

Certification in Data Science & AI

Session Plan

Duration: 300 Hours

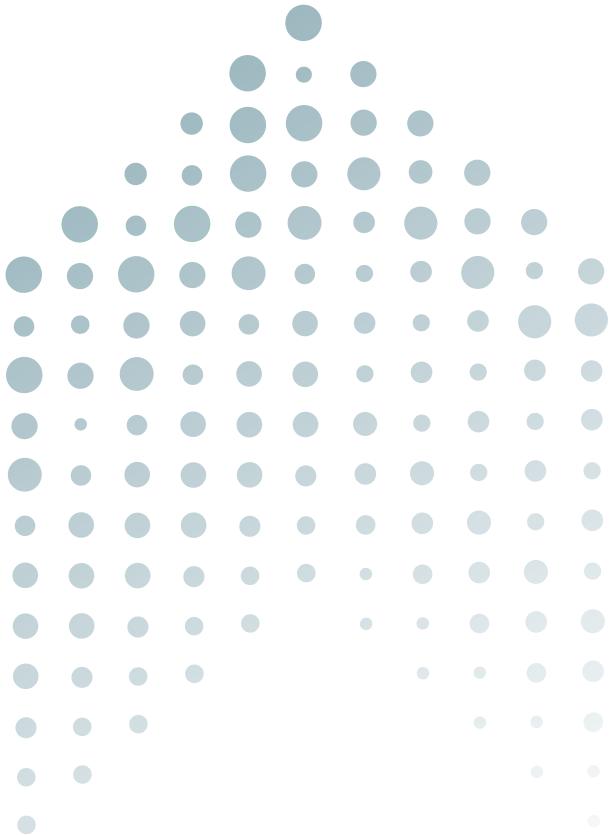


Table of Contents

S. No.	Modules	No. of Hours
1	Module 1: Excel-based EDA & Introductory Statistics	40
2	Module 2: Python Programming for Data Science	50
3	Module 3: Supervised Machine Learning (Track 1)	60
4	Module 4: Feature Engineering & Model Tuning (Track IV)	40
5	Module 5: Deep Learning (ANN, CNN, RNN)	70
6	Module 6: Capstone Projects & Industry Exposure	40

Module 1: Excel-based EDA & Introductory Statistics (Prep Course - Non-Programming)

Duration: **40 Hours**

Objective: **Build foundation in data understanding and statistical thinking using Excel.**

Understanding Data & Problem Statement

- What is data? Types (structured/unstructured)
- Understanding business problems and converting them to analytical problem statements
- Importance of data context in decision making

Categorical & Numerical Data Analysis

- Univariate vs multivariate analysis
- Measures: mean, median, mode, percentiles
- 3M analysis: Mean, Median, Mode on numerical features

Data Visualization in Excel

- Histograms, boxplots, bar charts
- Cross-tabulation & pivot tables
- Creating dashboards for summary views

Descriptive & Inferential Statistics

- Central tendency, dispersion, skewness, kurtosis
- Normal distribution, Bernoulli, binomial distributions
- Covariance vs correlation, causation

Hypothesis Testing

- Population vs sample, confidence intervals
- T-test, Chi-Square, ANOVA
- P-values and statistical significance

Week 1

Day	Session	Topic	Description	Dur (hrs)
Day 1	Session 1	Understanding Data & Problem Statement	Exploring types of variables and levels of measurement.	2
	Session 2		Understanding what constitutes data and its different types: structured vs unstructured.	2
Day 2	Session 3		Learning how to define a problem statement in analytics.	2
	Session 4		Discussing real-world scenarios and mapping them to data problems.	2
Day 3	Session 5		Understanding business objectives and converting them into measurable metrics.	2
	Session 6		Differentiating between nominal, ordinal, interval, and ratio scale variables	2
Day 4	Session 7	Categorical & Numerical Data Analysis	Introduction to univariate analysis for categorical data using frequency tables.	2
	Session 8		Analyzing numerical data using mean, median, and mode.	2
Day 5	Session 9		Working with percentiles, ranges, and IQR for numerical summaries.	2
	Session 10		Identifying skewness and kurtosis to describe distribution shapes & Comparing univariate vs multivariate analysis techniques.	2

