

## Data Science Techniques (MAT 339)

### Homework 4 - - 10 Points

Submit a **hard copy** of your work at the beginning of class on Wednesday, February 11th. There is no **electronic submission** for this homework.

As with the previous homework, since we will be printing, we will try to conserve ink again by not using screen shots of the terminal or folders in dark mode. You can either change the terminal and folder windows to light mode through options, or copy the command you issued and the contents of its output into one document you use to organize the homework assignment. This can be LaTeX, Word, markdown, text, etc. Just be sure it has a white or light background so as not to overuse printer toner.

1. Pick a **csv** file that is of interest to you and large enough that basic data processing produces non-trivial results (for example, filtering or sorting yields more than a handful of rows). Use **csvkit** tools from the command line to construct four different examples where you inspect and process your file. You must use each of the following commands in at least one of your examples: **csvstat**, **csvlook**, **csvcut**, and are encouraged to use others such as **csvsort**, **csvstack**, **csvgrep**. At least two of your examples must involve piping commands together using the **|** operator. In your submission:
  - Include the exact commands you ran.
  - Include the corresponding output, either pasted from the terminal or from files generated by your commands.
  - Briefly (1–2 sentences per example) explain what each command is doing and why it is useful for understanding your dataset.
2. Following the guidance from the *structured-data-queries-terminal* class slides, open **sqlite3** from the terminal.
  - (a) Load a database (you may use *feline.db* from the *structured-data-queries-terminal* class folder in the repo), run a query on one of its tables and display the results.
  - (b) Look up two additional commands to use in **sqlite3** either from the *structured-data-queries-terminal* slides or the references linked to from those. Execute those commands in **sqlite3** in the terminal.

Include the commands you ran and the results as your answer to this question. Avoid dark screen shots.