Assignment 3 - RL Questions

1. **Define the following properties of the FishingDerbyRL MDP:**

State Space S: ***The total number of states of the environment (represented by a grid world).***
The total number of states is the subdivisions squared. In this case, it is 10^2 = 100.

Action Space A: ***All possible actions carried out by the diver.***

The possible actions carried out by the diver are: “Left”, “Right”, “Up”, “Down”, and “Stay”

2. **Define and test at least one interval for each of them accordingly, to achieve the following desirable policies:**

Not improving/learning:
Low learning (Alpha α) results in a low variance, thus not improving/learning.

High variance but fast learning:
High learning (Alpha α) results in a high variance, thus fast learning.

Low variance and high long-term return:
Low learning (Alpha α) results in a low variance, and a high discount (Gamma γ) results in a more long-term focused agent.

High variance and high long-term return:
High learning (Alpha α) results in a high variance, and a high discount (Gamma γ) results in a more long-term focused agent.

3. **However, if the reward structure of an MDP is simple enough, the optimal policy degenerates in a simple heuristic. Given the 3_2_3.yml reward structure and initial position of jellyfish/kingfish/diver, what is the value of the long-term return of the optimal policy?**

The long-term return policy is based on the diver catching the kingfish while avoiding jellyfish and minimizing steps. This means that the path the diver takes calculates the specific value of the long-term return, and the number of steps to the kingfish while avoiding jellyfish.