



Name of the Subject: DATA ANALYSIS & INFO. EXT. Subject Code: IT-704

Seat No: IT076 Student ID: 18JITURN116 Branch/Sem: IT-VII

Q3

(9).

TID	Item bought
T100	{K, A, D, B}
T200	{D, A, C, E, B}
T300	{C, A, B, E}
T400	{B, A, D}

Min. Sup = 2 & Min. Conf = 80%

C ₁	Item	Count
	A	4
	B	4
	C	2
	D	3
	E	2
	K	1

As, Min. Sup = 2 so, remove K & arrange in descending order.

L ₁	Item	Count
	A	4
	B	4
	D	3
	C	2
	E	2

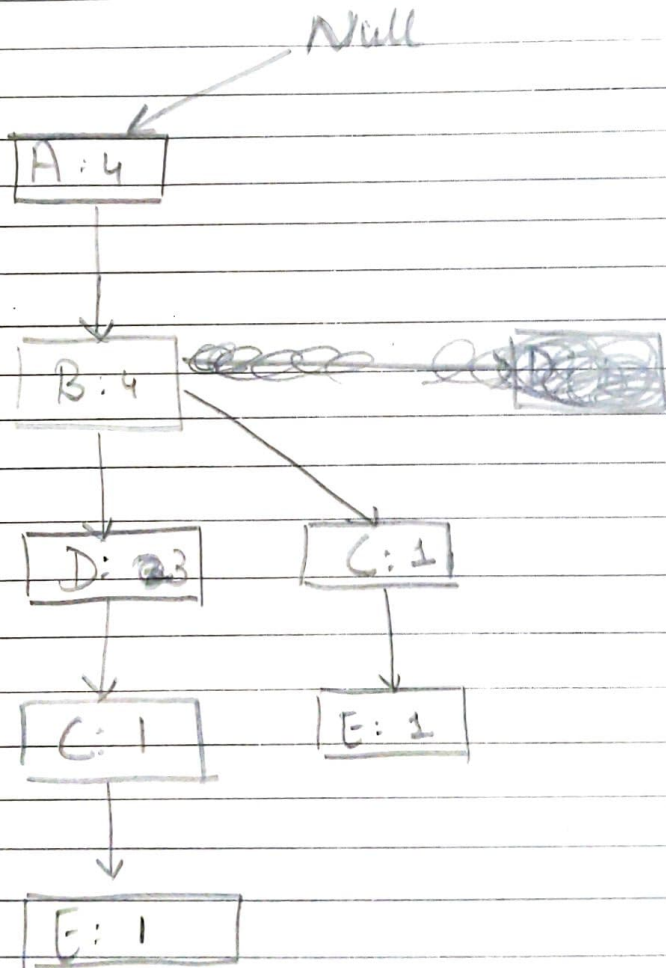
PTO
→



Name of the Subject: DATA ANALYSIS & INFO. EXT. Subject Code: IT 704
Seat No: IT076 Student ID: 18ITUBN116 Branch/Sem: IT-VII

Transaction ID	Item.	Order Itm - sel.
T ₁₀₀	K, A, D, B	A, B, D
T ₂₀₀	D, A, C, E, B	A, B, D, C, E
T ₃₀₀	C, A, B, E	A, B, C, E
T ₄₀₀	B, A, D	A, B, D.

FP Tree.





Name of the Subject: DAIE Subject Code: IT-704

Seat No: IT076 Student ID: 18ITURN116 Branch/Sem: IT-VII

Item.

Condition Pattern Base.

E

$\{A, B, D, C: 1\}, \{A, B, C, 1\}$

C

$\{A, B, 1\}, \{A, B, D, 1\}$

D

$\{A, B, 3\}, \{A, B, 3, 3\}$

B

$\{A, 4\}$

A

Item.

Condition Freq. Pattern Tree.

E

$\{A, B, C, 2\}$

C

$\{A, B, 2\}$

D

$\{A, B, 3\}$

B

$\{A, 4\}$

A

As, Min-Sup = 2 so, removing value below 2.

Item

Frequent Pattern generated.

E

$\{ \langle A, B, C, E: 2 \rangle, \langle A, B: 2 \rangle, \langle B, E: 2 \rangle, \langle C, E: 2 \rangle \}$

C

$\{ \langle A, C: 2 \rangle, \langle B, C: 2 \rangle, \langle A, B, C: 2 \rangle \}$

D

$\{ \langle A, D: 3 \rangle, \langle B, D: 3 \rangle, \langle A, B, D: 3 \rangle \}$

B

$\{ \langle A, B: 4 \rangle \}$

A



Name of the Subject: DAIE Subject Code: IT-704

Seat No: IT076 Student ID: 18ITUBN116 Branch/Sem: IT-VII

103

(b)

$$\text{MIN} = -\$351,976.00$$

$$\text{MAX} = \$4700,896.50$$

$$\text{LOW} = -\$159,876$$

$$\text{HIGH} = \$1838,761$$

— The most significant digit (msd).

$$\text{LOW}' = -\$1,000,000$$

$$\text{HIGH}' = +\$2,000,000$$

— The Interval Range over.

$$\text{Range} = \text{High} - \text{Low}$$

$$= (2,000,000 - (-1,000,000)) / 1,000,000$$

$$= 3$$

So, the Interval ar.

