horizontal line

**Blossom Academy**

Accra-Ghana

Kigali-Rwanda

Lagos-Nigeria

Data Science Curriculum

# **OVERVIEW**

Blossom Academy’s Data Science Course is a transformative course that prepares professionals to break into data careers.

As a graduate, you’ll leave poised to succeed in various Data Science and Analytics roles, creating predictive models that drive decision-making and a strategy throughout organizations.

**Throughout the expert-designed course, you’ll:**

* Collect, extract, query, clean, and aggregate data for analysis
* Store and organize data using SQL and Git, enabling scalable workflows
* Analyze, manipulate, and preprocess data using Pandas and NumPy, while identifying key insights and patterns.
* Craft and share compelling narratives through data visualizations.
* Apply descriptive and inferential statistics, probability distributions, hypothesis testing, and A/B testing for data-driven decision-making.
* Build and implement appropriate machine learning modules and algorithms to evaluate data science problems in finance, e-commerce, and other fields.
* Design and deliver forecasting models and recommendation systems.
* Deploy Machine Learning Models and APIs.
* Prepare for work, compiling a professional-grade solo, group, and client projects portfolio.

**PREREQUISITE**

This course is designed for intermediate learners. We recommend that students arrive with a mathematical foundation or familiarity with Python and programming fundamentals. Some of our successful graduates joined our program with technical backgrounds, such as a degree in computer science or mathematics or work experience in research or analysis. Other students engage in self-learning to build a foundation ahead of class.

**The course also requires**

* Your laptop (PC or Mac). It must be no more than four years old and able to run the most recent operating system.
* If remote, a webcam, headphones, and reasonable access to the internet are required.

We recommend students new to data science pursue our [**Data Analytics**](https://docs.google.com/document/d/16QM_5DbTvI_KVh6SHJq_iqszqc0xNJLrWD_b0S6rp-k/edit?usp=sharing)course.

**CAPSTONE PROJECT**

The Capstone Project spans the entire training phase of the program (16 weeks) in which students pursue independent projects on a question or problem of their choice, subject to approval from Blossom Academy.

Students are strongly encouraged to choose a project

1. that has 70%+ relevant topics covered within the program,

2. in which data availability is 70%+ specific.

This is to ensure the project can be completed within the stipulated period.

The capstone project is a valuable intellectual experience and a vehicle for students to demonstrate their competency to prospective employers.

Note: Completion of the Capstone Project is a requirement for graduation.

Each project is designed to provide hands-on experience with as many data science concepts as possible.

**DURATION**

7 - 10 months (Full-Time)

**PRE-WORK**

**Data Science Introduction.**

**Before the course begins, dive into a series of self-paced lessons on Python programming essentials and applied math for data science.**

* Explore fundamental Python programming concepts, including variables, lists, loops, dictionaries, and data sets.
* Leverage programming tools like GitHub and the command-line interface to manage data science projects.
* Practice solving coding challenges similar to the questions used in task-based data science interviews.
* Write and run Python functions using multiple arguments.
* Discover how key math concepts like statistical significance and probability distribution are applied throughout data science.

