

Chapter 4

RESULTS AND DISCUSSION

4.1 Exploratory Data Analysis (EDA) Results

This section presents a high-level exploratory data analysis (EDA) of the dataset covering the period from January 2021 to April 2025. The objective is to identify underlying patterns in demand behavior, including demand distribution across product types, individual SKU performance, temporal trends, and seasonality driven by calendar events or fiscal cycles. These insights form the empirical foundation for feature engineering and modeling decisions in the subsequent stages.

4.1.1 Demand Trends by Year

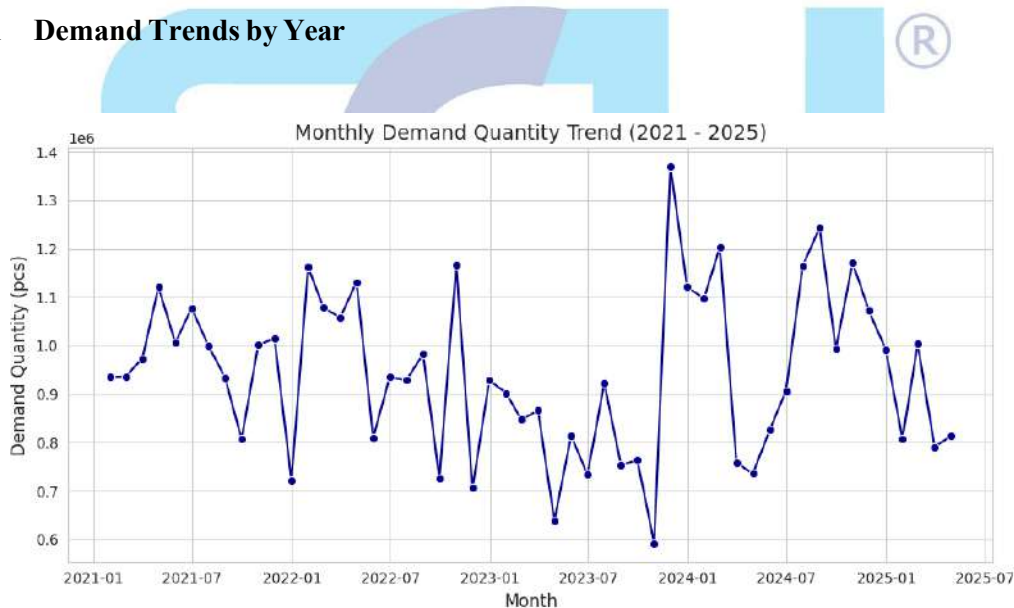


Figure 4.1: Monthly demand quantity trend across all SKUs (2021–2025).

Figure 4.1 shows the total monthly demand quantity in all SKUs from 2021 to 2025. Although the pattern displays periodic peaks and troughs, a significant upward shift in demand is observed in early 2024. Fluctuations are evident around mid-year and year-end periods, suggesting possible alignment with business cycles or external events.

4.1.2 Demand Distribution by Item Type

Demand Quantity Distribution by Item Type (Top 4 + Others)

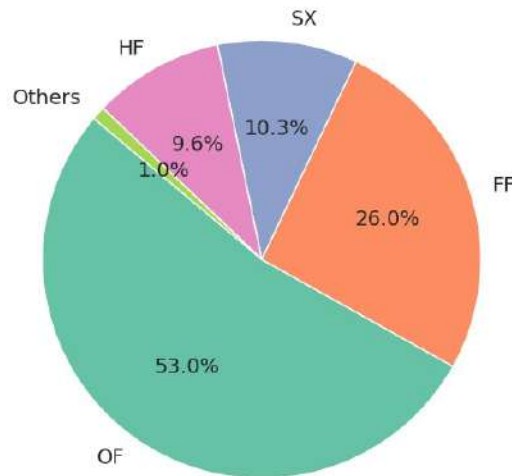


Figure 4.2: Demand quantity distribution by item type (Top 4 + Others).

To understand which types of products dominate the order volume, Figure 4.2 presents a pie chart of the cumulative demand quantity grouped by item type. The analysis confirms that the oil filter category (OF) contributes more than 50% of the total demand, followed by the fuel filter (FF), the special design filter (SX) and the hydraulic filter (HF). Other item types collectively contribute less than 1%, indicating that they play a marginal role in overall forecasting.

4.1.3 Monthly and Yearly Seasonality Patterns

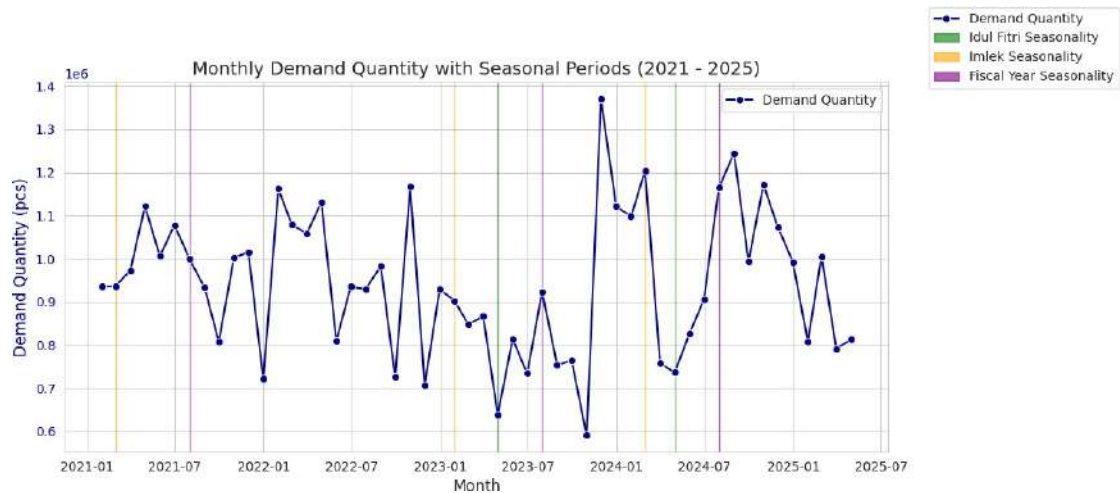


Figure 4.3: Monthly demand quantity with seasonal calendar flags (2021–2025).

To assess the existence of systematic patterns in demand, Figure 4.3 plots monthly demand over time while highlighting known seasonal events: Idul Fitri, Imlek (Chinese New Year) and fiscal year-end periods. In particular, several spikes in demand align with these seasonal flags, especially in the second quarter (April-May) and fourth quarter (October-December), consistent with festive procurement cycles and year-end budget utilization.

4.1.4 Macro-Economic Context: IHSG and Currency Trends

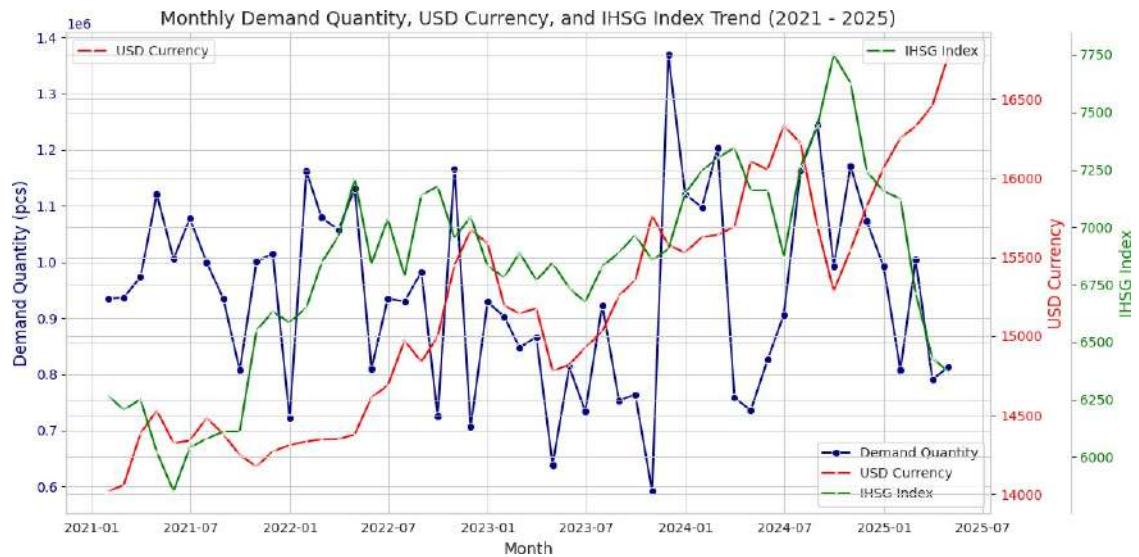


Figure 4.4: Monthly demand quantity vs. USD exchange rate and IHSG index.

To explore potential macroeconomic influences on order behavior, Figure 4.4 overlays the quantity of demand with the USD exchange rate and the Indonesian stock index (IHSG). Although direct causality is not inferred at this stage, preliminary trends suggest some degree of correlation. For example, demand tends to decline during periods of exchange rate depreciation or stock market contraction, indicating the relevance of these external variables for further model training and feature construction.

The EDA confirms that demand patterns are influenced by both internal (product and calendar structure) and external (macroeconomic) factors. These observations will guide the inclusion of appropriate features in the subsequent modeling framework.

4.2 Overview of the Experiment Process

This chapter presents the empirical results of the forecasting experiments conducted on heavy-duty vehicle filter products. The analysis begins with the baseline models, including the Naive and Moving Average approaches, to establish reference performance levels. Subsequently, the results of the machine learning models, Random Forest, LightGBM, and Linear Regression, are reported and compared. Each model is evaluated using the standard accuracy metrics: Mean Absolute Percentage Error (MAPE) and the coefficient of determination (R^2). The results are organized into three stages of experimentation: (i) baseline performance without feature engineering, (ii) performance after incorporating feature engineering, and (iii) performance after applying feature selection via the Feature Take-Out Simulation. This structure enables a clear assessment of how methodological enhancements contribute to the accuracy and robustness of demand forecasting.

4.2.1 Summary of Dataset Used (2021–2025)

The primary dataset utilized in this study comprises customer order records from January 2, 2021 to April 29, 2025. This dataset was curated from historical sales transactions and filtered to ensure consistency, reliability, and relevance for the machine learning forecasting task. The final dataset contains **54,229 rows** and **14 variables**, covering transactional, categorical, seasonal and economic dimensions.

- **sales_date**: The transaction date of each sales record.
- **item_code**: A unique identifier for each Stock Keeping Unit (SKU).

- **item_type**: Product groupings, e.g., FF (Fuel Filter), HF (Hydraulic Filter), OF (Oil Filter), SX (Special Design Filter).
- **sales_qty**: The quantity of product ordered.
- **channel, region, sales_market, customer_name**: Business attributes representing the source of demand.
- **year**: Extracted year from the sales date variable.
- **is_seasonal_idul_fitri, is_seasonal_imlek, is_seasonal_fiscal_year**: Binary flags indicating seasonal or calendar-based influences.
- **x_usd_currency, ihsg_index**: External economic indicators relevant to the market context.

No missing values were found for any of the variables, indicating that the dataset is clean and does not require imputation.

The dataset predominantly comprises four major item types:

- **OF (Oil Filter)**: 23,395 records
- **FF (Fuel Filter)**: 15,007 records
- **HF (Hydraulic Filter)**: 7,425 records
- **SX (Special Design Filter)**: 7,363 records

Other item types such as WC, AY, AX, AF, and GF contribute less than 2% of the total observations and are excluded from the primary forecast analysis.

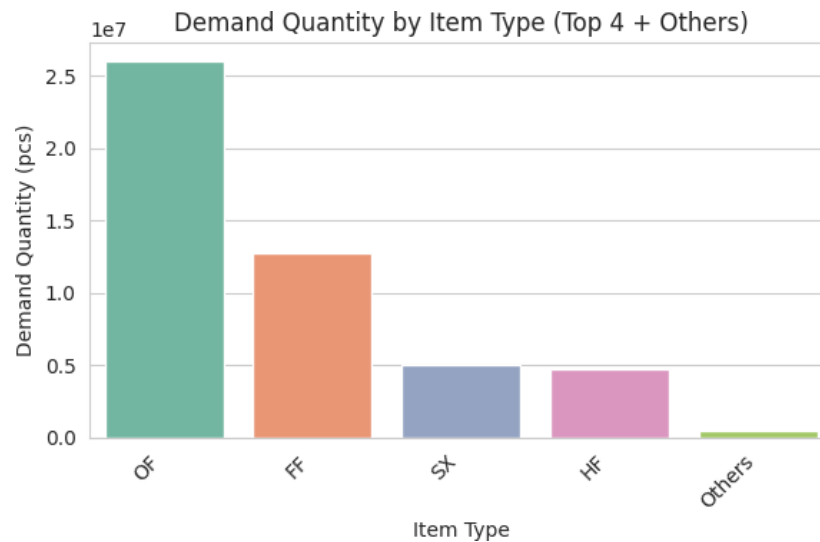


Figure 4.5: Bar chart of item type distribution within the dataset. Major product types (OF, FF, HF, SX) dominate the transaction records.

In Figure 4.5, this distribution justifies the focus on the four major item types (OF, FF, HF, SX) for model training and evaluation. The presence of temporal features and economic variables further supports multivariate modeling and time-sensitive forecasting tasks in later stages.

