CareerByteCode
Learning Made simple

CareerByteCode Admissions
support@careerbytecode.in
www.careerbytecode.in



CAREERBYTECODE PROGRAM

Predictive Intelligence: Applied Data Science & ML

The CareerByteCode Predictive Intelligence - Applied Data Science & ML is an intensive, practical program designed for learners who want to build a foundation in data science and applied machine learning. It starts with programming and data skills, develops statistical thinking and analytical ability, and progresses into core machine learning workflows.

Build up your skills.
Step up your career.

Mentors: Jenefer Rexee George & Daisy Grace Thomas

8 weeks · 24 live classes · 48 instructor-led hours · 4 major projects · Hybrid: classroom + online

Course Overview

LEARNING MADE SIMPLE

The CareerByteCode Predictive Intelligence - Applied Data Science & ML is an intensive, practical program designed for learners who want to build a foundation in data science and applied machine learning. It starts with programming and data skills, develops statistical thinking and analytical ability, and progresses into core machine learning workflows. Learners also receive an introduction to Deep Learning, NLP and Generative AI.

The program is designed for absolute beginners. It prioritises practical understanding, guided coding, industry-style datasets, project work and the ability to explain analytical decisions - not simply memorising algorithms or library commands.

8	24	48	3/wk	4
WEEKS	LIVE CLASSES	HOURS	CLASSES	PROJECTS

Your Mentors

LEARN FROM PRACTITIONERS



Jenefer Rexee George

AI ML DATA CONSULTANT

Jenefer Rexee George is an AI and ML Data Consultant and Software Engineer (AI/ML) who builds intelligent systems that combine data, machine learning and modern AI to solve real-world problems.



Daisy Grace Thomas

AI CONSULTANT · AI PRODUCT & AUTOMATION ENGINEER

Daisy Grace Thomas is an AI Product and Automation Engineer and Business Analyst with 9+ years delivering enterprise solutions across IT services, EdTech, finance and aerospace.

CareerByteCode
Learning Made simple

Table of Contents

PREDICTIVE INTELLIGENCE

01 Program at a Glance

02 Who This Course Is for

03 Before You Start: Prerequisites

04 What You'll Achieve: Learning Outcomes

05 Week-Wise Curriculum and Detailed Class Plan

06 Four Major Practical Projects

07 Assessment and Evaluation

08 Tools and Learning Environment

09 Training Methodology and Delivery

10 Career Support and Portfolio

11 Commercial Details

12 Important Academic Notes

How every class works

Each class combines concept explanation, a live instructor demonstration, guided practice and a short recap. You practise between classes and finish four projects for your portfolio. Delivery: Hybrid, classroom + online. Next batch dates are shared on enrolment.

CareerByteCode

Learning Made simple

Program at a Glance

PREDICTIVE INTELLIGENCE

Duration	8 weeks
Schedule	3 classes per week; 24 classes total
Class duration	2 hours
Instructor-led hours	48 hours
Audience	Students, freshers, working professionals and career switchers
Entry level	Absolute beginners; no prior coding experience required
Delivery	Hybrid: classroom + online
Core tools	Python, SQL, NumPy, Pandas and Scikit-learn
Additional concepts	Statistics, probability, visualisation, introductory Deep Learning, NLP and Generative AI
Major projects	4
Fee	₹40,000 + applicable GST

Who This Course Is for

PREDICTIVE INTELLIGENCE

Students, freshers, working professionals and career switchers

CareerByteCode
Learning Made simple

Before You Start: Prerequisites

PREDICTIVE INTELLIGENCE

- No previous programming or Data Science experience is required.
- Basic computer and internet usage skills.
- Ability to use a laptop or desktop for coding exercises.
- A willingness to practise between classes and complete assignments.
- A laptop with a modern browser; Python, Jupyter/Google Colab and required tools will be installed or accessed during the course.
- Basic school-level mathematics is sufficient to begin; relevant statistics and probability will be taught from the ground up.

