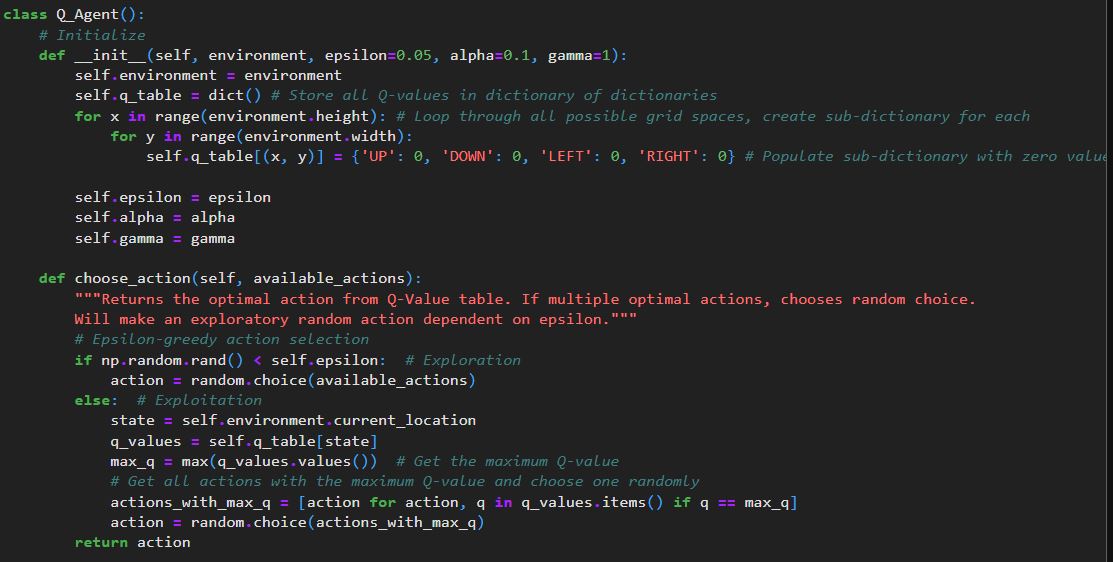
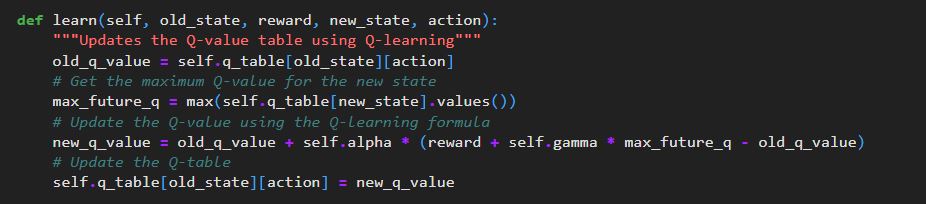
