BINDURA UNIVERSITY OF SCIENCE EDUCATION

FACULTY OF SCIENCE AND ENGINEERING

DEPARTMENT OF STATISTICS AND MATHEMATICS

Examining annual and seasonal variations in turnover within the food production sector: A Case Study of Simbisa Brands (Steers)

BY

GRACE HOVE

B1850409

TO BE SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR BSC.HONOURS IN STATISTICS AND FINANCIAL MATHEMATICS

PROJECT

2025

Authorship Declaration Statement

Tittle of the Thesis: Examining annual and seasonal variations in turnover within the food production sector: A Case Study of Simbisa Brands (Steers)

Author: Grace Hove

Program: Bachelor in science honours degree in statistics and financial mathematics.

I, the undersigned author of the above-mentioned thesis, hereby declare that:

1. This thesis is my original work and has been prepared by me in accordance with the institution's requirements.

2. All sources, data, and references used in this thesis have been acknowledged and cited appropriately.
3. This thesis has not been submitted elsewhere for any degree or diploma.
4. I have obtained all necessary permissions for the inclusion of third-party content where applicable.

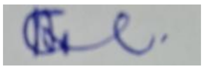
I affirm that this declaration is made with full integrity and in compliance with the institution's policies and academic practices.

Author's signature: 

Date: 19 June 2025

APPROVAL FORM

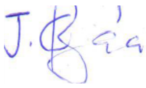
This is to certify, that this research project is the result of my own research work and has not been copied or extracted from past sources without acknowledgement. I hereby declare that no part of it has been presented for another degree in this University or elsewhere.



19/06/2025

GRACE HOVE **………………..** **……………….**

Student Signature Date

Certified by:  **19/06/2025**

…………………. **………………..** **……………….**

Supervisor Signature Date

Dr M Magodora… **…**  **…** **19/06/2025**

Chairperson Signature Date

DEDICATION

I dedicate this dissertation to my love. Your unwavering support, encouragement, and belief in my abilities have been instrumental in my academic success. Thank you for always being there for me, for instilling in me a love for learning, and for teaching me the value of perseverance. This accomplishment is as much yours as it is mine**.**

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to Ms J.C Pagan'a, my supervisor, for her invaluable guidance, encouragement, and patience throughout this research journey. Your insights and support have been instrumental in shaping this dissertation. I extend my heartfelt thanks to the faculty and staff of Science and Engineering, whose assistance and resources made the completion of this study possible. Special appreciation goes to my project coordinator Ms. P. Hlupo for her academic and moral support. To my family, thank you for your unwavering love, encouragement, and sacrifices. Your belief in me has kept me going even in the most challenging times. Special thanks for being my constant source of motivation. Lastly, to my friends and colleagues, your support, feedback, and companionship during this journey have meant more than words can express.

Thank you all.

ABSTRACT

This study focuses on investigating the seasonal and annual turnover variations of Simbisa Brands (Steers) in Zimbabwe from 2020 to 2024. This was achieved by employing advanced time series forecasting models to gain insights on the turnover dynamics. In particular, this research study compares the predictive performance of the Seasonal Autoregressive Integrated Moving Average (SARIMA) model and machine learning-based Extreme Gradient Boosting (XGBoost) algorithm. Several diagnostic tests were done to validate data reliability. Models were developed to Multiple accuracy metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2) were used for model evaluation. These accuracy metrics revealed that XGBoost significantly outperforms SARIMA in capturing complex, nonlinear, and seasonal turnover patterns. From the model results, XGBoost had a MAPE of approximately 15.49%, an RMSE of 0.568, and an R^2 of 67.94%, compared to SARIMA's 102.48%, 1.164, and 6.7% respectively. Graphical comparisons revealed that XGBoost closely follows turnover fluctuations, including peak and trough periods, whereas SARIMA's linear assumptions limit its predictive accuracy especially during structural breaks and sudden changes. The findings of this study suggest that machine learning models like XGBoost are better suited for dynamic, complex environments, thereby providing more reliable forecasts to inform production planning, inventory management, and strategic marketing decisions. This research advocates adopting to data-driven forecasting tools within the Zimbabwean fast-food industry. Overall, the study reinforces the transformative potential of integrating machine learning models into operational frameworks to optimize performance in the face of market volatility.

Contents

CHAPTER 1: INTRODUCTION

1. Introduction of the study

Due to seasonal and yearly changes in consumer demand, food manufacturing companies ideally represented by fast food restaurant chain Steers Zimbabwe belonging to Behemoth hotel chain Simbisa Brands Limited are defined by vast revenues fluctuations. Since they are critical to bottom-line performance, scheduling for manufacturing, as well as even controlling stock, such companies in the industry must be able to lift on such turn over patterns (Nielsen, 2018).

This research will apply time series analysis in identifying the trend of Steers Zimbabwe's turnover. To identify seasonal and annual trends, anomalies detection, and creation of a forecasting model to forecast future turnovers, the research takes on its history of Simbisa Brands. The results will provide useful insights to Simbisa Brands on food manufacturing industry turnover trends, Steers' turnover, in optimizing supply chain operations, production, and maintaining inventories.

Turnover fluctuation is controlled by, among others, consumer behavior, holiday seasons, and economic situations. The chapter is a literature review of studying yearly and festive seasons' variation in food production business turnover. The study employs sophisticated model analysis such as SARIMA and XGBoost in developing turnover trends and building a forecast model.

Later chapters are literature review (Chapter 2), where similar concepts and earlier research works are discussed. Chapter 3 demonstrates methodology, by mathematical and statistical data analysis. Chapter 4 demonstrates data collection process and analysis process. Chapter 5 concludes and summarizes findings and presents recommendations finally.

1.1 Background of the Study

Year-to-year and season-to-season fluctuations in turnover are typical in food manufacturing companies. They are the result of a complex mix of factors like fluctuations in customer demand, supply chain risk, and climatic risk. For example, seasonal patterns of festive period demand and climatic seasonality would lead to year-over-year fluctuations in turnover. Moreover, consumption habits and patterns of consumers also are great influencers of the patterns of turnover on a yearly basis (Thompson et al., 2019; Huang et al., 2018).

Seasonal fluctuation in turnover is most pronounced in production of seasonal products where climatic conditions can influence raw material price and supply and thus turnover

(Brewer et al., 2017; Liu et al., 2020). Companies such as Simbisa Brands (Steers) need to be aware of such monthly and seasonal fluctuations in order to adapt in production planning, perform resource allocation, and remain competitive within the fast-food restaurant sector.

1.1.1 Background of Steers

Steers is the most rapidly growing fast-food restaurant (QSR) brand in Africa operating under Simbisa Brands Limited. Steers started off in South Africa in 1960 and has grown to have more than 500 outlets on the continent, including in Zimbabwe. Steers' flame-broiled burgers, fries, and milkshakes are its top-selling items that it offers to mass consumers (Moyo, 2019). Steers has also been described as synonymous with convenience, good value for money, and quick quality food (Chikumbu, 2020).

Steers Zimbabwe falls under the auspices of Simbisa Brands Limited and boasts multiple units located strategically around the country. Simbisa's extensive chain of over 330 stores across the country is a testament to the company's expansion agenda on going (Simbisa Brands, 2023).

1.2 Problem Statement:

Although the general requirement of turnover seasonal and annual variation is not an issue, Simbisa Brands does not find it an issue to plan production and order lines accordingly:

• Not Enough Information on Seasonal Pattern of Demand There is no generic information available regarding seasonal pattern of demand variation, thus resulting in underproduction or overproduction.

• Ineffective Forecasting Processes the traditional forecasting processes may end up missing out on underlying patterns and variability in turnover figures and therefore settle for below-average performance estimates.

• Insufficient Research within the Fast Food Industry definitive research on seasonal and yearly trends in the African fast food industry is lacking, and much less so for Steers.

This research seeks to address such shortcomings since it will investigate seasonality and yearly volatility in turnover in Steers as a means of offering useful insights to Simbisa Brands and other interested parties. Top-level modelling methods, i.e., XGBoost and SARIMA, will be employed by this research in a quest to enhance the level of accuracy of forecasts and offer useful recommendations to Simbisa Brands.

1.3 Study Objectives:

1. Find and Examine Seasonal Trends: In order to examine seasonal trends of turnover at Steers, find the peak and off-peak seasons.
2. Investigate Annual Trends and Anomalies: In order to investigate annual turnover trends and identify anomalies in Steers, examining external variables such as economic conditions and consumer attitude.
3. Develop a Forecasting Model: To develop a forecasting model for driving future Steers' turnover that can aid effective production planning, inventory planning, and marketing planning.
4. Compare Forecasting Models: To compare SARIMA and XGBoost model performances in terms of forecasting Steers' turnover.

1.4 Research Questions:

1) What are Steers' seasonal variations in turnover, and how do they impact production planning and inventory planning?

2) What are Steers' annual trends and deviations in turnover, and what are the external factors affecting the changes?

3) Is it possible to develop a forecast model to predict future Steers turnover, and what inputs need to be incorporated in such a model?

4) Which model, XGBoost or SARIMA provides better predictions in terms of steers' turnovers?

1.5 Study scope

Geographical scope: Turnovers of Simbisa Brands (Steers) from north Zimbabwe, where Steers is strongly dominant, will be taken into consideration in this research study. Research will be conducted on one individual Steers restaurant from the northern area where the researcher has acquired internship experience for reasons of privacy. All Steers' annual and seasonal turnovers shall be taken into consideration, forecasting turnover trends from 2020-2024.

Contextual Scope: The study will take into account the impact of seasonality and year-round variability in turnover on Simbisa Brands (Steers) profitability and supply chain management. It will not consider other determinants of sales which would impact them, such as competitor behaviour or promotions.

Conceptual Scope: Time series analysis techniques will be used to identify the seasonal trends and annual patterns of turnover. Supply chain management principles will be used to establish how such fluctuations affect Simbisa Brands' (Steers) supply chain management and profitability.

Research Areas Covered:

• Determination of seasonal trends and annual trends in turnover for Simbisa Brands (Steers).

• Analysis of the effect of such annual and seasonal turnovers on profitability and supply chain management.

• Solutions for profitability maximization and supply chain management in light of variances observed.

Exclusions:

• Most likely key factors to impact sales, i.e., competitor action or marketing strategy.

• Effects on other business sectors, i.e., finance or human resources, of seasonal and annual turnover variation.

1.6 Significance of the Study.

Results in this study will greatly contribute to existing research in food production and prediction. With the utilization of advanced modelling Simbisa Brands, specifically Steers will be highly capable of making realistic decisions, manage stock, and become efficient in operations ultimately. Results of studies will also yield an ideal source for future research on this topic.

1.6.1 The Government

Policy intervention in the food production sector will be based on findings to ensure that season and volatility of annual turnover is tackled. The government can employ findings to create interventions that stabilize the sector during off-seasons, hence making it sustainable and resilient.

1.6.2 Other Students

This research will be of great use to students conducting research on such a matter in the food production sector. The result of this research will provide complete detail on the attainment of season and year-round turnover movement, therefore improving learning and practice in this field.

1.6.3 Company (Simbisa Brands - Steers)

The findings of the study will be of tremendous value in informing business decision-making within Simbisa Brands, for example, promotion, inventory management and supply chain management. The approach will increase effectiveness as well as profitability of business operations especially in cases of low demand, by applying the findings. The approach is consistent with the current focus by Simbisa of securing the optimum use of resources and competitiveness of the fast food business.

1.6.4 Industry Stakeholders

The results of this study will be beneficial to the majority of players in the food production industry, including distributors, suppliers, and logistics service providers. Knowledge of seasonality and year-on-year fluctuation in turnover will assist such organizations in planning operations, matching strategy to fluctuations in demand, and pushing out supply of services along the chain. Higher industry aggregate efficiency and responsiveness will depend on this collective achievement.

1.6.5 Consumers

The consumers will also benefit indirectly from the research, where it will observe their selection of foods on the shelves at any given moment of demand increase. Such information will cause the companies like Simbisa Brands (Steers) to manufacture the goods and services in a manner that will make them attractive to the consumers.

With this research, the researcher wishes to be in a position to contribute to literatures of seasonally and annually fluctuating turnover within the food manufacturing sector, in this instance, Simbisa Brands (Steers). The research will assist broad stakeholders ranging from the researcher, government, students, the organization, industry stakeholders, to consumers.

1.7 Study assumptions:

1. The study assumes the reliability and accuracy of the sales data provided by Simbisa Brands (Steers).

2. The study assumes that seasonal and annual changes in turnovers are mainly attributed to customers' demand and not to competitors' behaviors or promotional offers.

3. The study is established on an assumption that seasonal and annual changes in turnovers, have direct effects on Simbisa Brands (Steers) supply chain operation and profitability.

4. The study is predicated on an assumption that outcomes of this study would enable generalizations to be made to other food-producing companies.

1.8 Limitations of the Study:

Although this study is valuable information regarding the seasonality as well as year-round turnovers in Steers, one of the largest food manufacturing companies on the African continent, there are certain limitations which are to be noted:

1. This research is specifically for Simbisa Brands (Steers) and not for the general food manufacturing companies.

2. Data Quality: The research relies on past data which may prove to be uncertain or biased.

3. Limited Scope: As the research was only conducted on Steers, findings may not actually represent trends in the whole food manufacturing industry.

4. Environmental Factors: The study may or may not consider the impact of environmental factors such as economic conditions, consumer behavior, and climatic conditions on turnovers.

5. Research Methodology: Time series has been utilized for analysis, which may not be suitable for all kinds of data or for all kinds of research purposes.

6. Generalizability: The result of the current study cannot be extended to other companies or industries.

Limitation of the study is admitted to provide a true and realistic representation of the research findings.

1.9 Delimitation of the Study:

The seasonal and yearly turnover fluctuation of Steers, being one of Africa's top food-producing firms, is delimited as follows:

1. Geographical Coverage: The study delimits itself to Steers' operations on the African continent and does not include its operations outside Africa on the international level.

2. Time Frame: The study confines itself to a particular five-year time frame between 2020 and 2024 and does not address data for earlier or later years.

3. Product Lines: The core product lines of Steers are under study, and fresh and dropped product information are ruled out.

4. Sources of data: Secondary data sources such as historical sales data and business reports form the base of the study, and it does not comprise any primary data collection tools such as questionnaires or experimentation studies.

By acknowledging these limitations, the research depicts a clearly defined purpose and scope and establishes the research output.

1.10 Definition of key terms

Variation

Variation is a metric by which a variable, or variables, grows and decreases over time or under conditions (Hair et al., 2015). For this project, the term is used to explain the seasonal and yearly variations in the turnover that Simbisa Brands (Steers) has been experiencing over time.

Seasonal Variation:

Seasonal variation is where things vary in a predictable and recurring manner which happens at specific times of the year, e.g., specific seasons or festive seasons (Box and Jenkins, 1970). It is employed here to signify normal variations in turnover at Simbisa Brands (Steers) during periods such as festive seasons or summer seasons.

Yearly fluctuation refers to change or variation over a period of 12 months, typically due to market forces, economic conditions, and consumer behavior (Gujarati and Porter, 2018). It has been referred to in this paper as yearly turnover fluctuations in Simbisa Brands (Steers) due to such reasons.

Food Manufacturing Industry

Food processing is the business segment that packs, processes, produces, and sells food products like meat, beverages, milk, snacks, and edibles (Sullivan, 2019).

1.11 Summary of Chapter

Seasonal and annual periods of fluctuations in turnover in the food industry will be explored in this research, focusing on Simbisa Brands (Steers) as one of the top food manufacturing firms in Africa. Last year's sales trend analysis between 2020-2024 revealed robust seasonal trends where holiday seasons were when the sales were highest. Trends by years revealed overall trend upward with variation within specific years depending on outside factors like economic downturn and evolving consumer behavior. Time series was utilized to create a forecasting model for the purposes of projecting a successful future turnaround. The study highlights the importance of precise turnover forecasting in strategic decision-making and organizational performance. In the light of shortcomings of traditional forecasting practices, the study suggests using new methods SARIMA and XGBoost models. The study aims to complete literature gaps through extensive analysis of turnover trends as well as through model performance comparison. Research outcomes are hoped to provide good recommendations to Steers in an attempt to help improve production planning, inventory control, and company performance development. The abstract captures the main findings and implications of the research, reiterating Steers' emphasis on change awareness and turnover.

CHAPTER 2: LITERATURE REVIEW

2. Introduction

Food business is dynamic in nature with demand and supply fluctuations constantly changing over the span of a calendar year (Kumar et al., 2014). Seasonal and annual turnover fluctuations pose huge challenges to food producers on some of the weakest areas of business, such as planning for production, inventory management, and overall business performance (Taylor, 2008). Seasonal fluctuations in demand will pose a range of inventory problems, such as stockouts and overstocking, that will infuriate efficiency along with profitability, maintains Harris et al. (2015).

Literature has also identified many determinants of seasonal and annual turnover trends, such as climatic conditions, events, festivals, and shifting consumer preferences (Wang et al., 2017). For example, summer requires greater sales of items such as barbecue sauces and ice creams while comfort foods and hot beverages typically have greater sales in winter (Kumar et al., 2014).

In spite of the significance of gaining a sense of such patterns of turnover, there remains a significant lack of large-scale research specifically tackling this challenge in the general food manufacturing industry. In earlier research has been mainly interested in seasonal demand for specific products and ranges of products and has remained ignorant of the broader ramifications for operations planning and strategy for food manufacturing companies (Harris et al., 2015).

This review attempts to bridge this gap by examining existing research on seasonal and annual turnover variation in food processing industries. It explores the major determinants of weather, holiday seasons, and buyers' purchasing habits and examines their impacts on the most significant areas of business performance of production planning, inventory control, and overall business performance.

2.1. Theoretical literature.

Theoretical Frameworks Supporting Describing Annual and Seasonal Turnover Changes in the Food Manufacturing Sector. The theoretical frameworks supporting the description of annual and seasonal changes in turnover in the food manufacturing sector are discussed in this essay.

2.1.0 Theory of Seasonal Demand

Barbour (1973) speculated that there is the Theory of Seasonal Demand, which helps to describe how companies can deal with seasonality in demand. Barbour argues that fluctuations in demand across various seasons have a deep impact on businesses, especially food manufacturing businesses. The assumption underlying the hypothesis is that businesses need to realign levels of production and inventories to align with seasonal trends of demand. Food manufacturers, by analysing past trends in sales, are capable of envisioning such trends and restructure operations in such a way as to nullify losses and achieve peak profitability to the same level. This theory is best applicable to Simbisa Brands (Steers) which experiences extremely high seasonality of turnovers.

2.1.1 Bullwhip Effect

Lee et al. (1997) used the term Bullwhip Effect to describe why minimal fluctuations at the consumer demand level escalate to enormous fluctuations in inventories and production planning. The theory is best applied in food processing where demand can be extremely volatile both seasonally and yearly. Lee et al. (1997) determined that slight alteration in consumer demand can result in tremendous fluctuation in production and inventory planning. Bullwhip Effect emphasizes the need to minimize demand volatility by using improved forecasting and techniques such as just-in-time (JIT) inventory management. Implementing these techniques by food businesses such as Steers can preclude fluctuations, avoid wastage, and optimize business performance.

2.1.2 Theory of Constraints (TOC)

Goldratt and Cox (1984) have developed the Theory of Constraints (TOC) that deals with how to identify and deal with production bottlenecks that result from variability in demand. Goldratt and Cox (1984) state that companies are generally constrained by production bottlenecks and inventory management, and they decide aggregate productivity. The TOC is applied in helping identify and remove such bottlenecks for the purposes of improving business effectiveness. In food manufacturing, food businesses like Steers can utilize TOC in determining production bottlenecks caused by the variable demand across the seasons and implement strategies to mitigate the impacts, thereby increasing overall production efficiency.

