

About the Company

We aim to bridge the gap between academic learning and industry expectations through practical, industry-oriented education. We are building a community of industry professionals passionate about sharing their expertise and making high-quality, industry-relevant education accessible to aspiring professionals.

Benefits of Completing the Course

- **Production-Ready Portfolio:** Build and deploy a complete portfolio ranging from SQL and Power BI dashboards to scalable PySpark pipelines and LLM RAG agents.
- **Modern Tech Stack Mastery:** Gain hands-on expertise in Python, SQL, Snowflake, Power BI, Streamlit, FastAPI, MLflow, Airflow, PySpark, and Vector Databases.
- **Real-World Experience:** Solve complex enterprise challenges in customer segmentation, time series forecasting, automated MLOps, and generative AI.
- **Career Acceleration:** Position yourself strongly for high-demand roles like Data Scientist, Analytics Engineer, and AI/Data Product Developer.

Month-by-Month Course Overview

Month 1: Data Wrangling, Advanced SQL & EDA

- **Module 1: Python for Data Science & Scalable Wrangling:** Scalable Pandas dataframes and multi-indexing; Ingesting heterogeneous data (JSON, CSV, Parquet, REST APIs); Data hygiene, cleaning, and missing value remediation; Text analytics, regex pattern extraction, and modular cleaning pipelines.
- **Module 2: Enterprise SQL & Advanced Database Analytics:** Relational database schemas and normalization (1NF to 3NF); Advanced table joins and complex data aggregations; Common Table Expressions (CTEs) and modular SQL scripting; Window functions (ROW_NUMBER, RANK, LEAD, LAG) and performance tuning.
- **Module 3: Exploratory Data Analysis & Visual Analytics:** Automated data profiling and initial sanity checks; Univariate, bivariate, and multivariate visualizations; Interactive data visualizations using Plotly; Outlier analysis, anomaly detection, and insight storytelling.
- **Module 4: Applied Business Statistics & Decision Making:** Descriptive metrics, probability distributions, and confidence intervals; Hypothesis testing (Z-tests, t-tests, Chi-Square, ANOVA); Business experimentation and A/B testing design; Evaluating statistical significance and conversion lifts.

Month 2: Applied Machine Learning & Predictive Analytical Systems

- **Module 1: Feature Engineering & Dimensionality Reduction:** Categorical feature encoding and numeric normalization; Feature generation and automated selection strategies (RFE, Variance Thresholds); Dimensionality reduction using PCA and t-SNE; Handling class imbalances via SMOTE and resampling.
- **Module 2: Predictive Modeling: Regression & Tree-Based Algorithms:** Regularized regression (Ridge, Lasso) and coefficient interpretation; Decision Trees and ensemble methods (Random Forests, Extra Trees); Gradient boosting frameworks (XGBoost, LightGBM, CatBoost); Automated hyperparameter tuning using Optuna and GridSearchCV.
- **Module 3: Classification, Clustering & Customer Segmentation:** Classification algorithms (Logistic Regression, SVM, KNN, Naive Bayes); Evaluation metrics (Precision, Recall, F1-Score, ROC-AUC, PR Curves); Unsupervised clustering (K-Means, Hierarchical Clustering, DBSCAN); Customer segmentation profiling and RFM analytics framework.
- **Module 4: Time Series Analysis & Forecasting:** Time series decomposition (Trend, Seasonality, Residuals); Stationarity testing (Augmented Dickey-Fuller test) and differencing; Classical forecasting models (ARIMA, SARIMA, Exponential Smoothing); Modern forecasting baselines with Meta Prophet and backtesting frameworks.

Month 3: BI Dashboards, Web Apps & Milestone Capstone

- **Module 1: Business Intelligence & Interactive Dashboards:** Working with Power BI and Tableau for enterprise reporting; Star and Snowflake schema data modeling; Writing DAX and calculated fields for business metrics; Dashboard layout design, executive KPIs, and automated refreshes.
- **Module 2: Data Science Web Apps & API Integration:** Model serialization via Joblib, Pickle, and ONNX; Building lightweight REST endpoints with FastAPI and Pydantic; Interactive data app development using Streamlit; Frontend-backend integration and data product UI/UX design.
- **Module 3: Cloud Data Warehousing & Pipeline Orchestration:** Querying and storing data in Google BigQuery and Snowflake; Workflow automation and pipeline orchestration using Apache Airflow; Modern data stack architectures (ETL vs. ELT); Containerizing data applications with Docker.
- **Module 4: Milestone Capstone Project I:** End-to-end intermediate data product build and deployment; Multi-source dataset cleaning and predictive model pipeline creation; Interactive Streamlit application or Power BI reporting integration; Code repository structuring and technical presentation.

