

Welcome to [S&P 500 Time Series Data]

TEAM TUESDAY...

Muhammad Amjad
Helen Corbat
Jillian Howe
John Twomey
Nate Harris
Karina Mehta
Rameez Ali

November 13, 2025

10 year S&P 500



- Three related datasets describing the S&P 500 index and its constituent companies.
- S&P 500 is a market capitalization-weighted index of the 500 largest publicly listed companies trading in the United States.
- S&P 500 calculates a daily value based on the weighted average of its components and is rebalanced quarterly.
- S&P 500 generally sees ~5% annual turnover and ~33% turnover per decade.

Introduction

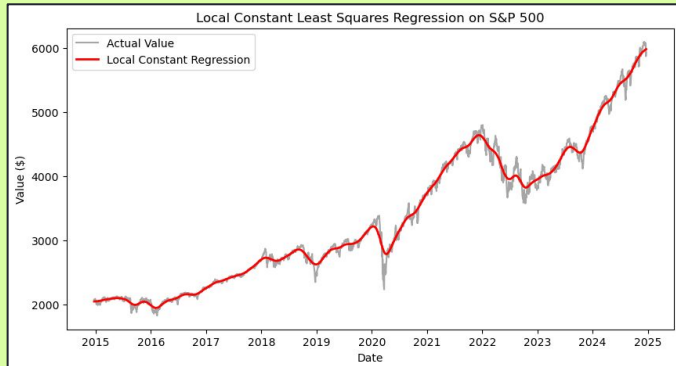
- Data from Kaggle, collected from Yahoo Finance and FRED.
- Version choice.
- Missing data.
- Data characteristics.

Data

LCLS

Also known as kernel regression:

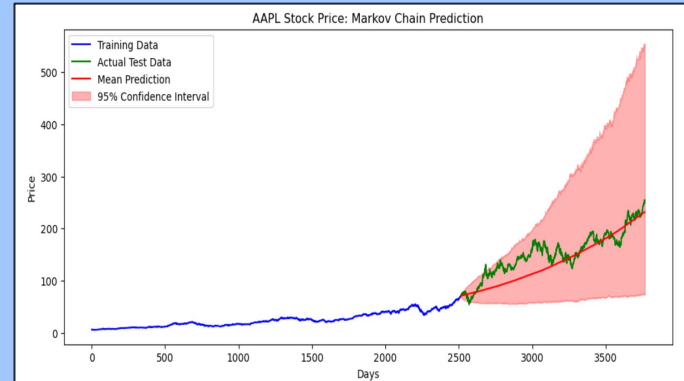
- Draws many small curves (one for each region, and stitch them together)
- Good for exploratory analysis
- Robust to nonlinearity
- Avoids overfitting large models

