Walmart Sale Time Series Foresting
Team 12

Jakob Chumtong (12986992), Dimitrios Papadopoulos (15240924), Efstathios Trigkas (14765667) and Dimitrios Tsiamouras (14896184)
Applied Forecasting in Complex Systems 2023
University of Amsterdam

ABSTRACT

The paper at hand presents a comprehensive analysis and forecasting approach for retail sales, specifically that of Walmart's M5 Forecasting competition dataset —the hierarchical sales data from the TX3 store in Texas in the Food3 category. The datasets provided contain item details, such as prices and sales, and calendar information in the form of special events and SNAP allowance. Our methodology consists of an in-depth Exploratory Data Analysis (EDA), followed by the implementation of three forecasting methods: Prophet, STL with multiple seasonality, and XGBoost. We conclude our work by discussing our results and providing our closing thoughts.

The EDA has been designed to provide insights into temporal patterns, item-level behaviors, and the impact of prices, special events, and SNAP days on sales. Through it, we aim to identify trends, time-series variation and the existence of one or more seasonal components. The forecasting methods then leverage these insights to bolster their predictive accuracy about future sales trends. Prophet, in particular, excels in capturing complex seasonality and handling irregular holidays, achieving an RMSE of 2.113774 and MAE of 1.642009. STL with multiple seasonality effectively models overlapping seasonal patterns, while XGBoost, a machine learning approach, demonstrates promising results with the integration of a plethora of diverse, exogenous variables.

Our paper highlights the strengths and limitations of each method and underscores the importance of exogenous variables in improving forecasting accuracy. We acknowledge challenges in handling abrupt changes, outliers, and external factors influencing sales. The exploration of advanced machine learning techniques, ensemble methods, and a broader set of exogenous variables is suggested for future research.

1 INTRODUCTION

Sales forecasting stands as a pivotal task in the retail domain, playing a crucial role in effective inventory management that simultaneously meets customer demand. The M5 Forecasting competition revolves around this very task, offering a challenging yet valuable opportunity to enhance predictive accuracy. Walmart, the world's largest company by revenue, generously provides an extensive datasets to this competition, of which we use the hierarchical sales data from its TX3 store in Texas.

Our specific focus within this expansive dataset is narrowed down to the Food3 category. This subset captures detailed information at the item level, along with store-specific details. Notably, the dataset incorporates various explanatory variables, including pricing, promotional activities, day-of-the-week trends, and special events. This rich and diverse dataset serves as a valuable testing ground for refining forecasting models.

In navigating the challenges of this project, our primary goal is to develop models that leverage the provided data to make precise predictions about future sales trends. The complexity of retail sales, influenced by dynamic factors such as pricing strategies, promotional events, and temporal patterns, adds intricacy to the forecasting task.

Most of the methodologies where based on the book by Rob J Hyndman and George Athanasopoulos [1]. The code and data used can be found on Github[1].

The structure of this paper unfolds as follows: Section 2 presents a thorough EDA to understand underlying patterns and valuable predictors, serving as the basis for defining our forecasting models. In Section 3, we delve into the details of the different models implemented, shedding light on their intricacies. In Section 4, we present each model's forecasts and compare them on the same metrics. Finally, the last two sections provide a comprehensive discussion on the project, the methodology employed, and insights derived from the results, followed by a conclusion that encapsulates the findings and potential avenues for future research.

2 EXPLANATORY DATA ANALYSIS

Description of the Available Data

To begin with, we present the datasets available and utilized for this project. The **calendar_afcs2023** file contains daily information on events and the availability of SNAP. The file **sell_prices_afcs2023** provides weekly information on product prices. Notably, in cases where a product was unavailable during a week, the price record is missing altogether, rather than being recorded as 0. The file

[1] https://github.com/Skoyntoyflis/AFCS-Project

sales_train_validation_afcs2023 contains daily information on product sales. It is essential to note that the data in this file are continuous, meaning that even if a product was unavailable during a period, the recorded sales were marked as 0.

To perform our EDA, we merged these datasets. This allowed us to have comprehensive records for a product's sales, price, and information on events and SNAP for each day. This consolidation simplified the process of analyzing and comparing the data.

