ME19B79 and ME19B177

Data Analysis and Plots

Code

In [1]:
```
filename = "CS6700_PA1_Data.tar.xz"
```

In [2]:
```
!ls -lh {filename} || gdown --fuzzy 'https://drive.google.com/file/d/1_jP8Ec6IpMdB10aJde
```

```
-rw-r--r-- 1 suraj suraj 13M Feb 24 22:54 CS6700_PA1_Data.tar.xz
```

In [3]:
```
!ls data > /dev/null || tar xf {filename}
```

In [4]:
```
!ls -1 data | wc -l
```

```
32
```

In [5]:
```
!ls -1 data/AA_qlearning_False_0_4_1_EpsilonGreedy_AA | wc -l
```

```
125
```

We have data from 32 different configurations.
Each configuration should have data from 125 different hyperparameter combinations.

In [6]:
```
import numpy as np

import matplotlib.pyplot as plt
import seaborn as sns
sns.set_style('darkgrid')
from IPython.display import display, Markdown, Latex
%matplotlib inline
```

In [7]:
```
from pathlib import Path

path = Path('./data')
best_expt_by_id = {}  # stores expt name, maps to object
for config_id, config in enumerate(path.iterdir()):
    best_expt_by_id[config_id] = []
    reward = -np.inf

    for expt in config.iterdir():
        data = np.load(expt)

        temp = str(expt).split('_')
        algorithm = temp[1]
        wind = temp[2] == "True"
        start_coord = np.fromiter(map(int, [temp[3], temp[4]]), int)
        p = float(temp[5])
        strategy = temp[6]
        alpha = float(temp[8])
        gamma = float(temp[9])
        heat_or_eps = float(temp[10])

        # tot_rewards = data['r_s'].sum()
```

```
        r = data['r'].sum()

        if r > reward:
            reward = r
            path = config / expt
            best_expt_by_id[config_id] = [algorithm, strategy, wind, start_coord, p, alp
```

In [8]:
```
def sort_order(item):
    key, (algorithm, strategy, wind, start_coord, p, alpha, gamma, heat_or_eps, path, to
    return algorithm == "sarsa", start_coord.sum(), p, wind,
best_expt_by_id = dict(sorted(best_expt_by_id.items(), key=sort_order))
```

In [9]:
```
len(best_expt_by_id)
```

Out[9]:
```
32
```

In [10]:
```
DOWN , UP, LEFT, RIGHT = 0, 1, 2, 3
x_direct = np.array((0, 0, -1, 1))
y_direct = np.array((-1, 1, 0, 0))

def plot_Q(Q, ax, vmax=None):
    sns.heatmap(Q.max(-1), edgecolors='k', linewidths=0.5, ax=ax)

    policy = Q.argmax(-1)
    policyx = x_direct[policy]
    policyy = y_direct[policy]
    idx = np.indices(policy.shape)
    ax.quiver(idx[1].ravel() + 0.5, idx[0].ravel() + 0.5, policyx.ravel(), policyy.ravel
```

In [11]:
```
visits_max = 4000
q_max = 1000

def print(algo):
    tick = 1
    for key in best_expt_by_id.keys():

        expt_info = best_expt_by_id[key]
        [algorithm, strategy, wind, start_coord, p, alpha, gamma, heat_or_eps, path, tot
        if algorithm != algo:
            continue

        display(Markdown(f'# Configuration {tick}'))
        outputs = (
            "| Reward | Algorithm | Exploration Strategy | Wind | Start Coors | P | . |
            "| :-: | :-: | :-: | :-: | :-: | :-: | :-: | :-: | :-: | :-: |",
            f"| {tot_rewards} | {algorithm} | {strategy} | {wind} | {start_coord} | {p}
        )
        display(Markdown("\n".join(outputs)))

        data = np.load(path)
        Q = data['q']
        Q = Q[::-1, :, :]
        visits = data['visits']
        visits = visits[::-1, ::]

        fig, axs = plt.subplots(2, 2, figsize=(10, 6))

        # Plot Optimal Policy
        plot_Q(Q, axs[0, 0], vmax=q_max)
        axs[0, 0].set_title('Q values with optimal policy')
```

```
        # Plotting Reward Curve
        episodes = np.arange(data['r_s'].shape[0])
        axs[0, 1].plot(episodes, data['r_s'])
        axs[0, 1].set_title('Rewards vs No of Episodes')
        axs[0, 1].set_xlabel('Episode')
        axs[0, 1].set_ylabel('Rewards')
        axs[0, 1].set_ylim(-200, 10)

        fig.show()

        # Plotting Steps Curve
        axs[1, 1].plot(episodes, data['s_s'])
        axs[1, 1].set_title('Steps to Steps vs No of Episodes')
        axs[1, 1].set_xlabel('Episode')
        axs[1, 1].set_ylabel('Steps')
        axs[1, 1].set_ylim(0, 100)

        # Plotting Heatmap of visited states
        sns.heatmap(visits, ax=axs[1, 0], vmax=visits_max)
        axs[1, 0].set_title('Heatmap of Visited States')

        fig.show()

        plt.pause(0.001)
        tick += 1
```

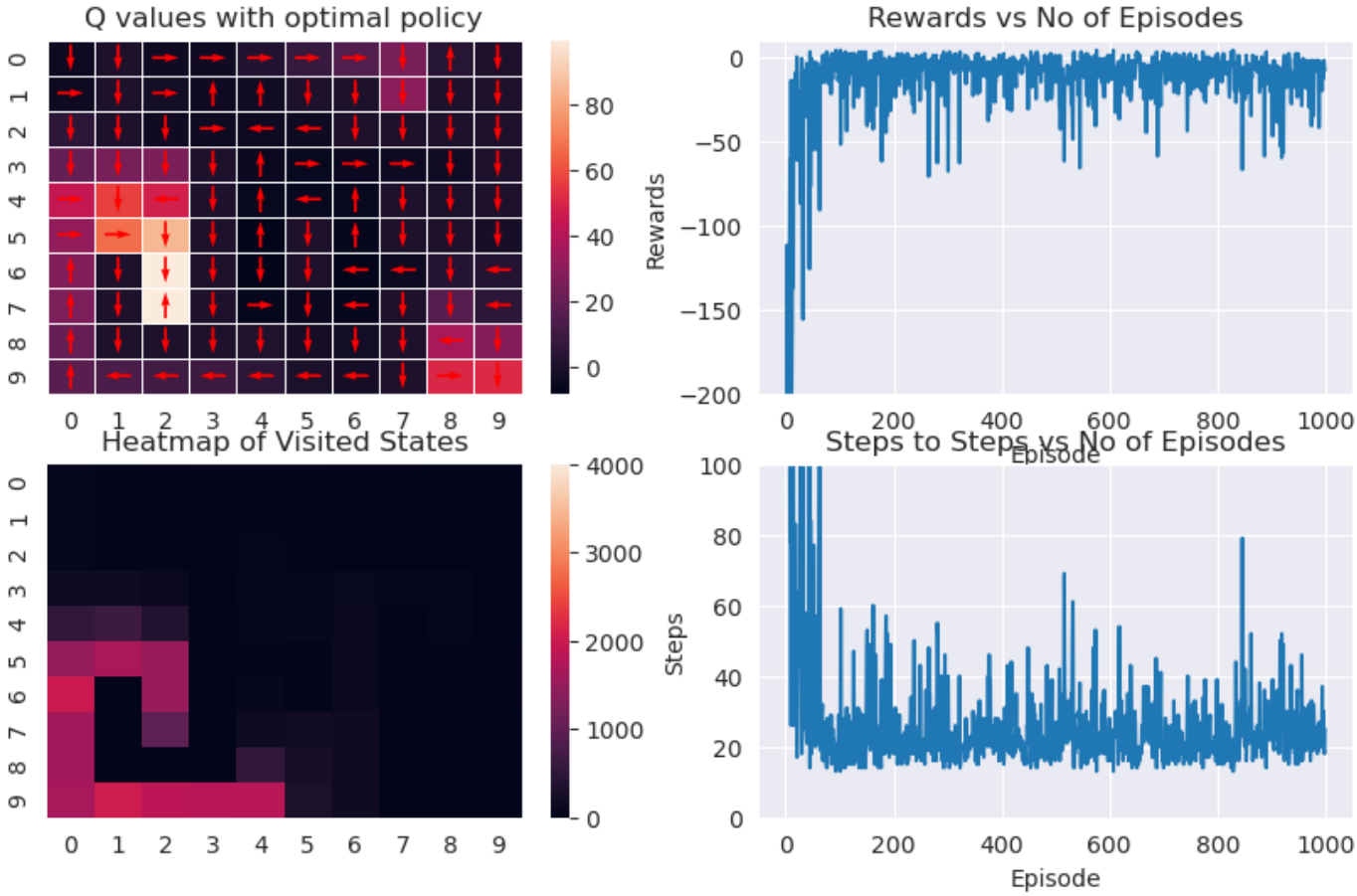
SARSA

In [12]:
```
print("sarsa")
```

Configuration 1

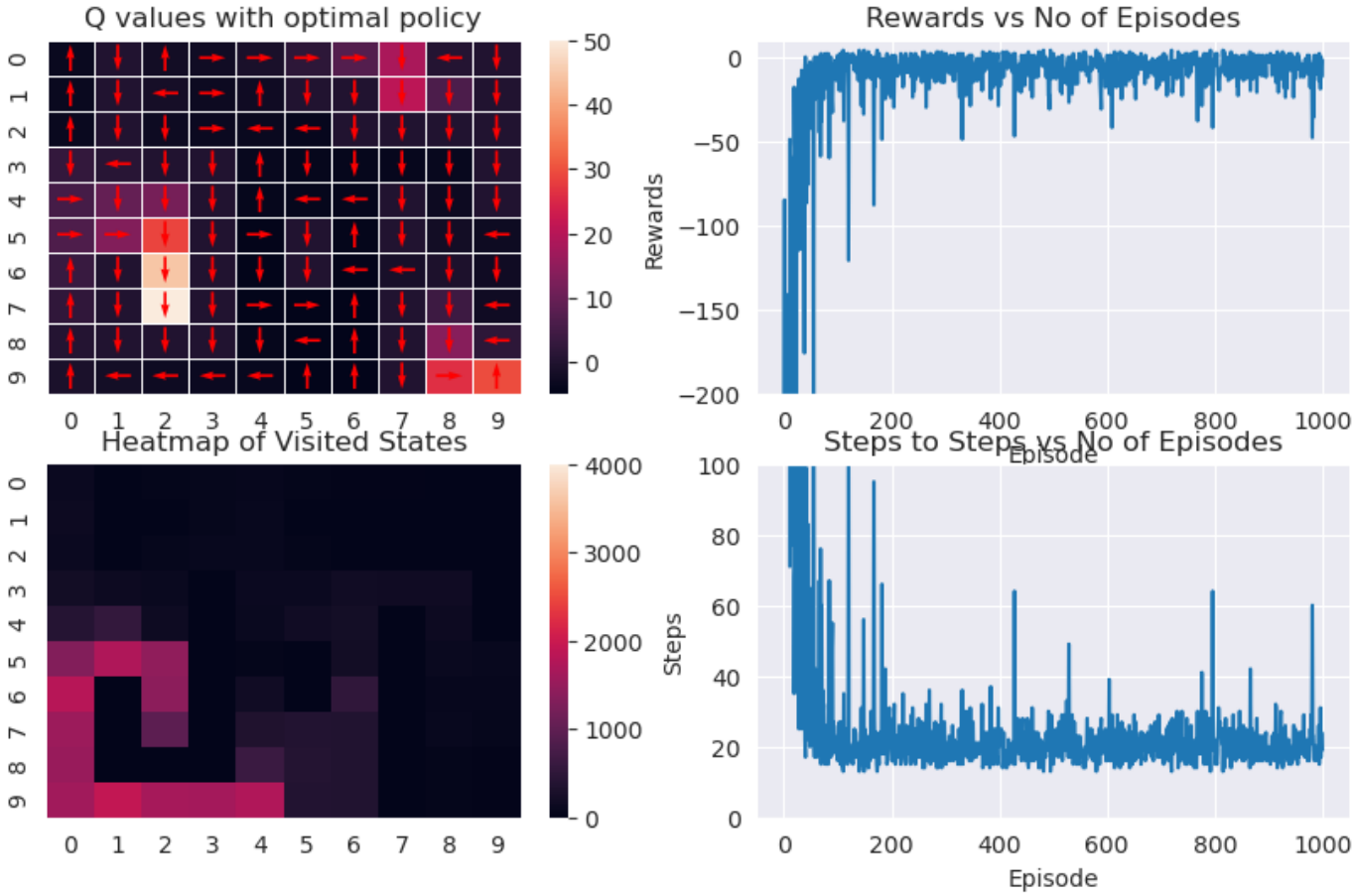
Reward	Algorithm	Exploration Strategy	Wind	Start Coors	P	.	Gamma	Alpha	Exploration Param
4.0	sarsa	EpsilonGreedy	False	[0 4]	0.7	.	0.9	0.5	2.0

```
/tmp/ipykernel_8072/2852134997.py:41: UserWarning: Matplotlib is currently using modul
e://matplotlib_inline.backend_inline, which is a non-GUI backend, so cannot show the fig
ure.
  fig.show()
/tmp/ipykernel_8072/2852134997.py:54: UserWarning: Matplotlib is currently using modul
e://matplotlib_inline.backend_inline, which is a non-GUI backend, so cannot show the fig
ure.
  fig.show()
```



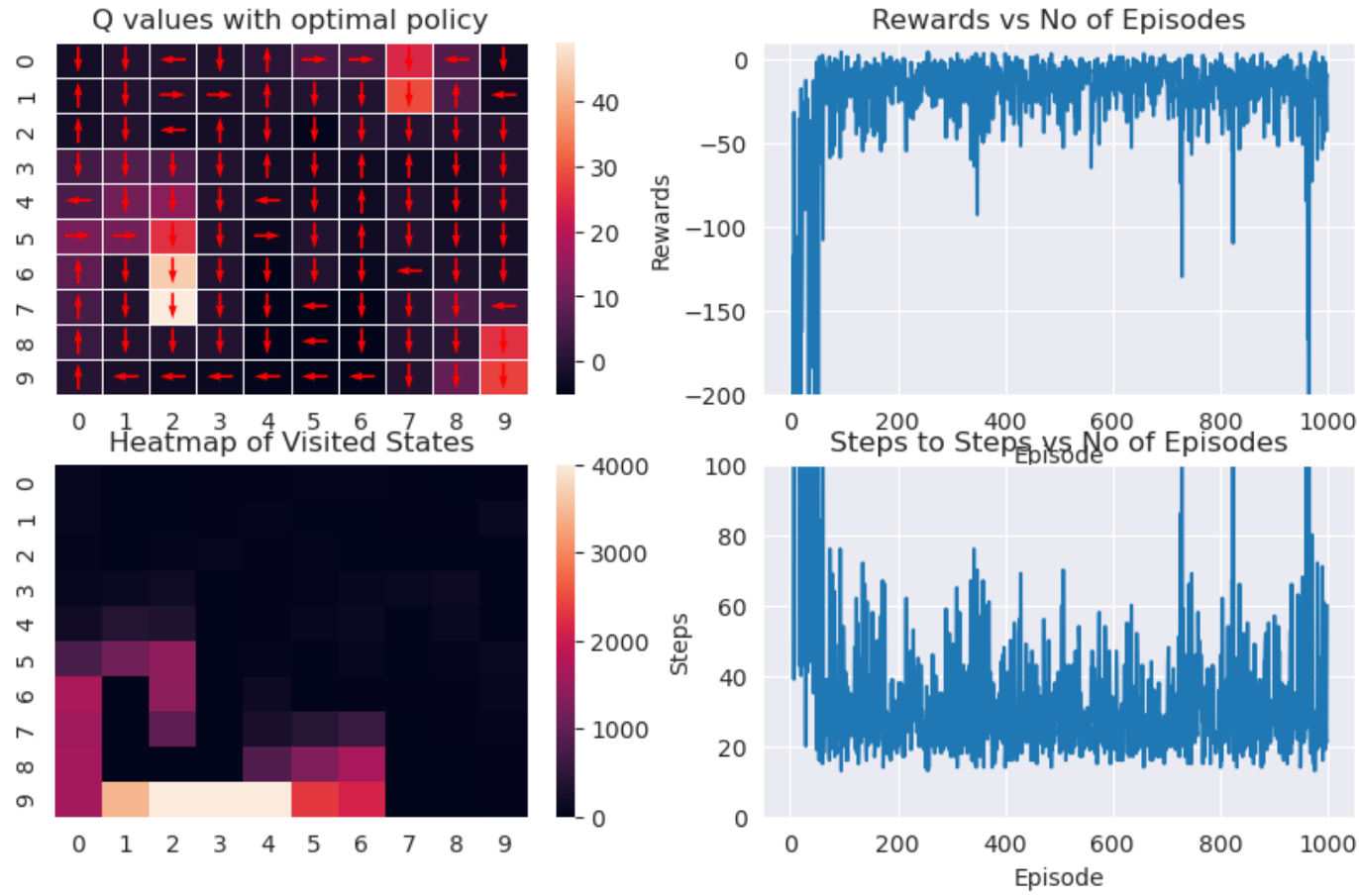
Configuration 2

Reward	Algorithm	Exploration Strategy	Wind	Start Coors	P	.	Gamma	Alpha	Exploration Param
4.0	sarsa	Softmax	False	[0 4]	0.7	.	0.8	0.2	0.0



