

Coding Challenge II

Learning Python Functions & Data Processing Relevant to Biomedical Questions

Due Tuesday, 10.13 at 11:59 pm

Please submit your answers in PDF form, and programs in .py format only. Screenshots can be saved and converted into a PDF.

NOTES: Justify each step in your code. Define the mathematical steps you took. Comment each line of code. Define input and output of your code. State all assumptions and simplifications.

1. **Using a Python program you create: Identify the two people in the Excel file ([Fit_Profile.2020.xlsx](#)) who are most similar across at least 5 numeric attributes.**
 - a. Which 5 numeric attributes did you pick? Justify why.
 - b. Did you normalize or weight the data beyond what was provided? Why or why not?
 - c. Define mathematically how you calculated similarity.
2. **Using Python code: Classify the people in the Excel file ([Fit_Profile.2020.xlsx](#)) via 5 numeric attributes into 4 groups, where each group contains people who are most similar to each other.**
 - a. Which clustering method did you use? Justify why.
 - b. Define mathematically how you calculated similarity.
3. **Plot in a heatmap (Biowheel or your choice of other data visualization) the data from your 4 cluster groups (Question 2) for the 5 numeric attributes.**
 - a. Briefly justify your choice of formatting and color scheme
 - b. How does the heatmap relate to your cluster groups?