



Data Scientist Launchpad

Zero to Hired

6-Month Data Science Programme • Syllabus & Programme Guide

Course Name	Data Scientist Launchpad — Zero to Hired
Duration	6 Months Total Programme
Technical Training	3 Months (Months 1-3) • 3 Sessions Per Week • 36 Sessions
Session Length	90 Minutes Each
Level	Complete Beginner → Job Ready
Internship	Starts Month 4 • With Mentor Support • 6-Month Experience Certificate
Placement Support	Starts Month 6
Alumni Access	Lifetime — after programme completion
Website	www.cadetx.co.uk

Month	Focus	Schedule
Month 1	Python Foundations	3 sessions/week • Sessions 1-12
Month 2	Data Handling, GitHub & Statistical Thinking	3 sessions/week • Sessions 13-24
Month 3	Statistics, SQL, ML & Deployment	3 sessions/week • Sessions 25-36
Month 4 onwards	Internship — Real Projects with Mentor Support	Weekly mentor guidance + project work
Month 6	Placement Support Activated	Applications, referrals & interviews

FOUNDATIONS • MONTHS 1-2 • Sessions 1-12 • 18 Hours

Python Foundations (Sessions 1-8)

SESSION 1 Python Part 1 — Introduction, Colab, Syntax & Variables

- Data Science fundamentals — 30-minute orientation
- What is Python and where it is used
- Google Colab and notebook cells
- Python syntax, indentation and comments
- Variables, assignment, input and output

SESSION 2 Python Part 2 — Data Types, Operators & Conditions

- Numbers, strings, Boolean values and None
- Type checking and type conversion
- Arithmetic, comparison and logical operators
- if, elif, else and nested conditions
- Formatted strings

SESSION 3 Python Part 3 — Loops

- for and while loops
- range, counters and accumulators
- Looping through strings and collections
- Nested loops
- break and continue

SESSION 4 Python Part 4 — Functions

- Defining and calling functions
- Parameters, arguments and return values
- Default and keyword arguments
- Local and global scope
- Breaking problems into reusable functions
- Brief introduction to lambda functions

SESSION 5 Python Part 5 — Lists, Tuples & Sets

- Lists — indexing, slicing and methods
- Nested lists and list comprehensions

- Tuples, packing and unpacking
- Sets, unique values and set operations
- Mutable vs immutable objects

SESSION 6 Python Part 6 — Dictionaries & String Manipulation

- Dictionaries, keys, values and methods
- Nested dictionaries
- String indexing and slicing
- String methods
- Splitting, joining, replacing and standardising text

SESSION 7 Python Part 7 — OOP Fundamentals

- Classes and objects
- Attributes and methods
- `__init__` and `self`
- Instance and class attributes
- Basic encapsulation and simple inheritance
- Polymorphism intuition

SESSION 8 Python Part 8 — File, Exception Handling & Debugging

- Relative and absolute paths
- Reading and writing simple text files
- `with open()`
- `try`, `except`, `else` and `finally`
- Reading tracebacks
- Input validation and debugging

NumPy & Pandas (Sessions 9–11)

SESSION 9 NumPy — Arrays and Vectorised Computing

- NumPy arrays and `ndarray`
- Shapes, dimensions and axes
- Array creation, indexing and slicing
- Boolean filtering
- Vectorisation and broadcasting
- Aggregation by axis and random numbers

SESSION 10 Pandas Part 1 — DataFrames and Selection

- Series and DataFrame
- Loading CSV and Excel files
- `head`, `info`, `describe`, `shape` and `dtypes`
- Selecting with `loc` and `iloc`
- Filtering, sorting and creating columns

SESSION 11 Pandas Part 2 — Manipulation and Data Cleaning

- Missing values and duplicates
- Data-type, text, category and date cleaning
- Invalid values and outlier checks
- `groupby` and aggregation
- Merging, concatenation and pivot tables

Git & GitHub (Session 12)

SESSION 12 Git & GitHub — Version Control

- Version control and Git vs GitHub
 - Repository, staging area, commit and history
 - `init`, `status`, `add`, `commit`, `log`, `push` and `pull`
 - `.gitignore`
 - Portfolio repository structure and README files
-

Statistics & Probability (Sessions 13-16)

SESSION 13 Descriptive Statistics

- Population, sample and variable types
- Mean, median and mode
- Variance and standard deviation
- Range, IQR and percentiles
- Distribution shape, outliers and grouped summaries

SESSION 14 Probability & Distributions

- Probability rules
- Independent, dependent and conditional events
- Bayes theorem intuition
- Random variables
- Binomial and normal distributions
- Z-scores

SESSION 15 Sampling, Standard Error & Confidence Intervals

- Random and biased samples
- Sampling variability and sampling distributions
- Central Limit Theorem
- Standard deviation vs standard error
- Point estimates and confidence intervals
- Bootstrap intuition and sample-size effect

