**SRR & CVR Govt. Degree College (A), Vijayawada**

**Department of Computer Science**

**Sem-II Mid-1**

**Introduction to Data Science and R**

**Time: 1hr Max Marks:20M**

**I Answer any ONE of the following 1 x 5 = 5 M**

1. Explain about the data science process
2. What are the different data structures in R? Explain
3. Explain about exploratory data analysis

**II Answer any FIVE of the following 5 x 2 = 10 M**

1. What are the 4 V’s of Big Data?
2. Explain about naming objects in R
3. Explain about data transformation.
4. Explain about data integration
5. How do you create a simple function in R?
6. List the types of data in R
7. How do I get help for a function in R?

**III Answer the following 10 x ½ = 5M**

Q1. \_\_ and \_\_\_ are types of categorical data

Q2. Match the following

1. List a. Converting one data type to another.

2. Data Frame b.Used to represent categorical data with levels.

3. Coercion c. A table-like structure

4. Factor d. An ordered collection that can be of different types.

Q3. What are the four display panes in RStudio?

Q4. What are the two types of assignment operators used in R?

Q5. TRUE Or FALSE are the values that can be stored in \_\_\_\_ data type

Q6. %/% is\_\_\_\_\_ arithmetic operator

Q7. RDBMS follows

1. Structured data
2. Unstructured data
3. Semi structured data
4. Quasi structured data

Q8. \_\_\_ function is used to get the number of rows and columns of a matrix.

Q9. Missing value in R is represented by

1. NULL b) NA c) Nan d) None

**SRR & CVR Govt. Degree College (A), Vijayawada**

**Department of Computer Science**

**Sem-II Mid-1**

**Introduction to Data Science With R**

**Time: 1hr Max Marks:20M**

**I Answer any ONE of the following 1 x 5 = 5 M**

1. Explain about creating, indexing and accessing vectors in R
2. Explain about different types of common databases or data sources.
3. What are the types of data classification?

**II Answer any FIVE of the following 5 x 2 = 10 M**

1. Explain about if else statement in R?
2. Explain about while loop in R?
3. How does break statement work in R?
4. Explain about vector recycling?
5. How do you create a simple function in R?
6. List the types of data in R
7. How do I get help for a function in R?

**III Answer the following 10 x ½ = 5M**

Q1. \_\_ and \_\_\_ are types of categorical data

Q2. Match the following

1. Spatial data a. Data from living organisms

2. Time series b. Data from social media

3. Biological data c. Data in time intervals

4. Social data d. Data specific to location

Q3. What are the four display panes in RStudio?

Q4. What are the two types of assignment operators used in R?

Q5. TRUE Or FALSE are the values that can be stored in \_\_\_\_ data type

Q6. %/% is\_\_\_\_\_ arithmetic operator

Q7. RDBMS follows

1. Structured data
2. Unstructured data
3. Semi structured data
4. Quasi structured data

Q8. Which statement will take the control out of a function

1. Break b) next c) return d) exit

Q9. Missing value in R is represented by

1. NULL b) NA c) Nan d) None

**SRR & CVR Govt. Degree College (A), Vijayawada**

**Department of Computer Science**

**Sem-II Mid-2**

**Introduction to Data Science and R**

**Time: 1hr Max Marks:20M**

**I Answer any ONE of the following 1 x 5 = 5 M**

1. How is R used in data visualisation? How to draw scatter and line charts in R?
2. What Is Data Preprocessing & What Are The Steps Involved?
3. Explain how linear regression and multiple linear regression is implemented in R?

**II Answer any FIVE of the following 5 x 2 = 10 M**

1. How do I make a pie chart in R?
2. How do you find outliers in boxplot in R?
3. What are descriptive methods in statistics?
4. How do you represent data in the form of vertical bars?
5. What is meant by inferential statistics?
6. What is a histogram used for?
7. What are the values getting displayed when we use summary () on a data frame?

**III Answer the following 10 x ½ = 5M**

1. \_\_\_\_\_\_\_\_&\_\_\_\_\_\_\_\_\_\_\_ are measures of central tendency?

2 . \_\_ & \_\_\_\_are preprocessing steps

3. \_\_\_\_\_\_\_\_ & \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ are techniques to fill missing values?

4. Recursive function is a function that calls\_\_\_

5. What does head() function do?

6. Which function is used for linear regression in R?

7. Which function is used to add points to a plot?

**SRR & CVR Govt. Degree College (A), Vijayawada**

**Department of Computer Science**

**Sem-II Mid-2**

**INTRODUCTION TO DATA SCIENCE AND R PROGRAMMING**

**Time: 1hr Max Marks:20M**

**I Answer any ONE of the following 1 x 5 = 5 M**

1. What is a perceptron? Explain with an example
2. Explain linear and multiple linear regression
3. Explain in detail about control structures in R with examples

**II Answer any FIVE of the following 5 x 2 = 10 M**

1. What does "training a model" and "validating a model" mean
2. What is root cause analysis?
3. What is a bottleneck in the context of handling large data?
4. What is data chunking
5. Identify two key functions from the dplyr package
6. What are Vectorised Operations?
7. Explain briefly about debugging in R.

**III Answer the following 10 x ½ = 5M**

**1. In R, which operator is commonly used to access columns within a data frame?**

a) [] b) [[]] c) $ d) %>%

**2 . \_\_ & \_\_\_\_are preprocessing steps**

**3. In R, which data structure is specifically designed for storing tabular data?**

a) list b) vector c) data.frame d) matrix

**4. What is a key characteristic of online learning algorithms for large datasets?**

a) They load the entire dataset into memory b) They require powerful hardware

c) They process data one observation at a time d) They always use batch updates

**5. Which R function is used to combine elements into a vector?**

a) list() b) data.frame() c) c() d) matrix()

**6. Which type of machine learning uses labelled data for training?**

a) Unsupervised b) Reinforcement c) Supervised d) Semi-supervised

**7. MapReduce is an algorithmic approach for large data that involves:**

a) Processing data entirely in memory. b) Primarily focusing on visualising data.

c) Dividing the task into map and reduce phases.

d) Compressing data to the smallest possible size before analysis.

**8. Match the following**

a) Dataframe. i) A type of vector that can contain elements of different classes

b) List ii)Used to store tabular data in R

c) Vectorised Operations iii) how R searches for the value associated with a symbol

d) Scoping Rules iv) Performing an operation on all elements of a vector simultaneously

Home Assignment1:

Q1. Adding and deleting vectors in R

Q2. Filtering

Q3. Vector recycling in R

Q4. Data Science Team

Q5. Data Scientist

Home assignment 2:

1. What are the different types of data in data science?

2. What are the different sources of data?

3. Explain about the data structures used in R?

4. What are the control structures used in R?

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

6. Explain about various operations on matrix rows and columns?

home assignment 3 & 4 question :1)How is R used in data visualisation?

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

3) What are the different types of learning?

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

Seminar Topics

1. How to handle missing data in R
2. mahesh: classification of digital data and types of data structures in R
3. Rakesh: sources of data or types of databases and data evolution
4. Mahendra: list and its operations
5. Chitti babu: vectors in R and it's types and operations
6. Sai : matrices and its operations
7. Shakeera: linear regression in R
8. Kowsar: data frame and its operations