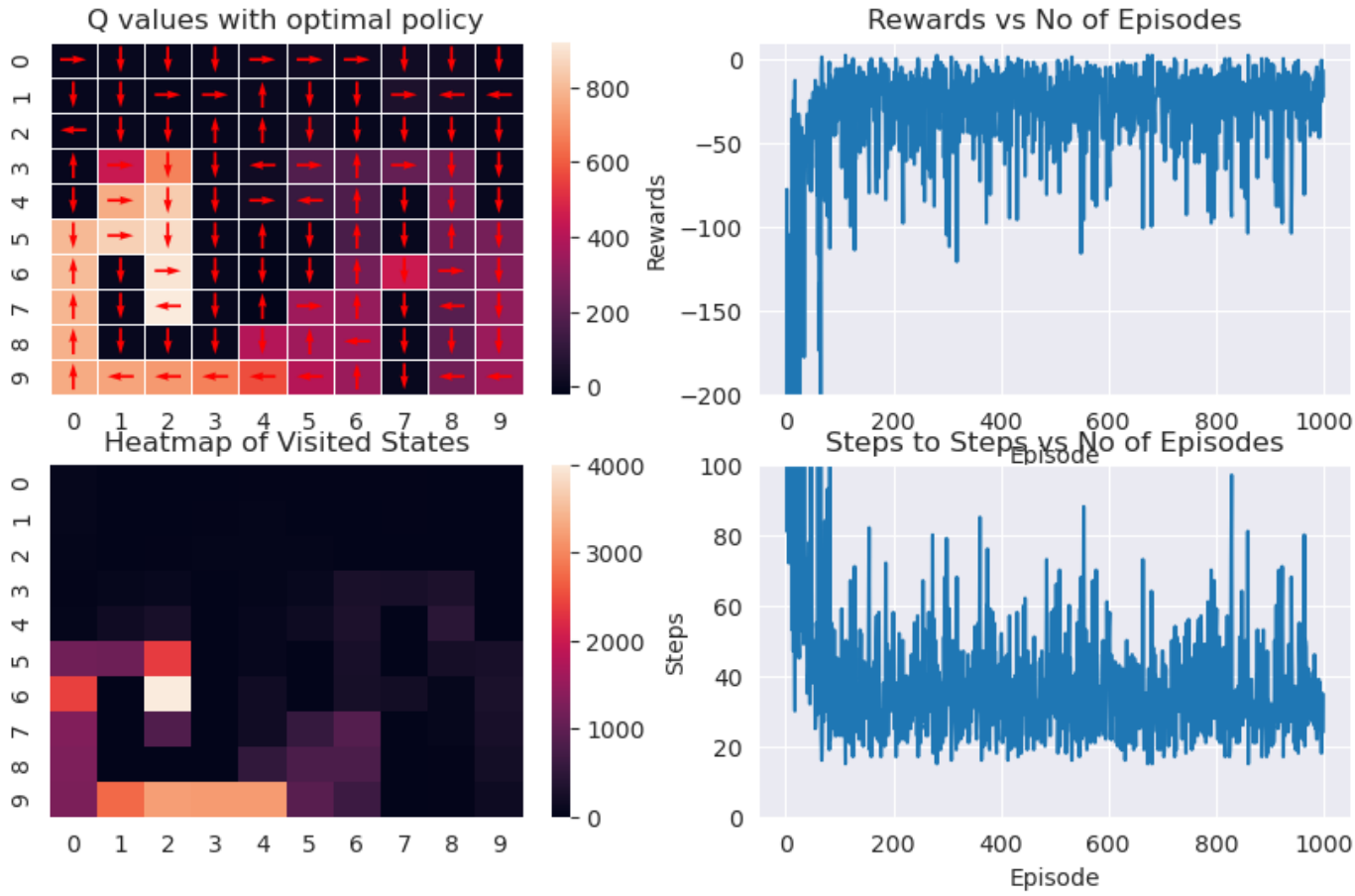
Configuration 3

Reward	Algorithm	Exploration Strategy	Wind	Start Coors	P	.	Gamma	Alpha	Exploration Param
2.0	sarsa	EpsilonGreedy	True	[0 4]	0.7	.	0.8	0.2	1.0



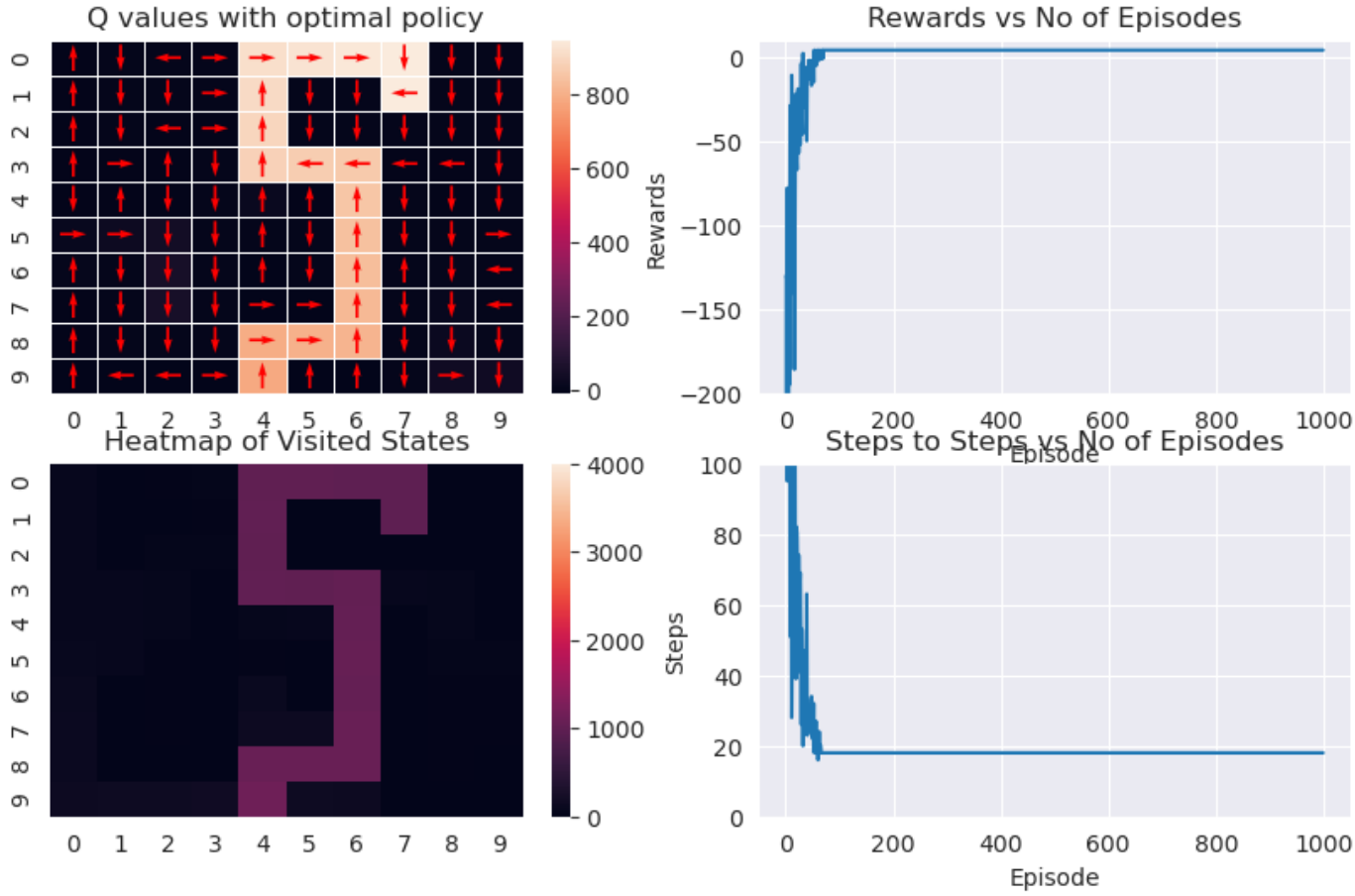
Configuration 4

Reward	Algorithm	Exploration Strategy	Wind	Start Coors	P	.	Gamma	Alpha	Exploration Param
2.0	sarsa	Softmax	True	[0 4]	0.7	.	0.99	0.3	0.0



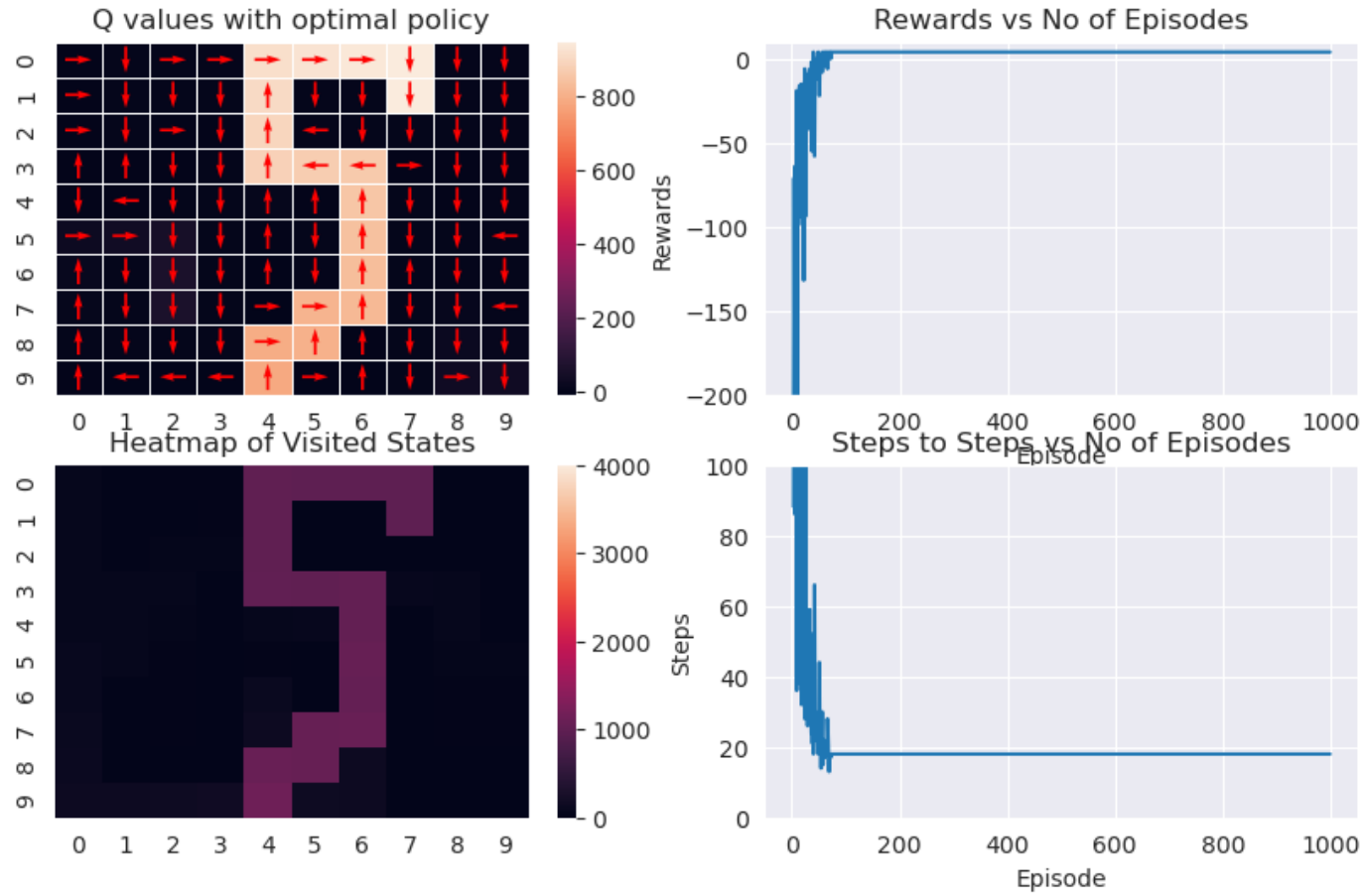
Configuration 5

Reward	Algorithm	Exploration Strategy	Wind	Start Coors	P	.	Gamma	Alpha	Exploration Param
4.0	sarsa	Softmax	False	[0 4]	1.0	.	0.99	0.3	0.0



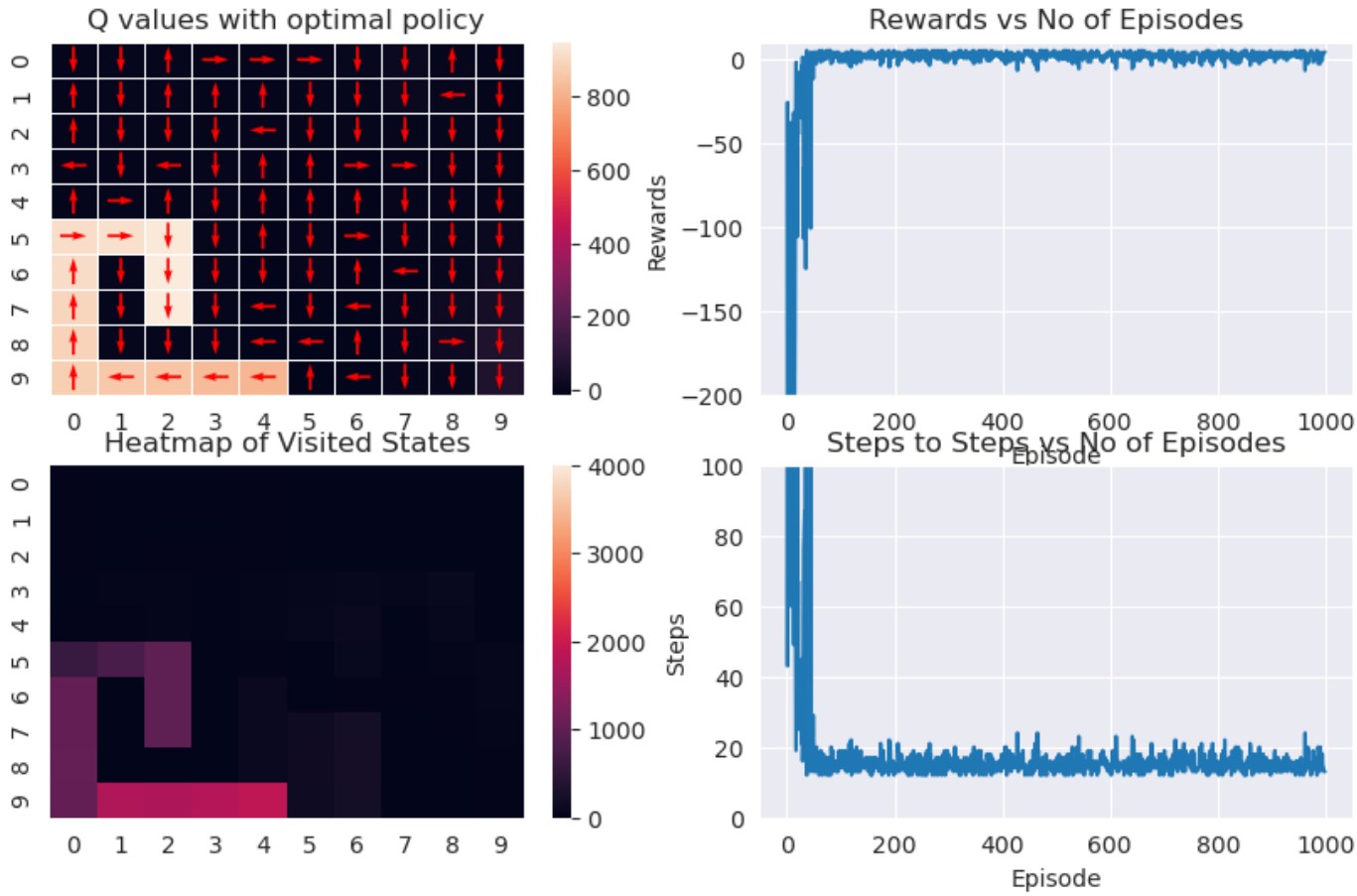
Configuration 6

Reward	Algorithm	Exploration Strategy	Wind	Start Coors	P	.	Gamma	Alpha	Exploration Param
4.0	sarsa	EpsilonGreedy	False	[0 4]	1.0	.	0.99	0.3	0.0



