Proleed
ACADEMY

Data Science Training Course

CURRICULUM

Module I – Introduction to Data Science

- What is Data Science? Lifecycle and ecosystem
- Data Science vs Data Analytics vs Data Engineering
- End-to-end pipeline: data insight impact
- Key concepts: data, information, knowledge, and insight
- Roles in a data team: Data Scientist, Analyst, Engineer, ML Engineer
- Data-driven decision making & experimentation culture
- Business understanding and problem framing
- Case studies: Netflix recommendations, Uber surge pricing, Amazon forecasting
- Tools & environments overview (Jupyter, Colab, VS Code, cloud notebooks)

Module II – Database Fundamentals

- What is a database and its role in analytics
- Relational vs non-relational databases
- Tables, records, keys, and relationships
- Database design and normalization
- ER diagrams and schema structure
- CRUD operations (Create, Read, Update, Delete)
- Data constraints and referential integrity
- Data warehousing basics
- Connecting Python, Tableau, and Power BI to databases

Module III – SQL for Analysts

- Understanding databases, schemas, and tables
- SELECT, WHERE, ORDER BY, GROUP BY, and HAVING clauses
- Data extraction and transformation using joins and subqueries
- Window functions (RANK, ROW_NUMBER, LAG, LEAD)
- Common Table Expressions (CTEs) for complex queries
- Aggregations, nested queries, and date/time manipulations
- Creating and using views for analytics
- SQL query optimization and indexing basics
- Real-world analytics use cases and mini-projects

Module IV – MySQL for Data Analytics

- Setting up MySQL and MySQL Workbench
- Writing and executing analytical queries
- Importing/exporting datasets (CSV, JSON)
- Stored procedures, triggers, and views
- User roles, permissions, and data security
- MySQL indexing and performance tuning
- Integrating MySQL with Tableau and Power BI

Module V – NoSQL for Data Analytics

- What is NoSQL and why it emerged
- Limitations of relational databases for modern data
- Core NoSQL database types (Document-based, Key-Value, Column-family, Graph)
- Differences between SQL and NoSQL (schema, scaling, flexibility)
- CAP theorem and BASE properties
- Data modeling strategies in NoSQL systems
- When to choose NoSQL over SQL in analytics workflows
- Overview of popular NoSQL databases (MongoDB, Cassandra, Redis, Neo4j)
- Connecting NoSQL systems to analytics tools

Module VI – MongoDB for Data Analytics

- Introduction to MongoDB and BSON structure
- Core components: databases, collections, documents, indexes
- Installing and setting up MongoDB & MongoDB Compass
- CRUD operations (insertOne, find, updateOne, deleteOne)
- Advanced queries and filtering
- Aggregation pipelines for analytical transformations
- Data modelling and schema design for analytics
- Indexing and performance tuning in MongoDB
- Integrating MongoDB with Python (PyMongo and Pandas)

Module VII – Python for Data Analytics

- Introduction to Python and Jupyter Notebooks
- Variables, control structures, and functions
- Working with files (CSV, JSON, Excel)
- Handling APIs and JSON data
- Using virtual environments and packages
- Automating data workflows
- Integrating Python with BI tools

Module VIII – NumPy for Numerical Computing

- Introduction to NumPy and its role in data analytics
- Creating and manipulating NumPy arrays
- Indexing, slicing, reshaping, and stacking arrays
- Understanding array data types and memory efficiency
- Vectorised operations and broadcasting
- Mathematical and statistical operations on arrays
- Random number generation and reproducibility
- Working with multi-dimensional (nD) data
- Integration of NumPy with Python and pandas

Module IX – Pandas for Data Manipulation

- Introduction to pandas and its importance in analytics
- Working with Series and DataFrames
- Importing and exporting datasets (CSV, Excel, JSON)
- Data inspection and exploration (info(), describe(), etc.)
- Data cleaning: handling missing values and duplicates
- Filtering, sorting, and conditional selections
- Grouping, aggregating, merging, and joining datasets
- Feature engineering and date-time operations
- Combining pandas with NumPy for advanced analysis

Module X – Matplotlib & Seaborn for Visualization

- Basics of Matplotlib plotting
- Seaborn for advanced statistical visuals
- Creating bar, line, scatter, heatmap, and box plots
- Customizing charts (titles, labels, grids, colors)
- Subplots and multiple visual layouts
- Storytelling through visuals
- Automating reporting with Python scripts

Module XI – scikit-learn for Predictive Analytics

- Machine Learning overview for analysts
- Train-test split, scaling, and encoding
- Regression models (Linear, Ridge, Lasso)
- Classification models (Logistic, KNN, Decision Tree)
- Model performance metrics (accuracy, precision, recall, F1-score)
- Cross-validation and hyperparameter tuning
- Model persistence and pipeline creation

