SODBRIDGE ACADEMY
ADVANCED PROFESSIONAL PROGRAMME

TRACK: DATA SCIENCE

Professional Training Curriculum

Programme Detail	Information
Duration	12 Weeks
Live Instruction	6 Hours Per Week
Total Live Instruction	72 Hours
Weekly Schedule	3 Sessions × 2 Hours
Programme Fee	₦200,000

1. Programme Overview

The Data Science track develops the statistical reasoning, programming and machine learning skills required to investigate data and build predictive models.
Learners strengthen their Python data analysis skills, prepare datasets, engineer features, establish baselines, train and compare machine learning models, validate performance, interpret results and communicate limitations.
The programme introduces supervised and unsupervised learning using practical datasets and scikit-learn. It also introduces Azure Machine Learning as a managed environment for organising experiments and model workflows.
The emphasis is on developing a sound and reproducible machine learning workflow rather than simply training models.

2. Learning Outcomes

By the end of the programme, learners should be able to:

- Apply statistical reasoning to understand datasets and modelling problems.
- Prepare datasets for machine learning.
- Perform feature engineering and preprocessing.
- Develop regression and classification models using scikit-learn.
- Apply appropriate train/test splits and cross-validation techniques.
- Evaluate models using suitable performance metrics.
- Recognise data leakage, overfitting, class imbalance and common validation pitfalls.
- Interpret model behaviour and communicate limitations.
- Compare models and justify model selection.
- Understand experiment and model workflow management using Azure Machine Learning.

3. Entry Requirements

Applicants should have:

- Working knowledge of Python and pandas.
- Familiarity with loading, filtering and transforming datasets.
- Basic algebra.
- Familiarity with descriptive statistics and probability.
- Comfort using Jupyter notebooks.
- Willingness to practise independently between live sessions.

Applicants who do not meet these requirements may be advised to complete the Foundations programme or an appropriate readiness assessment.

4. Tools and Technologies

The programme will primarily use:

- Python
- NumPy
- pandas
- Matplotlib
- Seaborn
- scikit-learn
- Jupyter Notebook
- Git and GitHub
- Microsoft Azure Machine Learning

5. Programme Structure and Weekly Curriculum

The programme begins with one shared professional core week, followed by 11 weeks of Data Science specialisation, applied practice and capstone development.

WEEK 1 — SHARED PROFESSIONAL CORE

Focus: Professional problem-solving and project practices.

Topics:

- Understanding professional data and machine learning projects.
- Translating business problems into measurable modelling questions.
- Identifying and assessing data sources.
- Introduction to data quality and reliability.
- Git and GitHub fundamentals.
- Project organisation and documentation.
- Reproducible analytical workflows.
- Responsible use of data.

WEEK 2 — STATISTICS AND PROBABILITY FOR MODELLING

Focus: Developing the statistical foundations required for machine learning.

Topics:

- Descriptive statistics.
- Measures of central tendency and spread.
- Understanding distributions.
- Sampling concepts.
- Introduction to probability.
- Correlation and relationships between variables.
- Statistical thinking and interpretation.
- Association versus causation.

WEEK 3 — PYTHON FOR DATA SCIENCE

Focus: Strengthening Python skills for analytical and modelling workflows.

Topics:

- NumPy fundamentals and array operations.
- Advanced pandas practice.
- Data manipulation and transformation.
- Exploratory visualisation.

- Reusable functions.
- Working with Jupyter notebooks.
- Organising experiments.
- Recording analytical decisions.

WEEK 4 — DATA PREPARATION AND FEATURE ENGINEERING

Focus: Preparing datasets for machine learning.

Topics:

- Handling missing values.
- Identifying and treating outliers.
- Categorical variable encoding.
- Feature scaling.
- Feature selection concepts.
- Train/test separation.
- Preprocessing pipelines.
- Preventing data leakage.

WEEK 5 — MACHINE LEARNING FOUNDATIONS

Focus: Understanding the machine learning workflow.

Topics:

- Introduction to machine learning.
- Supervised versus unsupervised learning.
- Understanding features and target variables.
- Establishing baseline models.
- Model assumptions.
- Selecting suitable algorithms.
- Understanding the machine learning lifecycle.
- Evaluating modelling objectives against business problems.

WEEK 6 — REGRESSION

Focus: Predicting continuous numerical outcomes.

Topics:

- Introduction to regression.
- Linear regression.
- Regularisation concepts.
- Model fitting.
- Residual analysis.
- Mean Absolute Error (MAE).
- Mean Squared Error (MSE) and Root Mean Squared Error (RMSE).
- R-squared.
- Interpreting model coefficients and limitations.