What You'll Achieve: Learning Outcomes

PREDICTIVE INTELLIGENCE

By the end of the program, learners should be able to:

- Explain the Data Science lifecycle and distinguish Data Science, Data Analytics and Machine Learning tasks.
- Write beginner-level Python programs and work with common data structures, functions and files.
- Use NumPy and Pandas to load, inspect, clean, transform and analyse datasets.
- Write SQL queries to retrieve, filter, aggregate and join data from relational tables.
- Apply descriptive statistics, probability concepts and basic statistical reasoning to real datasets.
- Perform exploratory data analysis (EDA), identify patterns and communicate findings using visualisations.
- Build and evaluate foundational regression, classification and clustering models using Scikit-learn.

- Understand decision trees, random forests, boosting concepts, feature engineering and model tuning at an introductory applied level.
- Explain the purpose and basic workflow of Deep Learning and NLP without claiming specialist-level implementation expertise.
- Use LLMs responsibly, write effective prompts and understand a basic Retrieval-Augmented Generation (RAG) workflow.
- Complete and present portfolio projects, explain methodology and discuss model limitations.

CareerByteCode
Learning Made simple

Week-Wise Curriculum and Detailed Class Plan

PREDICTIVE INTELLIGENCE

The curriculum contains 2 modules per week and 3 classes per week. Each class is planned for 2 hours. The instructor should reserve time in every session for a guided demonstration, learner practice and a short recap.

WEEK 01 — Data Science Foundations and Python Basics

2 MODULES I CLASSES 1-3

Module 1: Introduction to Data Science

- What is Data Science? Data, information and insights
- Typical Data Science lifecycle: problem definition, data collection, preparation, analysis, modelling and communication
- Industry use cases across finance, healthcare, retail, HR and e-commerce
- Data Science vs Data Analytics vs Business Intelligence vs Machine Learning
- Roles and responsibilities: Data Scientist, Data Analyst and ML Engineer
- Structured, semi-structured and unstructured data; datasets, rows, columns and data types

Module 2: Python Programming for Beginners

- Python installation or Google Colab/Jupyter setup; notebooks and cells
- Data types: integers, floats, strings and booleans; type conversion
- Conditional statements: if, elif and else
- Lists, tuples, sets and dictionaries
- Variables, naming rules, comments and basic input/output
- Operators and expressions
- Loops: for and while; range; break and continue
- Functions, parameters, return values and basic error handling

DETAILED CLASS PLAN

Class 1: Introduction to Data Science and the Data Lifecycle

- Data Science concepts and terminology
- Roles, lifecycle and real-world business questions
- Types of data and common industry datasets

✓ Hands-on: Classify example business problems and identify the data needed to solve them.

Class 2: Python Fundamentals: Syntax, Variables and Decisions

- Notebook environment and Python execution
- Variables, data types, operators and input/output
- if/elif/else, comparisons and logical operators

✓ Hands-on: Write small programs to calculate totals, discounts and eligibility conditions.

Class 3: Python Fundamentals: Loops, Collections and Functions

- for/while loops and iteration
- Lists, tuples, sets and dictionaries
- Functions, parameters, return values and debugging basics

✓ Hands-on: Build a simple data-entry and summary program using Python collections.

WEEK 02 — Python for Data Work, NumPy and Pandas

2 MODULES I CLASSES 4-6

Module 3: Python Skills for Data Handling

- Strings and common string operations
- Working with dates and basic data conversion
- Modules, packages and importing libraries
- Reading and writing text and CSV files
- List/dictionary comprehensions (introductory)
- Exceptions, debugging and writing readable code

Module 4: NumPy and Pandas

- NumPy arrays, dimensions, shape, indexing and slicing
- Pandas Series and DataFrames
- Inspecting data: head, tail, info, describe and value_counts
- Handling missing values and duplicate records
- Vectorised operations and basic aggregation
- Reading CSV and Excel files
- Selecting, filtering, sorting and creating columns
- Grouping, aggregation, merging and exporting results

DETAILED CLASS PLAN

Class 4: Python Data Handling and Working with Files

- Strings, dates and type conversion
- Read/write CSV files and inspect file paths
- Imports, packages, exceptions and debugging

✓ Hands-on: Load a small CSV file and calculate basic summary values.

