Use of Forward Indicators to Support Demand Planning for Tempur-Sealy International

by

Shane McGorty
BS Business and Economics, Lehigh University, 2022

and

Xiaoli Yang
MS Industrial Engineering, Northeastern University, 2016

SUBMITTED TO THE PROGRAM IN SUPPLY CHAIN MANAGEMENT
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE IN SUPPLY CHAIN MANAGEMENT
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2025

Signature of Author: _____
Shane McGorty
Department of Supply Chain Management
May 9, 2025

Signature of Author: _____
Xiaoli Yang
Department of Supply Chain Management
May 9, 2025

Certified by: _____
Selene Silvestri
Research Scientist at MIT Center of Transportation & Logistics
Capstone Advisor

Accepted by: _____
Prof. Yossi Sheffi
Director, Center for Transportation and Logistics
Elisha Gray II Professor of Engineering Systems
Professor, Civil and Environmental Engineering

Use of Forward Indicators to Support Demand Planning for Tempur-Sealy International

by

Shane McGorty

and

Xiaoli Yang

Submitted to the Program in Supply Chain Management
on May 9, 2025 in Partial Fulfillment of the
Requirements for the Degree of Master of Applied Science in Supply Chain Management

ABSTRACT

This study provides a framework for the evaluation of third-party datasets that can serve as valuable forward indicators for demand forecasting at Tempur Sealy International. Using a multi-step methodology involving exogenous feature selection, lag analysis, and baseline LSTM modeling, the study evaluates the extent to which third-party datasets enhance forecast performance across various product and channel combinations. While some indicators, such as consumer sentiment and retail activity, exhibit predictive value, the overall results are mixed with improvements in accuracy depending heavily on variable selection and context. These findings underscore the need for rigorous feature evaluation, model customization, and ongoing validation when incorporating exogenous data into demand planning processes.

Capstone Advisor: Selene Silvestri

Title: Research Scientist at MIT Center of Transportation & Logistics

ACKNOWLEDGMENTS

We would like to express our deepest gratitude to those who have supported us throughout the completion of this capstone project. First and foremost, we are truly grateful to our advisor, **Selene Silvestri**, for her trust, commitment, and encouragement. Her unwavering support, thoughtful guidance, and deep sense of responsibility were instrumental to our progress.

We are equally grateful to our sponsor company, Tempur Sealy International, particularly **Jimmy Rose, Nycole Barnes, and Travis McKenzie**, for their comprehensive, generous, and highly responsible support — from sharing the project's background and providing critical data to offering valuable and deep industry insights during each discussion and working closely with us on methodology alignment. Their insights, dedication, and responsiveness truly made a huge difference to our project.

We would like to extend our sincere gratitude to **Rodrigo Hermosilla** for his tremendous support and invaluable expertise in machine learning. His patient and detailed technical mentorship greatly strengthened the depth and rigor of our work.

We are immensely thankful to our writing coach, **Pamela Siska**, for her meticulous proofreading support and her continuous encouragement. Her thoughtful scheduling and guidance kept us on track and were instrumental in the successful completion of our project.

We would like to express our great appreciation to our mentors, **Maria Jesus Saenz** and **Jim Rice** for their invaluable guidance and mentorship.

We are deeply grateful to our families for their unwavering support and encouragement throughout this journey. Their sacrifices, caring, and belief in us made all the difference.

Last but not least, we would like to thank the CTL faculty for providing us with this incredible opportunity to work on such a meaningful capstone project.

– Xiaoli & Shane

TABLE OF CONTENTS

1. Introduction

1.1 Background

Since the start of the COVID-19 pandemic, there has been a rapid shift in consumer behavior with alternative channels to brick-and-mortar becoming popular for purchasing goods. Most consumer goods and retail companies have adopted this omnichannel strategy as an effective means to satisfy the modern consumer's expectation of convenience (Akwa-Mensah, 2023). However, this strategy introduces complexities to these organizations' supply chains in terms of their ability to anticipate demand across multiple channels, as well as across multiple regions and products. Demand can be difficult to predict when considering the breadth of a company's operation, and inaccurate forecasts increase the risk of lost sales due to inefficient allocation of resources to support customer fulfillment.

One cost-effective option for mitigating a large portion of this risk is through an improvement in demand forecasting methodologies. Demand forecasting plays a critical role in the life of a corporation as a basis for making decisions that relate to proactive asset allocation in anticipation of consumer demand. If a company forecasts an increase or decrease in their demand among nodes in their network, then the company can move their current inventory or source more product to the appropriate locations with enough advanced knowledge. Most companies leverage internal historical data as a baseline for generating these forecasts with the application of Machine Learning (ML) techniques growing in popularity for this practice; however, the inclusion of exogenous variables, within these techniques, remains in its introductory phase in many industries despite the importance of these variables to the macroenvironment in which a firm operates (Rocke and Zhang, 2024).

1.2 Motivation

This project's sponsor company, Tempur-Sealy International (TSI), believes there is value in applying these exogenous variables to their demand forecasts (TSI, personal communication, 2024). TSI leverages a robust forecasting system to aid in demand planning; however, in recent years, historical trends are not reflecting future trends and seasonality well across TSI's channels, leading to forecast accuracy challenges. Their forecast system was effective at predicting relatively stable bedding and mattress sales in the pre-pandemic environment. However, the rapid growth of customer engagement across multiple channels has introduced new complexities, making it more difficult for TSI to anticipate where and when demand will arise. As a result, TSI is exploring whether incorporating exogenous variables, particularly those related to the

bedding and mattress industry, into its demand planning processes can meaningfully improve forecast accuracy and help address the omnichannel challenges that have intensified since the pandemic.

1.3 Problem Statement

TSI is seeking to systematically identify areas of risk and opportunity within their demand forecasts through an evaluation of exogenous variables. These variables are derived through external datasets which are maintained by a third-party to reflect macro environmental factors related to the bedding and mattress industry. For TSI, Google Search Interest, related to their industry, as well as U.S. housing starts are interesting areas to focus on. However, while these areas are a good starting point, TSI is also interested in evaluating other external datasets as well to see which factors best improve their demand planning. TSI aims to determine the relevance of these variables on historical sales as well as the associated lag of influence of each variable.

With these goals in mind, TSI seeks to answer the following two key questions:

- Are third party data sets, both general economic and category/channel specific, valuable forward indicators for demand planning, and at what lag are they significant?
- How can the team best identify relationships between external datasets and historical sales to generate recommendations for demand planning?

1.4 Project Goals and Expected Outcomes

The goal of this project is to leverage machine learning techniques to help TSI evaluate the feasibility and value of incorporating third-party and publicly available datasets as leading indicators to enhance their omnichannel demand planning process for two brands individually. For confidentiality, we refer to them as Brand A and Brand B. Similarly, we anonymize product and channel names throughout the report (e.g. Product A, Channel A, Reseller A). The project aims to determine whether external datasets, with time lags, improve a baseline model performance metrics.

The deliverables for this project include:

- **Baseline Model:** A time series machine learning demand forecasting model using internal historical sales data and selected endogenous variables (e.g., promotions). This model serves as a benchmark for comparison with the advanced model.
- **Exogenous Variables Incorporated Model:** A model incorporating external variables to compare performance metrics with the baseline and assess their impacts on sales, both individually and in combination.

Ultimately, this study builds a framework to offer TSI a scalable tool to future-proof its demand planning and supports proactive, informed decisions in response to shifting demand across brands, products, and channels.

2. State of the Practice

The focus of this study is to develop a methodology for evaluating whether exogenous variables serve as forward indicators for the demand of TSI's mattresses, pillows, and bedframes. As a starting point to best determine a methodology, we perform research in three key areas: time series demand forecasting, ML forecasting approaches, and exogenous variables to be evaluated in this study.

2.1 Time Series Demand Forecasting Overview

Forecasting end-consumer demand is a critical component to any organization's supply chain. By anticipating where and when demand is likely to spawn, an organization can better coordinate supplier activities, proactively manage production, optimize inventory levels, and closely manage the timing of financial flows throughout their network. These are just some of the reasons forecasting can add value to an organization. However, while the benefits of accurate forecasting are clear, generating these forecasts is a relatively complex process that requires cross functional consensus and detailed analysis.

Forecasting typically includes a collaborative process between domain experts and key stakeholders as well as the analysis of historical demand and other data points. Historical data for forecasting often occurs along a time series. In this context, historical data is represented by numeric values recorded within equal time intervals (Seyedan and Mafakheri, 2020). Many forecasting models, ranging in levels of complexity, exist to predict future demand from this time series data. However, simple forecasting models often lack the computational depth that is needed to best identify non-linear leads and lags in a time series dataset despite its modern-day importance. This lack of linearity is becoming increasingly common "in demand behavior … due to competition among suppliers, the bullwhip effect, and mismatch between supply and demand" (Seyedan and Mafakheri, 2020). As a result, sophisticated ML approaches are being popularized for their ability to handle multiple large datasets simultaneously and uncover underlying patterns that exist between these datasets as a basis for their forecasts.

2.2 Approaches for ML Forecasting

With the popularization of multi-channel operations, consumer behavior is becoming more difficult to predict making "the adoption of big data analytics … a necessity for demand forecasting" (Seyedan and Mafakheri, 2020). Predictive ML approaches offer such a framework. These models go beyond the scope of

conventional forecasting in their ability to identify non-linear patterns across multiple variables in large datasets to aid in making predictions. This capability makes ML approaches an attractive partner for today's demand forecasting challenges (Feizabadi, 2020). However, the effective application of ML techniques requires a solid foundation in data preparation, as all predictive ML models depend on it in some form (Brownlee, 2020). Raw data is often messy and complex. If data is left in such an unorganized state, then additional computing power might be required to ensure accurate predictions. However, for most individuals seeking to leverage a predictive ML model, such access may be limited. Therefore, data preparation becomes one of the most critical steps in any ML project (Brownlee, 2020). This process is often specific to the type of ML problem being tackled, but it typically involves some combination of the tasks summarized in Table 1.

Table 1: General Data Preparation Process

Task	**Description**
Data Cleaning	Formatting and error identification/correction
Data Transforms	Scale and distribution adjustments
Feature Engineering	New variable creation
Feature Selection	Relevant variable determination

With this common base in mind, the following two subsections dive into an important distinction within the context of ML problems, regression vs. classification, as well as the selected model for this study, Long Short-Term Memory.

2.2.1 Regression v. Classification ML

It is important to frame the context of a proposed ML problem within one of two categories: classification or regression (Géron, 2019). Classification problems seek to place the output within distinct categories whereas regression problems predict specific values along a continuum, and this problem definition motivates all modeling decisions. Therefore, it is important to determine this portion of the task early in the scoping process.

A simple way to determine this problem type can be through understanding the downstream implications for the desired ML output (Géron, 2019). In the context of this study, TSI wants to identify value-add exogenous variables to their demand forecasting as such improvements allow for tighter production planning and better inventory allocation. The methodology portion of this study explains how a multiple

regression modeling approach, which outputs specific forecast values while using multiple independent features, can best identify those value-add variables.

2.2.2 Long Short-Term Memory

The Long Short-Term Memory (LSTM) network is a type of recurrent neural network (RNN). A major limitation of most neural networks is their lack of memory and cannot retain information about past inputs to make future predictions. While RNNs address this by iterating through sequences and maintaining cumulative information, they are primarily limited to short-term dependencies due to the vanishing gradient problem.

First developed by Hochreiter and Schmidhuber in 1997, LSTM soon became one of the most widely used RNN models as it introduces memory cells and gating mechanisms – Forget Gate, Input Gate, and Output Gate – that allow the network to selectively retain or discard information at each timestep and reintegrate relevant past inputs when needed. Unlike standard RNNs, LSTM excels at capturing long-term dependencies and managing the complexities of time-series data, making it well-suited for this project. Most importantly, it can effectively learn the impact of exogenous variables – including lagged features – on demand with minimal manual tuning. This aligns with a key focus of our study, providing TSI with early indicators of demand changes and insights into the lagged effects of external factors on sales.

2.3 Exogenous Variables

Most of the literature surrounding forecasting methodologies focuses on model evaluation along endogenous variables or the micro-level data internal to a firm (Yasir et al., 2022). However, when considering modern global supply chains, there is a clear opportunity to incorporate exogenous variables, macroeconomic factors, into forecasting practices to better define a firm's operating environment (Yasir et al., 2022). When considering potential variables to include, it is important to evaluate inclusion within the context of one's forecasting scope as an initial feature selection method (Yasir et al., 2022). When this extra consideration is taken, there are significant improvements in forecast accuracy throughout the supply chain (Sinaga et al., 2018).

For this study, the focus is to identify relevant exogenous variables to TSI's demand for mattresses, pillows, and bedframes across various operating channels within the US bedding and mattress market. As a starting point, we adopt a multi-faceted approach, combining logical reasoning and industry knowledge with insights from existing research, to pinpoint potential factors for further model evaluation. This approach

narrows the scope of this study's exogenous variables to the following channel-specific and US-economic indicators.

2.3.1 Web Traffic

Web Traffic refers to the number of visits to a website and the interactions visitors have with it, including average visit duration, pages per visit, bounce rate, and total page views. TSI is interested in exploring the relationship between the Web Traffic and demand forecasting to have an idea of the conversion rate, which measures the effectiveness of converting web visitors into purchasing customers.

2.3.2 Foot Traffic

Foot Traffic refers to the number of visits to a specific area within a certain period of time. In our project, the areas of interest include nationwide Mattress Stores, Home Improvement, Superstores, Furniture and Home Furnishings, Department Stores, Shopping Centers. The goal is to determine whether higher foot traffic correlates with higher potential sales.

2.3.3 Large Retail Stores

Large Retail Stores total sales represent another exogenous variable of research interest. We aim to determine whether higher overall sales at these stores translates into increased sales for TSI.

2.3.4 Google Trends

Google Trends is an online tool offered by Google that democratizes the availability of Google search data making it valuable for identifying trends in Google search activity across the past 20 years (Google, 2024). Academics and practitioners alike leverage the tool across a variety of analytical disciplines, including time-series forecasting, for its proven benefits in representing internet activity within the context of a given research question (Mavragani et al., 2018). This representation is possible through a normalization process which standardizes a sample of Google search data (Google, 2024). This process compares search interest across geographies and time horizons regardless of search volume. However, while this tool allows for standardized, limitless access to search data, one should be careful with how they employ it within their analyses.

Researchers have voiced some concern over reliability when noticing alterations in the output of this tool for the same queried term and time horizon across different querying sessions (Franzén, 2023). However, while these alterations appear problematic at first, they should be expected given the sampling nature of this tool (Medeiros & Pires, 2021). Therefore, one can mitigate this risk through averaging samples from different

querying sessions (Medeiros & Pires, 2021). This process instills more confidence in the application of Trends data and can help improve forecast accuracy (Medeiros & Pires, 2021).

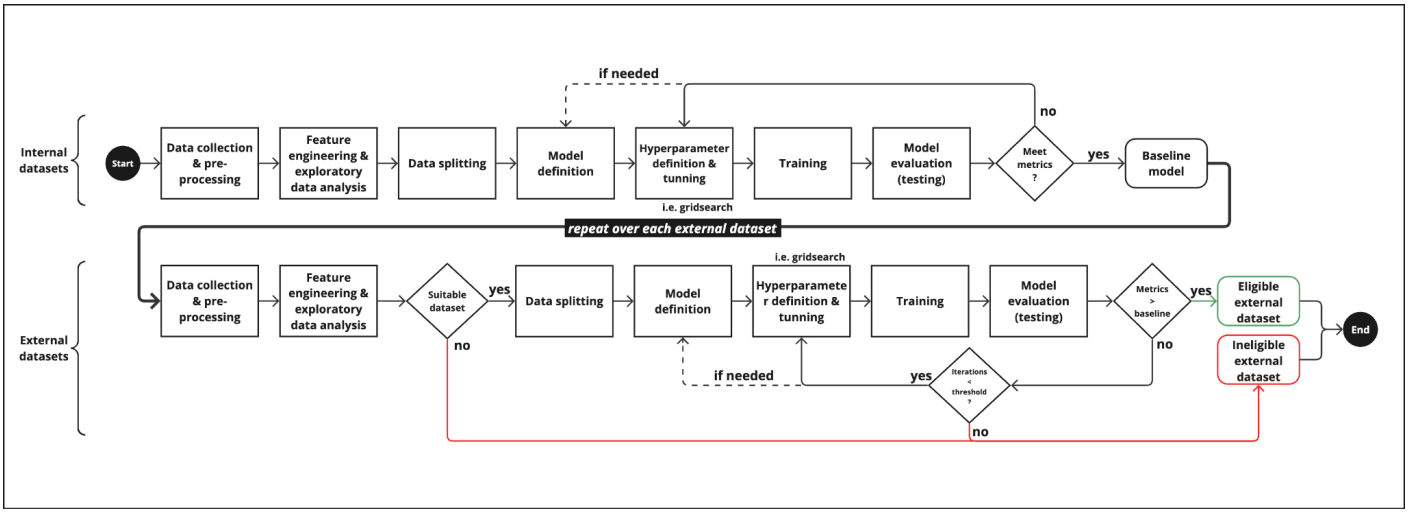
2.3.5 General Economic Indicators

General Economic Indicators are metrics that reflect the overall performance of the economy. TSI seeks to understand the direct impact of indicators such as U.S. housing starts, Customer Sentiment Index, Federal Interest Rate, Disposable Income, Inflation Rate, Consume Confidence on its sales. Based on TSI's past experience, U.S. housing starts appear to be an indicator that influences their sales. Through modeling, we test this hypothesis to verify whether the results align with their previous findings.

3. Methodology

Based on a review of the literature and discussions with our partner company, we determine that the LSTM model is the most appropriate for our needs. This approach seamlessly incorporates exogenous variables and captures the influence of lags on demand forecasting. To implement the methodology, we follow the steps outlined in Figure 1, which consist of two main branches: the top branch focuses on endogenous features, while the bottom branch systematically evaluates each exogenous variable. We first build a baseline model using only internal datasets, then test external datasets individually and in combination. If an external dataset improves model performance, it is marked as eligible; otherwise, it is excluded from further consideration.

Figure 1: Flow Process Chart



3.1 Data Preparation

Data preparation is a critical step for any machine learning model, as raw data must be transformed into a clean, structured format before being fed into the model to ensure accurate future demand

predictions. This is evident in our case, as we integrate internal data with a range of external variables – each with different time intervals and coverage. To ensure data integrity and alignment across datasets, we first apply preliminary preprocessing to both internal data and exogenous variables. Following exploratory data analysis and clustering, we run a structured preprocessing pipeline that includes imputation, normalization, and transformation to prepare the data for modeling.

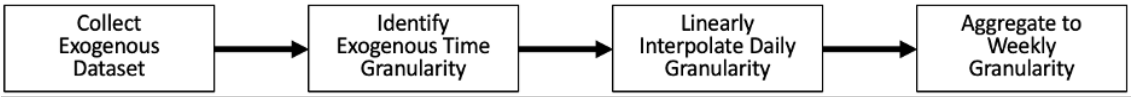
3.1.1 Internal Data Preprocessing

Preliminary preprocessing of internal data begins with consolidating all yearly files into a single dataset for both brands. To align with our sponsor company's strategic priorities and support efficient modeling, we narrow the product and channel scope to those most important to the business. In terms of granularity, we aggregate daily sales to the weekly level after parsing the date. As confirmed and corroborated by our sponsor company through data review, weekly intervals are preferred, as they balance dataset size while capturing important intra-month events. We also remove orders with a canceled shipping status to ensure the dataset reflects actual demand. Together, these steps produce a clean, focused dataset ready for exploratory data analysis.

3.1.2 Exogenous Variable Preprocessing

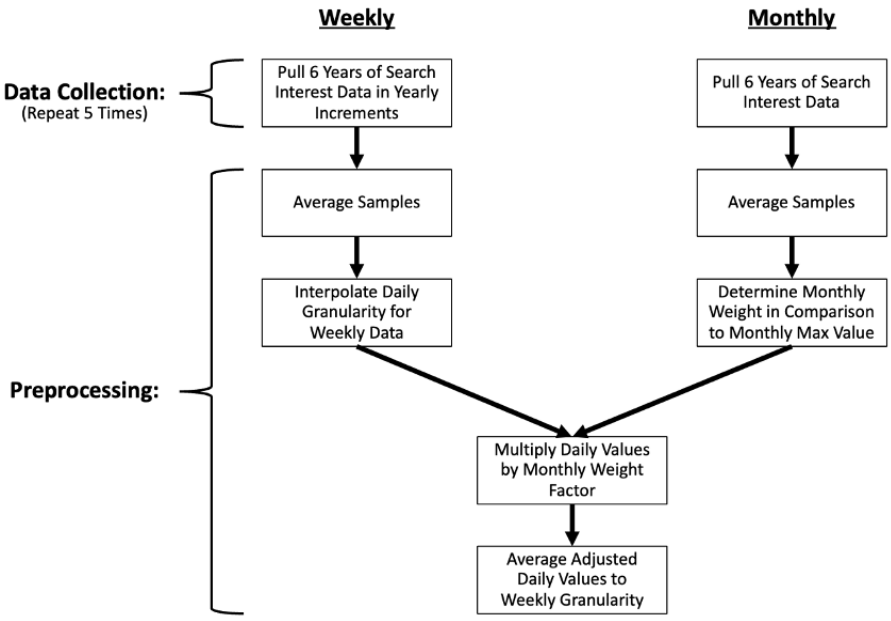
Exogenous variables are not often sourced in a format that is immediately suitable for machine learning models. These datasets are sourced generically in time granularities that differ from the desired time series model's requirements, which emphasizes the importance of taking additional preprocessing measures. In the case of this study, there is a need to place these exogenous variables into the same granularity as the target variable, TSI's weekly sales, to ensure statistical significance can be concluded between datasets and fair comparisons can be made between modeling steps. Apart from Google Trends data, the exogenous variables for this study follow a relatively consistent cleaning process as outlined in Figure 2. The first three steps of the process hold true for all exogenous variables; however, there is some differentiation within the final step of aggregation. The choice between averaging or summing during aggregation depends on the nature of the metric being represented. For example, the Consumer Price Index (CPI) is a normalized index representing relative price levels over time; therefore, taking the mean across days within a week preserves its intended interpretation. In contrast, U.S. housing starts represent cumulative counts of new residential construction projects, so a summation better reflects the weekly volume of economic activity. However, in the case of the latter, caution is exercised to ensure that linearly interpolated values are divided by an appropriate number of days factor to represent the time range between data points at a higher granularity.

Figure 2: General Exogenous Variable Preprocessing Procedure



In general, every exogenous variable interpolation is handled differently, and in the case of this study, Google Trends requires a more tailored process than the aforementioned procedure. This tool requires additional data collection and preprocessing. The tool extracts samples from Google's larger data ecosystem to normalize and to generalize interest across time and regions for a given search term (Google, 2024). During each querying session, a different sample is likely used to gain these insights; however, this design can lead to a lack of confidence that the data is representative of the entire population (Franzén, 2023). Furthermore, the tool automatically adjusts the time granularity with which data is provided depending on the time horizon requested by the user. This behavior introduces an additional challenge for representing multi-year trends in search term interest at a weekly granularity. To alleviate these limitations, this study employs the process as outlined in Figure 3. By following this process, some of the ambiguity in population representation is reduced via averaging five samples, and the high-level trends are largely maintained by pulling data at two different granularities: weekly and monthly.

Figure 3: Google Trends Exogenous Variable Preprocessing Procedure



In both the general and the Google Trends exogenous variable preprocessing procedures, final values are compared to original values to ensure that data integrity is maintained at the lower granularity. This evaluation involves comparing time plots and summary statistics before and after preprocessing with visual indicators being the primary level of acceptance criteria for sufficient preprocessing.

3.2 Exploratory Data Analysis

Exploratory Data Analysis (EDA) is a standard and essential step in machine learning, used to analyze and visualize datasets in order to gain a comprehensive understanding of their structure, patterns, trends, seasonality, anomalies, and distributions before applying advanced models.

3.2.1 Internal Data Overview

We begin by examining the internal dataset to uncover temporal patterns and sales behaviors. To understand overall trends, we first plot sales over time for both brands. To detect anomalies, such as outliers or missing values, we compare two methods: interquartile range (values beyond 1.5 times interquartile range from the 25^{th} or 75^{th} percentile) and standard deviation (values beyond three standard deviations). For accurate detection and imputation, we align with our sponsor company's guidance and impute both outliers and missing values using a rolling average over a defined sliding window before and after the affected point. To keep total sales unchanged, we offset the difference between the imputed and original quantities by evenly spreading it across the sliding window weeks before and after the affected week – dividing the gap by the total number of weeks in the window – at the product and channel level.

