DEMAND FORECASTING ANALYSIS INTERNSHIP ASSIGNMENT

PRESENTED BY : AMAN RAJ

Optimizing SKU-Level Demand Forecasting Through Weekly Transaction Data and Predictive Modeling

AGENDA

- Problem Statement
- Project Objective
- Data Overview
- Data Preprocessing
- Exploratory Data Analysis (EDA)
- Model Selection Strategy
- Forecasting for Each SKU (SKU1 → SKU6)
- Evaluation Metrics
- Key Insights & Challenges
- Recommendations
- Conclusion

INTRODUCTION TO DEMAND FORECASTING

What is Demand Forecasting?

- Demand forecasting is the process of predicting future customer demand for a product or service.
- It uses historical sales data, market trends, and analytics techniques to estimate future demand accurately.

Why Demand Forecasting Matters?

- Accurate forecasting leads to optimized supply chain, cost savings, and better customer satisfaction.
- Reduces costs associated with overstocking or stockouts.

PROBLEM STATEMENT

Lumax is facing challenges in accurately predicting weekly demand for its key SKUs.

This leads to:

- Frequent stock-outs, resulting in lost sales and poor customer service.
- Excess inventory, causing increased warehousing costs and working capital blockage.

Needs a systematic, SKU-level, short-term forecasting model to optimize inventory planning and improve supply chain efficiency.

PROJECT OBJECTIVE

- Forecast weekly demand for six key SKUs of Lumax.
- Predict 4 weeks ahead to support production and inventory planning.
- Try to Achieve MAPE (Mean Absolute Percentage Error) below 10%.
- Enable data-driven supply chain decision-making.

DATA OVERVIEW

- Source: Transaction data across 6 separate SKU sheets.
- Granularity: Transaction Level Data
- Fields Available: SKU ID, Date, Quantity (Qty), Sales, Category
- Time Range: April 2022 to September 2023 (mostly).

Material No.	Date of Invoice	Cat-III Desc	Cat-IV Desc	QTY	NET SALES
61017767	08/06/2022	4W PREMIUM BULB	4W HALOGEN BULB	8800	694968.32
61017767	09/06/2022	4W PREMIUM BULB	4W HALOGEN BULB	207	15257.54
61017767	10/06/2022	4W PREMIUM BULB	4W HALOGEN BULB	1132	87190.79
61017767	13/06/2022	4W PREMIUM BULB	4W HALOGEN BULB	600	43215.44

DATA PREPROCESSING

1. LOAD DATA: Read each SKU sheet from the LUMAX_Data.xlsx Excel File.

2. CLEAN COLUMN NAMES: Strip leading and trailing spaces from column headers.

3. CONVERT DATE COLUMN: Convert 'Date of Invoice' column to datetime format.

4. SET DATE AS INDEX: Set 'Date of Invoice' as the DataFrame index for time series operations.

5. IDENTIFY AND HANDLE MISSING DATES: Find missing dates and reindex the DataFrame to include them (NaNs).

6. OUTLIER DETECTION: Identify outliers in 'QTY' column using IQR (Interquartile Range Method).

7. SMOOTH OUTLIERS: Apply a 7-day Rolling Median to smooth the 'QTY'.

8. INTERPOLATION: Interpolate missing values after smoothing.

9. FORWARD FILL STATIC INFORMATION: Forward-fill ‘Material No.’ for continuity.

10. CLEAN DATASET: Drop unnecessary columns like 'Cat-III Desc‘ , 'Cat-IV Desc‘ , ‘Net Sales’.

11. RESAMPLE DATA TO WEEKLY FREQUENCY: Aggregate (sum) the smoothed and interpolated daily data to weekly demand, with weeks ending on Monday (w-Mon).

CHALLENGES IN DATA

- Missing weeks detected in time series across some SKUs.
- High demand fluctuations observed, especially in SKUs like SKU5.
- Strong multiplicative seasonality found across all SKUs (confirmed by STL decomposition).
- Applied Log Transformation and Box-Cox Transformation to stabilize variance.
- ACF and PACF plots were analyzed post-transformation to understand autocorrelation structure for model selection.

EXPLORATORY DATA ANALYSIS (EDA) – INITIAL OBSERVATIONS

- Separate EDA done for 6 SKUs individually.
- STL decomposition is done as multiplicative seasonality is present in all.
- Transformations (Box-Cox, Log) applied for stability.

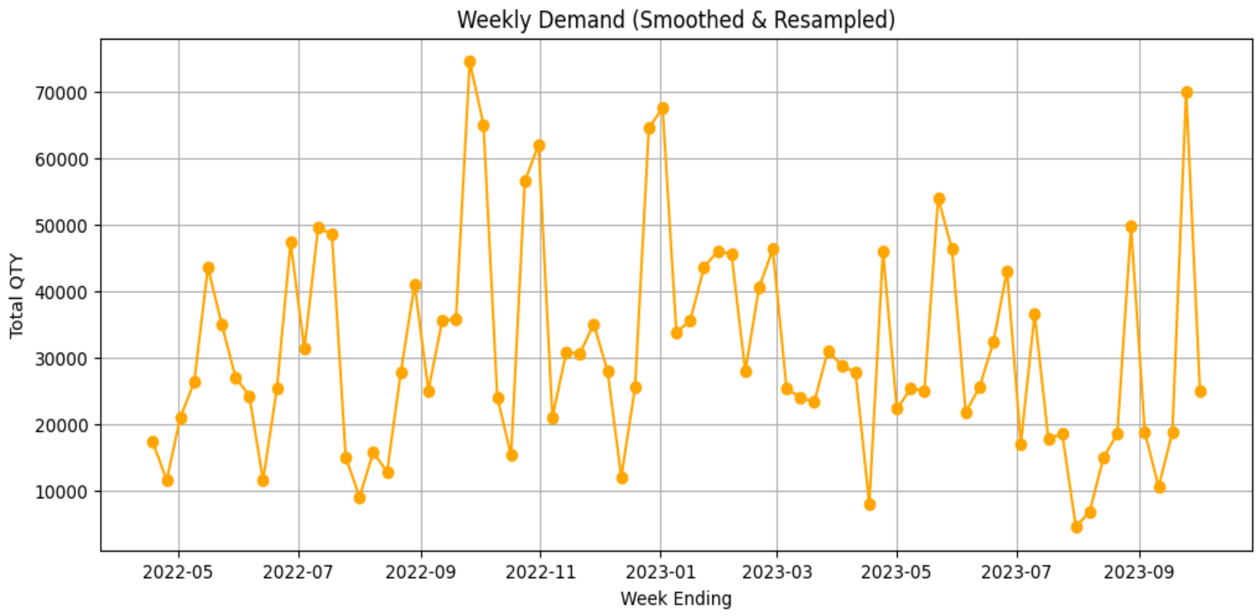


FIG: WEEKLY DEMAND PLOT OF SKU2

- ❑ Since in above graph magnitude of fluctuations increases with the level of demand.
- ❑ Multiplicative Decomposition is recommended to accurately model trend and seasonality.

- ADF Test proved stationarity post-transformation.
- Volatility seen in SKU2, SKU5.
- ACF-PACF guided modeling choice.

