Results…

Linear regression results:

Train MSE	Test MSE	Train R2	Test R2
0	0.1504	1	0.9002

This model has high predictive power, meaning that the linear relationships between the features and the target variable (prices) are captured effectively. It generalizes well, making it a strong contender for prediction. The model's limitation is that it assumes a linear relationship, which may miss non-linear price movements.

Random forest results:

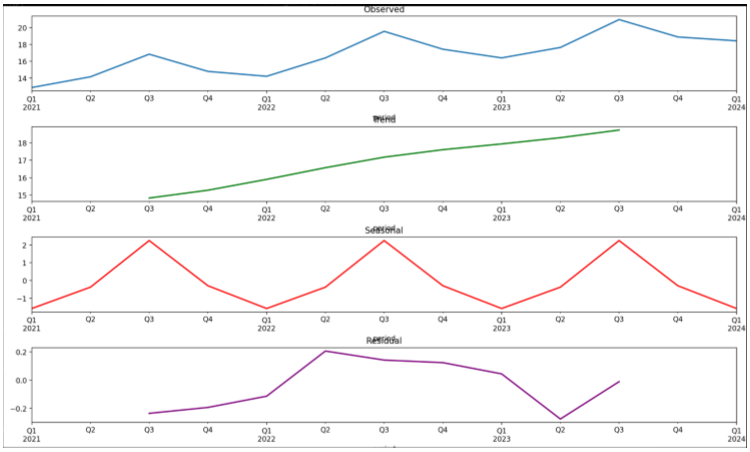
Train MSE	Test MSE	Train R2	Test R2
0.3416	2.4747	0.9080	-0.6425

Random Forest's complexity and potential to capture non-linear relationships cause it to overfit. It may have learned noise or minor fluctuations from the training set, leading to poor performance when applied to new data. To improve, we could tune the hyperparameters, increase the dataset size, or reduce model complexity to prevent overfitting.

Gradient boosting results:

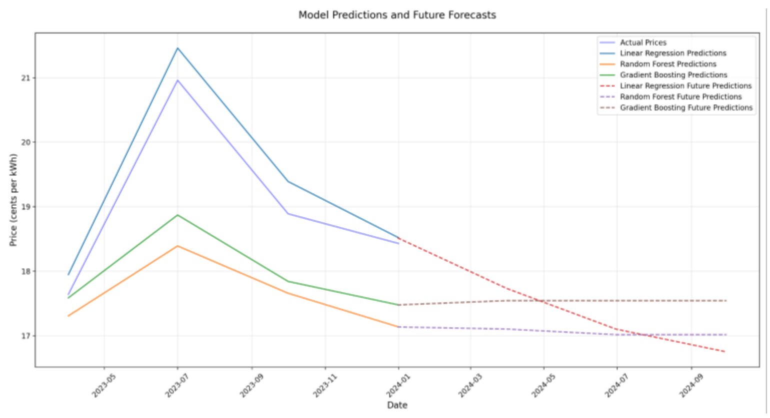
Train MSE	Test MSE	Train R2	Test R2
0	1.5932	1	-0.0575

Gradient Boosting is prone to overfitting when there is not enough data. Although it captures the trend in the training set very well, the model cannot generalize well enough to make accurate predictions for future periods. A more regularized model or a larger dataset would help reduce overfitting. Additionally, parameter tuning could smooth out the prediction curve.



1. **Observed**: The raw data, showing electricity prices over time. There's noticeable fluctuation with seasonal peaks and valleys each year.
2. **Trend**: The long-term upward movement of the data. Despite seasonal variations, prices generally increase over time.
3. **Seasonal**: Recurring patterns, where prices rise in certain quarters each year. We observe price spikes around the third quarter, followed by declines.
4. **Residual**: The remaining variation after removing the trend and seasonal components. It shows minor variations, indicating that most of the fluctuations are captured by the trend and seasonality.

This decomposition suggests the data follows both an increasing long-term trend and a strong seasonal cycle, with the residuals indicating little unexplained variation. For future improvements, we could incorporate external factors like energy demand or economic conditions to explain some of the residual variations.

