Q1: Using inbuilt packages to compare Apriori with FP-Growth.

transactions1 = [[1, 2, 5],

[2, 4],

[2, 3],

[1, 2, 4],

[1, 3],

[2, 3],

[1, 3],

[1, 2, 3, 5],

[1, 2, 3]]

transactions2 = [[1, 3, 7],

[2, 3, 7],

[1, 2, 3],

[2, 3],

[2, 3, 4, 5],

[2, 3],

[1, 2, 3, 4, 6],

[2, 3, 4, 6],

[1],

[1, 3]]

Minimum support=2 (both databases)

Output:

(asbd) dell@dell-Inspiron-3558:~/Desktop/Semester 8/Big Data/Lab/L4-18-2-20$ python 1.py

Size Apriori Time FP Growth time

9 5.4836273193359375e-06 0.00032806396484375

10 1.1444091796875e-05 0.0003178119659423828

Q2: Apriori with transaction reduction to improve time complexity.

transactions = [[1, 2, 5],

[2, 4],

[2, 3],

[1, 2, 4],

[1, 3],

[2, 3],

[1, 3],

[1, 2, 3, 5],

[1, 2, 3]]

i) Minimum support=2

Output:

(asbd) dell@dell-Inspiron-3558:~/Desktop/Semester 8/Big Data/Lab/L4-18-2-20$ python 2.py

0

{'5': 2, '3': 6, '4': 2, '1': 6, '2': 7} ['5', '3', '4', '1', '2']

[['2', '5', '1'], ['2', '4'], ['2', '3'], ['2', '4', '1'], ['3', '1'], ['2', '3'], ['3', '1'], ['2', '5', '3', '1'], ['2', '3', '1']]

1

{'3-5': 1, '3-4': 0, '4-5': 0, '1-5': 2, '1-3': 4, '1-4': 1, '1-2': 4, '2-5': 2, '2-3': 4, '2-4': 2} ['1-5', '1-3', '1-2', '2-5', '2-3', '2-4']

[['2', '5', '1'], ['2', '4', '1'], ['2', '5', '3', '1'], ['2', '3', '1']]

2

{'1-3-5': 1, '1-2-5': 2, '1-2-3': 2, '2-3-5': 1, '2-3-4': 0, '2-4-5': 0} ['1-2-5', '1-2-3']

[['2', '5', '3', '1']]

3

{'1-2-3-5': 1} []

[]

Frequent item sets

['5', '3', '4', '1', '2', '1-5', '1-3', '1-2', '2-5', '2-3', '2-4', '1-2-5', '1-2-3']

0.0006189346313476562

ii) Minimum Support =3

Output:

(asbd) dell@dell-Inspiron-3558:~/Desktop/Semester 8/Big Data/Lab/L4-18-2-20$ python 2.py

0

{'4': 2, '1': 6, '2': 7, '5': 2, '3': 6} ['1', '2', '3']

[['2', '1'], ['2', '3'], ['2', '1'], ['1', '3'], ['2', '3'], ['1', '3'], ['2', '1', '3'], ['2', '1', '3']]

1

{'1-2': 4, '1-3': 4, '2-3': 4} ['1-2', '1-3', '2-3']

[['2', '1', '3'], ['2', '1', '3']]

2

{'1-2-3': 2} []

[]

Frequent item sets

['1', '2', '3', '1-2', '1-3', '2-3']

0.00031948089599609375

Questions 3,4,6,7 use the same database and will all have the same output.

