alx

Data Science Programme Curriculum

In partnership with **EXPLORE AI** **ACADEMY**

alx

Data Science Programme Curriculum

Module 1: Explore 101 (1 Week)

Orientation

Setting up your learning environment	ExploreAI teaching philosophy and educational support framework	Troubleshooting at ExploreAI Academy

Introduction to Data and Data Analytics

What is data and how it is used to make data-driven decisions	An introduction to modern data practices and practitioners	Approaches to data analysis

Problem-solving

Mutually exclusive and collectively exhaustive statements and decisions	Design thinking and the scientific method	Introduction to solution-oriented communication

Programmatic Thinking

How to use algorithms and operators	Flowcharts, pseudocode, and conditional statements	Converting logic between statements, logic trees, pseudocode, and flowcharts

Module 2: Preparing Data (2 Weeks)

Introduction to Spreadsheets

Working with spreadsheets	Data types and formatting	Introduction to visualisation

Data Manipulation

Cleaning and analysing spreadsheet data	Working with various data types	Finding and fixing data anomalies

Introduction to Statistics

Summarising data using descriptive statistics	Measures of central tendency and spread	Samples and distributions

Introduction to Data Modelling

Basic spreadsheet functions and conditionals	Identifying patterns and the line of best fit	Testing assumptions and model accuracy

Data Science Programme Curriculum

Module 3: SQL (5 Weeks)

Introduction to SQL

Working with databases	Basic SQL data types and calculations	Aggregating, sorting, and grouping data

Relational Database Design

SQL schemas and entity relationships	Table normalisation, primary and foreign keys	Common table expressions and views

SQL in Practice

Set theory and SQL joins	Nested and subqueries	Improving query performance

Data Manipulation

Cleaning and analysing data	Working with numeric, time, and string data types	Data transformations and anomalies

Module 4: Data Visualisation and Storytelling (4 Weeks)

Data in Power BI

Loading and linking datasets in Power BI	Cleaning data and creating calculated columns and measures using DAX	Reports, data, and relationship views

Visuals in Power BI

Numeric visuals – cards, tables	Graphic visuals – line chart, bar chart, pie chart, column chart, treemap	Using slicers and custom visuals using DAX
Planning, designing, and prototyping	Working with various charts	Working with filters

Visual Storytelling

Telling a story with visuals	When to use which visuals	Presentation best-practice

www.alxafrica.com

Data Science Programme Curriculum

Module 5: Python (8 Weeks)

Python Programming Basics

Working in a Notebook environment	Pseudo code and debugging concepts	Working with primitive data types – variables, strings, integers, floating points, booleans

Functions and Control Flow

Creating and working with functions	Conditional statements	For loops and while loops

Data Structures

Lists, tuples, sets, and dictionaries	Working with DataFrames	Plots and graphs

Exploratory Data Analysis

Statistical measures, probabilities, and hypotheses	Algorithms and algorithmic complexity	Advanced interactive visual analysis

Module 6: Regression (5 Weeks)

Steps to Build a Model

Statistical learning, univariate and multivariate analysis	Training models, making predictions, testing accuracy	Variable significance and selection

Preparing Data for Modelling

Defining or engineering features and labels	Scaling, standardisation, and regularisation techniques	Splitting data for training, testing, and validation

Algorithms for Regression Models

K-nearest neighbours	Decision trees and random forests	Support vector machines

Model Tuning

Model performance metrics	Bias and variance	Hyperparameter tuning

Module 7: Natural Language Processing and Classification (5 Weeks)

An Overview of Natural Language Processing

Removing punctuation and symbols	Stopwords and regular expressions	Tokenizing text

Analysing Text

Lemmatisation of words	Bag of words	Sentiment analysis

Basic Classification

Logistic regression and binary classification models	Testing model output: confusion matrix, classification report	Feature engineering and selection

Advanced Classification

Hyperparameters and model validation	Dealing with imbalanced data and multi-class classification	Neural networks and image classification

Module 8: Unsupervised Learning (5 Weeks)

Dimensionality Reduction

Principal component analysis	Multidimensional scaling	Interpreting nonlinear transformations and embeddings

Hard and Hierarchical Clustering

What is clustering?	K-means clustering	Hierarchical clustering

Soft Clustering

Gaussian mixture models	Linear discriminant analysis and text clustering	Labelling data using cluster output

Recommender Systems

Measures of product similarity	Content and collaborative-based filtering	Evaluating a recommender system

Data Science Programme Curriculum

Module 9: AWS Foundations (5 Weeks)

Cloud Computing Basics

Introduction to cloud computing concepts	Pros and cons of cloud computing	Popular cloud service providers

Introduction to Amazon Web Services

Overview of AWS services	Networking and content delivery	Economics and billing

Storage and Compute Resources

Databases and object storage	Virtual machines	Serverless compute resources

Cloud Best Practice

Security, identity, and compliance	Cloud architecture framework	Automatic scaling and monitoring

Module 10: Integrated Exams (2 Weeks)

Review

Programme recap	Opportunity to review content in preparation for exams	Understanding the final assessment plan

Integrated Examination

Consolidated theory exam	Practical programming assessment	Applied machine learning exam