Sales Time Series Forecasting
Team 18

José Pérez Hernández (15750981), Pascal Venema (13697463), Stijn Hering (12377422) and Yop Xue (12064378)
Applied Forecasting in Complex Systems 2024
University of Amsterdam

ABSTRACT

This report aims to provide insightful findings from a supermarket's historical sales data to predict its future product sales over a 28-day period. The data includes both sales prices and quantities for each individual product, providing great versatility for data exploration for sales patterns and trends. During the initial stage of the project, the Exploratory Data Analysis (EDA) indicates strong seasonality and trends in the historical data, creating foundations for our approach to apply time series analysis on the dataset. To find out the best-fit model for the dataset, the performance of each model was evaluated based on its RMSE score. Additionally, due to the complexity that comes with the original dataset, the best model was also influenced by the objective goals, such as prioritizing predicting the total sales for the supermarket over predicting the sales for certain selected products. The report highlights the robustness and limitations of the analysis process and provides useful predictions that can help the management and decision-making processes of the retail industry strategically.

1 INTRODUCTION

Sales forecasting is a key part of running a successful retail business. It helps make sure that stores have just the right amount of stock—enough to meet customer demand without overstocking and wasting resources [1]. This project focuses on forecasting daily sales for the Food3 category at Walmart's TX3 store[1]. We'll use a subset of the M5 Forecasting dataset, which includes sales history, calendar events, pricing, and promotions. This data allows us to try out different forecasting techniques and see how they handle various patterns and external influences.

The goal is to predict sales for the next 28 days, using models that can pick up on trends, seasonality, and the effects of external factors like holidays or price changes. By combining both traditional and advanced forecasting methods, this project aims to create insights that help Walmart manage its inventory more efficiently. This goal leads to the following research questions:

RQ1 How accurately can sales for the Food3 category at Walmart's TX3 store be forecasted over the next 28 days?

- **RQ1.1** Which features (e.g., pricing, calendar events, promotions) have the most significant impact on forecasting accuracy?
- **RQ1.2** How can traditional methods like Simple Exponential Smoothing (SES) enhance results of methods like Prophet in terms of RMSE and MAE for products with few datapoints?
- **RQ1.3** How do external factors like holidays and price changes affect the RMSE and MAE of the forecasting models?

Motivation

Retail stores face a big challenge when it comes to balancing stock levels [4]. Too much stock means extra costs for storage or wasted items, while too little stock leads to empty shelves and unhappy customers. For Walmart, with its massive product range and a highly competitive market, getting this balance right is crucial.

Using additional information like prices, promotions, and holidays can make forecasts more accurate. For instance, a sale during a holiday weekend will likely drive higher sales than usual. Factoring in these variables lets us create smarter predictions that go beyond just looking at past sales trends. Ultimately, this project aims to build reliable forecasts that help Walmart keep shelves stocked and customers happy, all while avoiding unnecessary costs.

Problem Definition

The main goal of this project is to forecast daily sales for the next 28 days for the Food3 product category at Walmart's TX3 store. We'll use historical sales data along with other helpful variables like prices, promotions, and calendar events. Specifically, we want to:

(1) Understand the patterns in the sales data, such as trends over time, seasonal spikes, and any irregularities.
(2) Include extra information like prices and events to make the forecasts more accurate.

[1]https://corporate.walmart.com/about

(3) Compare different forecasting models and pick the one that gives the best results, using RMSE as our main measure of success.

Some challenges we expect to face include:

- Finding the best way to use additional variables like prices and promotions in our models.
- Making sure the model doesn't just fit the training data but also works well for new, unseen data.

By tackling these challenges, this project aims to build a forecasting model that can make reliable predictions while being practical enough to use in real-world retail planning.

2 DATASET DESCRIPTION

To generate 28-point sales forecasts for products, we will focus on three primary datasets. All datasets refer to the TX3 Walmart store, specifically to "FOODS_3" category products.

(1) "sales_train_validation_afcs2024.csv": Provides information about the amount daily sales of all products, starting 2011-01-29 and following 1913 days (2016-04-24). Given in a dataframe of 823 rows (products) and 1914 columns (days).
(2) "sell_prices_afcs2024.csv": Provides information about the weekly sell price of all products, covering same dates as dataset 1. Given in a dataframe of 185260 rows (product, per week) and 4 columns (product id, week id and sell price).
(3) "calendar_afcs2024.csv": Provides information about daily events (sporting, cultural, religious...) or SNAP purchases (government aid) occurring in same day range as dataset 1. Given in a dataframe of 1969 rows (days of the week) and 11 columns (characteristics of the event, if any)

3 EXPLORATORY DATA ANALYSIS

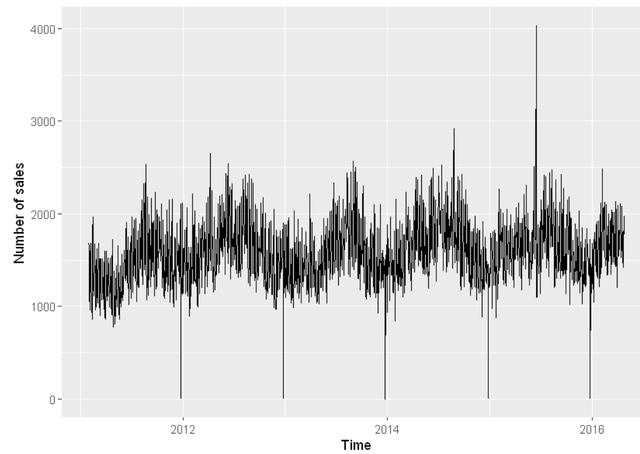
The Exploratory Data Analysis (EDA) will primarily involve examining the three datasets—sales, calendar, and pricing—to uncover underlying patterns and gain a deeper understanding of the data. This step is crucial as it provides insights into trends, seasonality, and irregularities that may influence sales, such as the impact of events, promotions, or pricing changes.

