Printed Page:- 04 Subject Code:- BMCA0313

Roll. No:

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA

(An Autonomous Institute Affiliated to AKTU, Lucknow)

MCA

SEM: III - THEORY EXAMINATION (2024- 2025)

Subject: Advanced Concepts of Analytics

Time: 3 Hours Max. Marks: 100

General Instructions:

IMP: *Verify that you have received the question paper with the correct course, code, branch etc.*

1. This Question paper comprises of three Sections -A, B, & C. It consists of Multiple Choice Questions (MCQ's) & Subjective type questions.

2. Maximum marks for each question are indicated on right -hand side of each question.

3. Illustrate your answers with neat sketches wherever necessary.

4. Assume suitable data if necessary.

5. Preferably, write the answers in sequential order.

6. No sheet should be left blank. Any written material after a blank sheet will not be evaluated/checked.

REG.JULY_DEC-2024

SECTION-A 20

1. Attempt all parts:-

1-a. Data integrity refers to: (CO1, K1) 1

(a) Speed of data processing

(b) Accuracy, consistency, and reliability of data

(c) Amount of data stored

(d) Cost of data storage

1-b. A data-cleaning technique involves: (CO1, K1) 1

(a) Ignoring missing values

(b) Standardizing data formats

(c) Increasing data redundancy

(d) Data storage

1-c. Data integration involves (CO2, K1,K2) 1

(a) Storing data in a single location

(b) Combining data from multiple sources

(c) Deleting duplicate entries

(d) Cleaning noisy data

1-d. A key benefit of correlation analysis is:(C02, K1 K2) 1

(a) Improved data redundancy

(b) Better understanding of variable relationships

(c) Increased data volume

(d) Simplified data entry

1-e. Data visualization primarily helps in:(CO3, K1 K2) 1

(a) Storing data
(b) Analyzing data
(c) Communicating data insights
(d) Creating databases

1-f. Effective use of typography in visualizations enhances:(CO3, K1 K2) 1

(a) Confusion
(b) Readability
(c) Data density
(d) None of the above

1-g. Power BI is primarily used for:(CO4, K1 K2 K3) 1

(a) Word processing
(b) Data visualization
(c) Audio editing
(d) Graphic design

1-h. The main purpose of Power BI Desktop is to:(CO4, K1 K2 K3) 1

(a) Write code
(b) Import images
(c) Create reports and dashboards
(d) Store data

1-i. Data processing models in Big Data primarily focus on:(CO5, K1 K2) 1

(a) Real-time processing
(b) Batch processing
(c) Both real-time and batch processing
(d) Manual data entry

1-j. A common tool used for Big Data analytics is:(CO5, K1 K2) 1

(a) Microsoft Excel
(b) Apache Hadoop
(c) Adobe Photoshop
(d) None of the above

REG.JULY_DEC-2024

2. Attempt all parts:-

2.a. Explain how data validation in spreadsheets contributes to data integrity. 2

2.b. Regular data cleaning practices can enhance overall data quality. Explain the benefits of this approach. 2

2.c. Define dynamic visualizations. 2

2.d. Explain how data relationships can be managed in Power BI. 2

2.e.	Define Big Data.	2

SECTION-B 30

3. Answer any five of the following:-

3-a.	Identify various types of dirty data and their sources. Explain how recognizing these issues can lead to more accurate data analysis.(CO1, K1)	6
3-b.	Describe an automated data cleaning tool, including its features and functionalities. Discuss how it improves data quality and analysis efficiency.(CO1, K1)	6
3-c.	Explain the process of data cleaning and its significance in preparing data for analysis. Include key steps involved.(CO2, K1 K2)	6
3-d.	Discuss the importance of data validation during the data cleaning process. Explain how validation techniques enhance data quality.(CO2, K1 K2)	6
3.e.	Examine the process of integrating multiple data sources into a single visualization. What challenges arise, and how can they be addressed?(CO3, K1 K2)	6
3.f.	Discuss the primary functions of Power BI in data analysis.(CO4, K1 K2 K3)	6
3.g.	Discuss the significance of Big Data in today's economy.(CO5, K1 K2)	6

SECTION-C 50

4. Answer any one of the following:-

4-a.	Identify different types of dirty data, explaining their origins and effects on data analysis. Discuss the importance of recognizing these data quality issues for accurate results.(CO1,K1)	10
4-b.	Analyze the role of data validation techniques in maintaining data integrity. Discuss the specific methods used in spreadsheets and their effectiveness in preventing errors.(CO1, K1)	10

5. Answer any one of the following:-

5-a.	Explain the concept of correlation analysis and its importance in identifying relationships between variables. Discuss how correlation coefficients can guide data-driven decisions.	10
5-b.	During data integration from multiple sources, duplicate entries for the same entity are discovered. Illustrate the steps necessary to identify these duplicates and provide a framework for resolving them.	10

6. Answer any one of the following:-

6-a.	Discuss the relationship between imagery and data in visualizations. How can effective use of images complement data and improve overall understanding?	10
6-b.	Discuss the role of context in interpreting visual data. How can designers incorporate context to aid audience comprehension and enhance the effectiveness of visualizations?	10

7. Answer any one of the following:-

7-a.	Describe the key features of Power BI and their significance in data analysis.	10

REG:JULY_DEC-2024

7-b.	Explain the steps for importing data from flat files and Excel files into Power BI.	10

8. Answer any <u>one</u> of the following:-

8-a.	Explain the concept of Big Data and its significance in the modern economy.	10
8-b.	Define classification in data analytics and discuss its application in various contexts.	10

REG:JULY_DEC-2024