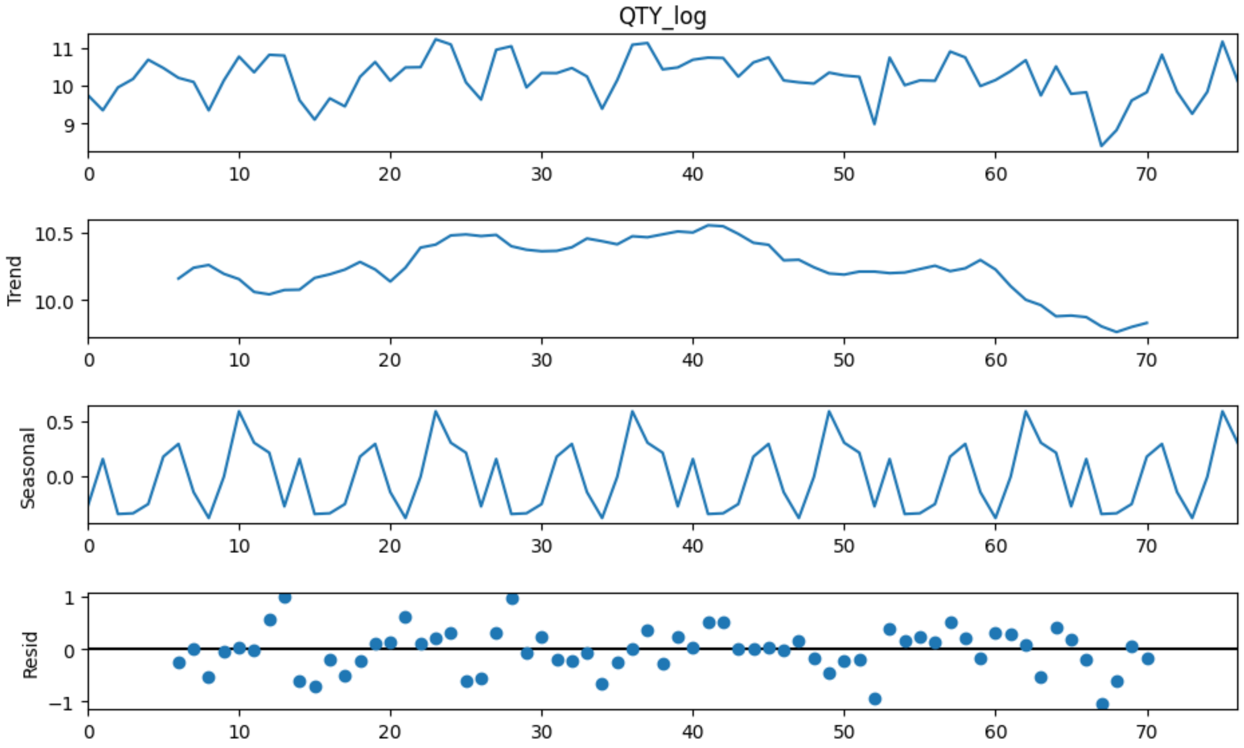


FIG: STL DECOMPOSITION OF SKU2

❑ **Similarly done all these steps for each SKU's**

❑ The STL decomposition of sales quantity (log-transformed) reveals a slight downward trend over time, along with strong and consistent seasonal patterns. The residuals are minimal, indicating that most of the variation is well explained by the trend and seasonality components. Overall, sales show predictable seasonal fluctuations with a marginal long-term decline.

EDA – TREND, SEASONALITY & WHITE NOISE

- **Visual inspection of weekly demand plots** indicated **multiplicative patterns** across SKUs.

- **Moderate, irregular seasonality detected** — visible but not strongly repetitive.

- **STL decomposition** applied to extract trend, seasonality, and residuals.

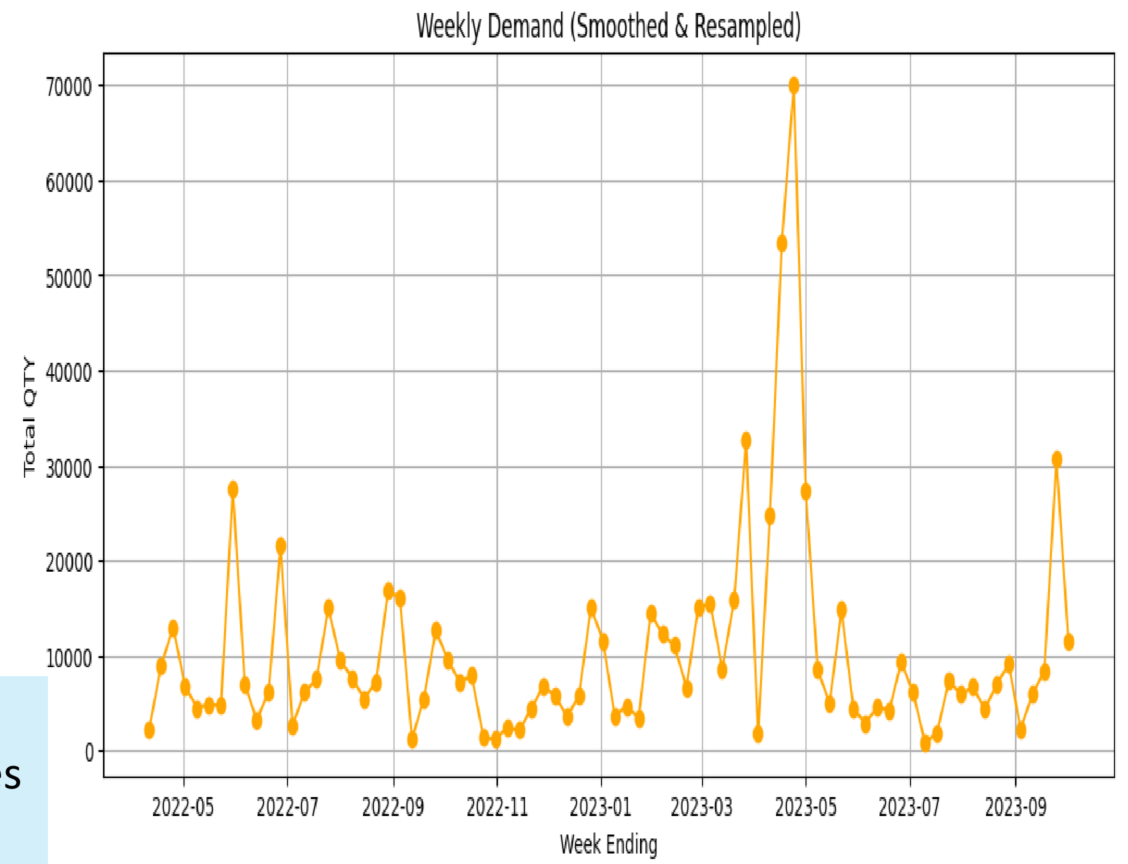


FIG: WEEKLY DEMAND OF SKU4

- ❑ Figure shows an overall increasing trend with significant seasonal spikes whose magnitude varies with the level of demand, indicating non-constant seasonality.

- **Ljung-Box test** applied to residuals for white noise checking.
- ✓ Only SKU4 and SKU6 showed **white noise residuals** (p-value > 0.05).

SKU	Ljung-Box p-value	White Noise?
sku1	2.25E-09	No
sku2	0.000287	No
sku3	0.001941	No
sku4	0.055	Yes
sku5	0.023577	No
sku6	0.0515	Yes

