top 10 movies with largest biases:
1. Last Days, The (1998): Bias = 0.7168
2. Sanjuro (1962): Bias = 0.7032
3. Lamerica (1994): Bias = 0.6948
4. Pushing Hands (1992): Bias = 0.6932
5. Live Flesh (1997): Bias = 0.6759
6. Harmonists, The (1997): Bias = 0.6746
7. Man from Laramie, The (1955): Bias = 0.6744
8. Sea Wolves, The (1980): Bias = 0.6733
9. Aparajito (1956): Bias = 0.6733
10. Apple, The (Sib) (1998): Bias = 0.6721
```

This ranking demonstrates the model's ability to identify movies that are skewed towards specific audiences but not necessarily mainstream, such as Last Days and Sanjuro. While these movies have strong appeal among certain audience segments, high bias values do not equate to broad popularity, and therefore need to be combined with traditional metrics such as overall ratings or views to balance the recommendations.

Exercise: Do you observe anything interesting from the movie embeddings?



The scatter plot illustrates the embeddings of the top 80 most-rated movies, where each point represents a movie. Some interesting phenomena can be observed in the distribution of movie embedding. Movies of similar genres tend to cluster in the same area, e.g., sci-fi movies such as Star Wars and Alien are close together, while classic dramas such as Shawshank Redemption and Forrest Gump show a similar distribution. In addition, some movies such as Titanic and The Wizard of Oz are far from others, possibly representing their unique characteristics in the embedding space. Overall, embedding captures the similarity of movies in a particular feature space well.

Question: How do you feel about the predicted ratings? Do you like these top 10 movies?

	movie_id	movie_name	predicted_rating
0	2210	First Blood (1982)	9.235331
1	705	Thinner (1996)	8.965622
2	1798	Friday the 13th Part V: A New Beginning (1985)	8.927766
3	831	Halloween: The Curse of Michael Myers (1995)	8.887348
4	1796	Friday the 13th Part 3: 3D (1982)	8.800321
5	1799	Friday the 13th Part VI: Jason Lives (1986)	8.721672
6	2219	Rocky V (1990)	8.678473
7	2364	Ravenous (1999)	8.600521
8	1805	Halloween 4: The Return of Michael Myers (1988)	8.597535
9	1803	Halloween II (1981)	8.588351

These 10 films include many classics such as First Blood, all of which have high artistic merit and audience recognition across a variety of genres and styles, and we love these recommendations.

Question:

- What is your estimated value of Toy Story (1995)?
  The estimated value for Toy Story (1995) is **$4,540,522.65**

- What are the top 10 mostly valued movies?

```
Estimated value of 'Toy Story (1995)': $4,540,522.65
Top 10 highest valued movies:
item_name
Godfather, The (1972)                                                        3.374008e+07
Raiders of the Lost Ark (1981)                                               2.254294e+07
Star Wars: Episode IV - A New Hope (1977)                                    2.083304e+07
American Beauty (1999)                                                       1.800831e+07
Pulp Fiction (1994)                                                          1.580941e+07
Shawshank Redemption, The (1994)                                             1.490603e+07
Schindler's List (1993)                                                      1.306673e+07
Casablanca (1942)                                                            1.033846e+07
Dr. Strangelove or: How I Learned to Stop Worrying and Love the Bomb (1963)  1.014664e+07
Sixth Sense, The (1999)                                                      9.762038e+06
Name: item_value, dtype: float64
```

According to the model prediction, the movie is valued at $4,540,522.65, which indicates its high audience appeal and potential commercial value, but fails to make it to the top 10 list of highest valuations. At the top of the list is “The Godfather (1972)” with a valuation of $33,740,080.00, which is far more than any other movie, reflecting its classic status and popularity.

Other highly valued films include classics such as Raiders of the Lost Ark (1981) and Star Wars: Episode IV - A New Hope (1977), which have historically had high commercial value and cultural impact.

Question:

- What are the movies that are top 30 rated but not in top 30 valued?

```
Movies in top 30 rated but not in top 30 valued:
1. Back to the Future (1985)
2. Men in Black (1997)
3. Shakespeare in Love (1998)
4. Ghostbusters (1984)
5. Star Wars: Episode I - The Phantom Menace (1999)
6. Toy Story (1995)
7. E.T. the Extra-Terrestrial (1982)
8. Groundhog Day (1993)
9. Jurassic Park (1993)
10. Terminator, The (1984)
11. Star Wars: Episode VI - Return of the Jedi (1983)
12. L.A. Confidential (1997)
13. Terminator 2: Judgment Day (1991)
```

  By comparing and finding the differences between the two ratings sets, we found these 13 movies.

- **What do you think about this approach of movie valuation? How does it compare with the approach based on the popularity of the movies?**
  This approach to movie valuation is based on predicted view time, which is influenced by both user preferences and ratings.
  Some highly rated movies may not appear in the top 30 valued because their predicted view time might be lower due to fewer active users.