

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

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA
(An Autonomous Institute Affiliated to AKTU, Lucknow)

B.Tech

SEM: IV - THEORY EXAMINATION (2025 - 2026)

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:-

- (1.1) A collection of information about a related topic is referred to as a _____. 1
 (CO1, K1)
- (a) Visualisation
 (b) Analysis
 (c) Conclusion
 (d) Data
- (1.2) Step that is not part of the Data Science process.(CO1, K1) 1
 (a) Discovery
 (b) Model Planning
 (c) Communication Building
 (d) Operationalize
- (1.3) Categorical data contains: (CO2, K1) 1
 (a) Labels
 (b) Heights
 (c) Salaries
 (d) Temperatures
- (1.4) _____ type of data format is JSON an example of. (CO2,K1) 1
 (a) Semi-structured data
 (b) Structured data
 (c) Unstructured data
 (d) Graphical data

- (1.5) Missing values can be handled using.(CO3,K2) 1
- (a) Deletion
 - (b) Imputation
 - (c) Both a and b
 - (d) None
- (1.6) Data integration involves. (CO3,K1) 1
- (a) Splitting data
 - (b) Combining multiple sources
 - (c) Removing duplicates
 - (d) Sorting data
- (1.7) Missing completely at random (MCAR) implies (CO4, K2) 1
- (a) Missing values depend on observed data
 - (b) Missing values depend on unobserved data
 - (c) Missing values occur randomly
 - (d) Missing values are systematic
- (1.8) A major objective of data integration is to improve: (CO4,K2) 1
- (a) Data inconsistency
 - (b) Data redundancy
 - (c) Data accessibility and consistency
 - (d) Data loss
- (1.9) The file format used by a Tableau workbook is: (CO5,K1) 1
- (a) .docx
 - (b) .ppt
 - (c) .twb
 - (d) .exe
- (1.10) Data blending combines data from _____ sources. (CO5,K1) 1
- (a) Multiple
 - (b) Single
 - (c) Hidden
 - (d) Local only

2. Attempt all parts:-

- (2.1) Differentiate between Structured and Unstructured Data. (CO1, K2) 2
- (2.2) Define transactional data. (CO2,K1) 2
- (2.3) Define KDD process. (CO3,K1) 2
- (2.4) Explain advantages and limitations of PCA.(CO4,K2) 2
- (2.5) Discuss the steps involved in connecting different data sources to Tableau. (CO5,K2) 2

SECTION-B 30

Answer any one of the following:-

- (3.1) Explain the five challenges of Big data in terms of V's. (CO1,K2) 6

(3.2)	Describe the six phases of data science lifecycle.(CO1,K2)	6
Answer any one of the following:-		
(3.3)	Differentiate between numeric and categorical data with suitable examples. (CO2,k2)	6
(3.4)	Explain social network data and mention its role in analytics with one use case.(CO2,K2)	6
Answer any one of the following:-		
(3.5)	Explain noisy data and discuss approaches used to reduce or remove it. (CO3,K3)	6
(3.6)	Explain data reduction and discuss its positive and negative aspects in data processing.(CO3,K3)	6
Answer any one of the following:-		
(3.7)	Explain virtualization and hypervisors in cloud computing.(CO4,K3)	6
(3.8)	Differentiate between EBS, EFS, and S3 storage services.(CO4,K2)	6
Answer any one of the following:-		
(3.9)	Description of Data Interpreter and its role in cleaning and preparing datasets in Tableau. (CO5,K2)	6
(3.10)	Explain the architecture of Tableau and discuss its main components in data visualization. (CO5,K2)	6
SECTION-C		50
4. Answer any <u>one</u> of the following:-		
(4.1)	Explain the concept of Data Science and discuss its importance in modern industries with suitable examples.(CO1,K3)	10
(4.2)	Discuss the applications of Data Science in different industries and explain use cases of companies like Facebook, Netflix, Amazon, Uber, and Airbnb. (CO1,K3)	10
5. Answer any <u>one</u> of the following:-		
(5.1)	Compare structured, semi-structured, and unstructured data in terms of storage, processing, scalability, and analysis techniques. (CO2,K4)	10
(5.2)	Compare primary and secondary data sources and analyze their advantages and limitations.(CO2,K4)	10
6. Answer any <u>one</u> of the following:-		
(6.1)	Describe methods used for identifying outliers in a dataset with examples.(CO3,K3)	10
(6.2)	Describe data munging and explain its importance in preparing raw data for analysis.(CO3,K3)	10
7. Answer any <u>one</u> of the following:-		
(7.1)	Differentiate between SQL and NoSQL databases and explain different types of NoSQL databases.(CO4,K2)	10
(7.2)	Explain edge computing in AWS and describe the role of edge locations and regional edge caches.(CO4,K2)	10
8. Answer any <u>one</u> of the following:-		
(8.1)	Explain the introduction and overview of Tableau and its role in data visualization.(CO5,K2)	10

- (8.2) Explain Tableau calculation features including SUM, AVG, and aggregate functions with suitable examples. (CO5,K2) 10

REG_JAN_JUNE_2026