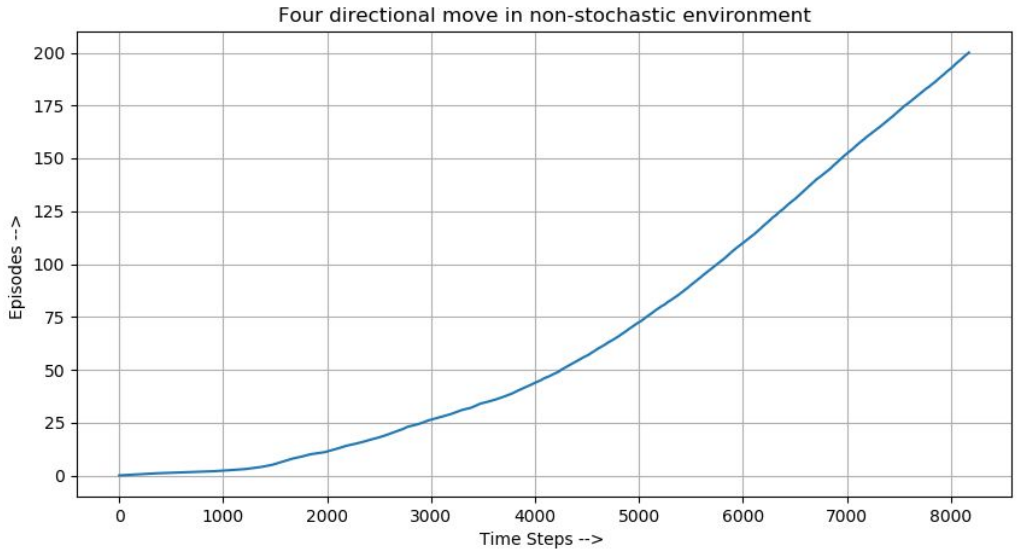
Assignment 4 By **Rupesh** [160050042]

Some design choices

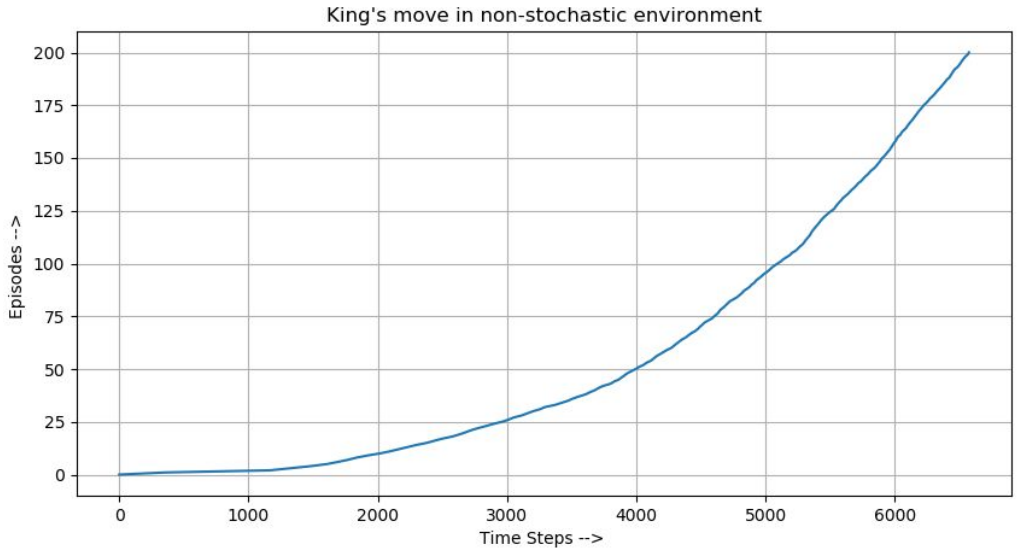
- When through the move the agent exist a wall, we move it back to the boundary for any axis it exists.
- The maze used is 10*7 same from the one in the book.
- X,y axis with origin at bottom left corner.
- Task repeated for 10 seeds and average value taken for every episode.

Task 1



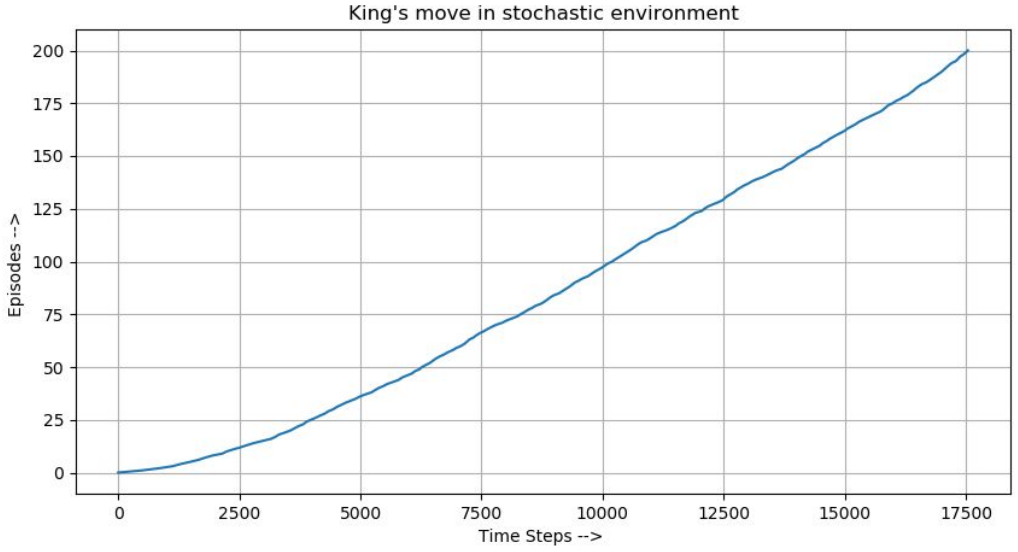
As the learning of Q happens, the better the agent becomes. The action picking is done E-greedy with no decay as described in the books. After learning, it takes around ~20 episode to reach goal and that achieves the constant slope.

Task 2



Making it do King's move i.e. 8 actions in all vertices or edge sharing square. This gives the king chance to reach early and hence the number of timesteps is lesser in this method to achieve the same goal early.

Task 3



Making the environment stochastic just makes learning difficult and environment unpredictable directly resulting in higher number of timesteps. This is because, even after finding the right direction to travel, you can never be sure of the wind and that stochasticity is tough to encounter by a simple Q matrix of the environment.