BINDURA UNIVERSITY OF SCIENCE EDUCATION

FACULTY OF SCIENCE AND ENGINEERING

DEPARTMENT OF STATISTICS AND MATHEMATICS

FORECASTING SALES DEMAND OF A MANUFACTURING COMPANY. A TIME SERIES APPROACH

BY

TANYARADZWA D DZVETERA

B202908B

A DISSERTATION SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR BSC.HONOURS IN STATISTICS AND FINANCIAL MATHEMATICS

SUPERVISOR: DR MAPUWEI

2024

APPROVAL FORM

The undersigned certify that they have read and recommended to the Bindura University of Science Education for acceptance of a dissertation entitled "Forecasting sales demand a time series analysis approach, submitted by B202908B in partial fulfillment of the requirements of the Bachelor of Science (Honours) Degree in Statistics and Financial Mathematics.

STUDENT

Name DZVETERA TANYARADZWA D

Signature: Date: June 10, 2024

DR. T. W. MAPUWEI ………………………….. ………………………

Supervisor Signature Date

DR. M. MAGODORA ………………………….. ………………………

Chairperson Signature Date

DECLARATION

I Dzvetera Tanyaradzwa D now declare that this submission is my work apart from the references to other people's work which has duly been acknowledged. I at this moment declare that this work has neither been presented in whole nor in part for any degree at this university or elsewhere.

Author: Dzvetera Tanyaradzwa D

Registration Number: B202908B

Signature:

Date: 10 June 2024

DEDICATION

This research project is dedicated to my beloved family who served as inspiration and encouragement during my study period.

ACKNOWLEDGEMENTS

Preparing a research project requires effort from several key individuals and stakeholders. Firstly, my gratitude goes to my supervisor, Dr. T. W Mapuwei who devoted a lot of his time and patience to this study. I am grateful for the constructive advice, selfless guidance, and concern he accorded me. Secondly, l wishes to convey my sincere gratitude to the Bindura University of Science Education for their invaluable input for this work. Thirdly, my heartfelt gratitude goes to my family and friends for the understanding moral support, and sacrifice throughout my engagement to this work. Lastly of special mention are my beloved parents Mr. V and Mrs. S Dzvetera. May the Lord almighty bless them.

ABSTRACT

Forecasting sales is crucial for manufacturing firms as it enables the identification of future sales patterns for products. This predictive analysis aids in optimizing retail operations to meet customer demand effectively, anticipate potential losses, and manage inventory levels efficiently. The data was collected from the manufacturing company Corked Spin Investment every month from the year 2017 up to 2023. The study investigates the trend analysis, sustainability, and effectiveness of the time series model in forecasting sales demand, it also compares multilayer Perceptron and SARIMA to show which method is the best in forecasting and lastly, it forecasts sales demand for the year 2024 on monthly bases. This research used a quantitative research design. A literature review is performed to identify a suitable model for forecasting sales demand for corked spin investment. Multilayer Perceptron (MLP) and Seasonal Autoregressive Integrated Moving Average (SARIMA) sales demand forecasting techniques were employed, and the outcomes revealed that Multilayer Perception has the best performance than SARIMA in forecasting Corked Spin Investment sales demand. This was done using R programming. From this research, it is concluded that Multilayer Perception can model nonlinear function compared to SARIMA. Hence MLP is chosen as the ideal model for predicting sales of Corked Spin Investment. The trend of MLP is increasing showing a better performance in the year 2024. It is advisable to use MLP to forecast sales of the company. Further studies should enhance the MLP model predictive power by incorporating additional variables such as macroeconomic indicators and competitor activity.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

SARIMA	Seasonal Autoregressive Integrated Moving Average
MLP	Multilayer Perception
ACF	Autocorrelation Function
AR	Autoregressive models
ARIMA	Integrated Autoregressive Moving Average Models
MA	Moving Averages
PACF	Partial Autocorrelation Function

CHAPTER 1

1.0 Introduction

As the saying goes, 'The past is often the best predictor of the future.' This notion holds in sales forecasting, capturing and comprehending the underlying patterns within data. Forecasting can make data-driven decisions and enhance sales performance. Forecasting sales accurately requires more than just intuition and experience; it necessitates a robust analytical approach. Time series analysis offers a comprehensive framework to model and predict sales patterns over time, enabling businesses to allocate resources effectively and capitalize on emerging opportunities.

This chapter provides an overview of the study, including the background, problem statement, study objectives, research questions, significance, assumptions, scope, and limitations. Additionally, it outlines the methodology employed for the study and concludes with a summary. After this introductory chapter, there are four more additional chapters to make them five. The second chapter examines the historical research literature by reviewing time series analysis and how it is useful in forecasting sales demand in Corked spin investment manufacturing companies. The third chapter consists of the research methodology. The main focus is on the methodology employed to achieve the goals of the study. Chapter four main focus is on the data presentation and analysis, discussion of the layout framework for the data process, and presentation of research results. Finally, the final chapter concentrates on drawing conclusions from the research and giving recommendations.

1.2 Background of the study

In the context of manufacturing companies, accurate sales demand forecasting plays a pivotal role in efficient resource allocation and operational planning (Keleperas, 2023). Several studies have highlighted the importance of accurate sales demand forecasting in the manufacturing industry. Since many businesses depend on demand forecasting to plan a company's long-term and make sure their supply chains are operating efficiently daily, forecasting is the best choice. The necessity for firms to integrate their internal and external chains of supply is largely dependent on their capacity to effectively foresee future demand. (Chong, 2014). For instance, Gonçalves (2021) conducted research on sales demand forecasting in the automotive industry and found that accurate forecasts significantly improved production planning, inventory management, and overall supply chain performance. Similarly, Preis, (2021) explored the impact of sales demand forecasting on the semiconductor manufacturing industry, emphasizing the need for advanced forecasting techniques to handle the industry's high volatility and demand uncertainty. Manufacturing companies can improve customer satisfaction by delivering products on time, streamlining production schedules, and successfully managing inventory levels when they can estimate sales demand (Småros, 2003)

Managers across all levels and industries have long been concerned with enhancing the effectiveness of demand forecasting (Kerkkänen, 2010). For manufacturing organizations to choose the best course of action for production planning, sales budgeting, transportation modes, and delivery planning, they need reliable forecasts for a variety of decision-making processes. Scientists and experts have focused in particular on how forecasts might be enhanced to boost accuracy to expedite manufacturing processes. (Caniato et al., 2002; Armstrong, 2001). Studies have indicated that enhanced forecasting methods are very helpful for different organizational divisions, such as planning and decision-making procedures (Winklehofer, et al., 1996)

The complexity and variability of sales demand patterns in the manufacturing sector present challenges for traditional forecasting methods (Brown et al., 2019). Therefore, the adoption of advanced time series analysis techniques has emerged as a critical approach for improving the accuracy of sales demand forecasts in manufacturing companies (Lalou,2020). Time series analysis provides the capability to identify and model underlying patterns and trends within historical sales data, thereby enabling more reliable predictions of future sales demand (Kilimci,2019). Hence, the objective of this project is to utilize different time series forecasting techniques to predict product demand and evaluate the demand forecasting approaches utilized by manufacturing enterprises, specifically focusing on Corked Spin Investments. The best approach for estimating the company's part needs is then chosen after a variety of approaches have been evaluated.

1.3 Statement of problem

Despite the critical importance of accurate sales demand forecasting in the manufacturing industry, many companies face significant challenges in effectively predicting future demand according to

Chase (2013). These challenges arise from factors such as seasonality, product lifecycle, market volatility, and changing consumer preferences. As a result, manufacturing companies often struggle with stockouts, excess inventory, inefficient resource allocation, and suboptimal production planning. Consequently, it's imperative to create resilient and dependable forecasting methodologies and models customized to handle the intricacies of sales demand forecasting within the manufacturing sector. The goal of this dissertation is to investigate and suggest fresh strategies that raise the precision and efficacy of sales demand forecasting, helping manufacturing firms increase customer happiness, cut expenses, and streamline operations.

1.4 Research objectives

1. To Identify patterns and trends in historical sales data to inform future sales projections.

2. to assess the appropriateness and efficiency of different time series models and techniques in analyzing and predicting sales demand within the manufacturing sector.

3. To compare the performance of times series models

4. To forecast sales for Corked spin investments for twelve months in 2024 using the best model.

1.5 Research questions

1. What trends and patterns can be identified in historical sales data, and how can this information be used to inform the forecasting of future sales demand in a manufacturing company?

2. Which statistical models and techniques are most suitable for analyzing and forecasting sales demand in the manufacturing industry?

3. Which is the best model for forecasting sales demand in a manufacturing company?

4. Can sales demand be forecasted over the study period?

1.6 Research hypothesis

Null Hypothesis: There is no meaningful correlation between past sales data and future sales demand within the manufacturing sector.

Alternative Hypothesis: Past sales data significantly influences and can be used to predict future sales demand in the manufacturing industry

1.7 Scope of the study

Forecasting sales demand on a manufacturing company. The study would involve collecting historical sales data, preferably over five years, to capture seasonal patterns, trends, and other relevant patterns in the data. The data is collected from Corked spin investment PVT manufacturing. The preprocessing of data is conducted. To guarantee the quality and suitability of the gathered data for time series analysis, this step includes cleaning and changing it. The purpose of exploratory data analysis is to obtain an understanding of the data, spot trends, and potentially uncover underlying linkages. This step may involve visualizations, statistical summaries, and other exploratory techniques.

