Hw 6 report

Experiment 1-

Description-

In this experiment we have performed Qlearning using the E-greedy selection for selecting the actions. We have created a 10*10 grid in which the cans are placed randomly by the probability of .5 and robby is also placed randomly. Robby has to take all the cans. Aftre each move Q-table is updated.The experiment is done or 5000 episodes and each episode has 200 steps. After training the testing is done using the last updated Q-table.

Results-

Chart Title
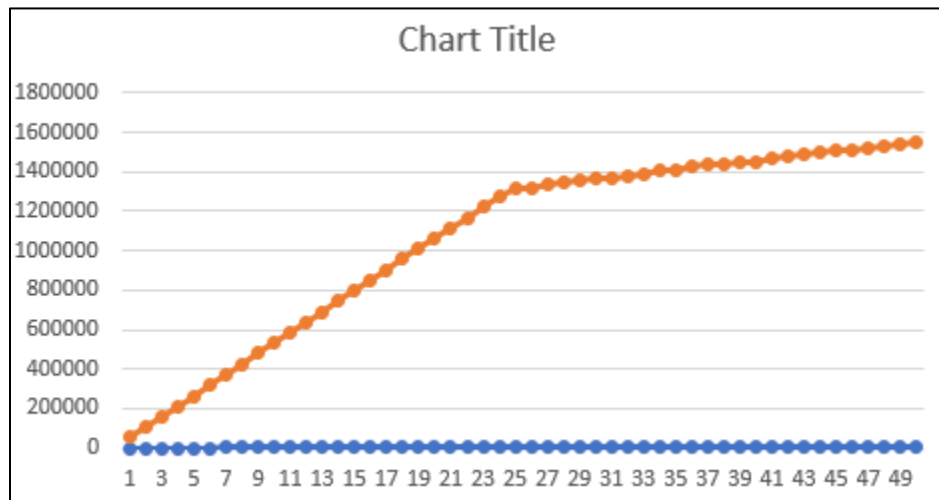
```
Mean:417
sd:1841064.2914309453
```
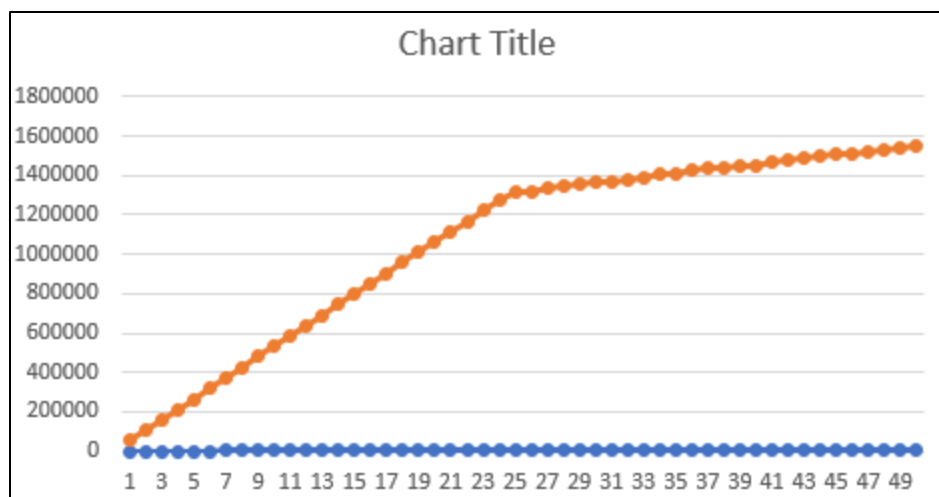
Experiment 2-