By visualising the data and analysing relationships between variables, EDA helps identify key drivers of sales and informs the selection of appropriate forecasting models [3]. Moreover, it enables the detection of potential anomalies or inconsistencies, ensuring cleaner and more reliable data for subsequent modelling efforts. Through EDA, we establish a solid foundation for building forecasts that accurately reflect the dynamics of the provided data.

Sales Variation

Regarding the number of sales, we can observe the daily total sales of all products combined throughout the whole time series in Figure 1. This shows a clear seasonal pattern, with sales being the highest in the summer months.

Figure 1: Daily Total Sales



We can now proceed to apply an STL decomposition in Figure 2 , which will allow us to visualize independently its major components regarding trend and seasonality. In this case, we observe there is both a considerable positive trend on the sales, as well as a clear seasonal pattern. Some in-week seasonality is also observed, but it is hard to interpret.

Figure 2: STL Decomposition of Total Sales

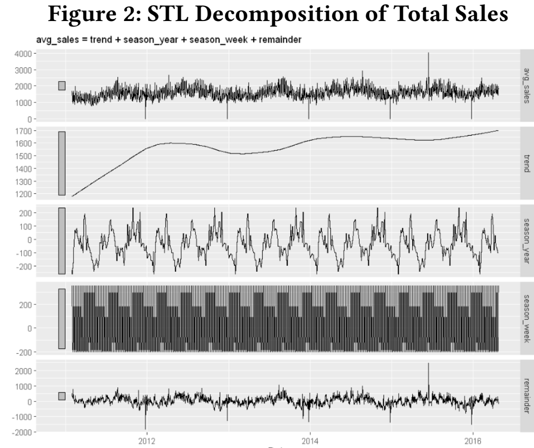
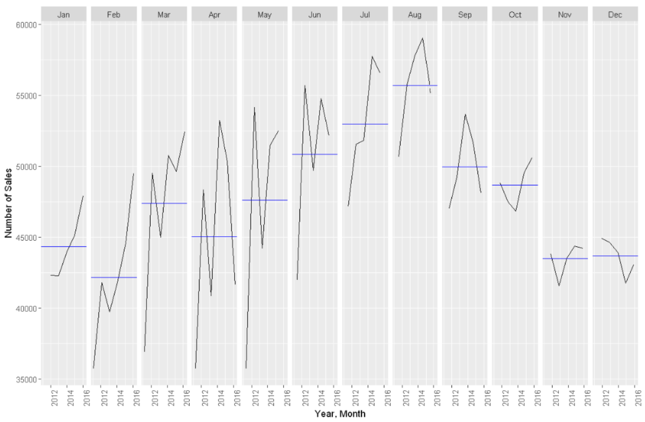


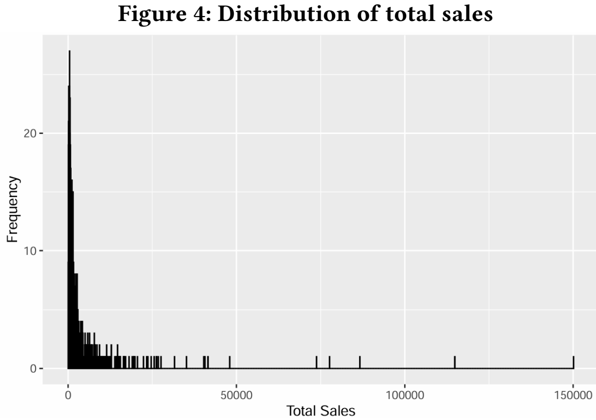
Figure 3 makes more clear the seasonal patterns observed. Monthly sales throughout all available years allows us to confirm how Summer is the season with the highest sales and winter with the least.

Figure 3: Yearly Sales per Month



One other interesting finding regarding sales comes when we observe the amount of units sold per product throughout the entire time series, in Figure 4. Even though observing a large variety of products, the distribution of sales stands largely deviated to the left, Specifically, mean sales are 3642, with maximum sales being 150.122 (product id 586), and lowest sales being 35 (product id 171).

Figure 4: Distribution of total sales



To better understand how product sales are distributed, we will select the top 10, middle 5 and bottom 5 products in terms of sales selling products and visualize their sales distributions. This allows us to get a visual insight on what pattern sales follow in these significant product categories.

Figure 10 (see appendix) shows the probability distribution of the amount of units sold per each of the selected products: first two rows indicate top 10 selling products, third row represents 5 middle and last row displays 5 bottom selling products.

Product with id "FOODS_3_586", which was shown to be the most sold of the whole dataset is observed to have one of the highest median values (around 75), but also one of the widest curves, with a considerable proportion between values of 50 and 100, which explains it being the most sold product. Most of these top selling products present a similar distribution, with a small proportion of 0 sales and a relatively normal distribution.

Products classified as middle selling have a much higher proportion of days with 0 sales, and none of them significantly surpasses 3 unit sales.

Finally, lower selling products present almost their whole distribution in the 0 sales mark, with small peaks in 1 unit sold.

Finally, in Figure 11 we can also observe the amount of sales throughout the whole time series, this time independently for each of the selected products.

Each category has a pattern which is common for almost all products classified in the same category (top, middle and bottom). Except product 090, 587 and 030 all of the top selling products have a relatively constant amount of sales (ignoring short term variance) which means top selling products tend to be sold during all seasons and throughout all years. Product 030 seems especially remarkable, since it has on average lowest sales than comparable products, but suffered an outburst of sales at the end of 2013 which made it position in this top 10.

