Data Science with R

Introduction **Duration 5 days**

What are the course objectives?

The Data Science with R has been designed to give you in-depth knowledge of the various data analytics techniques that can be performed using R. The data science course is packed with real-life projects and case studies, and includes R Studio Lab for practice.

- **Mastering R language:** The data science course provides an in-depth understanding of the R language, R-studio, and R packages. You will learn the various types of apply functions including DPYR, gain an understanding of data structure in R, and perform data visualisations using the various graphics available in R.

- **Mastering advanced statistical concepts:** The data science training course also includes various statistical concepts such as linear and logistic regression, cluster analysis and forecasting. You will also learn hypothesis testing.

Objectives:

This data science course will enable you to:

- Gain a foundational understanding of business analytics
- Install R, R-studio, and workspace setup, and learn about the various R packages
- Master R programming and understand how various statements are executed in R
- Gain an in-depth understanding of data structure used in R and learn to import export data in R
- Define, understand and use the various apply functions and DPLYR functions
- Understand and use the various graphics in R for data visualisation
- Gain a basic understanding of various statistical concepts
- Understand and use hypothesis testing method to drive business decisions
- Understand and use linear, non-linear regression models, and classification techniques for data analysis
- Learn and use the various association rules and Apriori algorithm
- Learn and use clustering methods including K-means, DBSCAN, and hierarchical clustering

Who can Attend This Course?

There is an increasing demand for skilled data scientists across all industries, making this data science course well-suited for participants at all levels of experience. We recommend this Data Science training particularly for the following professionals:

- IT professionals looking for a career switch into data science and analytics

- Software developers looking for a career switch into data science and analytics
- Professionals working in data and business analytics
- Graduates looking to build a career in analytics and data science
- Anyone with a genuine interest in the data science field
- Experienced professionals who would like to harness data science in their fields

Prerequisites:

There are no prerequisites for this data science online training course. If you are new in the field of data science, this is the best course to start with.

Day 1

Module 01- Introduction to Business Analytics

Introduction
Objectives
Need of Business Analytics
Business Decisions
Introduction to Business Analytics
Features of Business Analytics
Types of Business Analytics
Business Decisions
Business Intelligence (BI)
Data Science
Importance of Data Science
Data Science as a Strategic Asset
Big Data
Analytical Tools
Quiz
Summary

Module 02 - Introduction to R

Introduction
Objectives
An Introduction to R
Comprehensive R Archive Network (CRAN)
Cons of R
Companies Using R
Understanding R
Installing R on Various Operating Systems
Installing R on Windows from CRAN Website
Demo - Install R
Install R
IDEs for R
Install RStudio
Steps in R Initiation
Functions and Help in R
R Packages
Installing an R Package
Quiz
Summary

Module 03 - R Programming

Introduction
Objectives
Operators in R
Arithmetic Operators
Relational Operators
Logical Operators
Assignment Operators
Conditional Statements in R
Ifelse() Function
Switch Function
Loops in R
Break Statement
Next Statement
Scan() Function
Running an R Script
Running a Batch Script
R Functions
Use Commonly Used Functions
Demo - Use String Functions
Quiz
Summary

Day 2

Module 04 - R Data Structure

Introduction
Objectives
Types of Data Structures in R
Vectors
Matrices
Accessing Matrix Elements
Arrays
Data Frames
Elements of Data Frames
Factors
Lists
Importing Files in R
Read Data from a File
Exporting Files from R
Quiz
Summary

Module 05 - Apply Functions

Introduction
Objectives

Types of Apply Functions
Apply() Function
Lapply() Function
Sapply() Function
Tapply() Function
Vapply() Function
Mapply() Function
Dplyr Package - An Overview
Dplyr Package - The Five Verbs
Installing the Dplyr Package
Use the Select Function
Functions of Dplyr-Package - Filter()
Functions of Dplyr Package - Arrange()
Use Arrange Function
Functions of Dplyr Package - Mutate()
Functions of Dply Package - Summarise()
Use Summarise Function
Quiz
Summary

Module 06 - Data Visualization

Introduction
Objectives
Graphics in R
Types of Graphics
Bar Charts
Pie Charts
Histograms
Kernel Density Plots
Line Charts
Box Plots
Heat Maps
Word Clouds
File Formats for Graphic Outputs
Saving a Graphic Output as a File
Exporting Graphs in RStudio
Exporting Graphs as PDFs in RStudio
Demo - Save Graphics Using RStudio
Quiz
Summary

Day 3

Module 07 - Introduction to Statistics

Introduction
Objectives
Basics of Statistics
Types of Data
Qualitative vs. Quantitative Analysis

