**SRR & CVR GOVT DEGREE COLLEGE (A)::VIJAYAWADA**

**Revised Syllabus 2021-22**

| **Semester** | **Course Code** | **Course Title** | **Hours** | **Credits** |
| --- | --- | --- | --- | --- |
| **II** | **DCSC N-2352** | **INTRODUCTION TO DATA**  **SCIENCE WITH R** | **60** | **3** |

**Objective**

Data Science is a fast-growing interdisciplinary field, focusing on the analysis of data to extract knowledge and insight. This course will introduce students to the collection. Preparation, analysis, modeling and visualization of data, covering both conceptual and practical issues. Examples and case studies from diverse fields will be presented, and hands-on use of statistical and data manipulation software will be included.

**Outcomes**

1. Recognize various disciplines that contribute to a successful data science effort.

2. Understand the processes of data science - identifying the problem to be solved,data collection, preparation, modeling, evaluation and visualization.

3. Be aware of the challenges that arise in data sciences.

4. Develop and appreciate various techniques for data modeling and mining.

5. Be cognizant of ethical issues in many data science tasks.

6. Be comfortable using commercial and open source tools such as the R language and its associated libraries for data analytics and visualization.

7. Learn skills to analyze real time problems using R

8. Able to use basic R data structures in loading, cleaning the data and preprocessing the data.

9. Able to do the exploratory data analysis on real time datasets

10. Able to understand and implement Linear Regression

11. Able to understand and use - lists, vectors, matrices, dataframes, etc.

**Unit-1:**

Introduction to Data Science- Introduction- Definition - Data Science in various fields -

Examples - Impact of Data Science - Data Analytics Life Cycle - Data Science Toolkit - Data

Scientist - Data Science Team

Understanding data: Introduction – Types of Data: Numeric – Categorical – Graphical – High

Dimensional Data – Classification of digital Data: Structured, Semi-Structured and Un-

Structured - Example Applications. Sources of Data: Time Series – Transactional Data –

Biological Data – Spatial Data – Social Network Data – Data Evolution.

**Unit-2:**

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-3:**

Matrices - Creating Matrices - Adding or Removing rows/columns - Reshaping - Operations -

Special functions on Matrices. Lists - Creating List – General List Operations - Special

Functions - Recursive Lists. Data Frames - Creating Data Frames - Naming - Accessing - Adding

- Removing - Applying Special functions to Data Frames - Merging Data Frames- Factors and

Tables.

**Unit- 4:**

Input / Output – Reading and Writing datasets in various formats - Functions - Creating User-

defined functions - Functions on Function Object - Scope of Variables - Accessing Global,

Environment - Closures - Recursion. Exploratory Data Analysis - Data Preprocessing -

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.

**Unit- 5:**

Inferential Statistics with R - Types of Learning - Linear Regression- Simple Linear Regression

- Implementation in R - functions on lm() - predict() - plotting and fitting regression line.

Multiple Linear Regression - Introduction -comparison with simple linear regression -

Correlation Matrix - F-Statistic - Target variables Vs Predictors - Identification of significant

features - Implementation of Multiple Linear Regression in R.

**References**

1. Nina Zumel, John Mount, “Practical Data Science with R”, Manning Publications, 2014.

2. Jure Leskovec, Anand Rajaraman, Jeffrey D.Ullman, “Mining of Massive Datasets”,

Cambridge University Press, 2014.

3. Mark Gardener, “Beginning R - The Statistical Programming Language”, John Wiley &

Sons, Inc., 2012.

4. W. N. Venables, D. M. Smith and the R Core Team, “An Introduction to R”, 2013.

**SRR & CVR GOVT DEGREE COLLEGE (A)::VIJAYAWADA**

I BSc (Data Science) Semester -II

Revised Syllabus 2021-22

INTRODUCTION TO DATA SCIENCE WITH R

**MODEL PAPER**

**Time: 3hrs Max Marks:60**

**SECTION-A**

**Answer Any FIVE of the following Questions 5 X 4= 20 marks**

1. How is data science useful in our daily life?

2. What is the main role of a data scientist?

3. Write a short note on basic features of R programming ?

4. How do I get help for a function in R studio?

5. What does the table () function do in R?

6. What is the list in R? Explain with an example?

7. What are the different measures of central tendency and how do you measure them in R?

8. How do you read and write .csv files in R?

9. How do you make a correlation matrix in R?

10. How to Find the F Critical Value in R

**SECTION – B**

**Answer All the following questions 5 X 8=40M.**

11 a) What are the different types of data in data science?

OR

b)What are the different sources of data?

12.a) Explain about the data structures used in R?

OR

b) What are the control structures used in R?

13.a) What is a data frame and explain about common operations on Data Frames.

OR

b) Explain about various operations on matrix rows and columns?

14.a) How is R used in data visualisation?

OR

b) What Is Data Preprocessing & What Are The Steps Involved?

15. a) What are the different types of learning?

OR

b) Explain in detail about linear regression and multiple linear regression.

**SRR & CVR GOVT DEGREE COLLEGE (A)::VIJAYAWADA**

I BSc (Data Science) Semester -II

Revised Syllabus 2021-22

INTRODUCTION TO DATA SCIENCE WITH R

**Question Bank**

**SECTION-A**

**SHORT ANSWER QUESTIONS - 4 MARKS**

1.Define data science and applications of data science?

2.write short notes on data collection ?

3.What is a toolkit in data science?

4.Write about Help functions in R?

5.Briefly explain about R studio ?

6.How to declare Vectors in R?

7. How can missing values be handled in vectors?

8. How do you check for equality of a vector in R?

9. .Write short notes on Filtering in vectors?

10. Explain procedure to add and delete rows and columns in matrix ?

11. Write short notes on recursive lists R?

12.Explain about merging of Data Frames in R?

**LONG ANSWER QUESTIONS 8 MARKS**

1. What are the steps of the data science process?
2. What is the classification of digital data?
3. What is the role of a data scientist?
4. Explain about Data Cleaning methods ?
5. Explain about Data Characterisation and Analysis ?
6. Explain about data structures in the R language ?
7. Explain about Data Modelling and Mining techniques in R ?
8. What is Vector and explain about common vector operations ?
9. Explain about various functions applied on matrix rows and columns?
10. What is a List? Explain about various operations on Lists?
11. What is a Data Frame ? Explain procedure to create Data Frame with example?
12. Explain in detail about linear regression and multiple linear regression.