St. Francis Xavier University
Department of Computer Science

CSCI 435: Algorithms and Complexity
Assignment 1
Due February 5, 2025 at 2:30pm

Assignment Regulations.

- This assignment must be completed individually.
- Please include your full name and email address on your submission.
- You may either handwrite or typeset your submission. If your submission is handwritten, please ensure that the handwriting is neat and legible.

[8 marks] 1. The Antigonish Retail Council is analyzing the behaviour of students who purchase from local stores. They have collected data on purchases made by students in one month and stored this data in a two-dimensional array A whose rows are the names of students and whose columns are the items sold by various stores. Each entry $A[i, j]$ in the array corresponds to the quantity of item j that was purchased by student i that month. For example, here is part of the array A:

	Notebooks	**Pens**	**Frozen pizzas**	**Beer**
Jesse	1	2	6	0
Sean	0	0	0	12
Brooke	4	3	0	0
		⋮		

The council now wants to analyze this data; in particular, they want to solve the *diverse subset* problem. Say that a subset of students S is *diverse* if no two students in S have ever purchased the same items (i.e., for each item, at most one student in S has ever bought it). Given an $m \times n$ array A and an integer $k \leq m$, does there exist a subset of at least k students that is diverse?

Unfortunately, this is a hard problem to solve, and we can prove this. Show that Diverse-Subset is NP-complete by reducing from Independent-Set.

[8 marks] 2. Recall the QSAT problem from our lectures. Here, we will consider a variant of the problem where the given Boolean formula has no negated variables.

Let $\phi(x_1, x_2, \ldots, x_n)$ be a Boolean formula in conjunctive normal form; that is, a Boolean formula of the form $C_1 \wedge C_2 \wedge \cdots \wedge C_m$, where each clause C_i is a disjunction of three variables. Furthermore, suppose ϕ is *monotone*; that is, each variable in ϕ is of the form x_i and not $\overline{x_i}$.

The Monotone-QSAT problem asks whether the formula $\exists x_1 \; \forall x_2 \cdots \forall x_{n-1} \; \exists x_n \; \phi(x_1, x_2, \ldots, x_n)$ is true. Interestingly, while QSAT is in PSPACE, Monotone-QSAT is quite a bit easier to solve.

Give an algorithm that decides instances of Monotone-QSAT in polynomial **time** in terms of n.

[7 marks] 3. The HITTING-SET problem is defined as follows. Suppose $A = \{a_1, a_2, \ldots, a_n\}$ is a set, and let $B_1, B_2, \ldots, B_m$ be a collection of subsets of A. We say that a subset $H \subseteq A$ is a *hitting set* for the collection $B_1, B_2, \ldots, B_m$ if H contains at least one element from each subset B_i; in other words, H "hits" each of the subsets B_i.

HITTING-SET is well-known to be NP-complete, but much like we showed in lecture with VERTEX-COVER, we can make this problem tractable in special cases. Suppose we want to determine whether there exists a hitting set of size at most k, and furthermore, we know that each subset B_i contains at most c elements for some constant c.

Using ideas similar to those we used for VERTEX-COVER, describe an algorithm that solves HITTING-SET efficiently. Namely, just like our algorithm for VERTEX-COVER took time $O(2^k \cdot kn)$, your algorithm for HITTING-SET should take time $O(f(k, c) \cdot p)$, where f is an arbitrary function that depends on k and c and p is a polynomial expression. (You do not need to formally analyze your algorithm; a high-level justification of the runtime will suffice.)

[7 marks] 4. Suppose you are given a set of positive nonzero integers $A = \{a_1, a_2, \ldots, a_n\}$ and a positive integer B. A subset $S \subseteq A$ is called *feasible* if $\sum_{a_i \in S} a_i \leq B$; that is, if the sum of the integers in the subset S is not greater than B. Your goal is to find a feasible subset $S \subseteq A$ whose sum is as large as possible (naturally, without exceeding B).

Give a polynomial-time approximation algorithm that takes as input a set A and an integer B and produces as output a feasible subset $S \subseteq A$ whose sum is at least **half** the sum of the optimal feasible subset $S^* \subseteq A$.