# [CSCI 4502] HW-3

Michael Ghattas

9/19/2022

## Question-2

TID items_bought
T1 {A, K, T, X, Z}
T2 {A, H, X, T, Z}
T3 {A, B, D, R, S}
T4 {B, D, H, T, X}
T5 {B, C, H, M, S}

### Part(a)

-> Transactions Count = [T1, T2, T3, T4, T5] = 5

-> Items Scan: [(A = 3), (B = 3), (C = 1), (D = 2), (H = 3), (K = 1), (M = 1), (R = 1), (S = 2), (T = 3), (X = 3), (Z = 2)] = 12

-> L1: [(A = 3), (B = 3), (H = 3), (T = 3), (X = 3)] = 5

-> Item-sets Scan: [(A-B = 1), (A-C = 0), (A-D = 1), (A-H = 1), (A-K = 1), (A-M = 0), (A-R = 1), (A-S = 1), (A-T = 2), (A-X = 2), (A-Z = 2)]
[(B-C = 1), (B-D = 2), (B-H = 2), (B-K = 0), (B-M = 1), (B-R = 1), (B-S = 2), (B-T = 1), (B-X = 1), (B-Z = 0)]
[(C-D = 0), (C-H = 1), (C-K = 0), (C-M = 1), (C-R = 0), (C-S = 1), (C-T = 0), (C-X = 0), (C-Z = 0)]
[(D-H = 1), (D-K = 0), (D-M = 0), (D-R = 1), (D-S = 1), (D-T = 0), (D-X = 1), (D-Z = 0)]
[(H-K = 0), (H-M = 1), (H-R = 0), (H-S = 1), (H-T = 2), (H-X = 2), (H-Z = 1)]
[(K-M = 0), (K-R = 0), (K-S = 0), (K-T = 1), (K-X = 1), (K-Z = 1)]
[(M-R = 0), (M-S = 1), (M-T = 0), (M-X = 0), (M-Z = 0)]
[(R-S = 1), (R-T = 0), (R-X = 0), (R-Z = 0)]
[(S-T = 0), (S-X = 0), (S-Z = 0)]
[(T-X = 3), (T-Z = 2)]

[(X-Z = 2)]

-> L2: [(T-X = 3)] = 1

-> Frequent Item-sets Scan: [(T-X-A = 2), (T-X-B = 1), (T-X-C = 0), (T-X-D = 1), (T-X-H = 2), (T-X-K = 1), (T-X-M = 0),(T-X-R = 0), (T-X-S = 0), (T-X-Z = 2)]

-> L3: [] = 0

-> Since [(T-X = 3)] is the only set, the maximum number of possible frequent item-sets = 1

### Part(b)

-> Min. Support = 0.4 = (2 / 5)

-> Items Scan: [(A = 3), (B = 3), (C = 1), (D = 2), (H = 3), (K = 1), (M = 1), (R = 1), (S = 2), (T = 3), (X = 3), (Z = 2)] = 12

-> L1: [(A = 3), (B = 3), (D = 2), (H = 3), (S = 2), (T = 3), (X = 3), (Z = 2)] = 8

-> Item-sets Scan: [(AB = 1), (AD = 1), (AH = 1), (AK = 1), (AR = 1), (AS = 1), (AT = 2), (AX = 2), (AZ = 2)]
[(BC = 1), (BD = 2), (BH = 2), (BM = 1), (BR = 1), (BS = 2), (BT = 1), (BX = 1)]
[(DH = 1), (DR = 1), (DS = 1), (-X = 1)]
[(HM = 1), (HS = 1), (HT = 2), (HX = 2), (HZ = 1)]
[(TX = 3), (TZ = 2)]
[(XZ = 2)]

-> L2: [(AT = 2), (AX = 2), (AZ = 2), (BD = 2), (BH = 2), (BS = 2), (HT = 2), (HX = 2), (TX = 3), (TZ = 2), (XZ = 2)] = 10

-> Pruning: [(ATX => AT, AX, TX), (ATZ => AT, AZ, TZ)]

-> L3: [(ATX = 2), (ATZ = 2)] = 2

-> Frequent Item-sets: [ATX, ATZ]

### Part(c)

-> Total number of candidate item sets = 3

-> Rounds of database scans needed = 3

### Part(d)

-> First Method: $O(n)$

-> Second Method: $O(n*b*k^m)$

-> $O(n) <= O(nbk^m)$ Thus the first method is always better!

## End.