# Forecasting Dashboard KPI Report

Dashboard link:

## Objective

This report presents a comprehensive framework for evaluating forecast performance across different update frequencies (lag periods) using key performance indicators (KPIs). The analysis examines 10 SKUs using 5 forecasting models over 1,3,6 and 12 months lags, with particular focus on understanding how forecast update frequency impacts prediction accuracy.

Forecast accuracy is critical for effective supply chain management and inventory planning. However, a key question facing demand planners is: How frequently should forecasts be updated? Updating too frequently may capture noise rather than signal, while updating too infrequently may miss important market trends.

## Data Overview

* The weekly sales data for each SKU.

A graph of sales

Description automatically generated with medium confidence

A graph of sales distribution

Description automatically generated

Figure 1: Weekly Sales Distribution

* Monthly Sales for each SKU:

A graph of colorful lines

Description automatically generated

A graph of different colored lines

Description automatically generated with medium confidence

Figure 2: Monthly Seasonality Pattern - Per SKU for each month compared across years. Helps to capture and know seasonality or trends across years at same time.

Almost same patterns repeated across different years in same month for each SKU.

* Rolling Statistics

A graph with blue and green lines

Description automatically generatedA graph of a line graph

Description automatically generated with medium confidence

In essence, the rolling mean shows you the underlying trend, while the rolling standard deviation shows you the local variability around that trend.

## Model Experimentation:

* As the series were non-stationary, they contained trends, seasonality and some SKU’s even had small irregularity. This was seen by visual inspection but after running KPSS and ADF test, the series were non-stationary. \

A screenshot of a computer screen

Description automatically generated

Figure 3: ADF test results

A screenshot of a computer screen

Description automatically generated

Figure 4: KPSS test results

* Experimented between Holts – Winter Method, ARIMA, Prophet, and ML models like XGBoost and LightGBM.
* Now for this simple data and very less features, the traditional methods were better able to capture the relationships and better forecasting. ML models can perform great, but these boosting methods needs to be finetuned and need more features for training them.

## KPI Framework and Definitions:

To comprehensively assess forecast performance, we employ six complementary KPIs that measure different aspects of forecast quality:

### Mean Absolute Error:

The average absolute difference between actual and forecasted values.

where:

* n = number of forecast periods
* Actual\_i = actual sales in period i
* Forecast\_i = forecasted sales in period i

Interpretation:

* Lower is better
* Measured in original units (e.g., units sold)
* Provides intuitive understanding of average forecast error magnitude
* A MAE of 50 means forecasts is off by 50 units on average

Usefulness:

* Easy to interpret and communicate to stakeholders
* Not disproportionately influenced by outliers
* However, MAE doesn’t indicate error direction (over- or under-prediction) or penalize large deviations heavily.

### Root Mean Squared Error (RMSE):

The square root of the average of squared differences between actual and forecasted values.

Interpretation:

* **Lower is better**
* Measured in original units
* Penalizes large errors more heavily than MAE
* More sensitive to outliers than MAE

Usefulness:

* Highlights the presence of large forecasting errors
* Useful when large errors are particularly costly (e.g., stockouts)
* Standard metric in forecasting literature for model comparison

**MAE vs RMSE:**

The optimization of RMSE will seek to be correct on average, whereas the optimization of MAE will try to be as often overshooting the demand as undershooting the demand, which means targeting the demand median. When RMSE is much larger than MAE for a given lag period, it signals the presence of problematic outlier forecasts that could disrupt operations as RMSE is crucial for lag period analysis because it heavily penalizes large forecasting errors.

### Bias Percentage:

Bias Percentage measures the cumulative systematic tendency of forecasts to be consistently higher (over-forecasting) or lower (under-forecasting) than actual demand across all forecast periods.

where:

* sum(Forecast - Actual) = Total cumulative forecast error
* sum(Actual) = Total actual demand across all periods
* Result expressed as percentage

Optimal value: 0% (perfectly balanced forecasting)

Positive bias (+): Systematic over-forecasting

* Example: Bias% = +10% means forecasts are consistently 10% too high
* Result: Excess inventory, increased holding costs

Negative bias (-): Systematic under-forecasting

* Example: Bias% = -10% means forecasts are consistently 10% too low
* Result: Stockouts, lost sales, customer dissatisfaction

Near-zero bias (±5%): Balanced forecasting with roughly equal over- and under-predictions

Bias% is critical for evaluating forecast update frequencies because it reveals whether different lag periods introduce systematic directional errors. Longer lag periods may cause models to miss recent trend changes, leading to persistent over-forecasting (if trends are declining) or under-forecasting (if trends are growing). By comparing Bias% across lag periods (1, 3, 6, 12 months), we can identify which update frequencies maintain near-zero bias, ensuring inventory planning is neither consistently excessive nor deficient. Unlike MAE or RMSE which measure error magnitude, Bias% specifically detects the direction and consistency of errors, essential for avoiding the compounding costs of systematic forecasting mistakes over time.

