.

Printed Page:-3 **Subject Code:- BCSDS0301N**

Roll. No:

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA
(An Autonomous Institute Affiliated to AKTU, Lucknow)
B.Tech
SEM: III - THEORY EXAMINATION (2025- 2026)
Subject: Foundations of Data Science

Time: 2 Hours **Max. Marks: 50**

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 15

1. Attempt all parts:-

1-a. State whether data warehousing stores data mainly for? [CO1,K1] 1

(a) Transaction Processing
(b) Analytical Processing
(c) Audio Processing
(d) Image Compression

1-b. State whether CSV files are mainly used for? [CO2,K1] 1

(a) Image processing
(b) Audio data
(c) Video streams
(d) Tabular data storage

1-c. State the alternative name for raw data? [CO3,K1] 1

(a) destination data
(b) eggy data
(c) secondary
(d) machine learning

1-d. Select whether component mapped to the horizontal axis in a 2D chart is? [CO4,K1] 1

(a) X variable
(b) Legend
(c) Color scale
(d) Font size

1-e. State which High-level Python package that provides attractive statistical plots and 1

.

themes? [CO4,K1]

(a) Seaborn
(b) SciPy
(c) TensorFlow
(d) OpenCV

2. Attempt all parts:-

2.a.	Distinguish between traditional data and big data based on data volume and variety? [CO1,K2]	2
2.b.	Explain the use of social network data in analytics? [CO2,K2]	2
2.c.	Explain methods to maintain data quality? [CO3,K2]	2
2.d.	Explain the Linear Discriminant Analysis (LDA)? [CO3,K2]	2
2.e.	State suitability of pie chart for proportion display? [CO4,K1]	2

SECTION-B 15

3. Answer any three of the following:-

3-a.	Discuss all phases of Data Science lifecycle and explain any four applications of Data Science in different fields, such as healthcare, finance, transportation, e-commerce and social media? [CO1,K2]	5
3-b.	Discuss the future of Data Science and also Suppose that the mean and standard deviation of the values for the attribute income are $54,000 and $16,000, respectively.Transform the value of $73,600 for income using Z-score Normalization? [CO2,K2]	5
3-c.	Explain how Uber uses Data Science in it's platform and also Partition the given data into 4 bins using Equi-depth binning method and perform smoothing according to the following methods: a) Smoothing by bin mean b) Smoothing by bin boundaries Data:11,13,13,15,15,16,19,20,20,20,21,21,22,23,24,30,40,45,45,45,71,72,73,75 ? [CO2,K2]	5
3.d.	Explain the purpose of the dplyr package in R and list five of its most commonly used functions? [CO3,K2]	5
3.e.	Explain and write the basic Python code structure to create a simple line plot using Matplotlib and also mention any two types of plots that can be easily created using Seaborn? [CO4,K2]	5

SECTION-C 20

4. Answer any five of the following:-

4-a.	Interpret this statement critically, "Not all problems require Big Data or Machine Learning" and also provide three scenarios where traditional data and BI are sufficient and two scenarios where Big Data and ML are necessary? [CO1,K2]	4
4-b.	Compare DBMS and Data Warehouses based on architecture and functionality? [CO1,K2]	4
4-c.	Interpret and explain the types of variables present in a dataset also explain how Data Preprocessing, Data Reduction, and Data mining are interrelated within the KDD process? [CO2,K2]	4
4-d.	Explain multicollinearity and how it is related to redundant variables and also	4

.

	Demonstrate the process of reading data from a CSV file format? [CO2,K2]	
4-e.	Explain the use of Linear Discriminant Analysis (LDA) and its applications? [CO3,K2]	4
4-f.	Compare and contrast between Univariate and Multivariate EDA techniques in R used in analysis? [CO3,K2]	4
4-g.	Describe Heatmaps and Correlograms, and explain how they help in understanding the structure of multivariate data? [CO4,K2]	4
4-h.	Explain one visualization specifically suitable for geospatial data, also give one use case of 3D graphs in data visualization. State when would you use a Line plot instead of a Bar plot? [CO4,K2]	4