Week 2

Day	Session	Topic	Description	Dur (hrs)
Day 1	Session 11	Data Visualization in Excel	Creating basic visualizations like histograms and bar charts in Excel.	2
	Session 12		Using boxplots and density plots to analyze distributions.	2
Day 2	Session 13		Applying crosstab and pivot tables for categorical data exploration.	2
	Session 14		Building visual summaries using conditional formatting and charts & Creating simple dashboards for data insights.	2
Day 3	Session 15	Descriptive & Inferential Statistics	Exploring measures of central tendency: mean, median, mode.	2
	Session 16		Understanding dispersion: variance, standard deviation, and range.	2
Day 4	Session 17		Introduction to probability distributions: normal, binomial, log-normal.	2
	Session 18		Analyzing covariance and correlation and interpreting relationships & Explaining correlation vs causation with case examples.	2
Day 5	Session 19	Hypothesis Testing	Population and sampling techniques, Central Limit Theorem and its implications in hypothesis testing.	2
	Session 20		Performing t-tests for means and interpreting p-values, ANOVA and Chi-square tests	2

Module 2: Python Programming for Data Science

Duration: **50 Hours**

Objective: **Equip students with Python programming and data wrangling skills.**

Python Basics

- Setting up Anaconda, Jupyter, Google Colab
- Syntax, variables, data types, loops, control structures

Functions & Data Structures

- User-defined functions, lambda expressions, scope
- Lists, sets, tuples, dictionaries, strings

Modules, Exceptions, Debugging

- Importing modules, standard libraries (os, sys, math)
- Try/except blocks, debugging tips and tools

NumPy & Pandas

- NumPy arrays, operations, broadcasting, slicing
- Pandas Series & DataFrames, loading/saving CSVs and Excel

Data Manipulation

- groupby, merge, pivot, melt
- Handling missing data, filtering, sorting

Basic Visualization

- Matplotlib: line, bar, scatter, histogram
- Seaborn: countplot, boxplot, heatmap, pairplot

Week 3

Day	Session	Topic	Description	Dur (hrs)
Day 1	Session 21	Python Basics	Setting up the Python environment using Anaconda and Jupyter Notebook.	2
	Session 22		Learning basic Python syntax, variables, and keywords.	2
Day 2	Session 23		Understanding data types, arithmetic operations, and expressions, Implementing conditional statements (if-else) and logical operations.	2
	Session 24		Setting up the Python environment using Anaconda and Jupyter Notebook, Working with syntax and keywords	2
Day 3	Session 25	Functions & Data Structures	Creating custom functions and understanding scopes.	2
	Session 26		Exploring Python lists: indexing, slicing, and operations.	2
Day 4	Session 27		Understanding dictionaries, tuples, and sets.	2
	Session 28		Using string manipulation functions and loops on data structures.	2
Day 5	Session 29		Creating custom functions and understanding scopes.	2
	Session 30		Understanding list operations and access methods	2

Week 4

Day	Session	Topic	Description	Dur (hrs)
Day 1	Session 31	Modules, Exceptions, Debugging	Importing Python modules and exploring standard libraries.	2
	Session 32		Understanding the purpose of packages and how to structure Python projects.	2
Day 2	Session 33		Using `sys` and `os` modules for system-level tasks, Handling errors using try-except blocks and debugging techniques.	2
	Session 34		Importing modules and common Python tools, Understanding the purpose of packages and how to structure Python projects.	2
Day 3	Session 35	NumPy & Pandas	Creating and manipulating arrays using NumPy, Understanding vectorized operations and broadcasting.	2
	Session 36		Reading and exploring data using Pandas DataFrames.	2
Day 4	Session 37		Applying filtering, selection, and aggregation on tabular data.	2
	Session 38		Building and editing NumPy arrays.	2
Day 5	Session 39	Data Manipulation	Grouping and summarizing data using groupby, Joining datasets using merge and concat functions.	2
	Session 40		Pivoting and reshaping data using pivot and melt.	2