Description-

Done with 4 different learning rates- .20,.45,.65,.9

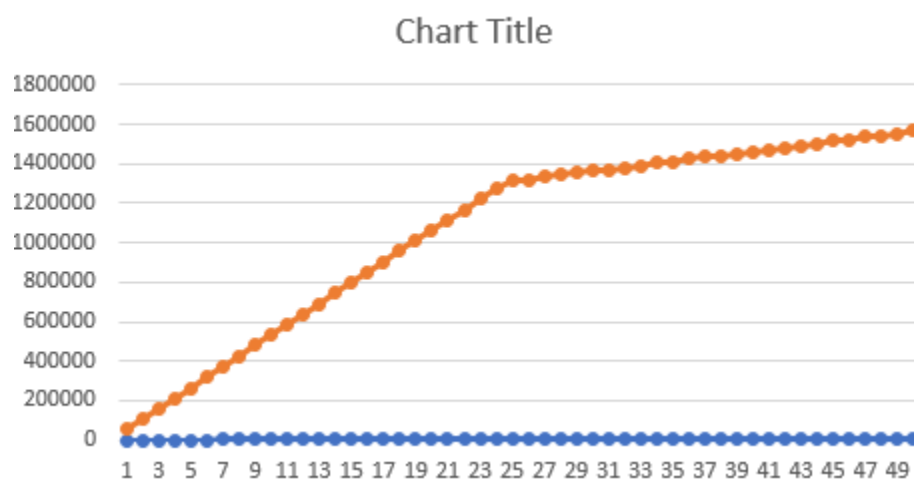Results-

With .20

Chart Title

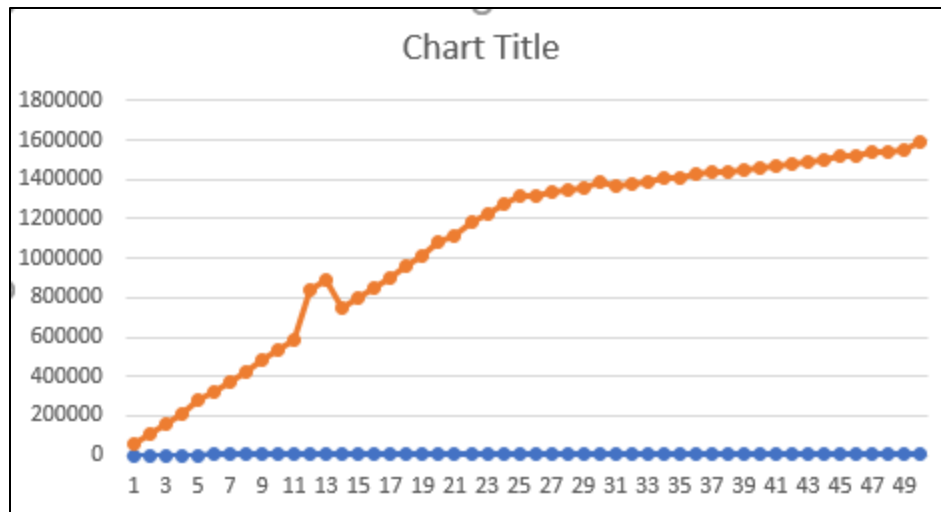Mean:417
sd:1841064.2914309453

With .45-

Chart Title

Mean:431
sd:1861842.2405606848

With .65-

Mean:428
sd:1881429.2200692766

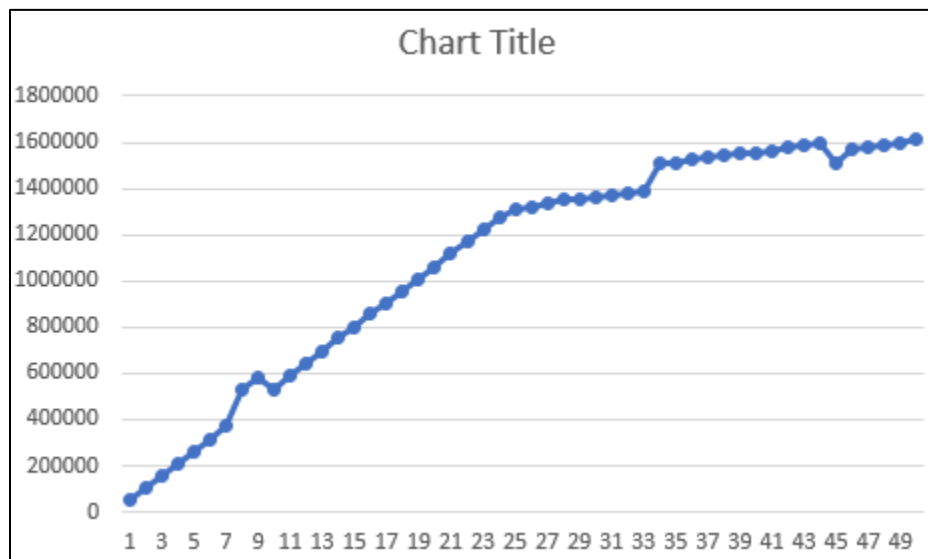With .9-

Mean:433
sd:1897040.5808423138

As the learning rate increase the reward also increases this can be seen from the difference in the mean and standard deviation of the different learning rates.