Q3 – Transaction Reduction using Vertical Transaction Database - Output:

(asbd) dell@dell-Inspiron-3558:~/Desktop/Semester 8/Big Data/Lab/L4-18-2-20$ python 3.py

9

{'1': 0, '2': 0, '3': 0, '4': 0, '5': 0}

{'1': 6, '2': 7, '3': 6, '4': 2, '5': 2}

['1', '2', '3', '4', '5']

[[1 1 0 0 1]

[0 1 0 1 0]

[0 1 1 0 0]

[1 1 0 1 0]

[1 0 1 0 0]

[0 1 1 0 0]

[1 0 1 0 0]

[1 1 1 0 1]

[1 1 1 0 0]] [3, 2, 2, 3, 2, 2, 2, 4, 3]

9

{'1-2': 0, '1-3': 0, '1-4': 0, '1-5': 0, '2-3': 0, '2-4': 0, '2-5': 0, '3-4': 0, '3-5': 0, '4-5': 0}

{'1-2': 4, '1-3': 4, '1-4': 1, '1-5': 2, '2-3': 4, '2-4': 2, '2-5': 2, '3-4': 0, '3-5': 1, '4-5': 0}

['1-2', '1-3', '1-5', '2-3', '2-4', '2-5']

[[1 1 0 0 1]

[1 1 0 1 0]

[1 1 1 0 1]

[1 1 1 0 0]] [3, 3, 4, 3]

4

{'1-2-3': 0, '1-2-5': 0, '1-3-5': 0, '2-3-4': 0, '2-3-5': 0, '2-4-5': 0}

{'1-2-3': 2, '1-2-5': 2, '1-3-5': 1, '2-3-4': 0, '2-3-5': 1, '2-4-5': 0}

['1-2-3', '1-2-5']

[[1 1 1 0 1]] [4]

1

{'1-2-3-5': 0}

{'1-2-3-5': 1}

[]

[] []

Frequent item sets

['1', '2', '3', '4', '5', '1-2', '1-3', '1-5', '2-3', '2-4', '2-5', '1-2-3', '1-2-5']

Q4 - ECLAT Algorithm - Output:

(asbd) dell@dell-Inspiron-3558:~/Desktop/Semester 8/Big Data/Lab/L4-18-2-20$ python 4.py

[[1 1 0 0 1]

[0 1 0 1 0]

[0 1 1 0 0]

[1 1 0 1 0]

[1 0 1 0 0]

[0 1 1 0 0]

[1 0 1 0 0]

[1 1 1 0 1]

[1 1 1 0 0]]

[0, 3, 4, 6, 7, 8]

[0, 1, 2, 3, 5, 7, 8]

[2, 4, 5, 6, 7, 8]

[1, 3]

[0, 7]

1 [0, 3, 4, 6, 7, 8]

2 [0, 1, 2, 3, 5, 7, 8]

3 [2, 4, 5, 6, 7, 8]

4 [1, 3]

5 [0, 7]

1 2 [0, 3, 4, 6, 7, 8] [0, 1, 2, 3, 5, 7, 8] [0, 3, 7, 8]

1 3 [0, 3, 4, 6, 7, 8] [2, 4, 5, 6, 7, 8] [4, 6, 7, 8]

1 4 [0, 3, 4, 6, 7, 8] [1, 3] [3]

1 5 [0, 3, 4, 6, 7, 8] [0, 7] [0, 7]

2 3 [0, 1, 2, 3, 5, 7, 8] [2, 4, 5, 6, 7, 8] [2, 5, 7, 8]

2 4 [0, 1, 2, 3, 5, 7, 8] [1, 3] [1, 3]

2 5 [0, 1, 2, 3, 5, 7, 8] [0, 7] [0, 7]

3 4 [2, 4, 5, 6, 7, 8] [1, 3] []

3 5 [2, 4, 5, 6, 7, 8] [0, 7] [7]

4 5 [1, 3] [0, 7] []

1-2 [0, 3, 7, 8]

1-3 [4, 6, 7, 8]

1-5 [0, 7]

2-3 [2, 5, 7, 8]

2-4 [1, 3]

2-5 [0, 7]

1-2 1-3 [0, 3, 7, 8] [4, 6, 7, 8] [7, 8]

1-2 1-5 [0, 3, 7, 8] [0, 7] [0, 7]

1-3 1-5 [4, 6, 7, 8] [0, 7] [7]

2-3 2-4 [2, 5, 7, 8] [1, 3] []

2-3 2-5 [2, 5, 7, 8] [0, 7] [7]

2-4 2-5 [1, 3] [0, 7] []

1-2-3 [7, 8]

1-2-5 [0, 7]

1-2-3 1-2-5 [7, 8] [0, 7] [7]

1 6

2 7

3 6

4 2

5 2

1-2 4

1-3 4

1-5 2

2-3 4

2-4 2

2-5 2

1-2-3 2

1-2-5 2

The frequent items instead of being stored in a dictionary, as in the other programs, are objects and at the the end of the algorithm each frequent item is printed along with it’s count.