## **CONTENT**

| **Unit 0** | **Foundations** | **Introduction to Data Science**   * What is Data Science? * Real-world applications in African tech and startups * Career tracks overview: Data Scientist, ML Engineer, Data Product Manager   **Development Environment Setup**   * Mastering Jupyter Notebooks and Visual Studio Code for scalable workflows * Comprehensive Git and GitHub workflow: version control, branching strategies and collaboration best practices * Managing project dependencies and virtual environments for reproducibility   **Python Fundamentals Refresher**   * Variables, Data Types, Control Flow (loops, conditionals) * Functions and Lambda Expressions * Working with Pandas and Numpy Basics   **Modular Programming & Code Structuring**   * Writing reusable, testable code in modules and packages * Benefits: maintainability, collaboration, code reuse   **Object-Oriented Programming (OOP)**   * Classes, objects, attributes, methods * Encapsulation and inheritance   **Functional Programming (FP)**   * Pure functions, immutability, side-effects * Higher-order functions: map( ), filter( ), reduce( ) * Lambdas and list comprehensions vs FP style * When and why to use FP in data workflows   **APIs & Web Scraping**   * Consuming REST APIs with requests * JSON parsing and data extraction * Web Scraping basics with BeautifulSoup   **AI Fundamentals & Generative AI for Data Science**   * History and Types of AI * AI vs ML vs Data Science * What is Generative AI? Why it matters now * Use cases in African industries: fintech, govtech, edtech * Strategic Uses of GenAI for Data Science * Understanding LLMs, multimodal models (text-to-image, code generation) * Risks: hallucinations, bias, ethics, regulatory context |
| --- | --- | --- |
| **Unit 1** | **Data Manipulation & Visualisation** | **Data Manipulation with Pandas & Numpy**   * Advanced data wrangling: reshaping, pivoting, merging datasets * Handling missing values: imputation, deletion, interpolation * Removing duplicates and cleaning data efficiently * Lambda functions and method chaining for clean, expressive code   **Exploratory Data Analysis (EDA)**   * Understanding feature types: categorical, numerical, datetime * Distribution analysis and visualization techniques * Outlier detection and treatment strategies * Correlation analysis and deriving domain-specific insights   **Data Visualization**   * Static plotting with Seaborn and Matplotlib: histograms, boxplots, scatterplots * Interactive visuals with Plotly: dashboards, drill-downs * Choosing the right chart for your data story |
|  | **Unguided Project** | Perform EDA and visualize insights from a structured dataset |
| **Unit 2** | **Scientific Computing & Quantitative Methods** | **Applied Business Statistics**   * Descriptive statistics: mean, median, mode in business contexts * Confidence intervals and margin of error interpretation for decision support   **Probability Distributions for Business**   * Application of normal and binomial distributions for forecasting and conversion analysis   **Experiment Design & A/B Testing Frameworks**   * Designing statistically valid experiments aligned with startup needs * Significance testing and interpreting results for product and marketing optimization |
|  | **Unguided Project** | Design, implement, and analyze an A/B test for a real-world business scenario |
| **Unit 3** | **SQL for Data Science** | **Introduction to SQL**   * Basics * CRUD Operations   **Queries & Aggregations**   * WHERE * GROUP BY * ORDER BY * Aggregates * HAVING * NULL handling * Pattern Matching * CASE WHEN statements (conditional logic inside queries) * UNION and UNION ALL   **Joins & Subqueries**   * INNER * LEFT * RIGHT * FULL JOIN * Subqueries   **Advanced SQL Techniques**   * Window Functions * CTEs   **SQL Integration in Python**   * Connecting to SQL databases with SQLAlchemy * Embedding SQL in business workflows: finance, marketing, operations |
|  | **Unguided Project** | Conduct advanced SQL analysis to extract actionable business insights |
| **Unit 4** | **Machine Learning for Data Science** | **Supervised Learning**   * Linear models: Linear Regression, Ridge, Lasso for prediction * Tree-based models: Decision Trees, Random Forests, Gradient Boosting * Classification techniques: Logistic Regression, Support Vector Machines * Model Evaluation Metrics: Train-test split, cross-validation frameworks, Precision, recall, F1-score, ROC-AUC and business relevance * Feature Engineering and Selection - Encoding categorical variables, feature scaling, Feature importance analysis and dimensionality reduction   **Gradient Boosting Frameworks**   * Implementation and tuning of XGBoost, LightGBM, CatBoost * Practical strategies for model optimization and handling class imbalance   **Introduction to Neural Networks**   * Perceptron * Activation functions * Forward/Backpropagation * Simple implementation using Scikit-Learn   **Unsupervised Learning**   * K-Means clustering for customer segmentation * Principal Component Analysis (PCA) for feature extraction |
|  | **Unguided Project** | Build, evaluate, and present a classification model with business-ready outputs |
| **Unit 5** | **Decision Science** | **Data-Driven Decision Making**   * Understanding data-driven strategies * Case study: Finance (Risk analysis, fraud detection) * Case study: Marketing (Customer segmentation, A/B testing) * Case study: Operations (Supply chain optimization, demand forecasting)   **Time Series Forecasting**   * Introduction to time series forecasting * ARIMA (AutoRegressive Integrated Moving Average) * SARIMA (Seasonal AutoRegressive Integrated Moving Average) * Exponential Smoothing for trend analysis * Leveraging Prophet for scalable business forecasting   **Recommendation Systems**   * Collaborative filtering: user-based and item-based approaches * Content-based filtering techniques * Real-world applications in e-commerce, edtech, and fintech |
|  | **Unguided Project** | Develop a time series forecasting model, recommendation engine with clear business value |
| **Unit 6** | **Model Deployment** | **Model Deployment & APIs**   * Introduction to Streamlit * Rapid ML app prototyping with Streamlit * Building production-grade ML APIs with FastAPI * Comparing Streamlit and FastAPI use cases for prototyping vs. production * Deployment strategies on Streamlit Cloud * Securing and testing ML endpoints for production readiness |
|  | **Capstone Project** | End-to-end ML project using real datasets (Finance, Retail, Healthcare) |