The Data Scientist

From First Notebook to Portfolio

The Python Quants

April 12, 2026

Contact: team@tpq.io

Enrollment: https://thedatascientist.dev

Program Mission and Promise

The Data Scientist is a self-paced, portfolio-driven training program by The Python Quants for absolute beginners, early-stage practitioners, and students from all disciplines who want to build practical Python and data science skills from the ground up [6]. A primer plus four tightly edited books, runnable labs, and weekly coaching guides stitch together into one learning path: from first Colab notebook and Python script, through calm data analysis and visualization, into statistics and machine learning, and finally to professional project habits with tests, environments, and dashboards [1, 2, 3, 4, 5]. The same visual language and house style span every module so delegates stay oriented while the technical stakes rise.

The program exists to answer a concrete question: *How do we help motivated beginners, working analysts, and graduate or post-graduate students turn their curiosity into interview-ready skills and a visible portfolio without overwhelming them with tools or theory?* Each module blends short explanations with focused exercises, labs, and coaching notes that support self-study, peer learning, and internal enablement. Delegates can join at different points depending on prior experience, yet the path rewards a full run-through because skills and portfolio artefacts keep compounding across the books.

TDS also serves as the natural basecamp for deeper journeys with The Python Quants: once delegates are comfortable with Python, data work, and project structure, they are well prepared to move into specialised programs such as *The AI Engineer* (`https://theaiengineer.dev`), *The Crypto Engineer* (`https://thecryptoengineer.dev`), or the flagship finance track built around the CPF program (`https://python-for-finance.com`).

Learning Arc at a Glance

Learning Outcomes

Ramp-Up — Primer [1] removes setup friction by guiding delegates through Colab, local Python environments, Git/GitHub, and first notebooks so that later modules can focus on skills rather than tooling surprises.

Module 1 — Python Programming Foundations for Data Science [2] builds solid Python fundamentals for data work: core types and control flow, functions, text and file handling, and early problem-solving patterns that show up in interviews and everyday scripts.

Module 2 — Python Data Analysis and Visualization [3] introduces NumPy, pandas, grouping, and storytelling plots around concrete datasets, turning messy tables into clear summaries, charts, and first narrative reports.

Module 3 — Statistics and Machine Learning with Python [4] develops practical statistical intuition, baseline models, and evaluation habits so delegates care about metrics, overfitting, and leakage before chasing complex architectures.

Module 4 — Software Engineering Practices for Data Scientists [5] turns notebooks into small, well-structured projects with tests, environments, and dashboards, closing the core TDS arc with portfolio-ready project work.

Program Scale

A few concrete numbers help anchor what you get with TDS:

- **5 books** (primer plus four coherent modules) that form the spine of the learning path.
- **9 labs** and focused drills that turn book chapters into runnable practice.
- **6 coaching guides** that translate the material into week-by-week study plans.
- **25 Colab-ready Jupyter notebooks** for hands-on exploration, from first scripts to full projects.
- **44 Python scripts** that reinforce habits around small, testable utilities and workflows.