Development and evaluation of an appropriate time series forecasting model is done. This involves exploring various models such as moving averages, autoregressive integrated moving averages (ARIMA), seasonal ARIMA (SARIMA), or more advanced techniques like state space models, or machine learning algorithms such as multilayer perception (MLP). It involves evaluating the forecasting models' dependability and accuracy using metrics such as forecast bias analysis, Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE).

Once trained and evaluated, the model can anticipate future sales demand. Assessing the accuracy and reliability of the model involves comparing actual sales data with the predicted figures. Forecasting models can be refined by adjusting model parameters and exploring different model configurations. The findings, methodology, and results of the study are documented and reported clearly and concisely. This also includes a description of the data used, the chosen model, the forecasting accuracy, and any recommendations or insights derived from the analysis.

1.8 Significant of the study

Businesses may make well-informed decisions about inventory management, production levels, resource allocation, and overall business strategies with the aid of accurate sales forecasting (Davis 2007). This study provides organizations with a robust methodology to forecast future sales with greater precision. Sales forecasts are instrumental in financial planning and budgeting processes according to Coveney (2017). By accurately predicting future sales, companies can better allocate financial resources, plan investments, and set achievable revenue targets. An in-depth analysis of sales trends allows businesses to optimize their resource allocation. By identifying periods of high or low sales, companies can adjust their staffing levels, reduce waste, and minimize costs. Understanding sales patterns and forecasting future demand enables businesses to tailor their marketing strategies accordingly. Identifying peak seasons or periods of lower demand, companies can design targeted marketing campaigns to maximize sales and customer engagement. Businesses can gain a competitive edge by accurately projecting sales since it enables them to predict market trends, respond to changing customer demands, and adapt their business strategies accordingly. It also helps companies stay ahead of the competition and make proactive decisions in a dynamic marketplace.

The importance of this study resides in its capacity to enable firms to improve financial planning, allocate resources optimally, make data-driven decisions, and obtain a competitive advantage in the marketplace. Over time, it helps to increase profitability and corporate performance.

1.9 Assumptions of the study

1. Stationarity: the presumption that the data's statistical characteristics won't change over time. This indicates that the data's mean, variance, and covariance are all constant across time.

2. Independence: It assumes that data points do not depend on one another, i.e., the value of one data point does not depend on the value of a data point that comes before or after it.

3. Linearity: The presumption that a linear model can adequately capture the connection between the variables in the time series.

4. Normality: The presumption that there is a normal distribution for the residuals or the variations between the values that were observed and those that were expected.**1.10 Limitations of the Study**

1. Assumption of Stationarity: The underlying data must demonstrate stationarity, or that its statistical characteristics hold over time, for time series analysis to be performed. Real-world sales data, however, could show seasonality, trends, or other non-stationary patterns, which could make the study

Other Factors: A wide range of other factors, including market competition, consumer preferences, economic conditions, and unanticipated occurrences, can have an impact on sales. A time series analysis might not fully account for all the complex dynamics at play, even though it can capture some of these elements.

3. Accuracy of Forecasts: While time series analysis can provide insightful forecasts, it is important to acknowledge that they are inherently probabilistic. Forecast accuracy may vary depending on the quality and completeness of the historical data, the suitability of the chosen model, and the complexity of the sales patterns.

4. Limited Predictive Power: It is crucial to recognize that sales forecasting using time series analysis does not guarantee absolute predictive accuracy. While it can provide useful estimates, there is always a degree of uncertainty associated with any forecast, and unexpected deviations from the projected sales can occur.

5. Data Availability and Quality: The forecasts' accuracy and dependability may be impacted by the availability and quality of previous sales data. Predictions that are biased or erroneous can result from incomplete or untrustworthy data sets.

1.11 Definition of terms

Demand forecasting: In the context of the supply chain process, demand forecasting primarily aims to forecast sales and product demand to facilitate the acquisition, stocking, or production of resources in the appropriate amounts to support the company's value-adding activities.
Time series: This is the series of numerical values obtained from consecutive periods usually with equal intervals between them. (Ruey, 2010)

1.12 Conclusion

This chapter introduces the study's background, problem statement, research questions, assumptions, significance, and limitations. It also outlines the objective of developing an efficient forecasting method through time series analysis. The subsequent chapter will delve into the study's methodology, including an examination of research gaps in theoretical literature.

CHAPTER 2: LITERATURE REVIEW

2.0 Introduction

The ability to forecast sales accurately is a cornerstone of robust business planning, particularly within the manufacturing sector (Vereecke, 2018). It shapes production schedules, informs inventory control, and has profound implications for financial strategies. This chapter is dedicated to dissecting time series analysis as a methodology for forecasting sales demand, an approach that decodes patterns in historical data to predict future outcomes. We begin by examining the theoretical underpinnings of time series analysis, tracing its evolution from simple historical averages to complex econometric models. The literature review explores various time series methods from classical approaches such as ARIMA (Autoregressive Integrated Moving Average), to cutting-edge techniques like machine learning algorithms that cater to non-linear and multifactorial analysis. The strengths and limitations of each method are dissected concerning their applicability to manufacturing contexts. Subsequently, the focus shifts to practical applications. Real-world cases where time series forecasting has been effectively utilized within manufacturing environments are highlighted to contextualize abstract mathematical concepts. The impact of forecasting accuracy on operational efficiency, customer satisfaction, and competitive advantage is discussed, providing an industry-specific backdrop against which the subsequent research will unfold.

2.1 Theoretical literature

2.1.1 Moving Average (MA) Method

To predict future sales demand, the moving average method includes averaging a predetermined number of previous periods (Khosroshahi, 2016). The moving average can be computed using the following formula: Sum of Sales Demand for Previous 'n' periods / 'n' is the formula for the Moving Average (MA). Choose the right number of periods (n) to take into account while calculating the moving average. The features of your data and the required degree of smoothing or responsiveness to variations in sales demand will determine this value. The moving average figure is utilized as the sales demand estimate for the following period after it has been calculated. Every time a new period comes up, this procedure is repeated, updating the moving average continuously.

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \qquad (1)$$

ε_t is a white noise.

2.1.2 Autoregressive (AR) Method

According to Smith (1993), the autoregressive model suggests that the time series' value at any given moment is linearly dependent on its prior values. The number of prior time steps taken into consideration is indicated by the autoregression order, or p. figuring out which order is best for your sales demand projection. Estimating the autoregressive model's coefficients is the next stage. Typically, methods such as maximum likelihood estimation or ordinary least squares regression are employed for this purpose. Once the coefficients are estimated, the autoregressive equation can be constructed. For example, an AR(1) model can be represented as:

$$y_t = c + \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \cdots + \emptyset_p y_{t-p} + \varepsilon_t \qquad (2)$$

Where:

- y_{it} represents the sales demand at time t.
- c is a constant term.
- $\emptyset_1$ is the coefficient associated with y_{t-1}
- ε_t is the random error term.

2.1.3 Autoregressive Integrated Moving Average (ARIMA) Method

Three parameters (p, d, and q) form the basis of ARIMA models: - p: The model's total number of autoregressive terms (AR). It speaks of the dependent variable's lagged values. - d: The model's degree of differencing (I). To attain stationarity, the data must be differenced a certain amount of times. Moving average terms (MA) in the model, denoted by q. It alludes to the model's error terms. The formula for the ARIMA model is as follows: = I(d) + MA(q) + AR(p) The time series variable at time t is represented by the variable AR(p), the autoregressive term of order p, the differencing term of order d by I(d), and the moving average term of order q by MA(q). The autoregressive component in ARIMA models

y_t= AR(p) + I(d) + MA(q)

2.1.4 Seasonal Autoregressive Integrated Moving Average (SARIMA) Method

The SARIMA model consists of three main components: Autoregression (AR), Differencing (I), and Moving Average (MA), denoted by parameters p, d, and q respectively. The seasonal component is represented by variables P, D, and Q, indicating seasonal differences. Specifically: d denotes the number of differencing operations required for data stationarity, p signifies the number of autoregressive terms, D indicates the seasonal differencing count for data stationarity, P represents the seasonal autoregressive terms, q denotes the number of moving average terms, and Q signifies the seasonal moving average terms. The notation SARIMA (p, d, q) (P, D, Q) s refers to the SARIMA model, where s denotes the frequency or number of time points per season.

$$(1-\emptyset_1\beta)(1-\emptyset\beta^4)(1-\beta)(1-\beta^4)y_t=(1-\theta_1\beta)(1+\theta_1\beta^4)e_t \qquad (3)$$

where,

- y_t is the time series data at time t,

- β is a backshift operator,

- α and φ are the coefficients for the Autoregressive (AR) terms,

- θ are the coefficients for the Moving Average (MA) terms,

- e_t is the error term at time t, and

The forecasting steps for SARIMA are similar to those of ARIMA models, with the additional step of accounting for the seasonal component in the data.

