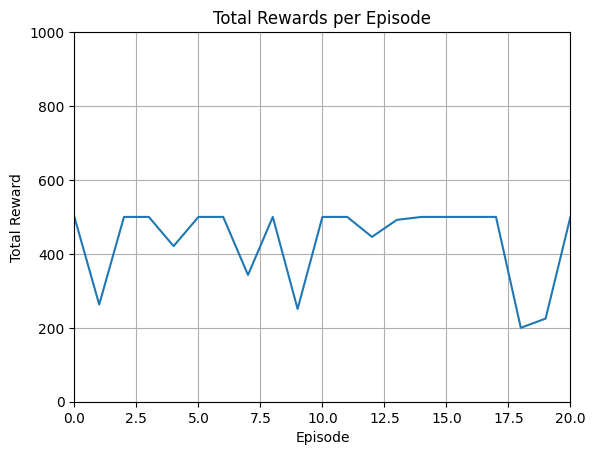
**DQN**

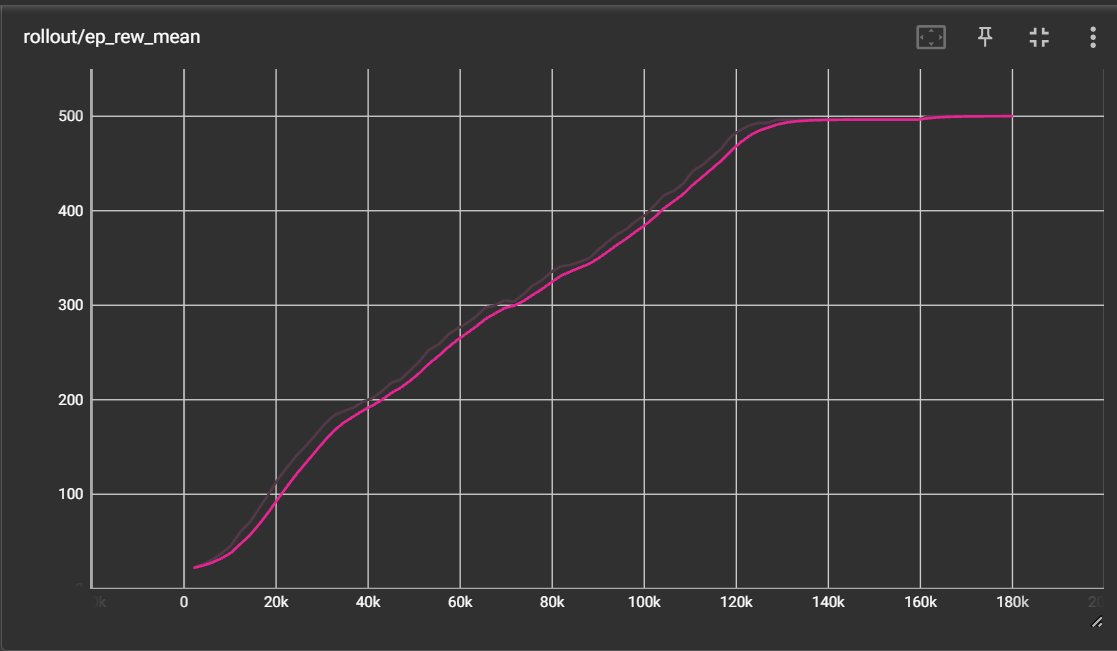
* I have used 3 layer neural network architecture with 24 neurons each in the hidden layer with ReLU activation function. Output layer has linear activation
* Training the Agent using BoltzmannPolicy(), ADAM Optimizer with learning rate of 0.001 and mean absolute error metric for evaluation
* Trained the model over 100000 steps.
* Tested this agent on 50 episodes and rendered the environment for visualization.
* **Strengths:** 1. Works well on discrete data thus favorable for Cartpole (action space = 2)
* 2. Reuses past experiences to stabilize learning
* 3. Boltzmann policy (softmax), chooses actions with probabilities proportional to their Q-values. This encourages exploration in proportion to the value of each action, thus more balanced exploration compared to epsilon-greedy.
* **Weaknesses:** 1. DQN struggles in environments with continuous or large action spaces.
* 2. DQN can require many episodes to converge to a good policy complex environments.

**Performance of Model over 20 episodes**

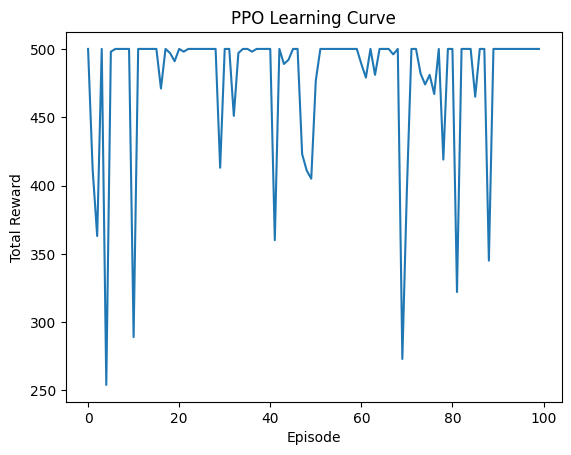


**PPO**

* I used stable\_baseline3 PPO with learning rate 0.0001, batch size 64 and trained over 100000 steps.
* **Strengths:** 1. Uses a clipped objective function to prevent large policy updates, resulting in more stable learning.
* 2. Works well in both continuous and discrete action spaces
* 3. Learns faster and more efficiently compared to state action based methods.
* **Weakness:** 1. Requires more resources because it uses both actor and critic networks, leading to higher computational costs
* 2. PPO can have slower performance per iteration due to more frequent updates.
* 3. PPO tends to exploit learned strategies quickly, which may reduce exploration compared to DQN with the Boltzmann policy.

 **Learning Curve**

**Performance of Model over 20 episodes**

****