



DATA SCIENCE PROSPECTUS

EXPLORE

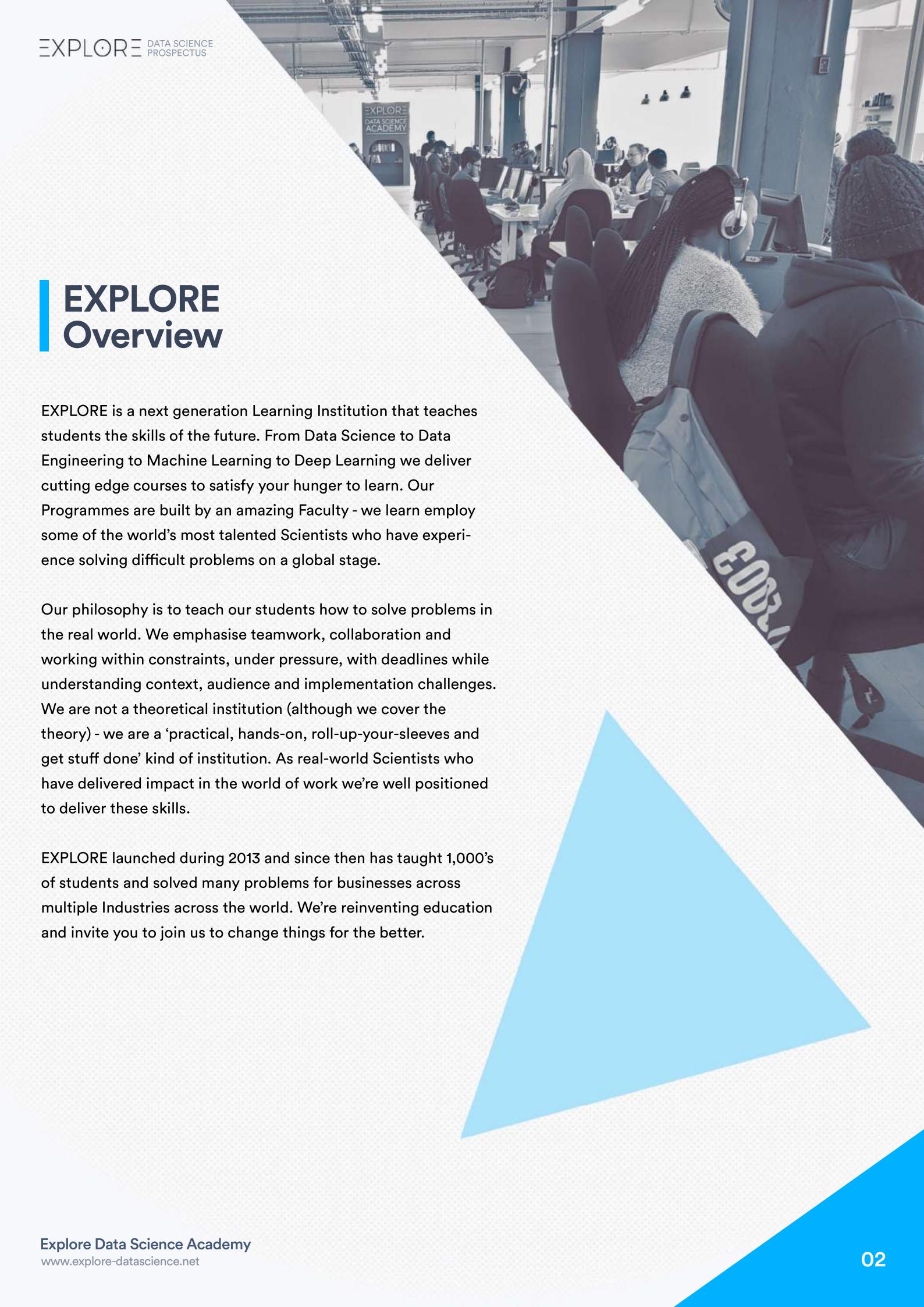
Explore Data Science Academy

www.explore-datasience.net

Table of Contents

EXPLORE Overview	02
Why Data Science?	03
This Course Is For You If You Want to	04
Who Should Take This Course?	05
How You Will Learn	05
Requirements	06
Your Success Team	07
Your Teaching Team	08
Curriculum Overview	09
— Module 1: EXPLORE 101	10
— Module 2: SQL for Data Science	11
— Module 3: Python for Data Science	13
— Module 4: Visualisation	15
— Module 5: Advanced Regression	17
— Module 6: Advanced Classification	19
— Module 7: Unsupervised Learning	21
EXPLORE Philosophy: Solving problems in the real world	23
Contact Information	25

EXPLORE Overview



EXPLORE is a next generation Learning Institution that teaches students the skills of the future. From Data Science to Data Engineering to Machine Learning to Deep Learning we deliver cutting edge courses to satisfy your hunger to learn. Our Programmes are built by an amazing Faculty - we learn employ some of the world's most talented Scientists who have experience solving difficult problems on a global stage.

Our philosophy is to teach our students how to solve problems in the real world. We emphasise teamwork, collaboration and working within constraints, under pressure, with deadlines while understanding context, audience and implementation challenges. We are not a theoretical institution (although we cover the theory) - we are a 'practical, hands-on, roll-up-your-sleeves and get stuff done' kind of institution. As real-world Scientists who have delivered impact in the world of work we're well positioned to deliver these skills.

EXPLORE launched during 2013 and since then has taught 1,000's of students and solved many problems for businesses across multiple Industries across the world. We're reinventing education and invite you to join us to change things for the better.

Why Data Science?

Four megatrends are fundamentally changing the shape of our world:



Unlimited Data

Vast amounts of data are being generated every minute.



Cloud Integration

We now have cloud providers who can store insane amounts of data for a few dollars.



Incredible Speed

The processing speed of our machines is increasing exponentially.



Open Source Algorithms

Powerful open source algorithms that can read, write, translate and see are now available to everyone.

