

# *SIGMA - X*

*Analysis of Reinforcement Learning Trajectories in Cliff Walking Environment*

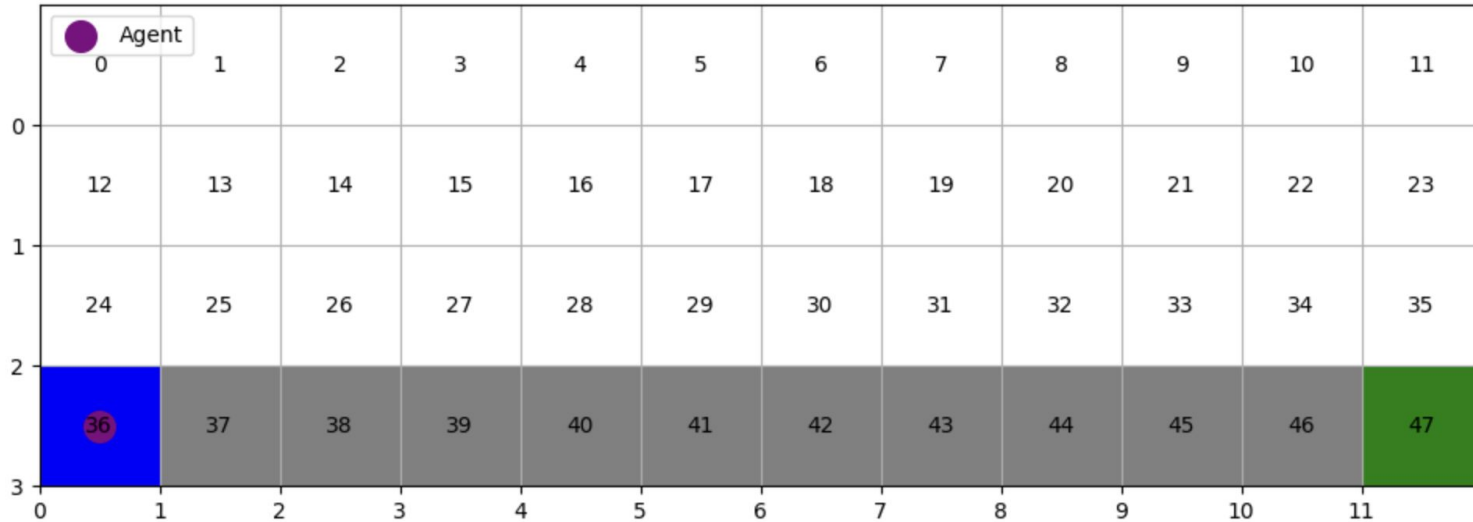
# *Dataset*

# General

- Research Objective: Analyze and compare trajectory patterns across multiple reinforcement learning algorithms
- Environment: Gym's Cliff Walking - 48 discrete states, 4 actions (up, down, left, right), with penalties for each step and cliff fall
- Algorithms Analyzed: SARSA, Q-learning, Expected SARSA, and a Random Policy

# Environment

Cliff Walking Grid World



# *Why this dataset?*

- *Uniform Start and Goal State*
- *Interpretability, we have 48 states which allows us to understand and analyze the behaviour of the agents*
- *Reinforcement Learning Path Convergence towards one path for a large number of episodes*

