2ⁿᵈ Semester MBA (Business Analytics)
SUPPLEMENTARY EXAMINATION- SEPT 2019

PAPER CODE: MB 204   TITLE OF PAPER- **Data Warehousing and Data Mining**

Time: *3:00 Hours*   *Max. Marks: 60*

Note: Write your Roll no. on the top of this question paper.
Marks are indicated against each question. Parts of a question must be answered together.

Q1. Answer **any five** the following questions: [5* 6 marks = 30 marks]

a) In real-world data, tuples with *missing values* for some attributes are a common occurrence. Describe various methods for handling this problem.

b) Suppose that a data warehouse consists of the three dimensions *time, doctor,* and *patient,* and the two measures *count* and *charge,* where *charge* is the fee that a doctor charges a patient for a visit.

   (i) Draw a star schema diagram for the above data warehouse.

   (ii) Starting with the base cuboid [*day; doctor; patient*], what specific *OLAP operations* should be performed in order to list the total fee collected by each doctor in 2004?

c) Differentiate between classification and clustering with examples.

d) Differentiate between OLAP and OLTP systems.

e) Using Equi-depth binning method, partition the data given below into 4 bins and perform the smoothing according to the following methods: i) smoothing by bin means ii) smoothing by bin median iii) smoothing by bin boundaries
   24, 25, 26, 27, 28, 56, 67, 70, 70, 75, 78, 89, 89, 90, 91, 94, 95, 96, 100, 102, 103, 107, 109, 112

f) Write short notes on the following: (i) Data discretization (ii) Numerosity reduction.

Q2. Attempt **any one** out of the following: [10 marks]

a) What is Data Warehouse? Explain the three-tier data warehouse architecture with the help of a diagram.

b) Describe the following approaches to clustering: *partitioning* methods and *hierarchical* methods. Give examples in each case.

Q3. Attempt **any four** out of the following questions: [4* 5 marks = 20 marks]

a) Describe ROLAP and MOLAP models.

b) Explain the representation of text documents, weighting scheme and retrieval of documents using vector-space model with an example.

c) Explain concept hierarchy with the help of an example.

d) Describe the operations of data cube with the help of suitable examples.

e) Suppose that the data mining task is to cluster the following eight points (with $(x; y)$ representing location) into three clusters.
   $A1(2; 10); A2(2; 5); A3(8; 4); B1(5; 8); B2(7; 5); B3(6; 4); C1(1; 2); C2(4; 9)$:
   The distance function is Euclidean distance. Suppose initially we assign $A1, B1,$ and $C1$ as the center of each cluster, respectively. Use the *k-means* algorithm to show *only* (a) The three cluster centers after the first round of execution and (b) The final three clusters.

***