

In partnership with



# Data Science

---

Prospectus | 2023



EXPLORE AI  
ACADEMY

# Contents

---

Introduction	<b>Page 01</b>
<b>Program Overview</b>	<b>Page 02</b>
Explore AI Academy Overview	<b>Page 03</b>
<b>Who is This Course For</b>	<b>Page 04</b>
How You'll Learn	<b>Page 05</b>
<b>Requirements</b>	<b>Page 06</b>
Weekly/Monthly Expectations	<b>Page 07</b>
<b>Program Timeline</b>	<b>Page 08</b>
Program Tools	<b>Page 09</b>
<b>Curriculum Overview</b>	<b>Page 10</b>
Program Modules	<b>Page 11 – 20</b>

# Introduction

---

Welcome to the Data Science Program student handbook! This handbook is designed to provide you with all the necessary information about the ALX <> Explore AI Data Science Program. It is intended to familiarize you with the program's curriculum, features, requirements and objectives.

# Program Overview

---

The ALX Data Science Program is a **11 month's learning experience**

designed to help you gain in-demand data science and machine learning skills, the ALX Data Science programme will equip you with the crucial tools and techniques to analyse visual and statistical data, create models, and communicate results to inform data-driven decisions. In addition, you will learn the universal programming language Python and solve real-world problems using machine learning.

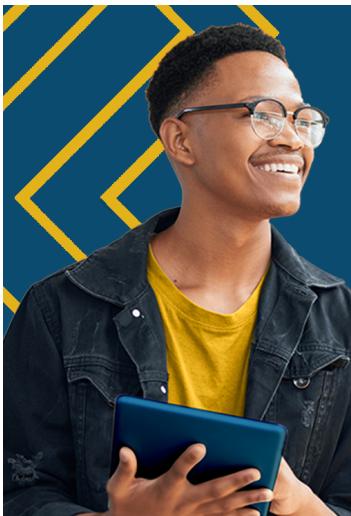
Delivered in partnership with ExploreAI, this globally-recognised certification will place you at the cutting edge of the digital economy, giving you a competitive advantage in the job market.

# ExploreAI Academy Overview

---

ExploreAI is a global data and AI solutions provider. We build AI-driven software for utilities, banks, insurers, and telcos. Within it is the ExploreAI Academy, whose mission is to transform the lives of talented African youth by equipping them with modern, relevant skills that enable them to find meaningful work.

We teach students the skills the global market is demanding, but that traditional education institutions are not producing. Our course content is curated and created by scientists with practical experience in the industry.



## Who is this course for?

You should consider doing this course if the following applies to you.

### 1 You want to learn new techniques in tech

The expert scientists at ExploreAI solve complex problems for big global companies. Their learnings are in turn used to refresh our Academy's course content at high velocity. The skills you'll learn here are thus modern, relevant, and used in production worldwide.

### 2 You want to improve your skills

You come from an adjacent or unrelated industry and want to futureproof your skillset. Or, you're in tech already but you feel your skills are out of date.

### 3 You want to solve real-world

We use an agile, project-based approach that immerses our students in the world of problem-solving and prepares them for the real world. You'll learn theory in each lesson, but quickly proceed to apply it.

# How you'll learn

Each module is broken up into manageable units called lessons.

---



Work through downloadable content and online instructional material.



You will continue to interact with your peers on the portal, where will also support you with non-technical questions. You will interact with tutors and your peers through the Answerbase platform for technical support.



Enjoy a wide range of interactive content, including video lectures and walk-throughs.



Investigate real-world case studies.



Apply what you learn each week in quizzes and ongoing project submissions, sharpening your ability to solve real-world problems.

# Program Requirements

This course and its subject matter are technical in nature. It is recommended that you have a basic understanding of mathematics and statistics.

## 1. Basic Requirements

- Basic computer literacy (using a web browser, operating an email account, downloading files, etc.).
- A current email account.
- Access to a computer, the internet, and PDF reader software.
- Access to the Google office productivity apps (Docs, Sheets, Slides – freely accessible to anyone with a Google account).
- Google Chrome to access the learning management system, though any popular browser should suffice.

## 2. Technical Requirements

- OS: Windows 10 recommended (Windows 7 minimum), in order to use Power BI; MacOS running Parallels for Windows will also suffice.
- Processor: Minimum i3, with a minimum clock speed of 2 GHz.
- RAM: Minimum 4 GB.
- Internet: A 10 Mbps line speed and 20 GB of data per month.