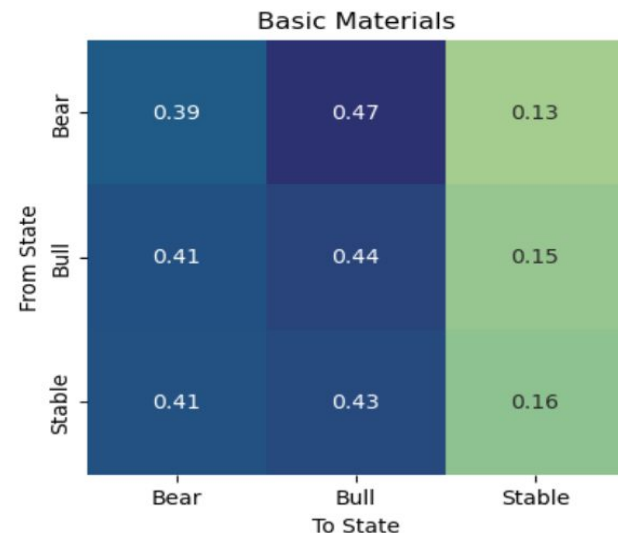
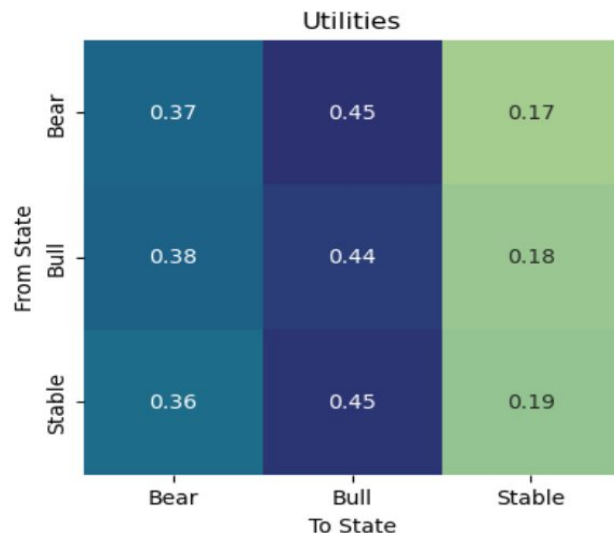
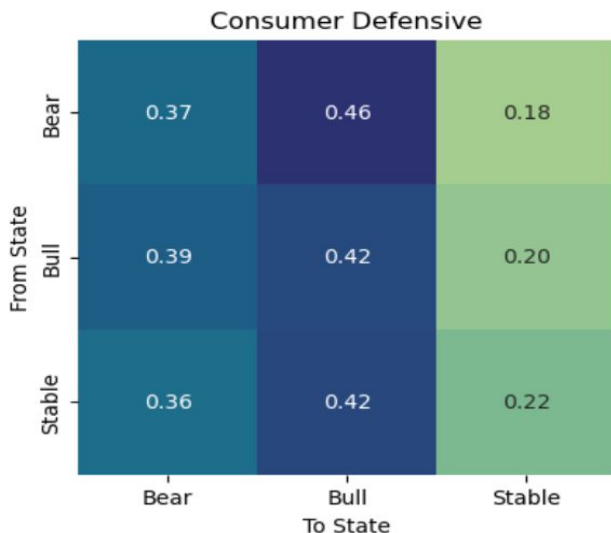
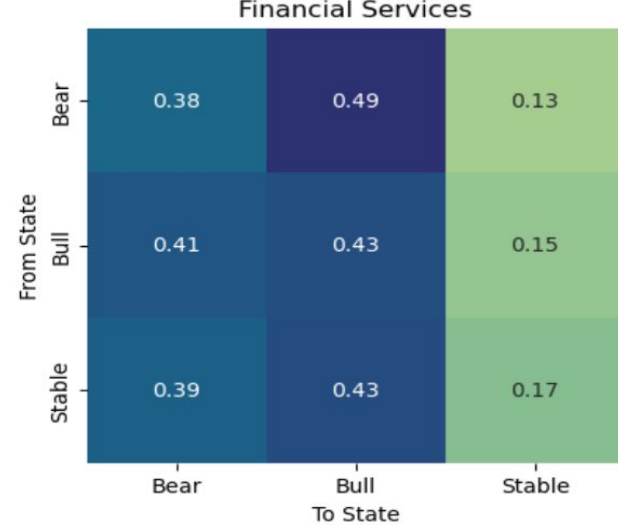
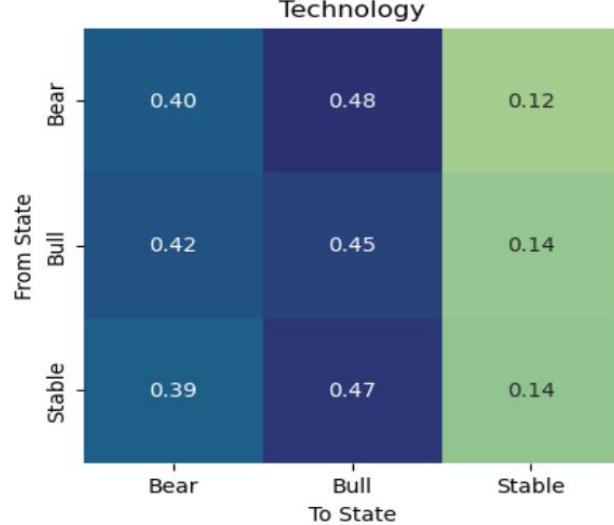
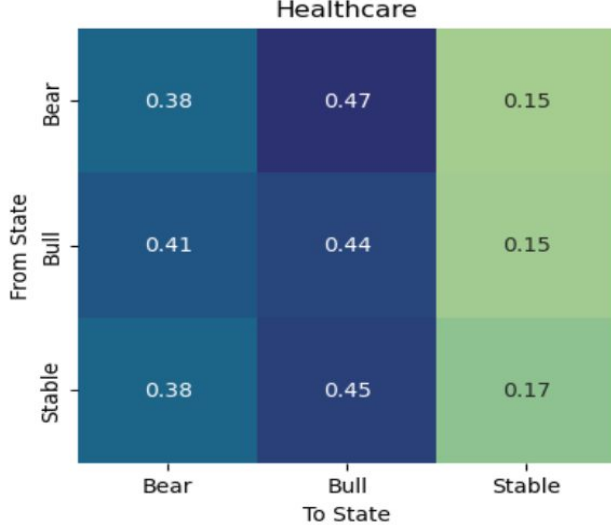


