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: The agent behavior appears to be completely random in this case. The agent is roaming in the environment without any sense of direction. Agent is accumulating negative reward because of breaking traffic rules and for causing accidents. The agent is also missing deadlines because of that.

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 have identified the following states, to track the state of a smart car:-

a) "OK" State: if the light is green, the vehicle will enter the "OK" state.
b) "Stop!" State: If the light is red the vehicle will enter the "Stop!" state.
c) "Deadline Missed!" If the time remaining in the deadline variable takes a negative value. The vehicle will enter the "Deadline missed!" state.

OPTIONAL: *How many states in total exist for the* ***smartcab*** *in this environment? Does this number seem reasonable given that the goal of Q-Learning is to learn and make informed decisions about each state? Why or why not?*

Answer: If we take the rough count values, the total number of states, which we will need to model this world, will be 48. This number doesn't seem reasonable since in order to learn a valid Q function by Q-learning we need to visit every state action pair almost infinite number of times, to model the true value of the Q function. In addition, in the real world we won't model all the intersection in a town as a state of the smart car.

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: The cars seems to be running in a more organized way compared to the way they were running previously. This behavior is happening because of the Q matrix, which is used to decide the next action, which decides the next action based on the current estimate of the long-term reward of the car.

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: I tuned the parameter value of gamme which is also known as the discount factor of the Q learning algorith. Currently, I am using gamma = 0.9.

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?*