



Certification in **Data Science**

Technologies covered R & Python
5 Months | Live online | 144 hours



Module 1: Introduction to Data Science

1. What is analytics & Data Science?
2. Common Terms in Analytics
3. Analytics vs. Data warehousing, OLAP, MIS Reporting
4. Relevance in industry and need of the hour
5. Types of problems and business objectives in various industries
6. Overview of analytics tools & their popularity
7. List of steps in Analytics projects
8. Identify the most appropriate solution design for the given problem statement
9. Project plan for Analytics project & key milestones based on effort estimates
10. Build Resource plan for analytics project
11. Why R and Python for data science?

Module 2: Introduction to R programming

1. What is R
2. What is S
3. History of R
4. Features of R
5. SAS versus R

Module 3: Installing and managing R

1. Installing R
2. Packages
3. Input/output
4. R interfaces
5. R Library

Module 4: R programming essentials

1. Basic operations in R
2. Different data types and data structures in R
3. Sub setting in R
4. Additional topics on data structures
5. Importing data sets in R
6. R loops and special functions
7. Calculation of commission and simple interest
8. Plots and charts in R
9. Merging and sorting functions in R
10. Summarising Data
11. Calculations of the measures of central tendency and measures of variability



Module 5: R Analytics- Testing of hypothesis

1. Concept of hypothesis
2. Null hypothesis
3. Alternative hypothesis
4. Type-I error
5. Type-II error
6. Level of Significance
7. Confidence Interval
8. Parametric Tests and Non Parametric Tests
9. One Sample T test
10. Two independent sample T test
11. Paired Sample T test
12. Chi square Test for Independence of Attributes.

Module 6: Analysis of variance

1. One Way Anova
2. Two Way Anova

Module 7: Exploratory Factor Analysis

1. Principal Component Analysis
2. Estimating the Initial Communalities
3. Eigen Values and Eigen Vectors
4. Correlation Matrix check and KMO-MSA check
5. Factor loading Matrix
6. Diagrammatic Representation of Factors
7. Problems of Factor Loadings and Solutions

Module 8: Cluster Analysis

1. Types of Clusters
2. Metric and linkage
3. Ward's Minimum Variance Criteria
4. Semi-Partial R-Square and R-Square
5. Diagrammatic Representation of clusters
6. Problems of Cluster Analysis



Module 9: Linear Regression and Multiple Linear Regression

1. Concept of Regression and features of Linear line.
2. Assumptions of Classical Linear Model
3. Method of Least Squares
4. Understanding the Goodness of Fit
5. Test of Significance of The Estimated Parameters
6. Multiple linear Regression with their Assumptions
7. Concept of Multocollinearity
8. Signs of Multicollinearity
9. The Idea Of Autocorrelation

Module 10: Logistic Regression

1. Concept and Applications of Logistic Regression
2. Principles Behind Logistic Regression
3. Comparison between Linear probability Model and Logistic Regression
4. Mathematical Concepts related to Logistic Regression
5. Concordant Pairs, Discordant Pairs and Tied Pairs
6. Classification Table
7. Graphical Representation Related to logistic Regression.

Module 11: Decision Tree & Time Series Analysis

1. Decision Tree
2. Concept of Time Series and its Applications
3. Assumptions of Time Series Analysis
4. Components of Time Series
5. Smoothing techniques
6. Stationarity
7. Random Walk
8. ARIMA Forecasting
9. Box Jenkins Technology
10. Merits and Demerits of BJ Technology

Module 12: Text mining and Market basket analysis

Module 13: Introduction to Python programming

1. Introduction to python
2. Installation of Anaconda

Module 14: Python variable & Conditional statements

1. Python Variable
2. Understanding more about variable
3. User input variable
4. Calculate discounted price
5. Conditional Statements



Module 15: Python functions

1. Python string functions
2. Python Numeric function

Module 16: Loops in Python

1. Introduction to Loop
2. FOR loop for Sum
3. FOR loop for multiplication
4. While loop

Module 17: Conditional statement & Python lists

1. Conditional statement(if elif else)
2. More of if else elif
3. Introduction to python list
4. Indexing of Python List
5. slicing of list
6. modifying a list
7. List methods

Module 18: Python Tuple & Python dictionary

1. Python Tuple
2. Functions on Tuple
3. Python Dictionary
4. Functions on dictionary
5. Methods of Dictionary

Module 19: Python Set and more of Python

1. Python Set
2. Create your own functions
3. Classes in python
4. Inheritance in Python
5. Time management function

Module 20: Introduction to Numpy

1. Introduction to numpy
2. Array in Numpy
3. Matrices using Numpy
4. Mathematical functions using numpy



Module 21: Introduction to Panda

1. Introduction to panda
2. Data frame in pandas
3. Group-by in pandas

Module-22: Linear Regression

1. Recap of linear regression theory
2. Application of linear regression using Python Library

Module 23: Logistic Regression

1. Recap of Logistic regression theory
2. Application of logistic regression using Python Library

Module 24: Support Vector Machine (SVM)

1. SVM theory: Linear, Dual SVM and Kernel Trick
2. SVM code along

Module 25: Decision Tree & Dimensionality Reduction

1. Curse of dimensionality
2. Principal Component Analysis
3. Singular Value Decomposition
4. Independent Component Analysis
5. Fisher Linear Discriminate Analysis

Module 26: Clustering

1. k-Means Clustering: Theory and Code-along
2. Hierarchical Clustering: Theory and Code-along
3. Gaussian Mixture Model: Theory and Code-along

Module 27: Project

