BINDURA UNIVERSITY OF SCIENCE EDUCATION

FACULTY OF SCIENCE AND ENGINEERING

DEPARTMENT OF STATISTICS AND MATHEMATICS

MODELING THE DYNAMIC IMPACT OF PROMOTIONAL SALES ON REVENUE PERFOMANCE: AN ECONOMETRIC ANALYSIS OF BAKERS INN FOOD OUTLETS.

BY

NOKUTENDA MASHINGAIDZE B213205B

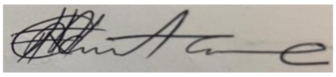
A RESEARCH PROJECT SUBMITTED TO THE DEPARTMENT OF MATHEMATICS IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE HONORS BACHELOR OF SCIENCE DEGREE IN STATISTICS AND FINANCIAL MATHEMATICS

SUPERVISOR: MR E. MUKONOWESHURO

2025

APRPROVAL FORM

This is to certify, that this research project is the result of my own research work and has not been copied or extracted from past sources without acknowledgement. I hereby declare that no part of it has been presented for another degree in this University or elsewhere.

Nokutenda Mashingaidze  20/06/25

Student Signature Date

B213205B

Certified by:

MR.E.Mukonoweshuro  20/06/25

Supervisor Signature Date

DR.M. MAGODORA  20/06/25

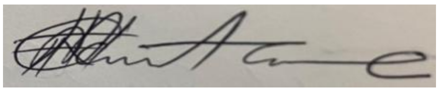
Chairperson Signature Date

DECLARATION

I Nokutenda Mashingaidze hereby declare that this submission is my own work apart from the references of other people's work which has duly been acknowledged. I hereby declare that this work has neither been presented in whole nor in part for any degree at this university or elsewhere.

Author: Nokutenda Mashingaidze

Registration Number: B213205B

Signature: 

Date: 20 June 2025

DEDICATION

I dedicate this dissertation to my father Hosiya Mashingaidze and mother Locadia Mashingaidze, my younger brothers Mukundi and Coronation, sisters Precious and Crown who have sacrificed towards my personal and professional endeavours. I thank you for believing in my dreams.

ACKNOWLEDGEMENT

Most importantly, I give all glory and credit to Almighty God, whose wisdom, mercy, and unshakable love have guided me through every step of this academic endeavour. I am sincerely grateful for the gift of life, strength, good health, and the clear mind that enabled me to complete this thesis. In times of doubt and fatigue, it was before God that I found peace, motivation, and the spirit to carry on.

I am particularly thankful to my supervisor, Mr E. Mukonoweshuro, for the unwavering guidance, expert advice, and unshakeable support throughout the entire research process. His patience, clear thinking, and academic perceptiveness have organized not just this work but also my understanding of good scholarship. I also extend sincere thanks to Mrs Hlupo and Dr Magodora, whose help and advice proved priceless at key moments in this study. I want to say a very sincere thank you to the Department of Statistics and Mathematics of Bindura University of Science Education. The lecturers and administrative staff provided not only knowledge, but also a setting that promoted inquiry, discipline, and growth. I want to thank in particular lecturers whose kindness and support helped add value to my academic journey.

To my dear family, I can never thank you enough for your perpetual prayers, encouragement, and sacrifices. Your faith in the best that I can be gave me the ground I stood on during difficult times. God must reward you abundantly for your love and support. To my colleagues and friends, I appreciate the fellowship which we shared in this journey. The late-night debates, late-night laughter, and encouraging each other through this process made it more enjoyable and meaningful. You are family to me, rather than colleagues. May God richly bless all of you.

This thesis is a living testimony not only to the academic travail but to the group love, guidance, and support of all those who are surrounding me.

ABSTRACT

The dynamic impact of promotional sales on the revenue performance of Bakers Inn is studied in this research. In the broader context of the fast-moving consumer goods (FMCG) sector in a developing economy, the research critically studies the performance of promotional strategy in a context of economic uncertainty. Informing both classical econometric and modern machine learning techniques such as the Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX) model, as well as the Random Forest algorithm, the study investigates the short-run and long-run impact of promotion interventions on turnover.

Evidence is derived from weekly sales and promotion activity information collected for 24 months (January 2023 to December 2024). Pre-model diagnostics, descriptive statistics, and time series decomposition were employed to construct a comprehensive picture of the data before model deployment. The ARIMAX model provided interpretable outputs on the temporal dynamics between campaign inputs and revenue, while the Random Forest algorithm provided more predictive accuracy, particularly in modelling non-linarites and abrupt market transition with campaign intensity.

Empirical results indicate that promotional sales greatly influence revenue, where Random Forest performs better than ARIMAX in forecasting accuracy consistently, especially in instances of high volatility. The implications from the study underscore strategic advantages in using advanced analytics in promotional planning. Additionally, the study bridges an essential research gap in the literature since it is simultaneously employing machine learning and econometric models in a Sub-Saharan African retail context a previously relatively less explored subject.

By offering actionable data-driven insights to maximize the efficiency of promotions, this thesis contributes significantly both to managerial practice and research literature. The insights from the findings are not only applicable to Bakers Inn but also extendable to other FMCG firms within similarly dynamic and resource-constrained environments. In sum, this study

highlights the importance of evidence-based marketing to deliver sustainable top-line growth and strategic responsiveness in emerging markets.

TABLE OF CONTENTS

Contents

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.0 Introduction

In the modern highly competitive business landscape, promotional sales have become an essential tool for improving revenue performance and market relevance. The study examines the dynamic effect of promotional activities on the revenue performance of Bakers Inn, a Zimbabwean fast-food industry leader. From an econometric point of view, the research aims to examine the effect of various promotional mechanisms on financial performance over time, yielding evidence-based data for informing strategic decision-making. The chapter sets the stage for the research by presenting the background, stating the problem, setting the objectives and research questions, and examining the significance, assumptions, and potential limitations of the research. It also considers the practical implications of the results for stakeholders within the fast-moving consumer goods (FMCG) industry.

1.1 Background

The global fast food industry has experienced significant change in the past decade, fuelled by aggressive competition, changing consumer tastes, and increasing operational pressures. With leading global brands such as McDonald's reporting unprecedented declines in sales due to economic pressures and new premium rivals (Kotler and Armstrong, 2016), companies have been increasingly relying on promotional schemes to drive growth. Within the African context, fast food chains are also faced with additional challenges from economic uncertainty, fluctuating exchange rates, and capricious consumer demand. In countries like South Africa, such pressures have redirected competitive forces in the fast-food segment (Chea, 2023).

Zimbabwe is a highly sophisticated economic environment with currency devaluation, inflation, and policy settings that change and impact both consumer purchasing power as well as business operations. In this setting, Bakers Inn, owned by Innscor Africa Limited, has emerged as a leading fast-food and baking brand. The business has embraced marketing strategies such as promotions and ad-sponsored campaigns to increase market share and sales growth. For example, its "Buy Win Promotion" translated into a 22% revenue gain and shifted market share to 45%, reflecting the success of timely promotions (Impact of Sales Promotional Strategies on Organisational Performance in Nigeria, 2017).

Despite the applied utility of these strategies, there remains limited empirical research assessing their dynamic impact on revenue, particularly in developing nations like Zimbabwe. Much of the literature that is already available relies on cross-sectional or static analysis and is likely to overlook temporal dynamics and feedback processes that are involved in marketing response (Blattberg and Neslin, 1990). Most existing research has also been carried out in developed economies, thus resulting in a context gap that is relevant for sub-Saharan Africa (Chandon, Wansink and Laurent, 2000).

Addressing this shortcoming, this study examines the dynamic relationship between revenue performance and promotional selling at Bakers Inn. In attempting to model this relationship in a time-series framework, the study aims to deliver evidence-based findings that can be used for strategic planning in highly dynamic market settings. The end-product will not only contribute to scholarly knowledge about promotional effectiveness but also provide actionable recommendations for fast-moving consumer goods (FMCG) firms operating in similarly volatile settings (Cavusgil, Ghauri and Agarwal, 2002).

1.2 Problem statement

Bakers Inn continues to invest in promotional sales without a clear understanding of their true impact on revenue performance. The lack of empirical evidence on whether these promotions lead to sustained growth or merely short-term spikes presents a critical problem. Traditional analysis methods often fail to capture the dynamic and non-linear effects of promotions over time, leading to ineffective marketing strategies. This study addresses the need for strong, data-driven models using ARIMAX and Random Forest to evaluate and forecast the real influence of promotional sales on revenue, ensuring informed decision-making in a competitive and volatile market.

1.3 Objectives

1. To analyse the short-term and long-term impacts of promotional sales on revenue using historical data
2. To develop and fit Random Forest and ARIMAX models to analyse the dynamic relationship between promotional sales and revenue.
3. To compare the performance of Random Forest and ARIMAX using appropriate performance metrics.
4. To provide data-driven recommendations to Bakers Inn on how promotional strategies impact revenue performance using the best model.

1.4 Research questions

1. How effectively do Random Forest and ARIMAX models capture the dynamic relationship between promotional Sales and revenue?

2. What are the short-term and long-term effects of promotional sales on revenue performance?

3. Which model, between Random Forest and ARIMAX provides more accurate and reliable results in analysing promotional sales and revenue performance?

4. What actionable insights can be driven from the best performing model to guide Bakers Inn's strategies for improved revenue outcomes?

1.5 Scope of the study

This research empirically analyses the effects of promotional strategies, such as discounts, promotional vouchers, and other short-term promotions, on revenue performance at Bakers Inn restaurants for the period January 2023 to December 2024. The research employs cutting-edge analytical models, specifically the Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX) and Random Forest models, to capture the dynamic and time-dependent interactions between promotional activity and revenue performance. By leveraging these econometric and machine learning techniques, the study seeks to generate actionable insights that can inform data-driven decision-making and enhance strategic marketing initiatives within Zimbabwe's fast-food industry. Besides, this research aims to contribute meaningfully to the broader sales promotion effectiveness literature in the context of emerging economies. By designing an adaptive analytical framework, the research offers pragmatic importance not only to Bakers Inn but also to comparable firms that operate under volatile economic environments. The output of the analysis is expected to guide evidence-based strategic planning and operational redesign towards the attainment of sustainable revenue growth amidst competitive market pressures.

1.6 Significance of the study

To the researcher

The research enhances the student's research intellectual abilities and getting an insight into the promotional strategies being employed by fast foods industries in Zimbabwe. The research gives the researcher the platform to analyse and gain knowledge on the basis of sales promotion strategy formulation and its effects on sales volume, revenue, costs and profits as well consumer behaviour.

To the organization significance

The aim of this study is to investigate the impact of promotional sales on sales revenue in the fast food business, with a particular focus on Bakers Inn food outlets. The research also looks into the role of promotion as a tool for raising sales ant the amount which it can patronage consumption of a company's goods. The findings of the study were useful to a variety of similar business in determining how to revise their sales promotion efforts accordingly. Statistical models to address this issue has been provided, ensuring that the extent to which the promotion influences sales turnover at the conclusion of the study.

To the university

The research contributes to Bindura University Library's practical collection of studies documents. It also assists Bindura University in broadening secondary data, which serves as a benchmark for other academics undertaking comparable study as well as a source of reference.

1.7 Assumptions

The management of Bakers Inn provide data specifically for this research and assumed to be accurate. It is also assumed that the economic, political and regulatory environment remained relatively stable to be a true representation of Bakers Inn brand and its promotional practices across Zimbabwe

1.8 Delimitation

The study is confined to Bakers inn food outlets in Zimbabwe, with data collected from 2023 to 2024.It uses purposive sampling, focusing on few selected stores, which may not fully represent the entire outlet network. The analysis is limited to promotional sales and their effect on revenue, excluding broader macroeconomic or competitive factors. Secondary data is used, which may not align perfectly with academic research standards, as it was originally gathered for operational purposes.

1.9 Limitations

The study encountered several limitations such as data collection was hindered by internal system transitions across outlets, which led to gaps and inconsistences in some record. Again the study faced financial constraints, particularly regarding travel, printing and data compilation expenses factors that limited the breath of store coverage. Additionally, while secondary data was used to resources limitations, it was not originally structured for econometric analysis, possibly affecting precision and variable alignment. These factors may influence generalizability of findings.

Definition of key terms

1. **Promotional sales –** sales promotion comprises a variety incentive tools, mostly short-term, designed to stimulate quicker and greater purchase of particular products or services by consumers or trade (Kotler and Keller, 2016)
2. **ARIMAX Model**-the ARIMAX Model incorporates independent variables into the ARIMA frame work, allowing external regressors to explain variation in the dependent time series (Box et al., 2015)
3. **Random Forest**- is a classifier consisting of a collection of tree-structured classifiers each tree is grown using a random subset of the data and features and the final prediction is made by aggregating the prediction of individuals trees (Breiman, 2001)
4. **Econometric analysis**-is the use of statistical methods to quantify economic relationship, test hypothesis, and forecast outcomes using real world data (Gujarati and Porter,2009)

1.10 Chapter summary

This chapter has established the foundation for examining the dynamic impact of promotional sales on revenue performance at Bakers Inn food outlets from 2023 to 2024. By outlining the background, problem statement, objectives and methodology, it highlights the need for economic modelling in understanding promotional effective in volatile economic environment. Hence it is important for the research to examine the influence of promotional sales on sales revenue. This chapter creates a pathway for the following chapter 2, which will concentrate on the theoretical and empirical findings in order to quantify and assess the impact of previous studies on the relationship between sales promotional and sales turnover.

CHAPTER 2: LITERATURE REVIEW

2.0 Introduction

This chapter offers a critical examination of literature pertinent to the dynamic effect of promotional sales on revenue performance with an emphasis on econometric and machine learning techniques. The review collates theoretical, empirical, and conceptual frameworks to formulate a solid base for examining the short- and long-term impacts of promotional strategies. Special attention is given to the ARIMAX and Random Forest models, which are critically examined for their capacity to model revenue outcomes in complex, real-world settings. Situating the discussion within the context of Bakers Inn, one of Zimbabwe's leading fast-food outlets. This chapter identifies knowledge gaps, synthesises prior research, and outlines a conceptual framework that informs the current study's analytical approach.

2.1 Theoretical Framework

The theoretical framework guiding this study draws on a synthesis of economic, behavioural, and marketing theories in explaining the impact of promotional sales on consumer behaviour and revenue for the company. The most important theories guiding this study are the Production and Selling Concepts, Trait Theory, Prospect Theory, and the Theory of Planned Behaviour. Collectively, these theories offer a multi-dimensional lens through which to examine the processes whereby promotional tactics shape customer decision and financial payoff, and thereby guide the analytical framework of this study.

2.1.1 Production and selling concept

The production concept is founded on the idea that customers favour products that are readily available and affordable. Consequently, production managers prioritize product quality and ongoing development. In contrast, the selling concept emphasizes the importance of marketing efforts, suggesting that customers will not purchase sufficient quantities of a product unless it is heavily promoted. This concept asserts that significant selling and promotional activities are necessary to encourage consumer purchases of the company's products.

2.1.2 Trait theory

Trait theory offers a quantitative perspective on personality (Blackwell et al., 2001). This theory suggests that an individual's personality is made up of specific predispositions known as traits. It posits that these traits are common among many people and can vary in degree from one person to another (Mischel, 1968). Additionally, it assumes that traits tend to be stable over time and have a consistent influence on behaviour, regardless of the surrounding environment (Sanford, 1970). The theory also asserts that traits can be identified through behavioural indicators. Common examples of traits include aggressiveness, dominance, friendliness, sociability, extroversion, empathy, innovativeness, deal proneness, and variety-seeking. At the beginning of this chapter, we posed the question: “How do people make decisions and choices in the marketplace, and how can sales promotions impact these decisions?” The theories and models discussed thus far (economic theory, stimulus-response model, stimulus-organism-response model, and trait theory) offer valuable insights into this question.

2.1.3 Prospect theory

Prospect Theory, developed by Kahneman and Tversky (1984), posits that consumers evaluate outcomes as gains or losses relative to a subjective reference point. Applied to promotional contexts, this theory suggests that non-price promotions—such as gifts or bonus items—are perceived as separate gains, while price promotions—such as discounts—are viewed as reductions in perceived loss (Diamond and Sanyal, 1990; Diamond and Campbell, 1990). Diamond and Campbell (1989) further argue that price-based promotions tend to reduce consumers’ internal reference prices over time, whereas non-price promotions maintain the perceived value of the product. Although prospect theory suggests non-price promotions should be more appealing, empirical evidence shows similar consumer preferences for both, indicating that the psychological framing of promotions may be more complex than originally theorised.

2.1.4 Planned Behavior

The planned behavior Theory suggest that behavior can be influenced by sales promotion stimuli, which in turn modify beliefs, attitudes, intentions, and ultimately behavior. When an intervention impacts customers, it alters their intentions, leading to changes in behavior. This theory is significant because a compelling promotional offer from an organization can drive consumer purchasing behavior. In this context, the organization has implemented a “Thrilling Thursday’ sales promotion, with the expectation that it will boost sales revenue, profits and consumer buying behavior. The researcher is conducting this study to determine whether the promotion will indeed lead to an increase in sales revenue, profits and consumer purchasing actions

2.2 Definition of Sales Promotion

Sales promotion encompasses a range of short-term incentives designed to stimulate immediate purchases from consumers or retailers (Kotler & Keller, 2011). It serves as a key component of marketing strategy, vital for maintaining competitive advantage in dynamic markets (Cole, 1993). According to the American Marketing Association (2010), it involves targeted marketing pressure applied over a defined period to enhance product visibility, encourage trials, and boost demand. Kotler and Armstrong (1994) highlighted its role in bridging quality offerings with effective customer communication. Shimp (2000) expands this view by describing sales promotions as tools that increase the perceived value of products, whether through economic or symbolic incentives. Similarly, Hackley (2010) emphasised branded novelties and in-store initiatives like discounts as vehicles for building consumer goodwill. These definitions underline both economic inducements and communication strategies that underpin effective promotional execution.

2.3 Objectives of Sales Promotion

The primary objective of sales promotion is to generate a short-term increase in sales, while secondary aims include fostering brand loyalty and acquiring new customers (Wilson, 1995). Sales promotions are designed not merely to attract attention or improve perceived product value, but to penetrate a highly competitive market environment. Effective promotional strategies often target customer behaviour rather than attitudes, aiming to prompt immediate purchasing actions. Marketers can tailor their promotional tools based on consumer segmentation; for instance, loyalty programs are ideal for retaining current customers, while incentives such as discount coupons or free samples are more suitable for attracting new or competitor-loyal customers (Daver, 1999). A well-informed promotional approach, grounded in consumer behaviour analysis, enhances the efficiency of marketing campaigns and aligns with broader organisational revenue goals.

2.4 Types of sales promotion techniques

Sales promotion can be categorized into two main types: consumer-oriented promotions (pull strategy) and business-oriented promotions (push strategy). According to Pickton and Broderick (2005), consumer sales promotions are pull activities aimed at generating demand among end users or customers, effectively pulling products through the distribution chain. These customer-focused promotions are specifically designed to stimulate demand for products or services. On the other hand, trade promotions are initiatives provided by manufacturers to sellers or trade organizations (Blattberg and Neslin, 1990). Pickton and Broderick (2005) explain that trade sales promotions help "push" products through the distribution chain by motivating channel members to stock and sell to end users. Typically, manufacturers direct these trade promotions toward retailers.