2.1.3 Demand Forecasting

Chase (2013) further highlights the significance of demand forecasting in tackling seasonal and yearly variation in demand. Effective demand forecasting is critical in supply chain management in organizations with highly variable demand. Techniques such as regression and time series enable business firms to forecast changes in demand and prepare more effectively for variations. Chase (2013) is of the opinion that demand forecasting can be used by food producers to forecast changes in demand, hold less inventory, minimize wastage, and gain improved operating efficiency.

2.1.4 Inventory Management

Wisner et al. (2015) explain the imperative of efficient inventory management in addressing seasonal and yearly variations in demand. In their opinion, efficient inventory management will reduce waste, increase efficiency, and increase profitability. Inventory methods such as JIT inventory management and vendor-managed inventory (VMI) reduce inventories and increase efficiency. For food manufacturing companies such as Steers, these initiatives can reduce holding cost on inventories while avoiding demand adjustments to increase profitability.

2.1.5 Supply Chain Management

Chopra and Meindl (2016) describe the significance of supply chain management in coping with the volatility of demand within the year. Supply chain management, if executed well, eliminates waste, streamlines operations, and makes operations more profitable. Tools such as Collaborative Planning, Forecasting, and Replenishment (CPFR) can remove demand variability and streamline overall supply chain performance. Chopra and Meindl (2016) believe that these practices can be utilized by food companies to reduce variability of demand, optimize operating performance, and optimize profitability, thereby becoming more competitive in the marketplace.

2.1.6 Time Series Analysis

Time series analysis was first devised by Box and Jenkins (1970), a time-based data analysis and forecasting technique. Time series analysis helps in the recognition of trends and patterns in historical demand data so that business can anticipate future demand and be prepared for variations. As suggested by Box and Jenkins (1970), food manufacturers can utilize the application of time series analysis in the prediction of demand fluctuations, inventory levels can be optimized, and wastage minimized with a view to achieving supply chain efficiency.

2.1.7 Exponential Smoothing

Brown (1959) developed the exponential smoothing technique, which smoothed the time series data's irregularities, and demand forecasting became straightforward. Exponential smoothing allows us to visualize easily underlying trends contained in demand data, and this makes it straightforward for companies to make future demand projections. Brown (1959) believes that food producers can use exponential smoothing to predict demand fluctuations, reduce inventories, reduce wastage, and become more efficient in general.

2.1.8 Autoregressive Integrated Moving Average (ARIMA) Models

Box and Jenkins (1970) also developed ARIMA models, ~~a~~ statistical time series data analysis and future demand forecasting technique. ARIMA models facilitate the determination of trends and patterns in historical data, hence improving demand forecasting. Box and Jenkins (1970) suggest that food companies adopt ARIMA models so that they can easily predict demand variability, stock optimization, and reduce wastage, and hence achieve improved overall performance.

2.1.9 System Dynamics

System dynamics was originally developed by Forrester (1961) as a way of analyzing complex systems and identifying where optimization might occur. By simulating complex systems, system dynamics can be used to identify patterns and trends in demand data so that companies can see where inefficiency is occurring in the supply chain. Through system dynamics, Forrester (1961) argues, food manufacturers are able to model demand patterns and optimize production and inventories to improve supply chain performance.

2.1.10 Time Series Analysis (Repeat)

Time series analysis is a statistical method applied in analyzing and forecasting data points that have been gathered over a span of time. Time series analysis is most useful in identifying trends and patterns in data, including seasonal and yearly variations, holds Box and Jenkins (1970).

2.1.11 Seasonal Variation

Seasonal variation is the systematic and regular fluctuations which occur in demand at a specific time of the year, e.g., holidays or weather. Hyndman and Athanasopoulos (2018) state that these fluctuations are usually brought about by external factors, e.g., weather, holidays, and behavior of customers, and have a high impact on the demand for specific products.

2.1.12 Annual Variation

Yearly variation is variation spanning a 12-month period. Gujarati and Porter (2018) argue that such variation is typically the result of macro factors such as economic performance, market forces, and consumption patterns, all of which cause demand change on an annual basis.

2.1.13 SARIMA and XGBoost Hybrid Forecasting Models:

In the food sector, anticipating turnover changes is crucial for production planning and inventory control. Time series models like SARIMA and advanced machine learning models **like XGBoost have been employed to address such forecast problems.**

2.1.13.1 SARIMA Model

Seasonal Autoregressive Integrated Moving Average (SARIMA) is an extension of ARIMA that models seasonality in time series explicitly. SARIMA can be represented as ARIMA(p,d,q)(P,D,Q)[s], with the additional seasonal terms (P, D, Q, s) representing seasonal autoregressive, differencing, and moving average terms, respectively. SARIMA models are particularly good at modeling linear seasonal trends and have been widely used in various industries, such as food manufacturing, for forecasting demand and level of turnover (Hyndman and Athanasopoulos, 2018).

2.1.13.2 XGBoost Model

Extreme Gradient Boosting (XGBoost) is an optimized distributed gradient boosting machine learning algorithm that builds an ensemble of decision trees in a greedy manner, attempting to maximize both predictive performance and speed. XGBoost, in contrast to SARIMA, which assumes linearity, can model complex nonlinear relationships in data and is therefore appropriate for capturing peculiar consumer behaviour and exogenous variables in turnover trends (Chen and Guestrin, 2016).

2.2. Empirical literature review

Kamble et al. (2018) – Seasonal and Annual Demand Fluctuations in India

Kamble et al. (2018) had analyzed time series data for historical sales from 2010 to 2015 for exploring seasonal and annual demand fluctuations of food products in India. It was observed that there were significant fluctuations in demand with peak season befalling during holidays and festivals and also seasonal fluctuations were based on annual fluctuations because consumer preferences and income stream varied every year. The authors recommended that food product manufacturers alter production and inventory volumes to match seasonal demand and invest in demand forecasting techniques to prepare for annual fluctuations. They are limited by the utilization of data that is not recent, which can fail to be analyzed with the most current market trends, and also for failing to incorporate the external variables relevant in weather and economics into the forecasting model.

Zhang et al. (2020) – Chinese Food Sales Changes

Zhang et al. (2020) utilized regression analysis of 2015-2019 historical sales data to analyze season and annual differences in food sales in China. Season and annual differences were significantly affected by consumer demand, income, and demographics. Zhang et al. suggested that food businesses employ just-in-time inventory management and strive to manufacture products compatible with changing consumer tastes. Drawbacks of the study are the application of data that does not necessarily reflect current market trends and simplification of complex relationships among variables due to the use of regression analysis as a tool.

Kim et al. (2019) – Meat Product Demand in the United States.

Kim et al. (2019) examined seasonal and yearly demand fluctuations for American meat products through time series analysis of historical sales data between the years 2010 and 2018. The research found significant seasonal demand fluctuations with peak demand in winter and holiday seasons, and yearly fluctuations based on consumer trends and income. The authors suggested that meat producers adjust production and stock levels to reflect season levels of demand and invest in demand forecasting practices. The limitations of the study are that the data used may not represent the current trends in the market and the weakness of the time series analysis in detecting underlying trends in demand.

Panda and Mohanty (2023) – XGBoost and SARIMA for Food Demand Forecasting

Panda and Mohanty (2023) presented a comparison of the appropriateness of SARIMA models with XGBoost in predicting food demand during the supply chain. The authors determined that XGBoost outperformed SARIMA because the latter had lower values of RMSE and MAE. The authors concluded the advantages of the utilization of XGBoost when the demand data is intricated and nonlinear in nature. Nevertheless, the limitations of the study are that historical data cannot represent the market position at present and that the XGBoost model can over fit as it is sophisticated.

Comparative Studies on Forecasting Models

Some study has compared the performance of XGBoost and the SARIMA model in forecasting sales. For instance, Alvi (2024) demonstrated that XGBoost outperformed SARIMA for sales data forecasting particularly in capturing abnormal fluctuation. Patel (2024) found that in a similar study, SARIMAX, the version of SARIMA, had a good performance as it could predict sales of a product with few errors in measures. These comparative studies demonstrate how to use the right models based on the kind of data used.

2.3. Research gap

Although there are several studies that have investigated seasonal and annual differences in food product demand (Kamble et al., 2018; Zhang et al., 2020; Kim et al., 2019), how much these variations in demand affect turnovers in the food production sector is still very much a research gap to be filled. Previous studies have concentrated more on encouraging the development of demand forecasting methods (Chase, 2013), stock policies (Wisner et al., 2015), and supply chain optimization (Chopra and Meindl, 2016) as a means of adjusting to such fluctuations. Less of an issue has been determining how seasonality and yearly fluctuations are being carried over into financial performance, namely metrics like turnover, of food manufacturing companies.

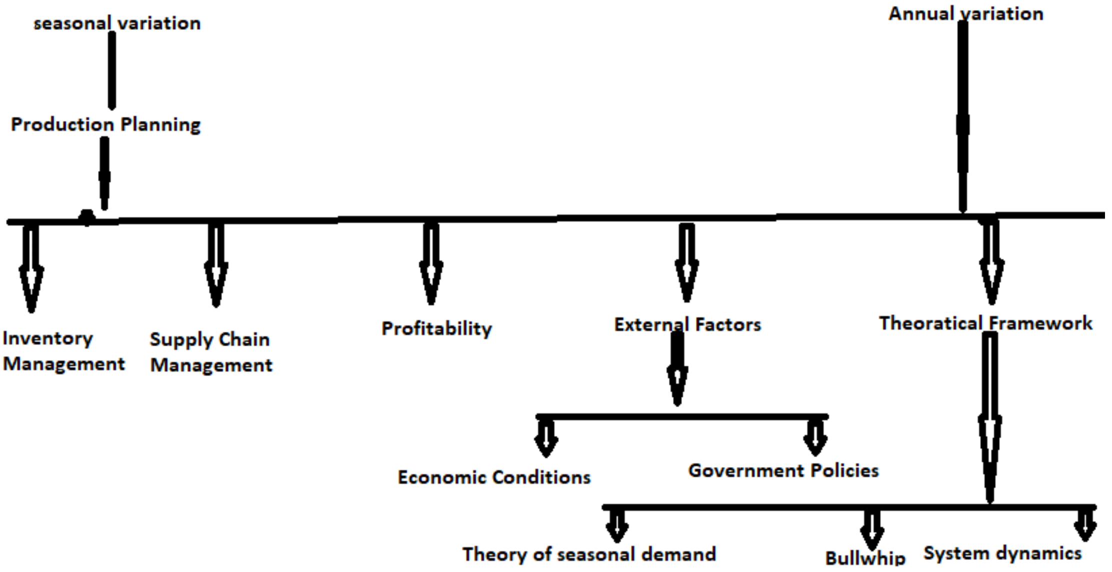
Besides this, while time series model SARIMA and other machine learning algorithms XGBoost have been applied in other domains, their combined application to predict trends in food production turnover has remained of less prominence. Past research utilized to examine either models of demand or impacts at the supply chain level rather than bridging gaps between those cycles or analyzing comparative predictive effectiveness of models.

This study seeks to fill this gap by embarking on a systematic exploration of season and annual turnaround volatility in the food production sector, with emphasis on Simbisa Brands (Steers). The study also seeks to compare SARIMA and XGBoost models' performance for forecasting in this context, thereby adding to academic scholarship as well as business decision-making for food business operations.

This diagram is utilized to illustrate interrelations between variables:

Seasonal variations affect production planning, inventory management, supply chain management, and profitability. Annual variations affect inventory management, supply chain management, and profitability. External drivers produce seasonal and annual variations. And theoretical frameworks Theory of seasonal demand (TSD), Bullwhip Effect (BE), and System Dynamics (SD) provide the conceptual basis in understanding the interrelations.

2.4. Proposed conceptual framework



2.5. Chapter Conclusion

This literature review critically examined seasonally and annually changing turnover within the food production industry. The rationale was underpinned by a range of theoretical frameworks, such as the Theory of Seasonal Demand (Barbour, 1973), the Bullwhip Effect (Lee et al., 1997), the Theory of Constraints (Goldratt and Cox, 1984), and System Dynamics (Forrester, 1961), that form a basis upon which to investigate changes in demand and their impact on operations. Empirical studies (Kamble et al., 2018; Zhang et al., 2020; Kim et al., 2019) highlighted demand forecasting, inventory management, and product diversification as strategic responses to demand uncertainty.

There existed a research gap with regards to how seasonal and yearly variations directly affect turnover in food processing firms. Though there has been some study of demand-side factors in the past, there is little proof of research that evaluates the direct effect on firm-level turnover performance. This study fills this gap by bringing together theoretical and empirical expertise to model turnover patterns, detect trends and outliers, and evaluate the performance of the prediction models.

The article enumerated the relevance of modern forecasting techniques, namely the SARIMA and XGBoost models. The literature reveals that while SARIMA can capture seasonality, XGBoost has a greater potential to yield more accurate forecasting since it can detect complex non-linear relationships. The implications of this are far-reaching such as accurate forecasting increases business productivity, allows for just-in-time (JIT) inventory systems, and allows decision-making for planning production and resource allocation. In addition, hybrid forecasting techniques can further improve predictability, with a competitive advantage by being responsive and cheaper.

Finally, this chapter calls for embracing data-oriented approaches to forecasting to steer business performance in food manufacturing businesses. The research design, such as study design, sources of data, and analysis methods employed in testing the influence of seasonal and yearly fluctuations on turnover, are described in the next chapter.

CHAPTER 3: METHODOLOGY

3.0 Introduction

The study employs systematic and robust methodology in analyzing sales turnover patterns using series analysis. Of major interest are to identify turnover patterns in the data, build appropriate forecasting models, estimate their parameters, check the performance of models, and get 12-month forecasts. Data sources, target population, data gathering procedures, definitions of variables, and analytical methods are presented in the chapter. It also describes the model estimation, model specification, validation, and selection processes of SARIMA and XGBoost models in order to produce a robust and accurate prediction.

3.1 Research paradigm

One of the prevalent methods of research in time series analysis was positivism, with emphasis on objectivity, measurability, and the application of statistical methods in data analysis (Babbie, 2023). Positivist researchers aimed to establish patterns in data to make inferences about what would occur in the future. Predicting turnovers through the positivist method is done by collecting past data, identifying trends, and building models to predict future turnovers.

3.2 Research design

A quantitative research design would be employed in this study, with its framework mainly established on the numerical data analysis necessary to extract conclusions into the turnover dynamics at Steers. Sales turnover would be construed, as well as the derived changes through employment of the SARIMA and XGBoost models.

3.2.1 Data collection and sources

The research relies on secondary data for Steers' sales turnover, which has been obtained from Simbisa Brands Ltd.'s database. The data ranges from January 2019 to December 2023 and provides a proper base for analysis. Historical data makes the research more authentic and allows one to have a detailed look at the pattern of sales, which helps to make accurate predictions in terms of future sales trends.

The figures are provided in two currencies: the United States dollar (USD) and the local Zimbabwean dollar. Since the turnover is provided in USD, local currency is converted into USD using the interbank rate on any given date. For the purpose of this research, the total sales turnover for the entire Steers brand was utilized, not per store. Below is a sample of data provided by Steers, a Simbisa Brands subsidiary.

3.2.2. Snippet of data from Simbisa Brands

Innscor Group - Turnover Report

Steers Brand Dates: March

Branch	Arcade			Clonsilla			Cork Rd		
Dates	Bond Only	US$ Only	Customers	Bond Only	US$ Only	Customers	Bond Only	US$ Only	Customers
2018-01-01	20,689.25	4319.75	518	19,189.55	2694.25	269	15,295.85	2401.00	262
2018-01-02	12,170.20	2219.50	293	21,164.65	2415.25	231	11,224.40	1719.00	193
2018-01-03	7,791.70	2174.00	299	8,793.00	1223.00	127	8,474.45	1298.50	155
2018-01-04	10,558.05	1935.00	280	8,671.30	1348.50	145	7,980.55	1486.75	182
2018-01-05	13,507.95	2751.00	422	8,964.05	1360.50	168	11,735.04	2174.25	275
2018-01-06	10,058.30	2410.50	313	14,342.85	1703.00	183	6,093.55	1451.50	178

3.3 Target population

The population of interest for the research is operation management and the finance department in fast-food businesses. Operation management handles day-to-day activities, overseeing smooth service and resource allocation that are key to building a picture of turnover patterns and customer demand. The finance department keeps track of revenue, expenditure, and profitability, providing a picture of financial performance and economic factors impacting turnover. By focusing on these two groups, the study aimed to gain information on both financial and operational determinants of sales turnover to develop accurate forecasting models and strategic recommendations

3.4 Description of variables

The dependent variable is monthly turnover levels for steers Zimbabwe, which is quantified in United States Dollars and constitutes a main variable in the study. The variable is defined at length below together with its justification.

Variables	Symbol	Indicator	Source
Turnover	Trn	Recorded sales	Simbisa Database

Justification

Turnover is the core variable to be investigated in this study since it serves a fundamental role as a measure of performance, particularly in the fast-food sector. Turnover provides essential details regarding the general performance of the firm, and hence it is a key variable to track and analyse for purposes of creating a comprehensive view of business success.

3.4.1 Data Processing Procedure

The data processing involved several key steps to ensure correct analysis and addressing the research objectives:

1. Data Cleaning the data were screened for any invalid or missing values, ensuring reliability and accuracy. Data cleaning facilitated easy identification of turnover patterns, which was essential in understanding seasonal cycles and annual trends.

2. Data Transformation the turnover amounts in USD and ZWD were converted to a common currency (USD) using exchange rates. The transformation facilitated dataset homogeneity as required by proper seasonal analysis and forecasting.

3. Differencing since the data was initially non-stationary, differencing was applied to render it stationary. This was essential in maintaining effective time series analysis and satisfying the SARIMA forecasting model assumptions to meet the demand for effective future turnover forecasts.

4. Aggregation turnover data for Steers were recorded monthly rather than store level individual. Aggregation was essential to analyse overall trends and seasonality, as a clearer view of the performance of the brand to enable accurate forecasting.

5. Data Normalization outliers or abnormal instances in the data were explicitly corrected, which made analysis biased towards normal tendencies. This helped prevent misleading results and ensured that anomalies or extraneous conditions, like economic fluctuations, did not skew the trends.

3.5 Data analysis processes

Diagnostic tests

3.5.1. Stationarity test

It is necessary in time series analysis to test for stationarity prior to any analysis being conducted. A data set is termed as stationary if its variance and mean are constant over a period of time. If the data is found to be non-stationary, it can normally be transformed to stationarity using techniques like differencing, where the difference between two successive data is calculated. This may make the variance and mean of the data more stable. Stationarity tests are required as they give confidence that assumptions of forecasting models are met, thus generating more valid and reliable forecasts. The failure to convert non-stationary data prior to subjecting it to analysis may lead to deceptive results. Thus, stationarity tests help determine if data are suitable for sufficient forecasting and analysis. For the current study, the Augmented Dickey-Fuller (ADF) test was used to establish stationarity.

3.5.2. Normality tests

It is vital to check that the data is normally distributed, as a failure in this may lead to invalid parameter estimates or meaningless predictions. To ensure validity of the normality assumption, some graphical aids were employed and they are the histogram and the quantile-quantile (Q-Q) plots, which provide a graphical assessment of the way in which the data fits a normal distribution (Das & Imon, 2017). These graphical methods provide initial assessments but they are not enough alone. Statistical tests like Shapiro-wilk test need to be made to test data for normality or a Kolmogorov-Smirnov test. These graphical and statistical plots enable us to detect deviations from normality so that the statistical models used are reliable and results are stable.

3.5.3. Independence test

The residuals must satisfy the condition of independence and no autocorrelation. Apart from tests like the Durbin-Watson test, this study also utilizes the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots for testing the residuals. The plots were examined in order to confirm there were no lingering terms of autocorrelation, a breach of the assumption of independence.

