Printed Page:- 04 Subject Code:- BCSDS0301

Roll. No:

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA

(An Autonomous Institute Affiliated to AKTU, Lucknow)

B.Tech

SEM: III - THEORY EXAMINATION (2024- 2025)

Subject: Foundations of Data Science

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. Which of the following uses relatively small amount of data to estimate about bigger population (CO1, K2) 1

(a) Inferential

(b) Exploratory

(c) Causal

(d) None of the Mentioned

1-b. Which type of data Hadoop can deal with is (CO1, K1) 1

(a) Structured

(b) Semi – structured

(c) Unstructured

(d) All of the above

1-c. Which of the following is characteristic of Processed Data? (CO2, K1) 1

(a) Data is not ready for analysis

(b) All steps should be noted

(c) Hard to use for data analysis

(d) None of the mentioned

1-d. Example for discrete data _____ (CO2, K1) 1

(a) height of children

(b) The number of children

(c) weight of children

(d) behaviour of children

1-e. Amongst which of the following step is performed by data scientist after acquiring the data? (CO3, K2) 1

(a) Deletion

(b) Data Replication

(c) Data Integration

(d) Data Cleansing

1-f. To remove noise and inconsistent data ____ is needed. (CO3, K4) 1

(a) Data Cleaning

(b) Data Transformation

(c) Data Reduction

(d) Data Integration

1-g. Which of the following implies no relationship with respect to correlation? (CO4, K2) 1

(a) Cor(X, Y) = 1

(b) Cor(X, Y) = 0

(c) Cor(X, Y) = 2

(d) All of the mentioned

1-h. Lists can be created using the _____ function. (CO4, K1) 1

(a) Matrix.li

(b) Matrix.lists

(c) Lists.matric

(d) List

1-i. Which of the following is a problem in Big Data Visualization? (CO5, K4) 1

(a) Structured Data

(b) Scaled Data

(c) Visual Noise

(d) Clustering

1-j. Which of the following is lattice command for producing a scatterplot? (CO5, K1) 1

(a) plot()

(b) lm()

(c) xyplot()

(d) All of the above

2. Attempt all parts:-

2.a. What is Business Intelligence? Name some BI tools ? (CO1, K2) 2

2.b. Explain High Dimensional data. (CO2, K2) 2

2.c.	How you treat the outliers in your data? (CO3, K3)	2
2.d.	Difference between positive and negative correlation? (CO4, k4)	2
2.e.	Name few visualization packages in R. (CO5, K1)	2

SECTION-B 30

3. Answer any five of the following:-

3-a.	Explain the Implicit and Explicit crowdsourcing with example. (CO1, K2)	6
3-b.	Differentiate between collaborative and content based filtering. (CO1, K2)	6
3-c.	What are the differences between Data Science, Machine Learning, and Artificial intelligence? (CO2, K2)	6
3-d.	Why should a data scientist understand different file formats? (CO2, K2)	6
3.e.	What do you understand about a fact table in the context of a data warehouse? What are the different types of fact tables? (CO3, K2)	6
3.f.	How do you remove whitespaces from the column? Give example. (CO4, K5)	6
3.g.	What is the difference between histogram and scatterplot? What are the functions used in the visualization these plots? (CO5, K1,K4)	6

SECTION-C 50

4. Answer any one of the following:-

4-a.	What do you mean by variety ,value and veracity in Big data? How do you relate telecommunication data as Big data? (CO1, K1, K3)	10
4-b.	What are the different types of data analysis and explain all in detail? (CO1, K2)	10

5. Answer any one of the following:-

5-a.	Explain the generation and analytics process of social network data with an example? (CO2, K3)	10
5-b.	What are the various advantages and disadvantages of data mining? (CO2, K2)	10

6. Answer any one of the following:-

6-a.	What are the missing values? and How do you handle missing values? (CO3, K3)	10
6-b.	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 (CO3, K3)	10

7. Answer any one of the following:-

7-a.	Explain PCA in detail and also explain how will you select the dimensions of the input vector in PCA? (CO4, K4)	10
7-b.	How do you make a frequency table in R? (ii)What is the use of Adorn function? (CO4, K3)	10

8. Answer any one of the following:-

8-a.	What is time series plot? Explain using example. (CO5, K4)	10

8-b. List down at least 5 libraries in R used for data visualization. Explain them briefly. 10 (CO5, K1)