In-depth Exploration and Findings

To initiate our exploratory data analysis (EDA), we delve into the time series representing aggregated sales. In Figure 1, a discernible upward trend in total sales captures our attention, accompanied by conspicuous weekly and yearly seasonal patterns, and a questionable monthly seasonality, which we will delve into shortly. Alongside these patterns, we pinpoint five notable sales troughs occurring on the 25th of December, a consequence of store closures on that day. Throughout the series, sporadic spikes emerge, potentially tied to circumstantial events. Noteworthy is the consistent overall variation, which remains constant and does not escalate with the series' magnitude.

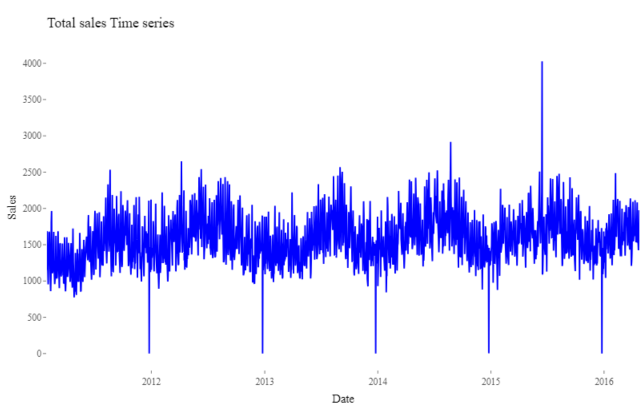

Figure 1: Aggregated sales

Although the general plot of the aggregated sales provides good preliminary information on temporal patterns, we deem it is necessary that we explore other possible underlying relations through Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots has yielded insightful observations. In Figure 2, it becomes evident that the data exhibit characteristics inconsistent with the properties of white noise. This conclusion is drawn from the notable concentration of autocorrelations outside the expected confidence intervals. Furthermore, a discernible pattern emerges in the form of pronounced weekly seasonality, with a conspicuous peak occurring every 7 records or days. The recurrent nature of this peak strongly supports the presence of a robust weekly seasonality within the dataset. Beyond establishing the non-random nature of the time series, these findings lay the groundwork for deeper investigations into the potential factors influencing this observed weekly rhythm. Such periodicity can be instrumental in refining forecasting models, providing a nuanced understanding of the temporal dynamics inherent in the sales data.

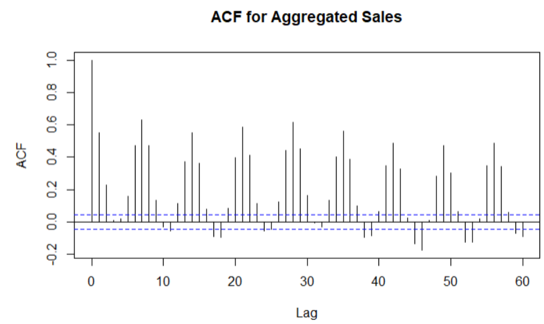

Figure 2: Acf plot

In conjunction with the ACF exploration, the PACF analysis of the aggregated sales data provides additional insight into the temporal dynamics. Contrasting with the discernible weekly seasonality highlighted in the ACF plot, the PACF plot, as depicted in Figure 3, does not reveal an overt seasonal structure. This absence of obvious seasonality in the PACF suggests a distinctive pattern in the data. Unlike the ACF, where the autocorrelations at specific lags revealed strong weekly periodicity, the PACF indicates that once the influence of shorter lags is considered, subsequent lags do not exhibit significant correlations. This nuanced observation implies that the effects of past observations on the current one are more immediate, with the direct autoregressive influence diminishing beyond shorter time intervals.

