1. Question

tid	Itemset
t_1	ABCD
t_2	ACDF
t_3	ACDEG
t_4	ABDF
t_5	BCG
t_6	DFG
t_7	ABG
t_8	CDFG

Transaction Database

(1)

tid	A	B	C	D	E	F	G
t_1	1	1	1	1	0	0	0
t_2	1	0	1	1	0	1	0
t_3	1	0	1	1	1	0	1
t_4	1	1	0	1	0	1	0
t_5	0	1	1	0	0	0	1
t_6	0	0	0	1	0	1	1
t_7	1	1	0	0	0	0	1
t_8	0	0	1	1	0	1	1

Binary Database.

(2.) $t(x)$

A	B	C	D	E	F	G
t_1	t_1	t_1	t_1	t_3	t_2	t_3
t_2	t_4	t_2	t_2		t_4	t_5
t_3	t_5	t_3	t_3		t_6	t_6
t_4	t_7	t_5	t_4		t_8	t_7
t_7		t_8	t_6			t_8
			t_8			

Vertical Database

(3.) Using minimum support 3, Apriori Algorithm find F(3).

Sol There are 3 main steps - Count, Pruning and Joining.

Step 1 → Count Items.

C_1 =

Itemsets	Count	
{A}	5	
{B}	4	
{C}	5	
{D}	6	
{E}	1.	→ remove as it is not satisfying min. support of 3.
{F}	4	
{G}	5	

Step 2 → Filter Items.

L_1 →

Itemset	Count
{A}	5
{B}	4
{C}	5
{D}	6
{F}	4
{G}	5

Step 3 → Join Items.

C_2 →

Itemsets	Count.
{A B}	3
{A C}	3
{A D}	4
{A F}	2
{A G}	2
{B C}	2
{B D}	2
{B F}	2
{B G}	2
{C D}	4
{C F}	2
{C G}	3
{D F}	4
{D G}	3
{F G}	2

Remove - because it did not meet the min. support of 3.

$L_2 \rightarrow$

Itemset	Count
{AB}	3
{AC}	3
{AD}	4
{CD}	4
{CG}	3
{DF}	4
{DG}	3

This is F(2).

Now, continue till F(3).

$C_3 \rightarrow$

Itemset	count
{ABC}	1
{ABD}	2
{ACD}	3
{ACG}	1
{ADf}	2
{ADG}	1
{CDG}	2
{CDF}	2
{DFG}	2

Removing all the C_3 itemsets except {ACD} = 3, because the rest did not meet the min. support of 3.

	Itemset	count
$L_3 \rightarrow$	{ACD}	3

This is F(3).

(4.) FP growth using minimum support of 2.

Frequency pattern set:

Itemset	Count (Frequency)
A	5
B	4
C	5
D	6
E	1
F	4
G	5

Arranging the Itemset in decreasing order of its counts.

Itemset	Count
D	6
A	5
C	5
G_1	5
B	4
F	4
E	1

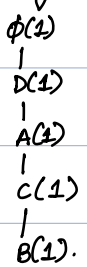
Now as the minimum support is 2, we can eliminate (E) and will not include it in our set. The set L will look like

$L = \{D:6, A:5, C:5, G_1:5, B:4, F:4, E:1\}$.

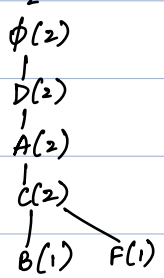
This is frequent pattern set. After this we will create ordered item set.

tid	Itemset	Ordered Itemset
t_1	$\{A, B, C, D\}$	$\{D, A, C, B\}$
t_2	$\{A, C, D, F\}$	$\{D, A, C, F\}$
t_3	$\{A, C, D, E, G_1\}$	$\{D, A, C, G_1\}$
t_4	$\{A, B, D, F\}$	$\{D, A, B, F\}$
t_5	$\{B, C, G_1\}$	$\{C, G_1, B\}$
t_6	$\{D, F, G_1\}$	$\{D, G_1, F\}$
t_7	$\{A, B, G_1\}$	$\{A, G_1, B\}$
t_8	$\{C, D, F, G_1\}$	$\{D, C, G_1, F\}$