FIG: TABLE OF LJUNG TEST

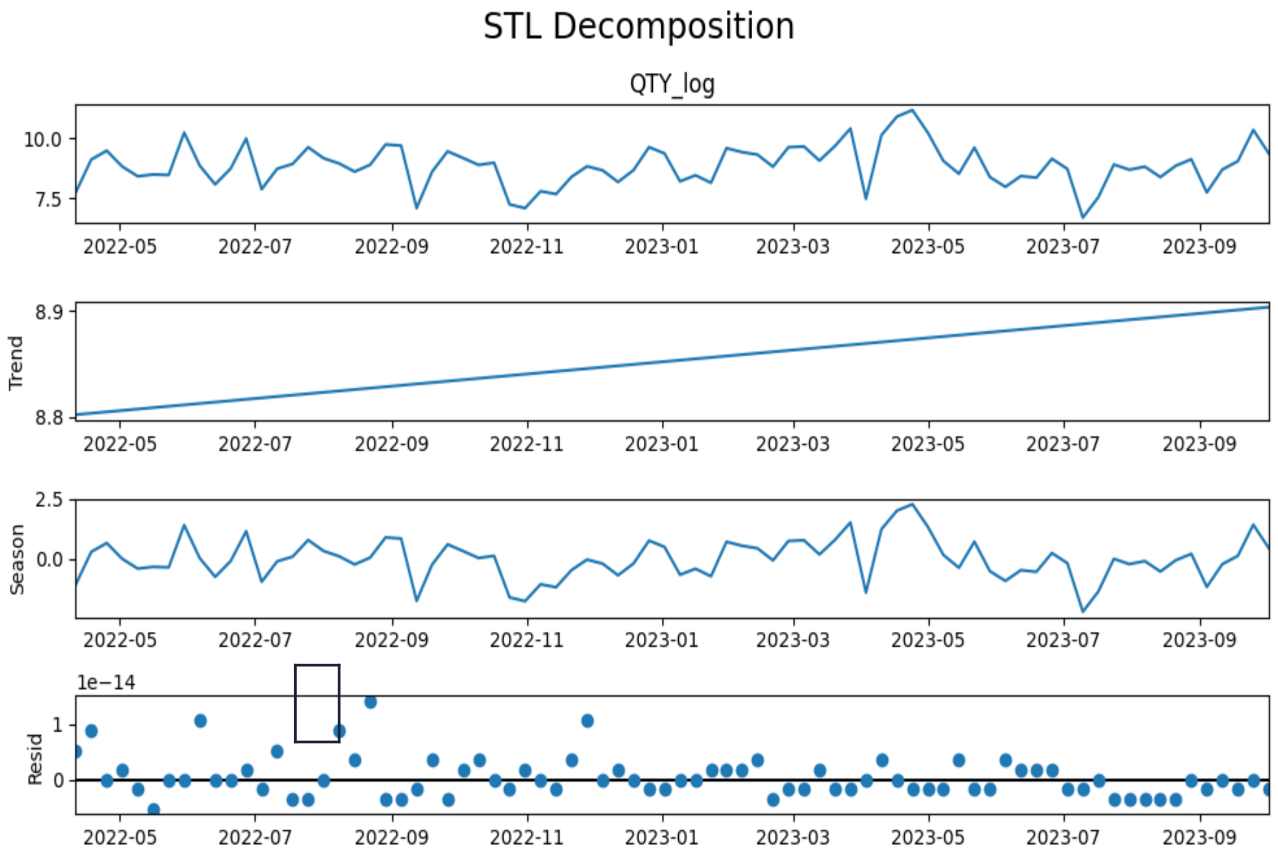


FIG: STL DECOMPOSITION OF SKU4

STATIONARITY VALIDATION AND TIME SERIES BEHAVIOUR

1. Stationarity Check using ADF Test: The raw demand series was mostly non-stationary (high ADF p-values). After differencing, p-values dropped below the threshold, confirming stationarity for modeling

```
ADF Test for Differencing Level: 0
ADF Statistic: -1.4420635751100954
p-value: 0.5619525763038824
Critical Value (1%): -3.53692771987915
Critical Value (5%): -2.907887369384766
Critical Value (10%): -2.591493291015625
The series is non-stationary, applying differencing...

ADF Test for Differencing Level: 1
ADF Statistic: -4.641883548177257
p-value: 0.00010805605331 57249
Critical Value (1%): -3.5386953618719676
Critical Value (5%): -2.9086446751210775
Critical Value (10%): -2.591896782564878

The series is stationary (reject null hypothesis)

Total differencing applied: 1
```

FIG: ADF TEST ON SKU2

2. Transformation for Stabilization: Log and Box-Cox transformations were applied, resulting in improved stationarity with lower p-values.

3. Autocorrelation Insights (ACF/PACF): ACF plots showed slow decay, and PACF plots revealed strong early lag correlations, indicating underlying autocorrelation structures.

- ❑ **Similar approach applied for other SKUs.**

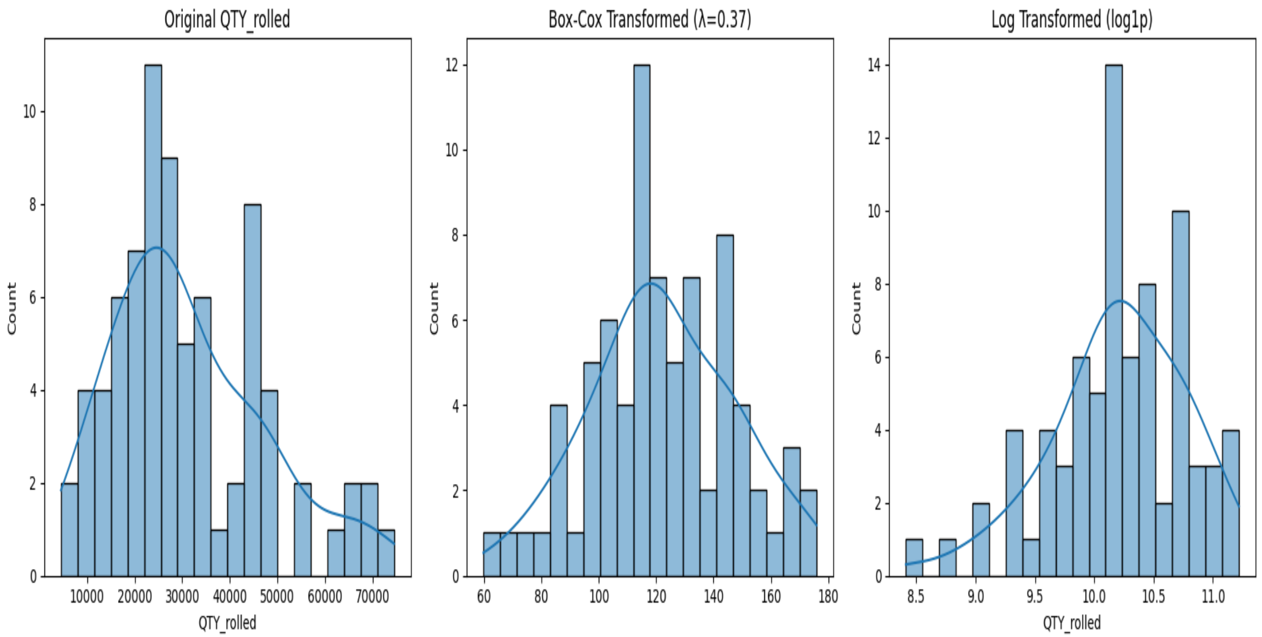


FIG: LOG & BOX-COX TRANSFORMATION OF SKU2

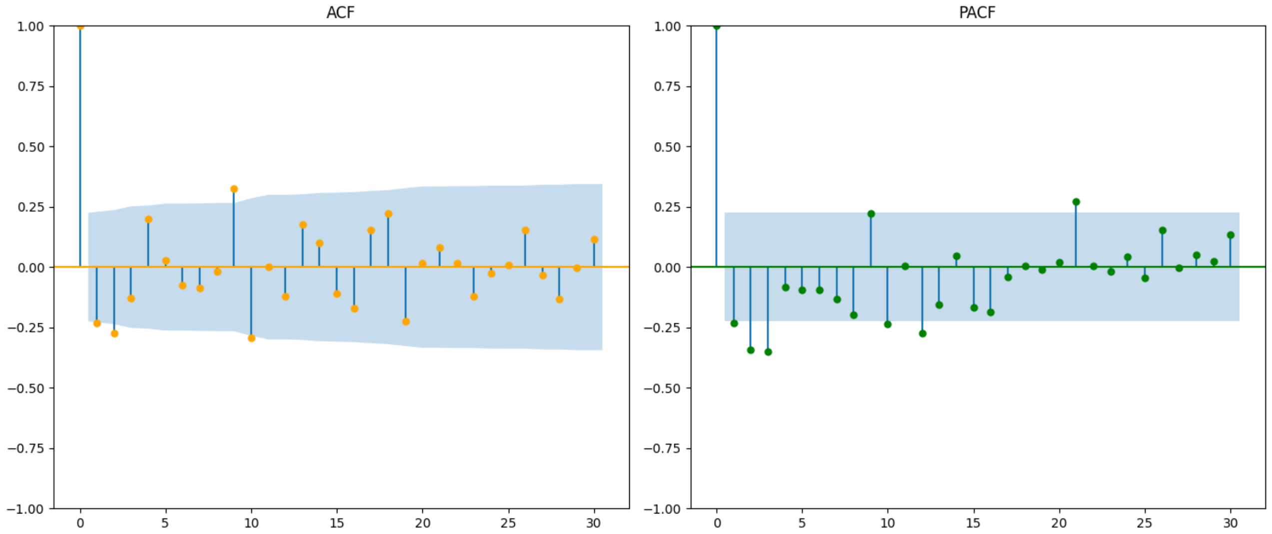


FIG: ACF&PACF PLOTS OF SKU2

MODEL SELECTION STRATEGY

1. **Multiple forecasting models evaluated** based on SKU behavior:

- NAÏVE MODEL
- MOVING AVERAGE
- ARIMA (AUTO)
- SARIMA (AUTO)
- SIMPLE EXPONENTIAL SMOOTHING (SES)
- HOLT-WINTERS SMOOTHING
- HOLT-LINEAR

2. **Selection based on:**

- Stationarity Test Results (ADF Test, White Noise - Ljung Test)
- Seasonality Presence (based on Weekly Demand Plots)
- Model Performance on Validation Set (MAPE)
- Model Simplicity and Interpretability

Sku's	Selected Model	Reason
sku1	Holt-Linear	Trend present, no strong seasonality
sku2	Holt-Winters	selected for fluctuating demand with weak seasonality.
sku3	Holt-Winters	selected for fluctuating demand with weak seasonality.
sku4	Holt-Winters	White Noise Present
sku5	Auto-Arima	Selected for Optimized Modeling of Non-seasonal, Transformed Data.
sku6	Holt-Winters	White Noise Present

3. Approach Summary:

- SKUs showing seasonality ➔ **SARIMA** selected.
- SKUs with white noise or irregular patterns ➔ **Holt-Winters Smoothing** applied.

Example:

- **SKU1:** Seasonality detected → SARIMA used.
- **SKU4:** White noise detected → Holt-Winters used.

✓ **Similar model selection strategy applied across all SKUs.**

MODEL TRAINING & EVALUATION STRATEGY

1. Training and Testing Split:

- Historical data was divided into training and testing sets.
- Last **3 weeks** were kept for **testing**.
- Remaining data used for **model training**.

2. Forecast Horizon:

- Forecasted **4 future weeks** of demand for each SKU.

