Lecturer:	*(Date)*	Approved by:	*(Date)*
(Signature & Fullname)		*(Signature, Position & Fullname)*	

(The above part must be hidden when copying for exam)

	FINAL EXAM	**Semester/Academic year**	**2**	**2019-2020**
BK TP.HCM **UNIVERSITY OF TECHNOLOGY - VNUHCM FACULTY OF CSE**		**Date**		**31/7/2020**
	Course title	**Data mining**		
	Course ID	**CO3029**		
	Duration	**80 mins.**	**Question sheet code**	**100**

Notes: - *Open book*
- *Do not use mobile phones,laptops or any electronic device*
- *Question sheet has 7 pages*
- ***Submit the question sheet together with the answer sheet***

ANSWER SHEET

SECTION 1 (7.0 points)

Question	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Answer	B	D	A	C	D	E	D	E	C	B	D	B	C	B	A

Question	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Answer	B	C	D	B	B	A	B	C	C	B	B	A	C	D	C

SECTION 2 (3.0 points)

Question 2.1 (1.0 point) Give a dataset as follows:

TID	Items bought
1	f, a, c, d, g, i, m, p
2	a, b, c, f, l, m, o
3	b, f , h, j, o
4	b, c, k, s, p
5	a, f, c, e, l, p, m, n

Draw the FP-tree from the above dataset with min_sup = 3?
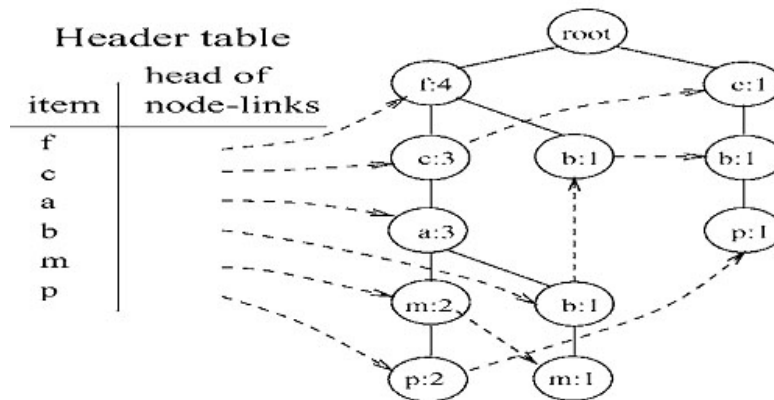
Question 2.2 (2.0 point): Given the ages of students in a data mining course as follows: 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70.

a) (1.0 point) Write down the values in the table.

Mean = 809 / 27 = 29.96

Median = 25

Mode = 25 or Mode = 35

Midrange = 83 / 2 = 41.5

Q1 = 20

Q2 = 25

Q3 = 35

b) (1.0 point) What are the outliers and why, based on interquartile range?

IQR = 15 → outlier is: 70 since 70 > Q3 + IQR * 1.5

--- END ---