5-2024

# IMPACT OF SEASONALITY ON DEMAND FORECASTING TECHNIQUES FOR SMALL-SCALE FOOD RETAILERS

INDU SREE GUTURU

IMPACT OF SEASONALITY ON DEMAND FORECASTING TECHNIQUES

FOR

SMALL-SCALE FOOD RETAILERS

---

A Project

Presented to the

Faculty of

California State University,

San Bernardino

---

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

in

Information Systems and Technology

---

by

Indu Sree Guturu

May 2024

# IMPACT OF SEASONALITY ON DEMAND FORECASTING TECHNIQUES

# FOR

# SMALL-SCALE FOOD RETAILERS

A Project

Presented to the

Faculty of

California State University,

San Bernardino

by

Indu Sree Guturu

May 2024

Approved by:

Dr.Conrad Shayo, Committee Chair and Member, Information Systems and Technology

Prof.William Butler, Committee Member, Information Systems and Technology

ABSTRACT

This study explores the impact of seasonality on SKU (Stock Keeping Unit) demand forecasting in small-scale food retailers through the analysis of historical sales data, focusing on four key products, labelled: Product A, Product B, Product C and Product D. The research aims to assess the accuracy of different forecasting models in capturing seasonal fluctuations. The research questions are: (Q1) What is the impact of seasonality on SKU demand forecasting in case of small-scale food retailers. (Q2) How does the efficacy of Traditional model compare to alternative technologically driven forecasting techniques? (Q3) To what extent can the impact of seasonality on SKU demand forecasting be differentiated between perishable and non-perishable goods? The data collected is from a local food retailer of LA county from the year 2021-2023. The collected dataset was broken down to weekly forecast sets to match our methodologies pick of small-time intervals as opposed to yearly or quarterly forecasts. The findings are:(Q1) There is a significant influence of seasonality on demand forecasting accuracy. Models incorporating seasonal components outperform those that do not. Specifically, the Winter-Holt's model, integrating seasonal components, yields more accurate forecasts for both Product A and Product B compared to the traditional Holt's model. (Q2) SARIMA (Seasonal Autoregressive Integrated Moving Average) emerges as the preferred forecasting model, outperforming techniques like Moving Average, Simple Exponential Smoothing, and ARIMA (Autoregressive Integrated Moving Average) for Product

A. SARIMA's superiority lies in its adept handling of seasonal dynamics and identification of stationarity within the data. (Q3) further differentiation between perishable and non-perishable goods reveals SARIMA's superiority in capturing seasonal demand patterns for non-perishable products like Product A. This study acknowledges limitations, such as reliance on past data and the exclusion of external factors impacting demand patterns. Future research should adopt a holistic approach, integrating historical sales data with external variables like market trends and consumer behavior. By embracing a multidimensional perspective and harnessing advancements in technology, future studies can pave the way for more sophisticated forecasting methodologies, enhancing overall performance and customer satisfaction in small-scale food retail environment.

# ACKNOWLEDGEMENTS

I extend my sincere gratitude to the faculty of California State University, San Bernardino, particularly the Department of Information Systems and Technology (IST), for their invaluable support.

I am deeply thankful to Dr. Conrad Shayo, Chair of the IST Department, for his encouragement and insightful feedback throughout this research. His expertise has been instrumental.

I also appreciate the contributions of Prof. William Butler, a dedicated Committee Member, whose feedback enriched the quality of this study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

# CHAPTER ONE: INTRODUCTION

## Demand Forecasting

In small-scale food retail, seasonality significantly affects SKU demand forecasting (Sriboonchitta,2019). This research aims to enhance understanding of this relationship. Seasonality, with its recurring demand variation, poses challenges for SKU demand forecasting, requiring accurate predictions to optimize inventory and meet customer expectations (Ziel, ,2022). In continuation to "Application of SARIMAX Model to Forecast Daily Sales in Food Retail Industry" (Fernandes,2016), this research aims to enhance understanding of the relationship between seasonality, external factors, and SKU demand forecasting in small-scale food retail (Sriboonchitta,2019). The sector's complexity demands a tailored approach to forecasting, with predictive data science playing a crucial role (Kolassa,2022). As per "Forecasting: Theory and Practice", Traditional methods often struggle to capture seasonality's subtleties, leading to overstocking and stockouts. Advanced forecasting models are proposed to address these issues of promotional Campaigns, Categories etc., however Seasonality remains as a future study topic (Ampountolas,2021) (Dunlea,2024). While causal models outperform basic benchmarks, there's limited evidence on the effectiveness of machine learning methods. Forecasting new products also lacks substantial evidence on diverse methods' efficacy according to Findings (Armstrong,2016).

Conventional data science forecasting techniques, such as multiple regression, exponential smoothing, the Holt-Winters model (also known as seasonal exponential smoothing), ARIMA, supervised regression and classification models, random forest, gradient boosting, and stochastic optimization, are frequently used for predicting food demand (Aalst, 2016; Fernandes, 2016). However, these methods have several drawbacks. For example, they have a relatively short "life cycle," lack the ability to adapt and learn, and struggle to cope with the volatile nature of the food market, leading to decreased usefulness of historical data for making forecasts (Blanch, 2009). Additionally, traditional methods are not capable of generalizing, meaning their predictions are only valid for a specific period. As the dynamics of the food market change, it becomes necessary to develop new models to maintain accurate predictions.

To overcome the limitations of traditional models an innovative hybrid forecasting model has been developed, seamlessly blending the strengths of linear autoregressive integrated moving average (ARIMA) and nonlinear artificial neural network (ANN) models (Cheikhrouhou,2019). Tailored to address the dynamic nature of small-scale food retail, this model incorporates volatility through a moving-average filter, demonstrating superior prediction accuracy across diverse experimental datasets (Udokwu,2022) However, the model's primary focus lies in the perishable goods sector, where the challenge of forecasting daily food sales is particularly formidable due to high volatility and skewness. To address this challenge, the study introduces the seasonal autoregressive integrated moving average with external variables (SARIMAX) model, aiming to surpass mean forecasts and provide precise predictions for the perishable goods sector, However, the same remains to be seen in terms of other forecasting techniques. ( Fernandes, 2022)

In "Food sales prediction: ‘If only it knew what we know", The major drawback of hybrid models/software is their high cost and lack of sustainability, especially with changing seasonality trends like during COVID-19. Small-scale retailers often lack funding and knowledge of advanced AI techniques, so it remains to explored as to which method has maximum accuracy (Pechenizkiy,2019) The objective is to identify cost-efficient models that consider both temporal considerations and seasonal influences (Sriboonchitta,2019). This study examines Winter-Holt's model for forecasting time series data with variations in both seasonality and trend, comparing it with  non-seasonal ARIMA model (Qi,2015). While these models typically use lagged values as inputs, incorporating processed variables like moving averages and annual seasonal indexes could enhance their prediction quality. This paper proposes the Winter-Holt's model for small-scale food retailers, integrating moving averages and annual seasonal indexes for local food retailer forecasting Additionally, the study highlights the issue of food wastage in catering services and suggests using machine learning models to improve food demand predictions. Implementing random forest and long short-term memory neural network models led to substantial reductions in wasted meals (ranging from 14% to 52%) (Schnorrenberger,2023). This research underscores the potential benefits of advanced forecasting techniques in reducing the environmental, social, and economic impacts of food waste.

This Culminating Experience Project seeks to explore the influence of seasonality on the precision of SKU demand forecasting within the realm of small-scale food retailers, with a specific focus on the application of conventional methodologies such as Winter-Holt's model as suggested (Boylan,2018). Further research is to compare accuracies of different models in presence of seasonality (Darbanian,2023).Additionally, the study aims to compare the efficacy of Winter-Holt's model against alternative forecasting techniques, including ARIMA, SARIMA  to identify which model exhibits superior accuracy in SKU demand forecasting for small-scale food retailers, especially when confronted with the challenges posed by seasonality(Mafakheri,2020). Furthermore, the research aims to discern whether the impact of seasonality on SKU demand forecasting can be distinguished among various categories of perishable and non-perishable goods within the small-scale food retail sector (Sriboonchitta,2019) Furthermore, the impact of seasonality on SKU demand forecasting for perishable goods has been done in European region. There is scope to replicate the same in North American region i.e, us and see if the results are identical (Garcia, 2020).

Based on the above problem statement of finding the best suited Time-series forecasting method to determine the SKU demand forecasting in presence of seasonality, the three research questions that this study aims to answer are:

(Q1) What is the impact of seasonality on SKU demand forecasting for small-scale food retailers?  (Machado, 2024), and (Boylan, J. E. 2023)

(Q2) How does the efficacy of Traditional model compare to alternative technologically driven forecasting techniques? (Kolassa, 2022), (Lachhab, A. 2018) and (Ampountolas, 2021)

(Q3) To what extent can the impact of seasonality on SKU demand forecasting be differentiated between perishable and non-perishable goods? (Bhardwaj, 2019), and (García, 2020)

All the mentioned references are available on IEEE, International Journal of Forecasting, Internation Journal of Information and Decision making and International Journal of Engineering business Management.

# CHAPTER TWO: LITERATURE REVIEW

## Background

In this chapter, we will study the research done so far and determine the approach and gaps in research to provide context for our project. We will discuss the relationship between seasonality and SKU demand forecasting accuracy within small-scale food retailers which is recommended as further study. We will explore how traditional forecasting methods, such as simple exponential smoothing and Holts' method struggle with seasonality, leading to challenges in predicting seasonal patterns has been done so far and determine how we can contribute to the same. Additionally, we will also discuss the impact of including external variables into forecasting models to improve accuracy. Furthermore, we will discuss traditional and advanced forecast methods and highlight the evolving landscape of demand forecasting. Through these discussions, we aim answer the challenges and opportunities in SKU demand forecasting, particularly in the context of limited historical data and evolving market dynamics in the food retail industry

Q1. How does seasonality affect the accuracy of SKU demand forecasting within small-scale food retailers? (Machado,2024) , (Boylan, 2023)

"Demand forecasting with high dimensional data: The case of SKU retail sales forecasting with intra- and inter-category promotional information," promotions were used a factor to Forecast demand and further research is required to address how Seasonality plays a crucial role in the accuracy of SKU demand forecasting within small-scale food retailers (Huang,2016). In these settings, long time series data are often scarce due to short product life cycles and evolving market dynamics. Traditional methods like exponential smoothing are popular due to their intuitive nature and ease of implementation, especially in businesses lacking extensive statistical expertise. However, these methods may struggle with short data histories and evolving seasonality (Boylan, 2023).

Despite the progress made, there are still opportunities for further research. Investigating the effectiveness of more restrictive methods on shorter data histories, exploring shrinkage of parameters, and examining seasonality at higher frequencies could deepen our understanding and improve forecasting accuracy in small-scale food retail settings (Machado,2024). The aim of this question is to develop a time series forecasting model that includes uncertainty in forecasts and accounts for the impact of external variables such as day-of-the-week seasonality, month-of-the-year seasonality, holidays, festivals, and price reductions outlined in “A hybrid seasonal autoregressive integrated moving average and quantile regression for daily food sales forecasting" (Ampountolas,2021).

Moreover, the study explores the impact of external factors, such as promotions, on forecasting accuracy. It finds that incorporating promotions into the forecasting models, especially with the SARIMA method, significantly improves results (Machado,2024). The study also highlights the importance of understanding the characteristics of the data and business constraints when selecting forecasting methods. While SARIMA models require a stronger statistical background, WH models rely more on computational performance. Additionally, the study suggests that implementing forecasting models at the level of single stores, while challenging, could be improved by using machine learning methods and hybrid analysis approaches. (Boylan,2023) and (Enaben ,2019).

Q2. How does the efficacy of Traditional model compare to alternative technologically driven forecasting techniques? (Kolassa,2022), (lachhab,2018) and (Ampontolas,2021)

The comparison between traditional forecasting models and alternative technologically driven techniques is a burgeoning area of research within forecasting. Traditional models, such as the Box-Jenkins method, Winter-Holts have been foundational in forecasting but may have limitations in accuracy when compared to newer, technologically driven approaches (Ziel, 2022). Data mining, for instance, utilizes computational algorithms to extract patterns from vast datasets, showing promise in improving accuracy, especially for complex data (Dunlea,2024). Similarly, neural networks, a type of machine learning algorithm, offer adaptability to changing patterns and have shown potential for surpassing traditional models in accuracy. Additionally, prediction markets, which aggregate information from groups to forecast events, are emerging as a novel alternative, although their effectiveness in forecasting requires further exploration (Lachhab, 2018). These comparisons underscore the need for continued research to ascertain the optimal approach for different forecasting contexts (Kolassa ,2022) and (Ampountolas,2021). This study aims to explore and provide recommendations for future research in the field of predictive analytics for demand forecasting in retail supply chain management (SCM) (Mafakheri, 2020).

Q3. To what extent can the impact of seasonality on SKU demand forecasting be differentiated between perishable and non-perishable goods? (Bharadwaj,2019), (Garcia,2020)

Demand forecasting with high dimensional data: The case of SKU retail sales forecasting with intra- and inter-category promotional information, Seasonality significantly influences SKU demand forecasting, especially in the food retail industry, where accurate predictions are crucial for managing perishable goods effectively (Bharadwaj,2019). Research is required to emphasize the challenges posed by seasonality, noting that perishable food items, such as bananas, exhibit high volatility and skewness in sales patterns. These characteristics vary over time and require specialized forecasting models to account for them (Garcia,2020) (Kenton,2019and Lachhab,2018). Application of SARIMAX model to forecast daily sales in food retail industry proposes a SARIMAX model that combines seasonal autoregressive integrated moving average (SARIMA) and external variables to forecast daily sales of perishable foods (Bharadwaj, 2019). Taking external variables is optional in these settings. This approach highlights the need to differentiate between perishable and non-perishable goods in demand forecasting, as the shelf-life constraints and seasonal demand variations of perishable items significantly impact forecasting accuracy in the European region (Garcia, 2020). Further study remains to be done in Northern American region with the same settings.

