Q.P. Code: 25960

Time: 03 Hours Marks: 80

Note: 1. Question 1 is compulsory

2. Answer any three out of remaining questions.

- Q1 A) Consider following dimensions for a Supermarket chain: Product, Store, Time and [10] Promotion. With respect to this business scenario, answer the following questions. Clearly state any reasonable assumptions you make.
 - (a) Design an information package diagram for this business scenario.
 - (b) Design a snowflake schema for the data warehouse, clearly depicting the fact table(s), Dimension table(s), their attributes and measures.
 - B) Consider the 5 transactions given below. If minimum support is 30% and minimum [10] confidence is 80%, determine the frequent itemsets and association rules using Apriori algorithm.

Transaction	Items		
T1	Milk, Jam, Butter		
T2	Milk, Butter		
T3	Milk, Cheese, Butter		
T4	Biscuit, Milk,		
T5	Biscuit, Cheese		

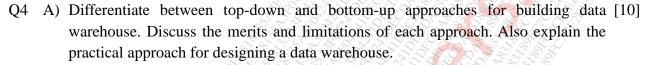
- Q2 A) Consider a Data Warehouse for a sport manufacturing company storing sales details [10] of various sports equipments sold, and the time of the sale. Using this example describe the following OLAP operations:
 - (i) Slice (ii) Dice (iii) Rollup (iv) Drill Down (v) Pivot
 - B) What is data mining? Describe the steps involved in the data mining when viewed [10] as a process of knowledge discovery. Present an example where data mining is crucial to success of business.
- Q3 A) What is Dimension Modeling? What is slowly changing dimensions? How this [10] problem is solved? Give example.
 - B) Given is the training data for height classification, classify the **tuple t= <Arvish**, [10] **M**, **1.97** > using Bayesian classification.

Name	Gender	Height	Output
Reena	F	1.6 m	Short
Mahesh	M	2 m	Tall
Tina	F	1.9 m	Medium
Meeta	F	1.88 m	Medium

Page 1 of 2

Q.P. Code: 25960

Siya	F	1.7 m	Short
Vikram	M	1.85 m	Medium
Lakshmi	F	1.6 m	Short
Andrew	M	1.7 m	Short
Henry	M	2.2 m	Tall
Akhil	M	2.1 m	Tall
Lata	F	1.8 m	Medium
Siraj	M	1.95 m	Medium
Rita	F	1.9 m	Medium
Kriti	F	1.8 m	Medium
Srishti	F	1.75 m	Medium



- B) What is clustering? Explain K means clustering algorithm. [10] Suppose the data for clustering is {2, 4, 10, 12, 3, 20, 30, 11, 25, 5, 36, 41, 14}. Assuming number of clusters to be 2 i.e. K = 2, cluster the given data using above algorithm.
- Q5 A) Describe different steps of ETL (Extraction, Transformation and Loading) cycle in [10] Data Warehousing for a pharmaceutical company.
 - B) What is Web Mining? Explain Web Usage Mining.

[10]

Q6 Write short note on the following (Answer any **FOUR**)

[20]

- a) Hierarchical Clustering Algorithms
- b) Metadata in Data Warehouse
- c) Decision tree Classification Model
- Snapshot and Transaction tables
- e) Data Exploration