Markov Chain

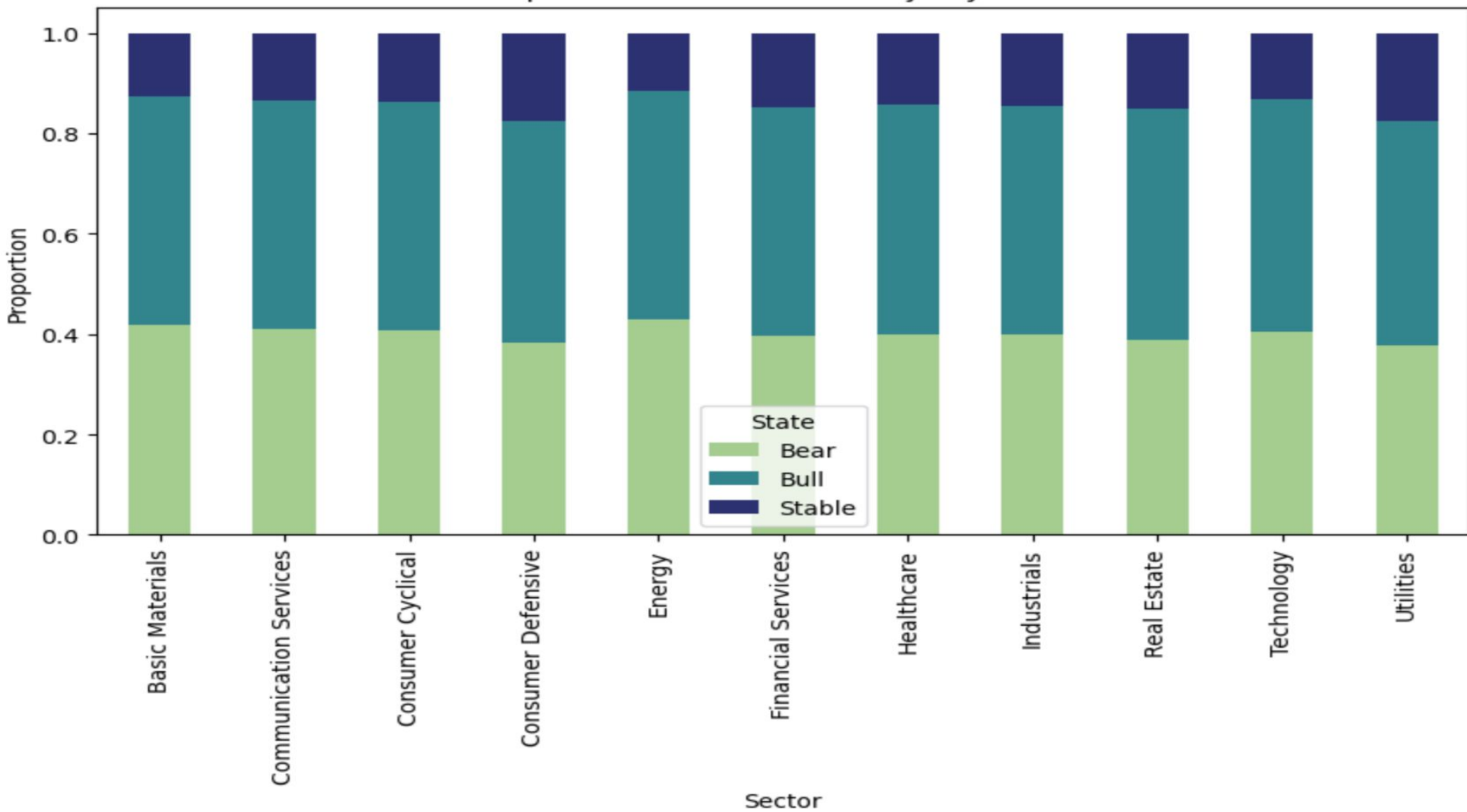
To model the likelihood of bearish, bullish, or stable days

