PROBLEM STATEMENT:

5.3 A database has five transactions. Let $min\_sup = 60\%$ and $min\_conf = 80\%$.

TID	*items_bought*
T100	{M, O, N, K, E, Y}
T200	{D, O, N, K, E, Y }
T300	{M, A, K, E}
T400	{M, U, C, K, Y}
T500	{C, O, O, K, I ,E}

(a) Find all frequent itemsets using Apriori and FP-growth, respectively. Compare the efficiency of the two mining processes.

(b) List all of the *strong* association rules (with support s and confidence c) matching the following metarule, where X is a variable representing customers, and $item_i$ denotes variables representing items (e.g., "A", "B", etc.):

$$\forall x \in transaction,\ buys(X, item_1) \wedge buys(X, item_2) \Rightarrow buys(X, item_3) \quad [s, c]$$

SOLUTION:

L1 :

(M)→3

(O)-→4

(N)→2

(K)→5

(E)→4

(Y)→3

(D)→1

(A)→1

(U)→1

(C)→2

(I)→1

L2 :

(M,O)→1

(M,K)→3

(M,E)→2

(M,Y)→2

(O,K)→3

(O,E)→3

(O,Y)→2

(K,E)→4

(K,Y)→3

(E,Y)→2

Rule1	Confidence	Rule2	Confidence
M→ K	3/3 = 100%	K→M	3/5 = 60%
O→K	3/4 = 75 %	K→O	3/5 = 60%
O→E	3/4-= 75%	E→O	3 /4 = 75%
K→E	4/5 = 80%	E→K	4/4 = 100%
K→Y	3/5 = 60%	Y→ K	3/3 = 100%

L3: (M,K,O,E,Y)

(M,K)→O = 1

(M,K)→E = 2

(M,K)→Y = 2

(K,O)→E = 3

(K,O) →Y = 2

(O,E)→Y = 2

(M,O)→E = 1

(M,O)→ Y = 1

(M,E)→Y = 1

(M,O)→ K = 1

(K,E)→ O = 3

(O,E)→K = 3

Rule1	Confidence	Rule2	Confidence
KO→ E	3/3 = 100%	E→KO	3/4 = 75%
OE→K	3/3 = 100 %	K→OE	3/5 = 60%
KE→O	3/4 = 75%	O→KE	3 /3 = 100%

Confidence:

KO→ E	3/3 = 100%
OE→K	3/3 = 100 %
O→KE	3 /3 = 100%

A database has five transactions. Let min_sup = 60% & min_conf = 80%.

TID	items_bought
T100	{M, O, N, K, E, Y}
T200	{D, O, N, K, E, Y}
T300	{M, A, K, E}
T400	{M, U, C, K, Y}
T500	{C, O, O, K, I, E}

Step 1:

Item		
M	III	3
O	IIII	4
N	II	2
K	~~IIII~~	5
E	IIII	4
Y	III	3
D	I	1
A	I	1
U	I	1
C	II	2
I	I	1

Step 2: min_sup = 60%

60% of 5 = 3

M	3
O	4
K	5
E	4
Y	3

Step 3:

MO	1
MK	3
ME	2
MY	2
OK	3
OE	3
OY	2
KE	4
KY	3
EY	2

Step 4: Rejecting pairs which are low than 60%

MK	3
OK	3
OE	3
KE	4
KY	3

Step 5:

MKO	1
MKE	2
MKY	2
OKE	3
OKY	2

Step 6:

Rejecting pairs which are low th 68;

OKE 3

Ed. self confidence

OK 3/3 = 100% ✓
OE 3/3 = 100% ✓
KE 3/4 = 75%
O 3/3 = 100% ✓
K 3/5 = 60%
E 3/4 = 75%

~~OK, OE~~

OK → E
OE → K
O → KE