FORECAST

SEP 2024

MMA 867 : PREDICTIVE MODELLING

SUPPLY OPTIMIZATION

Mitigate stock-outs worth $3 Billion annually

PREDICT.PRODUCE. PROFIT

SARIMA model for demand forecasting

SALES SURGE, ZERO WASTE

Smart Solution for a dynamic supply chain

0 24563 84926 54 2

Executive Summary:

In the world of constant change, Inventory management has still been one of the biggest challenges that fast-paced FMCG experience today. To shed light on the implications of this issue, in this project, we explore the realm of inventory operation from the lens of Walmart. As a world leader in the FMCG industry, Walmart has always strived to incorporate innovative approaches not only to enhance its existing supply chain but also to anticipate future developments in this space. Recently, with the introduction of e-commerce and political instability, the global supply chain network has experienced a big shift. E-commerce allows customers to purchase goods from the convenience of their homes, allowing faster sales and higher customer expectations of product delivery. On the other hand, political instability can lead to longer than usual time required to meet deliveries. Focusing our research on Walmart, we noticed that Walmart is faced with an inventory management problem which is acting as a double-edged sword. On one end Walmart is experiencing the problem of stockouts, a problem that is costing the company high costs in terms of opportunity cost, as we are not meeting sales requirement of the consumer. On the other hand, Walmart also experiences the problem of unproductive stock, a scenario in which it experiences high unsold stocks gathering dust and storage costs in its warehouses. To tackle this issue, we believe Walmart can best address the two problems by predicting future sales and anticipating the surge in demand in well in advance. We developed a SARIMA model that leverages past trends in demand to predict potential future sales. To further ease the application of our findings and model, we have three recommendations for Walmart: (1) Deepen the application of SARIMA (2) Maintain bufer stock before busy months (3) Establish a collaborative partnership among its dealer stores.

Objective:

As one of the largest global retailers, Walmart faces critical challenges in better forecasting sales to minimize stockouts and improve operational eficiency. According to the data, Walmart faced around $3 billion in opportunity costs due to shortages of products in high customer demand (Anderson, 2014). These shortages not only lead to missed potential sales but also impact on customer satisfaction. When customers are unable to find the products they need, their dissatisfaction grows, which could have negative efects on long- term loyalty. Moreover, the inventory turnover rate for Walmart has dropped from 9-10 in the early 2000s to 8.3 in recent years (Walmart Inc, 2000; Walmart Inc, 2024). This indicates that products are sitting in warehouses or on shelves longer than they should be, which creates bottlenecks in the supply chain. As a result, the whole operation slows down, and cash flow gets tied up.

Therefore, how to efectively forecast sales to ensure the right products are available at the right time to prevent both overstocking and understocking is important for Walmart. This not only helps reduce inventory holding costs and minimize opportunity costs but also benefits to company's long-term growth.

Data Exploration:

<u>Data source</u>: This project aims to build a model that predicts sales of the Walmart stores. The data has been sourced from Kaggle (https://www.kaggle.com/datasets/aslanahmedov/walmart-sales- forecast/data?select=stores.csv).

The data consisted of three csv files viz. features, stores, and train, with each file containing diferent columns. While the "features file" had 12 columns, the "train" and "stores" files had 5 and 3 columns respectively. The data from these three files was merged to form one dataset.

<u>Missing values</u>: There were 7 columns (Markdown 1, Markdown 2, Markdown 3, Markdown 4, Markdown 5, CPI, Unemployment) in the "features" file that had missing values. All the missing "Markdown" values were replaced with 0, and as there was no information in the dataset, we assumed that promotion had not been made in that week. For the remaining two columns – "CPI" and "Unemployment" the MICE imputation technique was used to fill the missing value.

<u>Merging the data</u>: Once all the missing values were dealt with, the "features" dataset, was merged with the "train" dataset on common columns ("store" and "date"). This dataset obtained above was once again merged with the "stores" dataset on the "store" column. Thus, the final dataset consisted of 16 columns and 421,570 rows.

<u>Trends Uncovered</u>: Through visualizations, the trends identified helped the team to understand the data in a better way.

1) Stores '20'and '4' have the highest weekly sales. Dept number 92 has the highest weekly sales followed by dept number 95 and 38.
2) Type A store has the highest weekly sales followed by B and C. Almost half of the stores are bigger than 150000 units and categorized as A.
3) CPI and Unemployment are negatively correlated to sales. A higher CPI indicates higher inflation resulting in less disposable income in hand for the consumer and vice versa. Similarly, the lower unemployment rate implies more people have purchasing power and hence, resulting in higher sales.