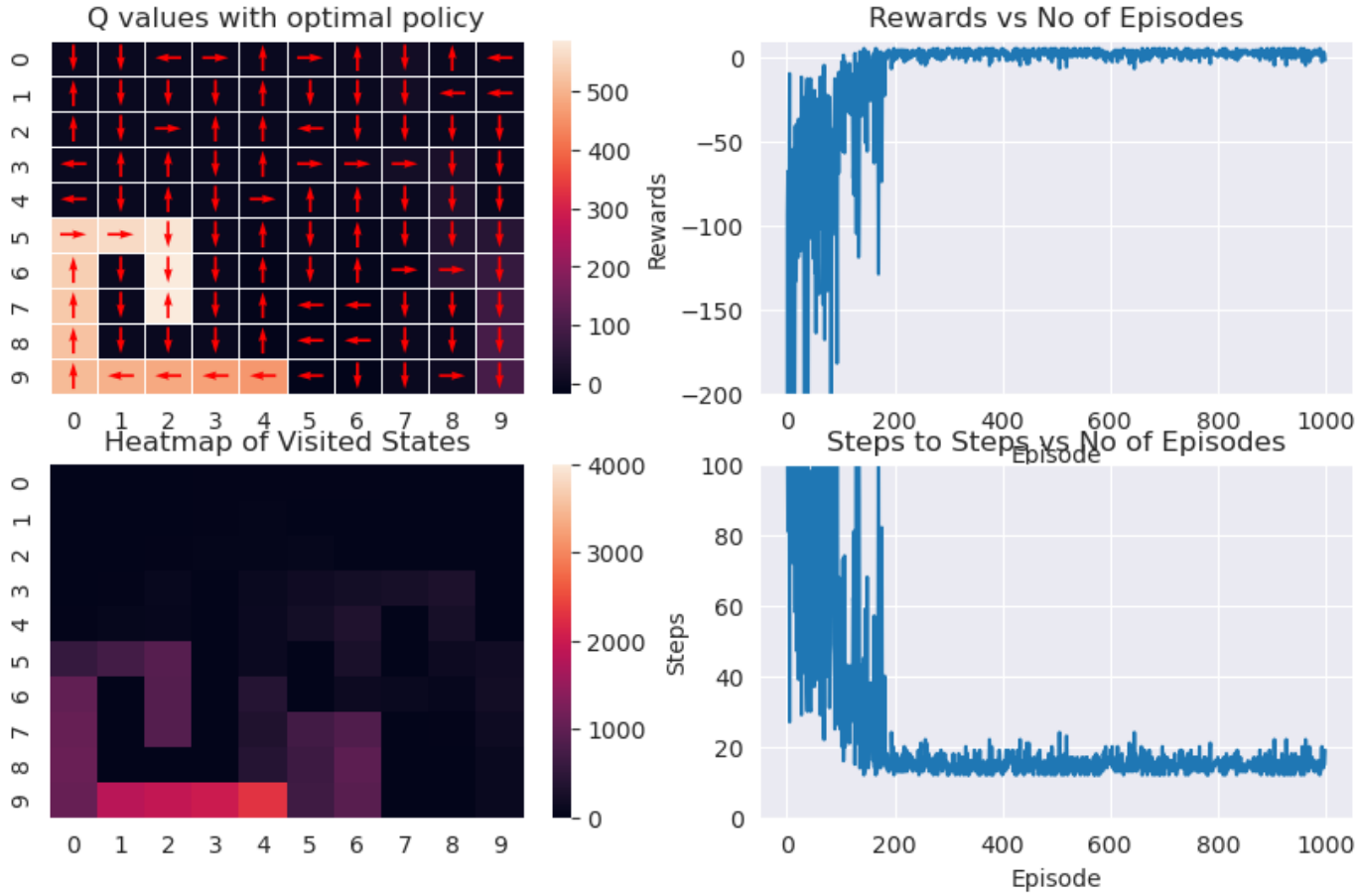
Configuration 7

Reward	Algorithm	Exploration Strategy	Wind	Start Coors	P	.	Gamma	Alpha	Exploration Param
5.0	sarsa	EpsilonGreedy	True	[0 4]	1.0	.	0.99	0.3	0.0



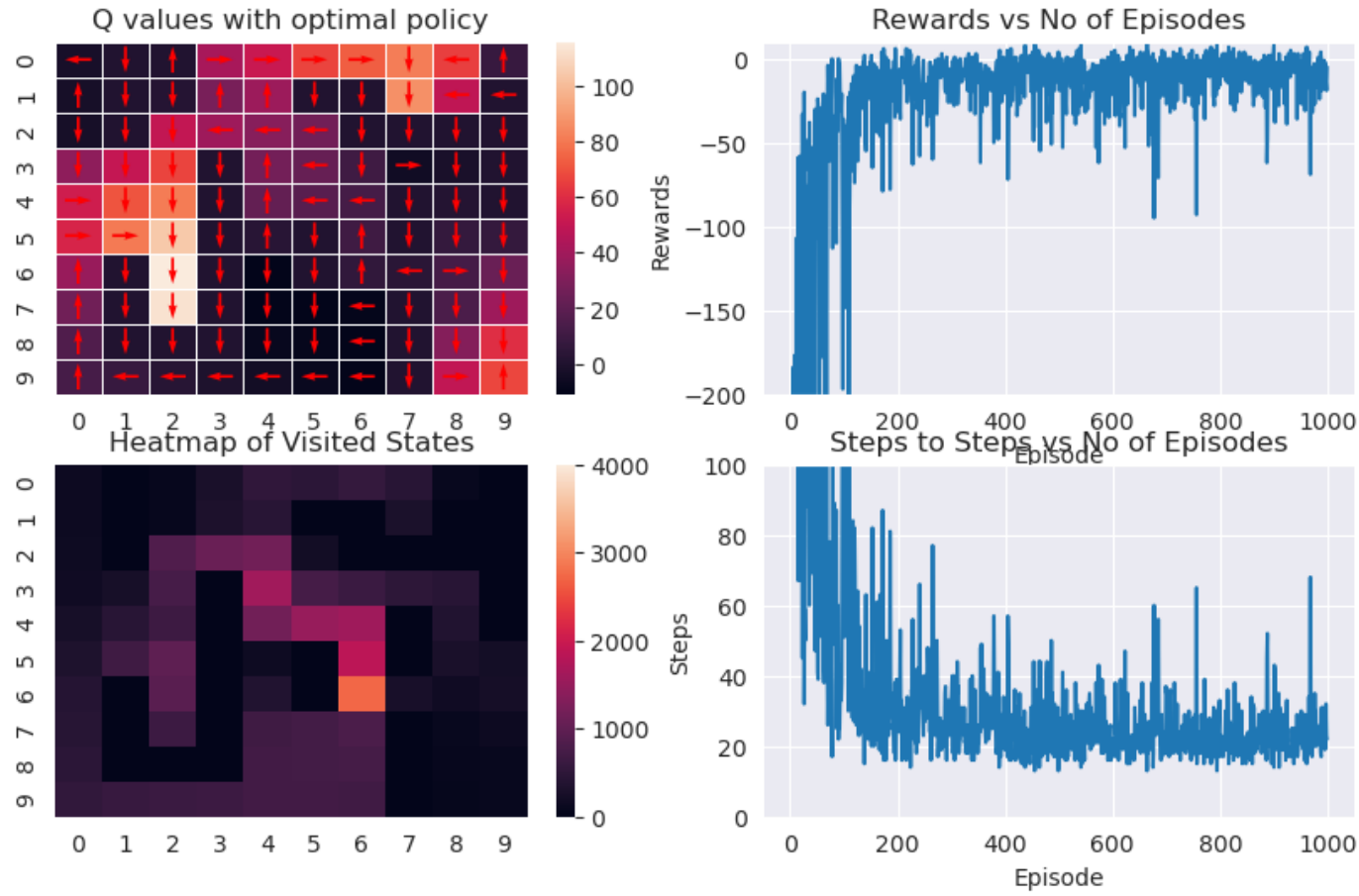
Configuration 8

Reward	Algorithm	Exploration Strategy	Wind	Start Coors	P	.	Gamma	Alpha	Exploration Param
5.0	sarsa	Softmax	True	[0 4]	1.0	.	0.99	0.1	1.0



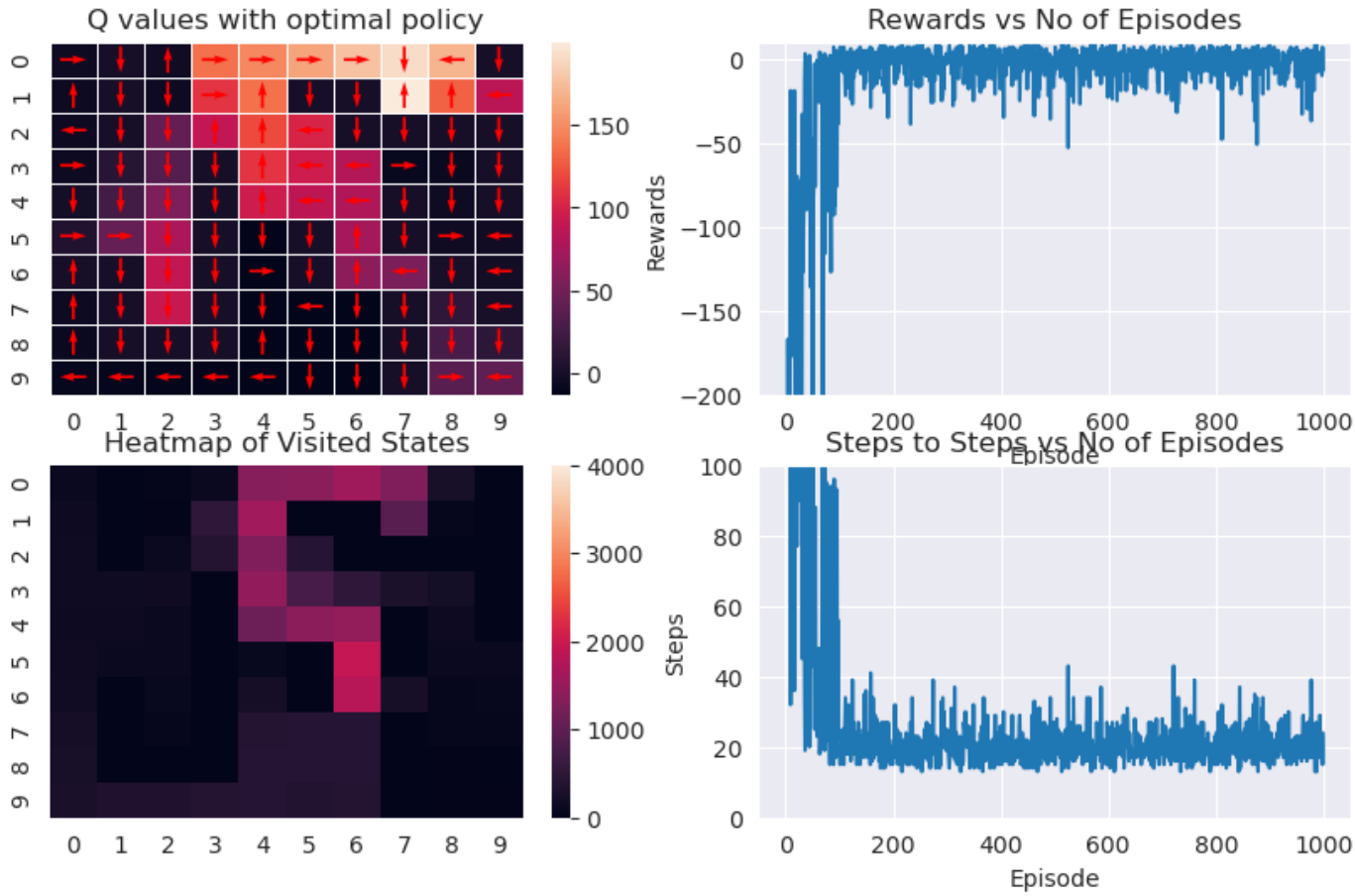
Configuration 9

Reward	Algorithm	Exploration Strategy	Wind	Start Coors	P	.	Gamma	Alpha	Exploration Param
9.0	sarsa	EpsilonGreedy	False	[3 6]	0.7	.	0.95	0.1	3.0



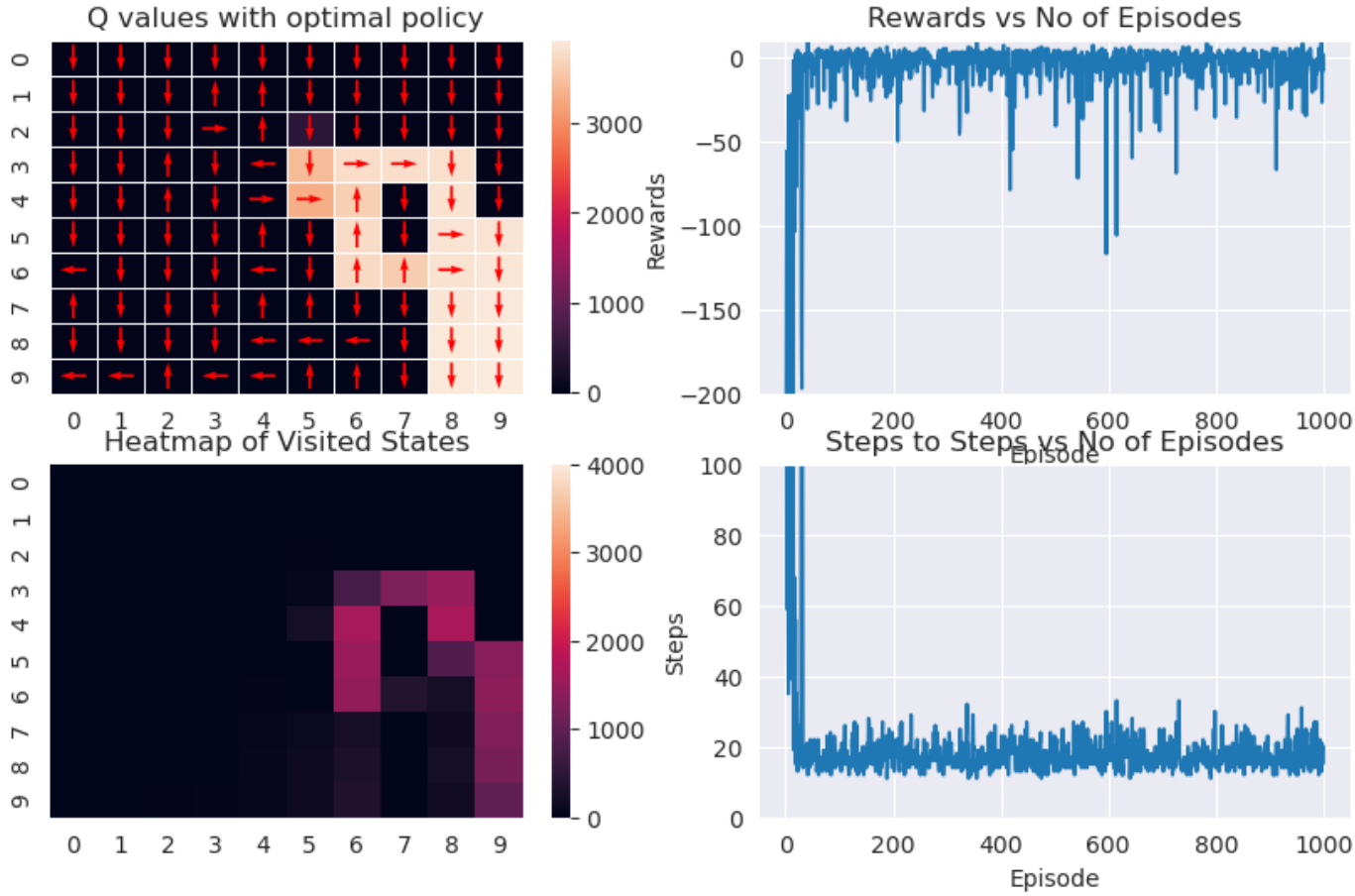
Configuration 10

Reward	Algorithm	Exploration Strategy	Wind	Start Coors	P	.	Gamma	Alpha	Exploration Param
9.0	sarsa	Softmax	False	[3 6]	0.7	.	0.95	0.2	1.0



