

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -1 EXAMINATION- 2016

B.Tech. (IT), 6th Semester

COURSE CODE: 10B22CI622

MAX. MARKS: 15

COURSE NAME: Data Mining

COURSE CREDITS: 4

MAX. TIME: 1 HR

Note: All questions are compulsory. Carrying of mobile phone during examinations will be treated as case of unfair means.

Q.1. (a) Explain in brief how the evolution of database technology led to data mining?
 (b) Write the name of steps involved in data mining when viewed as a process of knowledge discovery. [2 Marks]

Q.2. Suppose that the data for analysis includes the attribute *age*. The *age* values for the data tuples are (in increasing order) 13, 21, 22, 25, 25, 30, 30, 33, 35, 35, 35, 36, 40, 52, 70.

(a) Estimate the *mean* of the data?
 (b) Find the first quartile (Q_1) and the third quartile (Q_3) of the data?
 (c) Give the *five-number summary* of the data. [3 Marks]

Q.3. Suppose a group of 12 *sales price* records has been sorted as follows: 5, 10, 11, 13, 15, 35, 50, 55, 72, 92, 204, 215. Partition them into three bins by each of the following methods:

(a) equal-frequency (equal-depth) partitioning OR Equal-width partitioning
 (b) Clustering [2 Marks]

Q.4. A 2 X 2 contingency table for the data is shown below. Are *gender* and *preferred Reading* correlated?

Reading / Gender	Male	Female	Total
Fiction	250	200	450
Non-Fiction	50	1000	1050
Total	300	1200	1500

[3 Marks]

Q.5. Identify the prime features that distinguish data warehouses from other data repository systems. Briefly compare between OLTP and OLAP systems from the following perspective: (i) Users and system orientation, (ii) Data contents, (iii) Database design. Draw a figure for Star schema and Snowflake schema of a data warehouse (Consider any data warehouse of your choice) for sales records. [5 Marks]