21) write the apriori algorithm. Solve the following by Applying apriori algorithm

Consider the transaction below

TID	Item
T_1	1,3,4
T_2	2,3,5
T_3	1,2,3,5
T_4	2,5
T_5	1,3,5

Generate all possible Association rule by generating candidate set Considering minimum Support as 2 and the Confidence % for the rule should be >= 60%.

Solution

Candidate Set C_1

Item	Support count
1	3
2	3
3	4
4	1
5	4

L_1

F_1	Frequent Itemset
1	3
2	3
3	4
5	4

Candidate set C_2

Item	Support count
1, 2	1
1, 3	3
1, 5	2
2, 3	2
2, 5	3
3, 5	3

L_2

F_1	Frequent Itemset
1, 3	3
1, 5	2
2, 3	2
2, 5	3
3, 5	3

Candidate set C_3

Item	Support count
1, 3, 5	2
2, 3 5	2

L_3

F_1	Frequent Itemset
1, 3, 5	2
2, 3, 5	2

Candidate set C_4 can't be generated, thus the item set which one identified as two final frequent item sets are

1, 3, 5 and 2, 3, 5

Rule for {1, 3, 5}

{1, 3, 5} → {1, 3}, {1, 5} {3, 5} {1} {3} {5}

① {1, 3} ⇒ {1, 3, 5} - {1, 3}

1 ∩ 3 → 5

Support = $\frac{2/5}{3/5}$ = $\frac{2}{3}$ = 66%

② {1,5}

1∩5 → 3

Support = 2/2 = 100%

③ {3,5}

3∩5 → 1

Support = 2/3 = 66%

④ {1}

1 → 3∩5

Support = 2/3 = 66%

⑤ {3}

3 → 1∩5

Support = 2/4 = 50%

⑥ {5}

5 → 1∩3

Support = 2/4 = 50%

Rule for {2,3,5} → {2,3}, {2,5} {3,5} {2} {3} {5}

① 2∩3 → 5

Support = 2/5 / 2/4 = 100%

② {2,5}

2∩5 → 3

Support = 2/3 = 66%

③ {3,5}

3∩5 → 2

Support = 2/3 = 66%

④ {2}

2 → 3∩5

Support = 2/3 = 66%

⑤ {3}

3 → 2∩5

Support = 2/4 = 50%

⑥ {5}

5 → 2∩3

Support = 2/4 = 50%

Threshold Confidence :- {1,3} {1,5} {3,5} {1}

{2,3} {2,5} {3,5} {2}

Q2) Consider ~~many~~ transaction with minimum 15 transactions. Apply hash based technique and partitioning technique to identify frequent itemset.

TID	Items
T_1	Lipstic, Mascara, Eyeliner
T_2	Lipstic, Mascara
T_3	Lipstic, Foundation
T_4	Mascara, Nail polish
T_5	Lipstic, Mascara, Foundation
T_6	Nailpolish, Eyeliner
T_7	Lipstic, Eyeliner
T_8	Mascara, Foundation
T_9	Lipstic, Mascara, Eyeliner
T_{10}	Nailpolish, Foundation
T_{11}	Lipstick, Mascara, Foundation
T_{12}	Mascara, Eyeliner
T_{13}	Mascara, Eyeliner, Nailpolish
T_{14}	Lipstic, Foundation, Nail polish
T_{15}	Lipstic, Eyeliner, Foundation

Pass 1 : Counting pairs and Hashing

$h = 7$

Pair (Lipstick, Mascara) : h(Lipstick, Mascara) = (1+2) % 7 = 3

Pair (Lipstick, Eyeliner) : h(Lipstick, Eyeliner) = (1+3) % 7 = 4

Pair (Lipstick, Foundation) : h(Lipstick, Foundation) = (1+4) % 7 = 5

Pair (Lipstick, Nail polish) : h(Lipstick, Nail polish) = (1+5) % 7 = 6

pair (Mascara, Eyeliner) : h(Mascara, Eyeliner) = (2+3) % 7 = 5

pair (Mascara, Foundation) : h(Mascara, Foundation) = (2+4) % 7 = 6

pair (Mascara, Nail polish) : h(Mascara, Nailpolish) = (2+5) % 7 = 0

pair (Eyeliner, Foundation): h(eyeliner, Foundation) = (3+4)%7 = 0

Pair (Eyeliner, Nail polish): h(eyeliner, Nail polish) = (3+5)%7 = 1

pair (Foundation, Nail polish): h(Foundation, Nail polish) = (4+5)%7 = 2.

Hash Bucket	count	pair
0	2	(Mascara, Nail polish), (Eyeliner, Foundation)
1	1	(Eyeliner, Nail polish)
2	1	(Foundation, Nail polish)
3	1	(Lipstick, Mascara)
4	1	(Lipstick, eyeliner)
5	2	(Lipstick, Foundation), (Mascara, Eyeliner)
6	2	(Lipstick, Nail polish), (Mascara, Foundation).

pass 2 Filtering and counting

Support threshold ≥ 2

Pair	Count
Mascara, Nail polish	2
Eyeliner, Foundation	2
Lipstick, Foundation	4
Mascara, Eyeliner	3
Lipstick, Nail polish	2
Mascara, Foundation	3.

Assuming a minimum support count of 3.

1. partition the database

Partition 1

TID	Items
T1	Lipstick, Mascara, Eyeliner
T2	Lipstick, Mascara
T3	Lipstick, Foundation
T4	Mascara, Nail polish
T5	Lipstick, Mascara, Foundation

Partition 2

TID	Items
T6	Nail polish, Eyeliner
T7	Lipstick, Eyeliner
T8	Mascara, Foundation
T9	Lipstick, Mascara, Eyeliner
T10	Nail polish, Foundation

Partition 3

TID	Items
T11	Lipstick, Mascara, Foundation
T12	Mascara, Eyeliner
T13	Mascara, Eyeliner, Nail polish
T14	Lipstick, Foundation, Nail polish
T15	Lipstick, Eyeliner, Foundation

2. Local Frequent Itemsets in each partitions

Partition 1 frequent itemset

Itemset	Count
Lipstick	4
Mascara	4
Eyeliner	1
Foundation	2
Nail polish	1
Lipstick, Mascara	3
Lipstick, Foundation	2

Partition 2

Itemset	Count
Lipstick	3
Mascara	2
Eyeliner	3
Foundation	2
Nail polish	2
Lipstick, eyeliner	2
Mascara, Eyeliner	2

Partition 3

Itemset	Count
Lipstick	3
Mascara	3
Eyeliner	3
Foundation	3
Nail polish	2
Lipstick, Mascara	2
Lipstick, Foundation	2
Mascara, Eyeliner	2

Merging the frequent Itemset

Itemset	Global Count
Lipstick	4+3+3 = 10
Mascara	4+2+3 = 9
Eyeliner	1+3+3 = 7
Foundation	2+2+3 = 7
Nail polish	1+2+2 = 5
Lipstic, Mascara	3+0+2 = 5
Lipstic, Foundation	2+0+2 = 4
Lipstic, Eyeliner	0+2+0 = 2
Mascara, Eyeliner	0+2+2 = 4

frequent Itemset are : Lipstick, Mascara, Eyeliner, Foundation, Nail polish, (Lipstic, Mascara), (Lipstick, Foundation), (Mascara, Eyeliner).