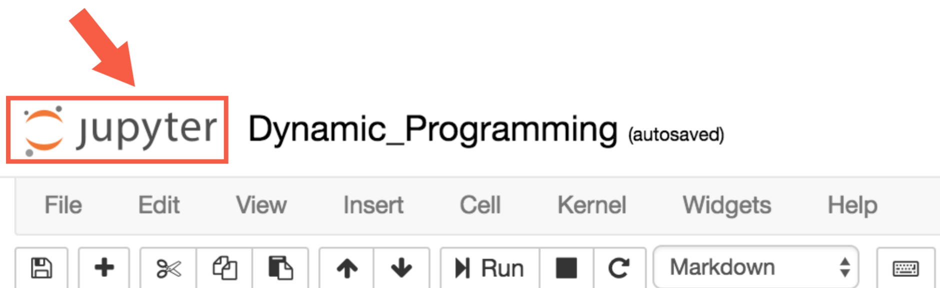
encouraged to open this sheet in a new window.

Feel free to check your solution by looking at the corresponding sections in `Dynamic_Programming_Solution.ipynb`. (*In order to access this file, you need only click on "jupyter" in the top left corner to return to the Notebook dashboard.*)



To find `Dynamic_Programming_Solution.ipynb`, return to the Notebook dashboard.

(Optional) Additional Note on the Convergence Conditions

To see intuitively *why* the conditions for convergence make sense, consider the case that neither of the conditions are satisfied, so:

- $\gamma = 1$, and
- there is some state $s \in \mathcal{S}$ where if the agent starts in that state, it will never encounter a terminal state if it follows policy π.

In this case,

- reward is not discounted, and
- an episode may never finish.

Then, it is possible that iterative policy evaluation will not converge, and this is because the state-value function may not be well-defined! To see this, note that in this case, calculating a state value could involve adding up an infinite number of (expected) rewards, where the sum may not converge.

In case it would help to see a concrete example, consider an MDP with:

- two states s_1 and s_2, where s_2 is a terminal state
- one action a (*Note: An MDP with only one action can also be referred to as a Markov*