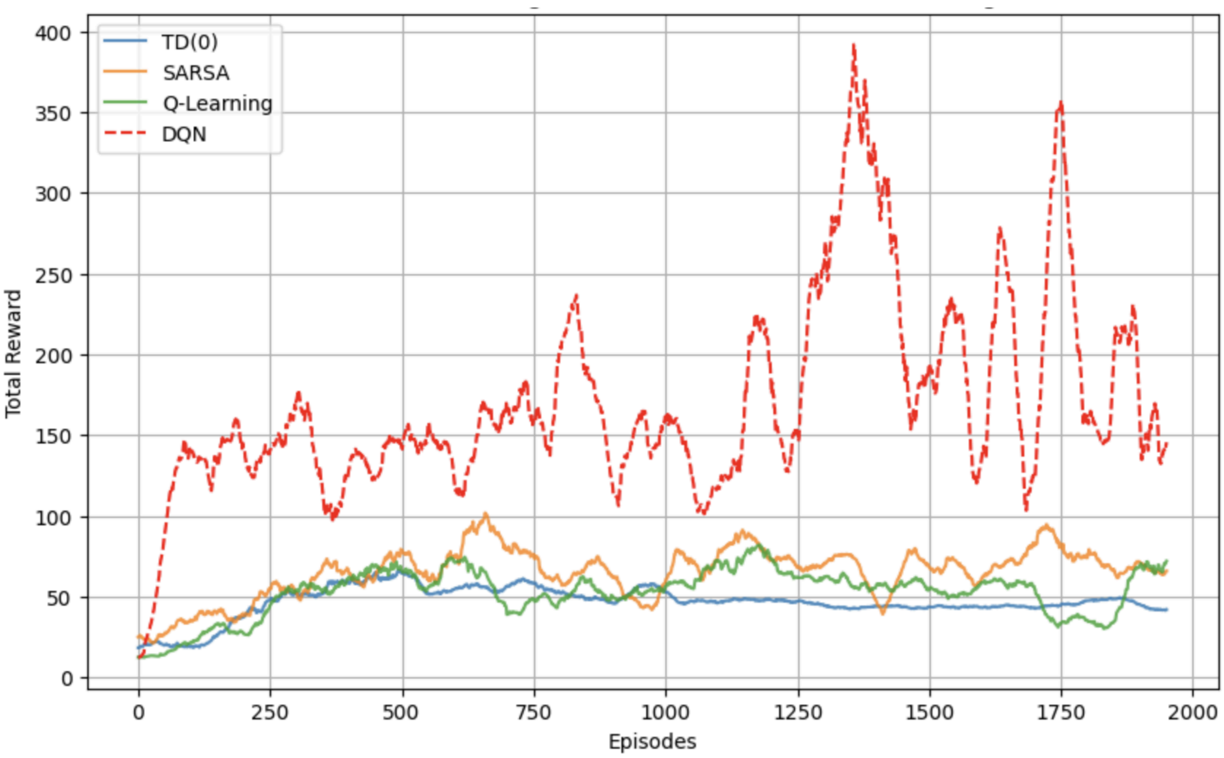
Question 2: Analysis of Experimental Results

