Introduction to Computing 250.205

Project Requirements

The Project is 8% of the total grade.

The goal of this project is to identify a topic and an interesting question, obtain a data set and analyze the data in a simple fashion using the Python libraries covered in the second part of Python (Matplotlib, pandas, NumPy) to answer the question.

Your first task is to **identify a question and find a data set to answer the question.** Here are some suggestions:

- You may adapt this project to your current lab interests, i.e., you can use your own data, or somebody else's data. The data does not need to be published.

- Explore datasets available online.
  You can look at the provided file:
  Datasets-Links.pdf, where you can find links to data sources**.**
  You can also use data sets from other sources.

- You could repeat an analysis performed in a published paper. You can start by going to Google Scholar and typing in subject that you are interested in, or the name of a researcher that you are interested in.

- If you have trouble locating a dataset, I can help you with that.

When choosing your data set, keep in mind that you will be asked to:

- Perform a numerical analysis: e.g., fitting your data, calculate statistical parameters, like average etc. Note that the data analysis should be meaningful.
- Make plots. For example: temperature vs time; voltage vs time, number of women vs number men in education.

You should submit:

- **Word or pdf -** This document should contain your project, which should include text and figures. It is like a paper. See document outline below.
- **Python script(s) – MUST BE PY FILES – If you use Jupiter Notebook, make sure to download the py file.**
- **Data file(s)**
- **Do not submit any zip file.**

Grading scheme (100 points):

- **(5) Submit the Data file(s)**

- **(55) Python script(s)**
  - (10) data import and (cleaning optional)

  - (20) Numerical analysis of the data.
    Example: a fit of the data, and possible prediction is acceptable as analysis.
    If you do not perform a fit, you need to perform some other sort of analysis with your data. Note that the analysis should be meaningful.

  - (20) Plots: any kind of plot: line, scatter, bar graph, histogram.
    Include errors if any.
    5 points for most basic plot.
    2 points for setting the figure size.
    5 points for title, axes labels (with units if any), legend, grid
    5 points for making the plot legible by setting:
    font size, marker size (if any), line width, line color.
    Text should be legible, and the figure should clearly show the data/results.
    3 points for saving the figure(s) in jpeg 300 dpi

  - (5) Write results to a text file (this would include fit results, or whatever you calculated for the analysis)

- **(40) Word/pdf document** – Here an outline of the document

OUTLINE of the document

Introduction to Scientific Computing
name last name, date

The number of paragraphs for each section is a suggestion.

(6) Abstract *1-3 paragraph*

Provide a summary of your findings. Here a short guideline
In this study, we conduct a comprehensive analysis of a [selected dataset] to explore [brief overview of the research question or objective]. Using [describe methodology or approach], we systematically examine the data to uncover [main findings or insights]. Our analysis offers insights into [specific aspects or trends], contributing to our understanding of [related field or topic].

This part might not apply to you - *Additionally, we introduce [any innovative approaches or techniques if any] to enhance our analysis. This research not only sheds light on*

[specific phenomenon or aspect] but also underscores the importance of data-driven inquiry in [relevant field or discipline].

(4) Background *1-3 paragraphs*

Briefly introduce the context of the dataset and its relevance to the research question. Discuss any pertinent background information necessary for understanding the dataset and the analysis process.

(6) Materials and Methods. *1-3 paragraphs*

State the source of information you are using.
Describe the methods and techniques applied in the data analysis process, such as data cleaning, visualization, and statistical analysis (i.e., standard statistics, regression analysis). Present any equations used in the analysis.

(8) Analysis - include at least one plot/figure *~1 paragraph per figure / table.*

In this section show your tables and plots. Tell the reader what the reported results are in words, i.e., describe the tables and plots. Describe the key insights derived from the analysis without interpretation. In the next section you will interpret what you state here.

(8) Discussion *1-4 paragraphs*

Interpret the results of the analysis, discussing their implications and relevance to the research question. Compare findings with existing literature if applicable and propose potential areas for further exploration or research.

(5) Conclusions *1-2 paragraphs*

Summarize the main findings of the study and their significance.

(3) References

List any sources or references consulted during the project.