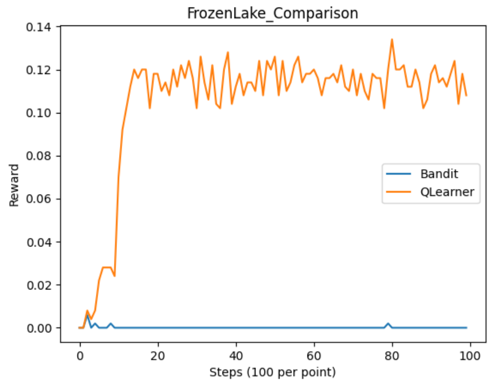
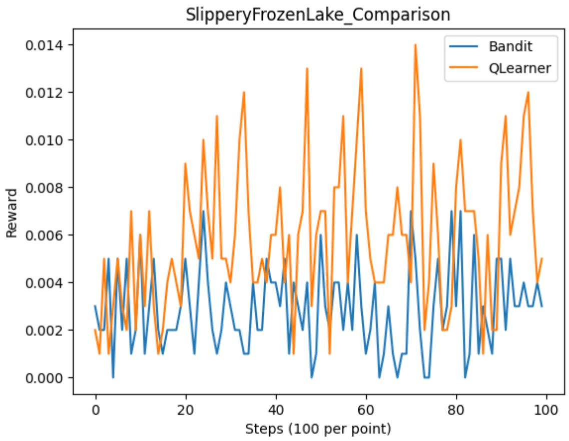
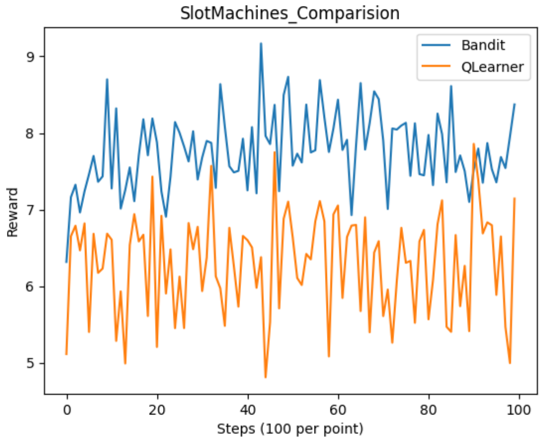
2a. The three plots are presented below:







For the FrozenLake task, we see that the QLearner quickly learns a path to obtain the reward, while the bandit is unable to learn this path. For the slippery frozen lake task, neither learner does particularly well, but the QLearner still outperforms the bandit. For the SlotMachines task, the Bandit outperforms the QLearner, although both are able to extract decent rewards from the game.

2b.

QLearning received higher rewards for the FrozenLake and SlipperyFrozenLake tasks. This is because QLearning takes into account the state of the system when calculating rewards and modifies Q(S,A) by an additional factor of gamma times the maximum achievable reward from the resulting state. It is therefore able to learn to “plan” a path across the lake that will get to the goal and avoid stepping in holes. Thus, given tasks with multiple states such as the FrozenLake and SlipperyFrozenLake tasks, Qlearner outperforms the MultiArmedBandit.

2c.

There is no way for MultiArmedBandit to take into account different states of the system, and without this ability it will always perform worse than QLearning at these tasks. Therefore there is no way to modify its existing hyperparameters to get it to perform as well as QLearning.

2d.

MultiArmedBandit appears to receive higher rewards than QLearning for the SlotMachines task. This is because MultiArmedBandit assumes a stationary world (which is accurate for the SlotMachines task) and is therefore better able to learn the payouts of each slot machine. QLearning by comparison assumes a nonstationary world and only factors in old information with a factor of (1-α). Since that old information is always as relevant as the new information for a stationary system, QLearning is doing itself a disservice in discounting it and as a result the MultiArmedBandit performs better.

2e.

You could change α to be equal to 1/N. This would make QLearning also assume a stationary environment and it would then perform as well as the MultiArmedBandit.