R Programming

Intermediate Level

Duration: 1 Day (9am – 4pm)

Course Pre-Requisites: Knowledge of R Programming at a beginner level is recommended for this course.

Course Overview

This intermediate R course dives deep into essential data science techniques. Students will refresh their understanding of R's core data types and structures, and master data import from various sources. They'll learn advanced data manipulation using the **tidyverse**, work with relational data, and explore initial data analysis techniques. By the course's end, students will be adept at data transformation, visualisation, and basic modelling in R.

1. Introduction

In the Introduction section, students will delve into a review of R data types, structures, and the common syntax for accessing data in data frames. This will equip them with the foundational knowledge of R's data handling capabilities.

2. Importing Data

In the 'Importing Data' section, students will explore importing data in RStudio, the various packages and functions to import data into R, and hands-on techniques for importing data from text files (like CSV) and Excel. This will provide them with the necessary tools to load datasets into R for analysis.

3. Workflow in R

In this unit, students will learn about creating reusable scripts, enhancing the reproducibility and efficiency of their R projects.

4. Manipulating Data

Students in this section will dive deep into the **tidyverse**, data summarisation, ordering, and date manipulations. They will also explore string operations, row and column manipulations

in data frames, and techniques to handle missing or erroneous data. This section will strengthen their data wrangling skills in R.

5. Working with Relational Data

In this section, students will uncover the methods to add new variables to a data frame from another source, apply mutating joins and merges, and filter joins. They'll also gain proficiency in exporting their manipulated data to files.

6. Basic Exploratory Data Analysis

This section of the training will guide students in choosing the right chart for their goals and data and embark on univariate analysis of numeric variables. This will give them a strong foundation in visualising and understanding the characteristics of their datasets.

7. Univariate Analysis

This section delves into the foundational aspects of data exploration. Students will learn to assess the overall distribution of data, focusing on central tendency measures like mean, median, and mode. The course will also introduce concepts related to the spread of data, such as range and standard deviation. Emphasis will be placed on identifying and handling outliers. Additionally, participants will gain insights into the shape of data distributions, enabling them to distinguish between different types of distributions such as normal, skewed, or bimodal.

8. Visual Representation of Distributions

This section delves into graphical methods for data visualisation. Students will master histograms for frequency views and boxplots for central tendencies and outliers. They'll also explore dot charts, stem and leaf plots, and the use of bar and column charts for categorical data representation.

9. Multivariate Analysis

In this module, students will dive into techniques for analysing multiple variables simultaneously. They'll learn to visualise relationships using scatterplots and the comprehensive scatterplot matrix. The study of correlations will elucidate the strength and direction of linear relationships between variables. Bar, column, and line charts will be covered to effectively represent varied data types..

10. Basic Modelling

This section introduces students to predictive modelling in R. They'll begin by constructing a linear model and then delve into evaluating its efficacy. Key aspects such as underlying assumptions of the model will be discussed. The culmination of this module equips students to make informed predictions using their crafted model.