Tutorial exercises:
Association Rule Mining.

Exercise 1. ***Apriori***

Trace the results of using the Apriori algorithm on the grocery store example with support threshold *s*=33.34% and confidence threshold *c*=60%. Show the candidate and frequent itemsets for each database scan. Enumerate all the final frequent itemsets. Also indicate the association rules that are generated and highlight the strong ones, sort them by confidence.

Transaction ID	Items
T1	HotDogs, Buns, Ketchup
T2	HotDogs, Buns
T3	HotDogs, Coke, Chips
T4	Chips, Coke
T5	Chips, Ketchup
T6	HotDogs, Coke, Chips

Solution:

Support threshold = 33.34% → minSup = 0.3334 = 2/6
The following letters will be used: H for Hotdog, B for Buns, K for Ketchup, Co fo Coke and Ch for Chips

Pass (k)	Candidate k-itemset and their support	Frequent k-itemsets
K=1	H (4/6), B (2/6), K (2/6), Co(3/6), Ch(4/6)	H(4/6), B(2/6), K(2/6), Co(3/6), Ch(4/6)
K=2	{H,B}(2/6), {H,K}(1/6), {H,Co}(2/6), {H,Ch}(2/6) {B,K}(1/6), {B, Co}(0), {B,Ch}(0) {K,Co}(0), {K,Ch}(1/6) {Co, Ch}(3/6)	{H,B}, {H,Co}, {H,Ch} {Co,Ch}
K=3	{H,Co,Ch}(2/6)	{H,Co,Ch}

Note that {H, B, Co} and {H, B, Ch} are not candidates when k=3 because their subsets {B, Co} and {B, Ch} are not frequent.

All Frequent Itemsets: {H}, {B}, {K}, {Co}, {Ch}, {H, B}, {H, Co}, {H, Ch}, {Co, Ch}, {H, Co, Ch}.

Association rules with confidence = 60% = 0.6
{H,B} generates:

- ~~H → B conf = sup({H,B})/sup({H}) = 2/4 = 0.5~~
- B → H conf = 2/2 = 1

{H, Co} generates:

- ~~H → Co conf = sup({H,Co}) / sup({h}) = 2/4 = 0.5~~
- Co → H conf = 2/3 = 0.66

{H,Ch} generates:

- ~~H → Ch conf = 2/4 = 0.5~~
- ~~Ch → H conf = 2/4 = 0.5~~

{Co,Ch} generates:

- Co → Ch conf = 3/3 = 1
- Ch → Co conf = ¾ = 0.75

{H, Co, Ch} generates:

- H, Co → Ch conf = 2/2 = 1
- H, Ch → Co conf = 2/2 = 1
- Co, Ch → H conf = 2/3 = 0.66
- ~~H → Co, Ch conf = 2/4 =0.5~~
- Co → H, Ch conf = 2/3=0.66
- ~~Ch → H, Co conf = 2/4=0.5~~

Exercise 2. ***FP-tree and FP-Growth***

a) Use the transactional database from the previous exercise with same support threshold and build a frequent pattern tree (FP-Tree). Show for each transaction how the tree evolves.
b) Use Fp-Growth to discover the frequent itemsets from this FP-tree.

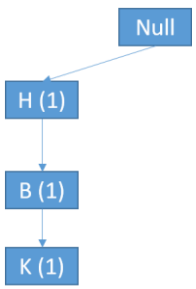
Solution

The first scan of the database generates the list of frequent 1-itemsets and builds the header table where the items are sorted by frequency.

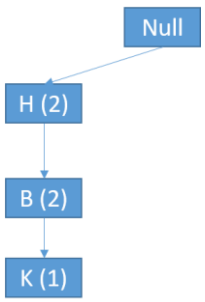
Item	support
H	4/6
Ch	4/6
Co	3/6
B	2/6
K	2/6

Transaction ID	Items	Ordered itemset
T1	HotDogs, Buns, Ketchup	{H,B,K}
T2	HotDogs, Buns	{H,B}
T3	HotDogs, Coke, Chips	{H,Ch,Co}
T4	Chips, Coke	{Ch,Co}
T5	Chips, Ketchup	{Ch, K}
T6	HotDogs, Coke, Chips	{H,Ch, Co}