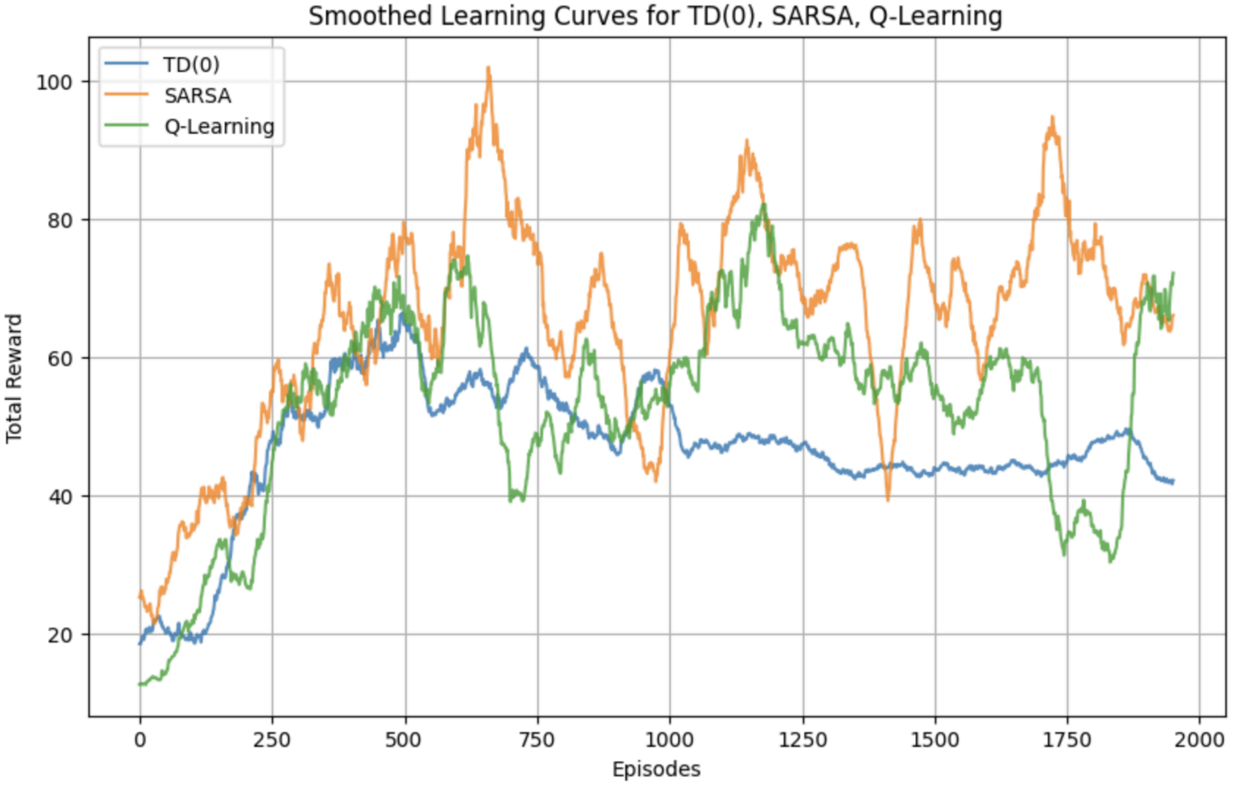
Write down what you have learned with your experimental results from Question 1. Analyze the performance of different algorithms in Question 1 by comparing the similarities and differences of these algorithms. Demonstrate your understanding and your insights.

I ran this experimentation over 30 times with different parameters values for alpha, gamma, epsilon and decay rate and what I have learned is that regardless of how I tweak them, DQN always ends up outperforming the TD(0), Sarsa, and Q learning. However, I noticed that different parameters affect the performance of TD(0), Sarsa, and Q learning differently. In addition, the **bins control the level of state discretization** in Q-Table-based learning methods and this also directly affects the learning of each algorithm. Some parameters and number of bin tend to be favorable for some of these algorithms depending on how you implement it.

I had some tests where Q learning tends to learn and converge better than Sarsa or the other way around with TD(0) included. So, after enough experimentation, I decided to use these parameters (**Alpha= 0.3, Gamma=0.9, Epsilon=0.3, Decay_rate=0.997**) with **number of bin** (**Bin =40**) and **2000 episodes** to ensure a fair comparison within the same figure (Note that DQN does not have alpha).

Alpha= 0.3, Gamma=0.9, Epsilon=0.3, Decay_rate=0.997) with **number of bin** (**Bin =40**) and **2000 episodes**





The learning curves above for **TD(0), SARSA, Q-Learning, and DQN** provide insights into their performance differences in solving the reinforcement learning task based on Cart Poole.

Key observations from the experiments based on the graph are:

1. SARSA Outperforms Q-Learning

- SARSA (orange) achieves higher rewards on average compared to Q-Learning (green).
- This result is kind of unexpected, as Q-Learning, being off-policy, usually converges to a more optimal policy. However, SARSA's policy-dependent updates led to more stable learning and better overall performance in this case as presented by the graph
- A possible explanation is that Q-Learning's update which maximize over estimated future rewards, may have led to more high-variance learning which makes it struggle to converge effectively.

2. Q-Learning vs. TD(0): Q-Learning Performs Better

- Q-Learning outperforms TD(0) consistently.
- This goes along with our expectations as Q-Learning actively learns from the best possible future rewards while TD(0) relies solely on immediate bootstrapped estimate as shown in these graph
- TD(0)'s weaker performance suggests that learning purely from temporal difference updates without off-policy corrections which is indeed not sufficient for effective learning in this environment.

3. DQN Outperforms All Algorithms

- DQN (red dashed) does significantly outperforms all other methods by achieving the highest rewards.
- This goes along with expectations since DQN utilizes a neural network for function approximation, experience replay, and target networks which leading to more efficient learning.
- The steep upward trend in rewards suggests that DQN does successfully generalize from experiences which allow it to achieve near-optimal performance over time.

Conclusion:

- The expected trend of improvement (SARSA > Q-Learning > TD(0)) holds true in this experiment, with the surprising exception that SARSA outperformed Q-Learning.
- DQN is indeed the best-performing method as expected from a deep learning approach.

.