67978: A Needle in a Data Haystack
Introduction to Data Science
Homework #3: Project
Due: **16 September, 2025 11:59pm, on Moodle (No late days)**

1 Instructions

Each group should submit two pdf files (writeups, see below) and one zip file (code). Filenames should include your **group member initials** (for example, group members DS and TH should turn in: writeup_DS_TH.pdf, noimages_DS_TH.pdf , code_DS_TH.zip). If you have links to a video or a demo, please make sure they are accessible to us (if on google drive – share with dafna.shahaf@mail.huji.ac.il). Points will be deducted for submissions which fail to comply.

1.1 Writeup

Writeup should address (in this order):

- Project title

- Team member info: name, email address, CS id

- Problem description: Give a brief but precise description or definition of the problem or hypothesis you set to evaluate.

- Data: What data did you use? Briefly describe the data, its size (number of records, number of GBs) and where you got it.

- Your solution! Tell us what method, algorithm, technique you used and how you modified them or scaled them up to your size of the data (if needed). Be as specific as you can, explain how parameters were chosen, etc.

- Evaluation:

    - Evaluation Criteria: Describe how you measure performance or success of your method. How do you define success? How did you make sure that your findings are not chance (significant)? *(Note: If your project includes a prediction model, against what baseline methods do you compare your algorithm? How did you obtain ground-truth labeled data so that you can then measure accuracy, precision, recall or some other metric?)*

    - Setup: Describe how did you setup your experiments. If there are human subjects, how did you recruit them? How would you avoid biases?

    - Results: Describe your evaluation results.

    - Visualization: Visualize the results (and explain your viz choice).

    - Impediments: What were some of the issues you encountered in your evaluation? How did you deal with them?

- Future Work: What extra fun things can be done? (Not by you, don't worry)

- Brief conclusion

Page limit: Your writeup (writeup_DS_TH.pdf) should be **5 pages (not taking figures into account)**. Please submit the same file without the images (noimages_DS_TH.pdf) so we and you know it's actually 5 pages (sorry about that).

67978: A Needle in a Data Haystack
Introduction to Data Science
Homework #3: Project
Due: **16 September, 2025 11:59pm, on Moodle (No late days)**

1.2 Code

Code needs to be readable (documentation is good for you). Pick whatever language you're comfortable with. If your project includes a demo (recommended!), it is highly recommended to either get it online (and provide the link) or record a video screen capture of you using it. There's a good chance your code won't run as-in on our computers, and we will not chase obscure package errors. :)

2 Grading

For your reference, you will eventually be graded on:

- Overall ambition, creativity, difficulty of project (15%)

- Execution and implementation (50%)

- Use of appropriate and compelling visualizations (25%)

- Writing (10%)