Using time series analysis for sales and demand forecasting

Alessandro Fasone
01/11/2025

Table of content

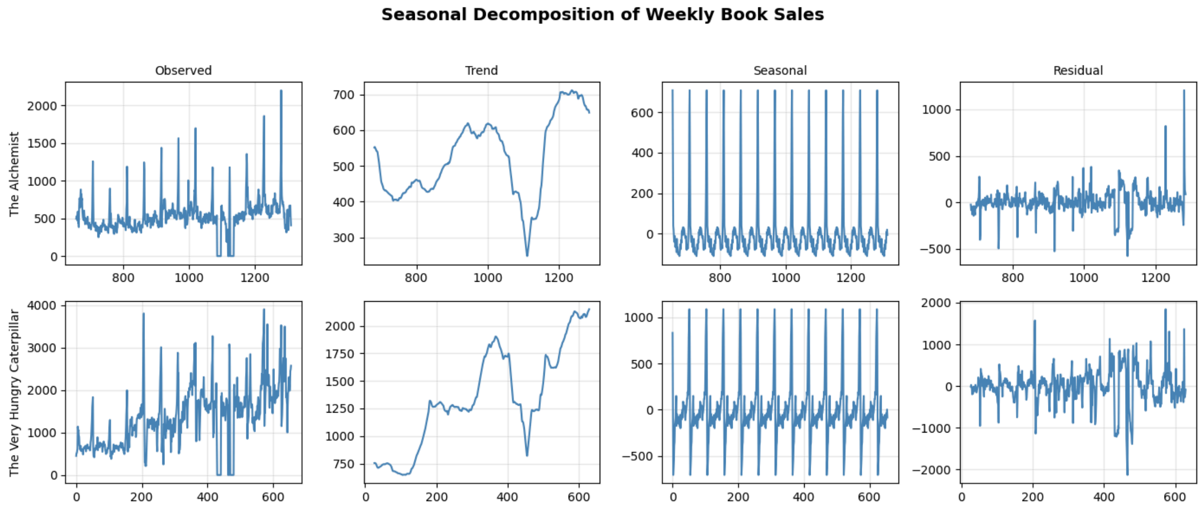
Problem statement

The core challenge for independent publishers is the inability to accurately forecast book demand, leading to inefficient print runs, missed sales, and excess inventory. This project leverages Nielsen BookScan's historical sales data and advanced forecasting models to identify underlying seasonal and trend patterns. The goal is to generate actionable, data-driven insights that guide reprinting, restocking, and investment decisions for long-term title performance.

Methods

The study employed a progressive, multi-model forecasting pipeline designed to balance statistical interpretability with predictive adaptability.
It began with classical time series models—seasonal decomposition, stationarity testing, and Auto-ARIMA—providing a robust understanding of long-term patterns and seasonality.
Building on this foundation, machine learning techniques (XGBoost) were applied to model nonlinear sales drivers using lagged features and grid-searched hyperparameters optimized through time-series cross-validation.
To further enhance temporal learning, LSTM neural networks were trained with KerasTuner, enabling the detection of complex, sequential dependencies in weekly sales data.
Finally, hybrid architectures were implemented: a sequential SARIMA–LSTM model capturing residual nonlinearities, and a parallel weighted combination merging interpretability with precision.
All models were extended to monthly forecasting, allowing short-term and strategic performance comparison using MAE and MAPE as standard accuracy metrics.

Figure 1 Seasonal decomposition of weekly book sales:



Results

Overview

The analysis forecasted sales patterns for The Alchemist and The Very Hungry Caterpillar using classical and machine learning models across weekly and monthly horizons.
Model accuracy and stability were evaluated using Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE), with lower values reflecting higher predictive reliability.

Weekly Models

The Auto ARIMA models served as classical baselines, effectively capturing seasonality but struggling with short-term fluctuations. For The Alchemist, ARIMA(1,0,2)(2,0,[],52) produced MAE = 141.83 and MAPE = 25.18%, while The Very Hungry Caterpillar's ARIMA(1,0,0)(2,0,[1],52) achieved MAE = 550.92 and MAPE = 24.79%. These results confirmed ARIMA's strength in modeling regular cycles but also its limitations in handling sudden demand changes.

The XGBoost models clearly outperformed traditional approaches. Leveraging nonlinear relationships and lag-based features, they achieved the lowest errors overall—MAE = 120.81 and MAPE = 20.46% for The Alchemist, and MAE = 346.19 and MAPE = 17.06% for The Very Hungry Caterpillar. This makes XGBoost the most reliable model for short-term, data-driven decision-making such as stock optimization and print management.

The LSTM models were optimized using KerasTuner, yielding deep architectures capable of capturing complex temporal dependencies in weekly book sales.
For The Alchemist, the final model included five stacked LSTM layers with 221K trainable parameters, incorporating dropout regularization to prevent overfitting.
The Very Hungry Caterpillar required a slightly larger network with over 300K parameters, enabling it to learn longer-term dependencies and smoother seasonal patterns.
This structure effectively captured both short-term fluctuations and broader seasonality, delivering competitive accuracy against classical models.

