

Printed Page:-

Subject Code:- ACSDS0301N

Roll. No:

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

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA

(An Autonomous Institute Affiliated to AKTU, Lucknow)

B.Tech

SEM: III - CARRY OVER THEORY EXAMINATION - APRIL 2023

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 focuses on the discovery of (previously) unknown properties on the data [CO1] 1
- (a) Data mining
 - (b) Big Data
 - (c) Data wrangling
 - (d) Machine Learning
- 1-b. Hadoop works in [CO1] 1
- (a) Centralized processing fashion
 - (b) master – slave fashion
 - (c) worker/slave fashion
 - (d) All of the mentioned
- 1-c. Which of the following is characteristic of Processed Data? [CO2] 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. What is the use of data cleaning? [CO2] 1
- (a) To remove the noisy data
- (b) Transformations to correct the wrong data.
- (c) Correct the inconsistencies in data
- (d) All of the above
- 1-e. A data warehouse is which of the following? [CO3] 1
- (a) Can be updated by end users.
- (b) Contains numerous naming conventions and formats.
- (c) Organized around important subject areas.
- (d) Contains only current data.
- 1-f. The operation of moving from finer-granularity data to a coarser granularity (by means of aggregation) is called a _____ [CO3] 1
- (a) Rollup
- (b) Drill down
- (c) Dicing
- (d) Pivoting
- 1-g. Which of following is a problem that missing data could cause? [CO4] 1
- (a) Some visualizations will not work with missing data
- (b) Some types of data cleaning steps won't work with null values
- (c) Machine learning models will break when they encounter null values
- (d) All of the above
- 1-h. Where can the data be updated? [CO4] 1
- (a) Informational Environment
- (b) Datawarehouse environment
- (c) Operational environment
- (d) Data mining environment
- 1-i. Which of the following is not a recommended type of graph? [CO5] 1
- (a) Pie chart
- (b) Histogram
- (c) Barchart
- (d) all of the above

1-j.	Velocity is the speed at which the data is processed [CO5]	1
	(a) TRUE	
	(b) FALSE	
	(c) All	
	(d) None	

2. Attempt all parts:-

2.a.	Name few Data Science tools ? [CO1]	2
2.b.	What is Primary Data? [CO2]	2
2.c.	List any five applications of data warehouse? [CO3]	2
2.d.	What are outliers? [CO4]	2
2.e.	What is the lattice package in R used for? [CO5]	2

SECTION B

30

3. Answer any five of the following:-

3-a.	What are the prerequisites for Data Science? [CO1]	6
3-b.	Discuss all phases of Data Science lifecycle? [CO1]	6
3-c.	What kind of data is HTML files, explain briefly? [CO2]	6
3-d.	Differentiate between categorical and graphical data? [CO2]	6
3.e.	Differentiate between data warehouse and database? [CO3]	6
3.f.	Difference between univariate and multivariate analysis? [CO4]	6
3.g.	Explain what is data visualization and its techniques? [CO5]	6

SECTION C

50

4. Answer any one of the following:-

4-a.	Briefly explain the need of data science with respect to its application areas? [CO1]	10
4-b.	Discuss various model planning tools in detail? [CO1]	10

5. Answer any one of the following:-

5-a.	Describe briefly the steps involved in data manipulation? [CO2]	10
5-b.	Write steps for reading a pdf, csv and json file in python? [CO2]	10

6. Answer any one of the following:-

6-a.	In real world data, tuples with missing values for some attributes are a common occurrence, describe various methods for handling this problem? [CO3]	10
------	---	----

6-b. Explain The Functionality Of OLAP with its types? [CO3] 10

7. Answer any one of the following:-

7-a. In real world data, tuples with missing values for some attributes are a common occurrence. Describe various methods for handling this problem? [CO4] 10

7-b. Distinguish between good and bad data ? What are the problems of data analysis? [CO4] 10

8. Answer any one of the following:-

8-a. What is time series plot. Explain using example? [CO5] 10

8-b. Explain, you want to a subset of employees who are above 30 years and who make salary less than 10000. How will you execute this in R, Explain with complete steps? [CO5] 10

COP 2022-23 July-Dec