But also, why it should be used with MAE and RMSE is, as a positive error on one item can offset a negative error on another item, a forecast model can achieve very low bias and not be precise at the same time. Obviously, the bias alone won't be enough to evaluate our forecast precision. But a highly biased forecast is already an indication that something is wrong in the model.

**NOTE**: As the denominator is sum of actual sales, some near zero or zero sales can shoot this high, so carefully check for those cases. Especially SKU 6,7 and 10 in our case.

### Service Level Percentage:

The percentage of periods where forecast met or exceeded actual demand.

where:

* n = Total Periods
* Forecast\_i = forecasted sales in period i
* Actual\_i = Actual sales in period i

Interpretation:

* Higher is better
* Target typically 85-95% depending on business requirements
* A service level of 90% means demand was met in 9 out of 10 periods

Now what happens is, we don’t want stockout for high running products, it’s usually helpful to capture high service level percentage of about 85-95%, as meaning forecast exceeded the demand and we had enough safety stocks. It can basically tell are forecast is continuously giving high percentage for all periods then we might want to cap the safety stock too, as it is overly forecasting.

**NOTE**: As the numerator is count (predicted – actual sales) in period i, and we have high bias percentage then we must look for those SKUs. Those SKUs might have zero or near zero sales. Especially SKU 6,7 and 10 in our case.

## Support KPIs:

### Symmetric Mean Absolute Percentage Error (SMAPE):

A scale-independent percentage error metric that treats over-forecasting and under-forecasting equally.

Interpretation:

* Lower is better
* Bounded between 0% and 200%
* Scale-independent (allows comparison across different SKUs)
* A SMAPE of 20% indicates 20% average relative error

Usefulness:

* Enables comparison across SKUs with different sales volumes
* Handles zero values better than traditional MAPE
* Symmetric treatment of over- and under-forecasting

### Tracking Signal:

A metric that detects forecast bias drift over time by comparing cumulative forecast error, also referred to as running sum of forecast errors, to mean absolute deviation.

Interpretation:

* Ideal range: -4 to +4
* Values outside ±4 indicate systematic bias requiring model recalibration
* Positive values = tendency to over-forecast
* Negative values = tendency to under-forecast

Usefulness:

* Early warning system for forecast model deterioration
* Monitors forecast bias evolution over time
* Prevents accumulation of systematic errors

**NOTE**: This is opposite from actual definition normally found on internet but used accordingly to the simple interpretation we have here. If Forecast > Actual consistently, then tracking signal will be positive so over forecasting and vice versa. It’s Actual – Forecast on internet I have used is Forecast – Actual. This was totally new KPI that I encountered while researching.

## Dashboard Functionality:

* **Filters:** Users can select model type (ARIMA, Holt-Winters, Prophet), SKU, and lag period (1, 3, 6, or 12 months) and which KPIs to showcase. By default, all are selected. These are shown on the left sidebar of dashboard. All lags should be included, it basically gives you option on main page to select which lag period KPI you need to use.\

