Every-Visit Monte Carlo prediction, for estimating V

Input: a policy π **to be evaluated**

Initialize:

    $V(s) \in \mathbb{R}$**, arbitrarily, for all** $s \in S$

    $Returns(s) \leftarrow$ **an empty list, for all** $s \in S$

Loop forever (for each episode):

    Generate an episode following $\pi : S_0, A_0, R_1, S_1 \ldots, S_{T-1}, A_{T-1}, R_T$

    $G \leftarrow 0$

    Loop for each step of episode, $t = T-1, T-2, \ldots, 0$

        $G \leftarrow \gamma G + R_{t+1}$

        Append G **to** $Returns(S_t)$

        $V(S_t) \leftarrow$ **average**$(Returns(S_t))$