R PROGRAMMING

Credits: 4 **Semester:IV**

SubjectCode:DS18402 **No. of Lecture Hours :60**

Objectives:

- To learn the statistical programming language R and to use it to manipulate data
- To reshape data to support different analysis and to explore data from a variety of sources by building inferential models, generating charts, graphs and other data representations.

Outcome: Students will be able to

CO1: Understand basic concept of R**.**
CO2: Demonstrate programming concepts and data structures in R.
CO3: Analyze a large problem by sub dividing it into smaller components using functions
CO4: Choose an appropriate graphic for analysis and analyze data using summary statistics.
CO5: Choose the type of regression based on data set.

UNIT – I

12hrs

1. **Introduction to Data Science's** in Various Fields, Impact of DS 1
2. Major Activities, Toolkit, Data Scientist, Data Science Team 2
3. **Introduction to R:** What is R, Features of R, Simple Math, R as a Calculator 2
4. **Reading and Getting Data into R:** Reading Data, Reading a File of Data from a Disk, Reading Bigger Data Files 3
5. The read.csv() Command, Alternative Commands for Reading Data in R 2
6. Saving Data Files to Disk and Reading Data from the Disk 2

UNIT – II

12hrs

1. **Series and Control Statements:** Assignment, Modes, Operators, Basic Functions, Generating Data sets, Control Structures 2
2. **Vectors:** Definition, Declaration, Generating, Indexing, Naming, Adding, And Removing Elements 5
3. **Operations on Vectors:** Recycling, Special Operators, Functions for Vectors Missing Values, Null Values, Filtering and Sub setting 5

UNIT – III

12hrs

1. **Data Structures in R-Arrays:** Creating Arrays, Dimensions and Naming, Indexing and Naming, Functions on Arrays 2
2. **Matrices:** Creating Matrices, Adding Rows/Columns, Removing Rows/Columns Reshaping Operations, Special Functions 2
3. **Lists:** Creating, Naming, Accessing Elements, Adding, Removing, Special Functions, Recursive Lists 2
4. **Data Frames:** Creating, Naming, Accessing, Adding, Removing, Special Functions, Merging Exercises 2
5. **Functions:**Creating, Functions on Function Object, Scope of Variable, Accessing Global Environment, Closures, Recursion, Creating New Binary Operator 4

UNIT-IV **12hrs**

1. **Descriptive Statistics:** Introduction, Descriptive Statistics, Central Tendency Variability, Mean, Median, Range, Variance, Summary, Exercises 6
2. **Graphics:** Introduction, Types, Packages, Basic Graph, Histograms, Stem Leaf Graph, Box Plots, Bar Plots 6

UNIT – V **12hrs**

1. **Linear Regression:** Inferential Statistics, Types of Learning, Linear Regression, Simple Linear Regression, Coefficients, Confidence Interval, RSE, R2 Implementation in R, lm, Functions on 1m, Predict, Plotting, Fitting Regression Line Exercises 6
2. **Multiple Linear Regression:** Introduction, Comparison with Simple Linear Regression, Correlation Matrix, F Statistic, Response Vs Predictors, Deciding Important Variables, Model Fit, Predictions, Generating a Model Interactive terms, Non Linear Transformations, ANOVA, lm with Polynomial Exercises 6