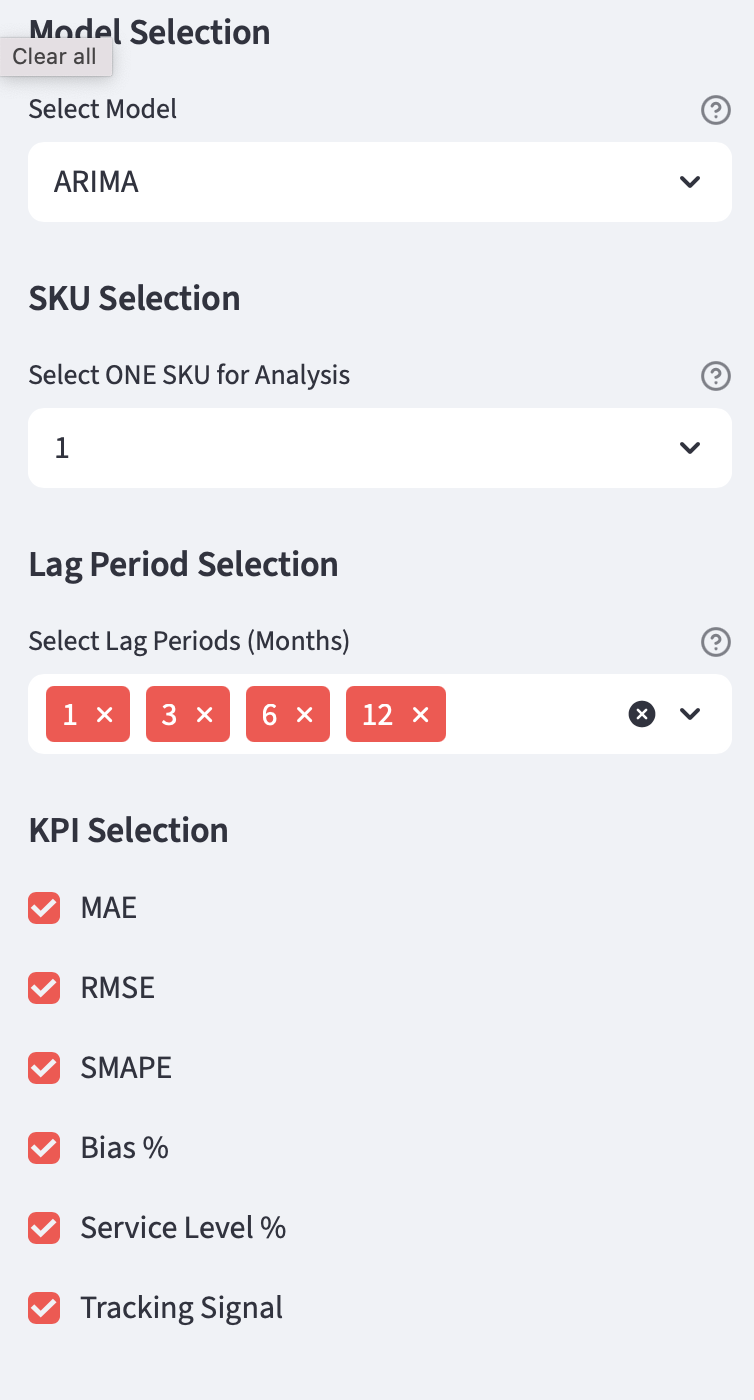


Figure 5: Sidebar – better to keep all lags here, default model is ARIMA, as it was observed best performance, however, we can select certain lag and cross comparative analysis on several subpages.

* **KPI Panels:** Display live metrics for the selected SKU and lag with hover-over definitions.