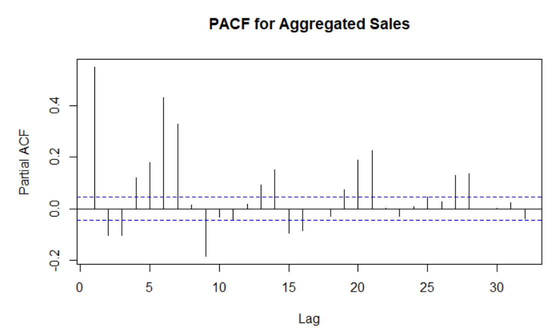

Figure 3: Pacf plot

The last seasonal component we will comment on is the monthly one. We already mentioned above that the presence of a monthly component is rather dubious. For this purpose we have Decomposed the series on a monthly level. Figure shows that, although there is a pattern on a monthly basis, this is rather inconsistent and appears to change at certain parts of the series. It requires further "zooming-in" on the series to get conclusive evidence of the reason for this behavior, which could be attributed to unavailability of items, changes in the seasons etc.

Importantly, the strong seasonal and trend components prompt the assertion that our data lacks stationarity. To further support our claim, we perform a KPSS test result, yielding a p-value of 0.01, which underscores the non-stationary nature of our data. This should be taken heavily into consideration when selecting appropriate forecasting models that can work with such a particularity.

To gain a better understanding of our data, we choose to plot a random set of items. Upon observation, we note the presence of numerous low values along with seemingly random spikes. If these spikes were entirely random or similar to white noise, they could pose challenges for our forecasting efforts. Another notable pattern is the intermittent unavailability of many products, as evidenced by the absence of their selling prices in certain weeks. It's important to differentiate this case from instances where the product was available (selling price recorded) but still had a recorded sale of 0, indicating no demand.

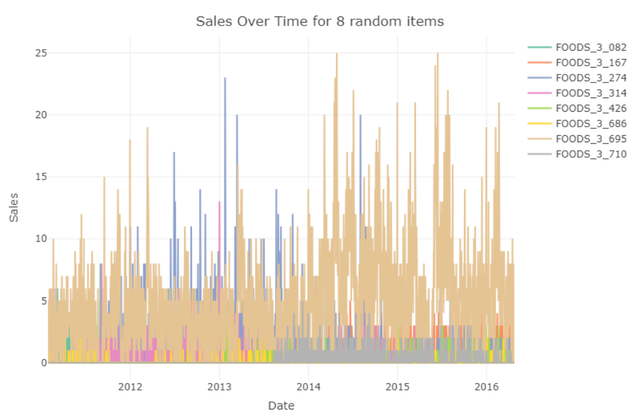

Figure 4: Time series for subset items

Continuing to focus on the behavior of individual items, we discover that item 586 had the highest mean sales throughout the whole recorded period, while 171 had the lowest. It is also worth noting that 586 was available every day, while 171 was out-of-stock for 1183 days. Lastly, looking at the last 365 days, both items follow an increase in their sales, keeping to the increasing overall trend we saw above.

At this juncture, we will delve into the underlying temporal patterns of our data, which may not be readily apparent in the aggregated plots. Specifically, we will examine two heatmaps. The first one illustrates the relationship between the day of the week and the month.

It becomes evident that sales are higher during the weekend, irrespective of the month. Mondays consistently show lower sales but remain above those of the other four weekdays. The highest number of sales is recorded on Sundays in March and August. Conversely, the lowest number of sales is noted on Thursdays in November, closely followed by Wednesdays in January. The weekly seasonality is prominently reflected in the heatmap, corroborating our earlier findings.

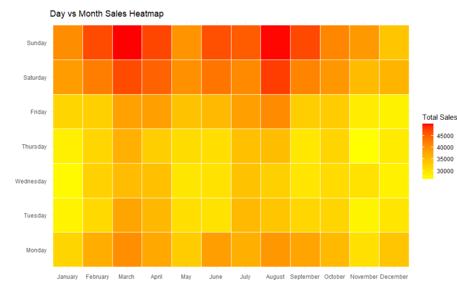

Figure 5: Day versus Month

As a complement to the previous heatmap, we will now examine how sales unfold across consecutive days of the month. Consistent with our earlier observations, a clear weekly seasonality is evident, with higher sales numbers during the weekends and on Mondays. However, a distinct pattern emerges: sales peak during the initial days of the month and gradually decrease as the month progresses.

