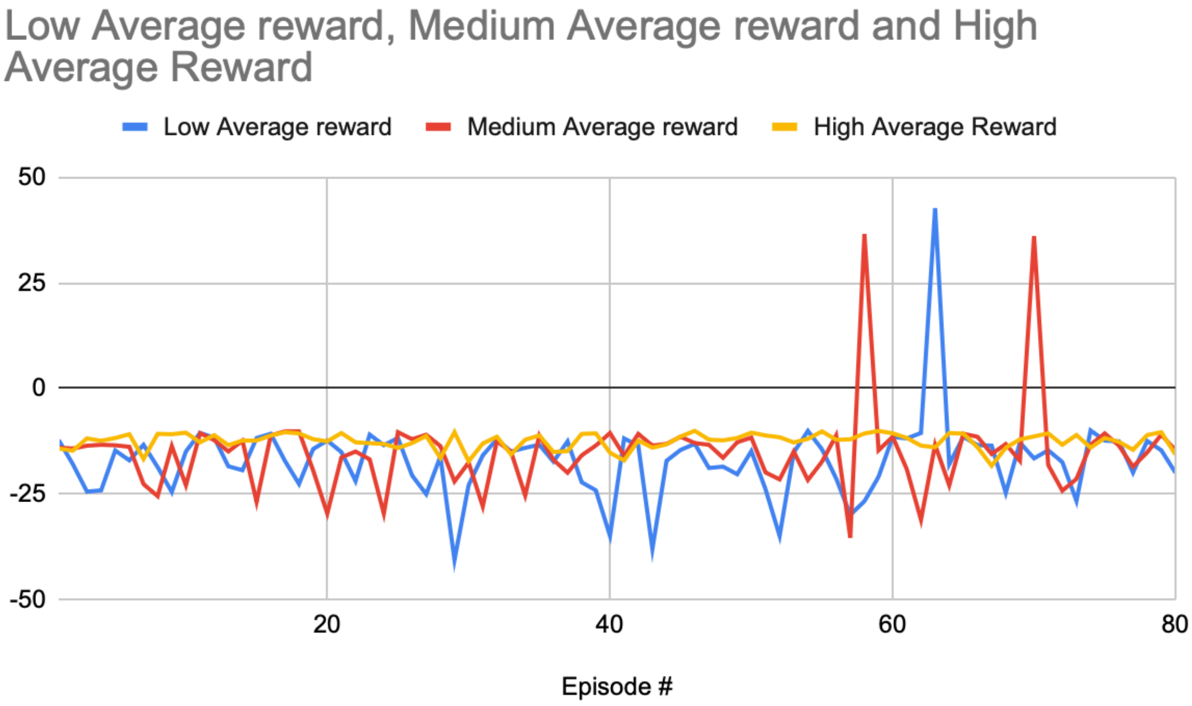
Sean Gor
Assignment 5
Dr. Forouraghi
Purpose: To experiment with different values of epsilon decay/gamma (discount factor values)

Part 1:

Graph:



The plot compares the agent's training performance under three different epsilon–decay rates: **Low decay (slow decay), Medium decay, and High decay (fast decay)**. Each curve shows the average reward obtained per episode over 80 training episodes.

Low Decay (value used = 50):

Low decay keeps ε high for a long time, meaning the agent continues to explore extensively before committing to any particular strategy. In the plot, the low-decay curve is noisy early on and improves slowly. However, this long exploration should, in theory, give the best chance of eventually discovering an optimal path. The occasional positive spikes indicate episodes where the agent reached the goal or significantly improved its route.

Medium Decay (value used = 200):

Medium decay provides a balanced trade-off between exploration and exploitation. Its curve is less noisy than low decay and tends to stabilize sooner. It gradually improves as the agent learns, and its rewards hover in a more consistent range. This decay rate often produces the most stable learning behavior, which matches the assignment description.

High Decay (value used = 600):

High decay reduces ε very quickly, so the agent stops exploring early and begins exploiting suboptimal routes it found at the start. The high-decay curve stabilizes the fastest, but it has the lowest long-term performance. This demonstrates the risk of collapsing to a suboptimal policy due to insufficient exploration.