2.1.5 Mean Absolute Error (MAE) Method

Divide your historical sales data into two sets: the training set and the testing set. The training set will be used to build the model, while the testing set will be used to evaluate the accuracy of the model's forecasts. Use a time series forecasting method, such as exponential smoothing or ARIMA, to generate forecasts for the sales demand in the testing set. Calculate the absolute difference between the forecasted values and the actual values of sales demand in the testing set. This can be done for each observation in the testing set. Sum up all the absolute errors calculated in step 3 and divide it by the number of observations in the testing set. This will give you the mean absolute error (MAE) of your forecasting model.

The formula for calculating MAE:

$$\text{MAE} = \frac{1}{N}\sum_{j=1}^{n} e_j \qquad (4)$$

2.1.6 Artificial neural network

Artificial neural networks (ANNs) are a subset of machine learning algorithms that mimic the structure and functionality of neurons in the human brain. They consist of interconnected nodes, or artificial neurons, organized into layers. Several types of neural networks exist, including Extreme Learning Machines (ELM), Multilayer Perceptron, Recurrent Neural Networks, and Long Short-Term Memory. ANNs can learn from data, identify patterns, and extrapolate this knowledge for future applications.

1. Input Layer: This is where the data enters the network. Each neuron corresponds to a single input variable.

2. Hidden Layer(s): The hidden layers process the data by performing computations that combine and transform the input data. Each neuron in the hidden layer corresponds to one or more computational variables, which are derived from the inputs.

3. Output Layer: The output layer produces the final output of the network based on the computations and transformations performed by the hidden layers.

ANN's performance is governed by several hyperparameters like the number of neurons, hidden layers, activation functions, and optimization algorithms. Thus, identifying the right combination of these hyperparameters is critical to obtaining high-quality forecasts.

2.1.7 Long Short Memory

Recurrent neural networks of the Long Short-Term Memory (LSTM) variety are intended to solve the vanishing gradient issue that conventional RNNs have. To do this, LSTMs incorporate a memory cell that can gradually recall or forget certain information. An LSTM is made up of several memory cells, each of which keeps track of a hidden state vector, h_t, which stands for its current memory. Every memory cell gets input from the one before it and produces an output that is sent to that cell's neighbor. Multiple gates that regulate the information flow into and out of the memory cell make up the cell. These gates are as follows:

1. Forget Gate: chooses which data from the prior cell state to disregard

2. Input Gate - determines which new information to store in the current cell state

3. Cell Gate - calculates the new candidate values for the cell state

4. Output Gate - determines which information to output to the current hidden state.

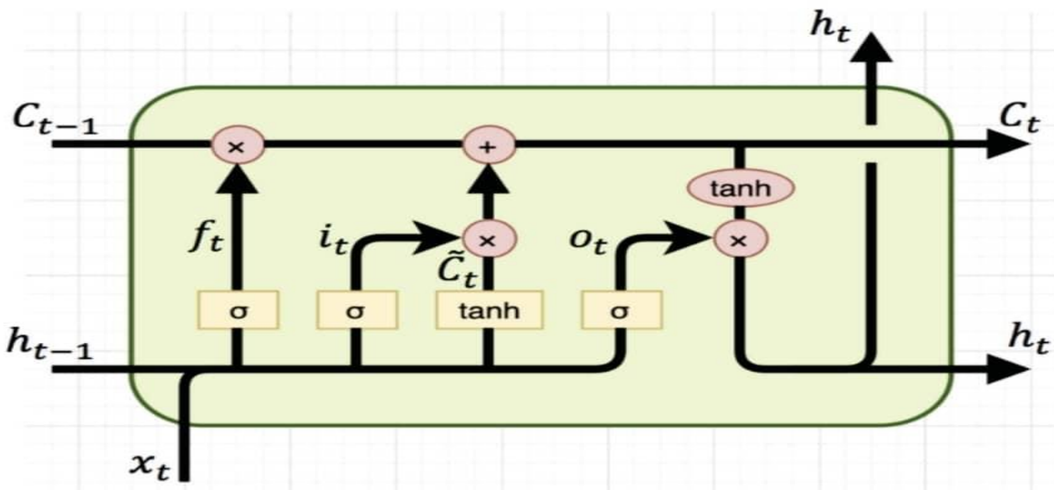

Figure 1: Long short memory

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + bi) \quad (5)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + bf \quad (6)$$

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + bc) \quad (7)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + bo) \quad (8)$$

$$h_t = o_t \tanh(c_t) \quad (9)$$

x_t is the input vector

o_t activation vector output gate

h_t is the output vector

i_t is the activation vector of the input gate

c_t is the cell state vector

σ is the sigmoid function

f_t is the activation vector of the forget vector

W_{xi} indicates the parameter between the input vector and the input gate

W_{xc} indicates the weight parameter between the cell and the input vector

W indicates the weight parameter between the input and forgotten gate

W_{xo} indicates the weight parameter between the input vector and output vector

2.1.8 Recurrent neural network

Recurrent Neural Networks (RNNs) are a class of artificial neural networks that are designed to deal with sequential data. Unlike traditional feedforward neural networks, which assume that each input is independent of all previous inputs, RNNs maintain a state or "memory" that can capture information from previous inputs and use it to make predictions. This makes RNNs well-suited for a variety of sequential data tasks, including forecasting. An RNN consists of a series of neural network units linked together in a chain-like structure. These units are called cells and are organized in one or more layers. Each cell has a state variable or memory, which captures and remembers information about previous inputs. The architecture of an RNN is designed to process sequential data one element at a time and update its internal state at each step of the sequence. The output of each cell is a function of the current input, the previous output of the cell, and its memory state.

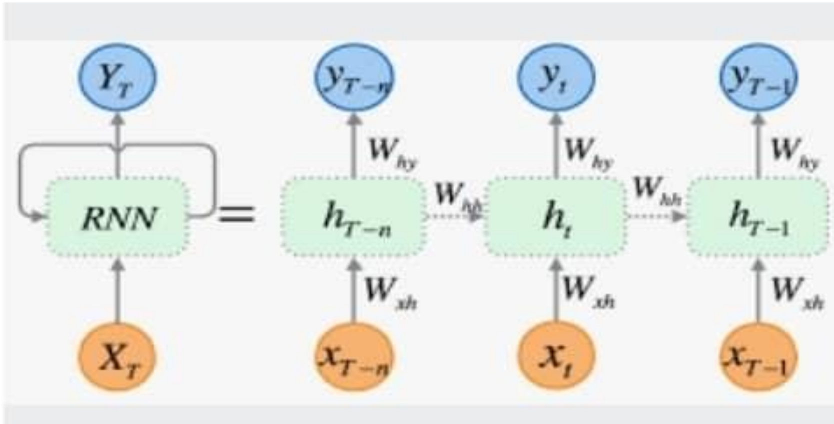

Figure 2: RNN

The mathematical formula that governs the computation in an RNN cell is as follows:

$$h^{(t)} = g_h\left(W_{xh}x^{(t)} + W_{hh}h_{T-n}\right) \qquad (10)$$

$$y^{(t)} = g_y\left(W_{hy}h^{(t)}\right) \qquad (11)$$

x is the input of the network

y output of the network

$g_y$activation function of the output layer

h hidden layer

g_h activation of hidden error

W_{hh} connection weight between hidden error

W_{xh} connection weight between the input layer and hidden layer

W_{hy} connection weight between the hidden layer and the output layer

t the time steps in the network

2.1.9 Multilayer perceptron

A particular type of feedforward neural network known as a multilayer perceptron (MLP) models intricate relationships between inputs and outputs through artificial neurons featuring nonlinear activation functions. These networks are proficient in various supervised learning tasks, including prediction. Comprising an input layer, one or more hidden layers, and an output layer, an MLP neural network assigns each neuron in the input layer to an input variable and each neuron in the output layer to an output variable. Through the hidden layers, input data undergoes combination and transformation before the output layer generates the network's final output. MLP-equipped neural networks serve as vital tools in sales demand forecasting due to their ability to accurately capture nonlinear interactions between input and output data. Optimization through fine-tuning hyperparameters like neuron count, hidden layers, activation functions, and optimization algorithms can significantly enhance their performance. Selecting input features with care is crucial for achieving highly accurate sales demand estimation using MLP neural networks.