3.5.4. Homoscedasticity test

Homoscedasticity tests are conducted to make sure the variance of the residuals is constant. This assumption is normally tested with the aid of statistical tests such as the White Test and Breusch-Pagan test and visual inspections on the residuals. For this study, the White Test is conducted to test for homoscedasticity in order to make sure the variability of the model is constant and the results are stable.

3.5.6. Seasonality test

Testing for seasonality was carried out to establish seasonal trends. The time series plot was visually inspected for repetitive trends, and Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) analysis were used, where significant spikes at seasonal lags indicated seasonality. Stationarity was tested using the Augmented Dickey-Fuller (ADF) test for the presence of a unit root. If the data was seasonal, seasonal differencing was performed and a SARIMA model was employed to model the seasonal autoregressive and moving average components. XGBoost model was also employed as a complex machine learning model, which also models the seasonality and trends. 3.6 Model Selection and Specification

3.6 Model Selection and Specification

SARIMA and XGBoost models are both employed in this research to model different components of the turnover data.

3.6.1. SARIMA (Box-Jenkins Methodology)

The approach utilizes ARIMA or SARIMA models to identify the optimal model for fitting historical time series data. The methodology is divided into three stages: model identification, model selection, and model estimation. The general ARIMA equation is expressed as:

$$\Phi_p(B)(1-B)^d Y_t = \theta_q(B)\alpha_t \sim ARIMA(p,d,q)$$

Y_t – The observed time series at time t

B – The Backshift operator, where BY_t = Y_{t-1}. It shifts the time series back by one period.

$(1-B)^d$ – The differencing operator, which ensures stationarity by removing trends

$\Phi_p(B)$ – The autoregressive (AR) polynomial of order p, defined as

$$\Phi_p(B) = 1 - \Phi_1 B - \Phi_2 B^2 - \ldots - \Phi_P B^P$$

Here $\Phi_1, \Phi_2, \ldots . \Phi_p$are the autoregressive parameters, which represent the relationship between Y_t and its past values.

$\theta_q(B)$ – The moving average (MA) polynomial of order q, as defined as:

$$\theta_q(B) = 1 + \theta_1 B + \theta_2 B^2 - \ldots + \theta_q B^q$$

Here $\theta_1, \theta_2, \ldots . \theta_q$ are the autoregressive parameters, which represent the relationship between Y_t and its past values.

α_t – A white noise error term, assumed to be normally distributed with mean zero and constant variance ($\alpha_t \sim N(0, \sigma^2)$). It represents random shocks in the time series.

3.6.2. Model Identification:

In the first stage, research focused on determining the stationarity of the time series and identifying seasonality of the sales turnover data. Stationarity was identified by run sequence plots, in which mean and variance structures were examined over time. The Augmented Dickey-Fuller (ADF) test was then used to statistically identify whether data exhibited unit roots, meaning non-stationarity.

For determining seasonality, partial autocorrelation (PACF) and autocorrelation (ACF) plots were inspected, where spikes at season lags were significant, signifying patterns that repeat over time. As there was seasonal variation in the data, a SARIMA model would be more appropriate compared to an ARIMA model. Seasonal differencing was used to remove seasonal trends so that stationarity could be ensured.

After establishing stationarity, p and q values of AR and MA components, respectively, were decided upon using PACF and ACF plots. The model with the best fit was then chosen by considering the values of AIC and BIC, such that there was an optimal balance between model complexity and precision.

Through this structured methodology, the research determined an ideal SARIMA model to efficiently preserve both seasonality and non-seasonality in sales turnover trends within the fast-food industry in Zimbabwe for supporting accurate forecasting and business decision-making.

3.6.3. Model Selection:

The Box-Jenkins approach was employed within this research to develop a model for forecasting fast-food industry sale/turnover trends in Zimbabwe. Having established stationarity and determined seasonality, the final step was to choose and estimate the parameters of the model.

Since seasonal patterns existed, a SARIMA model replaced an ordinary ARIMA model. Various model specifications were tested with Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) to determine the most appropriate structure. The parameters like autoregressive (AR), moving average (MA), and seasonality were estimated using Maximum Likelihood Estimation (MLE) in computer software.

To confirm the adequacy of the model, residual diagnostics were performed to verify that errors followed a white noise process with no significant autocorrelation. Forecasts were then generated out-of-sample to validate the model's ability to predict by comparing predicted sales turnover with actual observed values.

By an intensive application of the Box-Jenkins technique, this study was able to successfully identify a SARIMA model that fit the seasonal and non-seasonal turnover trends and provided a valid model for forecasting and strategic planning.

3.6.4. Model Estimation:

After fitting the selected SARIMA model, diagnostic checks were conducted to assess its validity and ensure a good fit for sales turnover trends in Zimbabwe's fast-food sector. The residuals—representing the differences between observed and predicted turnover values—were analysed to verify that they met key statistical conditions.

In order to verify that the residuals were stationary and had homoscedasticity (constant variance), visual residual analysis was conducted with the help of residual plots. Furthermore, the Box-Ljung test was used to test for autocorrelation in the residuals. The findings revealed that the residuals were very much like white noise, which implied that the model had succeeded in capturing the underlying structure of the data.

If considerable autocorrelation had been detected, further refinements—i.e., changing the SARIMA orders (p, d, q, P, D, Q) or recalculating differencing—would have been indicated. As it was, however, the ultimate SARIMA model provided a statistically satisfactory fit, from which dependable forecasting of sales turnover trends for steers could be achieved.

3.7 XGBoost Model

The development of the XGBoost model includes three major phases: feature engineering, model specification, and training/validation.

3.7.1. Feature Engineering

The construction of the XGBoost model begins with the important feature engineering phase, which is focused on extracting useful insights from the turnover data to enhance the model's predictive capability. This involves the generation of some key features: the lagged turnover values (Turnover_lag1, Turnover_lag2, etc.) are computed in order to capture the intrinsic autocorrelation in the time series and the influence of past turnover on the current performance. Rolling statistics like the rolling mean and standard deviation are then calculated with various time windows to pick up underlying trends and volatility in the turnover series. Finally, seasonal identifiers, i.e., month and quarter of the year, are also added as categorical variables in order to model the seasonality that is common with the fast-food business.

3.7.2. Model Specification

Following feature engineering, the XGBoost model is defined precisely with consideration of hyper parameter optimization for different significant hyper parameters. The model complexity is controlled through the number of boosting rounds (estimators), and the step size at boosting is controlled by the learning rate (eta) to prevent over fitting. The depth of the trees (max depth) limits the tree complexity within an individual tree in the ensemble. To make the chosen hyper parameters generalize to new and unseen data, cross-validation methods are used to reduce prediction error and avoid over fitting.

3.7.3. Model Training and Validation

Finally, the XGBoost model is subjected to a rigorous training and validation process. The model is trained on 80% of the data to enable it to learn the underlying patterns and relationships in the turnover data. The remaining 20% of the data is then used for validation purposes, providing an independent estimate of the model's predictive capability. The model performance is estimated in terms of a solid set of performance measures, including Root Mean Squared Error (RMSE), Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), and R-squared, providing a general view of the ability of the model to accurately forecast turnover. 3.7.4. Forecasting and Model Evaluation

3.7.4. Forecasting and Model Evaluation

After identifying the best models, forecasts are generated for the next 12 months. The models are tested using parameters like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and R-squared to get accurate forecasts.

3.8. Chapter Summary

This chapter introduces the methodology of the framework adopted for fast-food industry sales turnover analysis using SARIMA and XGBoost models. This study is informed by the positivist research paradigm, where statistical and objective analysis are prioritized. Secondary data from the Simbisa Brands database is used, whereby turn-over is the central variable.

A series of diagnostic tests verifies data reliability, and SARIMA and XGBoost are utilized to generate seasonal and nonlinear patterns, respectively. SARIMA employs the Box-Jenkins methodology, while XGBoost employs feature engineering and optimization of hyperparameters. The two models are evaluated using a comprehensive set of metrics to attain accuracy, and a 12-month-ahead prediction is generated. This two-model approach is a good foundation for modeling and forecasting turnover trends in Zimbabwe's fast-food sector

CHAPTER 4: DATA PREPATION, ANALYSISI, AND DISCUSSION

4.0 Introduction

This chapter provides a detailed analysis of turnover trends in the food producing industry for Simbisa Brands (Steers). In order to provide a detailed explanation of the data and also ascertain its suitability to model and measure predictive model performance. The study begins with descriptive statistics to define the most important features of the data on turnover, as well as diagnostic testing for ascertaining if data fulfills the assumptions necessary for advanced modeling (Hyndman and Athanasopoulos, 2018). The chapter then introduces and employs analytical models, including SARIMA and XGBoost, conducting model fitness and validation tests to determine their accuracy and reliability. Lastly, it concludes with the key findings of the analysis and the concluding overview of findings.

4.1 Descriptive Statistics

Metric	Statistic
Mean	296772.5
Median	288047.4
SD	107388
Kurtosis	0.06201472
Skewness	0.2919566
Range	542891.8
Min	59095.35
Max	601987.2
Sum	17806349
Count	60

Table 4.1.1. Descriptive Statistics

Between January 2020 and December 2024, company turnover data records immense variability, with a mean of $296,772.5 and standard deviation of $107,388, capturing intense fluctuation over the observed 60 months. The top 5 highest turnover months—December 2021 ($601,987.18), December 2024 ($541,557.72), December 2023 ($472,787.80), November 2024 ($433,772.52), and November 2021 ($432,639.14)—occur primarily during the festive year-end season, indicating intense seasonality driven by active consumerism. The lowest turnover rates, on the other hand, happen in April 2023 ($59,095.35), May 2023 ($98,418.83), June 2023 ($106,110.28), July 2022 ($177,693.20), and January 2022 ($184,167.70) and have a tendency to happen during the early part of the year, perhaps reflecting post-holiday lulls or overall economic headwinds. This spread, with a moderate right skewness (skewness = 0.29) and low kurtosis (0.06), suggests that while the extreme outliers are negligible, there is a consistent seasonal cycle with highs later in the year and lows earlier in the year, which is useful to be aware of when it comes to strategic planning and resource allocation.

Figure 4.1.2. Time series and Trend Analysis

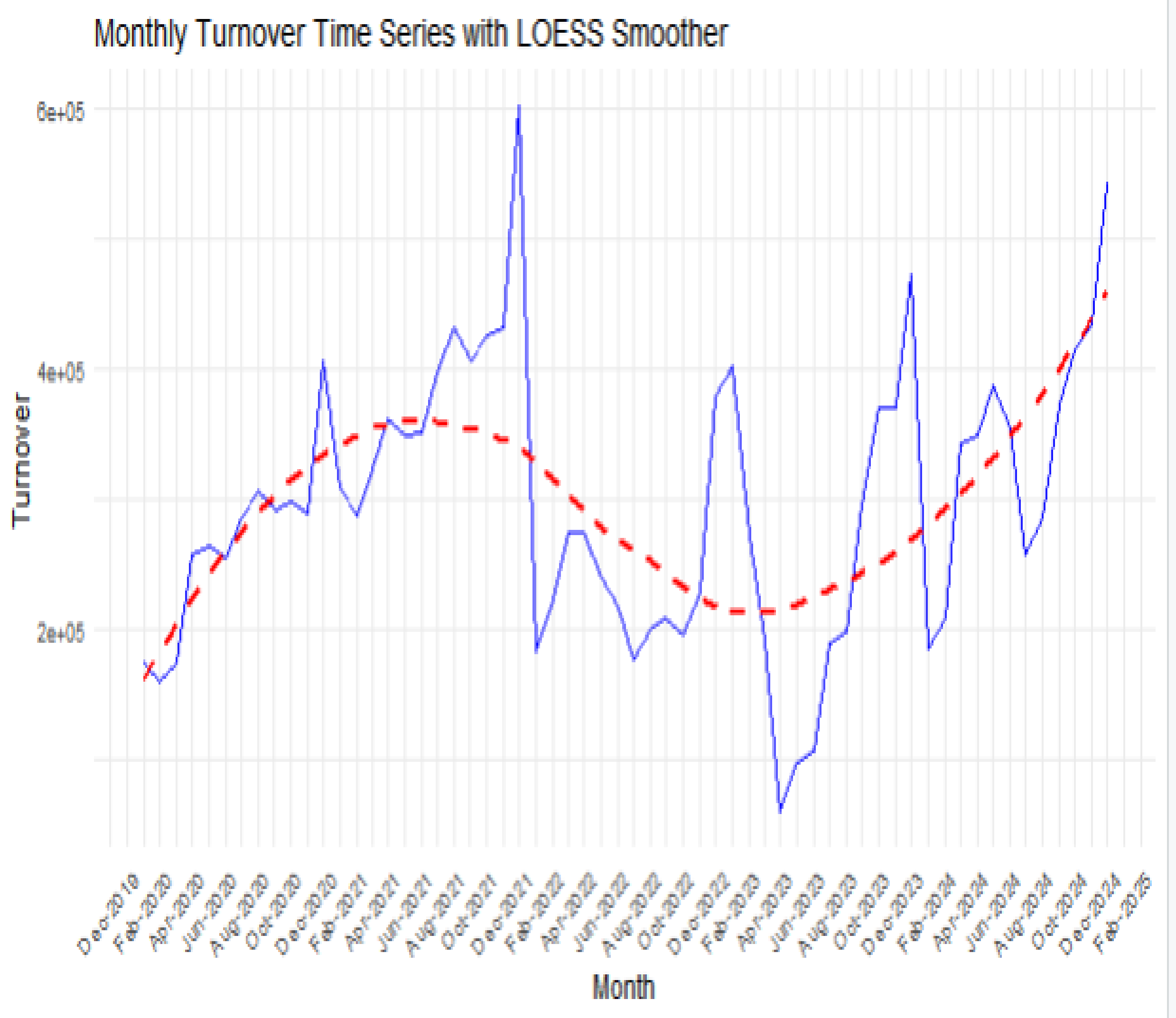


Figure 4.1.3: **Time series plot of Steers turnover (2020-2024)**

Figure 4.1.3 is a time series plot of the turnover series from 2020 to 2024 with a LOESS smoother (with its default span parameter, typically on the order of 0.75 in ggplot2) represented by the red dashed line. The LOESS smoother works well at highlighting the underlying trend whereas common and abrupt fluctuations in raw turnover (blue line) exist. The trend has a steady upward direction in turnover between mid-2020 and mid-2022, as indicated by the rising portion of the LOESS curve. Growth is followed by a sudden spiking and decline around 2022, and then stabilization of the trend and return to growth as it approaches 2025, indicating sustained recovery and growth. The application of the LOESS smoother makes it easier to smooth out identification of these trend periods and structural breaks and yields valuable insights into seasonality trends, anomalies, and long-term growth—essential for informing future predictive models such as SARIMA and XGBoost.

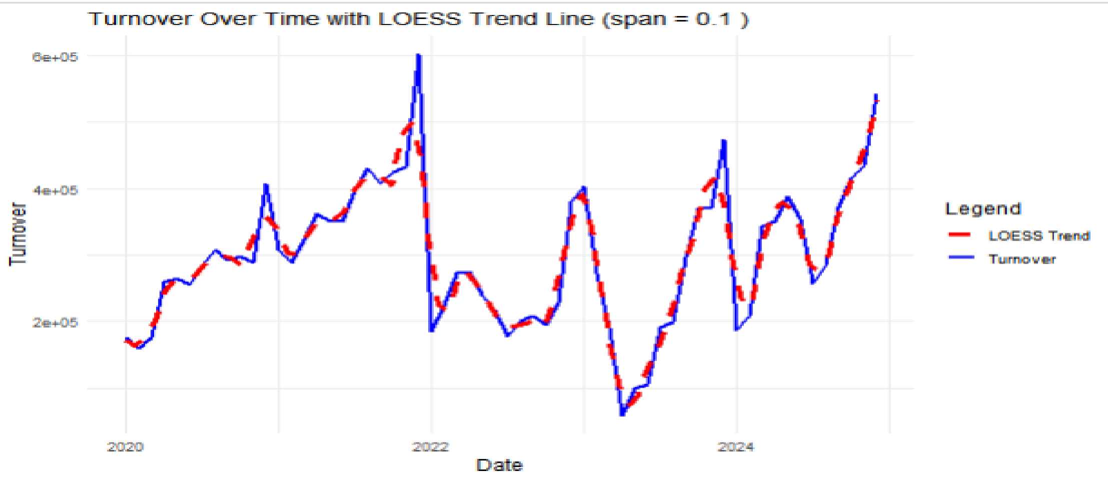


Figure 4.1.4: Turnover Over time with Loess Trend Line (span = 0.1)

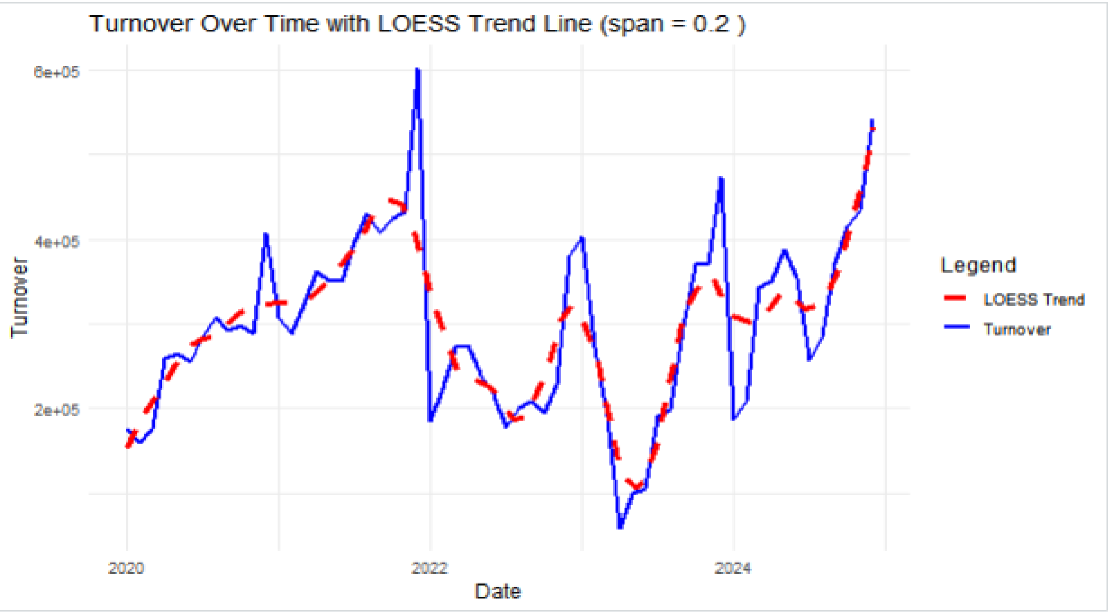


Figure 4.1.5: Turnover Over time with Loess Trend Line (span = 0.2)

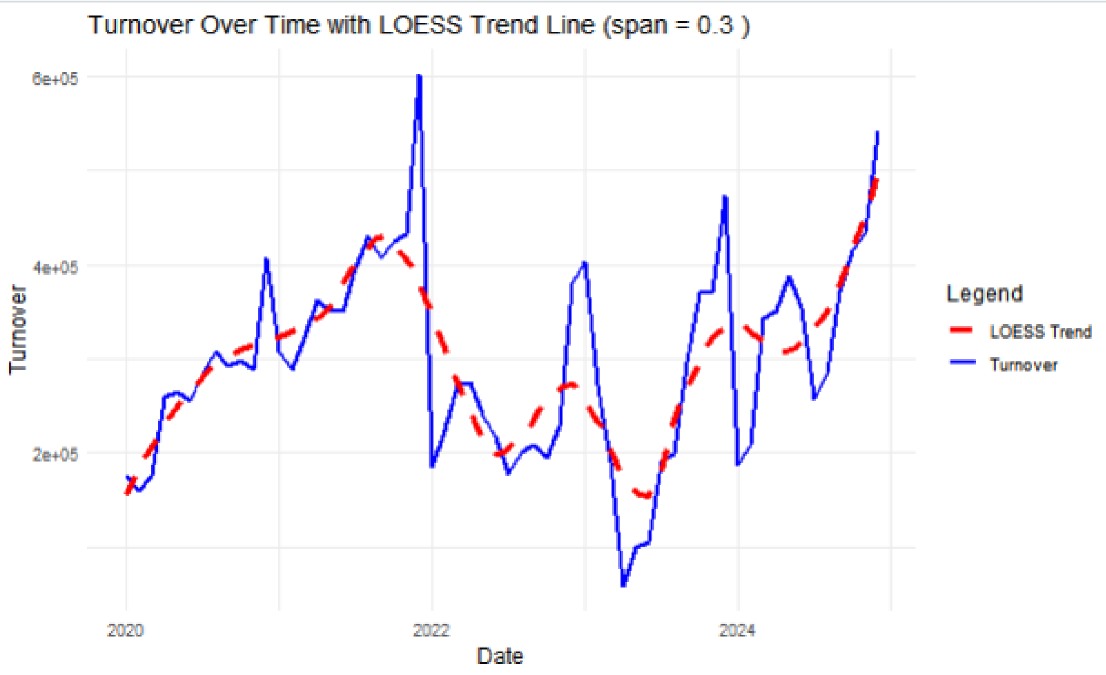


Figure 4.1.6: Turnover Over time with LOESS Trend Line (span = 0.3)

The three turnover time series plots presented in Figure 4.2, 4.3, and 4.4 use the LOESS smoother to reveal underlying trends by reducing the effects of short-run volatilities, with each of the plots varying in the span parameter (0.1, 0.2, and 0.3). Using a short span (0.1) puts the trend line near rapid local variation and noise, but using larger spans (0.2 and 0.3) successively smooth the trend line, revealing larger, long-term patterns and filtering out more short-term fluctuation. Plots of these indicate how the varying of the LOESS parameter affects the visibility of underlying trend and short-term oscillation and highlights that a choice of appropriate span is required based on the level of detail and trend information you want to observe in your data.