Class 5: NumPy and Pandas Foundations

- NumPy arrays, shapes, indexing and vectorised operations
- Pandas Series and DataFrames
- Load datasets; inspect columns, types and summary statistics

✓ Hands-on: Explore a sales dataset and answer basic questions using Pandas.

Class 6: Data Manipulation with Pandas + Project 1

- Filtering, sorting, selecting and creating columns
- Missing values, duplicates, grouping and aggregation
- Merge datasets and export cleaned data

✓ Hands-on: Project 1: Python-based Retail/Sales Data Exploration. Produce a cleaned dataset and a short insight summary.

WEEK 03 — Statistics, Probability and SQL

2 MODULES I CLASSES 7-9

Module 5: Mathematics, Statistics and Probability for Data Science

- Why statistics matters in Data Science
- Mean, median, mode, range, variance and standard deviation
- Probability basics, events and conditional probability
- Correlation vs causation; covariance
- Population, sample, variables and data distributions
- Quartiles, percentiles and interquartile range
- Common distributions: normal and binomial (conceptual and applied)
- Sampling, outliers and introductory hypothesis testing; p-value intuition

Module 6: SQL for Data Science

- Relational databases, tables, rows, columns, primary and foreign keys
- AND, OR, IN, BETWEEN, LIKE and NULL handling
- GROUP BY and HAVING
- Subqueries and Common Table Expressions (introductory)
- SELECT, DISTINCT, WHERE, ORDER BY and LIMIT
- Aggregate functions: COUNT, SUM, AVG, MIN and MAX
- INNER, LEFT and RIGHT JOIN concepts
- SQL use cases in analytical workflows

DETAILED CLASS PLAN

Class 7: Statistics: Descriptive Measures and Data Distributions

- Population, sample and variable types
- Central tendency and dispersion
- Quartiles, percentiles, outliers and distribution shape

✓ Hands-on: Calculate and interpret summary statistics for a business dataset.

Class 8: Probability and Statistical Thinking

- Probability rules and conditional probability
- Normal and binomial distribution intuition
- Correlation, causation, sampling and hypothesis-testing concepts

✓ Hands-on: Interpret simple scenarios and explain what a statistical result does - and does not - mean.

Class 9: SQL Foundations for Data Analysis

- Relational database concepts and table relationships
- SELECT, WHERE, ORDER BY, DISTINCT and filtering
- Aggregations, GROUP BY and HAVING

✓ Hands-on: Query a sample retail database to answer sales and customer questions.

WEEK 04 — SQL Applications, Data Cleaning and EDA

2 MODULES I CLASSES 10-12

Module 7: Applied SQL and Data Preparation

- INNER and LEFT JOIN with practical examples
- Subqueries and CTEs for readable analysis
- Date functions and text functions (introductory)
- Joining multiple tables and understanding duplicate rows
- CASE WHEN and conditional categorisation
- Data quality checks and reconciling SQL results with Pandas

Module 8: Exploratory Data Analysis and Visualisation

- Purpose and workflow of EDA
- Univariate, bivariate and multivariate exploration
- Visualisation principles and choosing an appropriate chart
- Writing observations, business questions and insight summaries
- Data quality: missingness, duplicates, inconsistent categories and outliers
- Distribution, relationship and category comparisons
- Matplotlib and Seaborn introduction

DETAILED CLASS PLAN

Class 10: Applied SQL: Joins, Subqueries and Business Questions

- INNER/LEFT JOIN and table relationships
- Subqueries, CTEs and CASE WHEN
- Date filtering and analytical query patterns

✓ Hands-on: Answer multi-table HR or e-commerce questions using SQL.

Class 11: Data Cleaning and Exploratory Data Analysis

- A repeatable data-quality checklist
- Missing values, duplicates, outliers and inconsistent formats
- Univariate and bivariate analysis using Pandas

✓ Hands-on: Clean and explore a healthcare or HR dataset, documenting data-quality decisions.

Class 12: Visualisation and Project 2

- Matplotlib/Seaborn chart fundamentals
- Histograms, box plots, bar charts, scatter plots and correlation heatmaps
- Translate patterns into clear findings

✓ Hands-on: Project 2: Healthcare/HR Exploratory Data Analysis. Submit a cleaned dataset, notebook, charts and a one-page insight report.

