22CA3E4: APPLIED DATA ANALYTICS

Course Name	Applied Data Analytics		L	T	P	C	CIA	SEE	TM
Course Code	22CA3E4		4	0	0	4	30	70	100
Year of Introduction: 2022	Year of Offering: 2022	Year of Revision: 2022			Percentage of Revision: NIL				
L-Lecture, T-Tutorial, P-Practical, C-Credits, CIA-Internal Marks, SEE-External Marks, TM-Total Marks									

Course Description and Purpose: Applied Data Analytics is a course that illustrates concepts of R-Programming, Data Structures, Descriptive Statistical Analysis, Basic Graphs, Analysis of ANOVA, Multivariate Analysis, Files & Databases.

Course Objective: The Applied Data Analytics course aims to equip students with a comprehensive understanding of R-Programming, Data Structures, Descriptive Statistical Analysis, Basic Graphs, Analysis of ANOVA, Multivariate Analysis, and Files & Databases, fostering proficiency in applying these concepts to solve real-world problems through machine learning algorithms and techniques.

Specific Objectives include:
▪ To *install*, *code* and *use* R Programming Language in R Studio IDE to perform basic tasks on Control Flow Statements, Data Structures and can invoke Operations on Data Structures.
▪ To understand the *Basic Terminologies*, *Concepts and Techniques* employed in Descriptive Statistical Analysis.
▪ To familiar with *Basic Graphics* and *Analysis of* ANOVA.
▪ To gain knowledge on *Basic Multivariate Analysis*.
▪ To apply how to import *Different Files* and *Connecting Databases to R*.

Course Outcomes:
CO1: Upon completing the course students will gain a comprehensive proficiency in utilizing R for data analysis, mastering R environment, working with packages, understanding, manipulating, and cleaning diverse datasets, employing various data types and structures, handling missing values, sorting and merging data, subsetting datasets, implementing control flow statements, and performing aggregation and restructuring operations, empowering them to apply advanced data analysis techniques for solving complex real-world problems.

CO2: Upon completion of the course, students will acquire a comprehensive understanding of measures of central tendency, dispersion, and shapes, various sampling techniques, hypothesis testing methods including parametric and non-parametric tests, enabling them to effectively analyse and interpret data, make informed decisions, and contribute meaningfully to statistical research and applications.

CO3: Upon completion of the course on "Basic and Advanced Data Visualization, and Analysis of Variance," students will proficiently create a wide array of graphical representations using bar plots, pie charts, histograms, line plots, dot plots, kernel density plots, and utilize advanced visualization techniques with the ggplot2 package, while also mastering the application of various ANOVA models, including one-way ANOVA, one-way ANCOVA, two-way factorial ANOVA, repeated measures ANOVA, and multivariate analysis of variance (MANOVA), enabling them to visually and statistically rrange complex datasets and draw meaningful insights for research and decision-making purposes.

CO4: Upon completion of the course on "Basic Multivariate Analysis, Time Series Analysis, and Forecasting," students will gain a comprehensive understanding of regression techniques including simple linear regression, multiple linear regression, and logistic regression, along with proficiency in time series analysis encompassing the creation and decomposition of time series, exponential models, and forecasting methods such as simple moving averages, weighted moving averages, and single

exponential smoothing, empowering them to rrange multivariate data, model time-dependent patterns, and make accurate predictions for diverse real-world applications.

CO5: Upon completion of the course on "Connecting R to External Interfaces," students will proficiently import and export data between R and various external sources including CSV files, Microsoft Excel spread sheets, databases (MySQL) for creating, querying, and managing tables, XML and JSON files for structured data exchange, as well as binary files, enabling them to seamlessly interface R with diverse data formats and sources for effective analysis and manipulation.

UNIT-I (12 Hours)

Introduction to R: Why use R?, R Environment, Working with R Packages, Understanding Datasets, Data Types, Data Structures (Operations on Data Structures), Missing Values, Sorting Data, Merging Datasets, Subsetting Datasets, Control Flow Statements, Aggregation and Restructurings.

UNIT-II (12 Hours)

Descriptive Statistics: Introduction to Descriptive Statistics (Measures of Central Tendency, Measures of Dispersion of Variability, Measures of Shapes (Skewness and Kurtosis)), Introduction to Sampling(Sampling Types), Hypothesis Testing with R(One Sample Test, One Sample Sign Test, Two Samples Test), Parametric Test(Correlations, Z-Test, T-Test), Non Parametric Tests (Wilcoxon Signed- Rank Test, Chi Square Test).