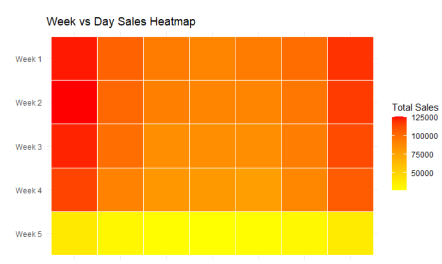

Figure 6: Week versus Day

Analysis of Prices. Regarding prices, the EDA reveals notable patterns. In Figure 7a), a significant portion of items exhibits a lack of pricing data for extended durations within the dataset. This phenomenon may be attributed to the intermittent availability of certain food items throughout the year.

Moving on to Figure 7b), spanning the 5-year timeframe, it becomes evident that substantial fluctuations in item prices are infrequent and the normalized variations in prices, depicted in Figure 7d), reach peaks of approximately 10%. Additionally, the mean item price across the dataset is observed to be 2.72 (Figure 7c)).

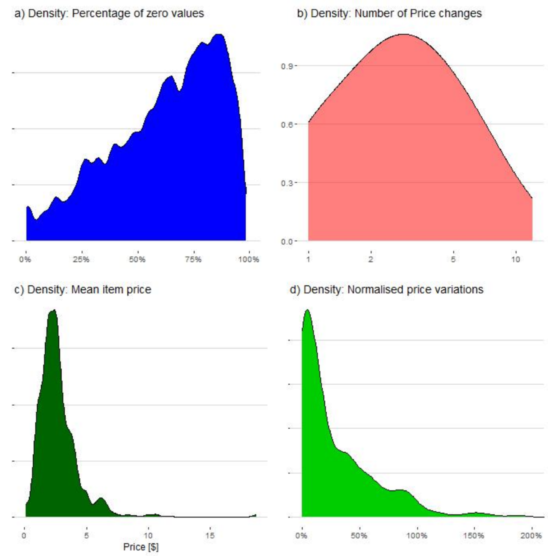

Figure 7: General price statistics

To further comprehend the impact of prices on sales, it is crucial to investigate whether both short-term and prolonged alterations in pricing influence consumer purchasing behaviours. The overall correlation between price fluctuations and sales is observed to be -0.1306, indicating a modest decrease in sales predominantly associated with price increases.

In Figure 8, the sales dynamics for three exemplar items are illustrated, with the orange line representing scaled prices. Notably, certain products, such as "FOODS_3_030" during 2013-2014, demonstrate a considerable influence of price changes on sales. Conversely, for items like "FOODS_3_520," consumer purchasing appears relatively stable, irrespective of price variations. Moreover, a noticeable trend emerges when there is a prolonged absence in sales followed by a price change, sales tend to recover in the majority of cases. This analysis provides valuable insights into the nuanced relationship between pricing dynamics and consumer behaviour within the dataset.

Analysis of Calendar. In our investigation of the dataset and its inherent dynamics, our focus shifted towards assessing the influence of special events, holidays, and Snap eligible days on product sales. We aimed to discern whether specific holidays or events had measurable impacts on product sales both on the day of the event and in the subsequent days.

Figure 8: Example plot of sales and prices for 3 items

Additionally, with regard to Snap, an acronym for the Supplemental Nutrition Assistance Program that extends benefits to eligible low-income individuals, we aimed to determine whether sales were influenced on these designated days.

As illustrated in the figure 9, the calendar indicates that approximately 8%. of the days are designated for special events. Among these events, 25%. are Cultural, 30% are National, 35% are Religious, and 10% are Sporting events. Upon examining the percentage of days associated with SNAP purchases, it becomes evident that around 33% of the days are SNAP days. Further investigation and research reveal that SNAP days consistently occur on the same days each month. There are always 10 SNAP days, and their specific dates for each month are outlined in the calendar figure 10.

After examining the distributions of the explanatory variables in the calendar, we aimed to analyze their relationships with product sales. Initially, we assessed the influence of snap days on sales. As illustrated in the figure 11, the average sales on snap days consistently surpass those on normal days, with an average increase of 13%. This indicates that snap days indeed have a significant impact on sales, suggesting that we should incorporate this variable into our forecasting models.