3.2.2 Clustering

To improve computational efficiency and maintain modeling focus, we implement K-means clustering. Rather than iteratively modeling every product-channel combination for the baseline and repeating the process with each exogenous variable, we focus on a representative subset derived through unsupervised learning. By analyzing historical sales at the product-channel level, we cluster combinations with similar demand behaviors and apply the silhouette score to determine the optimal number of clusters. From each cluster, we select the highest-volume combination as a representative for modeling. For Brand A, two clusters are segregated: Cluster 0 is represented by Product A, Channel A (A-A-A), and Cluster 1 is represented by Product B and Channel A (A-B-A). For Brand B, one cluster is formed and represented by Product A Channel A (B-A-A). This clustering-based approach reduces dimensionality, boosts efficiency, and preserves the diversity of demand patterns across combinations.

3.2.3 Distribution Box-Cox Transformation

After determining the product and channel combinations for each brand, we use histograms and box plots to examine the distribution of internal sales data and check for skewness. As noted by Kuhn and Johnson (2013), "An un-skewed distribution is one that is roughly symmetric. This means that the probability of falling on either side of the distribution's mean is roughly equal." Similarly, Brownlee (2020) highlights that

"many machine learning algorithms perform better when the distribution of variables is Gaussian." To reduce skewness, stabilize variance, and improve modeling performance, we apply Box–Cox transformation to make the data more Gaussian-like or bell-shaped.

3.3 Feature Engineering

Feature engineering refers to the process of creating new input variables from the available data (Brownlee, 2020). In our project, we apply feature engineering to internal sales by generating seasonality indicators, Fourier terms, and lag features. Based on guidance from our sponsor company, which runs promotions during holiday weeks (both the holiday week and three weeks prior), we include a binary promotion variable in the model. We test all feature combinations – along with internal sales – as inputs to the internal baseline model and select the combination that yields the best performance for each product–channel combination within each brand.

3.4 Exogenous Variable Lag Analysis

Given the potential for seemingly infinite combinations of exogenous variable shifts against target values, there is a need to reduce the solution space of lags for subsequent modeling. For this study, the sole focus is on lag relationships where an exogenous variable acts as a historical indicator for future sales (e.g., whether CPI from three weeks ago affects today's sales). This focus helps to reduce the initial scope; however, there remains a need to further isolate the exogenous variable shifts with the highest potential for improving model performance. To address this need, the study employs a process that identifies the strongest relationships between exogenous variables and target values through a mutual information (MI) test combined with minimum Redundancy Maximum Relevance (mRMR). This combination captures nonlinear dependencies between lagged exogenous variables and sales through mutual information, while mRMR promotes a parsimonious feature set by prioritizing relevance and minimizing redundancy. For mutual information to be appropriately assessed, each shifted exogenous variable must align exactly in time with the corresponding target sales observations, ensuring that the historical context of the lag is properly preserved. The MI-mRMR analysis is performed independently between each exogenous variable's potential lags (limited to less than three years) and the target demand values derived from the clustering analysis. For each brand-product-channel combination, the top ten lagged features, ranked by the mRMR optimization of mutual information and redundancy, are selected for inclusion in the modeling process.

To access the joint influence of exogenous variables on sales, we generate a correlation heatmap to identify highly collinear feature pairs (|correlation coefficients| > 0.5). For each pair, we retain the variable

with higher mutual information relative to sales and discard the other. This process is repeated iteratively until all remaining pairwise correlations fall below the 0.5 threshold. The final set of selected features is then combined for a separate model input.

3.5 Preprocessing Pipeline

Prior to model development, we build a preprocessing pipeline to standardize both internal and exogenous variables. We first apply mean imputation to fill in missing values, then rescale features to a 0–1 range using min–max normalization. This step is essential for machine learning models like LSTM, which rely on gradient-based optimizers that perform best when features are on a similar scale. Without proper scaling, variables with larger magnitudes may dominate the learning process and degrade model performance.

For the target sales variable, we apply the Box–Cox transformation between mean imputation and min–max normalization to reduce skewness and enhance predictive accuracy. The pipeline is fitted on the training set using fit_transform() and applied to the evaluation and testing sets using transform() to avoid data leakage.

3.6 Baseline Model Development

The baseline model for our project is developed using an LSTM model. The model leverages TSI's internal historical sales, along with feature-engineered endogenous variables that have demonstrated statistically significant improvements in performance metrics, to forecast future sales – serving as a benchmark for comparison with enhanced models that incorporate exogenous variables. Among all the endogenous features, we select the promotion, seasonality, and Fourier terms.

Unlike traditional machine learning projects that follow the 70-15-15 split rule for training, evaluation, and testing, we adopt a chronological partitioning strategy to prevent data leakage and ensure the model learns from past trends, seasonality, and patterns while being evaluated and tested on future periods.

- For Brand A, we use pre-2022 data for training, 2022 for evaluation, and post-2022 for testing. This setup allows the model to learn year-over-year patterns, evaluate performance on a complete year, and forecast an entire future year.
- For Brand B, which has fewer data points, we use data prior to 2024-04-30 for training, data from 2024-04-30 to 2024-08-04 for evaluation, and data after 2024-08-05 for testing.

For the modeling process, we import Python packages from TensorFlow Keras. We create sequences – fixed-length windows of consecutive time steps – as inputs to the LSTM model. For Brand A, we use a sequence length of 260 weeks (156 for training, 52 for evaluation and testing) from 2019 to 2023. For Brand

B, the sequence spans 90 weeks (68 for training, 11 for evaluation and testing), covering its initial sales data from 2023 to 2024.

Time steps refer to sliding windows of past data points grouped together for the model to learn from and use to predict the next point. We choose shorter time steps – 4 for Brand A and 2 for Brand B – to capture important patterns or sudden changes that occur within just a few time steps, such as event detection. Short windows allow the model to focus on local transitions, while the LSTM's internal memory can still accumulate knowledge beyond the current window. Another reason this approach works well is that the data has already been preprocessed through statistical aggregation, feature engineering, and signal transformation. Each step already encodes rich temporal information, making larger time windows unnecessary. The LSTM can then focus on learning patterns from these engineered features instead of raw data.

We then build models for both brands. The number of neurons in each LSTM layer is carefully chosen to balance model complexity; higher counts allow the model to capture more patterns.

- For Cluster 0 of Brand A (product A and channel A), we add one LSTM layer with 1024 neurons (using the default Sigmoid activation) to learn temporal patterns. This is followed by dense layers with 512, 256, and 128 neurons using ReLU activation to capture nonlinear relationships. A 20% dropout layer – within the standard range for RNNs – is placed after the LSTM layer to reduce overfitting by randomly deactivating a portion of neurons during training. One batch normalization layer is added after the dropout layer, and additional batch normalization layers are inserted between each dense and activation layer to improve generalization. The final dense output layer predicts sales.
- For Cluster 1 of Brand A (Product A and Channel B), we use one LSTM layer with 1024 neurons (using the default Sigmoid activation), followed by a stack of dense layers with 512, 256, 128, and 64 neurons (using ReLU activation). Batch normalization layers are placed between the dense layers to improve generalization. A single 20% dropout layer is included after the first dense layer to reduce overfitting. The final dense output layer produces the sales forecast.
- For Brand B (Product A and Channel A), given the smaller dataset, we reduce the capacity of the network to one LSTM layer with 512 neurons (sigmoid activation), followed by two dense layers with 256 and 128 neurons (ReLU activation). Batch normalization and 20% dropout layers are included between layers. A final dense output layer generates the forecast.

The model is compiled using an Adam optimizer to efficiently adjust the model's weights to minimize errors and a Huber loss function to measure prediction errors. The Huber loss function performs well when

the data contains spikes or noise. It behaves like MSE for small errors and like MAE for large ones, keeping the model stable without overreacting to sudden jumps. Training incorporates early stopping to avoid overfitting. After each epoch (a full pass through the training data), the model checks whether validation loss improves by more than a defined threshold, *min_delta* (the smallest improvement in validation loss needed to continue training). If there is no improvement after a specified number of epochs (*patience – the number of consecutive epochs without significant improvement before training stops*), training halts, and the model restores the best weights (which determine the influence of input variables on predictions) for optimal performance. For Brand A, we set a maximum of 2000 epochs with early stopping *patience* of 20 to balance learning and overfitting risk. We set *min_delta* to 0.05 for Product A and 0.01 for Product B to match the sensitivity required by their respective layer configurations. For Brand B, we use a smaller *min_delta* of 0.0001 to enable finer tuning, which is important for smaller datasets where even marginal improvements matter. *Batch size (the number of training samples processed before updating model weights)* is set to 24 for Brand A to speed up training and stabilize gradient updates for its larger dataset, and 16 for Brand B to allow more frequent weight updates and improve learning with its smaller dataset.

The final step is making predictions. At each step, the model takes the current input and previous hidden state to decide what information to retain in memory for improved forecasting. This process is applied to the testing set to generate the full weekly sales forecast for both brands.

3.7 Incorporating Exogenous Variables into Baseline Models

Next, we update our baseline LSTM model by incorporating the top-ranked exogenous features for each brand–product–channel combination, based on the lag analysis. The model structure and hyperparameter settings remain the same as in the baseline model but now include the exogenous variables both individually and in combination as a full feature set. By comparing performance metrics between the baseline and updated models, we can identify key indicators along with their corresponding lag effects.

3.8 Model Evaluation

Model evaluation is the final and crucial step to assess the performance of both our baseline and updated LSTM models. In this study, we use mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), and mean absolute percentage error (MAPE) for accuracy comparison. These metrics are calculated by comparing forecasted values with actual sales on the testing dataset for both the baseline model and the model incorporating exogenous variables. This allows us to determine whether

adding exogenous variables improves or worsens performance – and, if it improves, what insights and significance the inclusion brings to TSI prediction.

4. Results and Discussion

Following the completion of data preprocessing, the study proceeds with exploratory analysis on both internal and external datasets as well as model development and evaluation. In this chapter, the findings from these efforts are discussed, and their implications for demand forecasting are highlighted.

4.1 Results

This section presents the results of the data overview, outlier detection and imputation, and exogenous variable lag analysis. It also compares the performance metrics of baseline models against those of models incorporating exogenous variables.

4.1.1 Internal Data Overview

According to the aggregated-level normalized charts for Brand A and Brand B – covering trend, seasonality, and patterns – Brand A shows an upward trend (Figure 4) with clear and strong seasonality (Figures 5 & 6), and no specific patterns in the residual plots (Figure 6). In contrast, Brand B exhibits a downward trend (Figure 7), which may be due to the limited data available, but still demonstrates strong and consistent seasonality (Figures 8 & 9).

Brand A – Trend, Seasonality, and Patterns

Figure 4: Weekly Demand Trend Over Time – Brand A

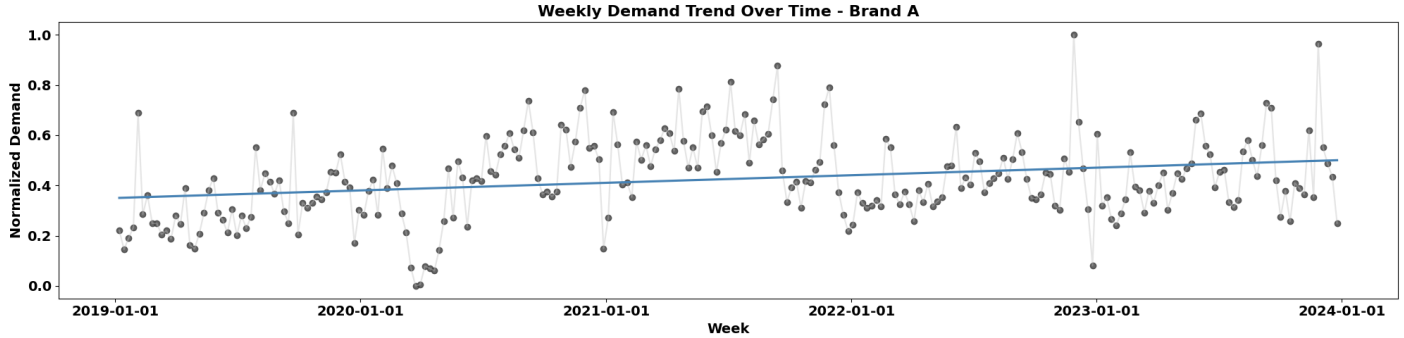


Figure 5: Weekly Demand Seasonality by Year – Brand A

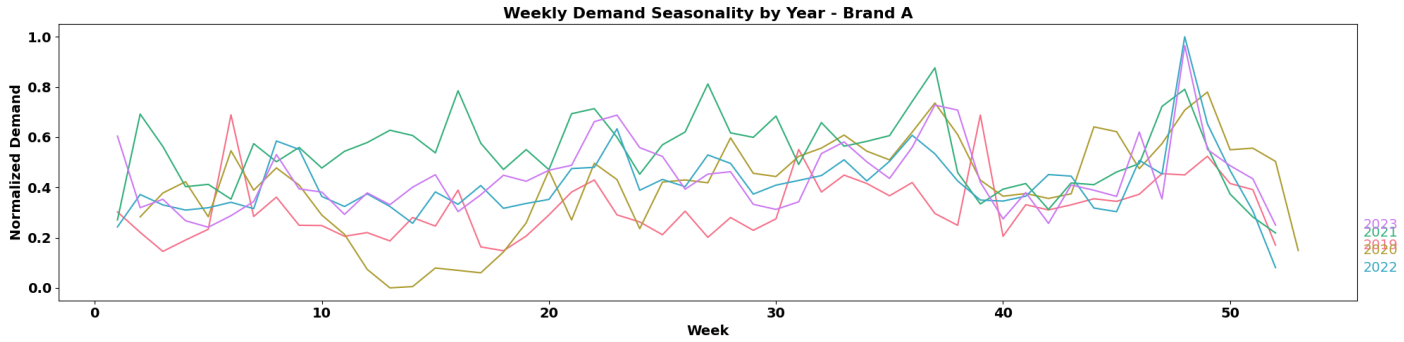
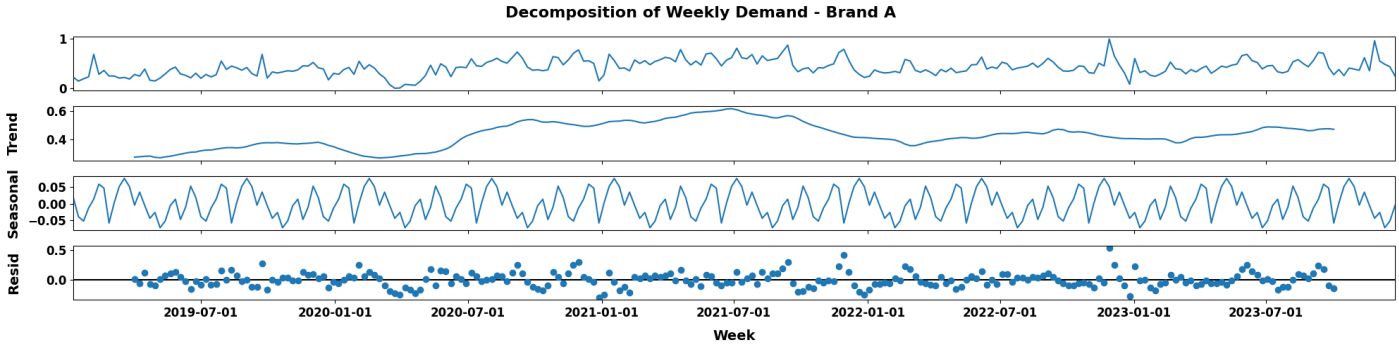


Figure 6: Decomposition of Weekly Demand – Brand A



Brand B – Trend, Seasonality, and Patterns

Figure 7: Weekly Demand Trend Over Time – Brand B

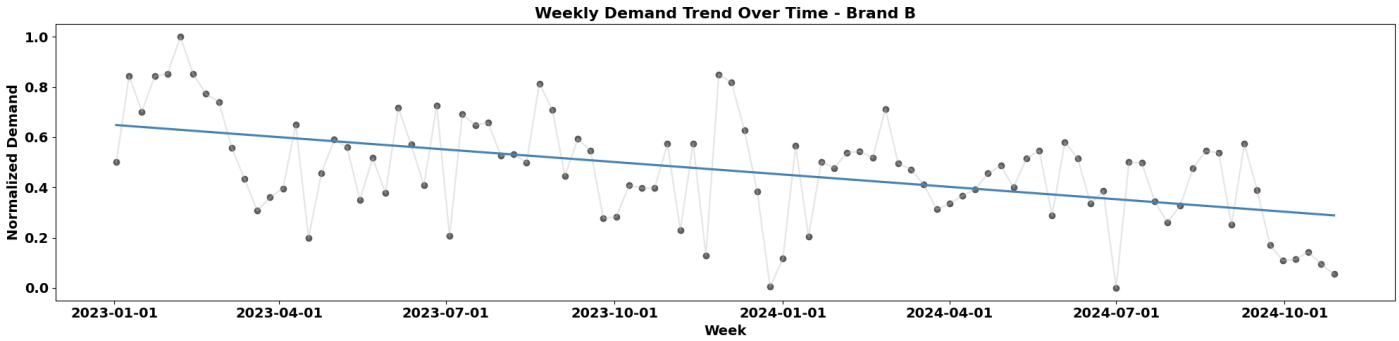


Figure 8: Weekly Demand Seasonality by Year – Brand B

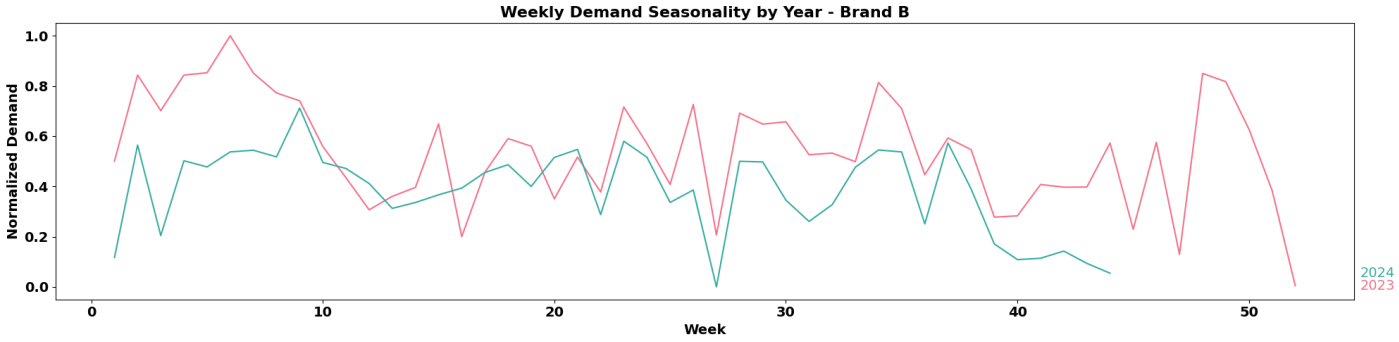
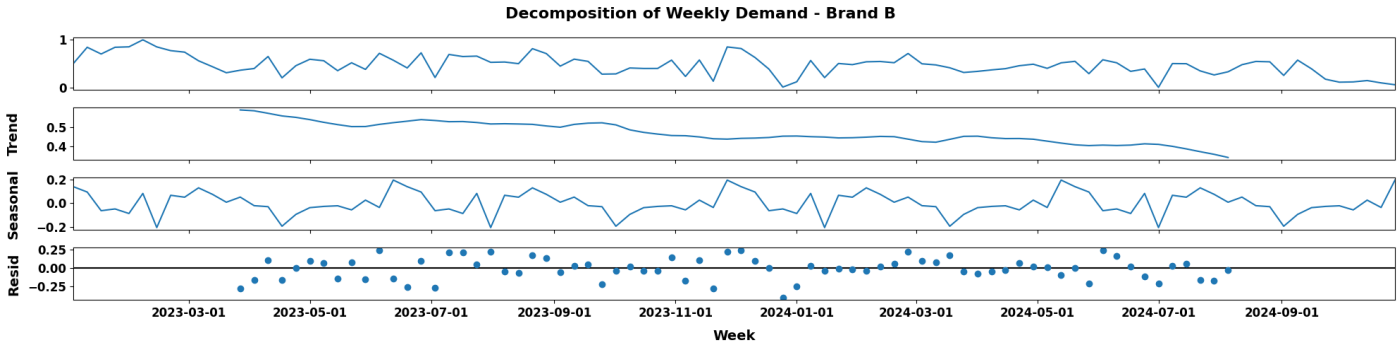


Figure 9: Decomposition of Weekly Demand – Brand B



After applying both the interquartile range and standard deviation methods, we identify outliers in the internal datasets, highlighted as red dots in the normalized line plot in Figure 10 for Brand A and Figure 12 for Brand B. Upon confirmation with our sponsor company, the circled red dots represent the confirmed outliers due to system issues, while the remaining red dots reflect regular seasonal sales patterns. We then impute the data as previously described. The post-imputation results are shown in Figure 11 for Brand A and Figure 13 for Brand B.