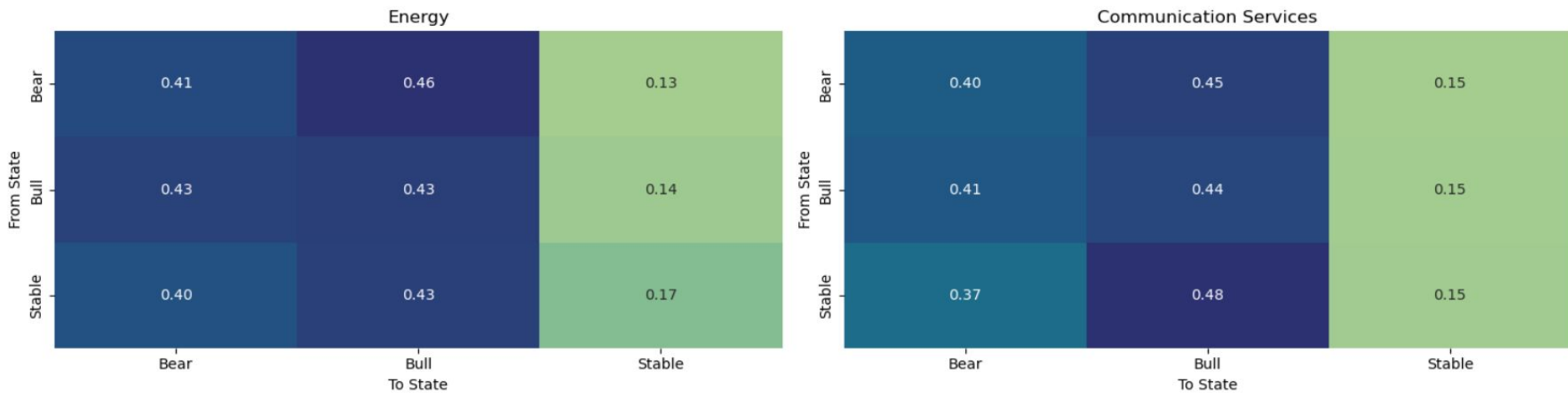
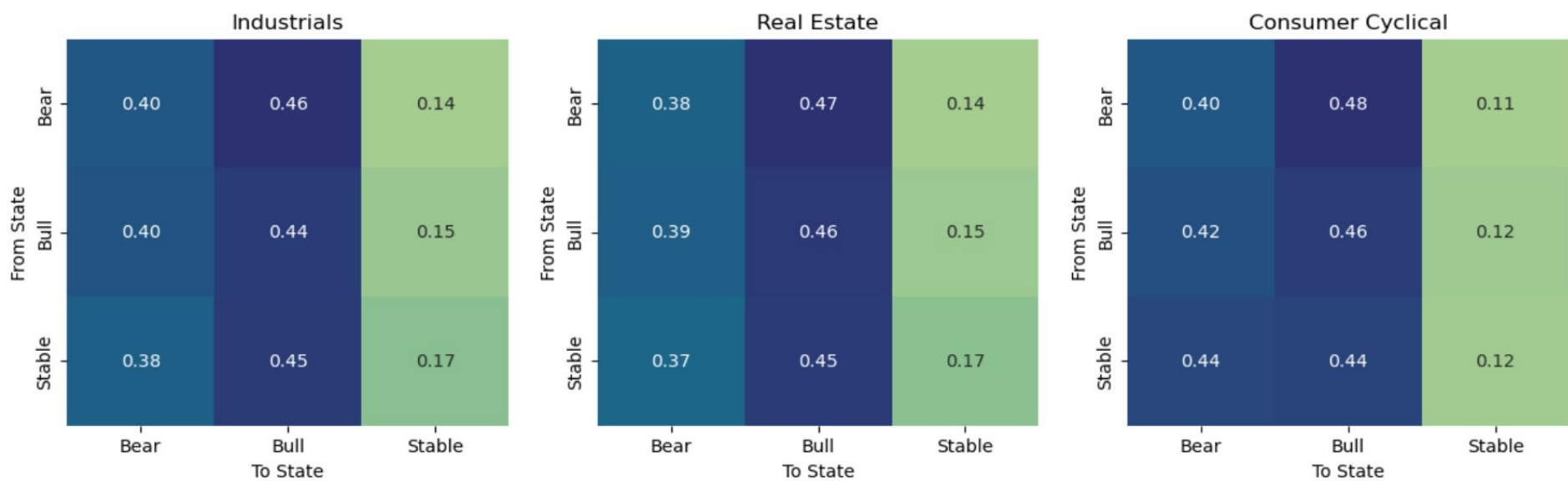
- **Bear:** negative percent change (market goes down)
- **Bull:** positive percent change (market goes up)
- **Stable:** percent change within a defined threshold