Data Science is the skillset used to harness the power of these tectonic shifts in our world. With the Data Scientist's tool-kit almost any problem can now be solved.

It is no surprise that Data Scientists are one of the most highly paid and in demand skills of the 21st Century. In fact, the job 'Data Scientist' was rated the 'sexiest profession of the 21st Century' by the Harvard Business Review.

This Course Is For You If You Want To

This course is 100% for you if these are what you are looking for:



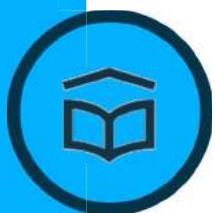
Improve Your Skills

Learn the latest advances to turbo-charge your career and set you up for success in the digital age.



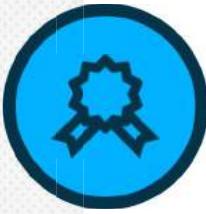
Learn New Techniques

Learn what matters - invest in your career by studying the latest algorithms, cloud solutions, open source tools and data science technologies.



Have a Support Team

EXPLORE has a large Faculty of Scientists with lots of Academic and real-world experience who are available to support you through the learning process.



Solve real-world problems

EXPLORE exposes you to real-world problems by delivering project-based deliverables where you need to apply your knowledge to solve problems.

Who Should Take This Course?

This course is geared towards newcomers to the field of data science who want to close the gaps in their analytical skills and knowledge. It will be beneficial to professionals who need to rapidly upskill and enhance their data science toolkit with demonstrable and practical skills.

It is also beneficial to professionals working in a range of business areas, including but not limited to marketing, sales, finance, and operations who want to learn how to use data and programming to increase efficiencies and identify new opportunities for their organisation.

This course is technical in nature. It is therefore recommended that you have a basic understanding of mathematics and statistics. Basic knowledge of at least one programming language would be beneficial and is recommended but is not required.

How You'll Learn

Every course is broken down into manageable, weekly units, designed to accelerate your learning process through diverse learning activities:



Work through downloadable content and online instructional material.



