Question 3

$$V(s) = \frac{\sum_{t \in \mathcal{T}(s)} \rho_{t:T(t)-1} G_t}{\sum_{t \in \mathcal{T}(s)} \rho_{t:T(t)-1}}$$

Given that we are at state S_t and perform action A_t, probability of the action trajectory $S_{t+1}, A_{t+1}, \ldots S_T$ occuring under any policy π is

$$Pr\{S_{t+1}, A_{t+1}, \ldots, S_T \mid S_t, A_t, A_{t+1:T-1} \sim \pi\}$$

$$\Rightarrow \frac{1}{\pi(A_t, S_t)} \prod_{k=t}^{T-1} \pi(A_k|S_k) p(S_{k+1}|S_k, A_k)$$

$$\rho'_{t:T-1} = \left[\prod_{k=t}^{T-1} (\pi(A_k|S_k)/b(A_k|S_k))\right] \times \frac{1}{\pi(A_t|S_t)} \times b(A_t|S_t)$$

$$\Rightarrow \boxed{\rho'_{t:T-1} = \prod_{k=t+1}^{T-1} \frac{\pi(A_k|S_k)}{b(A_k|S_k)}}$$

$\mathcal{T}(s,a)$ → all time steps where state s is visited and action a is performed

$$Q(s,a) = \frac{\sum_{t \in \mathcal{T}(s,a)} \rho'_{t:T(t)-1} G_t}{\sum_{t \in \mathcal{T}(s,a)} \rho'_{t:T(t)-1}}$$