# Mini project: VLookup and Grading Distributions

We know that Professor Darsow does grading differently than many other faculty members, allowing students to propose their own midterm and final grades, with justification. He is often asked questions about his grading schema, including “do students’ proposed grade distributions look much like the grade distributions in other classes?” Let’s see what we can find out in Excel!

### Filling in missing data

1. Access Eric Darsow's grading history (darsow\_ccac\_grade\_record.ods), and save it to your own machine. (You can use .ods or .xlsx—LibreOffice or Excel, respectively—whichever you’re most comfortable with.)
2. Study the data dictionary and the data sheets to conceptualize what the data set contains.
3. Use [VLOOKUP](https://support.office.com/en-us/article/vlookup-function-0bbc8083-26fe-4963-8ab8-93a18ad188a1) functions in your spreadsheet software of choice to populate the missing fields in the rows for both the mid-term grades and the final grades. Make sure to use the lookup tool instead of hand-classifying the data. Imagine there are millions of rows!
4. Once you have populated data in the core spreadsheet, copy the values into a new sheet such that there are no more formulas and just "plain" numbers and letters to work with.

### Comparing with external data

1. Access the anonymized spreadsheet of grades from a past semester of a course graded more conventionally, with points awarded for each assignment throughout the semester (grade\_comp\_student\_grades.ods).
2. Add the comparison grades to your spreadsheet with Professor Darsow’s grades. You’ll probably want to create a separate tab for the comparison data. (Do you see any problem with comparing these two sets of grades? If so, what tools do we have available to make comparison easier?)
3. Create a [pivot table](https://support.office.com/en-us/article/create-a-pivottable-to-analyze-worksheet-data-a9a84538-bfe9-40a9-a8e9-f99134456576) that computes the mean and standard deviation of grade points for each class’s final grades.
4. Create a frequency distribution (a histogram) which graphically shows the distribution of the midterm and final grades for at least two of the classes (Probably CIT-111, since there are the most students, and perhaps CIT-100 which was offered online.)

### Answering questions with the data

1. Comose a set of defensible statements comparing the grade distribution in each course based on the data revealed in the spreadsheet. Are the data skewed in a certain direction? Which course without point-based grading is most similar to the point-based counterpart (the comparison data)? Which is most different?