

Ratings of online statistics courses: 4 = Best, 1 = worst, blank = not taken

**In R Your Job is To:**

a user-based collaborative filter. This requires computing correlations between all student pairs. For which students is it possible to compute correlations with E.N.? Compute them.

Then:

Which single course should we recommend to E.N. based on the single nearest student to E.N.?Explain why.

Based on the cosine similarities of the nearest students to E.N., which course should berecommended to E.N.?

What is the conceptual difference between using the correlation as opposed to cosine similarities?[Hint: how are the missing values in the matrix handled in each case?

**Then:**

With large datasets, it is computationally difficult to compute user-based recommendations in real time, and an item-based approach is used instead. Returning to the rating data (not the binary matrix), let’s now take that approach.

If the goal is still to find a recommendation for E.N., for which course pairs is it possible and useful to calculate correlations?

Just looking at the data, and without yet calculating course pair correlations, which course would you recommend to E.N., relying on item-based filtering? Calculate two course pair correlations involving your guess and report the results.

Finally:

Apply item-based collaborative filtering to this dataset (using R) and based on the results, recommend a course to E.N.