Proportion of Bull vs Bear Days by Sector





Sector: Healthcare

Simulated: {'Bull': 0.43929712460063897, 'Bear': 0.3985623003194888, 'Stable': 0.1621405750798722}

Actual : {'Bull': 0.45330431341511673, 'Bear': 0.43110407597942224, 'Stable': 0.11559161060546103}

Sector: Technology

Simulated: {'Bear': 0.444888178913738, 'Bull': 0.43610223642172524, 'Stable': 0.11900958466453675}

Actual : {'Bull': 0.4750329348621615, 'Bear': 0.42772383508172723, 'Stable': 0.09724323005611124}

Sector: Financial Services

Simulated: {'Bull': 0.48083067092651754, 'Bear': 0.35942492012779553, 'Stable': 0.1597444089456869}

Actual : {'Bull': 0.469231319441133, 'Bear': 0.4195198130751991, 'Stable': 0.11124886748366793}

Sector: Consumer Defensive

Simulated: {'Bull': 0.4416932907348243, 'Bear': 0.36741214057507987, 'Stable': 0.19089456869009586}

Actual : {'Bull': 0.4476177912626696, 'Bear': 0.4103950926719873, 'Stable': 0.1419871160653431}

Sector: Utilities

Simulated: {'Bull': 0.4416932907348243, 'Bear': 0.3769968051118211, 'Stable': 0.18130990415335463}

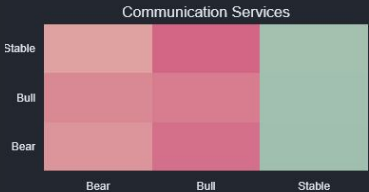
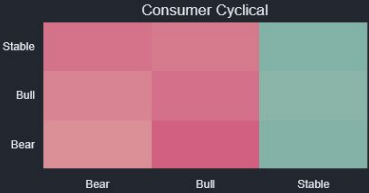
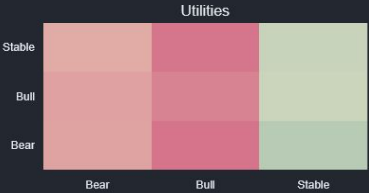
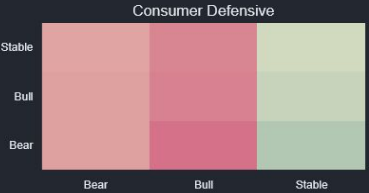
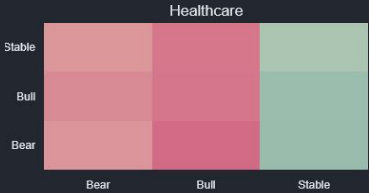
Actual : {'Bull': 0.4612385976766548, 'Bear': 0.41139322747472645, 'Stable': 0.12736817484861873}