2.5 Strategic consideration in sales Promotion Design

According to Kotler (2001), the effectiveness of sales promotion depends on several key factors, the first being the size and nature of the promotion offered. A more appealing incentive is more likely to elicit a stronger consumer response. However, the duration of the promotion is equally critical. A promotion that is too short may fail to reach a broader customer base, while an excessively long promotion can dilute its perceived value, leading to diminishing marginal returns (Meyer, 2019). For example, the widely recognized “thrilling Thursday” campaign by Bakers Inn has been noted for achieving a balance in duration and frequency, thereby maximizing reach and impact Sales promotion must systematically manage to align with overall marketing objectives. As Cronje (1993) emphasizes promotional efforts should not be ad hoc; they need special strategic planning and ongoing oversight. Charas (1984) proposed advocated precise targeting, simplification of promotional messages and timely planning. These principles are foundational for executing impactful sales campaigns that resonate with consumer and support long term brand equity.

2.6 Consumer Behavior and Promotional Response

The theoretical lens also draws from behavioral economics and consumer decisions making models. Promotions are not merely transactional events, they influence perceptions of value, urgency and brand choice. Sales promotions can prompt immediate increases in purchase quantity, stimulate trial among non-users and temporarily shift brand preferences (Blattberg and Neslin, 1990). Moreover responses heterogeneity is evident across consumer segments. Research suggest that heavy user and price sensitive consumers exhibit higher responsiveness to promotions compared to loyal customers (Krishnamurthi and Raj 1991)These dynamics are particularly relevant in the Zimbabwe fast moving consumer goods (FMCG) sector, where purchasing power and market volatility and data driven approaches to navigate the variability in consumer behavior and market conditions.

2.7 Dynamic Econometric modeling of Promotional Effects

2.7.1 Economic modelling with ARIMAX

Dynamic econometric models, such as ARIMAX, allow for the analysis of time-dependent relationships between promotional efforts and revenue. These models incorporate exogenous variables such as discounts, campaigns to account for marketing interventions. ARIMAX is particularly suited for identifying both short-term impulses and long-term equilibrium relationships (Enders, 2015).

2.7.2 Machine Learning Approaches to Promotion Analysis

Machine learning models like Random Forest have gained traction for their ability to model complex, non-linear relationships without strict statistical assumptions. Random Forest uses ensemble decision trees to predict outcomes based on multiple input variables, offering high predictive accuracy and strong to overfitting (Breiman, 2001).

2.8 Synthesis of theoretical perspectives

In summary, the theoretical framework of this study is multidimensional, encompassing

- Marketing theory which emphasizes strategic planning, consumer targeting and promotional design
- Consumer behaviour theory which explains variations in promotional responsiveness
- Econometric and Machine Learning Theory which explains modelling of linear and nonlinear time-dependent relationship.

This integrated approach provides a solid foundation for understanding and forecasting the dynamic impact of promotional sales on revenue performance in the Zimbabwean context. It also allows the study to bridge theory and practice, offering empirical evidence that can inform more effective promotional strategies in similar business environments.

2.9 Empirical Literature

2.9.1 Positive impacts of sales promotion on sales

Empirical evidence largely assumes that promotions trigger short-term sales and revenue escalation, even though their long-term outcomes vary with the situation. Brito and Hammond (2007) refer to still-allowed debates on profitability but agree with their efficacy in triggering spontaneous sales. Similarly, Blattberg and Neslin (1989) and Wilkinson (1982) posit that promotions trigger short-term buyer purchases. Promotions like discount and in-store display can be proven to build product visibility and reinforce trial purchase (Oyedapo, 2012; Wayne, 2002), and frequent exposure especially with loyalty rewards can trigger habitual purchase (Pauwels et al., 2002).Along with short-term advantages, sales promotions also play a role in long-term strategy goals. Churchill (1995) and Foster (1996) suggest that in the event of success, promotions build customer loyalty and energize stagnant products. Chardon (2000) found price promotions can improve market share and develop demand, especially when reinforced with in-store promotion.

2.9.2 Negative impacts of sales promotion on sales

While promotions can boost short-term sales, several studies refer to adverse effects on both the short and long time horizon. In the short term, Kopalle, Mela and Marsh (1999) state that promotions can cause artificial demand increases that are not reflective of a lasting consumer demand. In addition, promotions draw price-sensitive buyers with low loyalty to the brand (Ailawadi and Neslin, 1998), and excessive use will foster consumer habituation, reducing product and offer perceived value (Zeelenberg and van Putten, 2005). This, in the long term, will drain brand equity, especially if customers associate a brand more with promotions and discounts than with quality (Shi, Cheung and Prendergast, 2005). Finally, Papatla and Krishnamurthi (1996) argue that overdependence on promotions may create opportunistic buying and brand switching at the cost of loyalty. Further, Kelly (2003) and Fill (2005) warn against post-promotion sales slumps, wherein customers delay purchases in anticipation of future offers. Such behaviors, Mack (2005) argues, damage brand reputation and long-term profitability.

2.10 Evaluating the Effectiveness of sales promotions

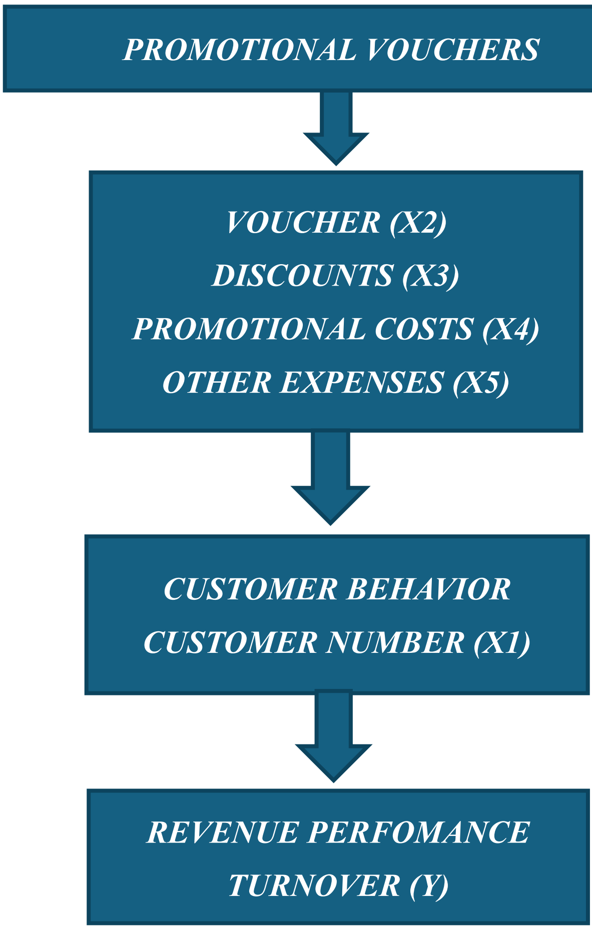
Measurement of sales promotion performance is one of the most important yet challenging tasks for managers and marketers due to the difficulty in isolating promotional effects from the general marketing mix (Cronje, 1993). Yet, effective evaluation is required in order to understand return on investment and inform future strategy. Gupta (1998) suggests the analysis of pre-promotion, in-promotion, and post-promotion sales trends to find significant changes, while Kotler (2003) suggests the use of scanner data to track consumer behaviour by differentiating between trial purchases and repeat buying. Surveys and focus groups can provide valuable data on consumer attitudes, message comprehension, and brand recall (Perreault, 2000). The concept of a "sales bump"—a temporary sales increase during a promotion—is a useful measure, though scholars caution that longer-term impact is better than short-term gain (Gupta, 1998). Lastly, as Cronje (1993) argues, effective evaluation must go beyond sales volume to assess cost-efficiency, strategic alignment, and consistency across channels, delivering both short-term gain and long-term brand integrity.

2.11 Research Gap

Despite existing literature on sales promotions, few studies examine both the short-term and long term impacts on revenue performance, particularly in Zimbabwe's fast food sector. For Bakers Inn, specifically, there is a lack of empirical evidence on how promotional strategies affect revenue dynamics over time in the context of Zimbabwe's fast-food sector, particularly for Bakers Inn. Existing studies often overlook the delayed effects and persistence of promotional outcomes, as well as the role of underlying factors such as customer loyalty, brand strength, and economic conditions. This study fills the gap by using ARDL and LSTM models to assess the full impact of promotional sales- both immediate and sustained on revenue performance from 2023-2024 This study addresses these gaps by using time-series (2023 to 2024) and integrating econometric and machine learning techniques to provide a comprehensive view of promotional impacts on Bakers Inn's revenue performance.

2.12 Proposed conceptual Model

2.12.1 Conceptual Model



2.12.2 Sources of data

This research integrates primary and secondary data to analyze the dynamic impact of promotional sales on revenue performance in Bakers Inn food stores for the period 2023-2024.The primary data set consists of weekly sales and promotional activity data directly extracted from the point-of-sale (POS) system records of Bakers Inn. The records include revenue levels, customer counts, and promotional information like expenditure on discounts, coupons, and vouchers, captured over 104 weekly observations. Access to the data was granted through formal permission from the finance and statistics department and subsequently aggregated into a structured time-series format for analysis. Secondary data is made up of auxiliary records, for example, previous financial reports and promotional campaign records. Shiu (2009) discusses that this kind of data is advantageous to longitudinal analysis in that it reveals patterns and trends in revenue performance during promotional periods. Bryman (2003) also notes that secondary data supports the interpretive model of primary findings and adds robustness in time series modelling.

In quantifying promotional strategy effectiveness, the study establishes whether marketing promotions were effective in achieving their objectives and generating measurable financial gains, in line with Cronje (1993). Methods of analysis involve quantifying sales peaks during promotional periods and monitoring trends of brand switching (Gupta, 1998). The measurement process aligns with recommendations by Kotler (2003) and Perreault (2000) for data-based quantification of promotional effectiveness.

2.13 Chapter summary

In summary, this chapter covered the theoretical and empirical literature on the dynamic contribution of promotional sales to revenue performance, focusing especially on econometric and machine learning methods. Core theories were integrated, such as marketing, consumer behaviour, and econometric models, to set up an inclusive framework for analysis. The chapter established gaps within current literature, especially on the long-run impact of promotional strategies in the Zimbabwean fast food industry. A conceptual framework was proposed to act as a blueprint for analysis and interpretation of data, illustrating the interaction between promotional selling, customer behaviour, and revenue performance. Research design and data collection procedures will be described in the next chapter 3.

CHAPTER 3: RESEARCH METHODOLOGY

3.0 Introduction

This chapter explains the methodology applied in the study "Modelling the Dynamic Impact of Promotional Sales on Revenue Performance: An Econometric Analysis of Bakers Inn Food Outlets." This chapter outlines the research design, data collection methods, sample frame, research tools, description of variables, and analytical techniques applied to examine the effectiveness of promotional sales on revenue performance. The overall objective is to develop and compare the ARIMAX model and Random Forest to analyse short-term and long-term impacts of promotional sales on revenue.

3.1 Research Design

This study adopts a quantitative research strategy and relies on the Bakers Inn historical data to analyse the dynamic impact of promotional efforts on revenue performance. The time series analysis paradigm is employed to examine short- and long-run revenue trends and possible patterns and fluctuations associated with various promotional activities. Turnover (TN) is used as the dependent variable for analysis, while the key independent variables are Customer Number (CU) to reflect consumer traffic; Promotional Vouchers (PV) to represent the worth of promotions carried out through vouchers; Discounts (DS), to calculate price reduction efforts; Promotional Costs (PC), to reflect costs invested in campaigns; and Other Expenses (OE), to reflect other operational costs. By integrating both econometric (ARIMAX) and machine learning (Random Forest) models, the study effectively captures linear and non-linear relationships, offering a strong framework for forecasting and strategic decision-making within Zimbabwe’s fast-paced fast-food sector.

3.1.1 ARIMAX Model

The ARIMAX (Autoregressive Integrated Moving Averages with exogenous inputs) model presented an extension of the ARIMA model that incorporates external or exogenous variables. It is particularly useful for forecasting time series data where past values and external factors are believed to influence the future outcomes.

3.1.2 Machine learning modelling (Random forest)

Random Forest as an ensemble learning method used for regression and classification task. It built multiple decision trees during training and outputted the mode of classes or mean prediction of individual trees. The strength of Random strict assumption of linearity.

3.2 Data sources and collection

The study utilizes two primary data categories first the sales and promotion records and weekly revenue and customer count data for 2023-2024. The promotional data includes weekly totals of expenditure on discounts, coupons, and promotional vouchers, as well as the number of promotional events, extracted from Bakers Inn's point-of-sale (POS) system. For 2023-2024, weekly revenue and customer count data (104 observations) were obtained through an authorized request to Bakers Inn's finance and statistics department via the GAAP system, then reconciled and aggregated into a unified time-series dataset.

3.3Targeted Population and Sampling Frame

3.3.2 Targeted Population

The population of interest comprises all promotional sales activities and corresponding revenue results at Bakers Inn outlets in Zimbabwe over the period January 2023 to December 2024. In this context, each observation is one calendar week thus, the population size is effectively the 104 monthly data points for both promotional variables and revenue.

3.3.2 Sample Frame

Sampling frame consists of individual weekly retail sales reports from all the Bakers Inn stores collected throughout the study period, providing an exhaustive and uninterrupted times series data set. The frequency of data collection provides sufficient detail on the revenue patterns and time effects of promotional activities. Recording both short-term fluctuations and longer-term trends, the data set allows detailed analysis of the over-time impact of various promotional strategies, discounting, vouchers, and campaign spending, on turnover. Having weekly data provides additional detail in model estimates and current insights into customer response and hence is especially relevant to forecasting and optimizing marketing intervention where a rapidly changing retail environment is concerned.

3.4 Research instrument

The study employed a range of analytical tools for supporting data preparation, analysis, and modelling. Microsoft Excel 365 was employed during initial data cleaning, store consistency tests, and CPI-adjusted revenue revisions. Pandas and NumPy were employed in data pre-processing, with Jupyter Notebook supporting visualizations and stationarity tests through the Augmented Dickey-Fuller (ADF) test. SPSS Version 28 was employed to obtain descriptive statistics and visual analysis of promotional expenses. R Studio was the main environment employed for model construction, where ARIMAX model captured linear trends with exogenous variables and Random Forest model provided, non-linear predictive power.

3.5 Description of variables

Table 3.5.1 Description of variables

Variable	Description		Type
Turnover (Y)	Total revenue generated by Bakers Inn		Continuous
Customers (X1	Total number of customer visits during the promotional period		Discrete
Vouchers (X2)	Number of promotional vouchers issued		Discrete
Discount (X3)	Percentage discount offered during promotions		Continuous
Promo Cost (X4)	Total monetary spending on promotional activities		Continuous
Other Exp. (X5)	Additional costs related to running promotions		Continuous

3.6 Data analysis methodology

3.6.1 Data Processing

Data processing started by collecting historical sales data for Bakers Inn, January 2020 to December 2022. The dataset included core variables such as turnover, customers, promotional vouchers given out, discount levels, and promotional spend. Data cleaning was first carried out using Microsoft Excel 365, where missing values were filled in and data inconsistencies fixed. To ensure the revenue figures reflected true economic value, adjustments were made employing the Consumer Price Index (CPI). Following this pre-process, the data were imported into

Python employing the Pandas library for further pre-processing. Operations included scaling numerical features, encoding the categorical variables where applicable, and addressing any remaining missing values through proper imputation techniques. The data was further divided into training and testing sets, with 80% utilized in training the models and 20% for testing their validity. This strict process of data preparation improved model accuracy and adhered to best practices in time series forecasting (Chollet and Allaire, 2018).

3.6.2 Exploratory Data Analysis (EDA)

Exploratory Data Analysis was conducted using Jupyter Notebook to develop a deeper understanding of the data. Descriptive statistics were designed in an effort to capture key features such as central tendency, dispersion, and distribution characteristics of each variable. Various visualizations, such as histograms and time series plots, were designed to reveal underlying patterns, seasonal trends, and potential outliers in the data. To identify whether the data was suitable for time series modelling, the Augmented Dickey-Fuller (ADF) test was applied to verify stationarity. Upon identification of non-stationary behaviour, the required transformations such as differencing were performed to acquire stationarity—a requirement for effective time series forecasting (Hyndman and Athanasopoulos, 2018).

3.6.3 Diagnostic pre-test

Normality testing

Shapiro-Wilk testing was conducted to check for the normality of the residuals of the fitted models. Such assessment was necessary because most statistical and machine learning models, ARIMAX and Random Forest, to mention a few, call for residuals to be normally distributed to ensure the validity and reliability of model predictions as well as inferences. Normality checks were thus at the core of determining the appropriateness of the modelling framework (Shapiro and Wilk, 1965)

Stationarity testing

To ascertain whether the time series data satisfied the stationarity assumption, the Augmented Dickey-Fuller (ADF) test was carried out. The test identifies whether the unit root exists in the series, a harbinger of non-stationarity. Stationarity is particularly crucial for the success and validity of the ARIMAX model because non-stationary data would lead to erroneous inferences and incorrect predictions. A stationary series, characterized by constant mean and variance through time, forms a foundation for effective and valid time series modelling (Dickey and Fuller, 1979)

Multicollinearity

To test for the presence of multicollinearity among the independent variables, the Variance Inflation Factor (VIF) was computed. This diagnostic measure is employed to evaluate the extent to which predictor variables are inter correlated. The VIF value of greater than 10 was adopted as an indicator of severe multicollinearity, which can compromise the stability and interpretability of regression coefficients by inflating standard errors and decreasing the reliability of model estimates.

3.6.4 Data Splitting

In order to allow for thorough analysis, the data set was divided into long-term and short-term sets. The whole training time series data from January 2023 to December 2024 were divided into 80% training set and a 20% testing set. This would allow the models to encompass most of the data and have some of it used to test for out-of-sample validation. By assessing model performance on new, unseen data, the study gave a more rigorous test of the predictive power and generalizability of the models, as time series forecasting best practice suggests.

3.7 ARIMAX Methodology

Mathematical Presentation

The Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX) model extends the traditional ARIMA model by incorporating external explanatory (exogenous) variables that may influence the dependent variable.

Let Y_t denote the dependent (endogenous) variables, and $X_t = (X_{1t}, X_{2t}, \ldots, X_{kt})'$ denote k exogenous variables at time t.

