**What’s the focus of this course?**

The Data Science with R training course has been designed to impart an in-depth knowledge of the various data analytics techniques which can be performed using R. The course is packed with real-life projects, case studies, and includes several for practice.  
  
**Mastering 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 graphics available in R.  
  
**Mastering advanced statistical concepts**:   The course also includes the various statistical concepts like linear and logistic regression, cluster analysis, and forecasting. You will also learn hypothesis testing, Data Visualization, Text data mining, Machine Learning,   
  
As a part of the course, you will be required to execute real-life projects. The compulsory projects are spread over four case studies in the domains of healthcare, retail, and Internet. Tech Vision has been provided to ensure a practical and hands-on experience.  Additionally, we have four more projects for further practice.

**What are the 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 DPLYP 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

**Day 1: Background, Getting Started, and Programing Concepts (4-5 hours)**

Day 1 convers the basics to get started up with R, people who want to start learning R from the beginner's level. We begin with the basics, discussing different programming constructs and structures of R, and move towards more advanced topics. We also give easy-to-follow examples to let you grasp the concepts introduced.

### R Introduction

* Overview
* R - Environment Setup
* R - Reserved Words
* R - Data Types
* R - Variables and Constants
* R - Operators

### R Decision and Loops

* R - Programming if...else
* R - ifelse() Function
* R - Programming for loop
* R - while Loop
* R - break & next
* R - repeat Loop

### R Functions

* R - Functions
* Function Return Value
* R - Environment & Scope
* R - Recursive Function
* R - Switch Function
* R - String Handling

### R Vector & Data Frame

* R - Vectors
* R - Lists
* R - Matrices
* R - Arrays
* R - Factors
* R - Data Frames
* R - Packages
* R - Data Reshaping

**Day 2: Data Import/Export (ETL), and Visualization (4-5 hours)**

Day 2 covers ETL and Data Visualization. ETL to read data from files or other DBMS/RDBMS sources and stored outside the R environment. We can also write data into files which will be stored and accessed by the operating system. R can read and write into various file formats like csv, excel, xml etc.

Also Covers numerous libraries to create charts and graphs. A pie-chart is a representation of values as slices of a circle with different colors. The slices are labeled and the numbers corresponding to each slice is also represented in the chart.

### R Data Import/Export (ETL)

* R - CSV Files
* R - Excel Files
* R - Binary Files
* R - XML Files
* R - Web Data
* R - Database

**R Basic Data Visualization**

* R - Pie Charts
* R - Bar Charts
* R - Boxplots
* R - Histograms
* R - Line Graphs
* R - Scatterplots

**R Advanced Data Visualization**

* Heat Map
* Mosaic Map
* Map Visualization
* 3D Graphs
* Correlogram

**Day 3: Statistics (4 hours)**

Day 3 covers Statistical analysis in R is performed by using many in-built functions. Most of these functions are part of the R base package. These functions take R vector as an input along with the arguments and give the result.

* R - Mean, Median & Mode
* R - Linear Regression
* R - Multiple Regression
* R - Logistic Regression
* R - Normal Distribution
* R - Binomial Distribution
* R - Time Series Analysis

**Day 4: Statistics - Extended & Text Mining (4-5 hours)**

**R Statistics – Extended**

* R - Nonlinear Least Square
* R - Decision Tree
* R - Random Forest
* R - Survival Analysis
* R - Chi Square Tests

**R Text Mining**

* [Loading Texts](https://rstudio-pubs-static.s3.amazonaws.com/31867_8236987cf0a8444e962ccd2aec46d9c3.html#loading-texts)
* [Preprocessing](https://rstudio-pubs-static.s3.amazonaws.com/31867_8236987cf0a8444e962ccd2aec46d9c3.html#preprocessing)
  + [Stage the Data](https://rstudio-pubs-static.s3.amazonaws.com/31867_8236987cf0a8444e962ccd2aec46d9c3.html#stage-the-data)
* [Explore your data](https://rstudio-pubs-static.s3.amazonaws.com/31867_8236987cf0a8444e962ccd2aec46d9c3.html#explore-your-data)
  + [Focus!](https://rstudio-pubs-static.s3.amazonaws.com/31867_8236987cf0a8444e962ccd2aec46d9c3.html#focus)
  + [Word Frequency](https://rstudio-pubs-static.s3.amazonaws.com/31867_8236987cf0a8444e962ccd2aec46d9c3.html#word-frequency)
  + [Plot Word Frequencies](https://rstudio-pubs-static.s3.amazonaws.com/31867_8236987cf0a8444e962ccd2aec46d9c3.html#plot-word-frequencies)
* [Relationships Between Terms](https://rstudio-pubs-static.s3.amazonaws.com/31867_8236987cf0a8444e962ccd2aec46d9c3.html#relationships-between-terms)
  + [Term Correlations](https://rstudio-pubs-static.s3.amazonaws.com/31867_8236987cf0a8444e962ccd2aec46d9c3.html#term-correlations)
  + [Word Clouds!](https://rstudio-pubs-static.s3.amazonaws.com/31867_8236987cf0a8444e962ccd2aec46d9c3.html#word-clouds)
* [Clustering by Term Similarity](https://rstudio-pubs-static.s3.amazonaws.com/31867_8236987cf0a8444e962ccd2aec46d9c3.html#clustering-by-term-similarity)
  + [Hierarchal Clustering](https://rstudio-pubs-static.s3.amazonaws.com/31867_8236987cf0a8444e962ccd2aec46d9c3.html#hierarchal-clustering)
  + [K-means clustering](https://rstudio-pubs-static.s3.amazonaws.com/31867_8236987cf0a8444e962ccd2aec46d9c3.html#k-means-clustering)

## Day 5: Data Science & Machine Learning with R

This course provides advanced-level training on Machine Learning applications and algorithms. It will give you hands-on experience in multiple, highly sought-after machine learning skills in both supervised and unsupervised learning. This machine learning training ensures you can apply machine learning algorithms like regression, clustering, classification,  and recommendation. The unique case study approach ensures you are working hands-on with data while you learn.

* Loading the dataset.
* Summarizing the dataset.
* Visualizing the dataset.
* Evaluating some algorithms.
* Making some predictions.

**Exam & certification FREE PRACTICE TEST**

**How to get certified?**

To become a Certified Data Scientist with R, you must fulfill the following criteria:

* + Complete any one project out of the four provided in the course. Submit the deliverables of the project.
  + Score a minimum of 80% in any one of the two simulation tests
  + Complete 85% of the course

Note:

* + When you have completed the course, you will receive a certificate for implementing the projects using R.
  + It is mandatory that you fulfill both the criteria i.e., completion of any one project and clearing the online exam with minimum score of 80%, to become a certified data scientist.