

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA

(An Autonomous Institute Affiliated to AKTU, Lucknow)

B.Tech

SEM: V - THEORY EXAMINATION (2024- 2025)

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-a. Point out the correct statement.(CO1, K1)

1

- (a) Raw data is original source of data
- (b) Pre-processed data is original source of data
- (c) Raw data is the data obtained after processing steps
- (d) None of the above

1-b. Choose the correct components of data science.(CO1, K2)

1

- (a) Domain Expertise
- (b) Data Engineering
- (c) Advanced Computing
- (d) All of the above

1-c. Pictorial representation of data using symbols is known as? (CO2, K2)

1

- (a) Bar Graph
- (b) Pictograph
- (c) Pie chart
- (d) None of the above

1-d. A teacher notes down the weight of each student in the class. What level of measurement is being used in this case? (CO2, K2)

1

- (a) Ordinal
- (b) Ratio

- (c) Nominal
- (d) Interval
- 1-e. Why do we need feature transformation?(CO3, K2) 1
- (a) Converting non-numeric features into numeric
- (b) Resizing inputs to a fixed size
- (c) Both A and B
- (d) None
- 1-f. To remove noise and inconsistent data.....is needed.(CO3, K2) 1
- (a) Data Cleaning
- (b) Data Transformation
- (c) Data Reduction
- (d) Data Integration
- 1-g. Find the outlier s in the given dataset: 21, 15, 18,11,13,71 (CO4, K2) 1
- (a) 11
- (b) 21
- (c) 71
- (d) 18
- 1-h. PCA reduces the dimension by finding a few.....(CO4, K2) 1
- (a) Hexagonal linear combination
- (b) Orthogonal linear combination
- (c) Octagonal linear combination
- (d) Pentagonal linear combination
- 1-i. The most popular data visualization library in python is ____ (CO5, K1) 1
- (a) matplotlib
- (b) matplotlib
- (c) matplotlib
- (d) pip
- 1-j. Which are pros of data visualization? (CO5, K1) 1
- (a) It can be accessed quickly by a wider audience.
- (b) It can misrepresent information
- (c) It can be distracting
- (d) None Of the above
2. Attempt all parts:-
- 2.a. Give three examples of Big data? (CO1, K1) 2
- 2.b. Describe secondary source of data.(CO2, K1) 2
- 2.c. What is Data Transformation? (CO3, K1) 2
- 2.d. What is the goal of outlier detection? (CO4, K2) 2

2.e.	Define Heat Map? (CO5, K1)	2
SECTION-B		30
3. Answer any <u>five</u> of the following:-		
3-a.	How data science can be used in medical industry? Explain in detail? (CO1, K3)	6
3-b.	Explain the role of Data Science in various fields.(CO1, K2)	6
3-c.	Explain various advantages of Transactional Data, also explain how can you convert text data to numerical data in data analytics? (CO2, K2)	6
3-d.	Differentiate between categorical data and numerical data. (CO2, K3)	6
3.e.	How data reduction helps in data preprocessing?(CO3, K3)	6
3.f.	Describe dimensionality, correlation, eigen vectors in data science perspective in detail? (CO4, K2)	6
3.g.	State the components of the dashboard and also explain it's importance in research and decision making with an example? (CO5, K3)	6
SECTION-C		50
4. Answer any <u>one</u> of the following:-		
4-a.	Explain Big Data, and where does it come from? How does it work? (CO1, K3)	10
4-b.	How would you build a model that sends bank customers a text message when fraudulent transactions are detected. Explain in detail? (CO1, K4)	10
5. Answer any <u>one</u> of the following:-		
5-a.	What are the different types of data sources, also explain Data preprocessing steps? (CO2, K3)	10
5-b.	Tabulate the differences between structured, unstructured and semi structured data.(CO2, K2)	10
6. Answer any <u>one</u> of the following:-		
6-a.	When using Python to clean dataset, what are some of the common issues that arise and how do you deal with them? (CO3, K3)	10
6-b.	Describe some common problems that occur during data processing? How can they be fixed?(CO3, K3)	10
7. Answer any <u>one</u> of the following:-		
7-a.	Can PCA be used to reduce the dimensionality of a highly nonlinear dataset? Explain various limitations of PCA.(CO4, K4)	10
7-b.	Differentiate between descriptive analysis, explanative analysis and exploratory analysis. (CO4, K3)	10
8. Answer any <u>one</u> of the following:-		
8-a.	Describe Tableau. What are the advantages of using Tableau? (CO5, K2)	10
8-b.	Explain the uses of calculated fields? What is the main difference between calculated fields and table calculations in Tableau? (CO5, K3)	10