WEEK 05 — Machine Learning Foundations and Regression

2 MODULES I CLASSES 13-15

Module 9: Introduction to Machine Learning

- What is Machine Learning and where it is useful
- Features, labels, observations and target variables
- Data leakage and why it creates misleading results
- Feature scaling and basic feature engineering
- Supervised, unsupervised and reinforcement learning (overview)
- Training, validation and test datasets
- Preprocessing pipelines and encoding categorical data
- Scikit-learn workflow: fit, predict and transform

Module 10: Regression Algorithms

- Regression problems and business examples
- Train/test split and baseline models
- Evaluation: MAE, MSE, RMSE and R-squared
- Interpreting model outputs and limitations
- Simple and multiple linear regression
- Predictions, residuals and regression assumptions (introductory)
- Underfitting and overfitting

DETAILED CLASS PLAN

Class 13: Machine Learning Concepts and Workflow

- ML types and common use cases
- Features, target, training and test data
- Data leakage, preprocessing and Scikit-learn workflow

✓ Hands-on: Prepare a modelling dataset and create a simple train/test split.

Class 14: Regression: Predicting Numeric Values

- Linear and multiple regression concepts
- Fit a regression model with Scikit-learn
- Predictions, residuals and feature interpretation

✓ Hands-on: Build a first model to estimate a numeric business outcome.

Class 15: Regression Evaluation and Project 3 Start

- MAE, MSE, RMSE and R-squared
- Baseline comparison, overfitting and limitations
- Feature preparation and experiment tracking basics

✓ Hands-on: Project 3: Finance/Retail Regression Model - for example, predicting sales, demand or a numeric risk-related measure using an approved teaching dataset.

WEEK 06 — Classification, Ensembles and Model Tuning

2 MODULES I CLASSES 16-18

CareerByteCode
Learning Made Simple

Module 11: Classification Algorithms

- Classification problems and class labels
- K-Nearest Neighbours (KNN)
- Naive Bayes (concept and demonstration)
- Class imbalance and threshold intuition
- Logistic regression
- Decision trees
- Confusion matrix, accuracy, precision, recall and F1-score
- Choosing metrics based on business impact

Module 12: Ensemble Learning and Model Improvement

- Why ensembles can improve model performance
- Bagging and boosting concepts
- Feature importance and interpretation
- GridSearchCV and RandomizedSearchCV (guided introduction)
- Random Forest
- Gradient Boosting and introductory model examples
- Hyperparameters vs model parameters
- Cross-validation, model comparison and reproducibility

DETAILED CLASS PLAN

Class 16: Classification: Predicting Categories

- Classification use cases in finance, HR and e-commerce
- Logistic regression, KNN and decision trees
- Train and test classification models

✓ Hands-on: Build a simple churn, approval or attrition classification model with a teaching dataset.

Class 17: Model Evaluation and Business Trade-offs

- Confusion matrix and classification metrics
- Precision vs recall; false positives and false negatives
- Class imbalance and choosing a metric for the use case

✓ Hands-on: Explain which error matters most in a selected business scenario.

Class 18: Ensemble Models, Tuning and Project 3 Completion

- Random Forest and boosting concepts
- Cross-validation and hyperparameter tuning
- Compare baseline and improved models; communicate limitations

✓ Hands-on: Complete Project 3 with a documented model comparison and evaluation report.

WEEK 07 — Unsupervised Learning, Deep Learning and NLP

2 MODULES I CLASSES 19-21

Module 13: Unsupervised Learning and Model Interpretation

- Supervised vs unsupervised learning
- K-Means algorithm and cluster interpretation
- Dimensionality reduction and PCA (conceptual introduction)
- Model interpretation, fairness, bias and limitations
- Clustering use cases
- Choosing the number of clusters; elbow method intuition
- Anomaly detection use cases (overview)

Module 14: Introduction to Deep Learning and NLP

- What is Deep Learning and how it differs from traditional ML
- Training loop, loss and epochs (conceptual)
- What is NLP; text preprocessing and tokenisation
- Embeddings and the relationship between NLP and modern LLMs
- Neurons, layers, weights, activation functions and neural-network intuition
- Where CNNs and sequence models are used (overview)
- Text classification and sentiment analysis concepts

DETAILED CLASS PLAN

Class 19: Unsupervised Learning: Clustering and Segmentation

- Clustering problems and business use cases
- K-Means workflow and cluster profiles
- Elbow method and interpreting segments

✓ Hands-on: Segment customers or products using a prepared dataset.