4.2.2 Reiteration of Modeling Phases: Baseline, Feature Engineering, and Feature Selection

To systematically assess the impact of input transformation on forecast accuracy, the modeling process in this study is divided into three sequential phases: (1) Baseline Modeling, (2) Feature Engineering, and (3) Feature Selection. Each phase builds on the previous one to allow for a clear measurement of the added value provided by new features or dimensionality reduction strategies.

4.3 Baseline Modeling

Two types of statistical baseline, Naïve and Moving Average models that assume demand repetition or averaging between past values are evaluated:

- **Naïve Model:** The Naïve model assumes that demand in the next period is equal to the most recent observed demand.

- **Moving Average Model:** The Moving Average (MA) model assumes that future demand can be estimated by averaging a fixed number of past observations. This model provides a simple but effective baseline for evaluating whether machine learning models significantly improve forecast performance.

Both models are evaluated in four key item types Fuel Filter (FF), Hydraulic Filter (HF), Oil Filter (OF), and Special Design Filter (SX) using the following performance metrics.

- **MAPE (%):** Mean Absolute Percentage Error
- R^2 : Coefficient of determination, indicating the variance explained

4.3.1 Performance Metrics Naïve Forecasting Model

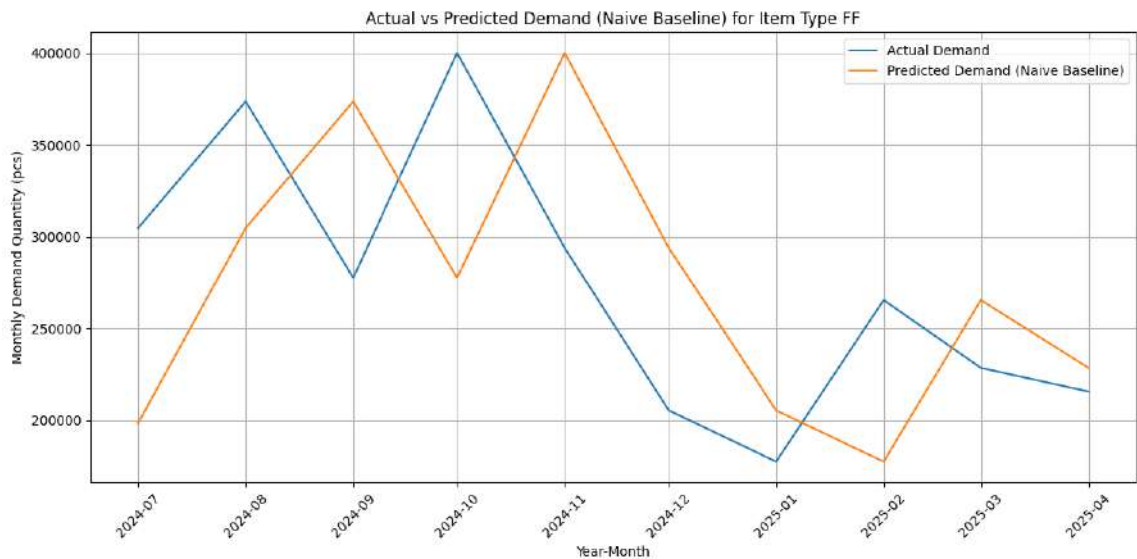


Figure 4.6: Actual demand for Item Type FF shows sharp fluctuations, while the naïve baseline lags behind and fails to capture turning points. This demonstrates the inability of the naïve method to represent volatile demand patterns.

In Figure 4.6, the comparison between actual demand and the naïve baseline for the type of FF item reveals that the naïve method consistently lags behind the trend of actual demand. Although the actual series exhibits sharp fluctuations, peaking in October 2024 and then declining steeply toward early 2025, the naïve baseline produces a smoother trajectory that simply repeats the value of the previous month. As a result, the baseline fails to capture turning points: for instance, it overestimates demand in September 2024 and underestimates demand during the subsequent decline from November 2024 onward. This lag behavior highlights that

the naïve approach cannot adequately represent the volatility of FF demand, reinforcing the need for more advanced forecasting models capable of capturing abrupt changes in trend.

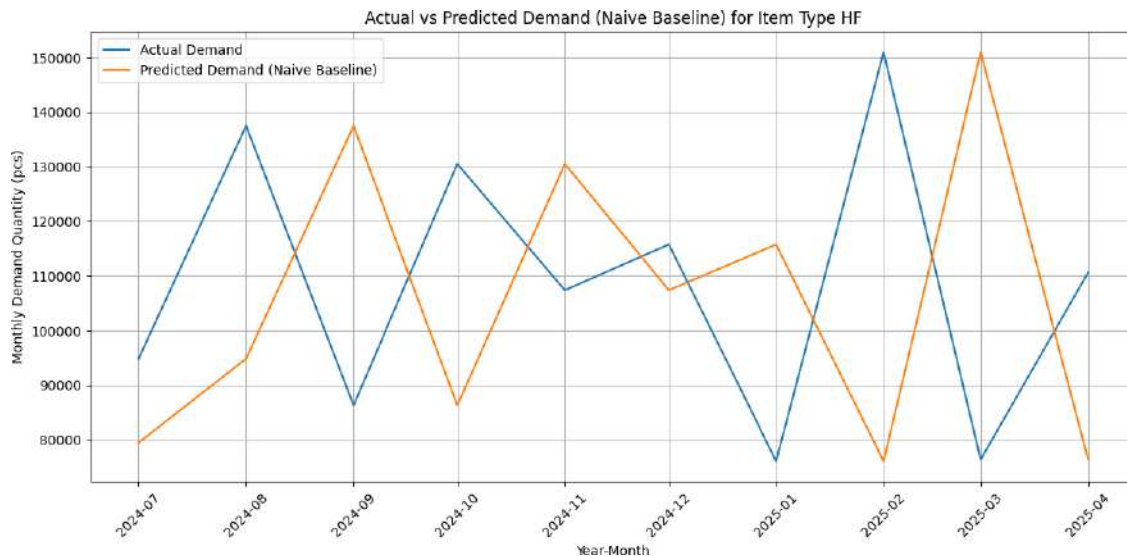


Figure 4.7: Actual demand for Item Type HF exhibits sharp peaks and troughs, while the naïve baseline produces a smoother, lagging response. This lag causes the baseline to miss sudden turning points and underestimate demand volatility.

In Figure 4.7, the comparison between actual demand and the naïve baseline for item type HF shows that the naïve method provides only a rough directional approximation, but fails to capture the high volatility of the series. Actual demand exhibits sharp peaks and deep troughs, such as the sudden rise in August 2024 and February 2025, followed by steep declines in subsequent months. In contrast, the naïve baseline produces a smoother and lag-prone response, as it simply repeats the value of the previous month. Consequently, the baseline systematically misses critical turning points and underestimates the magnitude of demand fluctuations. This reinforces the limitation of the naïve method when applied to volatile product categories such as HF.

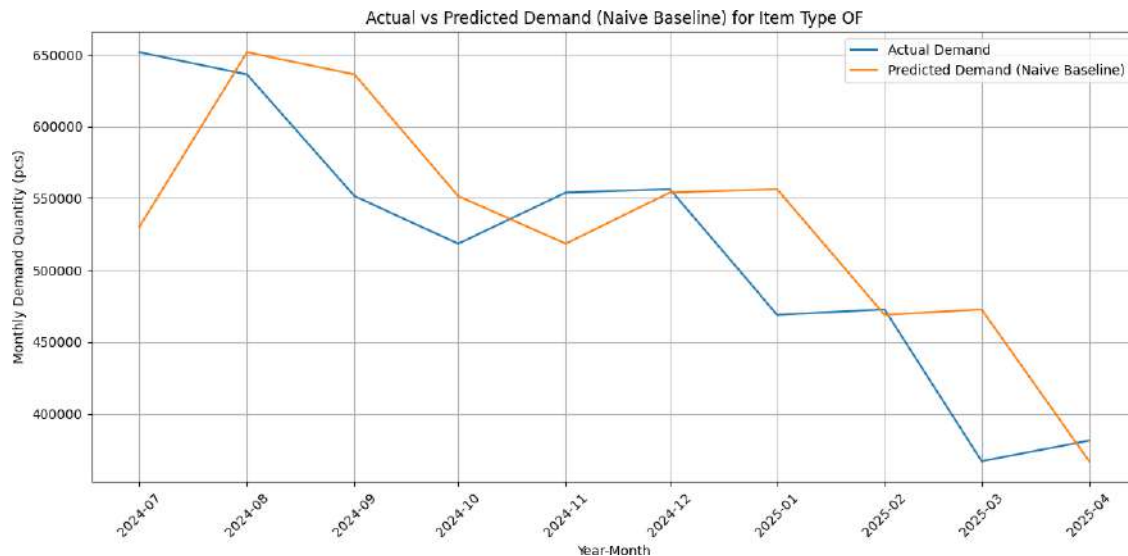


Figure 4.8: Actual and naïve baseline demand for Item Type OF both show a general downward trend, but the baseline tends to overestimate during declines and lags at turning points. While closer to actual demand than in other item types, the naïve method still fails to capture sharper shifts in the trend.

In Figure 4.8, the comparison between actual demand and the naïve baseline for the type OF OF item indicates that both series follow a generally downward trend from mid-2024 to early 2025. Unlike FF and HF, the naïve baseline aligns more closely with the actual demand because OF demand is relatively smoother and less volatile. However, the baseline tends to overestimate demand during periods of decline, such as from August to October 2024, and reacts with a delay when actual demand stabilizes or recovers slightly, as seen in November–December 2024. Although the naïve approach provides a rough approximation of the overall trajectory, it still fails to capture the sharper downturns and recovery points, limiting its usefulness for operational planning.

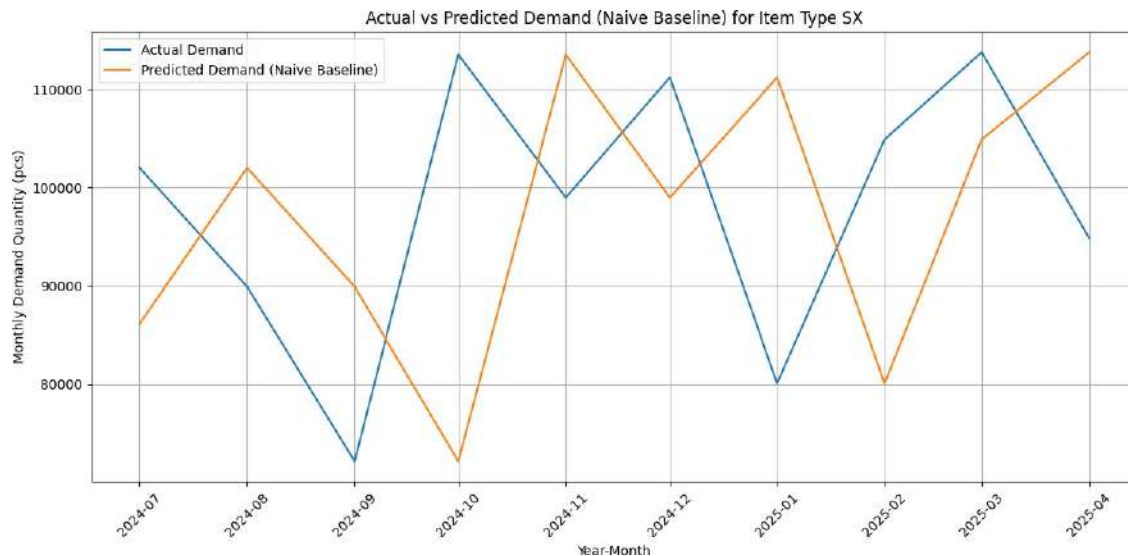


Figure 4.9: Actual demand for Item Type SX fluctuates sharply, while the naïve baseline provides a lagging and smoother approximation. This causes systematic underestimation during upswings and overestimation during downturns, showing the method’s weakness in volatile demand conditions.

In Figure 4.9, the comparison between actual demand and the naïve baseline for the type of item SX highlights the limitations of the naïve approach when applied to highly volatile demand patterns. The actual series exhibits strong fluctuations with sharp peaks and troughs, such as the steep decline in September 2024 followed by a sharp rebound in October 2024. In contrast, the naïve baseline lags behind, underestimating demand during sudden upswings and overestimating during steep declines. Although the baseline follows the general directional movement, it smooths out volatility and systematically misses turning points, demonstrating that it is inadequate to capture the irregular dynamics of SX demand.

Across all types of items, the comparison between actual demand and the naïve baseline reveals consistent weaknesses in the naïve approach. For FF and HF, where demand is highly volatile with sharp peaks and troughs, the naïve baseline systematically lags behind actual demand, leading to frequent overestimation during downturns and underestimation during upswings. SX exhibits even stronger volatility, where the baseline smoothing effect completely fails to capture erratic fluctuations, making it unsuitable for operational use. In contrast, OF demand follows a relatively smoother downward trajectory, allowing the naïve baseline to align more closely with the actual trend, although it still tends to overestimate during declines and react slowly at turning points. Overall, these results demonstrate that while the naïve method may provide a rough approximation in more stable series such as OF, it is inadequate to capture

the dynamics of volatile product families such as FF, HF and SX. This reinforces the need for more advanced forecasting models to achieve meaningful predictive accuracy across different filter categories.

