Stage 5

Dan Howe and Keith Funkhouser
CS838: Data Science
May 7, 2017

As discussed in Stage 1, we obtained a dataset containing CS course evaluations between the years of 2008 and 2013. The file contains 858 tuples, each of which summarizes the course evaluations for a given offering of a course. The schema of the table can be found in the first row of the table below.

`Q1` through `Q8` are the average (over all responders) responses, on a scale of 1 to 5, for each of the questions asked on the course evaluations (shown below the table). `Q_avg` is the average of these values over the eight questions.

Four examples of tuples are:

Semester	**Instructor**	**Course**	**Q1**	**Q2**	**Q3**	**Q4**	**Q5**	**Q6**	**Q7**	**Q8**	**Q_avg**
Fall 2009	Li,Jessie	CS310	2.11	2.89	2.78	2.56	3.33	3	3	2.11	2.72
Spring 2010	KADARUNDALAGI	CS302	2.19	2.75	2.31	1.67	1.33	1	2.62	1.67	1.94
Spring 2013	Viswanathan,G	CS310	2.33	2.79	2.79	2.67	3.29	3	3.96	2.04	2.86
Spring 2010	PELOQUIN	CS310	2.4	2.8	3.07	2.36	3.75	3.25	3.36	2.67	2.96

Questions asked:

- Q1 The usefulness of the instructor's presentations in learning the course matter
- Q2 The instructor's command of the course material
- Q3 The instructor's responsiveness to student questions
- Q4 The enthusiasm shown by the instructor in teaching
- Q5 The availability of the instructor for out-of-class contact
- Q6 Value of handouts prepared by instructor
- Q7 Would you recommend the course?
- Q8 Would you recommend the instructor?

The analysis task we performed was an OLAP-style aggregation of the ratings for each instructor in order to provide, for each instructor, a single estimate of student satisfaction. To do this, we essentially performed an aggregate operation to average the ratings of courses offered by each instructor. The top of the results, sorted by average rating, were:

Instructor	Average - Q_avg	
Brown	4.8825	1 course (302)
Arpaci-Dusseau,Remzi	4.82375	
Cobian	4.7865625	4 courses (302)
Naughton	4.7715	5 courses (564/764/784)
DeSmet	4.76	1 course (368)

Among these, 3 taught 5 or fewer courses (Brown, DeSmet, and Naughton). The remaining professor, Remzi Arpaci-Dusseau, has received most of the undergraduate teaching awards available at the university and is highly regarded by students.

When we look at the average ratings by course, interesting trends emerge. The following are the top courses, sorted by average rating:

Course	**Average - Q_avg**	**Course Title**
CS900	4.92375	Advanced Seminar in Computer Science
CS812	4.905	Arithmetic Algorithms
CS810	4.89	Mathematical Techniques for Analysis of Algorithms
CS747	4.855625	Advanced Computer Systems Analysis Techniques
CS402	4.77875	Introducing Computer Science to K-12 Students
CS809	4.755	Mathematical Techniques for Analysis of Algorithms
CS730	4.745	Nonlinear Optimization II
CS635	4.72	Tools and Environments for Optimization
CS764	4.7088888889	Topics in Database Management Systems
CS710	4.6990625	Computational Complexity
CS736	4.6975	Advanced Operating Systems

Among these, nearly all are graduate level courses, indicating overall satisfaction of graduate students with the teaching in the CS department. Specifically, many of the courses are in the Theory group, as well as significant representation by the Systems group.

Notably, we learned that CS302 (Introduction to Programming) and CS577 (Introduction to Algorithms) are some of the lowest-rated courses at 4.03 and 4.00 average ratings, respectively. We conclude that significant work can be done in order to increase student (primarily undergraduate CS major) satisfaction for these courses. However, this is a fairly high bar, as a 4.00 on the rating scale ("1 = DEFINITELY NOT ... 5 = YES, DEFINITELY") is already quite high.

If we had more time, we would like to determine if there are trends in instructor ratings over time. For example, do an instructor's ratings increase or decrease as he or she gets more experience (teaches more courses)? Does a given instructor have higher ratings in spring or fall semesters? Do certain events in an instructor's career, such as tenureship or the beginning/ending of a research project influence their ratings? This last question would require more data acquisition.