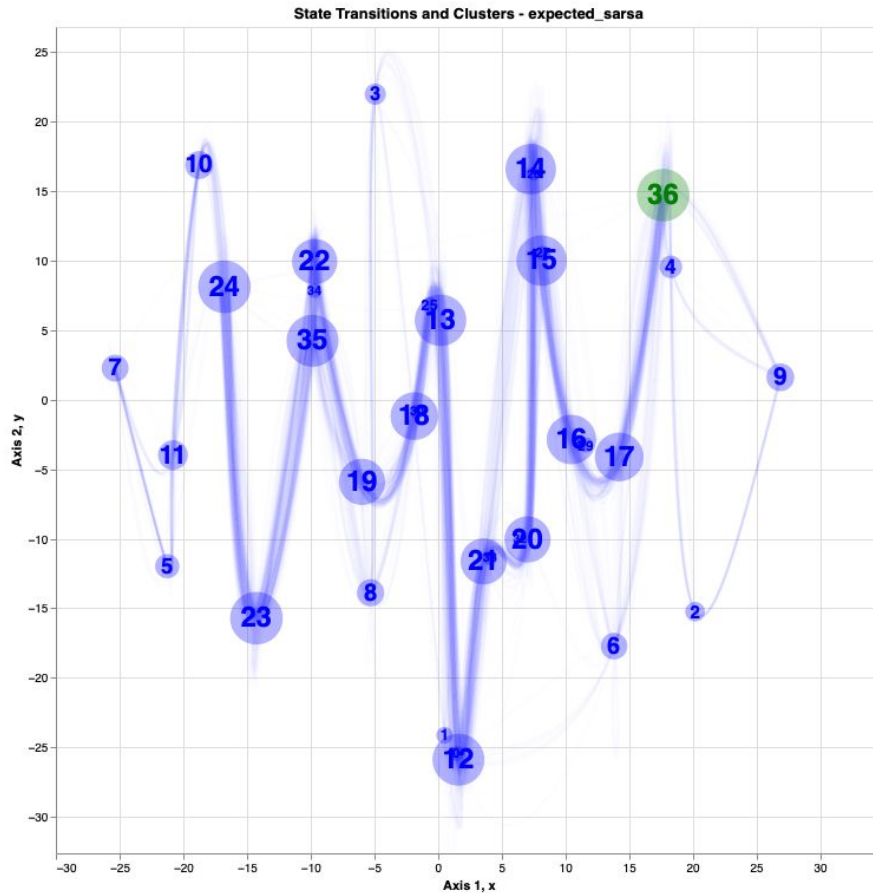
# *Projection Methods*

# Projection Methods

- *t-SNE*
- *PCA*
- *ICA*
- *UMAP - Yielded the best results*

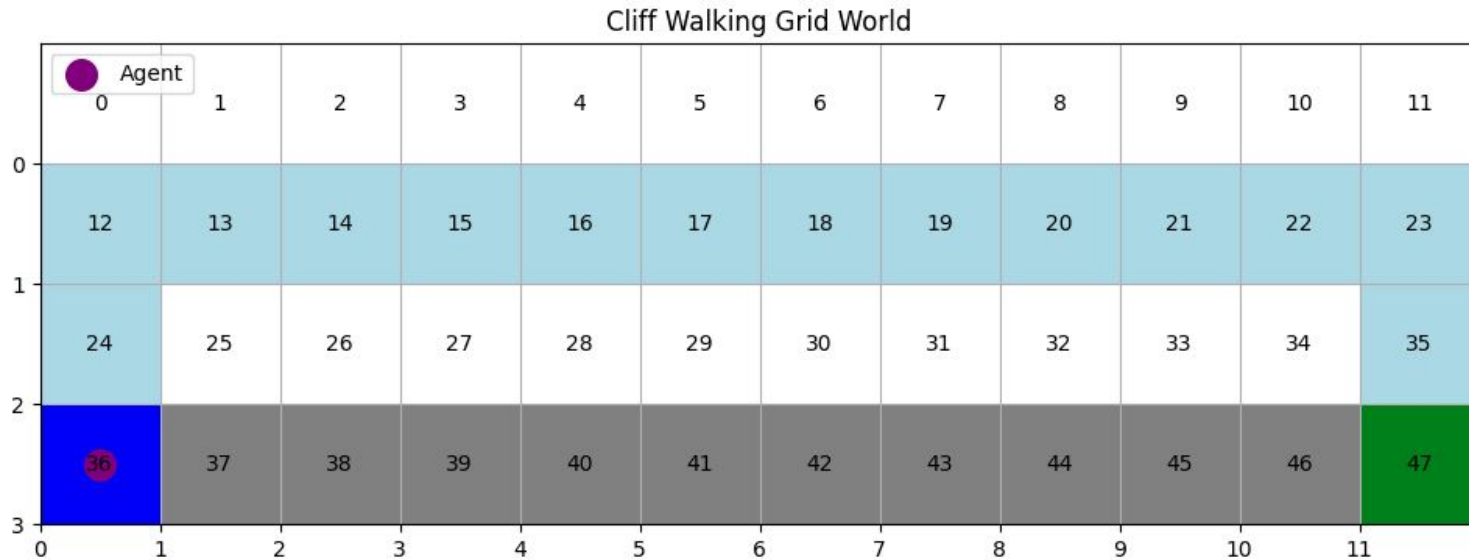