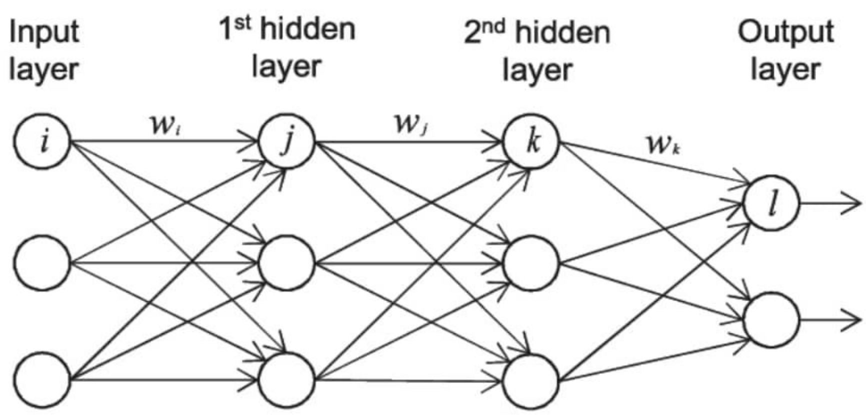

Figure 3: MLP

$$y = f(x_1\omega_1 + x_2\omega_2 + \cdots + x_n\omega_n \qquad (12)$$

$X\ and\ y$ input and output of the network

W_i connection weight between two input and hidden layer

W_J connection weight between the first and second hidden layer

W_K connection weight between the hidden layer and output layer

f activation function

i, j, k, l nodes in the network

2.2 Empirical literature

The usefulness of time series models and approaches in predicting sales demand in the manufacturing industry has been the focus of extensive research efforts. In a 2020 study, Adeyinka compared the effectiveness of modern methods like ARIMA and neural networks with more conventional techniques like exponential smoothing and moving averages. According to the study, ARIMA and neural networks performed better in terms of forecast error and accuracy than conventional techniques. This suggests that more intricate models are more appropriate for predicting the demand for products in the manufacturing sector. Sankaran (2014) conducted a second study that concentrated on the application of seasonal autoregressive integrated moving average (SARIMA) models for sales demand forecasting in the Indian automobile industry. According to the study, SARIMA models work well for forecasting in the manufacturing sector because they can accurately represent seasonal fluctuations and trends in sales demand. Similar outcomes were also noted in Zhao's (2022) study, which assessed how well SARIMA models predicted sales demand for a company that manufactures chemicals. For sales demand forecasting, machine learning techniques have become more and more common in addition to traditional time series models. In a study published in 2020, Güven examined the effectiveness of multiple machine learning algorithms, such as support vector machines, random forests, and artificial neural networks, in forecasting the sales demand of a manufacturing company. According to the study, machine learning models produced forecasts that were more accurate than those made by conventional time series models. In addition to examining specific models, some research has also compared the accuracy of various forecasting techniques. In a 2016 study, Uzzaman evaluated how well fuzzy time series, neural networks, and ARIMA predicted sales demand for a cement manufacturing company. All three of the methods were shown to be appropriate for demand forecasting; however, fuzzy time series had the lowest forecast error, making it the most practical approach for the industry.

Time series analysis has been widely used in the manufacturing industry to forecast sales demand. One such study was conducted by Firoz, (2014) to forecast sales demand for a steel manufacturing company in India. The authors used the Autoregressive Integrated Moving Average (ARIMA) model to analyze the monthly sales data from 2009 to 2017. The study found a significant upward trend in sales demand, with a seasonal component that was mainly influenced by the monsoon season. The authors concluded that using time series analysis can help the company make accurate demand forecasts and optimize their production and inventory levels. Another study by Guðmundsson, (2022) used time series analysis to forecast sales demand for a manufacturing company in Iceland. The authors collected monthly sales data for the past five years and used the Holt-Winters forecasting method to predict future demand. The study found a significant seasonal pattern in sales demand, with a peak during the summer months. The authors suggested that the company should adjust their production levels accordingly to meet the seasonal demand.

Seasonality, cyclical patterns, and trend analysis are the essential elements of time series analysis (Dagum, 2016). These elements are necessary to comprehend the fundamental causes influencing sales demand and to produce precise projections. The examination of long-term changes in data across time is known as trend analysis. It assists in determining the overall trend of the data, including whether it is rising, falling, or staying constant. Trend analysis can shed light on the general growth or fall of a company's sales in the context of sales demand forecasting. For instance, a business is likely to stay on the same course going forward if its sales have been rising consistently during the last several years, and vice versa. Recurring patterns in data that show up at regular intervals are referred to as seasonality. This information is crucial for companies to plan their production and inventory levels accordingly. A manufacturing firm that makes winter jackets will likely see increased sales in the colder months and decreased sales in the warmer ones.

Cyclical patterns refer to the fluctuations in data that occur over a longer period, typically more than a year. These patterns are influenced by economic conditions, market trends, and other external factors. In sales demand forecasting, identifying cyclical patterns can help companies prepare for periods of high and low demand and adjust their strategies accordingly. According to Yang's (2022) study, utilizing time series analysis to anticipate sales demand for a Taiwanese manufacturing company produced much more accurate forecasts when trend, seasonality, and cyclical patterns were included in the model. Researchers utilized various time series methods by analyzing the company's past sales data., such as autoregressive integrated moving average (ARIMA), moving average, and exponential smoothing. The findings demonstrated that the ARIMA model which takes into account all three factors performed better at estimating sales demand than the other techniques. Additionally, Chen's (2019) study on the use of artificial intelligence and time series analysis to estimate sales demand for an automotive manufacturing company discovered that adding trend, seasonality, and cyclical patterns to their model produced more accuracy. To establish a hybrid model, the study combined historical sales data with outside variables including market trends and economic indicators. The hybrid model beat other artificial intelligence techniques as well as conventional time series methods, according to the results, underscoring the need to take into account all time series analysis components when projecting sales demand.

2.3 Research gap

The scarcity of studies comparing the performance of various time series models leaves practitioners without clear guidance on which methods are most effective. This gap implies a need for comprehensive comparative analyses to identify the strengths and weaknesses of different approaches, aiding manufacturers in selecting the most suitable forecasting method for their specific needs. Many existing studies concentrate on particular industries or products, which restricts the generalizability of their findings. Since manufacturing companies operate in diverse contexts with unique demand patterns, there's a need for research that transcends industry boundaries to provide insights applicable across different manufacturing sectors. While large datasets have become increasingly available, their potential for improving forecasting accuracy in sales demand remains largely untapped. Research exploring the integration of Big Data analytics techniques, such as machine learning algorithms, into time series forecasting models, could offer significant advancements in prediction accuracy for manufacturing companies. Ensemble methods, which combine multiple forecasting models to enhance accuracy, are not extensively explored in sales demand forecasting research within the manufacturing domain. Investigating the effectiveness of ensemble methods, such as combining statistical models with machine learning algorithms, could lead to improved forecasting performance and robustness. Forecasting uncertainty, inherent in any prediction, is crucial for decision-making in manufacturing operations. However, few studies quantify uncertainty in sales demand forecasting, leaving decision-makers unaware of the reliability and confidence levels associated with forecasted values. Research focusing on methodologies to quantify and communicate forecast uncertainty could enhance the utility of sales demand forecasts for manufacturing companies.

2.4 Proposed conceptual model

The dependent variable is the amount of goods that the manufacturing company sells in a given amount of time. the sales data is arranged historically and usually measured at regular intervals (daily, weekly, monthly). the manufacturing company's historical sales figures for a given time frame. Time series analysis will be used for this. A statistical technique for examining and predicting trends in sequential data is time series analysis. It assumes that a variable's future values rely on its historical values. To create a time series forecasting model on the autoregressive, moving average, and differencing components, to name a few, will be incorporated into the proposed model. Other time series techniques that will be included in the model are trend analysis, which identifies and models the long-term upward or downward movement in sales demand, seasonality analysis, which captures and accounts for recurring patterns or seasonal fluctuations in sales demand, and Autoregressive Integrated Moving Average (ARIMA) modeling. Internal records from the manufacturing company will be consulted to gather historical sales data. The proper statistical software will be used to clean, preprocess, and analyze the gathered data. A variety of time series models will be put to the test and assessed according to how well they predict, fit, and are accurate. According to Hodson (2022) a variety of statistical indicators would be used to assess the suggested forecasting model against evaluate the model's efficacy, it will be compared against baseline models or other forecasting techniques currently in use. Sensitivity analysis can be used to assess how various factors or hypotheses affect the precision of the forecast.

2.5 Conclusion

In Chapter 2, time series analysis is used to examine empirical methods for sales demand forecasting in the manufacturing industry. It highlights how crucial precise forecasting is to financial success, inventory control, and strategic planning. Methods such as seasonal decomposition, exponential smoothing, and ARIMA are analyzed and their specific properties for various demand data types are revealed. In keeping with a holistic approach, the chapter also highlights the importance of taking into account outside variables such as economic statistics and market movements. It does, however, recognize the limitations of time series forecasting, including sudden changes in the market, the emergence of new rivals, technical advancements, and unanticipated international events. Subsequent investigations want to develop creative methods for predicting, such as utilizing real-time data, machine learning algorithms, and sophisticated statistical methodologies. The next chapter will show the modification of data and the collection methods.

CHAPTER 3: METHODOLOGY

3.0 Introduction

A key process for business success is the forecasting of sales demand in a manufacturing company. This includes forecasting future demand for the company's products or services, which is essential to making informed choices in areas such as production planning, inventory management, and marketing strategies. Effective sales forecasting is essential to the growth and sustainability of a company in today's competitive business environment, Doganis (2006). This article presents a detailed research design for forecasting sales demand in a manufacturing company using a time series analysis approach.

3.1 Research design

The research design gives key tools used by the researchers as it provides guidelines for how the research was conducted. This research pays attention to the development of quantitative research design. The goal of this research design is to be able to predict future sales demand. Therefore, this study uses a quantitative–descriptive research design.

3.2 Data source

Data that was used was historical data from Corked spin investments. The historical data was collected from the internal control system of SAP ONE BUSINESS. This data was for monthly sales from 2017 up to 2023.

3.3 Target Population and Sampling Procedures

The target population was all monthly sales from the year 2017 up to 2023. There were 84 observations in this study sample. The sampling procedure that was applied is probability sampling, which involves using a random selection method to ensure that each element has a probability of being chosen.

