### 1、In your report, mention what you see in the agent's behavior. Does it eventually make it to the target location?

Answer：It just move randomly like a headless chicken, sometimes it may luckily reach the target location but need many moves.

### 2、Justify why you picked these set of states, and how they model the agent and its environment

Answer：I pick 'next_waypoint' and 'light' as my state. 'next_waypoint' give a baisc direction to the target location and 'light' tell the agent whether take an action without punishment. I also try to take 'presence of cars' into consideration because it could give infomation to reduce punishment when take an action. But as a result it doesn't make a better result and take more time to exploration. As for the 'deadline', I think it's meanless. If the agent is trained successfully, it will always try to reach the destination as quickly as possible. The 'deadline' feature will cause confusion and make more punishment.

### 3、What changes do you notice in the agent's behavior

Answer：At first the agent still move like a headless chicken when it's not familiar with the environment. But quickly as the q table grows, it take less and less time to reach the target location

### 4、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?

Answer：I have tried many choice of learning rate, discount factor and exploration probability. I think the choice in my code is better compared to other choice. (alpha = 0.2, gamma = 0.8, epsilon = 0.8). It reach the destination quickly when familiar with the environment.

| learning rate | discount factor | exploration probability | average step for 10 trials | success trials | average rewards |
|---|---|---|---|---|---|
| 0.1 | 0.8 | 0.8 | 23.9 | 8 | 16.6 |
| 0.2 | 0.8 | 0.8 | 21.1 | 9 | 17.2 |
| 0.3 | 0.8 | 0.8 | 48.3 | 3 | 9.6 |
| 0.2 | 0.7 | 0.8 | 42.8 | 4 | 10.8 |
| 0.2 | 0.9 | 0.8 | 27.4 | 6 | 13.7 |
| 0.2 | 0.8 | 0.7 | 25.6 | 7 | 15.1 |
| 0.2 | 0.8 | 0.9 | 29.3 | 5 | 12.2 |

I think my agent is performing sub-optimally. Sometimes it incur penaltied because as I mentioned before, I ignore 'presence of cars' in my state. This make the agent not see cars around and sometimes not follow the rules of the road. And because my learning rate and exploration probability are static, this make my agent not always try to go the shortest path to the final destination, maybe I should change the learning rate and exploration probability as my agent become familiar with the environment.