Learnbay

Data Science and GenAI Program

9+
Months Program

60+
Hours of Project

In Collaboration with

IBM

1. Why Every Professional Must Learn **Data Science & GenAI**

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

WHY ACT NOW

50% - 60%
Higher Salaries
for Data + AI skills

4 in 5 Firms
Demand AI copilots

ORGANIZATIONS ARE ADOPTING:

AI COPILOTS

AI-POWERED DECISION SYSTEMS

INTELLIGENT ASSISTANTS

AUTOMATED WORKFLOWS

2. **Who** is this program for?

THIS PROGRAM IS DESIGNED FOR

- Beginners exploring AI
- Working professionals upskilling
- Non-tech professionals using AI tools
- Developers moving into AI systems
- Product, operations, and business roles

No advanced coding background required.

Program Overview

Data Science and Generative AI Program

9 Months
Flexible batches

100% Live
Instructor-led

IBM Certified
GenAI + Agentic AI

No Cost EMI
Interest-free

Career Services Pro

- Resume Help
- Mock Interviews
- Job Referrals

3. Can Professionals Grow **Without AI** Skills?

Yes — but growth stays linear.
With Data Science and GenAI — growth becomes exponential.

3-YEAR CAREER TRAJECTORY

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

You don't need to become a data scientist.
You need an **AI-first thinking and execution stack.**

KEY SKILLS TO ADD

GenAI Fundamentals

Agentic AI + LLMOps

YOUR AI-FIRST STACK ↑ Build AI

Agentic AI

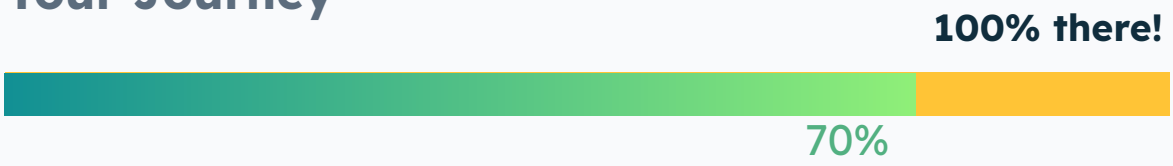
↑ GenAI Skills ↑

↑  Domain Foundation ✓ You have this ↑

5. Is this transition realistic for someone like you?

Yes. Absolutely.
You're not starting from zero - you're upgrading what you already have.

Your Journey



YOU HAVE

- ✓ Experience in your domain
- ✓ Business, technical, or operational thinking

+ JUST ADD

- \+ Analytics & Data Exploration
- \+ NLP & Machine Learning
- \+ LLM-Powered Systems & Gen AI

The question isn't if AI will reshape your domain
it's whether you'll lead it

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

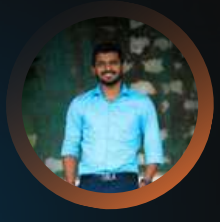
When I joined Learnbay I did not have any knowledge apart from the very basics. I gradually build my concept via various trainers and get trained in data science with strong knowledge/concepts.

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

Aravind

TheMathCompany → CATERPILLAR

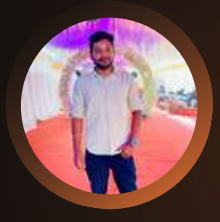
Data Scientist



Ritesh Kumar

unacademy → Capgemini

Senior Analyst

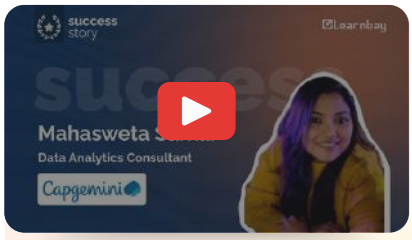


Ramki

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

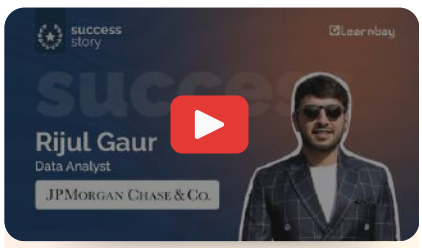
Data Quality Analyst

DATA POEM

Pinky Bhansali

Process Lead

AXA



Rahul Gaur

Data Analyst

JPMorgan Chase & Co.