Week 5

Day	Session	Topic	Description	Dur (hrs)
Day 1	Session 41	Data Manipulation	Cleaning missing values and handling duplicates & Applying groupby for data aggregation	2
Day 1	Session 42	Basic Visualization	Plotting data using Matplotlib: line, bar, and scatter plots.	2
Day 2	Session 43		Using Seaborn for enhanced visualizations: heatmap, boxplot, pairplot.	2
	Session 44		Customizing charts with labels, legends, and styles.	2
Day 3	Session 45		Creating multi-plot layouts for dashboards & Visualizing data using basic Matplotlib charts	2

*2nd session for Week 5 – Day 3 will start from next module, Refer module 3

Module 3: Supervised Machine Learning (Track 1)

Duration: **60 Hours**

Objective: **Introduce ML concepts, regression, classification, evaluation.**

ML Foundations & EDA

- Machine Learning vs AI vs Deep Learning
- Train-test split, data preprocessing recap
- EDA focused on predictive modeling

Linear Regression

- Simple and multivariate regression
- Assumptions: Linearity, homoscedasticity, multicollinearity
- Residual analysis, interpreting coefficients

Regularization Techniques

- Polynomial regression and overfitting
- Ridge and Lasso regression: bias-variance tradeoff
- Model selection using AIC, BIC

Logistic Regression & Classification

- Binary classification using logistic regression
- Model evaluation: confusion matrix, ROC-AUC
- Precision, recall, F1 score, accuracy

Week 5

Day	Session	Topic	Description	Dur (hrs)
Day 3	Session 46	ML Foundations & EDA	Understanding types of ML models and typical workflow in real-world scenarios, Splitting datasets into training and testing data.	2
Day 4	Session 47		Conducting exploratory analysis to identify data issues and relationships.	2
	Session 48		Basics of encoding categorical variables and scaling numeric features.	2
Day 5	Session 49		How supervised vs unsupervised learning apply in different use-cases.	2
	Session 50		Stratified and time-aware splitting strategies for robust model evaluation.	2

Week 6

Day	Session	Topic	Description	Dur (hrs)
Day 1	Session 51	ML Foundations & EDA	Using visualization to detect outliers and target imbalance, When to use label encoding vs one-hot encoding.	2
	Session 52		Common ML pipeline stages from raw data to model evaluation, Avoiding data leakage during train-test split.	2
Day 2	Session 53	Linear Regression	Linear regression fundamentals with single and multiple features, Checking linearity and equal variance assumptions in regression.	2
	Session 54		Plotting and interpreting residuals, R-squared, and RMSE, Identifying and mitigating the effect of outliers.	2
Day 3	Session 55		Multiple regression with feature interactions and transformations.	2
	Session 56		Using statistical tests to check regression assumptions.	2
Day 4	Session 57		Visual diagnostics with residuals and leverage plots.	2
	Session 58		Handling influence points using Cook distance.	2
Day 5	Session 59		End-to-end regression workflow from EDA to evaluation, Dealing with violations of normality and homoscedasticity.	2
Day 5	Session 60	Regularization Techniques	What is regularization and when to use it in ML.	2

Week 7

Day	Session	Topic	Description	Dur (hrs)
Day 1	Session 61	Regularization Techniques	Comparing Ridge and Lasso regression outcomes.	2
	Session 62		Cross-validation for tuning regularization strength.	2
Day 2	Session 63		Explaining AIC, BIC, and adjusted R-squared in model selection.	2
	Session 64		Fitting Ridge and Lasso with scikit-learn, Visualizing regularization impact on coefficients.	2
Day 3	Session 65		Using grid search with CV for hyperparameter tuning.	2
	Session 66		Model diagnostics using advanced performance metrics.	2
Day 4	Session 67		Use-case comparison of Ridge vs Lasso, Interpreting model complexity vs performance.	2
Day 4	Session 68	Logistic Regression & Classification	Applying logistic regression for binary classification tasks, Understanding the sigmoid function and log-odds logic.	2
Day 5	Session 69		Measuring performance using confusion matrix and ROC & Precision, recall, and F1-score interpretation.	2
	Session 70		Logistic regression pipeline for real-world data.	2

