Experiment by Experiment Analysis =>

1. Start state => [0,4] ; Wind => False ; P_transition = 1; Policy = Softmax:

Both SARSA and Q-learning manages to find the best possible path while avoiding all the bad states.

2. Start state => [0,4] ; Wind => False ; P_transition = 1; Policy =E-greedy :

Q-learning : Manages to find the best possible path avoiding all the bad and restart states.

SARSA : SARSA manages to find the nearest path to the nearest goal state but it doesn't avoid all the bad states.

3. Start state => [3,6] ; Wind => False ; P_transition = 1; Policy =Softmax :

Q-learning and SARSA algorithms find different goal states along different paths. Both of them are almost equivalent in their reward and number of steps to the goal. Both of these algorithms manage to avoid all the bad and restart states and get the best possible reward.

4. Start state => [3,6] ; Wind => False ; P_transition = 1; Policy =E-greedy :

Q-learning : Manages to find the best possible path avoiding all the bad and restart states.

SARSA : SARSA seems like it's equally exploring both of the goal states that are at equal steps from the start state.

5. Start state => [0,4] ; Wind => False ; P_transition = 0.7; Policy = Softmax :

From the state visit heatmap we can see that because of the p_transition = 0.7, both SARSA and Q-learning algorithms explore different goal states along different paths.

Q - learning => Q-learning still manages to find the best possible paths.
SARSA => SARSA goes for the farthest but the safest possible route to the safest possible goal state.

6. Start state => [0,4] ; Wind => False ; P_transition = 0.7; Policy = E-greedy :

Q-learning : Q-learning manages to find the path which reaches the nearest goal state.
SARSA : SARSA goes for the farthest but the safest possible route to the safest possible goal state .

7. Start state => [3,6] ; Wind => False ; P_transition = 0.7; Policy = Softmax/E-greedy :

Q-learning : Q-learning manages to find the best possible path to the best possible goal state.

SARSA : SARSA explores a lot of longer paths and manages to reach two different goal states along different paths.

8. Start state => [0,4] ; Wind => True ; P_transition = 1 ; Policy = Softmax/E-greedy:

Both Q-learning and SARSA algorithms follow similar or (same) paths to the same goal states.

9. Start state => [3,6] ; Wind => True ; P_transition = 1 ; Policy = Softmax :

Both Q-learning and SARSA algorithms follow similar or (same) paths to the same goal states.

10. Start state => [3,6] ; Wind => True ; P_transition = 1 ; Policy = E-greedy :

Even though Q-learning and SARSA algorithms follow similar or (same) paths to the same goal states , there seems to be a major difference.

Q-learning algorithm's convergence to optimal steps and average mean reward is more Stable than SARSA algorithm. SARSA seems lot more unstable in its final convergence than Q-learning.