Quiz 3b: Frequent Itemsets (Quiz type b)
There are two types of quizzes, see solution for b

1) Here is a collection of 6 baskets. Each contains three of the six items 1 through 6. {1, 2, 3}, {2, 3, 4}, {3, 4, 5}, {4, 5, 6}, {1, 3, 5}, {2, 4, 6}. The support threshold is 2. the hash function is j mod 7. Using the PCY Algorithm, you need to show 1. frequent buckets and 2. frequent pairs. (2 pts)

2) Consider the following basket data and a support threshold **s = 3**, answer the following questions.

$\mathbf{B_1 = \{m, c, b\}}$	$\mathbf{B_2 = \{m, p, j\}}$
$\mathbf{B_3 = \{m, b\}}$	$\mathbf{B_4 = \{c, j\}}$
$\mathbf{B_5 = \{m, p, b\}}$	$\mathbf{B_6 = \{m, c, b, j\}}$
$\mathbf{B_7 = \{c, b, j\}}$	$\mathbf{B_8 = \{b, c\}}$

Find all frequent itemsets with set size <= 3 (1pt). Write down the two association rules and **confidence** and **interest** numbers. One of your association rules should be derived from a frequent pair (i.e., X->Y), and the other one should be derived from a frequent triplet (i.e., X, Y->Z) (3pts).

3) In the SON algorithm, please write down the input/output for the Map/Reduce functions in both phases if we use a two-phase Map/Reduce (2pts total, 1pt for each phase).

Phase 1 Map Input:	Phase 1 Map Output:
Phase 1 Reduce Input:	Phase 1 Reduce Output:
Phase 2 Map Input:	Phase 2 Map Output:
Phase 2 Reduce Input:	Phase 2 Reduce Output:

4) Considering the Toivonen's algorithm, give one example of a singleton and one example of a pair in the negative border. You need to explain why your examples are considered as itemsets in the negative border (1pt). Explain how and why the Toivonen's algorithm uses the itemsets in the negative border (you can use your examples) (1pt).

5) If you have n types of items and x baskets, with support = s, at most how many pairs you need to count for finding frequent pairs (1pt)?