

Detailed Syllabus

Session-by-session topic breakdown for the Data Science & AI Programme – every day of all nineteen weeks, mapped to what you will actually learn.

95

sessions total

19

weeks

9

phases

1.5 hrs

per session

HOW TO READ THIS DOCUMENT

Day by day

Each session below is numbered D01 through D95 and grouped by week within its phase, matching the roadmap in the programme brochure.

PHASES COVERED

Python & Statistics, SQL, Machine Learning, Feature Engineering, Deep Learning, NLP & Generative AI, MLOps & Deployment, AI-Assisted Communication, and the Capstone Project.

PHASE 01 Python & Statistics for Data Science

WEEKS 1-3

WEEK 1

D01 Programme orientation & the data scientist's workflow —How raw data becomes a deployed model; the toolchain this programme covers

D02 Python refresher for data science —Variables, control flow, functions, working in notebooks and scripts

D03 NumPy essentials —Arrays, vectorised operations, broadcasting

D04 pandas essentials I —Series, DataFrames, reading and inspecting data

D05 pandas essentials II —Indexing, filtering, boolean masks

WEEK 2

D06 Data cleaning in pandas —Missing values, duplicates, data type conversion

D07 Descriptive statistics —Mean, median, mode, variance, standard deviation

D08 Data distributions —Normal distribution, skewness, identifying outliers

D09 Probability basics —Core rules, conditional probability, an intro to Bayes' theorem

D10 Sampling & estimation —Central Limit Theorem, confidence intervals

WEEK 3

D11 Hypothesis testing fundamentals —Null vs alternative hypotheses, p-values, significance

D12 Comparing groups —t-tests for two-sample comparisons

D13 Categorical data testing —Chi-square tests

D14 Linear algebra essentials —Vectors, matrices and dot products, the math behind ML

D15 Milestone project: Statistical Analysis & Hypothesis Testing —Apply hypothesis testing to a real dataset

PHASE 02 SQL for Data Science

WEEK 4

WEEK 4

D16 **Relational databases & SELECT basics** —SELECT, WHERE, ORDER BY, filtering fundamentals

D17 **Aggregation** —GROUP BY, HAVING, aggregate functions

D18 **Joins** —INNER, LEFT and multi-table joins

D19 **Subqueries & CTEs** —Nested queries, readable multi-step queries

D20 **SQL practice** —Sourcing and joining features for a model from a live database

PHASE 03 Machine Learning Essentials

WEEKS 5-7

WEEK 5

D21 **The machine learning workflow** —Train/test split, overfitting vs underfitting, the bias-variance tradeoff

D22 **Linear regression** —Simple and multiple regression, interpreting coefficients

D23 **Logistic regression** —Binary classification, decision boundaries

D24 **Evaluation metrics for regression** —MAE, MSE, RMSE, R-squared

D25 **Evaluation metrics for classification** —Accuracy, precision, recall, F1, confusion matrix

WEEK 6

D26 **Decision trees** —How trees split, overfitting in trees

D27 **Random forests** —Bagging and ensemble intuition

D28 **Gradient boosting** —Introduction to XGBoost

D29 **Cross-validation** —k-fold validation, avoiding data leakage

D30 **Model comparison practice** —Comparing multiple algorithms on the same problem

PHASE 03 Machine Learning Essentials

WEEKS 5-7

WEEK 7

- D31 **Handling categorical variables** —Encoding strategies for ML
- D32 **Regularisation** —Ridge, Lasso, avoiding overfitting
- D33 **Multi-class classification** —One-vs-rest and multinomial approaches
- D34 **Model selection practice** —Choosing the right algorithm for a given problem
- D35 **Milestone project: End-to-End Machine Learning Model** —Build, evaluate and compare ML models on a real problem

PHASE 04 Feature Engineering & Optimization

WEEK 8

WEEK 8

- D36 **Feature engineering fundamentals** —Creating meaningful features from raw data
- D37 **Handling imbalanced data** —Oversampling, undersampling, class weights
- D38 **Hyperparameter tuning** —Grid search and random search
- D39 **Model explainability with SHAP** —Feature importance and SHAP values
- D40 **Milestone project: Model Explainability** —Use SHAP to explain what drives a model's predictions

PHASE 05 Deep Learning

WEEKS 9-11

WEEK 9

- D41 **Neural networks from first principles** —Perceptrons and activation functions
- D42 **Building a neural network** —Forward pass, loss functions, backpropagation intuition
- D43 **Training neural networks** —Gradient descent, learning rate, epochs
- D44 **TensorFlow & Keras basics** —Building and training a simple network
- D45 **PyTorch basics** —Tensors and building a simple network