4) Promotions and fuel prices are positively correlated to sales. With the increase in promotions the price of the goods is reduced, thus, leading to higher sales. Similarly, an increase in fuel price increases the price of goods. With Walmart keeping its profit margins low relative to its competitors, the sale of goods at Walmart increases relative to its competitors.
5)
6) Holiday average sales are higher than normal dates.
7) Sales were highest in December followed by November month. The least sales occur in January owing to the purchases made in Dec and Nov.
8) Sales were highest for the events of Christmas (Week 51) followed by Black Friday (Week 47)
9) Walmart sales are heavily influenced by promotions and holiday periods particularly Black Friday and Christmas along with Markdown 5 and Markdown 2.

Methodology & Results:

We experimented with several models to predict sales for the Walmart dataset such as Linear regression, Lasso regression, ARIMA, etc., but couldn't predict the sales accurately. With an R-squared value of 0.062, linear and lasso regression could only explain 6.2% of the variability in the Weekly sales based on the independent variables.

ARIMA: Our analysis revealed that data is already stationary, and no diferencing is required (fig1). This was further validated by performing the Augmented Dickey-Fuller (ADF) test, which confirmed stationarity by rejecting the null hypothesis.

After plotting the ACF and PACF (fig2), a strong spike was observed at lag 1 in the ACF plot, indicating that there is a high autocorrelation at first lag, which means that the value at time 't' is highly correlated with the value at time 't-1'. Similarly, the significant spike at lag 1 in the PACF plot (fig2) indicates the presence of an autoregressive (AR) component. This points to a potential ARIMA (1,0,1) model.

Performance metrics: After running the ARIMA (1,0,1) model, we found that prediction accuracy was quite low with a low log-likelihood score of -1,503.363 and high AIC (3012.726) and BIC (3020.192) values, indicating a poor fit.

Key Findings: The actual vs. predicted graph (Fig 3) highlights the model's failure to capture the seasonality of Walmart sales, as seen by the smooth predicted line compared to the sharp spikes in actual sales during holiday periods. Our model is making predictions with a

significant deviation from the actual sales, especially in weeks with high variance or spikes in sales. This is apparent by our **RMSE** of over 7 million which indicates that this model is not accurately predicting sales given seasonality.

To address this issue, we introduced a seasonal component by using the SARIMA model.

SARIMA: SARIMA extends the ARIMA model by including the seasonal effects. The auto ARIMA function automatically selected the best combination of p,d, and q with the lowest AIC and identified SARIMA(1,0,0)(0,0,2)[52] as the optimal model for the data.

(1, 0, 0): Represents the non-seasonal part of the model.

p = 1: Indicates one non-seasonal autoregressive term, meaning the current value is influenced by the value from one time period earlier.

d = 0: No diferencing required, as the data is already stationary.

q = 0: No non-seasonal moving average term, indicating the model does not rely on past forecast errors at this level.

(0, 0, 2) [52]: Represents the seasonal part of the model, capturing patterns that repeat every 52 weeks (yearly seasonality).

P = 0: No seasonal autoregressive term, suggesting no dependence on past seasonal values.

D = 0: No seasonal diferencing needed, as the seasonal pattern is stable over time.

Q = 2: Two seasonal moving average terms, accounting for the errors from 52 weeks (one year) and 104 weeks (two years) ago. This helps the model adjust for recurring seasonal fluctuations.

Performance metrics: The SARIMA model scored a log likelihood score of -1,525.363 and high AIC (3061.954) and BIC (3074.508) values, indicating a less optimal model. The Jarque-Bera test (1271.17, p-value = 0.00) further suggested non-normally distributed residuals, highlighting the model's struggle to capture extreme variations in the data.

Key Findings: The actual vs. predicted graph (fig4) highlights that the SARIMA model better captures the seasonality of Walmart sales compared to the initial ARIMA model. However, the model still struggles to accurately predict high outliers specifically around Black Friday and Christmas but performs well in periods of less volatility as can be seen in the second half of the data.

The SARIMA model performs better when compared to ARIMA and Auto-ARIMA for predicting sales data and especially well in sections where volatility is not as drastic, especially near the end of the dataset. The autoregressive term and seasonal moving average terms are key

contributors to predicting the data, which is shown by the significance of the AR terms where the p-value is less than 0.001. Unfortunately, like the ARIMA and Auto-ARIMA, the SARIMA model fails to capture the large spikes in sales, especially around the holiday periods.