Sector: Basic Materials

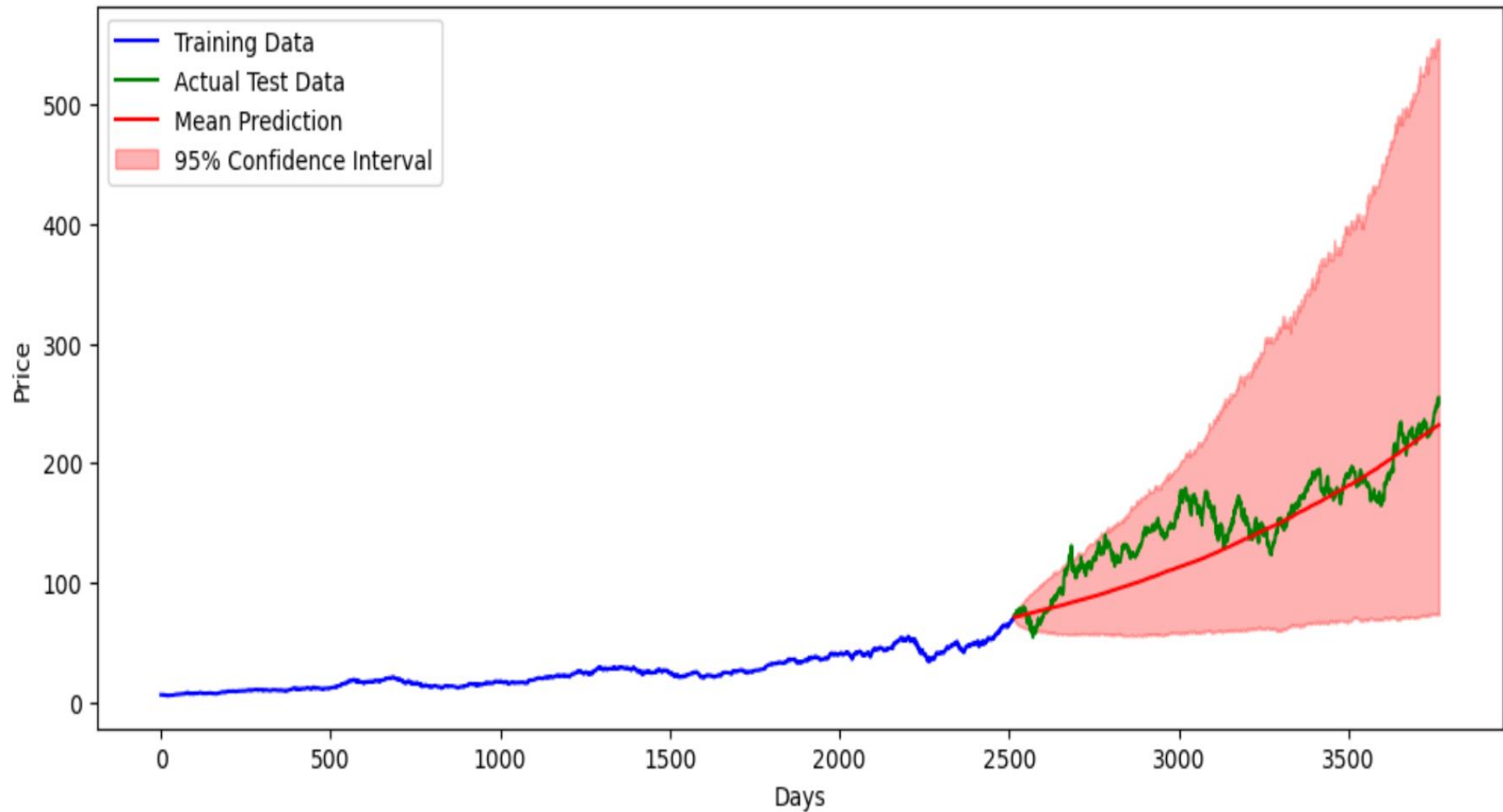
Simulated: {'Bull': 0.4289137380191693, 'Bear': 0.4057507987220447, 'Stable': 0.16533546325878595}