# *Analysis*

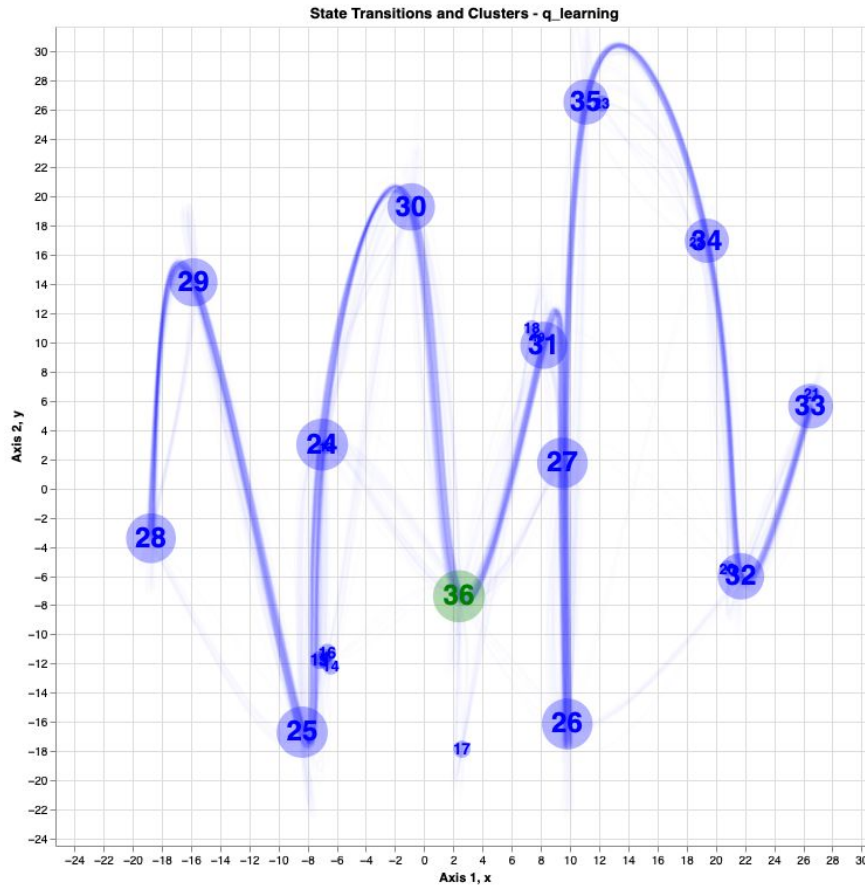




- *Clear convergence to the second most optimal policy*
- *Holds constant distance of 1 to the cliff where possible*
- *Relatively fast convergence*