SARIMAX: To better predict the sharp fluctuations in sales, especially during these holiday periods, we would recommend exploring the SARIMAX model which incorporates exogenous variables such as promotions and economic indicators to capture the external influences on Walmart sales. However, due to resource constraints and the high computational demands of this model, we were unable to implement it for this analysis. Investing in more robust computational equipment could enable us to fully leverage SARIMAX's capabilities for more accurate and comprehensive forecasting in the future.

Recommendation:

To address the rising need for eficiency and efectiveness in the supply chain we would encourage the company to focus on instilling 3 strategic initiatives in their supply chain operations.

First, Walmart must deepen the application of the SARIMA model. As mentioned in our results, incorporating a SARIMA model provides improved power of sales predictability than ARIMA and Auto-ARIMA models. We believe the further extension of this model both vertically, through feature engineering to add more features, and horizontally, applying the model on a store level to forecast sales on a store level, will add stability and depth to the dataset. Increasing the number of variables would also enable us to consider and evaluate more robust models such as SARIMAX, allowing for more precise and adaptive forecasting.

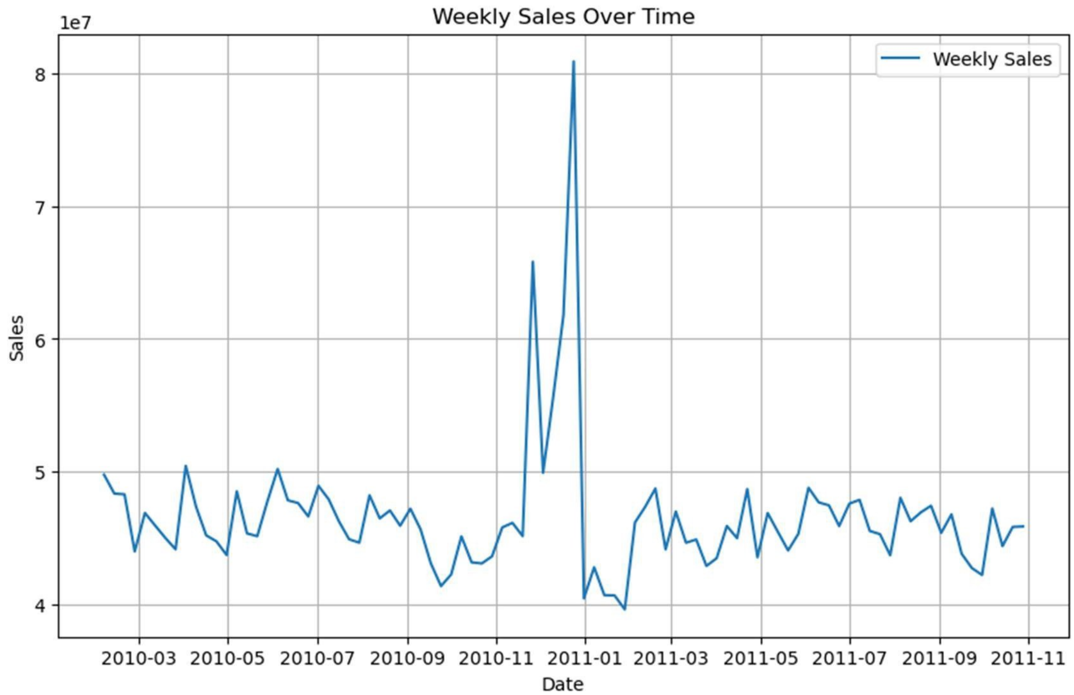
Second, Walmart must maintain additional stock to deal during months in which there is a high risk of stockouts. November and December are key months in which there has traditionally been a surge in customer demand. Therefore, by maintaining a bufer of additional stocks Walmart can not only meet customer demands but also serve additional customers who couldn't find products at a competing retailer's store.

Third, Walmart must establish partnerships among its retail stores to add an element of agility to its supply chain. This strategy will allow retailers to communicate and replenish stocks by leveraging extra stock at another fellow retailer's store. Establishing such a network will have synergistic value, a benefit that will lead to cost eficiency and robust flexibility in inventory management.