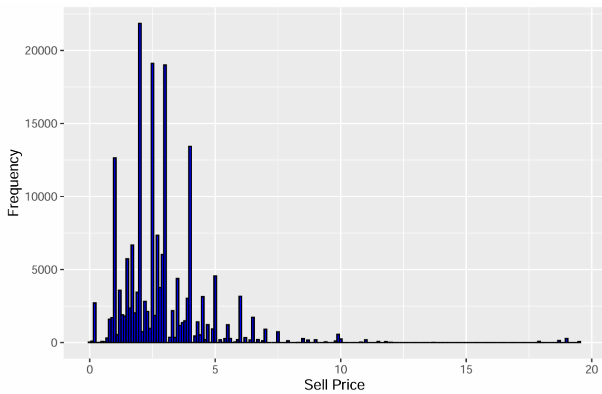
On the other hand, biggest difference between top selling products and bottom and middle products is that the two latest either have not been available during the whole time series (product 472, for instance, had zero sales until 2015, or have either suffered multiple periods with sales close to zero.

Price of Items

Regarding the price of items sold, we will also begin by trying to get an idea of how prices are distributed along products. Dataset includes information about weekly prices per product. To simplify the information, we will compute average, maximum and minimum price, which resulted in values \$2.84, \$19.84, and \$0.02, respectively.

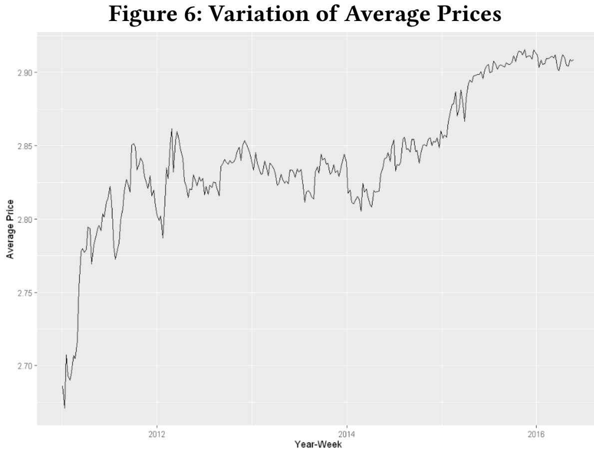
Figure 5 shows distribution of selling prices of all products. It is visible how most products are placed within the \$0 and \$5 price range.

Figure 5: Selling Price Distribution



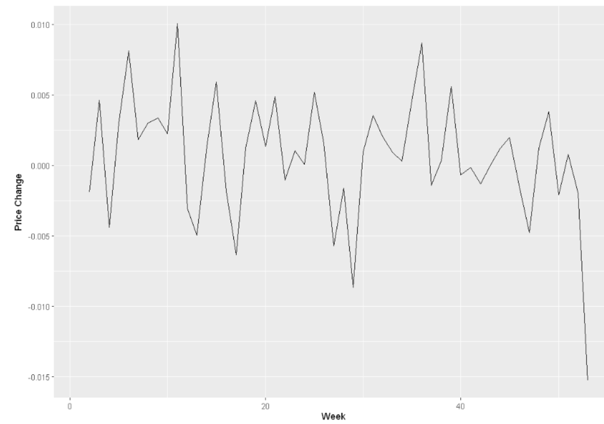
Another revealing visualisation is the variation of the price in Figure 6, in general, products have suffered throughout the whole time series. In this case, there is a steep increase in average price at the start of 2011 that can be observed. Prices stabilised relatively until the year 2014, after which a positive trend began again until the beginning of the year 2016. This trend shows how prices slowly increase over time, most likely due to inflation.

Figure 6: Variation of Average Prices



In terms of price change within weeks (for instance, in advance to holidays, promotions, etc.) it can also be insightful to visualize these patterns to observe if there is any specific week in which prices tend to increase or decrease more than the average. As seen earlier, prices generally have an upward trend, which makes inter-week changes (see Figure 7) generally positive and makes it hard to visualise clear times of the year in which prices vary relatively consistently.

Figure 7: Average Price Change per Week



It is also visible in the table (1) below how top 20 selling products all have an average price below $10 (in factt all of them except for product 691 have an average price below $5.

Furthermore, in a subset of the price-sales correlation table (2), we observe that high prices are negatively correlated with the number of sales.

Table 1: Price Statistics for Top 20 Products

id_trimmed	mean_price	sd_price	min_price	max_price
FOODS_3_691	7.480	0.000	7.480	7.480
FOODS_3_202	4.202	0.328	2.400	4.480
FOODS_3_693	3.480	0.000	3.480	3.480
FOODS_3_737	3.480	0.000	3.480	3.480
FOODS_3_220	3.196	0.306	1.000	3.280
FOODS_3_633	2.880	0.000	2.880	2.880
FOODS_3_088	2.780	0.000	2.780	2.780
FOODS_3_587	2.519	0.246	1.870	2.680
FOODS_3_171	2.465	0.110	1.680	2.480
FOODS_3_607	2.243	0.260	0.980	2.480
FOODS_3_512	1.980	0.000	1.980	1.980
FOODS_3_260	1.970	0.000	1.970	1.970
FOODS_3_472	1.970	0.000	1.970	1.970
FOODS_3_804	1.880	0.000	1.880	1.880
FOODS_3_811	1.870	0.031	1.780	1.880
FOODS_3_694	1.604	0.076	1.480	1.680
FOODS_3_586	1.593	0.072	1.480	1.680
FOODS_3_080	1.593	0.072	1.480	1.680
FOODS_3_555	1.593	0.072	1.480	1.680
FOODS_3_377	1.584	0.068	1.380	1.680