# Environment - Correlation

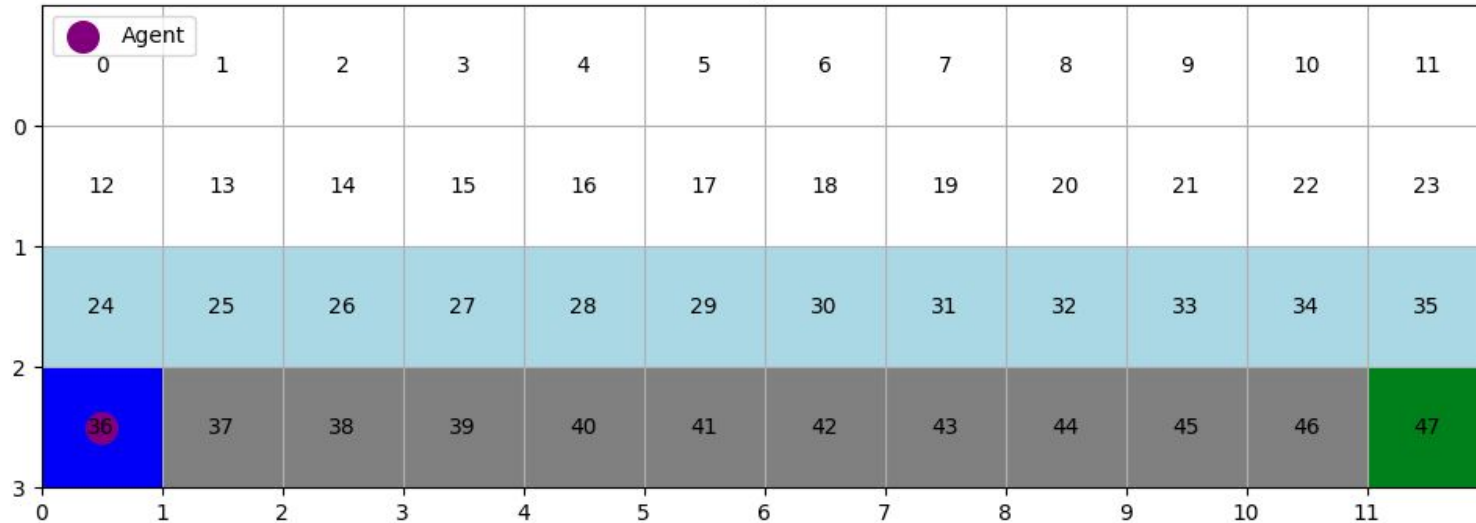


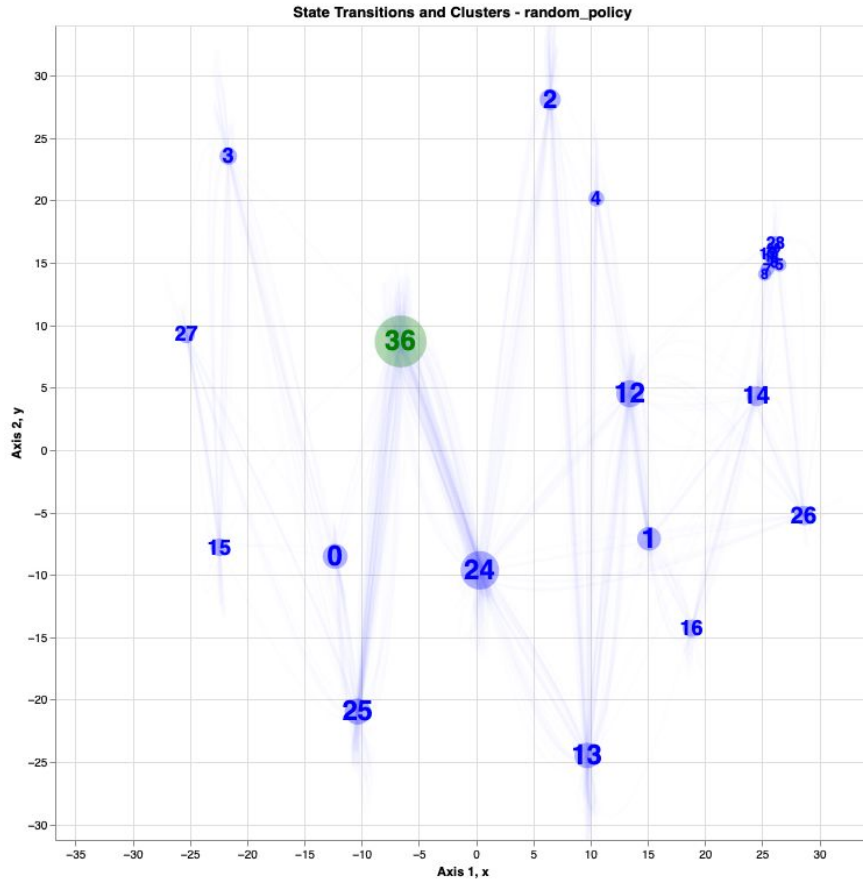


- *Clear convergence to the optimal policy*
- *Holds no distance to the cliff*
- *Extremely fast convergence*

