Assignment 1

1. Question: what is the data in this structure saying? Relate this to the course presentation of P.

   Example of the data:

   ```
   {0: {0: [(1.0, 0, 0.0, False)], 1: [(1.0, 4, 0.0, False)], 2:
   [(1.0, 1, 0.0, False)], 3: [(1.0, 0, 0.0, False)]}, 1: {0:
   [(1.0, 0, 0.0, False)], 1: [(1.0, 5, 0.0, True)], 2: [(1.0, 2,
   0.0, False)], 3: [(1.0, 1, 0.0, False)]}, 2: {0: [(1.0, 1,
   0.0, False)], 1: [(1.0, 6, 0.0, False)], 2: [(1.0, 3, 0.0,
   False)], 3: [(1.0, 2, 0.0, False)]}, 3: {0: [(1.0, 2, 0.0,
   False)], 1: [(1.0, 7, 0.0, True)], 2: [(1.0, 3, 0.0, False)],
   3: [(1.0, 3, 0.0, False)]},
   ```

   Answer: P is the probability of action and state or say the dynamics information for each state-action pair when computing the value function. Each key in the outer dictionary represents a state in the environment, and the inner dictionary contains information about the possible actions and their outcomes from that state.

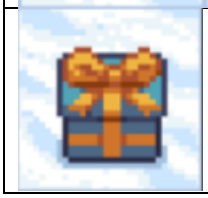
   For example, let's consider the first entry 0: {0: [(1.0, 0, 0.0, False)]}. It states that if the agent is in state 0 and takes action 0, there is a 100% probability (1.0) of staying in the same state (next state = 0), receiving a reward of 0.0, and the episode does not terminate (done = False).

2. Question: draw a table indicating the correspondence between the action you input (a number) and the logic action performed.

Input	Action performed
0	Move left
1	Move down
2	Move right
3	Move up
4	Exit()

3. Question: draw a table that illustrates what the symbols on the render image mean?

Symbols	Meaning
	Start of the game
	Agent

	Trap (part of environment) End of the game
	Gift/Rewards

4. Question: Explain what the objective of the agent is in this environment?
   Answer: The object is to guide the agent to obtain the rewards while skipping all the traps.

5. Question: How would you represent the agent's policy function and value function?
   The policy function will be a 1-dimensional array as **P[s]** to represent the policy function at state **s** or a 2-dimensional array as **P[s][a]**to represent the policy function at state **s** if action **a** is taken. Value function is simply v[s] to represent the value at state **s.**