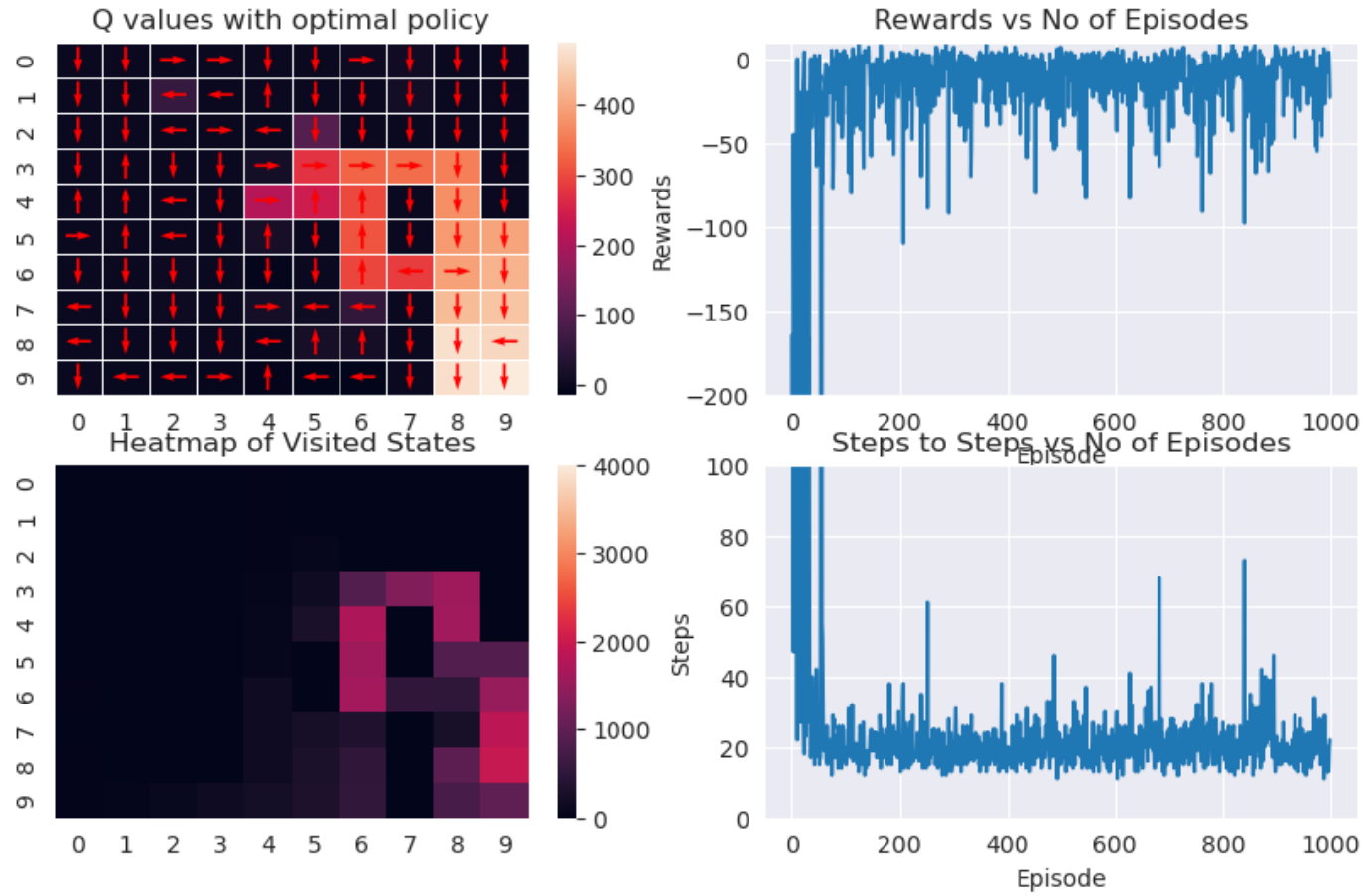
Configuration 11

Reward	Algorithm	Exploration Strategy	Wind	Start Coors	P	.	Gamma	Alpha	Exploration Param
6.0	sarsa	Softmax	True	[3 6]	0.7	.	0.999	0.5	3.0



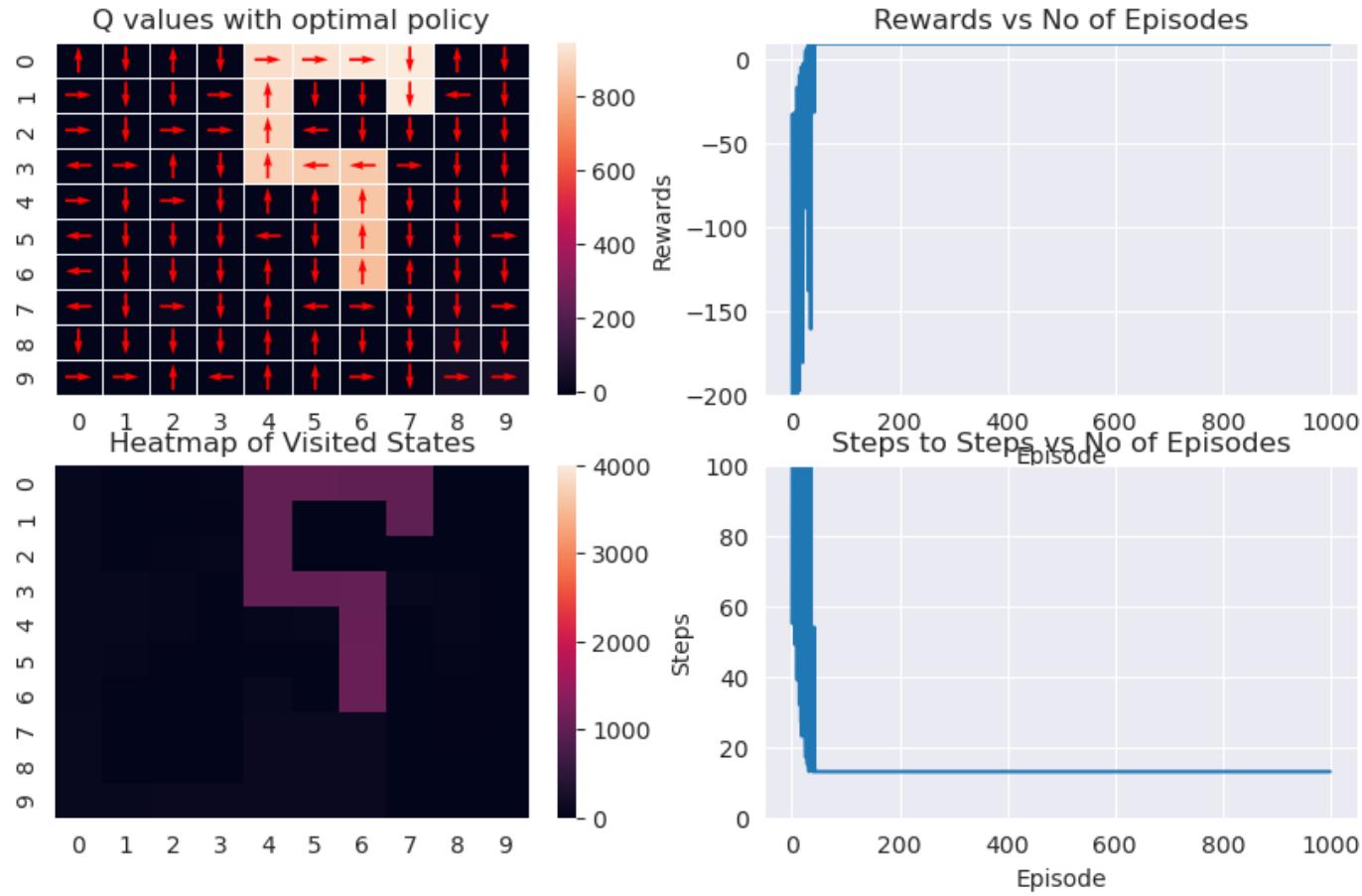
Configuration 12

Reward	Algorithm	Exploration Strategy	Wind	Start Coors	P	.	Gamma	Alpha	Exploration Param
9.0	sarsa	EpsilonGreedy	True	[3 6]	0.7	.	0.99	0.2	1.0



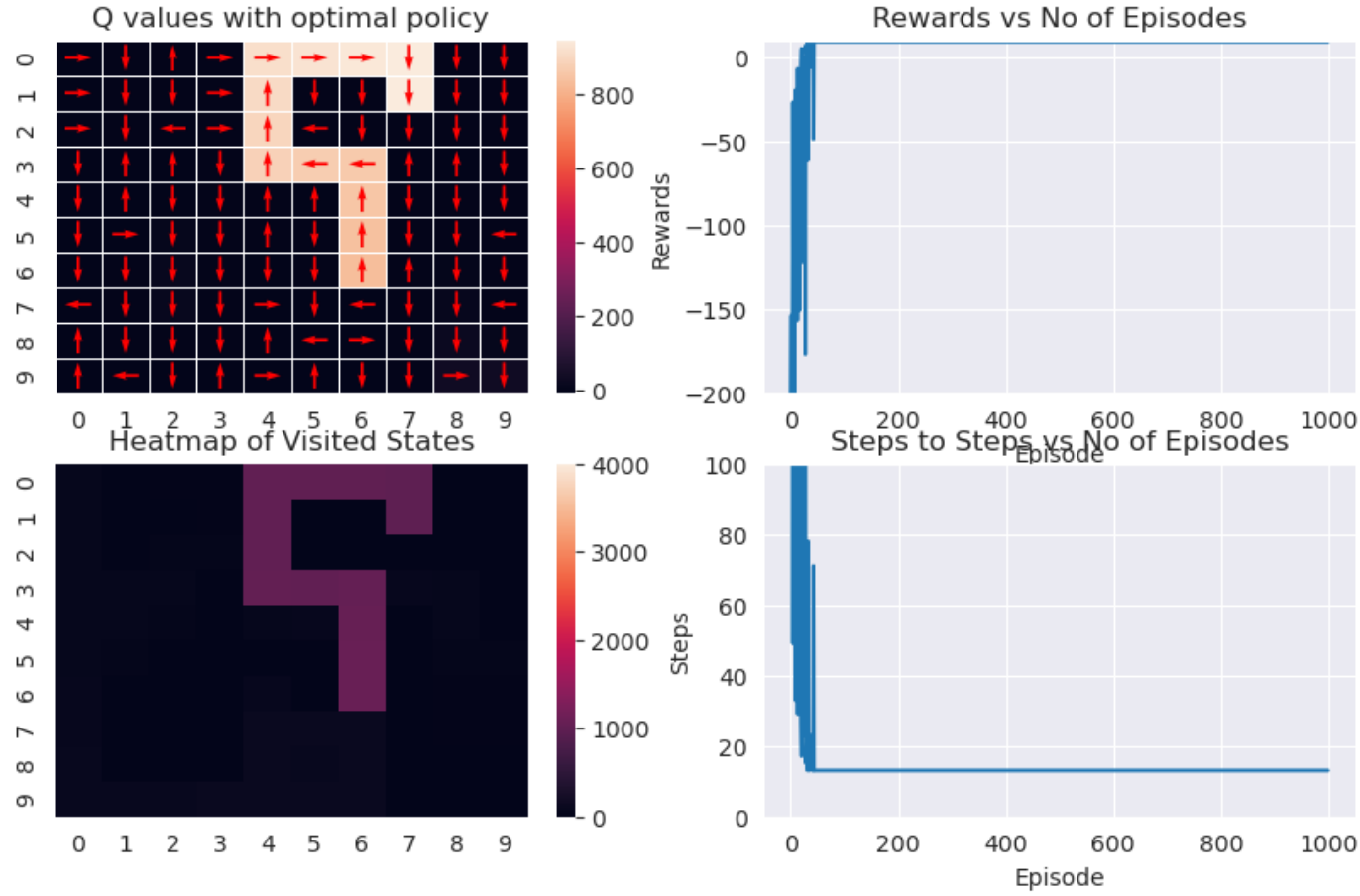
Configuration 13

Reward	Algorithm	Exploration Strategy	Wind	Start Coors	P	.	Gamma	Alpha	Exploration Param
9.0	sarsa	Softmax	False	[3 6]	1.0	.	0.99	0.3	0.0



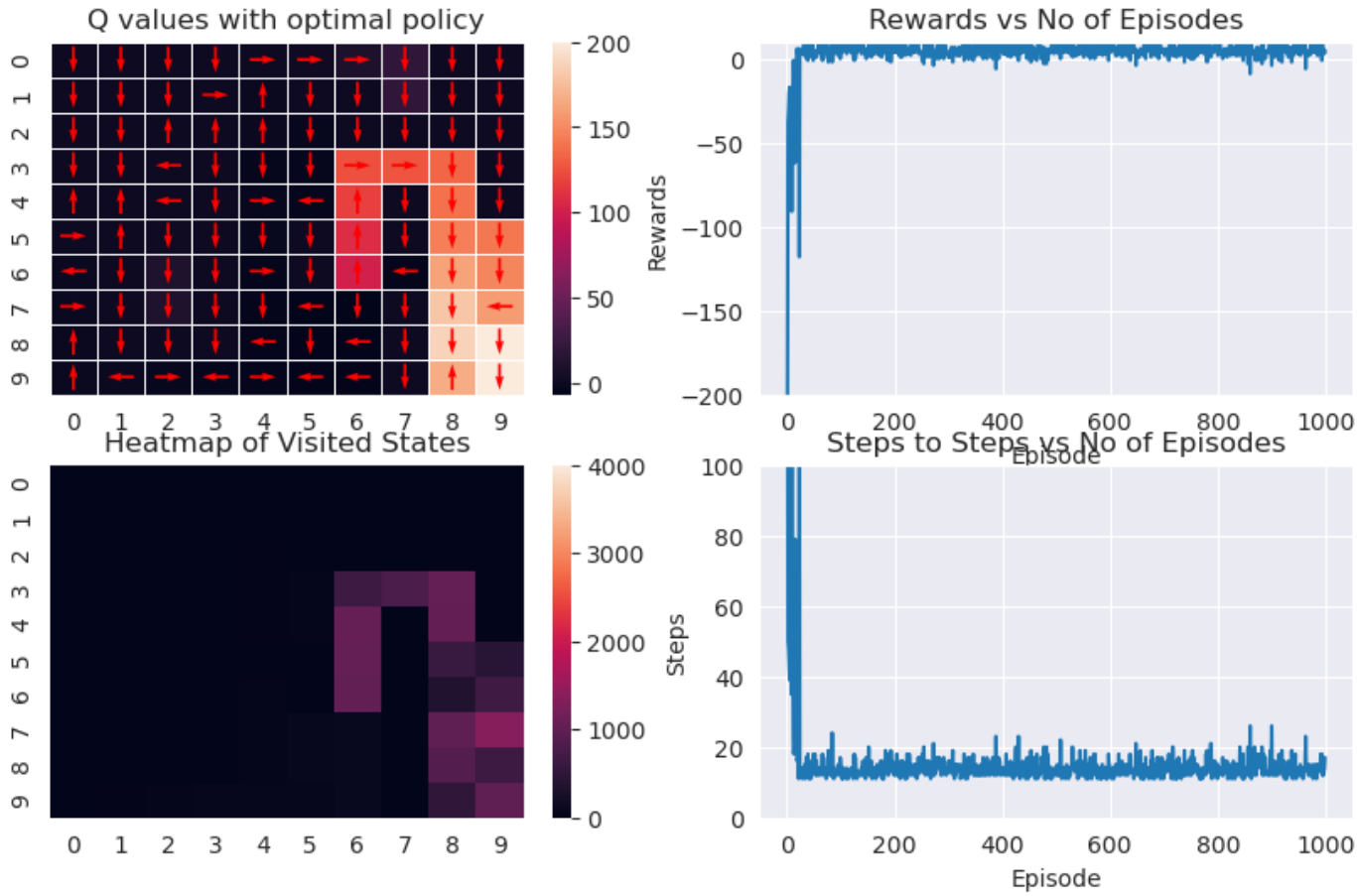
Configuration 14

Reward	Algorithm	Exploration Strategy	Wind	Start Coors	P	.	Gamma	Alpha	Exploration Param
9.0	sarsa	EpsilonGreedy	False	[3 6]	1.0	.	0.99	0.3	0.0