Q6 Outputs:

Hashing:

(asbd) dell@dell-Inspiron-3558:~/Desktop/Semester 8/Big Data/Lab/L4-18-2-20$ python 6\_a.py

0

{'1': 6, '2': 7, '3': 6, '4': 2, '5': 2} ['1', '2', '3', '4', '5'] 5 1

1

[{'1': 6, '2': 7, '3': 6, '4': 2, '5': 2}, ['1-2', '1-5', '2-5', '2-4', '2-3', '1-3']]

['1-2', '1-5', '2-5', '2-4', '2-3', '1-3'] ['1-2', '1-5', '2-5', '2-4', '2-3', '1-3'] 6 2

2

{'1-2-5': 2, '1-2-3': 2, '2-4-5': 0, '2-3-5': 1, '2-3-4': 0, '1-3-5': 1} ['1-2-5', '1-2-3'] 2 3

3

{'1-2-3-5': 1} [] 0 4

Frequent item sets

['1', '2', '3', '4', '5', '1-2', '1-5', '2-5', '2-4', '2-3', '1-3', '1-2-5', '1-2-3']

Partitioning:

(asbd) dell@dell-Inspiron-3558:~/Desktop/Semester 8/Big Data/Lab/L4-18-2-20$ python 6\_b.py

0

{} ['2', '4', '1', '3', '5'] 5 1

1

{'4-5': 0, '1-4': 1, '3-4': 0, '3-5': 1} ['2-4', '1-2', '2-3', '2-5', '1-3', '1-5'] 6 2

2

{'2-4-5': 0, '2-3-4': 0, '2-3-5': 1, '1-3-5': 1} ['1-2-5', '1-2-3'] 2 3

3

{'1-2-3-5': 1} [] 0 4

Frequent item sets

['2', '4', '1', '3', '5', '2-4', '1-2', '2-3', '2-5', '1-3', '1-5', '1-2-5', '1-2-3']

Sampling:

(asbd) dell@dell-Inspiron-3558:~/Desktop/Semester 8/Big Data/Lab/L4-18-2-20$ python 6\_c.py

['4', '1', '2', '3', '1-2', '1-3', '2-4', '2-3', '1-2-3']

['4', '1', '2', '3', '1-2', '1-3', '2-4', '2-3', '1-2-3', '5', '1-5', '2-5', '1-2-5']

['4', '1', '2', '3', '1-2', '1-3', '2-4', '2-3', '1-2-3', '5', '1-5', '2-5', '1-2-5']

['4', '1', '2', '3', '1-2', '1-3', '2-4', '2-3', '1-2-3', '5', '1-5', '2-5', '1-2-5']

['4', '1', '2', '3', '1-2', '1-3', '2-4', '2-3', '1-2-3', '5', '1-5', '2-5', '1-2-5']

['4', '1', '2', '3', '1-2', '1-3', '2-4', '2-3', '1-2-3', '5', '1-5', '2-5', '1-2-5']

['4', '1', '2', '3', '1-2', '1-3', '2-4', '2-3', '1-2-3', '5', '1-5', '2-5', '1-2-5']

['4', '1', '2', '3', '1-2', '1-3', '2-4', '2-3', '1-2-3', '5', '1-5', '2-5', '1-2-5']

['4', '1', '2', '3', '1-2', '1-3', '2-4', '2-3', '1-2-3', '5', '1-5', '2-5', '1-2-5']

