Programming for Language Technologists – 5LN429
Assignment 4

Johan Sjons

October 14, 2024

Contents

1 Introduction

In this assignment, you are given three fairly disparate tasks that aim to give you relevant practice in Python programming.

As usual, the assignment is anonymous. In other words, no information about who wrote the code is allowed.

1.1 Automatic correction

For automatic grading, follow instructions closely. Incorrect functions or file names will require re-submission.

1.2 Naming of files

Name each module (i.e., Python program) to `ans` followed by the number of the question and the `.py` extension. For example, the file with the code for question one should be `ans1.py`, the file with the code for question 2 should be `ans2.py`, and so on.

2 Assignment questions

2.1 Task 1

Write a function eq1() that takes a list as an input parameter and that is an implementation of the following equation:

$$Equation1 = \sum_{i=1}^{13} \frac{(x_i - x_i^3)^2(n-1)}{\sqrt{x_i}} \tag{1}$$

where x is a vector of numbers (a list of integers in the code) and n the number of elements in that vector.

2.2 Task 2

Write a program that tests the user on Swedish words. It should be possible to run the program and pass in a `.csv`-file using `sys.argv[1]`, in which commas separate English words and their Swedish translations. You can use any `.csv`, but if the program works with `body.csv`, `colors.csv` or `verb.csv`, it is sufficient.

The program should keep track of the correct and incorrect answers and report the results to the user. The program should also end each go with whether the user wants to try again. Follow PEP8, adhere to the "single responsibility principle" and DRY, and comment on your code properly. The program should not crash.

If you want color coding, you can use GLOBAL CONSTANTS (not to be confused with global variables).

2.3   Task 3

Using the knowledge you gained from Lab 6, conduct a small investigation using a dataset of your choice. Use any library you want (e.g., `pandas, seaborn, vega, altair, etc.`) Comment your code thoroughly (both regular comments and, where relevant, docstrings), and make sure you understand the code if you copy it from online.

Furthermore, write a short research report. No less than half a page, no more than one page. It is a mini-thesis, in a sense. One or a few sentences per section is all that is needed. The report must contain the following sections:

- Abstract

  - This should be a summary of your report

- Introduction

  - This should be a (somewhat) informal motivation for your work leading to a research question. The motivation need not be well-grounded and can, in fact, be semi-fabricated because the goal for Task 3 is not to conduct research but to learn a method and to practice reporting results.

    For example, you could say something like:

    "It has been shown that there is a correlation between weight and height. However, it is unclear whether this holds for every possible dataset and whether the relation is linear on the original scale. Thus, the research question is:

    Is there a relation between weight and height on the logarithmic scale?"

- Background

  - Find at least one study that reports results on the same research question or something close.

- Data and Method

  - This section describes the data (remember to include a link to the dataset in a footnote) and what you have done, basically taking this dataset and conducting a linear regression analysis.

- Results

  - This section should just describe what you found, with no interpretation other than saying what the R^2 value, p-value, intercept, and coefficients mean. Make use of the tutorial you saw in Lab 6 if you have difficulties understanding the meaning of these notions.

    You will include a plot of your data with a regression line.

- Discussion

  - Give a short discussion on how the results should be interpreted, for example, the implications. Again, nothing grand needs to be said.

- Conclusion

  - Summarize your findings in one sentence.

- References

  - A list of references

3   LaTeX

You can write the rapport in Word or any other word processor. However, if you want to step up your game, you should use **LaTeX**, which is something you need to learn later in the Master's programme anyway.

LaTeX is a typesetting system commonly used for creating and formatting documents, especially in academic and scientific writing.

LaTeX differs from traditional word processors like Microsoft Word in that LaTeX uses a markup language to define the structure and formatting of a document rather than a graphical interface where you directly manipulate the document's appearance. LaTeX is highly customizable and well-suited for documents with mathematical equations, scientific notation, and bibliographies.

See the assignment_4_rapport_template.tex for more information.

HINT: Use Overleaf (an online LaTeX editor and compiler).

4   What to hand in

Upload a zip file called `assignment4_5LN429_2024.zip` on Studium that contains a directory. The directory should be named `assignment_4` and must contain `ans1.py`, `ans2.py`, `ans3.py`, along with the report. The report must be a `.pdf`-file, called report.pdf **The deadline is October 22, 2024.**