Problem1

Load the Movielens 100k dataset (ml-100k.zip) into Python using Pandas data frames.

```
Rating data:
   user_id  item_id  rating  timestamp
0      196      242       3  881250949
1      186      302       3  891717742
2       22      377       1  878887116
3      244       51       2  880606923
4      166      346       1  886397596
```

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.

```
Utility matrix (top 5 users and top 5 items):
item_id    1    2    3    4    5
user_id
1        5.0  3.0  4.0  3.0  3.0
2        4.0  NaN  NaN  NaN  NaN
3        NaN  NaN  NaN  NaN  NaN
4        NaN  NaN  NaN  NaN  NaN
5        4.0  3.0  NaN  NaN  NaN

The top 10 users most similar to User 1:
user_id
773    0.204792
868    0.202321
592    0.196592
880    0.195801
429    0.190661
276    0.187476
916    0.186358
222    0.182415
457    0.182253
8      0.180891
dtype: float64
```

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

```
Ratings for item 508 from similar users:
user_id
592    5.0
880    4.0
429    4.0
276    5.0
222    3.0
Name: 508, dtype: float64

User 1's expected rating for item 508 is: 4.20
```

Problem2

Load the Movielens 100k dataset (ml-100k.zip) into Python using Pandas data frames. (Same as P1) .

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.

```
User 200's profile (partial data):
item_id
1    0.967593
2   -0.032407
3    0.000000
4    0.000000
5    0.000000
Name: 200, dtype: float64

User 15's profile (partial data):
item_id
1   -1.875
2    0.000
3    0.000
4    0.000
5    0.000
Name: 15, dtype: float64

Vector representation of item 95 (partial, around item 95):
item_id
93    0
94    0
95    1
96    0
97    0
dtype: int64

User 200 and Item 95:
  Cosine Similarity: 0.0768
  Euclidean Distance: 12.5634

User 15 and Item 95:
  Cosine Similarity: 0.0000
  Euclidean Distance: 13.6519
```

Which user would a recommender system suggest this movie to?

```
Recommendation Decision:
Based on cosine similarity, item 95 is more likely to be recommended to User 200.
Based on Euclidean distance, item 95 is more likely to be recommended to User 200.
Overall, the recommender system would suggest this movie to User 200.
```