BRAINWARE UNIVERSITY

Term End Examination 2024-2025
Programme – B.Tech.(CSE)-DS-2021/B.Tech.(CSE)-DS-2022
Course Name – Data Analytics and Reporting
Course Code - PCC-CSD503
(Semester V)

Full Marks : 60   **Time : 2:30 Hours**

[The figure in the margin indicates full marks. Candidates are required to give their answers in their own words as far as practicable.]

Group-A
(Multiple Choice Type Question)   1 x 15=15

1. *Choose the correct alternative from the following :*

(i) Identify the name of the first phase of data analytics lifecycle.

a) Model planning
b) Data preparation
c) Discovery/objective
d) Data visualization

(ii) Select the name of the tool that is NOT used for data preparation:

a) Power BI
b) Hadoop
c) Python
d) Tableau

(iii) Choose the name of the phase that the analyst team delivers final report, briefings and codes .

a) Operationalize
b) Model Building
c) Model Planning
d) None of these

(iv) Select the primary purpose of a histogram in data analysis.

a) To display individual data points
b) To show the relationship between two variables
c) To summarize the distribution of a dataset
d) To calculate the mean and median of the data

(v) Identify the type of data visualization that is best suited for showing the distribution of a continuous variable:

a) Bar chart
b) Scatter plot
c) Histogram
d) Pie chart

(vi) Choose the commonly used machine learning algorithm for classification tasks.

a) Linear regression
b) K-means clustering
c) Decision tree
d) Principal component analysis (PCA)

(vii) Select the primary goal of MapReduce in the context of big data processing.

a) Data storage
b) Data cleaning
c) Data visualization
d) Data processing and analysis

(viii) Identify the technology that is commonly used to store and manage large volumes of structured data in a distributed manner in big data systems.

a) NoSQL databases
b) Relational databases
c) Hadoop
d) Spark

(ix) Select the primary objective of data mining in Business Intelligence.

a) Data storage
b) Data presentation
c) Data discovery
d) Data transmission

(x) Choose the proper role of the "Reduce" function in MapReduce.

a) To split data into smaller chunks
b) To process and filter data
c) To aggregate and summarize results
d) To perform data extraction

(xi) Select the component in a big data stack that is often used for querying and analyzing data using a SQL-like language:

a) Apache Kafka
b) Apache Hadoop
c) Apache Hive
d) Apache Storm

(xii) Choose the correct five Cs in ethics:

a) Categorization, Compliance, Confidentiality, Communication, and Consequences
b) Care, Compassion, Courage, Communication, and Commitment
c) Character, Culture, Compliance, Consequences, and Context
d) Confidentiality, Consistency, Communication, Consequences, and Compliance

(xiii) Select the significant trend in ethical considerations regarding data privacy and cybersecurity:

a) Neglecting data protection in favor of business growth
b) Increased emphasis on securing personal data and privacy
c) Minimizing the importance of data breaches and cyber threats
d) Promoting unrestricted access to personal information

(xiv) Identify the primary goal of doing good data science.

a) Maximizing profit
b) Ensuring ethical standards are met
c) Collecting as much data as possible
d) Reducing computational costs

(xv) Identify the most critical aspect of valuing privacy.

a) Making data available to everyone
b) Respecting individuals' control over their data
c) Maximizing data utility
d) Minimizing the cost of data storage

Group-B
(Short Answer Type Questions)

3 x 5=15

2. Discuss some common data sources that BI tools utilize for analysis (3)
3. Give examples of data visualizations. (3)
4. Explain the role of model evaluation in the machine learning process. (3)
5. Explain Hadoop Distributed File System (HDFS). (3)
6. Explain YARN (Yet Another Resource Negotiator). (3)

OR

Explain MapReduce of Apache Hadoop open-source software framework. (3)

Group-C

(Long Answer Type Questions)   5 x 6=30

7. Explain the ways data context affects the interpretation and analysis of data.   (5)
8. Describe the role of Business Intelligence (BI) in data analytics.   (5)
9. Discuss the concept of data mining in big data analytics and provide examples of real-world   (5)
    applications.
10. Justify the importance of Data Correctness in Data Analytics.   (5)
11. Explain the various categories of Data Visualization.   (5)
12. Speculate the common causes of Data Incorrectness.   (5)

OR

Propose some ways to ensure data consistency.   (5)

****