4.3.2 Performance Metrics Moving Average Model

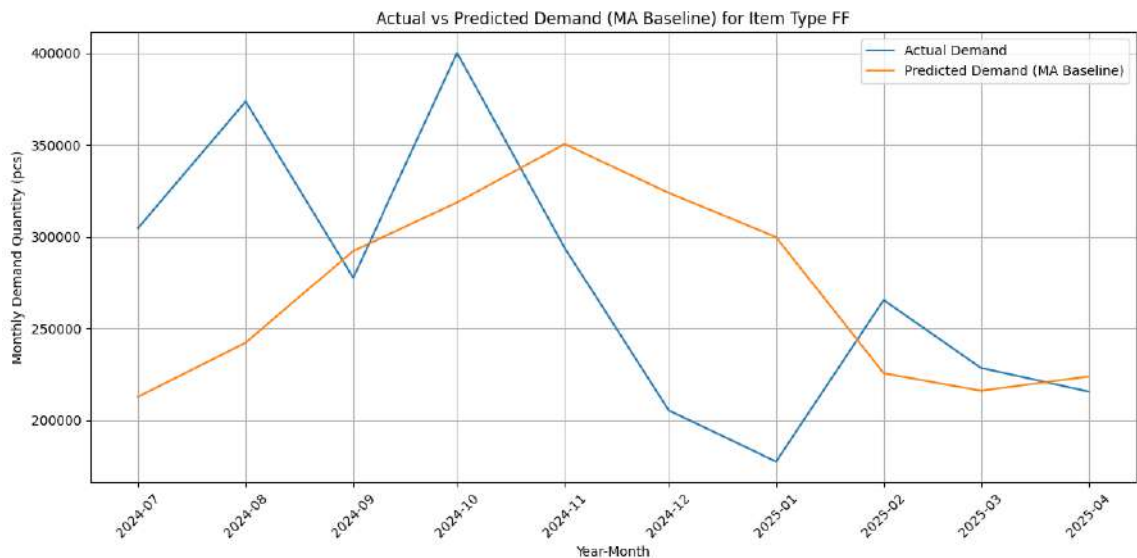


Figure 4.10: Actual demand for Item Type FF shows sharp peaks and troughs, while the moving average baseline smooths these fluctuations into a gradual trend. This results in underestimation during demand surges and overestimation during steep declines, limiting its usefulness in volatile demand conditions.

In Figure 4.10, the comparison between actual demand and the moving average (MA) baseline for Item Type FF shows that the MA baseline captures the overall long-term direction of demand but does not represent its volatility. The actual demand pattern is highly irregular, with sharp peaks such as in October 2024 and steep declines in December 2024 to January 2025. In contrast, the MA baseline produces a smoother trajectory by averaging past values, resulting in systematic underestimation during sudden surges and overestimation during rapid drops. Although the MA approach is useful for highlighting the general directional trend, it does not provide reliable short-term accuracy, making it unsuitable for operational planning in volatile product categories like FF.

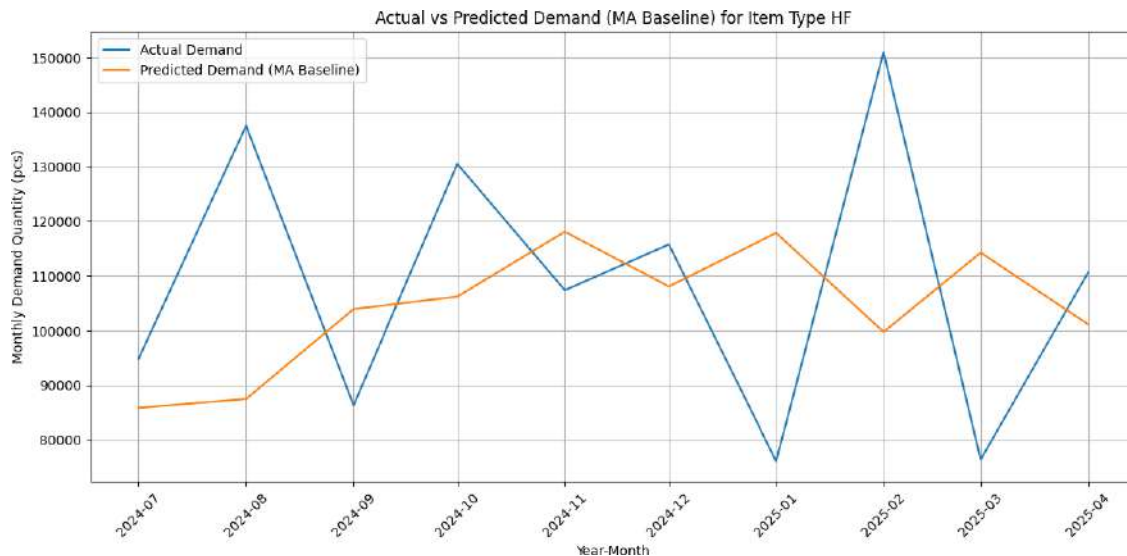


Figure 4.11: Actual demand for Item Type HF shows sharp volatility, while the moving average baseline smooths fluctuations into a gradual trend. This smoothing effect causes underestimation during spikes and overestimation during drops, limiting its suitability for volatile demand conditions.

In Figure 4.11, the comparison between actual demand and the moving average (MA) baseline for Item Type HF highlights the limitations of the MA approach when applied to volatile demand patterns. The actual demand series exhibits sharp fluctuations with abrupt peaks, such as in August 2024 and February 2025, and steep declines in September 2024 and March 2025. In contrast, the MA baseline produces a smoother and more gradual trajectory, dampening these extremes and responding more slowly to sudden changes. This results in systematic underestimation during rapid demand surges and overestimation during sharp declines. Although the MA baseline can reveal the general directional trend, it lacks the responsiveness required to capture the short-term volatility of HF demand, limiting its usefulness for operational forecasting.

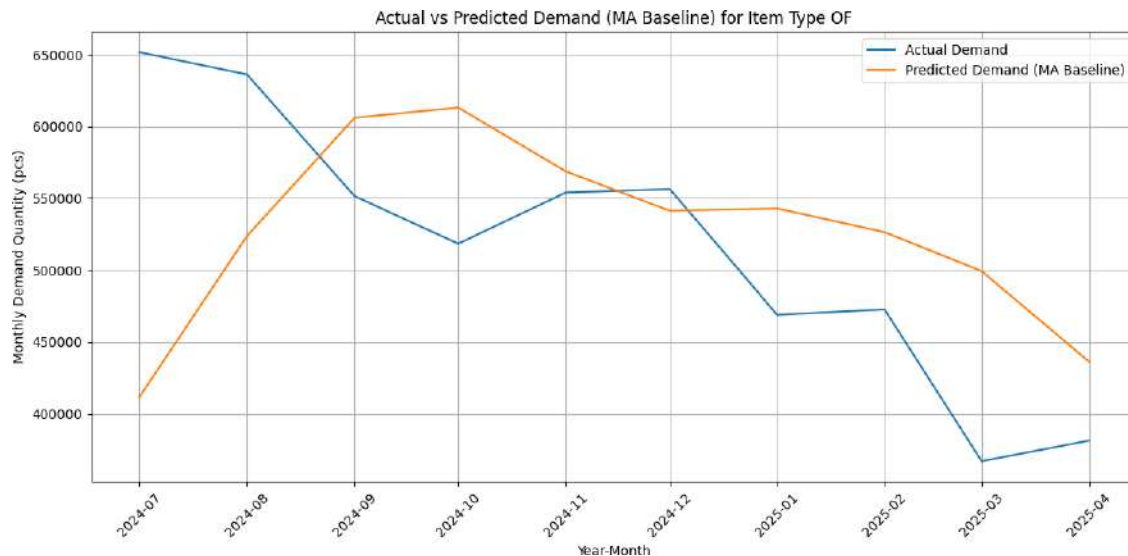


Figure 4.12: Actual and moving average demand for Item Type OF both show a downward trend, but the MA baseline smooths sharper fluctuations into a gradual decline. This leads to overestimation during steep drops and a lagging response during partial recoveries.

In Figure 4.12, the comparison between actual demand and the moving average (MA) baseline for Item Type OF shows that both series follow a general downward trend from mid-2024 to early 2025. However, actual demand exhibits more pronounced declines, particularly in September 2024 and March 2025, whereas the MA baseline smooths these changes into a gradual decline. This smoothing effect reduces the volatility, but also results in systematic overestimation during sharper downturns and a lagging response when actual demand partially recovers, as in November 2024 and April 2025. Although the MA baseline is more aligned with OF demand than in more volatile item types, it still fails to capture the timing and magnitude of sudden changes, limiting its effectiveness for precise forecasting.

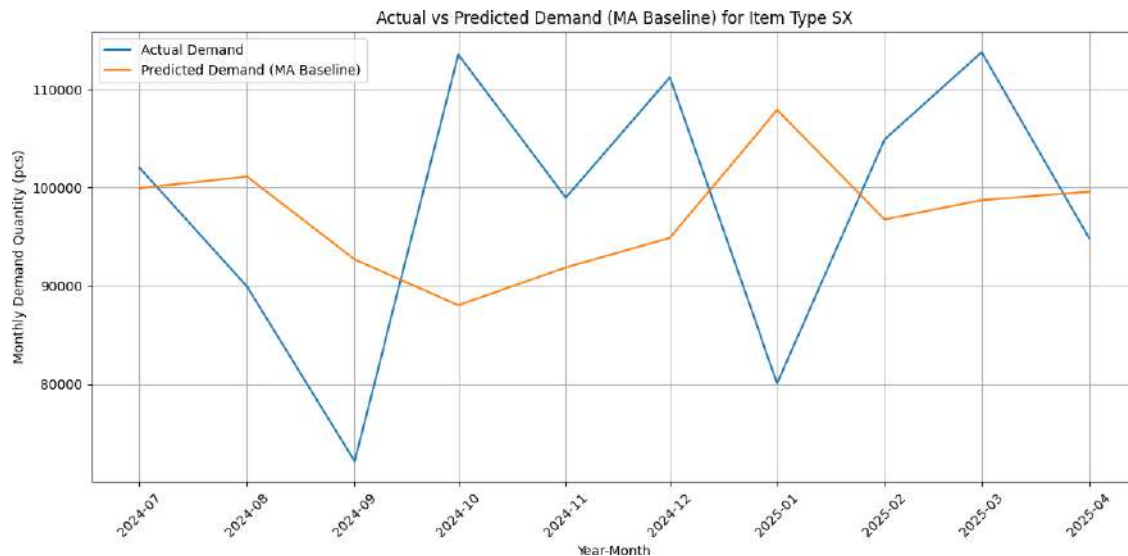


Figure 4.13: Actual demand for Item Type SX is highly volatile, while the moving average baseline smooths fluctuations into a gradual line. This results in underestimation during peaks and overestimation during troughs, showing that the MA approach cannot capture irregular demand patterns.

In Figure 4.13, the comparison between actual demand and the moving average (MA) baseline for Item Type SX shows that the MA baseline is unable to capture the highly volatile nature of SX demand. The actual series exhibits extreme fluctuations, with sharp peaks in October 2024 and March 2025 and steep troughs in September 2024 and January 2025. In contrast, the MA baseline produces a smoother trajectory that averages these abrupt changes, leading to systematic underestimation during demand surges and overestimation during sharp declines. This smoothing and lagging behavior makes the MA approach unsuitable for forecasting short-term demand in irregular and volatile product categories such as SX.

4.3.3 Baseline Selection for Model Evaluation

Table 4.1: Baseline Forecasting Results by Item Type

Item Type	Metrics	Naïve Model	Moving Average
FF (Fuel Filter)	MAPE (%)	26.84	26.08
	R^2	-0.4844	-0.4240
HF (Hydraulic Filter)	MAPE (%)	39.94	24.85
	R^2	-2.5562	-0.6064
OF (Oil Filter)	MAPE (%)	10.16	16.56
	R^2	0.4663	-0.3656
SX (Special Design Filter)	MAPE (%)	20.61	14.80
	R^2	-1.6821	-0.4544

The baseline results summarized in Table 4.1 provide the foundation for evaluating the performance of the proposed machine learning models. For each type of items, the choice of the baseline is determined by identifying the method (naïve or moving average) that produces the lowest forecast error, as indicated by the Mean Absolute Percentage Error (MAPE) and supported by the value R^2 . The selection is explained as follows:

- **FF (Fuel Filter):** The Moving Average model achieves a lower MAPE (26.08%) compared to the naïve model (26.84%), with a slightly better value R^2 . Therefore, the Moving Average is chosen as the baseline.
- **HF (Hydraulic Filter):** The Naïve model performs poorly with a MAPE of 39.94% and a negative R^2 of -2.5562. In contrast, the Moving Average significantly reduces the error (MAPE = 24.85%). Thus, the Moving Average is selected as the baseline.
- **OF (Oil Filter):** The Naïve model clearly outperforms the Moving Average with the lowest MAPE (10.16%) and a positive R^2 (0.4663). Therefore, the Naïve model is used as the baseline for this type of item.
- **SX (Special Design Filter):** The Moving Average demonstrates better performance with a lower MAPE (14.80%) compared to the Naïve model (20.61%). Hence, the Moving Average is selected as the baseline.

In summary, the baselines used for the next stage of model evaluation are as follows. Moving Average for FF, HF and SX, while the Naive model is retained for OF. These baselines provide appropriate benchmarks to assess the predictive improvements offered by machine learning models.