Experiment 3-

Description-

Epsilon is kept constant.

Results-

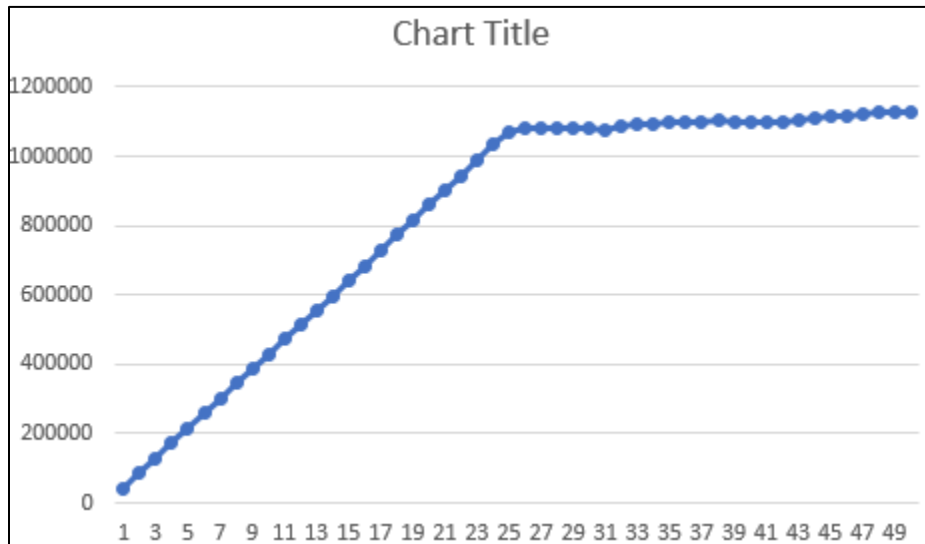
Mean:445
sd:1944551.4886369887

The mean and standard deviation have increased when epsilon is kept constant.

Experiment 4-

Description-

Negative reward is given for each action.

Results-

Chart Title

```
Mean:243.328
sd:1174476.9317023947
```

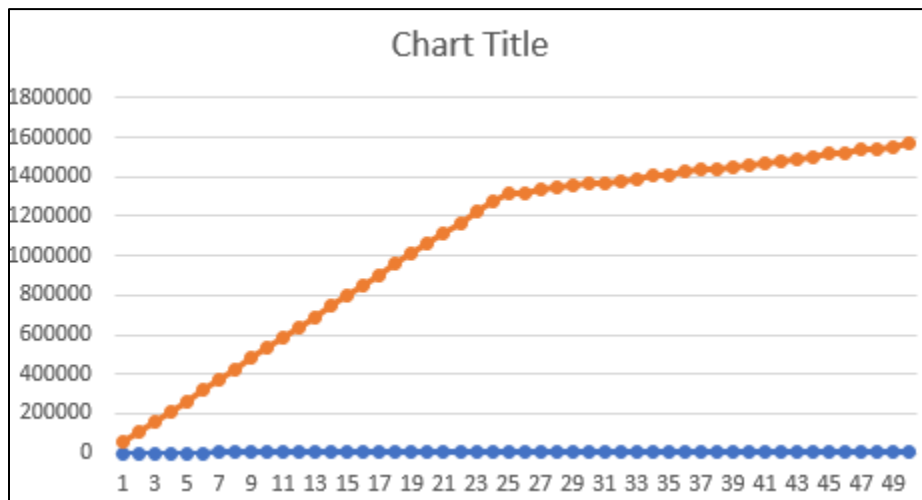The mean and standard deviation has decreased from the experiment 1.

Experiment 5

Description-

I have changed the gamma value.

Results-

Chart Title

```
Mean:438
sd:1904737.0066501454
```

Yes this is the expected result, the change in gamma has not affected the final values much because it is just the maximize the Q values.