Exercise 4

Claim

The Apriori algorithm enumerates the set of frequent itemsets in incremental polynomial time.

Proof

In the first iteration $\mathcal{F}_1$ gets printed because $\mathcal{C}_1 = \mathcal{I}$ and searching for all frequent 1-itemsets takes $|D|$ time. Now in general checking whether a k-itemset is frequent or not can be done in $|D|$ time and in each iteration this has to be done for all k-itemsets in $\mathcal{C}_k$ hence it is polynomial in $|D|$ and $|\mathcal{C}_k|$. So in order to show if each print has polynomial delay we have to proof that **CANDIDATE-GENERATION** is polynomial.
The **CANDIDATEGENERATION** algorithm get $\mathcal{F}_k$ of frequent k-itemsets as an input and return candidates $\mathcal{C}_{k+1}$. Lets note $l = \mathcal{F}_k$ the cardinality. We loop $\mathcal{O}(l^2)$ times over over pair X, Y of k-itemsets that differ only in the last element. Creating the set Z by concatinating strings has polynomial time complexity. And then we check if all k-subsets of Z are in $\mathcal{F}_k$ and this can be done in at most polynomial time in l since the cardinality of C_{k+1} is polynomial in the cardinality of $\mathcal{F}_k$
$\Rightarrow$ The Apriori algorithm is incremental polynomial.