In contrast, non-perishable goods may exhibit less pronounced seasonality in demand forecasting. However, it is yet to be determined. While they can still be affected by seasonal trends, the impact is likely to be less volatile and more predictable compared to perishable items. However, the literature suggests that the extent of differentiation between perishable and non-perishable goods in demand forecasting depends on various factors, including product characteristics, market dynamics, and forecasting techniques employed (Sriboonchitta,2019). Further research is needed to explore these differences comprehensively and develop tailored forecasting approaches for different product categories. To address the research recommendation outlined in the paper "A hybrid seasonal autoregressive integrated moving average and quantile regression for daily food sales forecasting," there is a critical need to develop a time series forecasting model. This model should effectively incorporate uncertainties in forecasts and consider the impact of various external variables. These variables include day-of-the-week seasonality, month-of-the-year seasonality, holidays, festivals, price reductions, and weather conditions on sales. (Garcia,2020)

# CHAPTER THREE: RESEARCH METHODS

The central objective of this research paper is to identify the most effective forecasting method for predicting SKUs in small-scale food retail establishments. This study addresses three specific research questions aimed at achieving this objective. The data used for this analysis is sourced from local resources covering the period from 2021 to 2023 and focuses on four selected products. Multiple forecasting models are employed to analyze the data and generate insights for the study's conclusions (Schnorrenberger ,2023)

## Q1. How does seasonality affect the accuracy of SKU demand forecasting within small-scale food retailers

To facilitate comparisons, this study aims to answer a specific question by applying Holts' and the Winter-Holts model to four distinct products: A and B .As per statistical analysis of the dataset which spans over multiple time intervals of monthly data, it was found that the data does not have central tendency and has significant increase over a period which allows us to identify trend and seasonality correlation using traditional methods rather than advanced methods( Lutoslawksy,2021). Holts is the top choice which incorporates trends in demand forecasting. To answer our research question, we are engaging Winter-holts as an extension to Holts model for predicting in presence of seasonality (Charles, 2021). The initial set of model runs will exclude the seasonality index.

Subsequently, a second set of runs will include the seasonality parameters for the same products. The outcomes of these runs will then be compared, with a focus on Accuracy and Error as key evaluation criteria (Mafakheri, 2020). Other attributes like MAPE, MAD, MSE are also considered for optimizing the model.

Holts Model:

Holt's linear trend model, which is also known as double exponential smoothing, with trend but no seasonal component. The model starts iteratively with only two components: level and trend. level, and trend. Level establishes a baseline around the variance in data whereas trend captures the variation in changes over time. The limitation of the model is that it does not account for seasonality and other cyclicity components in the data.  Below are the equations:

Level equation: $\alpha Y_t + (1-\alpha)(L_{t-1} + T_{t-1})$

Trend equation: $T_t = \beta(L_t - L_{t-1}) + (1-\beta) T_{t-1}$ Where:

$Y_t$ is the observed value at time t,

$L_t$ is the level at time t,

$T_t$ is the trend at time t,

$\alpha$ is the smoothing parameter for the level ($0 \le \alpha \le 1$),

$\beta$ is the smoothing parameter for the trend ($0 \le \beta \le 1$).

Winter - Holts Model:

The Winter-Holt model, or the Winter-Holt model, is a forecasting technique that predicts future values based on past data. The trend model includes seasonal components. When the model is run without considering seasonality, it essentially focuses on capturing the trend and level components of the time series data. This can be useful when the data does not exhibit clear seasonal patterns or when seasonality is not a significant factor in the forecasting process (Kolassa,2022). This model also includes variations such as additive and multiplicative seasonality, and methods for handling different types of trends, which are either linear or exponential. These parameters should be set depending on the type of data being analyzed. Holt-Winters method involves three smoothing equations:

Level equation: $L_t = \alpha (Y_t - S_{t-m}) + (1-\alpha)(L_{t-1} + T_{t-1})$

Trend equation: $T_t = \beta (L_t - L_{t-1}) + (1-\beta) T_{t-1}$

Seasonal equation: $S_t = \gamma (Y_t - L_t) + (1-\gamma) S_{t-m}$

$Y_t$ is the observed value at time t,

$L_t$ is the level at time t,

$T_t$ is the trend at time t,

$S_t$ is the seasonal component at time t,

m is the length of the seasonal cycle (e.g., 12 for monthly data),

$\alpha$, $\beta$, and $\gamma$ are smoothing parameters ($0 \leq \alpha, \beta, \gamma \leq 1$).

Q2. How does the efficacy of Traditional model compare to alternative technologically driven forecasting techniques?

This question is intended to be answered by comparing 3 Winter-Holts, ARIMA and SARIMA models. (Shah ,2022) Since Winter-Holts and SARIMA has already been discussed above, diving into ARIMA mode. Products A and B are utilized for this analysis. All three forecasting models—Winter-Holts, ARIMA, and SARIMA—are employed to predict the SKUs of both products. This approach allows for a comparison of their efficiencies. Winter-Holts represent the conventional forecasting technique. ARIMA involves exponential smoothing without considering seasonality and trends. SARIMA extends ARIMA by incorporating seasonality into the model.

ARIMA Model without Seasonality:

The Autoregressive Integrated Moving Average (ARIMA) model is a popular time series forecasting model that combines autoregressive (AR)(p), differencing (I)(q), and moving average (MA)(d) components. The ARIMA model is used to predict future values based on past observations, making it suitable for analyzing and forecasting time series data.

Auto-Regressive Component: It captures the linear relationship between an observation and its lagged component

$$Yt=c+\phi 1Yt-1+\phi 2Yt-2+\ldots+\phi pYt-p+\epsilon t$$

$Yt$ is the value of the time series at time t, c is a constant term,

$\phi_1, \phi_2, \ldots, \phi_p$ are the parameters of the model representing the effect of past values on the current value,

$\epsilon_t$ is white noise, representing random fluctuations in the data.

Moving Average (MA) Component: The moving average component models the relationship between the current observation and previous observation, also known as lag. An MA(q) process of order q can be represented as: $Y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q}$

Where: μ is the mean of the series, $\theta_1, \theta_2, \ldots, \theta_q$ are the parameters of the model representing the weights of past error terms and μ is the mean of the series, $\theta_1, \theta_2, \ldots, \theta_q$ are the parameters of the model representing the weights of past error terms.

Integration (I) Component: The integration component is used to make the time series stationary by differentiating the data. An integrated of order d, denoted as I(d), is applied to remove trends or non-stationarity in the series: $Y_t' = Y_t - Y_{t-1}$

Combining AR, I, and MA: The ARIMA model combines these components into a single model. An ARIMA (p, d, q) model consists of:

AR(p): Autoregressive component of order p.

I(d): Integration component of order d for differencing.

MA(q): Moving average component of order q.

The general equation for an ARIMA (p, d, q) model can be written as:

$Y_t' = c + \phi_1 Y_{t-1}' + \phi_2 Y_{t-2}' + \ldots + \phi_p Y_{t-p}' + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q} + \epsilon_t$

Q3. To what extent can the impact of seasonality on SKU demand forecasting be differentiated between perishable and non-perishable goods?

The dataset includes four different products: A, B, C, and D. Products A and B are perishable goods, while products C and D are non-perishable goods. The SARIMA (Seasonal Autoregressive Integrated Moving Average) model is employed to forecast the sales of all four products. The results are compared to determine which product has the best accuracy in forecasting. ( García,2020). The objective is to assess how the inclusion of a seasonality index affects the accuracy of forecasting for perishable and non-perishable goods, providing insights into the impact of seasonality on different types of products.

SARIMA MODELLING:

The Seasonal Autoregressive Integrated Moving Average (SARIMA) model is an extension of the ARIMA model that incorporates seasonality. It is used to forecast time series data that exhibit seasonal patterns. The SARIMA model is defined by three main components: the seasonal component(P) the autoregressive component(D), and the moving average component(Q).

1.Autoregressive (AR)Component:

Similar to ARIMA, the autoregressive component captures the linear relationship between an observation and its lagged component. In addition to that

SARIMA, this component will include seasonality as well. An AR process with an order p is represented as

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \theta_1 \epsilon_{t-1} + \epsilon_t$$

$Y_t$ is the value of the time series at time t, c is a constant term, $\phi_1$ ,$\phi_2$ ,$\phi_p$ are the autoregressive parameters, $\theta_1$ is the moving average parameter, $\epsilon_t$ is the error term.

Moving Average (MA) Componet:

Similar is what we do in ARIMA, the moving average component is relationship between the current and the previous lagged component error item. In SARIMA this component accounts for seasonality too. An MA process with an order q can be represented as

$$Y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q}$$

$\mu$ is the mean of the series, $\theta_1$ ,$\theta_2$ ,…,$\theta_q$ are the moving average parameters.

Integration (I) Component:

Again, like ARIMA, the integration component is used to make the time series stationary by differencing the data.

Seasonal Component: The SARIMA model introduces additional seasonal components to account for periodic fluctuations in the data. This includes seasonal autoregressive terms (SAR), seasonal differences (seasonal differencing), and seasonal moving average terms (SMA). The seasonal component can be represented as:

$$Y_t = c + \Phi_1 Y_{t-m} + \Phi_2 Y_{t-2m} + \ldots + \Phi_P Y_{t-pm} + \Theta_1 \epsilon_{t-m} + \Theta_2 \epsilon_{t-2m} + \ldots + \Theta_Q \epsilon_{t-qm} + \epsilon_t$$

where, m represents the seasonal period (e.g., 12 for monthly data), $\Phi_1, \Phi_2, \ldots \Phi_P$ are the seasonal autoregressive parameters, $\Theta_1, \Theta_2, \ldots, \Theta_Q$ are the seasonal moving average parameters.

<u>Combining AR, I, MA, and Seasonal Components:</u> The SARIMA model also combines these components into a single model. A SARIMA (p, d, q) (P, D, Q) m model consists of:

ARIMA (p, d, q) for the non-seasonal part.

Seasonal components (P, D, Q) m for the seasonal part

# CHAPTER FOUR: DATA COLLECTION, ANALYSIS AND FINDINGS

In this chapter, analysis carried out on the sales data collected from a local small-scale food retailer with six branches in LA county will be discussed. Since the data was not segmented, we have segregated and pooled in on a product and monthly level. The study is limited to 4 products which were chosen due to their steady demand pattern and have no irregularities like intermittent or lumpy demand. Product A has mean sales of $137,792 with a standard deviation of $91,784.14 with indicates a variation in sales. Upon visual inspection, the data overall has a period over period increase in sales. Product B has $15934.22 standard deviation, which means the data is not showing any central tendency, rather it is distributed across period over period. Also, product C and D shows similar statistical significance. Since the data is spread across multiple periods identifications of trends seasonality and patterns can be identified. Moreover, concentrating on these high performing products to provide insights into demand forecasting strategies. The sample data for the mentioned product can be found in the Appendix.

Q1. How does seasonality affect the accuracy of SKU demand forecasting within small-scale food retailers

To answer this research question with regards to impact of seasonality on demand forecasting, Winter-Holts model was used to products A, B. The initial simulation runs on Holts which excluded seasonality, followed by a second iteration we included these parameters. The outcomes were compared based on accuracy and error metrics such as MAPE, MAD, and MSE below.

1.1.1 Holts model for product A:

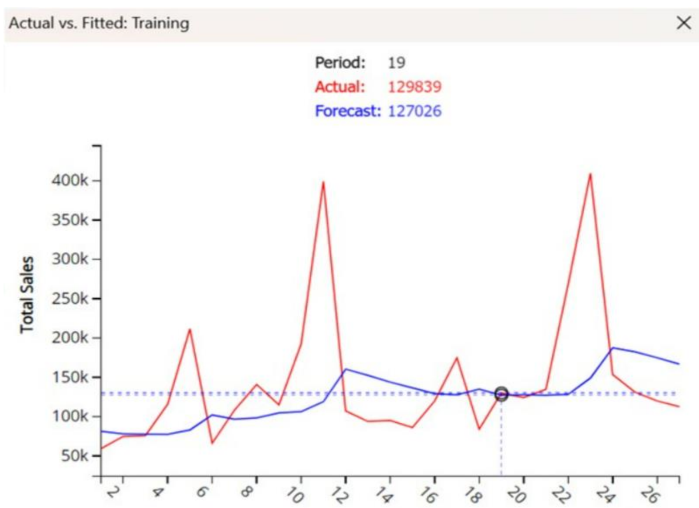



Figure 1. Total Sales vs Forecasted Sales – A using Holt's Method

The above forecast methodology does not consider seasonality component, considering the parameters this model has MAPE of 35% and MAD of 58845.87

1.1.2 Winters-Holts model for Product A:

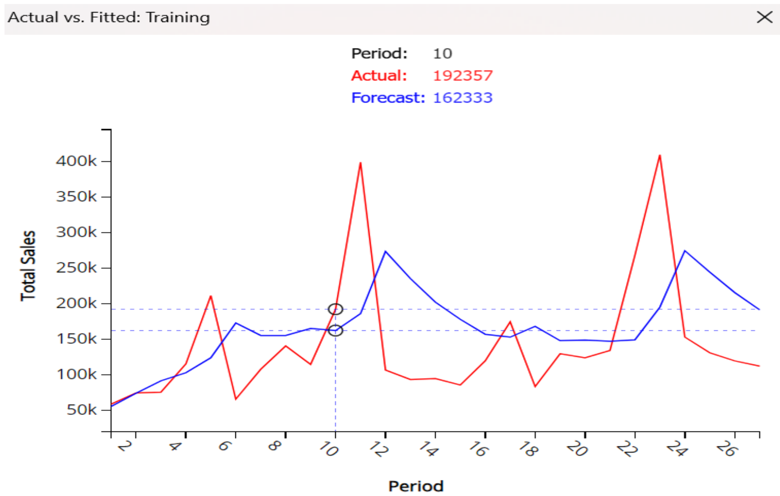



Figure 2. Total Sales vs Forecasted Sales – A using Winter-Holt's Method

The above model is a multivariate additive winter holts' model which accounts for trends and seasonality. MAPE is 57%, and MAD is 75342.05

