ABHINANDAN
IIT2020119

Data Mining
Tutorial Sheet 3

Q 2.) Min. Support = 60%
Min. Confidence = 80%

C1

Items	Support	
A	1	(20%)
C	2	(40%)
D	1	(20%)
E	4	(80%)
I	1	(20%)
K	5	(100%)
M	3	(60%)
N	2	(40%)
O	3	(60%)
U	1	(20%)
Y	3	(60%)

L1 → E, K, M, O, Y

C2

Items	Count (Support)
EK	4 (80%)
EM	2 (40%)
EO	3 (60%)
EY	2 (40%)
KM	3 (60%)
KO	3 (60%)
KY	3 (60%)
MO	1 (20%)
MY	2 (40%)
OY	2 (40%)

L2 → EK, EO, KM, KO, KY

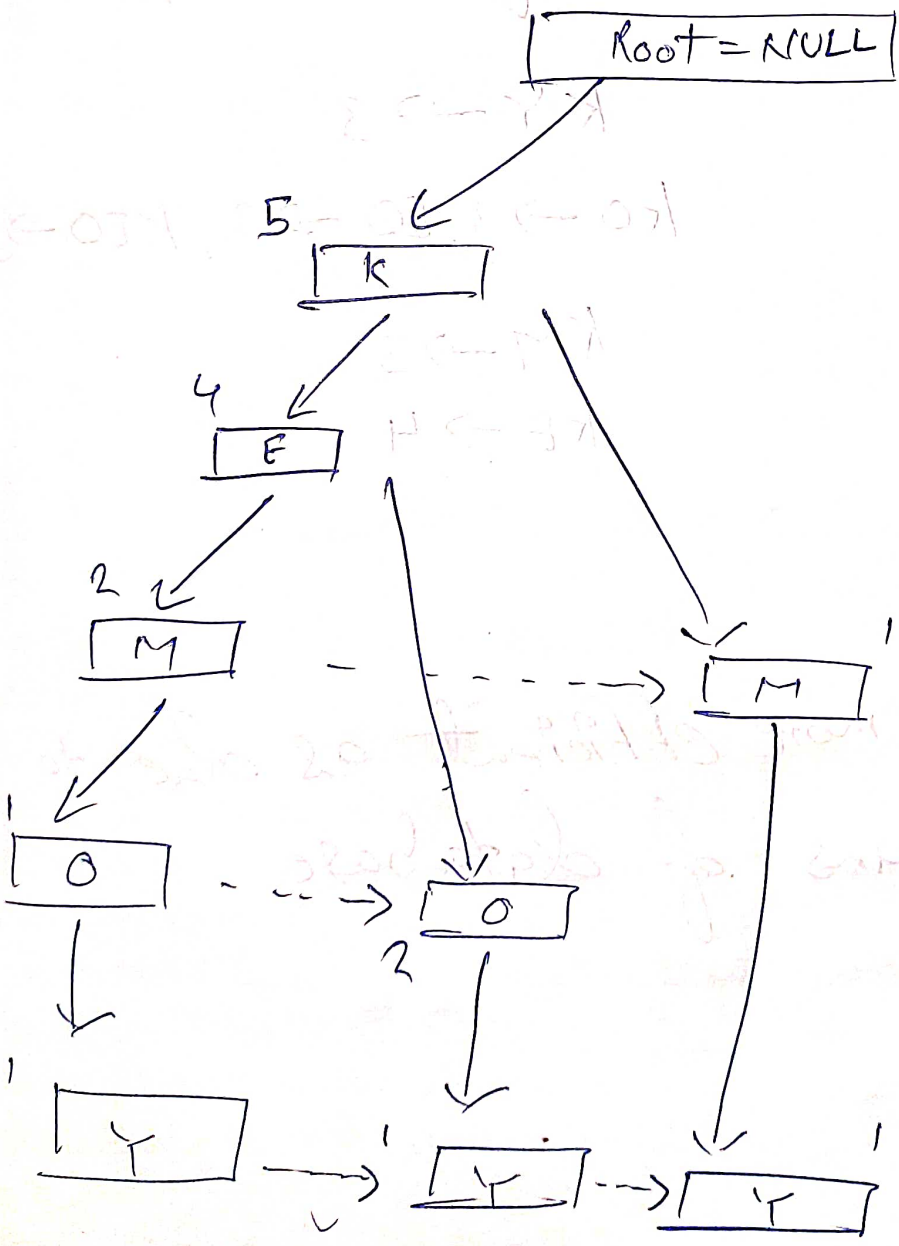
C3

Items	Support
E, K, O	3 (60%)
K, M, O	1 (20%)
K, M, Y	2 (40%)

L3 → EKO

As min. Support = 60%, so, f/p tree will be build using those elements itself.