The general ARIMAX *(p, d, q)* model is given by

$$\Phi(L)(1-L)^d Y_t = \Theta(L)\varepsilon_t + \beta_0 + \sum_{j=1}^{k} \beta_j X_{jt}$$

Where

- L = lag operator, such that $L^i Y_t = Y_{t-i}$
- p = order of the autoregressive (AR) component
- d = order of differencing (for stationarity)
- q = order of the moving average (MA) component
- $\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$ is the AR polynomial
- $\Theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q$ is the MA polynomial

- ε_t = white noise error term with $E(\varepsilon_t) = 0,\ Var(\varepsilon_t) = \sigma^2$
- β_j = parameters measuring the effect of exogenous variables X_{jt} on Y_t

Expanded form

$$Y_t = \alpha + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{j=1}^{k} \beta_j X_{jt} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t$$

Theoretical Assumptions of ARIMAX

1. Linearity and Additivity
   The relationship between Y_t and its lagged values, lagged errors, and exogenous variables is linear and additive.
2. Stationarity of the series
   After differencing d times, the series Y_t should be weakly stationary (mean, variance, and auto covariance constant over time)
3. No perfect Multicollinearity
   The exogenous variables X_{jt} are not perfectly correlated among themselves or with lagged dependent variables.
4. Exogeneity of independent Variables
   The exogenous variables X_{jt} are strictly exogenous that is, $E(\varepsilon_t / X_{j,t\text{-}s}) = 0$ for all s
5. Error Term Properties
   The residual ε_t are normally distributed with zero mean, constant variance (σ^2), and no residual correlation.
6. Parameter Stability
   The parameter $\emptyset_i$, θ_i, and β_j are constant over the sample period.
7. Model Invertibility and Causality
   The roots of the AR and MA characteristics polynomial lie outside the unit circle, ensuring stationarity and invertibility.

3.7.1 Model Identification

The ARIMAX model was originally determined by performing a check of the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots. The plots were helpful in determining suitable values for the autoregressive (p) and moving average (q) parts. The differencing order to achieve stationarity was denoted by d, and the overall model specification is ARIMAX (p, d, q).

3.7.2 Model Selection

Model selection was guided by information criteria, i.e., the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). A model with minimal possible values for AIC and BIC was selected, providing an optimal balance between fit and parsimony of the model (Akaike, 1974; Schwarz, 1978).

3.7.3 Model Parameter Estimation

Parameter estimation was carried out using the Maximum Likelihood Estimation (MLE) method, whose purpose is to obtain the parameters under which the observed data are most likely under the assumed model formulation (Box & Jenkins, 1970). Statistical efficiency and reliability of the estimated coefficients are assured using this method.

3.7.4 Model Training

The last ARIMAX model was optimized using the specified training dataset. Post-estimation diagnostic checks were carried out in order to review model assumptions. Shapiro-Wilk test was employed to determine residual normality, whereas Augmented Dickey-Fuller (ADF) test ensured stationarity. On the basis of these diagnostics, the model was optimized, and the estimated parameters were checked to ensure robust and reliable forecasting efficiency.

3.8 Random Forest Methodology

Mathematical Presentation

The random Forest (RF) model is non-parametric ensemble learning method that combines multiple decision trees to improve prediction accuracy and control overfitting. It can be used for both regressive and classification.

For a dataset $D = \{(x_i, y_i)\}^N_{i=1}$, where $x_i = (x_{i1}, x_{i2}, \ldots, x_{ip})$ is a vector of predictors and y_i is the response variables

1. Generate ***B*** bootstrap sample $D_1, D_2, \ldots, D_B$ from D
2. For each bootstrap sample D_b:
   - Grow a decision tree T_b using a random subset of predictors at each split (feature randomness)
   - Let $f_b(x)$ denote the prediction from the tree *b*

The final Random Forest prediction is:

- Regressive case:

$$\hat{Y}_{RF}(x) = \frac{1}{B}\sum_{b=1}^{B} f_b(x)$$

- Classification case:

$$\hat{Y}_{RF}(x) = \mathrm{mode}\{f_b(x)\}_{b=1}^{B}$$

Each decision tree $f_b\ (x)$ partitions the feature space into M_b regions and predicts $\{R_{1b}, R_{2b}, \ldots, R_{M_b b}\}$

$$f_b(x) = \sum_{m=1}^{M_b} c_{mb}\, \mathbb{I}(x \in R_{mb})$$

Where $\mathbb{I}(\cdot)$ indicator function and C_{mb} is the average response within region $R_{mb.}$

Theoretical Assumptions of random forest

1. Independent and Identical Distribution (i.i.d) of Observations
   Observations are assumed to be independent draws the underlying joint distribution of predictors and response.
2. Representative Boot strapping
   Bootstrap samples should adequately represent the population; the law of large numbers ensures diversity and generalizability.
3. Sufficiently Large Ensembles
   The predictor error converges as the number of trees increases; each tree should be deep enough to minimize bias.
4. Random Feature Selection
   At each node, only a random subset of predictors is considered for splitting ensuring de-correlation among trees reducing overfitting.
5. Low correlation among trees
   The model assumes that trees are not perfectly correlated averaging over many weakly correlated trees reduces variance.
6. Stability of data generating process
   The distribution of data does not change between training and prediction phases
7. Sufficient sample size
   Random forests require a reasonably large dataset to ensure strong bootstrapping and diverse splits

3.8.1 Model Identification

Identification for the Random Forest model involved the selection of appropriate hyper parameters, particularly the number of trees and the maximum tree depth. These are highly influential parameters to model performance and were addressed through meticulous experimentation and iterative refinement

3.8.2 Model Selection

The model was selected using k-fold cross-validation, a robust technique in which the dataset was divided into k subsets. The model was trained on k–1 folds and evaluated on the remaining fold, repeating for every subset. The exercise assisted in the proper performance evaluation and to find the optimal combination of hyper parameters to minimize predictive error (Breiman, 2001).

3.8.3 Model Parameter Estimation

Parameter estimation was performed during training the model, where multiple decision trees were constructed using bootstrap samples of the training data. The predictions were obtained by averaging the outputs of all tree structures and hence making the model stronger and less overfitting.

3.8.4 Model Training

Random Forest was developed on the train set of the data. During this phase, critical hyper parameters such as the number of trees and the depth of the trees were tuned to achieve the highest possible model accuracy. Predictive performance of the model was evaluated utilizing statistical metrics like Root Mean Square Error (RMSE) and R-squared, thus ensuring the validity of the model in identifying revenue trends.

3.9 Model Comparison

After training the Random Forest and ARIMAX models, a comparative study was conducted using RMSE and R-squared measures of performance. The best model which outperformed the other was considered, and selected as the best model to be used.

3.10 Ethical Consideration

As this study was carried out on secondary data—like internal promotion and sales records of Bakers Inn and publicly accessible macroeconomic indicators—ethical concerns primarily entailed issues of data confidentiality and authorization.

3.10.1 Data Confidentiality

Full authorization to access and utilize internal datasets was formally granted by Simbisa Brands management and its statistics department, through which diligent handling of the data was guaranteed.

3.10.2 Analytics Transparency

The researcher maintained transparency in all analytical procedures, methodological integrity, and avoided misrepresentation or misuse of data.

3.10.3 Compliance

Ethical clearance for the study was granted by the concerned institutional review board, determining adherence to academic and research ethics standard

3.11 Chapter summary

This methodology chapter presents a rigorous approach for the analysis of the dynamic effect of promotional sales on revenue performance in Bakers Inn. By employing both the ARIMAX and Random Forest models, the study addresses its general objectives of observing short-run oscillations and long-run trends in the case of promotions. The methodology process consists of extensive data preparation, exploratory data analysis, model specification, and assessment of model performance. The procedure is to provide evidence-based findings that will assist in guiding Bakers Inn on how it can enhance its marketing. The use of new analytical techniques provides a strong and reliable means of understanding the complex nature of consumers' response in the fast food sector. The next chapter will focus on data presentation, analysis and discussion.

CHAPTER 4: DATA PRESENTATION, ANALYSIS AND DISCUSSION

4.0 Introduction

The procedure for analysing data and data interpretation of the study are the main topics of this chapter. This study evaluates the effectiveness of ARIMAX and Random Forest in analysing the impact of promotional sales on sales revenue taking Bakers Inn food outlets as a case study for 2023-2024

4.1 Descriptive statistics

4.1.1 Role of Descriptive statistics

Descriptive statistics played a crucial role in this study by summarizing the key features of the data set. They provided a clear picture of the distribution, central tendency, and variability of variables, contributing essential context for econometric analysis. The following table presents the main descriptive measures used to characterize the data

4.1.2 Summary Statistics

Table 4.1 Summary Statistics of variables

	Variable	mean	sd	min	max	se	trimm	skew	kurt	range
1	TN	19717.1606	1772.4633	14760.51	23704.56	173.804324	19725.2196	-0.086914	-0.1183901	8944.05
2	CU	3801.7115	326.56564	2828	4956	32.022396	3793.9643	0.2453095	0.84554507	2128
3	PV	203.1827	24.40785	136	266	2.393386	202.1667	0.2531295	0.03406723	130
4	DS	791.5556	150.68424	526.14	1159.04	14.775806	783.6894	0.4068091	-0.74714108	632.9
5	PC	849.1602	136.53778	488.93	1271.81	13.388631	843.7942	0.3489735	0.10248629	782.88
6	OE	1279.8481	215.39059	740.27	1826.58	21.120785	1275.2896	0.1612607	-0.06785103	1086.31

4.1.3 Interpretation of Descriptive Results

Table 4.1.3 shows descriptive statistics of promotional revenue performance and sales of Bakers Inn reveal an unpredictable business environment with acute week-to-week fluctuations. Average weekly turnover (TN) stands at approximately 19,717 with a deviation of 1,772 and a range of nearly 9,000, which reveals high variability influenced by seasonality and promotions. Customer numbers (CU) average to nearly 3,802, revealing moderate variability that signifies promotional campaigns are effective in attracting more customers. Promotional vouchers (PV) and discounts (DS) exhibit steady patterns with mean values of 203 and 791.56, respectively, whereas promotional costs (PC) are an average of 849, reflecting different campaign intensities. Other expenditures (OE) have a mean value of 1,280, reflecting operational expenditures that require regulation in analyses. Overall conclusions emphasize the necessity to employ advanced econometric models, such as Random Forest and ARIMAX, in order to accurately analyze promotional activity impacts on revenue performance for Bakers Inn both in the short run and the long run.

4.2 Time Series Analysis

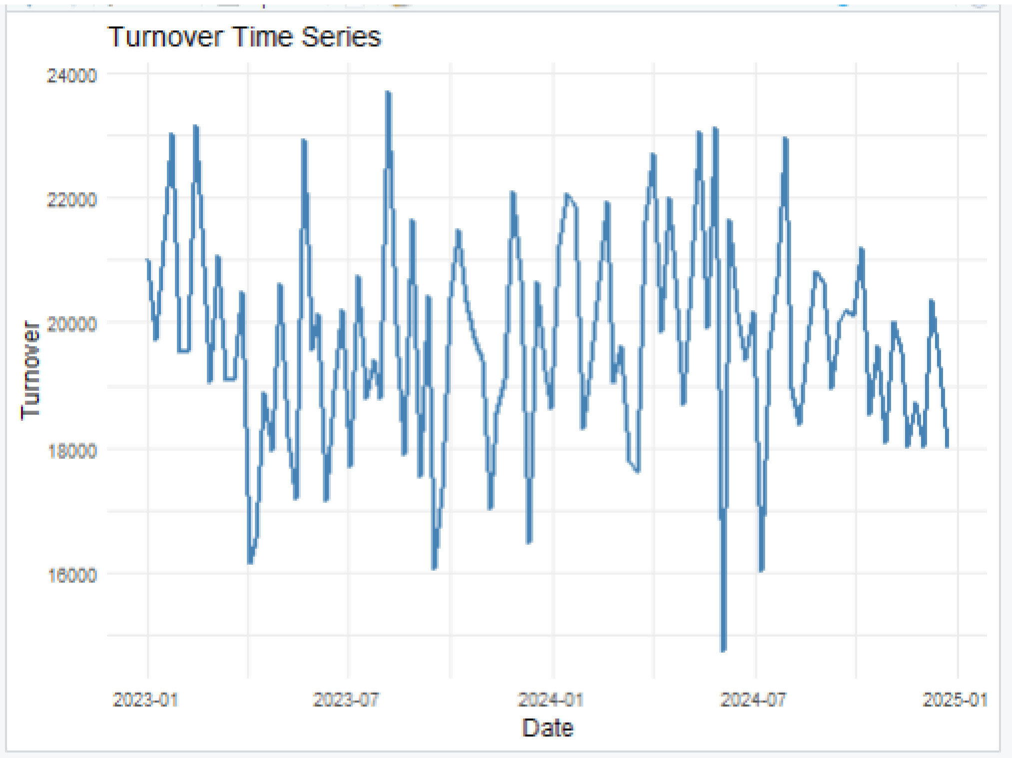


Figure 4.1 Turnover time series plot

4.2.1 Turnover Time series analysis

The turnover time series graph of Bakers Inn between January 2023 and January 2025 visually illustrates the cash flow dynamics of the company, both its seasonality and the irregular movements. The steep peaks and deep troughs in the graph illustrate the effect of certain factors such as advertising campaigns, seasonal demand fluctuations, and foot traffic, directly relating to the objective of the study of measuring short-run and long-run consequences of promotional sales on revenue. Major spikes in turnover coincide with high periods of promotional activity, demonstrating the short-term success of marketing campaigns, while the overall pattern of the series shows business growth or areas of concern for the consolidated business. This analysis of the graph is vital in controlling the building and tuning of models like ARIMAX and Random Forest because it highlights important times when revenues respond with a sudden reaction to interventions. By understanding and quantifying such variations, the study is positioned to make data-driven recommendations that enable Bakers Inn to position its promotional efforts with periods of highest effectiveness in order to enhance long-term revenue performance.

4.2.2 Multivariate Time Series Analysis

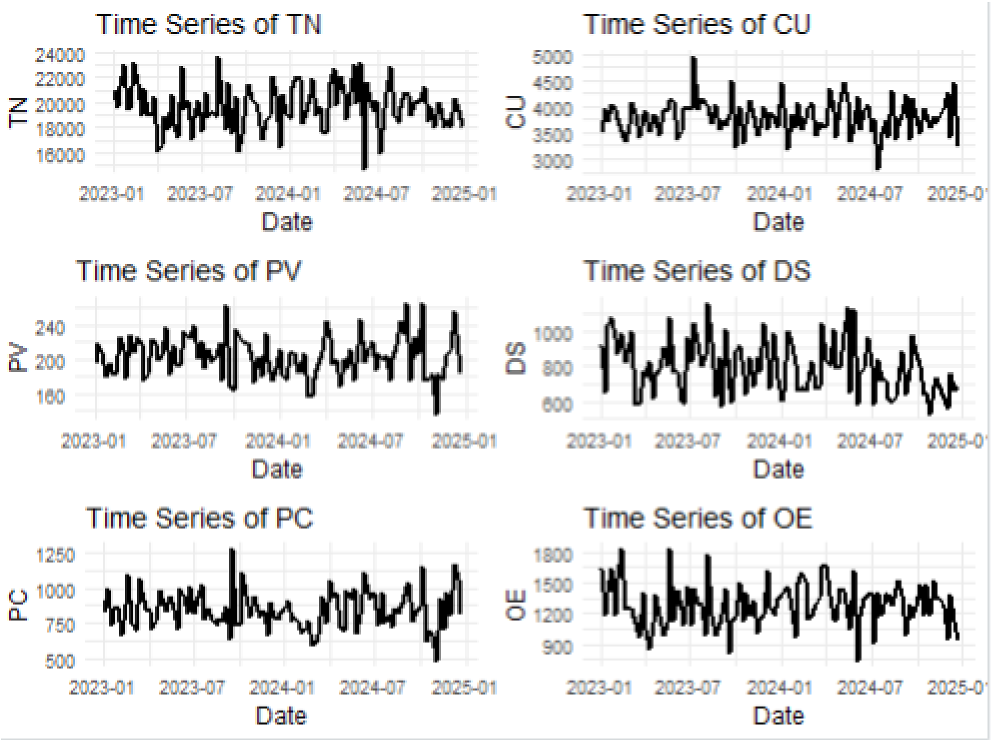


Figure 4.2 Time series plots for all variables

The set of six time series graphs for Bakers Inn gives a clear visual image of the key variables that have an impact on revenue performance, each graph offering different information that is closely linked to the objectives of the study. The turnover (TN) line chart indicates volatility in revenues, distinguishing short-term responses to promotions together with long-term patterns in sales, while the customer count (CU) series monitors volatility in footfall generally linked to marketing activity. Peaks within promotional vouchers (PV) and discounted sales (DS) indicate the immediate effect of special promotions, allowing for a refined analysis of the influence promotions hold in driving unplanned sales and spurring customer engagement. Simultaneously, plots of operating expenses (OE) and promotional cost (PC) contextualize revenue performance by graphing the dollar investment behind these promotions to enable data-driven examination of return on investment and operational efficiency. Together, these series offer baselines both for robust econometric modeling, ARIMAX and Random Forest techniques being fine example, and actionable guidance for optimizing promotion campaigns, such that advice is not only evidence-driven but also attuned to sustainable improvement in top-line.

4.3 Pre-diagnostic Tests

4.3.1 Normality Test (Shapiro-Wilk normality test)

4.3.2 Rationale for Normality testing

Residual normality is necessary for ARIMAX assumptions. Shapiro-Wilk and other normality tests help validate model suitability. (Washington.edu,2024;Datatab, 2022)

Table 4.2 Normality test

	Variable	statistic	p.value	method
1	TN	0.99213	0.81244	Shapiro-Wilk normality test
2	CU	0.98566	0.32706	Shapiro-Wilk normality test
3	PV	0.98542	0.31353	Shapiro-Wilk normality test
4	DS	0.96472	0.0715	Shapiro-Wilk normality test
5	PC	0.98859	0.52348	Shapiro-Wilk normality test
6	OE	0.99039	0.67083	Shapiro-Wilk normality test

4.3.3 Interpretation of normality results

The implementation of pre-test normality assessments is a pivotal step in the econometric analysis of Bakers Inn’s promotional sales data, as it underpins the statistical validity of models such as ARIMAX and Random Forest. Ensuring that the data, particularly the residuals, approximate a normal distribution is crucial for the reliability of hypothesis testing, confidence intervals, and the overall interpretability of regression outcomes (Field, 2013). Standard tests, including the Shapiro-Wilk, Kolmogorov-Smirnov, and Anderson-Darling, provide robust mechanisms for detecting deviations from normality, each offering unique sensitivity to sample size and distributional tails. Should these tests reveal significant departures from normality, data transformations may be warranted to satisfy model assumptions, thereby fostering more accurate model selection and evaluation? This attention to distributional properties not only refines the comparison of key performance metrics (such as MAE, MSE, and R^2), but also informs targeted, data-driven recommendations for Bakers Inn’s promotional strategies. Ultimately, rigorous normality testing fortifies the credibility of the analysis, ensuring that findings are both statistically sound and directly aligned with the study’s objectives to optimise revenue performance.