Brand A – Outlier Detection and Imputation

Figure 10: Weekly Demand Outlier Detection – Brand A

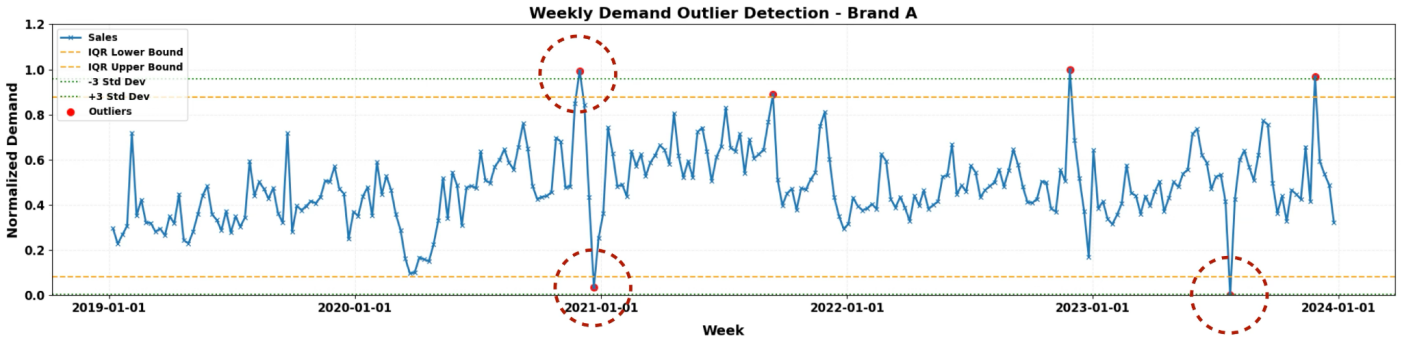
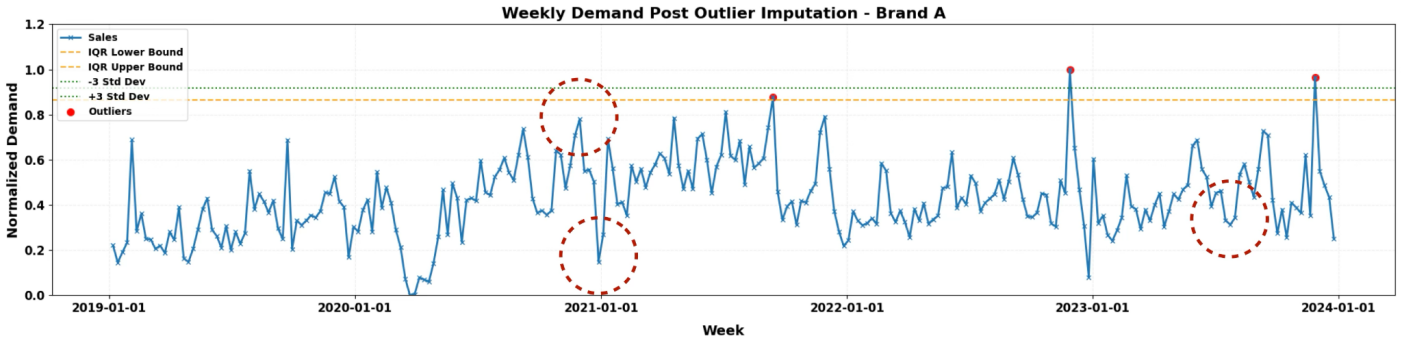


Figure 11: Weekly Demand Post Outlier Imputation – Brand A



Brand B – Outlier Detection and Imputation

Figure 12: Weekly Demand Outlier Detection – Brand B

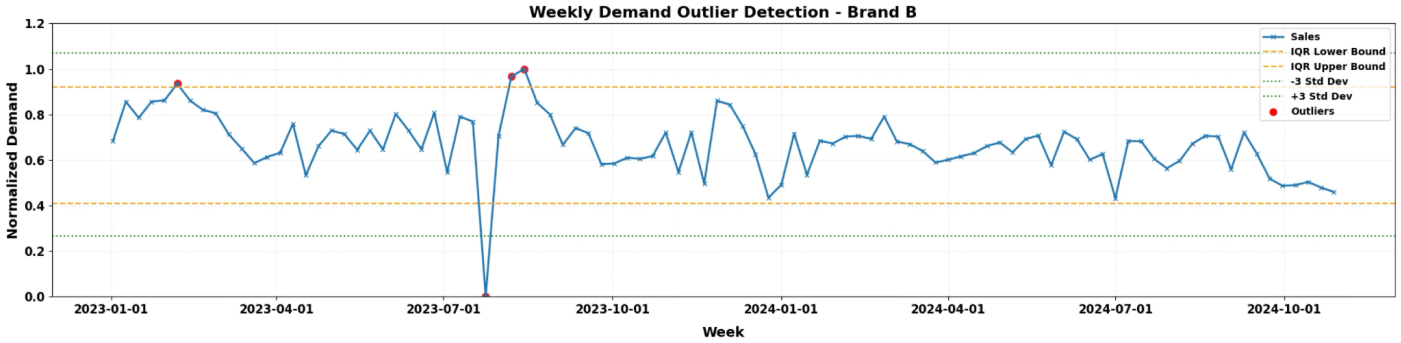
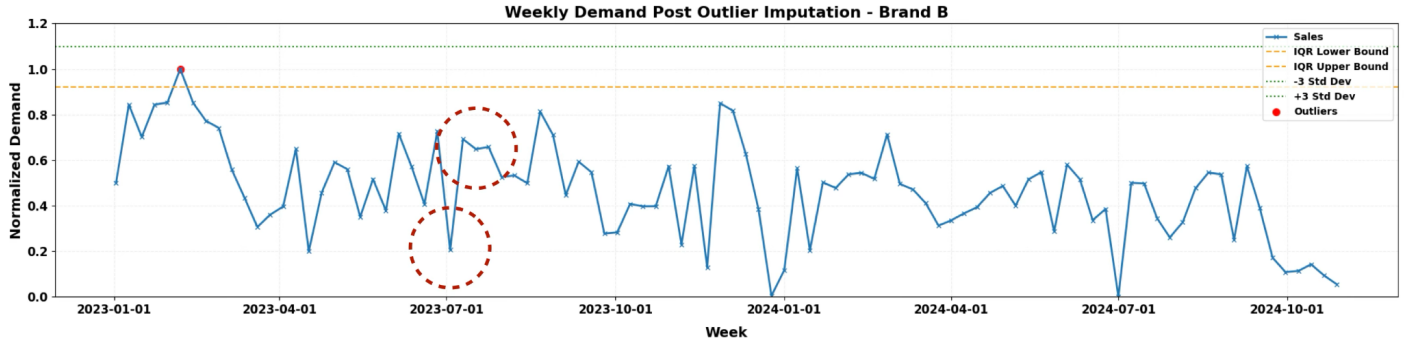


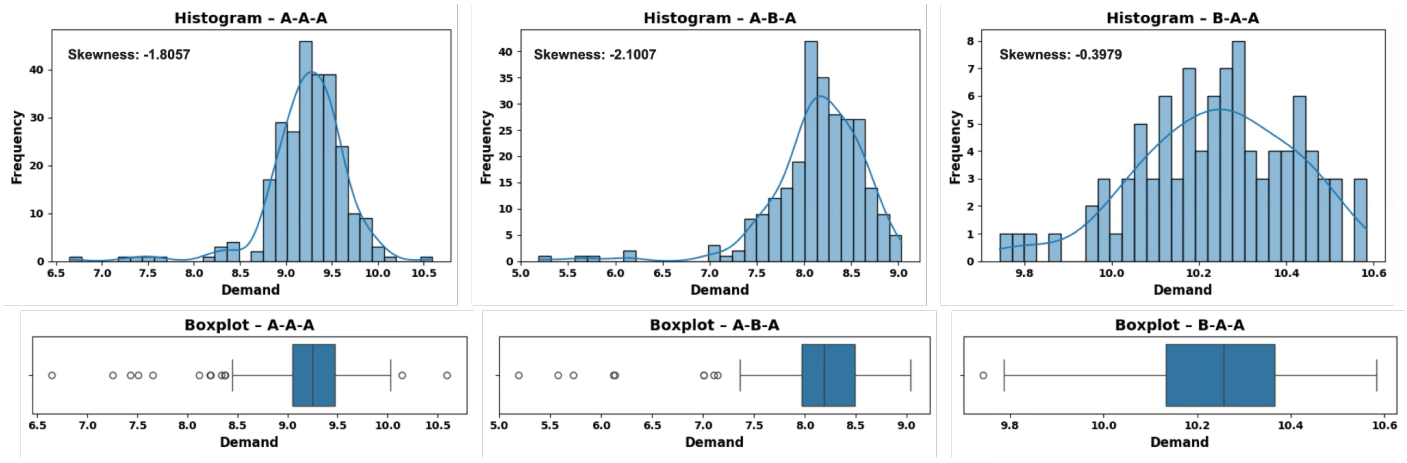
Figure 13: Weekly Demand Post Outlier Imputation – Brand B



Box–Cox Transformation

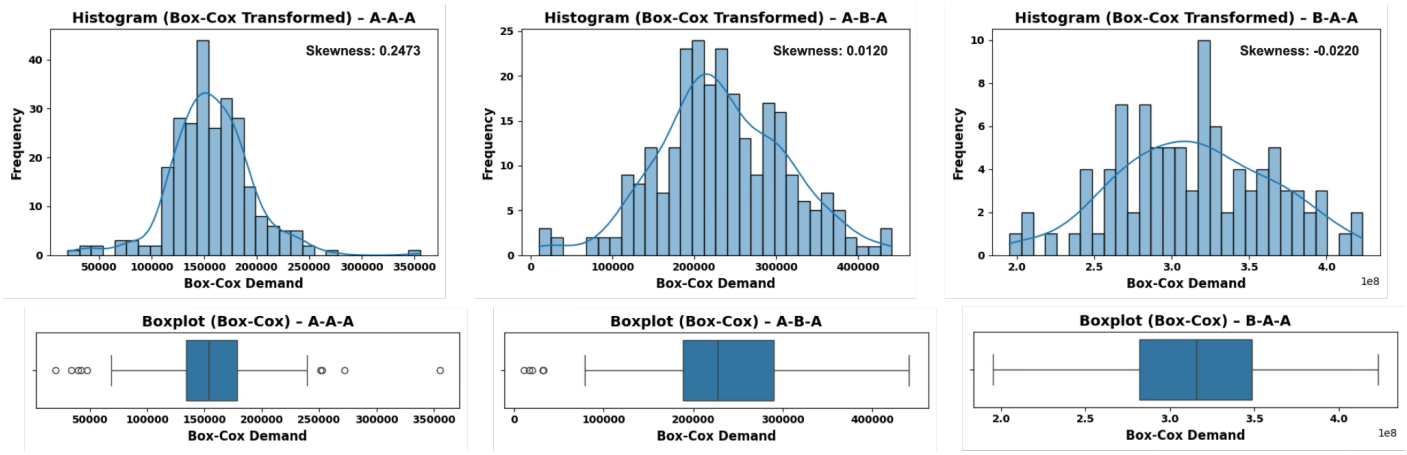
Based on the histogram in Figure 14, Brand A, Channel A sales data exhibits substantial left skewness, with a skewness value of -1.805 for Product A and -2.1007 for Product B. Since skewness below -0.5 indicates a left–skewed distribution, we apply the Box–Cox transformation to correct for asymmetry and stabilize variance. Although the distribution for Brand B, Channel A is more balanced, we apply the same transformation for consistency across preprocessing. Post–transformation, the distribution demonstrates improved symmetry and a more centralized tendency as shown in Figure 15.

Figure 14: Demand Distribution Before Box–Cox Transformation



Brand A Product A Channel A (A-A-A) | Brand A Product B Channel A (A-B-A) | Brand B Product A Channel A (B-A-A)

Figure 15: Demand Distribution Post Box-Cox Transformation



Brand A Product A Channel A (A-A-A) | Brand A Product B Channel A (A-B-A) | Brand B Product A Channel A (B-A-A)

4.1.3 Exogenous Variable Lag Analysis

Given the vast number of possible lag combinations with potential nonlinear relationships to the target variables, there is a clear need to reduce the feature space prior to modeling. To address this, the top ten lagged exogenous features for each brand–product–channel combination are selected based on MI-mRMR rankings. The results of this selection process are presented in Tables 2, 3, and 4.

Table 2: Brand A Product A Channel A (A-A-A) Exogenous Variable Lag Analysis Top Ten Features

Exog Source	Feature	Mutual Information
Consumer Price Index	21_wk_lag_cpi	0.394
Consumer Price Index	148_wk_lag_cpi	0.365
US Birth Rate	96_wk_lag_birth_rate	0.356
US Birth Rate	116_wk_lag_birth_rate	0.353
Google Trends	2_wk_lag_mattress_search_score	0.343
Disposable Personal Income	94_wk_lag_dspi	0.340
US Birth Rate	139_wk_lag_birth_rate	0.333
Disposable Personal Income	89_wk_lag_dspi	0.322
Disposable Personal Income	93_wk_lag_dspi	0.320
Disposable Personal Income	41_wk_lag_dspi	0.309

Table 3: Brand A Product B Channel A (A-B-A) Exogenous Variable Lag Analysis Top Ten Features

Exog Source	Feature	Mutual Information
Consumer Price Index	132_wk_lag_cpi	0.414
Consumer Sentiment Index	8_wk_lag_csi	0.353
Consumer Sentiment Index	20_wk_lag_csi	0.350
Disposable Personal Income	16_wk_lag_dspi	0.322
Disposable Personal Income	32_wk_lag_dspi	0.321
Consumer Sentiment Index	30_wk_lag_csi	0.315
US Existing Home Sales	26_wk_lag_us_existing_home_sales	0.314
Consumer Sentiment Index	12_wk_lag_csi	0.313
US Existing Home Sales	66_wk_lag_us_existing_home_sales	0.305
Federal Interest Rate	13_wk_lag_fir	0.299

Table 4: Brand B Product A Channel A (B-A-A) Exogenous Variable Lag Analysis Top Ten Features

Exog Source	Feature	Mutual Information
Disposable Personal Income	114_wk_lag_dspi	0.386
Disposable Personal Income	42_wk_lag_dspi	0.348
Google Trends	97_wk_lag_pillow_search_score	0.334
Reseller A Mattress Sales	79_wk_lag_reseller_a_unit_sales	0.333
Google Trends	19_wk_lag_mattress_search_score	0.291
Reseller B Mattress Sales	41_wk_lag_reseller_b_unit_sales	0.268
Web Traffic for Mattresses	1_wk_lag_visits	0.246
Reseller A Mattress Sales	17_wk_lag_reseller_a_unit_sales	0.239
Reseller A Mattress Sales	145_wk_lag_reseller_a_unit_sales	0.230
Web Traffic for Mattresses	43_wk_lag_visits	0.224

This analysis yields two notable findings regarding the relationship between lagged exogenous variables and sales. First, there is an abundance of short-lagged features, particularly those within a few weeks to a few months of the sales target. This finding suggests that many exogenous factors may have relatively immediate impacts on demand. For instance, short lags between web traffic and sales could reflect near-real-time customer behavior, where an increase in online activity quickly translates to higher sales activity. However, short-lagged relationships could also arise coincidentally from seasonality or operational effects, warranting cautious interpretation. As the key second finding, there is a prevalence of long-lagged features with some significant relationships appearing at lags approaching two to three years. These patterns suggest that while short-term responsiveness to external conditions is important, longer-term economic and market trends also shape sales behavior over extended horizons. Potential explanations for these long lags include delayed macroeconomic impacts, broader industry cycles, or post-purchase replacement behaviors, although the possibility of spurious correlations remains and highlights the need for future validation.

Across both short and long lags, the analysis identifies a diverse set of influential exogenous variables, including economic indicators such as Disposable Personal Income and the Consumer Price Index, as well as

category-specific signals like Google Trends and reseller sales. This variety reinforces the potential value of incorporating multiple types of external data into forecasting models to better capture both immediate consumer responses and slower-moving structural forces influencing demand.

Additionally, after conducting the iterative collinearity reduction process, we identify the following combined feature sets for each product–channel combination:

- Brand A Product A Channel A: 21-week lagged consumer price index, 96-week lagged US birth rate, 2-week lagged mattress Google search score
- Brand A Product B Channel A: 132-week lagged consumer price index, 26-week lagged US existing home sales, 66-week lagged US existing home sales
- Brand B Product A Channel A: all features except the 41-week lagged Reseller B mattress unit sales

4.1.4 Baseline vs. Models Incorporating Exogenous Variables

After running the baseline models and enhanced models incorporating the top-ranked correlated exogenous variables, we observe varying effects of these external factors on forecast performance across product segments.

For Brand A, Product A, Channel A, the inclusion of exogenous variables does not lead to improved RMSE forecast accuracy, as the baseline model already closely tracks actual sales. Among the tested inputs, the combined feature set yields the normalized metrics RMSE closest to the baseline (Figure 16), followed by the 41-week lagged disposable personal income and the 148-week lagged consumer price index. Similar patterns are observed across other normalized error metrics – MAE, MSE, and MAPE – as shown in the histogram in Appendix Figure 25. Figures 17–18 display the forecast versus actual sales plots for the baseline model and the best-performing model incorporating exogenous variables – combined feature set. For results of the remaining models, ranked by performance metrics from best to worst, see Appendix (Figures 26–35).

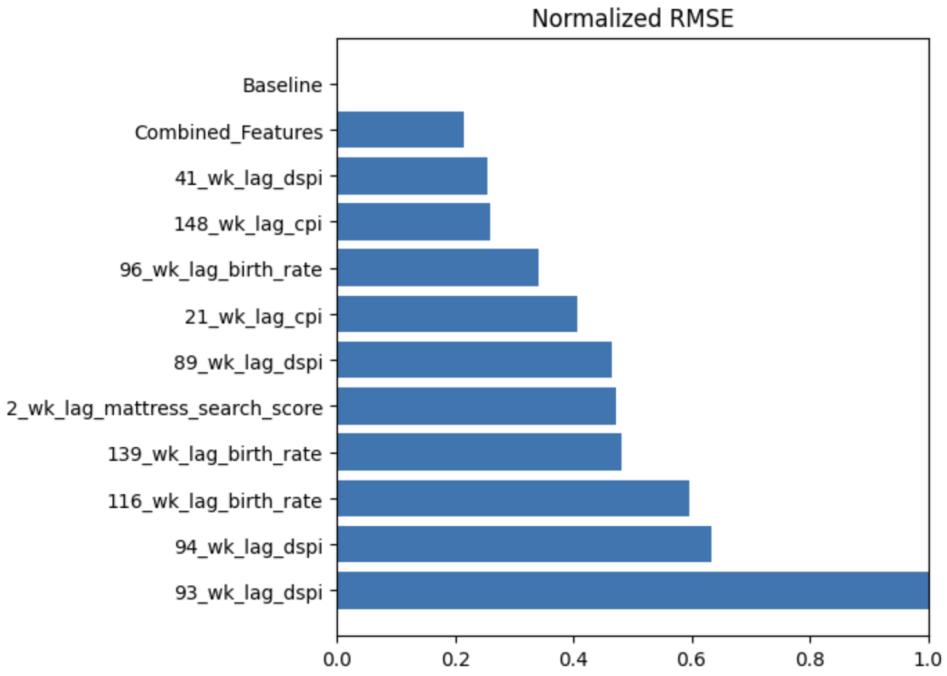
For Brand A, Product B, Channel A, the normalized RMSE results in Figure 19 exhibit a different pattern; most models incorporating exogenous variables outperform the baseline across standard error metrics, with the 12-week, 30-week, and 8-week lagged consumer sentiment index emerging as the top three normalized RMSE performers. The remaining metrics show consistent results, as presented in Appendix Figure 36. Figures 20–21 show the forecast versus actual sales plots for the baseline model and the top-performing exogenous variables model. For results of the other models with exogenous variables, see Appendix (Figures 37–46)

For Brand B, Product A, Channel A, as shown in Figure 22, a similar pattern to Brand A, Product A, Channel A is observed – baseline model outperforms others, with the combined feature set showing normalized RMSE values closest to the baseline. This is followed by the 114-week lagged disposable personal income and the 1-week lagged web traffic for mattresses. Figures 23–24 present the forecast versus actual sales plots for the baseline and the top-performing model with exogenous variables. For other metrics, see Appendix Figure 47. For results of the remaining models, see Appendix (Figures 48–57).

In summary, incorporating exogenous variables does not guarantee improved performance. While some combinations yield promising results, others show little to no benefit. Effectiveness varies by brand, product, and channel sales patterns, and may also shift depending on how strong the baseline model is in terms of feature engineering and hyperparameter tuning.

Brand A Product A Channel A (A-A-A) Normalized Metric Comparisons

Figure 16: Brand A Product A Channel A (A-A-A) Normalized Metric | RMSE



Brand A Product A Channel A (A-A-A): Forecasted vs. Actual Sales

Figure 17: Brand A Product A Channel A (A-A-A) Forecasted vs. Actual Sales | Baseline

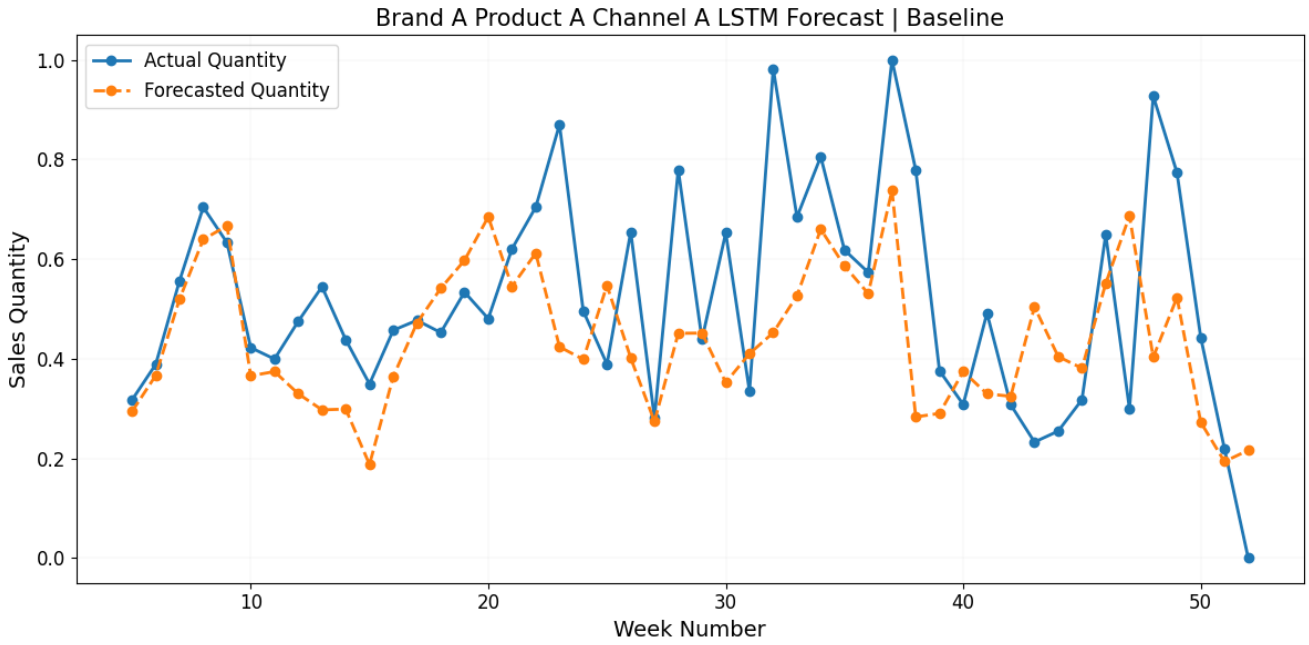
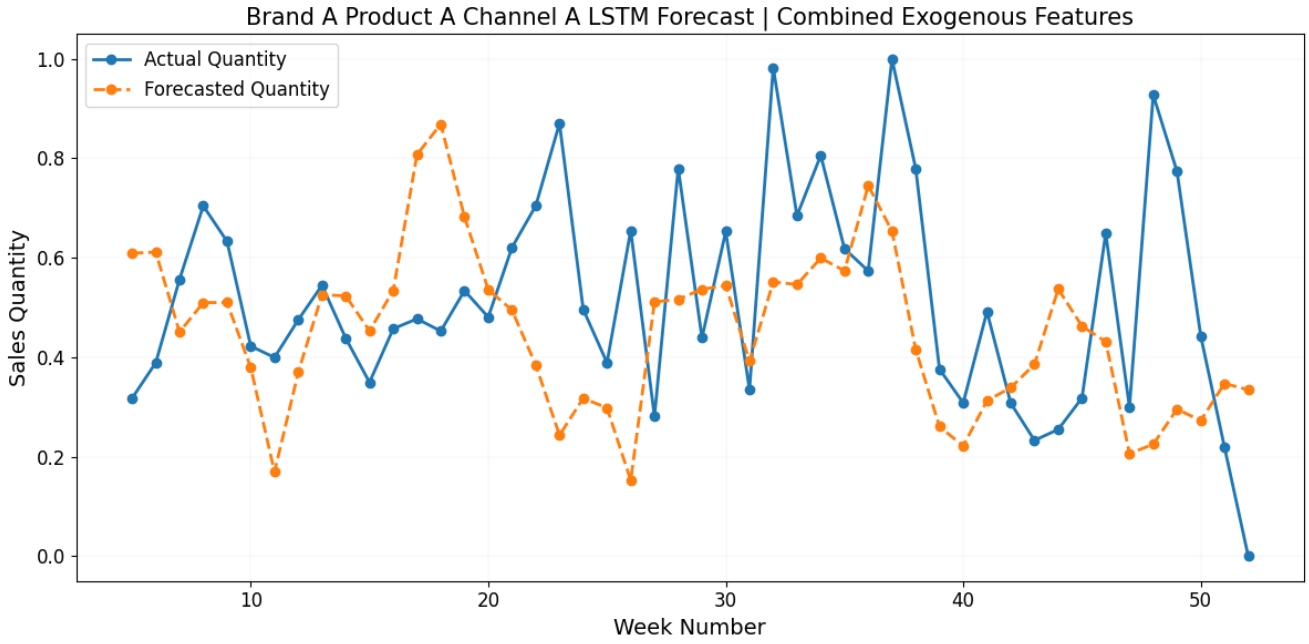


Figure 18: Brand A Product A Channel A (A-A-A) Forecasted vs. Actual Sales |Exogenous Variable: Combined Features



Brand A Product B Channel A (A-B-A) Normalized Metric Comparisons

Figure 19: Brand A Product B Channel A (A-B-A) Normalized Metric | RMSE

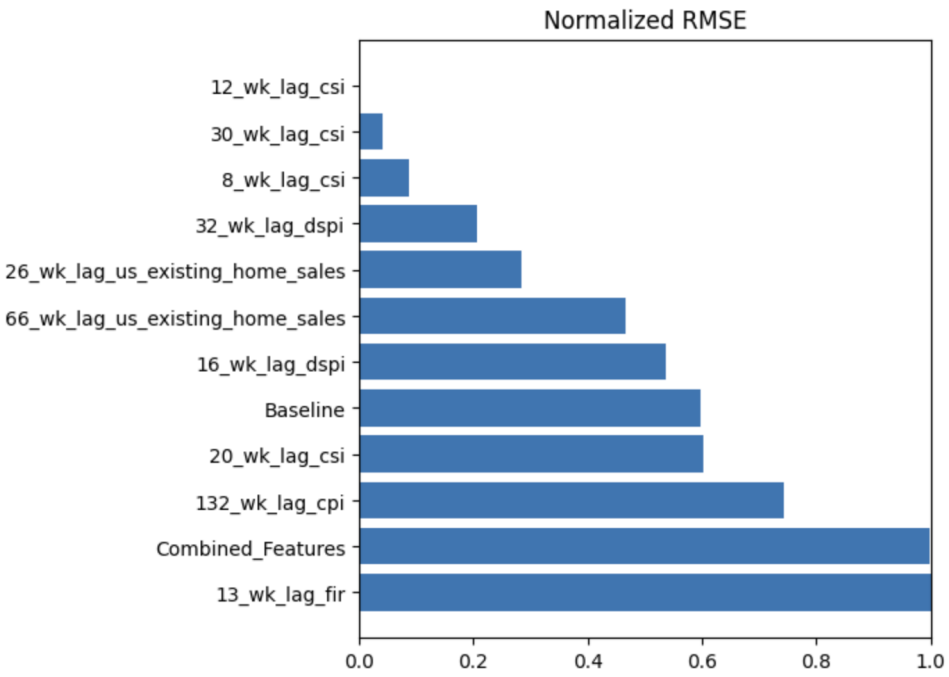


Figure 20: Brand A Product B Channel A (A-B-A) Forecasted vs. Actual Sales | Baseline