Hybrid architectures provided a balanced compromise. The Sequential SARIMA + LSTM model improved overall fit (MAE = 141.54, MAPE = 21.87% for The Alchemist; MAE = 362.12, MAPE = 16.09% for The Very Hungry Caterpillar), while the Parallel Hybrid achieved the best stability through weighted averaging.

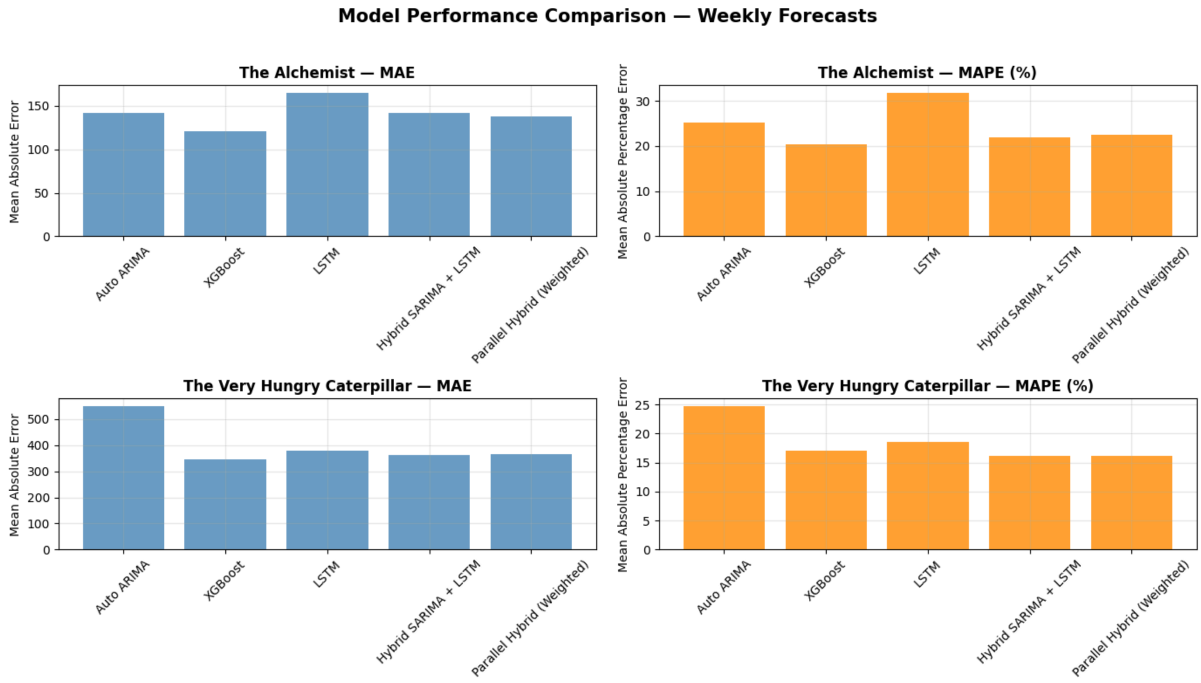
Overall, XGBoost delivered the best weekly forecasts for The Alchemist, while The Very Hungry Caterpillar performed slightly better with the Hybrid Sequential model.

Table 1: Weekly models and results

Book Title	Model	MAE	MAPE (%)
The Alchemist	Auto ARIMA	141.83	25.18

The Alchemist	XGBoost	120.81	20.46
The Alchemist	LSTM	165.14	31.86
The Alchemist	Hybrid SARIMA + LSTM	141.54	21.87
The Alchemist	Parallel Hybrid (Weighted)	137.86	22.51
The Very Hungry Caterpillar	Auto ARIMA	550.92	24.79
The Very Hungry Caterpillar	XGBoost	346.19	17.06
The Very Hungry Caterpillar	LSTM	378.22	18.57
The Very Hungry Caterpillar	Hybrid SARIMA + LSTM	362.12	16.09
The Very Hungry Caterpillar	Parallel Hybrid (Weighted)	364.88	16.17