## 3. Additional Requirements

Please note that Google, Vimeo, and YouTube may be used in our course delivery, and if these services are blocked in your jurisdiction or on your device, you may have difficulty accessing course content. Please check with us before registering for this course if you have any concerns about access restrictions affecting your experience with our learning management system.

# Weekly/Monthly Expectations

- **Weekly Assignment submissions due every Monday**
  - Learning Webinars;  
**Every Tuesday & Thursday.**
- 

## Monthly

- ALX Foundations Milestone submission.
  - ALX Foundations Sessions (Office Hours, Speaker Series).
  - Self-Placement Reporting.
- 

## End of Program

- Selections Phase 3 Interviews
- Data Science Integrated Exam

# Program Timeline

**Start**



## ALX Foundations Phase 1

First 15 weeks exclusively focused on **ALX Foundations**

**3 Months**

## Data Science + ALX Foundations Phase 2

Tech Specialization with light focus on ALX Foundations

**11 Months**

### The Room Fellowship

Different fellowship levels to unlock at different stages of the program(s), and participants to gain full fellowship access once they graduate from the Tech verticals.

# Program Tools

During the course of the program, you will be using the following tools for different purposes.

## alx | THE ROOM

### Community Platform

Day-to-day communication with peers and program teams.



### Canvas

Learning platform where ALX Foundations Phase self-paced learning materials and assessments will be accessed.



### Answer Base

Explore AI Technical support channel where you can get support with technical questions.



### YouTube

Video sharing service where live streaming of events will take place



### ZOOM

Video conferencing platform where we host all events.



### Athena

Learning platform where Explore AI Data Science learning materials and assessments will be accessed.



### G-Cal

Shared program calendar Used for scheduling all program events and important program milestones / deadlines.



### Slido

Online polling tool used to capture questions before and during online sessions



### LEA/Zendesk

ALX Program Support channels, you can engage with our AI, LEA or contact support on Zendesk.

# Curriculum overview

This course will provide students with the knowledge, skills, and experience to get a job as a data scientist, which requires a mix of programming and statistical understanding. The course will teach students to gather data, visualise data, apply statistical analysis to answer questions, and make their insights and information as actionable as possible. We cover the fundamentals of the data scientist's toolkit as well as a broad set of machine learning algorithms.

**Duration:** 11 months

**Pre-requisite skills:** Basic analytical background

**Course difficulty:** Advanced

**Tools learned:** Google Sheets, Python, Jupyter Notebooks, MySQL, Power BI

Phase	Module	Duration (Weeks)	Recommended time (Hours)
Fundamentals	Explore101	2	15
	Preparing data	2	70
	SQL	5	175
	Data visualisation and storytelling	4	140
	Python	8	280
Machine learning	Regression	5	175
	Natural language processing and classification	5	175
	Unsupervised learning	5	175
Cloud practitioner	AWS foundations	5	175
Consolidation	Integrated exams	2	70

# Module 1

# Explore101

## What is covered in this module:

### Orientation

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

### 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 statements and decisions
- Flowcharts, pseudocode, and conditional statements
- Converting logic between statements, logic trees, pseudocode, and flowcharts

# Module 2

# Preparing data

## What is covered in this module:

### Introduction to spreadsheets

- Working with spreadsheets
- Data types and formatting
- Introduction to visualisation

### Data manipulation

- Cleaning and analysing spreadsheet data
- Data types and formatting
- Introduction to visualisation

### 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

# Module 3

# SQL

---

## What is covered in this module:

### 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

## What is covered in this module:

### 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

### Dashboards

- 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

# Module 5

# Python

## What is covered in this module:

### 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

## What is covered in this module:

### 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

### What is covered in this module:

#### 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

### What is covered in this module:

#### 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

## Module 9

# AWS foundations

### What is covered in this module:

#### **Cloud computing basics**

- Introduction to cloud computing concepts
- Pros and cons of cloud computing
- Popular cloud service providers floating points, booleans

#### **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 and certification requirements

---

### What is covered in this module:

#### 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

Need Support? Contact us via Ask ALX/EXPLORE AI Team channel on The Portal, For Tech support you can post your question on **Answerbase**.