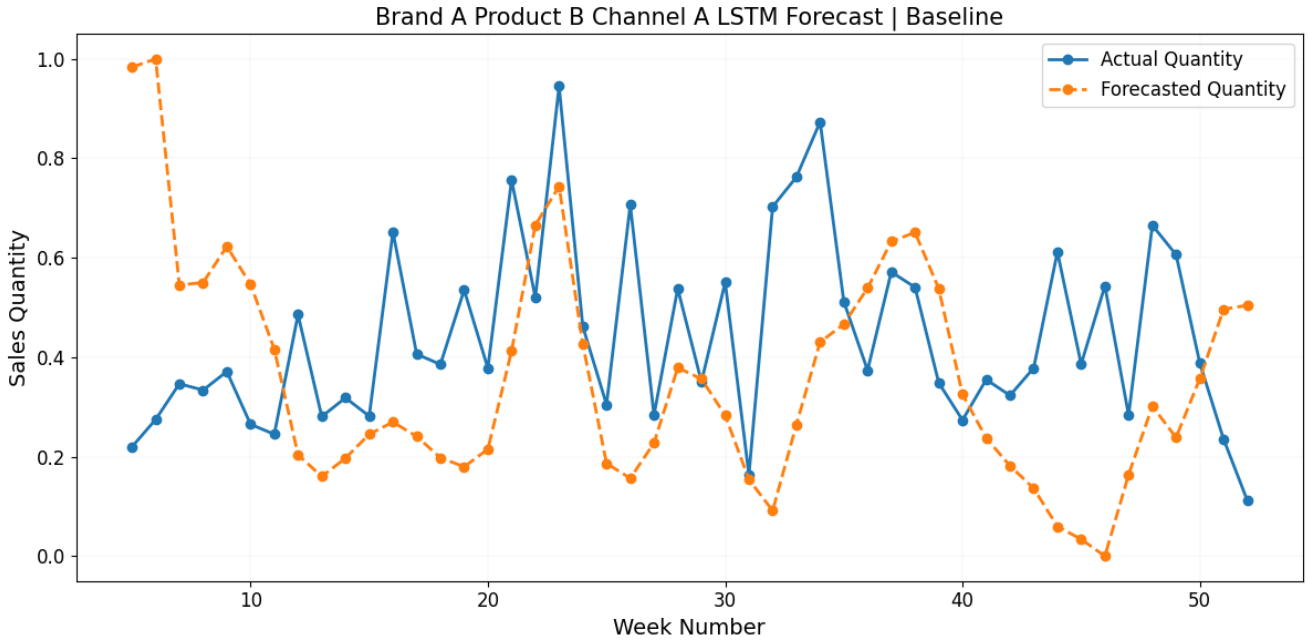
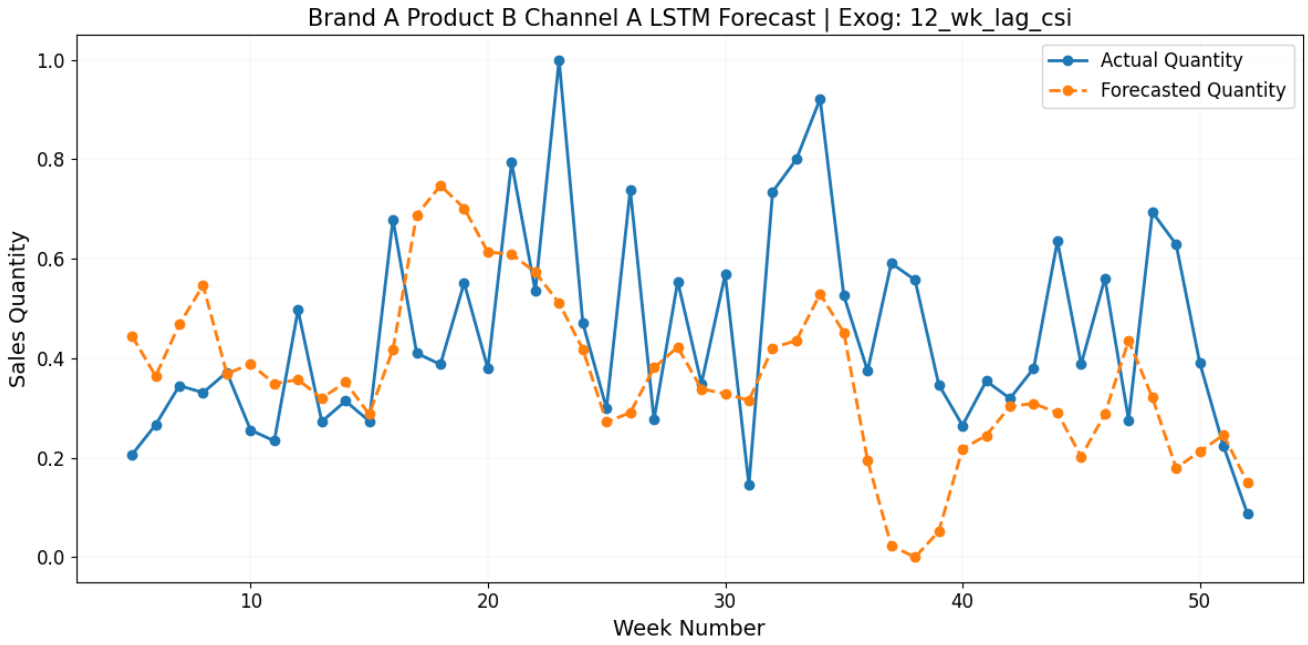
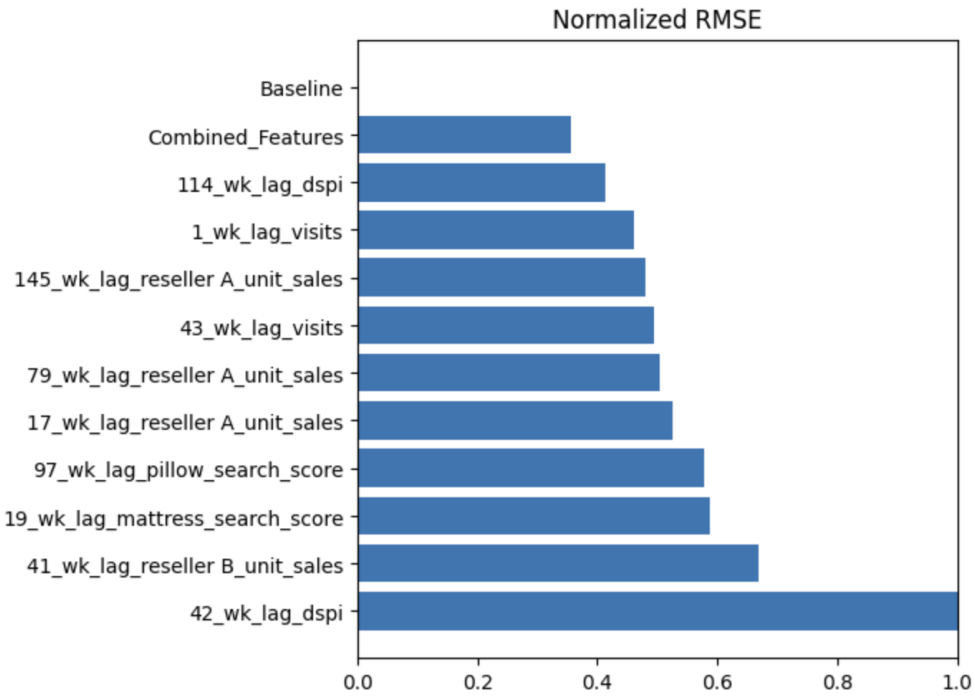


Figure 21: Brand A Product B Channel A (A-B-A) Forecasted vs. Actual Sales | Exogenous Variable: 12–week Lagged Consumer Sentiment Index



Brand B Product A Channel A (B-A-A) Normalized Metric Comparisons

Figure 22: Brand B Product A Channel A (B-A-A) Normalized Metric | RMSE



Brand B Product A Channel A (B-A-A): Forecasted vs. Actual Sales

Figure 23: Brand B Product A Channel A (B-A-A) Forecasted vs. Actual Sales | Baseline

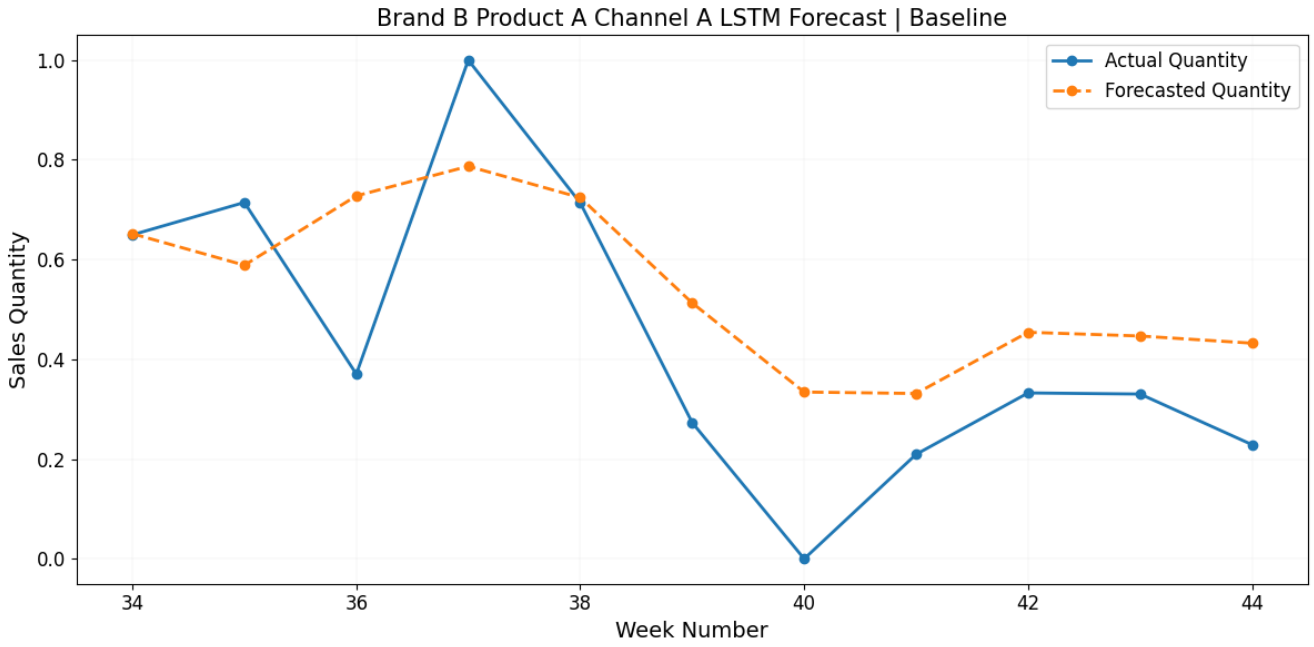
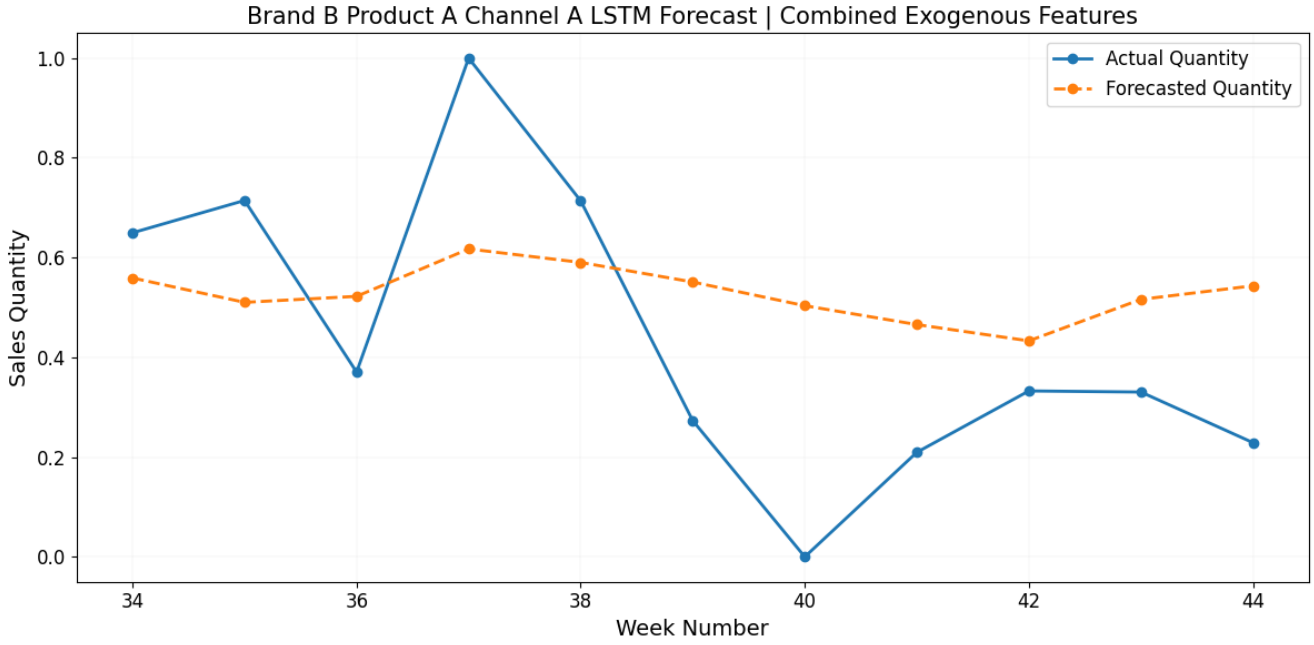


Figure 24: Brand B Product A Channel A (B-A-A) Forecasted vs. Actual Sales| Exogenous Variable: Combined Features



4.2 Implications

The following sections discuss the broader implications of this study's findings for the demand forecasting domain. These implications highlight the practical value of the research and offer insights for professionals seeking to enhance their forecasting models with exogenous variables.

4.2.1 Exogenous Variable Evaluation Methodology

The exogenous variable evaluation methodology in this study carries significant implications for demand forecasting practices. The results highlight the importance of a nuanced approach when incorporating external data into forecasting models. While the study demonstrates the potential value of integrating exogenous variables, it also underscores the necessity for careful selection and analysis. This necessity is seen in the finding that forecast accuracy does not uniformly improve with the inclusion of exogenous variables suggesting that the relationship between these variables and demand is complex and context-dependent. Therefore, the inclusion of variables requires domain expertise and rigorous statistical analysis to identify their appropriate relationships.

4.2.2 Exogenous Variable Lag Value

The exogenous variable lag analysis has several important implications for demand forecasting at TSI. First, the identification of both short- and long-lagged features indicates that external factors influence demand across multiple time horizons. Short-lagged relationships, often within a few weeks to a few months, suggest that certain external variables, such as consumer web activity or retail sales patterns, can have near-immediate effects on demand. These features present a promising opportunity for enhancing forecast responsiveness if they can be reliably monitored and incorporated in near real-time.

Conversely, the presence of meaningful long-lagged relationships, some approaching two to three years, highlights that broader macroeconomic trends and consumer behavior cycles also play a role in shaping future demand. However, caution is warranted when interpreting long-lagged features, as the potential for spurious correlations increases with greater time gaps. Without proper validation, these long lags could reflect coincidental historical patterns rather than true predictive signals. Thus, while long-lagged features offer potential strategic insights for medium-to-long-term planning, they should be used carefully and preferably supplemented with additional business knowledge or validation analyses.

Overall, the lag analysis reinforces the need for a flexible modeling approach that can capture both immediate and delayed external influences. By selectively integrating validated short- and long-lagged exogenous features into forecasting processes, TSI can better anticipate both rapid shifts and longer-term structural changes in demand.

4.3 Limitations

The results of this study offer promising insights into the value of incorporating exogenous variables for enhanced demand forecasting. However, the analysis is subject to certain limitations that should be

considered when interpreting the findings and implementing the proposed methodologies. The four most relevant limitations are discussed below.

4.3.1 Data Availability

Exogenous datasets are critical to enhancing demand forecasting models, yet their acquisition presents a substantial limitation. These datasets must be sourced externally, introducing challenges in ensuring their consistent availability and accessibility. The reliability of forecasts is highly dependent on the currency and completeness of the data used; therefore, any inconsistencies in data collection or updates can impede the forecasting process. Furthermore, access to certain datasets may be restricted or come at a high cost, limiting the scope of variables that can be included in the analysis. This constraint can force practitioners to make compromises in their model development, potentially affecting the accuracy and robustness of the final demand forecasts.

4.3.2 Data Leakage

Data leakage poses a significant challenge when incorporating exogenous variables, potentially compromising the validity of forecasting models. Data leakage occurs when information from the test set inadvertently influences the training of the model, leading to over-optimistic performance metrics. For instance, if there is a lagged significance of one week between Google Trends data and the target variable, the forecast horizon is limited to one week, unless future Google Trends data can be accurately predicted. Accurate forecasting of exogenous variables is difficult and introduces additional layers of complexity and potential error. This limitation necessitates careful management of lag effects and highlights the importance of rigorous validation techniques to ensure the model's predictive capability is not overestimated.

4.3.3 Mutual Information Behavior with Time Horizon Shifts

Another limitation is that the relationship between target variables and exogenous features may exhibit variability across different timeframes. The strength and direction of these relationships are not static and can shift due to evolving market dynamics, economic cycles, or unforeseen events. For example, the relationship between housing starts and demand for bedding products might be strong during periods of economic expansion but weaken during recessions. Similarly, changes in consumer behavior, technological adoption, or channel preferences could alter how predictive certain variables are over time.

Given that the data is dynamic, ongoing monitoring of exogenous variable behavior is essential. If an exogenous variable initially improves forecast accuracy, TSI must continue to observe whether its relationship with the target sales remains strong and consistent as new data becomes available. Metrics,

such as rolling mutual information scores, should be tracked to assess whether the variable continues to add value. If the dependency weakens significantly or if the feature begins to introduce noise rather than predictive signal, then it may be necessary to retrain the model, revalidate the feature set, or remove the variable from the forecasting process altogether. Exogenous variables should be treated as "living" data inputs, subject to change, for sustaining model reliability and avoiding forecast degradation over time.

4.3.4 Spurious Correlations

A notable limitation arises with the interpretation of features at very long lags. While the lag analysis may indicate significance, the practical relevance of such extended lags becomes questionable. In these instances, there is a higher risk of identifying spurious correlations, where the observed relationship is merely coincidental rather than indicative of a true predictive relationship. These relationships may not represent true causal links and could lead to unreliable forecasting models if given undue weight. Therefore, caution must be exercised when considering the implications of correlations with long lag times, and additional validation or domain expertise should be employed to discern meaningful relationships from spurious ones.

4.3.5 Alternative Model Configurations

One key limitation of this study is the reliance on a basic LSTM architecture to evaluate the impact of exogenous variables on forecast performance. While LSTM networks are well-suited for capturing temporal dependencies and nonlinear relationships, their performance is sensitive to architecture design and feature interaction modeling. In the current implementation, exogenous variables were introduced as direct inputs alongside engineered internal features. However, this unified treatment may have constrained the model's ability to disentangle the unique signal contribution of each external indicator, particularly when signals are weak or interact in complex ways.

Alternative modeling approaches may offer improved performance in this context; however, the model that is chosen for this task also must be able to handle multiple inputs, similar to the LSTM model in this study. Tree-based ML models, such as Random Forest or Gradient Boosting, provide robustness to multicollinearity and are better equipped to handle sparse or non-additive relationships between features. Additionally, simpler models such as regularized linear regression may serve as useful baselines, particularly in cases where interpretability is prioritized over predictive power. Beyond model selection, architectural refinements to the LSTM framework could also enhance performance. For example, assigning exogenous variables to dedicated subnetworks, where each is trained to extract temporal patterns specific to a particular indicator, could allow for more nuanced feature processing. These subnetworks could then be

combined via an attention layer that dynamically learns the relative contribution of each signal to the final sales prediction. Such techniques may be especially helpful in amplifying weak but meaningful exogenous effects that are otherwise diluted in a shared model input structure.

5. Conclusion

This study provides a framework to evaluate the significance of third-party datasets that could meaningfully enhance TSI's omnichannel demand forecasting capabilities. By developing a structured methodology, combining mutual information-based feature selection and LSTM modeling, the study tests whether exogenous variables, when appropriately shifted and selected, could serve as forward indicators of sales. As hypothesized, certain exogenous features do improve forecast performance under specific conditions while others do not guarantee better performance. Improvements vary by brand, product, and channel, and in some cases, exogenous variables underperformed the internal baseline. These mixed outcomes suggest that predictive value is data-dependent and highly sensitive to modeling choices. Nonetheless, the findings offer a valuable starting point for selectively integrating forward-looking signals into TSI's demand planning process, and the provided framework sets the stage for future exploration and refinement.

5.1 Management Recommendations

Several practical recommendations arise from this study for TSI's forecasting and planning teams. First, demand forecasting models should incorporate only those exogenous variables that demonstrate clear and timely predictive value. This means placing particular emphasis on variables whose effects align with the operational rhythms of the business. A targeted, evidence-based selection process ensures that only the most relevant signals are integrated, enhancing model reliability without introducing unnecessary complexity. For example, variables with short lag times, such as web activity or near-term retail signals, are more likely to reflect immediate shifts in consumer behavior and can be readily aligned with existing planning cycles. These features offer the most practical potential for improving agility in replenishment and production decisions. Second, variables with long lag effects, while occasionally showing a strong relationship, should be treated with caution. These features may reflect structural trends or replacement cycles but are also more susceptible to spurious relationships. Their inclusion in operational forecasts should be limited unless they are supported by domain knowledge or confirmed through continued validation over time. Third, TSI should invest in developing a system for monitoring variables and retraining the model. Relationships between external variables and internal demand can shift as consumer preferences and market dynamics evolve.

Without a mechanism for regularly reassessing variable relevance and retraining models accordingly, forecasting accuracy will degrade over time. Finally, the modeling strategy itself should remain flexible. While LSTM provided a solid foundation for capturing time-based patterns and nonlinear dynamics, the findings point to an even greater opportunity: exploring alternative model architectures. In particular, ensemble tree methods and attention-based deep learning frameworks warrant further investment, as they may offer stronger predictive performance, better interpretability, and improved handling of exogenous feature interactions—especially when variable relationships are complex or weakly expressed.

5.2 Future Work

Future research can build on this foundation in several ways. One promising direction involves experimenting with alternative model architectures, including those that explicitly assign exogenous variables to dedicated subnetworks and combine their outputs via attention mechanisms. This approach could better isolate the contribution of individual external signals and reduce interference from weak or noisy inputs. Further exploration of hybrid modeling techniques, such as combining deep learning with more interpretable models like Random Forests, may also provide a stronger balance between performance and explainability. These techniques could be especially valuable for business users seeking to understand the drivers behind model predictions. Additionally, future research should incorporate a portfolio of baseline models, rather than relying on a single architecture, when evaluating the impact of new exogenous variables. This strategy aligns with the framework developed in this study, but it enhances the robustness by identifying whether at least one model consistently benefits from the added feature. If even one model in the set demonstrates improved performance, the variable in question may merit inclusion. This approach provides a more resilient validation mechanism for feature selection, ensuring that valuable signals are not overlooked due to model-specific sensitivities. Another opportunity lies in expanding the feature space to include unstructured external data, such as news sentiment or social media indicators, which may offer early signals of shifts in consumer interest or macroeconomic conditions. Finally, future efforts should focus on integrating these models into TSI's broader operational workflows by ensuring that forecasting insights can be quickly translated into inventory, fulfillment, and capacity decisions. By continuing to test, refine, and operationalize the selective use of exogenous data, TSI can evolve its forecasting approach to meet the demands of a retail environment that increasingly rewards speed, precision, and insight.