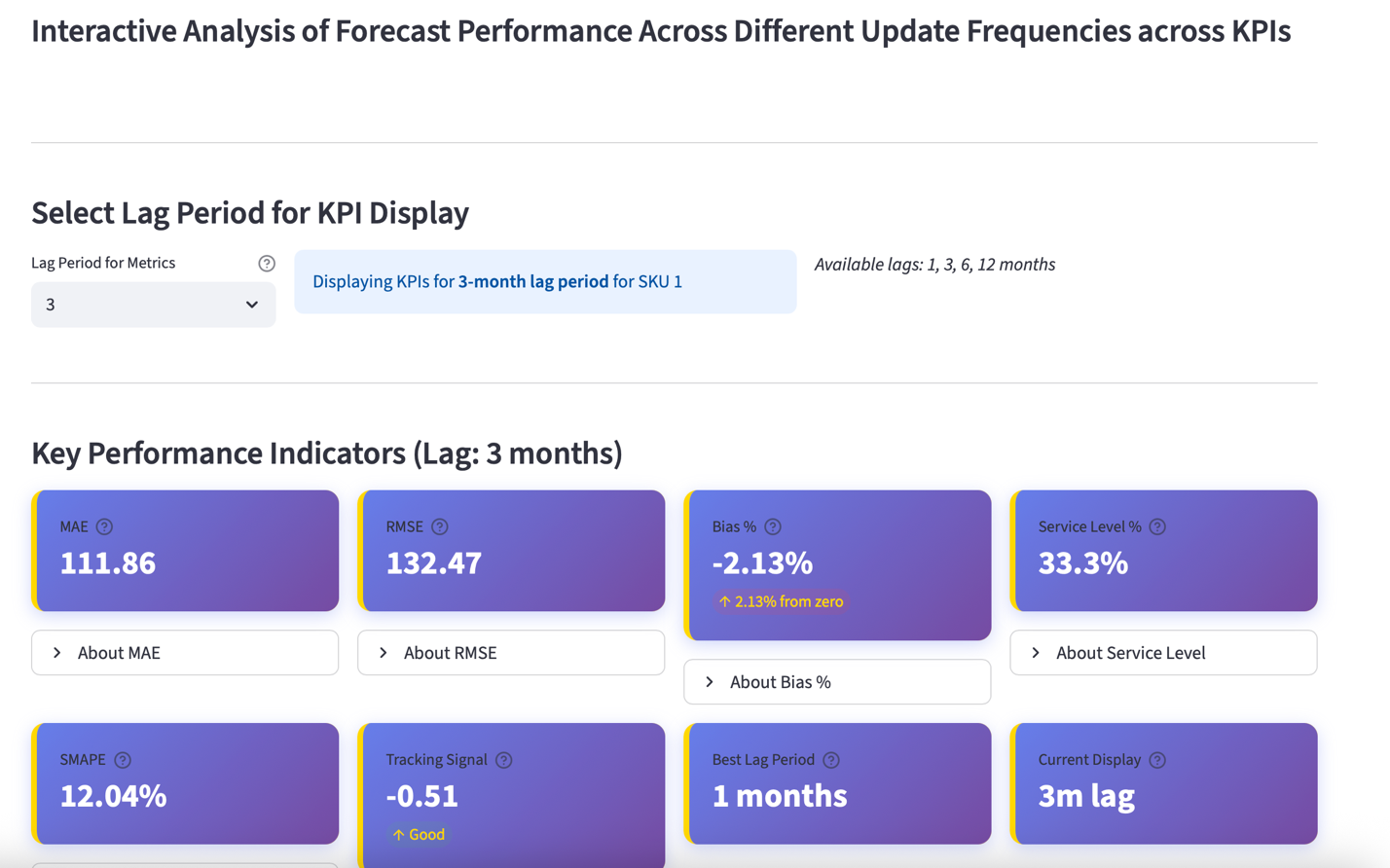
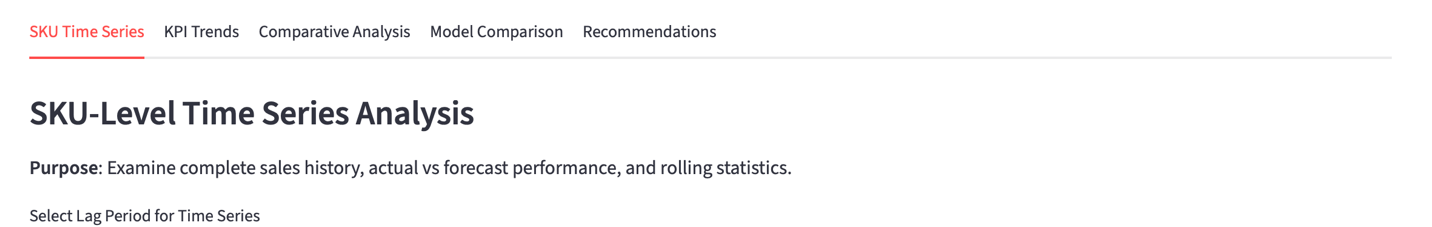


Figure 6: KPI Panels, shows KPI we selected from sidebar, there is option to select between available lags (1,3,6,12 – selected in sidebar) – so we can select out of this 4 available lags on this page, one at a time

