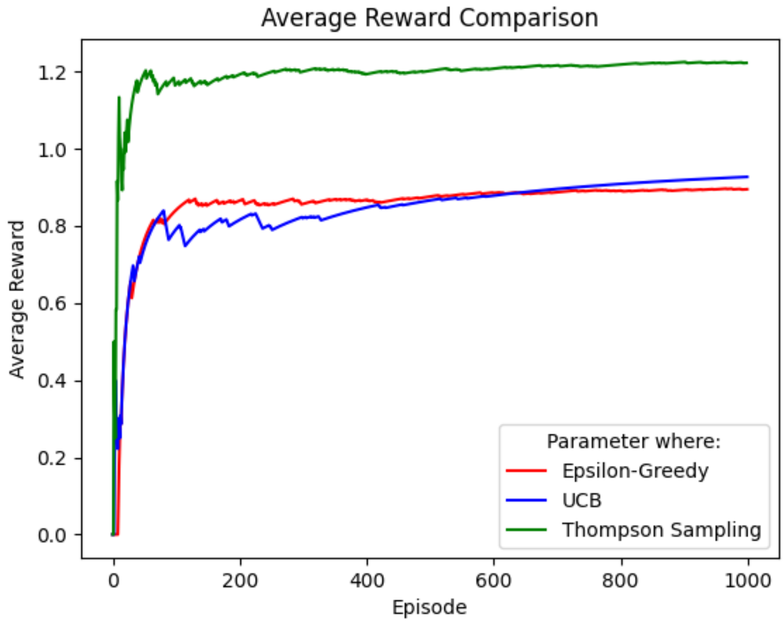
Assignment Chapter 2

61175024H 白偉辰 Nick

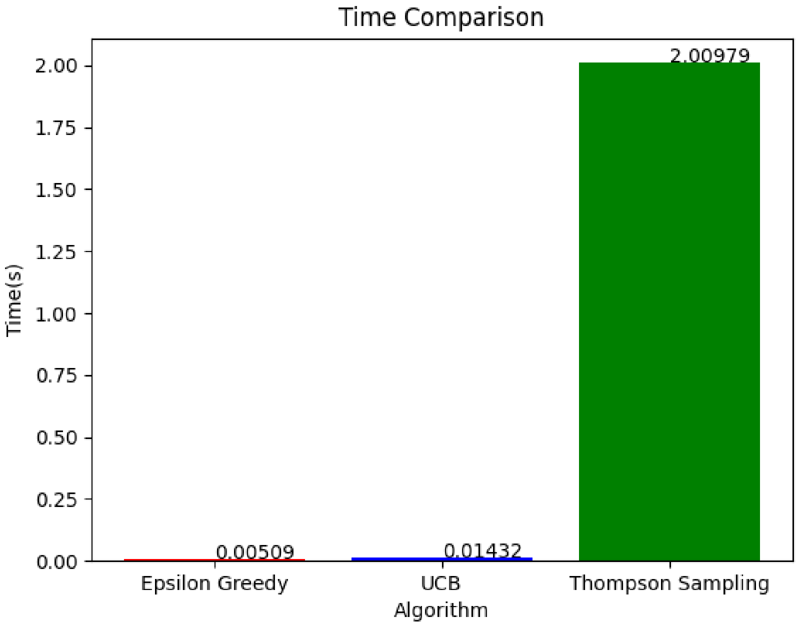
As figure shown below, Thompson Sampling has the the highest mean total reward, second UCB and at last Epsilon-Greedy.

Before 600 episodes, we can see that UCB's reward is a little bit lower than Epsilon-Greedy. Because UCB prefer actions that it hasn't had a confident value estimation for it yet. After enough exploration than UCB will focus on the one which has strong potential to have a optimal value. Therefore after 600 episodes, UCB gradually get better results than Epsilon-Greedy.



As figure shown below, Epsilon-Greedy spent the shortest time, second UCB and Thompson Sampling spent the longest time.

According to the result, more complex algorithm may get better result, but it may also cost more time on calculation. It's a trade-off between "Algorithm Complexity" and "Time Consumption" for developer.


Time Comparison
Time(s)
Epsilon Greedy
UCB
Thompson Sampling
Algorithm