Data Science Template:

Go from zero to hero with a fully initialised template research project in a minute

Python for Lunch || 6 March 2025 || Franka Jesse

Outline

- Why use a Data Science Template?
- How to use a Data Science Template?
  - An example: Cookiecutter
  - Using the Makefile
  - Linking to Git
  - Do’s and Don’ts

Why use a Data Science Template?

- To avoid having a messy project folder that gets more chaotic as time evolves

Why use a Data Science Template?

- To avoid having a messy project folder that gets more chaotic as time evolves

> DATA
> MODEL_src
> MODEL_src_v2
> results_run1
> compile_clean.csh

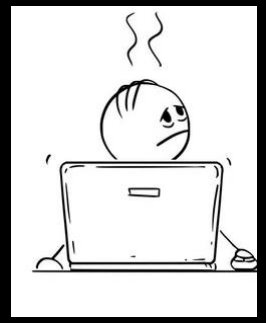
> OUTPUT

> OUTPUT_final
> OUTPUT_final_for_sure
> OUTPUT_final_for_sure_v2

> results_run1_v2
> results_run1_updated
> results_run2
> [...]

> ANALYSIS_output

> notebook_FIG1.ipynb
> notebook_FIG1_zoomed.ipynb
> notebook_FIG1_finalv5.ipynb
> notebook_FIG2_maybe_needs_update.ipynb
> [...]

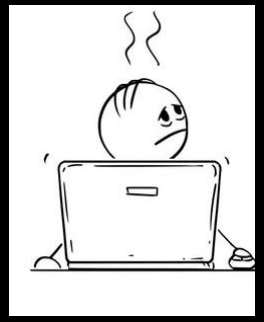


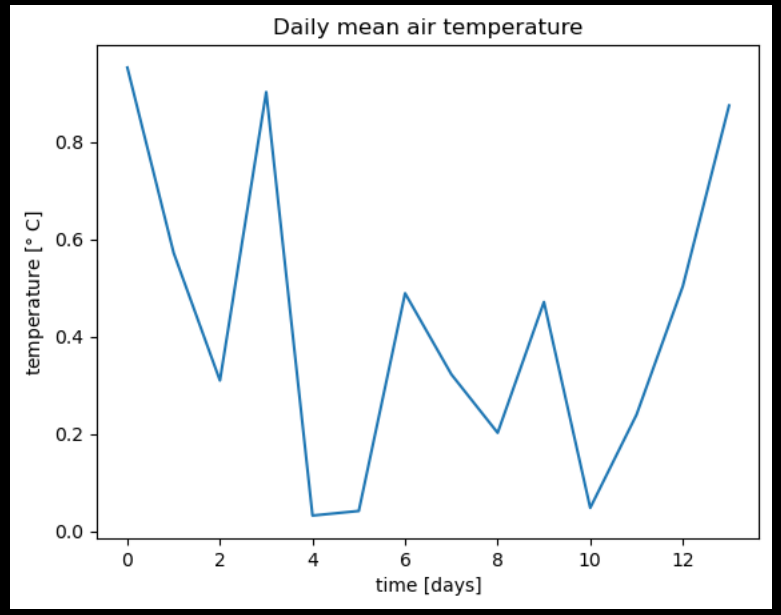
Why use a Data Science Template?

- To avoid having a messy project folder that gets more chaotic as time evolves
- To ensure reproducibility of your analysis and results

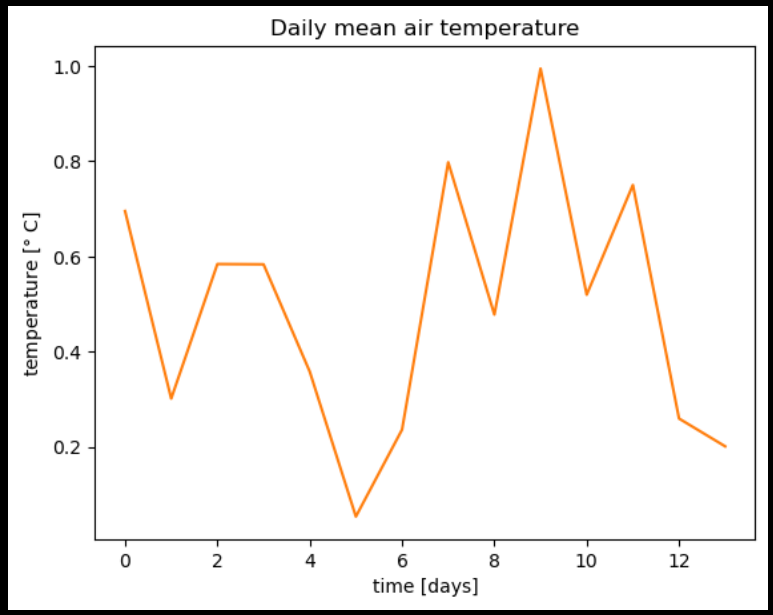
Why use a Data Science Template?

- To avoid having a messy project folder that gets more chaotic as time evolves
- To ensure reproducibility of your analysis and results





Plotting your results in January 2025



Adjusting the color in March 2025...

