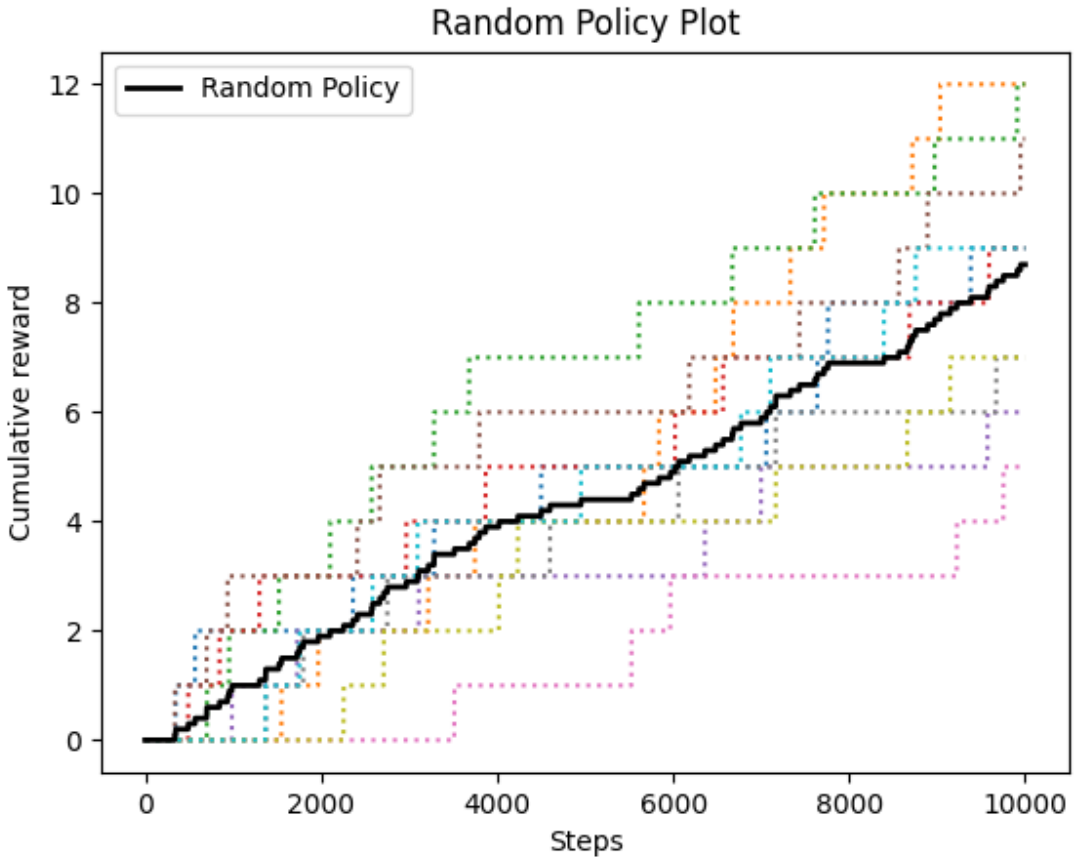
Reinforcement Learning Assignment - 1

Q3)How do you think this compares with your manual policy? (You do not have to run your manual policy for 104 steps!) What are some reasons for the difference in performance?

According to me, manual policy performs better than the random policy.
Reasons - (I am considering its Human who is taking manual actions)
1. As I will be controlling the agent manual, I can reach the goal state faster than random policy. As I am well aware of the goal_state location, I can easily route the agent towards the goal in minimum steps. But the random policy is more like exploration in a random direction for countless number of times.
2. Users can also eliminate the trial and error strategy, as we can observe the entire environment.
3. We also added some noise, 5% for each of the perpendicular actions, as the user will be incharge of taking actions, the agent will either go right/up directional, this is far better than 25% uniform probability of random policy.



Q4) Describe the strategy each policy uses, and why that leads to generally worse/better performance.

1. Manual Policy - agent's actions are controlled by the user via command inputs.
2. Random Policy - agent takes random action (uniform distribution)
3. Worse Policy - I removed the agent's Right command. Making sure that the agent never reaches the goal state. As we limit the actions taken by agents, it will never take a right turn. Hence , it leads to worse policy as it never reaches the goal state.
4. Better Policy - Since, the goal state is diagonally up and right side of the grid. I have given higher probability to these actions.
   Explanation - 70% probability to UP and Right , 35% each → UP / RIGHT
   15 % each → DOWN / LEFT. This will result in higher average cumulative rewards.