Jaya Sinha

Data Scientist

amazon

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

6. Why you should **Upgrade to MasterTrack Program**

This Program	MASTERTRACK PROGRAM
Fixed learning window, One-time curriculum, no refresh	**3-Year Flexi Subscription** Re-enter, upgrade, and refresh skills as data & AI evolve.
Generic projects, limited depth	**AI CoLab Experience** Work on real startup & enterprise-grade GenAI projects.
Generic projects, limited depth	**BYOP+Mentorship** Bring your own business problem. And get 1:1 expert guidance
No end-to-end capstone or business-driven projects	**Industry Certification** Earn globally recognised credentials backed by IBM and Microsoft.
LINEAR GROWTH **~10% career uplift**	

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.

Program **Fees**

Data Science and GenAI Program

INDUSTRY PROGRAM

Live online classes

- Instructor-led **live interactive** session
- **IBM** Certification Program
- Industry-grade projects

PROGRAM FEE

1,20,000/- +18 GST

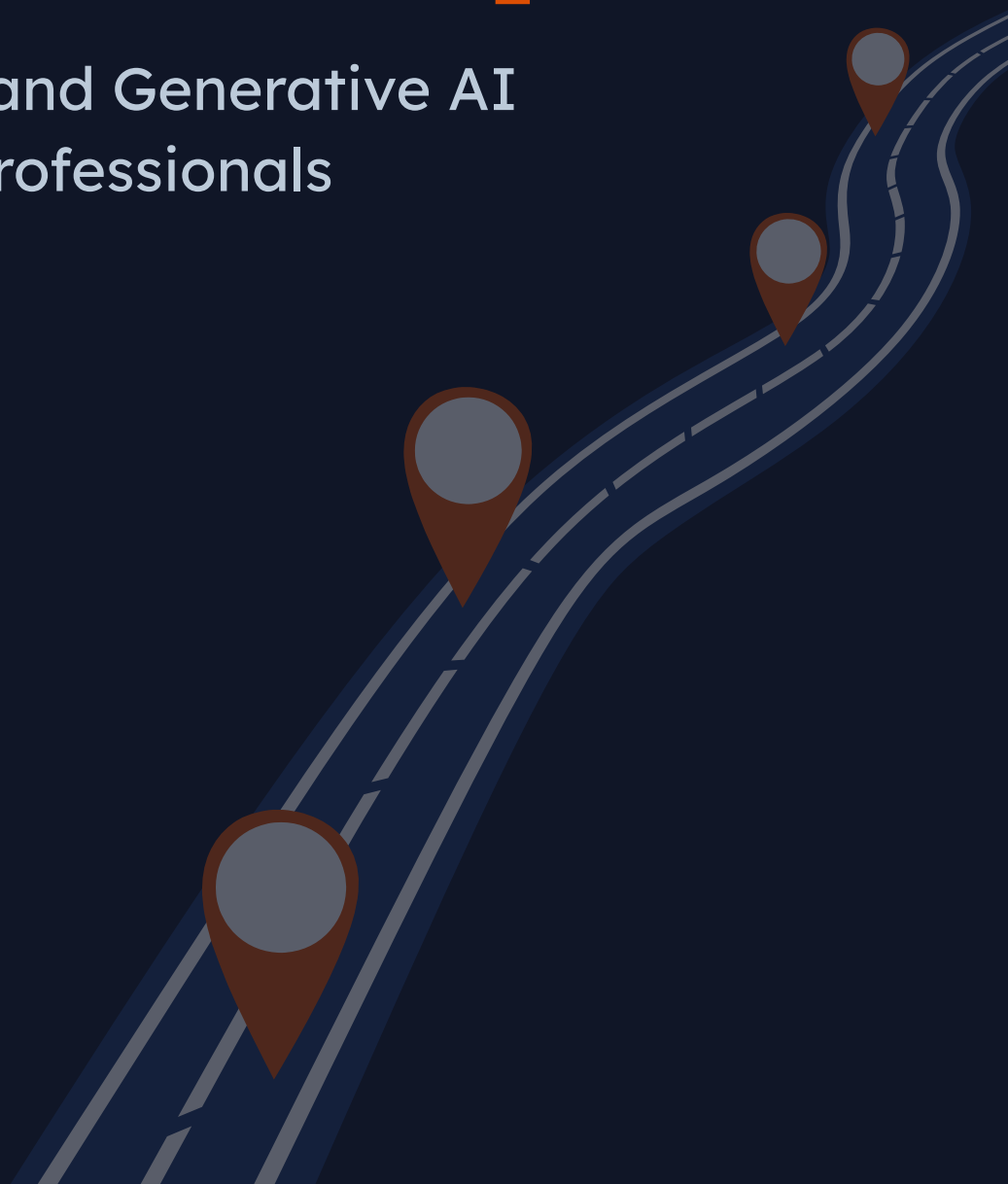
Payment Plan: No Cost EMI Available

For 6, 9 and 12 Months

9 Months Program

Program Roadmap

Data Science and Generative AI Program for Professionals



Program Roadmap

FOR DATA SCIENCE AND GENAI AI 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: 1.5 MONTHS