3. Evaluation Metric:

- **MAPE (Mean Absolute Percentage Error)** selected for accuracy measurement.
- **Target:** MAPE < 10% for good performance.

4. Validation Approach:

- **Hold-Out Method** (simple Train-Test split due to limited data).

Full Data ➔ Train Set ➔ Test Set ➔ Forecast ➔ Compare ➔ MAPE

MODEL EVALUATION METRICS

1. **Evaluation Strategy:**
- Used **MAPE** (Mean Absolute Percentage Error) as the primary evaluation metric.
- Lower MAPE indicates **better forecasting accuracy**.

2. **Test Period:**
- Last **3 weeks** data reserved for **testing** model predictions.

3. **Benchmarks:**
- MAPE **< 10%** considered **very good** for demand forecasting.
- Compared model forecasts vs actuals across 6 SKUs.

4. **Other Observations:**
- Analyzed **residual patterns** to check model performance (random noise = good model fit).

SKU	Model Used	MAPE (%)
SKU1	Holt Linear	19.62%
SKU2	Holt-Winters	8.79%
SKU3	Holt-Winters	21.81%
SKU4	Holt-Winters	7.32%
SKU5	Auto-Arima	10.40%
SKU6	Holt-Winters	19.30%

FORECAST RESULTS

1.Forecast Horizon:

- **4 weeks** future demand forecast generated for each SKU.

2.Visualization:

- Plotted **Actual vs Forecasted** values to compare.

3.Observation:

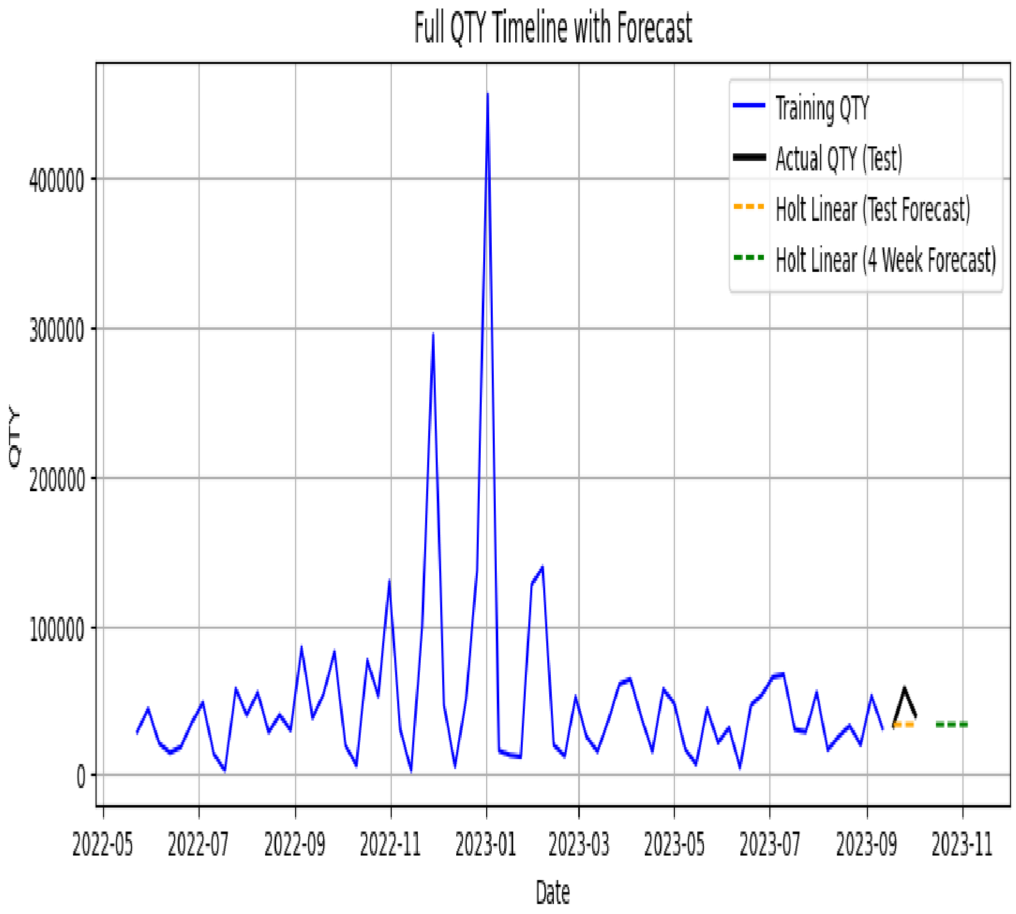
- Forecasts closely follow recent trends, ensuring reliability for planning.

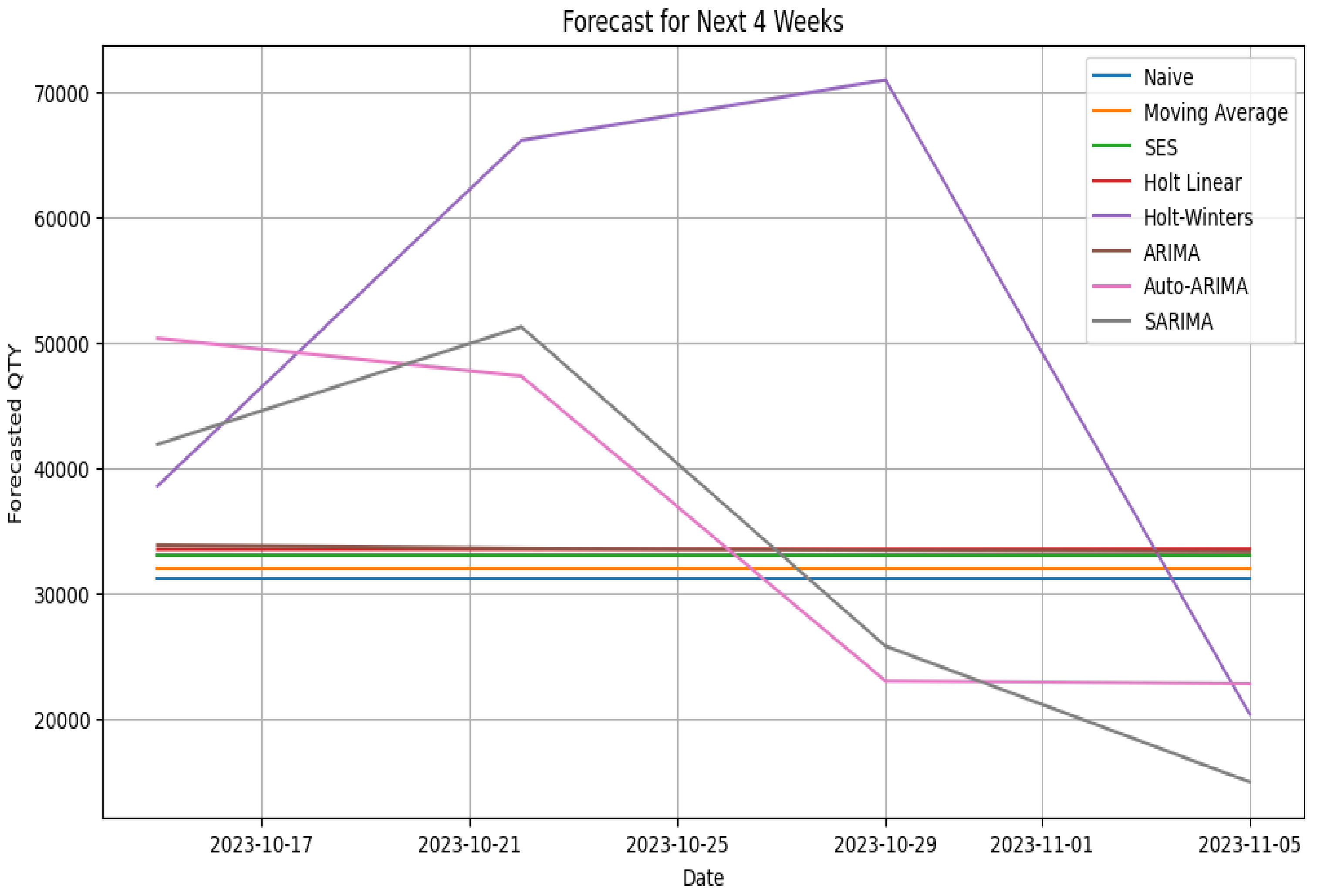
4.**Highlights:**

- For SKUs (sku2) with strong patterns the forecasts are very smooth.
- For SKUs (sku3) with irregular patterns forecast captures broad movements.