## 1.2.1 Holts Model for Product B:

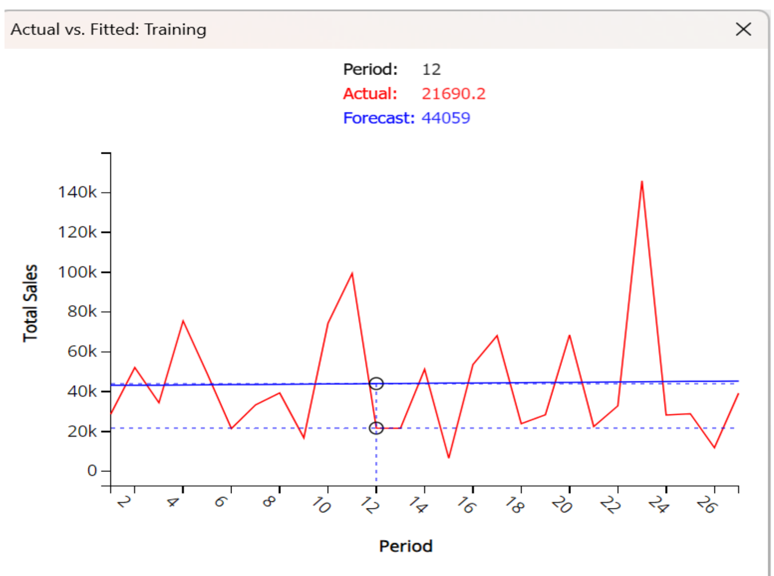



Figure 3. Total Sales vs Forecasted Sales for Product – B using Holt's Method

The above model incomplete does not incorporate any trend or seasonality, the model fails to predict future sales. MAPE is 81.91%, and MAD is 22618.04

<u>1.2.2 Winters Holts for Product B:</u>

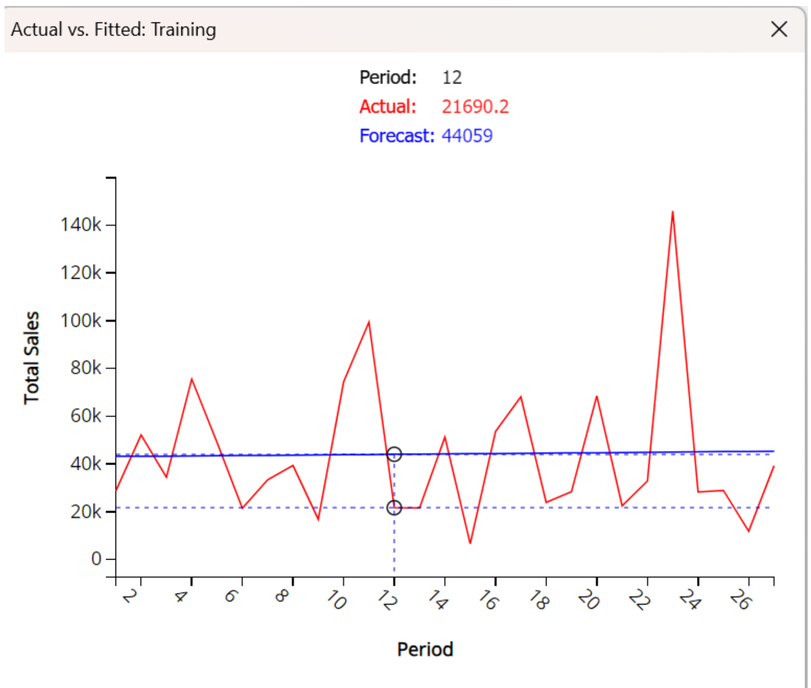



Figure 4. Total Sales vs Forecasted Sales – C using Winter - Holt's Method

MAPE of above multi-variate 113.48% and MAD 30214.52 using multivariate Winters-Holt’s method.

Q2. How does the efficacy of Traditional model compare to alternative technologically driven forecasting techniques?

For research question 2, we compare efficacy of traditional models versus alternative technologically driven forecasting techniques, we compared the Winter-Holts, ARIMA, and SARIMA models. Products A and D were used for this analysis. The results of these models were compared to determine their efficiencies in forecasting SKUs.

3.1.1 Product A – ARIMA :

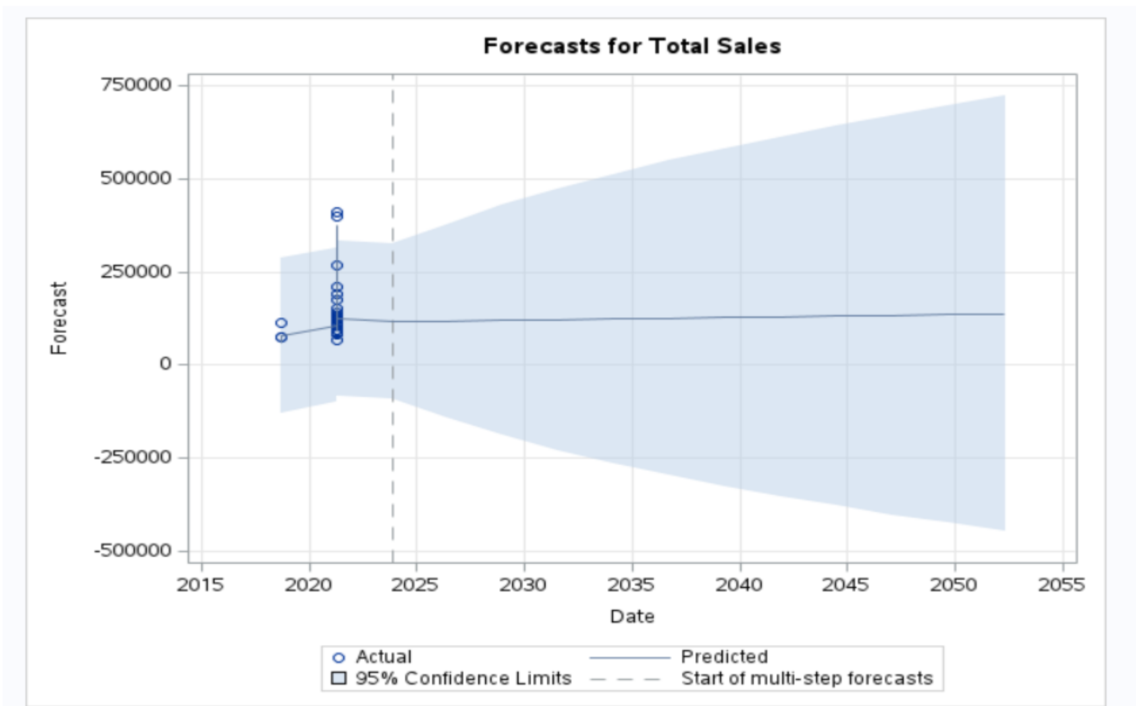



Figure 5. Total Sales vs Forecasted Sales for Product – A using ARIMA

| Forecasts for variable Total Sales | | | | |
|---|---|---|---|---|
| Obs | Forecast | Std Error | 95% Confidence Limits | |
| 28 | 116803.9 | 105869 | -90695.0 | 324302.8 |
| 29 | 118225.6 | 132074 | -140635.2 | 377086.3 |
| 30 | 120401.1 | 157498 | -188288.8 | 429091.0 |
| 31 | 122385.0 | 178528 | -227523.9 | 472293.9 |
| 32 | 124417.6 | 197518 | -262711.5 | 511546.7 |
| 33 | 126437.9 | 214793 | -294547.7 | 547423.4 |
| 34 | 128461.2 | 230788 | -323874.1 | 580796.6 |
| 35 | 130483.8 | 245741 | -351160.2 | 612127.9 |
| 36 | 132506.6 | 259836 | -376763.3 | 641776.6 |
| 37 | 134529.4 | 273205 | -400942.8 | 670001.6 |
| 38 | 136552.1 | 285950 | -423898.7 | 697002.9 |
| 39 | 138574.9 | 298150 | -445787.9 | 722937.6 |

| Outlier Detection Summary | |
|---|---|
| Maximum number searched | 1 |
| Number found | 1 |
| Significance used | 0.05 |

| Outlier Details | | | | |
|---|---|---|---|---|
| Obs | Type | Estimate | Chi-Square | Approx Prob>ChiSq |
| 11 | Additive | 263724.3 | 30.27 | <.0001 |

Table 1: Forecast Data Product A ARIMA

The above forecast data is a summarized table with chi-square 30.27 and significance level of 95%.The above iteration of ARIMA simulation has the AR, I, and MA parameters set at (1,1,0). As per ACF, Lag 0 = 1 and Lag 1 = -0.25; Data was stationary Residuals are normally distributed between -3E5 and 300000 as per Figure. Chi Square is 30.27 and approximate Probability of Chi Square is <0.0001

<u>2.1.2 Product A – SARIMA: Refer section 3.1</u>

<u>2.1.3 Product A – Winter Holts: Refer section 1.1.2</u>

<u>2.2.1 Product D – ARIMA</u>

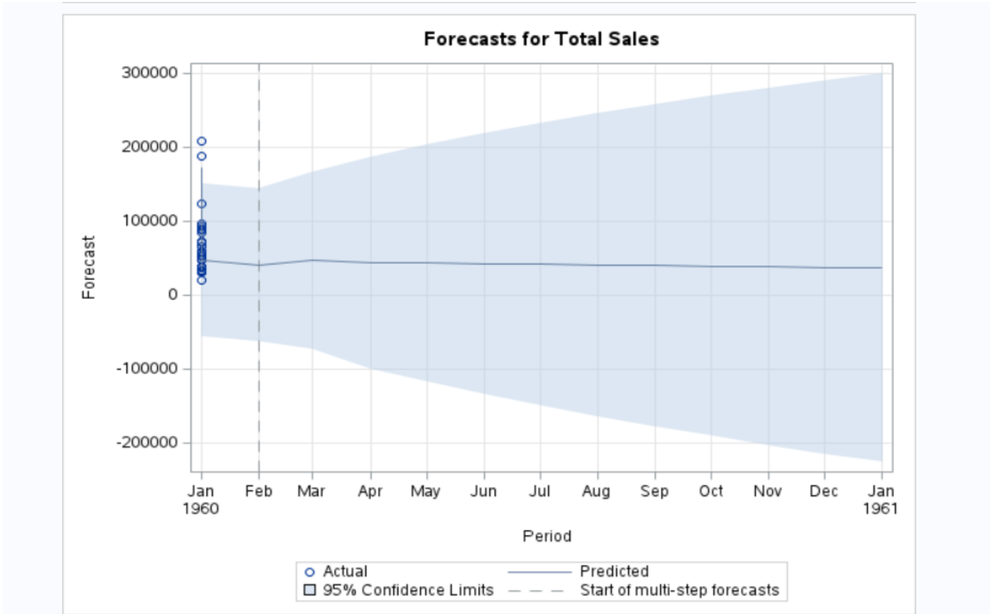



Figure 6. Total Sales vs Forecasted Sales for Product – D using ARIMA

| Forecasts for variable Total Sales | | | | |
|---|---|---|---|---|
| Obs | Forecast | Std Error | 95% Confidence Limits | |
| 28 | 116803.9 | 105869 | -90695.0 | 324302.8 |
| 29 | 118225.6 | 132074 | -140635.2 | 377086.3 |
| 30 | 120401.1 | 157498 | -188288.8 | 429091.0 |
| 31 | 122385.0 | 178528 | -227523.9 | 472293.9 |
| 32 | 124417.6 | 197518 | -262711.5 | 511546.7 |
| 33 | 126437.9 | 214793 | -294547.7 | 547423.4 |
| 34 | 128461.2 | 230788 | -323874.1 | 580796.6 |
| 35 | 130483.8 | 245741 | -351160.2 | 612127.9 |
| 36 | 132506.6 | 259836 | -376763.3 | 641776.6 |
| 37 | 134529.4 | 273205 | -400942.8 | 670001.6 |
| 38 | 136552.1 | 285950 | -423898.7 | 697002.9 |
| 39 | 138574.9 | 298150 | -445787.9 | 722937.6 |

| Outlier Detection Summary | |
|---|---|
| Maximum number searched | 1 |
| Number found | 1 |
| Significance used | 0.05 |

| Outlier Details | | | | |
|---|---|---|---|---|
| Obs | Type | Estimate | Chi-Square | Approx Prob>ChiSq |
| 11 | Additive | 263724.3 | 30.27 | <.0001 |

Table 2: Forecast Data Product D ARIMA

The forecast data has chi-square value of 30.27 with 95% confidence level and forecast sales around 138574 units. The above iteration of ARIMA simulation has the AR, I, and MA parameters set at (1,1,0).

As per ACF, Lag 0 = 1 and Lag 1 = -0.5; Data was stationary

Residuals are normally distributed between -15E4 and 150000 as per Figure 23.

Chi Square is 16.03 and Approximate Probability of Chi Square is <0.0001

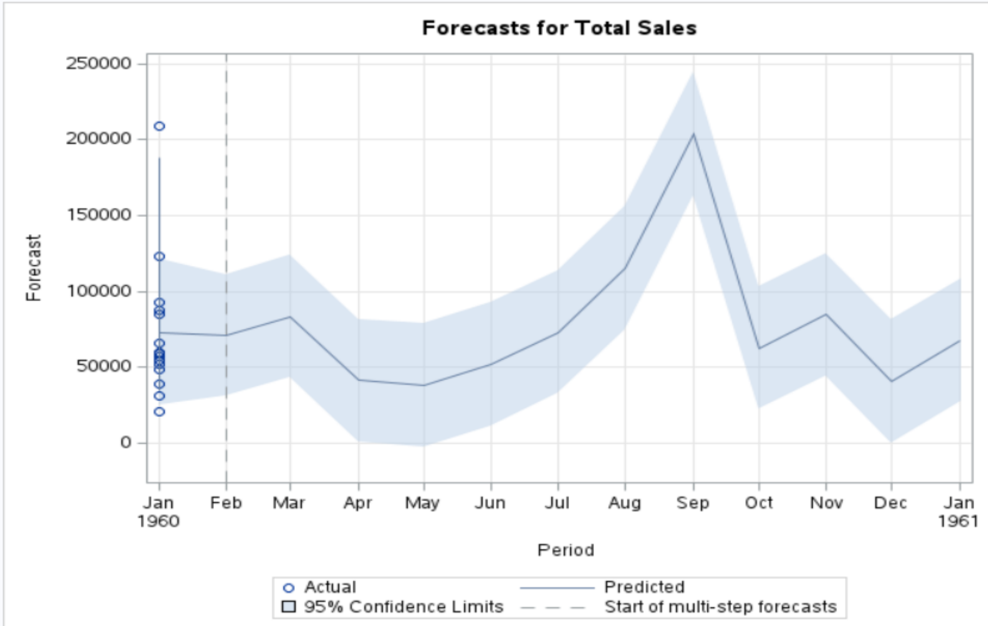
### 3.2.2. Product D SARIMA

| Autoregressive Factors | |
|---|---|
| Factor 1: | 1 - 0.1 B**(12) |

| Moving Average Factors | |
|---|---|
| Factor 1: | 1 - 0.14021 B**(1) |
| Factor 2: | 1 - 0.99988 B**(12) |

| Forecasts for variable Total Sales | | | | |
|---|---|---|---|---|
| Obs | Forecast | Std Error | 95% Confidence Limits | |
| 28 | 70879.6648 | 20499.492 | 30701.3978 | 111057.9317 |
| 29 | 83473.5239 | 20700.018 | 42902.2334 | 124044.8144 |
| 30 | 41137.2304 | 20700.018 | 565.9398 | 81708.5209 |
| 31 | 38026.7823 | 20700.018 | -2544.5082 | 78598.0729 |
| 32 | 51856.9499 | 20700.018 | 11285.6593 | 92428.2404 |
| 33 | 73031.3347 | 20700.018 | 32460.0442 | 113602.6252 |
| 34 | 115244.5500 | 20700.018 | 74673.2594 | 155815.8405 |
| 35 | 203885.0929 | 20700.018 | 163313.8024 | 244456.3835 |
| 36 | 62658.9569 | 20700.018 | 22087.6663 | 103230.2474 |
| 37 | 84583.8851 | 20700.018 | 44012.5946 | 125155.1756 |
| 38 | 40635.9385 | 20700.018 | 64.6480 | 81207.2290 |
| 39 | 67735.5041 | 20700.018 | 27164.2136 | 108306.7947 |

Table 3: Forecast Data for Product A- SARIMA




| Outlier Detection Summary | |
|---|---|
| Maximum number searched | 1 |
| Number found | 1 |
| Significance used | 0.05 |

| Outlier Details | | | | |
|---|---|---|---|---|
| Obs | Type | Estimate | Chi-Square | Approx Prob>ChiSq |
| 4 | Additive | -59166.7 | 6.18 | 0.0129 |

Figure 7. Total Sales vs Forecasted Sales for Product – D using SARIMA

The actual sales lie in the brackets of predicted sales. The forecast data has chi-square value of 6.9 with 95% confidence level and forecast units with std error 207000.