Following that, we delved into examining the correlation between events and product sales, both on the actual day of the event and the subsequent days. The analysis revealed that the average sales on event days and non-event days are more or less similar. We discovered that overall there is no significant increase or decrease in sales on special event days, either on the day of the event itself or in the days that follow.

After further analyzing the sales during events, the bar chart below Figure 12 illustrates that on Cultural and National event days, sales tend to be slightly higher than on

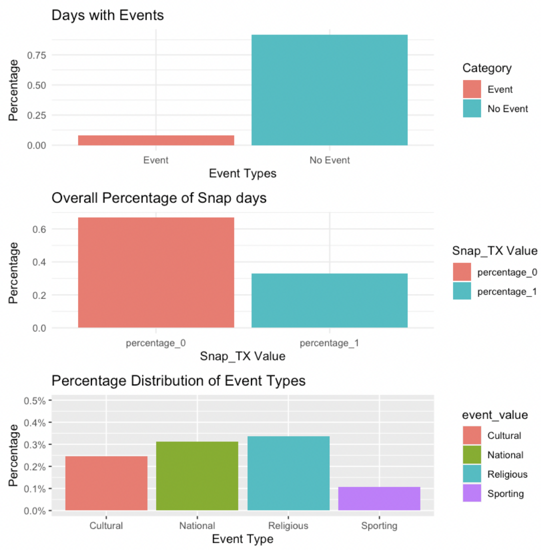

Figure 9: Distribution of Events and Snap Days

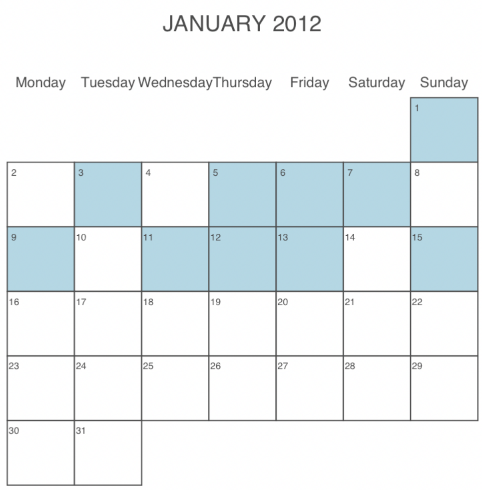

Figure 10: Specific Days with Snap

non-event days. Conversely, on Sporting and Religious event days, sales tend to be lower. Additionally, our analysis revealed that for specific Religious events, such as Christmas Day, sales register as 0, as it is the only day when Walmart stores are closed.

EDA Summary and Inspiration

In our thorough EDA, we delved into the daily retail sales data from Walmart's TX3 store, revealing intricate temporal

