**Scenario 1 Comparison**

After comparing two ε-greedy exploration strategies: ε = 0.1 and ε = 0.3. The agent with ε = 0.3 learned faster initially due to higher exploration but exhibited more variability across runs. The ε = 0.1 agent learned more slowly but showed more stable convergence in the later episodes. Over 500 episodes, both agents eventually discovered the optimal path, but ε = 0.1 consistently produced smoother and higher average rewards near convergence. The plot illustrates this trend, where the ε = 0.1 line gradually overtakes ε = 0.3, indicating more refined and consistent policy development with reduced randomness.