Calendar Events

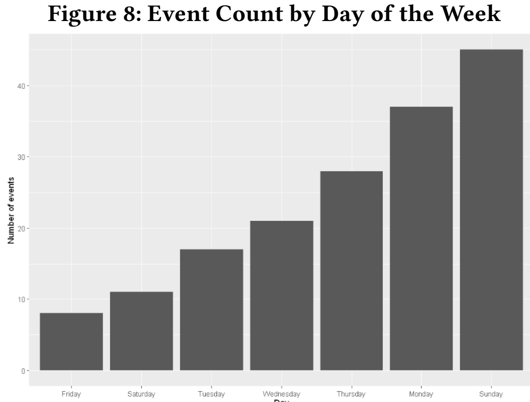
Special events such as celebrations or sports events are also factors that must be explored to better understand our data available for building future podcasts. To start with the analysis of this, we can get a simple bar chart (Figure 8 that allows

Table 2: Correlation Between Sales and Prices

id_trimmed	correlation
FOODS_3_464	0.16
FOODS_3_377	0.11
FOODS_3_760	0.11
FOODS_3_202	0.05
FOODS_3_714	0.05
FOODS_3_252	0.00
FOODS_3_171	-0.01
FOODS_3_725	-0.05
FOODS_3_721	-0.11
FOODS_3_090	-0.12
FOODS_3_694	-0.18
FOODS_3_586	-0.32
FOODS_3_587	-0.38
FOODS_3_030	-0.47
FOODS_3_555	-0.52

us to see how many events occur in the whole time series per day of the week.

Figure 8: Event Count by Day of the Week



This chart reveals how Sunday is the day of the week that has the highest number of events, followed by Monday and Thursday. On the other hand, Friday and Saturday have the least amount of events occurring. Knowing this distribution of events presents us useful information that can help us study if these days with higher amount of events have any impact with food product sales.

Since the dataset provides us with information about different types of events, it is insightful to observe all of the existing types of events and which of these types is the most common to occur.

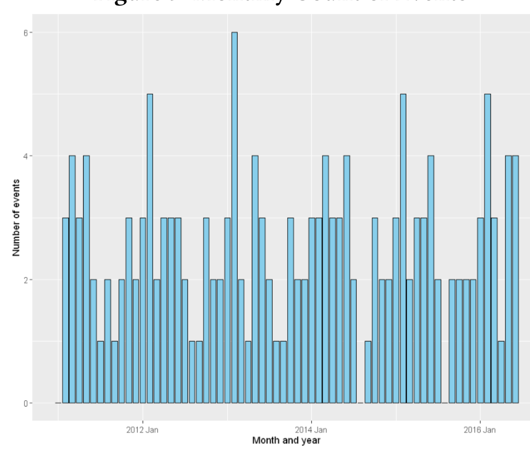
Table 3: Types of Events Count

Event Type	#Events
Religious	56
National	52
Cultural	41
Sporting	18

Religious is the most common event type (56 occurrences), whereas Sporting is the least common, with only 18 occurrences. National events are the second most common type of event with 52 occurrences, and cultural stands third with 41.

To get a clearer idea of how events are distributed along the time series, we can specify a bar graph that explains how many events of any type have taken place in each of the months for which we have data.

Figure 9: Monthly Count of Events



This shows how the end and the start of the year tend to be the periods with a higher concentration of events, whereas the summer season has on average the least events.

Product Statistics

When analyzing a subset of individual products categorized as top-10, middle-5, and bottom-5 based on summarized statistics table (9) and zero-sale statistics table (10), it becomes evident that most products do not sell well. Lower- to mid-performing products exhibit a high number of zero-sale days. For these middle- and bottom-tier products, the total mean sales are often close to zero, with the median frequently being zero as well. Additionally, for the middle-performing products, the standard deviation typically hovers around one. Overall, the data suggests a small selection of products sell consistently and perform exceptionally well, while the majority experience sporadic and random sales patterns.

When performing a seasonal decomposition (Figure 12) of the top-10 products, we observe subtle trends and, in some cases, yearly seasonality. However, the majority of the variance is attributed to weekly seasonality, while the remainder primarily consists of outliers. This insight provides valuable information that can guide and improve forecasting efforts.

4 METHODOLOGY AND IMPLEMENTATION

Data Preprocessing

Missing Data. To ensure the dataset is complete and free from missing values, the daily sales columns (`d_1`, `d_2`, ..., `d_1913`) were checked for missing data using Python's `isna()` method. The process involved calculating the total number of missing values for each column and summarising the results. Given the absence of missing data, no further action was required for this step, and the preprocessing pipeline proceeded as planned.

Data Transformation and Integration. To prepare the dataset for analysis and modelling, a transformation and integration process was undertaken. The primary objective was to merge and restructure data from multiple sources—sales data, calendar data, and price data—into a unified, analysis-ready format. This process included transforming the data structure, adding time-related features, and integrating supplementary information.

(1) **Restructuring Sales Data**
The sales dataset was transformed from a wide format, where daily sales were represented as separate columns (`d_1`, `d_2`, ..., `d_1913`), into a long format. Each row represented a specific product's sales on a given date, simplifying subsequent analysis and integration steps.
(2) **Adding Date Information**
The transformation used the starting date of the sales data, 29 January 2011, to compute exact dates for each `d_` column. This ensured proper alignment with the calendar data, enabling the addition of time-based features such as weekdays, months, and events. The starting date was verified against project documentation for consistency.
(3) **Integration with Calendar and Pricing Data**
The cleaned sales data was enriched with contextual variables from the calendar dataset, including holidays, event names, and SNAP eligibility. Similarly, pricing data was integrated by aligning product IDs and weekly identifiers, enabling the inclusion of explanatory variables such as price changes and promotions. While `item_price` and `SNAP` were unavailable in the validation and test datasets, event names were extrapolated into these datasets to allow the models to capture the potential impact of events on sales.
(4) **Validation Set Preparation**
Consistent transformations were applied to the validation dataset, ensuring alignment with the training data. The starting date for the validation set matched the most recent date in the training data to maintain temporal continuity.

