

Python Data Analysis, Visualization, and Storytelling

Module 2: Data Analysis, Visualization, and Storytelling Foundations

Dr. Yves J. Hilpisch¹

March 27, 2026



¹Get in touch: <https://linktr.ee/dyjh>. Web page: <https://thedata scientist.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 book is where Python begins to look unmistakably like data work. Module 1, *Python Programming Foundations for Data Science* [2], establishes fluency with values, loops, functions, files, and small artifacts. This book shifts the focus toward arrays, tables, plots, and short analytical stories that can actually support a claim.

That shift matters. In Module 1, the central question was often “Can I express this task clearly in Python?” In this volume, the more useful question becomes “Can I express this data problem clearly enough that its structure, patterns, and uncertainty are visible?”

Positioning

This book begins where basic Python fluency becomes analytical fluency. We are expected to be comfortable with simple functions, loops, files, and inspection. The new focus is arrays, tables, charts, and interpretation.

The book therefore has two parallel goals. The first is technical: learn how to work with NumPy arrays, pandas tables, and matplotlib figures. The second is judgmental: learn how to decide what a table means, what a chart actually says, and what kind of small story the evidence can support.

Structurally, the book is split into two parts. Part I, *Basic Data Work*, moves from Python objects into arrays, tables, cleaning workflows, and first plots. Part II, *Advanced Data Work*, adds a deeper pandas chapter, basic SQL awareness, and a small storytelling capstone that turns all of this into a portfolio-ready analytical artifact.

This volume, *Python Data Analysis, Visualization, and Storytelling* [3], is still a beginner book. It does not expect us to become full data analysts immediately. It does expect us to start reading data more calmly and producing more structured outputs than a plain script or notebook exercise from Module 1 [2].

By the end of the book, we should be able to:

- move from Python lists to NumPy arrays without confusion,
- perform a first layer of vectorized analytical work,
- read and shape small tabular datasets in pandas,
- generate and interpret basic matplotlib plots,
- and assemble the result into a small analytical artifact with a visible narrative.

How to Use This Book

Run the examples, inspect the outputs, and save the generated figures. In this module, the chart is not decoration. It is part of the argument.

The book also introduces a more deliberate visual workflow. Whenever matplotlib figures are generated for the teaching flow, they are saved in the local **figures/** folder as PDF. That

is not a technical flourish. It is part of the habit of treating visual output as a real artifact rather than a temporary screen event.

That habit matters because later books, especially *Statistics, Machine Learning, and Model Evaluation* [4] and *Software Engineering, Reproducibility, and Deployment Basics* [5], lean on analytical communication much more heavily. This book is the point where we stop only writing code and start producing evidence.

Contents

Preface	i
I Basic Data Work	1
1 From Python Objects to Analytical Structures	3
1.1 Why This Book Exists	4
1.2 From Objects to Analytical Structures	4
1.3 Why Arrays Matter	4
1.4 Why Tables Matter	5
1.5 Why Plots Matter	5
1.6 A Small Example of Why Shape Matters	5
1.7 From Numbers to Evidence	6
1.8 What the Delegate Should Learn to Notice	7
1.9 Where We Are Heading Next	7
2 NumPy Arrays and Vectorized Thinking	8
2.1 Why Arrays Feel Different	9
2.2 Creating a First Array	9
2.3 Inspect <code>.shape</code> and <code>.dtype</code> Early	9
2.4 Vectorized Arithmetic	10
2.5 Lists Versus Arrays	10
2.6 Summary Operations	10
2.7 Boolean Masks	11
2.8 A Distribution View	11
2.9 A Worked Example	11
2.10 Why This Matters for Data Science	12
2.11 Where We Are Heading Next	13
3 pandas DataFrames and Tabular Fluency	14
3.1 Why Tables Change the Work	15
3.2 From a Series to a DataFrame	15
3.3 Creating a First DataFrame	15
3.4 Inspect Before You Touch	16
3.5 Selecting One Column or Several	17
3.6 Filtering Rows with a Visible Rule	17
3.7 First Summaries	18
3.8 Sorting Makes Patterns Easier to See	18
3.9 A Small Worked Example	18
3.10 Why This Matters for the Capstone	19
3.11 Where We Are Heading Next	19

4	Cleaning, Grouping, and Summarizing Data	20
4.1	Running Example: Cleaning the Titanic Data	21
4.2	Why Cleaning Comes Before Clever Analysis	21
4.3	A Small Example Dataset	21
4.4	Handling Missing Data	22
4.5	Converting Types Calmly	23
4.6	Spotting Duplicates	23
4.7	Grouping and Summarizing	23
4.8	Multiple Aggregations with <code>.agg()</code>	24
4.9	Adding a Simple Feature	24
4.10	Tidy Data Habits	25
4.11	Where We Are Heading Next	25
5	Exploratory Data Analysis and Visualization	26
5.1	From Tables to First Questions	27
5.2	Summary Statistics as a First Lens	27
5.3	Line Charts for Trends Over an Ordered Axis	27
5.4	Bar Charts for Comparisons Between Categories	28
5.5	Histograms for Distributions	28
5.6	Titanic Survival Bar Charts	29
5.7	Scatter Plots for Relationships Between Two Variables	30
5.8	Choosing a Chart Type Deliberately	30
5.9	Describing Patterns Carefully	31
5.10	Where We Are Heading Next	32
II	Advanced Data Work	33
6	pandas Deep Dive	35
6.1	From Small Tables to Daily Workflows	36
6.2	Input and Output in Practice	36
6.3	Writing Out Clean Data	37
6.4	Joins and Relational Thinking	37
6.5	Grouping and Summarizing by Country	38
6.6	Reshaping with <code>pivot_table</code>	39
6.7	When <code>.apply()</code> Helps and When It Hurts	40
6.8	Method Chaining for Readability	40
6.9	A First Glance at Performance	41
6.10	Where We Are Heading Next	41
7	Basic SQL Awareness	42
7.1	Why SQL Matters Alongside pandas	43
7.2	A Small Example Schema	43
7.3	Setting Up SQLite and pandas Side by Side	43
7.4	Selecting Columns	44
7.5	Filtering Rows	44
7.6	Grouping and Aggregating	45
7.7	Simple Joins	46
7.8	When to Use SQL, When to Use pandas	47
7.9	Where We Are Heading Next	47
8	Analytical Storytelling with Small Data Projects	48

8.1	What We Are Building	49
8.2	Choosing a Dataset	49
8.3	Notebook and Repository Expectations	50
8.4	Shaping the Project	50
8.5	Questions, Tables, and Charts	51
8.6	Writing the Narrative	51
8.7	A Minimal Deliverables Checklist	51
8.8	Where We Are Heading Next	52
Glossary		54
Epilogue		57

List of Figures

1.1	Module 2 moves from Python objects to analytical structures, visual summaries, and small analytical stories.	4
1.2	A generated matplotlib line chart used in the opening analytical examples.	6
2.1	Vectorized thinking shifts attention from one element at a time to whole arrays of values.	9
5.1	A generated matplotlib bar chart showing total signups by channel for the small Module 2 example.	29
5.2	A generated matplotlib histogram that supports the first discussion of distributions.	30
5.3	A generated matplotlib scatter plot showing the relationship between site visits and signups.	31
6.1	A generated matplotlib bar chart showing total order amounts by country in the small orders example.	39

Contact

Python Data Analysis, Visualization, and Storytelling

Module 2: Data Analysis, Visualization, and Storytelling Foundations



Get in touch:

linktr.ee/dyjh

thedata scientist.dev