The above iteration is for the SARIMA model with (001,111)

AR = 0, I = 0, MA = 1

SAR = 1, SI = 1, SMA =1

ACF Lag difference between 1st and 0th is 1.5; Data is stationary

Residual are normally distributed between +/- 60000 in Fig 27.

Chi Square value is 6.18 and it is probability is 0.0129; Significant

Product D: Y = 3092.3 + 0.14*error(t-1) + 0.999*error(t-2)

3.2.3 Product D – Winter Holts

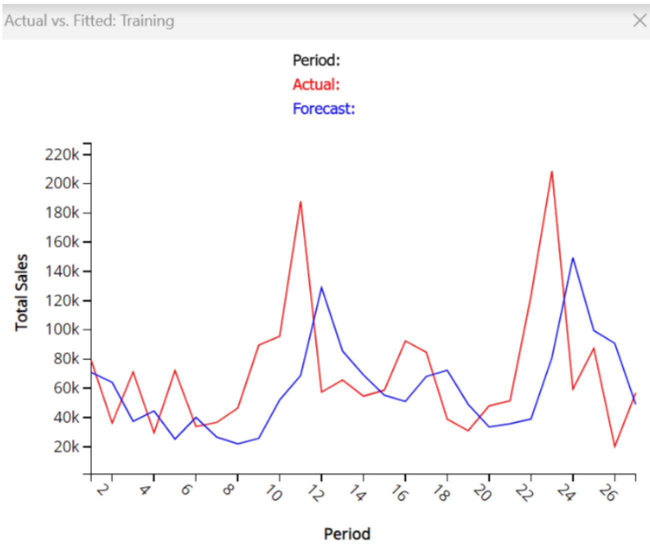



Figure 8. Total Sales vs Forecasted Sales Product – D (Winter-Holt's method);

Simulation results: MAPE 60%, MAD 38544.75

Q3. To what extent can the impact of seasonality on SKU demand forecasting be differentiated between perishable and non-perishable goods?

There has been research done to predict SKU for short- lifecycle goods, however the same has not been replicated for non-perishable food products which this question aims to do. The idea is to compare if the same results are obtained when SARIMA is used for both. Impact of seasonality is additionally explored for perishable goods to see how it effects. To answer research question 2, which investigates the differentiation of the impact of seasonality on SKU demand forecasting between perishable and non-perishable goods, we utilized the SARIMA model to forecast the sales of products A, C. Products A is perishable goods, while products C is considered non-perishable. By comparing the forecasting accuracy of these products, we aimed to assess the impact of seasonality on different types of goods.

Trend Analysis: Trend analysis is a technique that is used to analyze and identify patterns of data over time. It determines if there is any long-term upward, downward, or stable movement in the data. This analysis is useful for forecasting future trends or making informed decisions based on historical data.

Correlation Analysis: Correlation analysis examines the strength and direction of the relationship between two or more variables. It measures how changes in one variable are associated with changes in another variable. It identifies whether there is a linear relationship between variables and the extent to which they move together.

## 2.1 Product A SARIMA:

| Name of Variable = Total Sales | |
|---|---|
| Period(s) of Differencing | 12 |
| Mean of Working Series | 19787.85 |
| Standard Deviation | 25064.4 |
| Number of Observations | 15 |
| Observation(s) eliminated by differencing | 12 |

Table 4: Differencing Summary Table Product A SARIMA

The above forecast methodology does not consider seasonality component, considering the parameters this model has MAPE of 35% and MAD of 58845.87

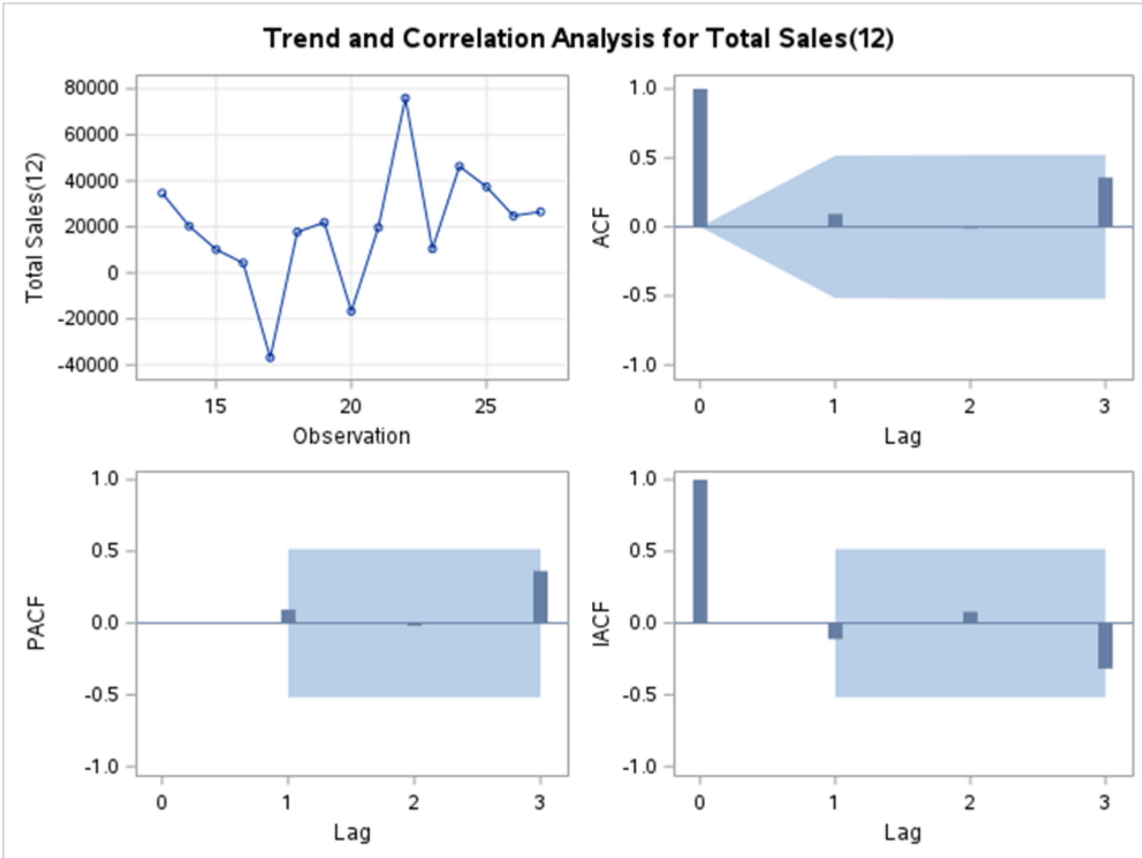



Figure 9: Trend and Correlation Analysis Product A SARIMA Method

| Maximum Likelihood Estimation | | | | | |
|---|---|---|---|---|---|
| **Parameter** | **Estimate** | **Standard Error** | **t Value** | **Approx Pr > \|t\|** | **Lag** |
| **MU** | 18390.1 | 3766.9 | 4.88 | <.0001 | 0 |
| **AR1,1** | 0.93109 | 0.04262 | 21.85 | <.0001 | 12 |

| | |
|---|---|
| **Constant Estimate** | 1267.171 |
| **Variance Estimate** | 1.1406E8 |
| **Std Error Estimate** | 10680.12 |
| **AIC** | 346.9091 |
| **SBC** | 348.3252 |
| **Number of Residuals** | 15 |

| Correlations of Parameter Estimates | | |
|---|---|---|
| **Parameter** | **MU** | **AR1,1** |
| **MU** | 1.000 | 0.078 |
| **AR1,1** | 0.078 | 1.000 |

| Autocorrelation Check of Residuals | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **To Lag** | **Chi-Square** | **DF** | **Pr > ChiSq** | **Autocorrelations** | | | | | |
| **6** | 3.16 | 5 | 0.6748 | 0.103 | 0.040 | 0.341 | -0.104 | -0.059 | -0.068 |

| Autocorrelation Check of Residuals | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **To Lag** | **Chi-Square** | **DF** | **Pr > ChiSq** | **Autocorrelations** | | | | | |
| **12** | 10.45 | 11 | 0.4904 | -0.355 | -0.057 | -0.016 | -0.246 | -0.022 | -0.014 |

Table 5: Characteristics Summary Table Product A SARIMA

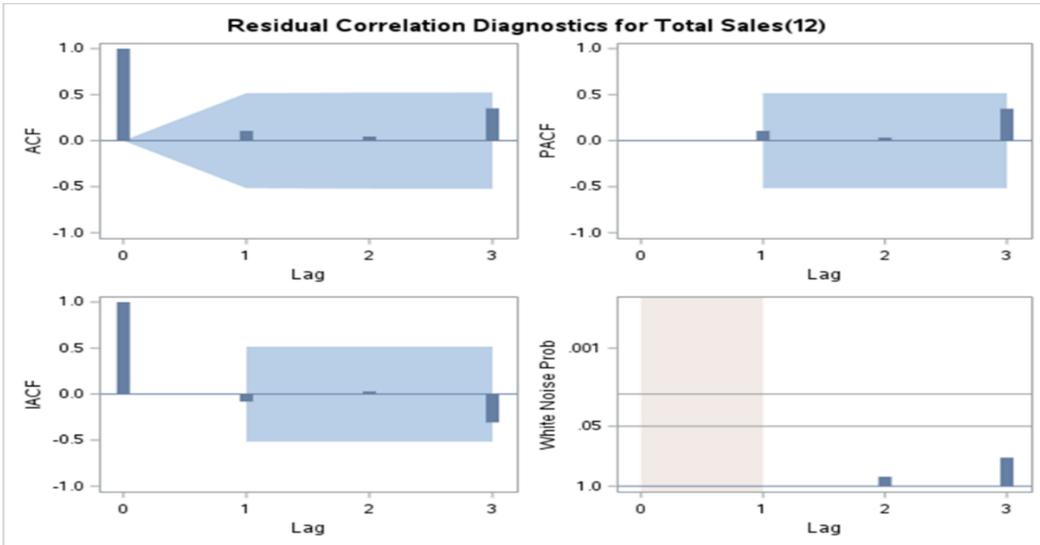



Figure 10. Residual Correlation Diagnostics: Total Sales Product A SARIMA

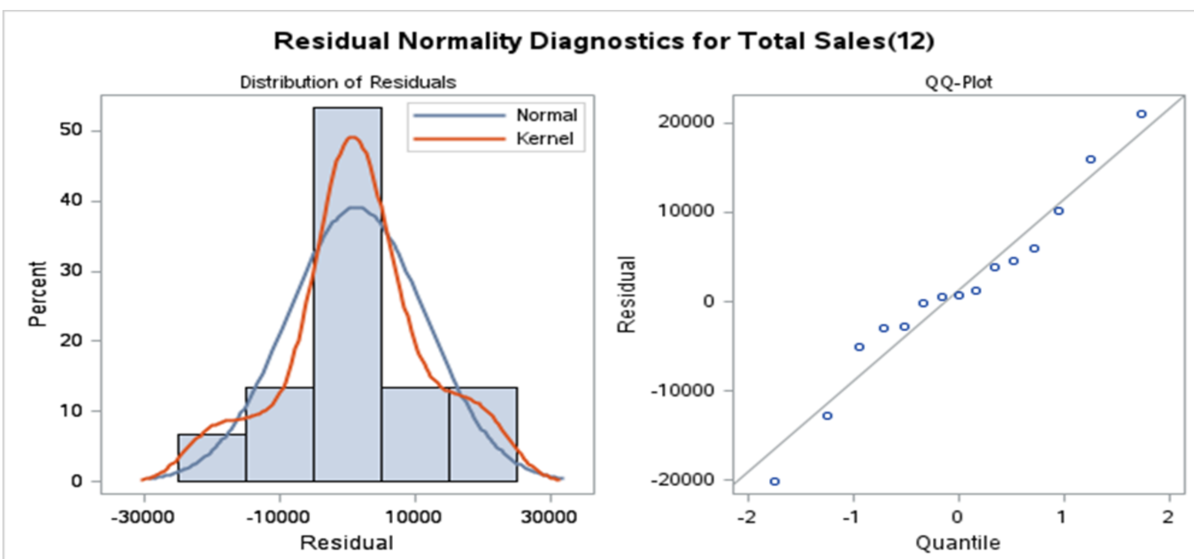



Figure 11. Residual Normality Diagnostics for Total Sales Product A

The above iteration is for the SARIMA model with (000,110)

AR = 0, I = 0, MA = 0

SAR = 1, SI = 1, SMA = 0

ACF Lag difference between 1st and 0th is 0.9; Data is stationary

Residual are normally distributed between +/- 30000 in Fig 7.