By merging and restructuring these datasets, the final dataset served as a robust foundation for analysis and modeling, incorporating temporal and contextual factors effectively.

Model Selection and Optimisation

Feature Engineering. Features were engineered to capture sales patterns at the product level. Key transformations included the extraction of trends, seasonality, and residuals using Seasonal-Trend Decomposition with LOESS (STL). Time-based features such as holidays, weekdays, and monthly seasonality were also included to account for external influences on sales.

Model Selection. Following data preprocessing and feature engineering, model selection was guided by the patterns observed during exploratory data analysis (EDA). A threshold-based classification framework was implemented to select appropriate models based on trends, seasonality, autocorrelation, and average weekly sales levels.

(1) **Threshold-Based Classification Using Percentiles**
Products were classified into two categories based on their average weekly sales:
- **Simple Patterns**: Products with average weekly sales below the 40th percentile were classified as "simple" and modeled using **Simple Exponential Smoothing (SES)**.
- **Complex Patterns**: Products with average weekly sales above the 40th percentile, strong seasonality, significant trends, or autocorrelation in residuals were classified as "complex" and modeled using **Prophet**.

The average weekly sales for each product were calculated by dividing total sales by the number of weeks in the training dataset. The 40th percentile threshold was dynamically computed to adapt to the sales distribution of the dataset, ensuring a flexible and data-driven classification.
(2) **Refinement Through Statistical Tests**
The classification process was further refined using the following statistical methods:

- **Augmented Dickey-Fuller (ADF) Test**: Applied to the trend component to assess stationarity. Products with p-values greater than 0.05 were considered to have significant trends.
- **Seasonality Variance**: Evaluated by examining the variance of the seasonal component. Products with significant seasonal variance (greater than 0.1) were flagged as having strong seasonality.
- **Ljung-Box Test**: Applied to the residuals to detect autocorrelation. Products with p-values below 0.05 were classified as having significant autocorrelation.

Products exhibiting significant trends, strong seasonality, or autocorrelation were categorized as "complex."

(3) **Simple Exponential Smoothing (SES)**
SES was chosen for products with simpler patterns. It effectively captured short-term trends while maintaining computational efficiency. Smoothing levels were optimized for each product to enhance forecast accuracy. Predictions were rounded to integers to ensure realistic sales values, and model performance was evaluated using Root Mean Squared Error (RMSE).

(4) **Prophet**
Prophet was selected for products with complex patterns, leveraging its ability to capture trends, seasonality, and external factors such as holidays. Event data, US holidays, and monthly seasonality components were integrated to improve accuracy. Cross-validation was conducted over a rolling horizon to validate performance and fine-tune parameters, ensuring robustness.

This hybrid classification approach, combining percentile thresholds with statistical tests, ensured that computationally efficient models were used for simpler patterns, while advanced methods addressed products with complex sales dynamics.

Model Training. The training pipeline followed a structured approach:

- **SES**: Each product's sales data was processed independently, with smoothing parameters optimized to ensure accurate forecasts.
- **Prophet**: Models were trained using data enriched with event information and custom seasonal components. Parameters such as seasonality strength, trend flexibility, and changepoints were tuned automatically.

Trained models and their configurations were saved for reproducibility. Forecast plots and decomposition components were stored for further analysis.

Validation. The validation framework ensured a comprehensive evaluation of model performance:

- **Metrics**: RMSE and MAE were used as primary evaluation metrics to measure forecast accuracy on the test dataset.
- **Cross-Validation**: Prophet's performance was validated using rolling-horizon cross-validation to ensure robust generalization to unseen data.
- **Model Comparison**: Comparative analysis of SES and Prophet results identified the best-performing model for each product category.

This validation framework ensured that the models effectively captured the unique characteristics of each product's sales patterns while maintaining a balance between computational efficiency and accuracy.

5 RESULTS

The performance of the models was summarized and compared using RMSE scores and sales metrics. Table 6 provides detailed metrics for the top 20 sold products, highlighting the models' forecasting capabilities for high-performing items. Prophet demonstrated superior performance for products with complex sales patterns, capturing seasonality and trends more effectively. SES, on the other hand, provided competitive forecasts for products with simpler and more stable sales patterns.

The comparison of aggregated metrics between SES and Prophet models is presented in Table 5. Prophet consistently achieved higher Test Weekly Average Sales and Test Total Sales, indicative of its ability to better align forecasts with actual sales for the top-performing products. However, SES achieved significantly lower RMSE scores on average, suggesting that for certain low-variance products, simpler models may suffice.

Table 4 summarizes the statistical characteristics of the forecast metrics. The higher variability in Prophet's RMSE values reflects its adaptation to more complex patterns, whereas SES displays more consistent, albeit less accurate, predictions.

Table 4: Summary Statistics of Metrics

Statistic	MAE	RMSE	Weekly Sales	Total Sales
Mean	1.605	2.133	2.29	65.73
Median	0.93	1.20	1.00	22.00
SD	2.39	2.96	5.45	152.13
Min	0.04	0.19	0.00	1.00
Max	25.61	29.35	73.00	2056.00

Table 5: Comparison of Test Data Metrics Between SES and Prophet Models

Metric (Test Data)	SES Model	Prophet Model
Average RMSE	0.78	3.03
Average Weekly Sales	0.32	3.61
Total Sales	4,073	50,021

Furthermore, Tables 7 and 8 present the top 10 and bottom 10 products ranked by RMSE for each model, providing a detailed comparison of their performance across different scenarios.