Week 8

Day	Session	Topic	Description	Dur (hrs)
Day 1	Session 71	Logistic Regression & Classification	Logistic regression pipeline for real-world data.	2
	Session 72		Decision thresholds and probability interpretation.	2
Day 2	Session 73		Explaining ROC-AUC and comparing classification models.	2
	Session 74		Using classification report for performance summary.	2
Day 3	Session 75		Advanced logistic regression with interaction terms, Sigmoid curve visualization and cut-off tuning.	2

*2nd session for Week 8 – Day 3 will start from next module, Refer module 4

Module 4: Feature Engineering & Model Tuning (Track IV)

Duration: **40 Hours**

Objective: Deep dive into preprocessing, feature transformation, and model optimization.

Feature Engineering

- Handling categorical variables: label encoding, one-hot
- Missing value imputation: mean, median, mode, KNN
- Outlier treatment and feature scaling

Dimensionality Reduction

- Feature selection using variance threshold, correlation matrix
- PCA (Principal Component Analysis) and its use cases

Model Tuning

- What are hyperparameters? Impact on learning
- GridSearchCV and RandomizedSearchCV with examples
- Cross-validation techniques: K-Fold, Stratified K-Fold

Model Selection

- Comparing models with metrics and complexity
- Bias-variance tradeoff visualization
- Learning curves and diagnostics

Week 8

Day	Session	Topic	Description	Dur (hrs)
Day 3	Session 76	Feature Engineering	Introduction to feature engineering and importance in model performance, Label encoding vs one-hot encoding with real examples.	2
Day 4	Session 77		Detecting and imputing missing values using mean, median, and KNN.	2
	Session 78		Outlier detection with boxplots, Z-score, and IQR.	2
Day 5	Session 79		Feature transformation: log, binning, and polynomial features, Scaling data with MinMaxScaler and StandardScaler.	2
	Session 80		Using visualization to detect outliers and target imbalance, When to use label encoding vs one-hot encoding & Feature construction: creating new features from raw data.	2

Week 9

Day	Session	Topic	Description	Dur (hrs)
Day 1	Session 81	Dimensionality Reduction	Why reduce dimensions? Benefits and risks, Applying PCA and interpreting explained variance.	2
	Session 82		Visualizing high-dimensional data using PCA/TSNE.	2
Day 2	Session 83		Combining feature selection with dimensionality reduction, Using correlation analysis to drop redundant features.	2
	Session 84		Building low-dimensional pipelines for model input & Understanding hyperparameters and how they affect models.	2
Day 3	Session 85	Model Tuning	Using GridSearchCV with scoring and parameter grids.	2
	Session 86		RandomizedSearchCV: faster tuning with randomness.	2
Day 4	Session 87		Cross-validation strategies: k-fold, stratified, and time series, Hyperparameter tuning for regularized models.	2
	Session 88		Tracking tuning results and managing overfitting.	2
Day 5	Session 89		Evaluating models post-tuning for robustness.	2
Day 5	Session 90	Model Selection	Introduction to model selection and why it's crucial.	2

Week 10

Day	Session	Topic	Description	Dur (hrs)
Day 1	Session 91	Model Selection	Comparing models with F1-score, accuracy, precision, and recall.	2
	Session 92		Understanding ROC-AUC and confusion matrix.	2
Day 2	Session 93		Bias-variance tradeoff and learning curves.	2
	Session 94		Model complexity vs generalization performance.	2
Day 3	Session 95		Choosing final model based on business and metric needs.	2

*2nd session for Week 10 - Day 3 will start from next module, Refer module 5

Module 5: Deep Learning (ANN, CNN, RNN, LSTM)

Duration: **70 Hours**

Objective: **Build and train deep learning models using Keras & TensorFlow.**

ANN Basics

- Neuron model, perceptron, multilayer perceptron (MLP)
- Feedforward and backpropagation
- Activation functions: sigmoid, ReLU, tanh, softmax

