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. 20**SECTION-A** 1. Attempt all parts:-1-a. Which of the following uses data on some object to predict values for other object 1 (CO1, K2) Inferential (a) (b) Exploratory Predictive (c) None of the mentioned (d) 1-b. Which of the following are the Data Sources in data science (CO1, K2) 1 (a) Primary Secondary (b) Both A and B (c) None Of the above (d) 1-c. Choose the incorrect property of the data warehouse. (CO2, K2) 1 (a) Collection from heterogeneous sources (b) Subject oriented (c) Time variant (d) Volatile 1-d. What is the use of data cleaning? (CO2, K3) 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.
 The generalization of cross-tab which is represented visually is \_\_\_\_\_\_ 1

 which is also called as data cube. (CO3, K1)
 1

- (a) ultidimensional cube
- (b) Two dimensional cube
- (c) N-dimensional cube
- (d) Cuboid
- 1-f.The operation of moving from finer-granularity data to a coarser granularity (by<br/>means of aggregation) is called a \_\_\_\_\_ (CO3, K3)1
  - (a) Rollup
  - (b) Drill down
  - (c) Dicing
  - (d) Pivoting

1-g.

- \_\_\_\_\_is a tool which is used to reduce the dimension of the data. (CO4, K2) 1
- (a) Principal components analysis
- (b) Product Components analysis
- (c) Principle Components analysis
- (d) Pre Complex analysis
- 1-h. A multivariate statistical technique for studying interrelationships among variables, usually for discovering underlying constructs or data reduction is known as: (CO4, K2)
  - (a) Multiple regression
  - (b) Factor analysis
  - (c) Discriminant analysis
  - (d) Canonical correlation analysis
- 1-i. Which of the following is appropriate to graph a single continuous variable? (CO5, K5)
  - (a) Waffle chart
  - (b) Histogram
  - (c) Pie Chart
  - (d) none

1-j. Which of the following is not a recommended type of graph? (CO5, K2)

1

1

1

- (a) Pie chart
- (b) Histogram
- (c) Barchart
- (d) to create histogram
- 2. Attempt all parts:-

2.a.	What is machine learning? What is its role in data Science? (CO1, K3)	2
2.b.	Why is Spatial Data Science important in business? (CO2, K5)	2
2.c.	List the types of OLAP servers. (CO3, K1)	2
2.d.	What is the use of summary()? (CO4, K3)	2
2.e.	Name few visualization packages in R. (CO5, K1)	2
<b>SECTIO</b>	<u>N-B</u>	30
3. Answe	r any <u>five</u> of the following:-	
3-a.	Explain how Uber and Facebook are using data science techniques for data analytics. (CO1, K3)	6
3-b.	Discuss the data science lifecycle model in brief? (CO1, K2)	6
3-c.	what are the different tasks in Data Mining. (CO2, K1)	6
3-d.	Compare and contrast between transactional and spatial data, with examples? (CO2, K4)	6
3.e.	Differentiate between data warehouse and database. (CO3, K4)	6
3.f.	Discuss a)Filter function b) Arrange function (CO4, K2)	6
3.g.	When will you use a histogram and when will you use a bar chart? Explain with an example. (CO5,K3)	6
<u>SECTIO</u>	<u>N-C</u>	50
4. Answe	r any <u>one</u> of the following:-	
4-a.	Explain Hadoop Distributed File System in detail. (CO1, K2)	10
4-b.	Why do you prefer the Big Data for social networking websites? What are the 5V's associated with it? (CO1, K5)	10
5. Answe	r any <u>one</u> of the following:-	
5-a.	Write steps for reading a pdf, csv and json file in python. (CO2, K2)	10
5-b.	What is data manipulation and what purpose it serves? (CO2, K4)	10
6. Answer any <u>one</u> of the following:-		
6-a.	<ul> <li>(i)Given the following data (in increasing order) for the attribute age: 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46,</li> <li>52, 70. Use min-max normalization to transform the value 35 for age onto the</li> </ul>	10
	range	
	[0.0,1.0]. (ii) Dicuss some of the methods for data transformation. (CO3, K5)	
6-b.	In real-world data, tuples with missing values for some attributes are a common occurrence. Describe various methods for handling this problem. (CO3, K2)	10
7. Answe	r any <u>one</u> of the following:-	
7-a.	What does the janitor package do in R and name some functions and explain with example. (CO4, K2)	10
7-b.	Explain PCA in detail and also explain how will you select the dimensions of the input vector in PCA? (CO4, K4)	10

•

8. Answer any one of the following:-

•

- 8-a. Discuss all 5 types of advanced Visualization plots. (CO5, K2) 10
- 8-b. List down at least 5 libraries in R used for data visualization. Explain them briefly. 10 (CO5, K2)

cop. tute