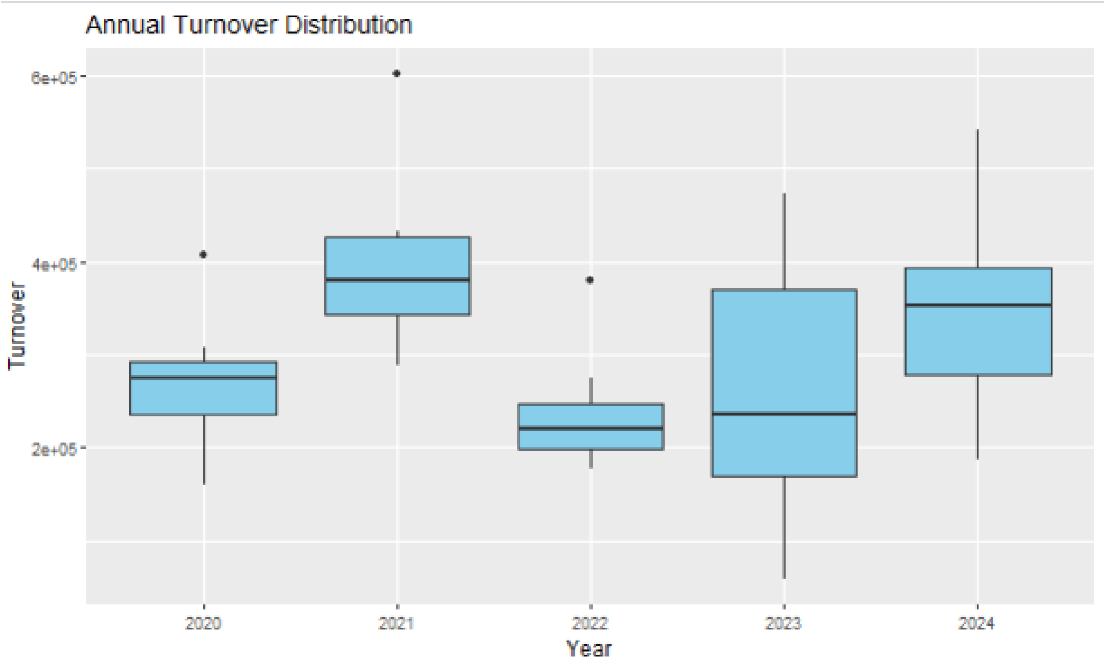
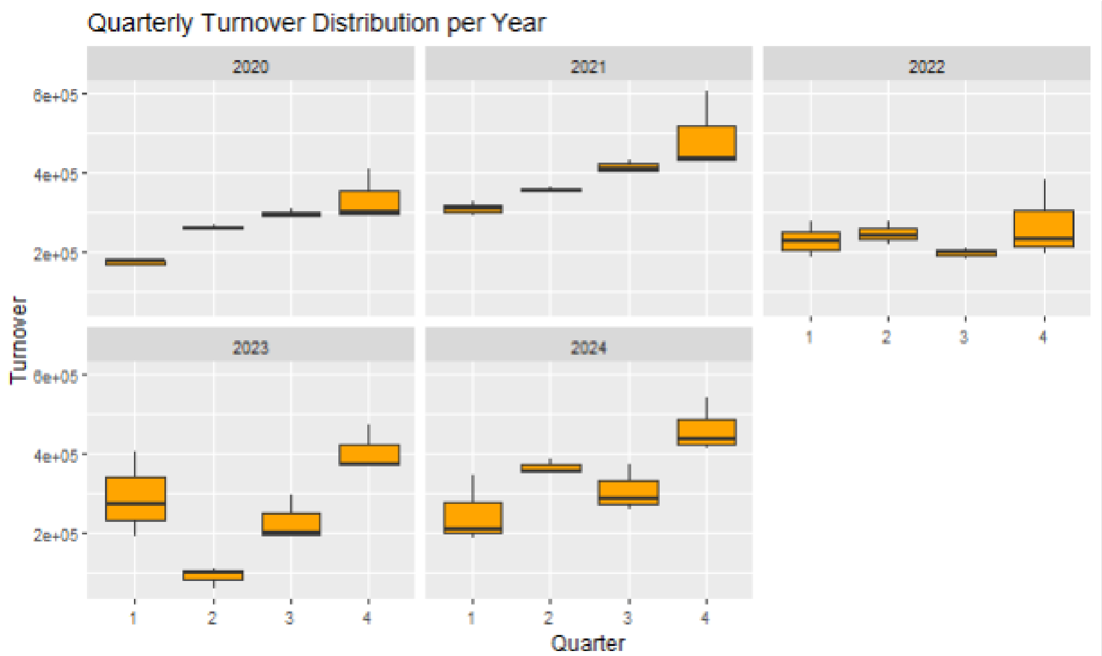


Figure 4.1.5 Annual Trend Analysis (Boxplots)

The above Figure 4.1.5 indicates a summary of the year-wise box plots of the distribution of turnover from 2020 to 2024, reflecting a clear increasing trend in turnover. The median turnover in 2020 was about $ 200,000 and gradually increased to about $ 40,000 in 2024. Every box depicts the interquartile range (IQR), i.e., turnover variability has generally increased over the years with a steep increase in 2022. Outliers exist in some years, indicating sporadic extreme turnover values.

Figure 4.1.6. Quarterly Turnover (by Year)



4.2. Quarterly Turnover Boxplots by year

Quarterly turnover box plot reflects evident seasonality over the years, where certain quarters, like Q4 in 2021 and 2024, consistently register higher turnover, possibly because of greater demand or specific business cycles. Quarters like Q2 in 2023 register lower turnover, indicative of off-season. This division is meant to provide insight into when turnover most typically rises or falls, assisting with strategic and operational planning. The box plot also illustrates how turnover distributions shift from year to year, highlighting the outliers like the very high turnover in Q4 2021 and the variability in 2023. These shifts are typical of outside forces like economic turmoil, changes in consumer behavior, or special promotions. By linking these annual patterns and deviations to thoroughly document outside forces, the study hopes to further explain their causes, thereby establishing a sound basis for future analysis.

4.2.1 Monthly trend analysis

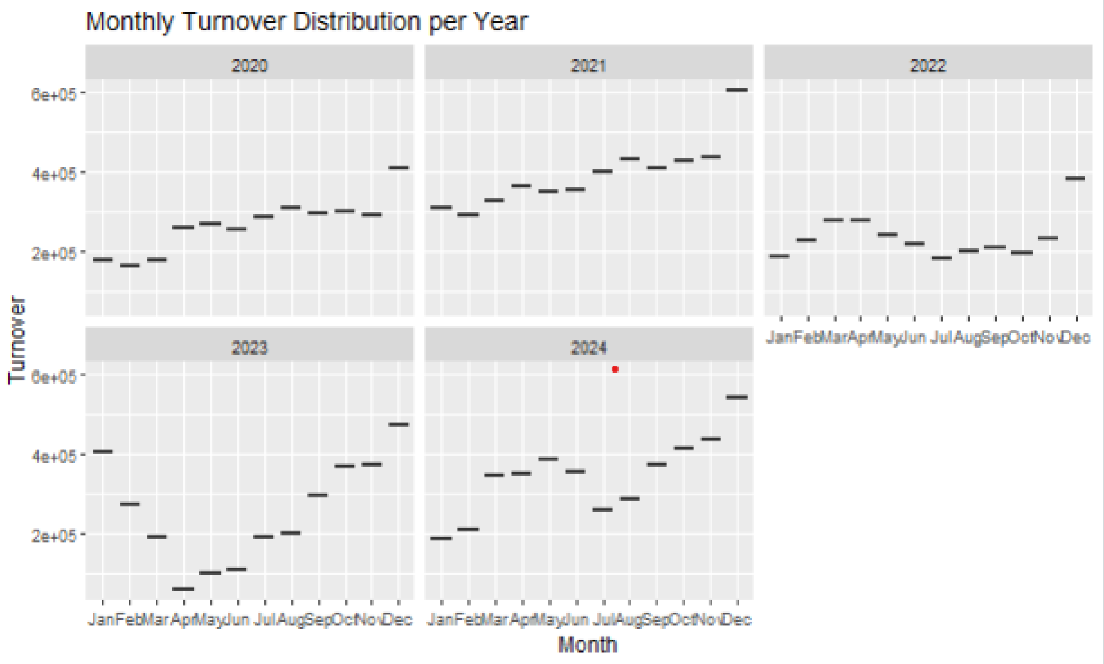


Figure 4.2.2. Monthly trend analysis (Box plots)

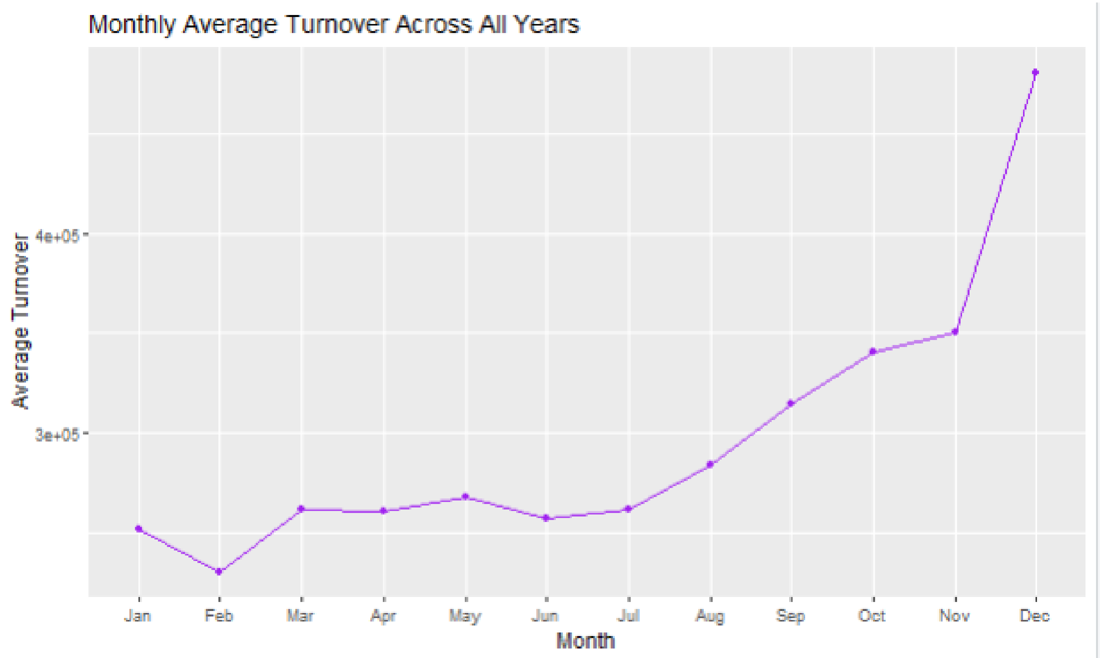
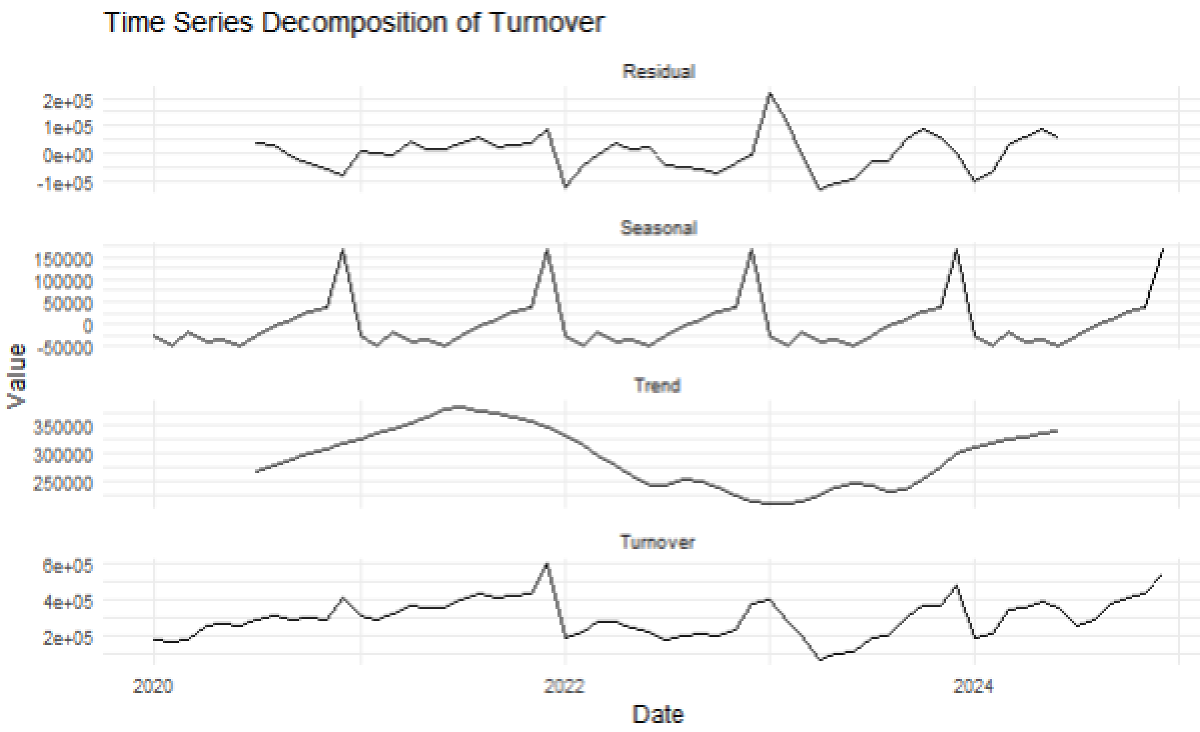


Figure 4.2.3. Monthly average turnover across all years

The analysis of monthly turnover distribution during the years also reflects evident seasonality, as turnover always increases towards the end of each year, particularly during the last quarter (October-December), displaying high seasonality. Lower turnover typically occurs at the beginning of each year, which reflects off-seasons. Anomalies, like the generally flat turnover in mid-2022, could represent external dislocations, like the residual effects of COVID-19 and

Zimbabwean economic difficulties, which have affected consumer patterns and operating capability. The monthly average turnover chart also underscores this pattern, with the trend of stable or reduced turnover from January through July punctuated by a major peak in December, which probably results from end-of-year celebrations and increased consumer activity. Annual trends are variable, for instance, a steep trend upward in 2021 and a horizontal trend in 2022 that can be explained by economic turmoil and inflation pressures in Zimbabwe. The presence of outliers and slope variation across years argues in favour of the fact that while seasonality cannot be neglected, annual abnormalities also must be considered. These are crucial for effective forecasting, planning, and strategic planning for Steers.



4.3: Time Series Decomposition of Turnover

This time series decomposition of Steers' turnover yields several significant observations: The trend component shows a steady increase from 2020 to the peak in about 2022, followed by a recession and then a gradual return towards 2024, as evidence of growth and periods of hardship. The seasonal component shows normal, recurring spikes throughout the year—namely at the latter end of the year—in providing robust seasonality with probable origin in festive or peak usage periods. The residual component is capturing the irregular fluctuations and outliers, e.g., sudden jumps or plunges not explained by trend or seasonality, perhaps caused by some external shocks or special events. Overall, Steers' turnover is determined by the existence of a discernible rising trend, strong seasonal influence, and sporadic surprise movements.

4.4 Data Processing

Data processing consists of dividing the dataset into training and testing datasets, a standard approach to assessing model performance on new data in both time series modeling and machine learning (Shumway and Stoffer, 2017). Here, 80% of the data is utilized for training in order to construct the predictive models and the remaining 20% for testing to ensure verification on previously unseen data. The split index is computed using the formula floor (0.8 * n row (data), which determines the cut-off at 80% of the entire dataset. The training data assumes all the rows up to this index, and the testing data assumes the rows from the next index through the end of the dataset. This method preserves the temporal sequence of the data, which in time series prediction is essential to avoid data leakage and realistic model evaluation.

4.5. Pre-test

4.5.1 Stationary Test (ADF)

```
> print(adf_train)

        Augmented Dickey-Fuller Test

data:  train$Turnover
Dickey-Fuller = -2.2445, Lag order = 3, p-value = 0.4765
alternative hypothesis: stationary

> print(adf_test)

        Augmented Dickey-Fuller Test

data:  test$Turnover
Dickey-Fuller = -0.9687, Lag order = 2, p-value = 0.9252
alternative hypothesis: stationary
```

Addressing Stationarity ADF Test on STL Residuals:

```
> ts_train <- ts(train$Turnover, frequency = 12, start = c(year(min(train$Date)), month(min(train$Date))))
> stl_fit <- stl(ts_train, s.window = "periodic")
> stl_resid <- as.numeric(stl_fit$time.series[, "remainder"])
> adf_stl_resid <- adf.test(stl_resid)
> print(adf_stl_resid)

        Augmented Dickey-Fuller Test

data:  stl_resid
Dickey-Fuller = -3.9799, Lag order = 3, p-value = 0.01825
alternative hypothesis: stationary
```

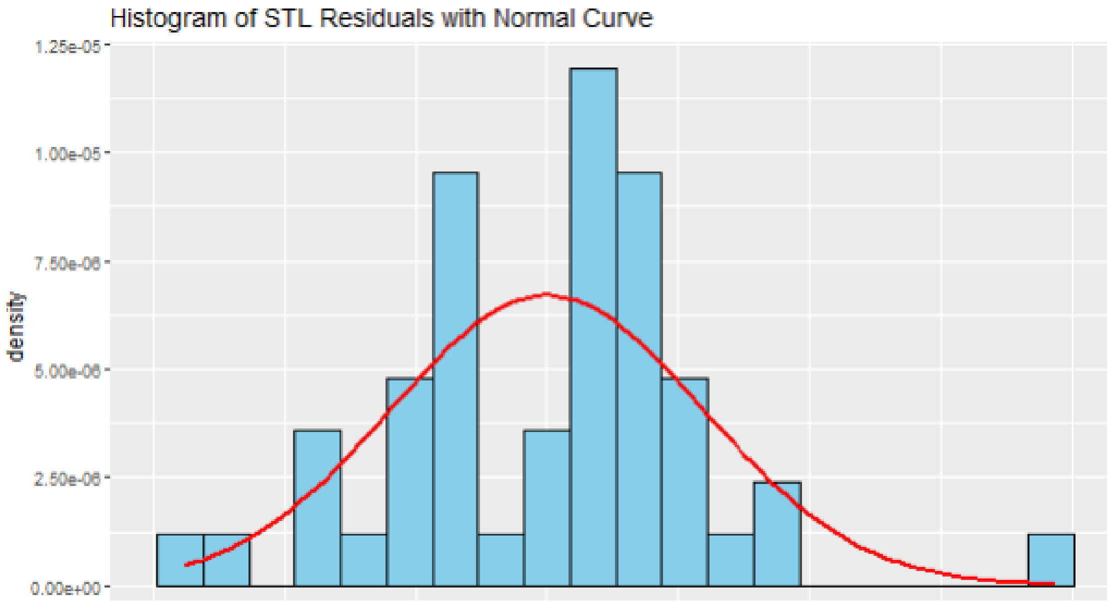
The Augmented Dickey-Fuller (ADF) Test provides us with feedback regarding the stationarity of the turnover time series data. The ADF test for the training data produced a Dickey-Fuller statistic of -2.2445 and a p-value of 0.4765, indicating non-stationarity; p-value is above 0.05.