4.4 Baseline Model Performance Using Raw Features

Table 4.2: Performance of ML Models Using Raw Features (2021–2025)

Item Type	Metrics	LightGBM +	Random	Linear
		Raw	Forest + Raw	Regression + Raw
FF (Fuel Filter)	MAPE (%)	20.67	21.39	21.10
	R^2	-0.2927	-0.3693	-0.3394
HF (Hydraulic Filter)	MAPE (%)	21.37	22.85	19.68
	R^2	-0.7458	-0.7971	-0.6915
OF (Oil Filter)	MAPE (%)	15.80	19.48	18.72
	R^2	-0.0797	-0.6058	-0.3254
SX (Special Design Filter)	MAPE (%)	12.53	11.75	13.46
	R^2	-0.0782	-0.5761	-0.2561

Before applying feature engineering and feature selection, three baseline models were constructed, LightGBM, Random Forest, and Linear Regression. Their performance was evaluated for each filter type using the mean absolute percentage error (MAPE) and the coefficient of determination (R^2). The results are presented in Table 4.2.

- **Fuel Filter (FF):** LightGBM achieved the lowest MAPE (20.67%), making it the baseline model for FF, although the values of R^2 remained negative.
- **Hydraulic Filter (HF):** Linear regression outperformed tree-based models with the lowest MAPE (19.68%), indicating it is the best baseline model for HF despite the negative R^2 .
- **Oil Filter (OF):** LightGBM showed the best performance with MAPE of 15.80%, while Random Forest produced weaker results.

- **Special Design Filter (SX):** Random Forest was superior with the lowest MAPE (11.75%) and is thus chosen as the baseline for SX.

Overall, MAPE was used as the main selection criterion, since the R^2 values were consistently negative, indicating that all the baseline models performed worse than predicted the mean of the actual demand. These baseline results serve as a reference point against which improvements in feature engineering and feature selection will be measured in subsequent sections.

Baseline selection summary or the models chosen are:

- FF (Fuel Filter) → LightGBM
- HF (Hydraulic Filter) → Linear Regression
- OF (Oil Filter) → LightGBM
- SX (Special Design Filter) → Random Forest

Transition to Feature Engineering: The baseline results confirm that all models trained with raw features exhibit limited predictive power, as indicated by consistently negative values of R^2 . Although LightGBM, Random Forest, and Linear Regression emerged as the best baseline choice for different filter types based on MAPE, their performance highlights the need to improve the input feature space. Consequently, the next stage of analysis applies *feature engineering* techniques to enrich the dataset with additional explanatory variables, in the expectation of improving model accuracy and capturing demand variability more effectively.

4.5 Feature Engineering

Feature engineering (FE) is a crucial step in enhancing the predictive capability of machine learning models by transforming raw data into more informative representations. In the context of demand forecasting, FE aims to generate features that better capture temporal dynamics, seasonality, and external influences that are not well represented in the raw dataset. Typical approaches include creating lagged features, moving averages, rolling statistics, and incorporating macroeconomic indicators. By systematically enriching the dataset, FE is expected to reduce forecasting errors and provide models with stronger explanatory power, thus addressing the limitations observed in the baseline models.

4.5.1 Feature Engineering Framework

To address the limitations observed in the baseline models, several categories of engineered features were developed. These features were designed to enhance the explanatory power of the dataset and improve the accuracy of demand forecasting models. The framework is structured as follows.

1. **Lag Features** Historical demand values were incorporated as lag features (e.g., demand at $t - 1$, $t - 2$, and $t - 3$). These features capture short-term temporal dependencies and help the model recognize recent demand fluctuations.
2. **Moving Averages** Moving averages (MA) of different window sizes (e.g. MA₁ to MA₆) were introduced to smooth irregularities and represent trends in medium-term demand. These features mitigate volatility and highlight the underlying demand patterns.
3. **Rolling Statistics** Rolling mean, variance, and standard deviation were created in selected windows to capture dynamic changes in demand stability. These statistics provide insight into demand variability and seasonality effects.
4. **Macroeconomic Variables** External variables such as the IHSG index, the USD currency rates were integrated to capture external demand drivers that may not be explained by historical sales alone. These features support the model in recognizing broader market influences.
5. **Calendar Features** Temporal markers such as month, quarter and year were included to represent potential seasonal effects or cyclical patterns in demand. These features enable the model to adjust for systematic calendar-related fluctuations.

This structured approach ensures that the dataset is enriched with meaningful features, balancing both the internal sales history and external environmental factors. The following subsections present the performance of machine learning models after incorporating these engineered features. Machine learning models are retrained on the expanded dataset, allowing an empirical comparison to the baseline performance. Improvements at this stage indicate the value of temporal and contextual signals in forecasting.

4.5.2 Baseline Selection after Feature Engineering

Table 4.3: Comparison of Model Performance: Raw Features vs Feature Engineering (FE). Bold values indicate the best-performing model per item type.

Item Type	Metrics	Best Raw Model	Best FE Model	(FE vs Raw)
FF (Fuel Filter)	MAPE (%)	20.67 (LGBM)	17.09 (LR)	↓ 3.58
	R^2	-0.2927 (LGBM)	0.3664 (LR)	↑ 0.6591
HF (Hydraulic Filter)	MAPE (%)	19.68 (LR)	12.74 (RF)	↓ 6.94
	R^2	-0.6915 (LR)	0.5834 (LR)	↑ 1.2749
OF (Oil Filter)	MAPE (%)	15.80 (LGBM)	8.22 (LR)	↓ 7.58
	R^2	-0.0797 (LGBM)	0.5622 (RF/LR)	↑ 0.6419
SX (Special Design Filter)	MAPE (%)	11.75 (RF)	11.12 (RF)	↓ 0.63
	R^2	-0.0782 (LGBM)	0.1163 (RF)	↑ 0.1945

The delta symbol shows the absolute change from Raw to FE; arrows indicate direction (better for MAPE, better for R^2).

The results demonstrate that the application of feature engineering (FE) significantly improves forecast performance across all item types. For fuel filter, oil filter, and hydraulic filter, FE reduced the MAPE values and transformed R^2 from negative to positive, indicating a significant explanatory power. Even for a special design filter, FE provided more stable results compared to the raw features. Based on these consistent improvements, models with FE are established as the new baseline before proceeding to feature importance evaluation and feature selection.

Feature-engineered models demonstrated superior performance compared to raw features, thereby establishing them as the baseline for subsequent analysis. The following section examines feature importance and applies feature selection to refine the models further.

4.6 Feature Importance Analysis

Feature importance analysis is a crucial step in understanding the relative contribution of different input variables to the predictive performance of the forecasting model. This stage provides interpretability by identifying which features significantly improve accuracy and which features may be redundant or even detrimental. In the context of demand forecasting, this analysis is particularly important because product demand is often influenced by multiple factors, including temporal dynamics, calendar events, and macroeconomic conditions.

In this study, feature importance is assessed through an ablation approach, where groups of engineered features (e.g., lags, moving averages, holiday indicators, fiscal year dummies, and macroeconomic variables) are systematically removed from the Linear Regression with Feature Engineering (LR+FE) model. The resulting error change, expressed as Δ MAE, quantifies the sensitivity of the model to each feature group. A positive Δ MAPE indicates that the removed feature group was important for predictive accuracy, while a negative Δ MAPE suggests redundancy or potential noise.

This analysis not only provides insights into which features drive forecasting performance for each product family, but also lays the foundation for the subsequent stages of feature selection and feature reduction, where the goal is to retain only the most impactful features and streamline the model for practical implementation.

4.6.1 Feature Importance in Linear Regression with Feature Engineering

The ablation analysis presented in Tables 4.4 and Table 4.5 evaluates the contribution of each feature group to the performance of the Linear Regression model with feature engineering (LR + FE). The table reports three sets of results for each product family (Fuel Filter, Hydraulic Filter, Oil Filter, and Special Design Filter): the baseline MAPE and R^2 obtained using all available features, the change in MAPE (Δ MAPE) when a group of features is removed, and the corresponding values of R^2 .

From these results, several important observations can be made.

Table 4.4: Feature Importance (Ablation) Analysis for Linear Regression with Feature Engineering

Item Type	Metrics	LR+FE (all)	(-lag)	(-ma)	(-imlek)	(-idul_fitri)	(-fiscal_year)	(-ihsg_index)	(-usd_currency)	(-month)
FF (Fuel Filter)	MAPE (%)	17.09	18.84	20.38	17.03	17.06	16.55	17.11	17.07	17.16
	Δ MAPE (%)		1.75	3.29	-0.06	-0.03	-0.54	0.02	-0.02	0.07
	R^2	0.3664	0.2649	-0.6981	0.3805	0.3670	0.4070	0.3638	0.3760	0.3790
HF (Hydraulic Filter)	MAPE (%)	13.07	15.34	20.07	12.49	12.41	13.15	12.23	13.04	12.79
	Δ MAPE (%)		2.27	7.00	-0.58	-0.66	0.08	-0.84	-0.03	-0.28
	R^2	0.5834	0.2868	-0.6729	0.5469	0.6151	0.5807	0.5704	0.5823	0.5889
OF (Oil Filter)	MAPE (%)	8.22	9.89	11.83	10.62	7.68	9.14	8.25	8.48	10.14
	Δ MAPE (%)		1.67	3.61	2.40	-0.54	0.92	0.03	0.26	1.92
	R^2	0.5622	0.6488	-0.0228	0.4174	0.5011	0.4420	0.5620	0.5631	0.5368
SX (Special Design Filter)	MAPE (%)	12.12	10.83	14.88	11.62	12.52	11.06	11.78	11.95	11.78
	Δ MAPE (%)		-1.29	2.76	-0.50	0.40	-1.06	-0.34	-0.17	-0.34
	R^2	-0.3248	0.1321	-0.7872	-0.2690	-0.3413	-0.0977	0.0984	-0.1672	-0.2730

Notes: Each ablation column removes the named feature group from the LR+FE model. Δ MAPE is computed as Δ MAPE = MAPE(LR+FE without group) – MAPE(LR+FE all). Positive values indicate worse performance when removed (group is important); negative values indicate improved performance (possible redundancy or noise).

Table 4.5: Feature Importance via Δ MAPE (Linear Regression + FE)

Item Type	(-lag)	(-ma)	(-imlek)	(-idul_fitri)	(-fiscal_year)	(-ihsg_index)	(-usd_currency)	(-month)	Max + Δ
FF (Fuel Filter)	1.75	3.29	-0.06	-0.03	-0.54	0.02	-0.02	0.07	(-ma) 3.29
HF (Hydraulic Filter)	2.27	7.00	-0.58	-0.66	0.08	-0.84	-0.03	-0.28	(-ma) 7.00
OF (Oil Filter)	1.67	3.61	2.40	-0.54	0.92	0.03	0.26	1.92	(-ma) 3.61
SX (Special Design)	-1.29	2.76	-0.50	0.40	-1.06	-0.34	-0.17	-0.34	(-ma) 2.76

Notes: Δ MAPE is computed as Δ MAPE = MAPE(LR+FE without feature group) – MAPE(LR+FE all features). Positive values mean performance worsens when that feature group is removed (i.e., the group is important). Negative values mean performance improves when removed (potential redundancy or noise).

- **Fuel Filter (FF):** Removing moving average (-ma) features leads to the highest degradation in performance (Δ MAPE +3.29, R^2 drops to -0.6981). This confirms that moving averages capture essential temporal patterns in Fuel Filter demand. In contrast, the removal of calendar variables such as Imlek or Idul Fitri has a negligible impact (Δ MAPE -0.06 and -0.03, respectively).
- **Hydraulic Filter (HF):** The sensitivity to moving average features is even stronger, with Δ MAPE rising to +7.00 and R^2 turning negative. This indicates that short-term temporal smoothing is critical for demand predictability in Hydraulic Filters. Fiscal year and currency variables, in contrast, contribute little.
- **Oil Filter (OF):** Model accuracy declines notably when moving averages are excluded (Δ MAPE +3.61), although holiday effects (e.g., Imlek +2.40) also exert some influence. The relatively stable R^2 values suggest that both temporal and event-related features have predictive relevance.
- **Special Design Filter (SX):** The impact of feature removal is mixed. Excluding moving averages again yields the largest deterioration (Δ MAPE +2.76), while removing

lag features unexpectedly improves performance ($\Delta\text{MAPE} -1.29$), suggesting that lag variables may introduce noise or multicollinearity for this item type.

Overall, the ablation results demonstrate that **moving average features consistently provide the strongest predictive power across all product families**, while the relevance of calendar and macroeconomic variables is more item-specific. This provides an empirical foundation for the next stage of the analysis, where feature selection and feature reduction methods will be applied.

4.6.2 Feature Importance in Random Forest with Feature Engineering

In addition to Linear Regression, feature importance analysis was also conducted using the Random Forest (RF) model combined with feature engineering (RF+FE). Random Forest is an ensemble learning method that is capable of capturing nonlinear interactions and variable importance through aggregated decision trees. By systematically removing groups of features and re-evaluating the model, the analysis provides insight into which inputs contribute most to forecasting accuracy.

Table 4.6 presents the results of this ablation study for four product families. Similarly to the linear regression case, performance metrics are expressed as MAPE, ΔMAPE , and R^2 . A positive ΔMAPE indicates a deterioration in accuracy when the feature group is removed, which means that the feature group is important. In contrast, negative values indicate that the removal of a feature group improved the model, suggesting possible redundancy or noise. This analysis is critical to understanding the robustness of Random Forest under different feature configurations and to identify which engineered features drive predictive performance.

The results of the feature importance analysis for the Random Forest model with feature engineering are presented in Tables 4.6 and Table 4.6. Similarly to the linear regression case, the evaluation is based on the Mean Absolute Percentage Error (MAPE), the change in MAPE (ΔMAPE) when a feature group is excluded, and the coefficient of determination (R^2). This analysis highlights which feature groups play the most critical roles in sustaining predictive accuracy across the four product families.