$$[-1,000,000, \$0], [\$0, \$1,000,000]$$

$$\& [\$1,000,000, \$2,000,000].$$

$$[-\$1,000,000, \$2,000,000]$$

$$[-\$1,000,000, \$0]$$

$$[\$0, \$1,000,000]$$

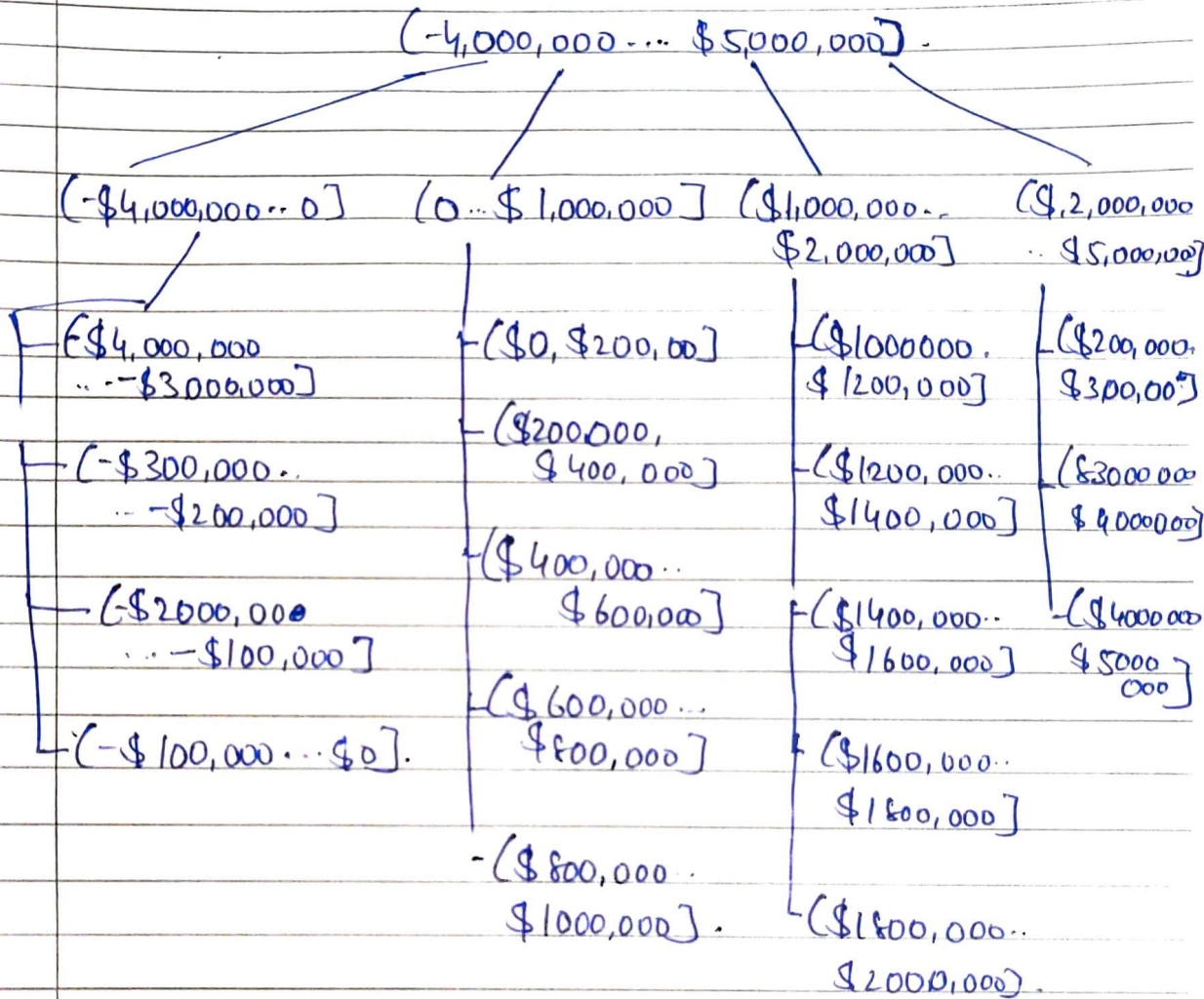
$$[\$1,000,000, \$2,000,000]$$



Name of the Subject: DAIE Subject Code: IT-704

Seat No: IT076 Student ID: 18ITURN116 Branch/Sem: II-VII

Rounding the MIN = $-\$4,000,000$
MAX = $\$5,000,000$





Name of the Subject: DAIE Subject Code: IT704
Seat No: IT076 Student ID: 18ITUBN116 Branch/Sem: IT-VII

102

(a) Limitation

- Major Computational Challenges.

- Multiple scans of transaction database.
- Huge number of candidates.
- Tedious workload of support counting for candidates.

- Improving Apriori

- Reduce passes of transaction database scans.
 - Shrink number of candidates.
 - Facilitate support counting of candidates.
- If there are 10^4 from one frequency, itself.
- So, it will check many set from candidate generation.



Name of the Subject: DAIE Subject Code: IT-703
Seat No: IT076 Student ID: 18ITUBN116 Branch/Sem: IT-VII

Q2

(b) Support

- The no. of transaction that include the $\{X\}$ & $\{Y\}$ part of the rule as a percentage of the total no. of transactn.

$$\text{Support} = \frac{\sigma(X+Y)}{\text{total.}}$$

Confidence (c)

- The no. of transaction that include all items in $\{B\}$ as well as no. of transaction include all items of $\{A\}$.

$$\text{Conf}(X \Rightarrow Y) = \frac{\text{Supp}(X \cup Y)}{\text{Supp}(X)}$$

- lift.