Class 20: Dimensionality Reduction, Responsible ML and Interpretation

- PCA intuition and why dimensionality reduction is used
- Anomaly detection overview
- Bias, fairness, explainability and limitations of predictive models

✓ Hands-on: Review a model's output and identify risks before business use.

Class 21: Deep Learning and NLP Foundations

- Neural-network building blocks and training concepts
- Deep Learning use cases; CNN/sequence-model overview
- NLP workflow, tokenisation, sentiment and text classification

✓ Hands-on: Run or inspect a guided introductory text-classification example.

WEEK 08 — Generative AI, RAG and Capstone

2 MODULES I CLASSES 22-24

Module 15: Generative AI for Data Science

- What is Generative AI; LLM concepts and common capabilities
- Prompt engineering: role, task, context, constraints and output format
- Hallucinations, verification and human review
- RAG basics: documents, chunking, embeddings, vector search and grounded answers
- Tokens, context windows and model limitations (high-level)
- Using LLMs to support data exploration, SQL/Python explanation and report drafting
- Privacy, sensitive data and safe use of AI tools
- Simple RAG workflow and practical limitations

Module 16: End-to-End Data Science Capstone and Career Readiness

- Problem statement, success criteria and dataset selection
- Feature preparation, baseline model and model selection
- Presenting insights to a technical and non-technical audience
- Resume project bullets, interview discussion and mock presentation
- Data understanding, cleaning and EDA
- Evaluation, error analysis and limitations
- Portfolio README, notebook organisation and project storytelling

DETAILED CLASS PLAN

Class 22: Generative AI, LLMs and Prompt Engineering

- LLM fundamentals, tokens and context
- Prompt structure and practical prompting patterns
- Using AI to assist analysis while validating every output

✓ Hands-on: Create prompts to explain a dataset, draft SQL/Python assistance and summarise verified findings.

Class 23: RAG Fundamentals and Capstone Build

- RAG architecture: source documents, chunking, embeddings and retrieval
- Grounded question answering and citation/verification concepts
- Capstone integration: problem, data, EDA, model and evaluation

✓ Hands-on: Build a small guided RAG prototype or integrate a RAG demonstration into the capstone where feasible.

Class 24: Capstone Presentation, Portfolio and Interview Preparation

- Final project walkthrough and business recommendation
- Explain data preparation, model choice, metrics and limitations
- Portfolio and GitHub presentation; resume project bullets
- Mock interview questions and feedback

✓ Hands-on: Present the final capstone and receive structured feedback.

Four Major Practical Projects

PREDICTIVE INTELLIGENCE

CareerByteCode
Learning Made simple

P1

Python Retail/Sales Data Exploration

Schedule: Week 2

Use Python, NumPy and Pandas to load, inspect, clean and summarise retail or sales data. Learners calculate key totals, compare categories and produce a concise insight summary.

Tools: Python, NumPy, Pandas

P2

Healthcare/HR Exploratory Data Analysis

Schedule: Week 4

Work with a synthetic or de-identified dataset. Clean missing and inconsistent data, explore distributions and relationships, create visualisations and communicate findings. No personal patient or employee data should be used.

Tools: Python, Pandas, Matplotlib/Seaborn, SQL

P3

Predictive Machine Learning Model

Schedule: Weeks 5-6

Build a regression or classification model using a finance, retail or e-commerce teaching dataset. Compare a baseline with one or more algorithms, evaluate results and explain model limitations.

Tools: Python, Pandas, Scikit-learn

P4 **End-to-End Data Science Capstone**

Schedule: Week 8

Choose an approved business problem and take it through problem framing, data preparation, EDA, feature preparation, modelling or segmentation, evaluation and presentation. Include a small Generative AI/RAG demonstration as an extension where appropriate.

