

| | | | | | | | | | | | | | | |
|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
| | | | | | | | | | | | | | | |
|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA

(An Autonomous Institute Affiliated to AKTU, Lucknow)

B.Tech.

SEM: V - THEORY EXAMINATION (2022 - 2023)

Subject: Data 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.

SECTION A

20

1. Attempt all parts:-

- | | | |
|---|---|---|
| 1 | Which of the following analysis is usually modeled by deterministic set of equations? (CO1) | 1 |
| | (a) Predictive | |
| | (b) Causal | |
| | (c) Mexhanistic | |
| | (d) All of the above | |
| 1 | In Big Data environments, Velocity refers- ? (CO1) | 1 |
| | (a) Data can arrive at fast speed | |
| | (b) Enormous datasets can accumulate within very short periods of time | |
| | (c) Velocity of data translates into the amount of time it takes for the data to be processed | |
| | (d) All of the above | |
| 1 | Which of the following statements is true about the source of data? (CO2) | 1 |
| | (a) The source of data that is collected and compiled by others is known as secondary data | |
| | (b) The source of data that is collected and compiled by others is known as tertiary data | |
| | (c) The source of data that is collected and compiled by others is known as primary data | |

- (d) None of the above
- 1 Blood Group of a person is an example of ? (CO2) 1
- (a) Numeric Data
 - (b) Categorical Data
 - (c) High Dimensional Data
 - (d) None of the above
- 1 The initial steps concerned in the process of knowledge discovery is: ? (CO3) 1
- (a) Data Selection
 - (b) Data Integration
 - (c) Data Cleaning
 - (d) Data Transformation
- 1 Methods of Dimensionality reduction involves: ? (CO3) 1
- (a) Wavelet Transform
 - (b) Principal Component Analysis
 - (c) Attribute Subset Selection
 - (d) All of the above
- 1 What is the role of exploratory graphs in data analysis? (CO4) 1
- (a) They are made for formal presentations
 - (b) They are typically made very quickly
 - (c) Axes, legends, and other details are clean and exactly detailed
 - (d) They are used in place of formal modeling
- 1 Descriptive analysis identifies patterns in time series data like trends, cycles. ? (CO4) 1
- (a) TRUE
 - (b) FALSE
- 1 The most popular data visualization library in python is _____ ? (CO5) 1
- (a) matplotlib
 - (b) matplotlib
 - (c) matplotlib
 - (d) pip
- 1 What Type Of Join Is Used In Blending? ? (CO5) 1
- (a) Right join

- (b) Outer Join
- (c) Left Join
- (d) None of the above

2. Attempt all parts:-

- | | | |
|------|---|---|
| 2.a. | What is Exploratory Analysis? (CO1) | 2 |
| 2.b. | Differentiate between data manipulation and data modification.? (CO2) | 2 |
| 2.c. | Explain Discrete Data Attribute.? (CO3) | 2 |
| 2.d. | What are the three ways of reducing dimensionality? (CO4) | 2 |
| 2.e. | Define Gantt chart.? (CO5) | 2 |

SECTION B

30

3. Answer any five of the following:-

- | | | |
|------|--|---|
| 3 | Tabulate the differences between analysis and reporting.? (CO1) | 6 |
| 3 | Explain the steps for a Data analytics project.? (CO1) | 6 |
| 3 | Differentiate between categorical data and numerical data ? (CO2) | 6 |
| 3 | Where is structured data used? Is excel a structured data? (CO2) | 6 |
| 3.e. | Differentiate between dimensionality reduction and Numerosity Reduction? (CO3) | 6 |
| 3.f. | Why organizations use time series analysis? Explain three examples of time series analysis.? (CO4) | 6 |
| 3.g. | How Can You Optimize the Performance of a Dashboard? (CO5) | 6 |

SECTION C

50

4. Answer any one of the following:-

- | | | |
|---|---|----|
| 4 | Explain the three types of Big data. Describe the features of Big Data in detail ? (CO1) | 10 |
| 4 | Explain how Facebook uses data to revolutionize social networking and advertising.? (CO1) | 10 |

5. Answer any one of the following:-

- | | | |
|---|--|----|
| 5 | What is a file format? Why there is a need to understand various file formats? (CO2) | 10 |
| 5 | What is the function to find the day of the week for a particular date value? How does the AND() function work in Excel? (CO2) | 10 |

6. Answer any one of the following:-

- | | | |
|---|---|----|
| 6 | Explain data discretization and concept hierarchy generation? (CO3) | 10 |
| 6 | Why is data discretization important? Explain data discretization using decision tree analysis? (CO3) | 10 |

7. Answer any one of the following:-

- | | | |
|---|---|----|
| 7 | Explain Principal Component analysis, assumptions, equations? (CO4) | 10 |
| 7 | Explain various key components in Exploratory data analysis? (CO4) | 10 |

8. Answer any one of the following:-

- | | | |
|---|--|----|
| 8 | Explain the different filters in Tableau and how are they different from each other? (CO5) | 10 |
| 8 | Describe all the data sources systems that Tableau can connect. Explain in detail ? (CO5) | 10 |