References

Akwa-Mensah, H. (2023). *Examining the Sustained Adoption of Omnichannel Shopping Beyond the COVID-19 Pandemic* (Order No. 30866439). Available from ProQuest Dissertations & Theses Global. (2891878363).

Brownlee, J. (2020). *Data Preparation for Machine Learning: Data Cleaning, Feature Selection, and Data Transforms in Python.* Machine Learning Mastery.

Chollet, F. (2021). *Deep Learning with Python* (2nd ed.). Manning Publications.

Feizabadi, J. (2020). Machine learning demand forecasting and supply chain performance. *International Journal of Logistics Research and Applications*, 25(2), 119–142.

Franzén, A. (2023). Big data, big problems: Why scientists should refrain from using Google Trends. *Acta Sociologica*, 66(3), 343-347.

Géron, A. (2019). *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow.* O'Reilly Media, Inc.

Google (n.d.). FAQ about Google Trends data. Trends Help.

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling.* Springer.

Mavragani, A., Ochoa, G., & Tsagarakis, K. P. (2018). Assessing the methods, tools, and statistical approaches in google trends research: Systematic review. *Journal of Medical Internet Research*, 20(11).

McKinney, W. (2016). *Data wrangling with Python.* O'Reilly Media.

Medeiros, M. C. and Pires, H. F. (2021). The proper use of Google trends in forecasting models. arXiv preprint arXiv:2104.03065.

Rocke, R., & Zhang, L. (2024). *Impact of Exogenous Variables on AI/ML Predictive Algorithm Prophet, in the FMCG Industry* [MIT SCM Master's Capstone].

Seyedan, M., Mafakheri, F. Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities. *J Big Data 7*, 53 (2020).

Sinaga, O., Saudi, M. H. M., & Roespinoedji, D. (2018). The relationship between economic indicators, gross domestic product (GDP) and supply chain performance. *Polish Journal of Management Studies*, 18(1), 338–352.

Yasir, M., Ansari, Y., Latif, K., Maqsood, H., Habib, A., Moon, J., & Rho, S. (2022). Machine learning–assisted efficient demand forecasting using endogenous and exogenous indicators for the textile industry. *International Journal of Logistics Research and Applications*, 1–20.

Glossary

Table 5: Glossary of Relevant Acronyms

Acronym	Term
TSI	Tempur-Sealy International
ML	Machine Learning
LSTM	Long Short-Term Memory
RNN	Recurrent Neural Network
EDA	Exploratory Data Analysis
MI	Mutual Information
mRMR	minimum Redundancy Maximum Relevance
MAE	Mean Absolute Error
MSE	Mean Squared Error
RMSE	Root Mean Squared Error
MAPE	Mean Absolute Percentage Error
CPI	Consumer Price Index
DSPI	Disposable Personal Income
CSI	Consumer Sentiment Index

Appendices

Brand A Product A Channel A (A-A-A): Normalized Metrics and Forecasted vs. Actual Sales

Figure 25: Brand A Product A Channel A (A-A-A) Normalized Metrics | MAE, MSE, RMSE, MAPE

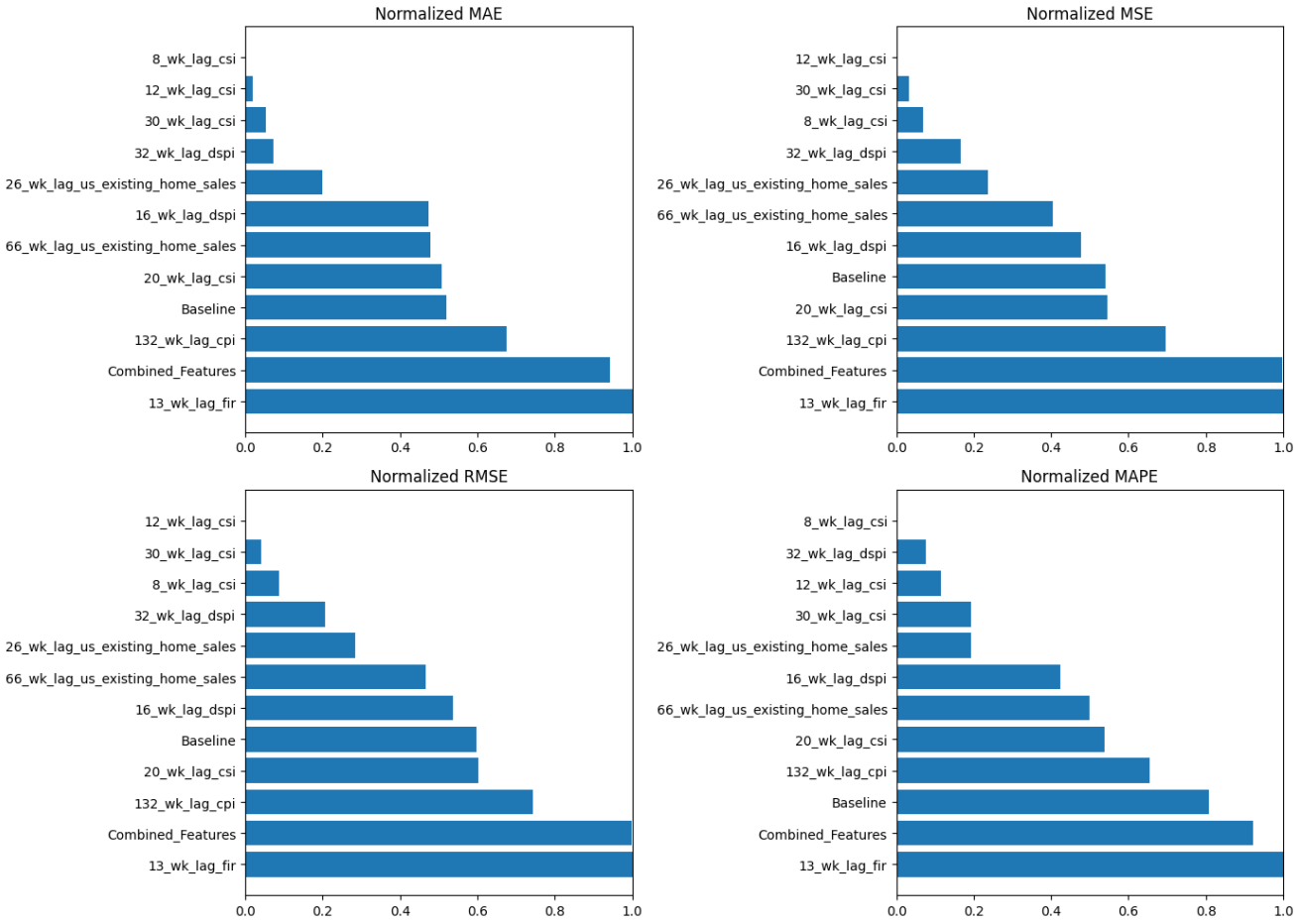


Figure 26: Brand A Product A Channel A (A-A-A) Forecasted vs. Actual Sales | Exogenous Variable: 41–Week Lagged Disposable Personal Income

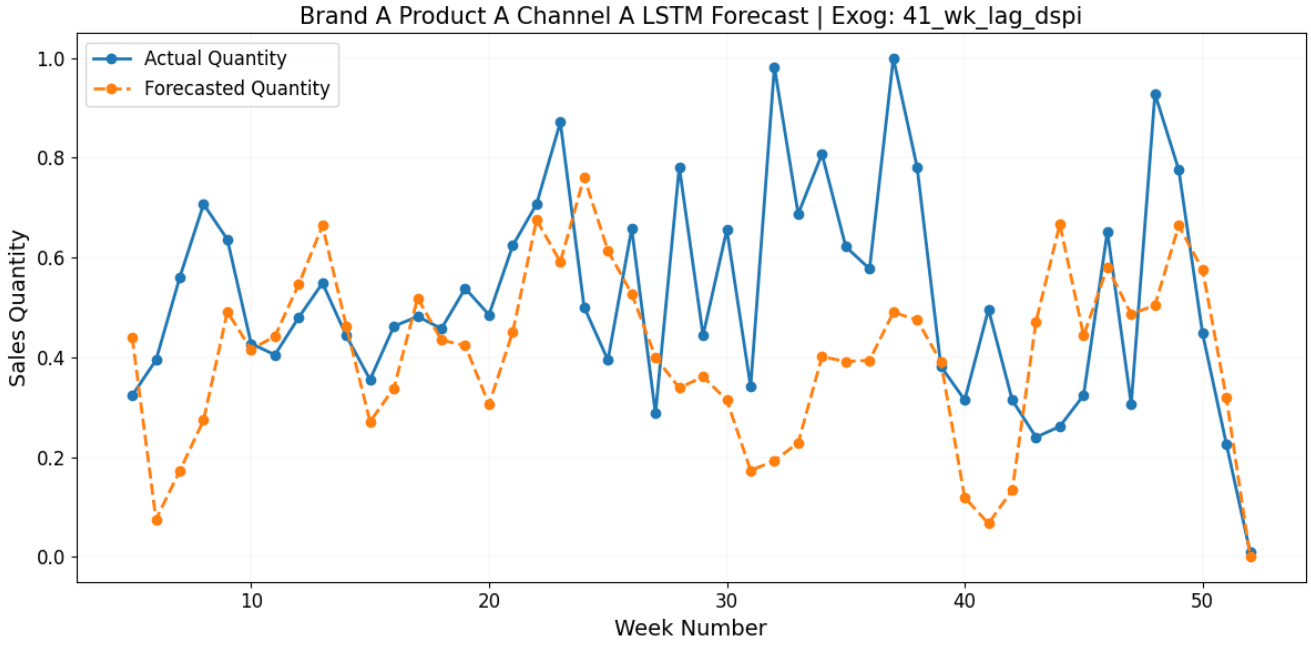


Figure 27: Brand A Product A Channel A (A-A-A) Forecasted vs. Actual Sales | Exogenous Variable: 148–Week Lagged Consumer Price Index

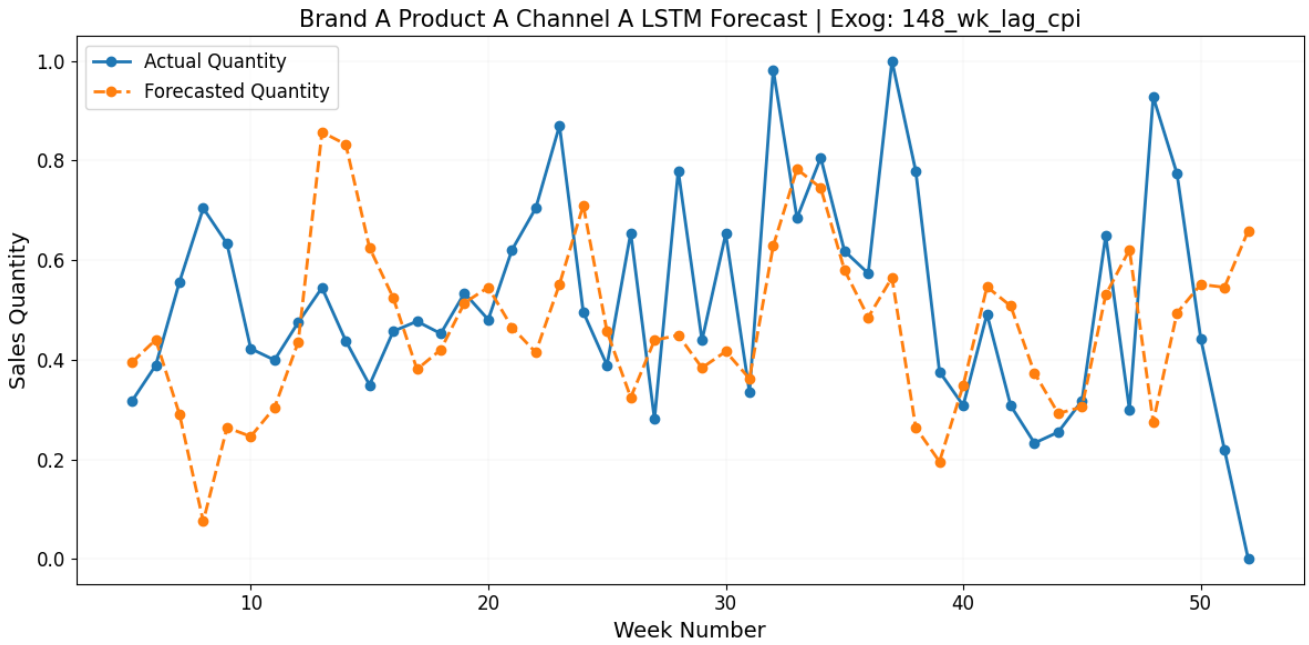


Figure 28: Brand A Product A Channel A (A-A-A) Forecasted vs. Actual Sales | Exogenous Variable: 96–Week Lagged Birth Rate

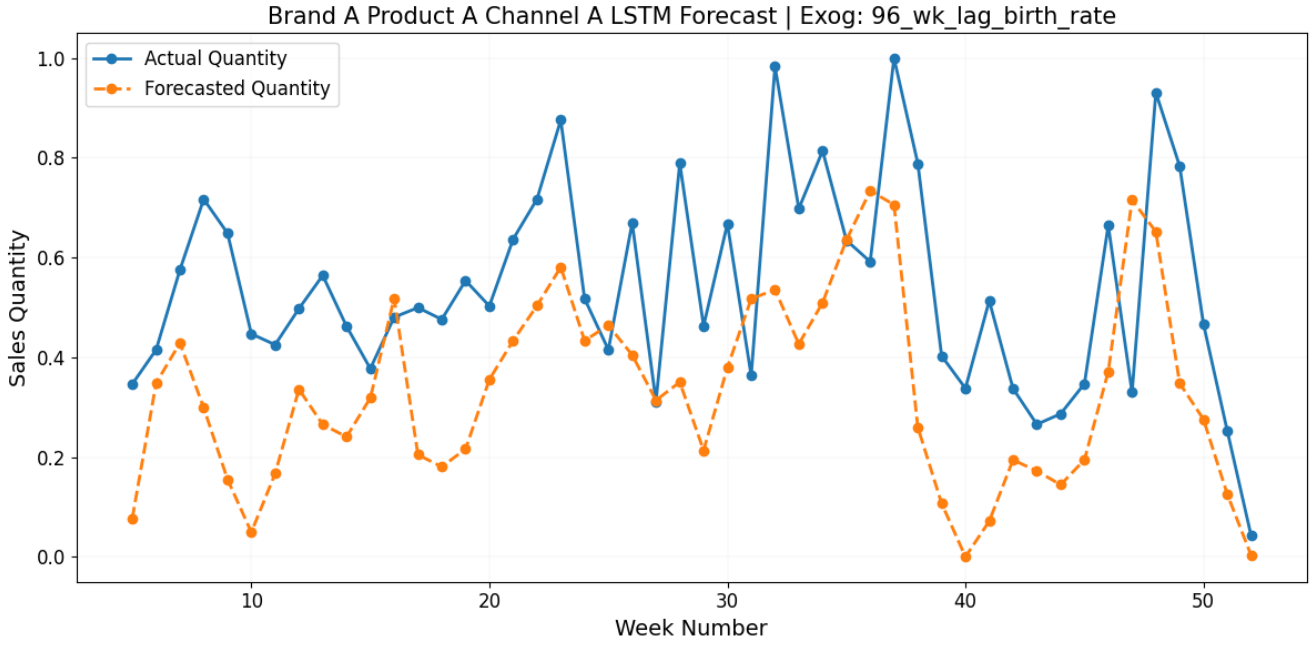


Figure 29: Brand A Product A Channel A (A-A-A) Forecasted vs. Actual Sales | Exogenous Variable: 21–Week Lagged Consumer Price Index

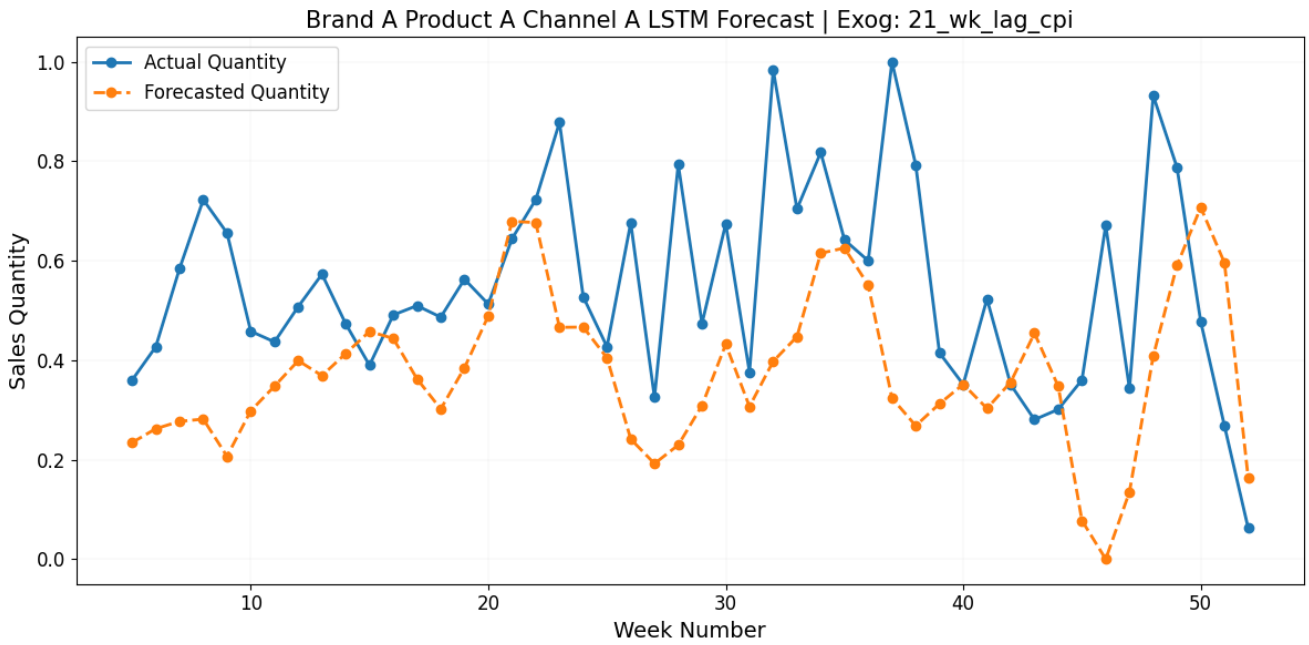


Figure 30: Brand A Product A Channel A (A-A-A) Forecasted vs. Actual Sales | Exogenous Variable: 89–Week Lagged Disposable Personal Income

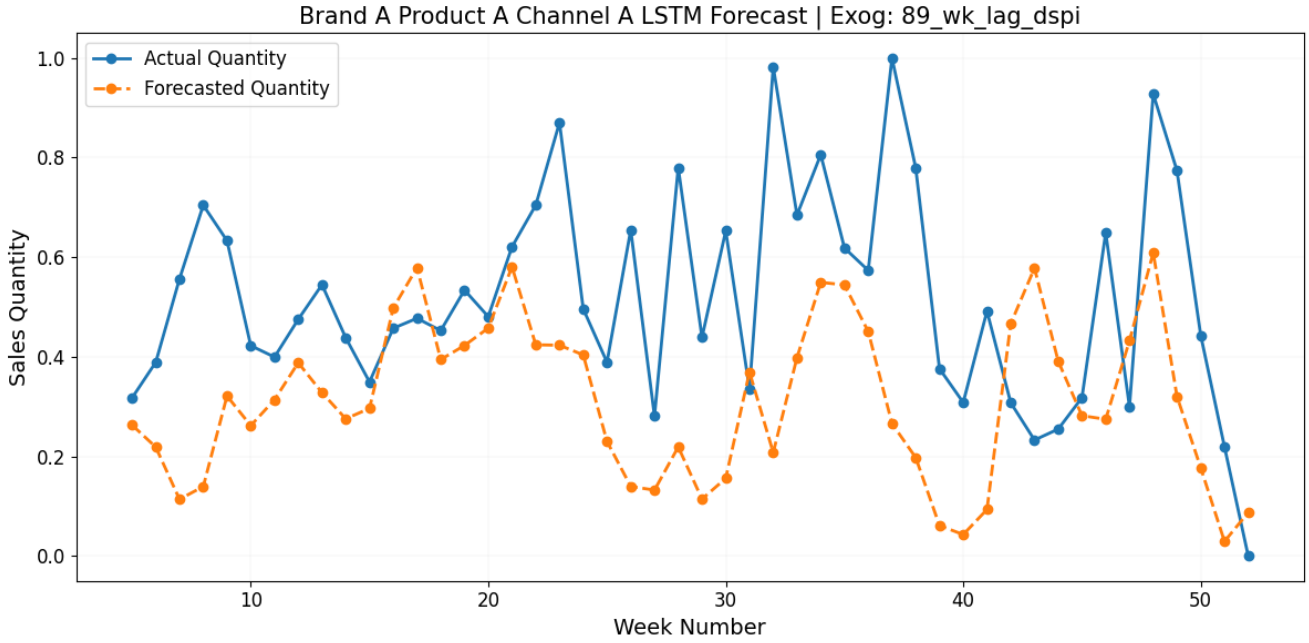


Figure 31: Brand A Product A Channel A (A-A-A) Forecasted vs. Actual Sales | Exogenous Variable: 2–Week Lagged Mattress Search Score

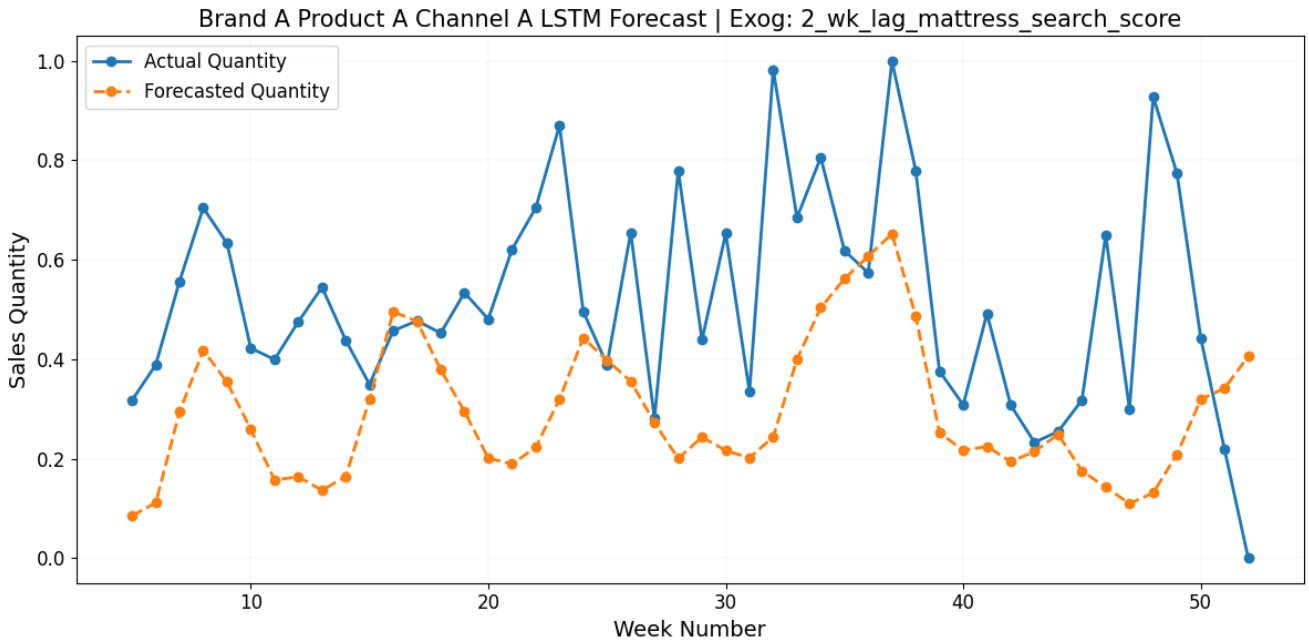


Figure 32: Brand A Product A Channel A (A-A-A) Forecasted vs. Actual Sales | Exogenous Variable: 139–Week Lagged Birth Rate