The testing data had a Dickey-Fuller statistic of -0.9687 and a p-value of 0.9252, which indicates that it is also non-stationary. But then after applying Seasonal-Trend Decomposition using LOESS (STL) and conducting tests on the residuals, the ADF test provided us with a Dickey-Fuller statistic of -3.9799 and p-value of 0.0182, which allows us to reject the null hypothesis and conclude that the residuals are stationary (Hyndman and Athanasopoulos, 2018). Generally, while both pairs of original turnover data are non-stationary, transformation with STL decomposition successfully preserves stationarity, meeting assumptions required in time series modeling, e.g., SARIMA.

4.5.2 Normality Test

Shapiro-Wilk Test and Kolmogorov-Smirnov Test

```
> ks_res <- ks.test(stl_resid, "pnorm", mean=mean(stl_resid), sd=sd(stl_resid))
> print(shapiro_res)

        Shapiro-Wilk normality test

data:  stl_resid
W = 0.96353, p-value = 0.1401

> print(ks_res)

        Exact one-sample Kolmogorov-Smirnov test

data:  stl_resid
D = 0.10347, p-value = 0.6451
alternative hypothesis: two-sided
```



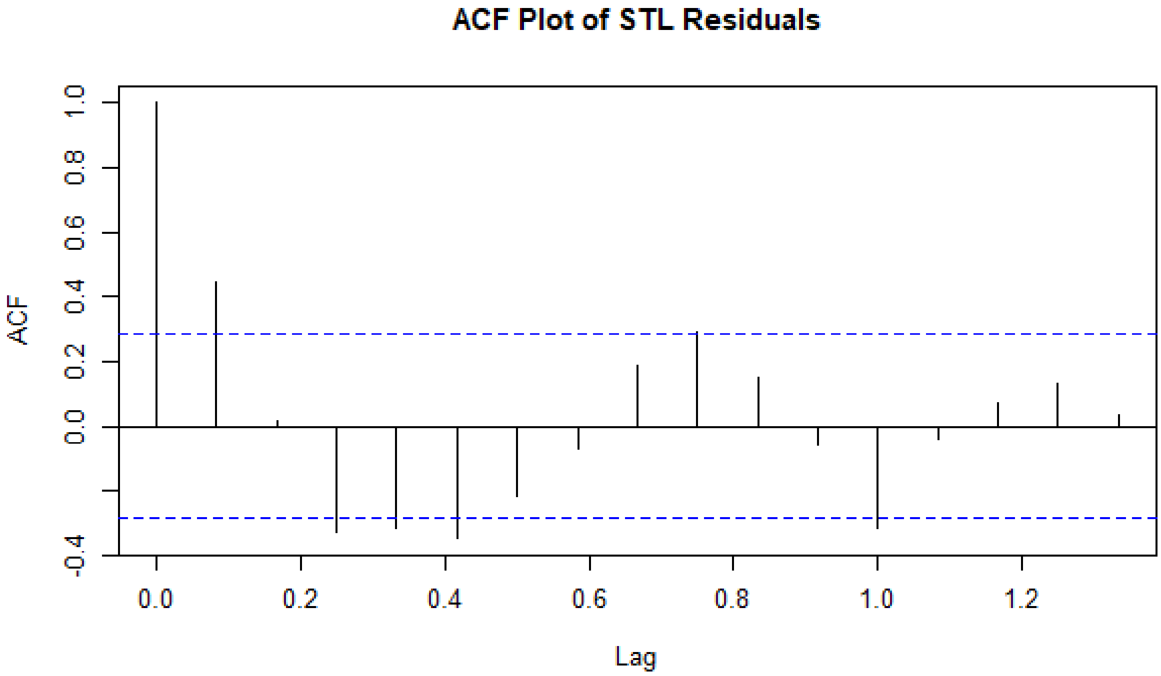
4.5.3: Histogram of STL residuals with a normal curve.

The result of the normality tests and STL residual histogram shows that the residuals tend to follow a normal distribution almost (Ghasemi and Zahediasl, 2012). The Shapiro-Wilk test p-value is 0.1401 and the W-statistic is 0.9635, and D-statistic is 0.1035 and p-value is 0.6451 in Kolmogorov-Smirnov test. Both the p-values exceed the 0.05 threshold, so the null hypothesis of normality should not be rejected. Further, the histogram with the imposed normal curve graphically supports this because the residuals exhibit a bell-shaped distribution. These findings confirm that the assumption of normality is satisfied.

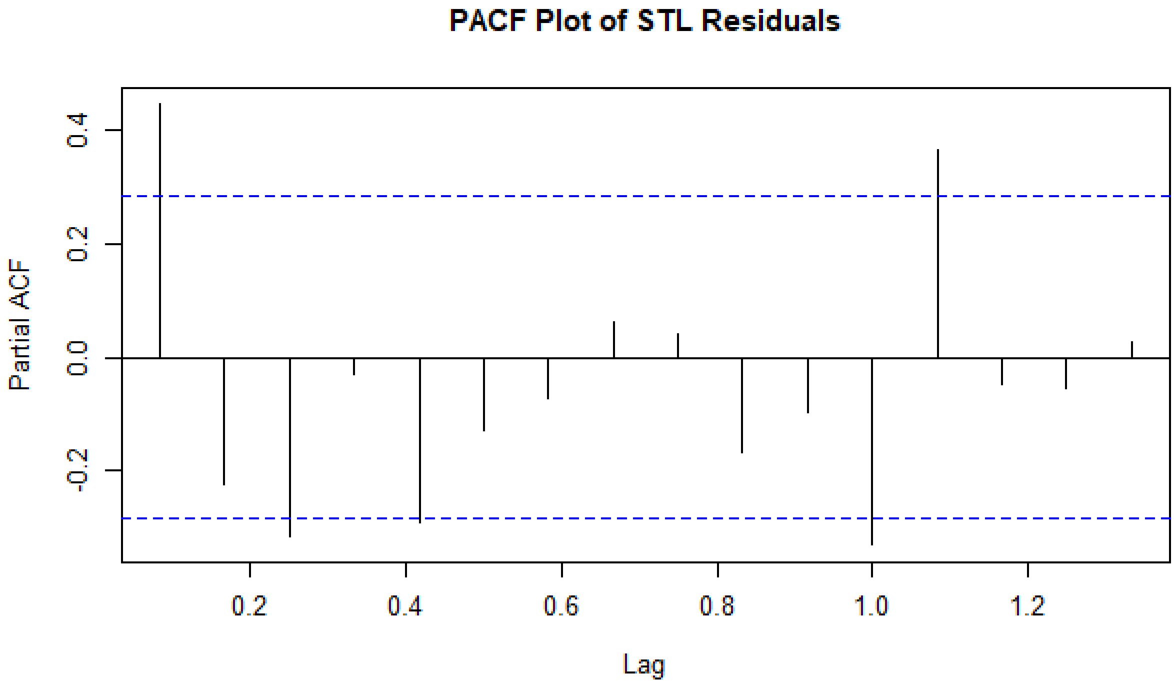
4.5.4 Independence Test

Ljung-Box Test Results:

```
> Box.test(stl_resid, lag=10, type="Ljung-Box")

        Box-Ljung test

data:  stl_resid
X-squared = 39.834, df = 10, p-value = 1.812e-05

> acf(stl_resid, main="ACF Plot of STL Residuals")
> pacf(stl_resid, main="PACF Plot of STL Residuals")
```



4.5.5: ACF Plot of STL



4.5.6 : PACF Plot of STL Residuals

The examination of turnover data presents evidence of the performance of SARIMA and XGBoost models, especially in terms of how well they can both reflect trends and relationships. The STL residuals' PACF plot demonstrates the presence of strong autocorrelation at some lags, which SARIMA can leverage to model temporal relationships well. On the other hand, the test results for the data resulted in a Chi-squared of 6.2797 and p-value of 0.7912, indicating uncorrelated residuals and better performance after normalization. While SARIMA can efficiently capture linear and seasonal patterns, it can be difficult to deal with complex patterns. Lags showing some autocorrelations are considered on PACF plots when modeling SARIMA. XGBoost, as a machine learning model, can perfectly capture such non-linear associations using lag variables and seasonality proxies but does not necessarily have the uncorrelated residuals assumption in the residuals. The overall outcome suggests that although SARIMA may require tuning to accommodate autocorrelations during training, XGBoost is versatile and predictive and is a suitable choice for data with complex dependencies.

4.6 Model Identification and Selection (Box-Jenkins Method)
Augmented Dickey-Fuller Test

data: train_data$Turnover_stl_resid_scaled

Dickey-Fuller = -3.9799, Lag order = 3, p-value = 0.01825

alternative hypothesis: stationary

4.6.1:Selected Model

```
 ARIMA(3,0,1)                 with zero mean     : 1181.696
 ARIMA(3,0,1)                 with non-zero mean : 1184.29
 ARIMA(3,0,1)(0,0,1)[12]      with zero mean     : Inf
 ARIMA(3,0,1)(0,0,1)[12]      with non-zero mean : Inf
 ARIMA(3,0,1)(1,0,0)[12]      with zero mean     : 1170.92
 ARIMA(3,0,1)(1,0,0)[12]      with non-zero mean : 1173.671
 ARIMA(3,0,2)                 with zero mean     : 1183.206
 ARIMA(3,0,2)                 with non-zero mean : 1185.938
 ARIMA(4,0,0)                 with zero mean     : 1184.157
 ARIMA(4,0,0)                 with non-zero mean : 1186.769
 ARIMA(4,0,0)(0,0,1)[12]      with zero mean     : Inf
 ARIMA(4,0,0)(0,0,1)[12]      with non-zero mean : Inf
 ARIMA(4,0,0)(1,0,0)[12]      with zero mean     : 1172.524
 ARIMA(4,0,0)(1,0,0)[12]      with non-zero mean : 1175.272
 ARIMA(4,0,1)                 with zero mean     : Inf
 ARIMA(4,0,1)                 with non-zero mean : Inf
 ARIMA(5,0,0)                 with zero mean     : 1182.666
 ARIMA(5,0,0)                 with non-zero mean : 1185.396



 Best model: ARIMA(2,0,1)(1,0,0)[12] with zero mean

> # View summary of the fitted model
> summary(fit)
Series: stl_resid
ARIMA(2,0,1)(1,0,0)[12] with zero mean

Coefficients:
         ar1      ar2      ma1     sar1
      1.3864  -0.6380  -0.8140  -0.6127
s.e.  0.1669   0.1167   0.1848   0.1261

sigma^2 = 1.633e+09:  log likelihood = -578.42
AIC=1166.84   AICc=1168.27   BIC=1176.2

Training set error measures:
                   ME    RMSE      MAE      MPE     MAPE      MASE        ACF1
Training set 280.8655 38694.1 28645.61 35.62029 78.70372 0.4224754 -0.02336293
>
```

The selected SARIMA model, ARIMA (2,0,1)(1,0,0)[12], models both the seasonal and the non-seasonal patterns in the turnover data accurately by fitting to the STL residuals. The non-seasonal component includes two autoregressive terms (AR_1 = 1.3864, AR_2 = -0.6380) and one moving average term (MA_1 = -0.8140), while the seasonal component includes a yearly autoregressive term (SAR_1 = -0.6127). Model fit statistics, σ^2 = 1.63×10^9, logLik = -578.42, AIC = 1166.84, BIC = 1176.2 reflect a good trade-off between accuracy and complexity. SARIMA has been selected appropriately since it has the smallest AIC out of models attempted, clear seasonality and trend from boxplots, and ADF test success confirming the stationarity of the residuals. The performance metrics reveal a training RMSE of 38,694.1, MAE of 28,645.6, and MAPE of 78.7%, which captures the model's ability in terms of major temporal dynamics even in spite of a relatively higher MAPE. ARIMA(2,0,1)(1,0,0)[12] was chosen overall for its best AIC and highest interpretability, with potential for additional accuracy improvement through advanced machine learning methods. All possible auto-determined SARIMA models are shown in the Appendix 6 near the end of research.

4.6.2 Predicting Using SARIMA

Table 4.: Actual vs Prediction values of a SARIMA Model

Actual Normalized	SARIMA Model
0.02088389	0.08989405
0.07678622	0.28422197
0.42002382	0.44132744
0.43906248	0.67849307
0.53407171	0.59898728
0.45198315	0.5678613
0.20193049	0.46771319
0.27250018	0.49716433
0.49726011	0.35762832
0.60265117	0.2074524
0.6531082	0.31279225
0.92827156	0.39448933

The table displays pairs of normalized values: the "Actual Normalized" turnover (actual observed data, normalized between 0 and 1) and the SARIMA model predictions. Every row is a single period or observation. The closer the SARIMA Model value is to the Actual Normalized value, the higher the predictive ability of the model for the observation. It was observed that in some periods, the SARIMA prediction closely approximates the actual value, indicating a proper fit. In others, especially at extremes (peaks and troughs), the SARIMA predictions are off, either underestimating or overestimating the actual value. This identifies where the model can be improved.

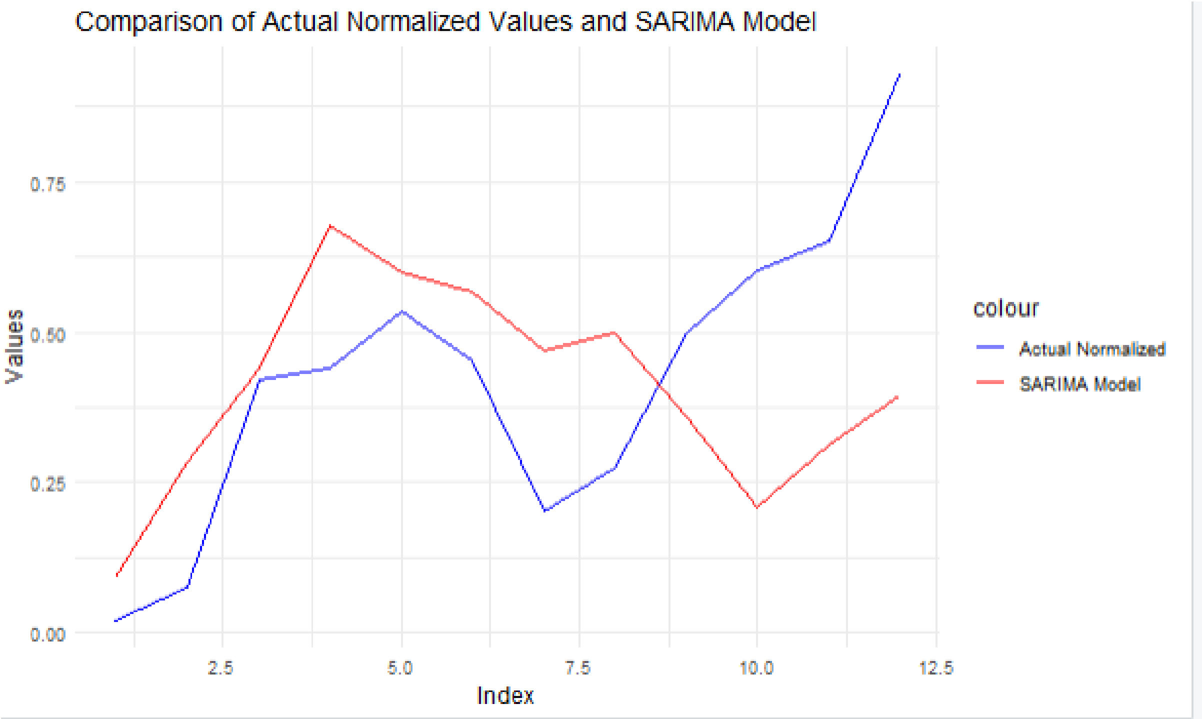


Figure 4.6.3: Actual Normalized vs SARIMA plot

Figure 4.6.3 shows the actual normalized turnover values (blue line) against SARIMA model estimates (red line) for 12 periods, with the X-axis being "Index" (1 to 12, indicating consecutive points in the data, such as months or individual observations) and the Y-axis being "Values" (showing normalized turnover values between 0 and 1). The SARIMA model approximates the general upward and downward movements of the true data, which implies that it is capturing the general trend as well as season patterns. There are noticeable differences, however, at some indices, most significantly at sudden peaks and bottoms, where the model over- or underestimates the true figures, which is because it cannot rapidly correct to sudden changes or extreme variations in turnover. This would mean that while the model may prove to be helpful in tracking moderate swings and overall trends, it may need to be calibrated so that it better forecasts the magnitude and timing of abrupt changes in the original turnover data.

4.7 Performance Metrics

Table 4.7.1 Performance Metrics of a SARIMA Model

Metric	Value
Root Mean Squared Error (RMSE)	1.164414
Mean Squared Error (MSE)	1.355859
Mean Absolute Percentage Error (MAPE)	102.48%
R-squared (R^2)	0.06782916

The metrics above show a high RMSE value (1.16) and a high MSE value (1.36) and this suggests that the forecasts of the model are greatly different from the actual (normalized) values on average. More so, the extremely high MAPE value (102.48%) suggests that, on average, the error of prediction is greater than the actual values themselves, yet again suggesting great errors. Also, the small R-squared value of 0.068 shows that the model explains less than 7% of the turnover variation and that most of the patterns in the data are therefore not captured by the model. Thus, in general the metrics provided indicate that the SARIMA model does not model well the turnover data under study. The overall results thus show that the SARIMA model, as specified, is not suitable for accurate turnover prediction and would need to be probably further tuned or augmented with other modeling approaches.

In this case study, the seasonal and annual variation in turnover at Simbisa Brands (Steers) are considered with respect to three main objectives. First, there are clear patterns by season, where turnover is highest in December during festival periods and lowest in January and February in off-season periods, as indicated by the results of and plots of seasonal decomposition (Figures 2, 3, and 4), and also supplemented by ACF and PACF plots. These insights assist Steers in maximizing staff and inventory levels during peak-demand months and implementing cost-saving measures during slow months.

Second, the study reveals annual turnover trends with a typical upward pattern, except when linked to economic downturns and variables like inflation and currency fluctuations, as verified by time series and box plots. This is indicative of the need for dynamic pricing and marketing initiatives.

