Learnbay

Data Science & GenAI

Master Certification

Get Job Ready in 5 Months

In collaboration with

IBM & Microsoft

Swipe to know more

1 Why Every **Professional** Must Learn **GenAI & AgenticAI ?**

Data is no longer just reporting.
It drives **decisions**, **automation**, and **intelligence**.

Those who can analyze **data + apply GenAI and automation** will lead teams, products, and businesses.

WHY ACT NOW

50–170%
Higher Salaries
for Data + AI skills

9 in 10 Teams
Expect data-driven decision making

The question isn't whether AI will impact your role.
It's whether you'll use it or be **replaced by those who do.**

2 How this course prepares you for Career transition in AI ?

It helps you transition into AI by building on your existing engineering experience. Build and Deploy Real AI Applications using GenAI & AgenticAI

The role isn't disappearing. The skill bar is moving up.

Today's high-impact professionals are expected to:

Translate business problems into data solutions

Build predictive & GenAI-Agentic AI powered systems

Use AI tools to multiply productivity

Program Overview

Data Science and GenAI for Professionals

13 Months
Flexible batches

100% Live
Instructor-led

IBM & Microsoft Certified
Data Science + GenAI

No Cost EMI
Interest-free

Career Services Pro

- Resume Help
- Mock Interviews
- Job Referrals

How long will it take me to become Job Ready?

JOB READY IN

4.5 – 5.5

Months

PROGRAM DURATION

12 – 13

Months

Complete **Python, GenAI, Agentic AI & Projects** in 4.5-5.5 months to unlock job referrals. Then continue mastering advanced AI concepts for the remaining months.

Learn Until You Master AI Like a Pro

No time pressure. Your learning doesn't stop at job readiness. We support you until you become an AI expert.

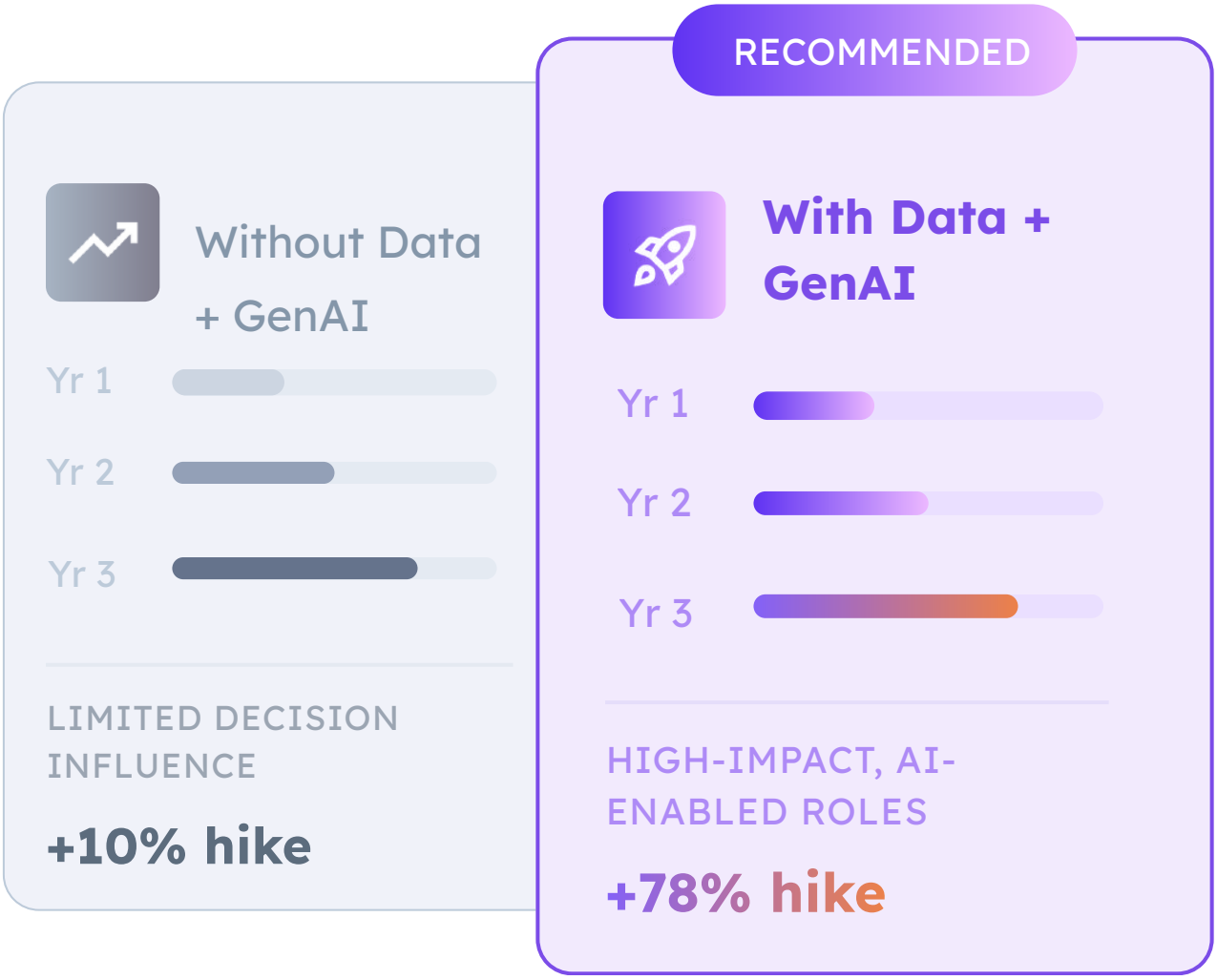
Your career shift starts at month 5, not 9.

We support you until you master AI completely.

3 How does this course accelerate your growth in the **AI-driven world ?**

Because it upgrades how you think, design and Deliver. This course **accelerates growth** by shifting you from task execution to AI-driven impact.

3-YEAR CAREER TRAJECTORY



4 How this **Learnbay Program** keeps you relevant in 2026?

We teach you Python + ML + GenAI + AgenticAI with Real industry Projects with **AI-first approach** that builds on your experience.

KEY SKILLS TO ADD

- Real Industry Projects from Industries
- Domain Specialised Track

YOUR AI-FIRST STACK ↑ Build AI

- Agentic AI + LLMOps
- GenAI Skills
- Python + ML + DL

4 Is this transition realistic for someone like you?

You're upgrading, not restarting.

Your Journey

YOU HAVE

- ✓ Domain knowledge
- ✓ Business understanding
- ✓ Problem Solving

+ JUST ADD

- + Data analysis & visualization
- + Machine learning, Deep Learning, NLP
- + Generative AI & automation

We help you add **Data + AI intelligence** to what you already know.

Alumni **Spotlight**

Mohd. Israr
Data Scientist

Thanks to the Learnbay data science course & excellent guidance, I was able to ace the TCS interview and secure a job with a 210% pay raise. The real-world time projects helped me develop my concepts as a data scientist.

Mechanical Domain → **Data Scientist @** TCS **230%** Salary Hike

Saurabh Kumar
Data Scientist

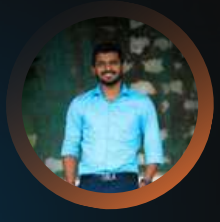
When I joined Learnbay I did not have any knowledge apart from the very basics. I gradually build my concept via various trainers and get trained in data science with strong knowledge/concepts.

Mathematics Professor → **Data Scientist @** Teleperformance **135%** Salary Hike

Aravind

TheMathCompany → CATERPILLAR

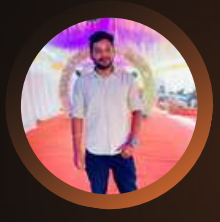
Data Scientist



Ritesh Kumar

unacademy → Capgemini

Senior Analyst

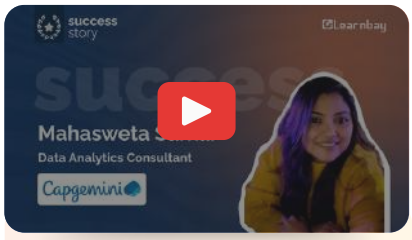


Ramki

cognizant → ANBSYSTEMS

Data Analyst

Success **Story**



Mahasweta Sarkar

Data Analytics Consultant

Capgemini

Manoj Kuna

Data Analysis Engineer

general motors

Bhavin Shah

Data Analyst

KPMG

Arvind K.

Sr. Data Scientist

TheMathCompany

Nandini Devi Muthu

Data Analyst

Rajashree B

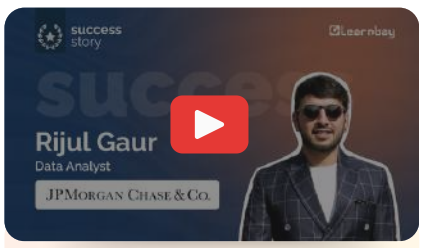
Data Quality Analyst

DATA POEM

Pinky Bhansali

Process Lead

AXA



Rahul Gaur

Data Analyst

JPMorgan Chase & Co.

Jaya Sinha

Data Scientist

amazon

Learnbay Advantage

JOB REFERRALS & ASSISTANCE

We stay with you until you're placed.