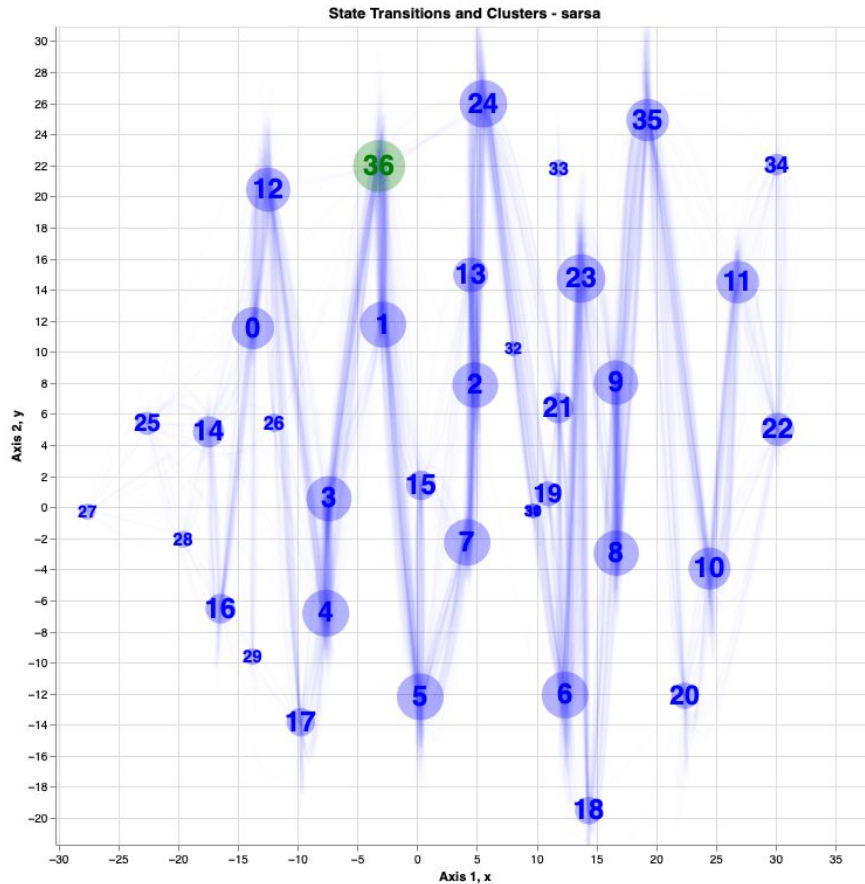
# Environment - Correlation

Cliff Walking Grid World





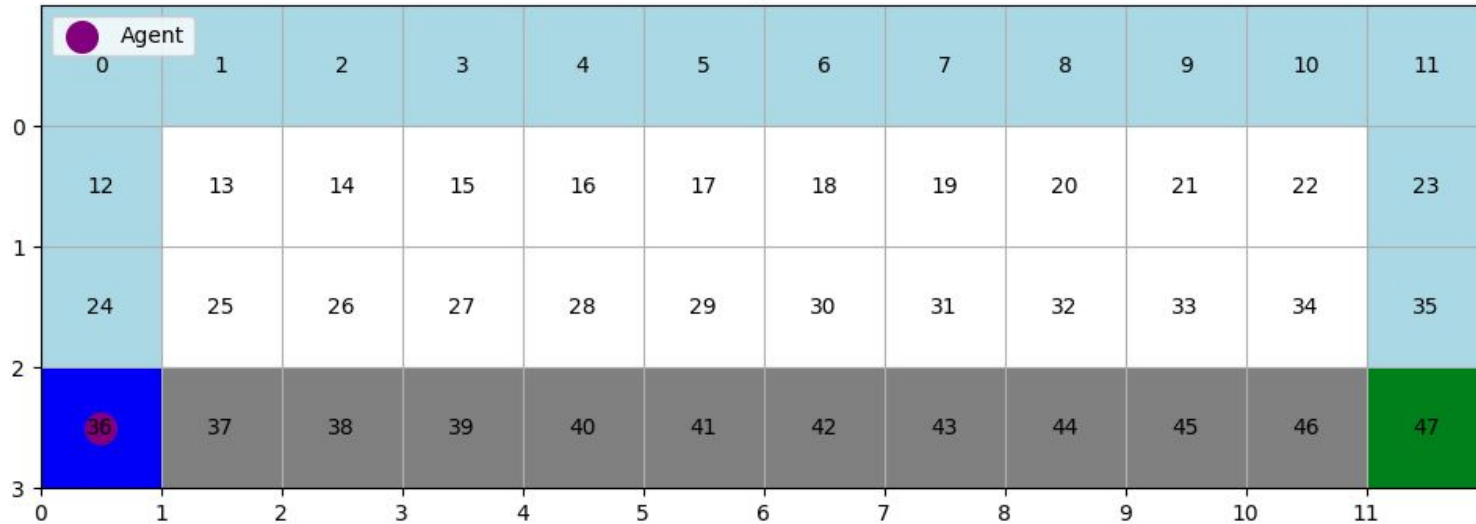
- *No convergence to a specific policy*
- *Random selection of states*
- *No clear convergence and significant entropy*

