This assignment highlighted the strengths of a structured approach to implementing CPU and GPU algorithms for reduction and histogram functions, enabling clear, organized development and validation. Implementing the CPU functions was straightforward, while the GPU functions benefited from parallelization but posed challenges in memory management, thread divergence, and ensuring efficient atomic operations. Debugging CUDA kernels was complex due to limited tools, and optimizing grid and block configurations required careful adjustment. If approaching this assignment again, I would incorporate profiling tools like NVIDIA Nsight earlier to identify bottlenecks sooner, and start with smaller, simpler GPU kernels before scaling, ensuring stable performance and more effective benchmarking across different input sizes.