Direct Referrals
To hiring partners

Mock Interviews
With FAANG mentors

AI Resume Lab
ATS-optimized

Career Desk
3-year support

Job Referrals in Top MNCs
Direct access to hiring managers

Unlimited	5+	3 Years
Referrals	Mock Sessions	Support

Job Referral & Career Acceleration

No cold applications. Your profiles goes straight to decision makers

We refer you **directly** to hiring companies

WHAT THIS MEANS FOR YOU

Skip the ATS Black hole

Your resume reaches Humans not Algorithms

Access Exclusive Opening

Roles not posted to job boards

Matched to your skills

Job which is aligned to your profile and domain

Prepare to ace the interview

Get interview ready with expert guidance

Mock Interviews

5 Session with Industry Expert

- Multiple Technical and HR Round with detailed feedback

AI-Optimized Resume

ATS-friendly and Job ready

- Optimized to highlight your technical skills and domain expertise

Complete Support until Placed

We’re with you in every step

5 What sets this program apart?

3-Year Flexible Subscription

Re-enter, upgrade, and refresh skills as data & AI evolve.

AI CoLab Experience

Work on real startup & enterprise-grade data + GenAI projects.

Earn 2 AI CoLab Certificates.

BYOP+Mentorship

Bring your own domain or business problem. Get 1:1 expert guidance to turn it into a data solution.

Industry Certification

Earn globally recognised credentials backed by IBM and Microsoft.

6 How do you transition as Expert, not Fresher?

Domain Specialization

Build depth in solving hard engineering problems.

+

AI Colab Expereince

Show impact through real projects and ownership

=

Expert

With proven portfolio

not

~~Fresher~~

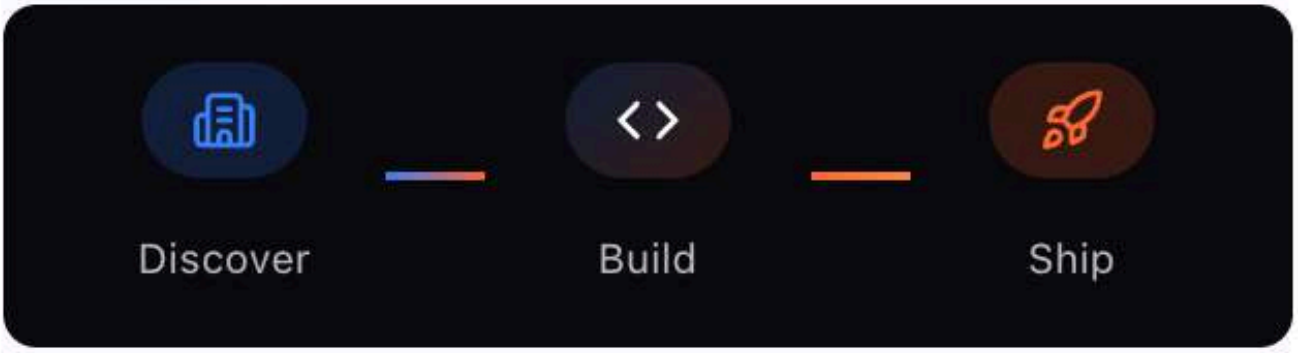
Starting over

From Domain Expert → to AI Expert

Work on Real Projects
Directly from Startups

Project certificate beats course completion certificate.

- Signed by the startup
- Code reviewed & verified
- Ready for your portfolio



"Course certificates say you watched.
Project proofs say you can build."

Real work. Real proof. Real jobs.

7 What kind of Projects will you build?

Real-World Data Science & GenAI Projects

Not classroom exercises.
Real industry-grade AI systems.

Data Science

Business Intelligence & Prediction Engine

End-to-end data science system to analyze trends, predict outcomes, and drive business decisions.

Python · Pandas · NumPy · SQL · Scikit-learn · Power BI / Tableau

Applied ML + GenAI

AI-Powered Forecasting & Decision System

Combine predictive modeling with GenAI explanations for business-ready forecasting dashboards.

Time Series Models · Machine Learning · LLMs · Visualization Tools

Learn From Basics Till AI model Deployment

LLMOps

Production-Grade LLM Deployment

Deploy, monitor, and scale GenAI applications with enterprise-ready controls.

AI Agents

Autonomous Incident Triage & RCA Agent

AI agent that analyzes logs, traces root cause, and recommends fixes during live incidents.

SKILLS YOU'LL MASTER

- Model versioning & controlled rollouts
- Production monitoring & alerting
- Prompt lifecycle management at scale
- Multi-agent orchestration workflows
- Cost optimization for LLM usage
- Safety guardrails, fallbacks & retries

Get Certified

and accelerate your career growth

IBM

Course Completion

Generative AI Certification

Earn an industry-recognized IBM GenAI Certification validating your expertise in enterprise-grade AI systems.

Microsoft Certifications

AI Co-Lab Project Certificate

This certification boosts your credibility in the IT sector and enhances your career prospects.

Program Fees

Data Science & GenAI Master Program for Tech Professionals

MASTER TRACK

Live online classes

- Instructor-led **live interactive** session
- **Dual** Certification Program
- **Mock technical interviews** for AI roles
- Industry-grade **capstone** projects

PROGRAM FEE

₹ 1,59,000/- +18% GST

Payment Plan: No Cost EMI Available

For 6, 9 and 12 Months

13 Months Program

Program Roadmap

Data Science and Generative AI Master Certification Program for Professionals

Program Roadmap

FOR DATA SCIENCE AND GENAI AI MASTER PROGRAM

DAY 0

DURATION: 3 DAYS

INDUCTION + ORIENTATION

Understand the GenAI & Agentic AI landscape, tools, expectations, and how this program aligns with your career goals.

MODULE 1

DURATION: 2.5 Months

FOUNDATION OF GENAI

Build Python, ML, DL, and NLP foundations for modern GenAI workflows.

MODULE 2

DURATION: 1 Month

ADVANCED GENERATIVE AI

Design and optimize LLM-based applications and autonomous AI agents using prompts, embeddings, RAG, and fine-tuning.

MODULE 3

DURATION: 1 Month

AGENTIC AI & AUTOMATION

Design autonomous AI agents that plan, reason, and automate complex workflows end to end.

MODULE 4

DURATION: 1 Month

STATISTICS & FOUNDATION OF DATA SCIENCE

Statistical concepts and analytical foundations essential for data science problem-solving.

MODULE 5

DURATION: 2.5 Month

ADVANCED DATA SCIENCE

Advanced data science methods including ML, time series, NLP, CV, and reinforcement learning.

AI Colab

DURATION: 8-10 Days

Work on AI Startup Projects

Work on real AI startup projects to build deployable systems and earn industry-recognized certifications.

MODULE 6

DURATION: 1 Months

DEPLOYMENT & MLOps

Foundational to advanced deployment and MLOps practices for production machine learning systems.

MODULE 7

DURATION: 1.5 Months

DATA ANALYTICS & VISUALIZATION- TOOLS

Platforms that simplify data analysis, dashboards, and visual exploration for better business insights.

MODULE 8

DURATION: 1 Month

DATA MANAGEMENT & SQL

Core techniques for storing, managing, and querying structured data using SQL and modern database systems.

MODULE 9

DURATION: 1 Months

BIG DATA ANALYTICS

Techniques and tools to process, manage, and analyze large-scale datasets across distributed computing systems.

MODULE 10

DURATION: 1 Months

DSA (Optional)

Foundational data structures and algorithms for efficient data processing and analytical tasks.

MODULE 11

DURATION: 15 Days

PROJECT MANAGEMENT (Skills and Tools)

Essential Agile project management skills and tools for planning, tracking, and delivering technical projects effectively.

Domain Specialization

You don't start over.

You **leverage** your domain, add **Data + GenAI skills**, and move into high-impact AI roles.

Choose any **2** domain electives
Build real **projects**. **Transition** with confidence.

Elective A

4 Project

Data Science & GenAI for BFSI

Design AI systems for risk, fraud, and financial intelligence—not generic models.

Move into Data Scientist / AI Analyst / Risk Analytics roles

Elective B

2 Project

Data Science & GenAI for Retail

Turn customer and sales data into revenue-driving intelligence.

Transition into Retail Data Scientist / GenAI Product Analyst roles

Choose any **2** domain electives
Build real **projects**. **Transition** with confidence.

Elective C

3 Project

Data Science & GenAI for Manufacturing

Apply AI where it matters most—operations, quality, and efficiency.

Move into Industrial Data Scientist / AI Engineer roles

Elective D

3 Project

Data Science & GenAI for Healthcare

Build AI solutions that support clinical, operational, and patient intelligence.

Transition into Healthcare Data Scientist / AI Analyst roles

Elective E

4 Project

Data Science & GenAI for SDE

Move beyond CRUD and APIs—build AI-first systems.

Advance into AI Engineer / Applied ML Engineer roles

Choose any **2** domain electives
Build real **projects**. **Transition** with confidence.

Elective F

3 Project

Data Science & GenAI for Managers

Lead AI initiatives—don't just approve them. Learn GenAI Use Cases for Business Automation

Step into AI Product, Analytics Leadership, or Strategy roles

Elective G

3 Project

Data Science & GenAI for Supply Chain

Turn supply chain complexity into AI-powered clarity. Master GenAI use cases for forecasting and flow optimization.

