Value Iteration Value and Policy Grid:

Time horizon 50:

```
[[ 72812.59881759  78797.90774269  71486.30736303  64831.58181241
   58907.28069181  54312.92472837]
 [ 78797.90774269  85233.23989507  77371.8951247   70216.81426576
   63847.0769631   58907.28069181]
 [ 71479.98275387  75744.05889076  70175.89402699 -52984.33804168
   57807.44599329  53290.00172602]
 [ 64793.52315964  60421.23825241  63600.8375357   57616.48487904
   52288.97660167  48157.43979959]
 [ 58707.77576755 -61909.71474336  57616.48487904 -64476.02522282
   47266.28793307  43485.9892193 ]
 [ 53290.00172602  49620.32661636  52288.97660167  47266.28793307
   42794.89226961  39327.28050891]
 [ 49088.46277949  45675.40856084  48157.43979959  43485.9892193
   39327.28050891  36102.16141037]]
v v v < < <
> < < < < <
^ ^ ^ < ^ <
^ < ^ < ^ ^
^ ^ ^ ^ ^ ^
^ < ^ < < ^
^ ^ ^ ^ ^ ^
```

Time horizon 100:

```
[[ 79063.53656654  85048.8454913   77737.2451092   71082.51955133
   65158.21841607  60563.86243363]
 [ 85048.8454913   91484.1776435   83622.83287176  76467.75200879
   70098.01469795  65158.21841607]
 [ 77730.92048136  81994.99660094  76426.83172508 -46733.40042743
   64058.38345927  59540.93901227]
 [ 71044.46082592  66672.1758528   69851.775096    63867.42213531
   58539.91332515  54408.37588053]
 [ 64958.71329661 -55658.77738789  63867.42213531 -58225.08875116
   53517.22303537  49736.92267395]
 [ 59540.93901227  55871.26353976  58539.91332515  53517.22303537
   49045.82454914  45578.20939655]
 [ 55339.39977135  51926.34496126  54408.37588053  49736.92267395
   45578.20939655  42353.08481407]]
v v v v v v
> < < < < <
^ ^ ^ ^ ^ ^
^ < ^ < ^ ^
^ < ^ ^ ^ ^
^ < ^ < ^ <
^ ^ ^ ^ ^ ^
```

Policy Iteration Value and Policy Grid:

Time horizon 50:

```
[[ 72812.59881759  78797.90774269  71486.30736303  64831.58181241
   58907.28069181  54312.92472837]
 [ 78797.90774269  85233.23989507  77371.8951247   70216.81426576
   63847.0769631   58907.28069181]
 [ 71479.98275387  75744.05889076  70175.89402699 -52984.33804168
   57807.44599329  53290.00172602]
 [ 64793.52315964  60421.23825241  63600.8375357   57616.48487904
   52288.97660167  48157.43979959]
 [ 58707.77576755 -61909.71474336  57616.48487904 -64476.02522282
   47266.28793307  43485.9892193 ]
 [ 53290.00172602  49620.32661636  52288.97660167  47266.28793307
   42794.89226961  39327.28050891]
 [ 49088.46277949  45675.40856084  48157.43979959  43485.9892193
   39327.28050891  36102.16141037]]
v v < < < <
> < < < < <
^ ^ ^ ^ ^ <
^ < ^ < < ^
^ ^ ^ ^ ^ ^
^ < ^ < < ^
^ ^ ^ ^ ^ ^
```

Time horizon 100:

```
[[ 79063.53656654  85048.8454913   77737.2451092   71082.51955133
   65158.21841607  60563.86243363]
 [ 85048.8454913   91484.1776435   83622.83287176  76467.75200879
   70098.01469795  65158.21841607]
 [ 77730.92048136  81994.99660094  76426.83172508 -46733.40042743
   64058.38345927  59540.93901227]
 [ 71044.46082592  66672.1758528   69851.775096    63867.42213531
   58539.91332515  54408.37588053]
 [ 64958.71329661 -55658.77738789  63867.42213531 -58225.08875116
   53517.22303537  49736.92267395]
 [ 59540.93901227  55871.26353976  58539.91332515  53517.22303537
   49045.82454914  45578.20939655]
 [ 55339.39977135  51926.34496126  54408.37588053  49736.92267395
   45578.20939655  42353.08481407]]
v v v v < <
> < < < < <
^ ^ ^ ^ ^ <
^ < ^ < ^ <
^ < ^ ^ ^ ^
^ < ^ < ^ ^
^ ^ ^ < ^ <
```

