HW5

Exercise 1

1.1- 15

1.1-15-(a). Which data set(s) will produce the most number of frequent itemsets?

Answer:

The E data set produces more frequent item sets, it has more frequent purchases with the items.

1.1-15-(b).. Which data set(s) will produce the fewest number of frequent itemsets?

Answer:

The dataset D has few frequent item sets, in that dataset items are not occurring oftenly.

1.1-15-(c).. Which data set(s) will produce the longest frequent itemset?

Answer:

The dataset E has the longest frequent itemset, in here the items are more often occurred in many purchases.

1.1-15-(d).. Which data set(s) will produce frequent itemsets with highest maximum support?

Answer:

The B dataset, the support is highest items purchased together in a single transaction in an dataset.

1.1-15-(e).. Which data set(s) will produce frequent itemsets containing items with wide-varying support levels (i.e., items with mixed support, ranging from less than 20% to more than 70%).

Answer:

The dataset E. each item has different support levels at 900.

1.2-1-(a) Using minsup = 3/8, show how the Apriori algorithm enumerates all frequent patterns from this dataset

tid	Itemset
T1	ABCD
T2	ACDF
T3	ACDEG
T4	ABDF
T5	BCG
T6	DFG
T7	ABG
T8	CDFG

Answer:

	A	B	C	D	E	F	G
T1	1	1	1	1	0	0	0
T2	1	0	1	1	0	1	0
T3	1	0	1	1	1	0	1
T4	1	1	0	1	0	1	0
T5	0	1	1	0	0	0	1
T6	0	0	0	1	0	1	1
T7	1	1	0	0	0	0	1
T8	0	0	1	1	0	1	1

Given minimum support = 3

Step1:

Count of each candidate

Itemset	Support , Count
A	5
B	4
C	5
D	6
E	1
F	4
G	5

Here the itemset E will be removed, its considered as infrequent

Step2:

Now compare the candidate support with the minimum support count

The E has the less support than the minimum support
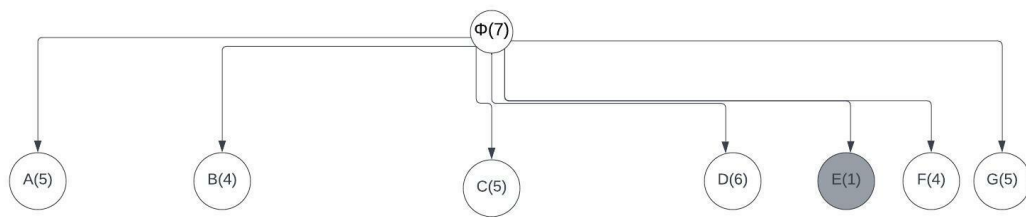
$= $ support(E) $<$ minimum support

$= 1 < 3$

Now this is considered as an infrequent item.

After removing E

Itemset	Support, count
A	5
B	4
C	5
D	6
F	4
G	5

Table L1

Step 3:

To know the set of two frequent itemset, we use

Itemset	Support, count
{A,B}	3
{A,C}	3
{A.D}	4
{A,F}	2
{A,G}	2
{B,C}	2
{B,D}	2
{B,F}	1
{B,G}	2
{C,D}	4
{C,F}	2
{C,G}	3
{D,F}	4
{D,G}	3
{F,G}	2

Now check support >= minimum support then, the itemsets with less support will be infrequent.

Here, these itemsets {A,F},{A,G},{B,C},{B,D},{B,F},{B,G},{C,F},{D,F},{F,G} are infrequent

After removing the infrequent itemsets,

Itemset	Support, Count
{A,B}	3
{A,C}	3
{A,D}	4
{C,D}	4
{C,G}	3
{D,F}	4
{D,G}	3
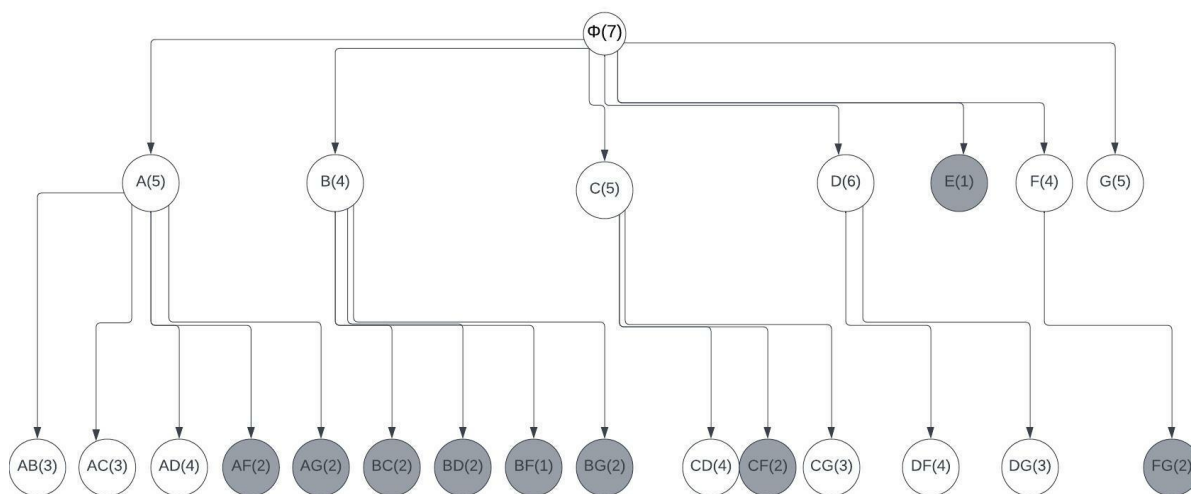
The frequent itemset final tree

Step 4:

Now we will join to generate the candidate set of 3 itemsets.