Investigate real-world case studies.



Interact with your peers and learning facilitators through real-time chat platforms and regular live webinars.



Learn how to use Jupyter Notebook, GitHub, Power BI, AWS, and various machine learning models.



Enjoy a wide range of interactive content, including video lectures, coding challenges, hackathons, and presentations.



Apply what you learn each week to quizzes, coding challenges and ongoing project submissions, culminating in the ability to use real-world data to solve a real-life problem.

Software Requirements

BASIC REQUIREMENTS

Basic computer literacy is essential for successful completion of the course. To complete the course, you will need a current email account and access to a computer and the internet, as well as a PDF Reader. You may need to make use of the Google office suite, which is free to anyone with a Gmail account. Alternatively, Microsoft Office applications (such as PowerPoint and Word) may be used. We recommend using Google Chrome to access the course, but any popular browser should suffice.



TECHNICAL REQUIREMENTS

- ✓ Recommended OS: Since we make use of Windows based apps (such as Power BI), Windows 10 is recommended (Windows 7 minimum)
- ✓ Processor: Minimum i3. Baseline higher than 2GHz
- ✓ RAM: Minimum 4GB
- ✓ Bandwidth: 20GB data per month should be sufficient, but will vary based on your personal activity
- ✓ Communication: Webcam and Microphone



ADDITIONAL REQUIREMENTS

Certain activities may require additional software and resources. In such cases, we will ensure that they will operate properly on a device with the above-mentioned specifications and will be clearly communicated to you during the completion of the course. Please note that Google, Vimeo, YouTube, Udemy and Datacamp may be used in our course delivery, and if these services are blocked in your jurisdiction, you may have difficulty in accessing course content. Please check with a Course Consultant before registering for this course if you have any concerns about this affecting your experience with the Online Campus.

Your Success Team

EXPLORE has a range of Faculty members at hand to assist you if you get stuck during your learning journey. We have experts readily available to assist you for when things get tricky - they have a wealth of experience with the material and are a friendly message away to help you on your way.



Head Tutor

A subject matter expert who'll guide you through content-related challenges.



Administration Support

Available to solve your software, tech and administrative queries and concerns.

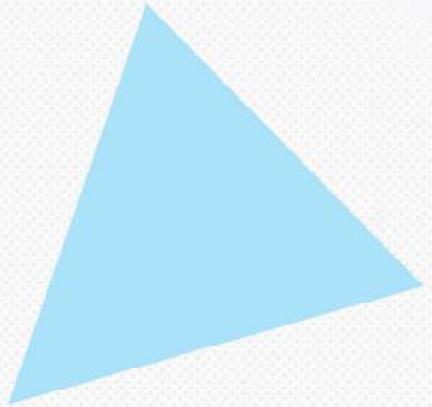


Success Manager

Your one-on-one support available during working hours (8am-5pm SAST) to resolve any challenges that you may have

Your Teaching Team

These EXPLORERs built the course you will go through - they have loads of experience working with the tools and technologies you will learn and had fun building something awesome for you to learn.



SUBJECT MATTER EXPERT



Jaco Jansen van Rensburg

Lead Instructor

Jaco is a Lead Data Scientist in the EXPLORE Data Science Academy. He has spent the bulk of his career on scientific and industrial research and holds a PhD in Mechanical Engineering with a focus on mathematical modelling and optimisation.

YOUR COURSE CO-CONVENORS

These subject matter experts guide the course design and appear in a number of course videos, along with a variety of industry professionals.



Jonathan Gerrand

Lead Instructor

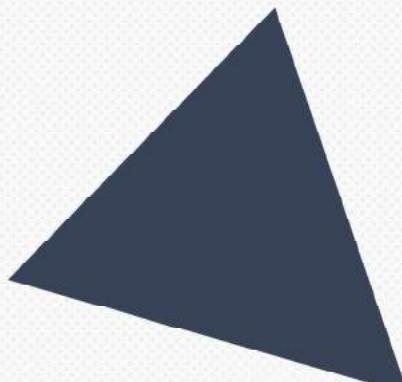
Jonathan is an electrical engineer specialising in biomedical image processing and machine learning (MSc). Having been a founding member of the Data Science for Impact and Decision Enablement (DSIDE) program within the CSIR, he is passionate about teaching, innovating for good, and equipping others to reason within a data driven world. Jonathan loves the outdoors, is unashamedly nerdy, and relishes a compelling debate.