Module XII – Data Cleaning & Pre-processing

- Identifying missing values and imputation techniques
- Outlier detection using statistical and visual methods
- Data normalization and standardization
- Encoding categorical variables (label, one-hot, ordinal encoding)
- Dealing with inconsistent data and duplicates
- Date-time formatting and time-series alignment
- Feature scaling and transformation
- Data validation and integrity checks
- Maintaining data quality across pipelines

Module XIII – Exploratory Data Analysis (EDA)

- What is EDA and why it's essential in analytics
- EDA process: understanding data → cleaning → exploring → interpreting
- Univariate analysis (distributions, histograms, box plots)
- Bivariate and multivariate analysis (correlation, scatter, pair plots)
- Identifying outliers and anomalies visually and statistically
- Data summarization using descriptive statistics
- Feature distributions and skewness detection
- Handling categorical vs numerical data
- Using pandas, Matplotlib, and Seaborn for EDA
- Project: Perform full EDA on a real-world dataset (sales, HR, or e-commerce)

Module XIV – Data Visualization

- Best practices in data visualization and storytelling
- Choosing the right visualization type for your data
- Data-to-insight storytelling with visuals
- Using Power BI, Tableau, or Looker for professional dashboards
- Interactive filters, drill-downs, and slicers
- Designing for non-technical audiences
- Visual perception and color psychology in dashboards
- Avoiding misleading or biased visuals
- Creating automated, real-time dashboards

Module XV – PowerBI (BI) Tools

- Power BI: DAX formulas, data modeling, and relationships
- Power Query for data transformation in BI
- Tableau: calculated fields, filters, parameters, and blending data sources
- Google Data Studio basics and integrations
- Dashboard design principles for executives
- Data refresh automation and scheduling
- Role-based dashboard access and permissions
- Embedding BI reports into business workflows
- Connecting BI tools to APIs and cloud databases

Module XVI – Tableau for Advanced Data Analytics

- Tableau interface, workspace, and data connection setup
- Data blending vs data joinin
- Tableau Prep for ETL (cleaning, merging, reshaping data)
- Dimensions, measures, and calculated fields
- Table calculations and level of detail (LOD) expressions
- Dynamic parameters and user-driven interactivity
- Advanced visualizations (heatmaps, treemaps, waterfall, box plots, bullet charts)
- Creating multi-source dashboards (Excel + SQL + API data)
- Story points for narrative data storytelling

Module XVI – Tableau for Advanced Data Analytics

- Dashboard actions (filter, highlight, URL, and sheet swapping)
- Performance optimization and data extracts vs live connections
- Publishing dashboards to Tableau Server and Tableau Cloud

Module XVII – Advanced Excel & Power Query

- Combining data from multiple workbooks & sources
- Using Power Query for ETL (Extract, Transform, Load)
- Advanced formulas with dynamic arrays (FILTER, SORT, UNIQUE)
- Dashboard interactivity with form controls & buttons
- Forecasting using Excel functions (TREND, FORECAST, LINEST)
- Creating scenario analysis models and sensitivity analysis
- Automating recurring reports with Power Query refresh
- Linking Excel with Power BI for hybrid dashboards
- Building Excel-based analytics projects

Module XVIII – Predictive Analytics

- Introduction to predictive modeling in analytics
- Regression and classification fundamentals
- Feature selection and model evaluation
- Time-series forecasting (ARIMA, Prophet) basics
- Predictive dashboards using Power BI or Python integration
- Using AutoML for business forecasting
- Interpreting model output for decision-making
- Model bias and ethical use of predictions
- Case study: Sales forecasting using regression

Module XIX – Big Data for Analysts

- Introduction to big data concepts and architecture
- Data warehousing vs data lakes
- Overview of cloud data platforms: BigQuery, Snowflake, Redshift
- Writing SQL on large datasets efficiently
- ETL pipeline concepts (Extract, Transform, Load)
- Data partitioning and optimization
- Integrating BI tools with big data sources
- Data governance and access management
- Real-world use case: analyzing millions of rows with cloud SQL

Module XX – Data-Driven Decision Making

- Identifying and defining KPIs & metrics
- Root cause analysis and problem framing
- Business experimentation and A/B testing
- Decision frameworks: RICE, ICE, Pareto analysis
- Communicating insights with data storytelling
- Building executive summaries from analytics findings
- Aligning insights with business goals
- Measuring ROI and performance impact
- Turning analytics into actionable strategies

Module XXI – Automation & Scripting

- Automating Excel workflows with VBA macros
- Using Python scripts to refresh reports automatically
- Integrating Google Sheets with APIs
- Automating BI dashboards with Power Automate or Zapier
- Sending automated email reports
- Scheduling jobs with Task Scheduler or Airflow
- Building reusable analytics scripts
- Version control and collaboration with Git
- Minimizing manual effort through smart automation

