

Data science foundations with Python

Course outline · issued 14 September 2026

Technology areas	Data Science, Data Analytics
Level	Intermediate
Short course	5 days
One-month track	4 weeks · 16 training days · 96 contact hours
Two-month track	8 weeks · 32 training days · 192 contact hours
Delivery	In person, Virtual, Hybrid
Class size	8 to 20 participants
Certification	Certificate of completion
Entry requirements	Strong spreadsheet skills. No programming experience assumed.

Who it is for

Analysts with spreadsheet experience moving into statistical analysis.

REVENUE USE CASE

A simple compliance risk score built from filing history features.

By the end, participants can

- Explore and clean datasets with pandas
- Apply descriptive statistics and simple predictive models
- Communicate findings with reproducible notebooks

Short course modules (5 days)

1. Python and notebooks
2. Data wrangling with pandas
3. Statistics for analysts
4. Regression and classification basics
5. Risk-scoring mini project

One-month track — weeks 1 to 4

4 weeks · 16 training days · 96 contact hours

Week 1 — Python and notebooks

- Python syntax for people who know spreadsheets
- Notebooks, environments and reproducibility
- Reading files and handling encodings
- Debugging without panic

Practical: Load a revenue dataset and produce first summary statistics.

Week 2 — Data wrangling with pandas

- Series, DataFrames and indexing
- Missing data, duplicates and type coercion
- Merging, grouping and reshaping
- Dates, periods and fiscal calendars

Practical: Clean and join three years of filing records.

Week 3 — Statistics for analysts

- Distributions, spread and what an average hides
- Correlation, confounding and causation
- Hypothesis tests and confidence intervals
- Visualising distributions with matplotlib

Practical: Test whether a compliance intervention changed filing behaviour.

Week 4 — Regression and classification basics

- Linear regression and interpreting coefficients
- Logistic regression for a yes/no outcome
- Train/test splits and overfitting
- Accuracy, precision, recall and why they differ

Practical: Build a first compliance risk classifier and evaluate it honestly.

Checkpoint: the one-month track is assessed [here](#).

Two-month track — continues through weeks 5 to 8

8 weeks · 32 training days · 192 contact hours

Week 5 — Feature engineering

- Turning filing history into model features
- Categorical encoding and scaling
- Leakage: the mistake that flatters every model
- Feature selection and parsimony

Practical: Engineer a feature set from raw assessment and payment history.

Week 6 — Model improvement and validation

- Tree-based models and ensembles
- Cross-validation and hyperparameter tuning
- Class imbalance in compliance data
- Calibration and threshold selection

Practical: Improve and properly validate the risk model.

Week 7 — Interpretability and ethics

- Explaining a model to a non-technical decision maker
- Feature importance and its pitfalls
- Fairness across taxpayer segments
- Documenting limitations and appropriate use

Practical: Write the model card, including what the model must not be used for.

Week 8 — Capstone and deployment thinking

- Packaging a reproducible notebook
- From notebook to a scheduled scoring job
- Monitoring drift after deployment
- Presenting findings to management

Practical: Present the risk-scoring model and its documentation to a panel.

Capstone assessment and post-course evaluation.

Capstone

A compliance risk-scoring model built from filing history, documented in a reproducible notebook with an honest account of its limitations.

About this outline

Indicative extended tracks. The one-month track runs weeks 1 to 4; the two-month track continues through week 8 with deeper labs, certification preparation and a capstone. Weeks, contact hours and cohort sizes are finalised with each client.

Prepared by	Dividend Consulting Nigeria Limited
RC number	463831
Contact	0806 022 0612 · imudiagra@gmail.com