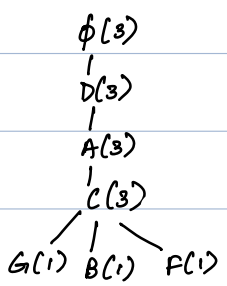
① Inserting set for t_1



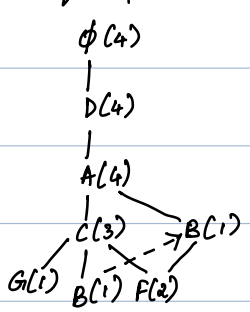
② For t_2



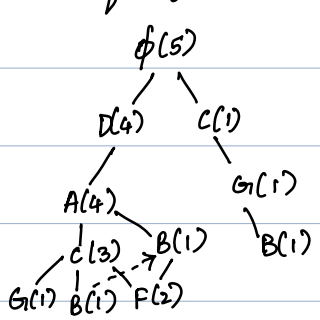
③ for t_3



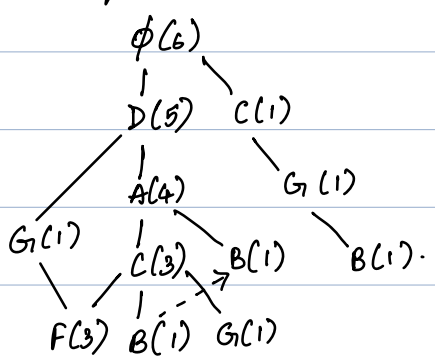
④ for t_4.



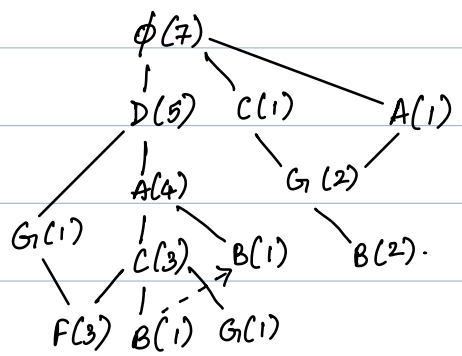
⑤ for t_5



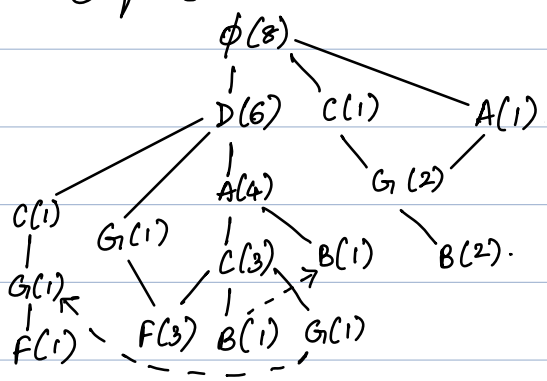
⑥ for t_6



⑦ for t_7



⑧ for t_8



Final FP Tree.

Conditional Pattern Base.

Items	Conditional Pattern Base	conditional Frequent Pattern Tree.
{F}	{(D,A,C:1), (D,A,B:1), (D,C,G1:1), (D,G1:1)}	{D:4}.
{B}	{(D,A,C:1), (D,A:1), (C,G1:1), (A,G1:1)}	–
{G1}	{(D,A,C:1), (D,C:1), (C:1), (D:1), (A:1)}	–
{C}	{(D,A:3), (D:1)}	{D:4}

$\{A\}$	$\rightarrow \{(D:4)\}$	$\longrightarrow \{D:4\}$
$\{D\}$	$\rightarrow$ —	$\longrightarrow$ —

Question

2. Given $x_1 = (0,3)$ $x_2 = (3,3)$ $x_3 = (0,0)$

Centroid $C_1 = (3.5, -1)$

1 → SSE :

$$\text{for } x_1 = \|x_i - C_i\|^2 = \|(0,3) - (3.5,-1)\|^2$$

$$= \|-3.5, 4\|^2 = 12.25 + 16 = \underline{28.25}.$$

$$\text{for } x_2 = \|(3,3) - (3.5,-1)\|^2 = 0.25 + 16 = \underline{16.25}$$

$$\text{for } x_3 = \|(0,0) - (3.5,-1)\|^2 = 12.25 + 1 = \underline{13.25}$$

$$SSE = SSE(x_1) + SSE(x_2) + SSE(x_3)$$

$$= 28.25 + 16.25 + 13.25$$

$$= 57.75$$

The sum of squared errors for the intial cluster assignment is 57.75.

2 → The location of next centroid can be calculated by taking mean of data points.

$$\text{Centroid} = \frac{x_1 + x_2 + x_3}{3} = \left(\frac{0+3+0}{3}, \frac{3+3+0}{3}\right) = (1,2).$$

The centroid after next iteration is $\underline{(1,2)}$.