SESSION 16 Correlation, Hypothesis Testing & A/B Testing

- Covariance; Pearson and Spearman correlation
- Correlation vs causation
- Null and alternative hypotheses
- p-value, significance level and Type I/II errors
- Effect size
- t-tests, chi-square test and A/B testing fundamentals

SQL for Data Science (Sessions 17-21)

SESSION 17 SQL Part 1 — Relational Databases & SELECT

- Tables, rows, columns and data types
- Primary and foreign keys
- Understanding a schema
- SELECT, aliases, DISTINCT and LIMIT
- Readable SQL

SESSION 18 SQL Part 2 — Filtering, Sorting & Functions

- WHERE and comparison operators
- AND, OR, NOT, IN and BETWEEN
- LIKE and NULL handling
- ORDER BY
- Essential string, numeric and date functions

SESSION 19 SQL Part 3 — Aggregation & Conditional Analysis

- COUNT, SUM, AVG, MIN and MAX
- GROUP BY, HAVING and nested aggregation
- CASE WHEN for conditional columns
- Rounding, null handling and subqueries

SESSION 20 SQL Part 4 — Joins & Subqueries

- INNER, LEFT, RIGHT and FULL OUTER joins
- Joining on multiple conditions
- Self-joins and aliasing
- Subqueries and CTEs (WITH clauses)
- Choosing between joins and subqueries

SESSION 21 SQL Part 5 — Window Functions & Advanced Queries

- ROW_NUMBER, RANK and DENSE_RANK
- PARTITION BY and ORDER BY in window context
- LAG, LEAD and running totals
- NTILE and percentile analysis
- Combining window functions with aggregation

EDA, VISUALISATION & FEATURE ENGINEERING • Sessions 22–27

SESSION 22 Exploratory Data Analysis Part 1 — Understanding Data

- EDA as a mindset, not a checklist
- Data types, shapes and missing-value patterns
- Univariate analysis — distributions, skew and outliers
- Bivariate analysis — relationships and conditional distributions
- EDA questions vs confirmatory analysis

SESSION 23 Exploratory Data Analysis Part 2 — Visualisation

- Matplotlib figure, axes and plot anatomy
- Line, bar, scatter, histogram and boxplot
- Seaborn statistical plots
- Pair plots and heatmaps
- Storytelling with visualisation

SESSION 24 Feature Engineering Part 1 — Encoding & Scaling

- Why feature engineering matters
- One-hot and label encoding for categorical variables
- Ordinal encoding
- Min-max scaling and standardisation
- When and why to scale

SESSION 25 Feature Engineering Part 2 — Feature Creation

- Interaction features and polynomial features
- Date and time feature extraction
- Binning and discretisation
- Domain-driven feature ideas
- Feature importance from tree models

SESSION 26 Pipelines — sklearn Pipeline & ColumnTransformer

- Why pipelines prevent data leakage
- sklearn Pipeline construction
- ColumnTransformer for mixed feature types
- Fitting and transforming train/test sets consistently
- Saving and reloading a pipeline

SESSION 27 Capstone EDA Project

- End-to-end EDA on a real dataset
- Business problem framing
- Full visualisation report
- Feature engineering decisions documented
- Presentation and peer review

MACHINE LEARNING • MONTH 3 • Sessions 28–34

Supervised Learning (Sessions 28–31)

SESSION 28 Supervised Learning Foundations & Regression

- Supervised vs unsupervised learning
- Train, validation and test sets
- Linear regression — OLS intuition and sklearn
- Polynomial regression and regularisation (Ridge, Lasso)
- MAE, RMSE, R^2 and residual analysis

- Overfitting vs underfitting and the bias-variance trade-off

SESSION 29 Classification — Logistic Regression & Tree Models

- Binary and multi-class classification
- Logistic regression — sigmoid, log-odds and probability threshold
- Decision trees — splitting criteria and depth
- Random Forest — ensemble intuition, feature importance
- Accuracy, precision, recall, F1 and ROC AUC
- Confusion matrix and classification report

SESSION 30 Cross-Validation, Hyperparameter Tuning & Model Selection

- k-fold and stratified k-fold cross-validation
- GridSearchCV and RandomizedSearchCV
- Scoring metrics in cross-validation
- Gradient boosting — XGBoost or LightGBM intuition
- Choosing a model for a given problem
- Final model evaluation on hold-out test set

SESSION 31 Unsupervised Learning — Clustering & PCA

- K-Means and appropriate clustering questions
- Scaling and selecting cluster count
- Elbow method and silhouette score
- Cluster profiling
- PCA and explained variance
- PCA visualisation and interpretation limits

Advanced Topics (Sessions 32–34)

SESSION 32 Time-Series Foundations

- Time-indexed data, trend and seasonality
- Temporal leakage
- Chronological train-test splitting
- Lag and rolling features
- Naive baselines
- MAE, RMSE and walk-forward validation intuition