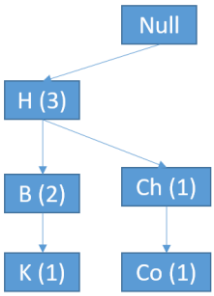
T1



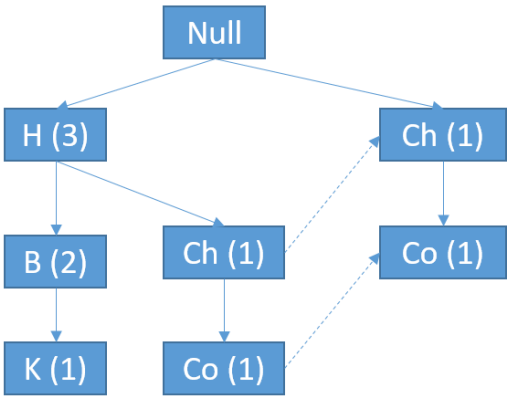
T2



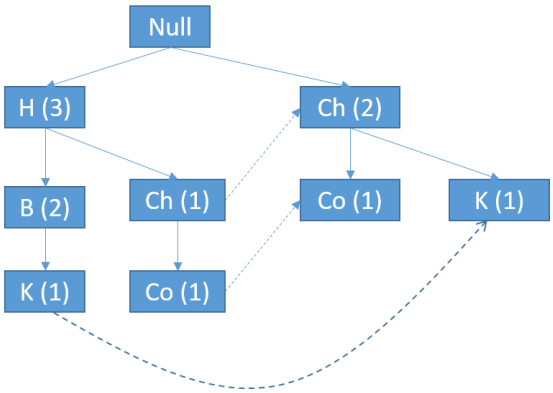
T3



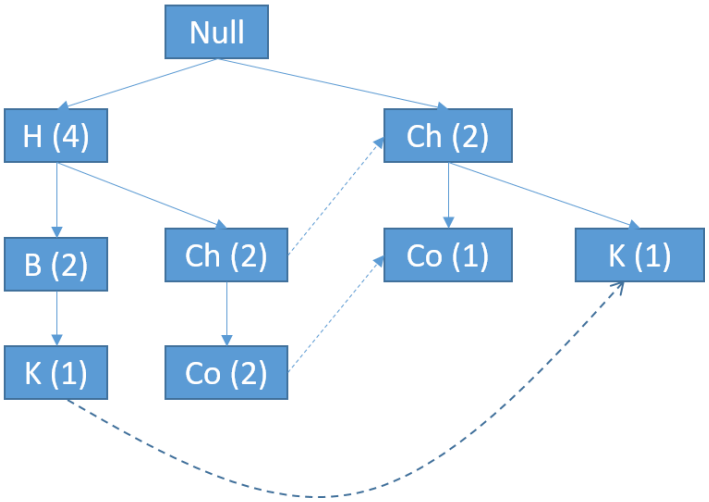
T4



T5



T6



Item	Conditional pattern base	Conditional frequent pattern tree	Frequent patterns generated
H (4)	{}	{}	{H:4}
Ch (4)	{H:2}	{H:2}	{H,Ch: 2}, {Ch:4}
Co (3)	{H,Ch:2}, {Ch:1}	{H: 2, Ch:3 }	{H,Co:2}, {Ch,Co:2}, {H,Ch,Co:2}, {Co:3}
B (2)	{H: 2}	{H: 2}	{H, B:2}, {B:2}
K (2)	{H,B:1} {Ch: 1}	~~{H:1, B:1, Ch:1 }~~ infrequent	{K:2}

Frequent patterns generated are the same as Exercise 1. Follow the same steps to generate the rules.

Exercise 4: **Apriori and FP-Growth (Implement the following in python)**

Giving the following database with 5 transactions and a minimum support threshold of 60% and a minimum confidence threshold of 80%, find all frequent itemsets using (a) Apriori and (b) FP-Growth. (c) Compare the efficiency of both processes. (d) List all strong association rules that contain "A" in the antecedent (Constraint). (e) Can we use this constraint in the frequent itemset generation phase?

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}

Solution: For python code, check the ipynb file uploaded on MS teams.

Apriori:

Support threshold = 60% → minSup = 0.6 = 3/5 (or minSup = 3, since there are 5 transactions)

Pass (k)	Candidate k-itemset and their support	Frequent k-itemsets
K=1	{A:3} {B:3}, {C:2}, {D:5}, {E:4}, {F:3}, {G:1}, {H:1}, {I:1}, {J:1}, {K:1}	{A:3}, {B:3}, {D:5}, {E:4}, {F:3}
K=2	{AB:1}, {AD:3}, {AE:2}, {AF:2}, {BD:3}, {BE:3}, {BF:2}, {DE:4}, {DF:3}, {EF:2}	{AD:3}, {BD:3}, {BE:3}, {DE:4}, {DF:3}
K=3	{{BDE:3}, {BDF:2}, {DEF:2}	{BDE:3}

Note: {ABD} {ADE} and {ADF} are note generated in K=2 because their subsets {AB}, {AE} and {AF} are infrequent (from k=2)

Frequent patterns are: {A:3}, {B:3}, {D:5}, {E:4}, {F:3}, {AD:3}, {BD:3}, {BE:3}, {DE:4}, {DF:3}, {BDE:3}

Association rules with confidence =80% = 0.8

{AD} generates:
- A → D conf = sup(AD)/sup(A) = 3/3 = 1
- ~~D → A conf = 3/5 = 0.6~~

{BD} generates:
- B → D conf = 3/3 = 1
- ~~D → B conf = 3/5 = 0.6~~

{BE} generates:
- B → E conf = 3/3 = 1
- ~~E → B conf = ¾ = 0.75~~

{DE} generates:
- D → E conf = 4/5 = 0.8
- E → D conf = 4/4 = 1

{DF} generates:
- ~~D → F conf = 3/5~~
- F → D conf = 3/3 = 1

{BDE} generates:
- B,D → E conf = 3/3=1
- B, E → D conf=3/3=1
- ~~D,E → B conf = ¾=0.75~~
- B → D, E conf = 3/3=1
- ~~D → B, E conf = 3/5 = 0.6~~
- ~~E → B, D conf = ¾ = 0.75~~