Itemset	Support, Count
{A,B,C}	1
{A,B,D}	2
{A,C,D}	3
{C,G,D}	2

Now check support >= minimum support then, the itemsets with less support will be infrequent.

In here these {A,B,C},{A,B,D},{C,D,G} are infrequernt.

The final tree is
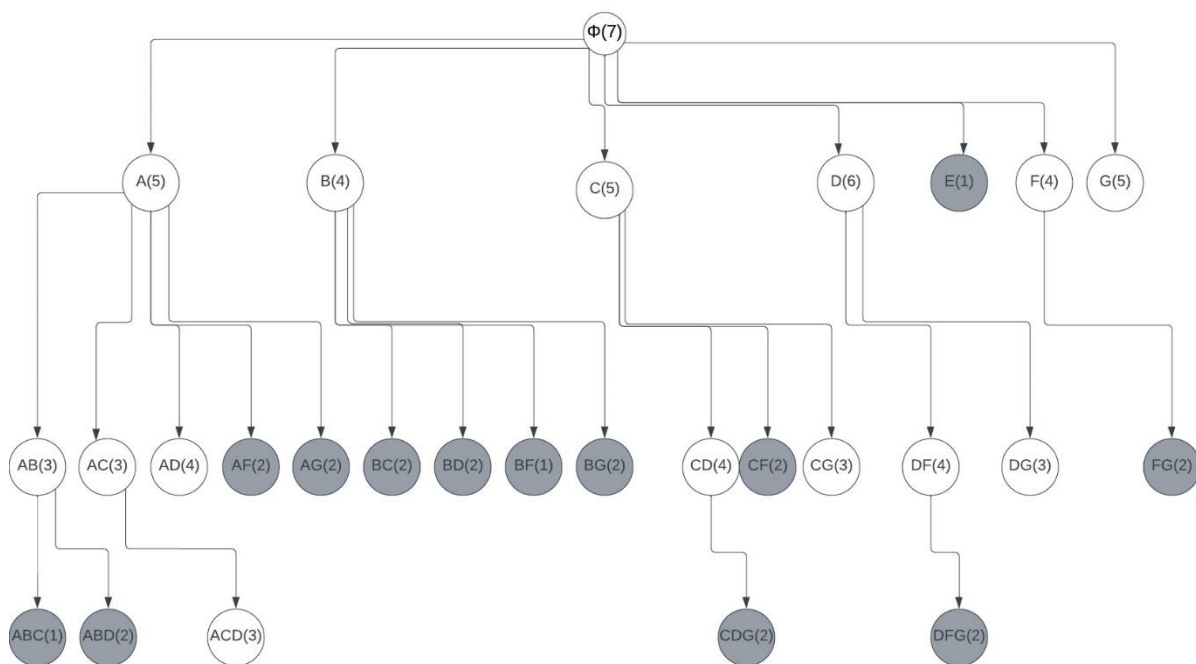
Note :The dark nodes are having less support than the minimum support.

The item sets with minus=3.

Sup	itemsets
6	D
5	A, C, G
4	B, F, AD, CD, DF
3	AB, AC, CG, DG, ACD

1.2-4. Given the database in Table 8.4. Show all rules that one can generate from the set ABE.

Answer:

	A	B	C	D	E
T1	1	0	1	1	0
T2	0	1	1	0	1
T3	1	1	1	0	1
T4	0	1	0	1	1
T5	1	1	1	0	1
T6	1	1	1	1	0

Itemset	Support count
ABE	2
AB	3
AE	2
BE	4
A	4
B	5
E	4

taking a frequent itemset ABE, the rules can be:

$$Confidence(X\text{->}Y) = \frac{Support(XUY)}{Support(X)}$$

$\{AB\} \text{-> } E = \frac{Sup(ABE)}{Sup(AB)} = \frac{2}{3} = 0.66 = 66\%$

$\{AE\} \text{-> } B = \frac{Sup(ABE)}{Sup(AE)} = \frac{2}{2} = 1 = 100\%$

$\{BE\} \text{-> } A = \frac{Sup(ABE)}{Sup(BE)} = \frac{1}{2} = 0.5 = 50\%$

$\{A\} \text{-> } BE = \frac{Sup(ABE)}{Sup(A)} = \frac{1}{2} = 0.5 = 50\%$

$\{B\} \text{-> } AE = \frac{Sup(ABE)}{Sup(B)} = \frac{2}{5} = 0.4 = 40\%$

$\{E\} \text{-> } AB = \frac{Sup(ABE)}{Sup(E)} = \frac{2}{3} = 0.5 = 50\%$