['4', '1', '2', '3', '1-2', '1-3', '2-4', '2-3', '1-2-3', '5', '1-5', '2-5', '1-2-5']

['4', '1', '2', '3', '1-2', '1-3', '2-4', '2-3', '1-2-3', '5', '1-5', '2-5', '1-2-5']

['4', '1', '2', '3', '1-2', '1-3', '2-4', '2-3', '1-2-3', '5', '1-5', '2-5', '1-2-5']

['4', '1', '2', '3', '1-2', '1-3', '2-4', '2-3', '1-2-3', '5', '1-5', '2-5', '1-2-5']

['4', '1', '2', '3', '1-2', '1-3', '2-4', '2-3', '1-2-3', '5', '1-5', '2-5', '1-2-5']

['4', '1', '2', '3', '1-2', '1-3', '2-4', '2-3', '1-2-3', '5', '1-5', '2-5', '1-2-5']

['4', '1', '2', '3', '1-2', '1-3', '2-4', '2-3', '1-2-3', '5', '1-5', '2-5', '1-2-5']

['4', '1', '2', '3', '1-2', '1-3', '2-4', '2-3', '1-2-3', '5', '1-5', '2-5', '1-2-5']

['4', '1', '2', '3', '1-2', '1-3', '2-4', '2-3', '1-2-3', '5', '1-5', '2-5', '1-2-5']

['4', '1', '2', '3', '1-2', '1-3', '2-4', '2-3', '1-2-3', '5', '1-5', '2-5', '1-2-5']

['4', '1', '2', '3', '1-2', '1-3', '2-4', '2-3', '1-2-3', '5', '1-5', '2-5', '1-2-5']

['4', '1', '2', '3', '1-2', '1-3', '2-4', '2-3', '1-2-3', '5', '1-5', '2-5', '1-2-5']

['4', '1', '2', '3', '1-2', '1-3', '2-4', '2-3', '1-2-3', '5', '1-5', '2-5', '1-2-5']

['4', '1', '2', '3', '1-2', '1-3', '2-4', '2-3', '1-2-3', '5', '1-5', '2-5', '1-2-5']

['4', '1', '2', '3', '1-2', '1-3', '2-4', '2-3', '1-2-3', '5', '1-5', '2-5', '1-2-5']

['4', '1', '2', '3', '1-2', '1-3', '2-4', '2-3', '1-2-3', '5', '1-5', '2-5', '1-2-5']

['4', '1', '2', '3', '1-2', '1-3', '2-4', '2-3', '1-2-3', '5', '1-5', '2-5', '1-2-5']

['4', '1', '2', '3', '1-2', '1-3', '2-4', '2-3', '1-2-3', '5', '1-5', '2-5', '1-2-5']

['4', '1', '2', '3', '1-2', '1-3', '2-4', '2-3', '1-2-3', '5', '1-5', '2-5', '1-2-5']

['4', '1', '2', '3', '1-2', '1-3', '2-4', '2-3', '1-2-3', '5', '1-5', '2-5', '1-2-5']

['4', '1', '2', '3', '1-2', '1-3', '2-4', '2-3', '1-2-3', '5', '1-5', '2-5', '1-2-5']

Frequent item sets

['4', '1', '2', '3', '1-2', '1-3', '2-4', '2-3', '1-2-3', '5', '1-5', '2-5', '1-2-5']

Q7 – Dynamic Itemset Counting

i)When M=5

Output:

((asbd) dell@dell-Inspiron-3558:~/Desktop/Semester 8/Big Data/Lab/L4-18-2-20$ python 7.py

['1', '2', '3', '4', '5']

