Name: Jardel Mitchell

Student ID# 816027213

Course: Introduction to Data Analytics

Assignment 2

Jardel Mitchell
816027213

Transactional data set D

TID	Items
1	J, M, S
2	J, R, S
3	G, M, R, S
4	G, J, M, R, S
5	G, M, S
6	G, M, R

unsorted	sorted
=> J = 3	G = 4
M = 5	J = 3
S = 5	M = 5
G = 4	R = 4
R = 4	S = 5

min_sup = 3

$C_1 = G_4, J_3, M_5, R_4, S_5$

Set of frequency 1-itemset

$L_1 = \{\{G\}:4, \{J\}:3, \{M\}:5, \{R\}:4, \{S\}:5$

$C_2 = L_1 \bowtie L_1$

$C_2 = \{\{G,J\}, \{G,M\}, \{G,R\}, \{G,S\}, \{J,M\}, \{J,R\}, \{J,S\}, \{M,R\}, \{M,S\}, \{R,S\}\}$

$C_2 = \{\{G,J\}:1, \{G,M\}:4, \{G,R\}:3, \{G,S\}:3, \{J,M\}:2, \{J,R\}:2, \{J,S\}:3, \{M,R\}:3, \{M,S\}:4, \{R,S\}:3\}$

Set of frequency 2-itemset

$L_2 = \{\{G,M\}:4, \{GR\}:3, \{GS\}:3, \{JS\}:3, \{MR\}:3, \{MS\}:4, \{R,S\}:3\}$

removed $\{\{G,J\}:2, \{JM\}:2, \{JR\}:2\}$ < min_sup

816027213

$C_3 = L_2 \bowtie L_2$

$C_3 = \{\{GMR\}, \{GMS\}, \{MRS\}\}$

$C_3 = \{\{GMR\}:3, \{GMS\}:3, \{MRS\}:2\}$

Set frequency 3-itemset

$L_3 = \{\{GMR\}:3, \{GMS\}:3\}$

$C_4 = L_3 \bowtie L_3$

$C_4 = \{\{GMRS\}\}$

$C_4 = \{\{GMRS\}:2\}$

The subset of {MSRG}{GMRS}, {GMR} {{GMS} are infrequent than {GMRS} which is also infrequent. {GMRS} = 2 in terms of support which is less than the min_sup, which is 3.

$L_4 = \emptyset$ and the algorithm ends

816027213

(b)

TID	Items	sup count
G	3, 4, 5, 6	4
J	1, 2, 4	3
M	1, 3, 4, 5, 6	5
R	2, 3, 4, 6	4
S	1, 2, 3, 4, 5	5

$C_1 = \{\{G\}:4, \{J\}:3, \{M\}:5, \{R\}:4, \{S\}:5\}$

Frequency 1-itemsets in vertical dat format

TID	Items
G	3, 4, 5, 6
J	1, 2, 4
M	1, 3, 4, 5, 6
R	2, 3, 4, 6
S	1, 2, 3, 4, 5

$C_2 = \{\{GJ\}, \{GM\}, \{GR\}, \{GS\}, \{JM\}, \{JR\}, \{JS\}, \{MR\}, \{MS\}, \{RS\}\}$

816027213

TID	Item	Sup-count	
GJ	{4}	1	→ removed
GM	{3,4,5,6}	4	
GR	{3,4,6}	3	
GS	{3,4,5}	3	
JM	{1,4}	2	- removed
JR	{2,4}	2	- removed
JS	{1,2,4}	3	
MR	{3,4,6}	3	
MS	{1,3,4,5}	4	
RS	{2,3,4}	3	