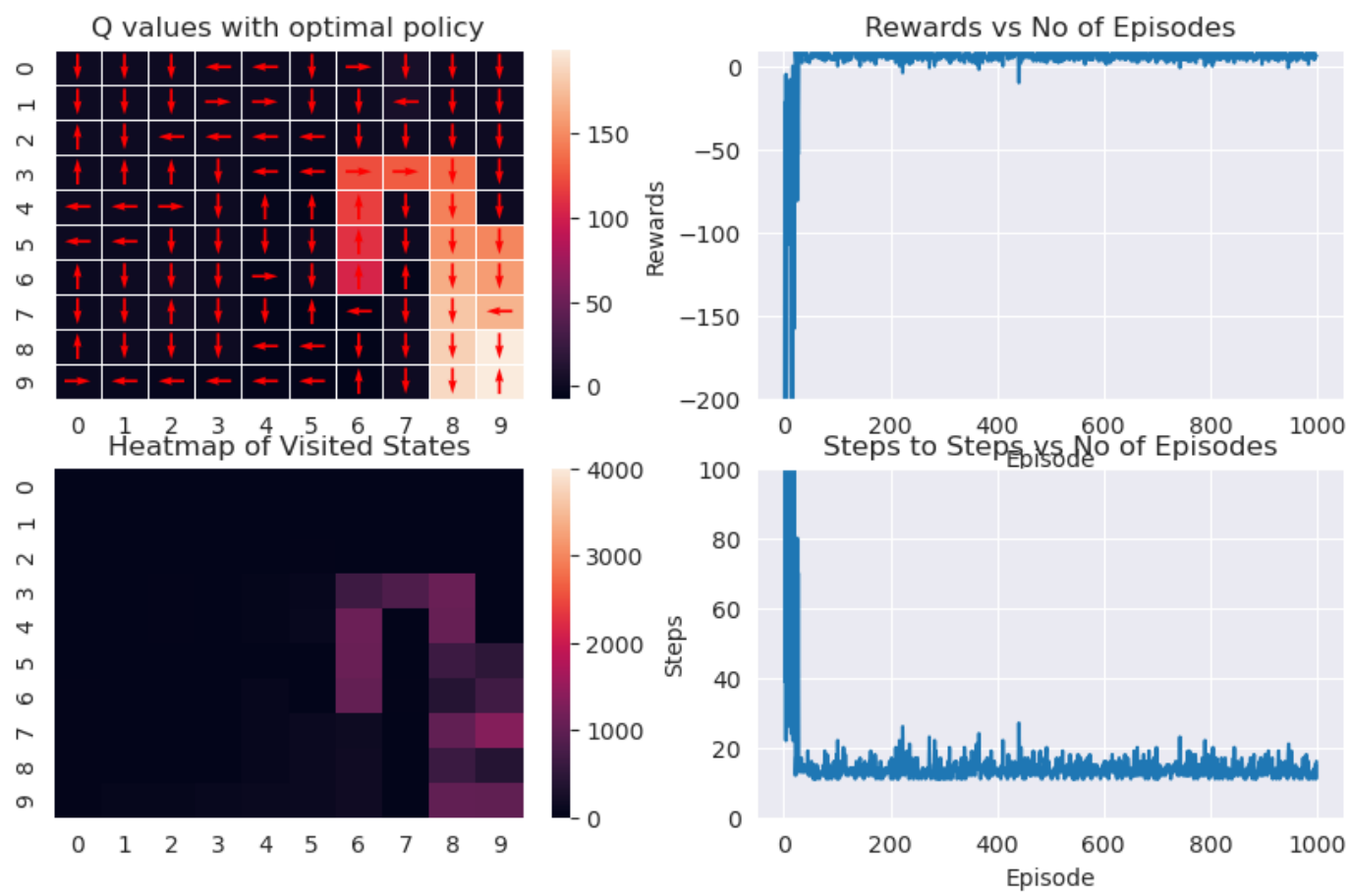
Configuration 15

Reward	Algorithm	Exploration Strategy	Wind	Start Coors	P	.	Gamma	Alpha	Exploration Param
10.0	sarsa	EpsilonGreedy	True	[3 6]	1.0	.	0.95	0.5	0.0



Configuration 16

Reward	Algorithm	Exploration Strategy	Wind	Start Coors	P	.	Gamma	Alpha	Exploration Param
9.0	sarsa	Softmax	True	[3 6]	1.0	.	0.95	0.3	0.0



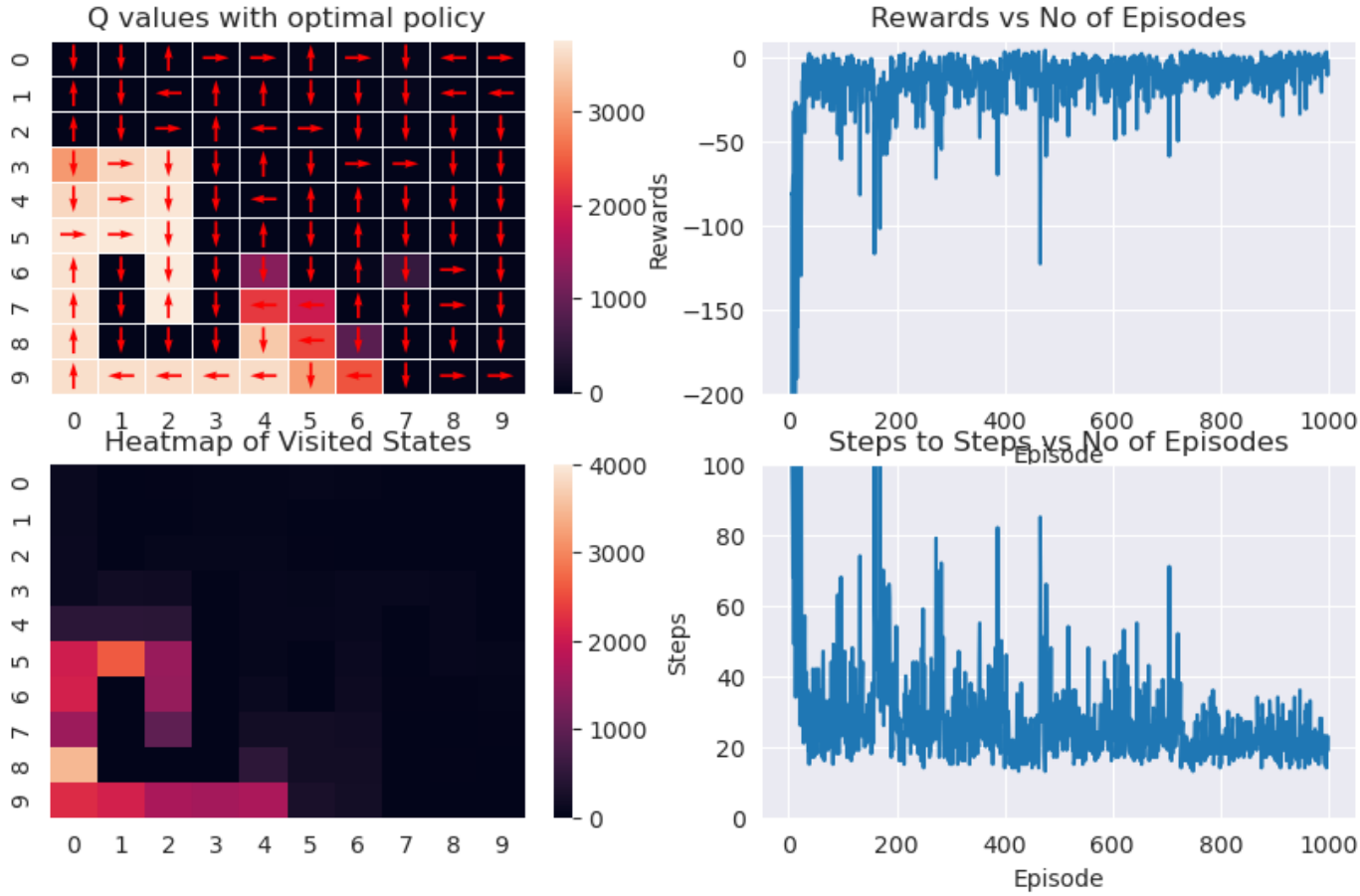
QLearning

In [13]:
```
print("qlearning")
```

Configuration 1

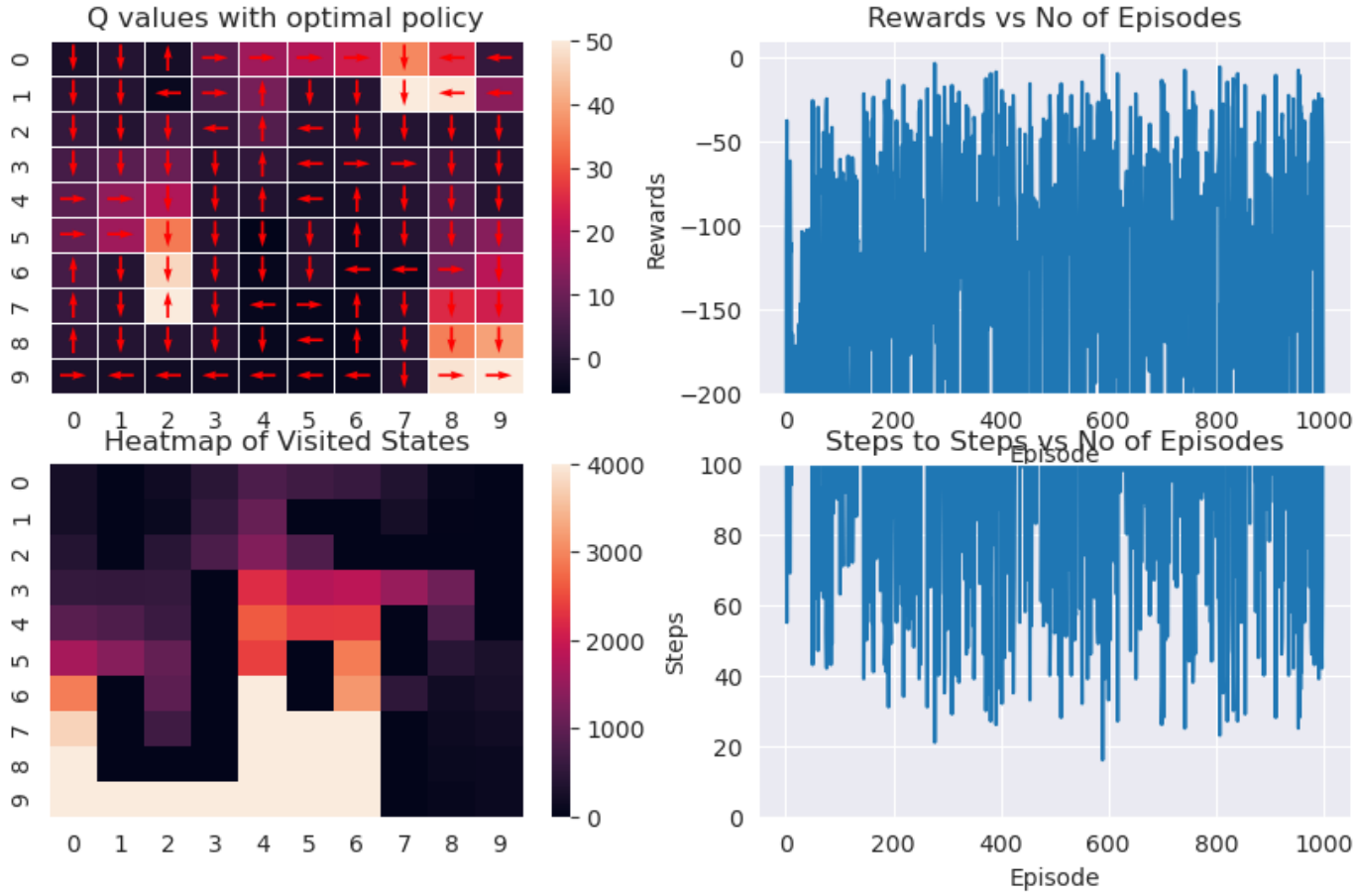
Reward	Algorithm	Exploration Strategy	Wind	Start Coors	P	.	Gamma	Alpha	Exploration Param
3.0	qlearning	EpsilonGreedy	False	[0 4]	0.7	.	0.999	0.5	2.0

```
/tmp/ipykernel_8072/2852134997.py:41: UserWarning: Matplotlib is currently using modul
e://matplotlib_inline.backend_inline, which is a non-GUI backend, so cannot show the fig
ure.
  fig.show()
/tmp/ipykernel_8072/2852134997.py:54: UserWarning: Matplotlib is currently using modul
e://matplotlib_inline.backend_inline, which is a non-GUI backend, so cannot show the fig
ure.
  fig.show()
```



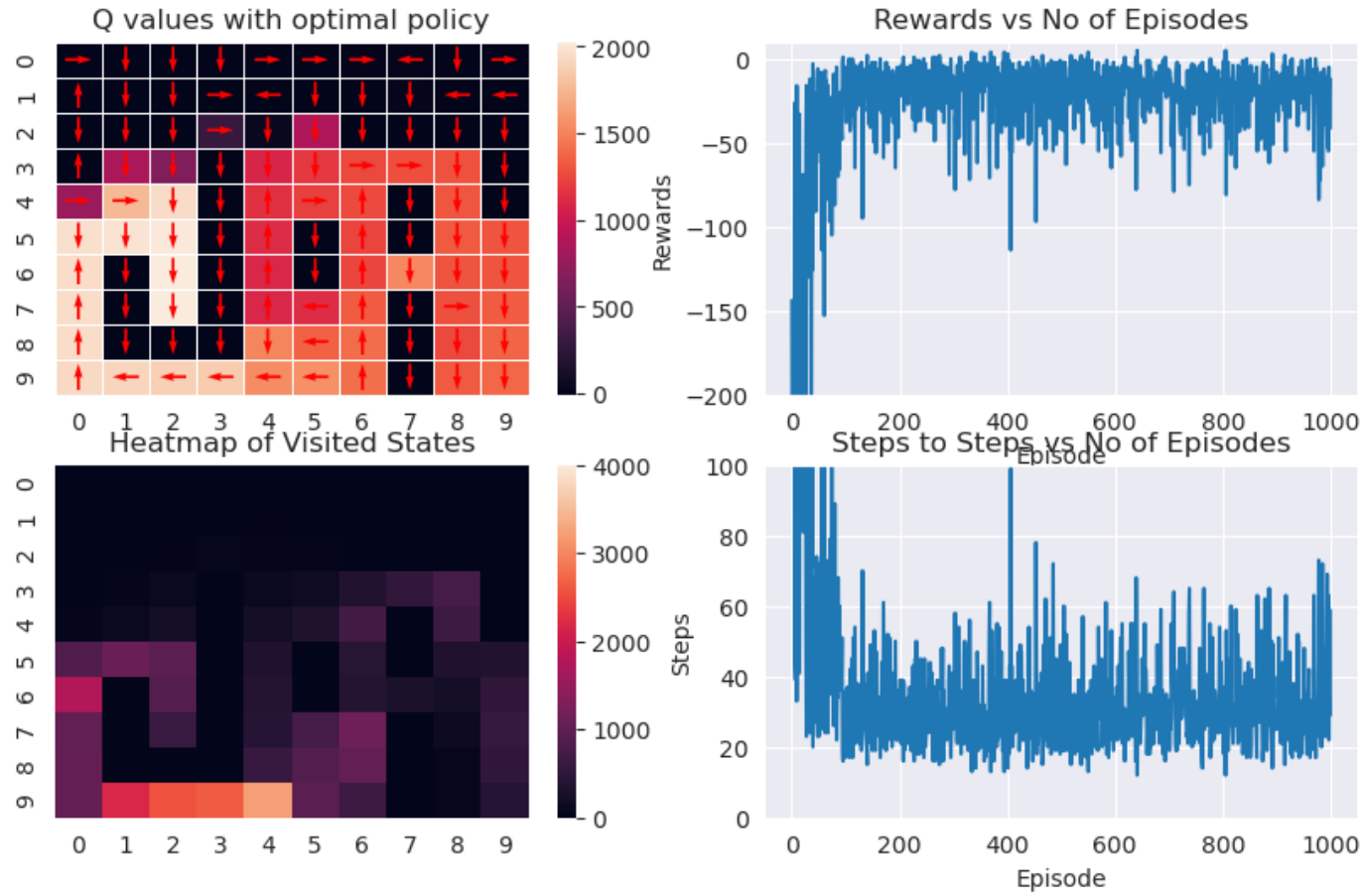
Configuration 2

Reward	Algorithm	Exploration Strategy	Wind	Start Coors	P	.	Gamma	Alpha	Exploration Param
4.0	qlearning	Softmax	False	[0 4]	0.7	.	0.8	0.4	4.0



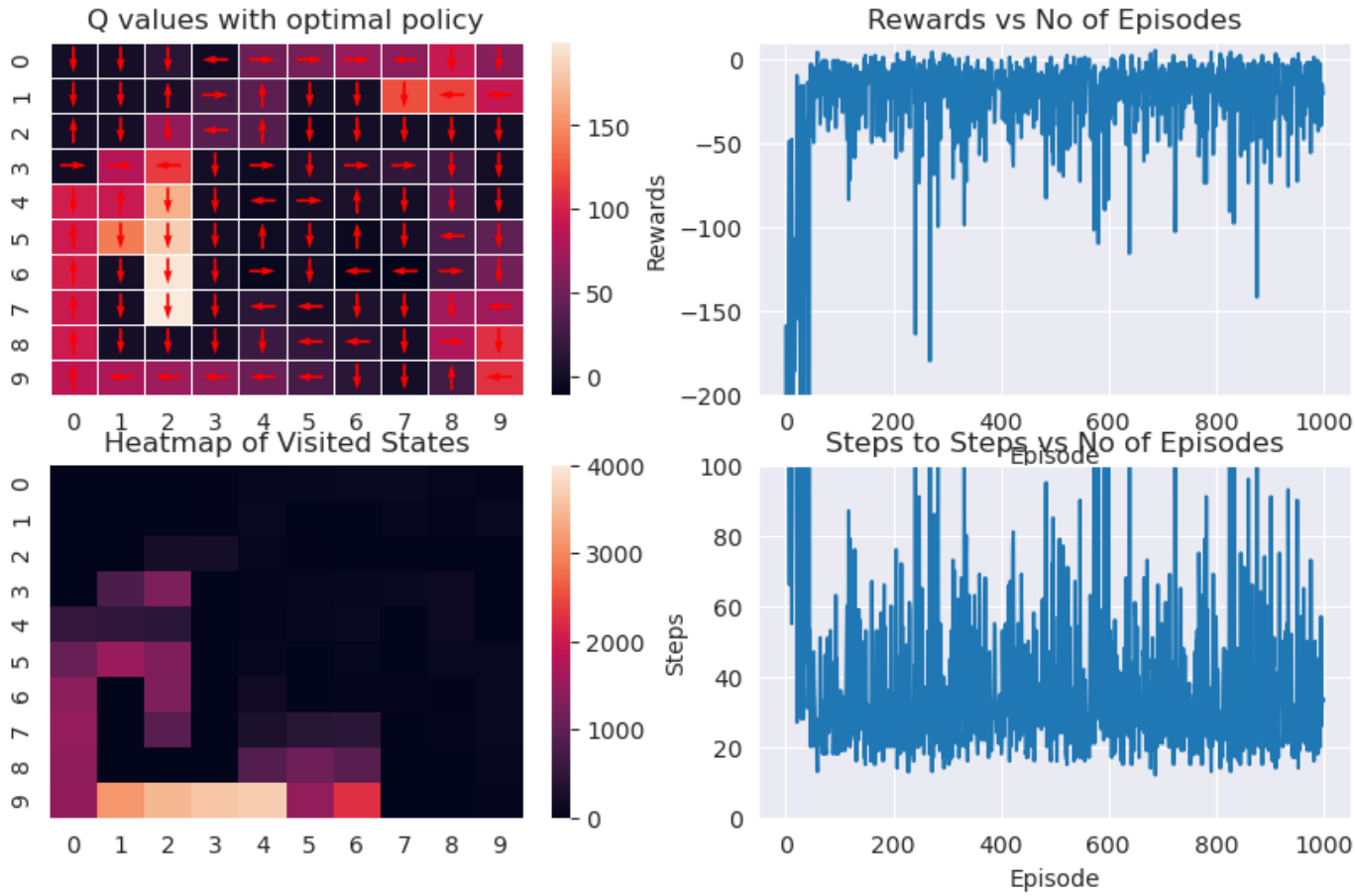
Configuration 3

Reward	Algorithm	Exploration Strategy	Wind	Start Coors	P	.	Gamma	Alpha	Exploration Param
3.0	qlearning	EpsilonGreedy	True	[0 4]	0.7	.	0.999	0.4	3.0