FORECAST GRAPHS

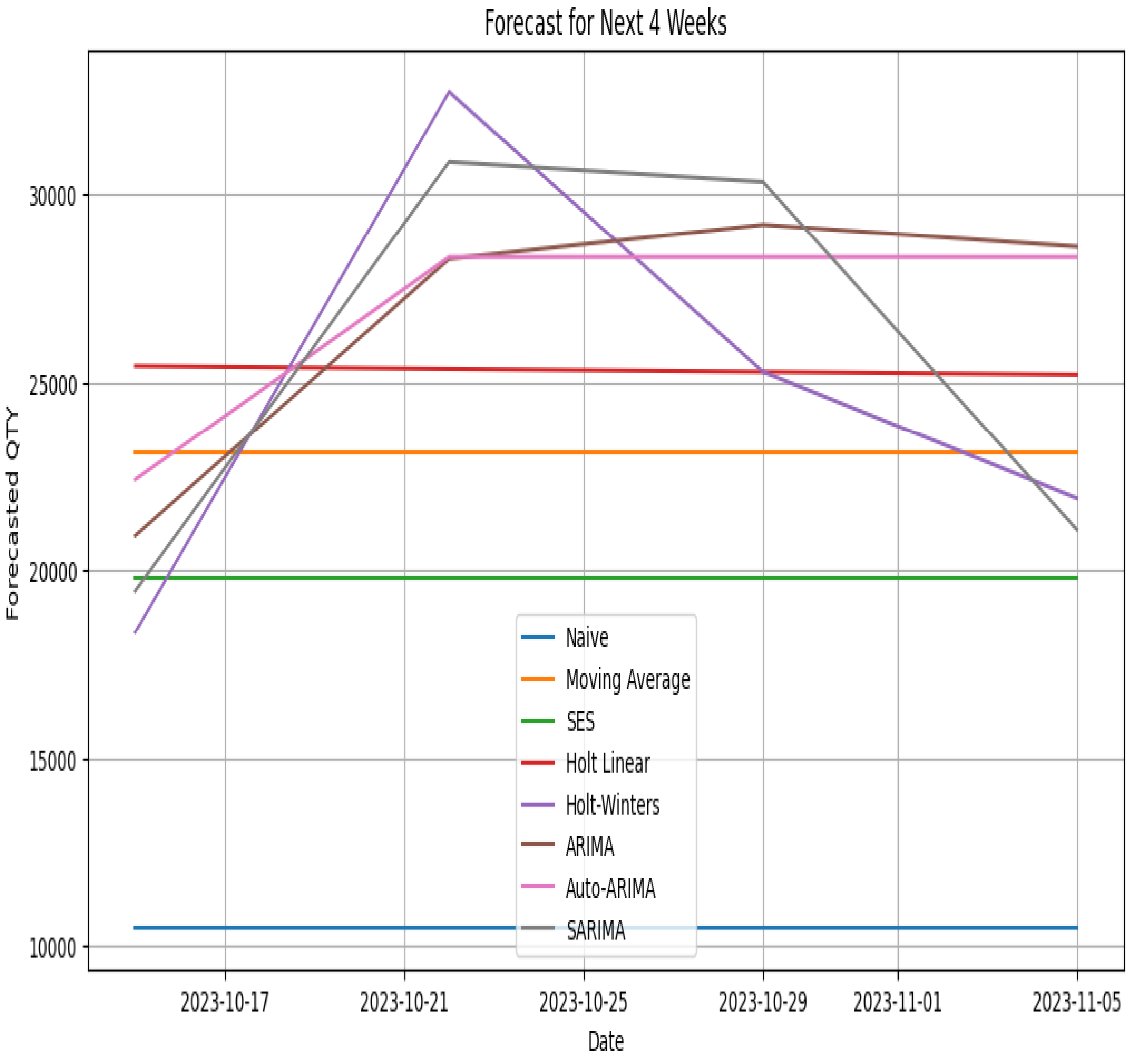
Forecasts for SKU1 :
Best Model (Holt Linear)

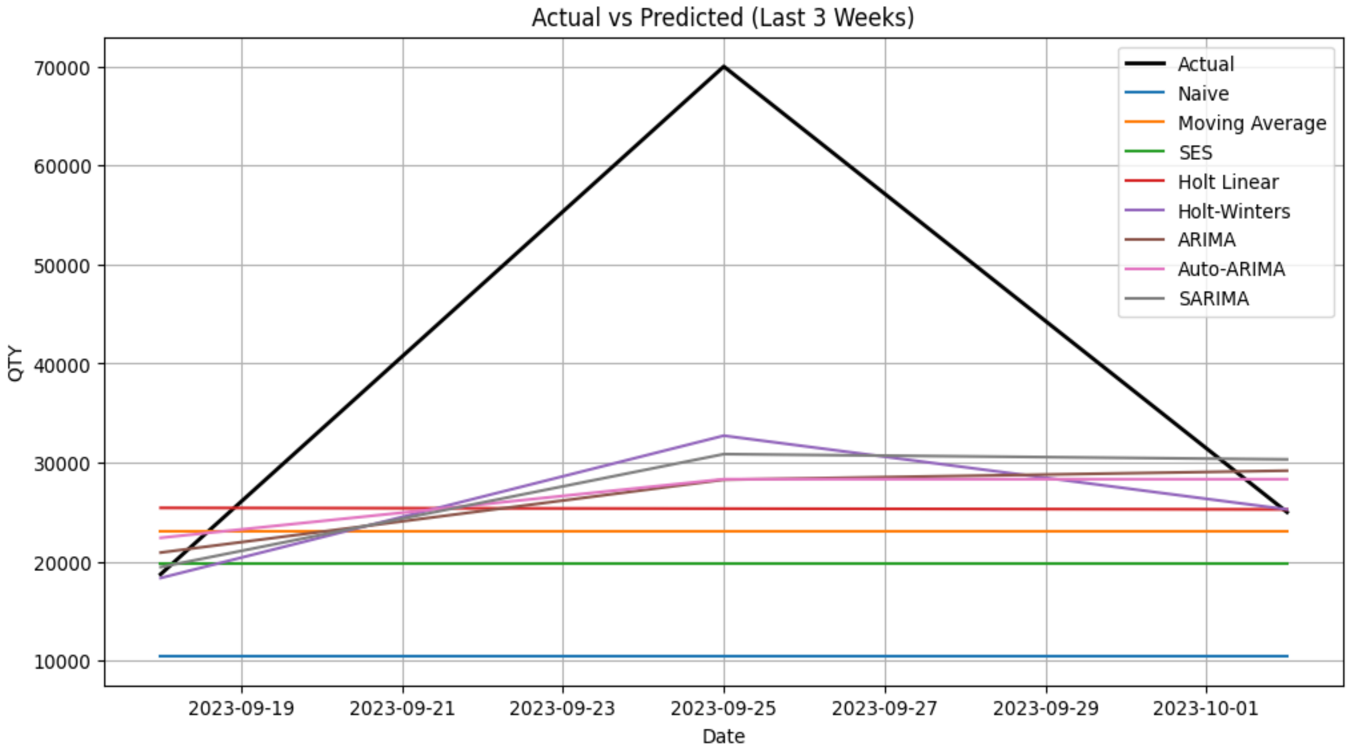




- Multiple models compared for short-term demand forecasting.