Frequency 2 itemset in vertical Data Format

IID	Item
GM	{3,4,5,6}
GR	{3,4,6}
GS	{3,4,5}
JS	{1,2,4}
MR	{3,4,6}
MS	{1,3,45}
RS	{2,3,4}

$C_3 = \{GMR\}, \{GMS\}, \{MRS\}$

816027213

Candidate 3 - item in Vertical Data Format

TID	Items	supcount
GMR	{3, 4, 6}	3
GMS	{3, 4, 5}	3
MRS	{3, 4}	2

Frequency 3 itemset in vertical Data Format

TID	Items
GMR	{3, 4, 6}
GMS	{3, 4, 5}

$C_4 = \{GMRS\}$

Candidate 4 - item in vertical Data Format

TID	Items	count
GMRS	{3, 4}	2

Frequency 4 - itemset in vertical Data Format

TID	item
Q	Q

$\therefore C_4 = Q \rightarrow F_4 = \emptyset$ & algorithm terminates

816027213

Q2

Point	x-coordinates	y-coordinates
p_1	1	1
p_2	1.5	2
p_3	4	4
p_4	5	5
p_5	6	4.6
p_6	4	3

(a) Formula → distance matrix $\sqrt{(y_2-y_1)^2+(x_2-x_1)^2}$

	p_1	p_2	p_3	p_4	p_5	p_6
p_1	0	1.118	4.243	5.657	6.161	3.606
p_2	1.118	0	3.202	4.610	5.197	2.693
p_3	4.243	3.202	0	1.414	2.088	(1.000)
p_4	5.657	4.610	1.414	0	1.077	2.236
p_5	6.161	5.197	2.088	1.077	0	2.561
p_6	3.606	2.693	(1.000)	2.236	2.561	0

The smallest value is 1.000 in p_3 & p_6, therefore merge p_3 & p_6

8160 2723

	p_1	p_2	p_3p_6	p_4	p_5
(b) p_1	0	1.118	4.243	5.657	6.161
p_2	1.118	0	3.202	4.610	5.197
p_3p_6	4.243	3.202	0	2.236	2.561
p_4	5.657	4.610	2.236	0	1.077
p_5	6.161	5.197	2.561	1.077	0

The smallest value is 1.077 ∴ merge p_4 & p_5

	p_1	p_2	p_3p_6	p_4p_5
p_1	0	1.118	4.243	6.161
p_2	1.118	0	3.202	5.197
p_3p_6	4.243	3.202	0	2.561
p_4p_5	6.161	5.197	2.561	0

The smallest value is 1.118 ∴ merge p_1 & p_2

	p_1p_2	p_3p_6	p_4p_5
p_1p_2	0	4.243	6.161
p_3p_6	4.243	0	2.561
p_4p_5	6.161	2.561	0

The smallest value is 2.561 ∴ merge p_3p_6 & p_4p_5

	p_1p_2	$p_3p_4p_5p_6$
p_1p_2	0	6.161
$p_3p_4p_5p_6$	6.161	0

The smallest value is 6.161 ∴ merge p_1p_2 & $p_3p_4p_5p_6$

816027213

p1 p2 p3 p4 p5 p6

p1 p2 p3 p4 p5 p6
0

c - Group Average Linkage

The distance matrix in part a indicates that the value nearest to 0 is 1.000. ∴ merge p3 & p6

	p1	p2	p3p6	p4	p5
p1	0	1.118	3.924	5.657	6.161
p2	1.118	0	2.948	4.610	5.197
p3p6	3.924	2.948	0	1.825	2.325
p4	5.627	4.610	1.825	0	1.077
p5	6.161	5.197	2.325	1.077	0

The smallest value is 1.077 ∴ merge p4 & p5

	p1	p2	p3p6	p4p5
p1	0	1.118	3.924	5.909
p2	1.118	0	2.948	4.904
pp6	3.924	2.948	0	2.075
p4p5	5.909	4.904	2.075	0

The smallest value is 1.118 ∴ merge p1 & p2

8160272C3

	p1 p2	p3 p6	p4 p5
p1 p2	0	3.436	5.407
p3 p6	3.436	0	2.075
p4 p5	5.407	2.075	0

The smallest value is 2.075 ∴ merge p3 p6 & p4 p5

	p1 p2	p3 p4 p5 p6
p1 p2	0	4.422
p3 p4 p5 p6	4.422	0

The smallest value is 4.422 ∴ merge p1 p2 & p3 p4 p5 p6

	p1 p2 p3 p4 p5 p6
p1 p2 p3 p4 p5 p6	0