**R Tools for Data Management and Analytics**

**Course Objectives:**

|  |  |
| --- | --- |
| This Course will enable the students to: | |
| 1. | Learn the fundamentals of R programming language. |
| 2. | Understand the data structures and functions for data manipulation. |
| 3. | Understand the concepts of data preprocessing and data wrangling. |
| 4. | Learn how to visualize the data using basic and advance tools like ggplot2 |
| 5. | Learn how R is easy to work with Continuous & Categorical response variable, Free text data. |

|  |  |
| --- | --- |
| **UNIT– I** | **07 Hours** |
| **Introduction:**  R, R Studio Introduction, Installation of R, Basic features of R Programming, Getting Started – Window Sections of RStudio, First Interaction, Command line versus Scripts, Comments, Help in R.  Variables in R – Naming Variable, Assigning Values to Variable, Finding Variable, Removing Variables. Input of Data – Input of Data from Terminal, Input of Data through R-Object. Output in R – using Print() function, cat() functions. Packages in R – Standard packages, Contributed Packages, Downloading & Installing R Packages, Importing Packages.  Inbuilt Functions in R – Mathematical, Logarithmic, Date & Time, Sequence, Repeat and String Functions. | |

|  |  |
| --- | --- |
| **UNIT– II** | **08 Hours** |
| **Data Types & Structures --** Basic Data types in R, **Vectors** – Class of a Vector, Elements of a Vector, Accessing Vector Elements, Functions for Vectors. **Matrices** – Creating a Matrix, Accessing Matrix Elements, Functions for Matrices. Array – Creating a List, Accessing List Elements, Functions for List. **Arrays –** Creating an array, Accessing Elements of an Array, Functions for Array. **Lists** – Creating List, Accessing List Elements, Functions for List. **Data Frame** – Creating a DataFrame, Accessing Data & Data from Data Frame, Functions for Data Frame.  **Decision-Making Structures** – If Structures, Switch statement.  **Loops** – For Loop, While Loop and Repeat Loop. | |

|  |  |
| --- | --- |
| **UNIT– III** | **08 Hours** |
| **Functions** – Functions without Arguments, Function with Arguments, **User defined functions** **–** creating functions with or without arguments, calling functions.  **Data Exploration and Manipulation -** Missing Data Management – Determining missing Data, Excluding Missing Values, Missing Data Imputation.  **Special Functions across Data Elements** – Data management using Apply functions, Data management with dplyr package.  **Import and Export Data** – Importing & Exporting data in Text, Excel, CSV, JSON file format in R. | |

|  |  |
| --- | --- |
| **UNIT– IV** | **08 Hours** |
| **Basic Visualization:**  Pie chart, Bar chart, Histograms, Line Chart, Box-and Whisker Plot, Scatter plot. ggplot2 for advance graphics. Perform basic visualization for CSV/Excel data.  **Basic Statistics:** Descriptive statistics – Measure of central tendency, variability, Quantile, Rank, Skewness and Kurtosis using R. Tables in R – Creating Tables, Marginal distribution, Calculation of Proportions,Outlier detections and management.  **Viewing data:** Display data in Terminal, Display data in rich format like Excel and various option to sort and filter. | |

|  |  |
| --- | --- |
| **UNIT– V** | **08 Hours** |
| **Correlation and Regression –** Correlation**,** Simple and Multiple Linear regression using Ordinary list square (OLS), Data preparation, accessing coefficients, accessing model accuracy (Residual standard error, R squared, F-statistics) in OLS summary. Accessing OLS model visually, Making predictions.  **Classification –** Logistic Regression, Why Logistic Regression**,** Data preparation, accessing coefficient, Making predictions, Model evaluation and diagnostics (Likelihood ratio, Pseudo R square, Validation of Predicted Values.  **Text Mining** – Introduction to Text based analysis, importing text file, reading data from websites, perform sentiment analysis for customer feedback. | |

**TEXTBOOKS:**

|  |  |
| --- | --- |
| 1. | Bharati Motwani, “Data Analytics With R”, Wiley |
| 2. | Gareth James , “An introduction to statistical learning with applications in R”, Springer; 2nd ed. 2021 |

**REFERENCEBOOKS:**

|  |  |
| --- | --- |
| 1. | Jared P. Lander, “R For everyone”  Pearson Education; Second edition (15 April 2018) |
| 2. | Hadley Wickham, “R for Data Science: Import, Tidy, Transform, Visualize, and Model Data” O'Reilly; First Edition |
| 3. | Garrett Grolemund, “Hands on Programming With R: Write Your Own Functions and Simulations” Shroff/O'Reilly; First Edition (1 January 2014) |

**WEBLINKS:**

|  |  |
| --- | --- |
| 1. | https://uc-r.github.io/ |
| 2. | https://www.bigbookofr.com/index.html |
| 3. | https://ds4ps.org/cpp-526-fall-2019/textbook/ |
| 4. | https://www.w3schools.com/r/s |
| 5. | https://cognitiveclass.ai/courses/r-101 |

**Course Outcomes:**

|  |  |
| --- | --- |
| **After the completion of this course, students will be able to:** | |
| CO1 | ***Interpret*** the R programming language fundamentals. |
| CO2 | ***Demonstrate*** the R data structures for data manipulation, *apply* decision control and loop techniques on Text, CSV/Excel data. |
| CO3 | ***Develop*** the functions and get familiar with apply features, techniques to handle missing data. |
| CO4 | ***Develop***the models to visualizethe data and ***apply*** the statistics techniques. |
| CO5 | ***Demonstrate*** the techniques of data preprocessing and data wrangling, predictive analytics, text mining. |