Question1

1.a

support count of {Beer} is **4**

support count of {Diaper} is **4**

support count of {Beer,Tylenol} is **2**

1.b

Confidence({Beer} ⇒ {Diaper})

P(D|B) = P(B,D)|P(D) = 3/4 = **0.75**

Confidence({Beer} ⇒ {Diaper})

P(B|D) = P(B,D)|P(B) = 3/4 = **0.75**

1.c

{Beer, Cheese, Diaper, Tylenol} is close pattern.

For X, if support(super(X)) < support(X), X is a closed cell.

{Beer, Cheese, Diaper, Tylenol} is the longest base cell, it did not have a super(X),while the count is higher than min_sup, so it is closed.

1.d

{Beer, Cheese, Diaper} is close pattern.

For X, if support(super(X)) < support(X), X is a closed cell.

{Beer, Cheese, Diaper, Tylenol} is the super(X) for {Beer, Cheese, Diaper}, while the count of super(X) is 1, 1 < 2, the count of X is 2, so it is closed.

1.e

{Beer, Cheese, Diaper} is **not** a max pattern,

For X, if support(super(X)) < min_sup, X is a max patten cell.

The super cell of {Beer, Cheese, Diaper} is {Beer, Cheese, Diaper, Tylenol}, it is count is 1 equal to the min_sup, so it is not a max pattern.

1.f

For a max pattern, which is frequent,

support(super(X)) < min_sup

min_sup <= support(X)

So a max pattern is also an close pattern.

Question2

2.a

C1: { {B}: 5 {C}: 3 {D}: 4 {E}: 3 {M}: 4 {T}: 3 }

since all element in C1 is higher OR equal than min_sup

L1: **{B,C,D,E,M,T}**

2.b

C2:{

{B,C}: 1; {B,D}: 3;{B,E}: 2;{B,M}: 4;{B,T}: 3

{C,D}: 2; {C,E}: 2; {C,M}: 2; {C,T}: 1;

{D,E}: 2; {D,M}: 3; {D,T}: 1;

{E,M}: 1; {E,T}: 0;

{M,T}: 3; }

Since L2 contain C2's element higher or equal to min_sup

L2: **{{B,D}, {B,M}, {B,T}, {D,M}, {M,T}}**

2.c

C3: {{B,M,T} :3; {B,D,M}:3}

Since count {B,M,T} is equal to 3,

L3:**{{B,M,T},{B,D,M}}**

2.d

No, a 4-itemset should be {B,D,M,T}, missing {B,D,T} and {D,M,T}.

Question3

3.a

C1 : { {A}: 4; {B}: 5; {C}: 3; {D}: 4; {T}: 2; {M}: 2; {Y}: 1; }

Since the min_sup is 3, L1 : **{B,A,D,C}**

TID Frequent items

1 BADC

2 BD

3 BA

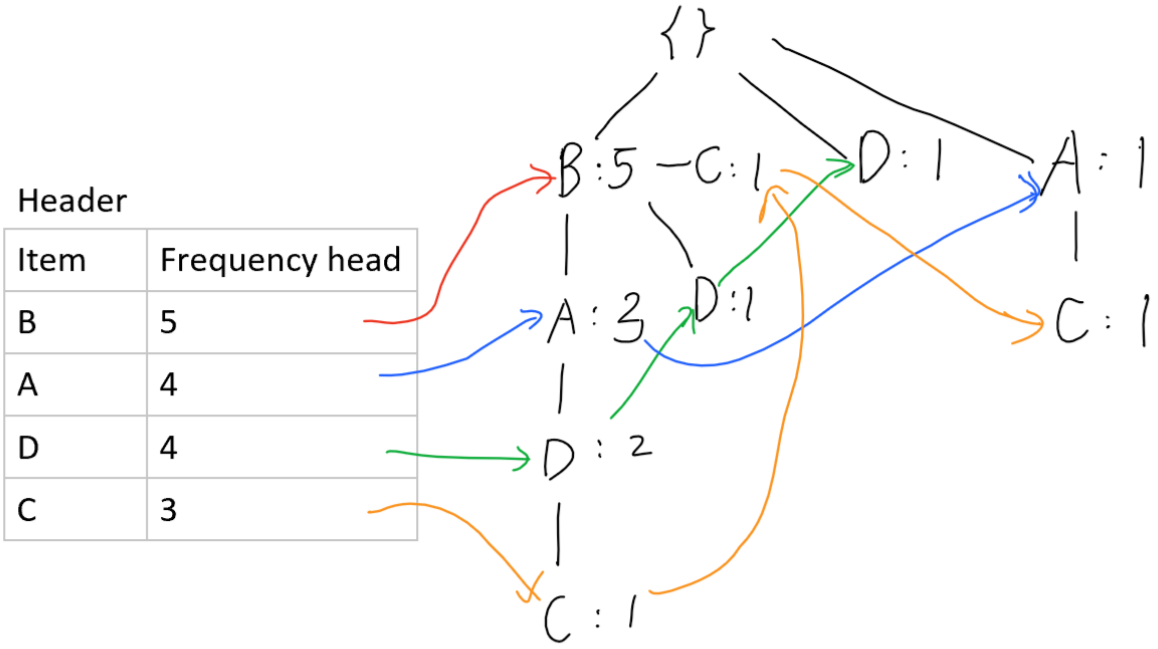
4 D

5 AC

6 BAD

7 BC

3.b



Header

Item	Frequency head
B	5
A	4
D	4
C	3

3.c

item	Conditional pattern base:	Frequent patterns
A	B:3	A,BA
C	BAD:1,B:1,A:1	C

3.d

If we list the nodes in the order, it will be more compression since the most frequent nodes are likely to be shared.

3.e

Yes, FP is faster than Apriori.

1st, Apriori scan the database **level by level**, while the FP go depth to find the **path** to combine all different posibility. Therefore, if FP would ignore those not necessary nodes.

2nd, Apriori scan the **entire** database **multiple** times to count each nodes, FP would scan merely the **projected** part of database and count **once**.

3.f

NO, FP-tree will omit lease frequent items that less than min_support.

Question4

4.a

eab:10; b:20

4.b

Since the total transaction is 150, and the relative min support is 0.5

the min support is 150*0.5 = 75

frequeny items are those more than 75

2 itemset is {ea:80};

No 3 itemset

Type *Markdown* and LaTeX: α^2