Figure 2: Weekly performance



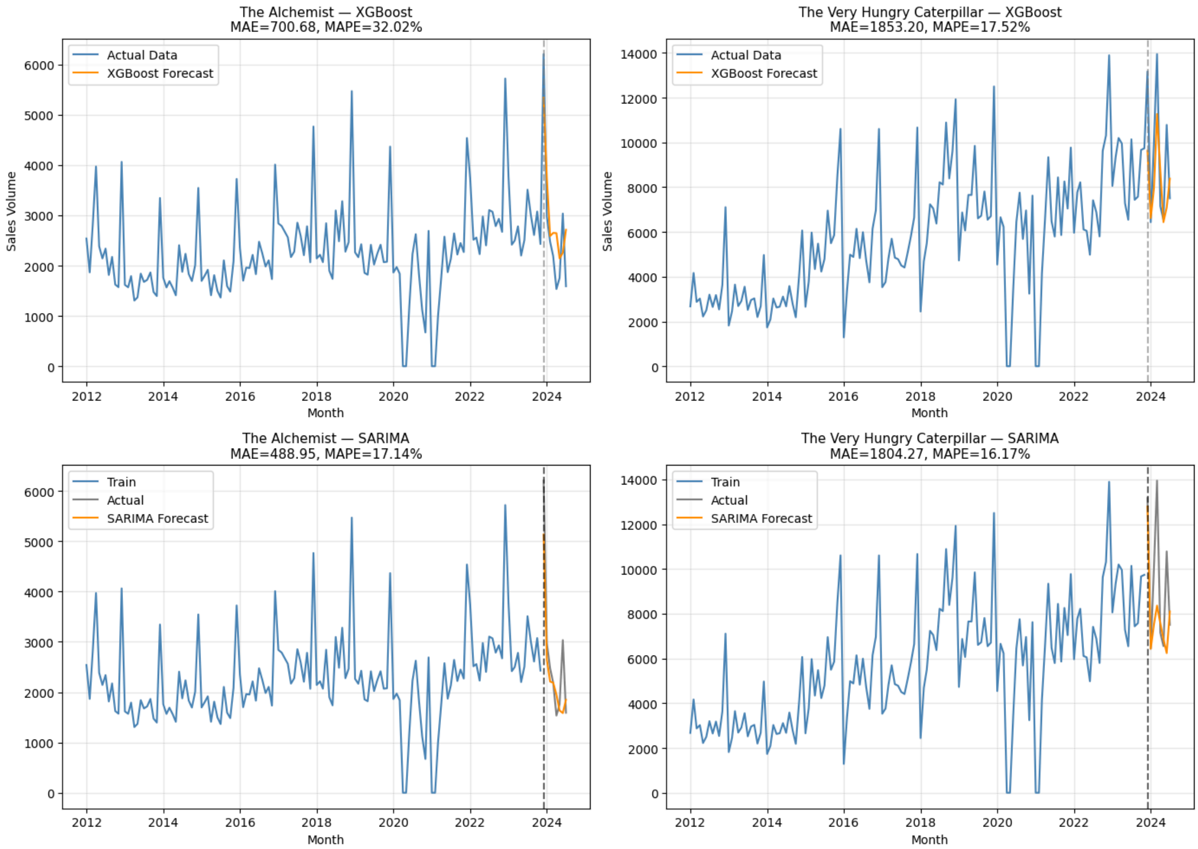
Monthly Models

When aggregating weekly data into monthly observations, the models produced smoother and less volatile patterns. The XGBoost monthly model achieved MAE = 700.7 (MAPE = 32.0%) for The Alchemist and MAE = 1853.2 (MAPE = 17.5%) for The Very Hungry Caterpillar. Despite higher absolute errors due to fewer data points, the forecasts followed the main trend with reduced noise.

The SARIMA monthly models showed superior performance, especially for The Alchemist (MAE = 488.9, MAPE = 17.1%), and remained competitive for The Very Hungry Caterpillar

(MAE = 1804.3, MAPE = 16.2%). This suggests that seasonality and cyclicality play dominant roles in monthly sales dynamics.

Figure 2: Monthly models forecast



Weekly vs Monthly Comparison

Comparing weekly and monthly forecasts for The Alchemist, the weekly XGBoost model effectively captured short-term volatility and promotional effects, while the monthly SARIMA smoothed fluctuations, exposing long-term sales consistency useful for planning print volumes and marketing cycles.

For The Very Hungry Caterpillar, the weekly Hybrid SARIMA–LSTM captured gradual recurring patterns, whereas the monthly SARIMA revealed clearer seasonality and sustained demand trends.

The shift from weekly to monthly aggregation thus enhances signal stability and interpretability, though at the cost of short-term sensitivity.

Conclusion

This analysis shows how advanced forecasting models can support data-driven publishing strategies.

SARIMA offered interpretability and long-term stability, while XGBoost and hybrid SARIMA–LSTM models captured short-term, nonlinear sales dynamics.

For The Alchemist, XGBoost proved most accurate for weekly forecasting, enabling agile stock and pricing management. For The Very Hungry Caterpillar, the hybrid SARIMA–LSTM model best identified recurring demand and seasonality. On a monthly scale, SARIMA remained the strongest performer, outlining broader sales cycles and guiding print and budget planning.

Business recommendation: adopt a dual-level forecasting approach — weekly hybrid models for tactical decisions like restocking and promotions, and monthly SARIMA models for strategic planning and resource allocation.

Together, these models combine short-term responsiveness with long-term foresight, helping Nielsen's clients optimize profitability, reduce waste, and align production with real market demand.