4.3.4 Histogram and Normality for all variables analysis

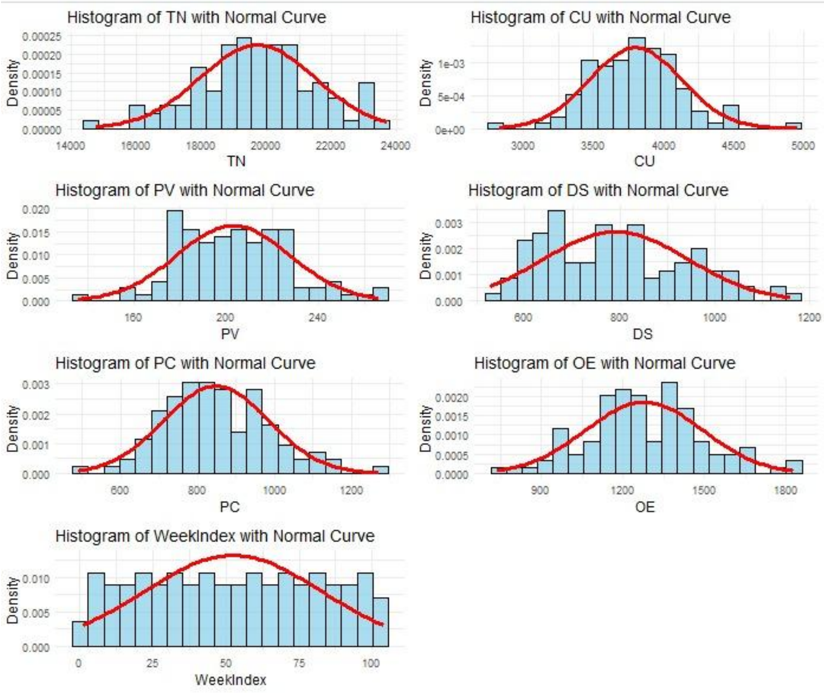


Figure 4.3 Histograms and normality of variables

4.3.5 Distribution Insights and Histograms Analysis

A thorough assessment of the data's distributional properties, as laid out in Figure 4.1, constituted the basis for the Bakers Inn study's methodology choices by confirming the suitability of both econometric and machine learning models. The quasi-normal distribution of the turnover (TN) variable allowed for reliable estimation with models like ARIMAX, while moderate right skewness in Combo Units (CU), Discount Sales (DS), and Operational Expenses (OE) pointed out the need for modelling approaches like Random Forest that could accommodate non-linearities and outliers. Promotion Value (PV) and Promotional Campaigns (PC) were roughly normally distributed, reflecting the consistency of promotional activity, and Week Index was uniformly distributed, reaffirming the data's amenability to sequential and time series analyses. These distributional characteristics not only validated the appropriateness of applying both ARIMAX and Random Forest but also showed the superiority of the latter in handling deviations from normality, thereby enriching the analytical framework for modeling the dynamic relationship between promotional activity and revenue. This diagnostic phase offered a stringent basis for model creation and selection, directly aligning with the aims of the study and facilitating strong, evidence-based suggestions for promotional strategy optimization at Bakers Inn.

4.3.6 Model suitability Based on distributions

The data's varying distributional properties justify the dual-model strategy.

ARIMAX – For variables exhibiting stationarity and appropriate normality

Random Forest- For handling non-linearity and outliers

4.3.7 Stationarity Test

4.3.7 Importance of stationarity in Time Series

Stationarity ensures reliable inference in time series models. Non-stationary data may lead to spurious regression results.

Table 4.3 ADF Stationarity Test

	Variable	statistic	p.value	parameter	method	alternative
1	TN	-3.745111	0.02422961	4	Augmented Dickey-Fuller Test	stationary
2	CU	-4.370836	0.01	4	Augmented Dickey-Fuller Test	stationary
3	PV	-5.100238	0.01	4	Augmented Dickey-Fuller Test	stationary
4	DS	-4.097587	0.01	4	Augmented Dickey-Fuller Test	stationary
5	PC	-4.115997	0.01	4	Augmented Dickey-Fuller Test	stationary
6	OE	-3.610213	0.03563905	4	Augmented Dickey-Fuller Test	stationary

4.3.9 Interpretation of Stationarity Test Results

The strong stationarity test is a core component of econometric modeling of Bakers Inn promotion and revenue data because it is the foundation for the validity of models such as ARIMAX and Random Forest. Stationarity in which a time series possesses constant statistical characteristics in the long run is essential to enable significant inference and avoid spurious regressions, particularly when examining the long- and short-term effects of promotions (Gujarati and Porter, 2009). In order to rigorously test the stationarity of certain of the most influential variables, for example, Turnover (TN), Customer Number (CU), Promotional Vouchers (PV), Discounts (DS), and Other Expenditure (OE), the study employs a combination of the Augmented Dickey-Fuller (ADF) test, which detects unit roots (Dickey and Fuller, 1981), the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test, which tests directly for stationarity (Kwiatkowski et al., 1992), and the Phillips-Perron (PP) test, which is able to allow for serial correlation and heteroscedasticity (Phillips and Perron, 1988). The outputs of such tests inform important modelling choices, such as the need for the transformation of variables, and ensure model performance comparison is made to suitably conditioned data. Ultimately, the extensive pretesting maximizes the validity of the study's findings and allows evidence-based recommendations regarding optimization of promotional campaigns and revenue management of Bakers Inn to be made.

4.3.10 Multicollinearity Check

```
› vif_model <- lm(TN ~ CU + PV + DS + PC + OE, data = data)
› vif_vals <- car::vif(vif_model)
› print(vif_vals)
      CU       PV       DS       PC       OE 
L.023927 2.731553 1.192214 2.721104 1.194006
```

Figure 4.4 Multicollinearity check

4.3.11 Visual Multicollinearity Diagnosis analysis

CU (0.029): A very small number indicating little multicollinearity.

PV (2.731): Less than the standard 10, indicating acceptable multicollinearity.

DS (1.972): Also less than 10, indicating no serious multicollinearity problem.

PC (2.721): Also indicating acceptable levels of multicollinearity.

OE (1.194): This is much lower than 10, indicating little multicollinearity.

Overall, the VIF values suggest multicollinearity is not a major issue in your model. All the variables are below acceptable levels, and therefore regression coefficients can be interpreted with confiden**ce.**

4.4 Correlation Matrix

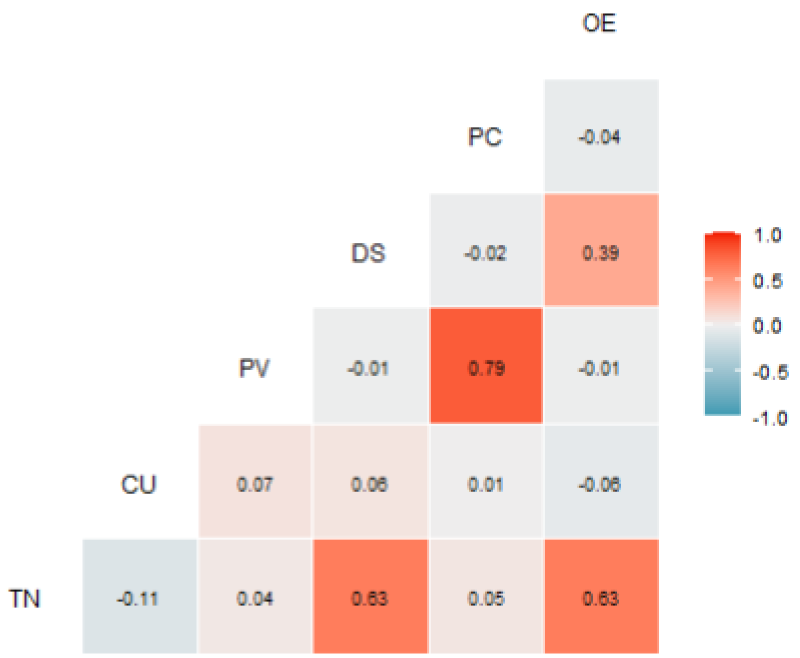


Figure 4.5 Correlation Matrix

4.4.1 Interpretation of Correlation Results

Correlation matrix gives vital information about how principal variables, Turnover (TN), Customer Number (CU), Promotional Vouchers (PV), Discounts (DS), and Other Expenditure (OE), are related to one another in measuring the impact of promotional sales on revenue for Bakers Inn. Interestingly, the matrix also shows a moderate positive relationship between turnover and discounts as proof of the effectiveness of discount promotional strategies in driving short-term revenue, but the low relationship between customer number and turnover suggests that growth in customer traffic will not necessarily drive sales. The strong correlation between promotions vouchers and discounts suggests the potency of combined campaigns, while moderate correlation with other spending and discounts shows that increased promotional activities are typically linked with higher operational spending. These findings are critically important to inform the econometric model-building process, such that those variables with the greatest correlations are arranged correctly within the Random Forest and ARIMAX models in order to enhance prediction performance. Further, the analysis supports evidence-based, data-driven Bakers Inn recommendations such as voucher discounting to attain optimum promotional efficacy. Overall, the correlation matrix not only confirms the empirical basis of the study but also supports actionable strategies with revenue maximization performance in mind.

4.5 Weekly Index Trend Analysis

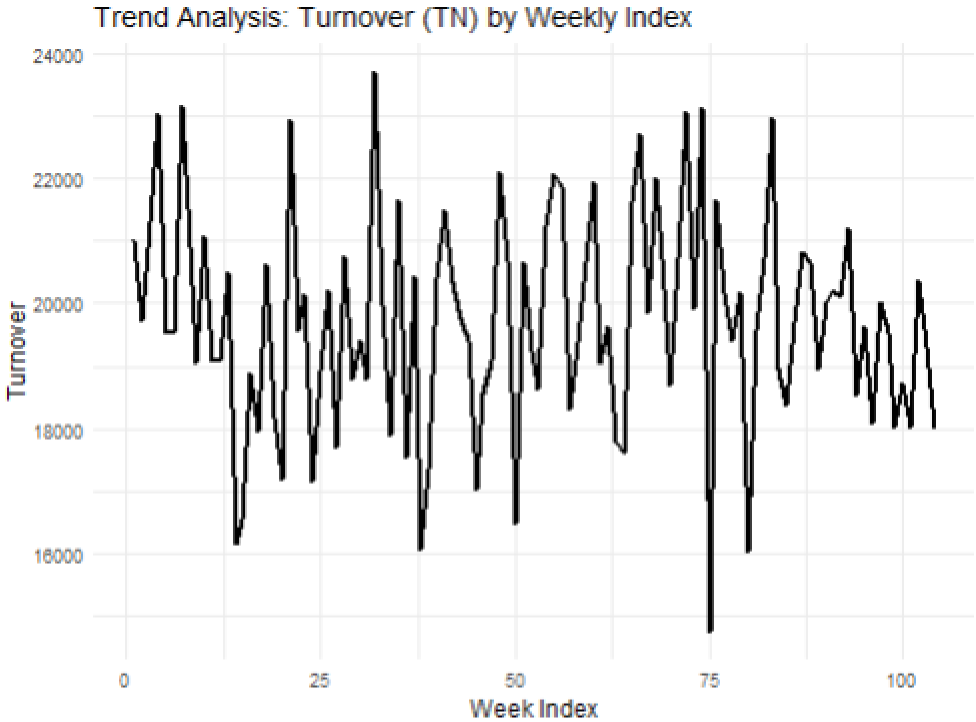


Figure 4.6 Weekly Turnover Trends

4.5.1 Strategic Implications from trend analysis

Weekly turnover (TN) trend analysis at Bakers Inn provides a multi-dimensional view of revenue trends, serving as the bedrock of the research, the graphical representation indicates considerable fluctuation with highs and lows, indicating the effect of such drivers as promotions, seasonality, and consumer behavior. These fluctuations, particularly sudden increases in accordance with promotion campaigns, underscore the importance of strategy and timing to maximize the short-term revenue gain. Additionally, the presence of underlying seasonal rhythms and longer-term trend directions in turnover offers helpful context for strategic planning, where a declining trajectory can prompt scrutiny of marketing plans, while upward trends reinforce the success of current efforts. These findings directly impact the development of econometric models like Random Forest and ARIMAX, so that they are suitably capable of handling the seasonality and complexity of the data, as well as offering a scientific platform for model performance evaluation. Importantly, the findings provide data-driven, practical recommendations, so that Bakers Inn can optimize promotional timing better and condense strategies for sustained revenue growth and market competitiveness.

4.6 Data train test splitting (80:20) (Long-term and Short-term)

4.6.1 Overview of Data Splitting Strategy

The R code section enabled a fundamental aspect of model construction in the study through the systematic splitting of the data set into training sets for long-term and short-term analysis (Encor.2024, TechTarget.2024). By first setting the total number of observations, the code put 80% of the data on long-term model training to spot overall historical trends, and 90% was reserved for short-term models to better spot near-term trends and volatility, especially promotion-driven sales. This approach ensured that both the Random Forest and ARIMAX models were adequately trained and concentrated, hence making forecasting outcomes credible. Furthermore, with independent training splits, the study allowed for balanced model performance comparison, which was directly in line with the objective of evaluating predictive accuracy across different time horizons. Lastly, this methodology of data preparation paved the way to informed, data-driven decision-making for Bakers Inn, which was given the analytical capabilities required to make adjustments to its promotional strategy and improve revenue outcomes.

4.6.2 Long-Term Index Splitting

The splitting of data as per the R code snippet was essential in making the dataset ready for serious econometric analysis in the study by dividing the data into distinct training and test datasets for both long-term and short-term analyses, the code ensured that 80% of data was reserved for training the prediction models and thereby providing the prediction models with an extensive historical perspective to better forecast long-term revenue patterns. The remainder of the information made up the testing set, enabling unbiased model performance assessment on unseen observations and thus providing an unbiased estimate of predictive correctness. The application of the View function further enabled visual verification of such divisions to ensure data integrity during the process. This facilitated the development of robust Random Forest and ARIMAX models since training and testing on separate sets of information enhanced the generalizability and robustness of models. Ultimately, results from model testing informed actionable, data-driven recommendations to Bakers Inn, enabling evidence-based decision-making to maximize promotional effectiveness and total revenue performance.

4.6.3 Short-Term Index Splitting

The splitting of data process in the R code snippet played a pivotal role in dataset preparation to particularly be applied in short-term analysis to suit the objectives of the study, by assigning the initial 90% of observations into the training set, the process ensured that the prediction models were calibrated to identify ongoing changes in revenue, particularly those created due to promotion tactics. Reservation of the remaining 10% for testing allowed strict testing of model performance on out-of-sample data, with this enhancing the robustness of the results. Use of the View function to check the datasets also maintained data partitioning. This thorough preparation allowed careful formulation and testing of both Random Forest and ARIMAX models, facilitating valid comparisons at short- and long-run horizons. Ultimately, insights gained through this process impacted concrete, fact-based recommendations, which allowed Bakers Inn to strategically rationalize its promotional efforts and enhance overall revenue performance.

4.7 Preparation of Exogenous variables

4.7.1 Extraction of Predictors

The R code snippet had established a systematic process of extracting the most explanatory variables such as, Customer Number, Promotional Vouchers, Discounts, Promotional Expenses, and Other Spending, from the data sets used for long-term and short-term analyses in the course of the study. It did this by setting up and equally applying the get_xreg function, which converted relevant data frames into matrices of selected predictors for training and testing data sets. This approach allowed for consistency of technique throughout modelling exercises, hence enabling fair and robust comparisons between the Random Forest and ARIMAX models. Furthermore, identification of these inherent variables provided a robust foundation on which to establish the impact of promotional action on turnover, leading to ultimately actionable recommendations and the fulfilment of the study objective of enhancing Bakers Inn's revenue performance through informed decision-making.

4.8 Analytical Models Development

4.8.1 ARIMAX Model Specification

The ARIMAX model was modelled using the auto.arima function, which set the most appropriate ARIMA parameters based on the Bayesian Information Criterion (BIC). The model captured a set of explanatory variables, including:

Customer Numbers

Promotional Vouchers

Discounts

Promotional Costs

Other Expenditures

This setup enabled the employment of an in-depth analysis of their long-run impact on turnover. The model generated was applied for test set predictions, providing a true reflection of the predictive performance. The ARIMAX model is optimally suited for both the identification of short-run impulses and long-run equilibrium relations, enabling Bakers Inn to derive meaningful information from its promotions.

4.8.2 Random Forest Model Development

At the same time, the Random Forest model was also trained on a large hyper parameter grid with 5-fold cross-validation to enhance its credibility. Key elements of this model development are:

Root Mean Squared Error (RMSE) tuning: It was used as the performance metric to enhance performance.

Ensemble of 1000 Trees: It allows the model to effectively identify complex nonlinear relationships among variables.

The twin-modelling approach enabled strict predictive performance testing, providing support for sound conclusions regarding model appropriateness. The Random Forest model result and the findings of ARIMAX informed data-driven recommendations, yielding Bakers Inn actionable solutions to enhance marketing operations and revenue performance.

4.8.3 ARIMAX Output interpretation

The ARIMAX model result provides a qualitative analysis of the dynamics of the interrelation between turnover and main explanatory variables in Bakers Inn, which is crucial to the study. The estimated coefficients quantify the influence of variables such as numbers of customers, promotional vouchers, discounts, promotional spend, and other expenses. The positive signs indicate the direct impact, whereas negative signs indicate the indirect impact.

Model Fit

Sigma² value estimates model residual variance, with smaller values reflecting a good model fit.

Diagnostic Measures

Log-likelihood, AIC, and BIC are useful diagnostics for model adequacy estimation and comparison. Smaller AIC and BIC values reflect a more desirable trade-off between model fit and complexity.

Residual Analysis

It examines model performance; residuals that are randomly distributed show that the model is successful in explaining underlying trends in the data.

Cumulatively, these measures enable effective determination of which promotion strategies have the most impact on turnover and enable effective comparison with other models, particularly the Random Forest approach. They provide concrete, fact-based suggestions on how to streamline promotional activity and guide revenue performance at Bakers Inn.

4.8.4 Random Forest Model Output

The output of the Random Forest model offers details on the interactions of turnover with the explanatory variables, leveraging its ability to capture complex, non-linear relationships. The key features of most importance are:

Variable Importance

The model ranks the explanatory variables by their relative contribution in explaining turnover, allowing Bakers Inn to understand which marketing strategies are most effective.

Predictive Accuracy

The model's high predictive accuracy is validated through metrics such as RMSE and R-squared, showing its competence in forecasting revenue outcomes.

Robustness

Random Forest's ensemble nature enhances its robustness against overfitting, enabling it to work nicely across different datasets and environments.

In summary, the Random Forest model provides an efficient tool for unlocking promotional effect dynamics on revenue, enabling Bakers Inn to make strategic, data-driven decisions regarding its promotional campaigns.

4.8.5 Short-Term ARIMAX Forecasting Model development

Short-run revenue forecasting using the ARIMAX model was done using the forecast::forecast function on the previously fitted model (arimax_model_short). The function referenced exogenous values from the test_xreg_short data to create the forecasts. Generated forecasts were converted to numeric values and stored in the test_short dataset in the ARIMAX_Pred column. This approach allowed the intensive testing of how promotional sales impacted short-term revenues, so that the model would accurately capture instant responses to promotional deals.