Forecasts for SKU2

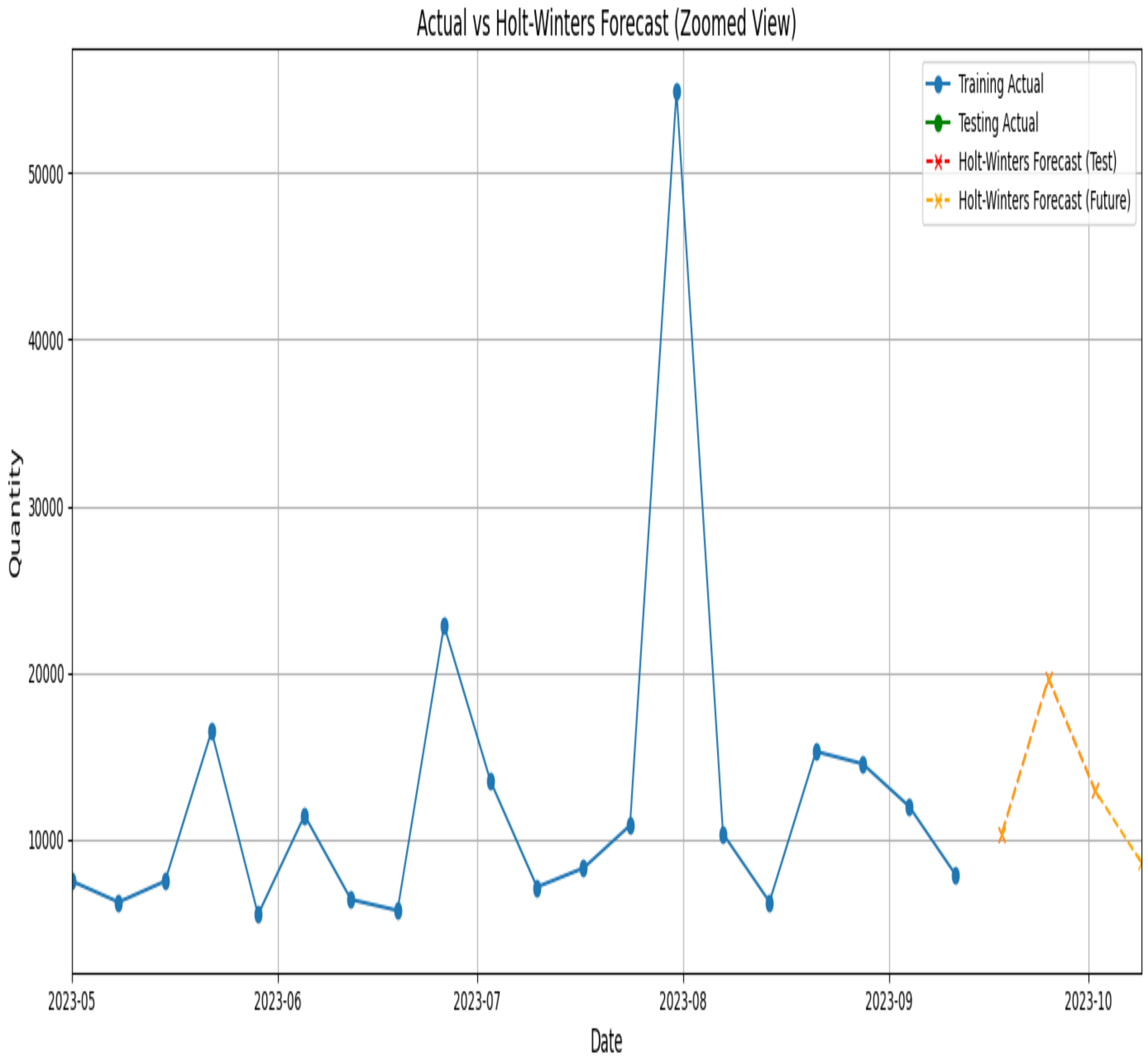


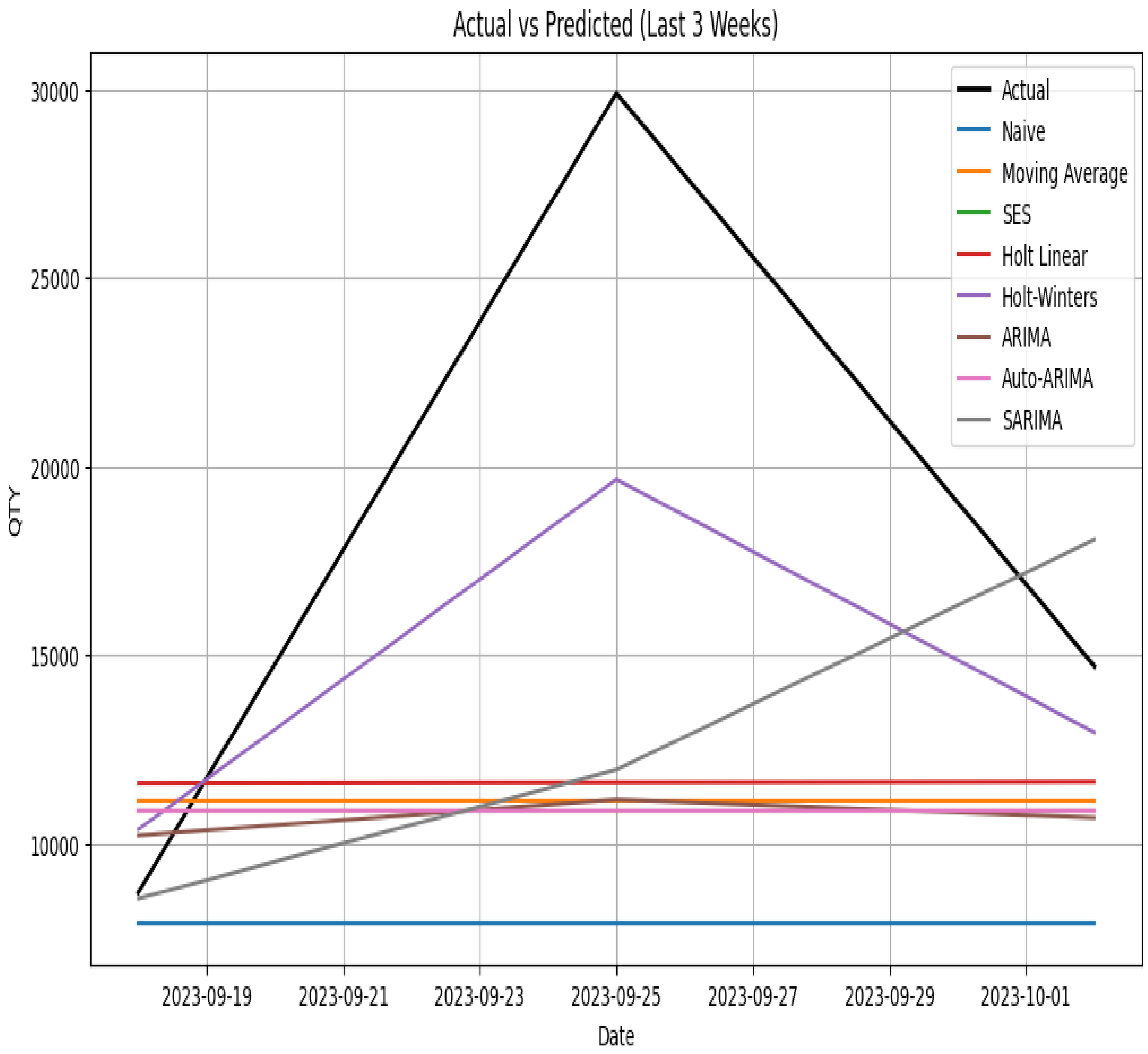


Conclusion: SARIMA/ARIMA are trend-sensitive; SES/Naive are static

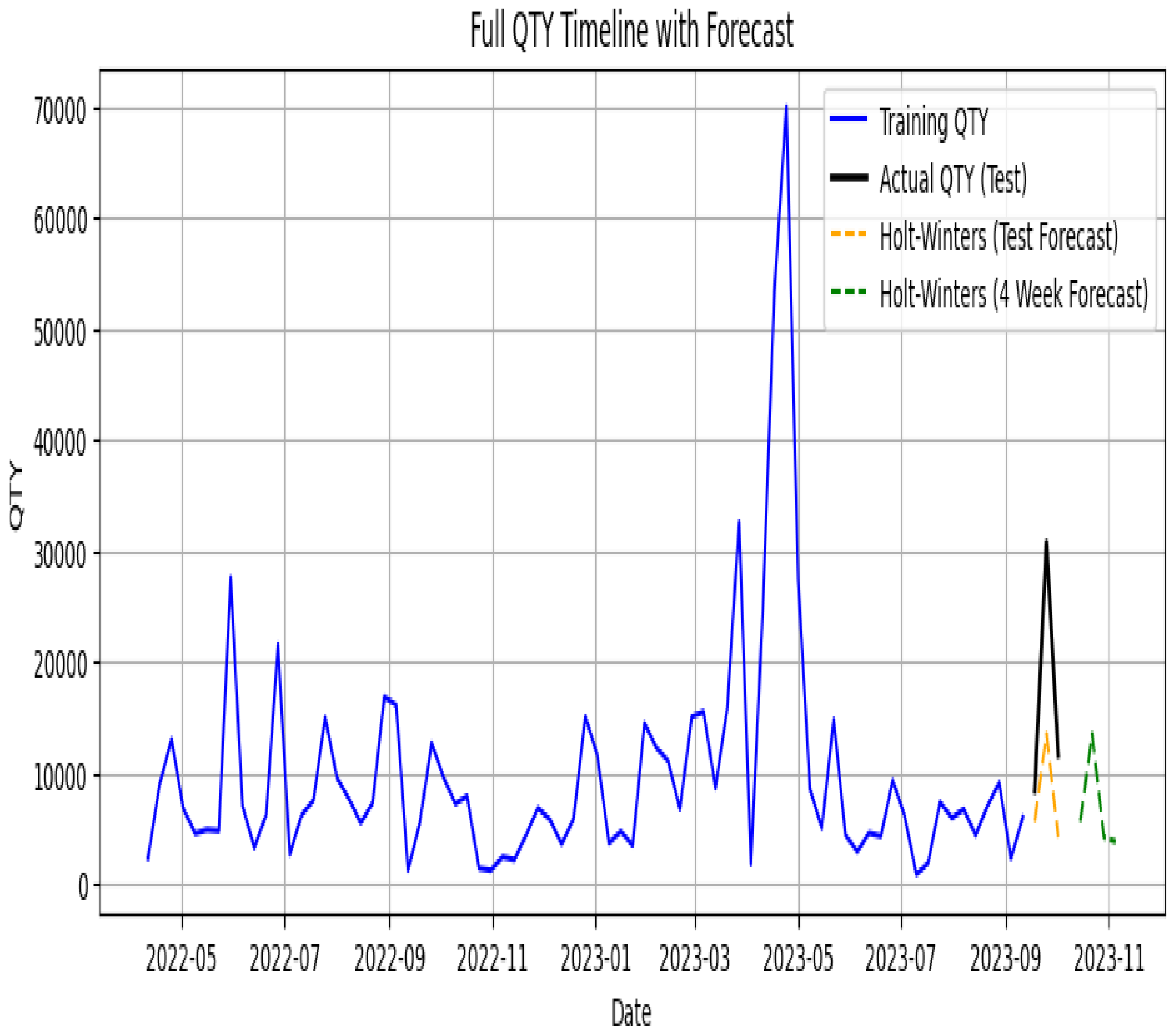
Forecasts for SKU3

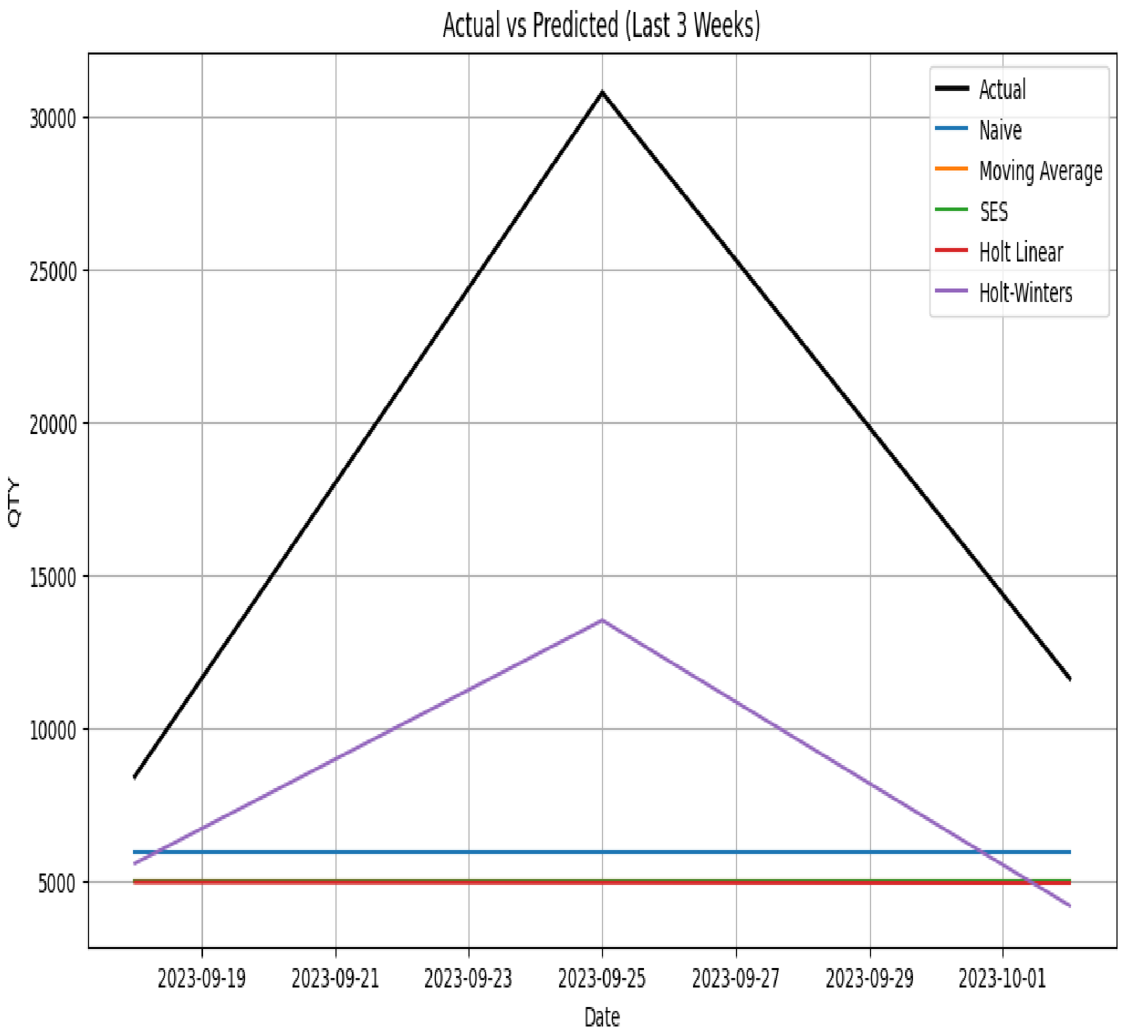