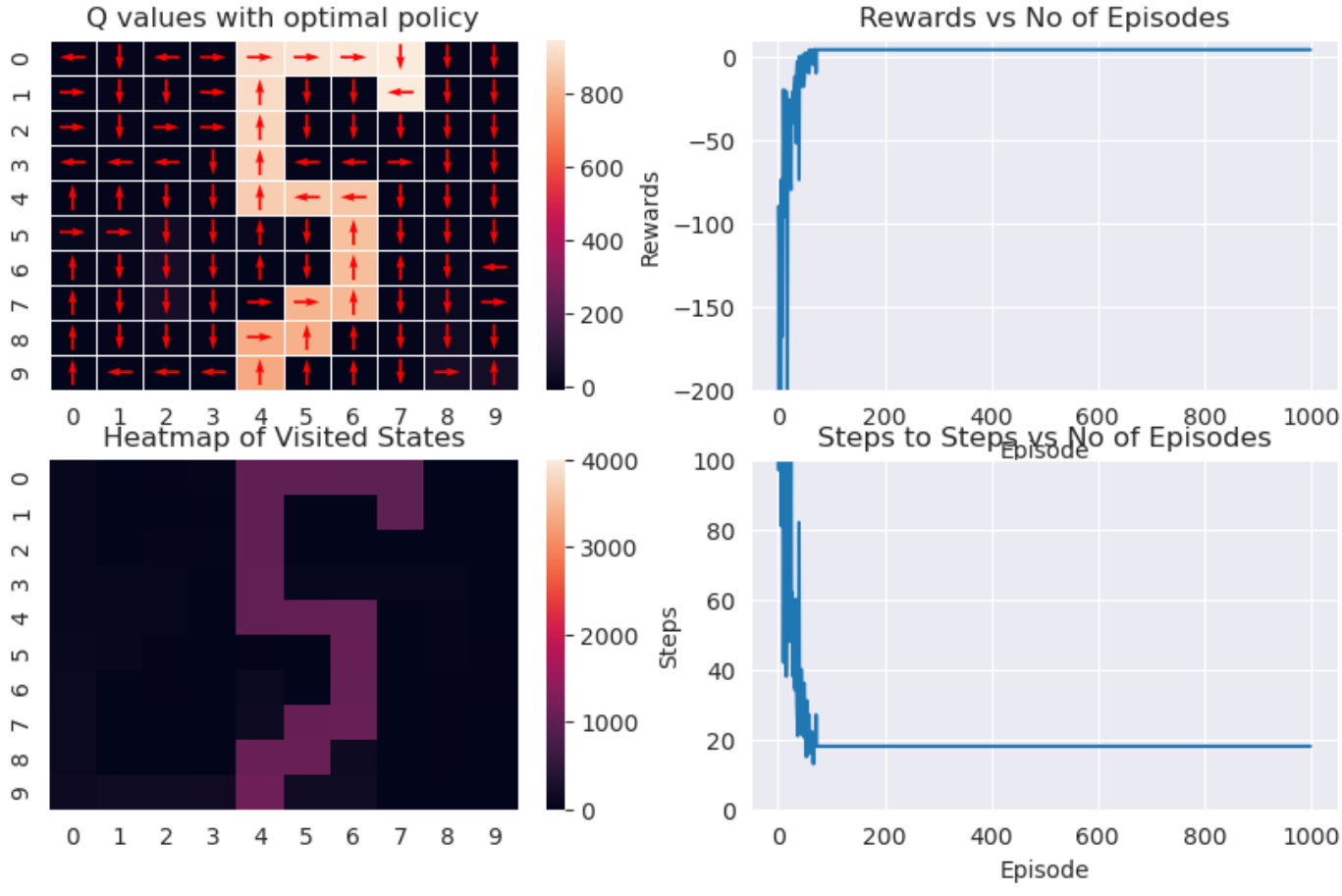
Configuration 4

Reward	Algorithm	Exploration Strategy	Wind	Start Coors	P	.	Gamma	Alpha	Exploration Param
3.0	qlearning	Softmax	True	[0 4]	0.7	.	0.95	0.4	0.0



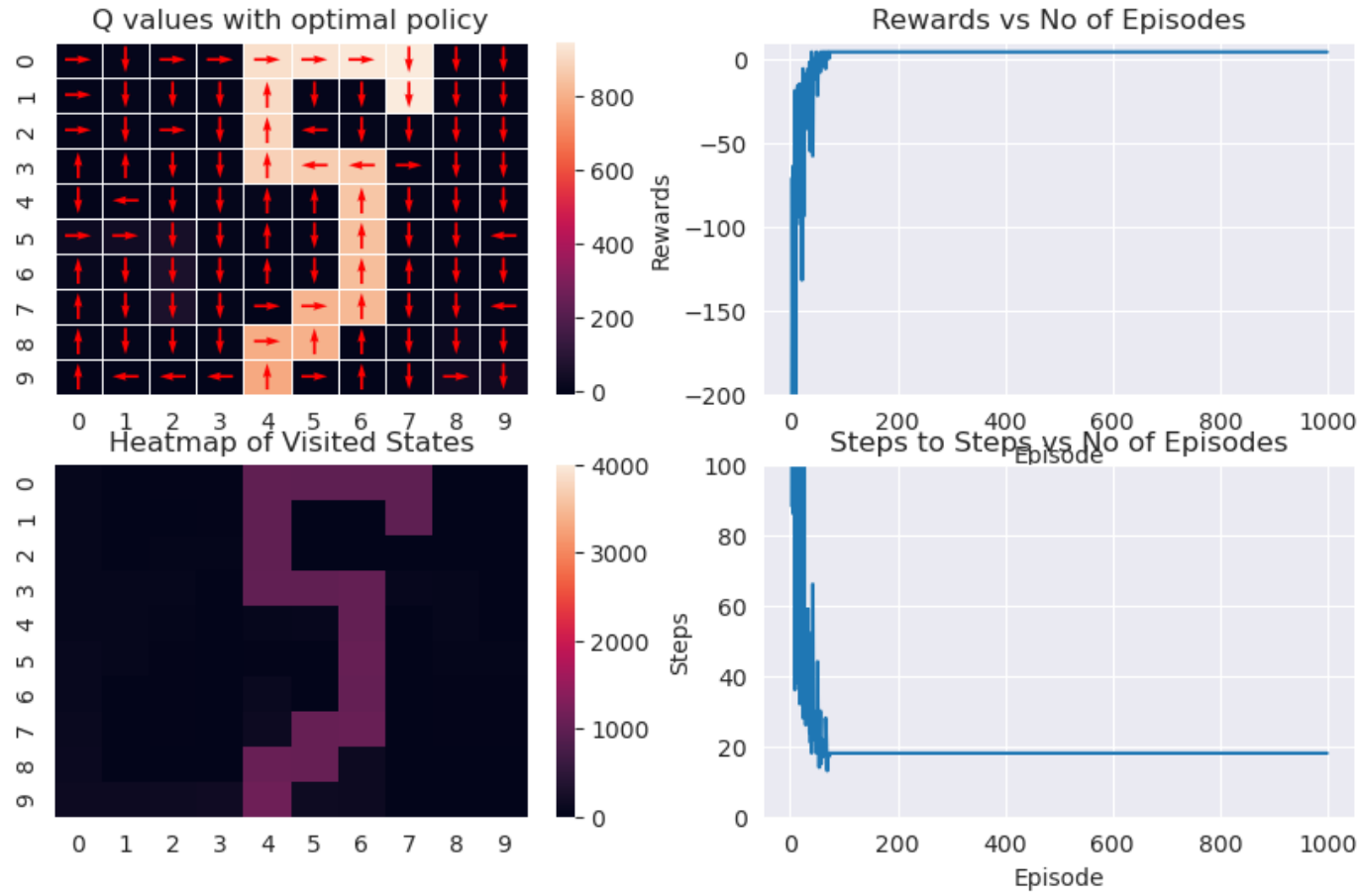
Configuration 5

Reward	Algorithm	Exploration Strategy	Wind	Start Coors	P	.	Gamma	Alpha	Exploration Param
4.0	qlearning	Softmax	False	[0 4]	1.0	.	0.99	0.3	0.0



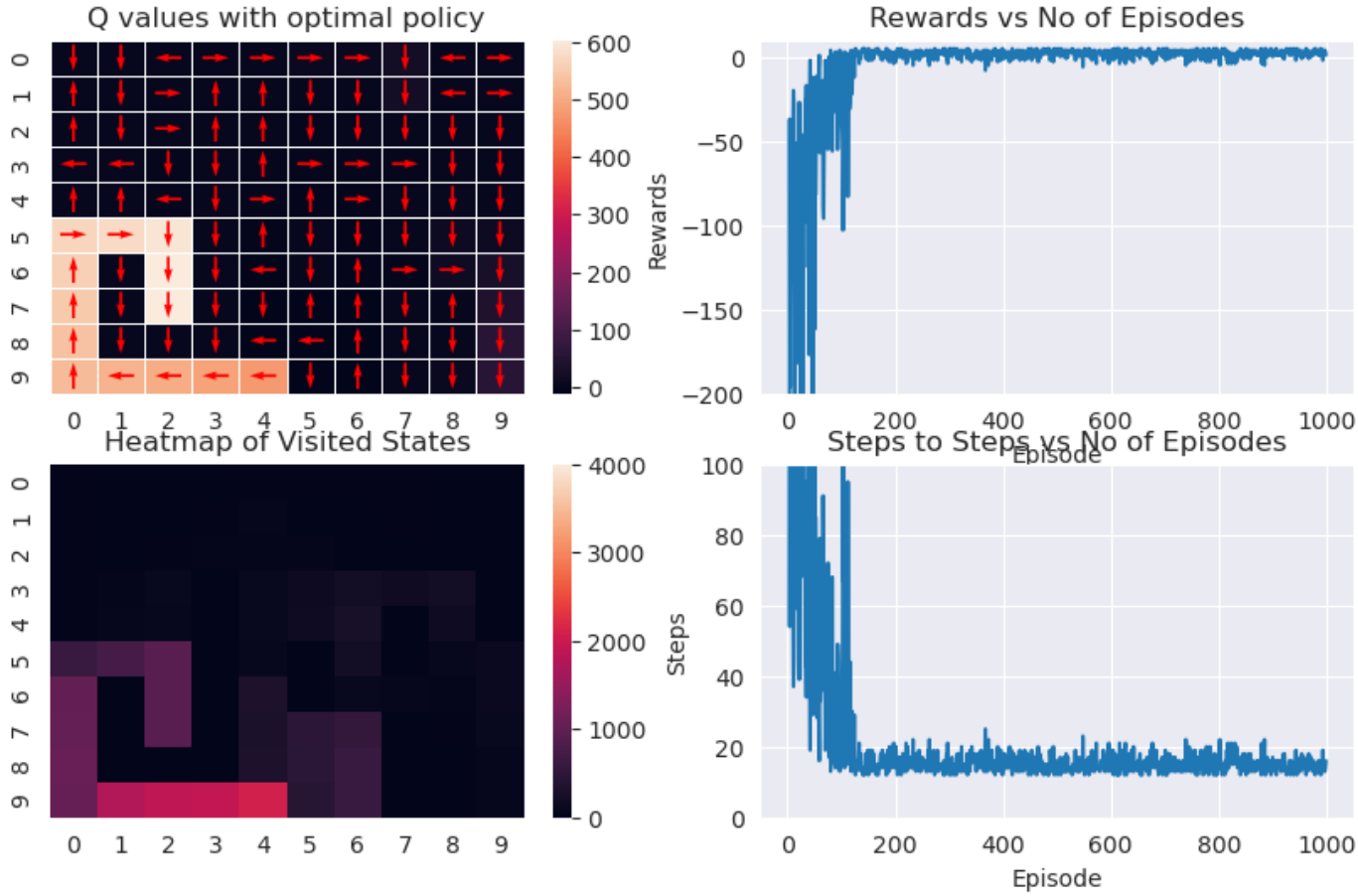
Configuration 6

Reward	Algorithm	Exploration Strategy	Wind	Start Coors	P	.	Gamma	Alpha	Exploration Param
4.0	qlearning	EpsilonGreedy	False	[0 4]	1.0	.	0.99	0.3	0.0



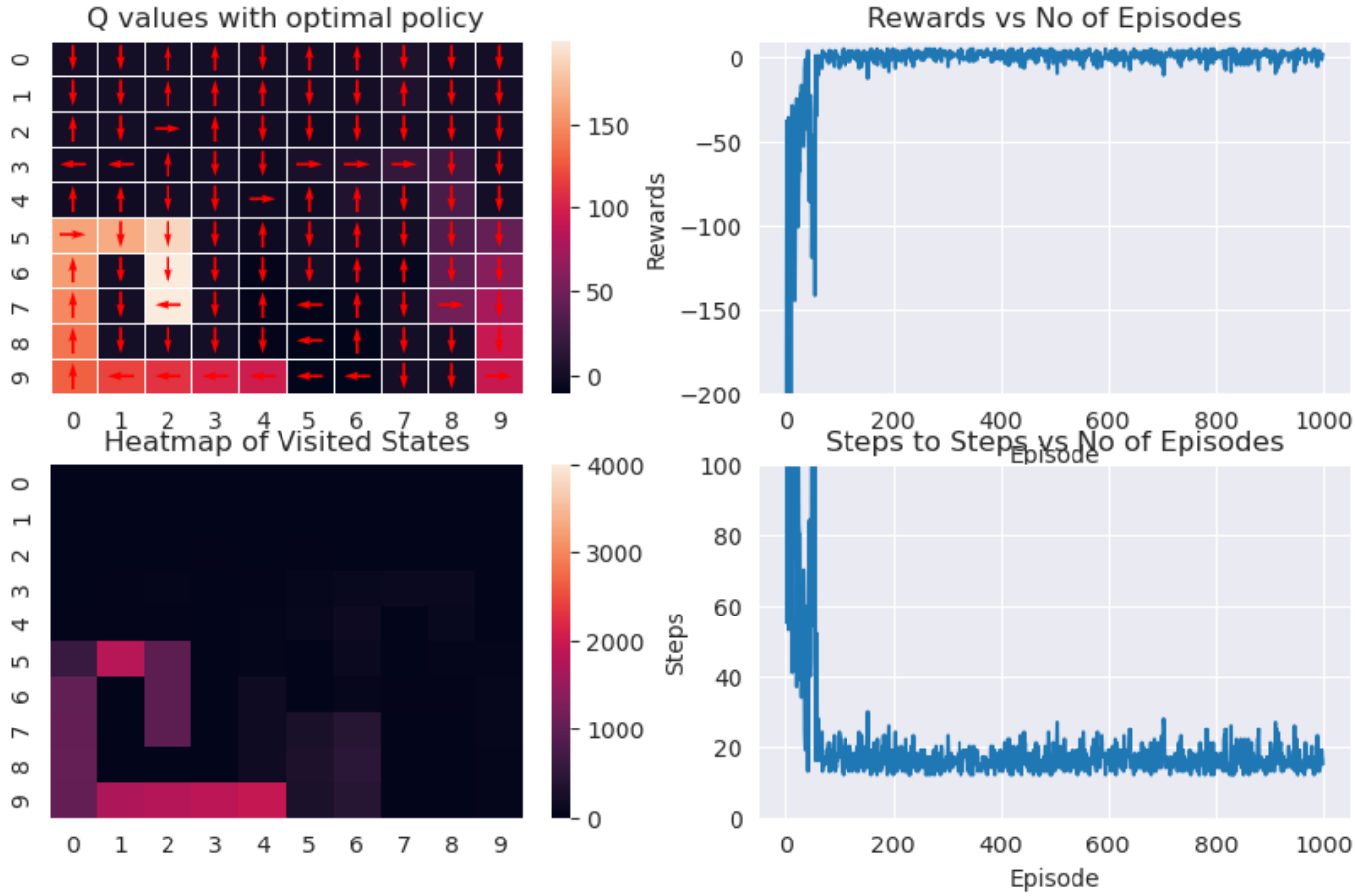
Configuration 7

Reward	Algorithm	Exploration Strategy	Wind	Start Coors	P	.	Gamma	Alpha	Exploration Param
5.0	qlearning	EpsilonGreedy	True	[0 4]	1.0	.	0.99	0.1	0.0



