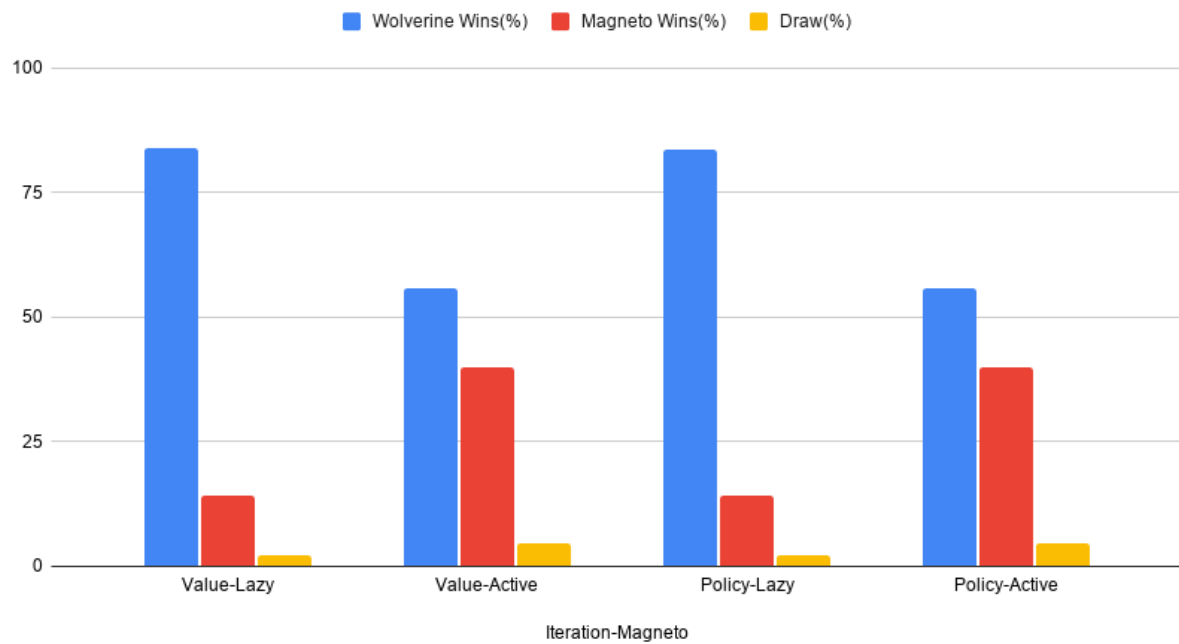
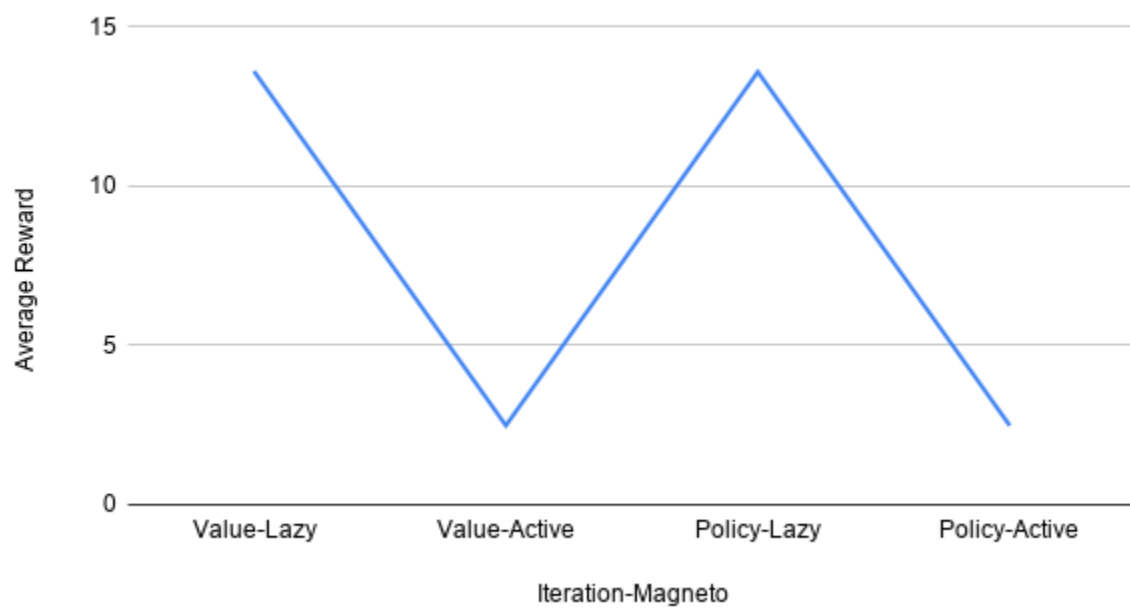


- All the observations were done using seed 0.
- All the games were run 100,000 times and results are averaged.

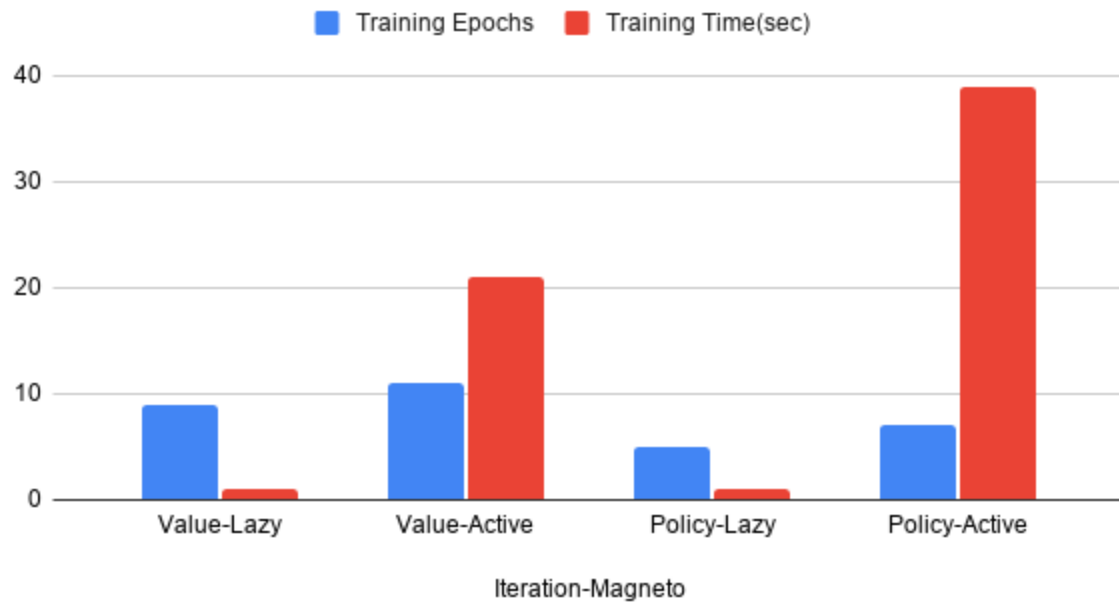
Wolverine Wins(%), Magneto Wins(%) and Draw(%)



Average Reward vs. Iteration-Magneto



Training Epochs and Training Time(sec)



Results

1. It can be observed that training times and epochs increases from lazy to active magneto
2. It can be observed that active magneto wins 25% more times than lazy magneto in both policy and value iterations
3. Also the average reward falls to % of that of lazy magneto when using active magneto