

$\epsilon$ -greedy optimal policy - last 500 steps

gamma

0.95