From Comprehensive Theory to Robust Practice

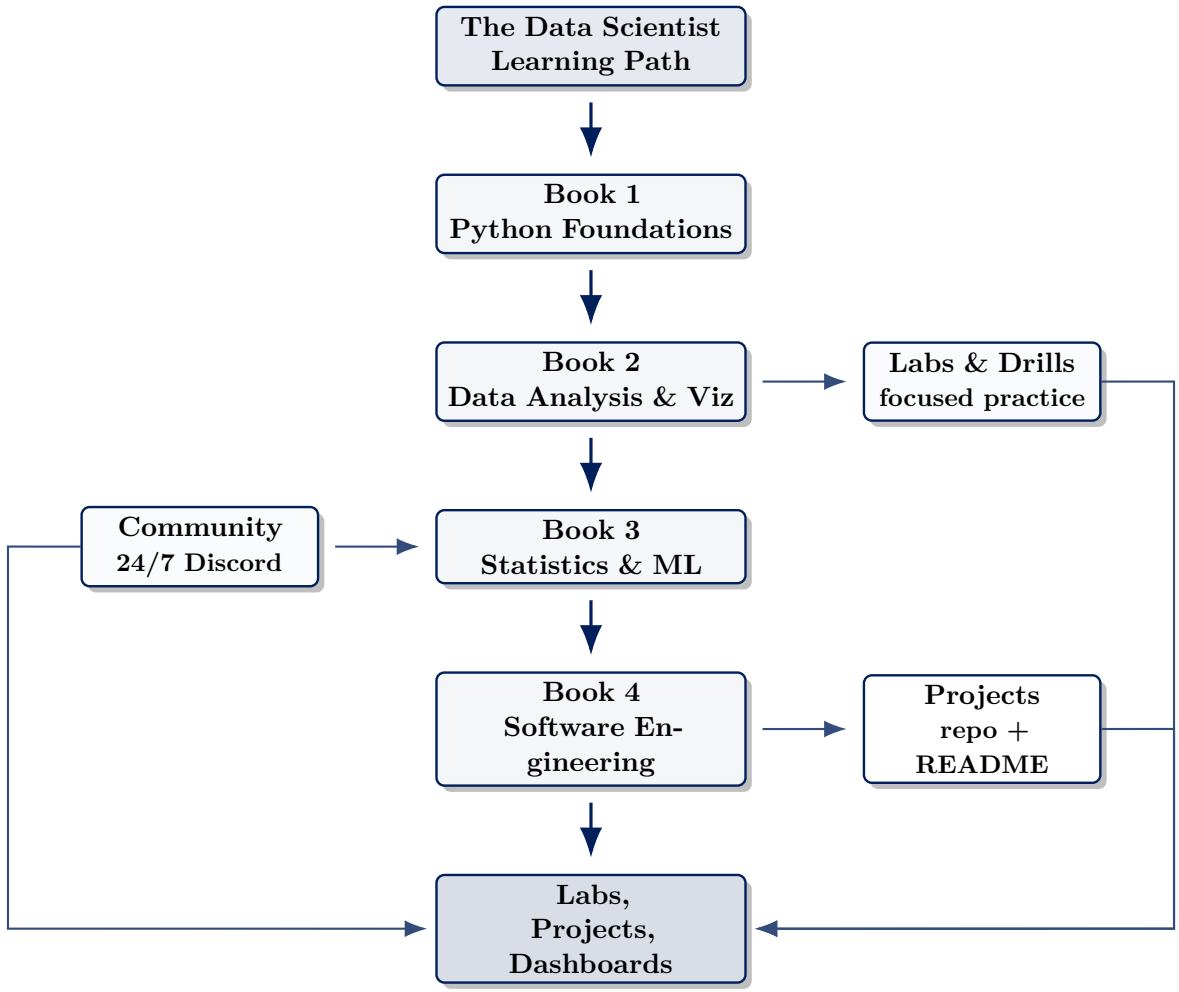

Figure 1: Each module in *The Data Scientist* feeds into labs, projects, and dashboards, creating a clear pipeline from first notebooks to portfolio-ready work.

The broader learning journey treats the books as an always-on reference layer. Labs, notebooks, and coaching notes lean on the same notation, examples, and diagrams, allowing delegates to

oscillate between self-paced reading, focused drills, and project work without constant context switching.

Exclusive Early Access

Delegates receive early access to the upcoming third edition (700+ pages) of Dr. Yves J. Hilpisch's O'Reilly book *Python for Finance*, providing a deep reference that complements the TDS curriculum and supports long-term practice.

Practice Labs

Beyond the primer and four core books, *The Data Scientist* ships a growing catalogue of short notes that double as lab scripts [12, 13, 14, 15, 16, 17, 18, 19, 20]. Each lab focuses on a tightly scoped skill and comes with runnable notebooks or Python scripts so you can rehearse critical moves before applying them in interviews, day-to-day work, or capstone projects.

