

U.G. 6th Semester Examination - 2021

COMPUTER SCIENCE

[HONOURS]

Discipline Specific Elective (DSE)

Course Code : COM.SC-H-DSE-L-603

(Introduction to Data Science)

Full Marks : 60

Time : 2½ Hours

The figures in the right-hand margin indicate marks.

Candidates are required to give their answers in their own words as far as practicable.

GROUP - A

1. Answer any **ten** questions: 2×10=20
- a) What is data frame in R?
 - b) Name one method each to visualize one-dimensional data and two-dimensional data.
 - c) What do you mean by exploratory data analysis?
 - d) *"Data science is multi-disciplinary"*: Justify the statement.
 - e) How to store two-dimensional and multi-dimensional data in R?
 - f) Name a few techniques that we use to deal with missing values in data.

[Turn Over]

- g) How to make fonts bold and italicized in markdown?
- h) When is box plot used?
- i) What is the procedure of taking input from .csv file in R? Explain with an example.
- j) What does reproducibility mean in data science?
- k) What will be the properties of a tidy dataset?
- l) How can you create bullet lists in markdown?
- m) What do you mean by outliers?
- n) How can you create a matrix in R in row major form?
- o) What is knitr?

GROUP-B

- Answer any **four** questions: 5×4=20
- 2. Define data science. Briefly mention the stages of data science. 3+2
 - 3. Briefly explain data cleansing. 5
 - 4. Differentiate between confirmatory and exploratory data analysis. 5
 - 5. Name the functions used in R to convert strings of date format into date type data. Now consider on 31st Aug, 2021 at 10 PM IST you will be catching a flight from Kolkata to London. You will reach London on 1st Sep,

775/Comp.Sc.

(2)

2021 at 3:30 AM GMT. Use R to store two dates and find the time taken in transit. 2+3

6. Briefly describe different categories of data that we deal with in data science. 5
7. What is multivariate statistical analysis? What are its the advantages? 2+3

GROUP-C

Answer any **two** questions: 10×2=20

8. a) What do you mean by data visualization?
b) Why is data visualization an important aspect of data science?
c) Describe a few data visualization techniques for different dimensionality of data with the help of suitable examples and diagrams. 2+2+6
9. a) What do you mean by dimensions of data?
b) What is dimensionality reduction of data?
c) Why is reduction of dimensions an important aspect in data science?
d) Explain one technique of dimensionality reduction. 2+2+2+4
10. a) Why is reproducibility a major concern in science?
b) What is reproducibility crisis?

- c) Markdown is a tool to report the result of research and analysis. Explain the method of reporting using markdown in R. You can choose an example document that has title; author; some texts in bold and italicized fonts; list of items and bibliography. 3+2+5

11. a) Write a function in R that will find if a number is prime or not. The number will be passed to the function as an argument.
b) Data visualization is a powerful tool. Explain the role of plot function in R to visualize data with some examples. 4+6