Name: Enrolment No:	

UPES

End Semester Examination, DEC 2024

Course: Data Environment **Semester : III**
Program: BBA_ABD **Time : 03 hrs.**
Course Code: DSQT2003_3 **Max. Marks: 100**

Instructions:

1. **All questions must be answered except for Questions 17 and 19.**
2. **For Questions 17 and 19, choose one option each:**
   Question 17: Answer either 17a or 17b.
   Question 19: Answer either 19a or 19b.

SECTION A
10Qx2M=20Marks

S. No.		Marks	CO
Q1	Which of the following is **NOT** a data classification based on nature/format? a) Structured b) Passive c) Unstructured d) Semi-structured data	2	CO1
Q2	Which one of the following is not suited to non-relational databases? a) Customer demographic details b) Social Media posts c) Product manuals d) Both b) and c)	2	CO1
Q3	Which of the following includes all forms of data? a) Unstructured data b) Big-data c) Semi-structured data d) Metadata	2	CO1
Q4	Determine the semi-structured data type among the following. a) A social media post b) A customer review c) A marketing email d) All of the above	2	CO1
Q5	What distinguishes meta-data from other forms?	2	CO1

	a) Highest in volume b) Highest in complexity c) Describes other forms of data d) Analyzes other forms of data		
Q6	Choose the correct order of general hierarchy between storage systems. a) Database -> Data warehouse -> Data mart b) Data mart -> Data warehouse -> Database c) Data warehouse -> Database -> Data mart d) All of the above	**2**	**CO3**
Q7	What is the difference between a data warehouse and a data lake? a) Data warehouses store data, while lakes manage data. b) Data warehouse manages data while lakes analyze data. c) Data warehouses deal with formatted data, while lakes store unformatted data. d) Data warehouses deal with unformatted data, while lakes store formatted data.	**2**	**CO3**
Q8	Which of the following is an example of descriptive analytics? a) Predicting sales for the next quarter b) Identifying reasons for a decline in sales c) Summarizing last year's sales trends d) Optimizing supply chain operations	**2**	**CO4**
Q9	What is the primary purpose of diagnostic analytics? a) To summarize past performance b) To identify the root cause of an event c) To predict future outcomes d) To automate decision-making processes	**2**	**CO4**
Q10	Which type of data analytics is used to answer the question, "What is likely to happen in the future?" a) Descriptive analytics b) Diagnostic analytics c) Predictive analytics d) Prescriptive analytics	**2**	**CO4**
SECTION B **4Qx5M= 20 Marks**			
Q11	State and describe the five Vs of Big Data.	**5**	**CO1**
Q12	Compare and contrast tier 1, 2, and 3 of DBMS architectures.	**5**	**CO3**
Q13	Describe the key components of a data warehouse. How is it different from a database in terms of purpose and structure?	**5**	**CO3**

Q14	Differentiate between predictive analytics and prescriptive analytics. Provide an example of where each type can be applied in a business scenario.	**5**	**CO4**
SECTION-C 3Qx10M=30 Marks			
Q15	What are the key challenges businesses face in collecting data? Discuss any two techniques for data collection, highlighting their advantages and disadvantages.	**10**	**CO2**
Q16	Discuss the steps involved in building a predictive analytics model. Highlight the importance of training and testing datasets and the evaluation metrics used to measure model performance.	**10**	**CO4**
Q17a	Explain the ETL (Extract, Transform, Load) process in the context of data warehousing. Why is it critical, and what are the main challenges businesses face during each stage?	**10**	**CO3**
	(Or)		
Q17b	What is "bad data," and how does it impact businesses? Discuss different types of bad data and provide examples for each.		
SECTION-D 2Qx15M= 30 Marks			
Case	NexiCare Health Solutions is a rapidly growing healthcare company specializing in telemedicine and wearable health devices. The company collects large amounts of patient data through wearable devices, online consultations, and surveys. NexiCare aims to leverage this data to provide personalized healthcare recommendations and improve operational efficiency. Recently, the company has faced challenges in its data management and analytics processes, impacting its ability to make accurate predictions and strategic decisions. NexiCare has identified inconsistencies in patient records. Duplicate entries, missing values in health metrics (e.g., heart rate and blood pressure), and outdated contact information are common. This has resulted in delayed consultations and reduced customer satisfaction. Data is collected from multiple sources, including wearable devices, telemedicine sessions, and feedback surveys.The storage system includes a combination of databases for operational data and a centralized data		**CO1, CO2, CO3, CO4**

	warehouse for analytics. However, the ETL process used to integrate data from various sources has been inefficient, leading to delays in analysis. NexiCare launched a predictive model based on historical data to identify patients at high risk of chronic illnesses. However, issues with data quality have led to false positives and negatives in predictions. For example, the model incorrectly flagged 15 healthy patients as high-risk while missing 10 truly high-risk cases out of a total of 100 patients evaluated. Despite having a data warehouse and analytics tools like Tableau, the company's descriptive and diagnostic analytics reports often fail to provide actionable insights due to inconsistent data inputs and limited staff expertise in analytics. NexiCare uses a relational database for patient records but has recently considered adopting a non-relational database for wearable device data due to its unstructured nature. The decision has been stalled because of a lack of understanding of these systems' differences and use cases. Based on these challenges, answer the following questions.		
Q18	a) Discuss the differences between and impact of true positives, true negatives, false positives, and false negatives. b) Define the formulas for precision, recall, and F1-score and calculate the same based on the following details: Total patients evaluated: 100 True Positives: 20 False Positives: 15 False negatives: 10	**15**	**CO1, CO2, CO3, CO4**
Q19a	Analyze the challenges NexiCare Health Solutions faced regarding data collection and storage. Discuss how their ETL process and data warehouse integration inefficiencies have impacted their ability to deliver accurate analytics. Suggest improvements for their data storage and ETL processes.	**15**	
	(Or)		
Q19b	Discuss the importance of descriptive, diagnostic, predictive, and prescriptive analytics in healthcare, using NexiCare's situation as an example. Propose how NexiCare could leverage these analytics to improve decision-making and provide better healthcare services to patients.		