Mining massive Datasets WS 2017/18

Problem Set 8

Rudolf Chrispens, Marvin Klaus, Daniela Schacherer

December 21, 2017

Exercise 01

MISSING

Exercise 02

MISSING

Exercise 03

```
{1, 2, 3} {2, 3, 4} {3, 4, 5} {4, 5, 6}
{1, 3, 5} {2, 4, 6} {1, 3, 4} {2, 4, 5}
{3, 5, 6} {1, 2, 4} {2, 3, 5} {3, 4, 6}

Threshold: 4 PCY Algorithm we use a has table with 11 buckets
```

Table 1 – 3a) Items

Items	1	2	3	4	5	6
Support	4	6	8	8	6	4

c) 2,3 and 4 are Frequent, because their count exceeds our threshold of 4.

d) The count of each pair element in the Bucket.

Exercise 04

MISSING

Table 2 – 3a) Itempairs

Baskets	pairs	support
1,2,3	(1,2)(1,3)(2,3)	2/3/3
2,3,4	(2,3)(2,4)(3,4)	3/4/4
3,4,5	(3,4)(3,5)(4,5)	4/4/3
4,5,6	(4,5)(4,6)(5,6)	3/3/2
1,3,5	(1,3)(1,5)(3,5)	3/1/4
2,4,6	(2,4)(2,6)(4,6)	4/1/3
1,3,4	(1,3)(1,4)(3,4)	3/2/4
2,4,5	(2,4)(2,5)(4,5)	4/2/3
3,5,6	(3,5)(3,6)(5,6)	4/2/2
1,2,4	(1,2)(1,4)(2,4)	2/2/4
2,3,5	(2,3)(2,5)(3,5)	3/2/4
3,4,6	(3,4)(3,6)(4,6)	4/2/3

Table 3 – 3b) Hashbucket

Bucket	1	2	3	4	5	6	7	8	9	10	11
Count	2	10	11	12	0	0	0	0	0	0	0
Pairs	(1,5) (2,6)	(1,2) (5,6) (1,4) (2,5) (3,6)	(1,3) (2,3) (4,5) (4,6)	(2,4) (3,4) (3,5)							