Future forecast suggests mild seasonality and trend continuation.





Forecasts for SKU4

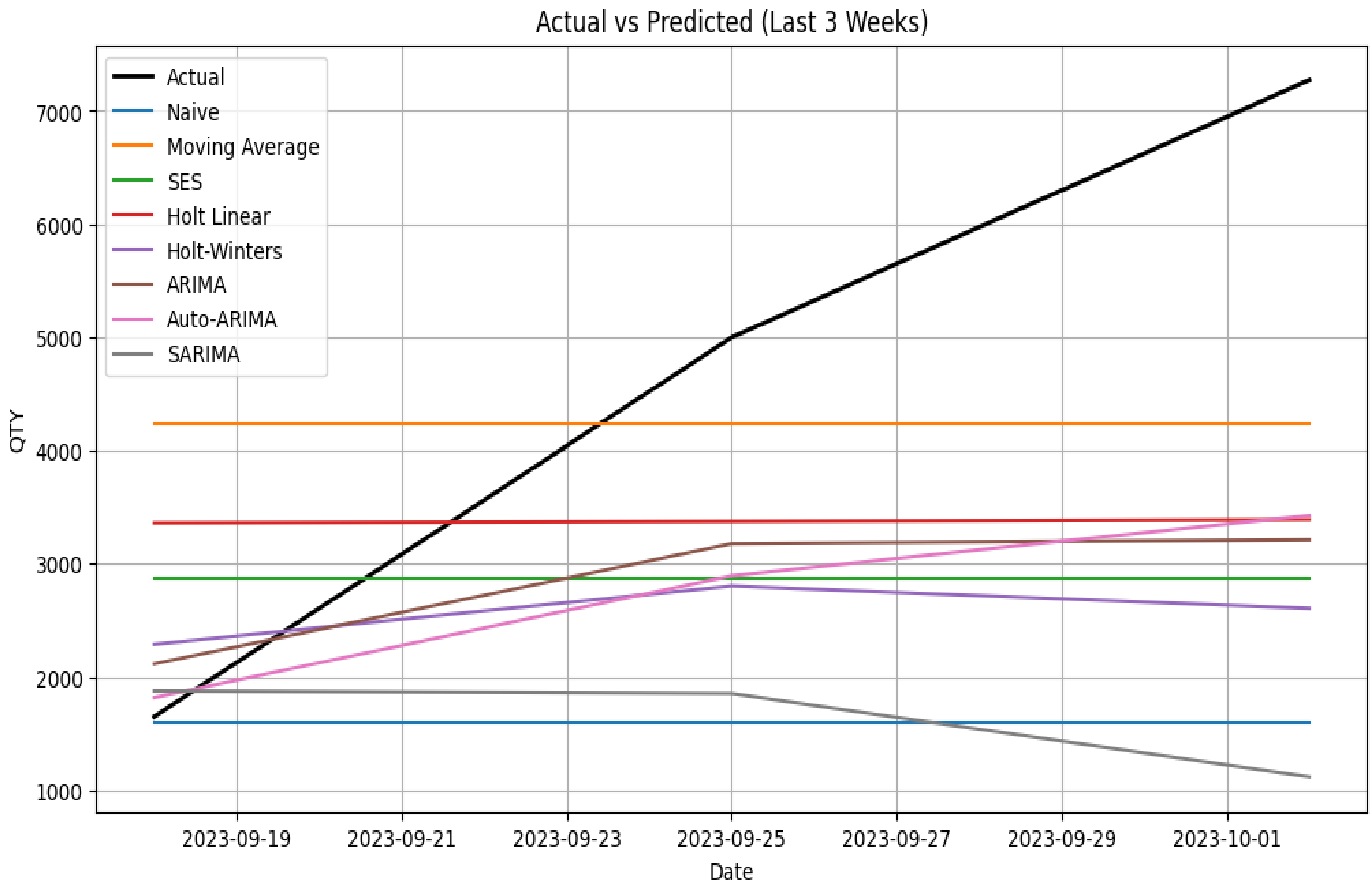




Actual values: Show a sharp peak mid-period, then decline.
Holt-Winters: Closest to actual trend but underestimates the peak.