Keras & TensorFlow

- TensorFlow vs PyTorch overview
- Model creation in Keras: Sequential & Functional API
- Dropout, BatchNormalization, Optimizers

Convolutional Neural Networks (CNNs)

- Convolution layer, filters, padding, strides
- Max-pooling, feature maps, RGB image processing
- CNNs in Keras: MNIST case study, Transfer Learning

Recurrent Neural Networks (RNNs) & LSTMs

- Why RNNs? Sequential data processing
- LSTM/GRU vs simple RNNs
- Bidirectional RNNs, time series modeling

Week 10

Day	Session	Topic	Description	Dur (hrs)
Day 3	Session 96	ANN Basics	Understanding how artificial neurons mimic biological neurons.	2
Day 4	Session 97		Structure of perceptrons and multilayer perceptrons.	2
	Session 98		Forward propagation and activation functions, Exploring activation: sigmoid, ReLU, and tanh.	2
Day 5	Session 99		Loss functions and how they measure model performance, Gradient descent and weight updates.	2
	Session 100		Backpropagation: how errors are propagated, Building and training a basic neural network.	2

Week 11

Day	Session	Topic	Description	Dur (hrs)
Day 1	Session 101	ANN Basics	Overfitting and regularization in neural networks.	2
	Session 102		Introduction to batch normalization, Applying dropout and its effect on training.	2
Day 2	Session 103		Case study: classification with MLP using real-world data.	2
Day 2	Session 104	Keras & TensorFlow	Getting started with Keras and TensorFlow.	2
Day 3	Session 105		Keras Sequential and Functional APIs.	2
	Session 106		Compiling models: loss functions and optimizers.	2
Day 4	Session 107		Training and evaluating models in Keras, Model callbacks: EarlyStopping and ModelCheckpoint.	2
	Session 108		Saving and loading models, Custom layers and model debugging.	2
Day 5	Session 109		Monitoring with TensorBoard, Hyperparameter tuning in Keras, Transfer learning basics in TensorFlow.	2
	Session 110		Hands-on mini-project using Keras and TensorFlow.	2

Week 12

Day	Session	Topic	Description	Dur (hrs)
Day 1	Session 111	Convolutional Neural Networks (CNNs)	Why use CNNs for image data.	2
Day 1	Session 112		Convolution operations and feature detection, Understanding filters, padding, and strides.	2
Day 2	Session 113		Pooling layers and their purpose.	2
	Session 114		Building a CNN from scratch using Keras.	2
Day 3	Session 115		Activation functions in CNNs.	2
	Session 116		Data augmentation for better generalization.	2
Day 4	Session 117		CNN regularization techniques.	2
	Session 118		CNN architecture: LeNet, AlexNet overview.	2
Day 5	Session 119		Implementing transfer learning in CNNs.	2
	Session 120		Using pretrained models: VGG16, ResNet & Visualizing CNN filters and feature maps.	2

Week 13

Day	Session	Topic	Description	Dur (hrs)
Day 1	Session 121	Recurrent Neural Networks (RNNs) & LSTMs	Why standard neural networks fail on sequences.	2
	Session 122		How RNNs work: architecture and flow.	2
Day 2	Session 123		Training RNNs and vanishing gradient problem.	2
	Session 124		Introducing LSTM and GRU networks.	2
Day 3	Session 125		Architecture of LSTM cells.	2
	Session 126		Bidirectional RNNs and their advantages.	2
Day 4	Session 127		Stacked RNNs and time series modeling.	2
	Session 128		Embedding layers and sequence preprocessing.	2
Day 5	Session 129		Sequence classification with LSTM in Keras.	2
	Session 130		LSTM for text generation & Sentiment analysis using LSTM.	2

Module 6: Capstone Projects + Industry Exposure

Duration: **40 Hours**

Objective: **Apply all concepts to real-world problems and prepare for job market.**

Capstone Planning

- Choosing the right problem statement
- Defining scope, timeline, dataset sources