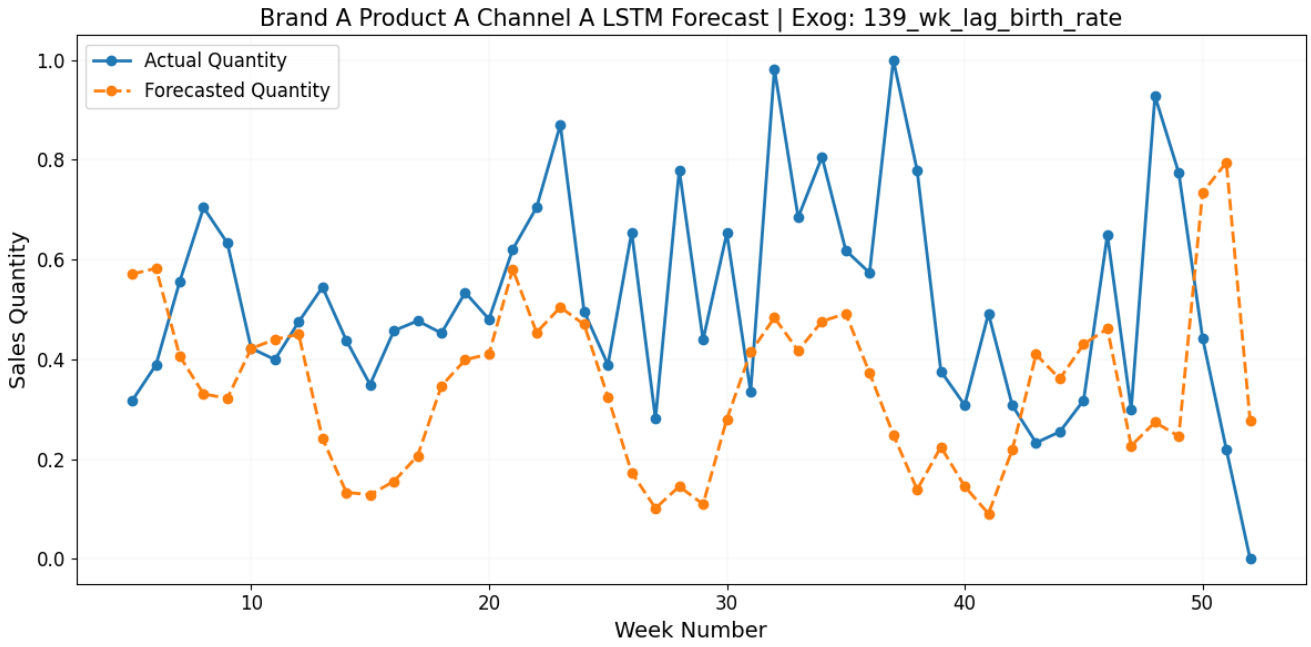


Figure 33: Brand B Product A Channel A (B-A-A) Forecasted vs. Actual Sales | Exogenous Variable: 116–Week Lagged Birth Rate

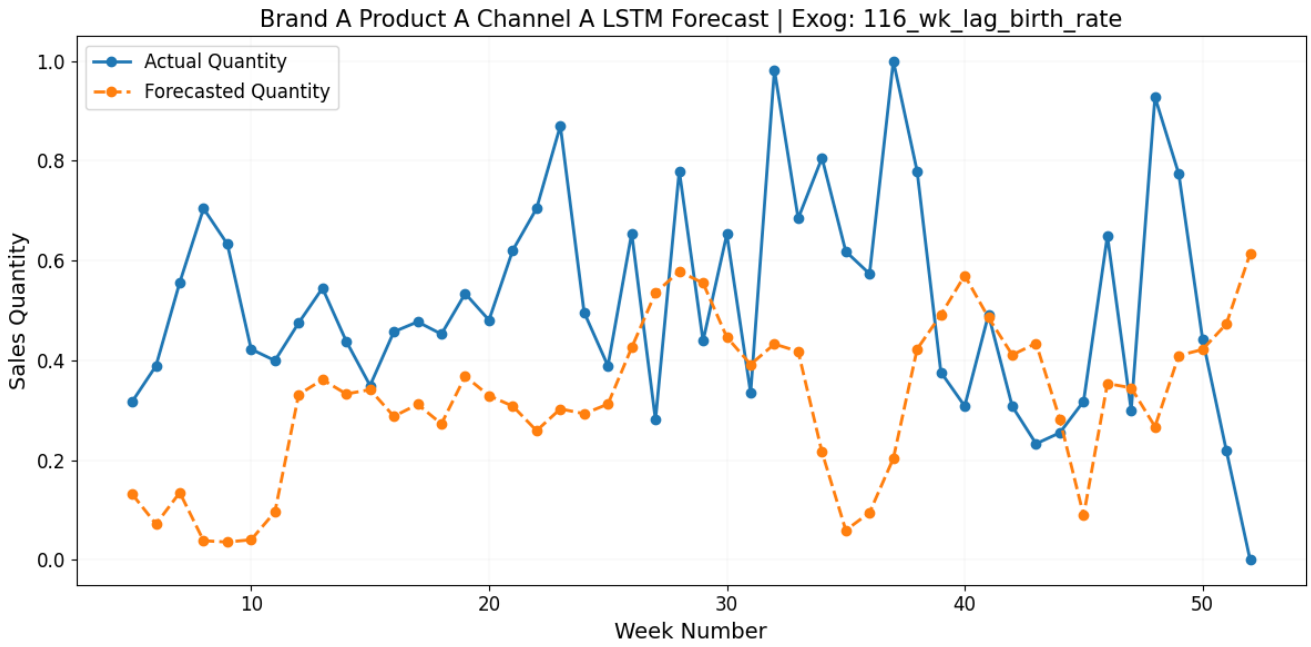


Figure 34: Brand A Product A Channel A (A-A-A) Forecasted vs. Actual Sales | Exogenous Variable: 94–Week Lagged Disposable Personal Income

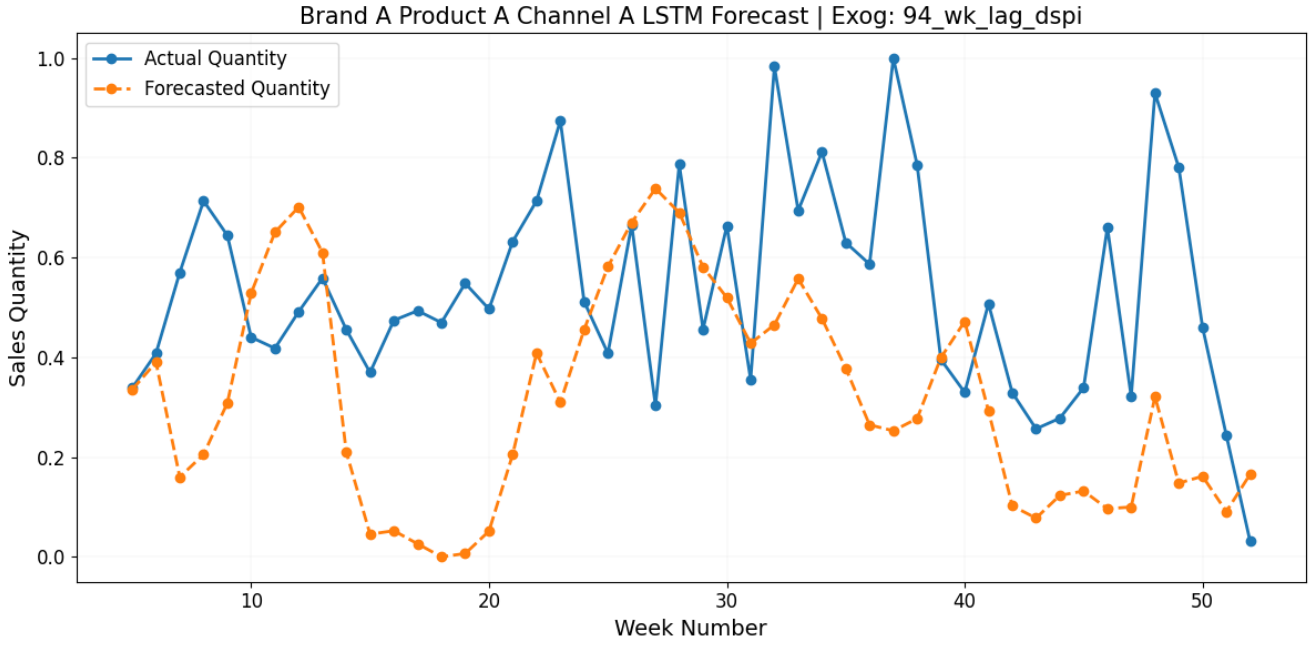
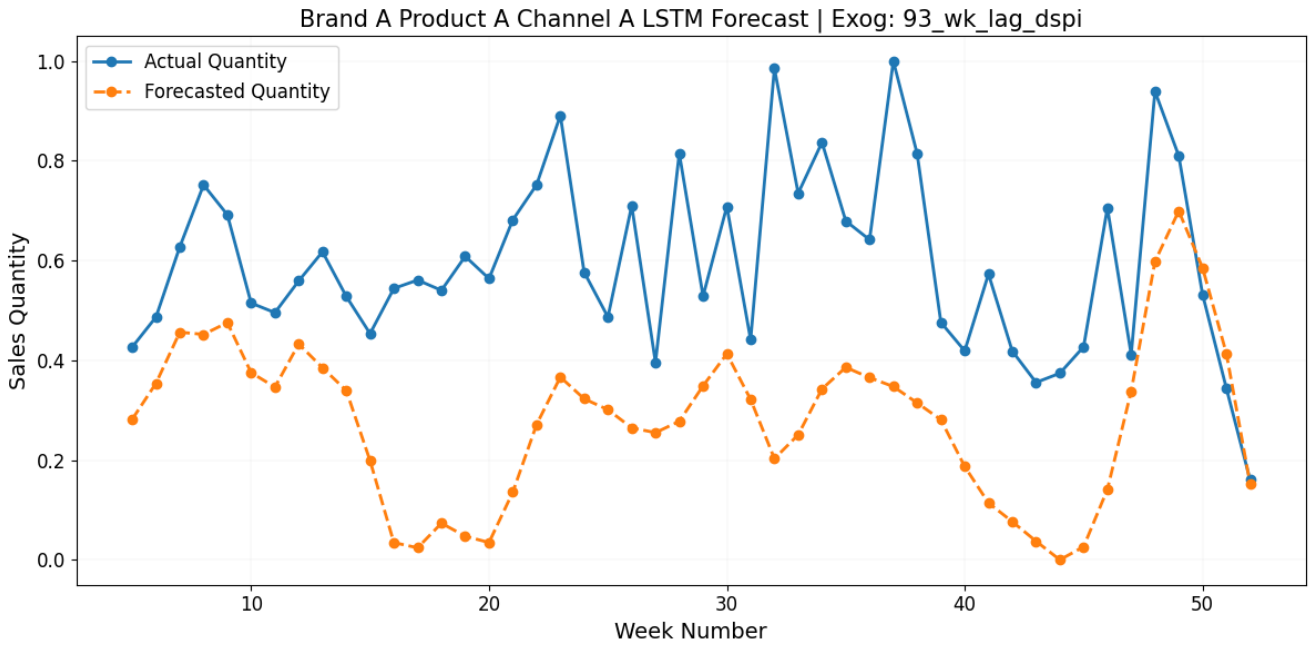


Figure 35: Brand B Product A Channel A (B-A-A) Forecasted vs. Actual Sales | Exogenous Variable: 93–Week Lagged Disposable Personal Income



Brand A Product B Channel A (A-B-A): Normalized Metrics and Forecasted vs. Actual Sales

Figure 36: Brand A Product B Channel A (A-B-A) Normalized Metrics | MAE, MSE, RMSE, MAPE

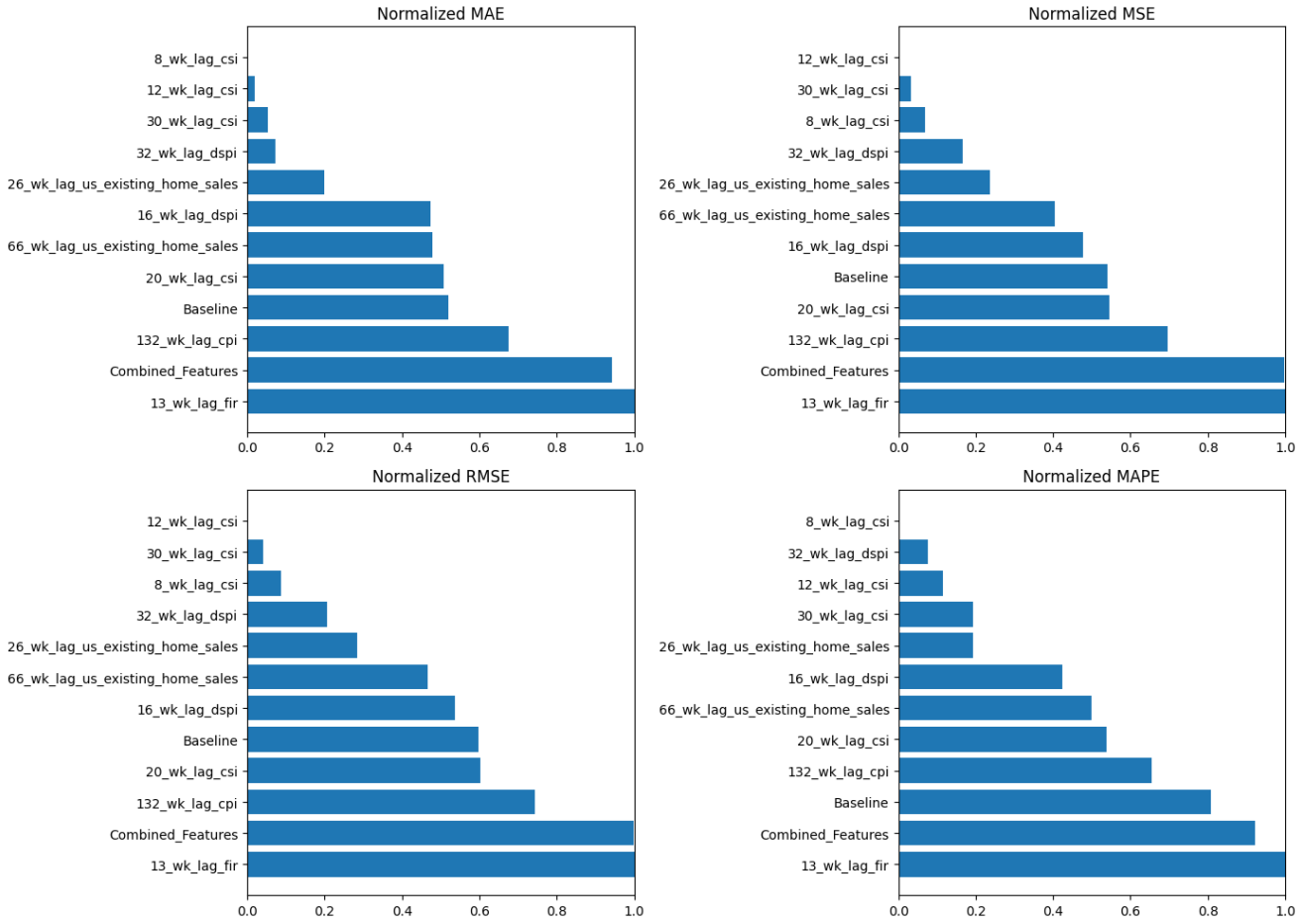


Figure 37: Brand A Product B Channel A (A-B-A) Forecasted vs. Actual Sales | Exogenous Variable: 30–Week Lagged Consumer Sentiment Index

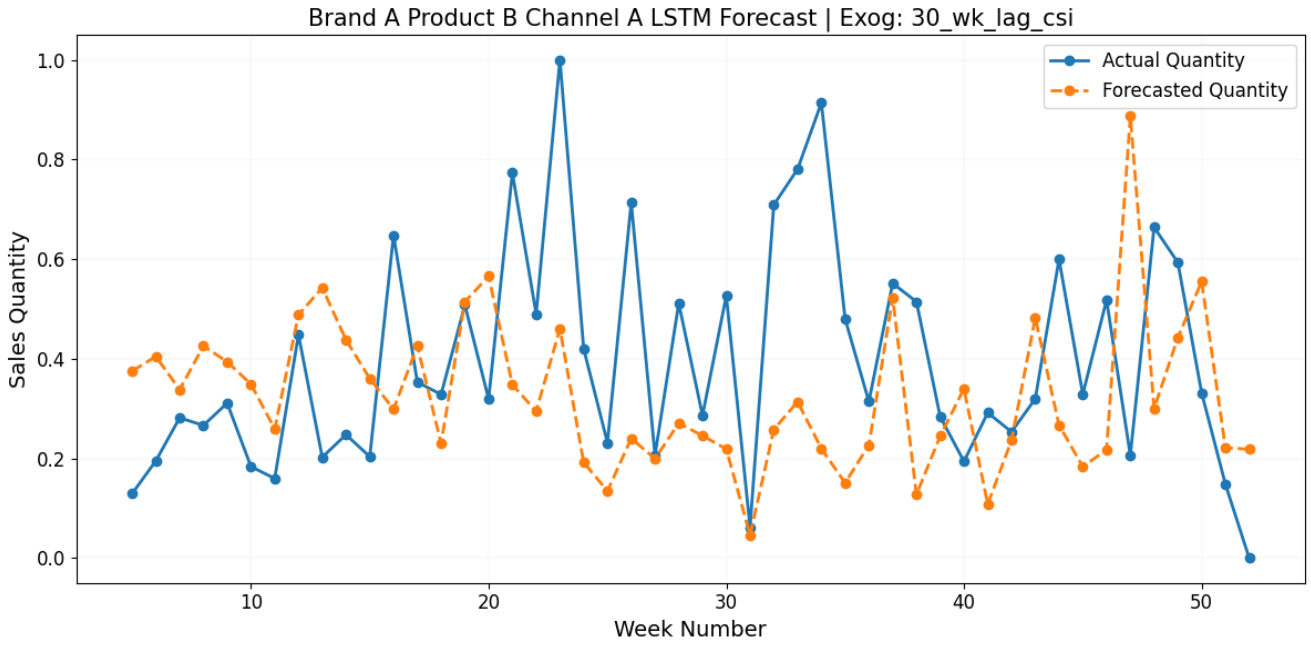


Figure 38: Brand A Product B Channel A (A-B-A) Forecasted vs. Actual Sales | Exogenous Variable: 8–Week Lagged Consumer Sentiment Index

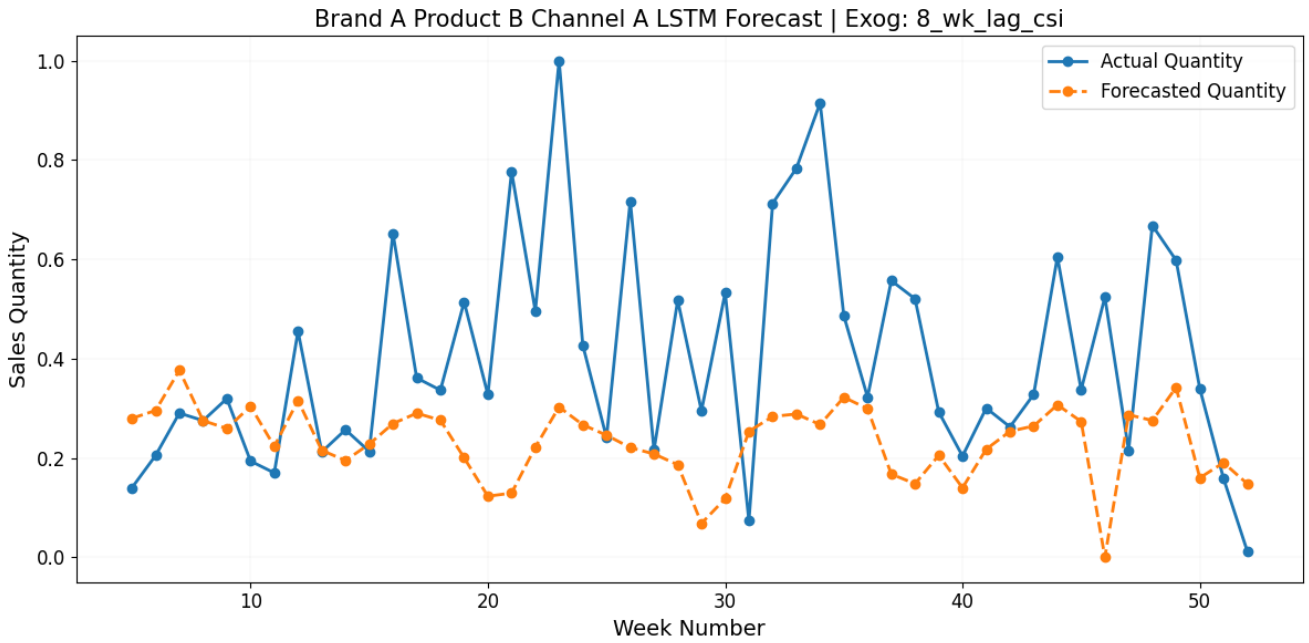


Figure 39: Brand A Product B Channel A (A-B-A) Forecasted vs. Actual Sales | Exogenous Variable: 32–Week Lagged Disposable Personal Income

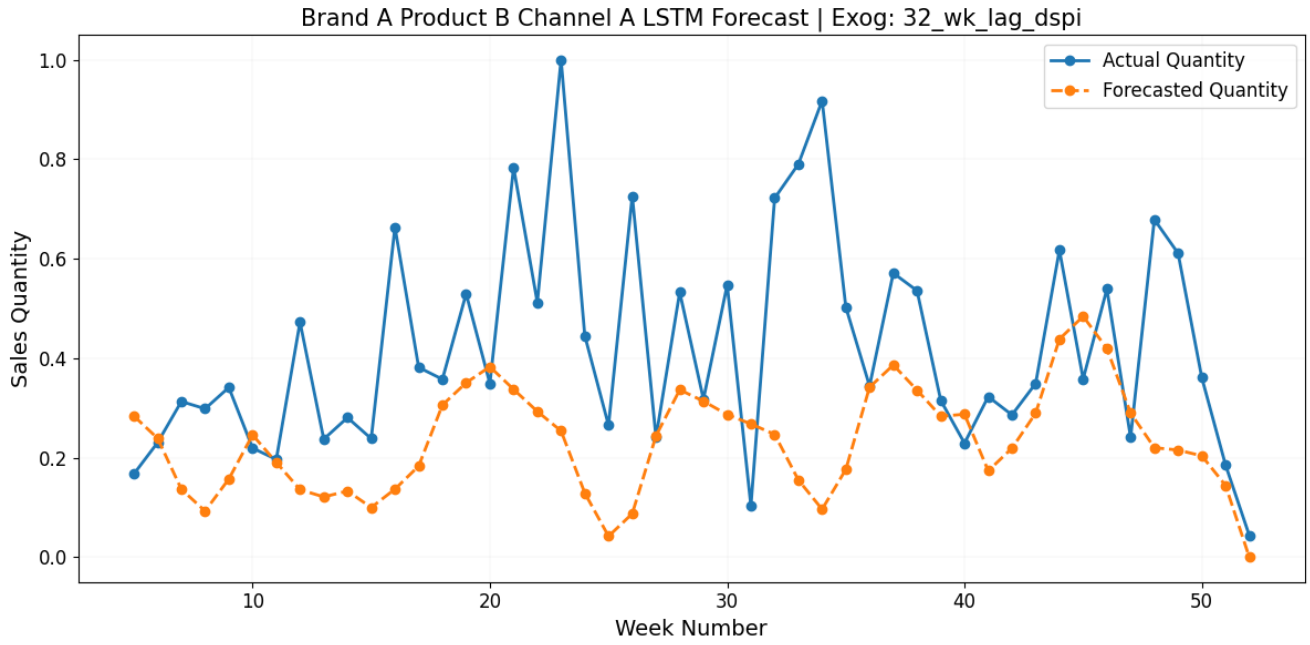


Figure 40: Brand A Product B Channel A (A-B-A) Forecasted vs. Actual Sales | Exogenous Variable: 26–Week Lagged US Existing Home Sales