Tools: Python, SQL, Pandas, Scikit-learn, Generative AI tools

Assessment and Evaluation

PREDICTIVE INTELLIGENCE

ASSESSMENT	WEIGHT	EVALUATION FOCUS
Class exercises and coding practice	20%	Short guided exercises covering Python, SQL, Pandas, statistics and ML concepts.
Weekly assignments and quizzes	15%	Concept checks, code reading, query writing and short analytical tasks.
Projects 1-3	30%	Data preparation, correctness, analysis quality, model selection and communication.
Final capstone	25%	Problem definition, end-to-end workflow, evaluation, practical value and presentation.
Participation and presentation	10%	Engagement, explanation of decisions, peer discussion and professional communication.

Suggested completion standard: learners should attend the required sessions, submit all four projects and achieve defined passing threshold in assignments and capstone evaluation. CareerByteCode can set the exact attendance and pass criteria before enrolment.

Tools and Learning Environment

PREDICTIVE INTELLIGENCE

- Python 3 and Google Colab or Jupyter Notebook for coding.
- NumPy and Pandas for numerical operations and data manipulation.
- SQL with a beginner-friendly relational database environment.
- Matplotlib and Seaborn for introductory data visualisation.
- Scikit-learn for classical machine learning.
- GitHub (introductory) for organising and presenting project work.
- An approved LLM interface for prompt engineering and RAG demonstrations; access and usage limits may depend on the provider.

Training Methodology and Delivery

PREDICTIVE INTELLIGENCE

- Hybrid delivery: the same learning objectives and assignments should be available to classroom and online learners.
- Recommended class rhythm: approximately 25-35 minutes concept explanation, 45-55 minutes instructor-led demonstration, 25-35 minutes learner practice and 5-10 minutes recap.
- Use small, guided datasets first, then progressively introduce realistic multi-column industry datasets.
- Provide starter notebooks, datasets, class notes, assignment instructions and project rubrics.
- Use synthetic or properly de-identified finance, healthcare, HR, retail and e-commerce datasets. Do not upload confidential or personally identifiable data to public AI tools.
- Provide recordings or catch-up materials where operationally possible, plus a defined doubt-clearing process.

Career Support and Portfolio

PREDICTIVE INTELLIGENCE

- Course completion certificate based on attendance and assessment requirements.
- Resume guidance focused on demonstrable skills and completed projects.
- Portfolio structure for notebooks, SQL scripts, project reports and presentations.
- Mock interviews covering Python basics, SQL, statistics, EDA and ML fundamentals.
- Interview preparation with scenario-based questions and project explanation practice.
- Guidance on presenting project outcomes honestly, including assumptions, limitations and metrics.

Commercial Details

PREDICTIVE INTELLIGENCE

Program fee: ₹40,000 + applicable GST per learner.

Duration and delivery: 8 weeks I 24 classes I 48 instructor-led hours I Hybrid classroom + online.

Included learning support: Course materials, guided exercises, four project briefs, assessment, certificate subject to completion criteria, resume guidance, mock interviews, portfolio and interview preparation.

Important Academic Notes

PREDICTIVE INTELLIGENCE

- The program is intensive: 48 instructor-led hours are not sufficient by themselves for mastery. Learners should plan regular independent practice between classes.
- Advanced mathematics, deep neural-network implementation, advanced NLP, distributed data engineering, cloud deployment and MLOps are intentionally outside the core scope. They can be offered as advanced follow-on courses.
- The Generative AI module introduces LLMs, prompting and RAG architecture. A production-grade RAG application, fine-tuning or agentic system is not a promised outcome of this short module.
- Project datasets should be curated for beginner accessibility and should not contain identifiable personal, patient, employee or financial account data.
- The instructor should adapt the pace based on learner progress, especially during Python, statistics and SQL weeks.

CareerByteCode
Learning Made simple

CareerByteCode
Learning Made simple

THANK YOU

Join the next batch, learn from practitioners, and leave with projects you can show and explain.

CareerByteCode Admissions · support@careerbytecode.in · www.careerbytecode.in/courses/predictive-intelligence-data-science/

CareerByteCode is a brand of Cloudnloud Technologies Pvt Ltd, Chennai. CareerByteCode Team.