Software Engineering Practices for Data Scientists

Module 4: Software Engineering, Reproducibility, and Deployment Basics

Dr. Yves J. Hilpisch[1]

March 27, 2026

[1] Get in touch: https://linktr.ee/dyjh. Web page: https://thedatascientist.dev. Research, structuring, drafting, and visualizations were assisted by GPT 5.x as a co-writing tool under human direction. Comments and feedback are welcome.

Preface

This preface explains how Module 4 completes the core arc of `The Data Scientist`. The earlier books — *Python Programming Foundations for Data Science* [2], *Python Data Analysis, Visualization, and Storytelling* [3], and *Statistics, Machine Learning, and Model Evaluation* [4] — teach Python, data analysis, and baseline modeling. This final module, *Software Engineering, Reproducibility, and Deployment Basics* [5], turns those separate skills into cleaner, more reproducible, and more visible project work.

Many beginners can already write notebooks that answer a question. Fewer can turn those notebooks into a small project that someone else can run, review, and understand. Module 4 focuses on that gap.

The goal is not to turn the reader into a full-time software engineer. It is to build working habits that make data-science projects easier to trust, share, and present:

- organizing folders and files with intention,
- using Git and GitHub calmly rather than fearfully,
- separating exploration from reusable code,
- checking data and logic with small tests and assertions,
- and packaging the result as a portfolio artifact.

This module also introduces lightweight dashboards and deployment awareness. The scope remains practical. The reader does not need to become a cloud specialist. They do need to understand why environments, app entry points, and clear documentation matter in real projects.

The study rhythm remains familiar:

- read a short section,
- run or adapt a small example,
- reflect on what would make the work easier to reuse,
- and save the result in a project folder that could be shown publicly.

By the end of this book, the target outcome is simple. The reader should be able to say: this project is not only interesting; it is also organized, documented, reproducible, and ready to share.

Contents

List of Figures

Contact

Software Engineering Practices for Data Scientists

Module 4: Software Engineering, Reproducibility, and Deployment Basics

Get in touch:

linktr.ee/dyjh

thedatascientist.dev