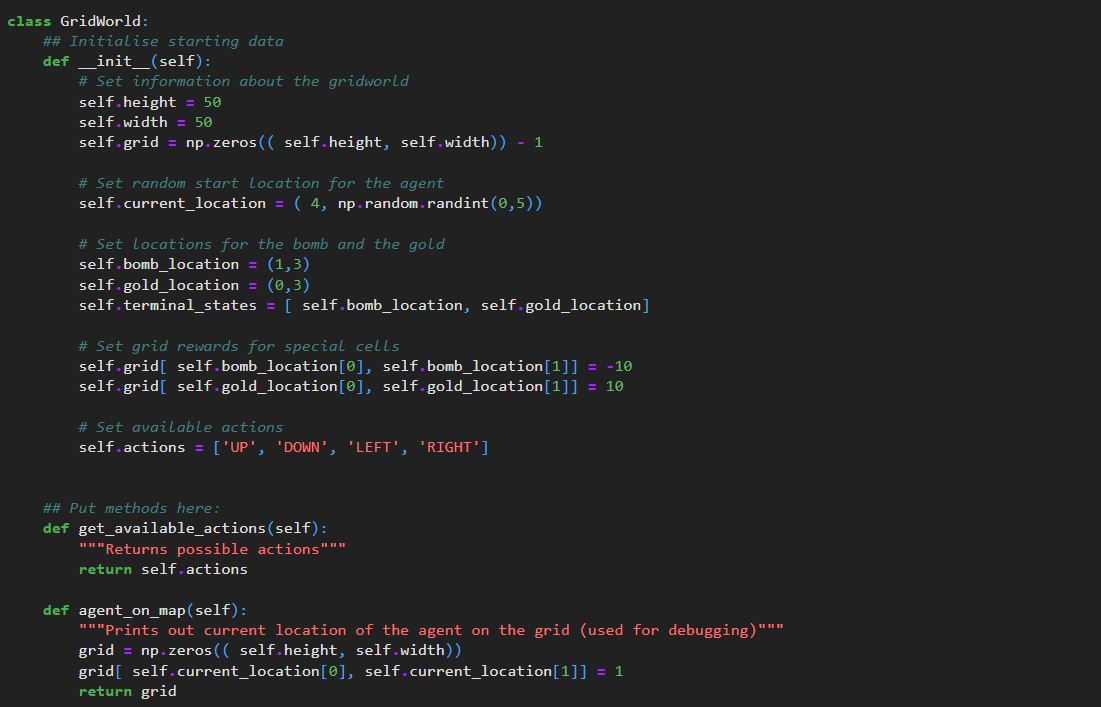
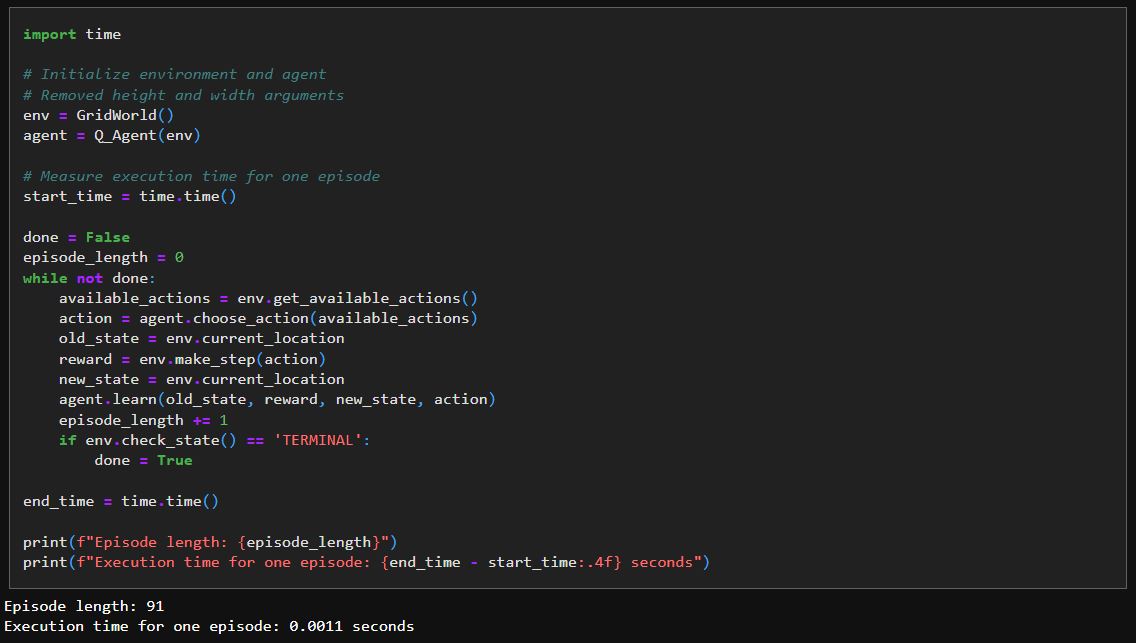
## Question 01