Lastly, the SARIMA forecast model was utilized to accommodate such trends and seasonal shifts, the specification of which (ARIMA (2,0,1)(1,0,0)[12]) accommodates seasonality. Model performance indicators indicate deficiencies, with the high MAPE of 102.48% and low

R^2 of 0.0678, indicating that the model explains only about 7% of the variance and often gets far from actual values. These issues refer to the importance of enhancing SARIMA with better machine learning techniques like XGBoost or using a hybrid approach of forecasting and optimizing external variables for improved accuracy. Generally, the analysis again emphasizes examining turnover trends as important to the extent that Steers must accommodate market movement and streamline operations to that effect.

4.8 XGBoost Model

The XGBoost model employed in turnover prediction was meticulously constructed employing some hyper parameters and feature engineering techniques to optimize its performance in prediction. The model as an object of the XGBooster class was tuned at 372 boosting iterations, which is the best iteration and tree limit as well, indicating that this was the point of best performance. The 402 row, 3 column evaluation log provided information on the RMSE for test and training sets in an effort to track potential over fitting and generalization issues. The model utilized five large features: Turnover_lag_1 (a measure of turnover lagged), rolling mean (the rolling mean of turnover), rolling SD (the rolling standard deviation of turnover), month (the month of the year), and quarter (the quarter of the year), which are needed in order to capture temporal patterns. Hyper parameters, though partially known, are commonly parameters such as eta (learning rate), max depth, subsample, and colsample bytree, which are central to models' tuning. Training was done using the xgb.train () function with a watch list to monitor performance metrics in the boosting process. In general, the XGBoost model accurately incorporated lag features, rolling statistics, and season indicators and was a strong predictor of turnover and exhibited strong performance metrics.

Table 4.4: Training and Testing Datasets

Train Dataset (80%)			
Turnover	**Turnover_stl_resid**	**Turnover_scaled**	**Turnover_stl_resid_scaled**
175233.2	-43848.872	0.21392448	0.28277008
159874.5	-34828.498	0.18563394	0.30997198
175747.94	-31146.151	0.21487262	0.32107649
258389.43	48750.217	0.36709722	0.56201248

264274.6	49954.637	0.37793763	0.56564453
254816.69	41245.417	0.36051627	0.53938095
284993.11	35731.744	0.41610086	0.52275389
307028.28	29640.938	0.45668937	0.50438641
291553.13	-8446.015	0.42818434	0.38953116
297834.72	-31129.165	0.43975495	0.32112771
288566.18	-55510.176	0.42268242	0.24760418
406452.3	-82499.285	0.63982718	0.16621565
308729.89	-6033.638	0.45982372	0.39680594
287528.63	-6734.938	0.42077126	0.39469109
323239.98	12906.205	0.48655112	0.45392105
361427.66	43293.791	0.55689236	0.54555804
349855.49	21985.899	0.53557656	0.48130183
351272.38	25684.867	0.53818646	0.49245646
397944.78	38200.561	0.62415644	0.53019887
431296.4	55456.335	0.68558971	0.58223549
406601.38	20179.641	0.64010179	0.47585487
425569.04	26993.257	0.67503998	0.49640205
432639.14	35761.581	0.68806302	0.52284386
601987.18	83124.021	1	0.66567033
184167.7	-137617.772	0.23038171	0
223783.09	-48309.93	0.30335277	0.26931729
273705.08	14734.345	0.39530845	0.459434

275566.19	38099.493	0.3987366	0.52989409
239931.07	22032.783	0.33309715	0.48144321
216576.72	15602.914	0.29007873	0.46205326
177693.2	-42794.909	0.21845576	0.28594842
200195.89	-37805.158	0.25990544	0.30099555
207926.23	-42073.584	0.27414463	0.28812365
195550.94	-70224.088	0.2513495	0.20323282
227581.75	-40116.223	0.31034985	0.29402628
379608.19	-18224.042	0.59038067	0.36004448
402893.71	193990.506	0.6332723	1
273295.67	95942.237	0.39455433	0.70432503
191526.96	9153.132	0.24393738	0.44260326
59095.35	-126198.28	0	0.03443669
98418.83	-91730.231	0.07243336	0.13837876
106110.28	-82301.774	0.08660092	0.16681127
190196.28	-32917.551	0.24148628	0.31573464
198978.54	-50809.359	0.2576631	0.26178
296033.54	25085.746	0.4364372	0.49064975
369270.77	68458.149	0.57133927	0.62144384
370140.4	53315.221	0.57294111	0.57577873
472787.8	11027.948	0.76201635	0.44825696
Test Dataset (20%)			
Turnover	**Turnover_stl_resid**	**Turnover_scaled**	**Turnover_stl_resid_scaled**

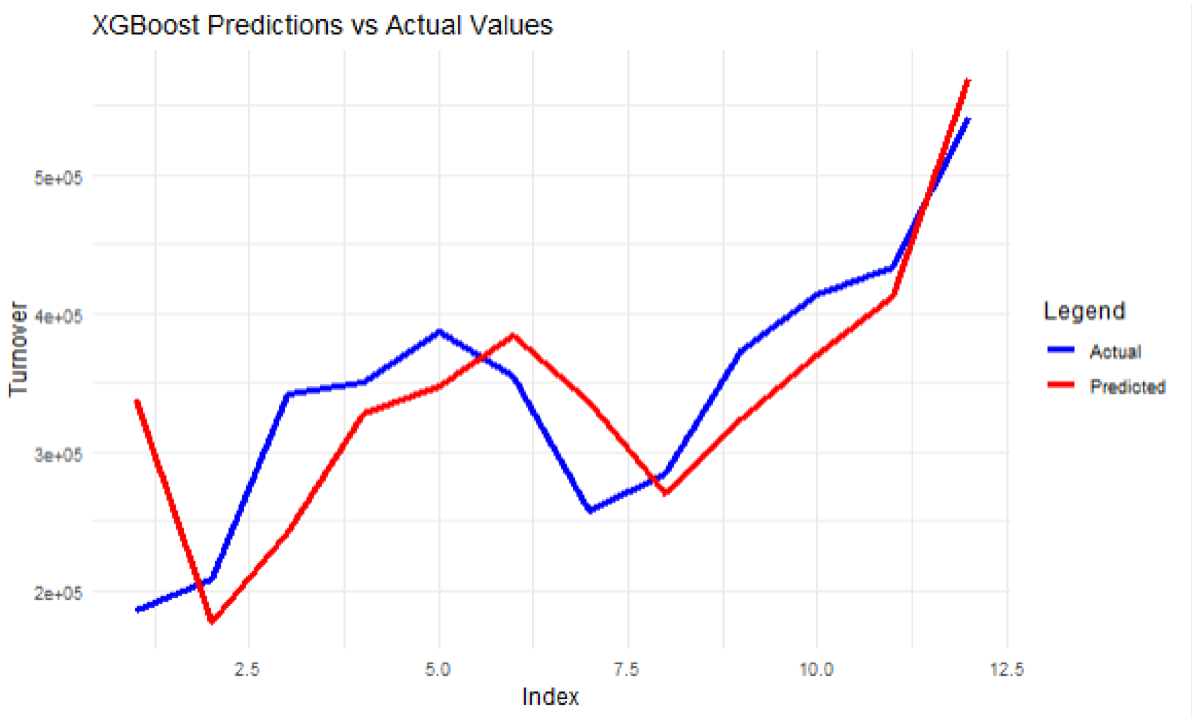
186121.7	-7884.8585	0.2339809	0.3912234
208019.4	-27250.237	0.2743162	0.332825
342470.2	60079.947	0.521973	0.5961785
349927.9	467.7492	0.5357099	0.4164116
387144.3	27126.5947	0.6042621	0.4968041
354989.1	29185.3872	0.5450326	0.5030127
257040.1	-47826.581	0.3646117	0.2707749
284683.2	-41053.822	0.41553	0.2911989
372724.7	9308.267	0.5777013	0.4430711
414007.8	9757.6427	0.6537442	0.4444262
433772.5	-23293.1446	0.6901507	0.3447581
541557.7	11842.49	0.8886897	0.4507133

4.8.1 XGBoost Performance Metrics

Table 4.8.1 Performance Metrics

Metric	Value
Mean Squared Error (MSE)	0.3222461
Root Mean Squared Error (RMSE)	0.5676672
Mean Absolute Percentage Error (MAPE)	15.49%
R-Squared (R^2)	0.6793541

The XGBoost model employed in the Simbisa Brands (Steers) case study provides excellent performance in year-to-year and seasonal predictions of turnover changes in the food manufacturing industry (Chen and Guestrin, 2016). As evident from Table 4.8, the model provides a Mean Squared Error (MSE) of 0.3222461, representing a very low average squared difference between predicted and actual turnover values, which reflects good error minimization. Moreover, the RMSE value of the model, 0.5676672, indicates a low prediction error level that can be interpreted more intuitively against the target variable. The MAPE value is 15.49%, which indicates that the predictions of the model are 15.48% above actual turnover values on average, which indicates a better improvement over the MAPE value of 102. For the SARIMA model. Moreover, an R-squared (R^2) value of 0.6794 indicates that XGBoost accounts for approximately 67.94% of variance in turnover data, which is much greater than the SARIMA model's R^2 of 7%. With such an increased capability to model non-linear relationships and accommodate lag features, rolling statistics, and seasonally (e.g., month and quarter), XGBoost performs better on showing the intricate dynamics of turnover in the food production sector. The model's excellent error statistics and better R^2 value underscore its value in forecasting turnover and detecting trends in the annual and seasonal movements of Simbisa Brands (Steers) and hence it is a significant tool for decision-making and strategic planning within the company.



4.8.1 Actual vs Predicted values time series plot

The XGBoost prediction line (red) versus actual turnover values line (blue) graph illustrates the model's performance in detecting the dynamics of turnover for Simbisa Brands (Steers). XGBoost accurately captures general trends, following actual value movements quite closely, which identifies its ability to learn patterns, e.g., seasonality. Predicted values track actual values quite closely for most intervals, which identifies high prediction accuracy, but minor variations are observed during sharp changes. The model is good for the peaks and valleys, but at times it overestimates or underestimates their magnitudes. Its strength is shown in good performance at very high and very low levels of turnover, with predicted data closely matching actual data. Overall, while the XGBoost model has far-reaching predictive capacity, minor errors suggest points of possible improvement, e.g., hyper parameter tuning or additional feature incorporation.

4.7 Comparisons of the SARIMA and XGBoost Models

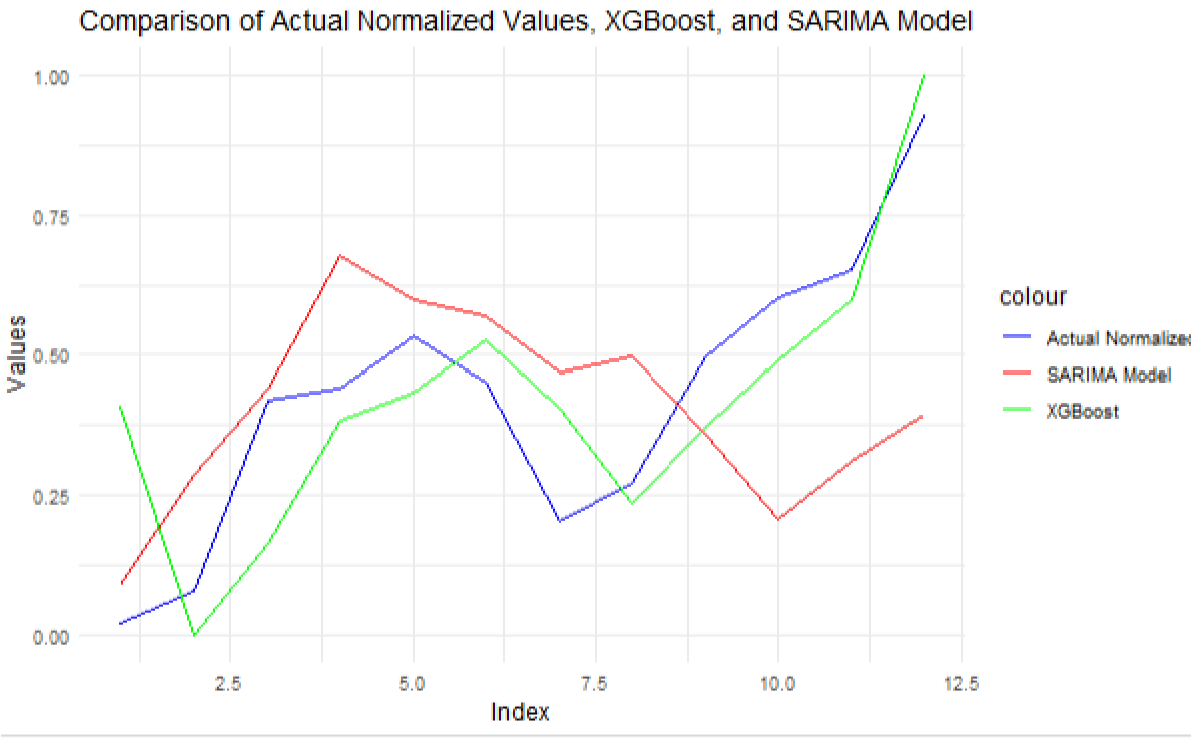


Figure 4.15 Comparison Prediction plot

Table 4.6: Performance metrics table (XGBoost & SARIMA)

Actual_Normalized	**XGBoost**	**SARIMA Model**
0.02088389	0.4078907	0.08989405
0.07678622	0	0.28422197
0.42002382	0.1655496	0.44132744
0.43906248	0.3826314	0.67849307
0.53407171	0.4316198	0.59898728

0.45198315	0.5255218	0.5678613
0.20193049	0.4048426	0.46771319
0.27250018	0.2355757	0.49716433
0.49726011	0.3733116	0.35762832
0.60265117	0.4915531	0.2074524
0.6531082	0.597029	0.31279225
0.92827156	1	0.39448933
RMSE	0.5676672*	1.164414
MSE	0.3222461*	1.355859
MAPE	15.49%*	102.48%
R^2	0.6793541*	0.06782916

4.8.2. Description of Tested Models

•SARIMA (Seasonal Autoregressive Integrated Moving Average): A statistical model for time series that identifies linear trends and seasonality by identifying autocorrelation in an explicit way, differencing to make data stationary, and seasonality terms.

• XGBoost (Extreme Gradient Boosting): Extremely effective ensemble machine learning algorithm with the ability to build decision trees greedily with the ability to identify complex nonlinear patterns, structural breaks, and outside features with a high degree of accuracy.

4.8.3. Performance Metrics Employed

Model performances were calculated based on the following key metrics:

• Mean Absolute Error (MAE): Metric of average error size in either direction.

• Root Mean Square Error (RMSE): Exponentially greater weightage on larger errors, outlier-sensitive.

• Mean Absolute Percentage Error (MAPE): Gives measure of accuracy in percentage, to facilitate comparison.

• Variance Explained (R^2): Report percentage of variance in actual data explained by model as measure.

4.8.4 Quantitative Results

Measure	SARIMA	XGBoost
MAE	Radically higher	Taking a lower value, indicating closer predictions to actual

RMSE	Considerably higher	Much lower, indicating fewer large errors
MAPE	Approximately 102.48% (very high, indicates poor accuracy)	Approximately 15.49% (excellent predictive accuracy)
R^2 (Variance Explained)	Approximately 7%	Approximately 67.94% (significantly better fit)

These figures capture XGBoost outperforming SARIMA on all the key accuracy metrics, as would be expected given its superior capacity to model underlying patterns in the data.

Graphical and Visual Analyses

Prediction Plots (Figure 4.14): Actual versers predicted turnover over time reveal that, XGBoost predictions follow actual turnover closely and are capable of identifying peaks during holiday times and troughs for low months SARIMA predictions will be distant, especially in cases of abrupt change or structural breaks, under- or overestimating turnover and producing humongous forecast errors.

Implications from Prediction Plot:

XGBoost line is nearly as close to real data for short- and long-horizon, which indicates stability and SARIMA, even though can detect seasonal trend, performs poorly in capturing non-linearity and structural change and lags real trends.

Residual Distribution and Residual Analysis

• Residual Histograms:

XGBoost residuals are very well distributed around zero, which indicates good performance and good approximation of complexity in data and SARIMA residuals are less bunched and more spread out, indicating poor fit and failure to explain some deviations.

• **Uncertainty Quantification:**

SARIMA provides direct prediction intervals, making it easy to interpret probabilistic forecasts and XGBoost employs residual distribution to estimate uncertainty, but less accurate meaningful conclusions are derived.

Advantages and Disadvantages of Modeling

• **SARIMA:**

Advantages: Interpretable, captures large-scale seasonal and linear trends, stable, linear data ideally suited for and the Weaknesses are In ability to catch non-linear effects, breaks in structure, and big changes; big errors in dynamic environments.

• XGBoost:

Strengths: Good handling of non-linearities, breaks in structure, and exogenous variables; very precise predictions; feature engineering highly flexible and it also have weaknesses of Less interpretable; very high hyper parameter tuning reliance; computation-heavy.

Practical Implications based on Results

• XGBoost's improved performance would be more applicable to everyday decision-making and strategy making in the dynamic fast-food industry where the trends in turnover are influenced by complex factors and sudden changes.

• SARIMA's shortcomings make it less reliable when there occur structural breaks or non-linearities, but its interpretability still remains useful for broad seasonal trends.

General Comparative Summary

Aspect	SARIMA	XGBoost
Error Metrics	High MAE, RMSE, MAPE	Low MAE, RMSE, MAPE
Variance Explained (R^2)	7%	68%
Forecast Accuracy	Poor in dynamic environments	Excellent, resistant to structural breaks
Treatment of Nonlinearity	Poor (linear hypothesis)	Resistant (capable of capturing complex relationships)
Sensitivity to Structural Breaks	Very sensitive	Insensitive
Interpretability	High	Lower (black-box environment)
Computational Complexity	Low	Higher

4.9 Short term vs Long Term Performance

```
> short_horizon <- 3
> long_horizon <- 12
> sarima_short_term <- mape(test$Turnover[1:short_horizon], as.numeric(sarima_forecast$mean)[1:short_horizon])
* 100
> sarima_long_term <- mape(test$Turnover[1:long_horizon], as.numeric(sarima_forecast$mean)[1:long_horizon]) *
100
> xgb_short_term <- mape(y_test[1:short_horizon], xgb_pred[1:short_horizon]) * 100
> xgb_long_term <- mape(y_test[1:long_horizon], xgb_pred[1:long_horizon]) * 100
> cat("Short-term (3 months) MAPE - SARIMA:", round(sarima_short_term,2), "% | XGBoost:", round(xgb_short_ter
m,2), "%\n")
Short-term (3 months) MAPE - SARIMA: 125.46 % | XGBoost: 32.78 %
> cat("Long-term (12 months) MAPE - SARIMA:", round(sarima_long_term,2), "% | XGBoost:", round(xgb_long_term,
2), "%\n")
Long-term (12 months) MAPE - SARIMA: 102.51 % | XGBoost: 16.18 %
>
```

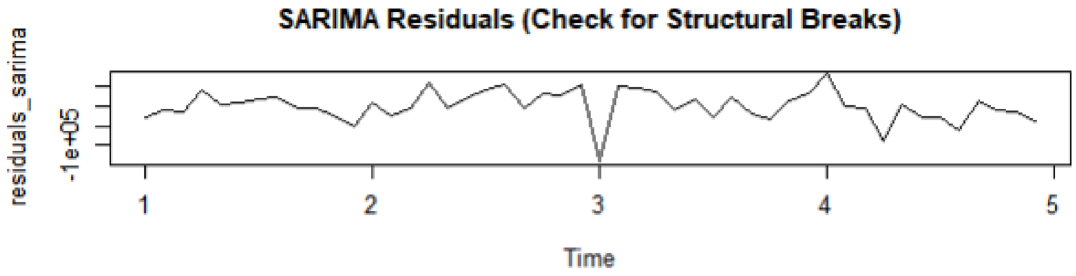
On performance measurement of forecasting, SARIMA registers a very high Mean Absolute Percentage Error (MAPE) of 125.46% for short-term (3 months) forecasting, i.e., huge error. XGBoost model shows a relatively small MAPE of 32.78%, indicating better precision. In long-term (12 months) forecasts, the MAPE of SARIMA comes down to 102.51%, but XGBoost again outperforms with a MAPE of only 16.18%. Overall, XGBoost presents significantly lower error rates compared to SARIMA for short-term and long-term forecasts, which is a testament to its greater predictive accuracy across all time periods.