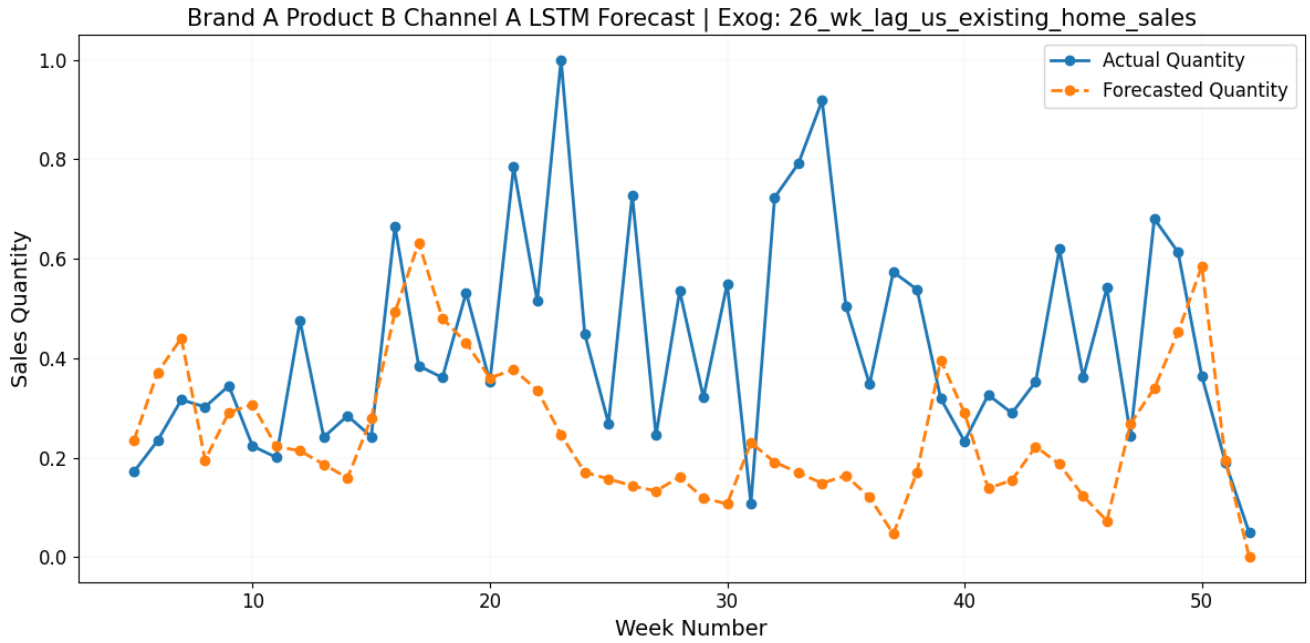


Figure 41: Brand A Product B Channel A (A-B-A) Forecasted vs. Actual Sales | Exogenous Variable: 66–Week Lagged US Existing Home Sales

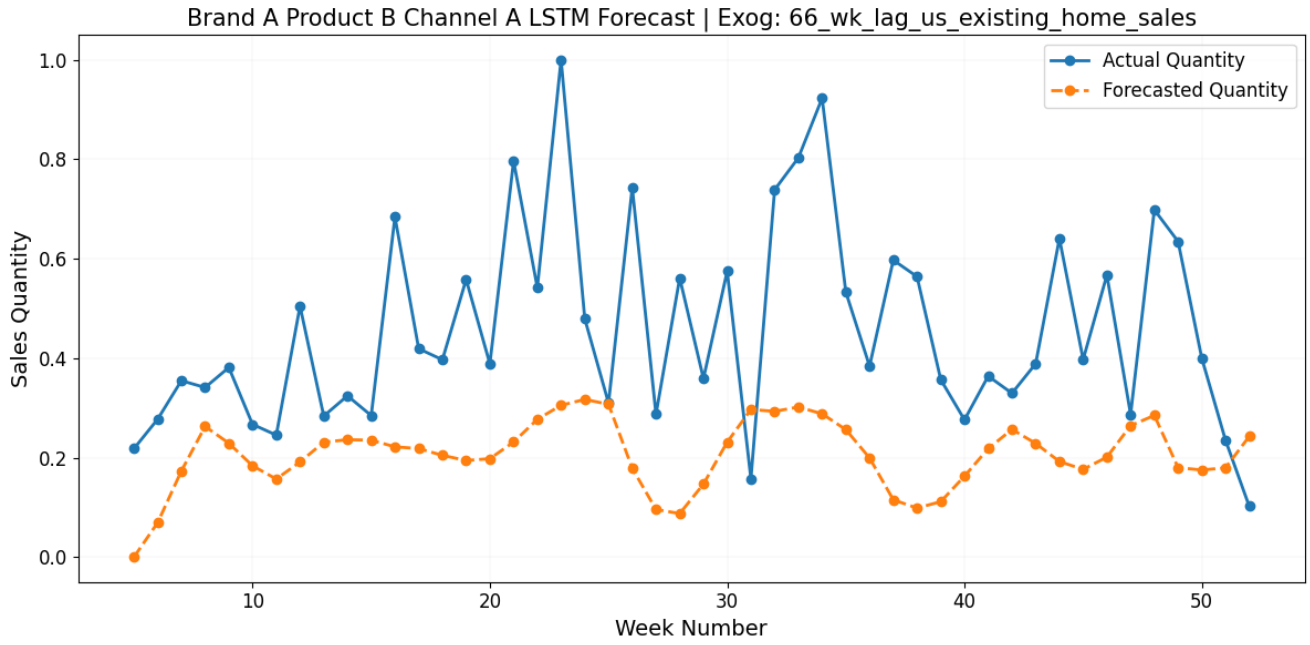


Figure 42: Brand A Product B Channel A (A-B-A) Forecasted vs. Actual Sales | Exogenous Variable: 16–Week Lagged Disposable Personal Income

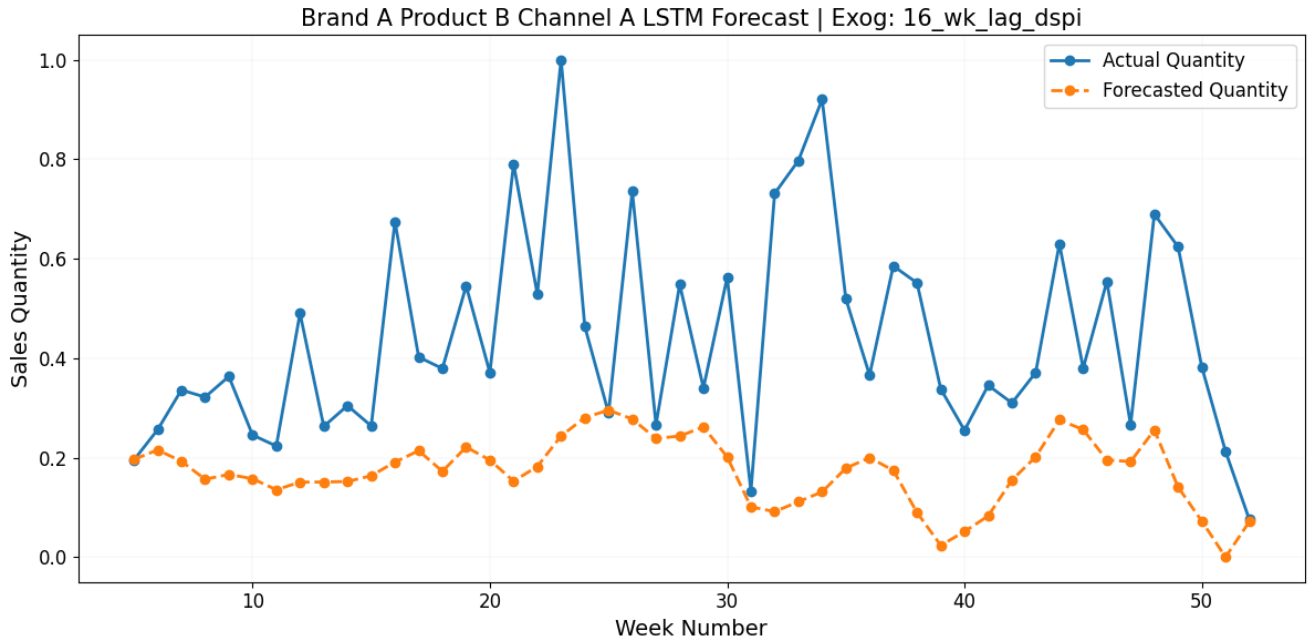


Figure 43: Brand A Product B Channel A (A-B-A) Forecasted vs. Actual Sales | Exogenous Variables: 20–Week Lagged Consumer Sentiment Index

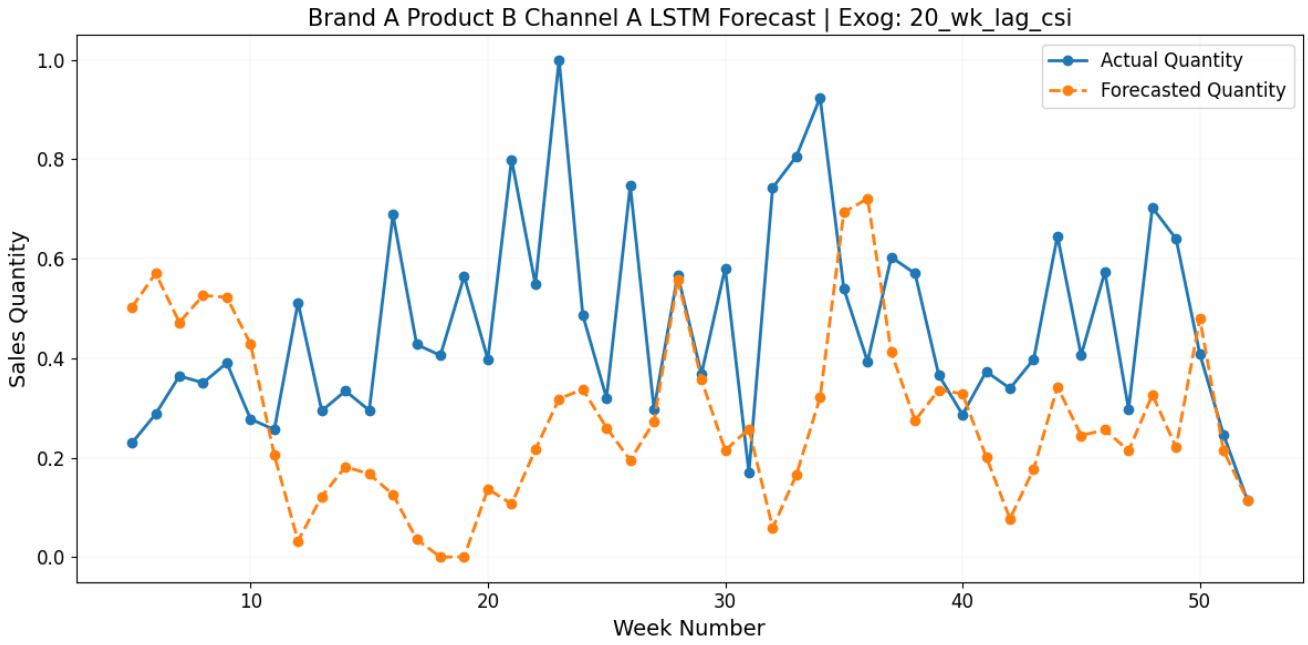


Figure 44: Brand A Product B Channel A (A-B-A) Forecasted vs. Actual Sales | Exogenous Variable: 132–week Lagged Consumer Price Index

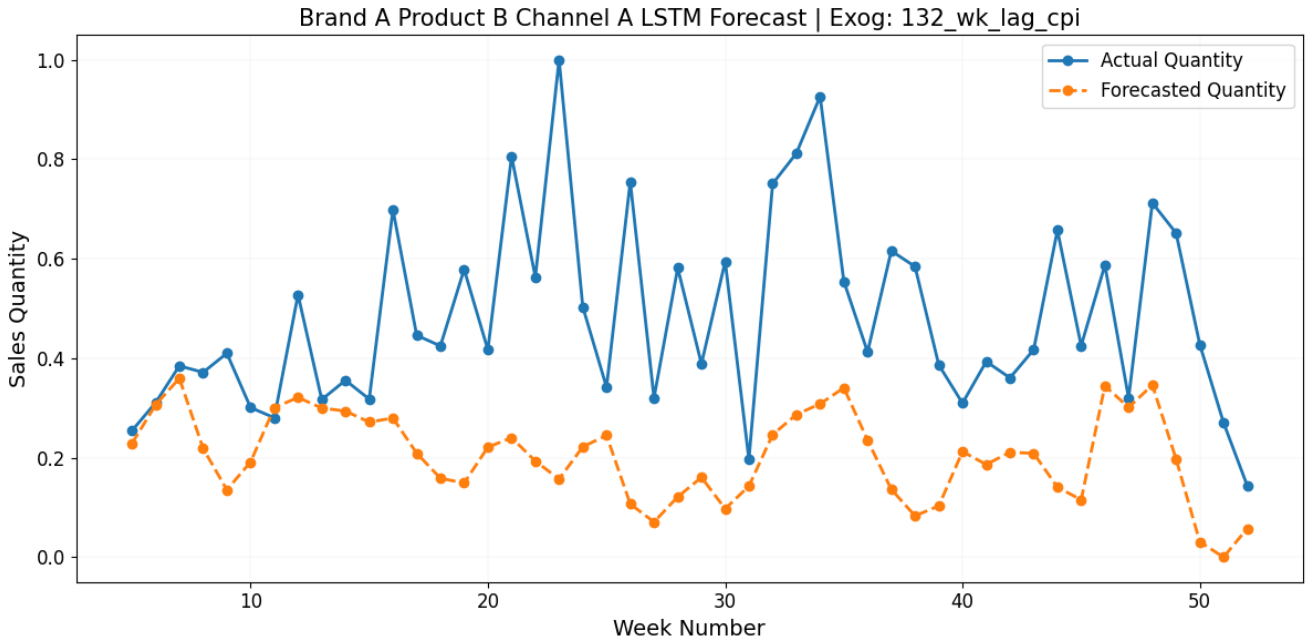


Figure 45: Brand A Product B Channel A (A-B-A) Forecasted vs. Actual Sales | Exogenous Variable: Combined Features

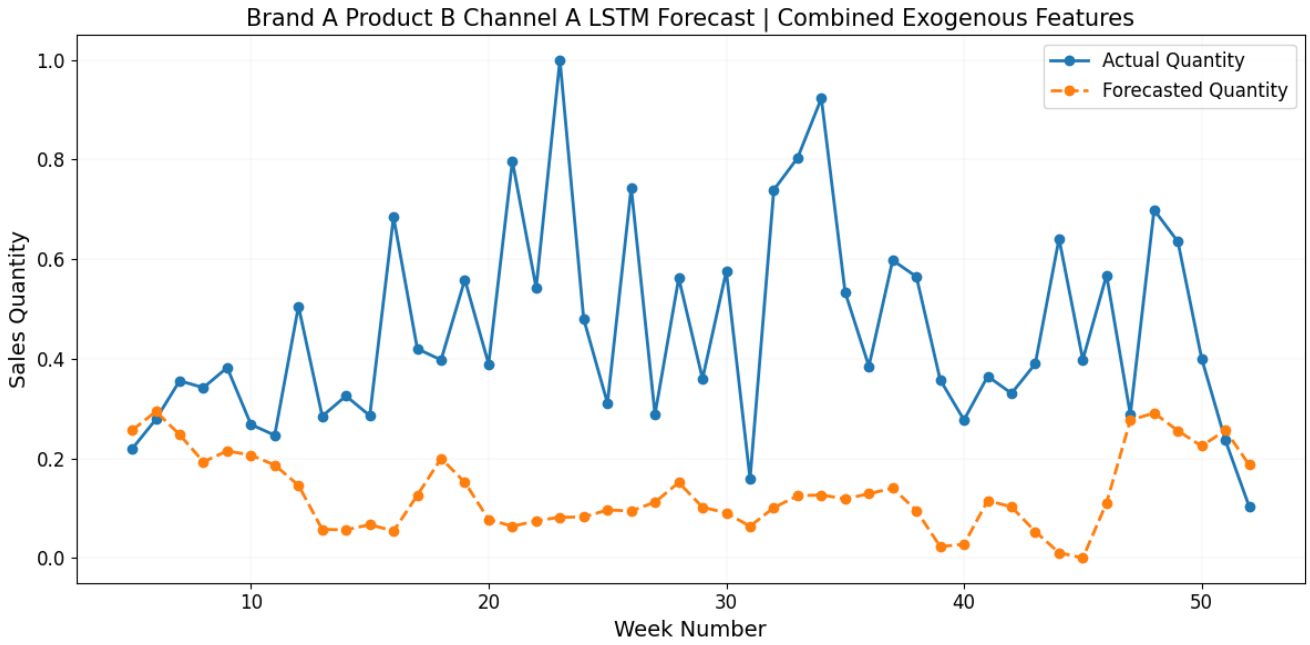
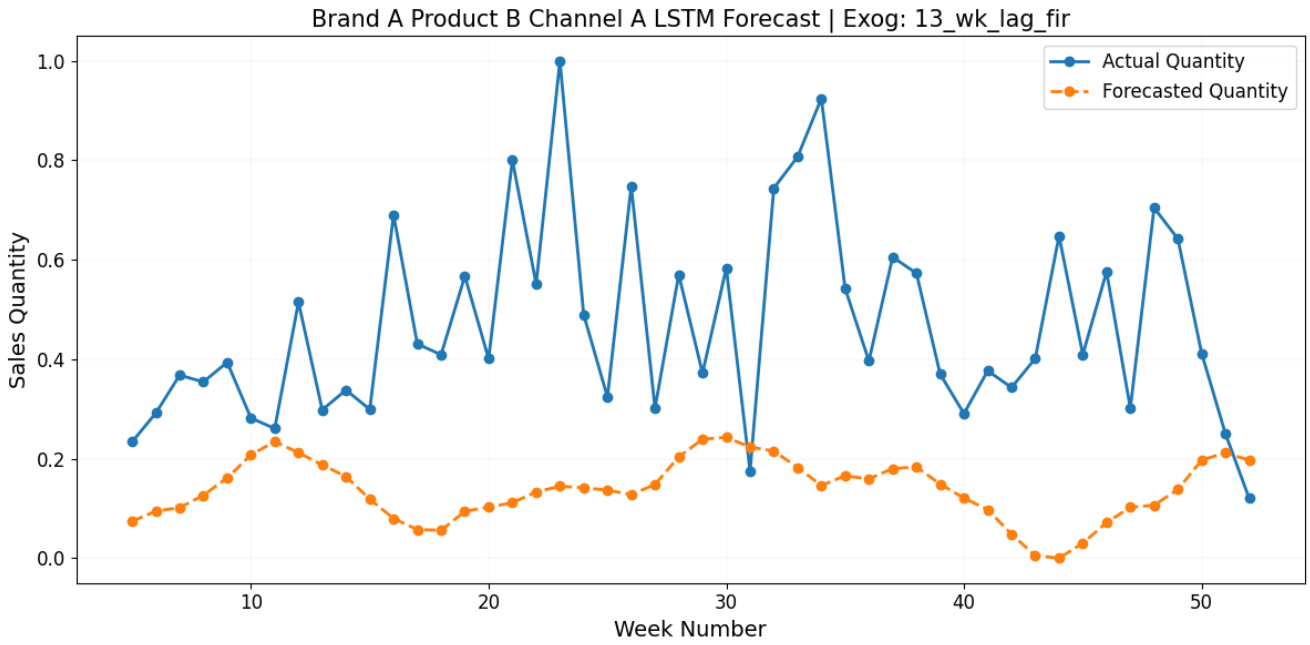


Figure 46: Brand A Product B Channel A (A-B-A) Forecasted vs. Actual Sales | Exogenous Variable: 13–Week Lagged Federal Interest Rate



Brand B Product A Channel A (B-A-A): Normalized Metrics and Forecasted vs. Actual Sales

Figure 47: Brand B Product A Channel A (B-A-A) Normalized Metrics | MAE, MSE, RMSE, MAPE

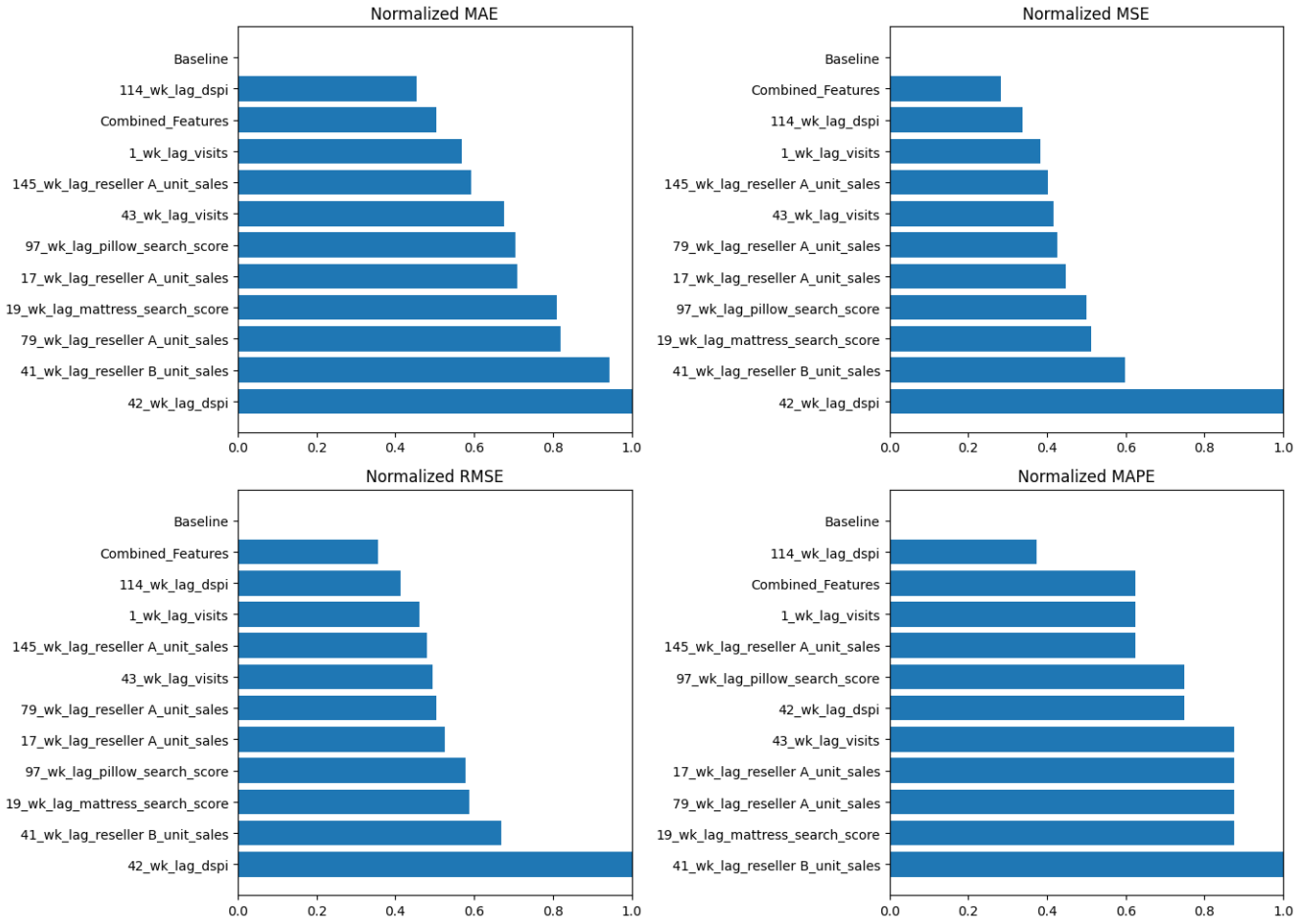


Figure 48: Brand B Product A Channel A (B-A-A) Forecasted vs. Actual Sales | Exogenous Variable: 114–Week Lagged Disposable Personal Income