TID	Ordered Items
T100	KEMOY
T200	~~KEMO~~ KEOY
T300	KEM
T400	KMY
T500	KEO



Items	Conditional Pattern Base	Conditional Freq. Pattern Tree
Y	KEMO:1, KEO:1, KM:1	K:3
O	KEM:1, KE:2	KE:3
M	KE:2, K:1	K:3
E	K:4	K:4
K		

Items	Freq. Pattern
Y	KY → 3
O	KO → 3, EO → 3, KEO → 3
M	KM → 3
E	KE → 4
K	

FP Tree is more efficient as due to only 2 scans of database

b)

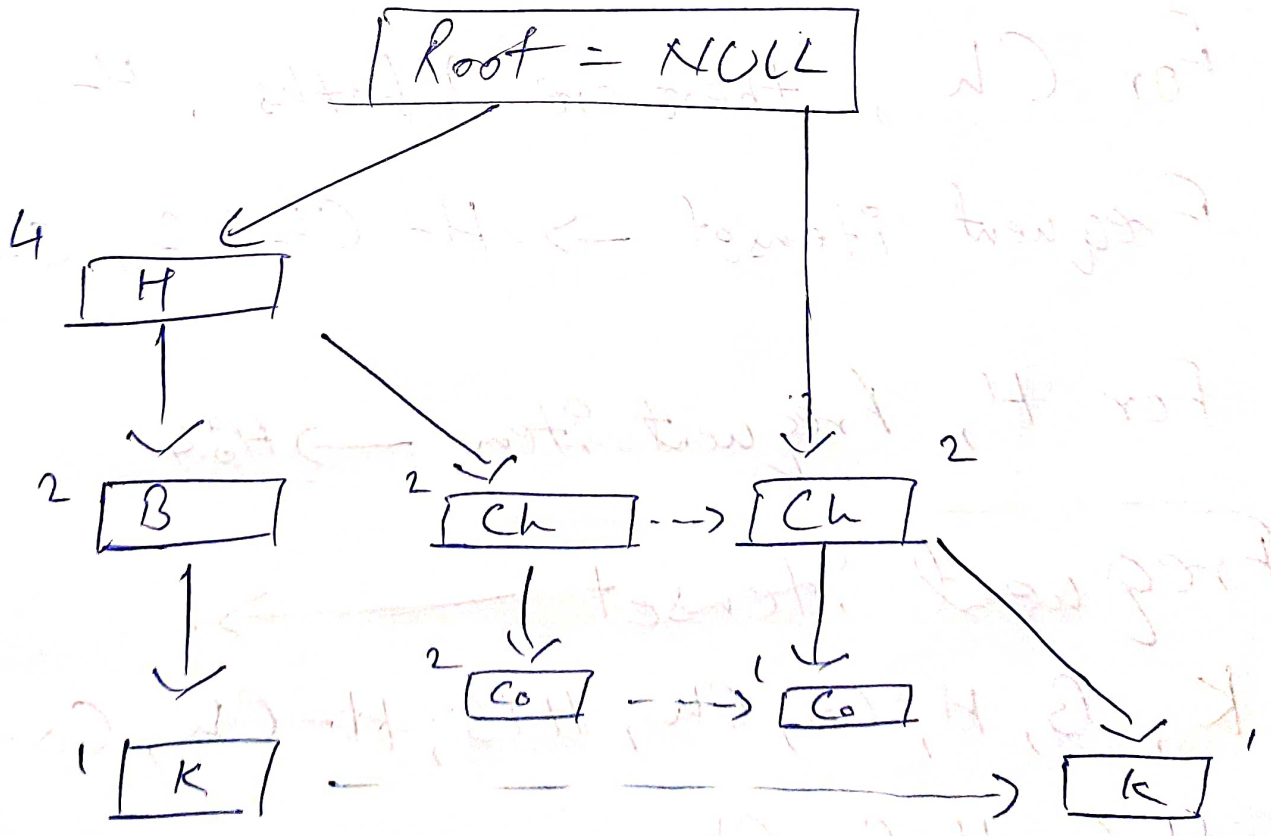
Association Rules	Confidence	
$[E, K] \rightarrow O$	75%	
$[K, O] \rightarrow E$	100%	Strong
$[E, O] \rightarrow K$	100%	Strong
$E \rightarrow [K, O]$	75%	
$K \rightarrow [E, O]$	60%	
$O \rightarrow [E, K]$	100%	Strong

Q3) Min. Support = 33.33%

Min. Confidence = 60%

Items	Support
Hotdogs (H)	4 (66%)
Chips (Ch)	4 (66%)
Coke (Co)	3 (50%)
Buns (B)	2 (33%)
Ketchup (K)	2 (33%)

FP-Tree



For K, there are 2 branches HB K, Ch K but it doesn't satisfy minimum support.