6 DISCUSSION

Insights and Observations

The findings of this study highlight the importance of aligning model selection with product-specific sales characteristics. Models like Prophet, which incorporate seasonality and trend detection, demonstrated robust performance for products with identifiable seasonal patterns or consistent variability. However, for products with erratic or sporadic sales patterns, Prophet's complexity sometimes led to less reliable forecasts, as reflected in its higher RMSE scores for certain products.

In contrast, Simple Exponential Smoothing (SES), with its straightforward approach, provided competitive forecasts for products with relatively stable or low-variability sales patterns. SES consistently achieved low RMSE scores for its top-performing products, underscoring its suitability for datasets without strong seasonal or trend components. However, SES struggled to handle products with moderate complexity, where more sophisticated modeling techniques might have been beneficial.

These results suggest that a dynamic model selection framework could be advantageous. Products with high sales variability or clear seasonal trends could be better suited for Prophet, while SES might remain the preferred choice for simpler, stable patterns. Future studies could refine criteria for model allocation by incorporating metrics such as sales variability, seasonality indices, or product importance.

Lack of Comparison

One limitation of this study is the absence of a detailed comparison with other classical forecasting approaches, such as those described by Hyndman [2]. While Prophet incorporates many of the strengths of classical models, including ARIMA and Holt-Winters Exponential Smoothing, a broader evaluation against these methods could provide valuable insights. For instance, ARIMA's ability to model stationary time series and Holt-Winters' triple exponential smoothing could serve as benchmarks to further evaluate Prophet's and SES's performance.

Future studies could include these classical methods to provide a broader perspective on Prophet's and SES's relative strengths and limitations, especially for datasets with distinct characteristics.

Performance

Prophet's top-performing products, such as `FOODS_3_141_TX_3` (RMSE: 0.60) and `FOODS_3_082_TX_3` (RMSE: 0.38), highlight its ability to effectively generalize for products with very low or sporadic sales. The top-performing products, based on RMSE, tend to have minimal sales (e.g., Test Total Sales: 5, 6, or even 2), suggesting that Prophet can handle datasets with sparse activity. However, these results are less insightful for practical applications, as such products do not contribute significantly to overall sales or forecasting challenges. The performance of Prophet on frequently sold products provides a more meaningful evaluation of its robustness.

Similarly, SES's top-performing forecasts are largely for products with extremely low sales variability or stable, negligible demand (e.g., `FOODS_3_539_TX_3`, RMSE: 0.19, Test Total Sales: 1). While SES excels in these cases, its results also indicate a limited challenge in forecasting such products. More meaningful insights can be drawn from its performance on products with moderate to high sales, where patterns or variability are more pronounced.

For both models, examining results for frequently sold products is more valuable in understanding their true forecasting capabilities. Products with higher Test Total Sales, such as `FOODS_3_586_TX_3` for Prophet (RMSE: 16.25, Test Total Sales: 2056), provide clearer indications of how well a model handles complex seasonality or variability. SES's performance on products like `FOODS_3_147_TX_3` (RMSE: 4.47, Test Total Sales: 74) highlights similar challenges in maintaining accuracy for higher-demand products.

To improve overall forecasting performance, future work could focus on integrating domain-specific knowledge, such as promotional effects, holidays, or external shocks, into forecasting models. Additionally, exploring hybrid approaches that combine SES and Prophet could balance simplicity and complexity. For example, using SES for initial trend smoothing followed by Prophet for residual modeling could leverage the strengths of both methods. Such an approach may also improve forecasting accuracy for products with higher sales activity, where both models currently show room for improvement.

Generalisability

This study's scope was limited to the Food3 category at Walmart's TX3 store. While this focus allowed for detailed analysis, it raises questions about the generalisability of the findings. Testing the models across a broader range of product categories and store locations would help validate the observed patterns and insights.

Furthermore, extending the dataset to cover longer time horizons could uncover long-term trends and changes in seasonality that were not captured in this analysis. Such an extension would also allow for a more comprehensive evaluation of model robustness over extended forecast periods.

Model Evaluation Metrics

This study primarily evaluated models using RMSE and MAE, standard metrics for point forecast accuracy. While these metrics are effective, they do not account for uncertainty in forecasts. Future research could benefit from incorporating probabilistic evaluation metrics, such as the Continuous Ranked Probability Score (CRPS) or Pinball Loss, to better understand forecast reliability. Additionally, prediction intervals, which were not extensively analyzed, could provide deeper insights into the robustness of the models under varying conditions.

Future Directions

Future studies should consider setting stricter thresholds for model selection, especially for products with low average sales or sporadic sales patterns. Refining criteria to allocate complex models like Prophet only to products with significant seasonality or variability could reduce overfitting and improve overall performance. Moreover, hybrid approaches combining SES and Prophet may further enhance forecasting accuracy across diverse product types.

Conclusion

These insights align with the research questions posed in this study (**Section 1**). Prophet demonstrates a strong capability to handle high variability and complex seasonality, making it well-suited for products with dynamic sales patterns. In contrast, SES offers a lightweight and efficient solution for simpler, more stable cases. By leveraging the complementary strengths of both models, this study highlights an effective strategy to optimize forecast accuracy across diverse product types, providing a solid foundation for future research to expand upon.

REFERENCES

[1] Robert Fildes, Shaohui Ma, and Stephan Kolassa. 2022. Retail forecasting: Research and practice. *International Journal of Forecasting* 38, 4 (2022), 1283–1318. https://doi.org/10.1016/j.ijforecast.2019.06.004 Special Issue: M5 competition.