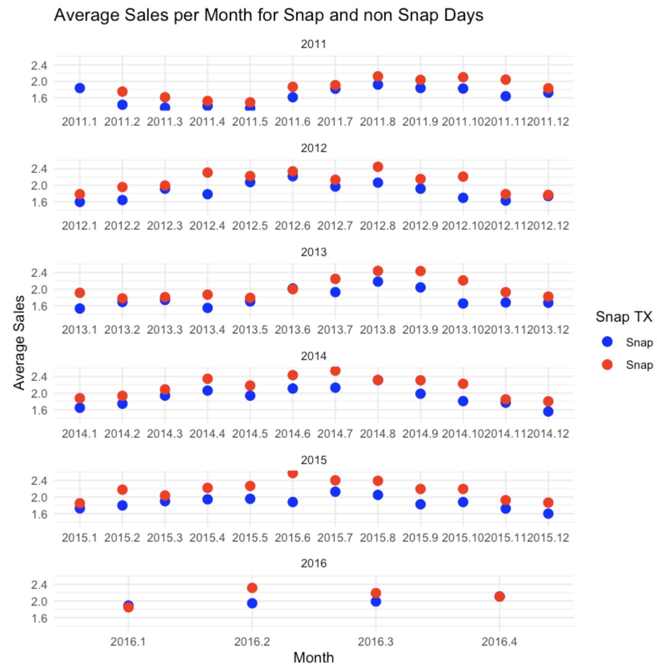

Figure 11: Average sales on Snap days

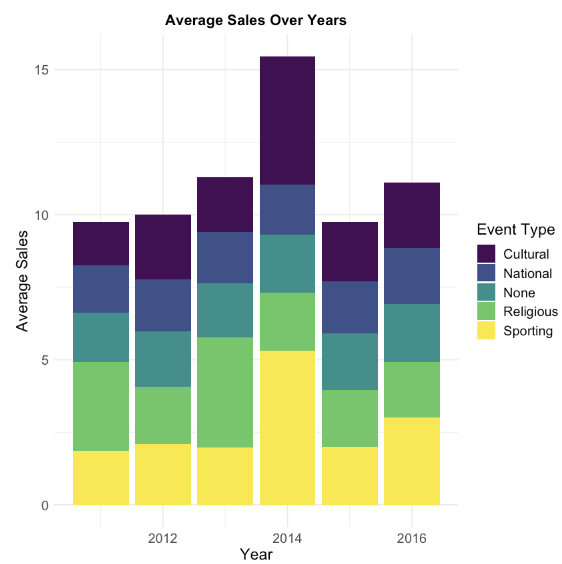

Figure 12: Overall

patterns. Notably, a discernible rising trend in total sales, coupled with pronounced weekly, monthly, and yearly seasonality, emerged as significant predictors crucial for enhancing forecast accuracy. To accommodate these diverse seasonal patterns, we will make use of Forecasting models capable of handling Multiple Seasonality.

Individual item analysis uncovered unique behaviors, including inconsistent availability, which could potentially hinder a model's ability to learn consistent patterns. To address this, we will consider robust forecasting methods such as Prophet and XGBoost, designed specifically to handle such peculiarities.

In order to further enhance our models' adaptability, we will incorporate lagged values for the last 28 days. These historical observations of sales, events, and Snap days enable the model to capture temporal dependencies and patterns, especially crucial in addressing the inconsistency in item availability.

Despite the sparsity of Events, we identified a consistent increase in sales leading up to these occasions. Similarly, SNAP days exhibited heightened sales. To model these effects, we will employ Dummy variables for Events and include SNAP as a predictor, having in mind that forecast period encompasses both Events and SNAP days.

All these insights guide our approach to forecasting, where we aim to leverage these predictors strategically. The next step involves implementing forecasting models aligned with these findings to achieve accurate and robust predictions.

3 FORECASTING METHODS

In this section, we present the various models implemented, each with its own intricacies and advantages.

STL with multiple seasonality

Given the daily nature of our dataset, it is plausible that multiple seasonal patterns are at play. This could include both short-term, weekly fluctuations and longer-term, annual patterns. To address this complexity and enhance our forecasting accuracy, we have chosen to explore the application of Seasonal and Trend decomposition using Loess (STL) with multiple seasonality. The STL() function, tailored to handle multiple seasonality, will provide us with several seasonal components in addition to trend and remainder components. These components will serve as the foundation for our forecasting approach. Specifically, we plan to employ seasonal naïve methods to forecast each of the seasonal components, while the seasonally adjusted data will undergo forecasting using the Exponential Smoothing State Space Model (ETS). Incorporating exogenous variables, including snap days, days of the week, event types, event names, and sell prices, further bolsters the predictive power of our models. These variables will be integrated into the forecasting process, allowing us to capture their impact on sales patterns and thus improve the overall performance of our forecasts.

Prophet

In contrast to traditional time series forecasting methods such as ARIMA, Dynamic Harmonic Regression, and STL with multiple seasonality, the Prophet model from the *fable* package operates on a different premise. Prophet, developed by Facebook, is specifically designed to handle time series data with strong seasonal patterns and irregular holidays. It is well-suited for datasets that exhibit daily observations, making it particularly applicable to retail scenarios.

Thus, the Prophet's intuitive design and flexibility make it particularly well-suited for the forecasting task at hand. It accommodates datasets with strong seasonal patterns and irregular holidays, aligning seamlessly with the characteristics of the retail sales data from Walmart's TX3 store. The model's ability to capture complex seasonality and incorporate user-defined events aligns with the specific requirements of the project, providing a robust solution for predicting daily sales in the given context.

