# Data Mining HW5.0

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?

①Here is the difference between Original and Centered user data:
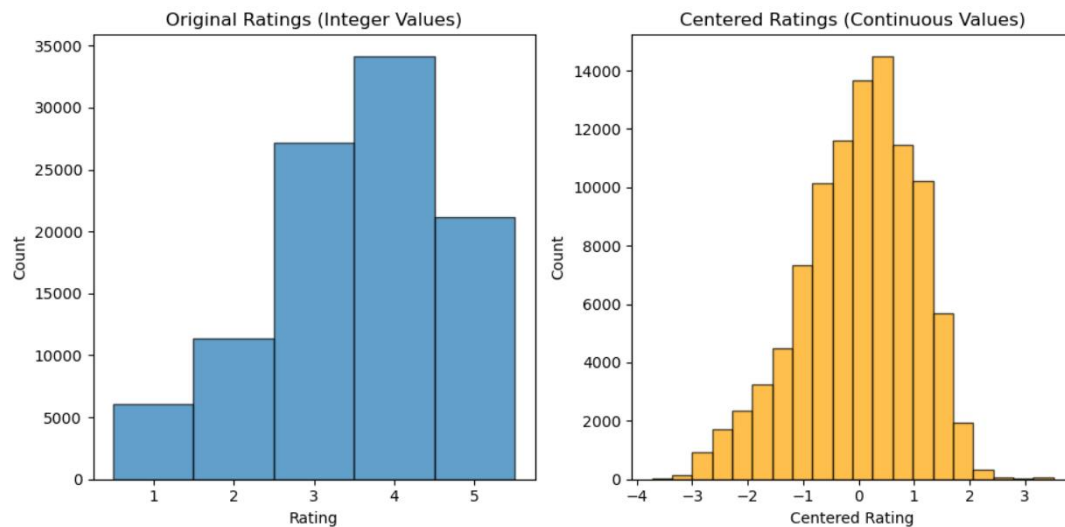
②We can get similarities for user 1 and get top 10 similar users:
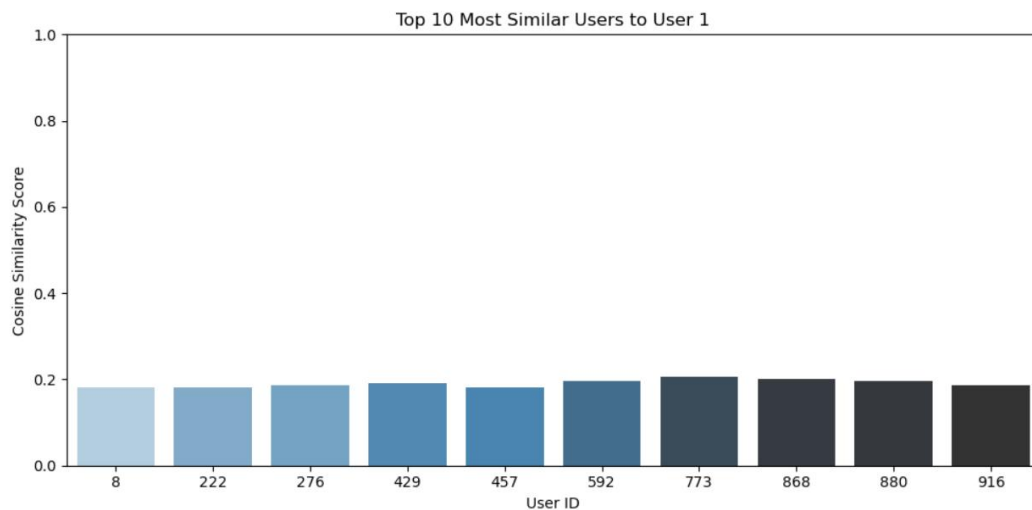
We can get the result:

```
The 10 most similar users to user 1 are: [773, 868, 592, 880, 429, 276, 916, 222, 457, 8]
The predicted rating for user 1 on item 508 is: 4.20
```

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?

Cosine Similarity and Distance:

```
User 15 - Cosine Similarity: 0.1215, Distance: 0.8785
User 200 - Cosine Similarity: 0.3647, Distance: 0.6353
The recommender system would suggest movie 95 to User 200
```

Analysis Results:
Based on the computed cosine similarity and distance metrics:
- User 15 has a similarity of 0.1215 and a distance of 0.8785 with Movie 95.
- User 200 has a similarity of 0.3647 and a distance of 0.6353 with Movie 95.

Key Findings
1. Similarity Comparison:

   -User 200's similarity (0.3647) is significantly higher than User 15's (0.1215), indicating that User 200's rating patterns align more closely with users who enjoyed Movie 95.

   -User 15's similarity is near zero, suggesting their rating behavior differs substantially from Movie 95's audience.

2. Distance Metric Validation:

   - The cosine distance (1 - similarity) confirms the similarity results.

   - User 200's smaller distance (0.6353) compared to User 15's (0.8785) further supports their stronger match with Movie 95.

3. Recommendation Decision:

   - The recommender system prioritizes higher similarity (or lower distance).

   - Therefore, Movie 95 should be recommended to **User 200**.