Printed Page:- 03

Subject Code:- ACSDS0301N

Roll. No:

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA

(An Autonomous Institute Affiliated to AKTU, Lucknow)

B.Tech

SEM: III - THEORY EXAMINATION (2023 - 2024)

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 step is performed by data scientist after acquiring the data. (CO1) 1

(a) Data Cleansing

(b) Data Integration

(c) Data Replication

(d) All of the mentioned

1-b. Which is the most significant language for Data Science? (CO1) 1

(a) R

(b) Ruby

(c) Java

(d) None of these

1-c. Point out the correct statement. (CO2) 1

(a) Raw data is original source of data

(b) Preprocessed data is original source of data

(c) Raw data is the data obtained after processing steps

(d) None of the mentioned

1-d. What does "mode" mean? (CO2) 1

(a) average

(b) middle

(c) Most frequent Number
(d) variation

1-e. What are some examples of data quality problems: (CO3) 1

(a) Noise and outliers
(b) Duplicate data
(c) Missing values
(d) All of the Above

1-f. _____ studies the collection, analysis, interpretation or explanation, and presentation of data (CO3) 1

(a) Statistics
(b) Visualization
(c) Data Mining
(d) Clustering

1-g. Joining two lists can be achieved either using the _____ function. (CO4) 1

(a) Concat
(b) Join
(c) Reduce
(d) Delete

1-h. PCA is a _____. (CO4) 1

(a) Non linear method
(b) Linear method
(c) Continuous method
(d) Repeated method

1-i. Which of the following are Use of data visualization. (CO5) 1

(a) See context of data
(b) Clear data understanding
(c) finding pattern in data
(d) all of above

1-j. A useful chart for displaying multiple variables is the ..... (CO5) 1

(a) scatter chart
(b) scatter chart matrix
(c) two-dimensional graph
(d) stacked column and bar chart

2. Attempt all parts:-

2.a. What are the goals of data science? (CO1) 2

2.b. Define Secondary Data. (CO2) 2

2.c. What are major tasks in data Preprocessing? (CO3) 2

2.d. What does linear correlation represent? (CO4) 2

2.e.	What information could you gain from a box-plot? (CO5)	2

SECTION-B 30

3. Answer any five of the following:-

3-a.	Differentiate between analysis and analytics? (CO1)	6
3-b.	What is recommendation system, Give types supported by examples? (CO1)	6
3-c.	Differentiate between discrete and continuous data? (CO2)	6
3-d.	What is data manipulation and how is it performed? (CO2)	6
3.e.	What do you understand about metadata and why is it used for? (CO3)	6
3.f.	What is the importance of dimensionality reduction? (CO3)	6
3.g.	What are the basic plots used in R programming? (CO5)	6

SECTION-C 50

4. Answer any one of the following:-

4-a.	Describe the skillset needed for Data scientist and explain some tools used in visualization process? (CO1)	10
4-b.	What are the current data security risk while performing data science processes with examples? (CO1)	10

5. Answer any one of the following:-

5-a.	Write steps for reading a pdf, csv and json file in python? (CO2)	10
5-b.	What is data manipulation and what purpose it serves? (CO2)	10

6. Answer any one of the following:-

6-a.	What are the missing values? and How do you handle missing values? (CO3)	10
6-b.	What is an outlier? How you detect outliers in your data? (CO3)	10

7. Answer any one of the following:-

7-a.	Distinguish between dimensionality reduction and numerosity reduction? (CO4)	10
7-b.	What is Multivariate analysis? What are the common ways to do Multivariate analysis? (CO4)	10

8. Answer any one of the following:-

8-a.	Explain about the benefits of interactive data visualization. What are some important features of a good data visualization? (CO5)	10
8-b.	Compare R and Python with respect to visualization. (CO5)	10