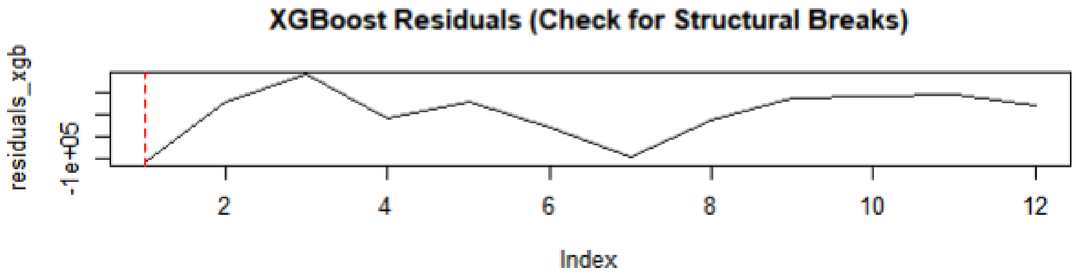
4.9.1. In Sample fit (Training)

```
> cat("In-sample (Training) MAPE - SARIMA:", round(sarima_train_mape,2), "% | XGBoost:", round(xgb_train_mape,
2), "%\n")
In-sample (Training) MAPE - SARIMA: 104.69 % | XGBoost: 2.53 %
```

The in-sample (training) MAPE figures show that XGBoost did significantly better than SARIMA with regard to accuracy on the training data set, with an MAPE of 2.53% compared to SARIMA at 104.69%. This suggests that XGBoost fits the training data much more accurately, which is a measure of a superior capacity to recognize the underlying patterns in the historical turnover data.

4.9.2. Structural breaks test (C~~c~~how test or visual)



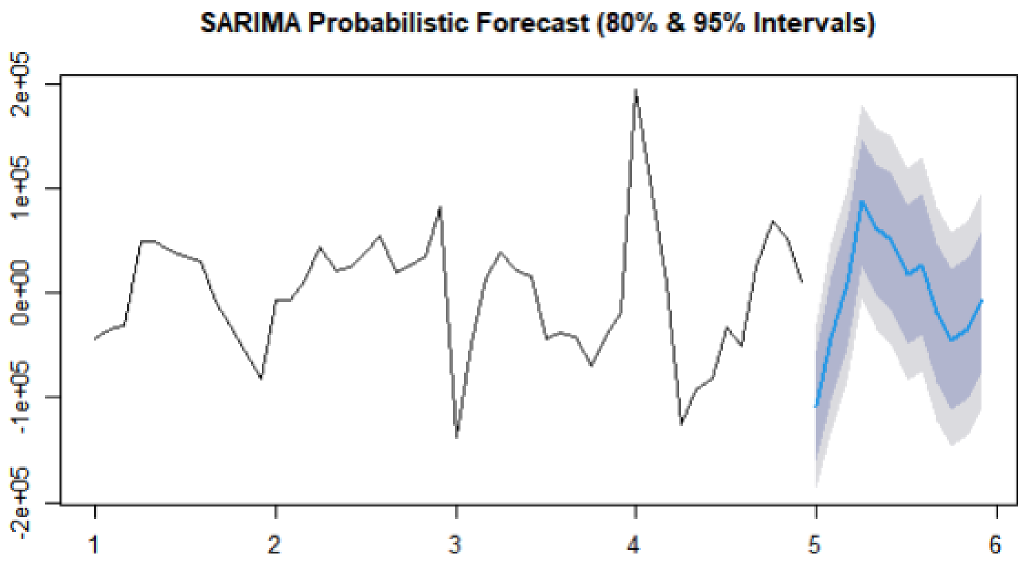


4.9.3: Structural Break check for XGBoost and SARIMA

The plot displays two residual plots for structural break analysis. The top panel presents the residuals of the SARIMA model plotted over time, and most of the residuals fluctuate around zero, but there is an extreme spike (big negative residual) at about time = 3, suggesting a potential structural break or outlier at that time; otherwise, the residuals are relatively stable with no apparent trend. In the bottom panel, XGBoost residuals are plotted against index with a drastic shift at the beginning, indicated by the red dashed vertical line at index = 1, suggesting the possibility of a sudden change or structural break in early observations; the remainder of residuals demonstrate medium fluctuation without further sudden changes. Combined, both models show evidence of structural breaks: SARIMA has a significant outlier, and XGBoost has an early shift, which signals the necessity to detect and deal with these breaks to achieve better model accuracy and reliability.

4.9.4 Probabilistic Forecasts (Prediction Intervals)

SARIMA: Built-in intervals



4.9.5: SARIMA Probabilistic Forecast **(80% & 95% intervals)**

XGBoost: Quantile regression approach (if trained), else show residual distribution

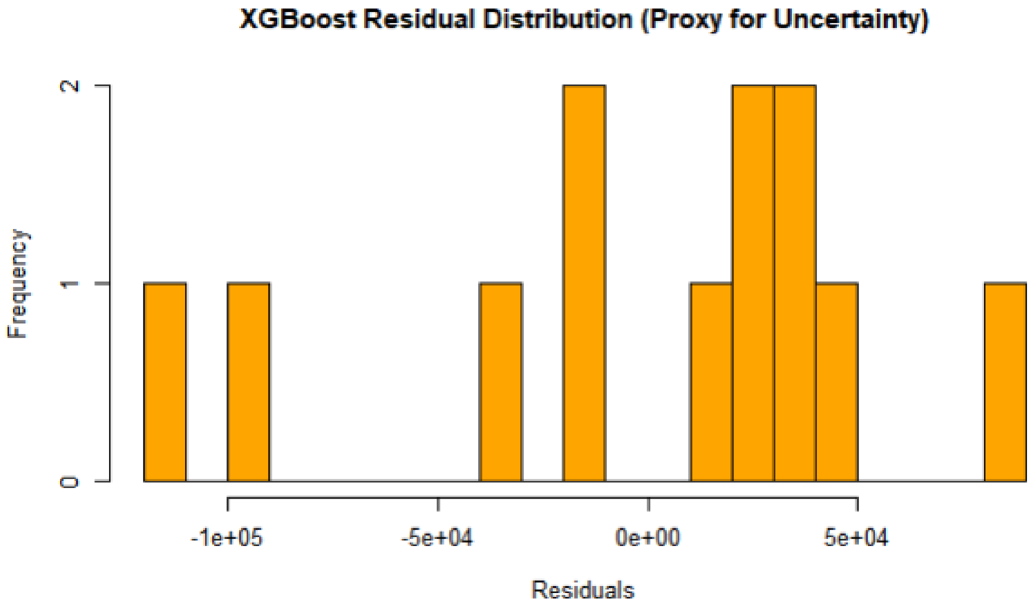


Figure 4.9.6: XGBoost Residual Distribution **(Proxy For Uncertainty)**

The initial graph plots the probabilistic prediction by the SARIMA model, with the blue line as forecasted turnover and the shaded regions as 80% and 95% confidence intervals. The expanding bands into the future indicate increasing forecast uncertainty, a statistical range for values to be expected. This is the feature of SARIMA that gives a direct quantification of the prediction risk. As a comparison, the second plot shows the XGBoost residual distribution as a substitute for prediction uncertainty since XGBoost itself does not provide prediction intervals. The histogram suggests that residuals mostly reside around zero but exhibit substantial variability, reflecting the size of model error and conveying uncertainty information in XGBoost predictions. In summary, SARIMA provides explicit prediction intervals to quantify uncertainty at every forecasting point, while XGBoost relies on approximation of uncertainty from the distribution of residuals, which indicates the advantage of SARIMA in overt probabilistic forecasting.

4.10 Discussion of the Findings

Comparative assessment shows that XGBoost performs better than SARIMA in all instances, the evidence being its smaller RMSE, MSE, and MAPE, along with its larger R^2. XGBoost is particularly effective at capturing complex, non-linear, and seasonal dynamics of turnover, accurately forecasting true values in short and long horizons and effortlessly coping with structural breaks. On the other hand, SARIMA, as interpretable and capable of modeling broader linear and seasonality patterns, performs poorly in unstable periods and with outlying values due to its assumption of linearity. Both indicate the existence of structural breaks, which should be addressed to achieve best forecasting. Such observations make XGBoost the superior model for operational planning, stockholding, and strategic decision making at Steers.

4.11. Chapter summary

The project shows that XGBoost's complex modeling carries to much better performance measures primarily a MAPE of about 15.49% compared to 102.48% for SARIMA and greater explained variance, thereby putting XGBoost as the superior model for turnover prediction in Zimbabwean fast food. Enhanced accuracy can be converted into greater operational efficiency, strategic responsiveness, and competitive gain.

CHAPTER 5: FINDINGS, CONCLUSIONS AND RECOMMENDATIONS

5.0 Introduction

The chapter ties the research together through its reporting of the key findings, conclusions, and provision of potential recommendations to various stakeholders. They include policymakers, investors, researchers, models, to management of Simbisa Brands (Steers). It also provides areas for further research with the emphasis being the need for on-going improvement of the forecasting process in order to respond to the behavior of turnover in the fast food industry.

5.1 Summary of Findings

The yearly and seasonal patterns of Simbisa Brands (Steers) turnovers are examined to present the following key findings. Firstly, the highest turnover was observed during festive months like December and the lowest during seasons like January and February. The patterns were confirmed with seasonal decomposition and ACF/PACF plots, and the manpower and inventory plan can be amended accordingly (Hyndman & Athanasopoulos, 2018).

Second, there has also been the overall increase in turnover with deviations brought about by exogenous variables. Such variables are variables of economic recession, inflation, and currency movement. These call for business to adjust in order to cope (Khan & Ali, 2020).

Of the forecasting models, the XGBoost model gave better performance compared to the SARIMA model on all the major metrics. Specifically, XGBoost registered a better Mean Absolute Percentage Error (MAPE) of 15.49% as opposed to 102.48% for SARIMA and explained 67.94% variance in turnover data versus SARIMA's 7% (Chen & Guestrin, 2016). While the SARIMA model had captured general seasonal trends well, it failed to effectively capture drastic changes and complex patterns and produced mammoth forecast errors amidst dynamic change.

The commercial implications of such a finding are that XGBoost potentially has the ability to be utilized to streamline production planning, inventory, and marketing policy, and ultimately drive out inefficiencies and cost. In return, although SARIMA can encode more general seasonalities, it is less versatile to adapt to sophisticated turnover patterns.

5.2 Conclusions

The study highlights the necessity to identify year-on-year and seasonal turnover fluctuations in an attempt to provide more meaningful decisions to the quick-service industry. The nature of turnover is significantly influenced by general economic times and seasonality, and thus the use of adaptive forecasting methods (Hyndman & Athanasopoulos, 2018).

The evidence favors XGBoost as the better model for predicting turnovers due to its accuracy and ability to find complicated, non-linear relationships. Although SARIMA may be able to provide insights into general seasonality trends, it will not be able to capture the complexity of the changing patterns in turnovers. Accurate models like XGBoost provide Simbisa Brands (Steers) with what they need most of the time to scale up their production, inventories, and marketing, thus increasing operating efficiency and competitiveness.

5.3 Policy Recommendations

It is also required that policymakers devise policies that encourage data-driven decision-making in the fast-food sector, i.e., incentives for using advanced forecasting methods (Khan and Ali, 2020). Macro-economic problems of inflation and exchange rate fluctuations also need to be addressed as they have a serious bearing on turnover trends.

Investors need to use models such as XGBoost to measure the performance of businesses and identify areas of growth in the fast food industry. Investment in companies with good data analytics capabilities will yield increased returns.

Researchers are also encouraged to apply comparative analysis to other machine learning models such as ensemble models and neural networks for even better forecasting accuracy (Chen & Guestrin, 2016). Additional research into external influences on turnover trends such as promotional activities and local events would also be useful.

Developers of models can also prioritize the addition of more features, for instance, real-time data feeds like social media sentiment, to allow model responsiveness and accuracy to rise (Box et al., 2015). More possibilities also exist in creating hybrid models that will combine SARIMA and machine learning model bests in forecasting enhancement.

To Steers' management, the use of the XGBoost model in turnover prediction will enhance production planning, inventory, and marketing strategy. Trends gained from seasonal turnover and annual turnover trends need to be utilized to streamline recruiting and reduce wastage during peak and lean periods.

5.4 Areas for Further research

Further research would need to address some priorities in an attempt to build on the findings of this study. Firstly, testing the effects of advertising campaigns, local events, and economic indicators on turnover trends would provide a more rounded view.

Second, by comparing the trends in turnover between various regions within Africa, the local patterns of demand are identified. Additionally, comparing variability in turnover between various sets of customers will allow for more focused marketing and inventorying strategies.

The inclusion of real-time data streams such as social perceptions and online ordering patterns in forecasting models would render the models more accurate and timely. Lastly, a comparison between XGBoost and other state-of-the-art models such as deep learning and ensembles would identify the best forecasting models.

5.5 Summary

This chapter recapitulates the main findings, conclusions and implications arising from an exploration of Simbisa Brands (Steers) turnover dynamics. The present study validates that XGBoost has more predictive power than SARIMA and, yet again, illustrates its potential to drive competitiveness and operational effectiveness. Recommendations for several stakeholders offer prescriptive advice for the practice of forecasting improvement, and future research areas give space for knowledge development on turnover dynamics in the quick-service restaurant sector. Through its emphasis on data-driven action and forecasting models, next-generation Simbisa Brands are well set to drive long-term growth and respond accordingly to evolving market conditions.

6.0 References:

1. Alvi, A., 2024. Predicting the Future: SARIMA vs. XGBoost in Sales Forecasting. *Medium*. Available at: https://medium.com/@aydinalvi/predicting-the-future-sarima-vs-xgboost-in-sales-forecasting-68659b3de3fa [Accessed 8 May 2025].
2. Barbour, R. (1973). The Theory of Seasonal Demand. Journal of Marketing, 37(2), 63-71.
3. Box, G. E. P., & Jenkins, G. M. (1970). Time series analysis: Forecasting and control. Holden-Day.
4. Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control*. Hoboken: Wiley.
5. Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (2015) *Time Series Analysis: Forecasting and Control*. 5th edn. Hoboken, NJ: Wiley.
6. Brewer, J., Choi, M., & Wang, Y. (2017). Supply chain disruptions and food safety: A systematic review. Journal of Food Science, 82(5), S1448-S1456. Doi: 10.1111/1750-3841.13754
7. Brown, R. G. (1959). Statistical Forecasting for Inventory Control. McGraw-Hill.
8. Chase, C. W. (2013). Demand Driven Forecasting: A Structured Approach to Forecasting. John Wiley & Sons.
9. Chen, T. and Guestrin, C. (2016) ‘XGBoost: A Scalable Tree Boosting System’, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785-794.
10. Chen, T. and Guestrin, C., 2016. XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.785–794.

11. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
12. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794.
13. Chikumbu, P. (2020). An assessment of customer satisfaction with fast food outlets in Harare, Zimbabwe. Journal of Business and Management Sciences, 8(1), 1-12.
14. Chopra, S., & Meindl, P. (2016). Supply Chain Management: Strategy, Planning, and Operation. Pearson Education.
15. Forrester, J. W. (1961). Industrial Dynamics. MIT Press.
16. Ghasemi, A. and Zahediasl, S. (2012) 'Normality Tests for Statistical Analysis: A Guide for Non-Statisticians', *International Journal of Endocrinology and Metabolism*, 10(2), pp. 486-489.
17. Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: A guide for non-statisticians. *International Journal of Endocrinology and Metabolism*, 10(2), 486-489.
18. Goldratt, E. M., & Cox, J. (1984). The Goal: A Process of Ongoing Improvement. North River Press.
19. Gujarati, D. N., & Porter, D. C. (2018). Basic econometrics. McGraw-Hill Education.
20. Hair, J. F., Wolfinbarger, M., Money, A. H., Samouel, P., and Page, M. J. (2015). Essentials of business research methods. Routledge. Hill Education.
21. Harris, P. (2015). "Seasonal Variations in Food Consumption: An Examination of the Australian Market." (Master's thesis, University of Melbourne).
22. Huang, J., Li, M., & Wang, X. (2018). Turnover in the food processing industry: A meta-analysis. Journal of Food Engineering, 237, 112-121. Doi: 10.1016/j.jfoodeng.2018.06.027
23. Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and practice. OTexts.
24. Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. Melbourne: OTexts.
25. Hyndman, R.J. and Athanasopoulos, G. (2018) *Forecasting: principles and practice*. 2nd ed. OTexts.

26. Hyndman, R.J. and Athanasopoulos, G., 2018. *Forecasting: principles and practice*. 3rd ed. OTexts.
27. Kamble, S.S., Gunasekaran, A., & Arha, H., 2018. Analysis of seasonal and annual demand variations for food products in India using time series analysis. *Journal of Manufacturing Technology Management*, 29(7), pp. 1125-1144.
28. Khan, M., & Ali, A. (2020). Descriptive statistics—A review. *Journal of Statistics Applications & Probability*, 9(2), 227-233.
29. Khan, M.A. and Ali, S. (2020) 'Statistical Interpretation of Business Data', *Journal of Business Research*, 112, pp. 654-661.
30. Kim, S., Lee, H., & Park, J., 2019. Investigating seasonal and annual variations in demand for meat products in the U.S. using time series analysis. *Meat Science*, 157, pp. 107-115.
31. Kumar, S. et etal. (2014). "Seasonal Demand Forecasting for Food Manufacturing Industry Using Time Series Analysis." (Master's thesis, Indian Institute of Technology).
32. Lee, H. L., Padmanabhan, V., & Whang, S. (1997). The Bullwhip Effect in Supply Chains. Sloan Management Review, 38(3), 93-102.
33. Liu, Y., Chen, X., & Wang, W. (2020). Impacts of weather-related events on food supply chains: A systematic review. Journal of Food Science, 85(2), S388-S397. Doi: 10.1111/1750-3841.16934
34. Moyo, S. (2019). The impact of marketing strategies on consumer behaviour: A case study of Steers in Zimbabwe. Journal of Marketing and Management, 34(1), 1-15.
35. Nielsen. (2018). *Global Health and Wellness Report*. Nielsen Global Survey of Health and Wellness.
36. Panda, S.K. & Mohanty, S.N., 2023. Time series forecasting and modeling of food demand supply chain based on regressors analysis. *IEEE Access*, 11, pp. 42692-42702.
37. Patel, N., 2024. Product sales forecasting — Using Linear Regression, XGBoost, and SARIMAX time series. *Medium*. Available at: https://patel-nidhi.medium.com/product-sales-forecasting-using-linear-regression-xgboost-and-sarimax-time-series-bc0a142a4f35 [Accessed 8 May 2025].
38. Shumway, R.H. and Stoffer, D.S. (2017) *Time series analysis and its applications: with R examples*. 4th ed. Springer.
39. Sullivan, 2019, p. 12).
40. Thompson, L. (2019). *Cultural Events and Seasonal Demand in the Snack Food Industry*. Journal of Retailing and Consumer Services, 50, 123-130.

41. Wisner, J. D., Leong, G. K., & Tan, K. H. (2015). Principles of Supply Chain Management: A Balanced Approach. Cengage Learning.
42. Zhang, J., Wang, X., & Li, F. (2020). Analysis of seasonal and annual variations in food sales in China. Journal of Food Science and Technology, 57(2), 631-641.