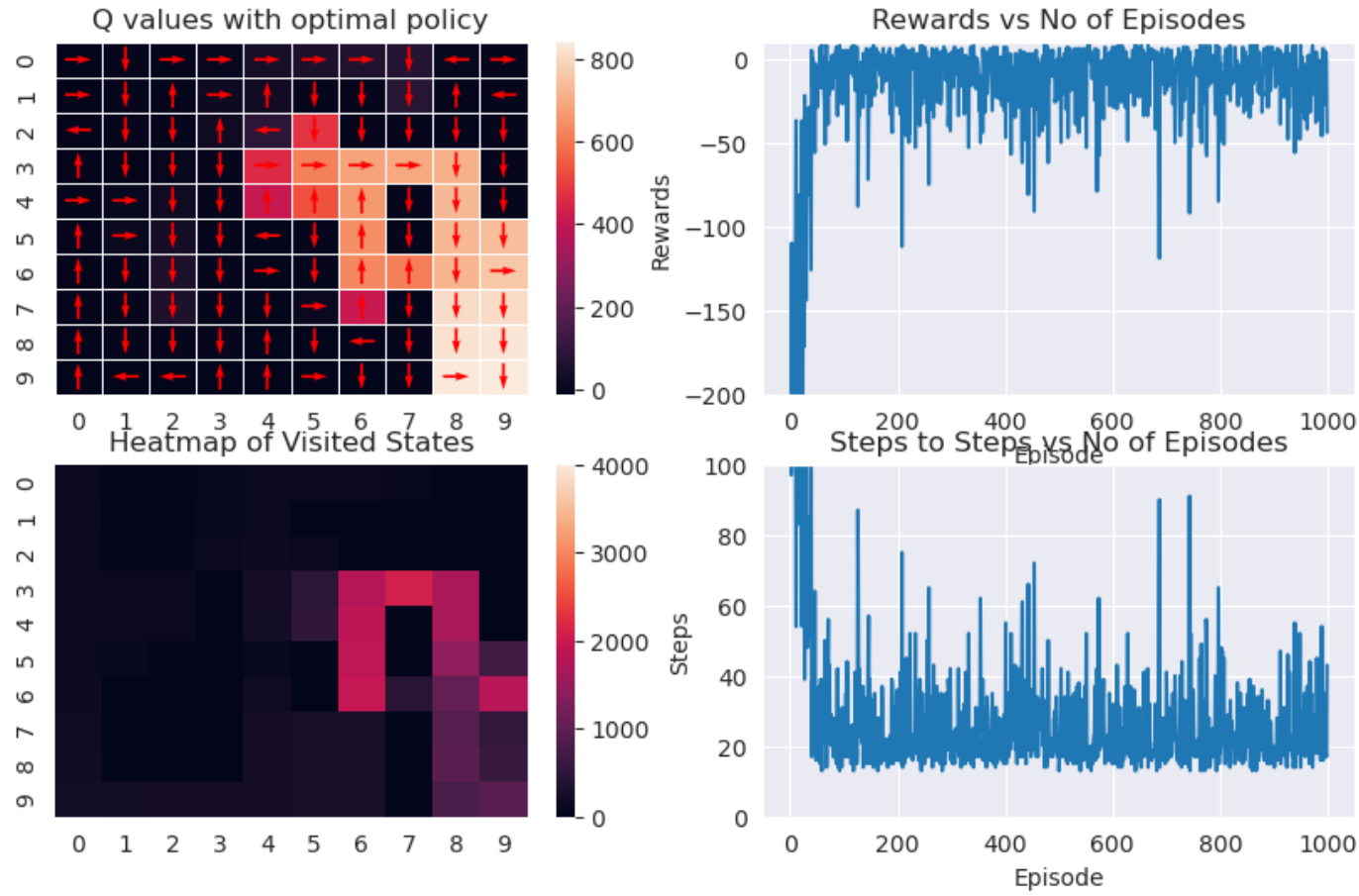
Configuration 8

Reward	Algorithm	Exploration Strategy	Wind	Start Coors	P	.	Gamma	Alpha	Exploration Param
5.0	qlearning	Softmax	True	[0 4]	1.0	.	0.95	0.3	1.0



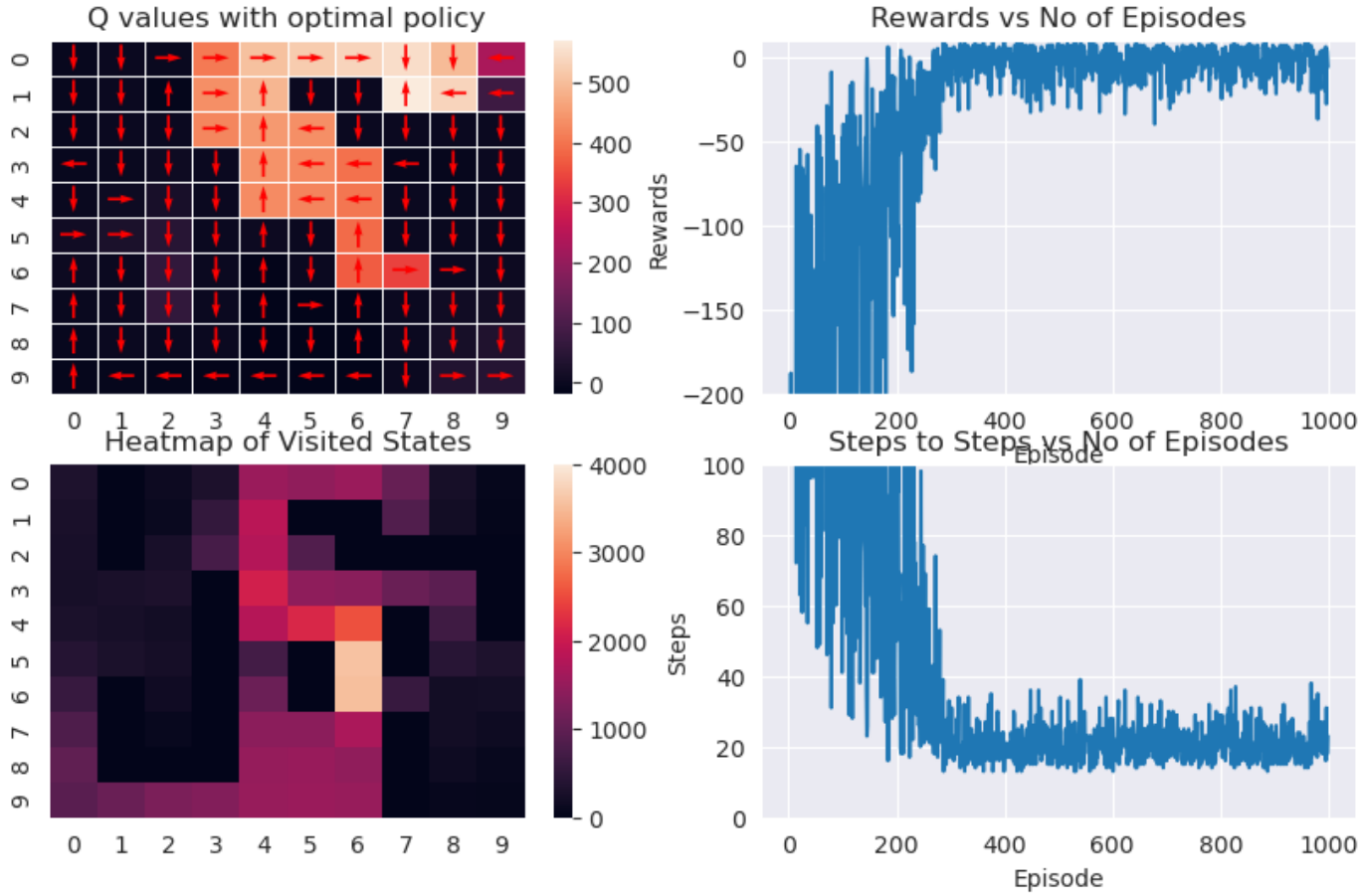
Configuration 9

Reward	Algorithm	Exploration Strategy	Wind	Start Coors	P	.	Gamma	Alpha	Exploration Param
8.0	qlearning	EpsilonGreedy	False	[3 6]	0.7	.	0.99	0.2	1.0



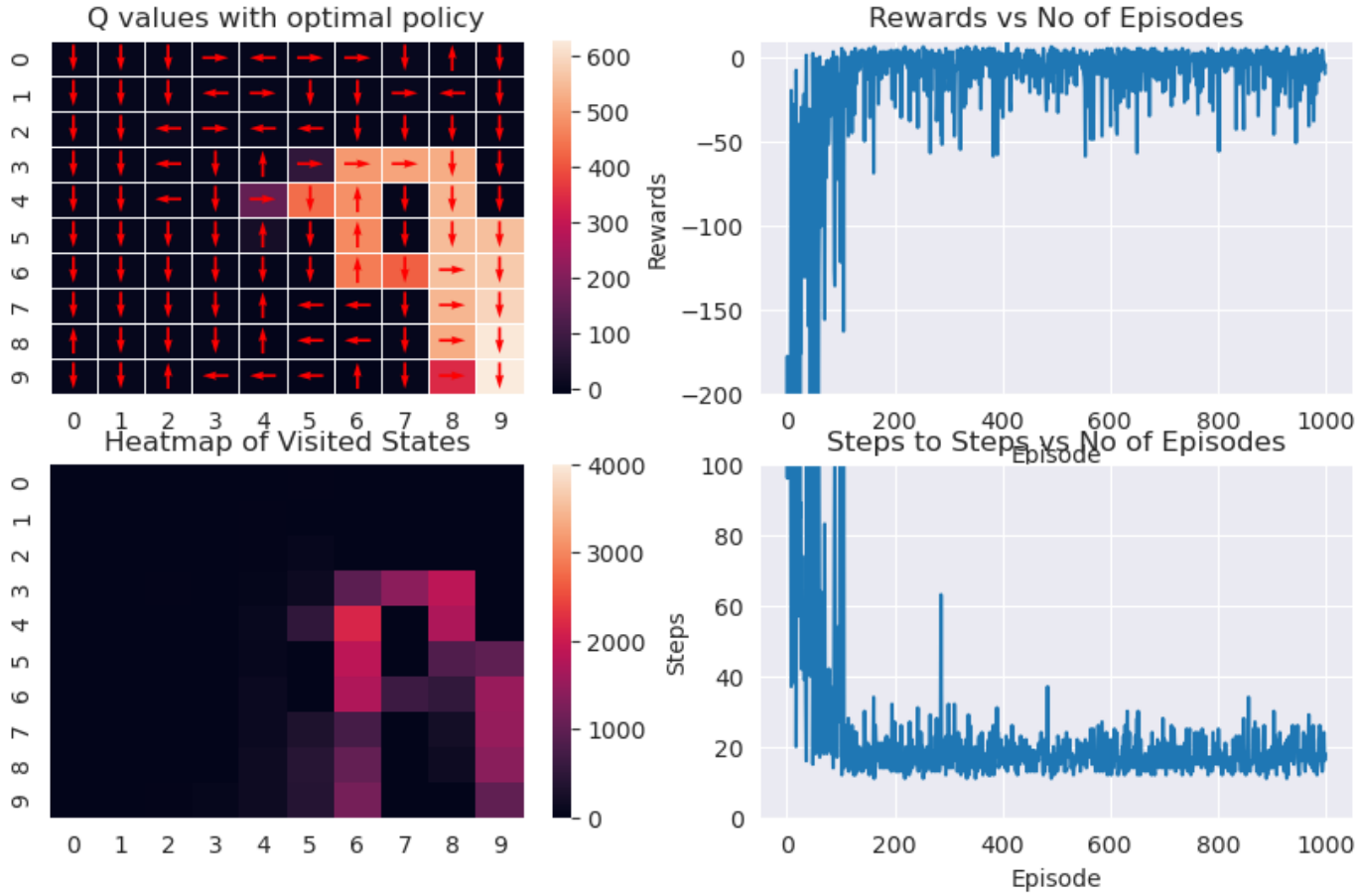
Configuration 10

Reward	Algorithm	Exploration Strategy	Wind	Start Coors	P	.	Gamma	Alpha	Exploration Param
9.0	qlearning	Softmax	False	[3 6]	0.7	.	0.99	0.1	4.0



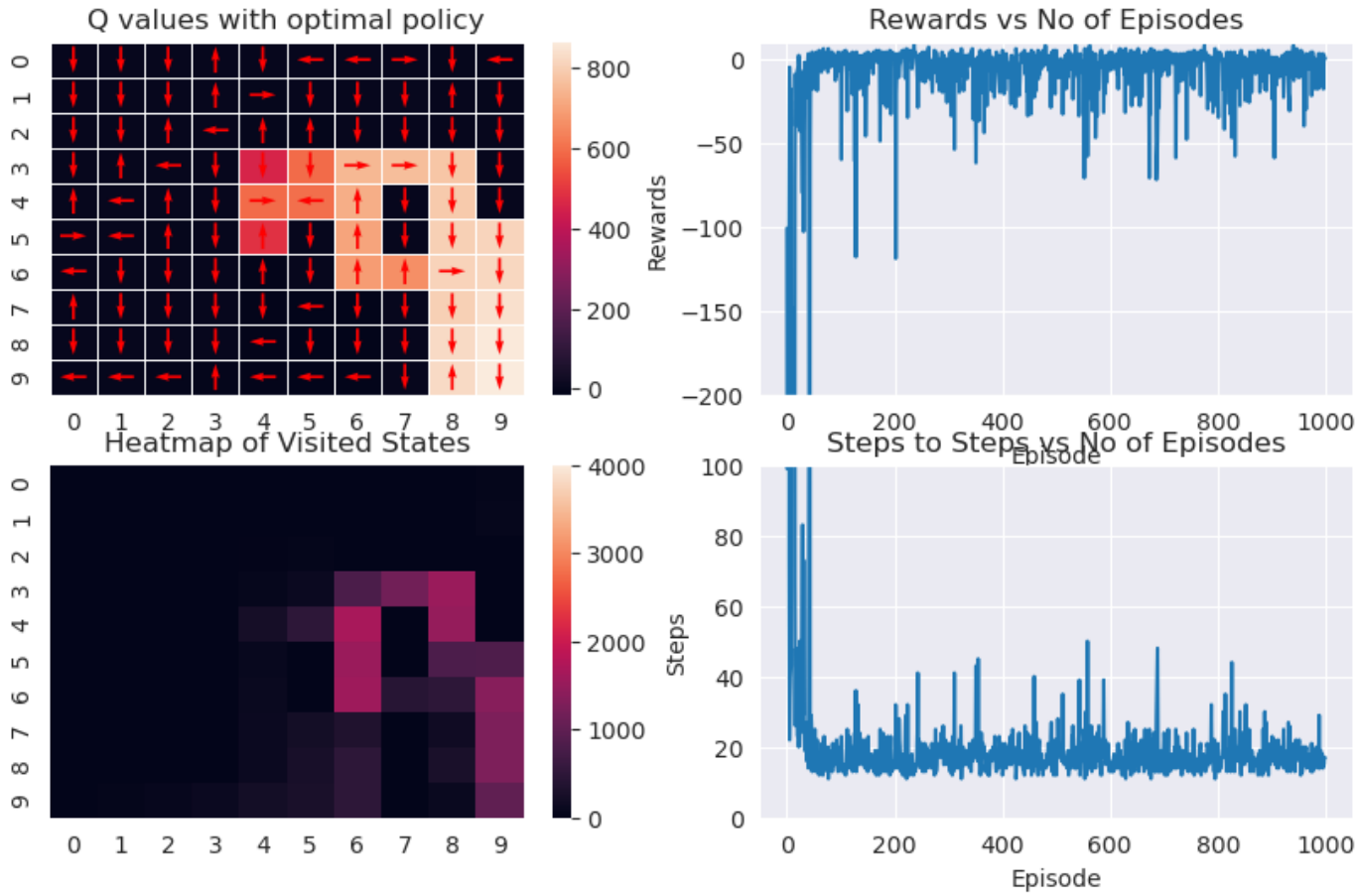
Configuration 11

Reward	Algorithm	Exploration Strategy	Wind	Start Coors	P	.	Gamma	Alpha	Exploration Param
5.0	qlearning	Softmax	True	[3 6]	0.7	.	0.99	0.1	3.0



