PSG COLLEGE OF ARTS & SCIENCE
(AUTONOMOUS)

BSc DEGREE EXAMINATION MAY 2025
(Sixth Semester)

Branch – **COMPUTER SCIENCE**

DATA SCIENCE

Time: Three Hours Maximum: 50 Marks

SECTION-A (5 Marks)
Answer **ALL** questions
ALL questions carry **EQUAL** marks (5 x 1 = 5)

1 What is Entropy in the context of information theory?
(i) The measure of how fast data can be processed
(ii) The measure of randomness or impurity in the dataset
(iii) The number of unique values in a feature
(iv) The ratio of true positives to false negatives

2 Which of the following statements best describes the role of NoSQL databases in Data Science?
(i) They are used only for storing relational data
(ii) They require strict schema definitions like traditional SQL databases
(iii) They do not support distributed storage
(iv) They are suitable for handling large volumes of structured and unstructured data

3 Which phase involves delivering the final results to stakeholders in a meaningful way?
(i) Deployment (ii) Model Building
(iii) Communication of Results (iv) Data Preparation

4 Which function is commonly used to read a CSV file into R?
(i) read.csv() (ii) read.xlsx()
(iii) read.table() (iv) read.json()

5 In Tableau, which chart type is best used to compare parts of a whole?
(i) Line Chart (ii) Pie Chart
(iii) Histogram (iv) Scatter Plot

SECTION - B (15 Marks)
Answer **ALL** Questions
ALL Questions Carry **EQUAL** Marks (5 x 3 = 15)

6 a Explain how programming helps a data scientist do their job?
OR
b Describe a situation where cluster sampling might be used.

7 a How would you use SQL joins and aggregation functions to analyze customer purchases from two different tables: Customers and Orders? Write an example.
OR
b How would using a columnar database help when analyzing large datasets in data science? Give an example.

8 a Assume any real-world problem, such as predicting customer churn or analyzing sales performance, and justify how you would apply data analytics to solve it.
OR
b After building a machine learning model, how would you operationalize it to ensure it provides real-time insights and updates in a production environment?

Cont...

9 a Narrate on writing Data into File using R.

OR

b Compare and contrast between normal distribution and probability distribution.

10 a Describe Tableau Basics.

OR

b Classify special chart types in Tableau.

SECTION -C (30 Marks)

Answer **ALL** questions

ALL questions carry **EQUAL** Marks (5 x 6 = 30)

11 a How would you combine business intelligence tools with data science techniques to provide insights that can help improve sales performance? What specific tools and techniques would you use, and why?

OR

b Outline Bayes Theorem and Inferential Statistics.

12 a Given a raw dataset with missing values, duplicate records, and inconsistent formatting, how would you use SQL to clean and prepare the data for analysis? Explain the steps and SQL techniques you would use.

OR

b Summarize the concepts of preparing Data for Analytics Tools with instances.

13 a Elucidate the process of Data Discovery. Infer the steps required to prepare it for analysis or modeling.

OR

b After building a predictive model, how would you analyze its performance and explain the results to a non-technical audience? Why is it important to communicate results clearly?

14 a Describe usage of built-in datasets in R to practice data analysis. Give an example of a dataset and explain what kind of analysis you could perform with it.

OR

b How would you use a list and a data frame in a data science project? Give an example of what each one might be used for?

15 a Examine the measures and descriptive statistics in Tableau to analyze a dataset. Using an example explain why are they important for understanding your data?

OR

b When designing a dashboard in Tableau, how would you decide what charts and layout to use to make the data easy to understand? Also explain the key design principles.

Z-Z-Z END