Jonathan Botha

Data Scientist

Jonathan has his PhD in Genetics. He decided to leave the world of Academia and pair his unique perspective and insight with Data Science in order to see what strange and interesting things unfold. He also unashamedly prefers tea to coffee.



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 toolkit as well as a broad set of machine learning algorithms.



Duration: **12 Months**



Recommended Time: **440 hours**



Pre-Requisite Skills: **Basic analytical background**



Course Difficulty: **Advanced**



Tools Learnt:



Phase	Module	Time
Fundamentals	EXPLORE 101 SQL for Data Science Python for Data Science Visualisation	20 hours 60 hours 60 hours 60 hours
Machine Learning	Advanced Regression Advanced Classification Unsupervised Learning	80 hours 80 hours 80 hours

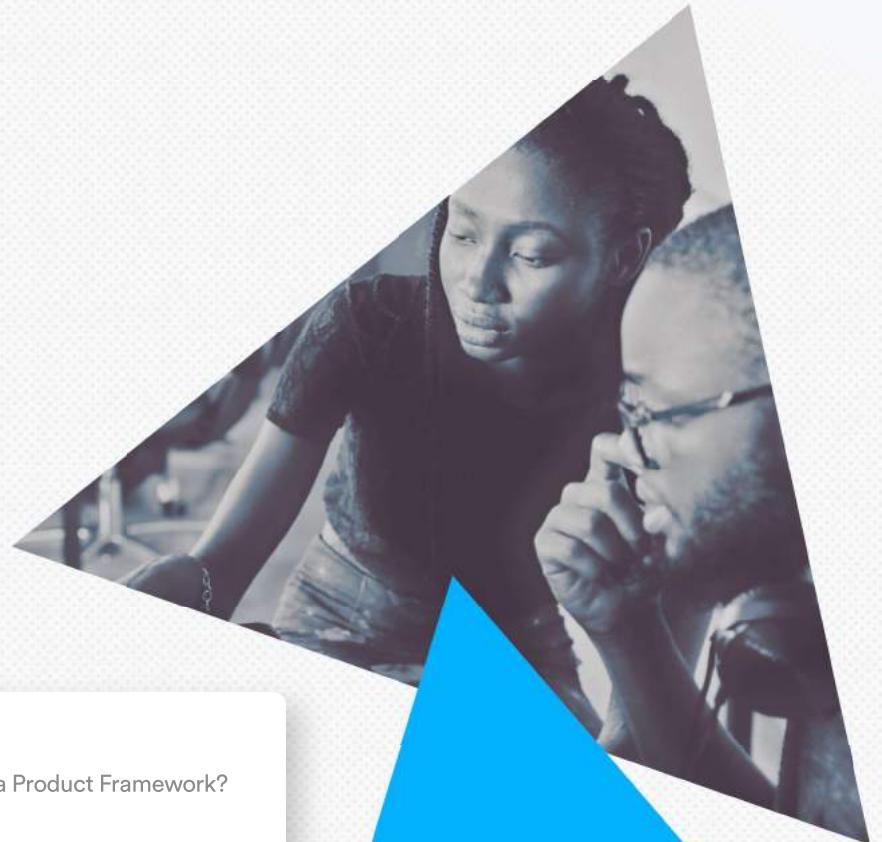
Module 1

EXPLORE 101

⌚ Recommended Time: **20 hours**

Learn about the EXPLORE Data Product Framework

What is covered in Module 1:



EXPLAIN

What is “explaining” in the EXPLORE Data Product Framework?

- ✓ Defining a Problem Statement
- ✓ Introduction to Communication and Visual Storytelling



GATHER

What is “gathering” in the EXPLORE Data Product Framework?

