Sample Question Format

KIIT Deemed to be University
Online Mid Semester Examination(Spring Semester-2021)

Subject Name & Code: Data Mining and Data Warehousing(IT 3031)
Applicable to Courses: 6TH SEMESTER CSE / IT/CSSE/CSCE

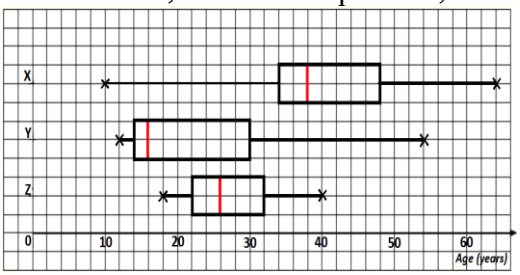
Full Marks=20 **Time:1 Hour**

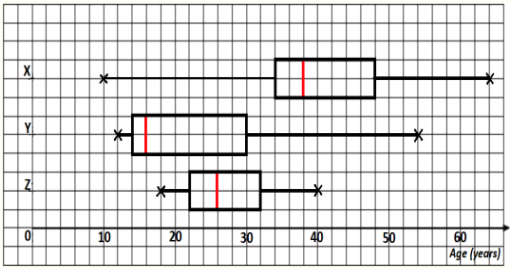
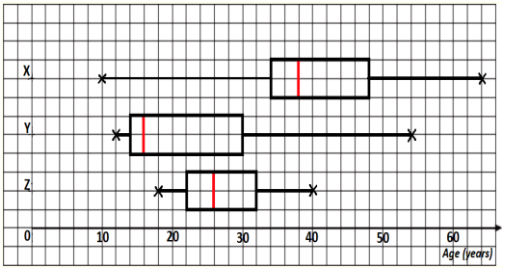
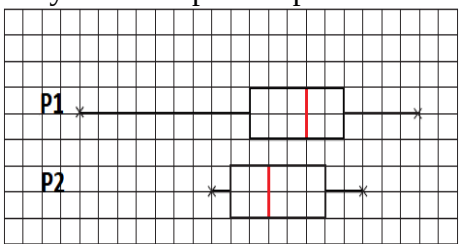
SECTION-A(Answer All Questions. All questions carry 2 Marks)

Time:20 Minutes **(5×2=10 Marks)**

Question No	Question Type(MCQ/SAT)	Question	Answer Key(if MCQ)	CO Mapping
Q.No:1(a)	MCQ	Essential process where intelligent methods are applied to extract data patterns. i) Data warehousing ii) Data mining ii) Text mining iv) Data selection	ii	CO1
	MCQ	The out put of KDD is: i) Data ii)Information iii) Query iv) Useful information	ii	CO1
	MCQ	The analysis performed to uncover interesting statistical correlations between associated-attribute-value pairs is called? i)Mining of Association ii) Mining of Clusters iii) Mining of Correlations iv) None of the above	ii	CO1
	MCQ	Data objects that do not comply with the general behavior or model of the data available. i)Outlier Analysis ii) Evolution Analysis iii) Prediction iv) Classification	i	CO1
Q.No:1(b)	MCQ	A snow-flake schema is which of the following type of tables: i) Fact ii) Dimension iii) Helper **iv)** All of the above	i	CO2
	MCQ	Strategic value of data mining is ……. i) cost-sensitive ii) work-sensitive iii) time-sensitive iv) technical-sensitive	iii	CO2
	MCQ	Classification is i) A subdivision of a set of examples into a number of classes	iii	CO2

		ii) A measure of the accuracy, of the classification of a concept that is given by a certain theory iii) The task of assigning a classification to a set of examples iv) None of these		
	MCQ	A star schema has what type of relationship between a dimension and fact table? i)Many-to-many ii)One-to-one iii)One-to-many iv)All of the above	iii	CO2
Q.No:1(c)	**MCQ**	Choose the correct option that focuses on multiple data item processing which is carried out in data warehouse process is: (A) On Line Analytics Processing (B) On Line Transactional Processing (C) Pre-Processing (D) Visualization on Information Processing	A	CO-2
	MCQ	The process of Data Discretization is known as (A) to scale up data (B) to convert continuous attribute to discrete attribute (C) to convert discrete attribute to continuous attribute (D) None	B	CO-2
	MCQ	The term 'Concept Hierarchy' is useful in (A) data warehouse (B) data normalization (C) multiple level of abstraction (D) to create histogram	C	CO-2
	MCQ	To make analysis better which of the following type of data is typically stored in a specific format in data warehouse? (A) Informational Data (B) Operational Data (C) Data Mart (D) Normalized Data	A	CO-2
Q.No:1(d)	**MCQ**	In a shoping mall the products are numbered from 01 to 30 are Sampoo's and numbered from 31 to 50 are Conditioners. Which product numbers would you include in a systematic sample of size 10 ? A. 10, 20, 30, 40, 50 B. 01, 06, 11, 16, 21, 26, 31, 36, 41, 46 C. 01, 11, 21, 31, 41 **D.** 05, 10, 15, 20, 25, 30, 35, 40, 45, 50	D	**CO2/CO3**