Move into Industrial Data Scientist / AI Engineer roles

Ritesh Kumar

From **Consultant** → To **Senior Data Analyst**

Placed at Capgemini

110% Salary Hike

13 Months Program

Detailed Program Curriculum

Data Science and Generative AI Master Certification Program

170+ Hours of Industrial Projects

32+ Hours of Capstone Projects

Unlimited Doubt Clearing Sessions

Induction + Orientation

Duration: 3 Days

DAY 0

Duration: 3 Days

- Overview Of AI, Generative AI, And Agentic AI.
- Career Paths And Role Mapping For AI Professionals.
- LinkedIn Profile Optimization For AI & Tech Hiring.
- Using **Python Colab** For Hands-On Learning.
- Learning Workflow, Tools, And Program Onboarding.

Foundation of GenAI

Duration: 2.5 Months

Module 1

Topic 1: **Core Python**

- **Variables**
  - Store values in various data types
  - Dynamic assign / reassign variables
  - Python naming conventions for maintainable code
- **Data types**
  - Numeric types: **int, float, complex.**
  - Text handling using str and string methods.
  - Collections: **list, tuple, set, dict.**
  - Mutable vs immutable data behavior.
- **Loops**
  - Iteration using for and while loops.
  - Sequence traversal with **range(), enumerate(), zip().**
  - Optimizing iterations for data processing tasks.

Topic 1: **Core Python**

- **Control Statements**
  - Conditional logic using if, elif, else.
  - **Comparison operators** (==, !=, <, >, <=, >=).
  - Logical operators (and, or, not).
  - Flow control using break, continue.
  - Building **decision trees** for real-world logic handling.

- **Functions**
  - Creating reusable logic blocks using def and return.
  - Passing parameters using positional and keyword arguments.
  - Anonymous function with **Lambda**

Project 1 **Simple Calculator**

1 hour

Create a console-based calculator that takes user input for two numbers and performs mathematical operations

Topic 2: **Advanced Python**

- **File Handling**
  - Reading and writing files using open(), read(), write().
  - Handling text, **CSV**, and **JSON** files for data pipelines.
  - Working with file paths using os and pathlib.

- **Regex**
  - Pattern matching using **re.search(), re.findall(), re.sub()**.
  - Cleaning and extracting text data from logs, emails, and documents.
  - Using regex tokens: \d, \w, +, *, ^, $, groups.

- **Exceptional Handling**
  - Custom exceptions handling using try, except, finally.
  - Common exceptions.
  - Preventing pipeline failures in production code.

Topic 2: **Advanced Python**

- **OOP's**
  - Creating classes and objects using class and __init__.
  - Instance variables vs class variables.
  - Methods and object behavior modeling.
  - Inheritance and method overriding.

Project 2 **Banking System Simulation**

BFSI

Simulate a simple banking system where users can create accounts, deposit, withdraw, and check balances.

Python | Data Structure | Function

Project 3 **Movie Night Recommender**

Entertainment

Recommend movies based on genre preferences using simple rule-based logic and conditional checks.

Data Types | Filtering Logic | User Input Handling

Project 4 **Smart Fitness Habit Tracker**

Healthcare

Track daily exercises, calories burned, and generate habit summaries using basic loops and functions.

Python | Functions | Conditional logic

Topic 3: **Essential Libraries for GenAI**

- **Numpy**
  - Creating arrays **using array, arange, linspace**.
  - **Array operations:** indexing, slicing, reshaping.
  - Vectorized computations for performance.
  - **Mathematical functions:** mean, sum, std, dot.
  - Broadcasting and handling multi-dimensional data.

- **Pandas**
  - Working with Series and DataFrame.
  - Reading data from **CSV, Excel, JSON**.
  - Data cleaning: missing values, duplicates, type casting.
  - Filtering, sorting, and conditional selections.
  - **Grouping** and **aggregation** using **groupby**.
  - Feature preparation for ML and GenAI pipelines.

- **Matplotlib**
  - Creating line, bar, scatter, histogram plots etc.
  - Customizing labels, titles, legends, and axes.
  - **Visualizing** trends and distributions in datasets.

- **Seaborn**
  - **Statistical plots:** boxplot, violinplot, countplot.
  - Relationship analysis using **pairplot** and **heatmap**.
  - Visualizing correlations and feature importance.
  - Styling plots for reports and dashboards.

Project 5 **Retail Sales Data EDA**

Retail

Analyze a retail sales dataset: clean data, plot trends (time series, bar charts), and derive insights.

Python Pandas Geopandas Matplotlib

Topic 4: **ML Fundamentals**

- **Basic of ML**
  - **Supervised** vs **Unsupervised** learning paradigms.
  - Features, labels, training, validation, and test sets.
  - Bias–variance trade-off and model generalization.
  - **Overfitting** and **underfitting** with real examples.
  - End-to-end ML workflow used in industry projects.

- **Regression Analysis**
  - **Multiple Regression** with feature interactions.
  - Model evaluation using **RMSE, MAE, R^2 score.**

Project 6 **Energy Consumption Forecasting**

Energy

Predict household energy consumption using linear regression by analyzing temperature, appliances, and daily usage patterns.

Feature engineering

RMSE/R^2 evaluation

- **Classification Analysis**
  - **Logistic Regression** for binary classification.
  - Handling class imbalance using sampling techniques.
  - **Evaluation metrics: Precision**, **Recall**, **F1-Score**, **ROC-AUC**.

Project 7 **Loan Approval Prediction**

Finance

Build a classification model (e.g. logistic regression, random forest) to predict loan approvals from applicant data.

scikit-learn

Pandas

Topic 5: **Deep Learning Fundamentals**

- **Perceptron**
- **Backpropagation and Optimization**
- **Attention Mechanism**

Project 8 **Smart Document Relevance Classifier**

HR

Compute attention weights on simple sentences to understand token importance during context interpretation.

Numpy Perceptron

Topic 6: **NLP Fundamentals**

- **Text Pre-Processing for Gen AI**
- **Text Embedding Techniques**
  - Bag-of-Words representation.
  - TF-IDF vectorization.
  - **Word2Vec: CBOW** and **Skip-Gram**.
- **Named Entity Recognition (NER)**
- **Topic Modeling**
  - **Latent Dirichlet Allocation (LDA).**
  - **Latent Semantic Techniques (LSA/LST).**
- **Part-of-Speech Tagging (POS)**
- **NLP Evaluation Metrics**

- **Transformer - BERT Model**
  - **Transformer architecture:** Encoder–Decoder blocks.
  - Multi-head self-attention mechanism.
  - Positional encoding concepts.
  - BERT pre-training objectives: **MLM** and **NSP**.
  - Use cases of **BERT** in real NLP systems.

Project 9 **Sentiment Analysis Using BERT**

Marketing

Use pretrained BERT embeddings to convert sentences into vectors and classify sentiment using a simple classifier.

BERT | Manual tokenization | Dense classifier

Project 10 **Legal Entity Extraction**

Legal

Perform named entity recognition and classification in legal documents using spaCy, and visualize results

spaCy | Pandas | Matplotlib

Advanced Generative AI

Duration: 1 Month

Module 2

Topic 1: **Fundamentals of GenAI**

- Difference Between Predictive AI And Generative AI.
- Generative Modeling Concepts And Probability Distributions.
- Pre-training vs Fine-tuning Paradigms.
- Foundation Models And Their Role In Modern AI Systems.

Topic 2: **Use cases**

- **Text Generation** For Chatbots And Assistants.
- Document Summarization And Report Automation.
- **Code Generation** And Developer Productivity Tools.
- Content Generation For Marketing And Media.
- Knowledge Assistants For Enterprises.
- **Search Augmentation** And **Question Answering Systems**.

Topic 3: **Prompt Engineering**

- **Zero-shot**, **One-shot**, And **Few-shot** Prompting.
- Instruction-based Prompt Design.
- Chain-of-Thought Prompting Techniques.
- Prompt Templates And Reusability.
- **Prompt Evaluation** And **Iterative Optimization**.

Topic 4: **LLM**

- **GPT Family** (GPT-3.5, GPT-4, GPT-4o).
- **LLaMA**, **Claude** and **Open-Source** LLM Ecosystem.
- **Tokenization** and **Context Windows**.

Topic 4: **LLM**

- **Inference Parameters: Temperature**, **Top-p**, **Max Tokens**.
- Hallucination And Limitations Of LLMs.
- Cost And Latency Considerations In Production.

Project 11 **AI-Powered Sales Email Writer**

Sales

Use an LLM to generate personalized sales outreach emails using customer behavior and CRM attributes.

OpenAI | Prompt Templates | Langchain

Topic 5: **Hugging Face**

- **Transformers Library** For NLP And GenAI.
- Pre-trained Models Hub And Model Cards.
- **Tokenizers** And **Pipeline** APIs.
- Fine-tuning Models Using **Hugging Face Trainer**.

Project 12 **Product Review Sentiment Classifier**

Retail

Build an NLP model to classify customer product reviews as positive or negative (using word embeddings and an LSTM or transformer).

