```
    end
    π'(s) ← arg max_{a∈A(s)} Q(s, a)
end
return π'
```

In the event that there is some state $s \in \mathcal{S}$ for which $\arg\max_{a \in \mathcal{A}(s)} Q(s, a)$ is not unique, there is some flexibility in how the improved policy π' is constructed.

In fact, as long as the policy π' satisfies for each $s \in \mathcal{S}$ and $a \in \mathcal{A}(s)$:

$\pi'(a|s) = 0$ if $a \notin \arg\max_{a' \in \mathcal{A}(s)} Q(s, a')$,

it is an improved policy. In other words, any policy that (for each state) assigns zero probability to the actions that do not maximize the action-value function estimate (for that state) is an improved policy. Feel free to play around with this in your implementation!

Please use the next concept to complete **Part 3: Policy Improvement** of `Dynamic_Programming.ipynb`. Remember to save your work!

If you'd like to reference the pseudocode while working on the notebook, you are encouraged to open **this sheet** in a new window.

Feel free to check your solution by looking at the corresponding section in `Dynamic_Programming_Solution.ipynb`.