

Data Analytics with R Certification Training

Course Fees – 15,000/-

Why should you take Data Analytics with R course?

- R is the most popular data analytics tool owing to it being open-source, its flexibility, packages and community
- "R" wins on Statistical Capability, Graphical capability, Cost, rich set of packages and is the most preferred tool for Data Scientists.
- The average salary for a Senior Data Scientist skilled in R is \$123k (Pay scale salary data)

Course Curriculum

Introduction to Data Analytics

- **Learning Objectives** - This module introduces you to some of the important keywords in R like Business Intelligence, Business Analytics, Data and Information. You can also learn how R can play an important role in solving complex analytical problems. This module tells you what is R and how it is used by the giants like Google, Facebook, Bank of America, etc. Also, you will learn use of 'R' in the industry, this module also helps you compare R with other software in analytics, install R and its packages.
- **Topics** - Introduction to terms like Business Intelligence, Business Analytics, Data, Information, how information hierarchy can be improved/introduced, understanding Business Analytics and R, knowledge about the R language, its community and ecosystem, understand the use of 'R' in the industry, compare R with other software in analytics, Install R and the packages useful for the course, perform basic operations in R using command line, learn the use of IDE R Studio and Various GUI, use the 'R help' feature in R, knowledge about the worldwide R community collaboration.

Introduction to R Programming

- **Learning Objectives** - This module starts from the basics of R programming like datatypes and functions. In this module, we present a scenario and let you think about the options to resolve it, such as which datatype should one to store the variable or which R function that can help you in this scenario. You will also learn how to apply the 'join' function in SQL.
- **Topics** - The various kinds of data types in R and its appropriate uses, the built-in functions in R like: seq(), cbind(), rbind(), merge(), knowledge on the various subsetting methods, summarize data by using functions like: str(), class(), length(), nrow(), ncol(), use of functions like head(), tail(), for inspecting data, Indulge in a class activity to summarize data, dplyr package to perform SQL join in R

Data Manipulation in R

- **Learning Objectives** - In this module, we start with a sample of a dirty data set and perform Data Cleaning on it, resulting in a data set, which is ready for any analysis. Thus using and exploring the popular functions required to clean data in R.
- **Topics** - The various steps involved in Data Cleaning, functions used in Data Inspection, tackling the problems faced during Data Cleaning, uses of the functions like grepl(), grep(), sub(), Coerce the data, uses of the apply() functions.

Data Import Techniques in R

- **Learning Objectives** - This module tells you about the versatility and robustness of R which can take-up data in a variety of formats, be it from a csv file to the data scraped from a website. This module teaches you various data importing techniques in R.
- **Topics** - Import data from spreadsheets and text files into R, import data from other statistical formats like sas7bdat and spss, packages installation used for database import, connect to RDBMS from R using ODBC and basic SQL queries in R, basics of Web Scraping.

Exploratory Data Analysis

- **Learning Objectives** - In this module, you will learn that exploratory data analysis is an important step in the analysis. EDA is for seeing what the data can tell us beyond the formal modeling or hypothesis. You will also learn about the various tasks involved in a typical EDA process.
- **Topics** - Understanding the Exploratory Data Analysis(EDA), implementation of EDA on various datasets, Boxplots, whiskers of Boxplots. understanding the cor() in R, EDA functions like summarize(), llist(), multiple packages in R for data analysis, the Fancy plots like the Segment plot, HC plot in R.

Data Visualization in R

- **Learning Objectives** - In this module, you will learn that visualization is the USP of R. You will learn the concepts of creating simple as well as complex visualizations in R.
- **Topics** - Understanding on Data Visualization, graphical functions present in R, plot various graphs like tableplot, histogram, Boxplot, customizing Graphical Parameters to improvise plots, understanding GUIs like Deducer and R Commander, introduction to Spatial Analysis.

Data Mining: Clustering Techniques

- **Learning Objectives** - This module lets you know about the various Machine Learning algorithms. The two Machine Learning types are Supervised Learning and Unsupervised Learning and the difference between the two types. We will also discuss the process involved in 'K-means Clustering', the various statistical measures you need to know to implement it in this module.
- **Topics** - Introduction to Data Mining, Understanding Machine Learning, Supervised and Unsupervised Machine Learning Algorithms, K-means Clustering.

Data Mining: Association Rule Mining & Collaborative filtering

- **Learning Objectives** - In this module, you will learn how to find the associations between many variables using the popular data mining technique called the "Association Rule Mining", and implement it to predict buyers' next purchase. You will also learn a new technique that can be used for recommendation purpose called "Collaborative Filtering". Various real-time based scenarios are shown using these techniques in this module.
- **Topics** - Association Rule Mining, User Based Collaborative Filtering (UBCF), Item Based Collaborative Filtering (IBCF)

Linear and Logistic Regression

- **Learning Objectives** - This module touches the base of 'Regression Techniques'. Linear and logistic regression is explained from the basics with the examples and it is implemented in R using two case studies dedicated to each type of Regression discussed.
- **Topics** - Linear Regression, Logistic Regression.

Anova and Sentiment Analysis

- **Learning Objectives** - This module tells you about the Analysis of Variance (Anova) Technique. The algorithm and various aspects of Anova have been discussed in this module. Additionally, this module also deals with Sentiment Analysis and how we can fetch, extract and mine live data from Twitter to find out the sentiment of the tweets.
- **Topics** - Anova, Sentiment Analysis.

Data Mining: Decision Trees and Random Forest

- **Learning Objectives** - This module covers the concepts of Decision Trees and Random Forest. The algorithm for creation of trees and classification of decision trees and the various aspects like the Impurity function Gini Index, Pruning, Entropy etc are extensively taught in this module. The

algorithm of Random Forests is discussed in a step-wise approach and explained with real-life examples. At the end of the class, these concepts are implemented on a real-life data set.

- **Topics** - Decision Tree, the 3 elements for classification of a Decision Tree, Entropy, Gini Index, Pruning and Information Gain, bagging of Regression and Classification Trees, concepts of Random Forest, working of Random Forest, features of Random Forest, among others.

Project Work

- **Learning Objectives** - This module discusses various concepts taught throughout the course and their implementation in a project.
- **Topics** - Analyze census data to predict insights on the income of the people, based on the factors like: age, education, work-class, occupation using Decision Trees, Logistic Regression and Random Forest. Analyze the Sentiment of Twitter data, where the data to be analyzed is streamed live from twitter and sentiment analysis is performed on the same.