

Data Science

Cleaning, exploring and reasoning about data. You work in pandas and SQL, test claims by simulation instead of formulas, and present a result someone can act on.

Foundational

4 phases

22 sessions

33 contact hours

Syllabus

4 phases, 22 sessions. Each phase ends in something you have built.

1

The Notebook and the Table

Get Python running on your own machine and read a real data file without help.

6 sessions

\$72 alone

1.1 Setting Up Without Losing a Week

uv, a virtual environment, JupyterLab running locally, and reading the error when it breaks.

[FREE PREVIEW](#)

1.2 Python You Actually Need

Variables, lists, functions, and loops, taken only as far as data work requires.

1.3 Cells, Kernels, and Hidden State

Why out-of-order execution quietly ruins results, and restart-and-run-all as a reflex.

1.4 Tables as a Data Type

Rows, columns, dtypes, and loading a CSV that fights back on encoding and headers.

1.5 Your First Real Questions

Filtering, sorting, and summarizing until you can describe a dataset in five sentences.

1.6 Code You Did Not Write

Using an AI assistant for pandas, and the three checks that catch a column it invented.

By the end: **A notebook that runs clean from a fresh kernel and answers three questions about a dataset you chose.**

2

Getting Data into Shape

Turn a messy export into a dataset you would defend line by line.

6 sessions

\$72 alone

2.1 Cleaning Reality

Missing values, duplicates, wrong types, and other people's spreadsheets.

2.2 Assignment That Behaves

Selecting with `.loc`, copy-on-write in pandas 3.0, and why chained assignment no longer works.

2.3 Group, Pivot, Reshape

`groupby`, aggregation, long against wide, and split-apply-combine as a single idea.

2.4 Joins and the Rows You Silently Gained

Merge keys, join types, and checking row counts before you believe the output.

2.5 SQL Against Files with DuckDB

Querying CSV and Parquet in place, and when SQL or Polars beats a chain of pandas calls.

2.6 Charts That Argue

Encodings, honest axes, and one figure that makes its point before the caption does.

By the end: **A reproducible EDA notebook on a public dataset with every cleaning decision written down.**

3

Saying Something Is True

Make a claim about a population from a sample and defend it under questioning.

6 sessions

\$72 alone

3.1 Chance and Sampling

Random samples, sampling variability, and simulating ten thousand coin flips to see it.

3.2 A Null You Can Simulate

Stating a hypothesis as a process you can run, then choosing a test statistic for it.

3.3 Permutation and A/B Tests

Shuffling group labels to compare two groups without a formula or a lookup table.

3.4 The Bootstrap

Resampling one sample to get an interval, and what 95% emphatically does not mean.

3.5 Where the Normal Curve Comes From

The central limit theorem as something you observe in simulation, not a theorem you memorize.

3.6 Correlation, Confounders, Causation

Randomized against observational data, and the traps that end analyst careers.

By the end: **A designed and analyzed A/B test with a written recommendation you would sign.**

4

From Analysis to Decision

Turn an analysis into a decision somebody actually makes.

4 sessions

\$72 alone

4.1 Regression as Description

Fitting a line, reading residuals, and refusing to call the result a prediction.

4.2 Metrics That Survive Incentives

Defining a metric people cannot quietly game, and naming the ones already gamed.

4.3 The Analysis Report

Turning a notebook into a document with Quarto so a stranger can rerun it next year.

4.4 The Stakeholder Readout

Structuring the argument, presenting it in ten minutes, and surviving the questions.

By the end: **A stakeholder memo and a live readout that drives one concrete decision.**

Tools you will use

Python

JupyterLab

uv

pandas

NumPy

DuckDB

Polars

seaborn

What you will build

1 Messy Data Rescue

A public dataset cleaned into a defensible EDA notebook

2 The A/B Test Call

Permutation test, bootstrap interval, one recommendation

3 The Decision Memo

Question to query to chart to a call you defend live

Registration

Phase 1: The Notebook and the Table	\$72
-------------------------------------	------

Phase 2: Getting Data into Shape	\$72
----------------------------------	------

Phase 3: Saying Something Is True	\$72
-----------------------------------	------

Phase 4: From Analysis to Decision	\$72
------------------------------------	------

Whole track, all 4 phases

\$240

Buying every phase separately costs \$288, so the whole track saves \$48.

No payment is taken online. Registering creates an order and payment instructions are sent with it. Access opens once payment is confirmed.