SESSION 33 Deep Learning and ANN Foundations

- Machine learning vs deep learning
- Perceptron; input, hidden and output layers
- Weights, biases and activation functions
- Forward propagation, loss and gradient descent
- Backpropagation intuition, epochs and batches
- Validation, overfitting and early stopping
- Brief CNN, RNN and transformer comparison

SESSION 34 NLP and Text Classification

- Text normalisation and tokenisation
- Stop words, stemming and lemmatisation
- Bag of Words, n-grams and TF-IDF
- Sparse feature matrices
- Naive Bayes or logistic regression for text
- Text-classification metrics
- Embeddings, attention and transformers conceptually

DEPLOYMENT & CAPSTONE • Sessions 35–36

SESSION 35 Building a Streamlit Data Science Application

- Training vs inference
- Refactoring notebook code
- Saving and loading the complete pipeline
- Prediction function
- Streamlit inputs and layout

- Input validation, predictions, probabilities and limitations
- Local testing

SESSION 36 Deploying and Maintaining the Application

- Application file structure
- venv, pip and requirements.txt
- Environment variables and secret-management principles
- GitHub-based deployment to one selected platform
- Deployment logs and dependency debugging
- Basic logging, model versions and drift intuition

INTERNSHIP & PLACEMENT PROGRAMME

Internship Programme (Months 4–5 • with Mentor Support)

Starting Month 4: Students begin working on real company-sourced projects with dedicated mentor support alongside ongoing learning.

Starts Month 4	Real company-sourced data science projects begin. Weekly mentor guidance, collaboration with global peers via the CadetX portal, task-based learning, structured problem-solving, and project reports.
Mentor Support	Each student is paired with a mentor throughout the internship. Regular 1:1 check-ins, code reviews, and career advice to bridge theory and practice.
Certificate — 6-Month Experience Letter	Students who complete the internship receive a 6-Month Experience Certificate documenting their role, projects, skills, and tools — ready to add to a CV and LinkedIn profile.

Placement Support (Month 6 onwards)

From Month 6, CadetX activates full placement support. We do not just teach you data science — we help you get hired.

Starts Month 6	Full placement support activated for all students who complete the technical training and actively engage with the internship.
CV Review & ATS Optimisation	Tailored for data science roles
LinkedIn Profile Optimisation	Headline, about, skills, featured
GitHub Portfolio Review	Project quality, README, commit activity
Mock Technical Interviews	Python, SQL, ML case studies
Behavioural Interview Preparation	STAR method, common DS questions
Job Recommendations	Matched to your skills and location
Referral to CadetX Partner Companies	Direct introductions to hiring teams
Interview Shortlisting Support	We advocate for you with employers
CadetX Applies for Jobs on Your Behalf	Our team actively submits applications to matched roles — saving you time and ensuring your profile reaches the right employers directly
We Help You Until You Get a Job	Our placement support does not expire. We stay with you, keep recommending you, and keep opening doors until you secure your role
Portfolio Website	We build and host a professional portfolio website showcasing your projects, skills, and experience — a live URL you can share with any employer
Resume Preparation	We build your resume for you — professionally written, ATS-optimised, and delivered as a polished PDF ready to send to employers

CadetX Job Recommendation Service: We actively recommend students to hiring companies in our partner network. When a role matches your profile — skills, location, experience level, and target salary — we put your name forward directly. CadetX applies for jobs on behalf of students. You do not just apply into the void. We open doors and we do not stop until you are hired.

CERTIFICATION & EXPERIENCE LETTER

Course Completion Certificate	Data Scientist Launchpad — issued by CadetX
--------------------------------------	---

6-Month Internship Experience Letter	Role, projects completed, and skills demonstrated
Skills & Tools Documentation	Formal record of every technology covered — ready for employers

ALUMNI ACCESS

Lifetime Portal Access	CadetX student portal and updated course materials forever
Alumni Community	Connect with peers who completed the same programme
Ongoing Job Referrals	If a matching role appears after the programme, we still recommend you
Events and Webinars	Invitations to CadetX industry sessions and alumni events
New Module Access	Updates to course content as the field evolves
Peer Mentoring	Alumni now working in data science giving back to current students

TOOLS & TECHNOLOGIES COVERED

Programming Language	Python 3.x
Version Control	Git & GitHub
Data Manipulation	NumPy, Pandas
Visualisation	Matplotlib, Seaborn
Statistics	SciPy, Statsmodels
Machine Learning	Scikit-Learn, XGBoost or LightGBM
Deep Learning	Keras or PyTorch
NLP	NLTK or SpaCy; classical text modelling
Databases	SQL
Application & Deployment	Streamlit
Development Environment	Google Colab, Jupyter Notebook, VS Code

CadetX Data Scientist Launchpad — Zero to Hired

36 Sessions • 3 Months Technical Training (3 sessions/week) • Internship from Month 4 • Placement from Month 6 • Alumni Access for Life
www.cadetx.co.uk