In order to fit the model and feed the categorical variables of events, dummy variables had to be created. Then, the dummy variables, the future prices of the items and the future events and SNAP days in the calendar from the datasets provided were included in the model. Finally, two periods of seasonality were selected, week and year, both with 10 Fourier terms for flexibility, as this was most apparent during the EDA in Section 2. After fitting the model to the training data, the sales of each item for 28 days were forecasted using the test validation set, including data from the future prices, events and SNAP-allowed days. The forecast produced achieved a Root Mean Square Error (RMSE) of 2.039923 and a Mean Absolute Error (MAE) of 1.576549257.

Machine Learning Methods

In our quest to employ machine learning techniques for sales prediction, we explored various methodologies and discovered that models based on decision trees exhibit excellent forecasting performance.

Consequently, we opted to implement XGBoost model in Python, utilizing the mlForecast library. During the model construction phase, we incorporated a set of explanatory variables. These included events, prices, Snap days, as well as lagged values spanning from 1 to 28. Additionally, we integrated various lagged transformations, such as rolling mean, into our modelling framework.

4 RESULTS

In this section, we present the performance evaluation of three forecasting models: Prophet, STL with multiple seasonality, and XGBoost. The primary objective is to assess their effectiveness in predicting future sales. The evaluation metrics employed include Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).

The results are depicted in Table 1. Upon examination, it is evident that the STL model with multiple seasonality exhibits the lowest RMSE, while XGBoost demonstrates the

Model	RMSE	MAE
Prophet	2.03	1.57
STL with MS	2.00	1.56
XGBoost	2.95	1.40

Table 1: Evaluation Metrics

lowest MAE. This discrepancy in evaluation metrics arises because RMSE tends to penalize larger residuals.

After careful consideration of the various accuracy metrics, we have opted to designate STL as our primary forecasting model. Subsequent to this decision, we have provided additional insights through the plotting of residuals in the Figure 13. Notably, the residuals of the STL model appear to follow a normal distribution. Figure 14 further illustrates the comparison between predicted and actual sales values.

It is imperative to note that while XGBoost had the lowest MAE, the choice of STL as our primary model takes into account the holistic evaluation of multiple metrics and its ability to capture the complex seasonality patterns in the sales data. The figures showcase the distribution of residuals in the STL model, providing assurance that the model is capturing the underlying patterns effectively. The comparison between predicted and real sales values in Figure 14 adds a visual dimension to the performance evaluation.

In conclusion, our comprehensive assessment of forecasting models, considering both MAE and RMSE, positions STL as the preferred choice for predicting future sales. The normal distribution of residuals and the close alignment of predicted and actual sales values further validate the robustness of the selected model.

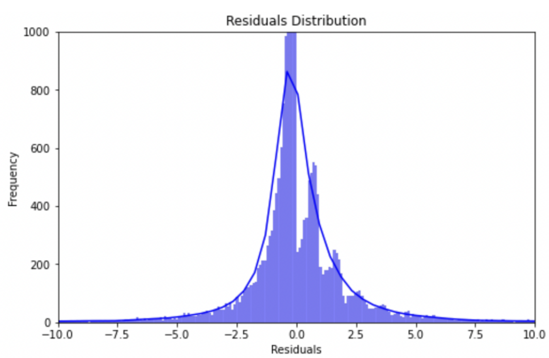

Figure 13: STL residuals

5 DISCUSSION

Sales forecasting is a complex task that plays a pivotal role in a company's operational and strategic decisions. In the context of the M5 Walmart competition, we employed three distinct forecasting methods—Prophet, STL with multiple seasonality, and XGBoost—in an attempt to accurately predict future sales for various products in Walmart's TX3 store. Our approach involved the integration of exogenous variables, such as sell price, snap days, events, and day of the week, to capture additional factors that influence consumer purchasing behavior. In this discussion, we reflect on the effectiveness of our methods, acknowledge their limitations, and explore avenues for future improvements.