Month 4: Advanced Production ML & MLOps Infrastructure

- **Module 1: MLOps & Experiment Tracking with MLflow:** Tracking parameters, metrics, and model artifacts; Managing centralized model registries; Ensuring pipeline reproducibility and version control.
- **Module 2: Model Monitoring & Production Maintenance:** Detecting data drift and concept drift in production; Performance degradation identification and automated alerts; Continuous monitoring strategy implementation.
- **Module 3: Containerization & Serverless Cloud Deployment:** Packaging ML models and APIs into secure Docker images; Deploying to cloud environments (AWS ECS, Google Cloud Run); Managing serverless architecture patterns.
- **Module 4: Scalable Data Product Architecture:** Designing fault-tolerant microservices for analytics; Handling asynchronous background tasks with Celery and Redis; Securing inference endpoints and managing environment secrets.

Month 5: Large-Scale Data Engineering & Cloud Platforms

- **Module 1: Distributed Computing with PySpark:** Spark core architecture and Resilient Distributed Datasets (RDDs); Spark DataFrame operations and transformations; Parallel execution models and cluster management.
- **Module 2: Big Data Performance Tuning & Optimization:** Spark cluster configuration and tuning strategies; Shuffle optimizations and memory management; Caching strategies for massive, petabyte-scale datasets.
- **Module 3: Production Data Pipelines & Data Quality:** Building production Apache Airflow DAGs; Pipeline idempotency, data lineage, and error handling; Automated data quality assertions using Great Expectations.
- **Module 4: NoSQL, Vector DBs & Enterprise Storage Caching:** Working with document stores (MongoDB); Caching query responses with Redis; Storing high-dimensional embeddings in vector databases (Pinecone, Milvus).

Month 6: Generative AI, LLMs & Final Capstone

- **Module 1: Generative AI Foundations for Data Scientists:** Large Language Model architectures and tokenization mechanics; Prompt engineering patterns and structural design; Evaluating cost-performance trade-offs in generative models.
- **Module 2: Retrieval-Augmented Generation (RAG) Systems:** Enterprise document ingestion and parsing; Text chunking strategies and embedding generation; Vector similarity search and internal QA microservice construction.
- **Module 3: LLM Fine-Tuning & Advanced Evaluation Frameworks:** Parameter-efficient fine-tuning (PEFT) using LoRA and QLoRA; Custom dataset preparation for fine-tuning workflows; Automated RAG evaluation metrics using Ragas and TruLens.
- **Module 4: Final Capstone Project: Production AI & Data Product:** Enterprise-grade multi-service system design; Integrating data ingestion, ML predictions, and LLM automation; Cloud orchestration and containerized production deployment; Final system demonstration and code review.

Learning Journey Visual Roadmap (Step-by-Step Process Flow Layout)

1

Month 1 Roadmap: Data Wrangling, Advanced SQL & EDA

Deliverable: Pandas processing pipelines, SQL CTEs/Window functions, & Plotly dashboards

2

Month 2 Roadmap: Applied Machine Learning & Forecasting

Deliverable: Scikit-Learn models, XGBoost/CatBoost, and ARIMA/Prophet forecasting

3

Month 3 Roadmap: BI Dashboards & Milestone Capstone

Deliverable: Power BI reports, FastAPI/Streamlit web apps, & Milestone Capstone I

4

Month 4 Roadmap: MLOps & Production Infrastructure

Deliverable: MLflow tracking, Docker containerization, & FastAPI microservices

5

Month 5 Roadmap: Large-Scale Big Data & Airflow

Deliverable: PySpark distributed pipelines, BigQuery/Snowflake, & Airflow orchestration

6

Month 6 Roadmap: GenAI, LLM RAG & Final Capstone

Deliverable: Vector databases, LoRA fine-tuning, Ragas evaluation, & Final Enterprise Capstone

Course Pricing & Investment

Course Name	Original Price	Discount	Price After Discount
First 3 Months (Core ML)	₹50,000	30%	₹35,000
Next 3 Months (Advanced ML)	₹50,000	30%	₹35,000
6-Month Course (Full Course)	₹70,000	30%	₹49,000