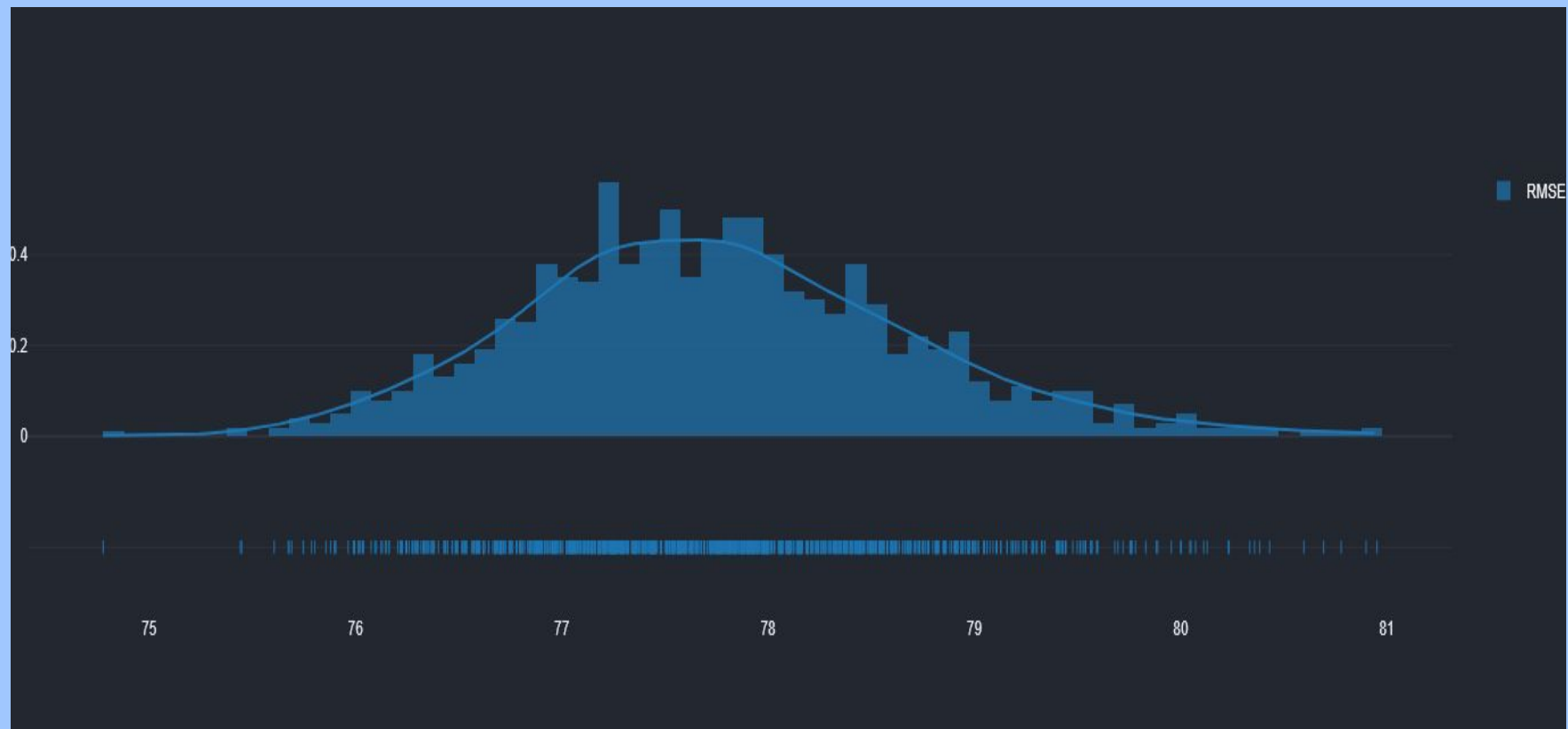
Actual : {'Bull': 0.46028173104850423, 'Bear': 0.43995062445541677, 'Stable': 0.099767644496079}

Sector Transition Matrices



AAPL Stock Price: Markov Chain Prediction





1. Data
2. What are we modeling?
3. Model description
4. Model
5. Issues
6. Conclusion

With 6 minutes, I would use this format:

1. Introduction: Describe the problem at a high level in intuitive and visceral terms (30 seconds)
2. Data: Describe the data and any missing data problems (1 minute)
3. Methods: What did you do? (1 minute)
4. Results: What did you learn? (2 minutes)
5. Conclusion: What was your main finding, and what would future work look like? (30 seconds)