* **Tabs:**

**Tab 1: SKU Time series, shows forecasted vs Actual sales for the given SKU, option to select lag periods.**



A graph of a graph

Description automatically generated

Figure 7: For lag 1 - SKU Time series

A graph with red and blue lines

Description automatically generated

Figure 8: For lag 3 - detailed zoomed in forecast v/s actual

A graph of sales

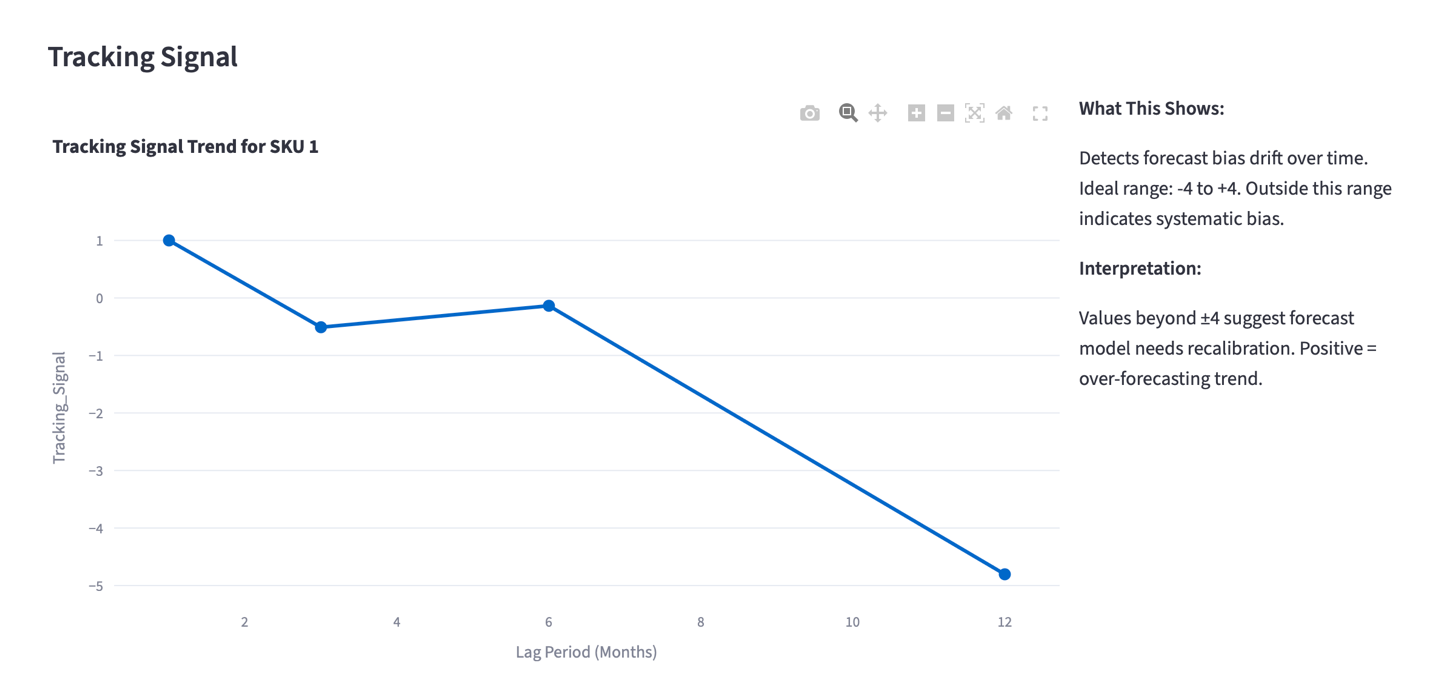
Description automatically generated

Figure 9: Rolling 3-month statistics

**Tab 2: KPI Performance Trends Across Lag Periods - Visualize how each KPI changes for SKU 1 as lag period increases.**

**A graph with a line

Description automatically generated**

****

**Tab 3: Comparative Lag Analysis**

**Purpose: Compare KPIs across different lag periods for selected SKU.**

A screenshot of a graph

Description automatically generated

**A screenshot of a graph

Description automatically generated**

Figure 10: All SKUs v/s Lag Periods, for a metric selected (MAE here).

*Tip: For given SKU, see KPIs’ performance for multiple lag period, and then use heatmap for comparison across all SKUs and Lag periods, for a given KPI***.**

**TAB 4: Model Performance comparison across different KPIs for a given SKU and given lag period.**

**A screenshot of a graph

Description automatically generated**

**TAB 5: Recommendations – Click the button to generate recommendations, used claude-3.7 model to give recommendations. The dynamically available data of dashboard feeds to the model and given context and gives customized recommendations. Provided in a way easily understood by the viewer from non-technical understanding too.**

**A screenshot of a computer

Description automatically generated**

**A screenshot of a web page

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

### Example usage of the KPI dashboard from the dashboard:

Model Used: ARIMA

*Let’s see how to see performance of KPIs – for lag 3 and lag 6.*

Accuracy (MAE & RMSE):

* At 3-month lag: Forecasts are off by 27.24 units on average (MAE)
* At 6-month lag: Forecasts are off by 125.98 units on average - that's 4.6x worse!