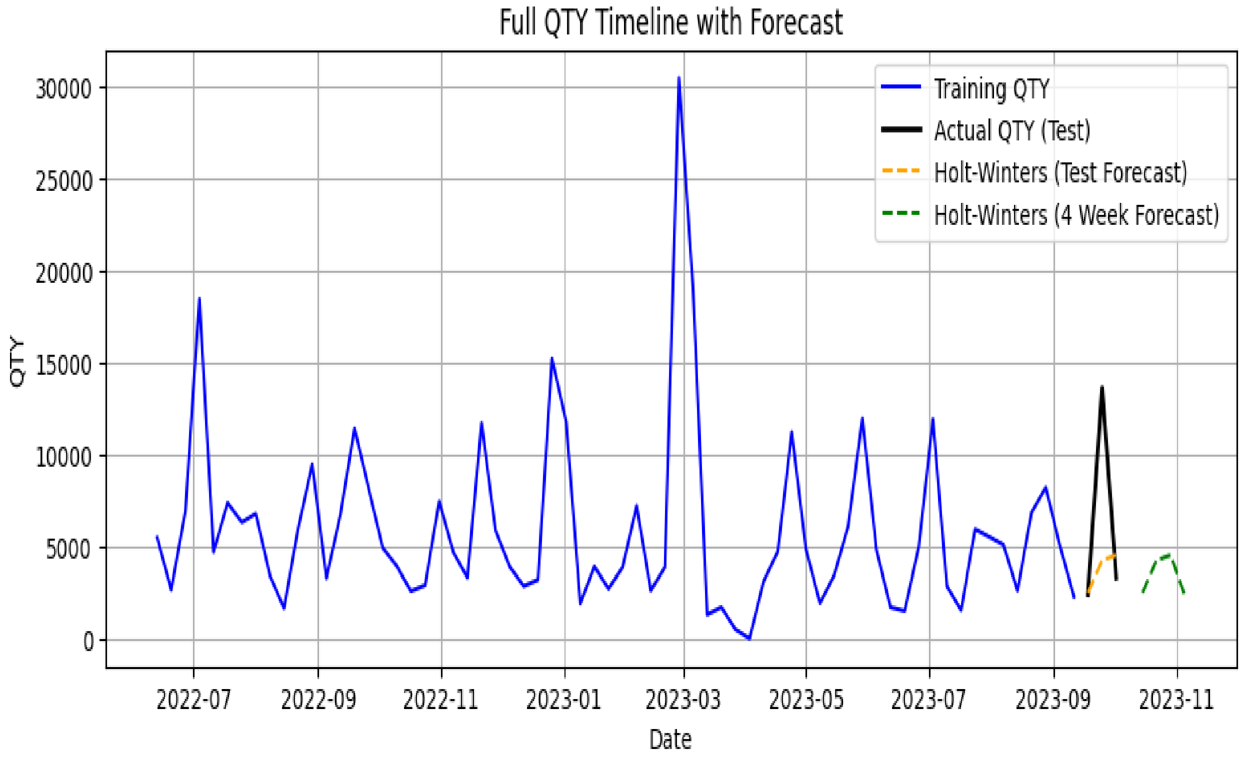
Forecasts for SKU5

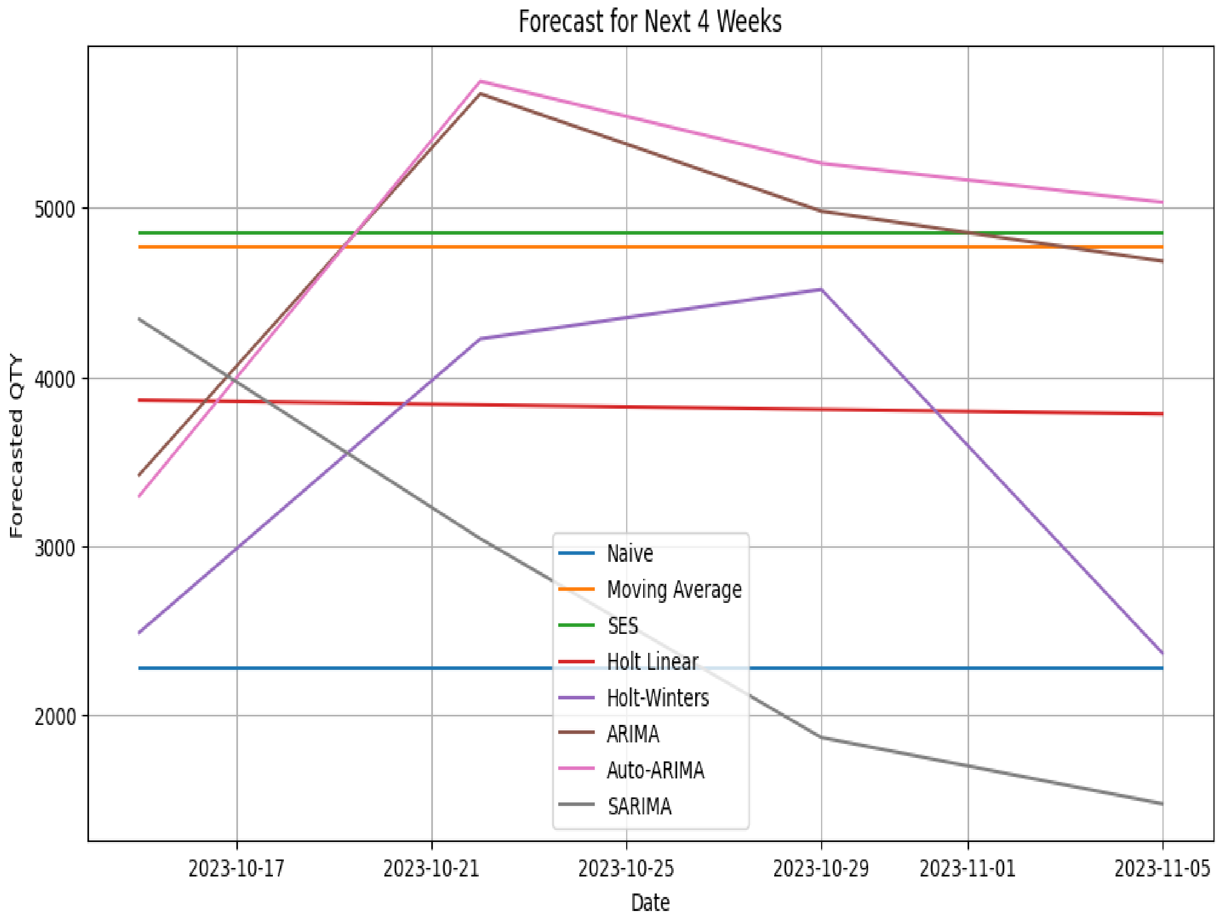
Conclusion: ARIMA-based models perform better; others fail to capture the upward trend



Forecasts for SKU6 :

Holt-Winters effectively captures seasonality and trend, especially visible in full timeline and short-term forecasts, showing it's reliable for cyclic demand.





GLIMPSES OF 4-WEEK FORECAST BY BEST MODELS

SKU 2	
WEEK	QTY
18/09/2023	18353
25/09/2023	32724
02/10/2023	25280
09/10/2023	21911

SKU 4	
WEEK	QTY
18/09/2023	5557
25/09/2023	13500
02/10/2023	4175
09/10/2023	3779

SKU 6	
WEEK	QTY
18/09/2023	2489
25/09/2023	4224
02/10/2023	4516
09/10/2023	2367

MODEL= HOLT WINTERS

BUSINESS IMPACT

Accurate demand forecasting led to:

- **Improved forecasting accuracy** across SKUs, reducing planning errors.
- **Optimized inventory planning**, preventing overstock and stockouts.
- **Reduced lost sales and carrying costs**, directly impacting profitability.

CHALLENGES FACED

While delivering the solution, I have encountered:

- **Data sparsity** for low-volume SKUs, affecting forecast reliability.
- **Outlier handling** without compromising genuine sales spikes.
- **Selecting the right forecast horizon** to balance accuracy and usefulness.
- **Selecting correct transformation** for better result.

FUTURE SCOPE / RECOMMENDATIONS

- **Develop a Streamlit-based forecasting dashboard** to enable real-time, interactive demand insights for planners and business users.

- **Adopt advanced ML algorithms** such as LSTM, XGBoost, or Prophet to capture complex and non-linear demand behaviors.

- **Build SKU-level forecast monitoring** with alert systems for proactive issue resolution.

Thank you for your time and attention!