Types of Measurements in Order
Statistical Investigation
Normal Distribution
Use Probability Distribution Functions
Distance Measures
Euclidean Distance
Manhattan Distance
Minkowski Distance
Mahalanobis Distance
Cosine Similarity
Correlation
Correlation Measures Explained
Pearson Product Moment Correlation (PPMC)
Pearson Correlation - Case Study
Dist() Function in R
Demo - Perform the Distance Matrix Computations
Perform the Distance Matrix Computations
Quiz
Summary

Module 08 - Hypothesis Testing

Introduction
Objectives
Hypothesis
Need of Hypothesis Testing in Businesses
Null Hypothesis
Alternate Hypothesis
Null vs. Alternate Hypothesis
Chances of Errors in Sampling
Types of Errors
Contingency Table
Decision Making
Critical Region
Level of Significance
Confidence Coefficient
Bita Risk
Power of Test
Factors Affecting the Power of Test
Types of Statistical Hypothesis Tests
Upper Tail Test
Test Statistic
Factors Affecting Test Statistic
Critical Value Using Normal Probability Table
Quiz
Summary

Module 09 - Hypothesis Testing II

Introduction
Objectives
Parametric Tests

Z-Test
T-Test
Use Normal and Student Probability Distribution Functions
Testing Null Hypothesis
Objectives of Null Hypothesis Test
Three Types of Hypothesis Tests
Hypothesis Tests About Population Means
Decision Rules
Chi-Square Test
Steps of Chi-Square Test
Important Points of Chi-Square Test
Degree of Freedom
Use Chi-Squared Test Statistics
Introduction to ANOVA Test
One-Way ANOVA Test
The F-Distribution and F-Ratio
F-Ratio Test
F-Ratio Test in R - Example
Perform ANOVA
Quiz
Summary

Day 4

Module 10 - Regression Analysis

Introduction
Objectives
Introduction to Regression Analysis
Use of Regression Analysis - Examples
Types Regression Analysis
Simple Linear Regression Model
Correlation
Find Correlation
Method of Least Squares Regression Model
Coefficient of Multiple Determination Regression Model
Standard Error of the Estimate Regression Model
Dummy Variable Regression Model
Interaction Regression Model
Non-Linear Regression Models
Demo - Perform Regression Analysis with Multiple Variables
Perform Regression Analysis with Multiple Variables
Non-Linear Models to Linear Models
Algorithms for Complex Non-Linear Models
Quiz
Summary
Conclusion

Module 11 - Classification

Introduction
Objectives
Introduction to Classification
Examples of Classification
Classification vs. Prediction
Issues Regarding Classification and Prediction
Data Preparation Issues
Evaluating Classification Methods Issues
Decision Tree
Classification Rules of Trees
Overfitting in Classification
Tips to Find the Final Tree Size
Basic Algorithm for a Decision Tree
Statistical Measure - Information Gain
Enhancing a Basic Tree
Demo - Model a Decision Tree
Model a Decision Tree
Naive Bayes Classifier Model
Bayesian Theorem
Naive Bayes Classifier
Naive Bayes Classifier - Advantages and Disadvantages
Nearest Neighbor Classifiers
Computing Distance and Determining Class
Choosing the Value of K
Support Vector Machines
Geometric Margin in SVMs
Linear SVMs
Non-Linear SVMs
Demo - Support a Vector Machine
Support a Vector Machine
Quiz
Summary

Module 12 - Clustering

Introduction
Objectives
Introduction to Clustering
Clustering vs. Classification
Use Cases of Clustering
Clustering Models
K-means Clustering
Pseudo Code of K-means
K-means Clustering - Case Study
Perform Clustering Using Kmeans
Hierarchical Clustering
Requirements of Hierarchical Clustering Algorithms
Agglomerative Clustering Process
Perform Hierarchical Clustering
DBSCAN Clustering

Quiz
Summary

Day 5

Module 13 - Association

Introduction
Objectives
Association Rule Mining
Application Areas of Association Rule Mining
Parameters of Interesting Relationships
Association Rules
Limitations of Support and Confidence
Apriori Algorithm
Step 1 - Mine All Frequent Item Sets
Algorithm to Find Frequent Item Set
Finding Frequent Item Set - Example
Ordering Items
Candidate Generation
Step 2 - Generate Rules from Frequent Item Sets
Perform Association Using the Apriori Algorithm
Demo - Perform Visualization on Associated Rules
Problems with Association Mining
Quiz
Summary
Conclusion