When you wait 6 months to update forecasts for SKU 3, your prediction errors become massive. The model is using very old data (6 months back) and completely missing recent changes in demand patterns.

With 3-month lag: You might order 100 units when you need 127 (27 unit error)

With 6-month lag: You might order 100 units when you need 226 (126 unit error).

SMAPE shows percentage error. At 3-month lag, you're only 3.63% off - that's incredibly accurate. At 6-month lag, you're 15% off - still not terrible, but 4x worse.

Both lags show balanced forecasting (bias near zero). This means errors aren't one-sided - you're not consistently predicting too high or too low. The negative sign means you're slightly under-forecasting (forecasts are 2-3% lower than actuals).

At 3-month lag: Forecast covers demand in only 1 out of 3 months (33.3%).

At 6-month lag: Forecast covers demand in only half the months (50%).

The forecast is insufficient 2 out of 3 months with 3-month lag. If this were real inventory, we would have constant stockouts. Even though 6-month lag has worse accuracy (higher MAE), it has better service level because it's under-forecasting less severely.

Both are healthy - no concerning bias drift over time. The model isn't systematically getting worse by seeing the tracking signals.

Overall, the 3-month lag performs better, there is low service level but acceptable.

*Comparison of Lag 3 months - SKU 3 vs SKU 6:*

Accuracy (MAE & RMSE):

* SKU 3: Errors of 27 units (MAE) and 39 units (RMSE)
* SKU 6: Errors of 40 units (MAE) and 68 units (RMSE)

SKU 3 is more accurate - forecasts are closer to actual sales. SKU 6 has 47% higher average error.

Bias\_%:

* SKU 3: -2.57% (nearly perfect balance)
* SKU 6: +85.13% (massive over-forecasting).

SKU 6 forecasts are 85% too high. Now this can be due to zero or near zero sales and model didn’t go to that level. As this is low running products as seen from the trend, we can easily cap below the forecasted values, to save inventory costs.

Now SKU 6 has higher service level, but that means we are ordering more than demand every time. SKU 6’s SMAPE can be bit overlooked as it has near zero sales and declining performance, which overshoots the metric.

SKU 6's tracking signal of +4.47 indicates accumulating positive bias - the model is consistently over-forecasting, and this bias is getting worse over time. The model needs to be looked.

This comparison demonstrates why multiple KPIs are essential. Looking at Service Level alone, SKU 6 appears superior (83% vs 33%). However, examining Bias% reveals SKU 6 achieves this through massive over-forecasting (+85%), leading to costly excess inventory. SKU 3, despite lower service level, has balanced forecasting (Bias% = -2.57%) and much better accuracy (MAE 27 vs 40, SMAPE 3.6% vs 129%). This illustrates the critical need for comprehensive KPI analysis rather than single-metric optimization.

### Final Recommendations

There's no one-size-fits-all. Fast-moving products need frequent updates (1-3 months). Mature or seasonal products do well with longer windows (12 months). Compare KPIs across lags to find the right balance between responsiveness and stability.

The trend-growing SKUs (1–5, 8, 9) show a gradual increase in errors from 1 to 6 months followed by a moderate decrease at 12 months, confirming that longer-term seasonality benefits these SKUs.

Conversely, the declining-trend SKUs (6, 7, 10) exhibit relatively stable or improving accuracy at longer lags, since their downward sales trajectory is more predictable and less sensitive to recent fluctuations.

This contrast emphasizes that the optimal lag period depends on the SKU’s underlying trend behavior, shorter lags for rapidly changing or growing items, and longer lags for stable or declining ones.

A demand planner should use the dashboard to compare KPI trends across lags and select the optimal update frequency that balances accuracy with computational efficiency.

Like industry practices where planners evaluate accuracy at multiple lag points (e.g., 6-month and 12-month horizons), the results from this analysis suggest adopting a dynamic lag framework.

Shorter lags (1-3 months) capture recent market dynamics and should be prioritized for fast-moving or growing SKUs (e.g., SKUs 1–5, 8, and 9).

Longer lags (6-12 months), on the other hand, provide a more stable representation of slow-moving or declining SKUs (e.g., SKUs 6, 7, and 10), where sales are less affected by short-term fluctuations.