67978: A Needle in a Data Haystack
Introduction to Data Science
Homework #5: Project
Due: **8 March, 11:59pm, on Moodle (No late days)**

THE HEBREW UNIVERSITY OF JERUSALEM

1 Instructions

Each group should submit one pdf (writeup) and one zip file (code). Filenames should include your **group number** (for example, writeup_group15.pdf , code_group15.zip). If you're submitting handwritten answers, make sure they are crystal clear (both the handwriting and the scan – please avoid shaky cell pictures taken in a dark room :)). If you have links to a video or a demo, please make sure they are accessible to us. Submissions which fail to comply will not be graded.

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.
- Experiments:
  - 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 experimental results.
  - Visualization: Visualize the results (and explain your viz choice).
  - Impediments: What were some of the issues you encountered in your experiments? 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: 6 pages (not including figures).

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 (CamStudio used to work for me, but it's been a while). 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 (20%)
- Written part (15%)