### GridWorld

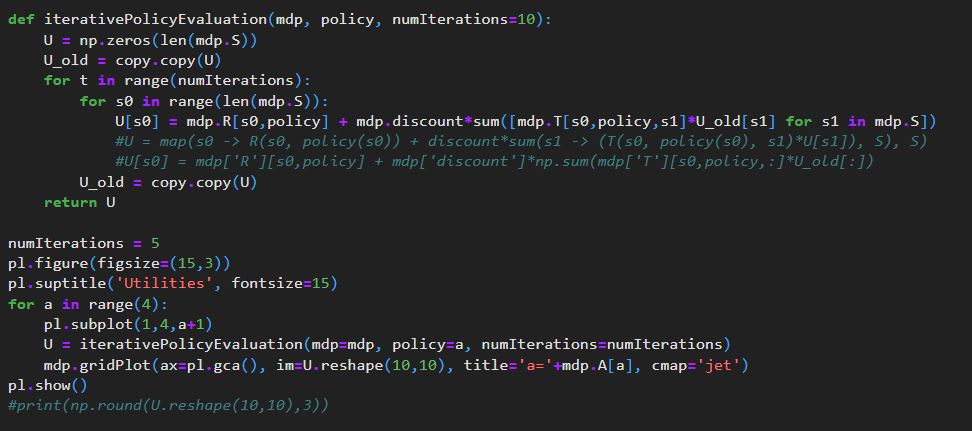


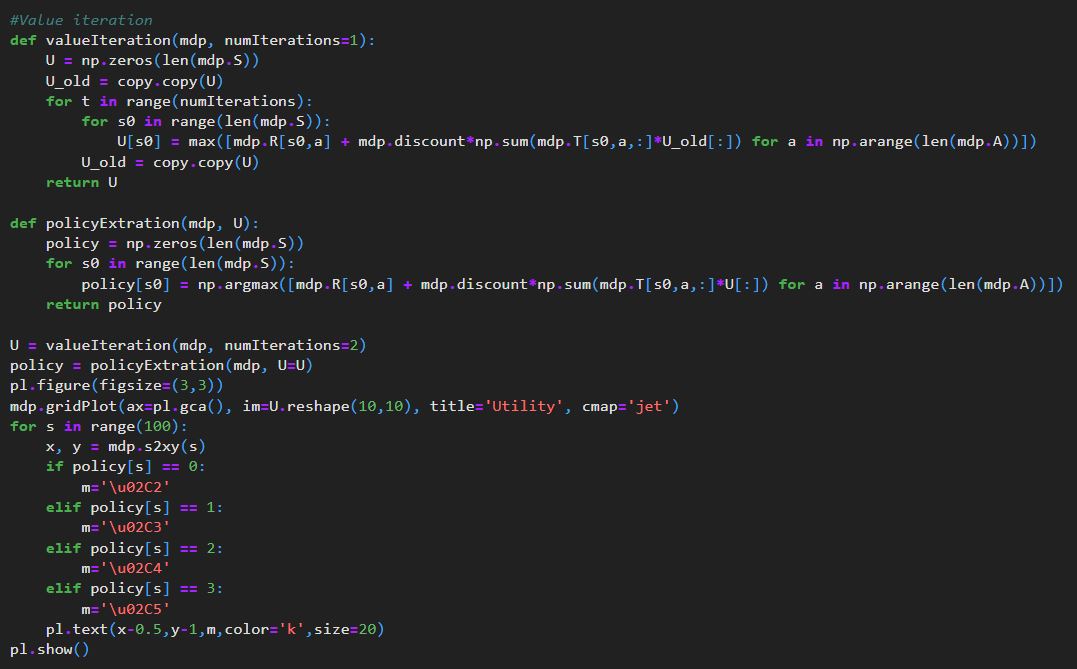


