QUESTION: *Observe what you see with the agent's behavior as it takes random actions. Does the* ***smartcab*** *eventually make it to the destination? Are there any other interesting observations to note?*

ANSWER: *Yes, smartcab evntually make it to the destination. Yes, there is some observations that is interesting like city is open from all side. Initial position of smartcab and destination is random.*

QUESTION: *What states have you identified that are appropriate for modeling the* ***smartcab*** *and environment? Why do you believe each of these states to be appropriate for this problem?*

ANSWER: *I am identified a state type in which I am store traffic signal and oncoming, left traffic , right traffic. Format of my state is*

state = (trafficSignal, oncomingtraffic, lefttraffic, righttraffic,waypoint, deadline)

*So, states is (‘red’,None,None,None,right,5),(‘green’,None,None,None,left,0),(‘red’,left,None,None,forward,4),(‘red’,None,left,None,None,4),(‘red’,None,left,None,left,2) etc. So, there is total 2*4*4*4*4*total_deadlines states in the problem. Each of these state identified every condition of the smartcab like if there is red signal and traffic is oncoming from left and take right turn and way_point for the destination is right and deadline remaining is 6 then state (‘red’,None,’right’,None,’right’,6) identified this condition. Because state of smartcab only depends on traffic signal light,traffic,waypoint and deadline. Since, our states representation contain all these variable. So, each of these states are appropriate for this problem.*

QUESTION: *What changes do you notice in the agent's behavior when compared to the basic driving agent when random actions were always taken? Why is this behavior occurring?*

ANSWER: *Initially agent’s behaviour is same as basic driving agent when random actions were always taken but after sometime agent behaviour changes slowly slowly. When random actions is always taken agent gets the positive reward sometimes but after q-learning agent get more rewards then earlier. For True enforce_deadline there is very few chance to make it to the destination for basic driving agent but for q-learning agent chance to make it to the destination increase.*

All of this is happening due to Q-learner because we give appropriate state, action q-value for each action taken by a state from particular state. If this action give good result it is stored else another action is choosen. After some iteration we have the agent which have some good policy for each state. So, that’s why initially our agent does not perform optimally but after some time it perform optimal.

QUESTION: *Report the different values for the parameters tuned in your basic implementation of Q-Learning. For which set of parameters does the agent perform best? How well does the final driving agent perform?*

ANSWER: *Different values of parameter tuned :*

1. (gamma = 0.1 ,epsilon = 1/number_of_trials,alpha = 0.8)

Total Success Rate: 84

Total Reward: 1868.5

Total Penalty Ratio: 0.308833010961

Last 10 success rate: 10

Last 10 total Reward: 198.5

Last 10 Penalty Ratio: 0.109243697479

2. (gamma = 0.2 ,epsilon = 1/number_of_trials,alpha = 0.8)

Total Success Rate: 85

Total Reward: 1864.5

Total Penalty Ratio: 0.299738219895

Last 10 success rate: 10

Last 10 total Reward: 193.0

Last 10 Penalty Ratio: 0.0853658536585

3. (gamma = 0.4 ,epsilon = 1/number_of_trials,alpha = 0.8)

Total Success Rate: 84

Total Reward: 1852.0

Total Penalty Ratio: 0.320638820639

Last 10 success rate: 10

Last 10 total Reward: 242.5

Last 10 Penalty Ratio: 0.0909090909091

4. (gamma = 0.1 ,epsilon = 1/number_of_trials,alpha = 0.9)

Total Success Rate: 86

Total Reward: 1904.5

Total Penalty Ratio: 0.300852618758

Last 10 success rate: 10

Last 10 total Reward: 209.0

Last 10 Penalty Ratio: 0.140350877193

5. (gamma = 0.2 ,epsilon = 0.3,alpha = 0.9)

Total Success Rate: 69

Total Reward: 1696.0

Total Penalty Ratio: 0.392680514342

Last 10 success rate: 8

Last 10 total Reward: 204.5

Last 10 Penalty Ratio: 0.20245398773

As, from above almost all set of parameter shows equal performance except 5th one which success rate is minimum.But I am select 3rd set of parameter, for 3rd set i.e. ***(gamma = 0.4 ,epsilon = 1/number_of_trials,alpha = 0.8)*** *last 10 success rate is maximum 10 and total rewards is 242.5 and penalty ratio is almost lower than all others and almost zero i.e. 0.09.*

So, I am select the parameter set ***(gamma = 0.4 ,epsilon = 1/number_of_trials,alpha = 0.8)*** *for my agent. It's performance is quite well its reach destination almost 84 times from 100 trials and its penalty ratio for last 10 iterations are almost zero and success rate is 100% for last 10 iterations.*

QUESTION: *Does your agent get close to finding an optimal policy, i.e. reach the destination in the minimum possible time, and not incur any penalties? How would you describe an optimal policy for this problem?*

ANSWER: Yes, our agent get close to finding an optimal policy, it reaches the destination almost 100% for last iterations in minimum possible time and 84 times in total 100 trials. It's penalty ratio for last iterations are almost zero that is no penalties in last iterations. I would describe an optimal policy for this problem in which there is no penalities and reach the destination in minimum possible time.