Assignment #13 Collaboration Filtering by Joshua Troup

**Q1. First consider a user-based collaborative filter. This requires computing correlations between all**

**student pairs. For which students is it possible to compute correlations with EN? Compute them.**

CORR (EN, LN) = .870388

CORR (EN, DS) = .003536

Student EN is nearest to LN

**Q2. Based on the single nearest student to EN, which single course should we recommend to EN?**

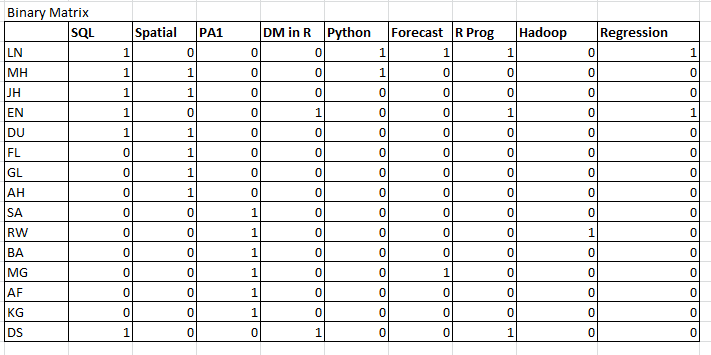
**Why?**

Since EN is closest to LN (.870388), LN has taken Python which EN has not. The course Python has been highly rated by LN so with the correlation being so high, Python is recommended for EN.

**Q3. Replace the ratings with a binary matrix indicating whether a student took the course or not.**

**Compute the cosine similarity of EN. from each of the other students for which such a calculation is**

**feasible.**



Cos Sim (LN, EN) = .6708

Cos Sim (DS, EN) = .8660

**Q4. Based on the cosine similarities of the nearest students to EN, which course should be**

**recommended to EN?**

The two closest students are LN and DS, then JH and DU which are tied. K=4 since they’re the closest users to EN. The course most purchased is Spatial Statistics and is recommended to EN.

**Q5. Just looking at the data, and without yet calculating course pair correlations, which course would**

**you recommend to EN, relying on item based filtering?**

SQL and Spatial match the best with more co-rated items. Spatial would be recommended based on item filtering.