- ✓ Understand the Problem Landscape
- ✓ Introduction to Databases and Data Engineering



ANALYSE

What is “analysing” in the EXPLORE Data Product Framework?

- ✓ Understand the project Equation of Value
- ✓ Introduction to Programming and Solution Governance



DEPLOY

What is “deploying” in the EXPLORE Data Product Framework?

- ✓ Understand the basics of Project Management
- ✓ Introduction to Version Control and Production

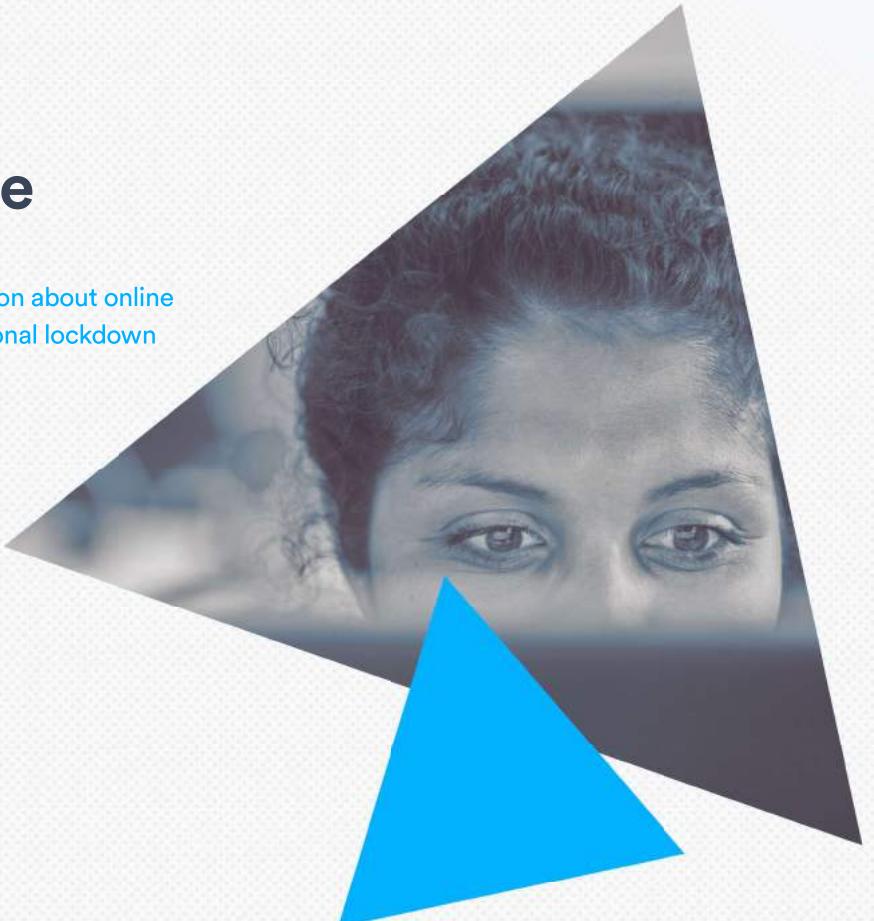
Module 2

SQL for Data Science

 Recommended Time: **60 hours**

 Project: Set up and extract valuable information about online retailers during a pandemic and national lockdown

What is covered in Module 2:



A. Working with SQL



Introduction to SQL

-  Working with databases
-  The value of well structured data
-  How to work with data



Basic SQL

-  SELECT, FROM, WHERE
-  JOINS
-  Aggregations

Module 2

SQL for Data Science



B. Working with databases



Database Management

- ✓ Normalisation of TABLES
- ✓ CREATE, ALTER commands
- ✓ Working with temporary tables
- ✓ Optimising performance



Data Manipulation

- ✓ INSERT, UPDATE, DELETE commands
- ✓ Cleaning data
- ✓ Writing complex SQL queries
- ✓ Real-world project