Module XXII – Domain Analytics

- Marketing analytics: campaign tracking, ROI, funnel metrics
- Financial analytics: budgeting, forecasting, variance analysis
- HR analytics: employee attrition, engagement, hiring metrics
- Operations analytics: process optimization, cost reduction
- Sales analytics: pipeline performance, conversion ratios
- E-commerce analytics: customer segmentation, basket analysis
- Product analytics: feature adoption, retention, churn
- Supply chain analytics: demand forecasting and logistics
- Healthcare, education, fintech analytics example

Module XXIII – Machine Learning Foundations

- Introduction to ML workflow (train, validate, test)
- Supervised vs Unsupervised learning
- Bias-variance tradeoff and model generalization
- Model evaluation (cross-validation, metrics)
- Feature engineering & selection basics
- Regularization (L1, L2) and overfitting prevention
- Data preprocessing pipelines (scikit-learn Pipelines)
- Model interpretability (feature importance, SHAP values)
- ML lifecycle: from idea to deployed model

Module XXIV – Regression Algorithms

- Linear Regression deep dive and assumptions
- Polynomial, Ridge, Lasso, and ElasticNet Regression
- Non-linear regression modeling
- Regularization techniques and cross-validation
- Model evaluation metrics (RMSE, R^2, MAE, MAPE)
- Residual analysis and diagnostics
- Feature transformations for regression
- Case study: Predicting house prices, sales forecasting

Module XXV – Classification Algorithms

- Logistic Regression & decision boundaries
- Decision Trees, Random Forests, and Gradient Boosting (XGBoost, LightGBM, CatBoost)
- KNN, Naive Bayes, SVM — when to use each
- Confusion matrix, precision, recall, F1-score
- ROC curves, AUC, and threshold tuning
- Cross-validation and hyperparameter optimization (GridSearchCV, Optuna)
- Feature importance and model explainability
- Handling imbalanced datasets (SMOTE, class weights)
- Real-world example: customer churn prediction

Module XXVI – Unsupervised Learning

- Clustering algorithms (K-Means, DBSCAN, Hierarchical)
- Dimensionality reduction (PCA, t-SNE, UMAP)
- Association rule mining (Apriori, FP-Growth)
- Anomaly detection and outlier analysis
- Evaluation metrics for unsupervised tasks (silhouette score)
- Visualizing high-dimensional data
- Real-world use: market segmentation, recommendation systems
- Combining clustering with supervised models
- Feature extraction using unsupervised methods

Module XXVII – Deep Learning

- Neural network architecture & backpropagation
- Activation functions and optimization algorithms
- TensorFlow and PyTorch essentials
- CNNs for image classification
- RNNs, LSTMs, GRUs for sequential data
- Transfer learning and pre-trained models
- Batch normalization, dropout, and regularization
- Hyperparameter tuning for deep learning
- Real-world use: image recognition, NLP tasks

Module XXVIII – Natural Language Processing (NLP)

- Text preprocessing: tokenization, stemming, lemmatization
- Stopwords removal and text normalization
- Bag-of-Words, TF-IDF, and word embeddings (Word2Vec, GloVe, FastText)
- Named Entity Recognition (NER) and POS tagging
- Sentiment analysis and text classification
- Topic modeling (LDA, NMF)
- Transformers and BERT architecture
- Hugging Face pipelines and fine-tuning models
- Real-world NLP projects: chatbot, review sentiment, keyword extraction

Module XXIX – Time Series Analysis

- Time series components: trend, seasonality, noise
- Stationarity, autocorrelation, and ACF/PACF analysis
- ARIMA, SARIMA, SARIMAX models
- Facebook Prophet for forecasting
- Feature engineering for temporal data
- Handling irregular time intervals and missing timestamps
- Rolling windows and lag features
- Deep learning for time series (LSTM, Temporal CNNs)
- Real-world use: stock prediction, demand forecasting

Data Science Training Course

Schedule

Duration

9 Months

(36 Weeks)

Training Schedule

Weekdays Training

Monday to Friday

Weekend Training

Saturday & Sunday

Daily / Weekly Session

Live Online Instructor Led Session



Data Science Training Course Fee

Limited Time Offer

Currency	Code	Offer Price	Regular Price
US dollar	USD	1580	~~2080~~
Canadian dollar	CAD	2170	~~2670~~
Australian dollar	AUD	2260	~~2760~~
Sterling pound	GBP	1200	~~1700~~
New Zealand dollar	NZD	2720	~~3220~~
Indian rupee	INR	135,000	~~165,000~~

Fee is inclusive of all applicable taxes and include examination and certification fee.

No other hidden charges of any type.

Thank You!

Thank you for downloading our curriculum! We're excited to be part of your learning journey. Feel free to reach out if you need more information.

Contact Us

enquiries@proleed.academy

+91-9988475050

www.proleed.academy