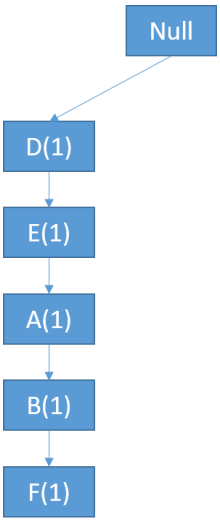
FP-Growth

The first scan of the database generates the list of frequent 1-itemsets and builds the header table where the items are sorted by frequency.

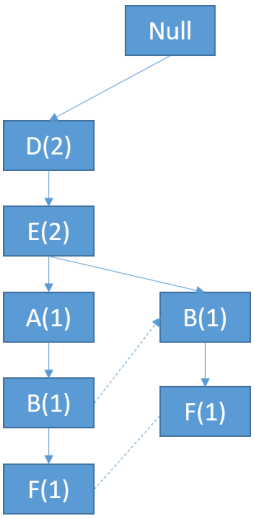
Item	support
D	5
E	4
A	3
B	3
F	3

TID	**Transaction**	**Ordered frequent item list**
T1	{A, B, C, D, E, F}	{D,E,A,B,F}
T2	{B, C, D, E, F, G}	{D,E,B,F}
T3	{A, D, E, H}	{D,E,A}
T4	{A, D, F, I, J}	{D,A,F}
T5	{B, D, E, K}	{D,E,B}

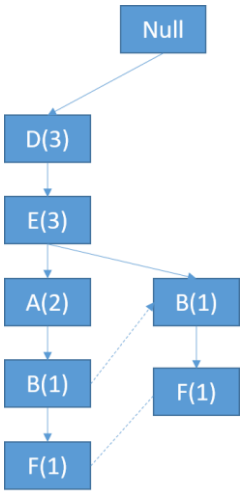
After T1



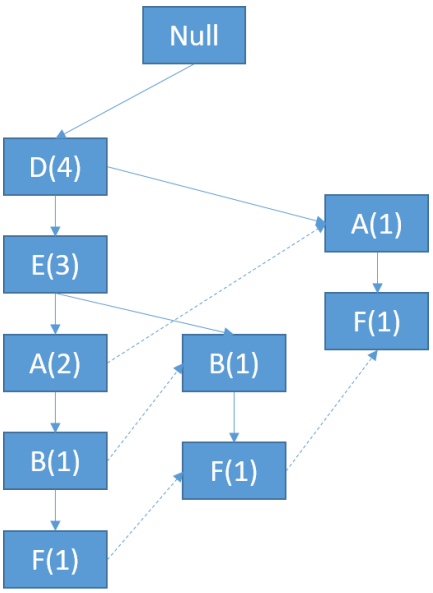
after T2



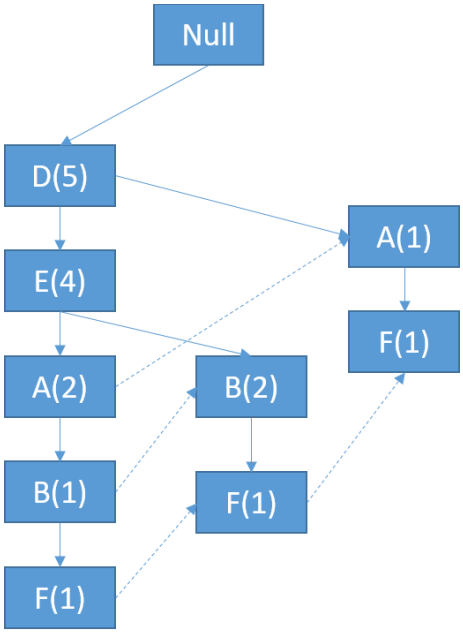
After T3



After T4



After T5



minSup = 3

Item	Conditional pattern base	Conditional frequent pattern tree	Frequent patterns generated
D (5)	{}	{}	{D:5}
E (4)	{D:4}	{D:4}	{D,E:4}, {E:4}
A (3)	{D,E:2}, {D:1}	{D:3~~,E:2~~} , {D:3}	{D,A:3}, {A:3}
B (3)	{D,E,A:1}, {D,E:2}	{D:3, E:3, ~~A:1~~} , {D:3, E:3}	{D,B:3}, {E,B:3}, {D,E,B:3}, {B:3}
F (3)	{D,E,A,B:1}, {D,E,B:1}, {D,A:1}	{D:3, ~~E:2, A:2, B:2~~}, {D:3~~, E:2, B:2~~}, {D:3, ~~A:2~~}	{D,F:3}, {F:3}

Frequent patterns generated: {D:5}, {E:4}, {A:3}, {B:3}, {F:3} {DE:4},{AD:3},{BD:3}, {BE:3}, {BED:3},{DF:3}

This is the same as Apriori. Follow the same steps to generate the rules