NLTK/spaCy | Hugging Face | NLP Libraries

Topic 6: **RAG (Retrieval Augmented Generation)**

- **Vector DB**
  - Embedding Generation Using OpenAI And Hugging Face Models.
  - Vector Stores: **FAISS**, **Pinecone**, **ChromaDB**, **Weaviate**, **Qdrant**
  - Similarity Search And Distance Metrics.
  - **Chunking Strategies** For Optimal Retrieval.
- **Multimodal RAG**
  - Text + Image Retrieval Pipelines.
  - Multimodal Embeddings (**CLIP**, **Vision Transformers**).
  - Use Cases In Documents And Media Search.
- **Graph RAG**
  - Knowledge Graph Construction From Documents.
  - Relationship-based Retrieval **Over Vector Search**.
- **RAG Evaluation Metrics**
  - Context Precision And Recall.
  - Answer Faithfulness And Groundedness.
  - **Latency** And **Cost Evaluation**.
  - Human-in-the-loop Evaluation Methods.

Project 13 **Custom Chatbot with Retrieval**

Tech Support

Build a domain-specific chatbot: fine-tune an LLM on company FAQs and implement retrieval-augmented generation (Chroma or Elasticsearch) to answer user queries.

OpenAI API | Langchain | Retrieval DB

Project 14 **HR Policy RAG Query Bot**

HR

Create a RAG-based chatbot that answers employee HR queries using internal company policies and guidelines.

Vector Store | Langchain | Embedding Models

Topic 7: **Langchain / Llama Index**

- **Prompt Templates** And **Output Parsers**.
- Chains And Sequential Workflows.
- Agents And Tool Calling.
- Integration With Vector Databases And APIs.
- Embed and store documents in vector DBs (**Pinecone**, **FAISS**, **Chroma**, etc.)
- Retrieve relevant chunks based on a user query
- Format the retrieved chunks as context for the LLM

Topic 8: **Fine tuning**

- **Full Fine-Tuning** vs **Parameter-Efficient Fine-Tuning (PEFT)**.
- **LoRA** And **QLoRA** Techniques.
- Instruction Fine-Tuning.
- Dataset Preparation And Labeling Remotely.
- **Overfitting Risks** And **Mitigation Strategies**.

Topic 9: **Fundamental Generative Models: GAN And VAE**

- **Generative Adversarial** Networks Architecture.
- Variational Autoencoders And Latent Space Learning.
- Use Cases In **Image And Data Generation**.

Project 15 **Synthetic Defect Image Generation** Manufacturing

Generate synthetic defect images to augment limited manufacturing datasets for improved model training performance.

GAN Pytorch TensorFlow

Topic 10: **Stable Defusion**

- **Diffusion** Model Fundamentals.
- **Text-to-Image** Generation Pipelines.
- Prompt Control And **Image Conditioning.**
- Fine-Tuning Diffusion Models (**DreamBooth, LoRA**).

Project 16 **Stable Diffusion Image Generator**

Marketing

Use a Stable Diffusion model to generate creative images from text prompts; experiment with fine-tuning or style transfer.

CUDA GPU

Diffusers library

Topic 11: **Flux Framework**

- Image And Creative **Generation Workflows.**
- High-quality **Diffusion-based Outputs.**

Topic 12: **LLMOp's**

- **Model Versioning** And Prompt Version Control.
- Monitoring **LLM** Outputs In Production.
- **Cost Tracking** And Optimization.
- **Evaluation Pipelines** For LLM Applications.
- Deployment Strategies For **GenAI Systems.**

Project 17 **Chatbot Deployment**

Tech Support

Build an LLMOps workflow to deploy, version, and monitor a chatbot using automated evaluation and logging.

FastAPI

Docker

LangSmith

Project 18 **RAG Q&A Pipeline**

Knowledge Management

Build a Retrieval-Augmented Generation (RAG) pipeline using a vector database and LLM for Q&A.

Pinecone | HuggingFace | LangChain

Topic 13: **Azure OpenAI**

- **Azure OpenAI** Service Architecture.
- Model Deployment And Endpoint Management.
- **Enterprise Security** And Compliance.
- Integrating Azure OpenAI With Applications.

Project 19 **Enterprise Document Assistant**

Enterprise IT

Build a secure document Q&A system using Azure OpenAI embeddings and enterprise-grade retrieval pipelines.

Azure OpenAI | Python SDK | LangChain

Project 20 **SCM Compliance RAG Assistant**

Supplychain

Build a RAG system that answers procurement and logistics compliance queries using supply chain documents.

OpenAI | FAISS/Chroma | HF Embeddings

Agentic AI & Automation

Duration: 1 Months

Module 3

Topic 14: **Understanding Agentic AI and Autonomous Agents**

- Difference Between **LLM Applications** And Agent-Based Systems.
- Autonomous Decision-Making And Goal-Oriented Agents.
- **Single-Agent vs Multi-Agent** Architectures.

Topic 15: **Key components of Agentic AI**

- Agent Roles, Goals, And Task Decomposition.
- Planning, Reasoning, And Execution Loops.
- **Tool Calling** And External System Integration.
- **Memory Types:** Short-Term, Long-Term, And Episodic Memory.
- Feedback Loops And Self-Reflection Mechanisms.
- Failure Handling And **Recovery Strategies.**

Topic 16: **AutoGen**

- **Multi-Agent** Conversation Design.
- Role-Based Agent Collaboration.
- Task Delegation And Message Passing.
- **Use Cases** For Autonomous Workflows.

Project 21 **AI Data Analyst Agent**

Finance

Develop an autonomous agent (e.g. using LangChain or LangGraph) that connects to a database and answers data questions by generating SQL queries and charts.

LLM (GPT-4/Gemini)

Mistral Agents

Topic 17: **CrewAI**

- Team-Based **Agent Orchestration.**
- Role Assignment And Hierarchical Agents.
- **Workflow Automation** Using Crews.

Project 22 **Quality Inspector Agent**

Manufacturing

Build agent team to analyze defect images, generate insights, and suggest corrective actions collaboratively.

Vision models | OpenCV tools | CrewAI

Topic 18: **LangGraph**

- **State-Based Agent** Workflows.
- Directed Graph Execution For Agents.
- Conditional Routing And **Branching Logic.**
- Multi-Step Planning And Execution Graphs.
- **Error Handling** And Retry Mechanisms.

Project 23 **RAG Product Query Bot**

E-Commerce

Build a RAG assistant that answers product comparisons and queries using catalog data and customer reviews.

FAISS / Chroma | OpeAI LLM

Topic 19: **MCP (Model Context Protocol) / ACP / A2A**

- Model Context Protocol (MCP) and Agent Communication Protocols (ACP)
- Agent-to-Agent (A2A) Interaction Patterns and Coordination
- Context Sharing and Secure Exchange Across Distributed Agents
- Enterprise Interoperability and Standards for Agent Systems

Project 24 **MCP Coding Assistant & Debugger**

IT

Build MCP-powered coding assistant that reads code files, runs tests, debugs errors, and suggests fixes automatically.

MCP SDK | OpenAI LLM | AutoGen

Topic 20: **No-Code/Low-Code AI Agent Basics**

- Build AI agents using **drag-and-drop workflow** builders.
- **Automate** tasks, approvals, and decision flows without coding.
- Evaluate feasibility, cost, and business **ROI** of AI agents.
- Work confidently with tech teams to scale AI faster.

Topic 21: **Ethics And Safety In Agentic AI**

- Autonomous Agent Risk Assessment.
- Bias, **Hallucination**, And Decision Transparency.
- Human-in-the-Loop Control Mechanisms.
- Compliance And Responsible AI Guidelines.
- **GuardRail**
  - Input/Output Validation and Policy Enforcement
  - Monitoring, Auditing, and Control of Agent Actions in Production

Project 25 **LLM Toxicity Monitoring & Safety Guardrail Engine**

Responsible AI

Create a safety pipeline to detect toxic content, enforce prompt rules, and generate safer LLM responses.

OpenAI | Perspective API | Rule-based guardrails

Topic 22: **AWS Bedrock**

- Foundation Models Available In Bedrock.
- Building Agents Using Bedrock APIs.
- Tool Invocation And Workflow Automation.
- **Security, IAM, And Enterprise Controls.**
- Integrating Bedrock With AWS Services.

Project 26 **Autogen Vehicle Diagnostics Agent**

Automotive

Create diagnostic agent that interprets sensor logs and suggests repair actions using reasoning and tool execution.

AutoGen | Bedrock Claude

Statistics & Foundation of DS

Duration: 1 Month

Module 4

Topic 1: **Introduction to Stats & Statistical Thinking**

- Types of variables: quantitative, and qualitative
- Role of descriptive vs inferential statistics.
- Understanding data distribution and variability.
- Probability basics including experiments, outcomes, and events.

Topic 2: **Probability Concepts**

- Probability distribution function (PDF) and cumulative distribution function (CDF).
- Conditional probability and real-life applications.
- Bayes' Theorem and its use in classification problems.
- Mutually exclusive, joint, dependent, and independent events.
- Problem-solving exercises to strengthen probability intuition.

Topic 3: **Population & Sampling Fundamentals**

- Difference between population and sample in real-world studies.
- Importance of sample size and its effect on confidence.
- Sampling methods: simple random, systematic, cluster, stratified.
- Non-probability sampling: convenience, quota, snowball, judgment sampling.

Topic 4: **Statistics**

- Measures of central tendency: mean, median, mode.
- Measures of dispersion: variance, standard deviation, range, IQR.
- Shape analysis: skewness and kurtosis interpretation.

- ANOVA / ANCOVA for comparing model performance.
- Central Limit Theorem and sampling distribution behavior.
- Point estimates vs interval estimates.
- Characteristics of Z-distribution and T-distribution.
- Type I and Type II errors with practical meaning.