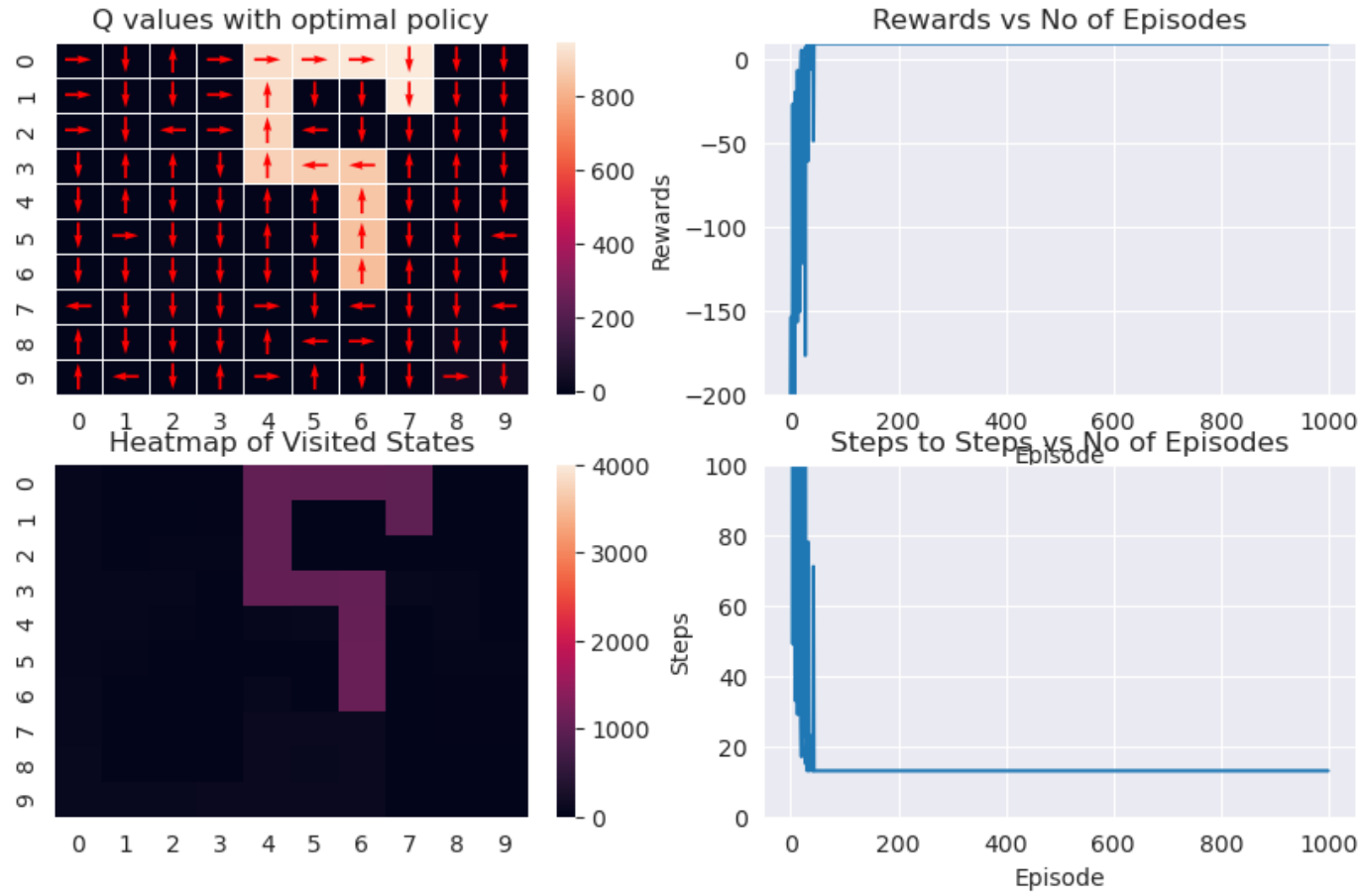
Configuration 12

Reward	Algorithm	Exploration Strategy	Wind	Start Coors	P	.	Gamma	Alpha	Exploration Param
10.0	qlearning	EpsilonGreedy	True	[3 6]	0.7	.	0.99	0.2	0.0



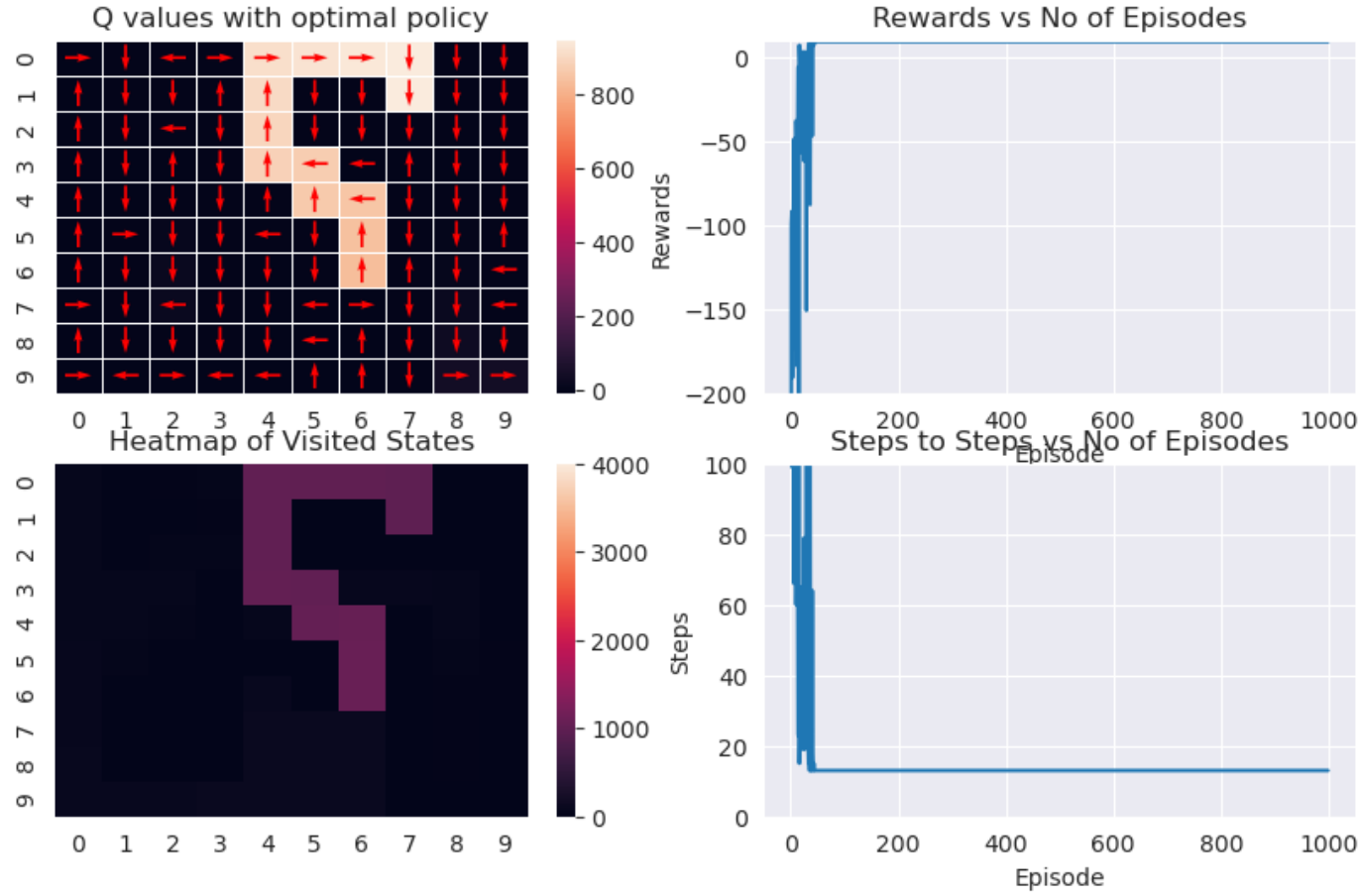
Configuration 13

Reward	Algorithm	Exploration Strategy	Wind	Start Coors	P	.	Gamma	Alpha	Exploration Param
9.0	qlearning	EpsilonGreedy	False	[3 6]	1.0	.	0.99	0.3	0.0



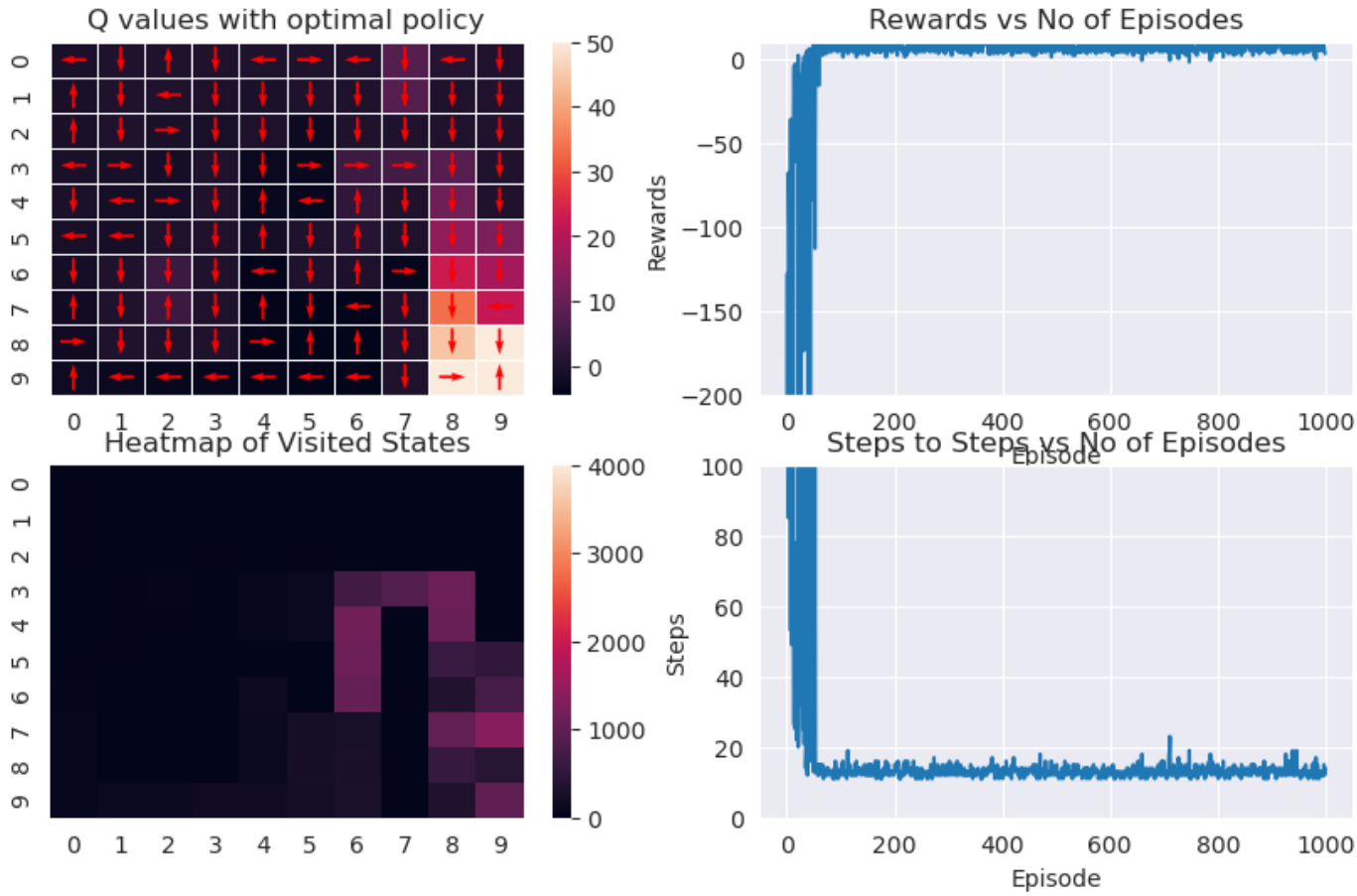
Configuration 14

Reward	Algorithm	Exploration Strategy	Wind	Start Coors	P	.	Gamma	Alpha	Exploration Param
9.0	qlearning	Softmax	False	[3 6]	1.0	.	0.99	0.3	0.0



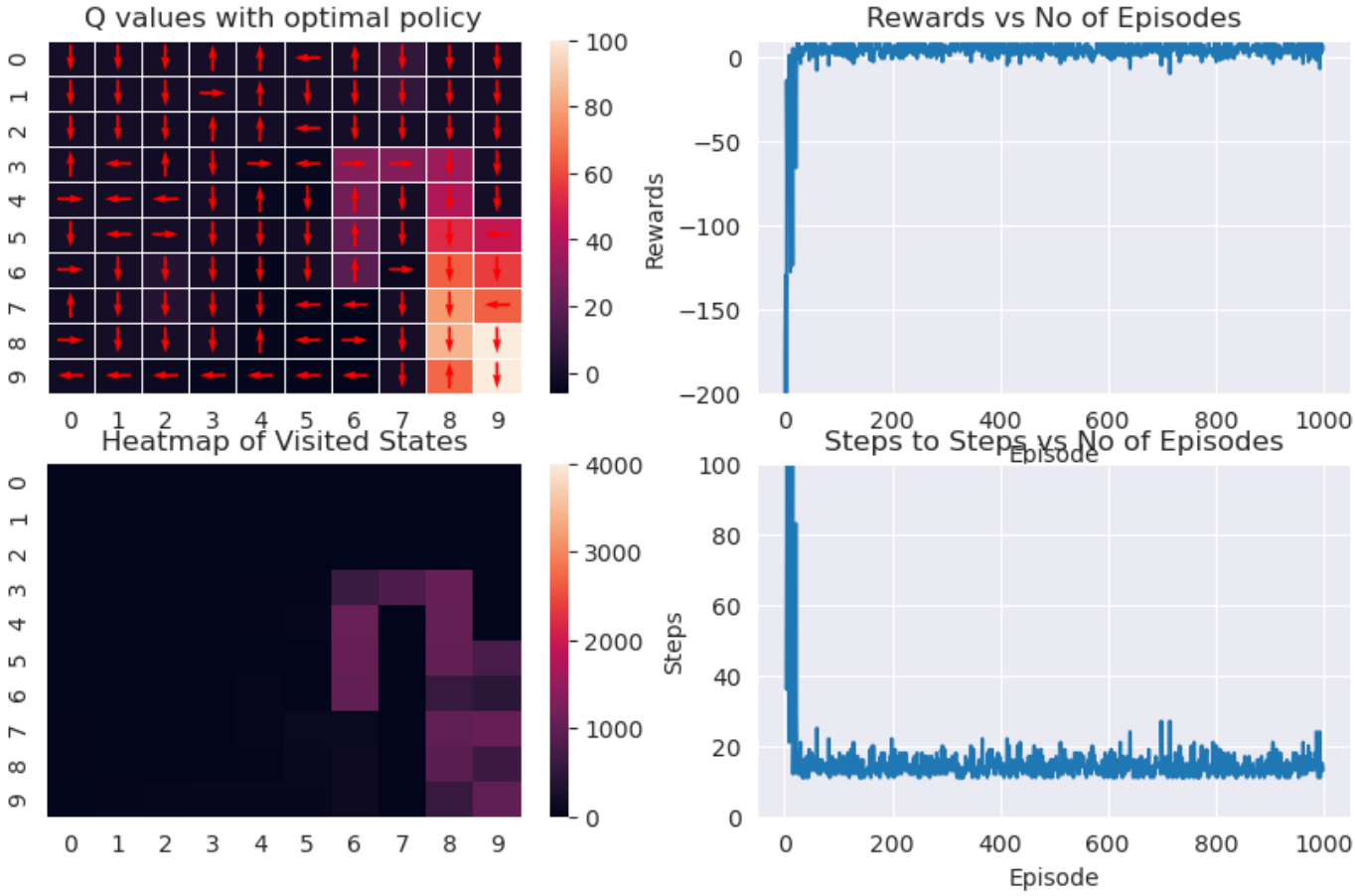
Configuration 15

Reward	Algorithm	Exploration Strategy	Wind	Start Coors	P	.	Gamma	Alpha	Exploration Param
9.0	qlearning	Softmax	True	[3 6]	1.0	.	0.8	0.2	0.0



Configuration 16

Reward	Algorithm	Exploration Strategy	Wind	Start Coors	P	.	Gamma	Alpha	Exploration Param
11.0	qlearning	EpsilonGreedy	True	[3 6]	1.0	.	0.9	0.4	0.0


Q values with optimal policy
Rewards vs No of Episodes
Rewards
Episode
Heatmap of Visited States
Steps to Steps vs No of Episodes
Steps
Episode

In [ ]: