Solution to exercise 7:

7. Use the dataset in Exercise 6. Let min_sup=3 and the minimum confidence min_conf=70%. Use the Apriori Algorithm to discover all the strong association rules. Note that in this case, a strong association rule A->B tells us that if a user visits certain set of pages A, the user has a high likelihood to visit the set of pages B.

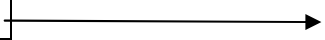
T1: {P1,P2,P3,P5,P7}
T2: {P1,P4,P5,P6,P7}
T3: {P1,P4,P6}
T4: {P1,P4,P5,P6,P7}
T5: {P3,P5}
T6: {P1,P2,P3,P7}
T7: {P2,P7}
T8: {P1,P2,P3,P4,P6,P7}

Transaction	Itemsets
T1	{1,2,3,5,7}
T2	{1,4,5,6,7}
T3	{1,4,6}
T4	{1,4,5,6,7}
T5	{3,5}
T6	{1,2,3,7}
T7	{2,7}
T8	{1,2,3,4,6,7}

Step1: generating 1 Item set frequent pattern

1-Itemset	Sup_count
{1}	6
{2}	4
{3}	4
{4}	4
{5}	4
{6}	4
{7}	6

Scan the data once to get the count of each item and remove the items that do not meet min_sup



1-Itemset	Sup_count
{1}	6
{2}	4
{3}	4
{4}	4
{5}	4
{6}	4
{7}	6

C1 **L1**

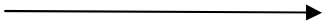
The set of frequent 1-itemsets, L1, consists of the candidate 1-itemsets satisfying minimum support.
In the first iteration of the algorithm, each item is a member of the set of candidate.

Generating 2-itemset Frequent Pattern

2-Itemsets	Sup_Count
{1,2}	3
{1,3}	3
{1,4}	4
{1,5}	3
{1,6}	4
{1,7}	5
{2,3}	3
{2,4}	1
{2,5}	1
{2,6}	1
{2,7}	4
{3,4}	1
{3,5}	2
{3,6}	1
{3,7}	3
{4,5}	2
{4,6}	4
{4,7}	3
{5,6}	2
{5,7}	3
{6,7}	3

C2 (L1 X L1)

Scan the data set again for the support count of C2 and remove the non frequent itemsets from c2 --> L2



2-Itemsets	Sup_count
{1,2}	3
{1,3}	3
{1,4}	4
{1,5}	3
{1,6}	4
{1,7}	5
{2,3}	3
{2,7}	4
{3,7}	3
{4,6}	4
{4,7}	3
{5,7}	3
{6,7}	3

L2

Step 3: Generating 3-item frequent set

From L2 to C3

3-itemsets	Sup_count
{1,2,3}	3
{1,2,7}	3
{1,3,7}	3
{1,4,6}	4
{1,4,7}	3
{1,5,7}	3
{1,6,7}	3
{2,3,7}	3
{4,6,7}	3

Reduce the size of C3 using the apriori property (any k-1) subset of a candidate must be frequent. Scan the dataset to get the support count

→

3-Itemsets	Sup_count
{1,2,3}	3
{1,2,7}	3
{1,3,7}	3
{1,4,6}	4
{1,4,7}	3
{1,5,7}	3
{1,6,7}	3
{2,3,7}	3
{4,6,7}	3

L3

C3 (L2 XL2)

Step 4: generating 4 item frequent sets:

4-itemsets	Sup_count
{1,2,3,7}	3
{1,4,6,7}	3

→

4- Itemsets	Sup_count
{1,2,3,7}	3
{1,4,6,7}	3

L4

C4

C5 =ϕ since the join of L4 and L4 doesn't generate any 5- itemsets and the algorithm terminates having found all the frequent item sets

These frequent itemsets are used to generate strong rules which satisfy both minimum support (3) and minimum confidence(70%).

Generating Association Rules from Frequent item sets
For each frequent itemset l, generate all nonempty subsets of l. For every nonempty subset of s of l, output rule s=> (l-s) if conf(s =>(l-s)) >= min_conf.
The frequent itemsets in this case are {1,2,3,7} and {1,4,6,7}
For itemset {1,2,3,7} all the non empty subsets are

{1}, {2}, {3},{7}, {1,2}, {1,3},{1,7}, {2,3}, {2,7}, {3,7} {1,2,3},{1,2,7}, {2,3,7}, {1,3,7}
Let's take l = {1,2,3,7}
Min confidence = 70%

Rule	s=>(l-s)	Confidence	Selected/Rejected
R1	1 =>(2,3,7)	sc{1,2,3,7}/sc{1} =3/6 = 50%	Rejected
R2	2 =>(1,3,7)	sc{1,2,3,7}/sc{2}= ¾ =75%	Selected
R3	3 =>(1,2,7)	sc{1,2,3,7}/sc{3} =3/4 =75%	Selected
R4	7=>(1,2,3)	sc{1,2,3,7}/sc{7} =3/6 =50%	Rejected
R5	(1,2) =>(3,7)	sc{1,2,3,7}/sc{1,2} =3/3 =100%	Selected
R6	(1,3) => (2,7)	sc{1,2,3,7}/sc{1,3} =3/3 =100%	Selected
R7	(1,7)=>(2,3)	sc{1,2,3,7}/sc{1,7}=3/5 =60%	Rejected
R8	(2,3) =>(1,7)	sc{1,2,3,7}/sc{2,3}= 3/3=100%	Selected
R9	(2,7) => (1,3)	sc{1,2,3,7}/sc{2,7} =3/4 =75%	Selected
R10	(3,7) => (1,2)	sc{1,2,3,7}/sc{3,7}=3/3 =100%	Selected
R11	(1,2,3) => 7	sc{1,2,3,7}/sc{1,2,3} }=3/3 =100%	Selected
R12	(1,2,7) => 3	sc{1,2,3,7}/sc{1,2,7}}=3/3 =100%	Selected
R13	(1,3,7)=> 2	sc{1,2,3,7}/sc{1,3,7}}=3/3 =100%	Selected
R14	(2,3,7) => 1	sc{1,2,3,7}/sc{2,3,7}}=3/3 =100%	Selected

In this way we have found 11 strong rules

For itemset {1,4,6,7} all the non empty subsets are
{1}, {4}, {6},{7}, {1,4}, {1,6},{1,7}, {4,6}, {4,7}, {6,7} {1,4,6},{1,4,7}, {1,6,7}, {4,6,7}
Let's take l = {1,4,6,7}

Rule	s=>(l-s)	Confidence	Selected/Rejected
R1	1 =>(4,6,7)	sc{1,4,6,7}/sc{1} =3/6 = 50%	Rejected
R2	4=>(1,6,7)	sc{1,4,6,7}/sc{4}= ¾ =75%	Selected
R3	6 =>(1,4,7)	sc{1,4,6,7}/sc{6} =3/4 =75%	Selected
R4	7=>(1,4,6)	sc{1,4,6,7}/sc{7} =3/6 =50%	Rejected
R5	(1,4) =>(6,7)	sc{1,4,6,7}/sc{1,4} =3/4=75%	Selected
R6	(1,6) => (4,7)	sc{1,4,6,7}/sc{1,6} =3/4 =75%	Selected
R7	(1,7)=>(4,6)	sc{1,4,6,7}/sc{1,7}=3/5=60%	Rejected
R8	(4,6) =>(1,7)	sc{1,4,6,7}/sc{4,6}= 3/4=75%	Selected
R9	(4,7) => (1,6)	sc{1,4,6,7}/sc{4,7} =3/3 =100%	Selected
R10	(6,7) => (1,4)	sc{1,4,6,7}/sc{6,7}=3/3 =100%	Selected
R11	(1,4,6) => 7	sc{1,4,6,7}/sc{1,4,6} }=3/4 =75%	Selected
R12	(1,4,7) => 6	sc{1,4,6,7}/sc{1,4,7}}=3/3 =100%	Selected
R13	(1,6,7)=> 4	sc{1,4,6,7}/sc{1,6,7}}=3/3 =100%	Selected
R14	(4,6,7) => 1	sc{1,4,6,7}/sc{4,6,7}}=3/3 =100%	Selected

In this way we have found, 11 strong rules.
There will be total of 22 strong rules.