3.4 Research instrument

The data was collected from the internal source from the manufacturing company's database, through SAP ONE BUSSINESS which provides valuable historical sales data, order data, customer profiles, and other relevant information. The information was transferred to Microsoft Excel for easy viewing and Microsoft R programming language containing statistical libraries was used for analysis of the data.

3.5 Method of data collection

The method that was used to collect the data was the source document where the historical data of past sales was collected. Historical sales data is a primary source of information for time series analysis. It includes records of past sales, time, and possibly other relevant attributes. This data was obtained from the manufacturing company's internal databases and sales records from SAP ONE BUSSINES. The data is copied to Excel for easy use and viewing purposes. The use of source documents serves as primary evidence, offering a clear audit trail for data sources and enhancing the creditability and trustworthiness of the data collected according to Appelbaum, (2016).

3.6 Description of variables and expected relationships

1. Dependent Variable - Sales Demand:

Sales demand is the dependent variable, which represents the quantity of products or services that customers are expected to purchase over a specific period. It serves as the primary variable of interest for forecasting. Sales demand can be measured in terms of units sold, revenue generated, or any other suitable metric.

2. Independent Variables:

Time is an independent variable in time series analysis. It represents the chronological order of observations and is essential for capturing patterns, trends, and seasonality in sales demand. Time can be measured in various units, such as years, months, weeks, or even shorter intervals.

Table 1: Variables

Variable	Symbol	Indicator	Source
Sales demand	SM	Monthly sales for Corked spin investment	Corked spin investment

Expected relationships

A positive relationship between historical sales and future sales:

It is generally expected trends of sales will continue in the future, therefore that higher historical sales will lead to higher future sales.

3.7 Data analysis procedures

Diagnostic test

1. Data Visualization:

To identify patterns, trends, and outliers in time series data, data visualization techniques such as line plots, scatter plots, and histograms may be useful. It is possible to gain a first indication of seasonal, trends or sudden changes in the data by visually examining them.

2. Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF):

Diagnostic methods such as ACF and PACF plots are used to determine whether there is autocorrelation in the time series data. While the PACF plot displays the correlation between the current observation and its lagged values after adjusting for intermediate lags, the ACF plot displays the correlation between the current observation and its lagged values. These graphs help to determine the proper lag times for the autoregressive AR and the moving average in the time series model.

3. Stationarity Tests:

In time series analysis, stationarity is an essential premise. Tests for stationarity, including the Augmented Dickey-Fuller (ADF) test, determine if the time series data shows a consistent mean and variation throughout time. To achieve stationarity for non-stationary data, the proper transformations and differencing are used.

4. Residual Analysis:

Analysing residuals entails looking at the time series model's residuals, or the variations between the anticipated and observed values. Plots of residuals, and autocorrelation plots of the residuals are used to evaluate how well the model fits the data. Residuals may show requirements for more fine-tuning to the model if they show patterns or strong autocorrelation.

5. Forecast Evaluation Metrics:

Metrics for evaluating forecasts, including mean absolute error (MAE), mean square error (MSE), or root mean square error (RMSE), compare the expected and actual numbers to determine how accurate the forecasting model is. These measures give a foundation for evaluating various forecasting models and quantifying the extent of forecast errors.

3.8 Analytical Model

Four stages must be followed in the development of the ARIMA model using the Box Jenkins approach: identification, estimation, diagnosis, and forecasting. In developing the model, the different components of the time series data, such as seasonality, trend, cyclicality, and irregularity, are crucial. Seasonality refers to predictable patterns that occur at regular intervals and are common in monthly sales demand data. Autoregressive and moving mean polynomials, which are identified by seasonal lags referred to as Multiplicative Seasonal Autoregressive Integrated Moving Mean Models are frequently applied for the development of models with these data. This model is written in a general form that includes white noise, an autoregressive element of degree p, a moving average element of degree q, an ordinary differenced component of degree d, a seasonal difference of degree D at lag s, and seasonal autoregressive and moving average differences of degrees P and Q at lag s.

$$(1-\emptyset_1\beta)(1-\emptyset\beta^4)(1-\beta)(1-\beta^4)y_t = (1-\theta_1\beta)(1+\theta_1\beta^4)e_t \qquad (13)$$

To compare the expected values of SARIMA, a feeding forward MLP shall be compared. The MLP has two secret layers, in addition to the backpropagation training methods. The past information in n months at the input level shall be used to predict future sales.

3.9 Model validation

3.9.1 Mean Absolute Percentage Error (MAPE):

To assess the precision of time series forecasts, MAPE serves as a widely utilized metric. It calculates the average percentage variance between the predicted values and the actual values. A lower MAPE indicates a more precise forecast.

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(y_t-\hat{y}_1)^2} \qquad (14)$$

3.9.2 Mean Squared Error (MSE) or Root Mean Squared Error (RMSE):

The Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) are metrics used to quantify the average squared difference between predicted values and actual values. These metrics gauge the magnitude of the prediction error.

MSE = $\frac{1}{N}\sum_{t=1}^{N}(y_t - \hat{y}_1)^2$ (15)

3.9.3 Forecast Error Distribution Analysis:

The performance of the model may be analyzed by analyzing the distribution of forecast errors. Histograms may be used to evaluate whether they follow a standard distribution, or if there is another pattern or bias.

3.10 Ethical consideration

Precise and accurate facts. The accuracy of the forecasts is directly impacted by the caliber of the data used in the research. As a result, it is essential to guarantee that the data is reliable, comprehensive, and devoid of prejudice. This involves making certain that the information is gathered and kept ethically, with the approval of the relevant parties, if any. Verifying the data and making sure it is representative of the total population are equally crucial.

The forecasting process's accountability and transparency. All of the procedures in the analysis, including the data sources, presumptions, and choices taken, must be recorded. This openness guarantees that the results may be independently confirmed and that the analysis to be replicated. Additionally, it encourages responsibility for the choices taken in light of the forecasts. Furthermore, to prevent misunderstandings or incorrect interpretations of the findings, stakeholders must be informed of the projections' limitations and uncertainties.

The projections' possible effects on many stakeholders, including consumers, workers, and investors. The financial performance of the organization can be significantly impacted by the projections' accuracy, and this can affect the stakeholders. As a result, it is imperative to make sure that the projections are used ethically and responsibly and that any potential dangers or restrictions are shared with the relevant parties.

The defense of confidentiality and privacy in data. When estimating sales demand, the business might have access to private data, including past client purchases and sales information, which needs to be managed carefully. It is crucial to make sure that this data is only accessed by authorized individuals and that the parties concerned have given their approval before it is shared or sold to third parties. To avoid any unwanted access or data breaches, the business must also have appropriate data security procedures in place. In addition to damaging the company's brand, violating data privacy and confidentiality policies may have legal repercussions.

Accountability and transparency. When using time series analysis to estimate sales demand, the business must be open and honest about the data sources, techniques, and underlying presumptions. This makes sure that everyone interested in the predictions may challenge any inconsistencies and understand how they are generated. The business must also take responsibility for any decisions made using the forecasts. In the case that the forecasts prove to be incorrect, the business must accept accountability and make the required changes to avoid repeating the same mistakes.

The forecasts. The projections must be used by the business to make responsible, well-informed decisions that benefit all parties involved, including shareholders, customers, and staff. Not only is it unethical, but using the forecasts to manipulate markets or take advantage of clients for financial gain may also have legal repercussions.

3.11 Conclusion

In a nutshell, this chapter is the study of research methodology. It answers all procedures that were taken during the research. Data analytics was incorporated, which will be utilized in the subsequent chapter. The following chapter will present the data and offer its interpretation.

CHAPTER 4:DATA PRESENTATION AND ANALYSIS

4.0 Introduction

This chapter's main focus is on the analysis of data that was collected for the study. The data was analyzed to evaluate the hypothesis and respond to the research questions mentioned earlier in the preceding chapters, Data is presented descriptively. This chapter encompasses the presentation and analysis of findings, alongside their interpretation. A retrospective time series analysis was performed to obtain meaningful data for the research objectives. The ultimate goal of the study forecast demand sales of Corked spin using the best model.

4.1 Descriptive statistics

Descriptive statistics shows the maximum, minimum, median, mean, and quantile values and reflects the outlies of the five years of data survey Descriptive statistics is an essential part of any statistical analysis, including forecasting sales demand using time series analysis. The data you have provided describes the statistical properties of the sales demand data over 7 years.

Table 2: Descriptive table

Mean	Median	Standard Deviation	Kurtosis	Skewness	Minimum	Maximum	1st Quartile	3rd Quartile
73852.7	66957.2	30064.8	-0.911	0.68	36330.1	135921	48881	97021

The mean sales demand value is approximately $73 851. This represents the average sales demand over the 7 years. This may indicate the general trend of sales and can provide a benchmark to measure future sales demand. The median sales demand value is approximately $66 957.15. This value indicates that there may be an even distribution of sales demand data, with an equal amount of sales demand above and below this mark. The standard deviation is approximately $30 064. This value measures the dispersion of the data set from the mean. The kurtosis value is -0.9113. This indicates how much of the variance of the data is due to infrequent extreme deviations, compared to frequent or moderate deviations from the mean sales demand value. The skewness value is 0.6800. Skewness describes the degree of asymmetry of the distribution of the data. In this case, the data is positively skewed, indicating that the data tends to cluster towards lower sales demand values, with the tail of the distribution extending to high sales demand values.