Chi Square value is 6.18 and it is probability is 0.0021; Significant

Product A: $Y = 18390 + 0.93109 * \text{error}(t-1)$

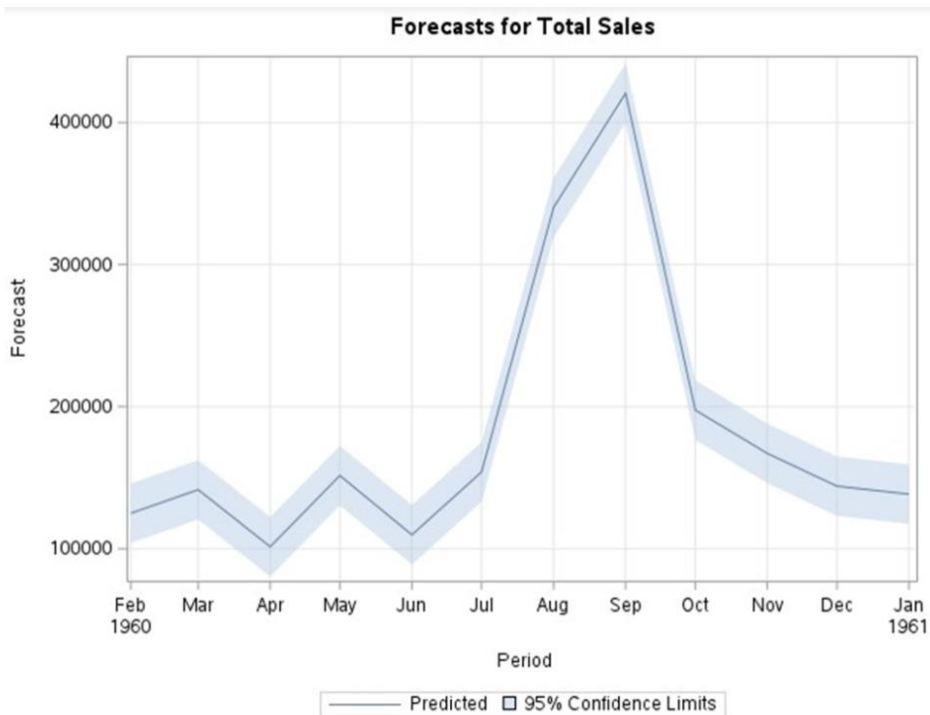


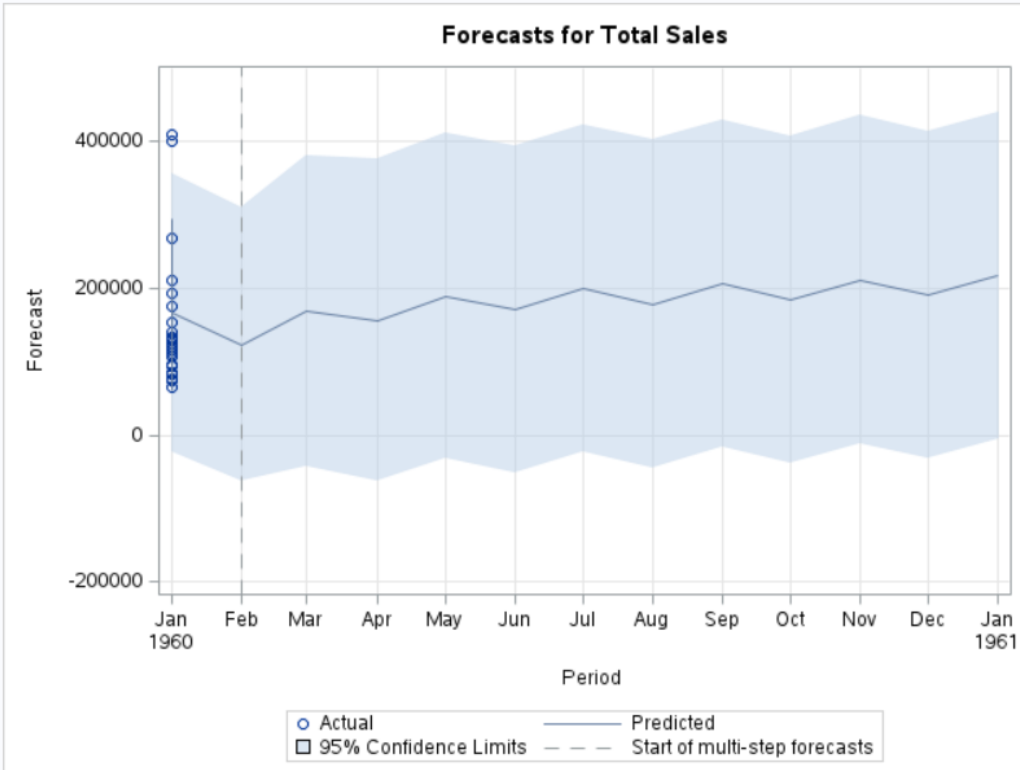

Figure 12. Total Sales vs Forecasted Sales for Product – A using SARIMA

| Forecasts for variable Total Sales | | | | |
|---|---|---|---|---|
| **Obs** | **Forecast** | **Std Error** | **95% Confidence Limits** | |
| **28** | 125012.5929 | 10680.121 | 104079.9394 | 145945.2463 |
| **29** | 141577.0354 | 10680.121 | 120644.3819 | 162509.6889 |
| **30** | 101442.3636 | 10680.121 | 80509.7101 | 122375.0171 |
| **31** | 151524.6680 | 10680.121 | 130592.0145 | 172457.3214 |
| **32** | 109810.9581 | 10680.121 | 88878.3046 | 130743.6115 |
| **33** | 154055.2394 | 10680.121 | 133122.5860 | 174987.8929 |
| **34** | 340236.1393 | 10680.121 | 319303.4859 | 361168.7928 |
| **35** | 420621.7235 | 10680.121 | 399689.0701 | 441554.3770 |
| **36** | 197600.3595 | 10680.121 | 176667.7061 | 218533.0130 |
| **37** | 167136.0652 | 10680.121 | 146203.4118 | 188068.7187 |
| **38** | 144042.3412 | 10680.121 | 123109.6877 | 164974.9946 |
| **39** | 138427.7332 | 10680.121 | 117495.0798 | 159360.3867 |

Table 6: Forecast Data Product A SARIMA

The above table is the summarized forecast data deduced from the above graphs. The standard error is 10680.121 with 95% confidence limits.

## 2.2 Product C- SARIMA:




| Outlier Detection Summary | |
|---|---|
| Maximum number searched | 1 |
| Number found | 1 |
| Significance used | 0.05 |

| Outlier Details | | | | |
|---|---|---|---|---|
| Obs | Type | Estimate | Chi-Square | Approx Prob>ChiSq |
| 11 | Additive | 242458.5 | 14.51 | 0.0001 |

Figure 13. Total Sales vs Forecasted Sales for Product – C using SARIMA

| Forecasts for variable Total Sales | | | | |
|---|---|---|---|---|
| Obs | Forecast | Std Error | 95% Confidence Limits | |
| 28 | 123229.6 | 94622.23 | -62226.5 | 308685.8 |
| 29 | 168163.5 | 107772 | -43065.1 | 379392.1 |
| 30 | 156507.3 | 111911 | -62833.9 | 375848.5 |
| 31 | 188921.1 | 113039 | -32632.0 | 410474.3 |
| 32 | 170431.4 | 113520 | -52062.9 | 392925.7 |
| 33 | 198974.1 | 113631 | -23738.2 | 421686.4 |
| 34 | 178434.4 | 113716 | -44444.4 | 401313.2 |
| 35 | 205752.1 | 113728 | -17151.3 | 428655.5 |
| 36 | 184625.9 | 113757 | -38333.4 | 407585.2 |
| 37 | 211528.3 | 113759 | -11434.9 | 434491.4 |
| 38 | 190263.1 | 113776 | -32733.8 | 413260.1 |
| 39 | 216998.0 | 113777 | -6000.1 | 439996.0 |

Table 7: Forecast Data Product C SARIMA

The above table is the consolidated forecast data of Product C. The standard error is stable at 113777 at 95% confidence level.

The above iteration is for the SARIMA model with (000,110)

AR = 0, I = 0, MA = 0

SAR = 1, SI = 1, SMA = 0

ACF Lag difference between 1st and 0th is 0.9; Data is stationary

Residual are normally distributed between +/- 30000 in Fig.

Chi Square value is 6.18 and it is probability is 0.0021; Significant

Product A: $Y = 18390 + 0.93109 * error(t-1)$

# CHAPTER FIVE: RESULTS

In This chapter delves deeper into the data-driven impact of seasonality on SKU demand forecasting in small-scale food retailers. We present the results of our data analysis, focusing on traditional forecasting methods' struggle with seasonality and the accuracy of technologically driven alternatives like SARIMA. Additionally, we discuss how seasonality affects perishable and non-perishable goods differently and offer insights for future studies.

Q1. How does seasonality affect the accuracy of SKU demand forecasting within small-scale food retailers?

Based on the results obtained for both product A and product B, it can be inferred that seasonality significantly affects the accuracy of SKU demand forecasting within small-scale food retailers.For product A, the Winter-Holt's model as per section 4 1.1.2, which incorporates seasonal components, yielded a higher MAPE and MAD compared to Holt's model as shown in section 4 1.1.1. The Winter-Holt's model provides a more accurate baseline forecast, suggesting the importance of considering seasonality in forecasting.

For product B, Holt's model resulted in a straight-line forecast with an MAPE of 81.91% and MAD of 22618.04, refer chapter 4 section 1.2.1. In contrast, the Winter-Holt's model, using a multivariate approach to predict seasonality, yielded a higher MAPE of 113.48% and MAD of 30214.52. Despite the higher errors, the Winter-Holt's model's ability to capture seasonality and adjust for these variations makes it a better choice for SKU demand forecasting in small-scale food retailers.

These results highlight the importance of considering seasonality in demand forecasting for small-scale food retailers, as ignoring seasonality can lead to less accurate forecasts and potential inventory management issues. By incorporating seasonal components, models like Winter-Holt's can provide more accurate and reliable forecasts, helping retailers optimize their inventory levels and meet customer demand more effectively

<u>Q2. How does the efficacy of Traditional model compare to alternative technologically driven forecasting techniques?</u>

This question deals with comparing three different models: Winter-holts, ARIMA and SARIMA for two different products, A and D. It is found that SARIMA (Seasonal Autoregressive Integrated Moving Average) decisively outperformed other forecasting models including Moving Average, Simple Exponential Smoothing, Holt's, Winter's Holt, and ARIMA for Product A).

Product A's sales pattern follows the seasonal sequence (000) (110), indicating clear fluctuations across different seasons or time periods. This pattern suggests a non-seasonal trend (000) with seasonal variations occurring every two time periods (110).

The superiority of SARIMA is attributed to several concrete factors:

Incorporation of Seasonality: SARIMA adeptly accounts for the seasonal dynamics in the data, essential for accurately forecasting products with seasonal demand. Failure to capture this seasonality often leads to significant forecasting errors.

Identification of Stationarity: Through rigorous analysis including examination of the Autocorrelation Function (ACF) chart and application of differencing techniques, the data for Products was confidently rendered stationary. Stationarity is a fundamental prerequisite for robust time series analysis and forecasting.

Diagnostic Analysis of Residuals: SARIMA models undergo meticulous diagnostic checks of residuals, ensuring that no underlying patterns remain unaddressed. Comprehensive evaluation of residual plots for autocorrelation and normality guarantees that the model captures all pertinent information within the data.

Significance of Model Parameters: SARIMA models meticulously select statistically significant parameters, substantiated through rigorous testing such as the Chi-square test. This ensures that the model parameters are robust and reliable for forecasting. Similar thorough analyses were conducted for Products

D, affirming SARIMA as the optimal forecasting model. The same approach was applied to identify the most suitable SARIMA models for each product, tailored to their unique sales patterns and characteristics.

The final SARIMA model equation for products, as mentioned, stands as:

Product A: $Y = 18390 + 0.93109 * error(t-1)$

Product D: $Y = 3092.3 + 0.14*error(t-1) + 0.999*error(t-2)$

Where Y represents total sales, and the equation adeptly captures the relationship between the forecasted values and the error term at the preceding time step.

The forecast table and graph, detailed in section 2.1, validate the forecast of the SARIMA model by accurately predicting future sales for Product A. These results underscore the critical importance of employing advanced forecasting techniques like SARIMA, particularly when addressing complex time series data characterized by seasonality and other significant trends.

Q3. To what extent can the impact of seasonality on SKU demand forecasting be differentiated between perishable and non-perishable goods?

After analyzing the impact of seasonality on demand forecasting, with a focus on perishable products, Product A and non-perishable Product C, in continuation of the research conducted (García, 2020), we have arrived at significant findings which are similar to the findings in European region.

In our iterative forecasting process utilizing SARIMA models, Product A has demonstrated superior forecast accuracy when employing the SARIMA model. This indicates that the seasonal components of Product A's demand patterns are well captured by the SARIMA framework, allowing for more precise predictions.

Conversely, Product C, a non-perishable item, has shown reasonable forecast accuracy when utilizing Winter's Holt method. While SARIMA and ARIMA models were also tested, they did not yield substantial results compared to the traditional Winter's Holt approach.

Based on these findings, we recommend the utilization of traditional forecasting methods, such as simple moving average, exponential smoothing for Product C. Despite the potential benefits offered by advanced models like SARIMA, the incremental improvement in forecast accuracy may not justify the additional complexity and computational resources required.

However, for Product A, which exhibits significant seasonality and demand fluctuations, the SARIMA model emerges as the preferred choice due to its ability to effectively capture the seasonal patterns and provide more accurate forecasts. These insights underscore the importance of tailoring forecasting methodologies to the specific characteristics of the products under consideration.

Conclusions and Area’s of Future Studies:

This section will discuss the findings from the research outlined above, addressing research questions and suggesting area of further studies.

Q1. How does seasonality affect the accuracy of SKU demand forecasting within small-scale food retailers?

