Homework 5.0

1. 1 Problem 1

Load the Movielens 100k dataset (ml-100k.zip) into Python using Pandas data frames. Convert the ratings data into a utility matrix representation and find the 10 most similar users for user 1 based on the cosine similarity of the centered user ratings data. Based on the average of the ratings for item 508 from similar users, what is the expected rating for this item for user 1?

The converted utility matrix is shown below.

item_id	1	2	3	4	5	6	7	8	9	10	...	1673	1674	1675	1676	1677	1678	1679	1680	1681	1682
user_id																					
1	5.0	3.0	4.0	3.0	3.0	5.0	4.0	1.0	5.0	3.0	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	4.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2.0	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5	4.0	3.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

5 rows × 1682 columns

The User ID and Cosine Similarity of 10 most similar users to User1 is shown below.

User ID	Cosine Similarity
266	0.7716
134	0.7644
413	0.6990
905	0.6542
199	0.6507
32	0.6046
66	0.6034
526	0.6010
69	0.5992
120	0.5843

Based on the average of the ratings for item 508 from similar users, the expected rating for this item for user 1 is 3.67.

1. 2 Problem 2

Load the Movielens 100k dataset (ml-100k.zip) into Python using Pandas data frames. Build a user profile on centered data (by user rating) for both users 200 and 15, and calculate the cosine similarity and distance between the user's preferences and the item/movie 95. Which user would a recommender system suggest this movie to?

The user profiles on centered data for both user 200 and 15 are shown below.

```
User ID: 200
User Profile (Centered Ratings):
item_id
1    0.967593
2   -0.032407
3    0.000000
4    0.000000
5    0.000000
Name: 200, dtype: float64

User ID: 15
User Profile (Centered Ratings):
item_id
1   -1.875
2    0.000
3    0.000
4    0.000
5    0.000
Name: 15, dtype: float64
```

The cosine similarity and distance between the user's preferences and the item/movie 95 are shown below.

```
User 200:
Cosine Similarity   : 0.0768
Euclidean Distance  : 12.8811
User 15:
Cosine Similarity   : 0.0000
Euclidean Distance  : 14.1387
```

Based on the cosine similarity and Euclidean distance metrics, Movie 95 is more likely to be recommended to User 200.

This is because User 200 has a higher cosine similarity and a lower Euclidean distance with the movie' s profile, indicating a closer alignment between the user's preferences and the characteristics of the movie.