The range of sales demand is approximately $99 591. This value indicates the spread of sales demand values and may provide insight into how much sales demand can fluctuate during this period. The minimum and maximum sales demand values are approximately $36330.1 and $135 920.5, respectively. These values indicate the lowest and highest sales demand values recorded during the seven years. 1st quartile and 3rd quartile, these values indicate that 50% of the sales demand values are distributed between these two values, which are less than the median value of the sales demand. This data suggests that the lower half of the sales demand data is less than the median, indicating there are potentially significant fluctuations in sales demand over the seven years.

4.2 Data Analysis and disscusssion

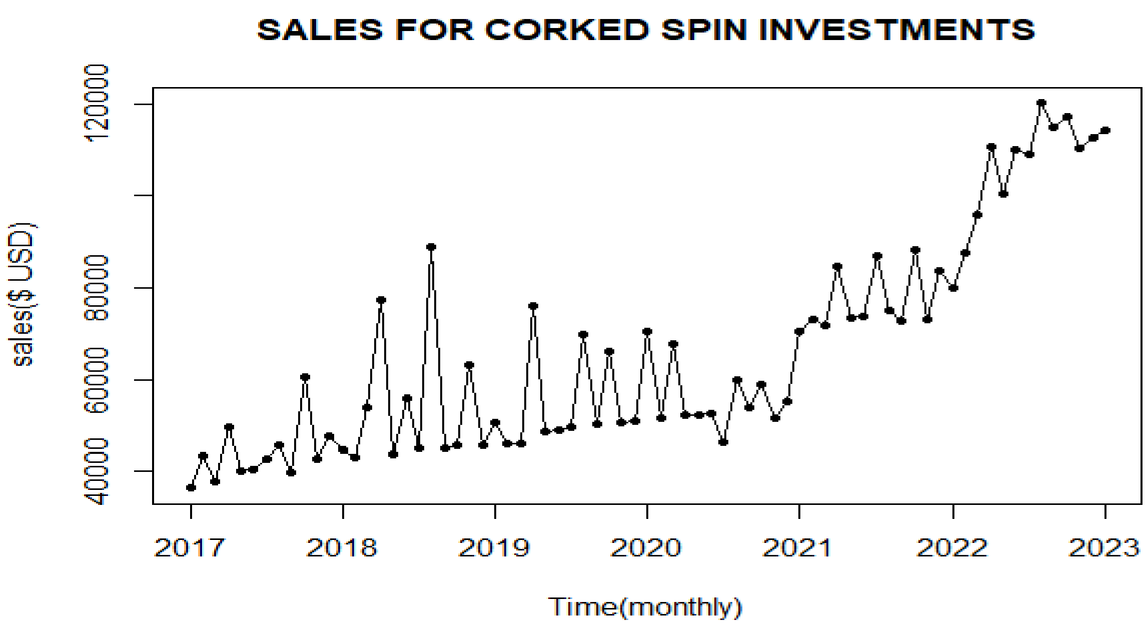

Figure 4: Row data time series plot

Corked spin investment sales were moderately high from the year 2017 -2020 they were a slow growth in sales. In the year 2021, they were a slight growth in sales this was caused by a change in currency as they were able to get imports at a cheaper price hence more production was done, and more sales were done. For the year 2022, the prices were reduced and they increased customers due to intense marketing, supplying more retail shops, and expanding their markets. They also started to do home deliveries hence sales rapidly growing.

4.3 Stationarity test

The stationarity test is performed utilizing ADF, which examines the origin of the unit root. To determine stationarity, the significant value or p-values is utilized

Table 3:Stationarity level test ADF

Augmented Dickey-Fuller Test
data: SM
Dickey-Fuller = -0.87549, Lag order = 4, p-value = 0.9507
alternative hypothesis: stationary

The null hypothesis was rejected because the p-value in the ADF test was greater than 0.05 hence concluded that it was not stationary and the time series variable was determined to be stationary.

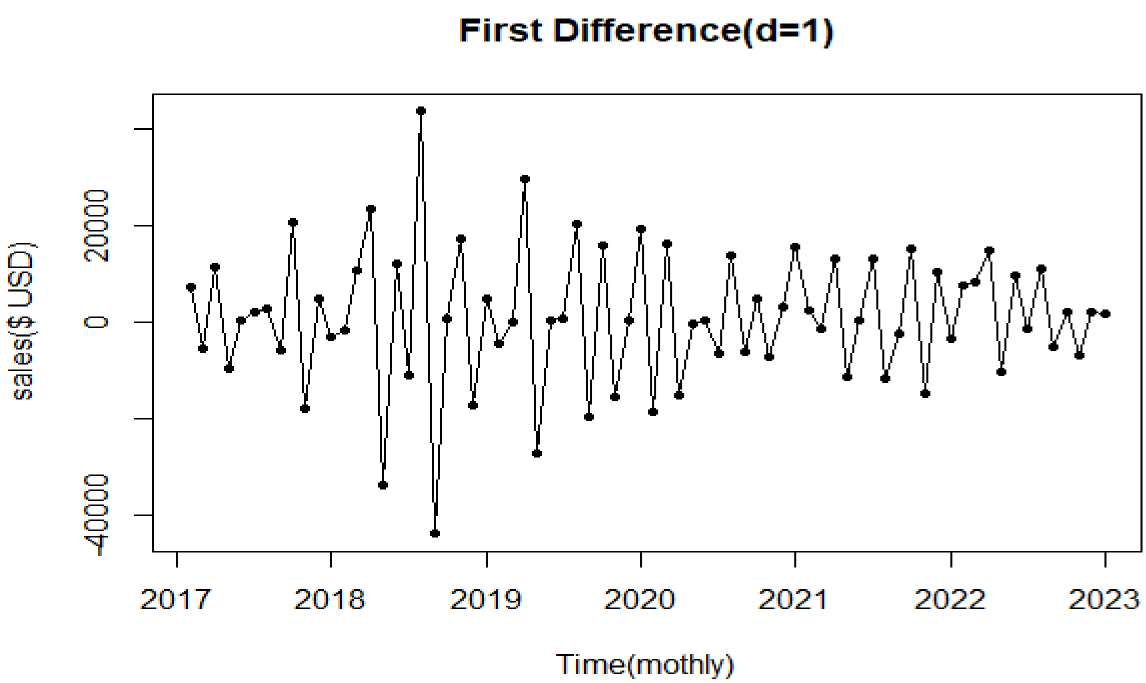

Figure 5: Stationary data time series plot

To assess stationarity, the observed data needs to undergo differentiation and testing. As observed, the sales data series seems stationary after the initial differentiation, yet some divergences from the mean persist. Additionally, the Augmented Dickey-Fuller (ADF) test is utilized to confirm whether the differenced data exhibits level or trend stationarity.

Table 4: Stationarity level test df 1 ADF

Augmented Dickey-Fuller Test
data: SM_Diff1
Dickey-Fuller = -4.9243, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary

Given that the time series variable was found to be stationary and the p-value in the ADF test was less than 0.05, the null hypothesis is now accepted.