Why use a Data Science Template?

- To avoid having a messy project folder that gets more chaotic as time evolves
- To ensure reproducibility of your analysis and results

- Other people will thank you!
- Moreover, you will thank you!

Setting up your project structure using the ‘Cookiecutter Data Science’ tool

→ https://cookiecutter-data-science.drivendata.org/

→ terminal

Linking to Git

- The ‘.gitignore’ file is already there, just need to ‘git init’
- Github desktop

Using the Data Science Template

- Using the Makefile to setup environment
- Adapting the ‘environment.yml’ file
- Formatting code
- Source vs. notebooks

Do’s and don’ts

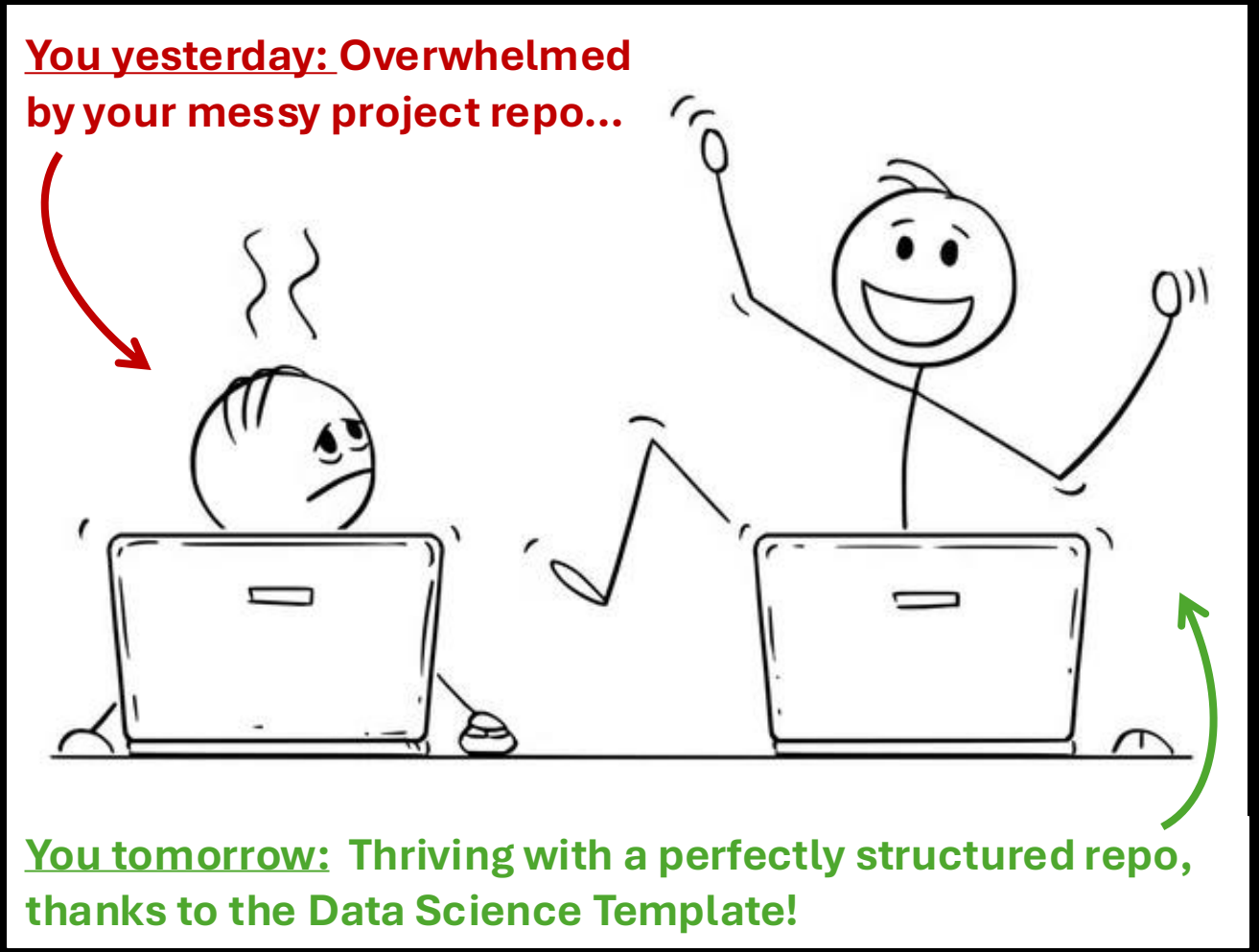
- Do’s:
  - Simplify the repository by removing unused folders.
  - Use source files for reusable code and notebooks for exploration.
  - Maintain organization with clear naming conventions and structured sub-folders.

Do’s and don’ts

- Do’s:
  - Simplify the repository by removing unused folders.
  - Use source files for reusable code and notebooks for exploration.
  - Maintain organization with clear naming conventions and structured sub-folders.

- Don’ts:
  - Rename folders carelessly, as it may break your Makefile.
  - Commit notebook changes without clearing outputs first.

Questions?



Find more details on Cookiecutter via:

https://cookiecutter-data-science.drivendata.org/