Seat No.: _____ Enrolment No._____

GUJARAT TECHNOLOGICAL UNIVERSITY

MBA– SEMESTER –III-EXAMINATION – WINTER-2023

Subject Code:4539251 **Date: 05-12-2023**

Subject Name: Data Warehousing and Data Mining

Time:10:30 AM TO 1:30 PM **Total Marks: 70**

Instructions:

1. **Attempt all questions.**
2. **Make suitable assumptions wherever necessary.**
3. **Figures to the right indicate full marks.**
4. **Use of simple calculators and non-programmable scientific calculators are permitted.**

Q.1	**(a)**	Give definition of following terms: 1. Tuple 2. Domain 3. Correlated subquery 4. Meta data 5. Dimensional modelling 6. Classification and prediction 7. Data mart	**14**
Q.2	**(a)**	What is database normalization, and why is it important in database design?	**07**
	(b)	Explain the difference between SQL's DDL (Data Definition Language) and DML (Data Manipulation Language).	**07**
		OR	
	(b)	What are the common data types used in SQL, and how do they affect database design?	**07**
Q.3	**(a)**	Discuss the significant distinctions between Data Warehousing (DWH) and Online Transaction Processing (OLTP) oriented database management systems (DBMS) environments.	**07**
	(b)	What is dimensional modeling, and how does it differ from entity-relationship modeling in traditional database design?	**07**
		OR	
Q.3	**(a)**	What is data extraction, and why is it a crucial step in the data warehousing process?	**07**
	(b)	What are the typical sources from which data is extracted in an ETL process, and why is data extraction important?	**07**
Q.4	**(a)**	Write short note on web mining & Multimedia with proper examples.	**07**

	(b)	What is frequent pattern mining, and how does it relate to data mining and association rule mining?	**07**

OR

Q.4	**(a)**	Write short note on Outlier analysis with examples.	**07**
	(b)	Write short note on K-means clustering with examples.	**07**
Q.5		A telecommunications company, TelecomConnect, is facing a growing problem of customer churn, where customers are switching to competing service providers. To combat this issue, the company has decided to leverage data mining techniques to predict and reduce customer churn. TelecomConnect has collected a dataset with information on customer demographics, usage patterns, billing history, and customer service interactions. The dataset includes both current customers and those who have churned.	**14**
	(a)	What types of data are available in the dataset, and what are the key attributes?	
	(b)	Are there any data preprocessing steps required before mining the data?	

OR

Q.5	**(a)**	Can you identify any patterns or trends in the data that may be associated with churn?	**07**
	(b)	How would you handle missing values and outliers in the dataset?	**07**

****