In conclusion, the present study has shed light on the critical role of seasonality in SKU demand forecasting within small-scale food retailers, emphasizing the importance of employing techniques like SARIMA and Winter-Holts to achieve accurate predictions. (Mafakheri,2020) Accuracy has significantly increased when seasonality was taken into consideration as (Kolassa, S,2022). Further studies must concentrate on looking at the effectiveness of different inventory management strategies, such as just-in-time (JIT) or economic order quantity (EOQ), in mitigating the effects of seasonality on inventory performance. Understanding the relationship between seasonality and inventory management could help small-scale food retailers optimize their stocking strategies and improve overall operational efficiency.

Q2. How does the efficacy of Traditional model compare to alternative technologically driven forecasting techniques?

In conclusion, while our study has provided valuable insights into the impact of seasonality on SKU demand forecasting that SARIMA offers the most accurate forecast model, there remains a significant scope for further exploration. Furthermore, the efficacy of traditional methods versus technologically driven approaches has been thoroughly examined, with SARIMA emerging as the superior choice for capturing complex time series patterns. (Shah,2022) However, by embracing a multidimensional approach and leveraging advancements in technology, future studies can contribute to the development of more sophisticated forecasting methodologies tailored to the unique challenges faced by small-scale food retailers in today's dynamic marketplace.

Moreover, exploring the applicability of emerging technologies such as machine learning and artificial intelligence in demand forecasting could yield valuable insights into improving forecast accuracy and responsiveness to dynamic market conditions.

Q3. To what extent can the impact of seasonality on SKU demand forecasting be differentiated between perishable and non-perishable goods?

The impact is minimal on perishable goods as compared to non-perishable goods. It is essential to acknowledge the limitations of our study. Primarily, our analysis has been restricted to past data, overlooking potential external factors such as market trends, economic conditions, and consumer behavior shifts. While our findings provide valuable insights into the impact of seasonality on perishable and non-perishable goods, (Seifert, 2022) further studies are warranted to incorporate a more comprehensive analysis encompassing these additional variables.

Future research around demand forecasting should include independent variables in the model, i.e. considering not only historical sales data but also external factors like sales promotions, competitive landscape, demographics, economic factors, cultural and social factor and market trends that may influence demand patterns, for instance in the United States Super Bowl has significant impact in Beer sales. By integrating these variables into forecasting models, researchers can increase the efficacy and accuracy forecasts, enabling small-scale food retailers to take informed decisions regarding inventory management and day to day chain operations.

# APPENDIX A

# SARIMA AND ARIMA MODELS DATA

## APPENDIX A

Sample Data of Product A:

| Period | Date | Product Line | Total Order | Total Sales |
|---|---|---|---|---|
| 1 | December-20 | Product - A | 631 | 58952.67 |
| 2 | January-21 | Product - A | 750 | 74583.97 |
| 3 | February-21 | Product - A | 643 | 75692.58 |
| 4 | March-21 | Product - A | 1020 | 115468.73 |
| 5 | April-21 | Product - A | 1634 | 211484.15 |
| 6 | May-21 | Product - A | 509 | 65864.04 |
| 7 | June-21 | Product - A | 812 | 107908.78 |
| 8 | July-21 | Product - A | 1231 | 140715.75 |
| 9 | August-21 | Product - A | 1039 | 114727.56 |
| 10 | September-21 | Product - A | 1517 | 192356.81 |
| 11 | October-21 | Product - A | 3784 | 399098.7 |
| 12 | November-21 | Product - A | 830 | 106899.03 |
| 13 | December-21 | Product - A | 889 | 93614.47 |
| 14 | January-22 | Product - A | 779 | 94918.43 |
| 15 | February-22 | Product - A | 842 | 85839.94 |
| 16 | March-22 | Product - A | 1083 | 119754.74 |
| 17 | April-22 | Product - A | 1427 | 174627.19 |
| 18 | May-22 | Product - A | 636 | 83631.76 |
| 19 | June-22 | Product - A | 1223 | 129838.68 |
| 20 | July-22 | Product - A | 972 | 124055.79 |
| 21 | August-22 | Product - A | 1095 | 134436.85 |
| 22 | September-22 | Product - A | 2228 | 268278.59 |
| 23 | October-22 | Product - A | 3644 | 409588.01 |
| 24 | November-22 | Product - A | 1412 | 153211.7 |
| 25 | December-22 | Product - A | 1170 | 131030.77 |
| 26 | January-23 | Product - A | 1007 | 119700.61 |
| 27 | February-23 | Product - A | 1046 | 112415.86 |

## APPENDIX – A CONTD.

Sample Data of Product B

| Period | Date | Product Line | Total Order | Total Sales |
|---|---|---|---|---|
| 1 | December-20 | Product - B | 252 | 28666.7 |
| 2 | January-21 | Product - B | 610 | 52156.22 |
| 3 | February-21 | Product - B | 305 | 34415.92 |
| 4 | March-21 | Product - B | 684 | 75557.79 |
| 5 | April-21 | Product - B | 543 | 48995.81 |
| 6 | May-21 | Product - B | 190 | 21488.96 |
| 7 | June-21 | Product - B | 333 | 33411.38 |
| 8 | July-21 | Product - B | 512 | 39351.32 |
| 9 | August-21 | Product - B | 177 | 16818.73 |
| 10 | September-21 | Product - B | 718 | 74378.24 |
| 11 | October-21 | Product - B | 928 | 99399.84 |
| 12 | November-21 | Product - B | 169 | 21690.17 |
| 13 | December-21 | Product - B | 210 | 21613.39 |
| 14 | January-22 | Product - B | 445 | 51249.32 |
| 15 | February-22 | Product - B | 50 | 6576.5 |
| 16 | March-22 | Product - B | 507 | 53637.81 |
| 17 | April-22 | Product - B | 597 | 68101.98 |
| 18 | May-22 | Product - B | 249 | 23928.81 |
| 19 | June-22 | Product - B | 283 | 28361.51 |
| 20 | July-22 | Product - B | 629 | 68547.41 |
| 21 | August-22 | Product - B | 247 | 22486.42 |
| 22 | September-22 | Product - B | 399 | 32858.3 |
| 23 | October-22 | Product - B | 1528 | 145905.62 |
| 24 | November-22 | Product - B | 298 | 28269.11 |
| 25 | December-22 | Product - B | 307 | 28964.95 |
| 26 | January-23 | Product - B | 184 | 11821.64 |
| 27 | February-23 | Product - B | 399 | 39250.06 |

**APPENDIX- A CONTD.**

Sample Data of Product C:

| Period | Date | Product Line | Total Order | Total Sales |
|---|---|---|---|---|
| 1 | December-20 | Product - C | 170 | 15443.8 |
| 2 | January-21 | Product - C | 409 | 42128.69 |
| 3 | February-21 | Product - C | 148 | 10781.32 |
| 4 | March-21 | Product - C | 512 | 48962.08 |
| 5 | April-21 | Product - C | 390 | 30820.4 |
| 6 | May-21 | Product - C | 359 | 27700.36 |
| 7 | June-21 | Product - C | 202 | 15235.63 |
| 8 | July-21 | Product - C | 559 | 52510.94 |
| 9 | August-21 | Product - C | 140 | 14124.76 |
| 10 | September-21 | Product - C | 592 | 52868.62 |
| 11 | October-21 | Product - C | 904 | 86111.08 |
| 12 | November-21 | Product - C | 503 | 45237.39 |
| 13 | December-21 | Product - C | 256 | 20617.81 |
| 14 | January-22 | Product - C | 474 | 37081.22 |
| 15 | February-22 | Product - C | 423 | 43010.38 |
| 16 | March-22 | Product - C | 735 | 71145.33 |
| 17 | April-22 | Product - C | 483 | 45200.41 |
| 18 | May-22 | Product - C | 426 | 39212.95 |
| 19 | June-22 | Product - C | 224 | 24159.2 |
| 20 | July-22 | Product - C | 265 | 19851.83 |
| 21 | August-22 | Product - C | 285 | 24805.86 |
| 22 | September-22 | Product - C | 460 | 43162.46 |
| 23 | October-22 | Product - C | 810 | 74338.18 |
| 24 | November-22 | Product - C | 350 | 30393.37 |
| 25 | December-22 | Product - C | 146 | 13943.96 |
| 26 | January-23 | Product - C | 354 | 36279.73 |
| 27 | February-23 | Product - C | 221 | 21817.86 |

**APPENDIX – A CONTD.**

Sample Data of Product D

| Period | Date | Product Line | Total Order | Total Sales |
|---|---|---|---|---|
| 1 | December-20 | Product - D | 151 | 11485.95 |
| 2 | January-21 | Product - D | 382 | 32354.04 |
| 3 | February-21 | Product - D | 326 | 31038.48 |
| 4 | March-21 | Product - D | 157 | 11929.22 |
| 5 | April-21 | Product - D | 270 | 23279.42 |
| 6 | May-21 | Product - D | 235 | 24363.63 |
| 7 | June-21 | Product - D | 155 | 16389.91 |
| 8 | July-21 | Product - D | 372 | 35964.95 |
| 9 | August-21 | Product - D | 377 | 35103.15 |
| 10 | September-21 | Product - D | 418 | 40285.92 |
| 11 | October-21 | Product - D | 1146 | 92692.52 |
| 12 | November-21 | Product - D | 219 | 19347.04 |
| 13 | December-21 | Product - D | 81 | 6667.03 |
| 14 | January-22 | Product - D | 122 | 11328.67 |
| 15 | February-22 | Product - D | 155 | 11843.14 |
| 16 | March-22 | Product - D | 192 | 17084.96 |
| 17 | April-22 | Product - D | 379 | 41784.71 |
| 18 | May-22 | Product - D | 311 | 24637.44 |
| 19 | June-22 | Product - D | 147 | 12315.04 |
| 20 | July-22 | Product - D | 306 | 22432.79 |
| 21 | August-22 | Product - D | 89 | 5616.08 |
| 22 | September-22 | Product - D | 509 | 38214.16 |
| 23 | October-22 | Product - D | 570 | 52332.81 |
| 24 | November-22 | Product - D | 143 | 13059.48 |
| 25 | December-22 | Product - D | 357 | 30845.81 |
| 26 | January-23 | Product - D | 309 | 26997.39 |
| 27 | February-23 | Product - D | 276 | 26873 |

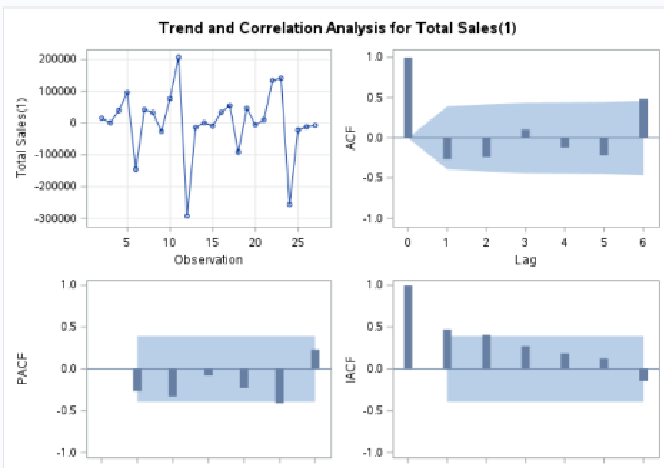
## APPENDIX- A CONTD.

Product A- ARIMA

| Name of Variable = Total Sales | |
|---|---|
| Period(s) of Differencing | 1 |
| Mean of Working Series | 2056.277 |
| Standard Deviation | 105442.8 |
| Number of Observations | 26 |
| Observation(s) eliminated by differencing | 1 |

| Autocorrelation Check for White Noise | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| To Lag | Chi-Square | DF | Pr > ChiSq | Autocorrelations | | | | | |
| 6 | 14.77 | 6 | 0.0222 | -0.264 | -0.239 | 0.104 | -0.122 | -0.216 | 0.484 |

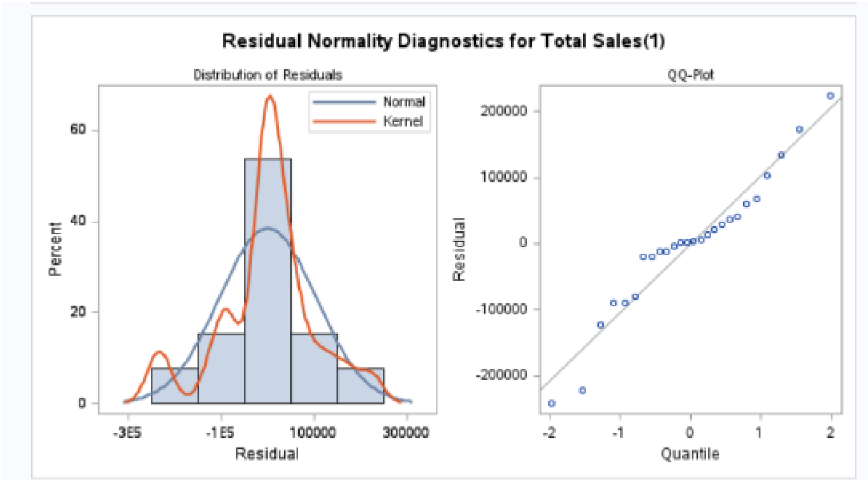
Table 7. Differencing Summary Table Product A ARIMA




| | |
|---|---|
| Constant Estimate | 2536.785 |
| Variance Estimate | 1.121E10 |
| Std Error Estimate | 105868.7 |
| AIC | 677.4081 |