Implementation

- EDA, data preprocessing, model selection
- Training, evaluation, fine-tuning

Presentation & Review

- Building a project report & visual dashboard
- Demoing models using Streamlit or notebooks
- Peer review and mentor feedback

Career Readiness

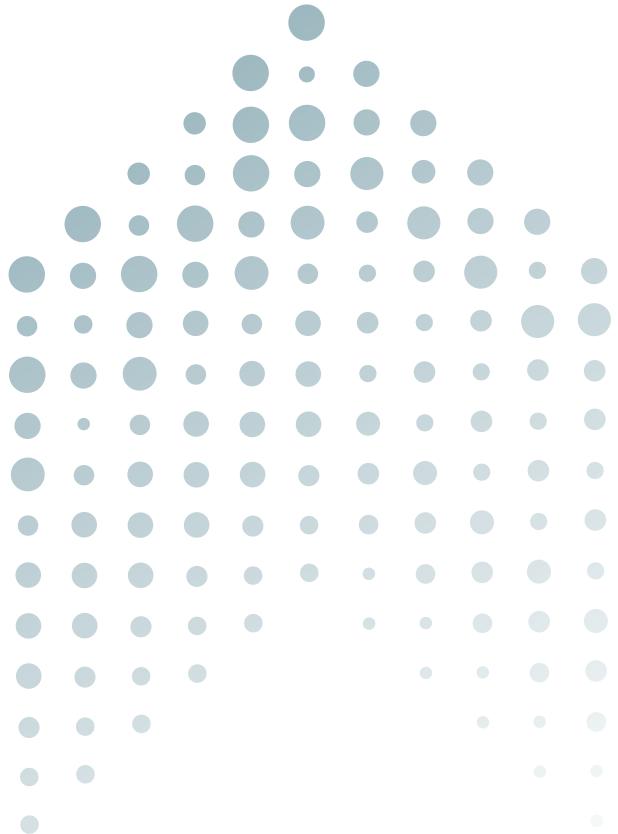
- Mock interviews, ML & Python questions
- Resume & LinkedIn profile building
- Guest sessions by industry experts

Week 14

Day	Session	Topic	Description	Dur (hrs)
Day 1	Session 131	Capstone Planning	Identifying a real-world problem suitable for a capstone project.	2
	Session 132		Defining objectives, KPIs, and problem scope, Dataset sourcing and feasibility assessment.	2
Day 2	Session 133		Outlining timelines, tools, and technology stack, Risk identification and mitigation planning.	2
	Session 134		Structuring capstone proposal and milestone tracking & Performing EDA to derive initial insights from the dataset.	2
Day 3	Session 135	Implementation	Cleaning and transforming data for modeling, Choosing appropriate ML/DL models for the problem.	2
	Session 136		Training models and evaluating initial performance, Refining models through hyperparameter tuning.	2
Day 4	Session 137		Ensembling models for improved accuracy.	2
	Session 138		Documenting modeling choices and performance metrics & Developing dashboards or interfaces for presentation.	2
Day 5	Session 139	Presentation & Review	Conducting validation and peer review of results.	2
	Session 140		Creating a structured capstone report with visuals.	2

Week 15

Day	Session	Topic	Description	Dur (hrs)
Day 1	Session 141	Presentation & Review	Storytelling with data and final recommendations.	2
	Session 142		Live demo preparation using notebooks or Streamlit.	2
Day 2	Session 143		Presenting to mentors or faculty for feedback.	2
	Session 144		Final project polish: packaging code and results.	2
Day 3	Session 145	Career Readiness	Scoring, evaluation, and showcasing best projects.	2
	Session 146		Resume building for Data Science and AI careers.	2
Day 4	Session 147		Preparing a strong LinkedIn and GitHub profile.	2
	Session 148		Crafting your capstone story for interviews.	2
Day 5	Session 149		Mock interviews: technical and behavioral.	2
	Session 150		Networking tips and how to approach hiring managers.	2



Business Toys Pvt. Ltd.

#360, Krishna Kaveri Tower 3rd Floor, Double Road,
Geleyara Balaga Layout, Chikkabanavara, Bengaluru,
Karnataka(India)- 560 090