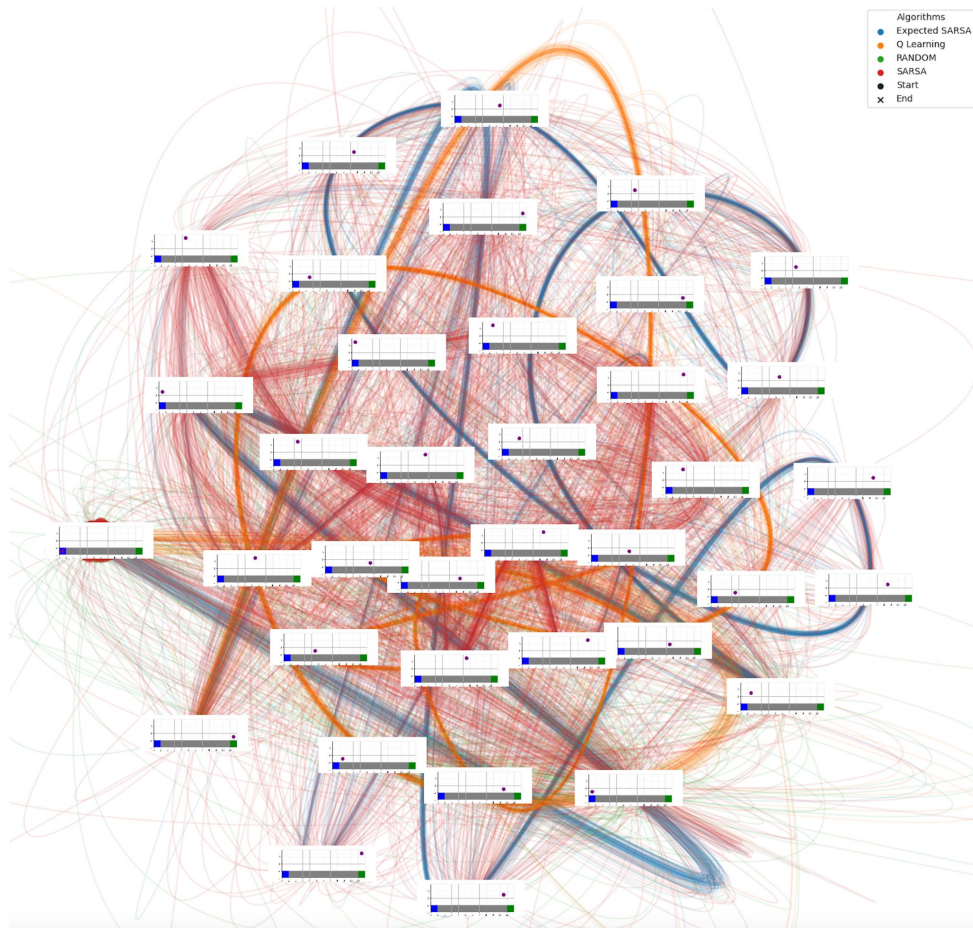


- *Clear convergence to the third most optimal policy*
- *Holds constant distance of 2 to the cliff where possible*
- *Relatively slow convergence*

# Environment - Correlation

Cliff Walking Grid World







# Summary

# Summary

- *Dataset worked very well for the visualization of learning trajectories*
- *Clear patterns were visible in the down-projections*
- *As expected, spatial information in the down-projection does not correlate with the actual spatial information in the data*

# *Questions and Answers*