- **Fuel Filter (FF):** The exclusion of lag and moving average features produces the largest increases in error ($\Delta\text{MAPE} +2.54$ and $+2.03$, respectively). In particular, removing moving averages results in a sharp decline in R^2 to -0.7867 , indicating that Random

Table 4.6: Feature Importance (Ablation) Analysis for Random Forest with Feature Engineering

Item Type	Metrics	RF+FE (all)	(-lag)	(-ma)	(-imlek)	(-idul_fitri)	(-fiscal_year)	(-ihsg_index)	(-usd_currency)	(-month)
FF (Fuel Filter)	MAPE (%)	19.76	22.30	21.79	19.74	19.40	19.88	19.43	19.23	19.56
	Δ MAPE (%)		2.54	2.03	-0.02	-0.36	0.12	-0.33	-0.53	-0.20
	R^2	-0.0450	-0.0877	-0.7867	-0.0270	0.0117	-0.0473	0.0368	0.0171	-0.0973
HF (Hydraulic Filter)	MAPE (%)	12.74	14.96	19.06	13.56	13.94	13.15	13.18	13.65	12.20
	Δ MAPE (%)		2.22	6.32	0.82	1.20	0.41	0.44	0.91	-0.54
	R^2	0.3415	0.1629	-0.5897	0.2826	0.2634	0.3140	0.3264	0.2977	0.3426
OF (Oil Filter)	MAPE (%)	13.61	14.78	15.76	13.42	12.21	13.17	13.19	13.44	13.16
	Δ MAPE (%)		1.17	2.15	-0.19	-1.40	-0.44	-0.42	-0.17	-0.45
	R^2	0.2181	0.0571	-0.1738	0.2408	0.1716	0.2308	0.1968	0.3322	0.1992
SX (Special Design Filter)	MAPE (%)	11.12	12.15	13.74	11.70	11.11	10.64	11.21	10.39	10.70
	Δ MAPE (%)		1.03	2.62	0.58	-0.01	-0.48	0.09	-0.73	-0.42
	R^2	0.1163	-0.0447	-0.2606	0.0004	0.0868	0.1236	0.0557	0.1148	0.1488

Notes: Each ablation column removes the named feature group from the RF+FE model. Δ MAPE is computed as Δ MAPE = MAPE(RF+FE without group) – MAPE(RF+FE all). Positive values indicate worse performance when removed (feature group is important); negative values indicate improved performance (possible redundancy/noise).

Table 4.7: Feature Importance via Δ MAPE (Random Forest + FE)

Item Type	(-lag)	(-ma)	(-imlek)	(-idul_fitri)	(-fiscal_year)	(-ihsg_index)	(-usd_currency)	(-month)	Max + Δ
FF (Fuel Filter)	2.54	2.03	-0.02	-0.36	0.12	-0.33	-0.53	-0.20	(-lag) 2.54
HF (Hydraulic Filter)	2.22	6.32	0.82	1.20	0.41	0.44	0.91	-0.54	(-ma) 6.32
OF (Oil Filter)	1.17	2.15	-0.19	-1.40	-0.44	-0.42	-0.17	-0.45	(-ma) 2.15
SX (Special Design)	1.03	2.62	0.58	-0.01	-0.48	0.09	-0.73	-0.42	(-ma) 2.62

Notes: Δ MAPE is computed as Δ MAPE = MAPE(RF+FE without feature group) – MAPE(RF+FE all features). Positive values mean performance worsens when that feature group is removed (i.e., the group is important). Negative values mean performance improves when removed (potential redundancy or noise).

Forest strongly depends on temporal smoothing to capture demand dynamics. Calendar events and macroeconomic variables have smaller or negative Δ MAPE values, reflecting limited or redundant contributions.

- **Hydraulic Filter (HF):** Moving average features again emerge as the most influential, with their exclusion leading to a substantial error increase (Δ MAPE +6.32) and a drop in R^2 to –0.5897. Lag features also play an important role (Δ MAPE +2.22), though their effect is less severe. Other variables, such as fiscal year and USD currency, show only modest contributions, while removing “month” slightly improves accuracy (Δ MAPE – 0.54).
- **Oil Filter (OF):** The Random Forest model demonstrates greater stability for this product family. The largest deterioration is observed when moving averages are removed (Δ MAPE +2.15), while the removal of Imlek (Δ MAPE –0.19) or Idul Fitri (Δ MAPE – 1.40) indicates that certain calendar features may introduce noise rather than predictive value. The R^2 values remain relatively stable, suggesting robustness of the model for this item type.

- **Special Design Filter (SX):** The removal of moving average features produces the greatest performance loss ($\Delta\text{MAPE} +2.62$), underscoring their importance. The exclusion of lag features yields a smaller deterioration ($\Delta\text{MAPE} +1.03$), while some groups such as Idul Fitri, fiscal year, or USD currency yield negative ΔMAPE , suggesting redundancy. Notably, removing the “month” variable also improves accuracy ($\Delta\text{MAPE} -0.42$), pointing to possible collinearity with other temporal features.

A comparative view of the ablation studies between linear regression (LR + FE) and random forest (RF + FE) provides several important insights. Both models consistently identify **moving average features** as the most critical determinants of prediction accuracy in all product families, as their exclusion leads to the highest increases in MAPE and, in many cases, negative R^2 values. This underscores the fundamental role of temporal smoothing in capturing demand patterns.

However, notable differences emerge between the two models. The LR+FE model shows greater instability when key features are removed, with sharper fluctuations in R^2 and more extreme ΔMAPE values. In contrast, RF+FE demonstrates relatively higher robustness, maintaining moderate accuracy even when calendar or macroeconomic features are excluded. Lag features appear relevant for both models, particularly for Fuel and Hydraulic Filters, but they tend to introduce noise in the Special Design Filter case. Calendar events (e.g. Imlek, Idul Fitri) and macroeconomic indicators (e.g. IHSG, USD currency) show mixed effects: in some cases, they add predictive value, while in others they contribute redundancy.

Overall, the comparison confirms that while both linear and nonlinear models heavily rely on moving averages, Random Forest provides more resilience to noisy or redundant features. These findings highlight the importance of prioritizing temporal features in subsequent feature selection, while adopting a cautious and item-specific approach for calendar and macroeconomic variables.

4.7 Feature Selection and Reduction (FS)

In this study, feature selection and reduction are applied to refine the feature space generated during the feature engineering stage. The goal is to identify which variables contribute the most to forecast accuracy and to remove those that are redundant or provide limited predictive value. By narrowing down the feature set, the models become more interpretable, less complex, and more robust for practical implementation.

The complementary approaches are employed: **Feature Take-Out Simulation**. As a custom ablation analysis, this method systematically removes feature groups (e.g., lag variables, moving averages, calendar dummies, and macroeconomic indicators) and measures the change in error, expressed as ΔMAPE . Positive ΔMAPE values indicate that the removed feature group was important, while negative values suggest redundancy or noise. This approach provides direct interpretability and links technical results to business understanding.

In this section, the results of feature selection and reduction are presented for each of the four product families. The analysis is carried out in two complementary steps. First, Random Forest and LightGBM feature importance rankings are reported to show which individual predictors (such as lags, moving averages, calendar events, or macroeconomic variables) contribute the most strongly to forecast accuracy. Second, the results of the feature removal simulation are presented, where groups of features are removed one at a time and the resulting change in error (ΔMAPE) is measured. Positive ΔMAPE values indicate that the excluded group was important, while negative values suggest redundancy or noise. Together, these results provide a clear picture of which predictors are indispensable for reliable demand forecasting.

4.7.1 Feature Importance Rankings of Random Forest

The first step of feature selection was to examine model-based feature importance rankings. The Random Forest (RF) provides an internal measure of how strongly each predictor contributes to error reduction, allowing the identification of the most influential variables at the individual feature level.

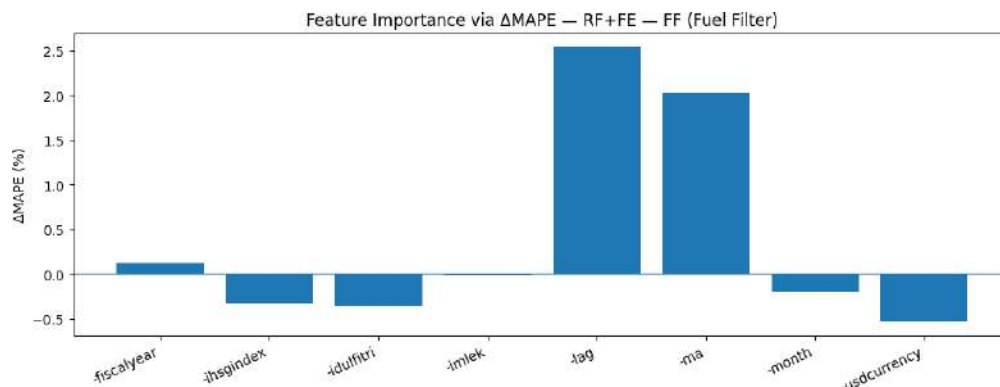


Figure 4.14: Feature importance via ΔMAPE for Fuel Filter (FF) using Random Forest with feature engineering. Temporal features (Lag and MA) dominate the model, while macroeconomic and holiday variables show limited or negative contribution.

As shown in Figure 4.14, the most influential predictors of fuel filter (FF) demand are the **Lag** and **Moving Average (MA)** features, which significantly increase the prediction error when removed (Δ MAPE of approximately +2.5% and +2.0%, respectively). This confirms that the recent demand history and smoothed demand trends are the key drivers of model accuracy.

In contrast, the **Fiscal Year** feature provides only marginal improvement, while **Imlek** contributes almost nothing. Interestingly, several features such as **IHSG Index**, **Idul Fitri**, **Month**, and **USD Currency** yield negative Δ MAPE values, which means that their inclusion introduces noise rather than predictive power.

Overall, the figure highlights that **temporal features (Lag and MA)** should be prioritized in the Random Forest model, while less relevant calendar and macroeconomic features can be excluded without reducing accuracy.

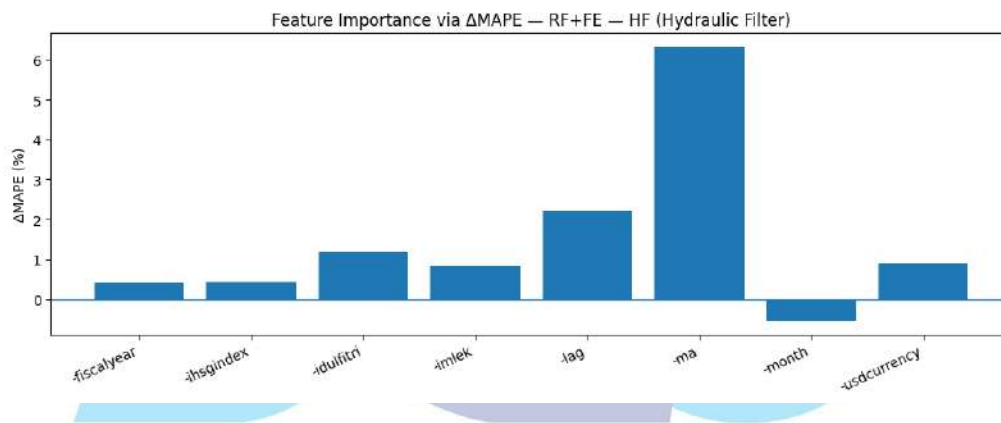


Figure 4.15: Feature importance via Δ MAPE for Hydraulic Filter (HF) using Random Forest with feature engineering. Moving Average and Lag are dominant features, while holiday and macroeconomic variables provide secondary contributions.

As shown in Figure 4.15, the most critical feature for the forecast of hydraulic filters (HF) is the **moving average (MA)**, whose removal increases MAPE by more than 6%. This indicates that smoothed demand trends are essential for capturing the order dynamics of hydraulic filters. The **Lag** feature is also highly influential (Δ MAPE of approximately +2.2%), confirming the importance of recent historical demand in the model.

Several features contribute moderately to accuracy, including **Idul Fitri** (+1.2%), **Imlek** (+0.9%), and **USD Currency** (+0.9%), suggesting that holiday periods and exchange rate movements have measurable effects on hydraulic filter demand. Smaller but positive contributions are observed from **Fiscal Year** and **IHSG Index**, which may capture weak economic and business cycles.

In contrast, the **Month** variable exhibits a negative Δ MAPE, indicating that it introduces noise rather than providing useful seasonal information.

Overall, the figure highlights that the **temporal features (MA and Lag)** should be prioritized in the Random Forest model for Hydraulic Filters, while certain calendar and macroeconomic factors may provide supportive but less dominant signals.

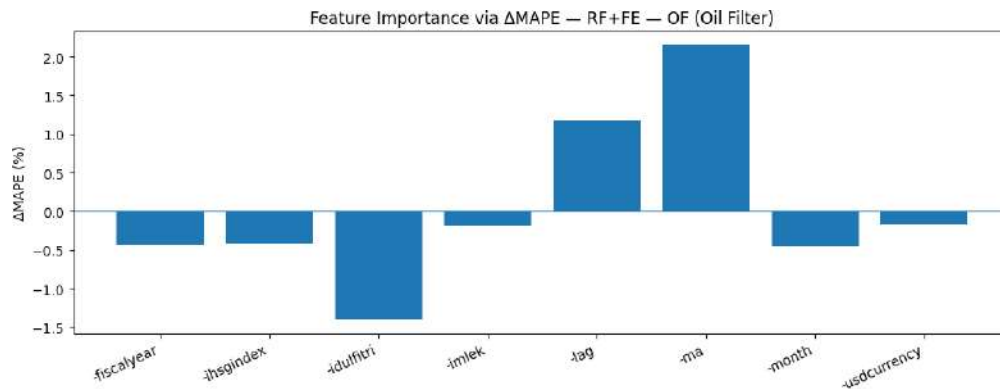


Figure 4.16: Feature importance via Δ MAPE for Oil Filter (OF) using Random Forest with feature engineering. Temporal features (Lag and MA) dominate the model, while calendar and macroeconomic variables provide little or negative contribution.

