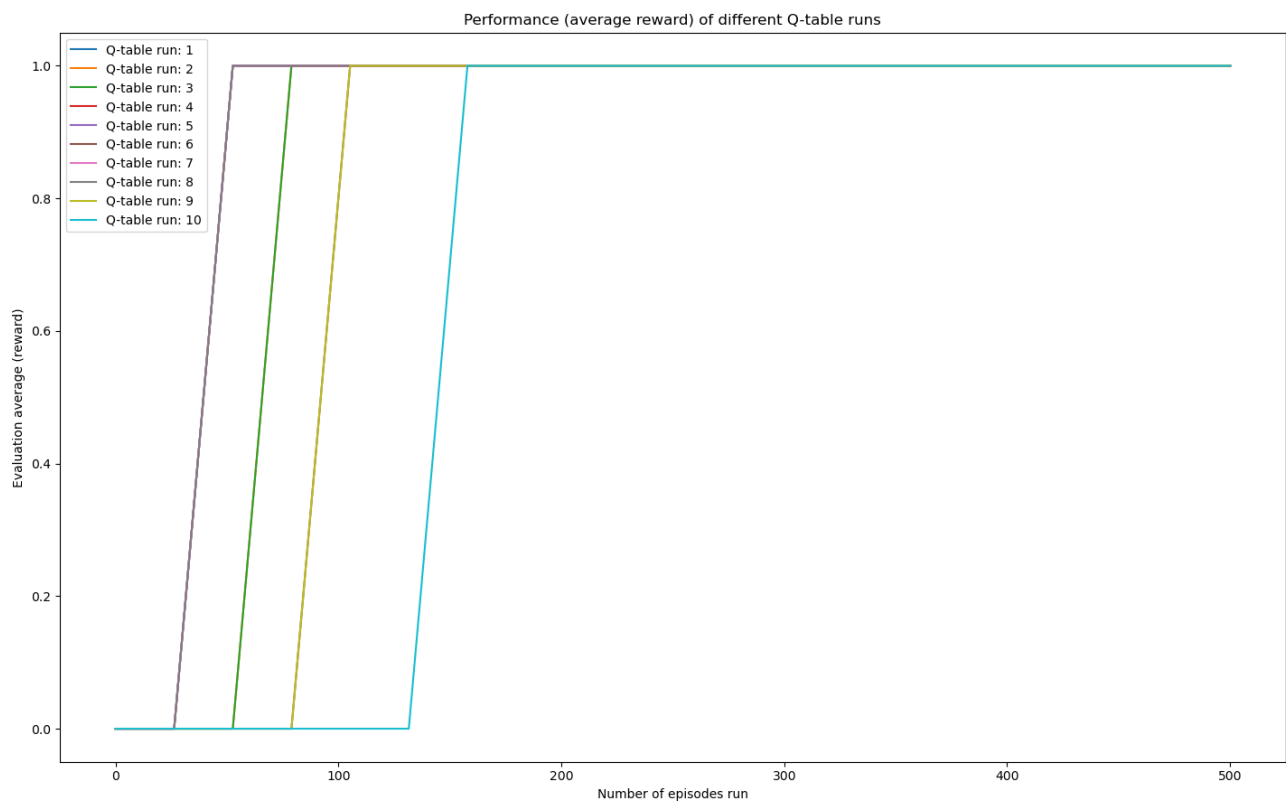
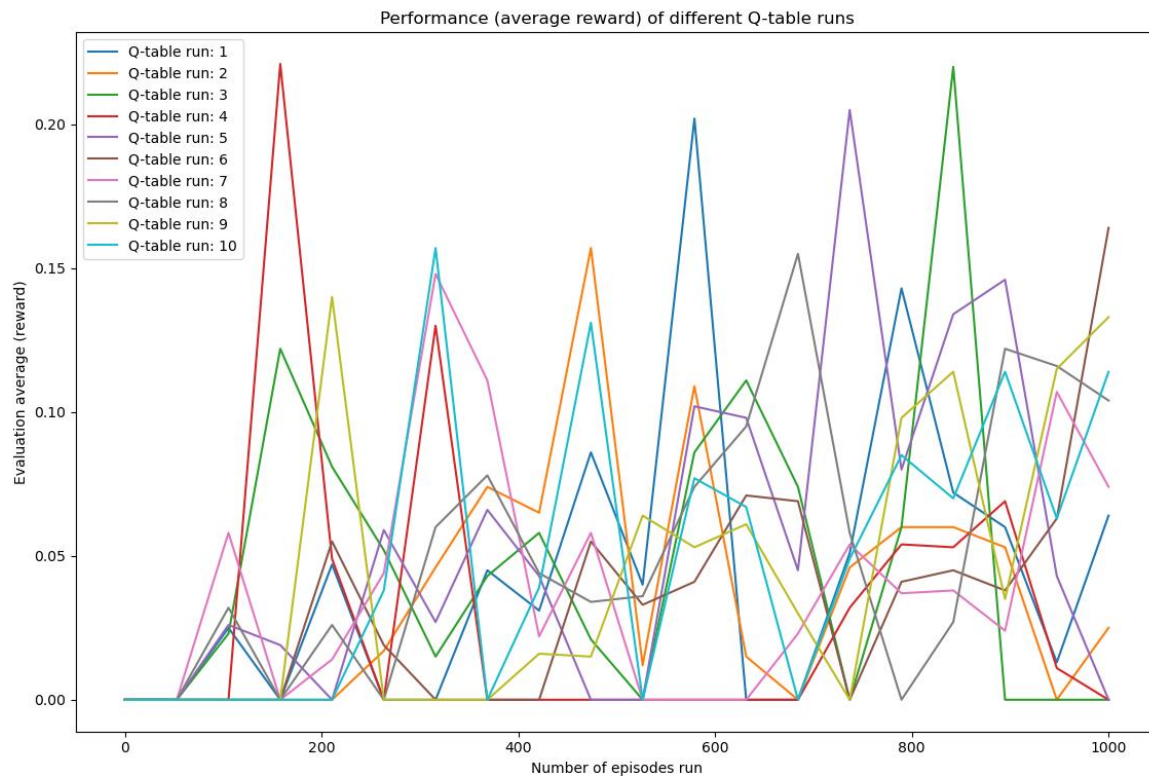