PHASE 05 Deep Learning

WEEKS 9-11

WEEK 10

D46 Convolutional Neural Networks (CNNs) —Convolutions, pooling, image classification

D47 Building a CNN —Training an image classifier

D48 Recurrent Neural Networks (RNNs) —Sequence data and memory in networks

D49 LSTMs & sequence modelling —Handling longer sequences

D50 Regularisation in deep learning —Dropout and batch normalisation

WEEK 11

D51 Transfer learning —Using pretrained models

D52 Introduction to Transformer architecture —The attention mechanism, intuitively

D53 Model evaluation for deep learning —Validation curves, avoiding overfitting

D54 Deep learning practice —Solving an image or sequence problem end to end

D55 Milestone project: Deep Learning Classifier —Train a CNN or RNN on a real dataset

PHASE 06 NLP & Generative AI

WEEKS 12-14

WEEK 12

D56 NLP fundamentals —Tokenisation, stopwords, stemming and lemmatisation

D57 Text representation —Bag of words and TF-IDF

D58 Word embeddings —Word2Vec and GloVe, intuitively

D59 Introduction to Transformers for NLP —BERT and the GPT family at a glance

D60 Using pretrained language models —Hugging Face Transformers basics

PHASE 06 NLP & Generative AI

WEEKS 12-14

WEEK 13

D61 **Prompt engineering fundamentals** —Writing effective prompts, few-shot examples

D62 **Embeddings for retrieval** —Semantic search and embedding models

D63 **Vector databases** —Storing and querying embeddings

D64 **Introduction to LangChain** —Chains, agents and tools

D65 **Building a simple retrieval pipeline** —Connecting a query to relevant documents

WEEK 14

D66 **Retrieval-Augmented Generation (RAG)** —Architecture and design patterns

D67 **Building a RAG application** —Connecting retrieval to generation

D68 **Fine-tuning basics** —When and how to fine-tune a model

D69 **Evaluating LLM applications** —Testing for accuracy, hallucination and relevance

D70 **Milestone project: RAG-Based Q&A Application** —A working RAG app using LangChain and a vector database

PHASE 07 MLOps & Deployment

WEEKS 15-16

WEEK 15

D71 **The MLOps lifecycle** —From notebook to production

D72 **Packaging a model** —Serialisation with pickle and joblib

D73 **Building an API with FastAPI** —Serving predictions over HTTP

D74 **API testing & validation** —Request and response design, input validation

D75 **Introduction to Docker** —Containers vs virtual machines

PHASE 07 MLOps & Deployment

WEEKS 15-16

WEEK 16

- D76 Containerising a model API** —Writing a Dockerfile, building an image
- D77 Experiment tracking with MLflow** —Logging parameters, metrics and models
- D78 Model versioning & registries** —Managing multiple model versions
- D79 Monitoring models in production** —Drift detection and logging predictions
- D80 Milestone project: Deployed Model API** —Package, containerise and deploy a live, monitored API

PHASE 08 AI-Assisted Data Science & Communication

WEEK 17

WEEK 17

- D81 AI copilots for EDA & modelling** —Accelerating exploratory work with AI tools
- D82 AI copilots for debugging & code review** —Using AI as a working part of the workflow
- D83 Building a Power BI dashboard from model outputs** —Presenting results to a business audience
- D84 Data storytelling** —Presenting technical results to non-technical stakeholders
- D85 Communication practice** —Turning a model into a business recommendation

PHASE 09 Capstone Project

WEEKS 18-19

WEEK 18

- D86 Capstone kickoff** —Choosing a track and planning the approach
- D87 Work session: data cleaning & EDA** —Preparing the chosen dataset
- D88 Work session: feature engineering & model building** —Building the core model
- D89 Work session: model evaluation & selection** —Choosing the best-performing approach
- D90 Work session: preparing for deployment** —Getting the model ready to ship

WEEK 19

- D91 Work session: building the deployment API** —Serving the model over HTTP
- D92 Work session: containerising & deploying** —Docker and a live deployment
- D93 Work session: monitoring setup & final testing** —Confirming the deployed system works end to end
- D94 Work session: refining the presentation** —Polishing the dashboard, docs and narrative
- D95 Final capstone presentations** —Presenting the completed capstone; programme wrap-up