

Reg. No.

**B.Tech. DEGREE EXAMINATION, MAY 2023**  
Fifth & Sixth Semester

**18CSE396T – DATA SCIENCE**

*(For the candidates admitted during the academic year 2018-2019 to 2021-2022)*

**Note:**

- (i) **Part - A** should be answered in OMR sheet within first 40 minutes and OMR sheet should be handed over to hall invigilator at the end of 40<sup>th</sup> minute.
- (ii) **Part - B & Part - C** should be answered in answer booklet.

Time: 3 hours

Max. Marks: 100

**PART – A (20 × 1 = 20 Marks)**

Answer ALL Questions

	Marks	BL	CO	PO
1. Who represents the business interests of the project? (A) Project sponsor (B) Client (C) Data scientist (D) Data architect	1	1	1	1
2. The process of learning to order items is referred as _____. (A) Ranking (B) Scoring (C) Characterizing (D) Classifying	1	1	1	1
3. Which operator assigns the value on the right to the symbol on the left? (A) ← (B) = (C) → (D) #	1	1	1	1
4. The process of handling invalid or missing values is referred as (A) Data cleaning (B) Data exploration (C) Data staging (D) Data warehousing	1	1	1	1
5. Which of the following is not a data cleaning technique? (A) Removing duplicates (B) Removing missing values (C) Scaling the data (D) Handling outliers	1	1	2	1
6. _____ is the process of identifying the problem and define the objectives and scope of the model. (A) Schedule (B) Model deployment (C) Evaluation (D) Model planning	1	1	2	1
7. What is the first step in model planning process? (A) Collecting data (B) Defining the problem (C) Selecting the modelling technique (D) Preparing the data	1	1	2	1
8. _____ is the most commonly used visualization for data exploration. (A) Histogram (B) Pie chart (C) Bar chart (D) Chart	1	1	2	1
9. What is the shape of two dimensional array with N rows and M columns? (A) (N, M) (B) (M, N) (C) (N + M) (D) (N – M)	1	1	3	1

10. Which of the following is a valid method to access a data frame column in R?	1	1	3	1
(A) []				
(B) [[ ]]				
(C) {}				
(D) {}				
11. What is the default ordering or levels in a factor in R?	1	1	3	1
(A) Alphabetical order				
(B) Numeric order				
(C) Random order				
(D) No default ordering				
12. Which of the following R functions is used to calculate the Chi-squared test statistic for a contingency table?	1	1	3	1
(A) Chisq.test ( )				
(B) T.test ( )				
(C) Cor.test ( )				
(D) Chi.test ( )				
13. The proportion of correct predictions among all predictions is called as	1	1	4	1
(A) Accuracy				
(B) Precision				
(C) Recall				
(D) Sensitivity				
14. A method for predicting continuous variables is called	1	1	4	1
(A) Linear regression				
(B) Clustering				
(C) Classification				
(D) Dimensionality reduction				
15. What is logistic regression used for?	1	1	4	2
(A) Regression analysis				
(B) Classification analysis				
(C) Clustering				
(D) Normalization				
16. What is the equation of simple linear regression model	1	1	4	2
(A) $y = mx + b$				
(B) $y = mx + c$				
(C) $y = ax + b$				
(D) $y = ax + c$				
17. The quantity to be predicted is called _____.	1	1	5	1
(A) Row				
(B) Column				
(C) Buzz				
(D) Topic				
18. Each _____ represents many different measurements of the popularity of a technical personal computer discussion topic	1	1	5	2
(A) Topic				
(B) Row				
(C) Measurement				
(D) Times				
19. With cache = false the code is always _____.	1	1	5	3
(A) Terminated				
(B) Run				
(C) Executed				
(D) Copied				
20. _____ controls whether the source code is copied in to the document.	1	1	5	1
(A) Cache				
(B) Echo				
(C) Eval				
(D) Message				

**PART – B (5 × 4 = 20 Marks)**  
Answer ANY FIVE Questions

21. Explain any two stages in data science project.	4	1	1	2
---	---	---	---	---

22. Explain the process of managing data.	4	1	1	2
23. Briefly discuss about data exploration.	4	1	2	3
24. Define precision, recall, accuracy and sensitivity with its formula.	4	2	4	3
25. Briefly discuss about DOT chart.	4	2	5	2
26. Explain structures and semi-structured data.	4	2	1	1
27. Discuss about hypotheses testing.	4	2	3	2

**PART – C (5 × 12 = 60 Marks)**  
Answer ALL Questions

28. a. Illustrate the relational and non relational database with an example.	12	2	1	3
(OR)				
b. Describe in detail about the roles in a data science project.	12	2	1	3
29. a. Explain in detail about model building and common tools for the model building phase.	12	1	2	2
(OR)				
b. Briefly discuss about data preparation and the common tools involved in data preparation phase.	12	1	2	2
30. a. Explain in detail about array, matrices, data frames and list using R function as an example.	12	2	3	3
(OR)				
b. Describe in detail about student-T-test and differences of means.	12	2	3	3
31. a. Briefly discuss the method of mapping problems to machine learning.	12	2	4	2
(OR)				
b. Explain logistic regression with an example.	12	2	4	2
32. a. Explain the following	12	3	5	3
(i) Presenting your model to end user				
(ii) Presenting your work to other data scientist				
(OR)				
b. Briefly discuss about visualizing a single variable and examining multiple variables.	12	3	5	3

\* \* \* \* \*