How the Labs Help

Taken together, these labs equip you with modern, practical data-science habits so you can:

- move from "I understand the chapter" to "I can run the drill in Colab or locally";
- connect exercises back to the exact chapters and coaching notes they reinforce, keeping learning loops short and repeatable;
- build a small but coherent portfolio where each project has clear data, code, and narrative that supports job-market conversations and interviews.

Ramp-Up and Module 1 Labs

This first cluster deepens the primer and early chapters of *Python Programming Foundations for Data Science* with hands-on text, file, and counting practice.

- *Workspace and Colab Check* [12] verifies that Colab, local Python, and GitHub are ready so later labs can focus purely on Python and data work.
- *Text and File Clinic* [13] walks through turning a small log file into a cleaned summary, reinforcing string methods, file handling, and problem decomposition.
- *Counting and Frequency Patterns* [14] drills list and dictionary patterns for counts, rankings, and simple aggregations that later reappear inside pandas.

Data and Modeling Labs

The second cluster aligns with *Python Data Analysis and Visualization* and *Statistics and Machine Learning with Python*, giving you a calm progression from messy tables to evaluated models.

- *Messy CSV to Clean Table* [15] focuses on turning a raw CSV into a tidy DataFrame with clear types and columns.
- *Storytelling with Plots* [16] exercises grouped summaries and charts around a concrete dataset, with an emphasis on narrative framing.

- *Baseline Models and Evaluation* [17] guides you through simple classification and regression baselines with metrics that reveal overfitting and calibration.
- *Overfitting and Leakage Clinic* [18] stages controlled mistakes so you can see, and then fix, common leakage and evaluation pitfalls.

Projects and Dashboards Labs

The final cluster leans on *Software Engineering Practices for Data Scientists* to turn notebooks into small, reproducible projects with first dashboards.

- *From Notebook to Project* [19] shows how to extract core logic into a `src/` layout, add tests, and wire up a simple CLI for your analysis.
- *Dashboard Quickstart with Streamlit* [20] walks from an existing analysis to a minimal but useful Streamlit app, ready to demo as part of a portfolio.

How We Designed the Experience

1. **Continuity of voice and visuals.** Every chapter opens with the same navy headings, callouts, and diagram language so readers never waste energy re-learning the interface across primer, books, and labs.
2. **Stackable mental models.** From core Python patterns to pandas workflows, evaluation checklists, and project layouts, later books keep referencing earlier habits so intuition and muscle memory accrete.
3. **Immediate practice hooks.** Exercises and labs become rehearsal prompts: text/file clinics, cleaning drills, baseline-model experiments, and project-structure checklists feed directly into the coaching guides.
4. **Evidence-backed storytelling.** Each volume mingles data snapshots, notebook outputs, and small case studies so delegates can see why methods matter before they are asked to implement them in their own projects.

Deep Dive into The Data Scientist

The Data Scientist is a comprehensive, beginner-friendly learning path built to make delegates fluent in Python, data analysis, statistics, and project packaging—not a collection of disconnected notebooks. Learners move at their own pace through tightly edited PDFs, runnable labs, and weekly coaching guides [7, 8, 9, 10, 11] so progress stays structured even when calendars are busy. The overall 100-hour path [6] provides a paced benchmark for delegates who want a clear finish line from primer to portfolio.

- **Content, paced.** Weekly coaching guides [7, 8, 9, 10, 11] break the material into sprints that mix reading with labs on setup, text/files, cleaning, modeling, and projects; the 100-hour overview path summarises the whole journey.
- **Learning goals.** Build enduring instincts for writing and reading Python, cleaning and exploring data, evaluating models, and structuring projects so you can contribute confidently to junior data and analytics roles.
- **Portfolio outcome.** Grow a GitHub portfolio of small but coherent projects from your labs and capstones, with clear READMEs, notebooks, and dashboards that showcase your skills to hiring managers and admissions committees.