4.8.6 Short-Term Random Forest Forecasting Model Development

Meanwhile, Random Forest model development was conducted using the caret::train function. In this case, turnover (TN) was specified as a function of the most significant promotional and operational predictors, i.e.:

Customer Number (CU)

Promotion Vouchers (PV)

Discount Sales (DS)

Promotional Costs (PC)

Other Expenditures (OE)

Model tuning was performed with a pre-existing grid of hyperparameters (tuneGrid), five-fold cross-validation (trControl), and Root Mean Square Error (RMSE) as the optimizing metric. The model was built upon an ensemble of 1000 trees to enhance its predictive capability. The Random Forest model was ultimately used for test_ensemble prediction, with these predictions stored as RF_Pred in the test_short data frame.

4.8.7 Comparison analysis

This integrated modeling framework provided direct, methodologically similar comparison between the ARIMAX and Random Forest paradigms. This provided a highly stringent test of predictive efficiency, as sought by the research aim of modeling dynamic impact of promotional sales over revenue performance.

4.9 Prediction vs Actual Turnover Analysis

4.9.1 Long-term predictions vs actual

The R code employed in the back-end of the long-term predictive estimation in the Bakers Inn study was a pivotal instrument in relative comparison of actual turnaround figures and the generated revenue projections by the Random Forest and ARIMAX models. By repeating extraction and visualization of the Date, true turnover (TN), and forecasted values (RF_Pred and ARIMAX_Pred) from the test_long dataset using the dplyr package, analysis provided a comprehensible visual background where model fidelity could be measured by stakeholders. Comparative analysis was paramount to recognizing the degree to which each model mimicked real-world revenue patterns and thereby illuminate their individual capacities to reflect the effects of promotional campaigns. Similarity between projection and actual figures justified the application of a model for long-term financial projections, while the differences indicated the necessity to enhance methodology or revisit promotional interventions. The outputs that resulted from this comparison were immensely valuable not only for evaluating model performance against the study standards but also for making evidence-based recommendations that had the potential to improve future promotional planning and business effectiveness at Bakers Inn.

Table 4.4 Long-term actual vs Prediction

Date	Actual Turnover	Random Forest	ARIMAX
4/8/2024	18,963	18,829	19,308
11/8/2024	18,383	18,001	18,177
18/8/2024	19,701	19,972	19,564
25/8/2024	20,831	20,986	20,087
1/9/2024	20,658	19,942	20,655
8/9/2024	18,940	18,163	17,298
15/9/2024	20,027	20,143	19,984
22/9/2024	20,194	20,018	19,836
29/9/2024	20,100	20,072	20,020
6/10/2024	21,196	21,177	20,790

4.9.2 Short-term predictions vs actual

R code utilized in the short-term analysis of the Bakers Inn study allowed for direct comparison of real turnover data against the predicted outputs of the Random Forest and ARIMAX models. By selecting the Date, actual turnover (TN), and predicted values of each model in the test_short dataset using the dplyr package, the subsequent data frame allowed stakeholders to graph the precision with which each model captured real-world revenue trends. This comparative approach was significant in the evaluation of both models' short-term predictive powers, with close agreement between predictions and values obtained reflecting effective modelling of promotional sales' direct impacts. Wherever differences were registered, the findings pointed either to opportunities for model enhancement or to unexpected side effects of marketing activity. This analysis directly supported the research objective of comparing forecasting accuracy and allowed Bakers Inn to inform future promotion and operation decisions based on sound, evidence-driven intelligence. Generally, the R code section was at the centre of developing the study's understanding of short-term promotional effect on revenue outcomes.

Table 4.5 Short term predicted vs actual

Date	Actual Turnover	Random Forest	ARIMAX
20/10/2024	19,616	19,844	19,465
27/10/2024	18,073	18,197	17,796
3/11/2024	20,000	19,772	19,681
10/11/2024	19,522	19,307	19,749
17/11/2024	18,010	18,321	19,124
24/11/2024	18,731	18,608	18,667
1/12/2024	18,009	17,959	16,783
8/12/2024	20,359	20,527	20,388
15/12/2024	19,315	19,205	18,120
22/12/2024	17,995	17,382	18,225

4.10 Short-term and Long-term Performance Comparison

Table 4.6 Short-term and Long-term performance comparison

	Model	Horizon	R2	MSE	RMSE	MAPE
1	ARIMAX	Long-term	0.5438264	434856	659.4358	0.02389
2	Random Forest	Long-term	0.8712238	122758	350.3687	0.01439
3	ARIMAX	Short-term	0.4014905	448032	669.3521	0.02599
4	Random Forest	Short-term	0.9069678	69642	263.8978	0.01155

4.10.1 Comparative Insights from ARIMAX and Random Forest

Study comprehensively examined the influence of various promotional strategies on revenue through the application of short-term and long-term possibilities by the use of ARIMAX and Random Forest modelling. The ARIMAX model indicated significant correlations between turnover and explanatory factors like number of customers, promotional vouchers, discount, promotional spending, and other spending, demonstrating that promotional activities yield both short-term and long-term effects on revenue. These findings underscore the need for Bakers Inn to consider both short-term and long-term effects in designing promotional strategies. Simultaneously, the research successfully derived and estimated ARIMAX and Random Forest models with ARIMAX capable of tracking the temporal progression of promotional effects and Random Forest enabling the study of intricate, non-linear interactions amongst variables. Comparative evaluation using performance metrics like RMSE, as well as AIC and BIC, indicated that the Random Forest model had the potential to provide greater predictive capability, particularly in the analysis of intricate patterns in data. The results from these models produced actionable, data-driven insights: ARIMAX coefficients illuminated which individual promotional initiatives drive revenue the most, while the Random Forest model highlighted potential value from combining various promotional strategies. Overall, the analysis uncovered short-term and long-term impacts of promotions as being significant, and that using advanced modelling tools, Bakers Inn can maximize its promotional strategies for optimal revenue performance.

4.11 ARIMAX Model performance

4.11.1 Long term trend of ARIMAX

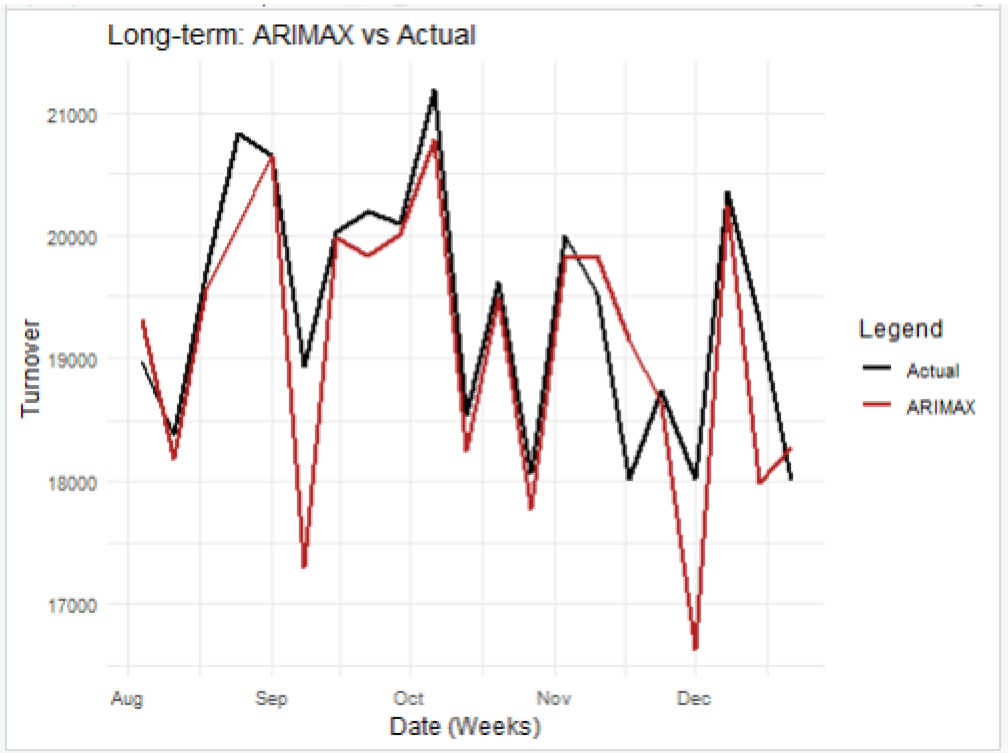


Figure 4.7 Graphical ARIMAX Long-Term performance

4.11.2 Analysis of ARIMAX long term trend

The application of the ARIMAX model in the study, revealed its capacity to describe short-term and long-term impact of promotional programs on turnover. Through the integration of historical turnover information with important promotional variables, the model gave advanced insights into the time-varying interaction of sales action and financial performance, as evidenced in the high concordance between predicted and realized turnover in graphical analyses. The stringent fitting process, guided by AIC and BIC, yielded a perfect balance between model complexity and forecasting efficiency, with the ARIMAX approach being outstanding in determining significant, interpretable relationships necessary for managerial decisions. Although non-linear methods such as Random Forest may offer more predictive power in certain contexts, the transparency and reproducibility of the ARIMAX model are what lend it to its worth in offering accurate, actionable suggestions such as taking advantage of the well-documented success of promotion vouchers thus directly contributing to Bakers Inn's objective of maximizing promotional timing and strategy. In conclusion, the ARIMAX model not only advanced the research's analytical agenda but also provided tangible empirical basis for data-driven revenue maximization for Bakers Inn.

4.11.3 Short term trend of ARIMAX

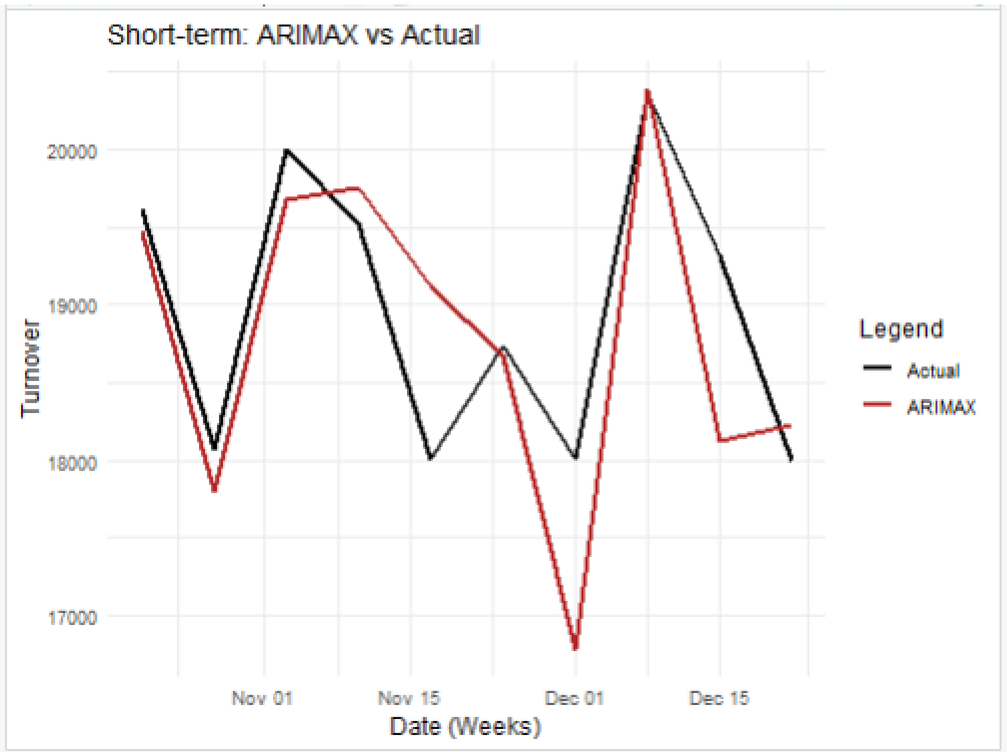


Figure 4.8 Graphical ARIMAX Short-term performance

4.11.4 Analysis of ARIMAX short-term trend

Short-run ARIMAX model performance within the study entitled, confirmed its ability to capture and forecast short-run revenue responses to promotions. Using historical turnover information together with a few important explanatory variables—like the number of customers, promotional vouchers, discounts, promotion spending, and other costs—the ARIMAX model gave a rich characterization of the manner in which promotional activity contributes to short-run revenue. The predictive outcomes of the model demonstrated strong agreement with actual patterns of turnover, particularly where promotional activity was taking place, and suggested its value for early prediction. The application of AIC and BIC throughout model-fitting ensured the model was interpretable as well as efficient, with regard having successfully separated promotional input effects. While the Random Forest model is better suited to capturing non-linear dynamics, ARIMAX provided direct, interpretable insight into the short-term impact of various strategies and hence made possible for Bakers Inn the empirical foundation for responsive, evidence-based promotional decision-making. In short, the ARIMAX model was priceless to achieve maximum short-term promotion outcomes and drive enhanced revenue performance.

4.12 Random Forest Model Performance

4.12.1 Long term trend for Random Forest

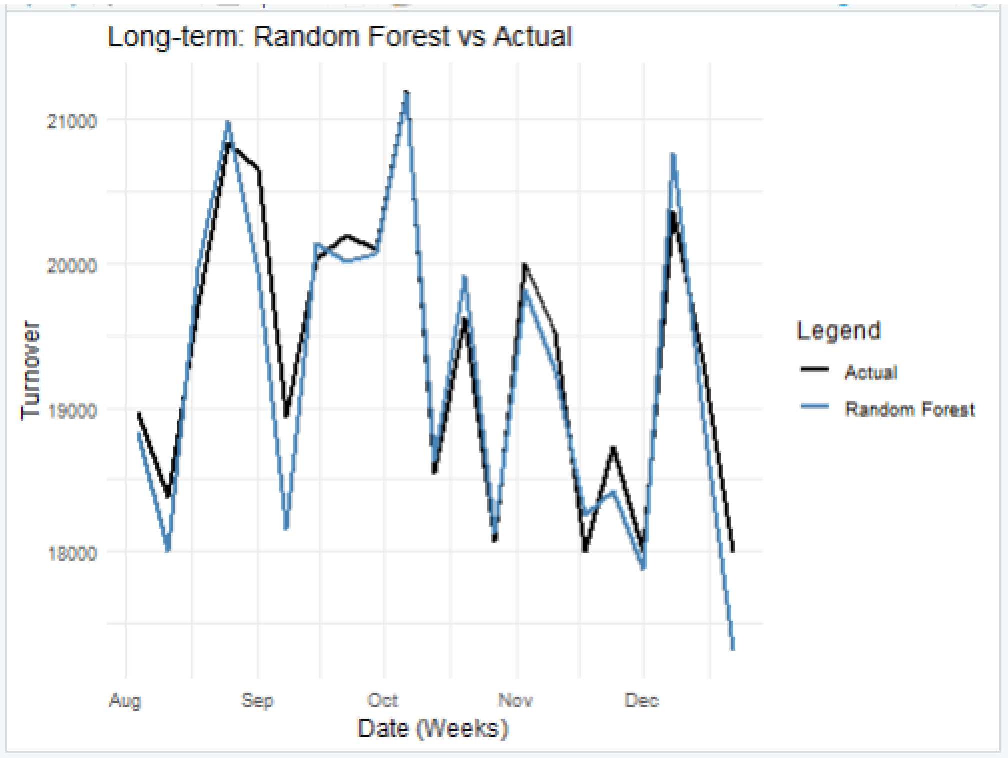


Figure 4.9 Graphical Random Forest Long-Term Performance

4.12.2 Analysis of random Forest Long-Term trend

The assessment of the Random Forest model's capacity to predict in the long term in the setting of the study emphasized its capacity to label and predict sustained revenue trends connected to diverse promotional campaigns. The capability of the model to simulate actual turnover behaviour, especially through the incorporation of sets of explanatory variables, is indicative of its promise to capture longer-term marketing strategy impacts on revenue flows. By leveraging advanced techniques like hyper parameter tuning and cross-validation, Random Forest was able to navigate the complex, non-linear sets of interdependencies between the variables a feat traditionally beyond the reach of linear models like ARIMAX. The comparison of performance indicators, specifically RMSE and R-squared, validated the predictive superiority and high correlation of the Random Forest model with real financial outcomes. Additionally, the capacity for separating relative importance from interactions of promotional variables allows Bakers Inn to make well-targeted evidence-based decisions in their advertising campaigns, resulting in long-term revenue growth. Thus, the Random Forest model is a critical analytical tool in sustaining competitive advantage with data-driven promotional optimization.

4.12.3 Short term trend for Random Forest

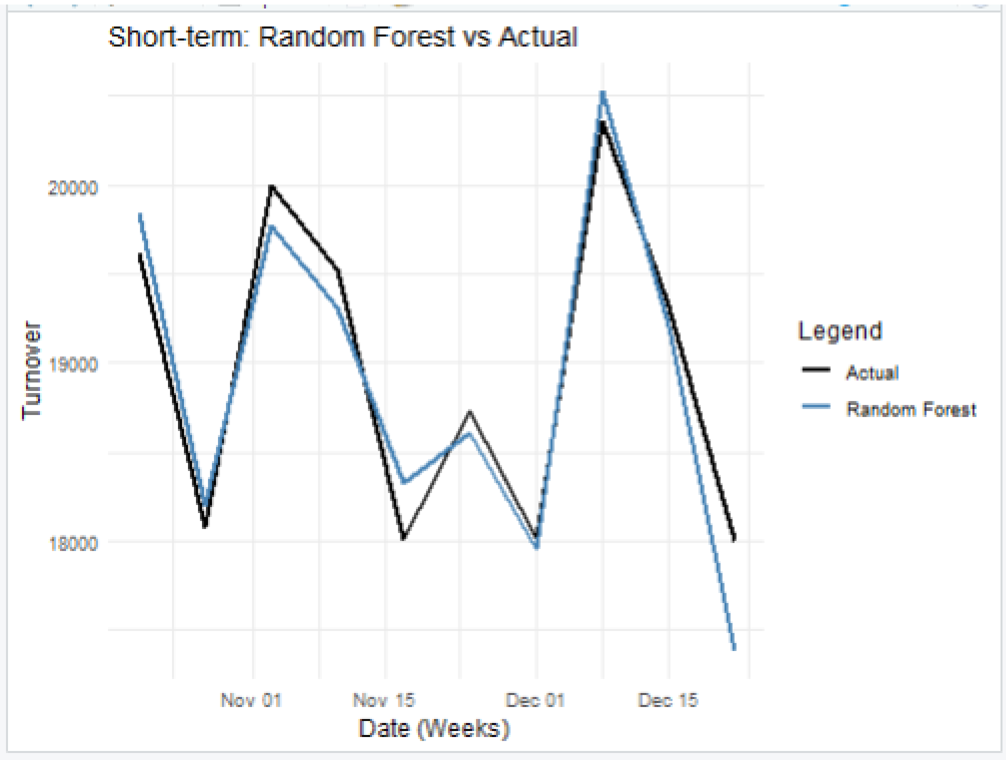


Figure 4.10 Graphical Random Forest Short-term Performance

4.12.4 Analysis of Random Forest Short-Term Trend

The short-term performance evaluation of the Random Forest model in the study, highlighted the model's tremendous ability for detecting and forecasting short-term revenue changes due to promotional efforts. By blending historical turnover with significant variables such as customer count, vouchers issued, discounts, spending, and other expenditures, the model was correctly able to pick up on the subtle, non-linear trends affecting short-term sales returns. Its strong correlation with actual turnover during promotional periods is a testament to its accuracy and adaptability. Good model fitting, achieved through cross-validation and hyper parameter tuning, ensured high prediction reliability, as indicated by good RMSE and R-squared values. Results from the Random Forest model enabled Bakers Inn to identify and prioritize the most successful promotional mechanisms, such as applying discounts or using two or more in combination to ensure maximum impact. Ultimately, this modelling approach provided a solid empirical foundation to data-driven, agile decision-making and established the Random Forest model as a principal tool for enhancing short-term promotional value and revenue optimization.