Topic 5: **Hypothesis Testing**

- Formulating null and alternative hypotheses.
- Z-test, t-test, chi-square test basics.
- Using the empirical rule and Z-table.
- Normal and binomial distributions with applications.

Project 27 **Survey Correlation Study**

Consumer Research

Analyze customer survey data for correlation patterns and test statistical significance of findings.

Numpy | Statsmodels | Pandas

Project 28 **KPI Statistical Analysis**

Operations

Compute confidence intervals and perform hypothesis tests for key business KPIs (e.g., conversion rates).

Scipy | Pandas | Numpy

Topic 6: **Exploratory Data Analysis (EDA)**

- Five-point summary: min, Q1, median, Q3, max.
- Box plot construction and interpretation.
- Outliers: causes, impact, and detection.
- Outlier treatment using IQR and Z-score methods.

Project 29 **Sales Dashboard EDA**

Retail

Perform exploratory analysis on retail data, derive insights, and build an interactive dashboard with charts.

Pandas | Seaborn | Plotly

Project 30 **Patient Data Insights**

Healthcare

Analyze patient health records to discover insights, trends, and potential anomalies with visualizations and charts.

Pandas | Seaborn | Matplotlib

Topic 7: **Regression Analysis - Supervised Learning**

- Correlation & regression assumptions
- Linear regression mechanics (cost, optimization, gradient descent)
- Model training workflow (train–test split, prediction)
- Multiple regression & feature interactions
- Residual analysis & diagnostics
- Evaluation, generalization & regularization (RMSE, MAE, R^2, L1/L2)

Project 31 **Housing Price Predictor**

Real Estate

Build an end-to-end pipeline for housing price prediction with automated feature engineering and hyperparameter tuning.

Scikit-learn | XGBoost | Pandas

Topic 8: **Classification Analysis - Supervised Learning**

- Logistic Regression, sigmoid function & MLE
- Linear vs logistic comparison
- Class imbalance handling (sampling techniques)
- Decision Trees & Random Forests
- KNN: distance metrics, K-value selection & curse of dimensionality
- Naive Bayes: assumptions & likelihood
- Evaluation metrics: Precision, Recall, F1, ROC-AUC & cross-validation
- **Project:** Classification problem with real business data.

Project 32 **Predictive Maintenance Classifier**

Manufacturing

Predict equipment failures using classification on sensor data with feature engineering and model evaluation.

Scikit-learn | Featuretools | Pandas

Topic 9: **Unsupervised Learning Introduction - Clustering**

- K-means clustering: centroid optimization and within-cluster sum of squares
- Cluster validation: elbow method, silhouette analysis, and optimal selection
- Distance metrics: Euclidean, Manhattan, and domain-specific measures
- Feature scaling and normalization for distance-based clustering
- Selecting optimal number of clusters using practical heuristics
- K-means initialization methods and convergence behavior
- Managing outliers and high-dimensional clustering challenges

Advanced Machine Learning

Duration: 2.5 Months

Module 5

Topic 1: **Advanced Clustering Techniques**

- Hierarchical approaches: agglomerative and divisive clustering strategies
- Linkage criteria: single, complete, and average linkage mathematical foundations
- Density-based clustering: DBSCAN algorithm and noise point handling
- Clustering comparison: K-means vs hierarchical vs DBSCAN
- Clustering evaluation: internal and external validation measures

Project 33 **Machine Sensor Behavior Clustering**

Manufacturing

Cluster machine sensor sequences to detect unusual operating behaviors indicating early equipment issues.

DBSCAN | Feature Scaling | Time-based Clustering

Project 34 **Credit Usage Behavior Clustering**

BFSI

Segment credit card users by spending behavior to support risk analysis and personalized offers.

Scikit-learn | K-Means | PCA

Topic 2: **Ensemble Learning Techniques (Bagging & Boosting)**

- Bootstrap aggregation (bagging): concept, implementation
- Variance reduction through sampling diversity
- Random Forest: out-of-bag estimation and feature importance calculation
- Boosting foundations: sequential learning and error correction mechanisms
- AdaBoost, gradient boosting
- Gradient boosting: loss function optimization and residual learning
- Advanced boosting: XGBoost, CatBoost, and LightGBM implementation strategies
- Bagging vs boosting: bias-variance implications

Topic 3: **Support Vector Machine (SVMs)**

- Linear SVM: maximum margin principle and support vector identification
- Soft margin formulation: slack variables and regularization parameter tuning
- Kernel methods: polynomial, radial basis function, and kernel trick mathematics
- Multi-class extensions: one-vs-one and one-vs-rest strategies

Topic 4: **Dimensionality Reduction Methods**

- Principal Component Analysis: eigenvalue decomposition and variance explanation
- Component interpretation: loading analysis and dimensionality selection
- Feature selection comparison: filtering, wrapper, and embedded approaches

Project 35 **Customer Segmentation**

E-commerce

Perform customer segmentation on purchase data using clustering and visualize segment profiles with PCA plots.

Scikit-learn | Pandas | Matplotlib

Topic 5: **Model Optimization Strategies**

- Hyperparameter tuning: GridSearchCV, RandomSearchCV, and Bayesian approaches
- K-fold cross-validation for parameter selection: nested validation and unbiased estimation
- Performance curve analysis: learning curves and validation curves
- ROC curve optimization, threshold tuning
- Handling imbalanced datasets: SMOTE, resampling
- Model interpretation: feature importance, partial dependence, and SHAP values

Project 36 **Loan Default Predictor**

Risk Analytics

Use ensemble models to predict loan default risk using borrower history and financial attributes.

LightGBM | Grid Search | Feature Importance

Topic 6: **Time Series Analysis**

- Time series components: trend, seasonality, noise
- Stationarity tests: ADF, KPSS
- Autocorrelation & partial ACF plots
- Seasonal decomposition (additive, multiplicative)

Topic 7: **Classical Forecasting Methods**

- TMoving averages & exponential smoothing
- ARIMA modeling: identification, estimation, diagnostics
- SARIMA for seasonal data
- Model comparison using AIC/BIC & backtesting

Topic 8: **Advanced Forecasting**

- Prophet framework for business series
- Feature-based regression models (lags, rolling stats)
- Tree-based regressors (Random Forest, XGBoost) on time features
- Time series cross-validation & walk-forward validation

Project 37 **Energy Anomaly Detection**

Energy

Detect anomalies in energy consumption time series using an LSTM autoencoder and visualize findings.

TensorFlow Pandas Matplotlib

Natural Language Processing

Topic 1: Text Preprocessing

- Tokenization, stopword removal, stemming, and lemmatization.
- Lowercasing, punctuation removal, and text normalization.
- Converting raw text into clean, usable form.

Topic 2: Word Embeddings

- Represent words as dense vectors capturing meaning.
- Word2Vec for context-based learning.
- GloVe for global co-occurrence patterns.
- FastText for handling subword information.

Topic 3: Transformers & BERT

- Attention-based architecture for understanding long text.
- BERT for bidirectional contextual embeddings.
- Fine-tuning BERT for NLP tasks like QA and classification.

Topic 4: Sequence-to-Sequence Models

- Encoder-decoder setup for translating sequences.
- Used in translation, summarization, and dialogue.
- Attention improves sequence alignment.

Topic 5: Text Classification

- Assign labels to text (spam, sentiment, categories).
- Use embeddings + neural networks or transformer models.
- Evaluate using accuracy, precision, recall, F1-score.

Topic 6: **Topic Modeling**

- Discover hidden topics in text collections.
- LDA and embedding-based topic extraction.
- Group documents by common themes.

Topic 7: **NLP Evaluation Metrics**

- Accuracy and confusion matrix.
- Precision, Recall, and F1-Score.
- BLEU score for text generation.
- Perplexity for language models.

Project 38 **Transformer Sentiment Pipeline**

Marketing

Build a sentiment analysis pipeline using transformer models (e.g., BERT) on social media data.

Transformers | Pandas | scikit-learn

Project 39 **QA Chatbot with RAG**

Customer Support

Implement a question-answering chatbot using a pre-trained LLM and retrieval-augmented generation (RAG) with a custom knowledge base.

HuggingFace | Pinecone | LangChain

Advanced Deep Learning

Topic 1: Computer Vision

- Image Fundamentals
- Image representation using pixels, channels, and encoding.
- Histogram analysis to study brightness and contrast.
- Image transformations: resize, rotate, flip, crop.
- Image filtering: Gaussian, median, bilateral for smoothing and noise removal.

Topic 2: Convolutional Neural Networks (CNNs)

- Convolution operation, padding, and stride basics.
- Pooling layers for reducing spatial size.
- CNN architectures

Topic 3: Image Classification

- Softmax classifier for predicting class probabilities.
- Transfer learning using pre-trained models.
- Data augmentation to improve generalization.

Topic 4: Object Detection & Localization

- Traditional methods: HOG + SVM.
- Deep learning-based detectors:
- R-CNN , Fast R-CNN, Faster R-CNN
- YOLO (v3–v8), SSD, RetinaNet

Topic 5: **Image Segmentation**

- Semantic segmentation – classify every pixel.
- FCN – fully convolutional networks.
- U-Net – encoder–decoder for medical images.
- DeepLab – atrous convolutions for better context.
- Instance segmentation – detect + segment each object.
- Mask R-CNN – adds mask prediction to Faster R-CNN.