WEEK 7 — CLASSIFICATION

Focus: Predicting categories and outcomes.

Topics:

- Introduction to classification.
- Logistic regression.
- Decision tree concepts.
- Classification workflows.
- Confusion matrix.
- Accuracy, precision and recall.

- F1-score.
- ROC-AUC concepts.
- Classification thresholds.
- Class imbalance.

WEEK 8 — MODEL VALIDATION AND IMPROVEMENT

Focus: Building reliable models and evaluating their performance.

Topics:

- Train, validation and test datasets.
- Cross-validation.
- Overfitting and underfitting.
- Hyperparameter tuning.
- Comparing models fairly.
- Model evaluation strategies.
- Reproducible experiments.
- Improving model performance responsibly.

WEEK 9 — UNSUPERVISED LEARNING

Focus: Discovering patterns and groupings in data.

Topics:

- Introduction to unsupervised learning.
- Clustering concepts.
- K-means clustering.
- Feature scaling and distance.
- Selecting and interpreting clusters.
- Evaluating clustering results.
- Introduction to dimensionality reduction.
- Practical applications and limitations.

WEEK 10 — MODEL INTERPRETATION AND RESPONSIBLE USE

Focus: Understanding and communicating model behaviour.

Topics:

- Feature importance.
- Permutation importance concepts.
- Model interpretation techniques.
- Error analysis.
- Subgroup performance.
- Explainability limitations.
- Bias and responsible model use.
- Privacy considerations.
- Communicating uncertainty and limitations.

WEEK 11 — AZURE MACHINE LEARNING WORKFLOW

Focus: Introduction to managed machine learning environments.

Topics:

- Introduction to Azure Machine Learning.
- Workspace and compute concepts.
- Organising machine learning experiments.
- Tracking experiment runs and metrics.
- Managing model artefacts.
- Model registration.

- Introduction to batch inference and endpoint deployment options.
- Understanding resource requirements and cost considerations.

WEEK 12 — CAPSTONE PRESENTATION AND DEFENCE

Focus: Demonstrating practical machine learning competence.

Activities:

- Presenting the modelling problem.
- Explaining data preparation and feature engineering.
- Demonstrating the baseline model.
- Comparing trained models.
- Explaining evaluation choices.
- Interpreting model errors and limitations.
- Demonstrating a reproducible workflow.
- Discussing possible improvements and next steps.

6. Capstone Project

Learners will complete a practical machine learning project demonstrating their ability to frame a problem, prepare data, train models, evaluate performance and communicate findings.

The expected deliverables include:

1. A clearly defined real-world prediction or grouping problem.
2. A documented dataset and data preparation process.
3. Exploratory data analysis.
4. A baseline model.
5. One or more trained machine learning models.
6. Model evaluation and comparison.
7. Interpretation of results and errors.
8. Discussion of model limitations.
9. A reproducible notebook or project structure.
10. A final presentation of findings and recommendations.

Azure Machine Learning may be used to organise experiments and register the selected model.

7. Assessment Approach

Learners will be assessed through:

- Statistical reasoning and data preparation exercises.
- Python and pandas implementation tasks.
- Regression and classification exercises.
- Model validation and comparison checkpoints.
- Model interpretation exercises.
- Capstone project assessment.

The capstone assessment will consider problem framing, reproducibility, evaluation quality, interpretation, responsible use and communication.

8. Learning Model

The programme follows three complementary stages:

LEARN

Instructor-led explanations, demonstrations and guided machine learning practice.

BUILD
Practical weekly exercises that progressively develop statistical, programming and modelling skills.
APPLY
An end-to-end machine learning project, review and presentation demonstrating the learner's ability to explain and defend their modelling decisions.

9. Cloud Technology and Cost Considerations
The programme introduces Microsoft Azure Machine Learning as a managed environment for organising experiments and model workflows.
Practical sessions may cover experiment tracking, model artefact management and model registration.
Azure compute, storage and related resources may incur charges depending on usage, region and service configuration.
Where available, learners may be guided to use eligible trial credits or learning allowances.
Guidance will be provided on cost controls and resource cleanup.
Deployment is introduced as a practical concept; the programme does not promise production-scale model hosting or unlimited cloud resources.

10. Additional Information
- Live instruction is delivered over three 2-hour sessions each week.
- Learners should plan for approximately 4–6 additional hours of independent practice weekly.
- Learners should have access to a suitable computer and Python development environment.
- Public, synthetic or appropriately authorised datasets should be used for practical projects.
- Programme tools and cloud service availability may change; equivalent learning environments may be used where necessary.
- Programme details are subject to the final cohort calendar and published enrolment terms.

SODBRIDGE ACADEMY
Real-world learning through practical skills, projects and professional development.