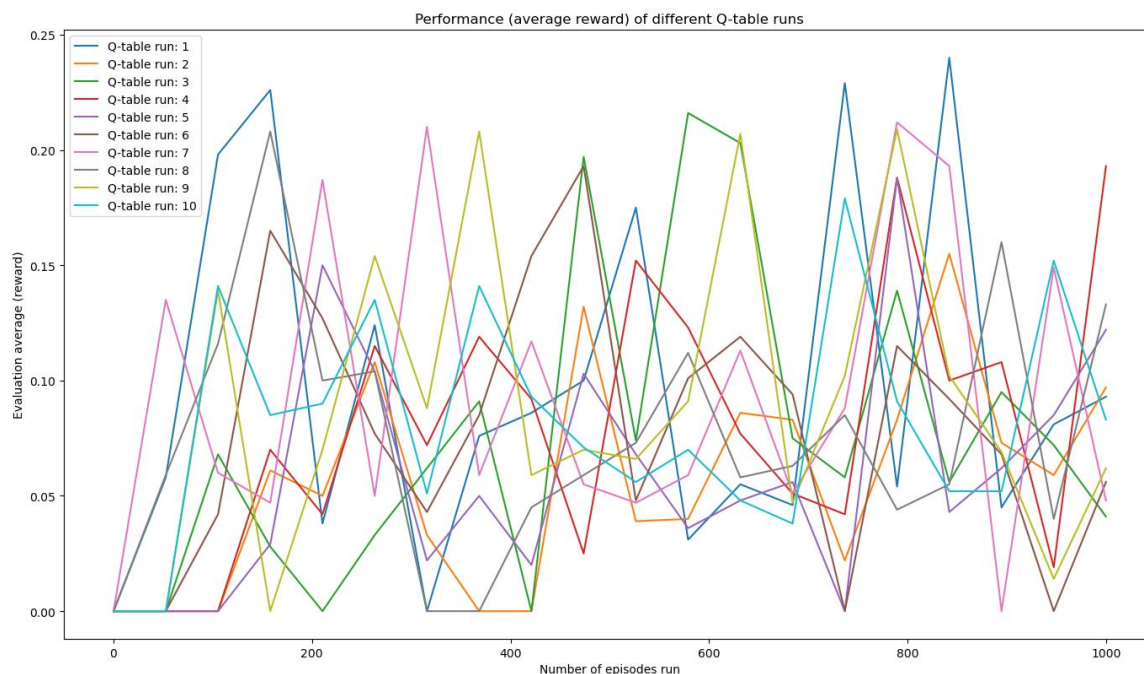
Following figures were obtained from the runs:



The deterministic run (`is_slippery=False`) finds the goal almost every single time. Especially when the number of episodes is increased.



The non-deterministic run (`is_slippery = True`) gives worse results as expected. The average reward does not seem to get noticeably better with more episodes run.



The non-deterministic run (`is_slippery = True`) with Q-table update algorithm improved. I think the result is marginally better than with old update

algorithm. I tried to play with alpha and gamma but could not find meaningful improvement with those tweaks.