For week 2, the lab interduces Markov decision process, which allows to formalize a sequential decision making for agent to take action to move to a state where it not just influences the immediate reward but also the state condition or value. The components of MDP are:

* States
* Models
* Possible actions
* Rewards
* Policy

The lab for week 2 contains a practical application of MDP using standard grid and negative grid, both grids have 3 rows and 4 columns, where each index is a state, actions that agent can take are: (left, right, up, down), state 0,0 is an initial state where the agent begins, state 2,2 is a terminal state with +1 rewards where state 1,2 has -1 which is also a terminal state. State 1,1 is a barrier state which the agent can’t take a decision to take an action that results to move to the barrier state. In the standard grid all other states other than states above have 0 rewards while the negative grid those states have -0.1 reward which is penalty, and the reason for that is to develop a procedure to take the optimum decisions that results to the optimal reward. However, in this lab, we only provide a policy, that we already specified in “ValuesAndAPolicy.ipynb file.”

This lab reminds me of a method I study in MSD688 class and in data structure class in my undergraduate degree where we compute the optimal path to a goal node (terminal state) from a tree and a graph. The concept of a negative reward and penalty to take an action is interesting because every decision is important and we want to reach the goal with minimum penalties possible, which will result to maximum reward