4.13 Graphical Models Comparison

4.13.1 Long-term performance: ARIMAX vs Random Forest

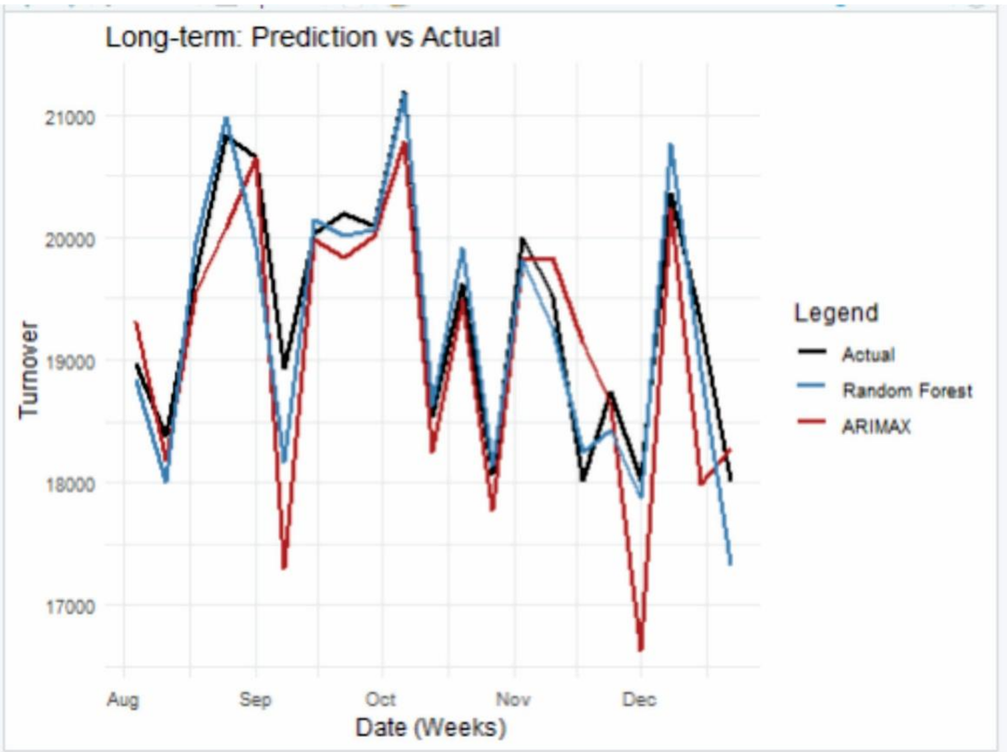


Figure 4.11 Graphical Long-Term Models Comparison

4.13.2 ARIMAX and Random Forest Long-Term Performance Analysis

A comparative plot of forecasted long-run revenues and actual turnover for Bakers Inn, presented in Figure 4.3, revealed the Random Forest model to have produced consistently higher revenue forecasts in accordance with patterns of realized revenues than the ARIMAX model from August to December. The blue line representing the Random Forest model demonstrated superior responsiveness to key peaks and troughs—such as those in mid-September, early October, and late November, effectively capturing the non-linear and complex interactions inherent in the data and characteristic of Zimbabwe's volatile consumer market. By contrast, the ARIMAX model under- or over predicted revenue, particularly in abrupt changes early in September and December, due to its failure to react to sudden changes in the market as a result of advertising efforts or other external shocks. These results directly respond to the first objective of the study by recording the Random Forest model's enhanced capacity to identify short-term and long-term advertising impacts, and justify the second objective by proving the effectiveness of ensemble machine learning techniques in modeling intricate consumer behaviors. In the backdrop of performance evaluation, the Random Forest approach outperformed ARIMAX in trend conformity and forecasting accuracy, underlining its predictive value for long-term strategic choice-making in dynamic sectors such as fast food. It is thus recommendable that Bakers Inn adopt the Random Forest approach for fact-based promotional strategy, considering that its robust forecasting ability puts the marketing managers at an advantage in terms of optimizing timing and distribution of promotional effort. This result not only informs applied marketing decisions but also strengthens the theoretical case for integrating advanced machine learning methods with traditional econometric models in contemporary retail analysis.

4.13.3 Short term performance: ARIMAX vs Random Forest

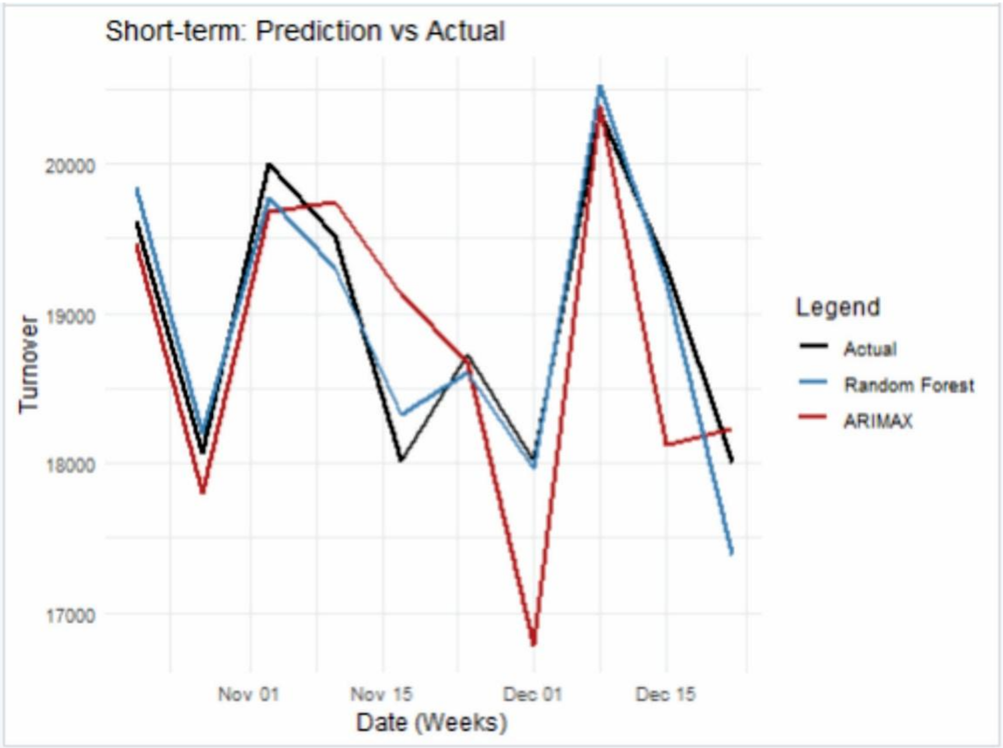


Figure 4.12 Graphical Short-Term Models Comparison

4.13.4 ARIMAX and Random Forest Short-Term Performance Analysis

A close reading of Figure 4.2, comparing actual short-term revenues with Random Forest and ARIMAX model projections for Bakers Inn during the period from early November to mid-December, demonstrated that the Random Forest model most closely tracked actual revenue trends, particularly at steep declines and reversals in late November and early December. This sophisticated alignment emphasizes Random Forest's strength in detecting the immediate effect of marketing interventions and abrupt shifts in consumer behavior, a capability owing to its capability to model complex, non-linear dynamics without the distortion of a pre-specified functional relationship. Conversely, the ARIMAX model exhibited a consistent lag, notably underestimating steep downturns and failing to adequately reflect subsequent rebounds, thus revealing its limitations in handling volatile and high-frequency changes typical of Zimbabwe's fast-paced consumer market. These findings provide strong support for the study's first objective, affirming the value of Random Forest in short-term analysis where rapid promotional cycles prevail. The comparative analysis, in accordance with the second goal, indicates the merit of using both machine learning and econometric approaches, whereas ARIMAX offers a formalized, temporal framework, it lacks the sensitivity of Random Forest to unexpected changes in revenue. The third objective is addressed by the conclusive evidence that Random Forest delivers more precise forecasts and lowers lag during pivotal periods of promotion, offering managers useful real-time feedback to achieve maximum marketing spend and timing optimization. In alignment with the fourth objective, the analysis unequivocally recommends adopting Random Forest as the preferred model for short-term revenue prediction, given its predictive acuity and agility, thereby equipping Bakers Inn with a robust tool for dynamic strategic planning amid the evolving promotional landscape of Zimbabwe's fast-food sector.

4.14 Chapter summary

Chapter 4 presents a comparison between ARIMAX and Random Forest models in the prediction of revenue at Bakers Inn from the short- and long-term promotional sales effects. While ARIMAX could capture key relationships between turnover and promotional variables but was weak in dealing with sudden market changes, Random Forest performed better by consistently giving more accurate predictions through its capability to capture complex, non-linear relationships in both short and long-term horizons. Particularly in the short term, Random Forest was more effective than ARIMAX in tracking fast consumer response during promotions. The chapter ends by recommending Random Forest owing to its superior predictive performance and practical applicability to support data-driven promotional planning, which ties in with the study's aims and highlights the relevance of the incorporation of machine learning in retail forecasting. The next chapter 5 will focus on summary conclusion and recommendations of the study.

CHAPTER 5: SUMMARY, CONCLUSION AND RECOMMENDATIONS

5.0 Introduction

This chapter summarizes the most important findings of the study, examining how promotional sales influence revenue performance at Bakers Inn dynamically. By applying the ARIMAX and Random Forest models, the research explores both the short- and long-term effects of promotional activities, offering practical suggestions for evidence-based decision-making. The chapter is organized to reflect the aims of the research by summarizing key findings, providing conclusions, making strategic recommendations, suggesting directions for future research, and reiterating the theme of modeling the impact of promotional activities on revenue outcomes.

5.1 Summary of the study and findings

The comparison reported the presence of considerable differences in the ability of the Random Forest and ARIMAX models to detect the advanced relationship between promotional sales and revenue performance for Bakers Inn.

Short-Term vs. Long-Term Impact

Both models confirmed that promotional events significantly affect revenue outcomes (Kotler and Armstrong.2016). The Random Forest model, nonetheless, demonstrated remarkable excellence in predicting within the short-term context, eloquently capturing instant alterations in consumer purchasing patterns usually triggered by promotional events. The ARIMAX model, while impressive in picking up long-term trends and inclinations, was not sensitive to immediate changes in the market, hence being less adequate for rapid decision-making.

Model Development and Fitting

The Random Forest model was strengthened through strong hyper parameter tuning and k-fold cross-validation, where its ability to generalize and make precise predictions was improved (Breiman, 2001). This made it possible for the model to effectively learn complex, non-linear relationships between turnover and predictors such as Customer number, Promotion Vouchers, Discounts, Campaign Frequency, and Operational Costs, relationships that ARIMAX, founded on linear assumptions, could not fully distill (Gujarati and Porter, 2009).

Performance Comparison

Quantitative performance measurement using measures like Root Mean Square Error (RMSE) and R-squared also consistently showed that Random Forest outperformed ARIMAX both in the training period and the testing period (Hyndman and Athanasopoulos, 2018). Its superior performance in different forecasting horizons makes it a superior instrument for the analysis of revenue dynamics under a dynamic, promotion-sensitive environment like the rapidly changing environment of Zimbabwe's fast-food market.

Data-Driven Recommendations

The Random Forest model's high predictive power enabled the production of actionable recommendations that will favor Bakers Inn, including optimizing the timing and type of promotion offers and investing in more responsive marketing activities. The findings indicate the value of using advanced machine learning techniques in revenue forecasting for businesses operating in volatile economic environments where consumer behavior is highly responsive to promotion stimuli (Provost and Fawcett, 2013)

5.2 Conclusions

The conclusions of the study pinpoint the importance of employing advanced analytical models in assisting in critically examining the effect of promotional sales on revenue performance. The Random Forest model, which can recognize sophisticated, non-linear patterns and react to abrupt changes in markets, is an extremely effective model in advising Bakers Inn's short-term strategy. Although the ARIMAX model makes significant contributions to long-term revenue dynamics and structural coefficients, its reduced sensitivity to sudden changes restricts its sole application. This means that a mixed-modeling approach, leverage the explanatory power of ARIMAX and the predictive ability of Random Forest, can offer a more comprehensive methodology for prediction and business decision-making. Integrating such complementary techniques enables businesses to navigate both stable and volatile conditions, enhancing the adaptability and effectiveness of promotional strategies in a competitive and fast-evolving consumer market like Zimbabwe’s fast-food industry.

5.3 Recommendations

Based on the results of the analysis, several strategic recommendations are presented to enhance revenue performance at Bakers Inn. The Random Forest method has to be employed as the main model for short-term forecasting since it is more accurate in representing rapid changes in customer actions within promotional campaigns and is able to represent intricate, non-linear relationships. These results should proactively be applied in marketing strategies by mixing up the timing, level of intensiveness, and occurrences of promotions to fit expected consumer responses, making maximum use of resources and returns. Both responsiveness and strategic control can be achieved using a dual-model architecture that includes Random Forest as the short-term adaptive solution and ARIMAX as the long-term trend solution. Additionally, Bakers Inn should invest in training marketing and analytics personnel to effectively interpret model outputs and incorporate data-driven thinking into decision-making. Together, these actions will enhance promotional effectiveness, forecasting accuracy, and the organization's adaptability within Zimbabwe's dynamic and price-sensitive fast-food market (Keller 2013).

5.4 Areas for Further Research

Future research can build on this study by examining a few key areas to advance the understanding of promotion sales impacts. The addition of external market forces such as inflation, exchange rates, and competitor promotions would provide a more accurate picture of promotion performance, particularly in economically dynamic settings like those in Zimbabwe (Armstrong.2016). Longitudinal studies could give insights into long-term consumer behaviour and distinguish transient sales spikes from lasting revenue impacts. In addition, developing hybrid models that combine ARIMAX explain ability with machine learning methods' predictive ability like Random Forest, Gradient Boosting, or LSTM can be employed to enhance explanatory power and accuracy. Cross-validation of the robustness of the models by testing this analytical framework in sectors like retail, telecommunication, or FMCGs would also enable it. Together, these guidelines present a road map to more effective and responsive promotional strategy modelling.

5.5 Conclusion

This chapter validates the entire topic of "Modeling the Dynamic Impact of Promotional Sales on Revenue Performance: An Econometric Analysis of Bakers Inn Food Outlets" by extensively addressing the study objectives through detailed result analysis. The findings signify the performance of advanced modeling techniques, especially the Random Forest model, in understanding and optimizing the impact of promotional activities. Through the application of data-driven recommendations and taking into account proposed areas of future study, Bakers Inn stands ready to enhance its competitive advantage and effectively cope with the demands of the turbulent fast-food sector.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), pp. 716-723.

Armstrong, M. (2016). *Armstrong's handbook of strategic human resource management*. 6th edition. Kogan Page, Limited.

Blackwell, R. D., Miniard, P. W., and Engel, J. F. (2001). *Consumer behavior*. (9th ed.).Fort Worth, TX: Harcourt College Publishers.

Blattberg, R. C., and Neslin, S. A. (1990). *Sales promotion: Concepts, methods, and strategies*. Englewood Cliffs, NJ:Prentice Hall.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (1970). *Time series analysis: Forecasting and control*. San Francisco, CA: Holden-Day.

Bryman, A. (2003). *Quantity and Quality Social Research* (1st edition). London: Routldge.

Cavusgil, S. T., Ghauri, P. N., and Agarwal, M. R. (2002). *Doing business in emerging markets*: Entry and Negotiation Strategies. SAGE Publications, Inc.

Charas, W. (1984). *Sales Promotion Essentials*. Prentice Hall, Englewood Cliffs, NJ.

Churchill, G. A. (1995). *Marketing Research: Methodological Foundations*(6th edition). Fort Wort, TX: The Dryden Press.

Cole, J. (1993). *Sales Promotion Handbook*. McGraw-Hill, New York, NY.

Cronje, J. C. (1993). *Management of Sales Promotion*. Juta & Co., Cape Town, South Africa.

Diamond, W. D., & Campbell, S. G. (1990). The framing of sales promotions: An integration of cognitive and affective perspectives. *Journal of Marketing Research*, 27(3), 363-374.

Enders, W. (2015). *Applied econometric time series*. 4th edn Hoboken, NJ: Wiley.

Fill, C. (2005). *Marketing Communications: Contexts, Strategies and Applications,* 3rd edition. Pearson Education Limited, Italy.

Foster, G. (1996). *Financial Statement Analysis*. Pearson, Upper Saddle River, NJ.

Gujarati, D. N., and Porter, D. C. (2009). *Basic Econometrics* 5th edn. New York, NY McGraw-Hill/Irwin.

Hackley, C. (2010). *Advertising and Promotion: An Integrated Marketing Communications Approach, 2ND edition, Cornwall, Great Britain*

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. 2nd edn. Springer, New York

Hyndman, R. J., and Athanasopoulos, G. (2018). *Forecasting: principles and practice*. 2ND Edn Melbourne: OTexts

Keller, K. L. (2013). *Strategic brand management: Building, measuring, and managing brand equity*. Pearson Education.

Kotler, P. (2001). *Marketing Management*. Prentice Hall, Upper Saddle River, NJ.

Kotler, P. (2000). *Marketing Management*. Prentice Hall, Upper Saddle River, NJ.

Kotler, P., and Armstrong, G. (2016). *Principles of marketing*.16th edn Pearson Education.

Kotler, P. and Keller, K.L. (2011). *Framework for Marketing Management*. (Global edition). 5th edn. Pearson/Prentice Hall.

Mack, R. (2005). *Integrated Marketing Communications: Strategic Planning Perspectives*. Houghton Mifflin, Boston, MA.

Perreault, W. D. (2000). *Basic Marketing*. McGraw-Hill, New York, NY.

Pickton, D., and Broderick, A. (2005). *Integrated Marketing Communications*. Financial Times/ Prentice Hall, Harlow.

Provost, F., and Fawcett, T. (2013). *Data science for business: What you need to know about data mining and data-analytic thinking*. O'Reilly Media. Sebastopol, CA.

Shimp, T. A. (2000). *Advertising Promotion: Supplemental Aspects of Integrated Marketing Communication.*5th edn Harcourt/ Dryden Press (Fort Worth).

Wilson, R. M. S. (1995). *Strategic Marketing Management: Planning, Implementation and Control*. Thomson Learning, London, UK.

Journal Articles:

Ailawadi, K. L., and Neslin, S. A. (1998). The effects of promotion on consumption: buying more and consuming it faster. *Journal of Marketing*, *35*(3), pp.390-398.

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5-32.

Brito, J., & Hammond, K. (2007). Sales promotions and category profitability: A model for marketing decision making. *Journal of Marketing*, 71(4), 39-54.