UNIT-III (12 Hours)

Basic Graphs: Bar Plots, Pie Charts, Histograms, Line, Dot Plots, Kernel Density Plots and Dot Plots.
The Advanced Graphics: The ggplot2 Package.
Analysis of Variance: Fitting ANOVA Models, One-way ANOVA, One-way ANCOVA, Two-way factorial ANOVA, Repeated Measures ANOVA, Multivariate Analysis of Variance (MANOVA)

UNIT-IV (12 Hours)

Basic Multivariate Analysis: Regression (Simple Linear Regression, Multiple Linear Regression, Logistic Regression), Time Series Analysis (Creating Time Series, Components of Time Series Analysis, Seasonal Decomposition, Exponential Models), Forecasting (Simple Moving Averages, Weighted Moving Averages, Single Exponential Smoothing.)

UNIT-V (12 Hours)

Connecting R to External Interfaces: CSV Files (Reading From a CSV File, Writing to a CSV File), Microsoft Excel (Reading from XLSX File, Writing to XLSX File), Databases (Connecting R to MYSQL ,Creating Tables, Inserting Rows, Updating Rows, Deleting Rows, Querying Rows, Querying Tables, Dropping Tables), XML Files (Reading From XML Files, JSON Files, Reading From JSON Files), Binary Files (Writing to Binary Files, Reading From Binary Files).

Prescribed Text Book			
	Author	Title	Publisher
1	Dr. Rob Kabacoff	R in Action : Data Analysis and Graphics with R. [UNIT-I ,UNIT-II ,UNIT-III]	Manning Publications Co, Edition 2011.
2	Dr.Jeeva Jose	A Beginners Guide For Data Analysis Using R Programming. (UNIT IV and UNIT V) UNI IV: Chapter-11 11.3 [11.3.1 to 11.3.3] 11.5,11.6 [11.6.1 to 11.6.3] UNIT V: Chapter-6 [6.1 to 6.6]	Khanna Book Publishing Co.(P) Ltd, Edition 2019.

Reference Text Books			
	Author	Title	Publisher
1	Dr. Dhaval Maheta	Data Analysis using R	Notion Press, September 2021
2	Michael J.Crawley	The R Book	Wiley, Edition: 2007
3	Ken Black John	Business Statistics for Contemporary Decision Making	John Wiley & Sons, Inc., Edition 2013

P.B. Siddhartha College of Arts & Science, Vijayawada – 520 010.
(An Autonomous College in the jurisdiction of Krishna University)
M.C.A – III Semester
Title: Applied Data Analytics
Course Code: 22CA3E4
(w.e.f admitted batch 2020-21)

Time: 3 Hours   **Max.Marks: 70**
    Answer ALL questions   **(5×4 = 20 Marks)**

1. (a) What are the different *Data Types* used in R. (CO1,L1)
                            (or)
   (b) Define *Subsetting and merging*. (CO1,L1)
2. (a) Explain *Hypothesis Testing* in R? (CO2,L2)
                            (or)
   (b) Explain Random *Sampling and cluser sampling* ? (CO2,L2)
3. (a) Explain about histograms with example using R (CO3.L5)
                            (or)
   (b) Define ANOVA and uses of ANOVA CO3,L5)
4. (a) Explain about *Logistic Regression using R*. (CO4,L2)
                            (or)
   (b) Explain *Time Series Analysis* and uses of time series analysis?(CO5,L2)
5. (a) What is the syntax used to read *XML Files*. (CO5,L1)
                            (or)
  (b) How we can insert data into R using MYSQL?(CO5,L1)

Answer ALL questions
   All Questions Carry Equal Marks.   **(5×10 = 50 Marks)**

6. (a) Outline the different *Data Structures* used in R. (CO1,L2)
                (or)
   (b) Explain *Control Flow Statements* in R. (CO1,L2)
7. (a) Explain the different statistical measures used in *Descriptive Statistics*. (CO2,L5)
                (or)
   (b) Explain *Non Parametric Test* and *Wilcoxon Signed-Rank Test* in R with example. (CO2,L5)
8. (a) List *Various Types of Charts* in R. (CO3,L4)
                (or)
   (b) Analyze *One-way ANOVA* and *Two-way factorial ANOVA*. (CO3,L4)
9. (a) Discuss *Simple* and *Multiple Regression* in R with Example. (CO4,L6)
                (or)
   (b) Elaborate different components used in *Time Series Analysis in R* with example. (CO4, L6)
10. (a) How do you *connect to a database* in R using MYSQL ? Give one example. (CO5,L5)
                (or)
   (b) How to do you import *csv file* and *binary file* in R with example? (CO5,L5)