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.

EXPLORE AI ACADEMY

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	1	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

EXPLORE AI ACADEMY

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

Introduction to data and data analytics

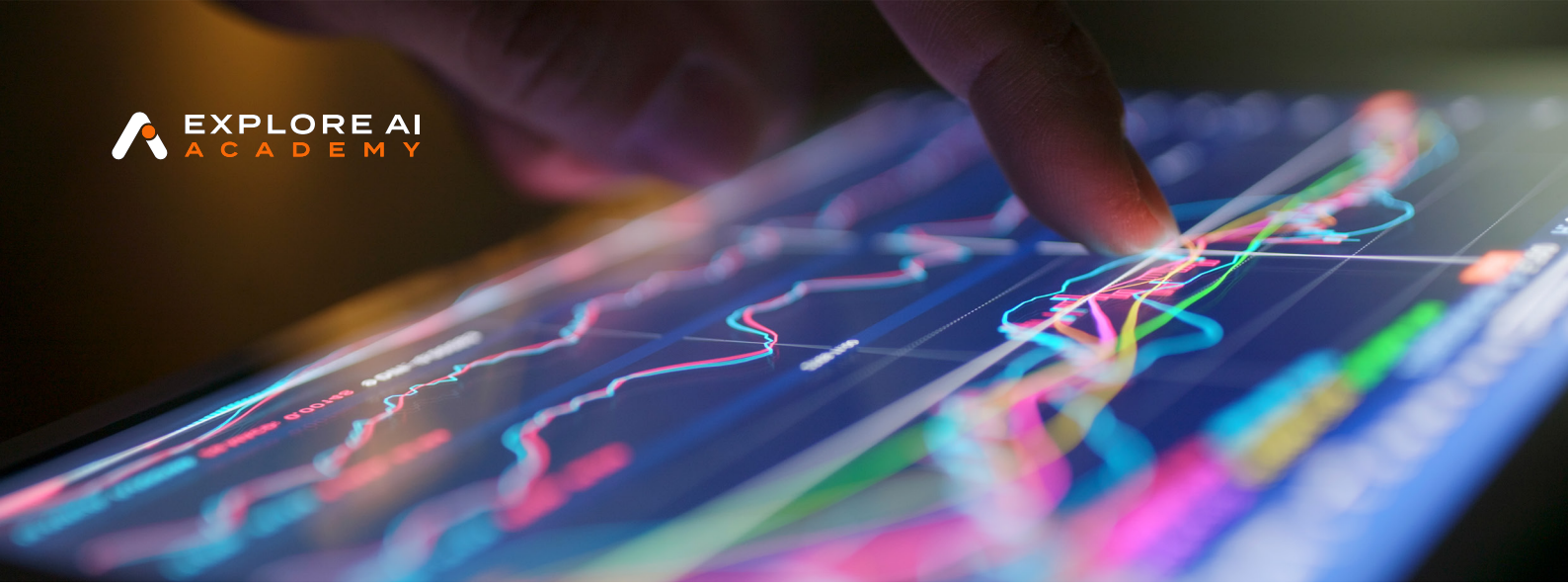
- What is data and how it is used to make data-driven decisions
- An introduction to modern data practices and practitioners
- Approaches to data analysis

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

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
- 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

EXPLORE AI ACADEMY

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

EXPLORE AI ACADEMY

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

EXPLORE AI ACADEMY

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

EXPLORE AI ACADEMY

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

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