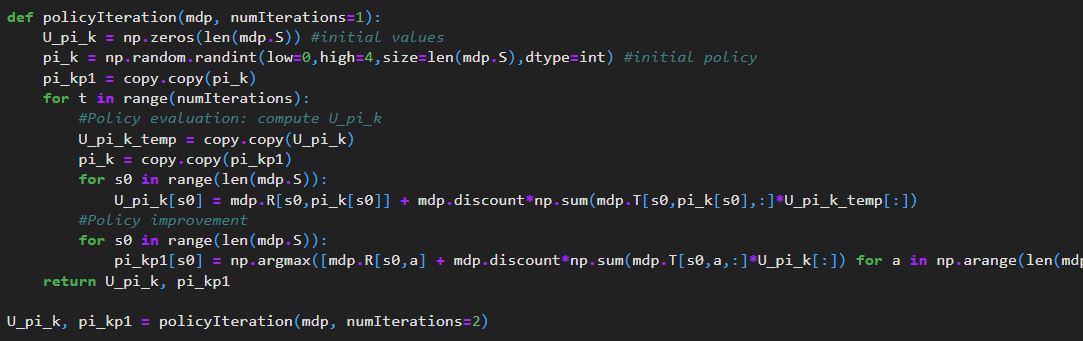




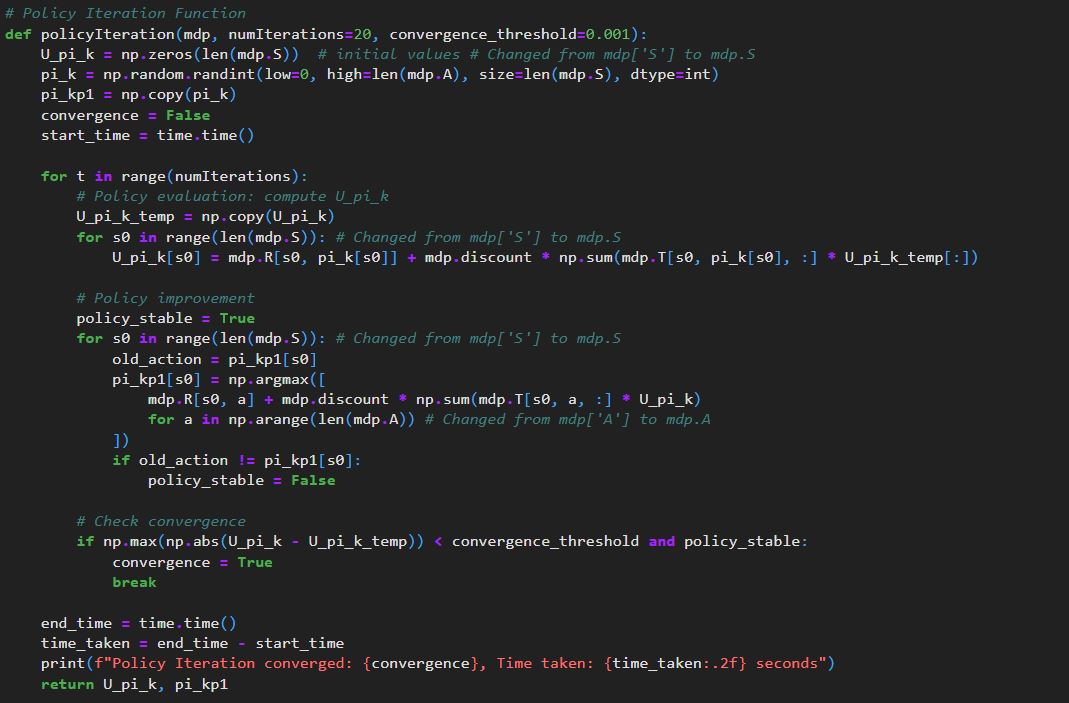
### Markov Decision