	MCQ	Our KIIT University employs the following numbers of faculty members in 3 different positions as; Professor: 10, Associate Prof.: 20, Asst. Prof: 20. How many from each positions should be included in a quota sample of size: 25? A. 3, 11, 11 B. 5, 10, 10 C. 10, 5, 10 D. 10, 10, 5	B	**CO1/ CO2/ CO3**
	MCQ	In a picnic trip consists of 40 members of whom 15 are gents. A quota of size 8 is to be selected for site visit. How many ladies and how many gents should be included in the sample? A. 6, 2 B. 5, 3 C. 2, 6 D. 3, 5	B	**CO1/ CO2/ CO3**
	MCQ	In a DMDW class 75 students are present of whom 15 are girls. A quota of size 15 is to be selected for site visit. How many boys and how many girls should be included in the sample? A. 5, 10 B. 10, 5 C. 3, 12 D. 12, 3	D	CO1
Q.No:1(e)	**MCQ**	Three different games are playing by three differnt age group (X, Y, and Z) of the people. Players behaviours are visualizing in the following box ploat with whisker, answer the question;  Which game do you think ***you*** (according to your age) would not be allowed to play? A. Game X B. Game Y C. Game Z D. None	C	CO-1
	MCQ	Three different games are playing by three differnt age group (X, Y, and Z) of the people. Players behaviours are visualizing in the following box ploat	B	CO-1

		with whisker, answer the question;  Which game would ***you*** probably enjoy most? A. Game X B. Game Y C. Game Z D. None		
	MCQ	Three different games are playing by three differnt age group (X, Y, and Z) of the people. Players behaviours are visualizing in the following box ploat with whisker, answer the question;  Which game would ***your parents*** probably enjoy most? A. Game X B. Game Y C. Game Z D. None	A	CO-1
	MCQ	In an industry, the quality inspector will check the two different types of product that are marked in P1 & P2. The box and whisker plots below shows the results of quality tests on the quality of the respective products.  "The inspector might prefer to use the product (P2) because that P2 has the smaller _____, and the larger _____ . , which means that industry is less likely to produce a poor product. A. range, minimum B. range, maximum	A	CO-1

		C. median, range D. lower quartile, median		

SECTION-B(Answer Any One Question. Each Question carries 10 Marks)

Time: 30 Minutes **(2×10=10 Marks)**

<table>
<tr><th>Question No</th><th>Question</th><th>CO Mapping</th></tr>
<tr><td>Q.No: 2</td><td>Given two objects represented by the tuples (22,1,42,10) and (20,0,36,8). Compute Euclidean distance, Manhatton distance and Minkowski distance(q=3) between these two given objects.</td><td>CO1</td></tr>
<tr><td>Q.No: 3</td><td>Use K-means algorithm and Euclidean distance to cluster the following 10 points into 3 clusters.
A1(2,10),A2(9,4),A3(8,4),A4(9,4),A5(5,8),A6(7,5),A7(6,4),A8(1,2),A9 (4,9),A10(6,10).Suppose the initial centers are A1,A4 and A9,run the K-means algorithm for 3 iterations at the end of each iterations show the cluster centers and also the final three clusters.</td><td>CO1</td></tr>
<tr><td>Q.No: 4</td><td>Suppose the task is to cluster the following 8 points (with (x, y) representing location) into 3 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, C1 as centre of each cluster, respectively. Use k-means algorithm to show only
a) The three cluster centres after the first round execution.
The final 3 clusters</td><td>CO1</td></tr>
<tr><td>Q.No: 5</td><td>State and discuss each step in the Apriori algorithm. A data base has following TEN transactions.
<table>
<tr><th>Trans ID</th><th>Items Purchased</th></tr>
<tr><td>101</td><td>Milk, bread, eggs</td></tr>
<tr><td>102</td><td>Milk, Juice</td></tr>
<tr><td>103</td><td>Juice, butter</td></tr>
<tr><td>104</td><td>Milk, bread, eggs</td></tr>
<tr><td>105</td><td>Coffee, eggs</td></tr>
<tr><td>106</td><td>Coffee</td></tr>
<tr><td>107</td><td>Coffee, Juice</td></tr>
<tr><td>108</td><td>Milk, bread, eggs, cookies</td></tr>
<tr><td>109</td><td>Cookies, butter</td></tr>
<tr><td>110</td><td>Milk, bread</td></tr>
</table>
Find all frequent item sets using Apriori algorithm. Use 0.3 for the minimum support value. Find all association rules from all frequent itemsets generated.</td><td></td></tr>
<tr><td>Q.No: 6</td><td>The 'database' below has four transactions. What association rules can be found in this set, if the minimum support (i.e coverage) is 60% and the minimum confidence (i.e. accuracy) is 80% ?
<table>
<tr><th>Transaction ID</th><th>Item List</th></tr>
<tr><td>T1</td><td>K, A, D, B</td></tr>
<tr><td>T2</td><td>D, A, C, E, B</td></tr>
<tr><td>T3</td><td>C, A, B, E</td></tr>
</table></td><td></td></tr>
</table>

		T4	B, A, D		
Q. No:7	Given the transnational Data for a shop. Find the frequent items using Apriori algorithm. Design a Hash Table H_2 for the candidate 2-itemsets.				

Tid	**Items**
T100	I_1, I_3, I_4
T200	I_2, I_3, I_5
T300	I_1, I_2, I_3, I_5
T400	I_2, I_5

Controller of Examinations