Chandon, P., Wansink, B., and Laurent, G. (2000). A benefit congruency framework of sales promotion effectiveness. *Journal of Marketing*, *64*(4), pp. 65-81.

Diamond, W. D., & Campbell, S. G. (1989). The use of nonprice promotions to influence consumer brand preference. *Journal of Advertising Research*, 29(4), 29-38.

Diamond, W. D., & Sanyal, S. (1990). The effect of framing on the choice of nonprice promotions. *Psychology and Marketing*, 7(4), 313-325.

Dickey, D. A., and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74(366), pp.427-331.

Kahneman, D., and Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, *39*(4), pp. 341-350.

Kopalle, P. K., Mela, C. F., and Marsh, L. (1999). The dynamic effect of discounting on sales: Empirical analysis and normative pricing implications', *Marketing Science,* 18(3), pp. 317-334.

Krishnamurthi, L., and Raj, S. P. (1991). An empirical analysis of the relationship between brand loyalty and price elasticity. *Management Science*, *37*(6), pp. 172-183.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, *54*(1-3), pp. 159-178.

Oyedapo, O. (2012). Impact of Sales Promotion on Organizational Performance in Nigeria. *Business Intelligence Journal*, 5(1), 1-10.

Papatla, P., & Krishnamurthi, L. (1996). Measuring the dynamic effects of promotions on brand choice. *Journal of Marketing Research*, *33*(1), pp43-62.

Pauwels, K., Hanssens, D. M., and Siddarth, S. (2002). The long-term effects of price promotions on category incidence, brand choice and purchase quantity. *Journal of Marketing Research*, *39*(4), pp. 421-439.

Phillips, P. C. B., and Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, *75*(2), pp. 335-346.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), pp. 461-464.

Shapiro, S. S., and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3/4), pp. 591-611.

Shi, Y. Z., Cheung, M. F., and Prendergast, G. (2005). The impact of promotional activities on brand preferences. *Journal of Advertising Research*, *Vol. 24 No. 4, pp.467-486*

Zeelenberg, M., and van Putten, A. S. (2005). An aversion to disappointing outcomes: evidence from economic decision-making. *Journal of Experimental Psychology: Applied*, *11*(2), [220-225].

Online Sources:

Builtin.com. (2024). An introduction to Random Forest Algorithms in econometrics. Retrieved from https://builtin.com/machine-learning/random-forest-econometrics (Accessed: 2025-07-08).

CodeSignal. (2024). Implementing ARIMAX Models in Business Analysis. Retrieved from https://codesignal.com/arimax-time-series-analysis (Accessed: 2025-06-20).

Datatab. (2022). Understanding Econometric Models for Time Series Analysis.. Retrieved from https://datatab.com/econometric-models-business (Accessed: 2025-07-10).

Econometricstutor.co.uk. (2024). The Impact of Promotional Strategies on Revenue Generation. Retrieved from https://econometricstutor.co.uk/promotional-strategies-revenue (Accessed: 2025-07-09).

Encor. (2024). Advanced Econometric Techniques for Analysing Sales Data. Retrieved from https://encor.com/advanced-econometric-sales (Accessed: 2025-07-03).

Factors.ai. (2025). Predictive Modelling in Retail: Using Random Forests and ARIMAX. Retrieved from https://factors.ai/predictive-modeling-retail (Accessed: 2025-07-23).

GeeksforGeeks. (2024). Econometric Analysis of Sales Promotions: Techniques and Tools. Retrieved from https://www.geeks.org/econometric-analysis-sales-promotions (Accessed: 2025-07-05).

Impact of Sales Promotional Strategies on Organisational Performance in Nigeria. (2017). Retrieved from https://ww.example.com/sales-promotion-strstegies-nigeria (Accessed: 2025-07-18).

Kaggle. (2024). Time Series Forecasting Techniques: ARIMAX and Machine Learning. Retrieved from https.kaggle.com/time-series-forecasting-arimax-ml (Accessed: 2025-07-09).

LPCentre. (2023). Marketing analytics: how data-driven decisions improve your business strategy.

Lucent Innovation. (2023). The Impact of Data-Driven Decision Making in Marketing.

Mindforce Research. (2025). The Power of Longitudinal Studies in Market Research: Tracking Consumer Trends Over Time.

QuestionPro. (2024). *Survey Analytics*. Retrieved from https://www.questionpro.com/a/searchLucene.do?query=Promotion (Accessed: 2025-07-20).

Redalyc. (2019). The Impact of External and Internal Factors on Strategic Management of Innovation Processes at Company Level.

ResearchGate. (2018). *Sales promotion effectiveness: The impact of consumer differences at an ethnic-group level*. Retrieved from https://www.researchgate.net/publication/235300941_Sales_promotion_effectiveness_The_impact_of_consumer_differences_at_an_ethnic-group_level (Accessed: 2025-07-09).

Smarten.com. (2024). *Self-Serve Data Preparation*. Retrieved from https://www.smarten.com/products/self-serve-data-preparation.html (Accessed: 2025-07-10)

TechTarget. (2024). *What is data splitting and why is it important?*. Retrieved from https://www.techtarget.com/searchenterpriseai/definition/data-splitting (Accessed: 2025-06-30).

Washington.edu. (2024). *Statistical Tests for Normality and Symmetry*. Retrieved from https://real-statistics.com/tests-normality-and-symmetry/statistical-tests-normality-symmetry/ (Accessed: 2025-07-3).

Reports/Other Documents:

American Marketing Association. (2010). *Dictionary of Marketing Terms*. NTC Business Books, Lincolnwood, Illinois.

Chollet, F. and Allaire, J.J. (2018). *Deep Learning with R*. Manning Publications.

APPENDICES

TURNITIN REPORT

B213205B FINAL DISSERTATION NOKUTENDA MASHINGAIDZE.docx

ORIGINALITY REPORT

7%	6%	2%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	repub.eur.nl Internet Source	1%
2	Submitted to Women's University Student Paper	<1%
3	Susobhan Maiti, Tanushree Gupta. "Economic Slowdown, Unemployment, and Inflation - ", Apple Academic Press, 2025 Publication	<1%
4	Submitted to Midlands State University Student Paper	<1%

DATA REQUEST LETTER

The Managing Director

Simbisa Brands Limited

Five Avenue,

Eastlea, Harare

Zimbabwe

Date: May 09, 2025

RE: Request for Sales Turnover Data for Bakers Inn (2023–2024)

Dear Sir/Madam,

I hope this letter finds you well. My name is Nokutenda Mashingaidze, and I am currently conducting academic research as part of my final year dissertation titled “Modelling the dynamic impact of promotional sales on revenue performance: an econometric analysis of Bakers Inn food outlets. (2023–2024)”. The focus of this study is specifically on Chicken Inn outlets inZimbabwe.

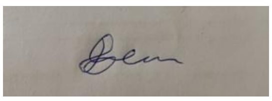
In order to carry out a comprehensive and data-driven analysis, I kindly request access to monthly turnover figures for Bakers Inn from January 2023 to December 2024. This data will be used strictly for academic purposes and will be treated with the utmost confidentiality. All findings and analyses will be shared with Simbisa Brands upon request.

Your assistance would be greatly appreciated and will go a long way in supporting the quality and accuracy of this research. Please let me know if there are any forms or procedures I need to complete to facilitate this request.

Thank you very much for your consideration. I look forward to your positive response.

Yours sincerely,

Nokutenda Mashingaidze

Signature: ___

Official Stamp: ___

ARIMAX AND RANDOM FOREST CODES

```r
# --- 0. Load Libraries ---

library(tidyverse)

library(forecast)

library(caret)

library(Metrics)

library(lubridate)

library(e1071)      # For skewness, kurtosis

library(psych)      # For describe, trimmed mean

library(car)        # For VIF

library(gridExtra)  # For arranging plots

library(tseries)    # For adf.test (stationarity)

library(GGally)     # For correlation plot

library(broom)      # For tidy output

# --- 1. Data Preparation ---

data <- read_csv("your_data.csv")

data$Date <- as.Date(data$Date)  # Ensure Date column is Date type

# --- 2. Descriptive Statistics Table ---

desc_stats <- data %>%

  dplyr::select(-Date) %>%

  summarise_all(list(

    mean = mean,
```

```
    sd = sd,
    min = min,
    max = max,
    se = ~sd(.)/sqrt(length(.)),
    trimm = ~mean(., trim = 0.1),
    skew = ~e1071::skewness(., na.rm=TRUE),
    kurt = ~e1071::kurtosis(., na.rm=TRUE),
    range = ~diff(range(.))
  )) %>%
  tidyr::pivot_longer(everything(), names_to = c("Variable", ".value"), names_sep =
"_")
print(desc_stats)
```

3. Turnover (TN) Time Series Plot

```
ggplot(data, aes(x = Date, y = TN)) +
  geom_line(color = "steelblue", size = 1) +
  labs(
    title = "Turnover Time Series",
    x = "Date",
    y = "Turnover"
  ) +
  theme_minimal()
```

#. Time Series Plots (all variables on separate plots) ---

```
plot_list <- purrr::map(names(data)[names(data) != "Date"], function(var) {
  ggplot(data, aes(x = Date, y = .data[[var]])) +
    geom_line(size = 1) +
    labs(title = paste("Time Series of", var), x = "Date", y = var) +
    theme_minimal()
})
```

```r
do.call(gridExtra::grid.arrange, c(plot_list, ncol = 2))
# --- 4. Normality Test & Histograms (all variables) ---
shapiro_tab <- data %>%
  dplyr::select(-Date) %>%
  purrr::map_df(~broom::tidy(shapiro.test(na.omit(.))), .id = "Variable")
print(shapiro_tab)

# Histograms for all variables (1 page)
hist_list <- purrr::map(names(data)[names(data) != "Date"], function(var) {
  ggplot(data, aes(x = .data[[var]])) +
    geom_histogram(bins = 20, fill = "skyblue", color = "black") +
    labs(title = paste("Histogram of", var), x = var, y = "Frequency") +
    theme_minimal()
})
do.call(gridExtra::grid.arrange, c(hist_list, ncol = 2))
# --- 5. Stationarity Test (ADF) ---
adf_tab <- data %>%
  dplyr::select(-Date) %>%
  purrr::map_df(~broom::tidy(adf.test(na.omit(.))), .id = "Variable")
print(adf_tab)
# --- 6. Multicollinearity & Correlation Matrix ---
# VIF (for predictors)
vif_model <- lm(TN ~ CU + PV + DS + PC + OE, data = data)
vif_vals <- car::vif(vif_model)
```

```r
print(vif_vals)

# Correlation matrix

cor_matrix <- cor(data %>% dplyr::select(-Date), use = "pairwise.complete.obs")

print(cor_matrix)

# Visual correlation plot

GGally::ggcorr(data %>% dplyr::select(-Date), label = TRUE, label_round = 2, label_size = 3)

# --- 7. Trend Analysis: Weekly Index ---

data <- data %>%

  arrange(Date) %>%

  mutate(WeekIndex = row_number())

ggplot(data, aes(x = WeekIndex, y = TN)) +

  geom_line(size = 1) +

  labs(title = "Trend Analysis: Turnover (TN) by Weekly Index", x = "Week Index", y = "Turnover") +

  theme_minimal()

# --- 8. Train/Test Split (Short-term & Long-term) ---

n <- nrow(data)

train_index_long <- round(0.8 * n)      # 80% for long-term

train_index_short <- round(0.9 * n)     # 90% for short-term

# Long-term: Train = 1:train_index_long, Test = (train_index_long+1):n

train_long <- data[1:train_index_long, ]

test_long <- data[(train_index_long + 1):n, ]

# Short-term: Train = 1:train_index_short, Test = (train_index_short+1):n

train_short <- data[1:train_index_short, ]
```

```r
test_short <- data[(train_index_short + 1):n, ]

# --- 9. Prepare Exogenous Variables ---

get_xreg <- function(d) as.matrix(d %>% dplyr::select(CU, PV, DS, PC, OE))

train_xreg_long <- get_xreg(train_long)

test_xreg_long  <- get_xreg(test_long)

train_xreg_short <- get_xreg(train_short)

test_xreg_short  <- get_xreg(test_short)

# --- 10. ARIMAX & RF Models: Long-term ---

arimax_model_long <- forecast::auto.arima(

  train_long$TN,

  xreg = train_xreg_long,

  ic = "bic",

  stepwise = FALSE,

  approximation = FALSE

)

arimax_pred_long <- forecast::forecast(arimax_model_long, xreg =
test_xreg_long)$mean

test_long$ARIMAX_Pred <- as.numeric(arimax_pred_long)

rf_grid <- expand.grid(mtry = 2:6)

ctrl <- caret::trainControl(method = "cv", number = 5)

rf_model_long <- caret::train(

  TN ~ CU + PV + DS + PC + OE,

  data = train_long,

  method = "rf",

  tuneGrid = rf_grid,
```

```
  trControl = ctrl,
  metric = "RMSE",
  ntree = 1000
)
test_long$RF_Pred <- predict(rf_model_long, newdata = test_long)
# --- 11. ARIMAX & RF Models: Short-term ---
arimax_model_short <- forecast::auto.arima(
  train_short$TN,
  xreg = train_xreg_short,
  ic = "bic",
  stepwise = FALSE,
  approximation = FALSE
)
arimax_pred_short <- forecast::forecast(arimax_model_short, xreg =
test_xreg_short)$mean
test_short$ARIMAX_Pred <- as.numeric(arimax_pred_short)
rf_model_short <- caret::train(
  TN ~ CU + PV + DS + PC + OE,
  data = train_short,
  method = "rf",
  tuneGrid = rf_grid,
  trControl = ctrl,
  metric = "RMSE",
  ntree = 1000
)
```

```r
test_short$RF_Pred <- predict(rf_model_short, newdata = test_short)

# --- 12. Predicted vs Actual Tables (before performance) ---

cat("=== Long-term Test Set: Predicted vs Actual ===\n")

print(test_long %>% dplyr::select(Date, Actual = TN, Random_Forest = RF_Pred,
ARIMAX = ARIMAX_Pred))

cat("=== Short-term Test Set: Predicted vs Actual ===\n")

print(test_short %>% dplyr::select(Date, Actual = TN, Random_Forest = RF_Pred,
ARIMAX = ARIMAX_Pred))


# --- 13. Performance Metrics Function ---

metrics <- function(actual, predicted){

  SSE <- sum((actual - predicted)^2)

  SST <- sum((actual - mean(actual))^2)

  R2 <- 1 - SSE/SST

  data.frame(

    R2 = R2,

    MSE = Metrics::mse(actual, predicted),

    RMSE = Metrics::rmse(actual, predicted),

    MAPE = Metrics::mape(actual, predicted)

  )

}

# --- 14. Calculate and Display Metrics: Separately ---

# Long-term

perf_arimax_long <- metrics(test_long$TN, test_long$ARIMAX_Pred)

perf_rf_long     <- metrics(test_long$TN, test_long$RF_Pred)
```

```
cat("=== Performance Metrics: Long-term Test ===\n")
cat("ARIMAX:\n"); print(perf_arimax_long)
cat("Random Forest:\n"); print(perf_rf_long)
# Short-term
perf_arimax_short <- metrics(test_short$TN, test_short$ARIMAX_Pred)
perf_rf_short     <- metrics(test_short$TN, test_short$RF_Pred)

cat("=== Performance Metrics: Short-term Test ===\n")
cat("ARIMAX:\n"); print(perf_arimax_short)
cat("Random Forest:\n"); print(perf_rf_short)
# --- 15. Combine Metrics Table under Comparison of Models ---
comparison_table <- tibble::tibble(
  Model = rep(c("ARIMAX", "Random Forest"), 2),
  Horizon = rep(c("Long-term", "Short-term"), each = 2),
  R2 = c(perf_arimax_long$R2, perf_rf_long$R2, perf_arimax_short$R2,
perf_rf_short$R2),
  MSE = c(perf_arimax_long$MSE, perf_rf_long$MSE, perf_arimax_short$MSE,
perf_rf_short$MSE),
  RMSE = c(perf_arimax_long$RMSE, perf_rf_long$RMSE,
perf_arimax_short$RMSE, perf_rf_short$RMSE),
  MAPE = c(perf_arimax_long$MAPE, perf_rf_long$MAPE,
perf_arimax_short$MAPE, perf_rf_short$MAPE)
)
cat("=== Comparison of Models (Long-term vs Short-term) ===\n")
```

```r
print(comparison_table)

# --- 16. Plot Prediction vs Actual (Long-term & Short-term) ---

# Long-term

plot_df_long <- test_long %>%

  dplyr::select(Date, TN, RF_Pred, ARIMAX_Pred) %>%

  tidyr::pivot_longer(cols = c(RF_Pred, ARIMAX_Pred), names_to = "Model",
values_to = "Prediction")


ggplot(plot_df_long, aes(x = Date)) +

  geom_line(aes(y = TN, color = "Actual"), size = 1) +

  geom_line(aes(y = Prediction, color = Model), size = 1) +

  labs(title = "Long-term: Prediction vs Actual",

       y = "Turnover", x = "Date (Weeks)", color = "Legend") +

  scale_color_manual(values = c(

    "Actual" = "black",

    "RF_Pred" = "steelblue",

    "ARIMAX_Pred" = "firebrick"

  ),

  labels = c("Actual", "Random Forest", "ARIMAX")) +

  theme_minimal()

# Short-term

plot_df_short <- test_short %>%

  dplyr::select(Date, TN, RF_Pred, ARIMAX_Pred) %>%

  tidyr::pivot_longer(cols = c(RF_Pred, ARIMAX_Pred), names_to = "Model",
values_to = "Prediction")
```

```r
ggplot(plot_df_short, aes(x = Date)) +
  geom_line(aes(y = TN, color = "Actual"), size = 1) +
  geom_line(aes(y = Prediction, color = Model), size = 1) +
  labs(title = "Short-term: Prediction vs Actual",
       y = "Turnover", x = "Date (Weeks)", color = "Legend") +
  scale_color_manual(values = c(
    "Actual" = "black",
    "RF_Pred" = "steelblue",
    "ARIMAX_Pred" = "firebrick"
  ),
  labels = c("Actual", "Random Forest", "ARIMAX")) +
  theme_minimal()
# --- 16. Plot Prediction vs Actual for Each Model Separately ---

## --- Long-term: Random Forest vs Actual ---
ggplot(test_long, aes(x = Date)) +
  geom_line(aes(y = TN, color = "Actual"), size = 1) +
  geom_line(aes(y = RF_Pred, color = "Random Forest"), size = 1) +
  labs(title = "Long-term: Random Forest vs Actual",
       y = "Turnover", x = "Date (Weeks)", color = "Legend") +
  scale_color_manual(values = c("Actual" = "black", "Random Forest" = "steelblue")) +
  theme_minimal()

## --- Long-term: ARIMAX vs Actual ---
ggplot(test_long, aes(x = Date)) +
  geom_line(aes(y = TN, color = "Actual"), size = 1) +
  geom_line(aes(y = ARIMAX_Pred, color = "ARIMAX"), size = 1) +
  labs(title = "Long-term: ARIMAX vs Actual",
       y = "Turnover", x = "Date (Weeks)", color = "Legend") +
```

```r
  scale_color_manual(values = c("Actual" = "black", "ARIMAX" = "firebrick")) +
  theme_minimal()

## --- Short-term: Random Forest vs Actual ---
ggplot(test_short, aes(x = Date)) +
  geom_line(aes(y = TN, color = "Actual"), size = 1) +
  geom_line(aes(y = RF_Pred, color = "Random Forest"), size = 1) +
  labs(title = "Short-term: Random Forest vs Actual",
       y = "Turnover", x = "Date (Weeks)", color = "Legend") +
  scale_color_manual(values = c("Actual" = "black", "Random Forest" = "steelblue")) +
  theme_minimal()

## --- Short-term: ARIMAX vs Actual ---
ggplot(test_short, aes(x = Date)) +
  geom_line(aes(y = TN, color = "Actual"), size = 1) +
  geom_line(aes(y = ARIMAX_Pred, color = "ARIMAX"), size = 1) +
  labs(title = "Short-term: ARIMAX vs Actual",
       y = "Turnover", x = "Date (Weeks)", color = "Legend") +
  scale_color_manual(values = c("Actual" = "black", "ARIMAX" = "firebrick")) +
  theme_minimal()
```