[{'1': 0, '2': 0, '3': 0, '4': 0, '5': 0}, {'1-2': 0, '1-3': 0, '1-4': 0, '1-5': 0, '2-3': 0, '2-4': 0, '2-5': 0, '3-4': 0, '3-5': 0, '4-5': 0}, {'1-2-3': 0, '1-2-4': 0, '1-2-5': 0, '1-3-4': 0, '1-3-5': 0, '1-4-5': 0, '2-3-4': 0, '2-3-5': 0, '2-4-5': 0, '3-4-5': 0}, {'1-2-3-4': 0, '1-2-3-5': 0, '1-2-4-5': 0, '1-3-4-5': 0, '2-3-4-5': 0}, {'1-2-3-4-5': 0}]

{} {} {'1': 3, '2': 4, '3': 2, '4': 2} {'5': 1, '1-2': 0, '1-3': 0, '2-3': 0, '1-4': 0, '2-4': 0, '3-4': 0}

{'1': 6, '2': 7, '3': 6, '4': 2, '5': 2} {} {'1-2': 2, '1-3': 3, '2-3': 3} {'1-4': 0, '2-4': 0, '3-4': 0, '1-5': 0, '2-5': 0, '3-5': 0, '4-5': 0, '1-2-3': 0}

{'1': 6, '2': 7, '3': 6, '4': 2, '5': 2, '1-2': 4, '1-3': 4, '2-3': 4, '2-4': 2} {'1-4': 1, '3-4': 0} {} {'1-5': 1, '2-5': 1, '3-5': 0, '4-5': 0, '1-2-3': 0}

{'1': 6, '2': 7, '3': 6, '4': 2, '5': 2, '1-2': 4, '1-3': 4, '2-3': 4, '2-4': 2, '1-5': 2, '2-5': 2, '1-2-3': 2} {'1-4': 1, '3-4': 0, '3-5': 1, '4-5': 0} {} {'1-2-5': 0}

{'1': 6, '2': 7, '3': 6, '4': 2, '5': 2, '1-2': 4, '1-3': 4, '2-3': 4, '2-4': 2, '1-5': 2, '2-5': 2, '1-2-3': 2} {'1-4': 1, '3-4': 0, '3-5': 1, '4-5': 0} {} {'1-2-5': 1}

{'1': 6, '2': 7, '3': 6, '4': 2, '5': 2, '1-2': 4, '1-3': 4, '2-3': 4, '2-4': 2, '1-5': 2, '2-5': 2, '1-2-3': 2, '1-2-5': 2} {'1-4': 1, '3-4': 0, '3-5': 1, '4-5': 0} {} {}

Frequent itemsets

{'1': 6, '2': 7, '3': 6, '4': 2, '5': 2, '1-2': 4, '1-3': 4, '2-3': 4, '2-4': 2, '1-5': 2, '2-5': 2, '1-2-3': 2, '1-2-5': 2}

ii)When M=3

Output:

(asbd) dell@dell-Inspiron-3558:~/Desktop/Semester 8/Big Data/Lab/L4-18-2-20$ python 7.py

['1', '2', '3', '4', '5']

[{'1': 0, '2': 0, '3': 0, '4': 0, '5': 0}, {'1-2': 0, '1-3': 0, '1-4': 0, '1-5': 0, '2-3': 0, '2-4': 0, '2-5': 0, '3-4': 0, '3-5': 0, '4-5': 0}, {'1-2-3': 0, '1-2-4': 0, '1-2-5': 0, '1-3-4': 0, '1-3-5': 0, '1-4-5': 0, '2-3-4': 0, '2-3-5': 0, '2-4-5': 0, '3-4-5': 0}, {'1-2-3-4': 0, '1-2-3-5': 0, '1-2-4-5': 0, '1-3-4-5': 0, '2-3-4-5': 0}, {'1-2-3-4-5': 0}]

{} {} {'2': 3} {'1': 1, '3': 1, '4': 1, '5': 1}

{} {} {'2': 5, '1': 3, '3': 3, '4': 2} {'5': 1, '1-2': 0, '1-3': 0, '2-3': 0, '1-4': 0, '2-4': 0, '3-4': 0}

{'2': 7, '1': 6, '3': 6, '4': 2, '5': 2} {} {'1-2': 2, '1-3': 3, '2-3': 2} {'1-4': 0, '2-4': 0, '3-4': 0, '1-5': 0, '2-5': 0, '3-5': 0, '4-5': 0, '1-2-3': 0}

