

**I. Learning Outcomes**

After learning this course the student will be able

1. Be comfortable using commercial and open source tools such as the R language and its associated libraries for data analytics and visualization.
2. Learn skills to analyze real time problems using R
3. Able to use basic R data structures in loading, cleaning the data and preprocessing the data.
4. Able to do the exploratory data analysis on real time datasets
5. Able to understand and implement Linear Regression
6. Able to understand and use - lists, vectors, matrices, data frames, etc

II. Syllabus**Unit – 1: Computer basics**

Basic applications of computer, components of computer system, Central Processing Unit (CPU), input and output units, computer memory and mass storage devices. Programming languages and their applications. Concept of files and folders. Software and types of software. Operating System like Windows and Linux.

Unit – 2: Data processing

Data processing using spreadsheets – Data entry and editing features in Excel, copy, paste, paste special options, sort and filter options, auto sum, steps of finding average and standard deviation of data using statistical functions. Matrix operations like transpose, multiply and inverse using Excel functions. Simple graphs like bar chart, line chart and pie chart in Excel. Exporting Excel output to word processors like MS-Word and slide presentations like Power Point.

Unit – 3:

Scatter diagram, fitting of straight line, polynomial and power curves using Excel – Reading R-square value and equation from the graph. Predicting future values using ‘forecast’ and ‘trend’ functions. Data Analysis Pak and its features. Performing Student’s t-test and one- way Analysis of Variance using Data Analysis Pak. P-value and its interpretation.

Unit – 4: R Programming

Introduction to R, Features of R – Environment – R Studio. Basics of R-Assignment - Modes - Operators - special numbers - Logical values - Basic Functions - R help functions - R Data Structures - Control Structures. Vectors: Definition- Declaration - Generating - Indexing - Naming - Adding & Removing elements - Operations on Vectors - Recycling - Special Operators - Vectorized if- then else-Vector Equality – Functions for vectors - Missing values - NULL values - Filtering & Subsetting.

Unit – 5:

Matrices - Creating Matrices, Adding or Removing rows/columns, Operations. Creating Data Frames, Naming, Accessing, Adding, and Removing, Applying Special functions to Data Frames, Merging Data Frames Factors and Tables.

Exploratory Data Analysis – Descriptive Statistics – Central Tendency - Variability - Mean - Median - Range - Variance - Summary - Handling Missing values and Outliers - Normalization Data Visualization in R : Types of visualizations - packages for visualizations - Basic Visualizations, Advanced Visualizations and Creating 3D plots.

**Practical Syllabus**

1. Installing R and R studio
2. Create a folder DS_R and make it a working directory. Display the current working directory
3. installing the "ggplot2", "caTools", "CART" packages
4. load the packages "ggplot2", "caTools".
5. Basic operations in r
6. Working with Vectors:
 - a) Create a vector v1 with elements 1 to 20.
 - b) Add 2 to every element of the vector v1.
 - c) Divide every element in v1 by 5.
 - d) Create a vector v2 with elements from 21 to 30. Now add v1 to v2.
7. Using the data present in the table given below, create a Matrix "M"

	C1	C2	C3	C4	C5
C1	0	12	13	8	20
C2	12	0	15	28	88
C3	13	15	0	6	9
C4	8	28	6	0	33
C5	20	88	9	33	0

Find the pairs of cities with shortest distance.

8. Consider the following marks scored by the 6 students

Section	Student	M1	M2	M3
A	1	46	54	45
A	2	34	55	55
A	3	56	66	64
B	1	43	44	45
B	2	67	76	78
B	3	76	68	37

- a) create a data structure for the above data and store in proper positions with proper names
 - b) display the marks and totals for all students
 - c) Display the highest total marks in each section.
 - d) Add a new subject and fill it with marks for 2 sections.
9. Three people denoted by P1, P2, P3 intend to buy some rolls, buns, cakes and bread. Each of them needs these commodities in differing amounts and can buy them in two shops S1, S2. The individual prices and desired quantities of the commodities are given in the following table

	Price						
	S1	S2		Demand.quantity			
Roll	1.5	1		Roll	Bun	Cake	Bread
Bun	2	2.5	P1	6	5	3	1
Cake	5	4.5	P2	3	6	3	2
Bread	16	17	P3	3	4	3	1

- a) Create matrices for above information with row names and col names.
 - b) Display the demand. quantity and price matrices
 - c) Find the total amount to be spent by each person for their requirements in each shop
 - d) Suggest a shop for each person to buy the products which is minimal.
10. Applying summary() to find the mean, median, standard deviation, etc
 11. Implementation of Visualizations - Bar, Histogram, Box, Line, scatter plot, etc.