Process

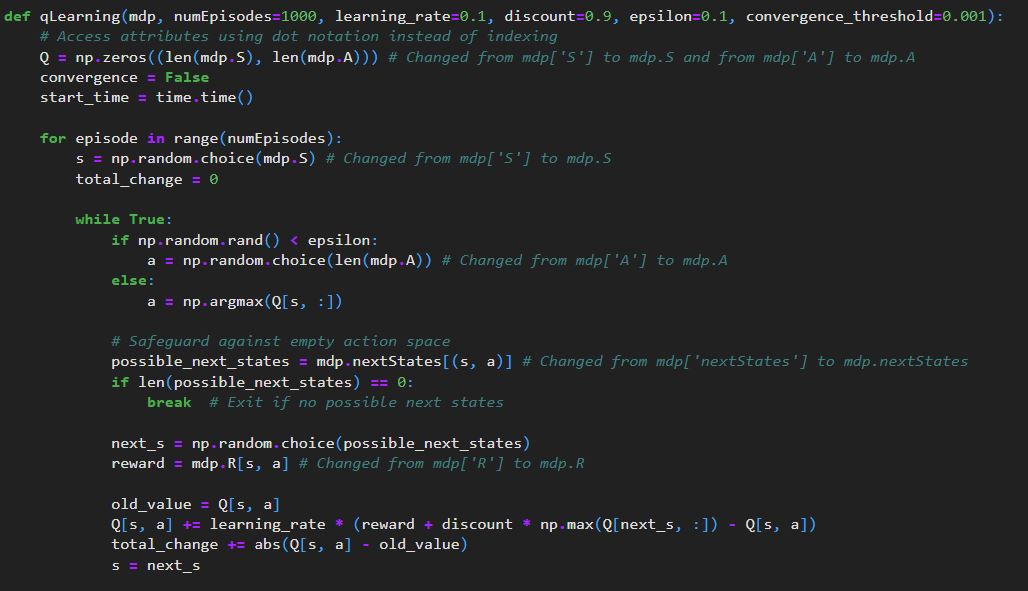


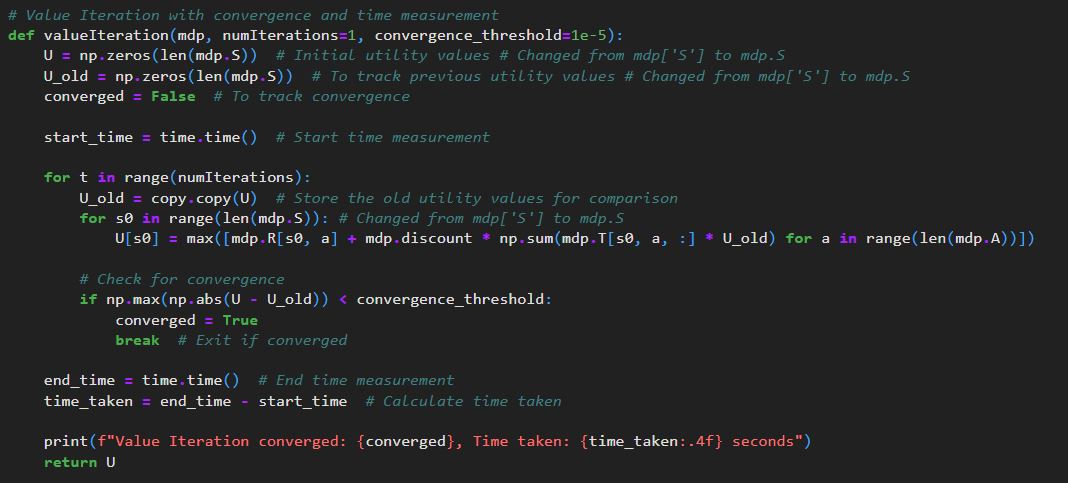