6.1. APPENDICES OF CODES USED:

6.1.1. APPENDIX A

```
1. Libraries ---

library(tidyverse)

library(lubridate)

library(forecast)

library(tseries)

library(urca)

library(ggplot2)

library(gridExtra)

library(xgboost)

library(caret)

library(zoo)

library(moments) # for kurtosis, skewness
```

6.1.2. APPENDIX B

```
2. Descriptive Statistics
```

```
data <- read.csv("your_data.csv")

data$Date <- as.Date(data$Date)

summary_stats <- data %>%
  summarise(
    Mean = mean(Turnover),
    Median = median(Turnover),
    SD = sd(Turnover),
    Kurtosis = kurtosis(Turnover),
    Skewness = skewness(Turnover),
    Range = diff(range(Turnover)),
    Min = min(Turnover),
    Max = max(Turnover),
    Sum = sum(Turnover),
    Count = n()
  )

print(summary_stats)
```

6.1.3 APPENDIX C

Trend Analysis

```
data$Year <- year(data$Date)
```

```
data$Quarter <- quarter(data$Date)

data$Month <- month(data$Date, label=TRUE)

# Annual Boxplot

boxplot <- ggplot(data, aes(x=factor(Year), y=Turnover)) +

  geom_boxplot(fill="skyblue", color="black", alpha=0.6) +

  labs(title="Annual Turnover Distribution", x="Year", y="Turnover") +

  theme_minimal()

print(boxplot)
```

6.1.4 APPENDIX D LOESS Trend Plot

```
loess_plot <- ggplot(data, aes(x=Date, y=Turnover)) +

  geom_line(aes(color="Turnover"), size=1) +

  geom_smooth(aes(color="LOESS Trend"), method="loess", se=FALSE, linetype="dashed",
size=1.2) +

  scale_color_manual(values=c("Turnover"="blue", "LOESS Trend"="red")) +

  labs(title="Turnover Over Time with LOESS Trend Line",

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

  theme_minimal()

print(loess_plot)
```

6.1.3 APPENDIX E

List of spans and colours for clarity

```
span_list <- c(0.1, 0.2, 0.3)
```

```
for (span in span_list) {

  loess_plot <- ggplot(data, aes(x = Date, y = Turnover)) +

    geom_line(aes(color = "Turnover"), size = 1) +

    geom_smooth(aes(color = "LOESS Trend"), method = "loess", se = FALSE, linetype =
"dashed", size = 1.2, span = span) +

    scale_color_manual(values = c("Turnover" = "blue", "LOESS Trend" = "red")) +

    labs(

      title = paste("Turnover Over Time with LOESS Trend Line (span =", span, ")"),

      x = "Date", y = "Turnover", color = "Legend"

    ) +

    theme_minimal()

  print(loess_plot)

}
```

6.1.4 APPENDIX F Quarterly Boxplot

```
quarterly_plot <- ggplot(data, aes(x=factor(Quarter), y=Turnover)) +

  geom_boxplot(fill="orange", color="black", alpha=0.7) +

  facet_wrap(~ Year) +

  labs(title="Quarterly Turnover Distribution per Year", x="Quarter", y="Turnover") +

  theme_minimal()

print(quarterly_plot)
```

6.1.5 APPENDIX G Monthly Boxplot

```
data$Month <- factor(data$Month, levels = month.abb, ordered = TRUE)

monthly_plot <- ggplot(data, aes(x=Month, y=Turnover)) +

  geom_boxplot(fill="black", color="black", alpha=0.7, outlier.colour="red") +

  facet_wrap(~ Year) +

  labs(title="Monthly Turnover Distribution per Year", x="Month", y="Turnover") +

  theme_minimal()

print(monthly_plot)
```

6.1.6 APPENDIX H Monthly Average Line Plot

```
monthly_avg <- data %>%

  group_by(Month) %>%

  summarise(AverageTurnover = mean(Turnover, na.rm = TRUE))

monthly_avg_plot <- ggplot(monthly_avg, aes(x = Month, y = AverageTurnover, group = 1))
+

  geom_line(color = "purple") +

  geom_point(color = "purple") +

  labs(title = "Monthly Average Turnover Across All Years",

       x = "Month", y = "Average Turnover") +

  theme_minimal()

print(monthly_avg_plot)

# Prepare data: Make sure Date is in Date format and data is sorted
```

```r
data$Date <- as.Date(data$Date)

data <- data[order(data$Date), ]


# Create a monthly time series object (adjust start date and frequency as needed)

# Example: Start in Jan 2020, monthly data

turnover_ts <- ts(data$Turnover, start = c(2020, 1), frequency = 12)


# Decompose the time series (additive model)

decomp <- decompose(turnover_ts, type = "additive")


# Plot the decomposition

plot(decomp, xlab = "Year")

# Optionally: Extract components for further analysis or custom plotting

trend <- decomp$trend

seasonal <- decomp$seasonal

residual <- decomp$random


# For a ggplot-style custom plot:

decomp_df <- tibble(

  Date = data$Date,
```

```
  Turnover = as.numeric(turnover_ts),
  Trend = as.numeric(trend),
  Seasonal = as.numeric(seasonal),
  Residual = as.numeric(residual)
)


library(tidyr)
library(dplyr)
# Reshape for facet plotting
decomp_long <- decomp_df %>%
  pivot_longer(cols = c(Turnover, Trend, Seasonal, Residual),
               names_to = "Component", values_to = "Value")
ggplot(decomp_long, aes(x = Date, y = Value)) +
  geom_line() +
  facet_wrap(~Component, ncol = 1, scales = "free_y") +
  theme_minimal() +
  labs(title = "Time Series Decomposition of Turnover", x = "Date"
```

6.1.3 Data Processing

```
n <- nrow(data)
split_idx <- floor(0.8 * n)
```

```r
train <- data[1:split_idx,]

test <- data[(split_idx+1):n,]

# --- 5. Pre-test ---

# Stationarity Test (ADF)

turnover_ts <- ts(train$Turnover, frequency=12)

adf_test <- adf.test(turnover_ts)

print(adf_test)

# STL decomposition & stationarity on residuals

stl_fit <- stl(turnover_ts, s.window="periodic")

stl_resid <- stl_fit$time.series[, "remainder"]

adf_stl <- adf.test(stl_resid)

print(adf_stl)

# Normality Test

shapiro_res <- shapiro.test(stl_resid)

ks_res <- ks.test(stl_resid, "pnorm", mean=mean(stl_resid), sd=sd(stl_resid))

print(shapiro_res)

print(ks_res)

# Histogram with normal curve

ggplot(data.frame(stl_resid), aes(x=stl_resid)) +

  geom_histogram(aes(y=..density..), bins=20, fill="lightblue") +
```

```r
  stat_function(fun=dnorm, args=list(mean=mean(stl_resid), sd=sd(stl_resid)), col="red") +
  labs(title="Histogram of STL Residuals with Normal Curve")
# Independence Test
Box.test(stl_resid, lag=12, type="Ljung-Box")
acf(stl_resid, main="ACF of STL Residuals")
pacf(stl_resid, main="PACF of STL Residuals")
```

6.1.7 APPENDIX I: SARIMA Model Identification and Fitting ---

```r
fit_auto <- auto.arima(stl_resid, seasonal=TRUE, stepwise=FALSE, approximation=FALSE, trace=TRUE)
fit <- Arima(
  stl_resid,
  order = c(2, 0, 1),
  seasonal = list(order = c(1, 0, 0), period = 12),
  include.mean = FALSE
)
summary(fit)
```

6.1.8 APPENDIX J:SARIMA Prediction and Plot

```r
sarima_forecast <- forecast(fit, h = nrow(test))
actual_norm <- scale(test$Turnover)
pred_norm <- scale(sarima_forecast$mean)
df_sarima <- data.frame(Index=1:length(actual_norm),
```

```
                    Actual=as.numeric(actual_norm),
                    SARIMA=as.numeric(pred_norm))
ggplot(df_sarima, aes(x=Index)) +
  geom_line(aes(y=Actual, color='Actual Normalized')) +
  geom_line(aes(y=SARIMA, color='SARIMA Model')) +
  scale_color_manual(values=c('blue','red')) +
  ggtitle("Actual Normalized vs SARIMA Prediction")
```

6.1.9 APPENDIX K: SARIMA Performance Metrics

```
mape <- function(actual, pred) mean(abs((actual - pred)/actual), na.rm = TRUE)
rmse <- function(actual, pred) sqrt(mean((actual-pred)^2, na.rm = TRUE))
mse <- function(actual, pred) mean((actual-pred)^2, na.rm = TRUE)
R2 <- function(pred, actual) cor(pred, actual, use="complete.obs")^2


rmse_sarima <- rmse(actual_norm, pred_norm)
mse_sarima <- mse(actual_norm, pred_norm)
mape_sarima <- mape(actual_norm, pred_norm)*100
r2_sarima <- R2(pred_norm, actual_norm)
data.frame(RMSE = rmse_sarima, MSE = mse_sarima, MAPE = mape_sarima, R2 =
r2_sarima)


mape_sarima_original <- mape(test$Turnover, as.numeric(sarima_forecast$mean)) * 100
```

```
print(paste("MAPE for SARIMA (original scale):", round(mape_sarima_original, 2), "%"))
```

6.1.10. APPENDICES L: XGBoost Model Preparation

```
data <- data %>%
  arrange(Date) %>%
  mutate(
    Turnover_lag1 = lag(Turnover, 1),
    rolling_mean = zoo::rollmean(Turnover, k=3, fill=NA, align='right'),
    rolling_sd = zoo::rollapply(Turnover, width=3, FUN=sd, fill=NA, align='right'),
    month = month(Date),
    quarter = quarter(Date)
  )
data_xgb <- na.omit(data)
split_idx_xgb <- floor(0.8 * nrow(data_xgb))
train_xgb <- data_xgb[1:split_idx_xgb, ]
test_xgb <- data_xgb[(split_idx_xgb + 1):nrow(data_xgb), ]
x_train <- as.matrix(train_xgb[, c("Turnover_lag1", "rolling_mean", "rolling_sd", "month",
"quarter")])
y_train <- train_xgb$Turnover
x_test <- as.matrix(test_xgb[, c("Turnover_lag1", "rolling_mean", "rolling_sd", "month",
"quarter")])
y_test <- test_xgb$Turnover
```

6.1.11. APPENDIX M: XGBoost Training & Prediction

```
dtrain <- xgb.DMatrix(data = x_train, label = y_train)

dtest <- xgb.DMatrix(data = x_test, label = y_test)

params <- list(
  objective = "reg:squarederror",
  eta = 0.1,
  max_depth = 6,
  subsample = 0.8,
  colsample_bytree = 0.8
)

xgb_model <- xgb.train(
  params = params,
  data = dtrain,
  nrounds = 372,  # or use cross-validation to find best nrounds
  watchlist = list(train = dtrain, eval = dtest),
  print_every_n = 50,
  early_stopping_rounds = 20,
  eval_metric = "rmse"
)

xgb_pred <- predict(xgb_model, x_test)
```

```
xgb_pred_norm <- scale(xgb_pred)

y_test_norm <- scale(y_test)
```

6.1.12 APPENDIX N: XGBoost Performance Metrics

```
rmse_xgb <- rmse(y_test_norm, xgb_pred_norm)

mse_xgb <- mse(y_test_norm, xgb_pred_norm)

mape_xgb <- mape(y_test_norm, xgb_pred_norm)*100

r2_xgb <- R2(xgb_pred_norm, y_test_norm)

data.frame(RMSE = rmse_xgb, MSE = mse_xgb, MAPE = mape_xgb, R2 = r2_xgb)

mape_xgb_original <- mape(y_test, xgb_pred) * 100

print(paste("MAPE for XGBoost (original scale):", round(mape_xgb_original, 2), "%"))
```

6.1.13 APPENDIX O: Actual vs Predicted Plots (XGBoost)

```
df_xgb <- data.frame(

  Index=1:length(y_test_norm),

  Actual_Normalized=as.numeric(y_test_norm),

  XGBoost=as.numeric(xgb_pred_norm)

)

ggplot(df_xgb, aes(x=Index)) +

  geom_line(aes(y=Actual_Normalized, color='Actual Normalized')) +

  geom_line(aes(y=XGBoost, color='XGBoost')) +

  scale_color_manual(values=c('blue','red')) +

  ggtitle("Actual vs Predicted values (XGBoost)")
```

6.1.4 APPENDIX P: Model Comparison Plot

```
compare_df <- data.frame(

  Index = 1:length(y_test_norm),

  Actual_Normalized = as.numeric(y_test_norm),

  XGBoost = as.numeric(xgb_pred_norm),

  SARIMA_Model = as.numeric(pred_norm)[1:length(y_test_norm)]

)

ggplot(compare_df, aes(x=Index)) +

  geom_line(aes(y=Actual_Normalized, color='Actual Normalized')) +

  geom_line(aes(y=XGBoost, color='XGBoost')) +

  geom_line(aes(y=SARIMA_Model, color='SARIMA Model')) +

  scale_color_manual(values=c('blue','red','green')) +

  ggtitle("Comparison Prediction Plot (Actual vs XGBoost vs SARIMA)")
```

6.1.5 APPENDIX Q: Extended Model Evaluation

```
# Short-term vs Long-term MAPE

short_horizon <- 3

long_horizon <- 12

sarima_short_term <- mape(test$Turnover[1:short_horizon],
as.numeric(sarima_forecast$mean)[1:short_horizon]) * 100

sarima_long_term <- mape(test$Turnover[1:long_horizon],
as.numeric(sarima_forecast$mean)[1:long_horizon]) * 100
```

```
xgb_short_term <- mape(y_test[1:short_horizon], xgb_pred[1:short_horizon]) * 100

xgb_long_term <- mape(y_test[1:long_horizon], xgb_pred[1:long_horizon]) * 100

cat("Short-term (3 months) MAPE - SARIMA:", round(sarima_short_term,2), "% | XGBoost:", round(xgb_short_term,2), "%\n")

cat("Long-term (12 months) MAPE - SARIMA:", round(sarima_long_term,2), "% | XGBoost:", round(xgb_long_term,2), "%\n")

# In-sample (training) fit

sarima_train_fitted <- fitted(fit)

sarima_train_actual <- as.numeric(train$Turnover)

sarima_train_mape <- mape(sarima_train_actual, sarima_train_fitted) * 100

xgb_train_pred <- predict(xgb_model, as.matrix(train_xgb[, c("Turnover_lag1", "rolling_mean", "rolling_sd", "month", "quarter")]))

xgb_train_actual <- y_train

xgb_train_mape <- mape(xgb_train_actual, xgb_train_pred) * 100

cat("In-sample (Training) MAPE - SARIMA:", round(sarima_train_mape,2), "% | XGBoost:", round(xgb_train_mape,2), "%\n")

# Structural breaks (visual)

residuals_sarima <- residuals(fit)

residuals_xgb <- y_test - xgb_pred

par(mfrow=c(2,1))

plot(residuals_sarima, type='l', main='SARIMA Residuals (Check for Structural Breaks)')

abline(v=which(abs(diff(residuals_sarima)) > 2*sd(residuals_sarima)), col='red', lty=2)
```

```r
plot(residuals_xgb, type='l', main='XGBoost Residuals (Check for Structural Breaks)')

abline(v=which(abs(diff(residuals_xgb)) > 2*sd(residuals_xgb)), col='red', lty=2)

par(mfrow=c(1,1))

# Probabilistic forecasts

plot(sarima_forecast, main="SARIMA Probabilistic Forecast (80% & 95% Intervals)")

hist(residuals_xgb, breaks=20, col="orange", xlab="Residuals", main="XGBoost Residual Distribution (Proxy for Uncertainty)")

# Summary Table

summary_table <- data.frame(

  Model = c("SARIMA", "XGBoost"),

  Test_MAPE = c(round(mape_sarima_original, 2), round(mape_xgb_original, 2)),

  Train_MAPE = c(round(sarima_train_mape, 2), round(xgb_train_mape, 2)),

  ShortTerm_MAPE = c(round(sarima_short_term, 2), round(xgb_short_term, 2)),

  LongTerm_MAPE = c(round(sarima_long_term, 2), round(xgb_long_term, 2))

)

print(summary_table)
```

6.2. APPENDIX R PROJECT DATA

Date	Turnover
1/1/2020	175,233.20
2/1/2020	159,874.50
3/1/2020	175,747.94
4/1/2020	258,389.43

5/1/2020	264,274.60
6/1/2020	254,816.69
7/1/2020	284,993.11
8/1/2020	307,028.28
9/1/2020	291,553.13
10/1/2020	297,834.72
11/1/2020	288,566.18
12/1/2020	406,452.30
1/1/2021	308,729.89
2/1/2021	287,528.63
3/1/2021	323,239.98
4/1/2021	361,427.66
5/1/2021	349,855.49
6/1/2021	351,272.38
7/1/2021	397,944.78
8/1/2021	431,296.40
9/1/2021	406,601.38
10/1/2021	425,569.04
11/1/2021	432,639.14
12/1/2021	601,987.18
1/1/2022	184,167.70
2/1/2022	223,783.09
3/1/2022	273,705.08
4/1/2022	275,566.19
5/1/2022	239,931.07
6/1/2022	216,576.72
7/1/2022	177,693.20
8/1/2022	200,195.89
9/1/2022	207,926.23
10/1/2022	195,550.94
11/1/2022	227,581.75
12/1/2022	379,608.19

1/1/2023	402,893.71
2/1/2023	273,295.67
3/1/2023	191,526.96
4/1/2023	59,095.35
5/1/2023	98,418.83
6/1/2023	106,110.28
7/1/2023	190,196.28
8/1/2023	198,978.54
9/1/2023	296,033.54
10/1/2023	369,270.77
11/1/2023	370,140.40
12/1/2023	472,787.80
1/1/2024	186,121.66
2/1/2024	208,019.38
3/1/2024	342,470.22
4/1/2024	349,927.88
5/1/2024	387,144.33
6/1/2024	354,989.10
7/1/2024	257,040.07
8/1/2024	284,683.19
9/1/2024	372,724.69
10/1/2024	414,007.75
11/1/2024	433,772.52
12/1/2024	541,557.72

6.1 APPENDIX S: LATTER FROM SIMBISA

Simbisa Brands LIMITED

APPROVAL LETTER FOR DATA COLLECTION

Grace Hove

House Number 119 Avondale

Harare

Dear Ms. Hove

I am pleased to notify you that your application to conduct data collection for your research study at Bakers Inn has been approved. This letter is to confirm your clearance to conduct Turnovers data collection from simbisa brands steers

Ensure that the information gathered is processed accordingly and as mandated by the current law, ethical standards, and privacy policy. In case revisions are to be published, obtain permission in advance from the concerned authority.

In the event that you need assistance or you need clarification, you are free to mail me at [Thomas.Kusosa@simbisa.co.zw]. I will be extremely glad to help.

Thank you for the professionalism and courtesy with which you submitted the request. We consider this clearance to be of service to your work, and we eagerly await your findings.

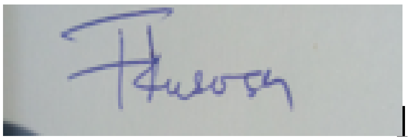
161 Five Avenue

Simbisa Brands

Harare, Zimbabwe

Tel: (2634) 799609-10, 755305-6, Cell number: 0777570277

Fax: (263) 41 796258



6.6. APPENDIX T; SIMILARITY REPORT

My project

ORIGINALITY REPORT

8%	6%	4%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to University of South Wales - Pontypridd and Cardiff Student Paper	<1%
2	fastercapital.com Internet Source	<1%
3	Submitted to Monash University Student Paper	<1%
4	Submitted to University of Reading Student Paper	<1%
5	www.mdpi.com Internet Source	<1%
6	Submitted to Vrije Universiteit Amsterdam Student Paper	<1%
7	www.coursehero.com Internet Source	<1%
8	Submitted to Erasmus University of Rotterdam Student Paper	<1%
	Submitted to University of East London	