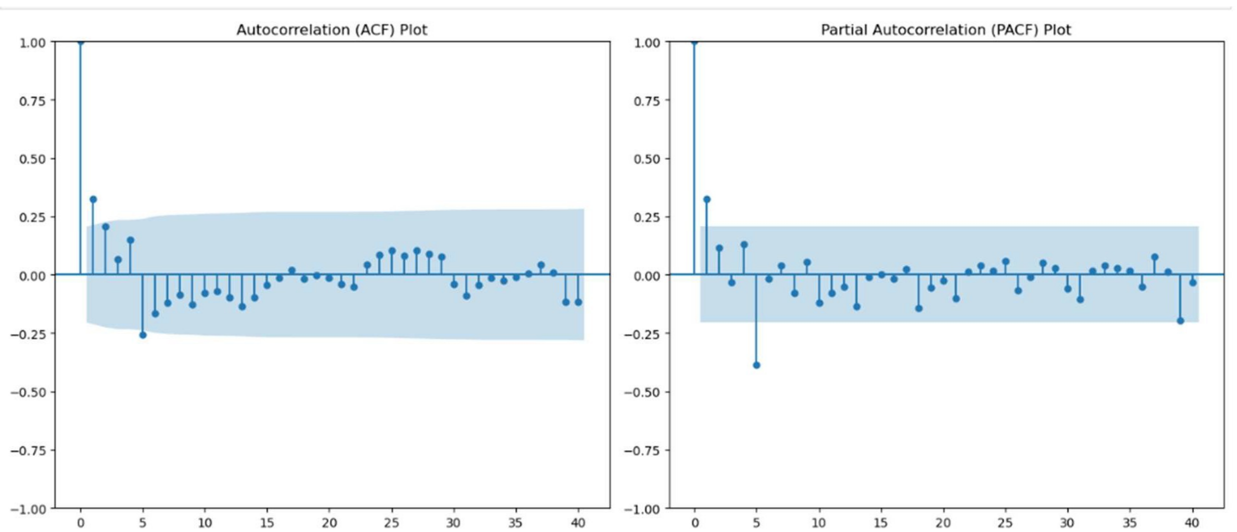
To summarize, Walmart can enhance its supply chain and overcome the current challenges of stockouts and unproductive inventory by following our three recommended strategies.

Walmart must deepen the application of our SARIMA model, maintain additional stock, and establish partnerships among its retail partners.

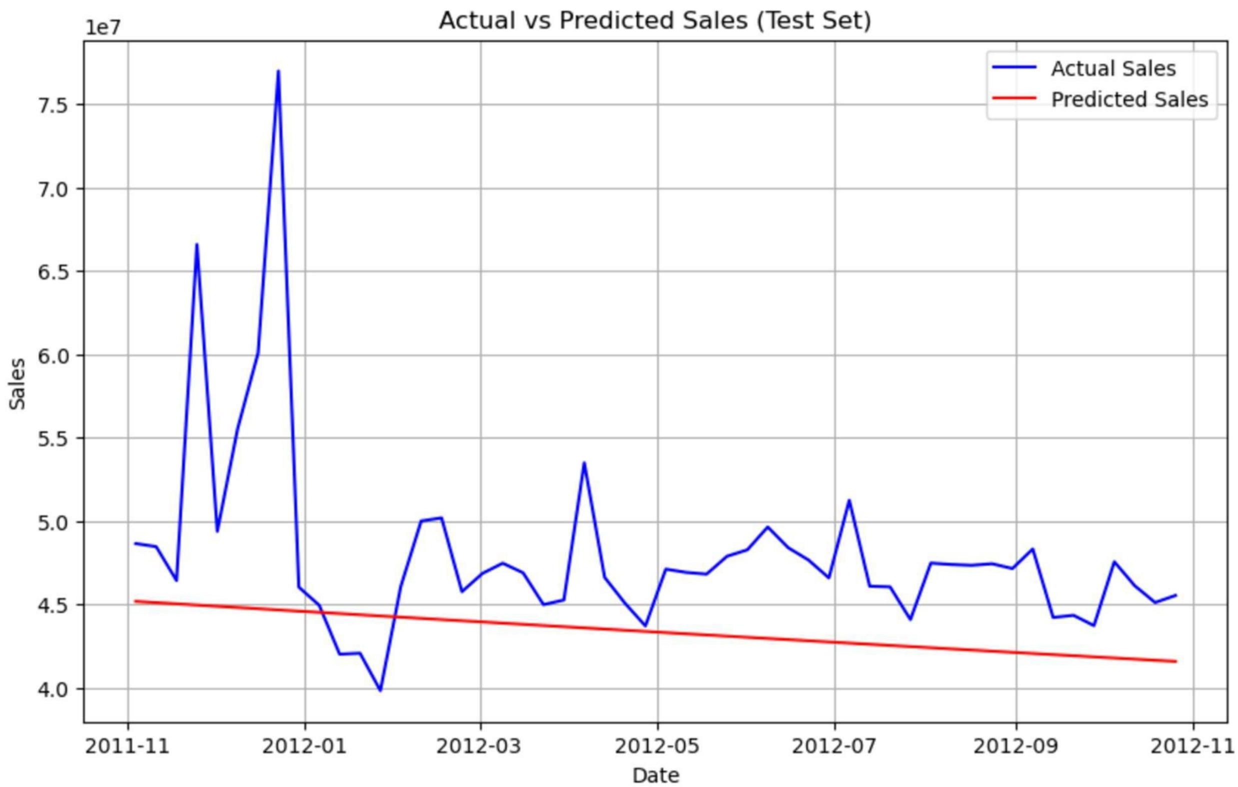
Appendix:



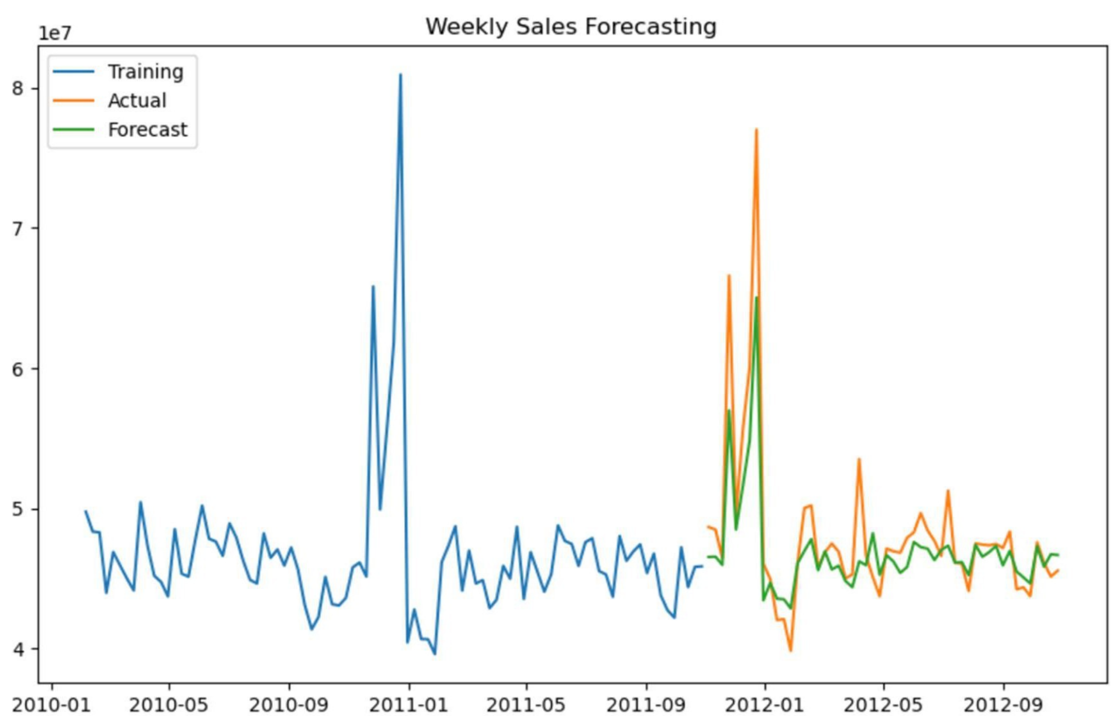
(fig1)



(fig2)


Actual vs Predicted Sales (Test Set)
1e7
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
Sales
Actual Sales
Predicted Sales
2011-11
2012-01
2012-03
2012-05
2012-07
2012-09
2012-11
Date

(fig3)


Weekly Sales Forecasting
1e7
8
7
6
5
4
Training
Actual
Forecast
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09

(fig4)

Anderson, G. (2014, April 15). *Walmart sees $3b opportunity in fixing out-of-stocks*. RetailWire. https://retailwire.com/discussion/walmart-sees-3b-opportunity-in-fixing-out-of-stocks/

Walmart Inc. (2020). *Walmart Annual report 2000*. https://s201.q4cdn.com/262069030/files/doc_financials/2000/ar/2000-annual-report-for-walmart-stores-inc_130199460083383743.pdf

Walmart Inc. (2024). *Annual report 2024*. https://s201.q4cdn.com/262069030/files/doc_financials/2024/ar/2024-annual-report-pdf-final-final.pdf