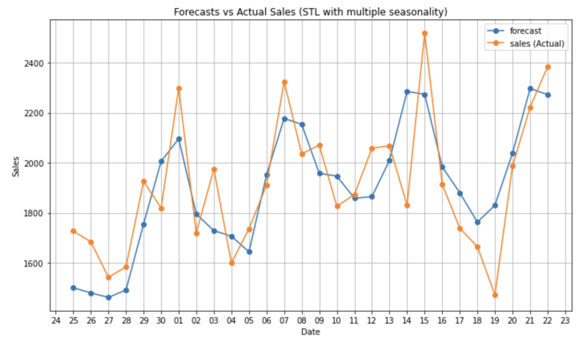

Figure 14: 28 days Forecast

Prophet Forecasting

Our use of the Prophet forecasting algorithm demonstrated its utility in capturing seasonality and trend components within the sales data. Prophet's flexibility allowed us to incorporate holidays and special events, further enhancing its forecasting capabilities. However, Prophet relies heavily on historical data and may not handle abrupt changes or outliers gracefully. In this context, the model may have struggled to adapt to unprecedented events, such as the COVID-19 pandemic. Future enhancements could involve fine-tuning Prophet's hyperparameters and exploring methods to make it more robust to sudden disruptions.

STL with Multiple Seasonality

The application of STL with multiple seasonality revealed its effectiveness in capturing complex and overlapping seasonal patterns within the sales data. By allowing for multiple seasonality components, this approach was particularly useful in modeling weekly and monthly patterns simultaneously. Nevertheless, this method may face challenges when dealing with long-term trends and large datasets. Our implementation addressed these concerns through robustness techniques, but further research into handling such issues is warranted.

XGBoost Machine Learning

Leveraging the power of XGBoost, a gradient boosting machine learning algorithm, we aimed to capture intricate relationships between sales and exogenous variables. The model showed promising results, especially when incorporating various features. However, XGBoost's performance can be highly dependent on feature engineering and hyperparameter tuning. A deeper exploration of feature importance and additional feature selection methods could enhance its accuracy.

Exogenous Variables:

Our inclusion of exogenous variables, such as sell price, snap days, events, and day of the week, was essential for improving the forecasting accuracy of all three methods. However, the real world is replete with diverse factors influencing sales—economic conditions, local events, competitor actions, and more. Our models may not have fully captured the complexity of these external influences. Future work could involve expanding the set of exogenous variables to encompass a broader range of factors.

In conclusion, sales forecasting is indeed a challenging task, and our efforts represent a valuable step forward. Our implemented models provided reasonable predictions, yet there is always room for improvement. The future of sales forecasting may lie in the integration of more advanced machine learning techniques, ensemble methods, and the incorporation of unstructured data sources, such as economic indicators. Moreover, exploring advanced statistical models that can handle evolving and volatile data environments could further enhance our forecasting capabilities.

6 CONCLUSION

In conclusion, our endeavor to forecast sales for TX3's Food3 category using a comprehensive analysis and diverse forecasting methods has provided valuable insights into the intricacies and peculiarities of the M5 Forecasting competition data. Through our EDA, we uncovered diverse temporal patterns, unique item-level behaviors, and the impact of multiple factors such as prices, special events, and SNAP days on sales. All of these findings served immensely in complexifying our models and enhancing their ability to identify latent relationships, that would otherwise be lost.

The forecasting methods that were elected, namely Prophet, STL with multiple seasonality, and XGBoost, each exhibited their own strengths and limitations. Prophet excelled in capturing complex seasonality and handling irregular holidays, while STL effectively modeled overlapping seasonal patterns. XGBoost, a machine learning approach, demonstrated promising results with the integration of diverse exogenous variables. This gives merit to the idea of the implementation of an ensemble of methods, in an attempt to cover each method's weakness, while leveraging their strength.

Our paper highlights the importance of exogenous variables in improving forecasting accuracy and acknowledges challenges in handling abrupt changes, outliers, and external factors influencing sales. We provided a detailed discussion of the results, emphasizing the nuances of each method and proposing avenues for future research.

In summation, we aspire that our work be used in the ongoing efforts to advance sales forecasting. We hope to have set a solid foundation for future research and warmly suggest that more advanced methods, such as ensemble or machine learning approaches, be employed towards these tasks.

REFERENCES

[1] [n.d.]. *Forecasting: Principles and Practice (3rd ed).* https://otexts.com/fpp3/