Module 3

Python for Data Science

 Recommended Time: **60 hours**

 Project: Write a set of functions into a module that calculates specific metrics and analyses a company

What is covered in Module 3:

A. Python Fundamentals



Python programming basics

- ✓ Working in a Notebook environment
 - ✓ Pseudo code and debugging concepts
 - ✓ Interactive vs scripting mode
 - ✓ Working with primitive data types - variables, strings, integers, floating points, booleans



Logic and functions

- ✓ Conditional statements - IF and ELSE IF
 - ✓ Working with lists
 - ✓ For loops and while loops
 - ✓ Break | Continue principles
 - ✓ Creating and working with functions

Module 3

Python for Data Science

B. Python Data Structures



Data Types

- ✓ Working with Strings, Numbers, Booleans
- ✓ Lists and Tuples Semantics
- ✓ Working with Comparisons
- ✓ Working with Statements



Dataframes and using libraries

- ✓ Sets and Dictionaries Semantics
- ✓ Working with Comparisons
- ✓ Importing Data - using Numpy and Pandas libraries
- ✓ Working with Data Frames

Module 4

Visualisation

⌚ Recommended Time: **60 hours**

💡 Project: **Fix a dashboard and drill down into the insights**

What is covered in Module 4:



A. Principles of Visualisation



Visualisation

- ✓ Telling a story with visuals.
- ✓ When to use which visuals.
- ✓ Tools for visualisation.



Power BI Basics

- ✓ Report, data and relationship view.
- ✓ Loading and linking datasets in Power BI.
- ✓ Cleaning data in query editor.
- ✓ Create calculated columns and measures using DAX.

Module 4

Visualisation



B. Building Dashboards



Visuals in Power BI

- ✓ Numeric visuals - cards, tables.
- ✓ Graphic visuals - line chart, bar chart, pie chart, column chart, treemap.
- ✓ Using slicers.
- ✓ Import custom visuals.



Dashboards

- ✓ Planning, designing and prototyping.
- ✓ Working with various charts.
- ✓ Working with filters.
- ✓ Real-world project.

Module 6

Advanced Regression

⌚ Recommended Time: **80 hours**

💡 Project: **Predict the prices of fresh produce to plan your inventory**

What is covered in Module 6:

A. Pre-processing and Model Building



Preparing data for modelling

- ✓ Basics of Scikit Learn library in Python
- ✓ Define features and labels
- ✓ Train/test split
- ✓ Encoding categorical features
- ✓ Standardising features



Steps to build a model

- ✓ Overview of regression
- ✓ Linear regression models
- ✓ Train models and make predictions
- ✓ Regularization techniques - Ridge, Lasso
- ✓ Test model output - MSE, MAE, RMSE

Module 6

Advanced Regression

B. Regression Algorithms



Algorithms for regression models

- ✓ K nearest neighbors
- ✓ Decision trees
- ✓ Random forests
- ✓ Support vector machines



Model Tuning

- ✓ Optimizing model performance by tuning
- ✓ hyperparameters
- ✓ Gridsearch techniques
- ✓ Which models to use when

Module 6

Advanced Classification

⌚ Recommended Time: **80 hours**

💡 Project: **Using tweets, predict if someone believes in Climate Change**

What is covered in Module 6:



A. Classification Algorithms



Steps to build a model

- ✓ Overview of classification
- ✓ Logistic regression models
- ✓ Build and train models in python
- ✓ Make predictions using trained model
- ✓ Test model output - confusion matrix, classification report



Algorithms for classification models

- ✓ K nearest neighbors
- ✓ Decision trees
- ✓ Random forests
- ✓ Support vector machines

Module 6

Advanced Classification

B. Natural Language Processing



Overview of NLP

- ✓ Overview of NLTK library in python - for NLP
- ✓ Removing punctuation and symbols
- ✓ Tokenizing text
- ✓ Stopwords and removing them



Analysing text

- ✓ Lemmatization of words
- ✓ Bag of words
- ✓ Sentiment analysis

Module 7

Unsupervised Learning

⌚ Recommended Time: 80 hours

💡 Project: Build a recommender engine that suggests movies based on your list of favourites