PYTHON FOR GENAI

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

MODULE 2

DURATION: 2 Months

FOUNDATION OF GENAI

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

MODULE 3

DURATION: 2 Months

GENAI & AGENTIC AI

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

MODULE 4

DURATION: 2 Months

STATISTICS & MACHINE LEARNING

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

MODULE 5

DURATION: 1.5 Months

DATA ANALYTICS & VISUALIZATION- TOOLS

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

MODULE 6

DURATION: 1.5 Months

Data Management & SQL

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

MODULE 7

DURATION: 1 Months

BIG DATA ANALYTICS

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

9 Months Program

Program Curriculum

Data Science and Generative AI Program

60+ Hours of Industrial Projects

18+ 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.

Python for GenAI

Duration: 1.5 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

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.

Topic 3: **Essential Libraries for GenAI**

- **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 3 **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

Mock Interviews

5 Session with Industry Expert

Multiple Technical and HR Round with detailed feedback

Foundation for GenAI

Duration: 2 Months

Module 2

Topic 1: **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 4 **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 5 **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 2: **Deep Learning Fundamentals**

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

Project 6 **Smart Document Relevance Classifier**

HR

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

Numpy Perceptron

Topic 3: **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**

Topic 3: **NLP Fundamentals**

- **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 7 **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

Ritesh Kumar

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

Placed at Capgemini

110% Salary Hike

Saurabh Kumar

From **Professor** → To **Data Scientist**

Placed at Teleperformance

135% Salary Hike

Generative AI & Agentic AI

Duration: 1.5 Months

Module 3

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 8 **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 9 **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 10 **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.

Topic 7: **Langchain / Llama Index**

- 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**.

Project 11 **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 09: **Azure OpenAI**

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

Agentic AI

Duration: 15 Days

Topic 1: **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 2: **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 3: **AutoGen**

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

Project 12 **AI Data Analyst Agent**

Fianance

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 4: **CrewAI**

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

Topic 5: **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 13 **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 6: **MCP (Model Context Protocol) / ACP / A2A**

- Model Context Protocol (**MCP**) Fundamentals.
- Agent Communication Protocol (**ACP**).
- **Agent-to-Agent** (A2A) Interaction Patterns.
- Context Sharing Across Distributed Agents.
- **Secure Context Exchange** Between Models.
- Enterprise Interoperability Standards For Agents.

Project 14 **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 7: **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 8: **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 And Output Validation.
  - Policy Enforcement And Rule-Based Constraints.
  - Monitoring Agent Actions In Production.

Project 15 **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 9: **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 16 **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 & Machine Learning

Duration: 2 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 17 **Survey Correlation Study**

Consumer Research

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

Numpy | Statsmodels | Pandas

Project 18 **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 19 **Sales Dashboard EDA**

Retail

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

Pandas | Seaborn | Plotly

Project 20 **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 21 **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 22 **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
- Cluster validation: elbow method, silhouette analysis
- Distance metrics: Euclidean, Manhattan, domain-specific measures

Topic 10: **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 11: **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 12: **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 23 **Customer Segmentation**

E-commerce

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

Scikit-learn | Pandas | Matplotlib

Topic 13: **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 24 **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 14: **Time Series Analysis**

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

Topic 15: **Classical Forecasting Methods**

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

Advanced Machine Learning

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 25 **Transformer Sentiment Pipeline**

Marketing

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

Transformers | Pandas | scikit-learn

Project 26 **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

Data Analytics & Visualization-Tools

Duration: 1.5 Months

Module 5

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 27 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 28 **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 29 Interactive Sales Dashboard

Retail

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

Tableau | SQL | CSV

DATABASE MANAGEMENT SYSTEM

Duration: 1.5 Months

Module 6

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 30 **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 31 MongoDB Catalog Design

E-commerce

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

MongoDB | PyMongo | NoSQL

Project 32 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 7

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 33 **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 34 **Spark Data Pipeline**

Data Engineering

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

Spark

Pandas

Numpy

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