Figure 14: Trend and Correlation Analysis Product A ARIMA Method

# APPENDIX A CONTD.




| Model for variable Total Sales | |
|---|---|
| Estimated Mean | 2022.75 |
| Period(s) of Differencing | 1 |

| Autoregressive Factors | |
|---|---|
| Factor 1: | 1 + 0.25413 B**(1) |

Figure 15. Residual Normality Diagnostics for Total Sales Product A ARIMA

Product C – SARIMA

| Maximum Likelihood Estimation | | | | | |
|---|---|---|---|---|---|
| Parameter | Estimate | Standard Error | t Value | Approx Pr > \|t\| | Lag |
| MU | 2681.9 | 4669.3 | 0.57 | 0.5657 | 0 |
| MA1,1 | 0.0091255 | 3.72740 | 0.00 | 0.9980 | 1 |
| MA1,2 | 0.98165 | 4.27331 | 0.23 | 0.8183 | 2 |
| AR1,1 | -0.44567 | 0.43660 | -1.02 | 0.3074 | 1 |
| AR1,2 | 0.55245 | 0.16390 | 3.37 | 0.0008 | 2 |

| | |
|---|---|
| Constant Estimate | 2395.512 |
| Variance Estimate | 8.9534E9 |
| Std Error Estimate | 94622.23 |
| AIC | 676.8541 |
| SBC | 683.1446 |
| Number of Residuals | 26 |

| Correlations of Parameter Estimates | | | | | |
|---|---|---|---|---|---|
| Parameter | MU | MA1,1 | MA1,2 | AR1,1 | AR1,2 |
| MU | 1.000 | 0.326 | 0.329 | -0.109 | -0.013 |
| MA1,1 | 0.326 | 1.000 | 0.492 | 0.232 | -0.542 |
| MA1,2 | 0.329 | 0.492 | 1.000 | -0.726 | 0.445 |
| AR1,1 | -0.109 | 0.232 | -0.726 | 1.000 | -0.892 |
| AR1,2 | -0.013 | -0.542 | 0.445 | -0.892 | 1.000 |

| Autocorrelation Check of Residuals | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| To Lag | Chi-Square | DF | Pr > ChiSq | Autocorrelations | | | | | |
| 6 | 13.26 | 2 | 0.0013 | -0.094 | -0.306 | 0.151 | -0.246 | -0.177 | 0.418 |
| 12 | 28.32 | 8 | 0.0004 | -0.118 | -0.140 | 0.065 | -0.244 | 0.013 | 0.457 |
| 18 | 35.05 | 14 | 0.0014 | -0.088 | -0.100 | 0.087 | -0.159 | -0.061 | 0.183 |
| 24 | 35.71 | 20 | 0.0166 | -0.026 | -0.042 | 0.019 | -0.039 | -0.013 | -0.001 |

# APPENDIX A CONTD.

| Name of Variable = Total Sales | |
|---|---|
| Period(s) of Differencing | 1 |
| Mean of Working Series | 2056.277 |
| Standard Deviation | 105442.8 |
| Number of Observations | 26 |
| Observation(s) eliminated by differencing | 1 |

| Autocorrelation Check for White Noise | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| To Lag | Chi-Square | DF | Pr > ChiSq | Autocorrelations | | | | | |
| 6 | 14.77 | 6 | 0.0222 | -0.264 | -0.239 | 0.104 | -0.122 | -0.216 | 0.484 |

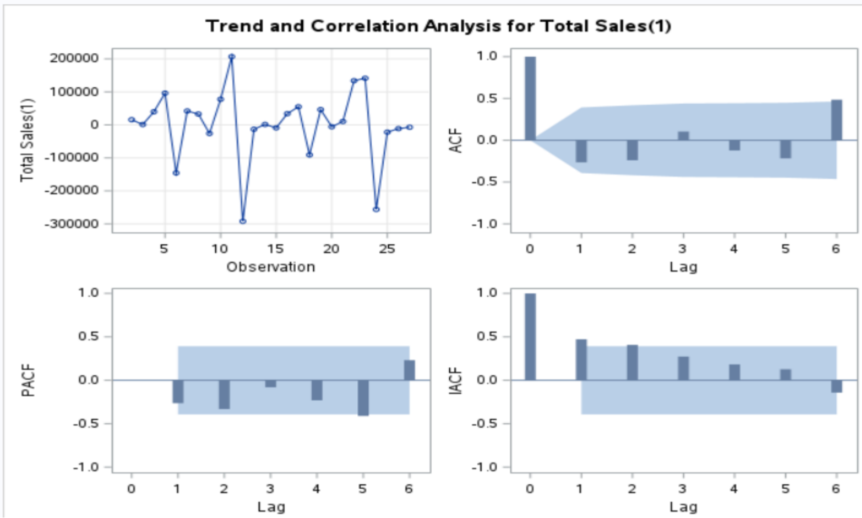
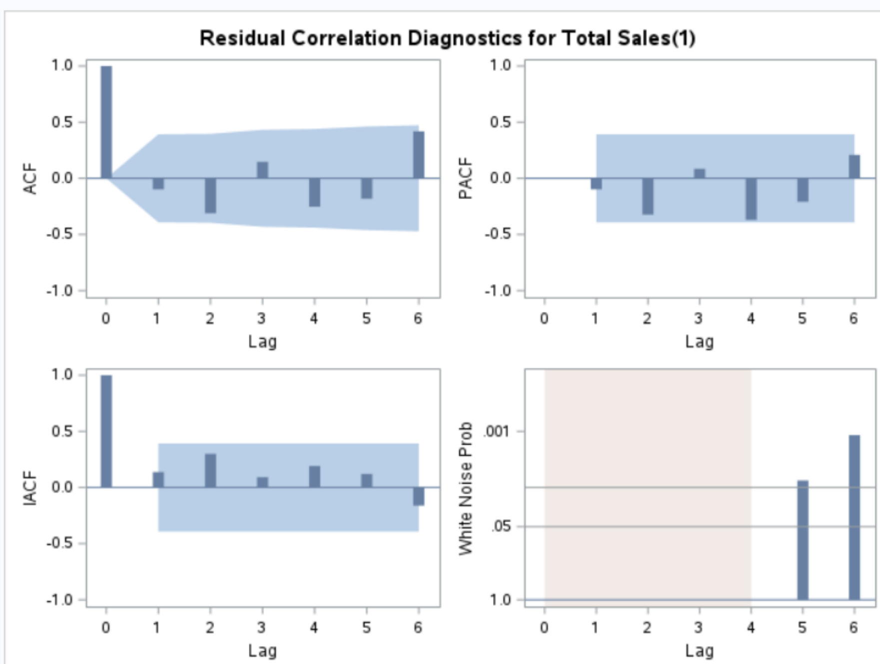



Table 8. Differencing Summary Table Product C; Figure 13: Trend and Correlation Analysis



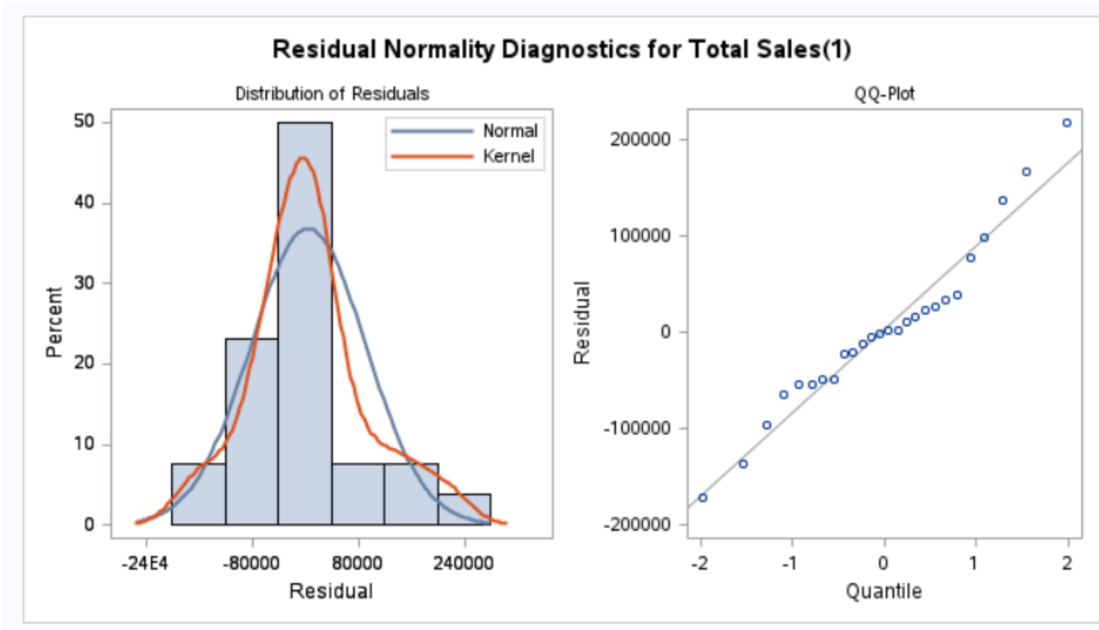



Figure 15. Residual Normality Diagnostics for Total Sales Product C SARIMA

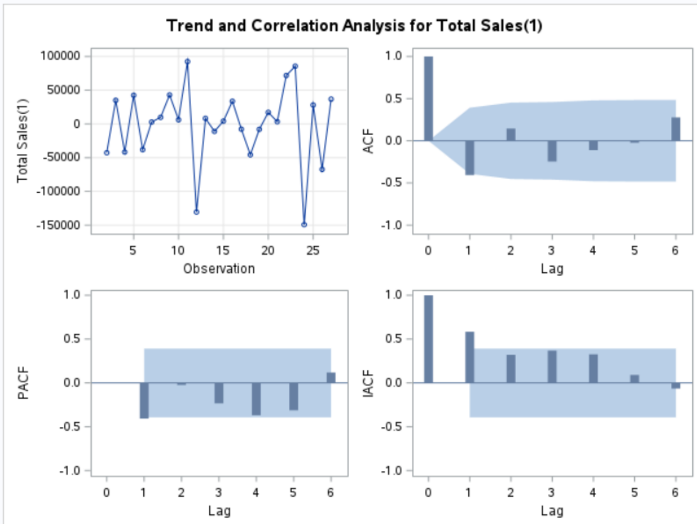
## Product D – ARIMA

| Name of Variable = Total Sales | |
|---|---|
| Period(s) of Differencing | 1 |
| Mean of Working Series | -846.328 |
| Standard Deviation | 55770.93 |
| Number of Observations | 26 |
| Observation(s) eliminated by differencing | 1 |

| Autocorrelation Check for White Noise | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| To Lag | Chi-Square | DF | Pr > ChiSq | Autocorrelations | | | | | |
| 6 | 10.58 | 6 | 0.1024 | -0.407 | 0.146 | -0.246 | -0.110 | -0.023 | 0.277 |

Table 9. Differencing Summary Table Product D ARIMA

# APPENDIX A CONTD.




| Constant Estimate | -1122.72 |
|---|---|
| Variance Estimate | 2.7909E9 |
| Std Error Estimate | 52828.76 |
| AIC | 641.3749 |
| SBC | 643.8911 |
| Number of Residuals | 26 |

| Maximum Likelihood Estimation | | | | | |
|---|---|---|---|---|---|
| Parameter | Estimate | Standard Error | t Value | Approx Pr > \|t\| | Lag |
| MU | -797.99073 | 7422.6 | -0.11 | 0.9144 | 0 |
| AR1,1 | -0.40693 | 0.18642 | -2.18 | 0.0290 | 1 |

Figure 16: Trend and Correlation Analysis Product D ARIMA Method

| Correlations of Parameter Estimates | | |
|---|---|---|
| Parameter | MU | AR1,1 |
| MU | 1.000 | 0.019 |
| AR1,1 | 0.019 | 1.000 |

| Autocorrelation Check of Residuals | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| To Lag | Chi-Square | DF | Pr > ChiSq | Autocorrelations | | | | | |
| 6 | 9.32 | 5 | 0.0970 | -0.011 | -0.104 | -0.320 | -0.262 | 0.022 | 0.308 |
| 12 | 21.10 | 11 | 0.0323 | 0.001 | -0.073 | -0.201 | -0.169 | 0.084 | 0.397 |
| 18 | 24.14 | 17 | 0.1157 | 0.007 | 0.061 | -0.183 | -0.072 | -0.012 | 0.048 |
| 24 | 28.03 | 23 | 0.2146 | -0.056 | 0.036 | -0.110 | 0.082 | -0.001 | 0.034 |

# APPENDIX A CONTD.

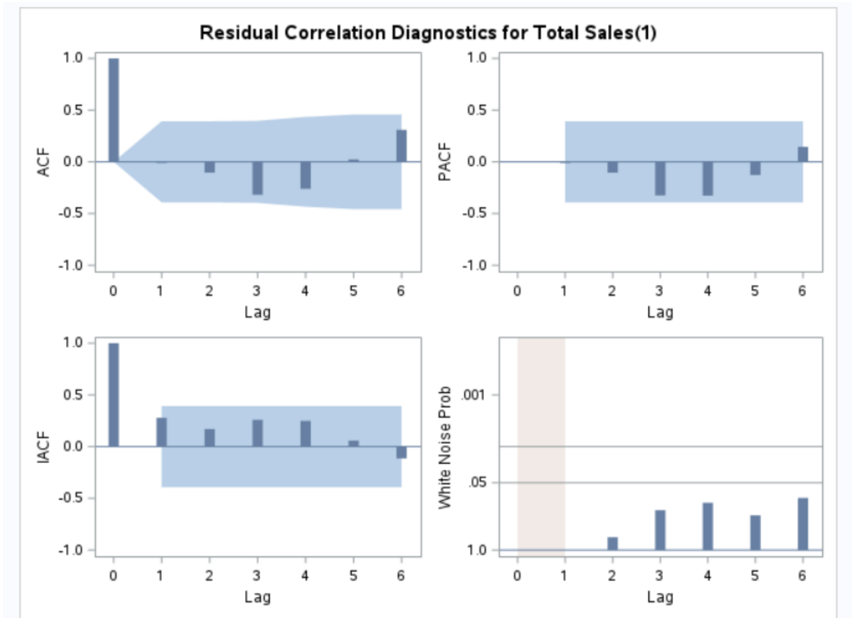



Figure 17. Residual Correlation Diagnostics for Total Sales Product D ARIMA

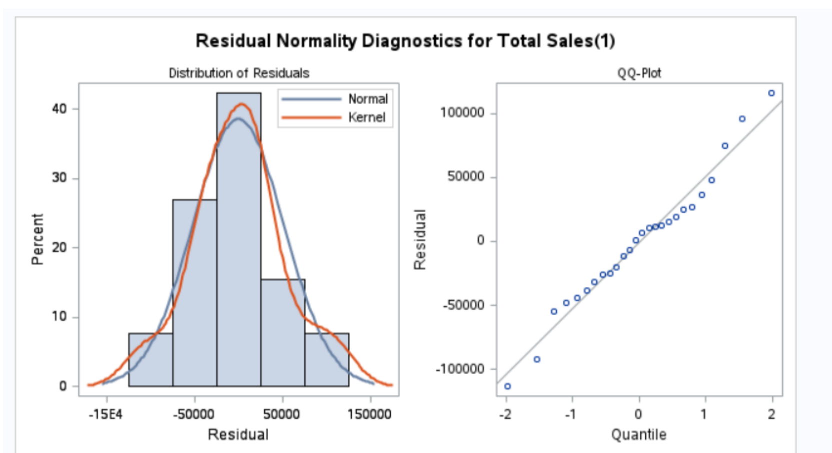



Figure 18. Residual Normality Diagnostics for Total Sales Product D ARIMA

| Model for variable Total Sales | |
|---|---|
| Estimated Mean | -797.991 |
| Period(s) of Differencing | 1 |

| Autoregressive Factors | |
|---|---|
| Factor 1: | 1 + 0.40693 B**(1) |

## APPENDIX A CONTD.

## PRODUCT D SARIMA

| Name of Variable = Total Sales | |
|---|---|
| Period(s) of Differencing | 12 |
| Mean of Working Series | 4438.208 |
| Standard Deviation | 24282.5 |
| Number of Observations | 15 |
| Observation(s) eliminated by differencing | 12 |

