



Data Science Curriculum

Learn Python, SQL, statistics and machine learning in a 24-week online data science bootcamp with guided projects and mentorship.

Duration: 24 weeks

Format: Online classes with mentorship; confirm cohort times on WhatsApp

Level: Beginner Friendly

Program Overview

A beginner-focused online data science bootcamp connecting Python programming, SQL, statistics and machine learning through guided practice and portfolio projects.

What Students Will Achieve

- Write and explain Python code for preparing and analysing data
- Use statistics to interpret evidence and uncertainty
- Build regression, classification and clustering workflows
- Evaluate models and recognise leakage, overfitting and limitations
- Present reproducible portfolio work with clear business context

Tools and Platforms

Python | Jupyter / Google Colab | pandas and NumPy | SQL | Matplotlib and Seaborn | scikit-learn | Git and GitHub

Curriculum Modules

Weeks 1–4: Python programming foundations

Start with variables, data types, conditions, loops and functions. Learn to read errors, work with files and organise a notebook so someone else can follow your code.

Practical Output:

A Python notebook that loads a practice dataset and answers simple business questions.



Weeks 5–7: Data preparation with pandas, NumPy and SQL

Select and transform columns, handle missing values and duplicates, join tables and check data types. Use SQL filtering, joins and aggregation to prepare structured data for analysis.

Practical Output:

A cleaned dataset, reusable preparation code and documented quality checks.

Weeks 8–10: Statistics and probability for analysis

Explore distributions, averages, variation, probability and sampling. Learn what confidence intervals and hypothesis tests can tell you, how assumptions affect conclusions, and why correlation does not establish causation.

Practical Output:

A statistical analysis notebook that explains uncertainty and the limits of its conclusions.

Weeks 11–12: Exploratory analysis and visualisation

Investigate patterns, outliers and relationships using pandas, Matplotlib and Seaborn. Choose charts for the question being asked and communicate findings without overstating the evidence.

Practical Output:

An exploratory report with annotated charts, findings and questions for further investigation.

Weeks 13–16: Supervised machine learning

Understand features, targets and baseline models. Build regression models for numerical predictions and classification models for categories using scikit-learn, including linear models and decision trees.

Practical Output:

A baseline and a trained model, with a clear explanation of the prediction task.

Weeks 17–19: Model evaluation and responsible use

Separate training, validation and test data; fit preprocessing on training data; and recognise data leakage and overfitting. Compare regression errors and classification precision, recall and F1. Discuss privacy, bias and when a model should not be used.

Practical Output:

A reproducible evaluation report that compares models and documents their limitations.



Weeks 20–21: Unsupervised learning and segmentation

Explore clustering, feature scaling and customer segmentation. Interpret groups carefully, check whether they are useful for the business question and distinguish clusters from proven categories.

Practical Output:

A segmentation notebook with visual profiles and a discussion of limitations.

Weeks 22–24: Capstone and portfolio communication

Bring the workflow together: define a question, prepare data, explore patterns, build and evaluate a model, then explain the result. Use Git and GitHub to document your work and make the steps reproducible.

Practical Output:

A portfolio repository with code, a README, evaluation results and a concise presentation.

Capstone Project

Complete a documented data science case study: define a question, prepare a practice dataset, explore patterns, compare a baseline with a trained model and evaluate the result on held-out data. Present the code, findings, limitations and recommendations in a GitHub repository and short presentation.

Student Support

- Mentorship and guided support throughout the learning journey
- Certificate of completion after successful participation
- Career guidance to help position students for opportunities