Topic 6: **Computer Vision with Transformers**

- Vision Transformers (ViT) – split image into patches, use attention.
- DETR – transformer-based end-to-end object detection.
- Swin Transformer – hierarchical vision transformer.

Project 40 **Defect Detection Classifier**

Manufacturing

Classify images of manufacturing defects using transfer learning (e.g., ResNet) and compare model accuracy.

TensorFlow | Opencv | Matplotlib

Project 41 **Medical Image Segmentation**

Healthcare

Perform semantic segmentation on medical images (MRI) using U-Net with data augmentation and compute IoU metrics.

TensorFlow | Matplotlib | Albumentations

Reinforcement Learning

Topic 1: **RL Fundamentals & Core Method**

- Agent–environment interaction, rewards, actions, policies.
- Markov Decision Processes (MDP) and Bellman equations essentials.
- Exploration–exploitation strategies (ε-greedy, softmax).
- Monte Carlo, TD Learning, SARSA, Q-Learning basics.
- Policy Gradients and Actor–Critic (REINFORCE, A2C/A3C).
- Classic use cases: game-playing AI and simulation environments.

Topic 2: **Deep RL & Advanced Algorithms**

- Convolution operation, padding, and stride basics.
- Deep Q-Networks (DQN) with neural networks + experience replay.
- Advanced Actor–Critic methods: PPO, DDPG, TD3, SAC.
- Continuous control with policy gradient deep models.
- Multi-Agent RL: cooperative, competitive, and mixed settings.
- Transfer learning, fine-tuning, and lifelong RL concepts.
- Applications: robotics, autonomous driving, finance, recommendation systems.

Project 42 **Content Personalization Agent**

Product

Train an RL agent to choose best-performing product content using user interaction signals as rewards.

Python | RL Environments | Deep Q-Learning

Deployment & MLOps

Duration: 1 Month

Module 6

Topic 1: **Model Serving & Packaging**

- FastAPI / Flask Serving: Build lightweight APIs to expose ML models.
- Model Packaging (BentoML / MLflow Models): Standardize models for easy deployment.
- Dockerization: Wrap the model + API into containers for consistent execution.
- Local Testing: Validate predictions, latency, and real-world behavior before cloud use.

Topic 2: **Cloud Deployment Essentials**

- AWS & Azure Basics: S3/Blob storage, compute instances, IAM essentials.
- Deployment Platforms: SageMaker / Azure ML for quick hosted endpoints.
- CI/CD for Deployment: Automate build → test → deploy workflows.
- Basic Monitoring: Track API uptime, latency, and failure rates.

Project 43 **Cloud Model Deployment**

Cloud Services

Deploy a model on AWS/GCP: use SageMaker endpoint or Google AI Platform with serverless API.

Docker | AWS SageMaker | Terraform

MLOps

Topic 1: **Reproducible ML Workflow**

- Data & Model Versioning: Manage dataset changes using DVC; track models via MLflow Registry.
- Experiment Tracking: Log metrics, hyperparameters, artifacts using MLflow or Weights & Biases.
- Pipeline Automation: Build training & inference pipelines with Airflow, ZenML, or Kedro.

Topic 2: **Continuous Training, Deployment & Monitoring**

- CT Pipelines: Auto-trigger re-training when new data arrives or drift is detected.
- Automated Deployment: CI/CD pipelines using GitHub Actions, Jenkins, or GitLab CI.
- Model Monitoring: Detect drift, data issues, and performance degradation (Evidently, WhyLabs).
- Alerts & Rollbacks: Automatically promote or revert model versions using a registry.

Project 44 **Model Monitoring with MLflow**

Tech Monitoring

Implement model monitoring by logging predictions to MLflow, tracking metrics, and storing models in registry.

MLflow | Prometheus | Python

Data Analytics & Visualization-Tools

Duration: 1.5 Months

Module 7

Tool 1: Excel

Topic 1: Analytics Essentials with Excel

- Excel Basics & Core Functions: Interface, data types, formulas
- Lookup functions, conditional functions, and text functions.
- Data Preparation & Cleaning: Import/export, missing data handling, validation, formatting
- Power Query for cleaning & automation.
- Statistical & Analytical Techniques: Descriptive statistics, exploratory analysis, filters/sorting, what-if analysis, goal seek.
- Pivot-Based Analysis: PivotTables for summarization, grouping, segmentation, slicers, and data-driven insights.

Topic 2: Visualization & Advanced Analytics

- Charts & Visualization Methods: Bar, line, pie, scatter, combo charts, advanced charting, conditional formatting for insights.
- Interactive Dashboards: PivotCharts, KPI visuals, storytelling layouts, mini dashboard creation.
- Data Modeling & Power Tools: Power Pivot, relationships, basic DAX concepts, Excel add-ins for extended analytics.
- Projects & Applications: Business case studies, forecasting, reporting mini-projects, and real-world analytics workflows.

Project 45 Employee Performance & Attrition Dashboard

Human Resources

Build an Excel dashboard to analyze employee performance, satisfaction, and attrition patterns for workforce decisions.

Excel | Pivot | Power Query

Tool 2: Power BI

Topic 1: **Power BI Essentials**

- Data Prep with Power Query: Clean, transform, merge, unpivot, format text/date/number fields.
- Data Modeling: Relationships, granularity, aggregations, time-series basics, calculated columns.
- Core Visuals & Interactivity: Bar, pie, line, maps, cards, slicers, filters, drill-down, bookmarks.
- Connecting & Working with Data: Load from Excel/SQL, model views, query editor, BI best practices.

Topic 2: **Visuals, Dashboards & DAX Analytics**

- Advanced Visuals: Funnel, ribbon, waterfall, treemap, gauge, scatter, KPIs & combos.
- DAX for Analytics: Measures, calculations, time-intelligence functions.
- Dashboard Design: Layouts, formatting, interactivity, storytelling, sharing & collaboration.
- Projects: Customer segmentation dashboard, exploratory BI reports, hands-on visual assignments.

Project 46 **Campaign Report**

Marketing

Build a Power BI marketing campaign report with DAX measures and embedded Python visuals.

Power BI | DAX | Python

Tool 3: Tableau

Topic 1: Tableau Essentials & Data Preparation

- Connecting & Preparing Data: Data types, joins, blending, pivots, transformations, SQL connections.
- Core Visual Building: Bar, line, pie, scatter, bubble, maps (geo, custom geocoding, polygon, WMS).
- Data Exploration Tools: Filters, groups, sets, parameters, drill-down, marks, dimensions & measures.
- View & Layout Tools: Shelves, cards, sorting, formatting, captions, exporting, metadata management.

Topic 2: Visual Analytics, Dashboards & Advanced Features

- Advanced Visuals: Tree maps, heat maps, bump charts, funnel, waterfall, histogram, bullet & motion charts.
- Calculations & Analytics: Calculated fields, LOD expressions, trends, forecasting, predictive visuals.
- Dashboards & Storytelling: Interactive dashboards, stories, actions, design best practices, publishing & collaboration (Tableau Online).
- Projects & Applications: Build multi-chart dashboards, map-based visuals, predictive model integration.

Project 47 Interactive Sales Dashboard

Retail

Create a Tableau sales dashboard with interactive filters, drill-down capabilities, and real-time data refresh.

Tableau | SQL | CSV

DATA BASE MANAGEMENT SYSTEM

Duration: 1 Months

Module 8

SQL

Topic 1: **SQL Fundamentals**

- Core SQL & RDBMS Basics: Syntax, data types, tables, CRUD operations, querying single tables.
- Filtering & Retrieval: WHERE, ORDER BY, DISTINCT, logical operators, formatting results.
- Joins & Multi-Table Queries: INNER, LEFT, RIGHT, FULL, CROSS joins, UNION, subqueries, nested queries.
- Aggregations & Analytics: GROUP BY, HAVING, summary stats, ranking functions, Top-N analysis.

Topic 2: **Advanced SQL for Analytics**

- Views & Modifications: INSERT/UPDATE/DELETE, creating views, inline views.
- Optimization: Indexes, execution plans, performance tuning.
- Advanced Database Concepts: Normalization, stored procedures, functions, triggers, user-defined objects.
- SQL in Practice: Integrating SQL with Python, advanced analytical queries for real datasets.

Project 48 **SQL Sales Aggregator**

Retail

Write complex SQL queries to aggregate and analyze sales data from a relational database.

SQL JOIN GROUPBY

MongoDB

Topic 1: MongoDB Essentials

- MongoDB basics, data structures, and CRUD operations
- Querying data, indexing strategies, and performance tuning
- Schema design, aggregation workflows, and data processing
- NoSQL principles, transactions, replication, and scaling concepts

Topic 2: NoSQL with MongoDB

- NoSQL basics and data models (key–value, document, columnar, graph)
- MongoDB vs RDBMS and schema flexibility for scalable systems
- Data modeling, consistency ideas (CAP), replication and sharding
- NoSQL applications in analytics, caching, and distributed systems

Project 49 MongoDB Catalog Design

E-commerce

Design a MongoDB schema for a product catalog and implement filtered queries and indexing.

MongoDB | PyMongo | NoSQL

Project 50 SQL Query Optimization

IT Operations

Optimize SQL joins and indexes for a given query workload in a sample database.