Table 9. Differencing Summary Table Product D SARIMA

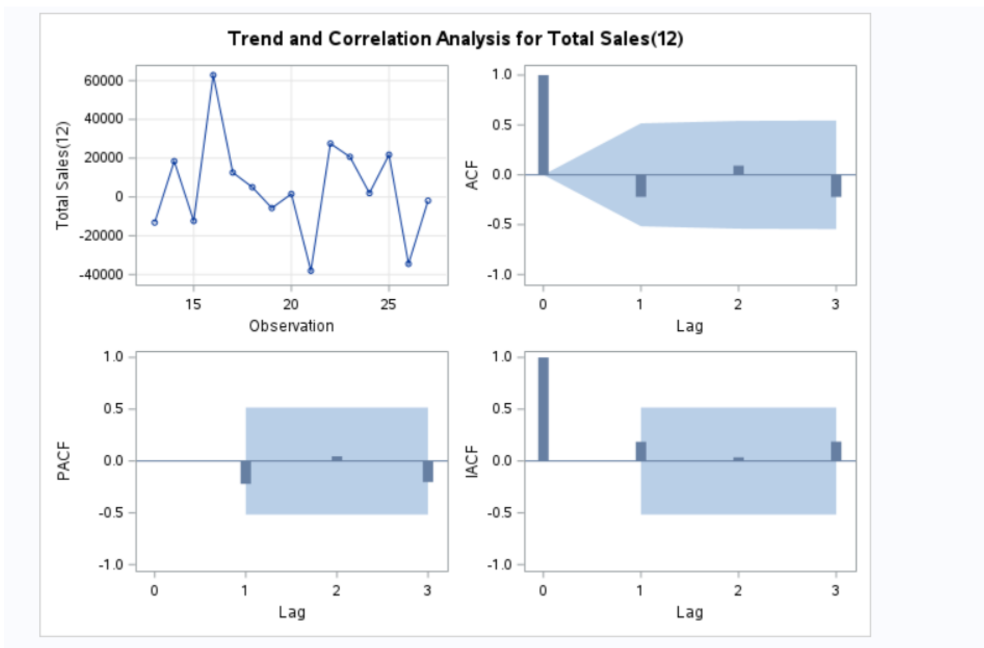



Figure 20: Trend and Correlation Analysis Product D SARIMA Method

## APPENDIX A- CONTD

| ARIMA Estimation Optimization Summary | |
|---|---|
| Estimation Method | Maximum Likelihood |
| Parameters Estimated | 4 |
| Termination Criteria | Maximum Relative Change in Estimates |
| Iteration Stopping Value | 0.001 |
| Criteria Value | 6455.475 |
| Maximum Absolute Value of Gradient | 2.4244E8 |
| R-Square Change from Last Iteration | 0.267649 |
| Objective Function | Log Gaussian Likelihood |
| Objective Function Value | -172.034 |
| Marquardt's Lambda Coefficient | 0.001 |
| Numerical Derivative Perturbation Delta | 0.001 |
| Iterations | 1 |
| Warning Message | Estimates may not have converged. |

| Autocorrelation Check of Residuals | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| To Lag | Chi-Square | DF | Pr > ChiSq | Autocorrelations | | | | | |
| 6 | 7.48 | 3 | 0.0581 | -0.016 | 0.033 | -0.267 | -0.309 | -0.139 | 0.337 |
| 12 | 13.85 | 9 | 0.1277 | 0.099 | 0.122 | 0.003 | -0.317 | -0.056 | -0.046 |

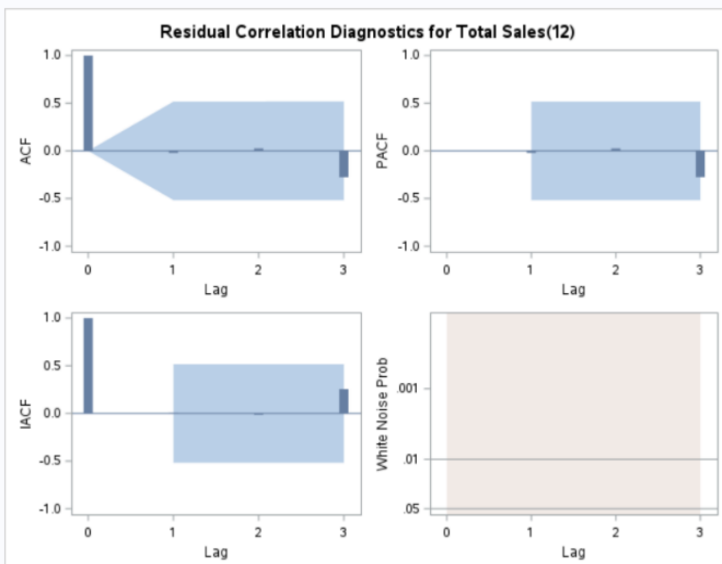



Figure 21. Residual Correlation Diagnostics for Total Sales Product D SARIMA

## APPENDIX A CONTD.

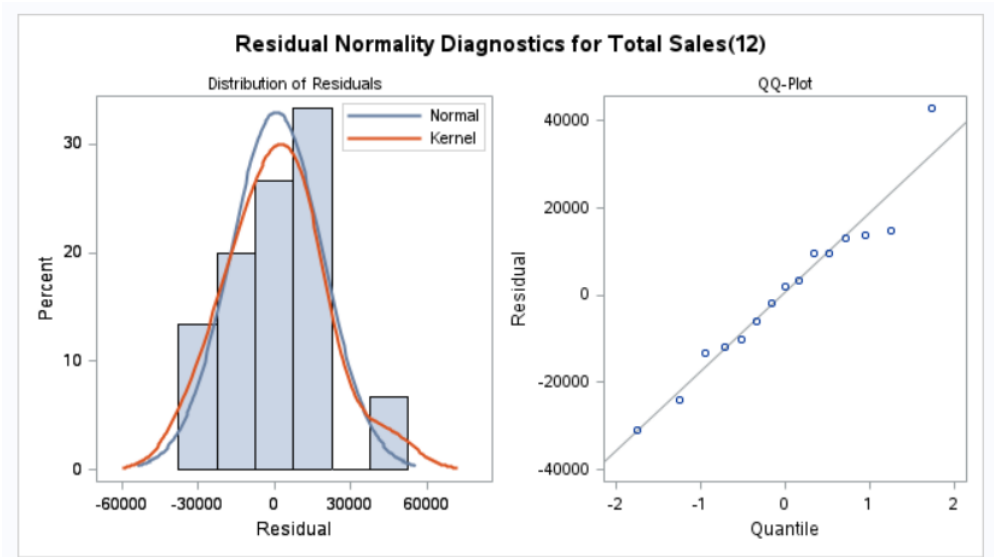



Figure 22. Residual Normality Diagnostics for Total Sales Product D SARIMA

## REFERENCES


1. Ampountolas, A. (2021). Modeling and forecasting daily hotel demand: A comparison based on SARIMAX, neural networks, and WH models. Forecasting, 3(3), 580–595. https://doi.org/10.3390/forecast3030037

2. Arunraj, N. S., Ahrens, D. (2015). A hybrid seasonal autoregressive integrated moving average and quantile regression for daily food sales forecasting. International Journal of Production Economics, 170, 321–335.

3. Arunraj, N. S., Ahrens, D., and Fernandes, M. (2016). Application of SARIMAX model to forecast daily sales in food retail industry. International Journal of Operations Research and Information Systems, 7(2), 1–21. https://doi.org/10.4018/ijoris.2016040101

4. Babaï, M. Z., Barrow, D. K., Taieb, S. B., Bergmeir, C., Bessa, R. J., Bijak, J., Boylan, J. E., Browell, J., Carnevale, C., Castle, J. L., Cirillo, P., Padgett, C., Cordeiro, C., Oliveira, F. L. C., De Baets, S., Dokumentov, A., Petropoulos, F., Apiletti, D., Assimakopoulos, V., . . . Ziel, F. (2022). Forecasting: theory and practice. International Journal of Forecasting, 38(3), 705–871. https://doi.org/10.1016/j.ijforecast.2021.11.001

5. Chaouch, F., Bouzidi, L., and Benaben, F. (2019). Revisiting the Holt-Winters' Additive Method for Better Forecasting. International Journal of Information Technology and Decision Making, 18(4), 1279–1298. https://doi.org/10.1142/S021962201940006X

6. Dunlea, J. (2024, January 4). What is ML-Powered Demand Forecasting? Akkio. https://www.akkio.com/post/demand-forecasting-using-machine-learning

7. Ensafi, Y., Amin, S. H., Zhang, G., and Shah, B. (2022). Time-series forecasting of seasonal items sales using machine learning – A comparative analysis. International Journal of Information Management Data Insights, 2(1), 100058. https://doi.org/10.1016/j.jjimei.2022.100058

8. Fattah, J., Ezzine, L., Aman, Z., Moussami, H. E., and Lachhab, A. (2018). Forecasting of demand using ARIMA model. International Journal of Engineering Business Management, 10, 184797901880867. https://doi.org/10.1177/1847979018808673

9. Fildes, R., Ma, S., and Kolassa, S. (2022). Retail forecasting: Research and practice. International Journal of Forecasting, 38(4), 1283–1318. https://doi.org/10.1016/j.ijforecast.2019.06.004

10. Food demand prediction using the nonlinear autoregressive exogenous neural network. (2021). IEEE Journals and Magazine | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/9585704

11. Kenton, W. (2019, May 27). Seasonality: What it Means in business and Economics, examples. Investopedia. https://www.investopedia.com/terms/s/seasonality.asp

12. Kırcı, M., Isaksson, O., and Seifert, R. W. (2022). Managing perishability in the fruit and vegetable supply chains. Sustainability, 14(9), 5378. https://doi.org/10.3390/su14095378

13. Llanes, R. P., Sala, H. V., and García, A. O. (2020, March 1). Models for predicting perishable products demands in food trading companies. https://www.redalyc.org/journal/3783/378365895005/html/

14. Ma, S., Fildes, R., and Huang, T. (2016). Demand forecasting with high dimensional data: The case of SKU retail sales forecasting with intra- and inter-category promotional information. European Journal of Operational Research, 249(1), 245–257.

15. Mor, R. S., Jaiswal, S. K., Singh, S., and Bhardwaj, A. (2019). Demand forecasting of the short-lifecycle dairy products. In Understanding the Role of Business Analytics, Singapore: Springer, pp. 87-117.

16. Pannakkong, W., Huynh, V., and Sriboonchitta, S. (2019). A novel hybrid autoregressive integrated moving average and artificial neural network model for Cassava export forecasting. International Journal of Computational Intelligence Systems, 12(2), 1047. https://doi.org/10.2991/ijcis.d.190909.001

17. Rodrigues, M. A., Miguéis, V. L., Freitas, S. V., and Machado, T. (2024). Machine learning models for short-term demand forecasting in food catering services: A solution to reduce food waste. Journal of Cleaner Production, 435, 140265. https://doi.org/10.1016/j.jclepro.2023.140265

18. Schnorrenberger, D. (2023, December 28). The future of AI demand forecasting in grocery retail. InvafreshTM. https://invafresh.com/the-future-of-ai-demand-forecasting-in-grocery-retail/

19. Scott Armstrong, J. (2006). Findings from evidence-based forecasting: Methods for reducing forecast error. International Journal of Forecasting, 22(3), 583-598. https://doi.org/10.1016/j.ijforecast.2006.04.006

20. Seyedan, M., and Mafakheri, F. (2020). Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities. Journal of Big Data, 7(1). https://doi.org/10.1186/s40537-020-00329-2

21. Svetunkov, I., Chen, H., and Boylan, J. E. (2023). A new taxonomy for vector exponential smoothing and its application to seasonal time series. European Journal of Operational Research, 304(3), 964–980. https://doi.org/10.1016/j.ejor.2022.04.040

22. Taha Falatouri, Farzaneh Darbanian, Patrick Brandtner, Chibuzor Udokwu. (2022). Predictive Analytics for Demand Forecasting – A Comparison of SARIMA and LSTM in Retail SCM. Procedia Computer Science, Volume 200, 993-1003. https://

23. Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babaï, M. Z., Barrow, D. K., Taieb, S. B., Bergmeir, C., Bessa, R. J., Bijak, J., Boylan, J. E., Browell, J., Carnevale, C., Castle, J. L., Cirillo, P., Padgett, C., Cordeiro, C., Oliveira, F. L. C., De Baets, S., Dokumentov, A., . . . Ziel, F. (2022b). Forecasting: theory and practice. International Journal of Forecasting, 38(3), 705–871. https://doi.org/10.1016/j.ijforecast.2021.11.001

24. Mitra A, Jain A, Kishore A, Kumar P. A Comparative Study of Demand Forecasting Models for a Multi-Channel Retail Company: A Novel Hybrid Machine Learning Approach. Oper. Res. Forum. 2022;3(4):58. doi: 10.1007/s43069-022-00166-4. Epub 2022 Sep 27. PMCID: PMC9514716.

25. Nasseri, M., Falatouri, T., Brandtner, P., and Darbanian, F. (2023). Applying Machine Learning in Retail Demand Prediction—A comparison of Tree-Based Ensembles and Long Short-Term Memory-Based Deep Learning. Applied Sciences, 13(19), 11112. https://doi.org/10.3390/app131911112

26. Pannakkong, W., Huynh, V.-N., and Sriboonchitta, S. (2019, September 26). A novel hybrid autoregressive integrated moving average and artificial neural network model for cassava export forecasting. International Journal of Computational Intelligence Systems. https://www.atlantis-press.com/journals/ijcis/125918240/view

27. ChatGPT (OpenAI) - For rephrasing