4.4 Autocorrelation function (ACF) and Partial autocorrelation function (PACF)

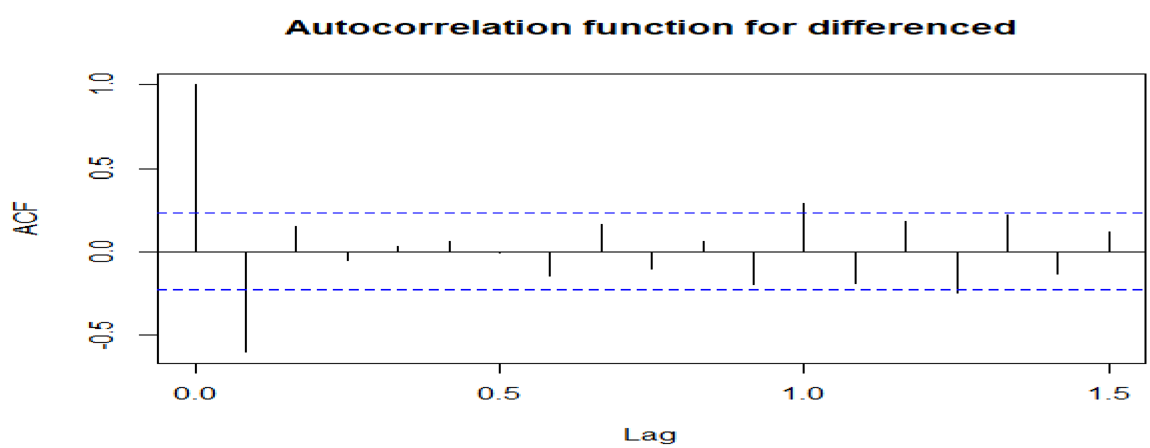

Figure 6: ACF

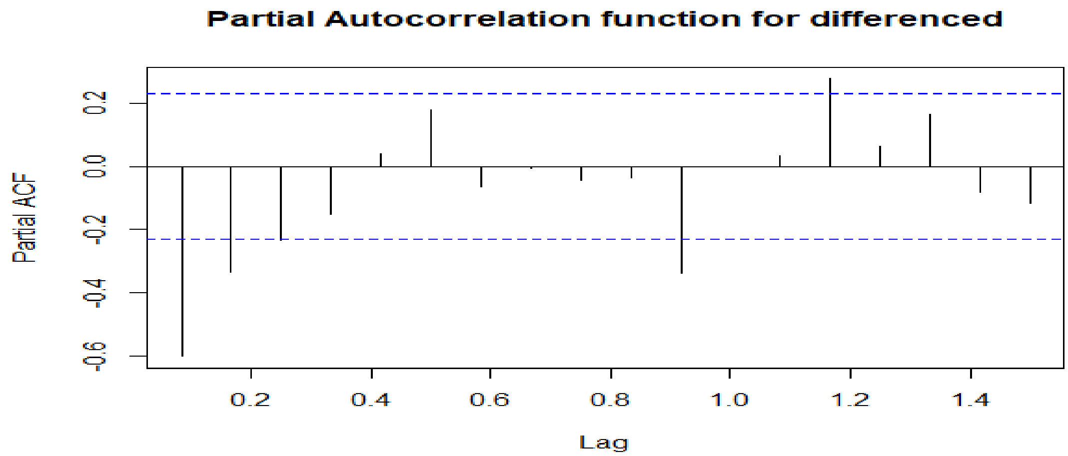

Figure 7: PACF

To determine the suitable model and ensure precise forecasts, we fitted multiple ARIMA models to the observed data, referencing the ACF and PACF plots. Given that the data was differenced, the fitted ARIMA models would be of order (p, d=1, q). Both ACF and PACF tails diminish to zero, and the lags remain within significant bounds. Hence, based on this analysis, it's viable to employ a SARIMA model of ARIMA 0,1,3 for the dataset.

4.5 Diagnostic Checking

The residuals were examined to see if the ARIMA model prediction was satisfied. The Ljung-Box test, ACF and PACF plots, histograms, and density charts were also done. The diagnostic analyses shown utilizing the residual plot demonstrate that the structure of sales appears to be constant, although certain residuals deviate from the mean zero and constant variance. As a result, the researcher is confident that ARIMA (0, 1, 3) is effective in forecasting sales of corked spin investment

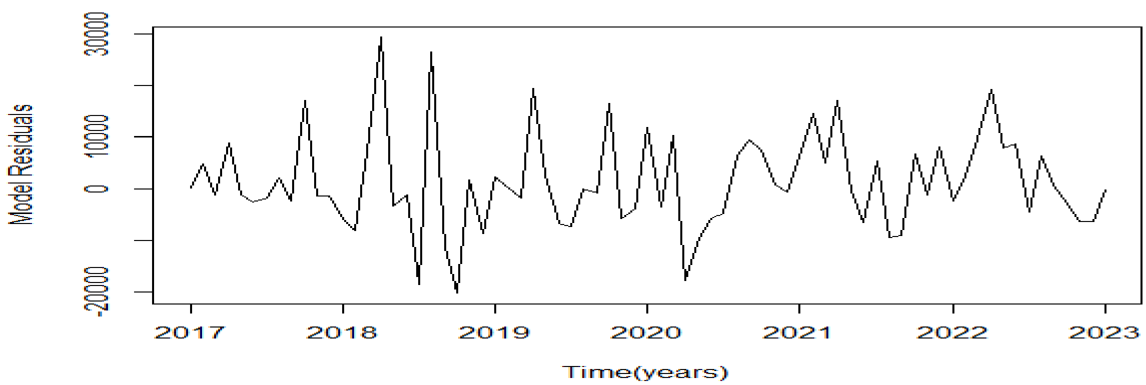

Figure 8: Model Residual Results

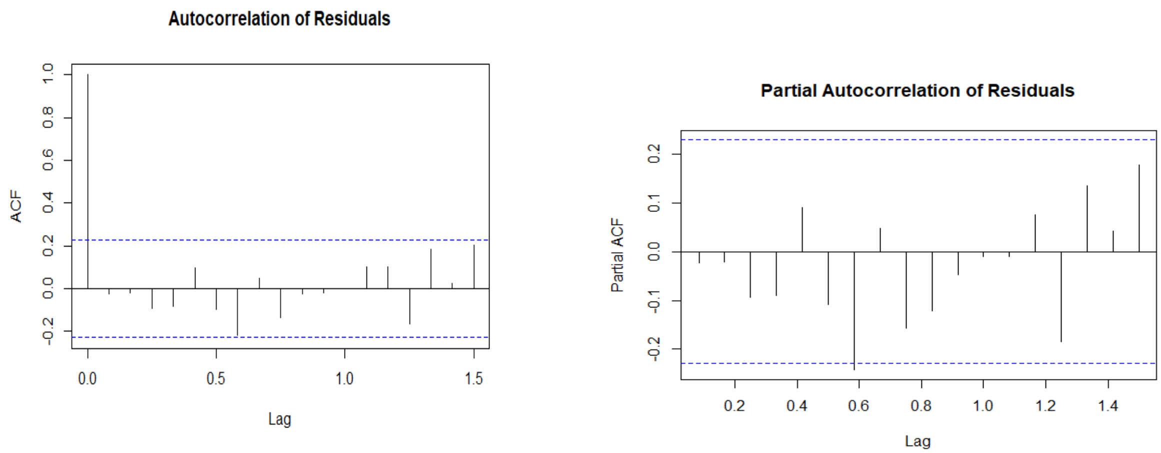

Figure 9: Testing for independence of residuals

The autocorrelation (ACF) plots show that, except lag 1, sample autocorrelation falls within the 95% confidence interval for the first 15 lags. The PACF of the residuals shows that the residuals' autocorrelation is all zeros, implying that all lags are uncorrelated. In conclusion, residuals have a constant, selected model SARIMA (0, 1, 3), and the real mean of the residuals is approximately zero. As a result, the chosen model meets all of the model assumptions.

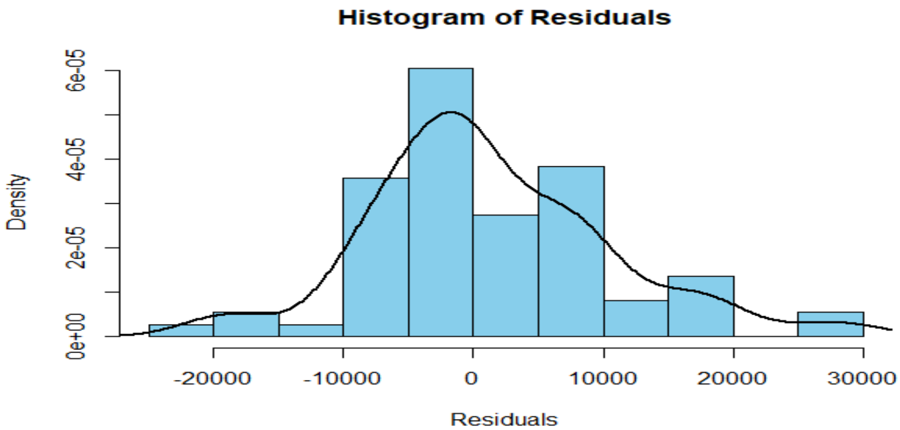

Figure 10: Testing for Normality of residuals

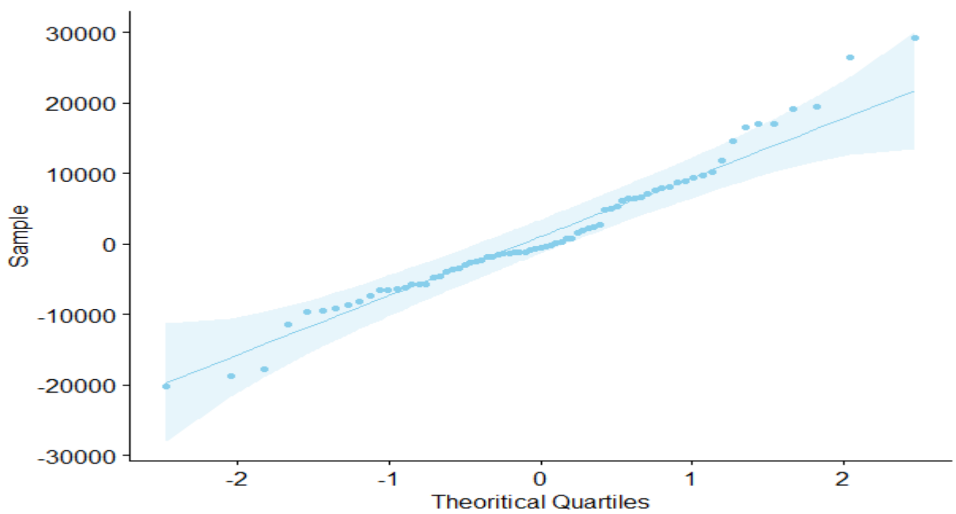

Figure 11: Testing for Normality of residuals

Normal Q plot, which illustrates that the normal distribution gives a satisfactory fit for this model since the extreme values somewhat tail off and the majority of the values are on the line, indicating normally distributed residuals.

Table 5: Box-Ljung test

Box-Ljung test
data: Residuals
X-squared = 2.071, df = 5, p-value = 0.8392

The p-values exceed the Q statistics, indicating that at a significance level of 5%, we do not reject the null hypothesis. This inference is drawn from the Ljung-Box test statistic table provided in the appendix. As a result, it is concluded that the residuals are independently established.