Conclusion:

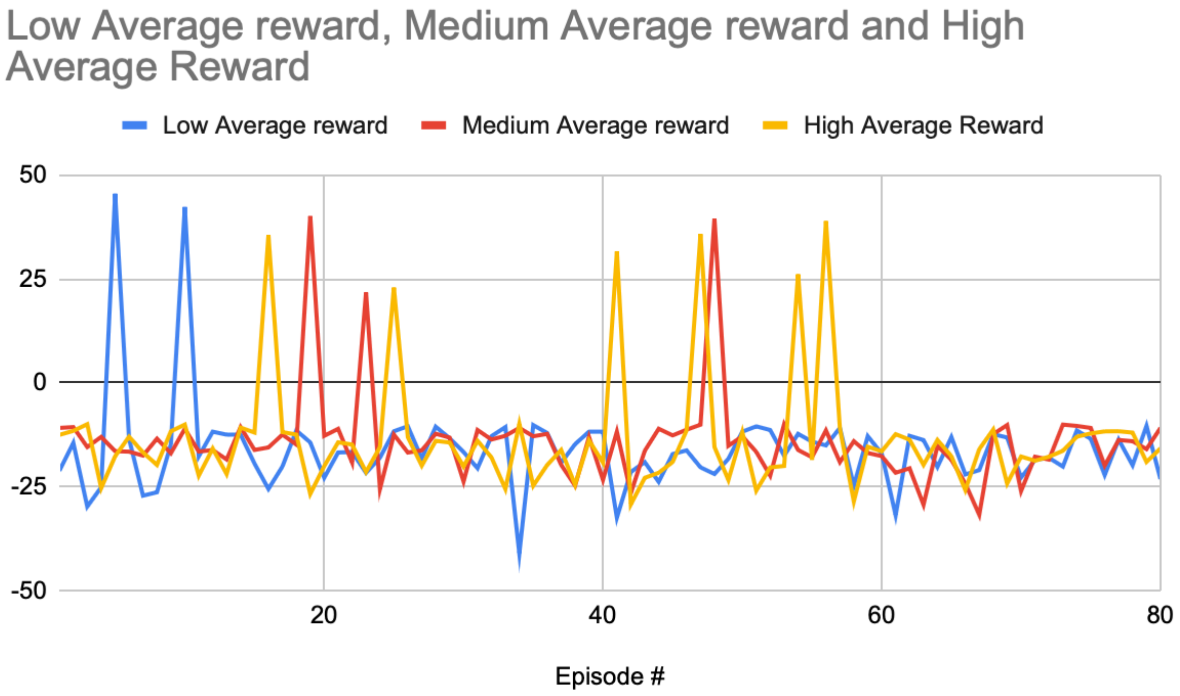
Overall, the results match the expected outcomes:

- **High decay** learns fast but often settles for a worse long-term policy.
- **Medium decay** is balanced and produces stable learning.
- **Low decay** is the slowest to improve but has the highest chance of discovering the best route.

Part 2:

Low Gamma (0.5) — Short-Sighted Behavior

Though not required, I also drew a graph to represent the relation between the different values of gamma, similar to the values of epsilon decay tested in part 1. I used the value 50 (the small one) for the value of the epsilon decay when testing the different values of gamma.



With Gamma = 0.5, the agent heavily discounts future rewards, meaning it focuses almost entirely on immediate penalties and short-term outcomes.

Observations:

- The reward curve was highly unstable, with large negative dips from frequent crashes.
- There were some occasional large positive spikes, meaning the agent sometimes reached the goal by chance.
- Overall, the performance was inconsistent and noisy.

This matches the assignment's explanation that a low-Gamma agent is *short-sighted*, may prefer avoiding immediate penalties over pursuing the long-term goal, and often ends up taking longer-than-optimal paths or getting stuck.

Medium Gamma (0.8) — Balanced Performance

At Gamma = 0.8, the agent begins valuing future rewards more, but not overwhelmingly so.

Observations:

- The reward curve became smoother and more stable than low Gamma.
- Extreme spikes were reduced, and the agent showed more consistent improvement.
- The performance represented a balance between short-term and long-term decision-making.

This aligns with the assignment's description of balanced behavior, where the agent neither overreacts to immediate penalties nor relies too heavily on distant rewards.

High Gamma (0.99) — Farsighted and Goal-Focused

Using Gamma = 0.99, the agent treats future rewards almost as strongly as immediate ones. Reaching the goal becomes a major priority.

Observations:

- The reward curve was the smoothest of all three settings.
- The agent achieved more frequent and higher positive reward spikes, indicating more successful episodes of reaching the goal.
- Fewer large negative drops appeared, showing more stable and optimal path-finding.

This matches the assignment's expected outcome: a high-Gamma agent is *farsighted*, prioritizes the large goal reward, and is most likely to find the shortest and most efficient path.

Overall Conclusion:

Across both experiments, the agent's performance aligned with the theoretical expectations from reinforcement learning. In Part 1, adjusting the epsilon-decay rate controlled how quickly the agent shifted from exploration to exploitation: high decay settled quickly but suboptimally, medium decay was balanced, and low decay explored longest and occasionally discovered better routes. In Part 2, varying the discount factor Gamma changed the agent's planning horizon: low Gamma produced short-sighted and unstable behavior, medium Gamma balanced immediate and future rewards, and high Gamma resulted in the most consistent and efficient learning. Overall, both experiments show how exploration strategy (epsilon decay) and planning horizon (gamma) jointly affect the stability, path quality, and long-term performance of a DQN-based agent.