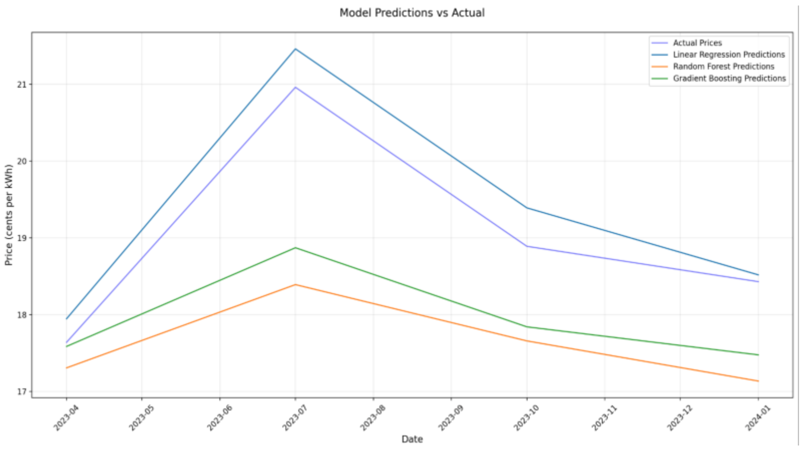


- **Actual Prices**: The solid blue line represents the true values of electricity prices up until Q1 2024.

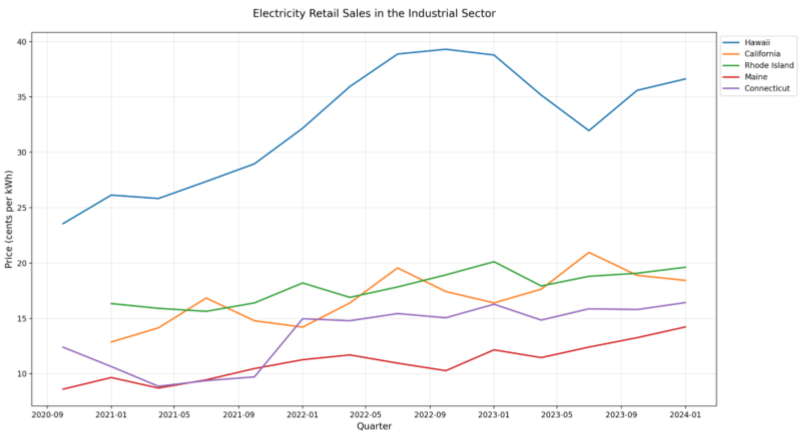
- **Model Predictions**: Solid lines in different colors show how each model fits the known data. Linear regression fits the closest to actual values, while Random Forest and Gradient Boosting underestimate prices.

- **Future Forecasts**: Dashed lines represent each model's forecast for electricity prices in future quarters. The Linear Regression model predicts a sharp decline in prices, while Random Forest and Gradient Boosting predict more stable or slightly increasing trends.

Overall, the **Linear Regression model** predicts a dramatic future drop, but this may indicate overfitting or sensitivity to short-term fluctuations. **Random Forest and Gradient Boosting models** provide more conservative estimates, although their test performances were less accurate. To improve, incorporating more time features or external factors could help refine the predictions.



This graph shows the comparison between actual electricity prices and the predictions made by three different models: Linear Regression, Random Forest, and Gradient Boosting. The linear regression model is closely aligned with actual prices, while the Random Forest and Gradient Boosting predictions exhibit a larger deviation.



This graph depicts electricity prices across five states in the industrial sector over time. Hawaii consistently has the highest prices, while the other states show more stable trends with occasional fluctuations. This highlights the variation in energy pricing across different states.

What Could Have Been Done Better:

- Dataset Size: Expanding the dataset could significantly improve model performance by providing more historical information for the models to learn from.

- Hyperparameter Tuning: Especially for Random Forest and Gradient Boosting, tuning hyperparameters such as tree depth, learning rate, and the number of estimators could help strike a balance between bias and variance.

- Feature Engineering: Creating more robust features (e.g., incorporating external factors like economic indicators or weather conditions) could improve model accuracy.