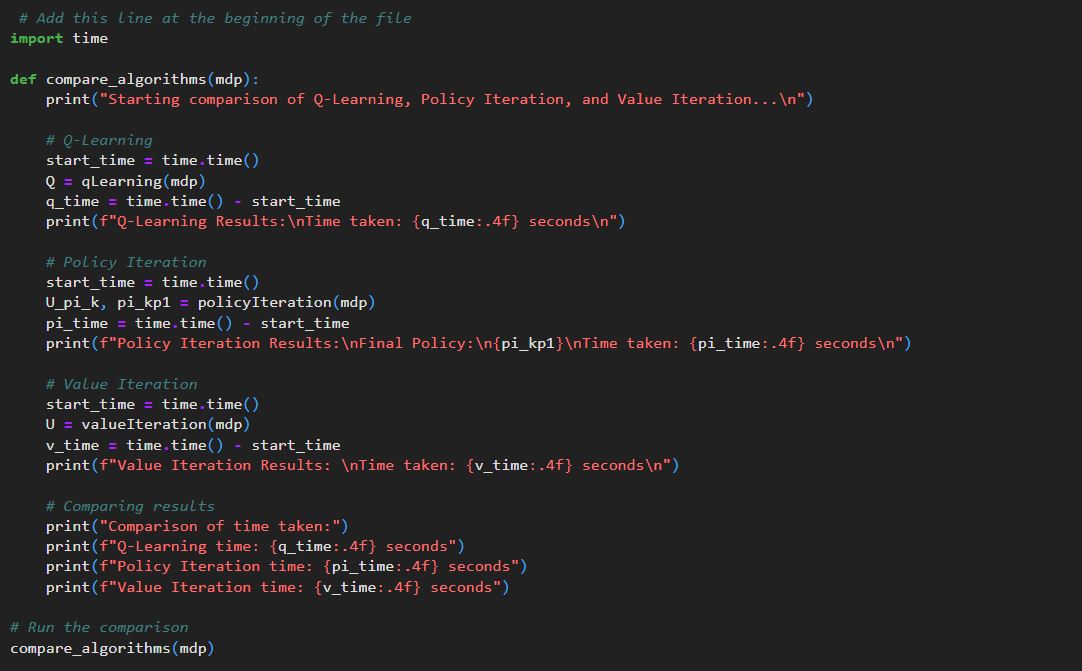


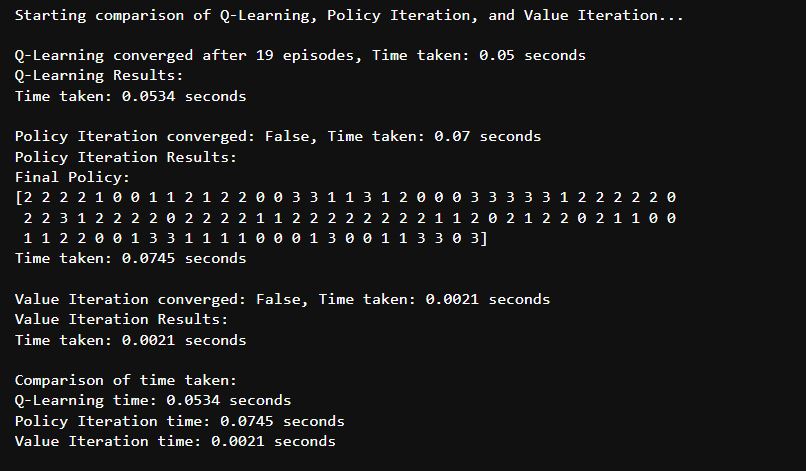
## Question 02:









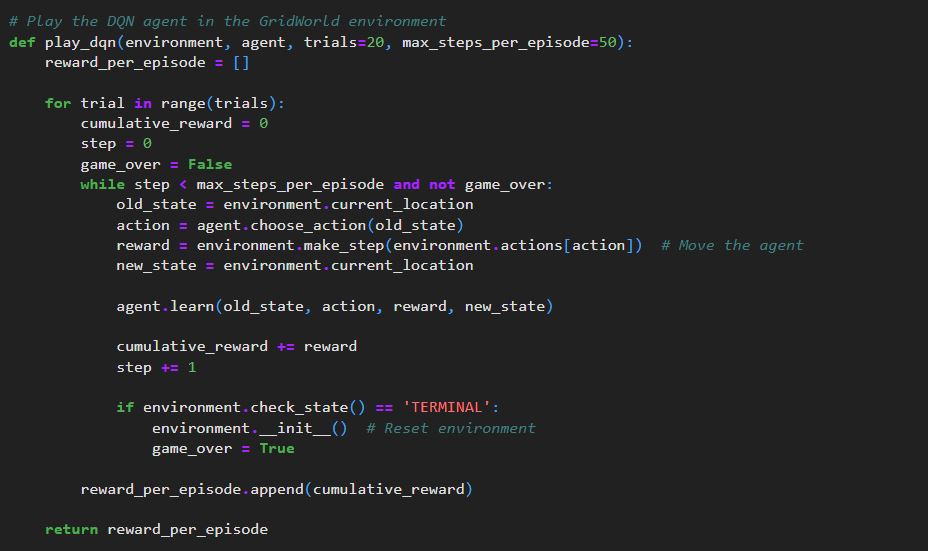


### Difference Between Model Based and Model Free Algorithms

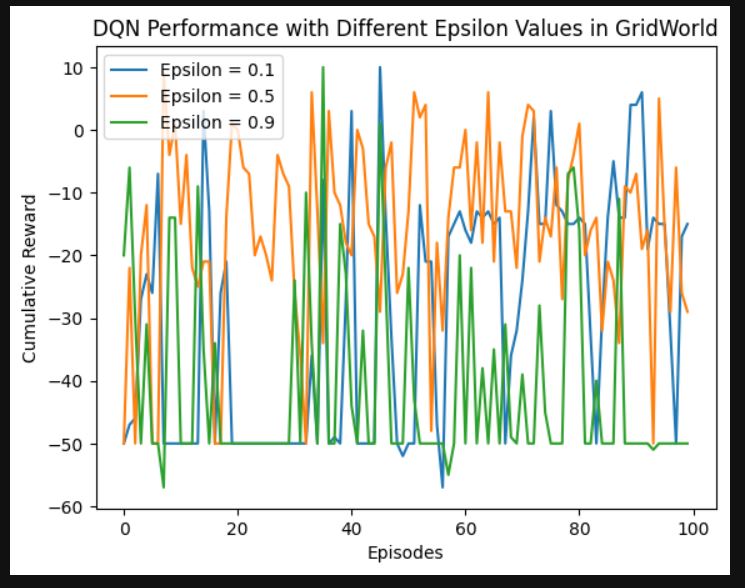
|  |  |  |
| --- | --- | --- |
| **Feature** | **Model-Free Algorithms** | **Model-Based Algorithms** |
| Learning Approach | Directly learns from interactions with the environment without a model. | Constructs an explicit model of the environment's dynamics to inform decisions. |
| Sample Efficiency | Generally, it requires more real-world interactions to learn an optimal policy. | More sample-efficient, as it can simulate experiences using the learned model. |
| Complexity | Simpler implementation does not require a model. | More complex due to the need to learn and maintain an accurate model of the environment. |
| Adaptability | Slower adapt to changes in the environment; relies on accumulated experience. | Faster adaptation with an accurate model, allowing for better handling of changes. |
| Environmental Utilization | Treats the environment as a black box; no internal model. | Actively builds and refines a model of the environment to predict outcomes and plan actions. |

## Question 03:

## 



## 



The efficacy of the Deep Q-Network (DQN) was assessed through the implementation of reinforcement learning via three different types of epsilon-greedy strategies (0.1, 0.5, 0.9 all). The findings provide evidence that alteration in the epsilon value impacts how the agent behaves as well as how well it performs. Thus, when epsilon = 0.1, very poor performance was observed in the model, of course with an equally dominant concentration of very extreme negative values, particularly around -50.0, which may be indicative of suboptimal exploration of the environment. On the other hand, at epsilon = 0.5, this improved even more as higher values of reward (up to 7.0) were recorded coupled with lower incidence of the most extreme penalties. This means that there was a better tradeoff between exploration and exploitation allowing the agent to find better states. However, when the value of epsilon increased to 0.9, the model registered a higher presence of even more extreme negative values like -58.0 further indicating that too much exploration results in the worst-case scenario which is excessive penalty in decision making. Overall, the findings demonstrate the need to carefully choose the epsilon parameter in order to achieve optimum performance levels of the DQN, advocating that both over and underscoring exploration can be detrimental to the learning process.