SQL | EXPLAIN | Indexing

BIG DATA ANALYTICS

Duration: 1 Months

Module 9

Hadoop

Topic 1: **Hadoop Fundamentals**

- Big Data & Distributed Systems: Big Data concepts, distributed architecture, Hadoop master–slave structure.
- HDFS Operations: File storage model, directories, manipulation commands, permissions, quotas.
- MapReduce Framework: Mapper–reducer flow, shuffle/sort, architecture, real-world use cases.
- Core Hadoop Components: YARN resource management, Hadoop ecosystem overview (Hive, Pig, HBase, Sqoop).

Topic 2: **Hadoop Ecosystem**

- HDFS basics, architecture, and fault-tolerant storage.
- YARN & MapReduce job execution and resource management.
- Ecosystem integration: Hive, Sqoop, and SQL/NoSQL connectivity.
- Cluster setup, configuration, scaling, and performance tuning.

Project 51 **Query Development**

IT

Use Hive to create tables and query a large dataset (e.g., logs or ecommerce data).

Hive | Hadoop | HiveQL

Spark

Topic 1: **Spark Fundamentals**

- Spark Basics: Core features, architecture, RDDs, DAGs, transformations & actions.
- Spark Environment Setup: SparkSession, standalone mode, YARN/Mesos cluster management.
- RDD & DataFrame Operations: Creation, transformations, actions, interoperability.
- Spark SQL: Structured data processing, SQL queries, DataFrame APIs.

Topic 2: **Spark Ecosystem & Advanced Tools**

- Spark Streaming: Real-time pipelines, DStreams/Structured Streaming, Kafka integration.
- Spark MLlib: Feature engineering, building & training ML models, model evaluation.
- Performance Optimization: Caching, partitioning, Tungsten & Catalyst optimizer basics.
- PySpark & Integration: Using Spark with Python, connecting Spark with external data sources.

Project 52 **Spark Data Pipeline**

Data Engineering

Build a Spark DataFrame pipeline to clean, transform, and process a large dataset at scale.

Spark

Pandas

Numpy

DSA (Optional) For SDE

Duration: 1 Months

Module 10

DSA

Topic 1: Core Concepts & Essential Data Structures

- Time & Space Complexity, Big-O analysis
- Arrays & Strings: Sliding window, prefix/suffix, searching
- Linked Lists: Operations, fast–slow pointers, cycle detection
- Stacks & Queues: Implementations, monotonic stack/queue, deques
- Hashing: Hash maps/sets, frequency tables, efficient lookups

Topic 2: Trees, Graphs & Fundamental Algorithms

- Trees & BSTs: Traversals, patterns, tree-based problem solving
- Heaps & Priority Queues: Top-K, scheduling, optimization tasks
- Graphs: BFS/DFS, shortest paths (Dijkstra), connectivity, cycle detection
- Advanced Graphs: Topological sort, bipartite checks, MST basics
- Sorting & Searching: Merge sort, quicksort, binary search, search-on-answer

Topic 3: Problem-Solving Techniques & Developer Applications

- Recursion & Backtracking: Permutations, combinations, constraint solving
- Dynamic Programming: Memoization, tabulation, common DP patterns
- Greedy Algorithms: Interval scheduling, optimal selection strategies
- Real-World Use Cases: Caching logic, rate limiting, pathfinding, optimization
- Coding Interview Skills: Patterns, debugging, performance tuning

Project 53 **Server Load Balancer Simulation**

Distributed Systems

Simulate task distribution across servers using queues to model fair load balancing.

Queues | Round Robin | Time Simulation

Project 54 **Text Autocomplete System**

Productivity Tools

Build autocomplete engine using Trie to suggest words from dictionary as user types.

Prefix Search | Python | Trie Implementation

Project 55 **Shortest Path Route Finder**

Logistics

Implement Dijkstra's algorithm to compute shortest delivery routes on a simulated road network graph.

Graphs | Priority Queue | Dijkstra's Algorithm

Project 56 **Interactive DSA Playground Tool**

SDE

Build tool to simulate stacks, queues, trees, and graphs with step-by-step visual operations.

Streamlit | Custom DSA | Python

Project Management (Skills & Tools)

Elective

Duration: 15 Days

Topic 1: **Agile Foundations & Practices**

- Scrum framework: roles, ceremonies & artifacts
- Sprint planning, daily standups, reviews & retrospectives
- Kanban boards, WIP limits & workflow optimization
- User stories, story points, estimation & backlog prioritization
- Agile metrics: velocity, burndown charts, cycle time & throughput
- Managing cross-functional AI/Data Science projects

Topic 2: **Tools & Practical Project Execution**

- Jira setup: projects, backlogs, sprints & issue tracking
- Scrum & Kanban boards, dashboards, filters & reports
- Collaboration features: comments, mentions, notifications & permissions
- Jira free-tier limitations & best practices
- Open-source PM tools: OpenProject, Taiga, iceScrum & Redmine
- Communication & team coordination tools (Slack, Teams, Confluence)

Project 57 **AI Sprint Planning & Jira Workflow Setup**

Agile Delivery

Plan a full AI sprint, create Jira boards, define epics, and break user stories into actionable tasks.

Jira | User Story Creation | Sprint Planning

Data Science & GenAI for BFSI Professionals

Duration: 1 Month

Elective A

BFSI Domain Fundamentals

- KYC → onboarding → transactions → credit → servicing → collections
- Financial services workflows: portfolio advisory → compliance → fraud → risk scoring
- Insurance workflows: claims → underwriting → policy servicing → grievance redressal
- Key BFSI metrics: default rate, fraud probability, risk score, NPA metrics, claim turnaround, compliance accuracy

Project 1 **GenAI Credit Underwriting Document Assistant**

Credit Risk

Build LLM assistant that reviews financial documents and generates eligibility summaries for credit decisions.

Azure | LLM Summarization | Document Parsing

Project 2 **GenAI Claims Processing Assistant**

Insurance

Summarize claim documents, detect missing information, and generate verification checklists for adjusters.

LLM Review | Structuring Templates | PDF Processing

Project 3 **GenAI Portfolio Insights Generator**

Financial Advisory

Create GenAI tool that summarizes portfolio performance and provides investment insights in simple language.

LLM Analysis | JSON Input Parsing | Prompt Templates

Project 4 **Agentic Compliance Review Agent**

Banking

Build agent that checks policy documents against regulatory guidelines and flags compliance deviations.

RAG | Agentic Workflows | RBI Rule Parsing

Project 5 **Agentic Fraud Detection & Explanation Engine**

Fraud Detection

Implement multi-agent workflow that detects anomalies, explains fraud likelihood, and drafts investigation reports.

CrewAI | LangGraph | Reasoning Tools

Project 6 **GenAI Digital Banking Q&A Assistant**

Digital Banking

Create LLM-powered chatbot that answers customer queries using banking FAQs and product details.

LLM Prompting | Retrieval | Conversational Pipelines

Data Science & GenAI for Retail

Duration: 1 Month

Elective B

Retail & E- com Domain Fundamentals

- Retail workflows:
- catalog → listing → pricing → promotion → order → delivery → returns
- E-commerce stack overview:
- search, recommendation, merchandising, demand planning, customer support
- Key KPIs: conversion, AOV, returns rate, search-to-cart %, fill rate, OTIF

Project 1 **GenAI Review Summarization Assistant**

Retail

Summarize thousands of customer reviews to produce sentiment insights and product improvement recommendations.

LangChain | Sentiment Classification | LLM Pipelines

Project 2 **GenAI Returns Insights & Recommendation Assistant**

E-commerce

Analyze return reasons using GenAI and generate improvement suggestions for products, logistics, or packaging.

LLM Summaries | RAG | Sentiment Models

Project 3 **Agentic Customer Query Resolution Agent**

Customer Support

Build multi-agent workflow that resolves customer queries, checks order status, and performs automated follow-ups.

Autogen | API Tool Calling | LangGraph

Project 4 **Agentic Catalog Enrichment Workflow**

E-Commerce

Create multi-agent workflow that extracts attributes, enriches product data, and validates outputs automatically.

CrewAI | LangGraph | CSV Tools

Project 5 **Agentic Pricing Recommendation Agent**

Retail

Build an agent that analyzes demand signals and competitor pricing to generate dynamic price suggestions.

LangChain | Data Tool Calls | LLM Reasoning

Project 6 **GenAI Bundle Recommendation Engine**

Growth & Analytics

Generate product bundle ideas using purchase patterns, browsing behavior, and LLM reasoning for higher conversions.

LLM Reasoning | Association Rules | Behavioral Data

Data Science & GenAI for Manufacturing

Elective C

Duration: 1 Month

Manufacturing & Supplychain Fundamentals

- Manufacturing workflows:
- production planning → machining → assembly → quality → packaging → maintenance
- Supply chain workflows:
- procurement → inventory → warehousing → shipping → logistics → last-mile
- Key metrics:
- OEE, MTBF, defect rate, cycle time, lead time, OTIF, fill rate

Project 1 **Assembly Troubleshooting RAG Agent**

Manufacturing

Build RAG system that retrieves fault, tool, and assembly instructions to help technicians resolve issues faster.

LangChain | Vector DB | LLM Retrieval

Project 2 **GenAI Inventory Insights RAG Assistant**

Supply Chain

Build a RAG assistant that answers inventory, stock-level, and replenishment queries using supply chain documents.