{'2': 7, '1': 6, '3': 6, '4': 2, '5': 2} {} {'1-2': 3, '1-3': 3, '2-3': 3} {'1-4': 0, '2-4': 1, '3-4': 0, '1-5': 1, '2-5': 1, '3-5': 0, '4-5': 0, '1-2-3': 0}

{'2': 7, '1': 6, '3': 6, '4': 2, '5': 2, '1-2': 4, '1-3': 4, '2-3': 4, '2-4': 2} {'1-4': 1, '3-4': 0} {} {'1-5': 1, '2-5': 1, '3-5': 0, '4-5': 0, '1-2-3': 0}

{'2': 7, '1': 6, '3': 6, '4': 2, '5': 2, '1-2': 4, '1-3': 4, '2-3': 4, '2-4': 2, '1-5': 2, '2-5': 2, '1-2-3': 2} {'1-4': 1, '3-4': 0, '3-5': 1, '4-5': 0} {} {'1-2-5': 0}

{'2': 7, '1': 6, '3': 6, '4': 2, '5': 2, '1-2': 4, '1-3': 4, '2-3': 4, '2-4': 2, '1-5': 2, '2-5': 2, '1-2-3': 2} {'1-4': 1, '3-4': 0, '3-5': 1, '4-5': 0} {} {'1-2-5': 1}

{'2': 7, '1': 6, '3': 6, '4': 2, '5': 2, '1-2': 4, '1-3': 4, '2-3': 4, '2-4': 2, '1-5': 2, '2-5': 2, '1-2-3': 2} {'1-4': 1, '3-4': 0, '3-5': 1, '4-5': 0} {} {'1-2-5': 1}

{'2': 7, '1': 6, '3': 6, '4': 2, '5': 2, '1-2': 4, '1-3': 4, '2-3': 4, '2-4': 2, '1-5': 2, '2-5': 2, '1-2-3': 2, '1-2-5': 2} {'1-4': 1, '3-4': 0, '3-5': 1, '4-5': 0} {} {}

Frequent itemsets

{'2': 7, '1': 6, '3': 6, '4': 2, '5': 2, '1-2': 4, '1-3': 4, '2-3': 4, '2-4': 2, '1-5': 2, '2-5': 2, '1-2-3': 2, '1-2-5': 2}

Q5: Apriori which will differentiate ab from ba

transactions = [[10, 20, 50],

[20, 40],

[20, 30],

[10, 20, 40],

[10, 30],

[20, 30],

[10, 30],

[20, 10, 30, 50],

[20, 10, 30],

[40,50]]

Minimum support=2

Output:

(asbd) dell@dell-Inspiron-3558:~/Desktop/Semester 8/Big Data/Lab/L4-18-2-20$ python 5.py

0

{'40': 3, '30': 6, '20': 7, '10': 6, '50': 3} ['40', '30', '20', '10', '50'] 5 1

1

{'40-30': 0, '40-20': 0, '40-10': 0, '40-50': 1, '30-40': 0, '30-20': 0, '30-10': 0, '30-50': 1, '20-40': 2, '20-30': 4, '20-10': 2, '20-50': 2, '10-40': 1, '10-30': 4, '10-20': 2, '10-50': 2, '50-40': 0, '50-30': 0, '50-20': 0, '50-10': 0} ['20-40', '20-30', '20-10', '20-50', '10-30', '10-20', '10-50'] 7 2

2

{'20-40-50': 0, '20-30-40': 0, '20-30-50': 1, '20-10-40': 0, '20-10-30': 2, '20-10-50': 1, '10-30-50': 1, '10-20-30': 0, '10-20-50': 1} ['20-10-30'] 1 3

Frequent item sets

['40', '30', '20', '10', '50', '20-40', '20-30', '20-10', '20-50', '10-30', '10-20', '10-50', '20-10-30']