What is covered in Module 7:



A. Clustering



Hard and Hierarchical Clustering

- ✓ What is clustering?
- ✓ Various forms of clustering - hard, hierarchical, soft
- ✓ K-Means Clustering
- ✓ Hierarchical Clustering



Soft Clustering

- ✓ Gaussian Mixture Models
- ✓ LDE clustering of text
- ✓ Labelling data using cluster output

Module 7

Unsupervised Learning



B. Exploiting Relationships



Recommender Systems

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



Dimensionality Reduction

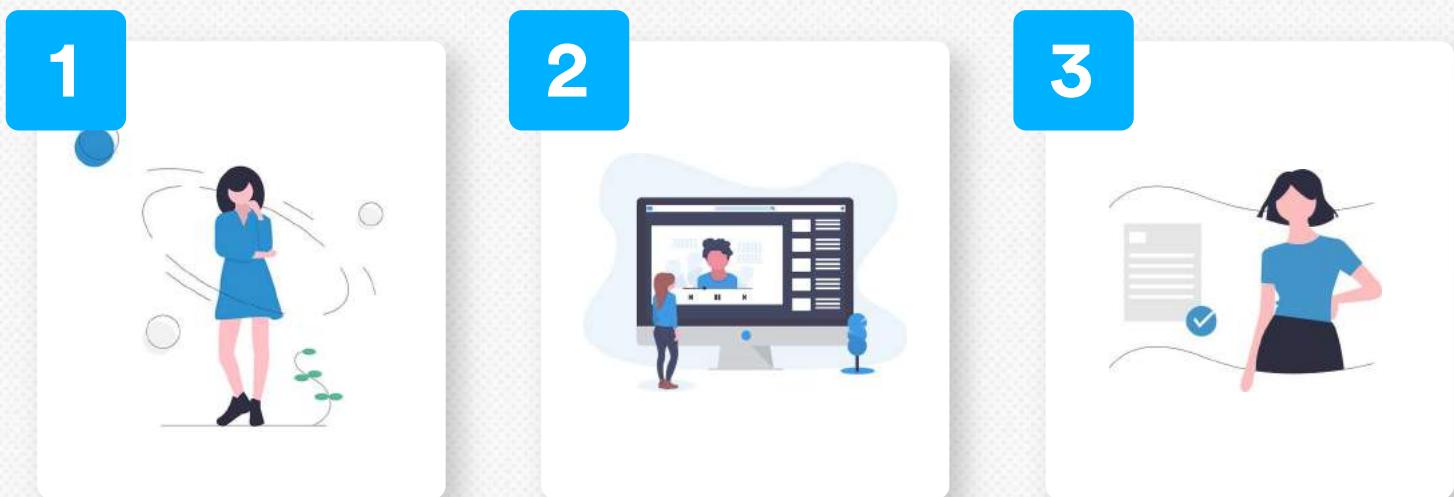
- ✓ Principal component analysis
- ✓ Linear discriminant analysis
- ✓ Interpreting extracted features and transforming them back

EXPLORE Philosophy: Solving problems in the real world

At EXPLORE we focus on building our student's ability to solve problems in the real world. Building things that work and make a difference is hard - that's what we teach.

We're not a traditional learning institution that spends weeks teaching matrix multiplication on a whiteboard (although understanding that is useful) - we're a practical, solution-oriented institution that teaches our students to work in teams, under pressure, with deadlines while understanding context, constraints and the audience.

Our courses are typically broken into Sprints where we teach a core set of concepts within the framework of solving a problem in a team with a tight deadline.



Students cycle from Sprint to Sprint solving different problems in different teams as they build this core muscle over the course.



EXPLORE

Start a new career today by enrolling in one of our
Data Science, Data Analytics or
Data Engineering courses.

Admission Related Enquiries Mail:
admissions@explore-ai.net

General Enquiries Mail:
general@explore-ai.net

Website:
www.explore-datasience.net