Assignment 2

Brady Mitchelmore - 202112249

Question 2.1:

`min_sup = 0.60`

`min_conf = 0.80`

TID	items_purchased
T1	{A, B, C, D, E, F}
T2	{G, B, C, D, E, F}
T3	{A, H, D, E}
T4	{A, I, J, D, F}
T5	{J, B, B, D, K, E}

Exercise 2.1.1:

Apriori:

First we scan the table for count of each candidate

Itemset	Support Count	Support
{A}	3	0.60
{B}	4	0.80
{C}	2	0.40
{D}	5	1.00
{E}	4	0.80
{F}	3	0.60
{G}	1	0.20
{H}	1	0.20
{I}	1	0.20
{J}	2	0.40

Next compare relative candidate support with the minimum support of 0.60. Here only 5 candidates in the first table satisfy the minimum support

Itemset	Support Count	Support
{A}	3	0.60
{B}	4	0.80
{D}	5	1.00
{E}	4	0.80
{F}	3	0.60

Itemset	Support Count	Support
{K}	1	0.20

Next we generate frequent 2-itemsets from the table above

Itemset	Support Count	Support
{A, B}	1	0.20
{A, D}	3	0.60
{A, E}	1	0.20
{A, F}	2	0.40
{B, D}	3	0.60
{B, E}	3	0.60
{B, F}	2	0.40
{D, E}	4	0.80
{D, F}	3	0.60
{E, F}	2	0.40

Next we compare relative candidate support with the minimum support of 0.60. Here only 5 candidates in the table above satisfy the minimum support

Itemset	Support Count	Support
{A, D}	3	0.60
{B, D}	3	0.60
{B, E}	3	0.60
{D, E}	4	0.80
{D, F}	3	0.60

Next we generate frequent 3-itemsets from the table above.

Itemset	Support Count	Support
{B, D, E}	3	0.60
{D, E, F}	2	0.40

Next we compare relative candidate support with the minimum support of 0.60. Here only 1 candidate in the table above satisfy the minimum support

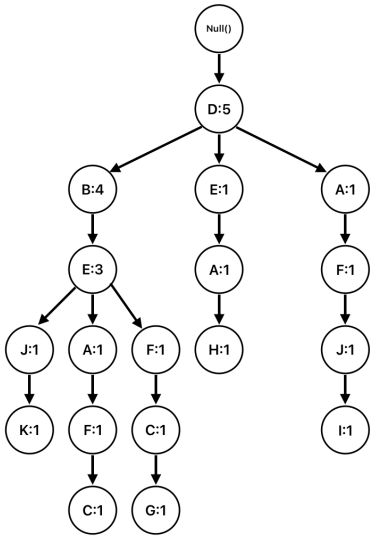
Itemset	Support Count	Support
{B, D, E}	3	0.60

No frequent 4-itemsets can be generated so we stop here with a frequent itemset of **{B, D, E}** which has a support count of 6

FPGrowth:

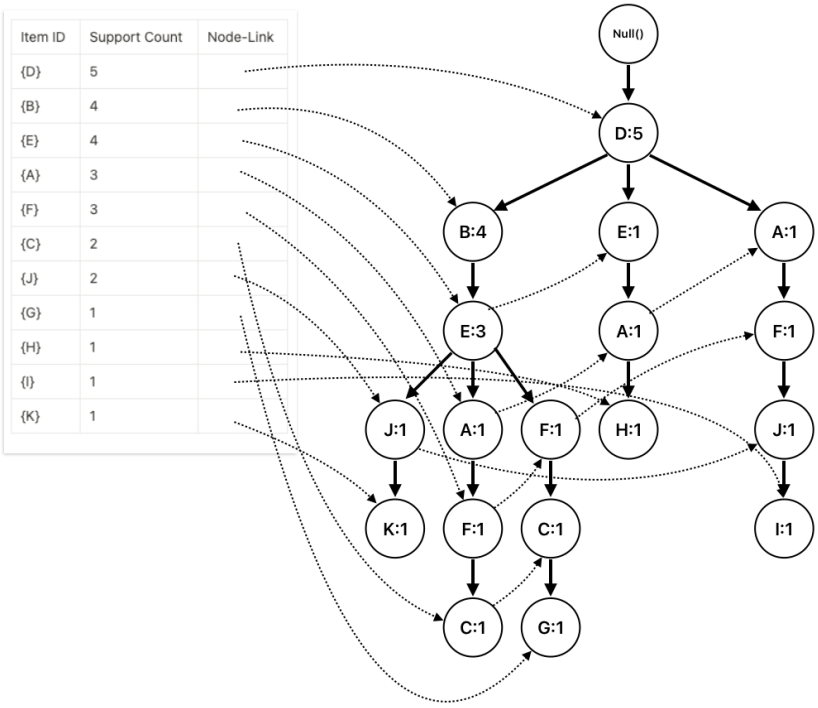
First we scan the table for count of each candidate sorted in descending order

Itemset	Support Count	Support
{D}	5	1.00
{B}	4	0.80
{E}	4	0.80
{A}	3	0.60
{F}	3	0.60



Itemset	Support Count	Support
{C}	2	0.40
{J}	2	0.40
{G}	1	0.20
{H}	1	0.20
{I}	1	0.20
{K}	1	0.20

Then we construct the node links between the itemsets and nodes



Mining the FP-tree by creating conditional pattern bases

Item	Conditional Pattern Base	Conditional FP-tree	Frequent Patterns Generated
{K}	{{D, B, E, J: 1}}	{}	{}
{I}	{{D, A, F, J: 1}}	{}	{}
{H}	{{D, E, A: 1}}	{}	{}
{G}	{{D, B, E, F, C: 1}}	{}	{}
{J}	{{D, B, E: 1}, {D, A, F: 1}}	{}	{}
{C}	{{D, B, E, A, F: 1}, {D, B, E, F: 1}}	{}	{}
{F}	{{D, B, E, A: 1}, {D, B, E: 1}, {D, A: 1}}	{D: 3}	{D, F: 3}
{A}	{{D, B, E: 1}, {D, E: 1}, {D: 1}}	{D: 3}	{D, A: 3}
{E}	{{D, B: 3}, {D: 1}}	{D: 4}	{D, E: 4}
{B}	{{D: 4}}	{D: 4}	{D, B: 4}
{D}	{{}}	{}	{}

Exercise 2.1.2:

From Exercise 2.1.1 we have the following frequent 3-itemset.

Itemset	Support Count	Support
{B, D, E}	3	0.60

With {B, D} = 3, and {B, D, E} = 3, we can calculate the confidence.

$$confidence = \{ B, D, E \} / \{ B, D \} = 3/3 = 1.00$$

Itemset	Support Count	Support	Confidence
{B, D, E}	3	0.60	1.00

This is the only strong association rule with support 0.60 and confidence 0.80 which matches the metarule {B, D} $\rightarrow$ {E}.