As shown in Figure 4.16, the most influential predictors of oil filter (OF) forecasting are the features **moving average (MA)** and **lag**. Their removal increases MAPE by approximately +2.1% and +1.2%, respectively, demonstrating that smoothed demand trends and recent historical sales are the strongest drivers of forecast accuracy.

In contrast, most calendar and macroeconomic features provide negligible or negative contributions. The **Idul Fitri** variable shows a notably negative Δ MAPE (around -1.4%), indicating that its inclusion reduces accuracy, as this holiday does not affect oil filter demand. Similarly, features such as **Fiscal Year**, **IHSG Index**, **Month**, **Imlek**, and **USD Currency** yield small negative values, suggesting that they add noise rather than predictive power.

Overall, the figure highlights that **temporal features (MA and Lag)** should be prioritized in the Random Forest model for oil filters, while vacation and macroeconomic factors can be excluded to simplify the model and improve robustness.

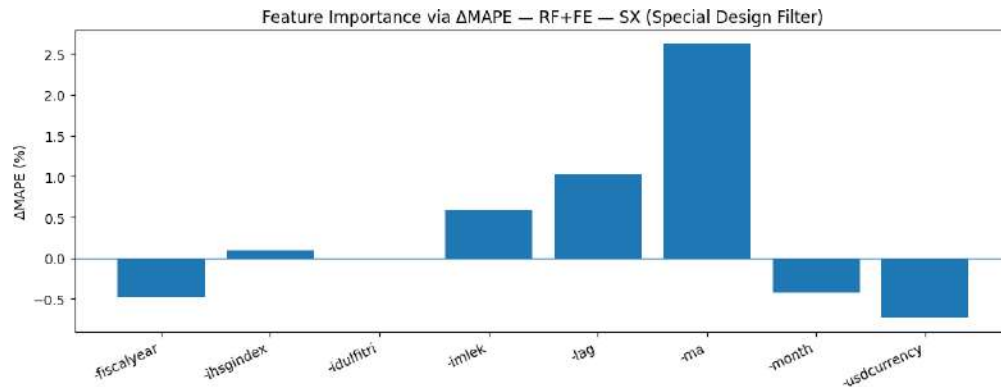


Figure 4.17: Feature importance via Δ MAPE for Special Design Filter (SX) using Random Forest with feature engineering. Temporal features (MA and Lag) dominate, with Imlek providing some seasonal contribution, while fiscal, monthly, and currency features reduce accuracy.

As shown in Figure 4.17, the most influential predictors for Special Design Filter (SX) demand are the **Moving Average (MA)** and **Lag** features. Their removal increases MAPE by approximately +2.6% and +1.0%, respectively, demonstrating that temporal demand patterns are the key drivers of the forecast accuracy for this product family. The **Imlek** holiday also provides a measurable contribution (+0.6%), suggesting some seasonal influence on SX demand.

Other features, such as the **IHSG Index** (+0.1%) and **Idul Fitri** (0.0%), have negligible effects and do not significantly improve accuracy. In contrast, variables such as **Fiscal Year**, **Month**, and **USD Currency** yield negative Δ MAPE values, indicating that they introduce noise rather than a predictive value for this type of product.

Overall, the figure highlights that the **temporal features (MA and Lag)** should be prioritized in the Random Forest model for Special Design Filters, with **Imlek** providing supportive seasonal information, while other calendar and macroeconomic features may be excluded to improve model robustness.

4.7.2 Feature Importance Rankings of Linear Regression

In addition to the tree-based models, this study also evaluates feature importance within the Linear Regression framework. Unlike ensemble methods such as Random Forest, which can capture non-linear relationships and complex interactions, Linear Regression assumes a linear relationship between predictors and the target variable. The importance of features in this context is derived from the magnitude and direction of the estimated regression coefficients,

standardized to allow a meaningful comparison between variables. By examining these coefficients, it becomes possible to identify which features exert the strongest positive or negative influence on demand forecasting performance and to contrast their relative impact with the results obtained from tree-based approaches. This subsection presents the feature rankings from the Linear Regression models for each product family (Fuel Filter, Hydraulic Filter, Oil Filter, and Special Design Filter), providing insights into the explanatory role of individual predictors under a linear modeling assumption.

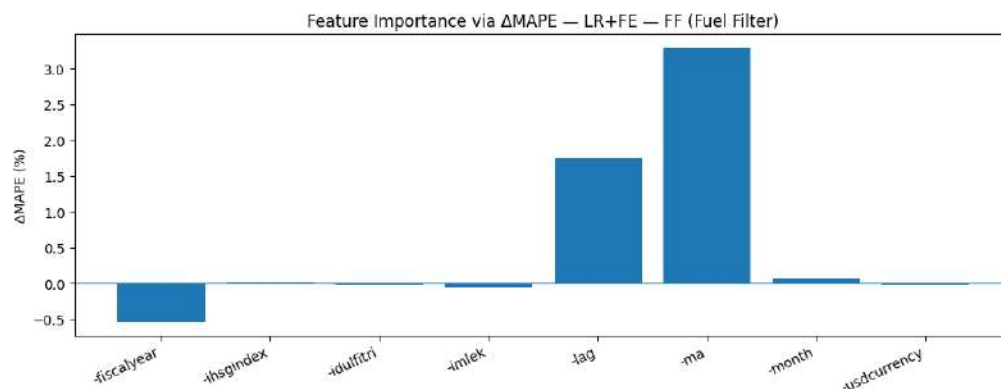


Figure 4.18: Feature importance rankings for Fuel Filter (FF) using Linear Regression. Lag and Moving Average are the most influential predictors, while macroeconomic and calendar features contribute weakly.

As shown in Figure 4.18, the most influential predictors of fuel filter (FF) demand in the Linear Regression model are the **Lag** and **Moving Average (MA)** features. Both exhibit strong positive standardized coefficients, confirming that recent historical sales and smoothed demand patterns are the primary linear drivers of forecast accuracy.

In contrast, the coefficients for macroeconomic indicators such as **IHSG Index** and **USD Currency**, as well as calendar-related features including **Idul Fitri** and **Imlek**, are relatively small and in some cases statistically insignificant. This indicates that their contribution to the accuracy of the forecast in a linear framework is minimal. The **Month** variables likewise do not provide meaningful explanatory power, suggesting that seasonal patterns in Fuel Filter demand cannot be effectively captured by a simple linear specification.

Overall, the Linear Regression results for FF underscore the dominant role of **temporal demand features (Lag and MA)**, while also revealing the limitations of Linear Regression in handling non-linear relationships and feature interactions that are better captured by ensemble methods such as Random Forest.

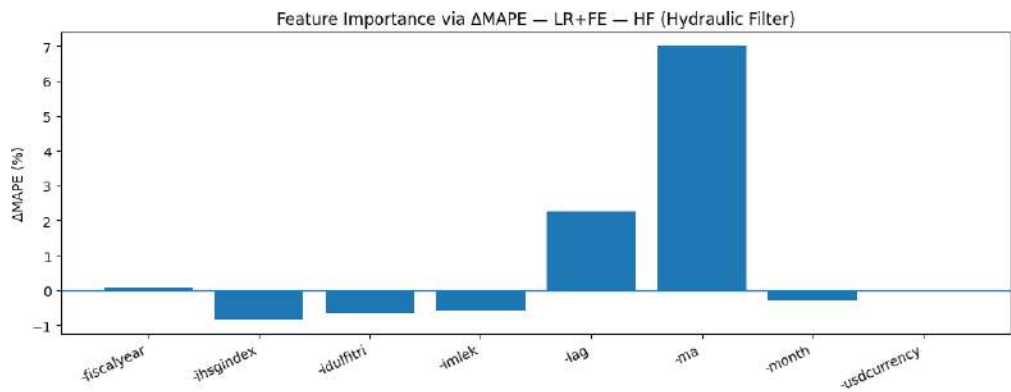


Figure 4.19: Feature importance rankings for Hydraulic Filter (HF) using Linear Regression with feature engineering. Moving Average and Lag dominate the model, while calendar and macroeconomic variables show limited or negative contributions.

As shown in Figure 4.19, the most influential predictors of hydraulic filter (HF) demand in the Linear Regression model are the **Moving Average (MA)** and **Lag** features. Their removal increases MAPE by approximately +7.0% and +2.3%, respectively, confirming that smooth demand patterns and recent historical sales are the strongest linear drivers of forecast accuracy.

In contrast, the contributions of macroeconomic and calendar-related variables are limited or even detrimental. The **Fiscal Year** provides only a marginal improvement, while **Month**, **IHSG Index**, **Idul Fitri**, and **Imlek** all yield negative ΔMAPE values, indicating that their inclusion introduces noise rather than improving predictive power. The **USD Currency** variable shows no measurable effect in the linear framework.

Overall, the results highlight that the **temporal demand features (MA and Lag)** should be prioritized in Linear Regression models for Hydraulic Filters, while most external or seasonal variables can be excluded to simplify the model without loss of accuracy.

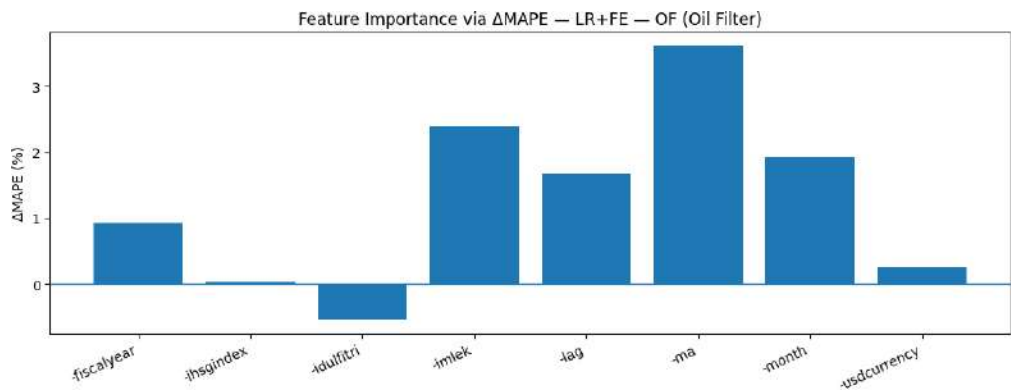


Figure 4.20: Feature importance rankings for Oil Filter (OF) using Linear Regression with feature engineering. Moving Average and Lag dominate the model, while Imlek and Month also emerge as meaningful contributors.

As shown in Figure 4.20, the most influential predictors of oil filter (OF) demand in the Linear Regression model are the **moving average (MA)** and **lag** features. Their removal increases MAPE by approximately +3.6% and +1.7%, respectively, confirming that smooth demand patterns and recent sales history are the strongest linear drivers of forecast accuracy.

Unlike other product families, **Imlek** (+2.4%) and **Month** (+2.0%) also emerge as important predictors, suggesting that Oil Filter demand is more sensitive to seasonal and calendar effects under a linear specification. The **Fiscal Year** (+0.9%) and **USD Currency** (+0.3%) provide small but positive contributions, while the **IHSG Index** has a negligible impact. In contrast, **Idul Fitri** shows a negative ΔMAPE (around −0.5%), indicating that its inclusion reduces the accuracy of the model.

Overall, the results highlight that the **temporal features (MA and Lag)** remain central to Oil Filter forecasting, but the Linear Regression model also places a notable weight on specific seasonal effects, particularly **Imlek** and **Month**, which distinguishes this product family from others.

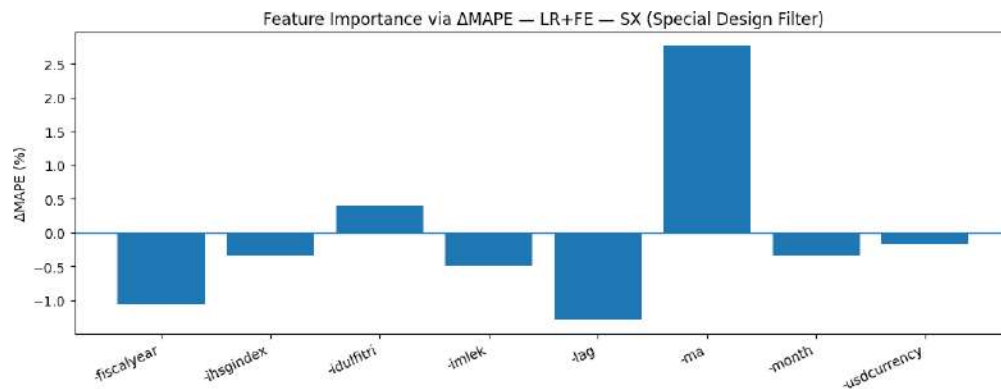


Figure 4.21: Feature importance rankings for Special Design Filter (SX) using Linear Regression with feature engineering. Moving Average is the only dominant predictor, while most other features contribute little or negatively.