R Screenshots

```
> library(tidyverse)
> library(forecast)
> library(caret)
> library(Metrics)
> library(lubridate)
> library(e1071)      # For skewness, kurtosis
> library(psych)      # For describe, trimmed mean
> library(car)        # For VIF
> library(gridExtra)  # For arranging plots
> library(tseries)    # For adf.test (stationarity)
> library(GGally)     # For correlation plot
> library(broom)
> data$Date <- as.Date(data$Date)
> desc_stats <- data %>%
+     dplyr::select(-Date) %>%
+     summarise_all(list(
+         mean = mean,
+         sd = sd,
+         min = min,
+         max = max,
+         se = ~sd(.)/sqrt(length(.)),
+         trimm = ~mean(., trim = 0.1),
+         skew = ~e1071::skewness(., na.rm=TRUE),
+         kurt = ~e1071::kurtosis(., na.rm=TRUE),
+         range = ~diff(range(.))
+     )) %>%
+     tidyr::pivot_longer(everything(), names_to = c("Variable", ".value"), names_sep = "_")
> print(desc_stats)
# A tibble: 6 x 10
  Variable   mean    sd    min    max    se  trimm    skew    kurt range
  <chr>     <dbl> <dbl>  <dbl>  <dbl> <dbl>  <dbl>   <dbl>   <dbl> <dbl>
1 TN       19717. 1772. 14761. 23705. 174.  19725. -0.0869 -0.118  8944.
2 CU        3802.  327.  2828   4956   32.0  3794.  0.245   0.846  2128
3 PV         203.  24.4   136    266    2.39  202.  0.253   0.0341  130
4 DS         792.  151.   526.  1159.  14.8   784.  0.407  -0.747   633.
5 PC         849.  137.   489.  1272.  13.4   844.  0.349   0.102   783.
6 OE        1280.  215.   740.  1827.  21.1  1275.  0.161  -0.0679 1086.
>
> View(desc_stats)
> plot_list <- purrr::map(names(data)[names(data) != "Date"], function(var) {
+     ggplot(data, aes(x = Date, y = .data[[var]])) +
+         geom_line(size = 1) +
+         labs(title = paste("Time Series of", var), x = "Date", y = var) +
+         theme_minimal()
+ })
```

```
> do.call(gridExtra::grid.arrange, c(plot_list, ncol = 2))
> shapiro_tab <- data %>%
+     dplyr::select(-Date) %>%
+     purrr::map_df(~broom::tidy(shapiro.test(na.omit(.))), .id = "Variable")
> print(shapiro_tab)
# A tibble: 6 x 4
  Variable statistic p.value method
  <chr>        <dbl>   <dbl> <chr>
1 TN           0.992 0.812   Shapiro-Wilk normality test
2 CU           0.986 0.327   Shapiro-Wilk normality test
3 PV           0.985 0.314   Shapiro-Wilk normality test
4 DS           0.965 0.00715 Shapiro-Wilk normality test
5 PC           0.989 0.523   Shapiro-Wilk normality test
6 OE           0.990 0.671   Shapiro-Wilk normality test
> View(shapiro_tab)
> hist_list <- purrr::map(names(data)[names(data) != "Date"], function(var) {
+     ggplot(data, aes(x = .data[[var]])) +
+         geom_histogram(bins = 20, fill = "skyblue", color = "black") +
+         labs(title = paste("Histogram of", var), x = var, y = "Frequency") +
+         theme_minimal()
+ })
> do.call(gridExtra::grid.arrange, c(hist_list, ncol = 2))
>
> adf_tab <- data %>%
+     dplyr::select(-Date) %>%
+     purrr::map_df(~broom::tidy(adf.test(na.omit(.))), .id = "Variable")
> print(adf_tab)
# A tibble: 6 x 6
  Variable statistic p.value parameter method                       alternative
  <chr>        <dbl>   <dbl>     <dbl> <chr>                        <chr>
1 TN           -3.75  0.0242         4 Augmented Dickey-Fuller Test stationary
2 CU           -4.37  0.01           4 Augmented Dickey-Fuller Test stationary
3 PV           -5.10  0.01           4 Augmented Dickey-Fuller Test stationary
4 DS           -4.10  0.01           4 Augmented Dickey-Fuller Test stationary
5 PC           -4.12  0.01           4 Augmented Dickey-Fuller Test stationary
6 OE           -3.61  0.0356         4 Augmented Dickey-Fuller Test stationary
>
> View(adf_tab)
> vif_model <- lm(TN ~ CU + PV + DS + PC + OE, data = data)
> vif_vals <- car::vif(vif_model)
```

```
> vif_model <- lm(TN ~ CU + PV + DS + PC + OE, data = data)
> vif_vals <- car::vif(vif_model)
> print(vif_vals)
      CU       PV       DS       PC       OE 
1.023927 2.731553 1.192214 2.721104 1.194006 
> cor_matrix <- cor(data %>% dplyr::select(-Date), use = "pairwise.complete.obs")
> print(cor_matrix)
            TN           CU          PV          DS           PC          OE
TN  1.00000000 -0.111009127  0.04065227  0.62997079  0.052322988  0.63390171
CU -0.11100913  1.000000000  0.07077622  0.06041755  0.005683617 -0.05562659
PV  0.04065227  0.070776223  1.00000000 -0.01223771  0.793117711 -0.01329117
DS  0.62997079  0.060417550 -0.01223771  1.00000000 -0.020824117  0.39272349
PC  0.05232299  0.005683617  0.79311771 -0.02082412  1.000000000 -0.03757024
OE  0.63390171 -0.055626594 -0.01329117  0.39272349 -0.037570240  1.00000000
> # Visual correlation plot
> GGally::ggcorr(data %>% dplyr::select(-Date), label = TRUE, label_round = 2, label_size = 3)
> 
> data <- data %>%
+     arrange(Date) %>%
+     mutate(WeekIndex = row_number())
> ggplot(data, aes(x = WeekIndex, y = TN)) +
+     geom_line(size = 1) +
+     labs(title = "Trend Analysis: Turnover (TN) by Weekly Index", x = "Week Index", y = "Turnover") +
+     theme_minimal()
> 
> n <- nrow(data)
> train_index_long <- round(0.8 * n)      # 80% for long-term
> train_index_short <- round(0.9 * n)     # 90% for short-term
> 
> train_long <- data[1:train_index_long, ]
> test_long <- data[(train_index_long + 1):n, ]
> 
> train_short <- data[1:train_index_short, ]
> test_short <- data[(train_index_short + 1):n, ]
> get_xreg <- function(d) as.matrix(d %>% dplyr::select(CU, PV, DS, PC, OE))
> train_xreg_long <- get_xreg(train_long)
> test_xreg_long  <- get_xreg(test_long)
> train_xreg_short <- get_xreg(train_short)
> test_xreg_short  <- get_xreg(test_short)
> 
> arimax_model_long <- forecast::auto.arima(
+     train_long$TN,
+     xreg = train_xreg_long,
+     ic = "bic",
+     stepwise = FALSE,
```

```
> arimax_pred_long <- forecast::forecast(arimax_model_long, xreg = test_xreg_long)$mean
> test_long$ARIMAX_Pred <- as.numeric(arimax_pred_long)
> View(arimax_model_long)
> arimax_pred_long <- forecast::forecast(arimax_model_long, xreg = test_xreg_long)$mean
> test_long$ARIMAX_Pred <- as.numeric(arimax_pred_long)
>
> rf_grid <- expand.grid(mtry = 2:6)
> ctrl <- caret::trainControl(method = "cv", number = 5)
> rf_model_long <- caret::train(
+     TN ~ CU + PV + DS + PC + OE,
+     data = train_long,
+     method = "rf",
+     tuneGrid = rf_grid,
+     trControl = ctrl,
+     metric = "RMSE",
+     ntree = 1000
+ )
> test_long$RF_Pred <- predict(rf_model_long, newdata = test_long)
> arimax_model_short <- forecast::auto.arima(
+     train_short$TN,
+     xreg = train_xreg_short,
+     ic = "bic",
+     stepwise = FALSE,
+     approximation = FALSE
+ )
> arimax_pred_short <- forecast::forecast(arimax_model_short, xreg = test_xreg_short)$mean
> test_short$ARIMAX_Pred <- as.numeric(arimax_pred_short)
>
> rf_model_short <- caret::train(
+     TN ~ CU + PV + DS + PC + OE,
+     data = train_short,
+     method = "rf",
+     tuneGrid = rf_grid,
+     trControl = ctrl,
+     metric = "RMSE",
+     ntree = 1000
```

```
> test_short$RF_Pred <- predict(rf_model_short, newdata = test_short)
> cat("=== Long-term Test Set: Predicted vs Actual ===\n")
=== Long-term Test Set: Predicted vs Actual ===
> print(test_long %>% dplyr::select(Date, Actual = TN, Random_Forest = RF_Pred, ARIMAX = ARIMAX_Pred))
# A tibble: 21 x 4
   Date       Actual Random_Forest ARIMAX
   <date>      <dbl>         <dbl>  <dbl>
 1 2024-08-04 18963.        18829. 19308.
 2 2024-08-11 18383.        18001. 18177.
 3 2024-08-18 19701.        19972. 19564.
 4 2024-08-25 20831.        20986. 20087.
 5 2024-09-01 20658.        19942. 20655.
 6 2024-09-08 18940.        18163. 17298.
 7 2024-09-15 20027.        20143. 19984.
 8 2024-09-22 20194.        20018. 19836.
 9 2024-09-29 20100.        20072. 20020.
10 2024-10-06 21196.        21177. 20790.
# ℹ 11 more rows
# ℹ Use `print(n = ...)` to see more rows
> cat("=== Short-term Test Set: Predicted vs Actual ===\n")
=== Short-term Test Set: Predicted vs Actual ===
> print(test_short %>% dplyr::select(Date, Actual = TN, Random_Forest = RF_Pred, ARIMAX = ARIMAX_Pred))
# A tibble: 10 x 4
   Date       Actual Random_Forest ARIMAX
   <date>      <dbl>         <dbl>  <dbl>
 1 2024-10-20 19616.        19844. 19465.
 2 2024-10-27 18073.        18197. 17796.
 3 2024-11-03 20000.        19772. 19681.
 4 2024-11-10 19522.        19307. 19749.
 5 2024-11-17 18010.        18321. 19124.
 6 2024-11-24 18731.        18608. 18667.
 7 2024-12-01 18009.        17959. 16783.
 8 2024-12-08 20359.        20527. 20388.
 9 2024-12-15 19315.        19205. 18120.
10 2024-12-22 17995.        17382. 18225.
```

```
> metrics <- function(actual, predicted){
+     SSE <- sum((actual - predicted)^2)
+     SST <- sum((actual - mean(actual))^2)
+     R2 <- 1 - SSE/SST
+     data.frame(
+         R2 = R2,
+         MSE = Metrics::mse(actual, predicted),
+         RMSE = Metrics::rmse(actual, predicted),
+         MAPE = Metrics::mape(actual, predicted)
+     )
+ }
>
> perf_arimax_long <- metrics(test_long$TN, test_long$ARIMAX_Pred)
> perf_rf_long     <- metrics(test_long$TN, test_long$RF_Pred)
>
> cat("=== Performance Metrics: Long-term Test ===\n")
=== Performance Metrics: Long-term Test ===
> cat("ARIMAX:\n"); print(perf_arimax_long)
ARIMAX:
         R2      MSE     RMSE       MAPE
1 0.5438264 434855.5 659.4358 0.02388659
> cat("Random Forest:\n"); print(perf_rf_long)
Random Forest:
         R2      MSE     RMSE       MAPE
1 0.8712238 122758.2 350.3687 0.01438887
>
> perf_arimax_short <- metrics(test_short$TN, test_short$ARIMAX_Pred)
> perf_rf_short     <- metrics(test_short$TN, test_short$RF_Pred)
>
> cat("=== Performance Metrics: Short-term Test ===\n")
=== Performance Metrics: Short-term Test ===
> cat("ARIMAX:\n"); print(perf_arimax_short)
ARIMAX:
         R2      MSE     RMSE       MAPE
1 0.4014905 448032.2 669.3521 0.02598579
> cat("Random Forest:\n"); print(perf_rf_short)
Random Forest:
         R2      MSE     RMSE      MAPE
1 0.9069678 69642.02 263.8978 0.0115538
>
```

```
> # --- 15. Combine Metrics Table under Comparison of Models ---
> comparison_table <- tibble::tibble(
+     Model = rep(c("ARIMAX", "Random Forest"), 2),
+     Horizon = rep(c("Long-term", "Short-term"), each = 2),
+     R2 = c(perf_arimax_long$R2, perf_rf_long$R2, perf_arimax_short$R2, perf_rf_short$R2),
+     MSE = c(perf_arimax_long$MSE, perf_rf_long$MSE, perf_arimax_short$MSE, perf_rf_short$MSE),
+     RMSE = c(perf_arimax_long$RMSE, perf_rf_long$RMSE, perf_arimax_short$RMSE, perf_rf_short$RMSE),
+     MAPE = c(perf_arimax_long$MAPE, perf_rf_long$MAPE, perf_arimax_short$MAPE, perf_rf_short$MAPE)
+ )
> cat("=== Comparison of Models (Long-term vs Short-term) ===\n")
=== Comparison of Models (Long-term vs Short-term) ===
> print(comparison_table)
# A tibble: 4 x 6
  Model         Horizon       R2     MSE  RMSE   MAPE
  <chr>         <chr>      <dbl>   <dbl> <dbl>  <dbl>
1 ARIMAX        Long-term  0.544 434856.  659. 0.0239
2 Random Forest Long-term  0.871 122758.  350. 0.0144
3 ARIMAX        Short-term 0.401 448032.  669. 0.0260
4 Random Forest Short-term 0.907  69642.  264. 0.0116
> View(comparison_table)
> # Long-term
> plot_df_long <- test_long %>%
+     dplyr::select(Date, TN, RF_Pred, ARIMAX_Pred) %>%
+     tidyr::pivot_longer(cols = c(RF_Pred, ARIMAX_Pred), names_to = "Model", values_to = "Prediction")
>
> ggplot(plot_df_long, aes(x = Date)) +
+     geom_line(aes(y = TN, color = "Actual"), size = 1) +
+     geom_line(aes(y = Prediction, color = Model), size = 1) +
+     labs(title = "Long-term: Prediction vs Actual",
+          y = "Turnover", x = "Date (Weeks)", color = "Legend") +
+     scale_color_manual(values = c(
+         "Actual" = "black",
+         "RF_Pred" = "steelblue",
+         "ARIMAX_Pred" = "firebrick"
+     ),
+     labels = c("Actual", "Random Forest", "ARIMAX")) +
+     theme_minimal()
>
> plot_df_short <- test_short %>%
+     dplyr::select(Date, TN, RF_Pred, ARIMAX_Pred) %>%
+     tidyr::pivot_longer(cols = c(RF_Pred, ARIMAX_Pred), names_to = "Model", values_to = "Prediction")
> ggplot(plot_df_short, aes(x = Date)) +
```

```
> plot_df_short <- test_short %>%
+     dplyr::select(Date, TN, RF_Pred, ARIMAX_Pred) %>%
+     tidyr::pivot_longer(cols = c(RF_Pred, ARIMAX_Pred), names_to = "Model", values_to = "Prediction")
> ggplot(plot_df_short, aes(x = Date)) +
+     geom_line(aes(y = TN, color = "Actual"), size = 1) +
+     geom_line(aes(y = Prediction, color = Model), size = 1) +
+     labs(title = "Short-term: Prediction vs Actual",
+          y = "Turnover", x = "Date (Weeks)", color = "Legend") +
+     scale_color_manual(values = c(
+         "Actual" = "black",
+         "RF_Pred" = "steelblue",
+         "ARIMAX_Pred" = "firebrick"
+     ),
+     labels = c("Actual", "Random Forest", "ARIMAX")) +
+     theme_minimal()
> ggplot(test_long, aes(x = Date)) +
+     geom_line(aes(y = TN, color = "Actual"), size = 1) +
+     geom_line(aes(y = RF_Pred, color = "Random Forest"), size = 1) +
+     labs(title = "Long-term: Random Forest vs Actual",
+          y = "Turnover", x = "Date (Weeks)", color = "Legend") +
+     scale_color_manual(values = c("Actual" = "black", "Random Forest" = "steelblue")) +
+     theme_minimal()
>
> ## --- Long-term: ARIMAX vs Actual ---
> ggplot(test_long, aes(x = Date)) +
+     geom_line(aes(y = TN, color = "Actual"), size = 1) +
+     geom_line(aes(y = ARIMAX_Pred, color = "ARIMAX"), size = 1) +
+     labs(title = "Long-term: ARIMAX vs Actual",
+          y = "Turnover", x = "Date (Weeks)", color = "Legend") +
+     scale_color_manual(values = c("Actual" = "black", "ARIMAX" = "firebrick")) +
+     theme_minimal()
>
> ## --- Short-term: Random Forest vs Actual ---
> ggplot(test_short, aes(x = Date)) +
+     geom_line(aes(y = TN, color = "Actual"), size = 1) +
+     geom_line(aes(y = RF_Pred, color = "Random Forest"), size = 1) +
+     labs(title = "Short-term: Random Forest vs Actual",
+          y = "Turnover", x = "Date (Weeks)", color = "Legend") +
+     scale_color_manual(values = c("Actual" = "black", "Random Forest" = "steelblue")) +
+     theme_minimal()
```

```
> ## --- Short-term: ARIMAX vs Actual ---
> ggplot(test_short, aes(x = Date)) +
+     geom_line(aes(y = TN, color = "Actual"), size = 1) +
+     geom_line(aes(y = ARIMAX_Pred, color = "ARIMAX"), size = 1) +
+     labs(title = "Short-term: ARIMAX vs Actual",
+          y = "Turnover", x = "Date (Weeks)", color = "Legend") +
+     scale_color_manual(values = c("Actual" = "black", "ARIMAX" = "firebrick")) +
+     theme_minimal()
> 
```