QP CODE

C1089

Enrollment Number: ……

Name: ……

B.SC DEGREE EXAMINATIONS, AUGUST 2025
Second Semester
B.Sc Data Science
B24DS03DC – Introduction to Data Science and Analytics
(2024 July admissions)

Time: 3 Hours **Max Marks: 70**

Section A

Answer any ten of the following questions in a word or sentence each. Each question carries 1 mark.

1. What is the primary objective of data science?
2. What does the DIKW pyramid represent in the context of data transformation?
3. Why is data standardization important?
4. What are the techniques used to simplify complex datasets?
5. What are structural attributes of data?
6. What is feature extraction in text analytics and what is the Bag of Words model?
7. What are the key components of a data warehouse?
8. What is Z-scores?
9. Differentiate OLAP and OLTP systems.
10. How does data storytelling enhance business communication?
11. Write any two libraries used for text data analytics.
12. How is data integrated into a data warehouse?
13. What is Multidimensional Analysis in OLAP Systems?
14. What are the key components of the Snowflake Schema in a data warehouse?
15. What is a Data Management Plan?

(1X10=10)

Section B

Answer any five of the following questions in two or three sentences each. Each question carries 2 marks.

16. Discuss the importance of data collection in the data science process.
17. Explain normalization techniques used in data preprocessing.
18. What are categorical and numerical data types?
19. What are the different data aggregation methods?

20. Explain One-Hot Encoding and how it used in machine learning for categorical data?
21. What is the difference between normalization and standardization in data processing?
22. What is the importance of the General Data Protection Regulation in data privacy?
23. What are the different types of dimensions in a data warehouse with examples?
24. Explain the ethical principles in data science with examples.
25. Define Data Governance and explain its key components.

(2X5=10)

Section C

Answer any five of the following questions in one page each. Each question carries 4 marks.

26. Explain different data types and how it differs in terms of structure and analysis.
27. What are the key components of administrative attributes and how they contribute to effective data management?
28. Differentiate the Continuous Wavelet Transform and the Discrete Wavelet Transform.
29. How can we apply different data discretization methods to a dataset?
30. Discuss various key feature engineering methods used in machine learning and how they improve data for modelling.
31. Write the major steps involved in text preprocessing for text analytics.
32. Explain the different types of sampling techniques used for data reduction.
33. Differentiate Word2Vec and GloVe in text analytics.
34. Compare and contrast the different types of data warehouse architectures.
35. Explain the key provisions and objectives of the Indian Digital Personal Data Protection Act.

(5X4=20)

Section D

Answer any two of the following questions in three pages each. Each question carries 15 marks.

36. Describe in detail a sample scenario related to the Data Science Project Life Cycle.
37. Explain the importance of data cleaning in the data analysis process and describe the key steps involved in ensuring high-quality, reliable data.
38. What are the various data transformation techniques and how they improve the efficiency of data analysis?
39. Discuss the significance of data storytelling in decision-making and how these techniques improve text analytics.

(15X2=30)