As shown in Figure 4.21, the most influential predictor of the demand for special design filters (SX) in the linear regression model is the **Moving Average (MA)** feature. Its removal increases MAPE by approximately +2.8%, confirming that smoothed demand trends are the primary linear driver of the forecast accuracy for this product family. A small positive contribution is observed from **Idul Fitri** (+0.4%), indicating a limited seasonal effect.

In contrast, most other features reduce the accuracy of the model when included. The **Lag** variable shows a negative Δ MAPE (around -1.2%), which differs from its importance in other product families, suggesting that recent sales history does not linearly explain the SX demand. Similarly, **Fiscal Year**, **Imlek**, **Month**, **USD Currency**, and **IHSG Index** all contribute negatively, indicating that annual cycles, holiday effects, macroeconomic factors, and monthly dummies add noise rather than predictive value.

Overall, the linear regression results for SX emphasize that only the feature of **moving average** provides meaningful predictive power, while most other variables are irrelevant or harmful. This highlights the irregular and less predictable demand pattern of SX, which is better suited for non-linear ensemble methods such as Random Forest or LightGBM.

4.7.3 Results of Feature Removal Simulation

The feature removal simulation was conducted to evaluate the robustness of each forecasting model after excluding variables that showed limited or negative contributions to predictive accuracy. By systematically removing weak or noisy predictors, the models were retrained using only the most relevant subsets of features, as determined by the Δ MAPE

analysis. This approach aimed to assess whether reduced feature sets could maintain or improve forecast accuracy while also improving model interpretability and efficiency. The results are presented in Table 4.8, which reports the best performing combinations of features for linear regression and random forest for all types of items, together with their corresponding MAPE values and R^2 .

Table 4.8: Results of Feature Removal Simulation: Best Models and Feature Sets per Item Type

Item Type	Model	Feature Set	MAPE (%)	R^2
FF (Fuel Filter)	Linear Regression	MA, Lag, Month	16.32	0.4387
HF (Hydraulic Filter)	Linear Regression	MA, Lag, Fiscal_Year	10.82	0.6020
OF (Oil Filter)	Linear Regression	MA, Imlek, Month, Lag, Fiscal_Year	7.67	0.5262
SX (Special Design)	Linear Regression	MA, Idul_Fitri	10.43	0.2134
FF (Fuel Filter)	Random Forest	Lag, MA, Fiscal_Year	19.62	0.0260
HF (Hydraulic Filter)	Random Forest	MA, Lag, Idul_Fitri, USD_Currency, Imlek, Fiscal_Year	11.54	0.3827
OF (Oil Filter)	Random Forest	MA, Lag	11.59	0.3848
SX (Special Design)	Random Forest	MA, Lag, Imlek	9.71	0.2489

The results in Table 4.8 highlight that Linear Regression consistently outperformed Random Forest for most product families, particularly FF, HF and OF, where it achieved lower MAPE values and higher R^2 scores. For example, the Oil Filter (OF) obtained the best overall accuracy with a MAPE of 7.67% and a R^2 of 0.5262 when using a reduced feature set consisting of MA, Imlek, Month, Lag, and Fiscal Year. Similarly, the Hydraulic Filter (HF) achieved strong performance with linear regression (MAPE 10.82%, R^2 0.602), confirming the robustness of the temporal and annual cycle features. In contrast, for the Special Design Filter (SX), Random Forest yielded a slightly better configuration (MAPE 9.71%, R^2 0.2489), although overall predictive accuracy remained weaker compared to other item types, reflecting the more irregular demand patterns of this product family.

Overall, these findings indicate that while ensemble methods such as Random Forest can capture complex relationships, Linear Regression with carefully selected features provides more stable and accurate forecasts for most filter types. This emphasizes the importance of feature removal in improving both accuracy and interpretability while also revealing differences in model suitability between product families.

4.8 Selected Best Model Performance

After evaluating all candidate models for both full and reduced feature sets, the next step was to identify the best-performing model for each product family. The selection was based primarily on the Mean Absolute Percentage Error (MAPE) as the main accuracy metric, with R^2 serving as a supporting indicator of model reliability. Models that achieved the lowest MAPE while maintaining reasonable explanatory power were considered the most suitable for forecasting. Table 4.9 presents a summary of the best model selected for each type of item, together with the associated set of characteristics, MAPE and values R^2 .

Table 4.9: Selected Best Model Performance per Item Type

Item Type	Best Model	Selected Feature Set	MAPE (%)	R^2
FF (Fuel Filter)	Linear Regression	MA, Lag, Month	16.32	0.4387
HF (Hydraulic Filter)	Linear Regression	MA, Lag, Fiscal_Year	10.82	0.6020
OF (Oil Filter)	Linear Regression	MA, Imlek, Month, Lag, Fiscal_Year	7.67	0.5262
SX (Special Design)	Random Forest	MA, Lag, Imlek	9.71	0.2489

The results in Table 4.9 indicate that **linear regression emerged as the best performing model for three of four product families** (FF, HF, and OF), consistently achieving lower MAPE values and higher R^2 scores compared to Random Forest. This highlights the effectiveness of a simpler linear framework when combined with carefully selected temporal and seasonal features. In contrast, the **Special Design Filter (SX) performed best under Random Forest**, suggesting that its more irregular demand patterns benefit from the non-linear capability of ensemble methods. Overall, these findings underscore the importance of tailoring the selection of models to the characteristics of each family of products rather than applying a uniform single approach.

Identifying the best-performing models for each product family provides a solid foundation for subsequent analysis. By establishing that Linear Regression delivers superior results for FF, HF, and OF, while Random Forest is more suitable for SX, the study ensures that the forecast accuracy is optimized according to the product-specific demand characteristics. Building on these selected models, the next section evaluates the research hypotheses by examining whether the application of machine learning techniques and feature selection strategies significantly improves forecast accuracy compared to baseline approaches. This transition from model performance to hypothesis testing strengthens the empirical validation of the study framework and its implications for demand forecasting in the heavy-duty filter industry.

4.8.1 Best Model per Item Type

Based on the results of the feature removal simulation, the best-performing model was identified for each product family. The selection was determined by the lowest MAPE values and the highest R^2 scores. The summary of best models per item type is as follows:

- **FF (Fuel Filter):** The best model is *linear regression* with the feature set *MA, Lag, Month*, achieving a MAPE of **16.32%** and an R^2 of **0.4387**.
- **HF (Hydraulic Filter):** The best model is *linear regression* with the feature set *MA, Lag, Fiscal_Year*, achieving a MAPE of **10.82%** and an R^2 of **0.6020**.
- **OF (Oil Filter):** The best model is *linear regression* with the feature set *MA, Imlek, Month, Lag, Fiscal_Year*, achieving a MAPE of **7.67%** and an R^2 of **0.5662**.
- **SX (Special Design Filter):** The best model is *Random Forest* with the feature set *MA, Lag, Imlek*, achieving a MAPE of **9.71%** and an R^2 of **0.2489**.

These results demonstrate that Linear Regression dominates as the most suitable model for three item types, whereas Random Forest provides better performance for the Special Design Filter, which exhibits more irregular demand patterns.

4.9 Cross Model Comparison

The preceding subsection identified the best-performing model for each product family. To complement these findings, this section compares the models between item types using two key indicators. Mean Absolute Percentage Error (MAPE) as the primary measure of accuracy and the coefficient of determination (R^2) as the model fit measure. This cross-model comparison provides an integrated view of how Linear Regression and Random Forest perform relative to each other across different filter categories. The purpose of this analysis is not only to confirm the consistency of model selection but also to prepare for the subsequent step of visual evaluation. In the next section, forecast plots will be presented (Actual vs. Best Model vs. Baseline).

Table 4.10: Cross Model Comparison: Baseline vs. Selected Best Model (MAPE and R^2)

Item Type	Model	MAPE (%)	R^2
FF (Fuel Filter)	Moving Average (baseline)	26.08	-0.4240
FF (Fuel Filter)	Moving Average (best model)	16.32	0.4387
HF (Hydraulic Filter)	Moving Average (baseline)	24.85	-0.6064
HF (Hydraulic Filter)	Moving Average (best model)	10.82	0.6020
OF (Oil Filter)	Naïve (baseline)	10.16	0.4663
OF (Oil Filter)	Moving Average (best model)	7.67	0.5262
SX (Special Design Filter)	Moving Average (baseline)	14.80	-0.4544
SX (Special Design Filter)	Random Forest (best model)	9.71	0.2489

Table 4.10 presents the comparison between the baseline methods (Naïve or moving average) and the best models selected for each family of products using MAPE and R^2 as evaluation metrics. The results demonstrate clear improvements across all item types when moving from baseline to the best performing model. For the **Fuel Filter (FF)**, the best model reduced the MAPE from 26.08% to 16.32% (approximately a 37% improvement), with R^2 increasing from -0.4240 to 0.4387. The **hydraulic filter (HF)** shows the largest gain, since the MAPE decreased from 24.85% to 10.82% (approximate improvement 56%, while R^2 improved substantially from -0.6064 to 0.6020. In the case of the **Oil Filter (OF)**, the Naïve baseline produced a MAPE of 10.16% with R^2 of 0.4663, while the best model further improved performance to a MAPE of 7.67% and R^2 of 0.5262. For the **Special Design Filter (SX)**, the Random Forest model achieved better accuracy than the baseline, lowering the MAPE from 14.80% to 9.71% (about 34% improvement) and increasing R^2 from -0.4544 to 0.2489.

Overall, these findings confirm that the best selected models significantly outperform the baselines. In particular, for FF, HF, and SX, the negative values R^2 of the baselines indicate that the forecasts were worse than simply predicting the mean demand, while the best models corrected for this problem by producing positive values R^2 , providing meaningful explanatory power. This validates the effectiveness of model selection and feature refinement in improving forecast accuracy.

4.9.1 Visual Forecast Comparison

Beyond numerical metrics such as MAPE and R^2 , it is equally important to assess how models capture the demand dynamics over time. This section presents a visual comparison between the **actual demand**, the **best-performing model forecasts**, and the **baseline forecasts** for each product family. The visualizations allow for a clearer interpretation of how well the selected models track fluctuations, trends, and irregular patterns in customer orders compared to simple benchmark approaches. This approach highlights not only the accuracy but also the practical reliability of the chosen forecasting models in replicating real-world demand behavior.

Forecast plots for each type of item (actual versus predicted) are shown in Figures 4.22, Figures 4.23, Figures 4.24, Figures 4.25 illustrating the ability of selected models to track demand cycles over time.

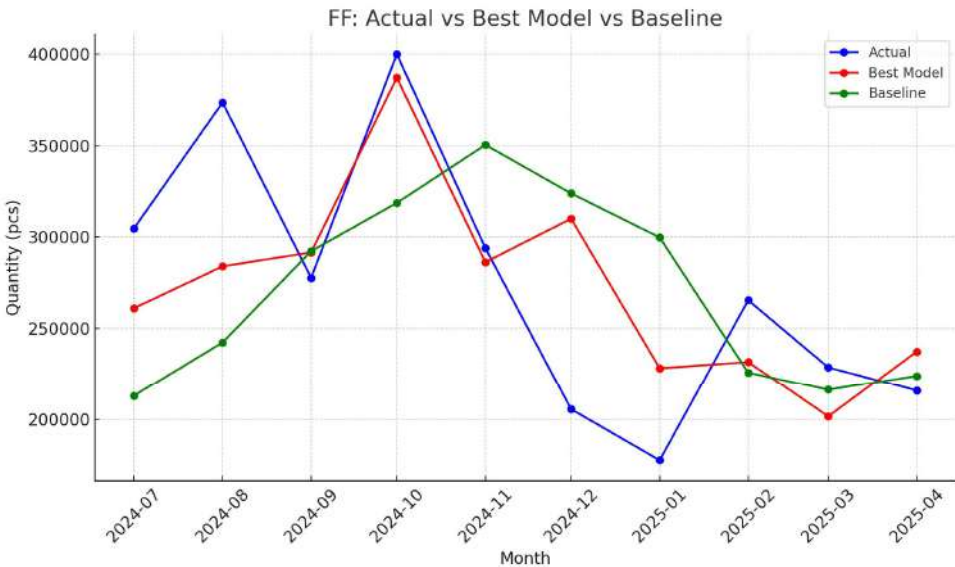


Figure 4.22: Actual demand for Item Type FF fluctuates sharply, while the baseline shows only a gradual rise and fall. The best model follows the actual trend more closely, capturing peaks and declines with greater responsiveness than the baseline.

In Figure 4.22, the comparison between actual demand, baseline, and the best model for Item Type FF demonstrates clear differences in how each approach represents the trend. The actual demand is highly volatile, with sharp peaks in August and October 2024 followed by steep declines in December 2024 and January 2025, and a partial recovery in February 2025. The baseline produces a smoother trajectory, rising gradually until October 2024 and then declining steadily, but it fails to capture the magnitude and timing of sudden turning points. In contrast, the best model follows the overall shape of actual demand more closely, capturing

the October 2024 peak and subsequent downward movement with greater accuracy, although it still underestimates the extremes. This indicates that the best model is substantially more effective than the baseline in representing the dynamics of FF demand, particularly during periods of rapid change.

