



Data Analytics with R

Data Analytics, in today's world, has opened up a plethora of opportunities for industries and businesses across the world. It has made it possible to find answers to solutions, uncover trends and patterns that were impossible to have been noticed otherwise.

Travel and hospitality, Retail, Healthcare, Government, Banking and Financial sectors are some of the industries that find wide use of Data Analytics. R is a great open source tool that offers the most comprehensive statistical analysis packages. This course will teach you to use R for Big data analytics while exploring and exploiting the features that this tool has to offer!

Why R for Data Analytics?

- Like we said, the wide statistical analysis package collection including standard statistical tests, models and analysis make it a great choice with Data Analysts.
- Flexibility in managing and manipulating data.
- R offers excellent graphical capability.
- Integration with different programming languages like Java, Ruby, C++, Python.

Why should you learn R?

- R runs on all platforms – Windows, PC, Mac, Linux and more.
- R makes statistics easy to learn and fun
- Learn R and become the in-demand Data Scientist, now!

Course Duration

- 30 Working Days

Curriculum

Introductory Session

This will be an introduction session with a brief explanation about Data Analytics scope of this field and introduction to R platform.

- Briefing about Analytics domain
- How insights from data can help business solve day-to-day problems and find solution
- Various platforms which can help you in the journey of becoming Data Scientist
- Introduction to R as a platform

R Nuts and Bolts

This session will be an introduction to Basics of coding on R Studio platform

- Understanding different windows of R Studio
- Basics of R Programming and some important rules for coding in R
- Installing predefined packages
- Entering inputs and R objects (Vector, Matrix, Dataframes and Factors)
- R Datatypes
- Using dplyr package
- Text Manipulations using Strings
- Reading data (csv file) in R

DATA MANIPULATIONS AND LOOPING IN R

In-depth understanding about data manipulation using different packages and functions & conditional loopings in R.

- Subsetting dataset
- Date and Time in R
- Loops: while & for
- Conditionals: if-else
- Functions: Defining functions, Anonymous functions
- Apply family of functions
- Sampling in R

EXPLORATORY ANALYSIS IN R

Exploratory Analysis will help you know more about the features of datasets, statistically. For understanding real-time data in the industry, this is the first step

- Central Tendencies
- Measurements of Dispersion
- Test of Normality
- Null Value Treatment
- Outlier Treatment
- Correlation Analysis
- Reshaping Data
- Merging Data

R Studio Visualisations

Creating basic as well as interactive visualisation in R.

- Categorical Data: Barplot, Pie Chart
- Numeric: Boxplot, Histogram, Scatter Plot, Line Chart
- Using different libraries to make graph presentable (ggplot2, Rcolorbrewer)

INFERENCEAL ANALYSIS IN R

Inferential Analysis is very useful in knowing underline information of data. It is generally used in the industry for A/B or Test/Control group comparisons

1) Parametric Statistical Tests

- Basic theory of Inferential Statistics
- Hypothesis tests using Z Test
- T-statistics Test
- Two sampled Z Test and T Test
- ANOVA
- Post-hoc Test

2) Non-Parametric Statistical Tests

- Wilcoxon Test
- Mann-Whitney U Test

- K.S. Test
- Runn Test
- Chi-Square Test

DATA LOADING AND FILE FORMATS

This section begins with loading and bringing data from different data sources in R.

Descriptive Statistical Analysis

- Data loading and file formats
- Loading JSON files
- XML and HTML Web Scraping
- Interacting with HTML and Web APIs
- Interacting with databases
- Text Mining/Text Analytics in R

MACHINE LEARNING

Introduction to machine learning and its further bifurcations. Learning most of the industry-wise used machine learning techniques.

- What is Machine Learning
- Machine Learning real-world examples
- Assumptions for Linear Regression

SUPERVISE LEARNING TECHNIQUES

- Linear Regression Assumptions checks in R
- Building Linear Regression Model in R
- Stepwise Method

Case Study- Linear Regresssion

- Exploring Data
- Dividing data into Test and Train
- Model Building and R
- Predicting on Test Data using model

Logistic Regression

- Understanding Logistic Regression

- Classification Model Building using Logistic Model
- Confusion Matrix

Random Forest

- Decision Tree
- Random Forest

SVM and Naive Bayes

- SVM
- Naïve Bayes

UNSUPERVISED LEARNING TECHNIQUES

Unsupervised Learning

- Clustering
- K-means
- Hierarchical Clustering
- Time Series Analysis

