DATA 601 : Project 1

Fall 2025

This will be a group project where you would use introductory data science skills on a dataset to gather, process, visualize and present insights from that dataset. Students are required to form groups of size 3-4. Any groups smaller or larger will be merged or broken respectively on the discretion of the instructor.

Each group would need to pick a topic for the project. Essentially what you will need to do it:

- Gather data regarding the project (web-scraping, readily available data, APIs, etc).
- Visualize the data.
- Draw meaningful insights from the data.

NOTE: Use of external tools like Power BI or Tableau for visualization is only acceptable if majority of your diagrams are created using python based packages (matplotlib, seaborn, etc)

Other criteria required in the project (3 out of 12):

1. Object oriented code
2. Intricate file management/manipulation (not just simply reading and writing of 1-2 files)
3. Data cleaning and/or Data transformation
4. Numerical computation / Mathematical processing
5. Regular Expression
6. Intricate data gathering process (e.g through Web-scrapping)
7. Statistical analysis
8. Linear Regression and variants (e.g ridge regression, lasso regression)
9. Logistic Regression and variants (e.g perceptron)
10. Graphical User Interface
11. Using APIs in your project.
12. One ML model which is not linear or logistic regression (or variants of them). If you use this you will have to explain what and how the model works. If you have questions about this, please ask the instructor.

Deliverables:

Week 5: Proposal:

For Week 5, with respect to the Project, this is what is going to be expected of you per group:

(i) A 1-2 page document (in the IEEE conference template) that covers the following:
   i. Title
   ii. Introduction : Tell us about the problem.
   iii. Proposed project: What are you going to do, and what is the timeline. Mention how are you planning to meet (at least) 3 out of 11 criteria.

iv.   Related work : What other work is out there around this topic, if something is similar, how is yours different (we are not looking at Ph.D level research, just some small differentiability).
    v.   References cited in the document.
(ii)   A 5-7 minute presentation covering the above

Week 10 : Progress

For Week 10, with respect to the Project, this is what is going to be expected of you per group:

(i)   A 2-3 page document (in the IEEE conference template) that covers the following
    i.   Title
    ii.   Introduction
    iii. Background : background information on the concepts involved in the project, you can have different subsections for individual concepts.
    iv.   Related work : What other work is out there around this topic, if something is similar,
    v.   Current Progress : describe what has been done so far (add further sections if needed)
    vi.   Acknowledgments: List the group members and their current contributions to the project, and how you are using AI tools in your project.
    vii. References cited in the document.
(ii)   A 10 minute presentation covering the above.

Week 15 : Final week

(i)   A 3+ page document (in the IEEE conference template) that covers the following
    i.   Title
    ii.   Introduction
    iii. Background: background information on the concepts involved in the project, you can have different subsections for individual concepts.
    iv.   Related work: What other work is out there around this topic, if something is similar,
    v.   Content: <label sections as necessary> describe your work, experiments and results (tables, plots etc). Mention how you have met the three criteria
    vi.   Future Work: suggest future work based on observations.
    vii. Acknowledgments: List the group members and their contributions to the project, and how you have used AI tools in your project.
    viii. Conclusion : Write a conclusion paragraph where you summarize your entire work (from idea, to results, to your final thoughts).
    ix.   References cited in the document.
(ii)   A 10-15 minute presentation covering the above.
(iii)   The final code

Peer evaluation:

At the end of the semester students will email me about their evaluations of their peers. They will grade their peers on a 0-20 scale about their contributions to the project. The final peer evaluation grade will be the average of the grades a student gets from all their peers.

Weightage of those deliverables :.

- Week 5: Proposal (10% of project)
- Week 10: Progress (30% of project)
- Week 15: Final (40% of project)
- Peer evaluation (20% of the project)

Final Code:

- 1 main_notebook.ipynb, you may have other python scripts that you may import in your main ipynb. Please use markdown cells in your jupyter notebook to explain the code beyond just comments.
- README file if necessary to tell me which code file contains what.