



About Data Science with R

The Data Science with R course focuses on imparting in-depth knowledge of various techniques for data analytics using R. The course includes real-life projects, case studies, and includes R CloudLabs for practice.

Master R language: The 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 visualizations using the various packages available in R.

Master advanced statistical concepts: Learn various statistical concepts like linear and logistic regression, cluster analysis, and forecasting with predictive data analytics. You will also learn hypothesis testing techniques.

Case Study Based Project: As a part of the course, you will be required to execute real-life projects using CloudLab. There are four case study based projects.

Course Objectives

This course will enable you to:

- Gain a foundational understanding of business analytics
- Install R, R-studio, and workspace setup. You will also learn about the various R packages
- Master the 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 visualization
- Gain a basic understanding of the 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



Curriculum

Module 1: Introduction to the Concepts of Data Analytics and R

Module 2: R Data structures

Module 3: Using R for Statistics

Module 4: Data Visualization

Module 5: Data Manipulation

Module 6: Introduction to Machine Learning, DRT

Module 7: Clustering, MBA

Module 8: Linear, Logistic regression with R

Module 9: Decision Tree, Naive Bayes

Module 10: SVM, NN, Ensemble Learning

Module 11: Project Explanation

Module 12: Time series Analysis with R

Module 13: Text Analytics with R

Instructor

Instructors are handpicked from a selected group of industry experts and mentors and trained to deliver the best online learning experience. All instructors have at least ten years of industry experience and extensive functional expertise in the field they train.

Certification

- The assessment will be done on the basis of an online test and Project Evaluation at the end of the course.
- On completion of the Assessment (Project + Test) with a minimum of 70% marks, a certificate of successful completion will be issued from NIIT.



Pre-requisites

Learner should have basic knowledge of statistics and computer programming. Preferably a reasonable level of proficiency in any data handling tool like MS Excel.

FAQs

What is the Case Study based Project which I need to develop as part of this program?

Project Specifications

Background: A finance company deals in all home loans. They have presence across all urban, semi urban and rural areas. Customer first applies for home loan after that company validates the customer eligibility for loan.

Problem: Company wants to automate the loan eligibility process (real time) based on customer detail provided while filling online application form. These details are Gender, Marital Status, Education, Number of Dependents, Income, Loan Amount, Credit History and others. To automate this process, they have given a problem to identify the customers segments; those are eligible for loan amount so that they can specifically target these customers. You are required to use your “R” skills to deliver a model for predicting loan eligibility of a customer. In other words, you will tell the company by your solution that whether customer applied for loan will default or not in future!

Why should I join this course?

The term “data scientist” is the most happening and the hottest job title in the IT field. Starting salaries match upto \$100,000, which is 22% more than other job roles (Source: <http://www.burtchworks.com>). R programming is the best tool for most jobs that involve data. It is the most popular programming language that data scientists use. It is in heavy use at several large organizations like Microsoft, Google and Facebook. R programming in academia is important because it creates a pool of talent that feeds the data industry. The best way to learn R is by doing. This course is designed to master various R programming concepts for data analytics by implementing various real-life and industry-based projects.