[2] Rob J. Hyndman and George Athanasopoulos. 2021. *Forecasting: Principles and Practice* (3rd ed.). OTexts, Melbourne, Australia. https://otexts.com/fpp3

[3] Matthieu Komorowski, Dominic C. Marshall, Justin D. Salciccioli, and Yves Crutain. 2016. *Exploratory Data Analysis*. Springer International Publishing, Cham, 185–203. https://doi.org/10.1007/978-3-319-43742-2_15

[4] Jose-Luis Guerrero Ricardo Ernst and Alan Roshwalb. 1993. A Quality Control Approach for Monitoring Inventory Stock Levels. *Journal of the Operational Research Society* 44, 11 (1993), 1115–1127. https://doi.org/10.1057/jors.1993.184 arXiv:https://doi.org/10.1057/jors.1993.184

Table 6: Results: Top 20 Sold Products

Product-ID	MAE	RMSE	Test_Weekly_Average_Sales	Test_Total_Sales
FOODS_3_030_TX_3_validation	12.32	15.16	23	641
FOODS_3_080_TX_3_validation	3.14	3.97	11	305
FOODS_3_090_TX_3_validation	25.61	29.35	64	1799
FOODS_3_202_TX_3_validation	8.39	9.94	32	892
FOODS_3_252_TX_3_validation	15.68	19.66	51	1440
FOODS_3_376_TX_3_validation	8.46	10.50	15	426
FOODS_3_377_TX_3_validation	8.07	10.27	42	1168
FOODS_3_541_TX_3_validation	10.36	11.40	7	190
FOODS_3_555_TX_3_validation	9.50	11.66	30	829
FOODS_3_586_TX_3_validation	13.96	16.25	73	2056
FOODS_3_587_TX_3_validation	9.68	11.55	23	636
FOODS_3_607_TX_3_validation	7.71	10.36	14	385
FOODS_3_635_TX_3_validation	14.04	17.86	7	188
FOODS_3_681_TX_3_validation	10.11	12.26	18	507
FOODS_3_694_TX_3_validation	6.46	7.42	25	687
FOODS_3_714_TX_3_validation	6.89	8.07	25	706
FOODS_3_723_TX_3_validation	6.36	8.02	17	484
FOODS_3_804_TX_3_validation	9.46	11.90	26	727
FOODS_3_808_TX_3_validation	12.36	13.75	6	162
FOODS_3_811_TX_3_validation	6.32	9.28	13	376

Table 7: Top 10 and Bottom 10 Prophet Metrics Based on RMSE

Product ID	**MAE**	**RMSE**	**Weekly Avg Sales**	**Total Sales**
FOODS_3_141_TX_3_validation	0.29	0.60	0	5
FOODS_3_199_TX_3_validation	0.21	0.60	0	6
FOODS_3_128_TX_3_validation	0.25	0.57	0	5
FOODS_3_274_TX_3_validation	0.25	0.57	0	9
FOODS_3_194_TX_3_validation	0.32	0.57	0	6
FOODS_3_577_TX_3_validation	0.25	0.57	0	6
FOODS_3_050_TX_3_validation	0.21	0.46	0	2
FOODS_3_294_TX_3_validation	0.21	0.46	0	2
FOODS_3_045_TX_3_validation	0.21	0.46	0	6
FOODS_3_082_TX_3_validation	0.14	0.38	0	3
FOODS_3_090_TX_3_validation	25.61	29.35	64	1799
FOODS_3_444_TX_3_validation	21.46	23.20	23	653
FOODS_3_120_TX_3_validation	20.14	22.67	15	409
FOODS_3_252_TX_3_validation	15.68	19.66	51	1440
FOODS_3_580_TX_3_validation	7.50	19.20	9	241
FOODS_3_635_TX_3_validation	14.04	17.86	7	188
FOODS_3_586_TX_3_validation	13.96	16.25	73	2056
FOODS_3_295_TX_3_validation	9.96	15.68	15	421
FOODS_3_501_TX_3_validation	10.61	15.65	14	391
FOODS_3_234_TX_3_validation	10.71	15.56	15	425

Table 8: Top 10 and Bottom 10 SES Metrics Based on RMSE

Product ID	MAE	RMSE	Weekly Avg Sales	Total Sales
FOODS_3_539_TX_3_validation	0.04	0.19	0	1
FOODS_3_328_TX_3_validation	0.04	0.19	0	1
FOODS_3_796_TX_3_validation	0.04	0.19	0	1
FOODS_3_793_TX_3_validation	0.04	0.19	0	1
FOODS_3_713_TX_3_validation	0.04	0.19	0	1
FOODS_3_758_TX_3_validation	0.04	0.19	0	1
FOODS_3_597_TX_3_validation	0.04	0.19	0	1
FOODS_3_779_TX_3_validation	0.04	0.19	0	1
FOODS_3_260_TX_3_validation	0.04	0.19	0	1
FOODS_3_553_TX_3_validation	0.04	0.19	0	1
FOODS_3_147_TX_3_validation	2.64	4.47	3	74
FOODS_3_747_TX_3_validation	1.96	4.32	2	59
FOODS_3_047_TX_3_validation	1.96	2.32	2	67
FOODS_3_618_TX_3_validation	1.64	2.05	2	62
FOODS_3_278_TX_3_validation	1.64	1.96	2	50
FOODS_3_152_TX_3_validation	1.75	1.92	1	25
FOODS_3_827_TX_3_validation	1.50	1.83	2	42
FOODS_3_317_TX_3_validation	1.36	1.81	1	26
FOODS_3_545_TX_3_validation	1.39	1.78	2	51
FOODS_3_603_TX_3_validation	1.25	1.78	2	49