Frequent item → K : 2

For B, there are 2 branches, HB. So,

Frequent item → HB : 2, B : 2

For Co, there are 2 branches H - Ch - Co & Ch - Co.

Frequent item → Co : 3, Co - Ch - H : 2, Co - Ch : 3

For Ch, there are 2 paths, H - Ch & Ch

Frequent itemset → H - Ch : 2, Ch : 4

For H, Frequent item → H : 4.

Frequent itemset →

K, B, H, Co, Ch, HB, H - Ch, Co - Ch, H - Co, H - Co - Ch

Q4)

TID	Transaction
T1	A B C D E F
T2	B C D E F G
T3	A D E H
T4	A D F I J
T5	B D E K

Min. Support → 60%. (0.6 * 5 = 3)

Min. Confidence → 80%

Items	Support
A	3
B	3
C	2
D	5
E	4
F	3
G	1
H	1
I	1
J	1
K	1

L1 → A, B, D, E, F

C2

Items	Support
AB	1
AD	3
AE	2
AF	2
BD	3
BE	3
BF	2
DE	4
DF	3
EF	2

L2 → AD, BD, BE, DE, DF

C3

Items	Support
BDE	3
DEF	2

L3 → BDE

Frequent itemset ⟶ BDE

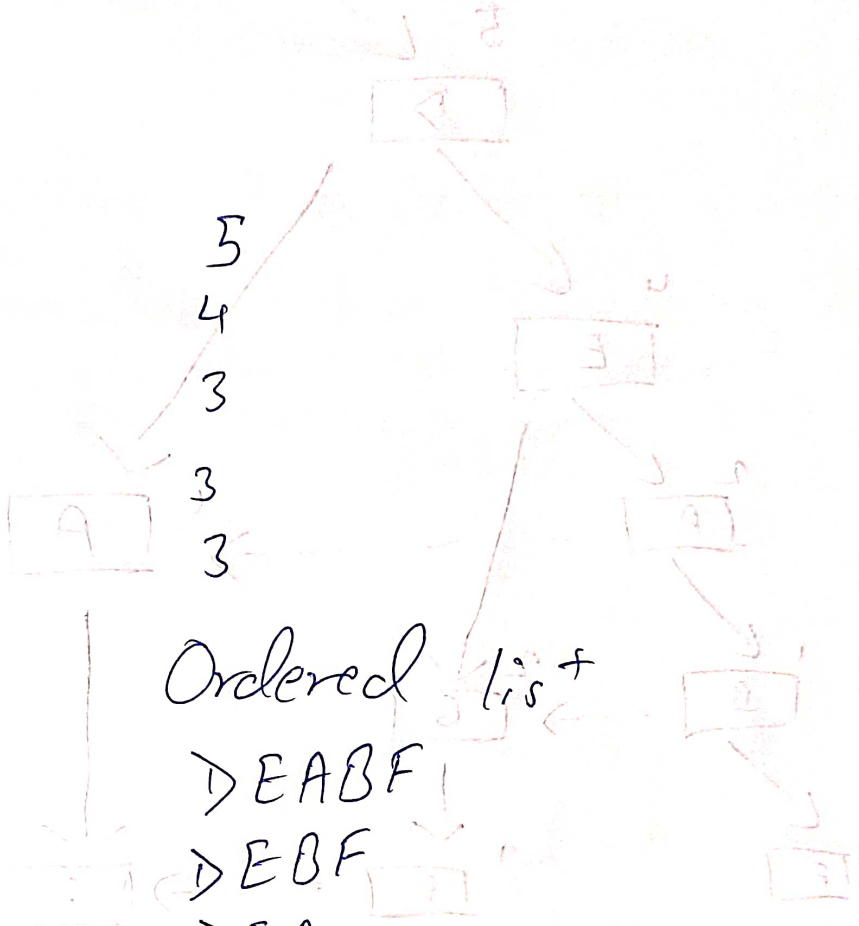
b) FP-tree

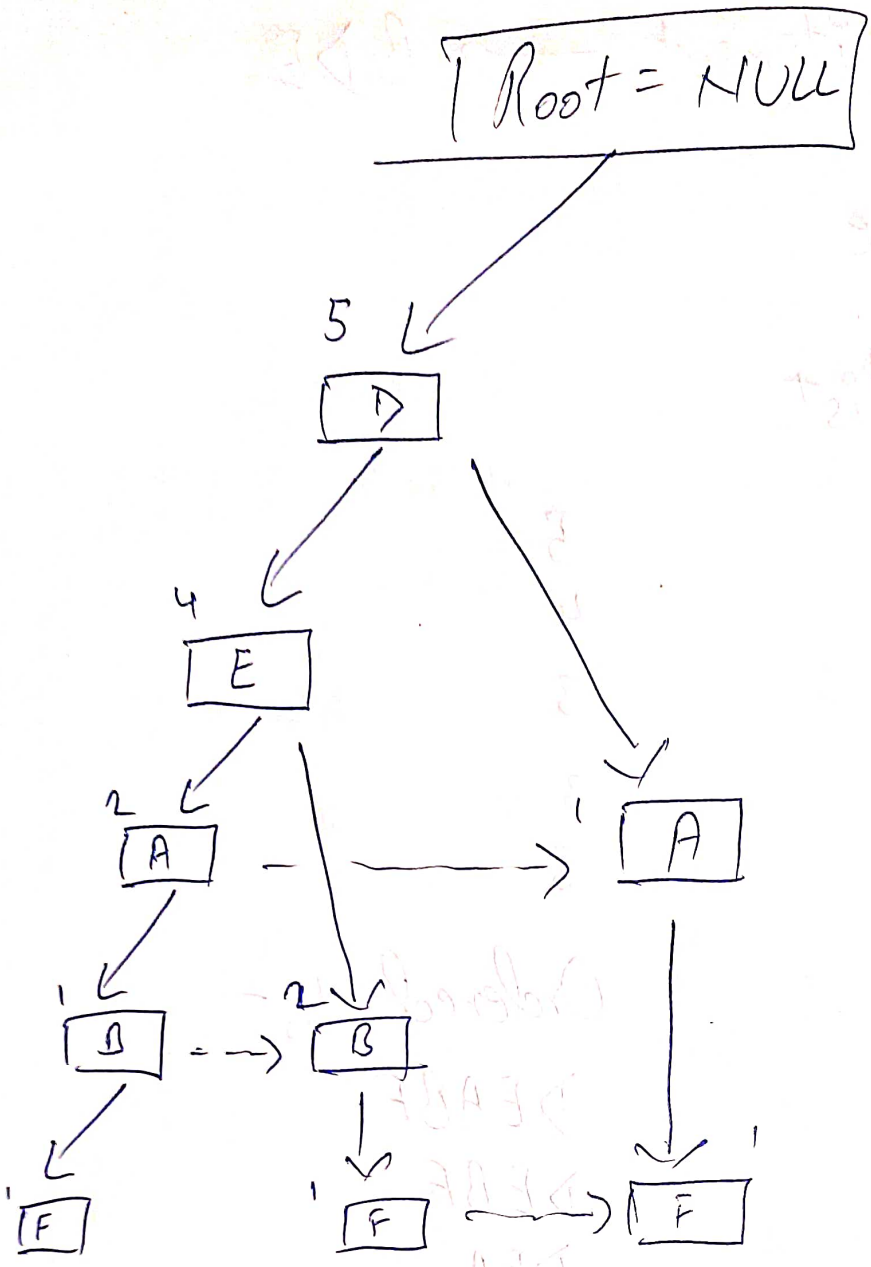
F-list

D	5
E	4
A	3
B	3
F	3

TID	Ordered list
1	DEABF
2	DEBF
3	DEA
4	DAF
5	DEB

Sequence ⟶ FBAED





Item	Frequent Pattern
A	DE, D
B	~~DEA~~, DEB
F	DF
E	DE

c) Apriori algorithm has high time complexity, while FP tree is fast as it uses only 2 scans.

d) Strong association Rules with A as antecedent

	Confidence
$A \rightarrow D$	100%
$AE \rightarrow D$	

e) The constraint cannot be used to generate because it is used to narrow down frequent itemsets.