CERTIFICATION IN
DATA ANALYTICS USING R

Course Code : OCIT0009

R is a programming language software environment for statistical computing and graphics supported by the R Foundation for Statistical Computing. The R language is widely used among statisticians and data miners for developing statistical software and data analysis. It is an approach of data analysis employed for summarizing and visualizing data set, and the focus of the approach is to analyze data's basic structures and variables to develop a basic understanding of the data set. in this data science certification course, students will learn data exploration, data visualization, predictive analytics and descriptive analytics techniques with the R language

Curriculum

Module 1: Introduction to R Environment

History and development of R Statistical computing programming language, installing R and R studio, getting started with R, creating new working directory, changing existing working directory, understanding the different data types, installing the available packages, calling the installed packages, arithmetic operations, variable definition in R, simple functions, vector definition and logical expressions, matrix calculation and manipulation using matrix data types, workspace management.

Module 2: Data Structures and Control Statements

Introduction to different data types, vectors, atomic vectors, types and tests, coercion, lists, list indexing, function applying on the lists, adding and deleting the elements of lists, attributes, name and factors, matrices and arrays, matrix indexing, filtering on matrix, generating a covariance matrix, applying function to row and column of the matrix, data frame-creating, coercion, combining data frames, special types in data frames, operations in data frame, applying functions: lapply ( ) and supply ( ) on data frames, control statements, loops, looping over non vector sets, arithmetic and Boolean operators and values, branching with if, looping with for, if-else control structure, looping with while, vector based programming.

Module 3: I/O Operations and String Manipulations

Introduction to I/O functions in R, accessing I/O devices, using of scan( ), readline ( ) function, comparison and usage of scan and readline function, reading different format files into R: text file, CSV file, Statistical package files, xls and xlsx files, reading data frame files, converting from one format to another using in built function, writing different file format in to the local machine directory, getting file directory information, accessing the internet : overview of TCP/IP, sockets in R, implementation of parallel R, basics of string manipulations – grep ( ), nchar ( ), paste( ), sprintf( ), substr( ), regexpr( ), strsplit( ), testing of file name with given suffix.

Module 4: R for Summary and Parametric Tests

Descriptive statistics – summery statistics for vectors, making contingency tables, creating contingency tables from vectors, converting objects in to tables, complex flat tables, making 'Flat' contingency tables, testing tables and flat table objects, cross tables, testing cross tabulation, recreating original data from contingency tables, switching class, mean (arithmetic, geometric and harmonic) median, mode for raw and grouped data, measure of dispersion-range, standard deviation, variance, coefficient of variation, testing of hypothesis – small sample test, large sample test – for comparing mean, proportion, variance (dependent and independent samples) correlation and regression – significance of correlation and regression coefficients.

Module 5: R for Graphs, Non parametric Tests and ANOVA

Introduction To Graphs, Box-Whisker Plot, Scatter Plots, Pairs Plots, Line Chart, Pie Chart, Cleveland Dot Charts, Bar Charts, Customization Of Charts, Non Parametric Tests: The Wilcoxon U-Test (Mann-Whitney: One And Two Sample U-Test, Tests For Association: Chi Square Tests.

Learning Outcomes

- Illustrate the concept the importance of multivariate analysis in data analysis
- Recognise the importance of classification rule mining in data mining business and engineering.
- Understand the importance of R in customising the analytics in
- Understand the different options in I/O operations in R programming
- Understand the basic concepts of statistical functions in R for the analysis.