

HW01

1. Recall your course-specific PIN (a two-digit number, refer Grade Center). Select the corresponding university (with rank same as your PIN) from the list at <https://www.topuniversities.com/where-to-study/north-america/united-states/ranked-top-100-us-universities>; if there's a tie, treat them in the order given, e.g., two universities are ranked #30, so use #30 for the first and #31 for the second.
2. Look up the webpage of the selected university's computer science department (if there's no CS department, use the most similar department, e.g., EECS or CSE or CE).
3. Find ten of the current faculty members whose ranks are Assistant Prof., Associate Prof., Prof., or Distinguished Prof. (Don't select those whose titles are Lecturer, Adjunct Prof., Emeritus Prof., Teaching Prof., or Prof. of Practice.), whose first & last name initials are not identical.
4. Go to the Google Scholar Citations page for each of those ten faculty members, and list the following information for each: univ. rank, first initial, last initial, numbers of citations in each of the years 2017-2022, h-index, and i-10 index. For example, if your PIN is 22, then you may have the following entry for Prof. Lav Varshney from Univ. of Illinois: 22, L, V, 633, 688, 732, 861, 946, 1054, 35, 100
5. Combine all ten records into a CSV file with the name of the university, e.g., Illinois.csv
6. Generate a noisy version of this file, with a different filename preceded by "noisy", e.g., noisyIllinois.csv, in which each of the citation counts is randomly modified by adding an integer in the range [-10, +10], e.g., the previous entry may be modified to 22, L, V, 631, 678, 742, 866, 942, 1050, 35, 100
7. Generate a missing data version of the noisy file, with a different filename preceded by "missing", e.g., missingnoisyIllinois.csv, randomly omitting one of the citation numbers in each record, e.g., the previous entry may be modified (omitting the third citation number entry) to 22, L, V, 631, 678, , 866, 942, 1050, 35, 100
8. Submit all three files.

Due Date 9/13/2023 11:59PM