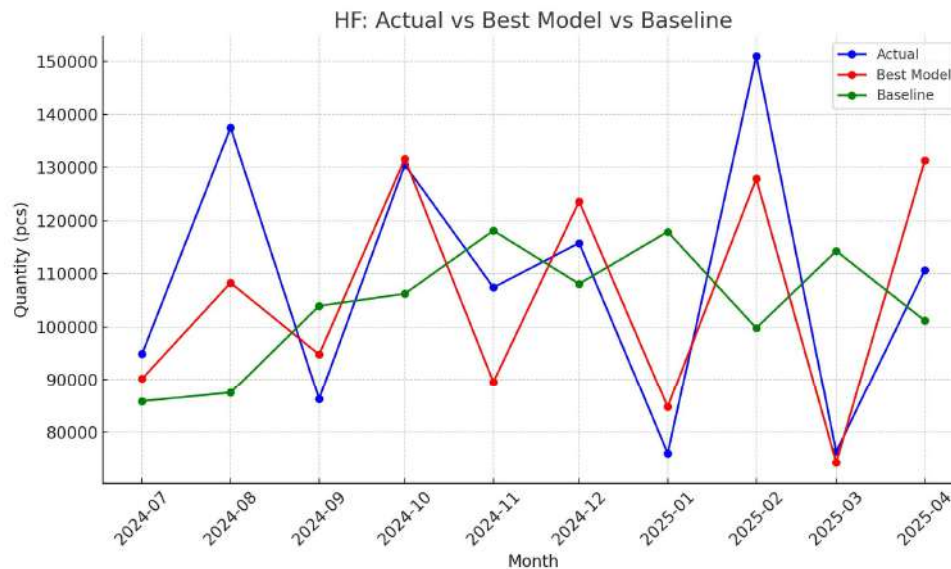


Figure 4.23: Actual demand for Item Type HF is highly volatile, with sharp peaks and troughs. The best model follows these turning points more closely than the baseline, which smooths the series and fails to capture sudden shifts.

In Figure 4.23, the comparison between actual demand, baseline, and the best model for Item Type HF highlights the different ways each approach represents the trend. Actual demand is highly volatile, with sharp peaks in August 2024, October 2024, and February 2025, and steep troughs in September 2024, January 2025, and March 2025. The baseline produces a smoother and more stable trajectory, gradually increasing until late 2024 and then fluctuating slightly, but it fails to capture the magnitude and timing of sudden surges or collapses. In contrast, the best model follows the direction of actual demand more closely, reproducing the timing of several major peaks and troughs with greater accuracy, although the extremes are still underestimated. This shows that while the baseline oversimplifies demand dynamics, the best model provides a more realistic representation of HF demand, particularly during periods of rapid change.

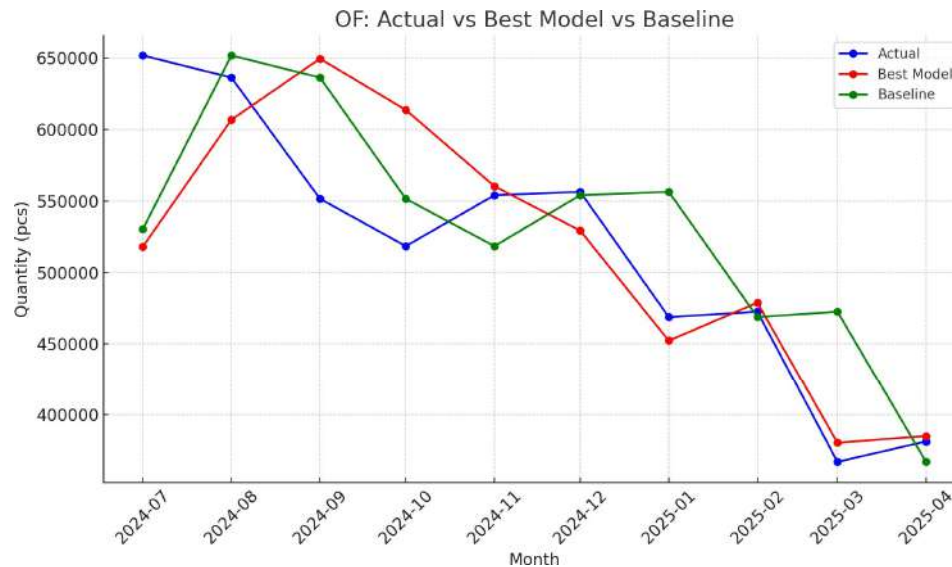


Figure 4.24: Actual demand for Item Type OF shows a general downward trend with mild fluctuations. The best model follows this trend more closely than the baseline, which smooths the series and tends to overestimate during sharper declines.

In Figure 4.24, the comparison between actual demand, baseline, and the best model for Item Type OF shows that all three series follow a general downward trend from mid-2024 to early 2025. The actual demand declines steadily with some fluctuations, while the baseline produces a smoother trajectory that slightly overestimates demand during sharper downturns, such as in September 2024 and March 2025, and lags at minor recovery points. The best model, in contrast, tracks actual demand more closely than the baseline, particularly during mid to late 2024, and captures both the overall decline and the small rebounds observed in November to December 2024. However, it still underestimates the steep decline in early 2025. This shows that while the baseline provides only a broad directional trend, the best model offers a more responsive and reliable approximation of OF demand.

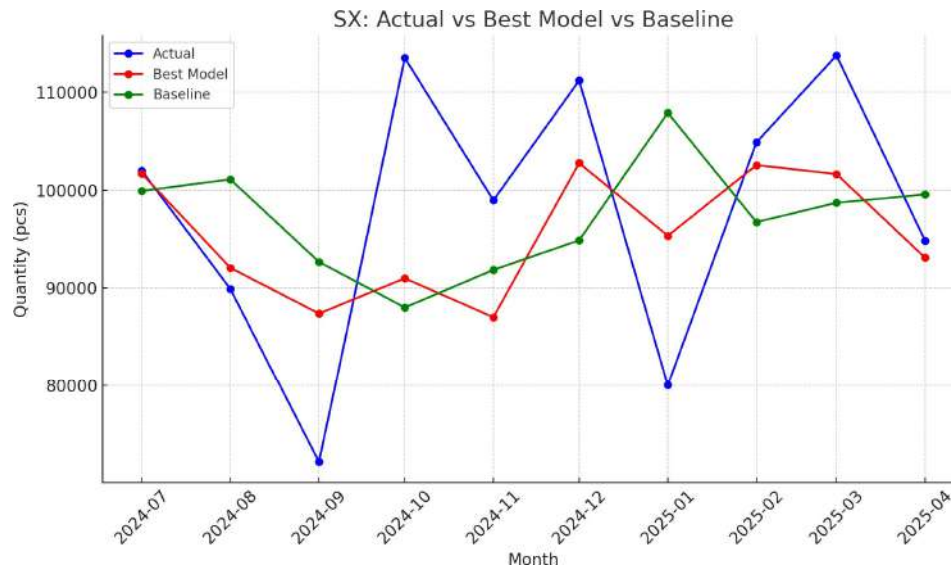


Figure 4.25: Actual demand for Item Type SX is highly volatile with sharp peaks and troughs. The best model tracks these fluctuations more closely than the baseline, which smooths the series and misses extreme variations.

In Figure 4.25, the comparison between actual demand, baseline, and the best model for Item Type SX highlights the challenges of forecasting highly volatile demand. The actual demand exhibits sharp fluctuations, with steep troughs in September 2024 and January 2025 and pronounced peaks in October 2024 and March 2025. The baseline produces a smoother trajectory, staying around the mid range of values, leading to consistent overestimation during troughs and underestimation during peaks. In contrast, the best model follows the overall direction of actual demand more closely, capturing the timing of major peaks and troughs with greater accuracy, although the magnitudes of extreme changes remain underestimated. This shows that while the baseline oversimplifies the irregular behavior of the SX demand, the best model provides a more realistic approximation of its volatility.

4.10 Insight Generation

This section synthesizes the findings of the best-performing models across all filter product families. The discussion highlights the most influential contributing factors that shaped the accuracy of the forecast and provides information on the observed forecast behavior for each type of item.

4.10.1 Top Contributing Factors

The analysis confirms that temporal and contextual features play different roles in product families.

- **FF (Fuel Filter):** The feature set of *MA*, *Lag*, and *Month* was the most effective to capture the variation in demand. The reliance on these features indicates that demand patterns for fuel filters are largely shaped by short-term historical sales trends (moving average and lag effects), with additional influence from seasonal monthly cycles.
- **HF (Hydraulic Filter):** The inclusion of *Fiscal_Year* alongside *MA* and *Lag* significantly improved the model accuracy. This suggests that the demand for hydraulic filters follows fiscal and budget cycles in addition to short-term consumption trends.
- **OF (Oil Filter):** The best results were achieved using a richer set of features for *MA*, *Imlek*, *Month*, *Lag*, and *Fiscal Year*. The strong contribution of *Imlek* (Chinese New Year) implies that cultural and festive events drive increases in the demand for oil filters. Coupled with monthly and fiscal cycles, these factors highlight the multifaceted drivers of oil filter consumption.
- **SX (Special Design Filter):** The most effective model leveraged *MA*, *Lag*, and *Imlek*, but achieved only moderate explanatory power ($R^2 = 0.2489$). Although short-term dynamics and festive events provide partial signals, the relatively low R^2 suggests that the demand for special design filters is heavily project-based or customer-specific, reducing the relevance of standard temporal features.

4.10.2 Forecast Behavior Insight

The forecast performance across models provides meaningful insights into the nature of demand behavior:

- **FF (Fuel Filter):** The forecast captured seasonal variation and short-term fluctuations, but residual volatility remained. This reflects that fuel filter demand is somewhat predictable but subject to irregular order surges, likely linked to maintenance schedules or operational cycles.

- **HF (Hydraulic Filter):** The forecasts aligned closely with actual sales, supported by the integration of fiscal-year effects. This shows that the demand for hydraulic filters is relatively structured, shaped by annual procurement cycles and planned maintenance, making it more predictable compared to fuel filters.
- **OF (Oil Filter):** The model captured both long-term trends and culturally driven demand peaks. The forecast accuracy was the strongest among all item types (MAPE of **7.67%**), highlighting that the demand for oil filters follows recognizable temporal and cultural drivers, which the model effectively leveraged.
- **SX (Special Design Filter):** Although forecast errors were relatively low (MAPE of **9.71%**), the explanatory power remained limited. This indicates that forecasts capture average demand levels but cannot account for variance driven by bespoke customer projects or one-off orders. As such, the forecast is useful for planning at an aggregate level, but less reliable for precise month-to-month projections.

Overall, the results emphasize that the tailoring of the features to each product family is essential to maximize forecast accuracy. Although moving averages and lag features consistently form the backbone of predictive power, fiscal cycles, and cultural events provide critical complementary signals, especially for oil and hydraulic filters. In contrast, special design filters require alternative approaches, potentially incorporating project-specific or customer-order data, to improve explanatory performance.

4.11 Limitations and Future Opportunities

Although the forecasting framework presented in this study demonstrates substantial improvements over traditional methods, several limitations remain that open avenues for further research and development. This section discusses the known constraints of the current approach and outlines potential opportunities to extend the work in future implementations.

4.11.1 Limitations of the Current Modeling

- **Single-Step Forecasting Horizon:** The current models are designed to predict customer orders one month in advance. Although suitable for short-term tactical planning, this approach does not address medium to long-term planning needs, such as allocation of production capacity or strategic procurement.

- **Manual Feature Engineering Dependency:** The modeling pipeline relies heavily on manually engineered features (lags, moving averages, calendar flags), which require domain knowledge and iterative testing. Although effective, this approach limits scalability when applied across a broader product portfolio or changing operational contexts.
- **Static Modeling Assumptions:** Model training and evaluation are based on fixed historical data (2021–2025). Demand shifts caused by unforeseen events (e.g., economic shocks, regulation changes) are not dynamically reflected unless the model is re-trained.
- **Limited Granularity for Lead Time or Promotion Effects:** While calendar events (e.g. holidays) are considered, customer-specific lead times, sales campaigns, or urgent procurement patterns are not explicitly modeled, which can lead to underestimation during sudden demand spikes.

4.11.2 Suggestions for Improving External Data Integration

The inclusion of external variables such as macroeconomic indicators (e.g., IHSG index, USD exchange rate) has not shown any value. However, the use of external data in this study is preliminary and limited in scope.

Future studies may benefit from:

- **More granular and time-aligned external data:** Incorporating weekly or even daily economic indicators, commodity prices (e.g., steel, oil) and market sentiment indices could further explain demand volatility.
- **Incorporation of sector-specific indicators:** For example, construction permits, mining investment indices, or fleet maintenance cycles may provide contextually rich signals to forecast demand in B2B industrial sectors.
- **Customer Behavior Logs:** Including data such as request volumes, quotation frequencies, and ordering history at the customer level may improve forecast precision, especially for irregular or project-driven demand patterns.

4.11.3 Potential for Multistep and Real-Time Forecasting

To enhance the applicability of the forecasting system across different planning horizons and business use cases, future research can explore the following advancements:

- **Multistep Forecasting:** Extending the forecast window to 3 months, 6 months, or 12 months would support mid-term procurement, raw material planning, and financial forecasting. Sequence-based models such as recursive ML or encoder-decoder architectures may be investigated for this purpose.
- **Real-Time Forecasting Pipelines:** Integrating the model into an automated pipeline (e.g., through APIs or ERP connectors) allows real-time forecasting updates as new data become available. This enables adaptive planning and improves organizational agility.
- **Probabilistic and Interval Forecasting:** Rather than generating point estimates, future models could output confidence intervals or prediction distributions to reflect uncertainty, which is crucial for risk-aware inventory policies.

Collectively, these future enhancements can make the forecasting system not only more accurate but also more actionable, integrated, and resilient to uncertainty in dynamic industrial environments.