Table 9: Summary Statistics for Top 10, Middle 5, and Bottom 5 Products

id_trimmed	mean_sales	median_sales	sd_sales	total_sales	max_sales	min_sales	n_days
FOODS_3_586	78.47	76	25.16	150122	200	0	1913
FOODS_3_090	60.04	61	48.10	114854	380	0	1913
FOODS_3_252	45.29	43	20.57	86632	150	0	1913
FOODS_3_555	40.60	37	18.74	77673	128	0	1913
FOODS_3_377	38.58	38	11.76	73797	95	0	1913
FOODS_3_030	25.08	21	28.07	47976	320	0	1913
FOODS_3_202	21.72	21	13.87	41551	109	0	1913
FOODS_3_714	21.25	20	10.27	40650	88	0	1913
FOODS_3_587	21.07	21	14.65	40306	83	0	1913
FOODS_3_694	18.38	18	7.00	35152	54	0	1913
FOODS_3_721	0.66	0	1.05	1257	9	0	1913
FOODS_3_725	0.66	0	1.05	1256	8	0	1913
FOODS_3_731	0.66	0	1.12	1256	8	0	1913
FOODS_3_760	0.66	0	1.64	1256	16	0	1913
FOODS_3_464	0.65	0	1.02	1253	11	0	1913
FOODS_3_633	0.03	0	0.20	64	2	0	1913
FOODS_3_472	0.02	0	0.21	47	5	0	1913
FOODS_3_693	0.02	0	0.17	46	2	0	1913
FOODS_3_260	0.02	0	0.20	43	4	0	1913
FOODS_3_171	0.02	0	0.17	35	4	0	1913

Table 10: Proportion of Zero Sales for Top 10, Middle 5, and Bottom 5 Products

id_trimmed	total_days	zero_sales_days	zero_sales_pct
FOODS_3_260	1913	1883	98.43
FOODS_3_171	1913	1883	98.43
FOODS_3_472	1913	1877	98.12
FOODS_3_693	1913	1871	97.80
FOODS_3_633	1913	1857	97.07
FOODS_3_760	1913	1537	80.35
FOODS_3_731	1913	1236	64.61
FOODS_3_721	1913	1172	61.27
FOODS_3_725	1913	1166	60.95
FOODS_3_464	1913	1136	59.38
FOODS_3_090	1913	472	24.67
FOODS_3_030	1913	467	24.41
FOODS_3_587	1913	388	20.28
FOODS_3_202	1913	257	13.43
FOODS_3_555	1913	9	0.47
FOODS_3_377	1913	8	0.42
FOODS_3_714	1913	8	0.42
FOODS_3_694	1913	8	0.42
FOODS_3_586	1913	7	0.37
FOODS_3_252	1913	6	0.31

Figure 10: Distribution for top 10, middle 5, and bottom 5 sold products.

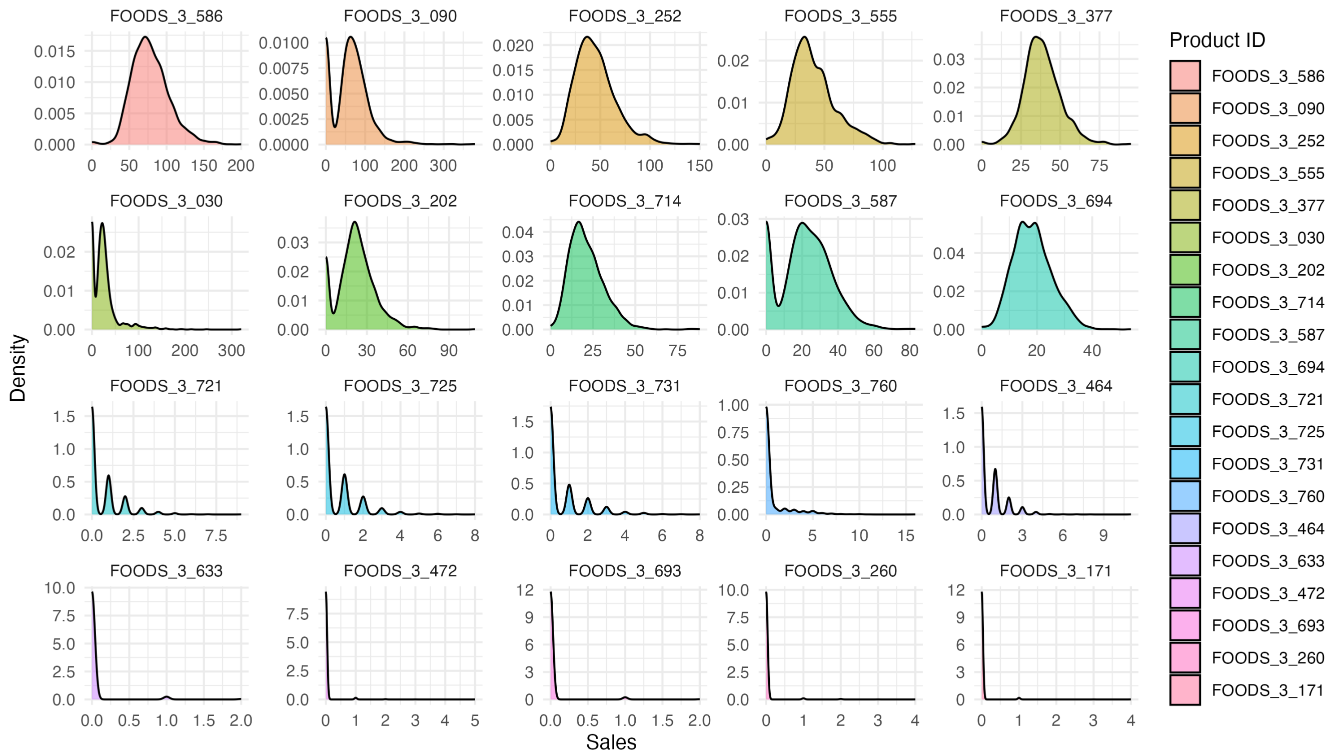


Figure 11: Sales Trend for Top 10, Middle 5, and Bottom 5 products.

Figure 12: Seasonal Decomposition Top 10 Products