- **Who it's for.** Motivated beginners, analysts, and graduate or post-graduate students from any discipline who want a structured path into data science without needing prior formal training. Light familiarity with spreadsheets or basic coding helps, but the primer keeps newcomers on track.
- **Benefits to delegates.** Self-paced materials, reproducible labs, and 24/7 community support translate into faster onboarding, clearer project narratives, and a repeatable workflow for future domain-specific work or follow-on programs such as *The AI Engineer*, *The Crypto Engineer*, or the CPF finance flagship.

Outlook: From TDS to Your Next Basecamp

For many delegates, *The Data Scientist* is not the end of the journey but the safest way to reach a solid basecamp: one coherent path that turns "I am curious about data" into concrete skills, projects, and interview stories. From that base, three natural directions open up:

- **The AI Engineer (`https://theaiengineer.dev`).** Once you are comfortable with Python, data work, and evaluation basics, *The AI Engineer* builds on that foundation and moves into modern machine learning and AI engineering practices: model lifecycle, prompt and evaluation workflows, deployment patterns, and MLOps.
- **The Crypto Engineer (`https://thecryptoengineer.dev`).** If you are drawn to crypto infrastructure and security, TDS gives you the data-science and coding confidence to tackle *The Crypto Engineer*, where the focus shifts to applied cryptography, Bitcoin systems, markets, and live operations.
- **CPF Program (`https://python-for-finance.com`).** If you want to go deep into quantitative finance, TDS is a low-friction way to test the waters. After building interview-ready Python and data skills here, you can step into the CPF flagship training program with more confidence and a clearer sense of your interests.

Whichever route you choose, the habits you build in TDS—structured notebooks, small projects, clean data workflows, and honest evaluation—remain the basecamp you can return to whenever you push into new domains or responsibilities.

Bibliography

[1] Yves J. Hilpisch. *Python Primer for Data Science*. The Python Quants, 2026.

[2] Yves J. Hilpisch. *Python Programming Foundations for Data Science*. The Python Quants, 2026.

[3] Yves J. Hilpisch. *Python Data Analysis and Visualization*. The Python Quants, 2026.

[4] Yves J. Hilpisch. *Statistics and Machine Learning with Python*. The Python Quants, 2026.

[5] Yves J. Hilpisch. *Software Engineering Practices for Data Scientists*. The Python Quants, 2026.

[6] Yves J. Hilpisch. *The Data Scientist*. The Python Quants, 2026.

[7] The Python Quants. *TDS Coaching Guide 00: Ramp-Up Week*. Internal program note, 2026.

[8] The Python Quants. *TDS Coaching Guide 01: Module 1 Week*. Internal program note, 2026.

[9] The Python Quants. *TDS Coaching Guide 02: Module 2 Week*. Internal program note, 2026.

[10] The Python Quants. *TDS Coaching Guide 03: Module 3 Week*. Internal program note, 2026.

[11] The Python Quants. *TDS Coaching Guide 04: Module 4 Week*. Internal program note, 2026.

[12] The Python Quants. *Workspace and Colab Check Lab*. Internal program lab note, 2026.

[13] The Python Quants. *Text and File Clinic Lab*. Internal program lab note, 2026.

[14] The Python Quants. *Counting and Frequency Patterns Lab*. Internal program lab note, 2026.

[15] The Python Quants. *Messy CSV to Clean Table Lab*. Internal program lab note, 2026.

[16] The Python Quants. *Storytelling with Plots Lab*. Internal program lab note, 2026.

[17] The Python Quants. *Baseline Models and Evaluation Lab*. Internal program lab note, 2026.

[18] The Python Quants. *Overfitting and Leakage Clinic Lab*. Internal program lab note, 2026.

[19] The Python Quants. *From Notebook to Project Lab*. Internal program lab note, 2026.

[20] The Python Quants. *Dashboard Quickstart with Streamlit Lab*. Internal program lab note, 2026.

Contact

The Data Scientist Program

From First Notebook to Portfolio

Stay connected:

team@tpq.io

linktr.ee/dyjh

thedatascientist.dev