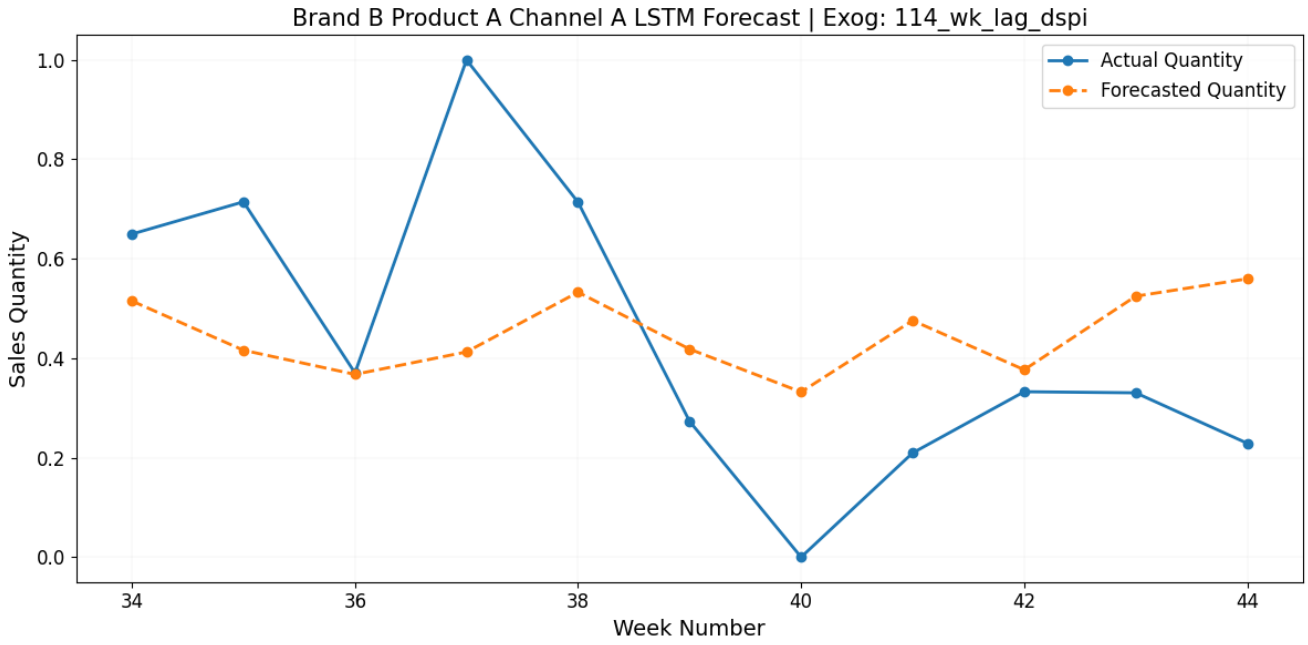


Figure 49: Brand B Product A Channel A (B-A-A) Forecasted vs. Actual Sales | Exogenous Variable: 1–Week Lagged Visits (Web Traffic)

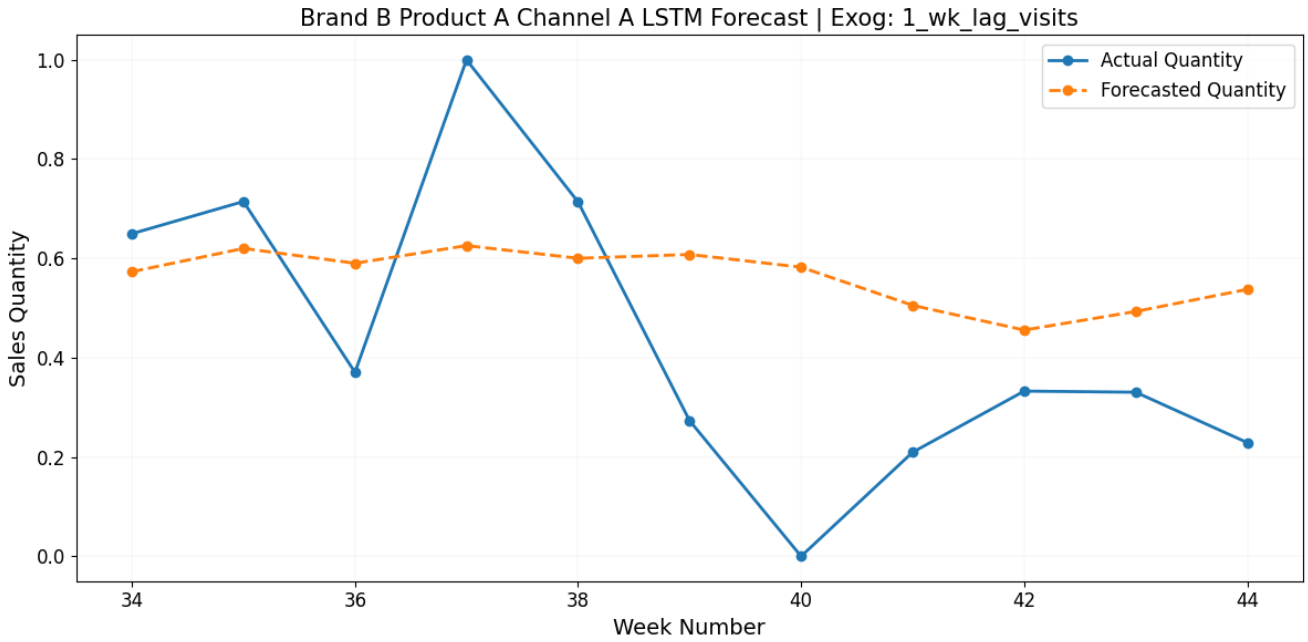


Figure 50: Brand B Product A Channel A (B-A-A) Forecasted vs. Actual Sales | Exogenous Variable: 145–Week Lagged Reseller A Unit Sales

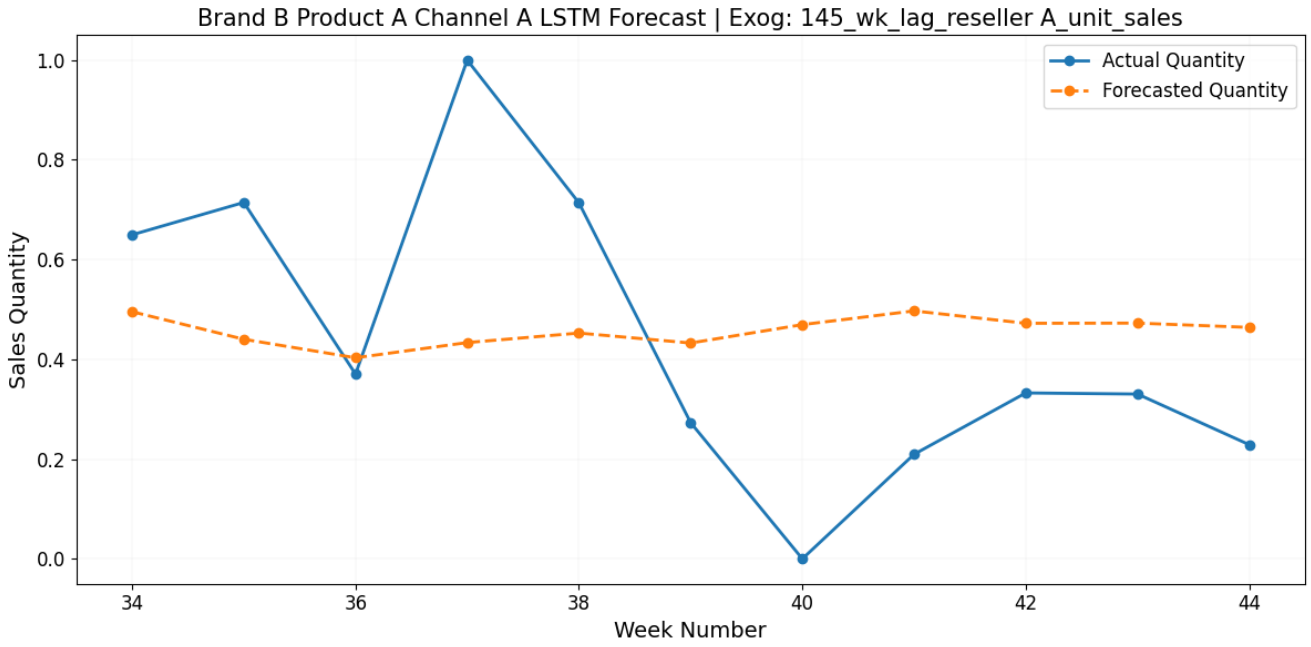


Figure 51: Brand B Product A Channel A (B-A-A) Forecasted vs. Actual Sales | Exogenous Variable: 43–Week Lagged Visits (Web Traffic)

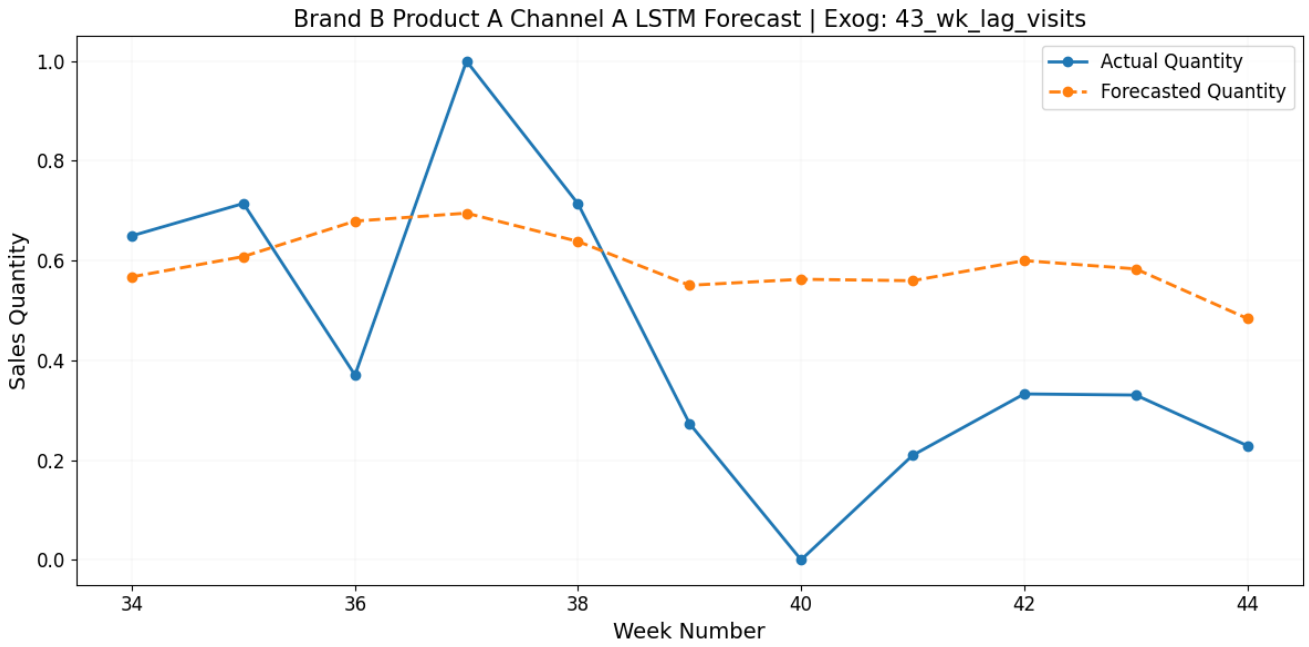


Figure 52: Brand B Product A Channel A (B-A-A) Forecasted vs. Actual Sales | Exogenous Variable: 79–Week Lagged Reseller A Unit Sales

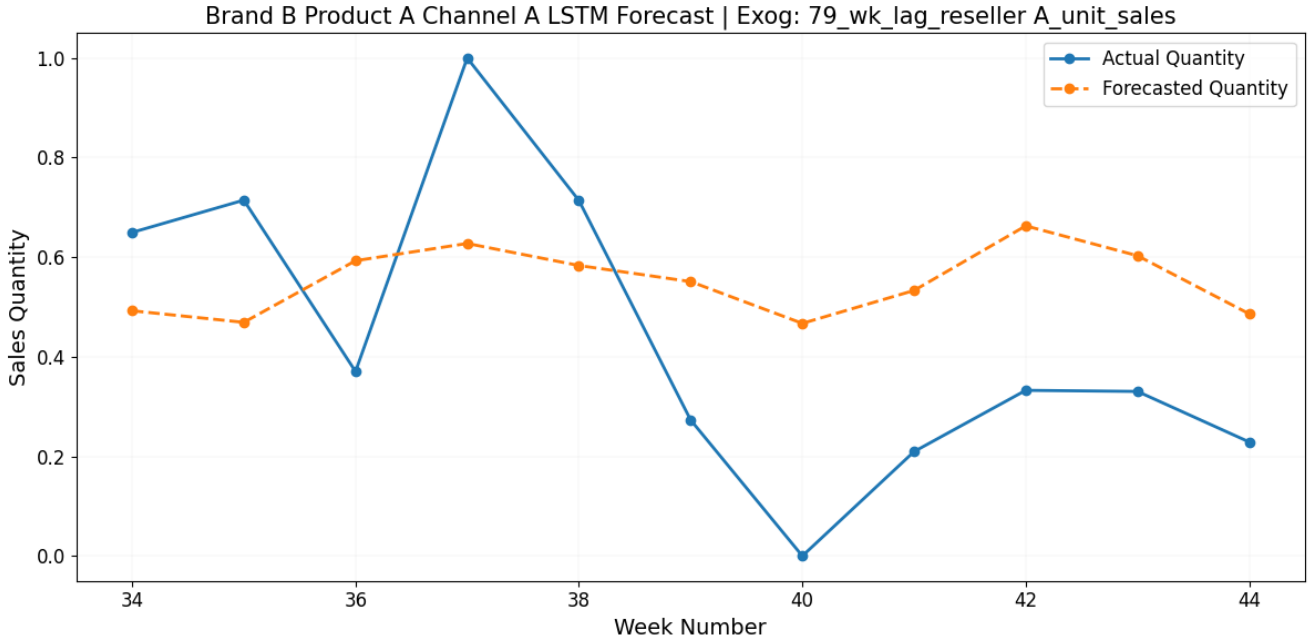


Figure 53: Brand B Product A Channel A (B-A-A) Forecasted vs. Actual Sales | Exogenous Variable: 17–Week Lagged Reseller A Unit Sales

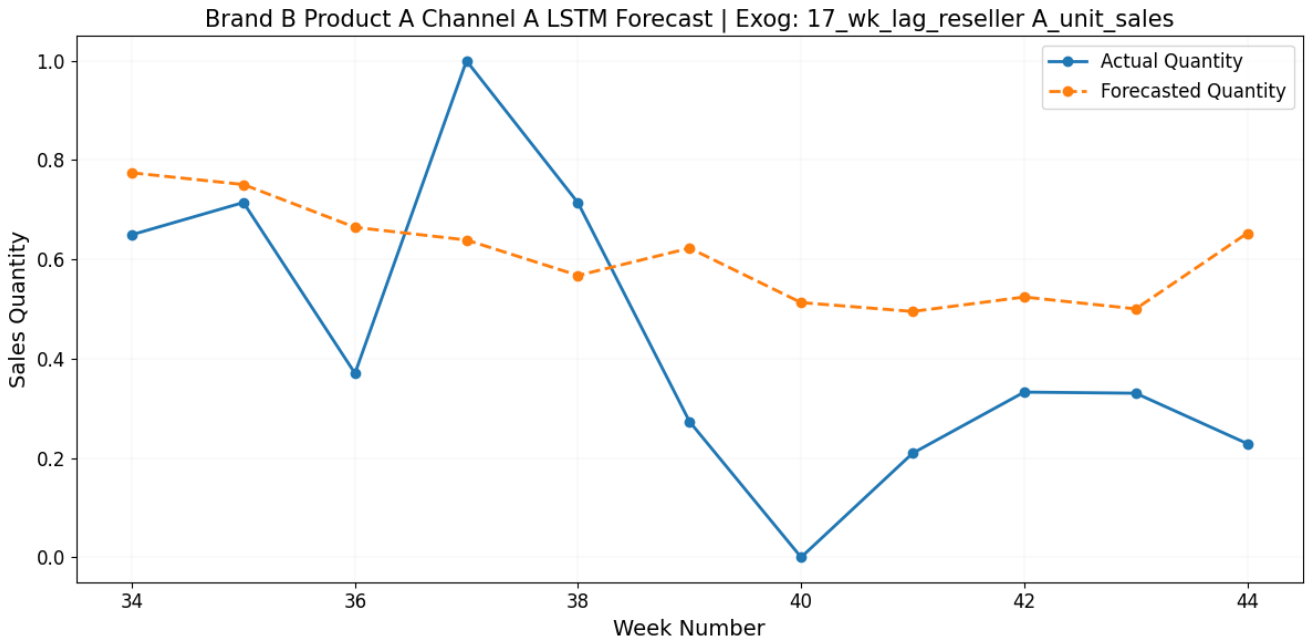


Figure 54: Brand B Product A Channel A (B-A-A) Forecasted vs. Actual Sales | Exogenous Variable: 79–Week Lagged Reseller A Unit Sales

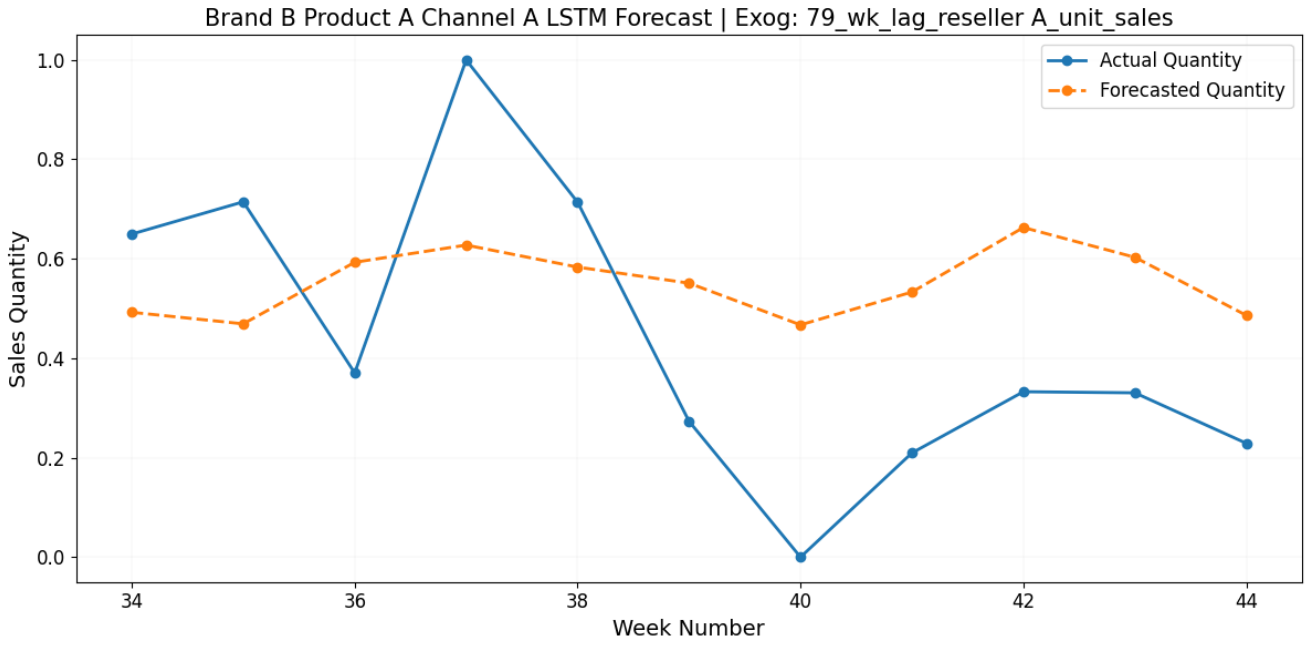


Figure 55: Brand B Product A Channel A (B-A-A) Forecasted vs. Actual Sales | Exogenous Variable: 19–Week Lagged Mattress Search Score

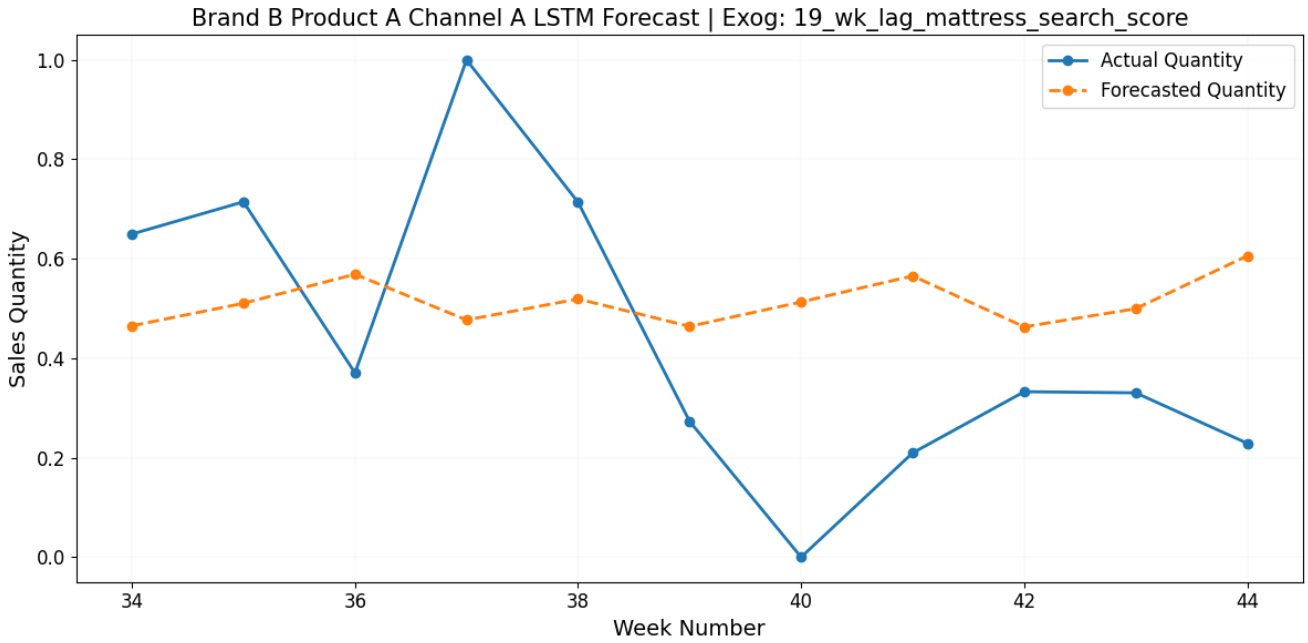


Figure 56: Brand B Product A Channel A (B-A-A) Forecasted vs. Actual Sales | Exogenous Variable: 41–Week Lagged Reseller B Unit Sales

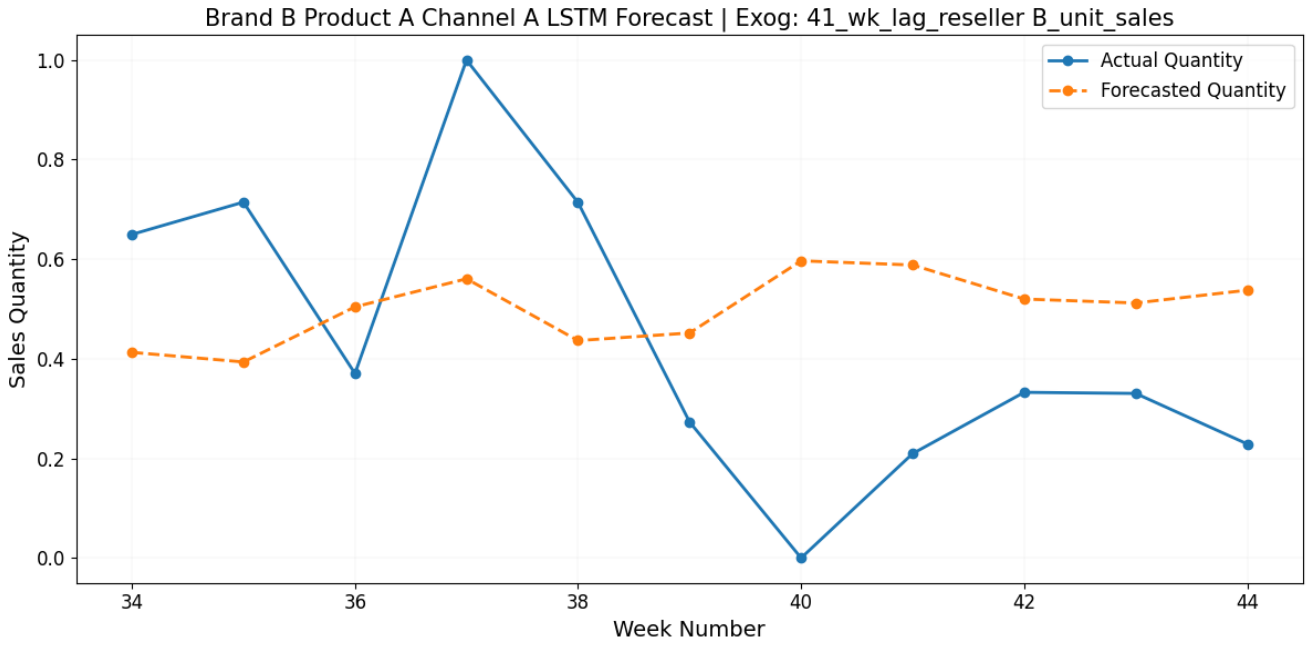


Figure 57: Brand B Product A Channel A (B-A-A) Forecasted vs. Actual Sales | Exogenous Variable: 42–Week Lagged Disposable Personal Income