Q-learning Value and Policy Grids

Run 1:

```
Value Grid:
9749.77 9492.72 4286.75 0.00 0.00 0.00
9973.84 0.00 9993.53 7330.24 0.00 0.00
9725.46 9389.76 9509.12 0.00 0.00 0.00
9392.88 9102.76 8978.78 8010.94 -1.12 -1.24
8818.80 0.00 8536.15 0.00 5459.40 566.80
8610.27 8330.10 8377.99 7021.78 6651.64 3827.41
8385.58 8182.70 8079.69 7494.76 7047.65 5876.12

Policy Grid:
v < v ^ ^ ^
> ^ < < ^ ^
^ < ^ ^ > <
^ < ^ < v <
^ ^ ^ ^ v <
^ < ^ v v <
^ ^ ^ < < <
```

Run 2:

```
Value Grid:
9732.64 9899.47 7626.59 6401.04 3089.13 657.43
9896.26 0.00 8939.45 8831.99 6094.48 592.98
9780.58 9998.13 8944.21 0.00 -1.79 -2.00
9522.77 9289.53 8894.60 6354.46 -2.35 -2.54
9272.12 0.00 8002.34 0.00 2329.70 -2.63
8917.94 8072.07 7983.64 7812.78 4562.44 3010.08
8557.31 8062.29 7906.46 7720.25 7542.48 3652.97

Policy Grid:
v v ^ v < <
> ^ v < < <
> ^ ^ ^ ^ ^
^ < ^ < < <
^ ^ ^ ^ v >
^ < < < v <
^ v ^ ^ < <
```

Run 3:

```
Value Grid:
9652.47 9824.81 6256.65 1780.37 421.13 0.00
9993.51 0.00 6721.13 4057.37 860.48 94.61
9588.29 9689.87 8088.60 0.00 3708.72 2907.31
9168.96 8941.91 7436.30 7017.76 6379.87 5883.40
9003.20 0.00 7745.00 0.00 6319.15 -1.24
8779.92 8203.86 7594.97 7209.94 6471.66 5127.73
8404.45 8016.79 7586.16 6969.00 5715.48 2068.97

Policy Grid:
v v < < v ^
> ^ v < v <
^ ^ v ^ v v
^ < v < < <
^ ^ ^ ^ ^ <
^ < v < < <
^ ^ ^ ^ ^ <
```

Run 4:

```
Value Grid:
9625.46 9960.94 -0.50 1187.51 353.26 -1.24
9969.59 0.00 9687.50 6117.88 758.65 -0.75
9732.04 9990.78 8165.36 0.00 2086.36 780.47
9465.77 9223.49 7646.19 4738.55 3838.70 1202.61
8769.81 0.00 7456.63 0.00 2039.82 2248.62
8184.97 8095.60 7397.35 7353.82 5983.02 4532.11
8081.71 8138.73 7703.33 7036.00 4377.49 3268.91

Policy Grid:
> v ^ v v v
> ^ < < v v
> ^ ^ ^ v <
^ < v < < <
^ ^ v ^ v <
^ < v v < <
^ < < < < ^
```

Run 5:

```
Value Grid:
9701.70 9708.54 6609.58 4375.85 -0.50 0.00
9998.43 0.00 8381.04 1793.26 -0.99 0.00
9529.38 9433.82 7328.86 0.00 915.85 -1.36
9483.89 7610.52 7207.68 7168.37 4486.78 1551.56
9267.84 0.00 7542.49 0.00 3054.55 -2.43
8997.94 8575.06 8115.11 7408.60 5904.42 4130.45
8677.09 8206.85 7824.88 7264.05 6864.54 5621.80

Policy Grid:
v ^ < < > >
> ^ < < ^ >
^ ^ v ^ v <
^ > ^ < < <
^ ^ v ^ v ^
^ < < < < v
^ ^ < < < <
```

Learning outcomes:

Student 1: I learned how to use Markov Decision processes efficiently and how to implement them while handling uncertainty that is inherent in every decision-making process in the real world. I also learned how the different algorithms discussed are built and what strengths and weaknesses they have in comparison. Value and policy iteration come to very similar conclusions even with value iteration being a simpler model, while an algorithm like q-learning requires more extensive training but is more versatile.

Student 2: With this project, I learned more deeply about q-learning. Whenever Dr. Harrison lectured about q-learning and explained it in a very simple way. For this assignment, I got to explore his simple explanation more in depth, and I learned what this algorithm actually does. I also learned more about programming bellman equations. I coded one in MA 417, but it was for a different type of problem (supply chain management).

Who-did-what:

Student 1: Policy evaluation and policy iteration functions.
Student 2: Value iteration, q-learning, and world initialization