LangChain | Vector DB | Azure Search

Project 3 **Agentic QA Inspection Workflow**

Quality Control

Implement multi-agent system where inspectors analyze images, classify defects, and generate corrective actions.

CrewAI | Vision Models | LangGraph

Project 4 **GenAI Logistics RCA Assistant**

Logistics

Use GenAI to analyze delivery failures and generate likely root causes with corrective action suggestions.

LLM Reasoning | Document Embeddings | RCA Prompt

Project 5 **Agentic Warehouse Task Planner**

Warehousing

Build an autonomous agents system that assigns warehouse tasks using SKU priority and worker availability.

CrewAI | LangGraph | Scheduling Logic

Project 6 **GenAI Predictive Maintenance Report Assistant**

Manufacturing

Build GenAI tool that explains sensor anomalies and generates human-readable maintenance recommendations.

LLM Reasoning | Time-Series Extraction | Summarization

Data Science & GenAI for Healthcare Professionals

Duration: 1 Month

Elective D

Healthcare Domain Fundamentals

- Healthcare workflows: patient intake → diagnosis → treatment → documentation → billing
- Pharma workflows: R&D → clinical trials → manufacturing → safety → pharmacovigilance
- Key metrics: ADRs, trial duration, readmission rate, LOS, claim turnaround, safety compliance

Project 1 **Agentic ICD/CPT Coding Assistant**

Healthcare

Build agent that extracts diagnosis/procedure codes from clinical notes using rule checks and LLM reasoning.

CrewAI | Rule-Based Validation | RAG optional

Project 2 **RAG-Based Drug Safety Signal Investigator**

Drug Safety

Build RAG system that identifies adverse drug reactions by retrieving evidence from case reports and literature.

LangChain | Vector DB | LLM Reasoning

Project 3 **GenAI Trial Protocol Summary Assistant**

Clinical Trials

Summarize complex clinical trial protocols into simplified inclusion criteria, procedures, and risk summaries.

JSON Output | LLM Summarization | Document Chunking

Project 4 **Agentic Differential Diagnosis Generator**

Diagnosis Support

Create agent that analyzes symptoms, retrieves guidelines, and suggests possible differential diagnoses safely.

CrewAI | RAG | Safety Filters

Project 5 **GenAI FDA/EMA Compliance Checker**

Compliance & QA

Build system that checks SOP content against regulatory standards and highlights missing compliance elements.

RAG + LLM | Validation Layers | Regulatory Documents

Project 6 **GenAI Patient Information Assistant**

Patient Support

Build chatbot that explains diagnoses, treatments, and care instructions in simple language for patients.

LLM Prompting | Safety Filters | Conversation Flow

Data Science & GenAI for Managers

Duration: 1 Month

Elective E

Topic 1: **Agile Foundations & Practices**

- Scrum framework: roles, ceremonies & artifacts
- Sprint planning, daily standups, reviews & retrospectives
- Kanban boards, WIP limits & workflow optimization
- User stories, story points, estimation & backlog prioritization
- Agile metrics: velocity, burndown charts, cycle time & throughput
- Managing cross-functional AI/Data Science projects

Topic 2: **Tools & Practical Project Execution**

- Jira setup: projects, backlogs, sprints & issue tracking
- Scrum & Kanban boards, dashboards, filters & reports
- Collaboration features: comments, mentions, notifications & permissions
- Jira free-tier limitations & best practices
- Open-source PM tools: OpenProject, Taiga, iceScrum & Redmine
- Communication & team coordination tools (Slack, Teams, Confluence)

Project 1 **AI Sprint Planning & Jira Workflow Setup**

Agile Delivery

Plan a full AI sprint, create Jira boards, define epics, and break user stories into actionable tasks.

Jira | User Story Creation | Sprint Planning

Project 2 **GenAI KPI Insights & Forecast Dashboard for Managers**

Business Strategy

Build AI-driven KPI dashboard that summarizes trends, identifies risks, and suggests business actions for leaders.

LLM Summaries | KPI Templates | Forecast Models

Project 3 **GenAI Marketing Campaign Design & Optimization Assistant**

Growth Strategy

Use GenAI to design campaigns, segment audiences, and generate strategic recommendations for managers.

LLM Reasoning | Prompt Templates | Customer Segmentation

Project 4 **AI Risk Assessment & Mitigation Planning Framework**

Compliance

Identify operational and financial risks, generate AI-assisted mitigation plans, and produce risk heatmaps.

LLM Analysis | Risk Scoring Matrix | Compliance Checklists

Project 5 **GenAI Root Cause & Action Plan Generator**

Operations

Use GenAI to analyze incidents, identify root causes, and auto-generate structured action plans for managers.

RCA Tree | LLM RCA Prompts | Workflow Templates

Data Science & GenAI for Software Developers

Duration: 1 Month

Elective F

AI-Powered System Design

Topic 1: AI-Powered System Design Foundations

- Principles for designing scalable GenAI and agentic applications.
- Patterns for context routing, retrieval layers, and orchestration.
- Latency, cost, safety, and reliability considerations in AI systems.
- Architecture choices for multimodal, RAG, and multi-agent workflows.

Topic 2: Production Architecture & Infrastructure Patterns

- Event-driven AI system patterns for enterprise workloads.
- Model gateways, inference routers, and distributed retrieval patterns.
- Monitoring, observability, fallback systems, and guardrail enforcement.
- Designing multi-environment deployments for regulated enterprises.

Project 1 AI Inference Gateway & Routing Layer

System Design

Create an inference router balancing requests across different LLMs based on cost and performance.

API Gateway

Routing Logic

Backend Engineering for LLM Applications & RAG Microservices

Topic 1: Backend Architecture for AI-Driven Applications

- Backend services orchestrate LLM workflows behind APIs.
- Routing layers manage dynamic prompt construction and responses.
- Middleware organizes context injection across pipelines.
- Modular design supports flexible AI integration.

Topic 2: High-Performance API Engineering for LLM Services

- Async servers handle concurrent AI requests efficiently.
- Request batching reduces model latency during peak load.
- Streaming responses improve perceived responsiveness significantly.
- API gateways unify authentication across all LLM endpoints.

Topic 3: Caching, Persistence & Storage for AI Microservices

- Caches reduce repeated model computation for identical prompts.
- Persistent layers store structured outputs and audit trails.
- Feature stores retain embeddings and conversation context.
- Distributed storage supports multi-service AI architectures reliably.

Topic 4: **Distributed RAG & AI Microservice Patterns**

- AI microservices collaborate through event-driven pipelines.
- Retrieval microservices expose optimized document search endpoints.
- Orchestration layers coordinate multi-step LLM tasks programmatically.
- Backend agents route tasks between independent functional services.

Project 2 High-Performance LLM API Service

API Architecture

Build a low-latency API service delivering LLM responses using async execution and intelligent routing.

FastAPI AsyncIO

Project 3 Intelligent Prompt Caching Microservice

Backend Systems

Implement a backend service that caches prompts and responses to reduce repeated computation and system load.

TTL Strategies Hash Keys

Capstone Projects

1 Google – Gmail AI Spam Filter

TensorFlow-based spam detector blocks ~100M+ malicious emails per day in Gmail

Python | Google Cloud APIs | TensorFlow

2 Copilot Security Auditor

NLP-powered assistant scans software configurations and code to flag security policy violations.

Python | GitHub Copilot | Azure Security Center

3 AutoGPT Shopping Assistant

Agentic AI assistant chains LLM prompts to auto-suggest products and handle user queries (orders, reviews).

LangChain | DynamoDB | Amazon Lex

4 Einstein Lead Scoring

ML model analyzes CRM data to score leads by conversion likelihood using past customer behavior.

CRM data | Python | Salesforce Einstein

Capstone Projects

5 JPMorgan Chase – LOXM Trade Execution (RL)

ML/RL trading system forecasts price impact and optimizes equity trade execution to minimize costs.

Python · RL libraries · PyTorch

6 Mastercard – AI Scam Alerting

AI sorts millions of “call before you dig” requests to identify high-risk excavations that could cut fiber.

Python · ML models · GIS data

7 Johnson & Johnson – Biopsy Image Biomarker

Computer vision on biopsy images detects FGFR gene alterations to guide cancer treatment.

Python · CNN models · Digital pathology data

8 Goldman Sachs – Credit Risk Forecasting

Predicts loan default probability using borrower data and macroeconomic indicators.

Python · SQL · XGBoost

Capstone Projects

9 Autonomous Fleet Simulation

Simulates autonomous driving scenarios (RL training environment) to improve Autopilot behavior.

Python PPO/DQN CARLA simulator

10 Zomato – Food Image Tagging

Computer vision model tags menu photos with dish names and attributes (cuisine, ingredients).

CNN Python PyTorch

11 Airtel – 5G Traffic Forecasting

Time-series model predicts 5G bandwidth demand by region/time to optimize base station deployment.

Python LSTM networks Apache Kafka

12 Emirates – VIP Sentiment Analysis

NLP analyzes social media and feedback of high-value customers to gauge satisfaction.

Python Hugging Face

Learnbay

Thank you!

A premium learning experience for professionals committed to growth, excellence, and impact.

For more queries and information please reach out to us at:

+91 77956 87988

Visit us at

www.learnbay.co