Recommendation System

The code starts by loading the movie and rating data from CSV files. It then filters the data to include only the top 200 movies and the top 200 users who have rated those movies. The filtered data is then used to create a user-movie matrix, which is standardized using the StandardScaler function from scikit-learn.

Next, the code calculates the similarity matrix using the cosine similarity metric. The similarity matrix is used to find the top 10 similar movies for a given movie ID using the get_similar_movies function.

The main function of the code is recommend_movies, which takes in a user ID, the user-movie matrix, the similarity matrix, the filtered movies, and the number of recommendations to provide. The function first retrieves the ratings of the given user from the user-movie matrix and creates a list of unrated movies. It then drops the unrated movies from the user-movie matrix and calculates a new similarity matrix based on the updated matrix.

The function then calculates a score for each unrated movie by taking the weighted average of the user's ratings for similar movies. The score is based on the updated similarity matrix and the user's ratings. The function then sorts the scores in descending order and returns a list of the top n movies with the highest scores.

There are opportunities to improve the code by adding more features such as genre-based recommendations and user-based collaborative filtering.

Main Functions used:

1. **get_similar_movies(movie_id, n=10)**
   This function takes in a movie ID and the number of similar movies to return. It first finds the index of the movie in the filtered movies data frame. It then calculates the similarity scores between the given movie and all other movies using the similarity matrix. The function returns the top n movies with the highest similarity scores.

2. **recommend_movies(user_id, user_movie_matrix, similarity_matrix, filtered_movies, n=3)**
   This function takes in a user ID, the user-movie matrix, the similarity matrix, the filtered movies data frame, and the number of recommendations to provide (default is 3). It first retrieves the ratings of the given user from the user-movie matrix and creates a list of unrated movies. It then drops the unrated movies from the user-movie matrix and calculates a new similarity matrix based on the updated matrix. The function then calculates a score for each unrated movie by taking the weighted average of the user's ratings for similar movies. The score is based on the updated similarity matrix and the user's ratings. The function then sorts the scores in descending order and returns a list of the top n movies with the highest scores.

3. **StandardScaler()**
   This function from the scikit-learn library is used to standardize the user-movie matrix. It scales the data so that each feature has a mean of 0 and a standard deviation of 1.

4. **cosine_similarity(user_movie_matrix.T)**
   This function from the scikit-learn library is used to calculate the similarity matrix. It calculates the cosine similarity between each pair of movies in the user-movie matrix.

Outputs of The Code:

Q1)

```python
similarity_matrix_ = pd.DataFrame(similarity_matrix, columns=user_movie_matrix.columns, index=user_movie_matrix.columns)
similarity_matrix_
```

[128] ✓ 0.1s Python

movieId	1	2	3	4	5	6	7	8	9	10	...	251	252	253	254	255	256
movieId																	
1	1.000000	0.364634	0.284825	0.000000	0.279775	0.327928	0.270701	0.128765	0.195017	0.352899	...	0.078029	0.207645	0.338312	0.000000	0.054509	0.193446
2	0.364634	1.000000	0.278446	0.102937	0.277072	0.278936	0.240488	0.130496	0.000000	0.342590	...	0.257005	0.218991	0.368741	0.205874	0.000000	0.150889
3	0.284825	0.278446	1.000000	0.150278	0.391915	0.341700	0.592157	0.406427	0.288534	0.231410	...	0.346341	0.317709	0.389201	0.300557	0.000000	0.278921
4	0.000000	0.102937	0.150278	1.000000	0.239259	0.176604	0.297614	0.253546	0.000000	0.129040	...	0.384111	0.132964	0.066749	0.500000	0.000000	0.126547
5	0.279775	0.277072	0.391915	0.239259	1.000000	0.256166	0.539392	0.404422	0.000000	0.275293	...	0.229755	0.234621	0.226247	0.299074	0.107000	0.262406
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
256	0.193446	0.150889	0.278921	0.126547	0.262406	0.134093	0.281211	0.484849	0.000000	0.300284	...	0.194433	0.496376	0.374480	0.253095	0.301833	1.000000
257	0.029922	0.171278	0.242671	0.272772	0.241474	0.226412	0.266273	0.276642	0.000000	0.272201	...	0.419099	0.413467	0.242764	0.545545	0.000000	0.322174
258	0.144217	0.391489	0.243856	0.253546	0.151658	0.104481	0.181101	0.257143	0.000000	0.232658	...	0.389559	0.235989	0.157957	0.507093	0.000000	0.171123
259	0.121886	0.000000	0.240445	0.000000	0.239259	0.000000	0.238091	0.676123	0.000000	0.240874	...	0.000000	0.199447	0.249197	0.000000	0.000000	0.539935
260	0.533147	0.374672	0.211501	0.051085	0.157366	0.421490	0.240978	0.000000	0.000000	0.308726	...	0.073584	0.158350	0.236703	0.000000	0.028557	0.100203

Q2)

```python
print(get_similar_movies(1))
```

[93] ✓ 0.0s

```
     movieId                                     title  \
0          1                          Toy Story (1995)
224      260  Star Wars: Episode IV - A New Hope (1977)
97       110                         Braveheart (1995)
123      150                          Apollo 13 (1995)
46        50                Usual Suspects, The (1995)
92       104                      Happy Gilmore (1996)
138      165         Die Hard: With a Vengeance (1995)
43        47               Seven (a.k.a. Se7en) (1995)
32        34                               Babe (1995)
176      208                         Waterworld (1995)
126      153                     Batman Forever (1995)
```

```python
print(get_similar_movies(4))
```

[107] ✓ 0.0s

```
     movieId                                   title  \
3          4                 Waiting to Exhale (1995)
100      113                  Before and After (1996)
153      181  Mighty Morphin Power Rangers: The Movie (1995)
151      179                         Mad Love (1995)
206      240                         Hideaway (1995)
209      243                            Gordy (1995)
187      219                        Cure, The (1995)
182      214       Before the Rain (Pred dozhdot) (1994)
180      212                      Bushwhacked (1995)
214      250      Heavyweights (Heavy Weights) (1995)
177      209                White Man's Burden (1995)
```

Q3)

```
[114]  ✓ 0.0s

...    Heat (1995)
       Seven (a.k.a. Se7en) (1995)
       Usual Suspects, The (1995)
```