4.6 Model validation

4.6.1 SARIMA

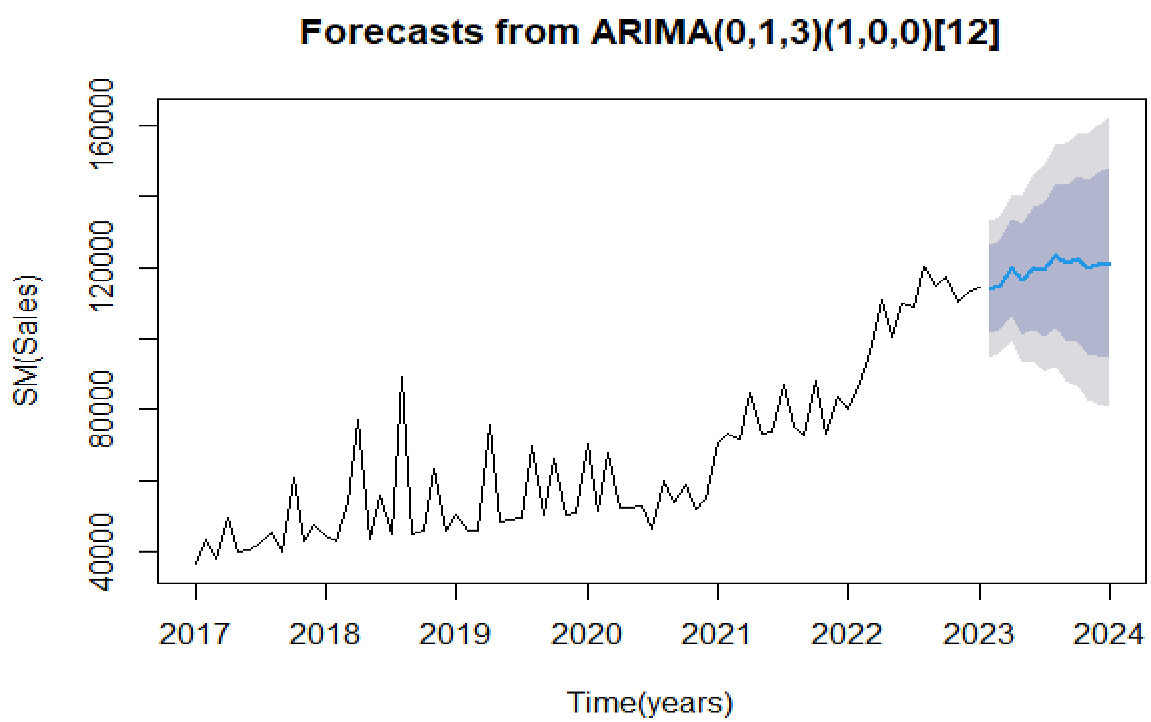

Figure 12: Forecasting 2023 sales demand

The above diagram shows the blue line's SARIMA forecasted values. The researcher observed that sales had been increasing constantly into the future from January 2023 to December 2023.

Table 6: Error measures

Training set error measures:							
	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	1292.51	9451.98	7001.19	0.23	11.04	0.43	-0.02

Table 7: SARIMA forecasted values

Forecast:					
	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Feb-23	113919.4	10368.73	126470.1	94724.84	133113.9
Mar-23	115093.4	102512.6	127674.3	95852.63	134334.2
Apr-23	1200068.5	106647.6	133489.5	99542.91	140594.2
May-23	116582.9	101101.5	132064.4	92906.15	140259.7
Jun-23	119867.6	102569.5	137165.8	93412.37	146322.9
Jul-23	119454.9	100513.5	138396.4	90486.54	148423.4
Aug-23	123292.9	102839.9	143746	92012.64	154573.3
Sep-23	121510.1	99649.54	143370.5	88077.3	154942.7
Oct-23	122282.2	99099.65	145464.8	86827.53	157736.9
Nov-23	119986.1	95552.87	144419.4	82618.69	157353.5
Dec-23	120743.7	95120.8	146366.7	81556.83	159930.6
Jan-24	121362.3	94602.53	148122.1	80436.75	162287.9

4.6.2 Multilayer perceptron

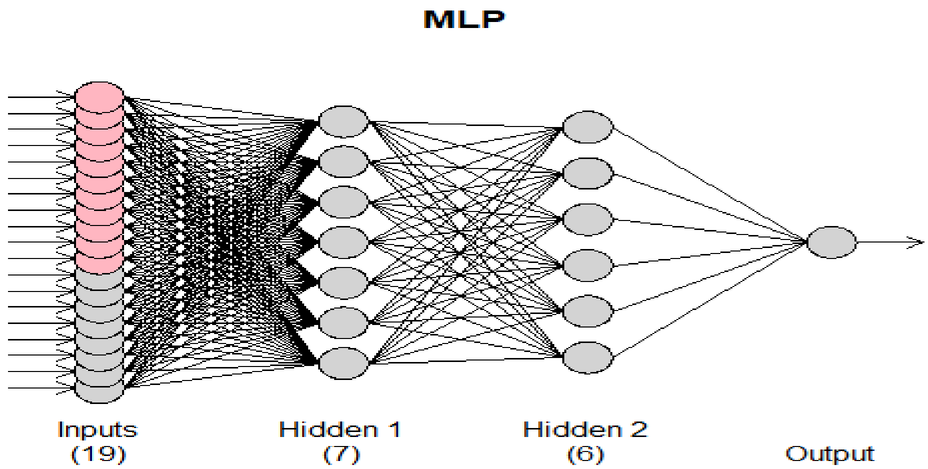

Figure 13: Nodes for MLP

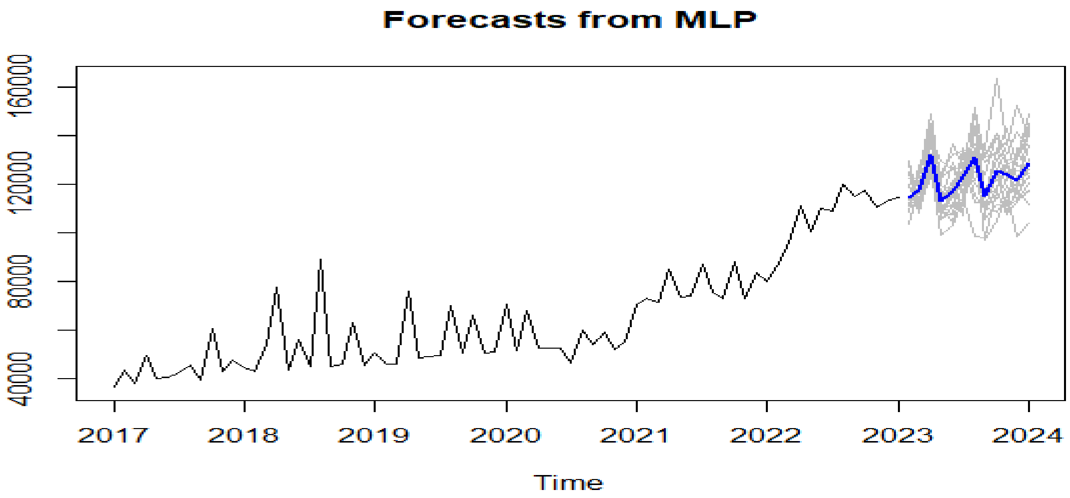

Figure 14: MLP Forecasted Sales Demand for 2023

The above diagram shows the blue line's forecasted values. The researcher observed that sales had been increasing constantly into the future from January 2023 to December 2023.

Table 8: Error measures MLP

Error measures:

	ME	RMSE	MAE	MPE	MAPE	ACF1
Training set	-7.06	336.26	170.62	-0.048	0.34	0.247

Table 9: Forecasted value for 2023 MLP

Forecast:	
	Point Forecast
Feb-23	114535
Mar-23	122643.9
Apr-23	128306.8
May-23	108487.4
Jun-23	110728.1
Jul-23	116727.4
Aug-23	118761.9
Sep-23	114206.9
Oct-23	124685.4
Nov-23	117593.8
Dec-23	120229.2
Jan-24	127519.3

4.7 Discussion of findings

Table 4.8 and Figure 4.12 below compare the actual demand for 2023, forecasted SARIMA values and Multilayer Perception forecasted values.

Table 10: Comparison of SARIMA and MLP

MONTH	ACTUAL	MLP	SARIMA
Jan-23	114468.46		
Feb-23	111986.58	114535	113919.4
Mar-23	111547.32	122643.9	115093.4
Apr-23	129677.73	128306.8	1200068.5
May-23	122136.21	108487.4	116582.9
Jun-23	122786.02	110728.1	119867.6
Jul-23	129650.77	116727.4	119454.9
Aug-23	133982.02	118761.9	123292.9
Sep-23	130806.22	114206.9	121510.1
Oct-23	132904.82	124685.4	122282.2
Nov-23	125459.25	117593.8	119986.1
Dec-23	135920.52	120229.2	120743.7
Jan-24		127519.3	121362.3
MAPE		0.378	11.042
RMSE		391.99	9451.97
MAE		183.69	7001.192

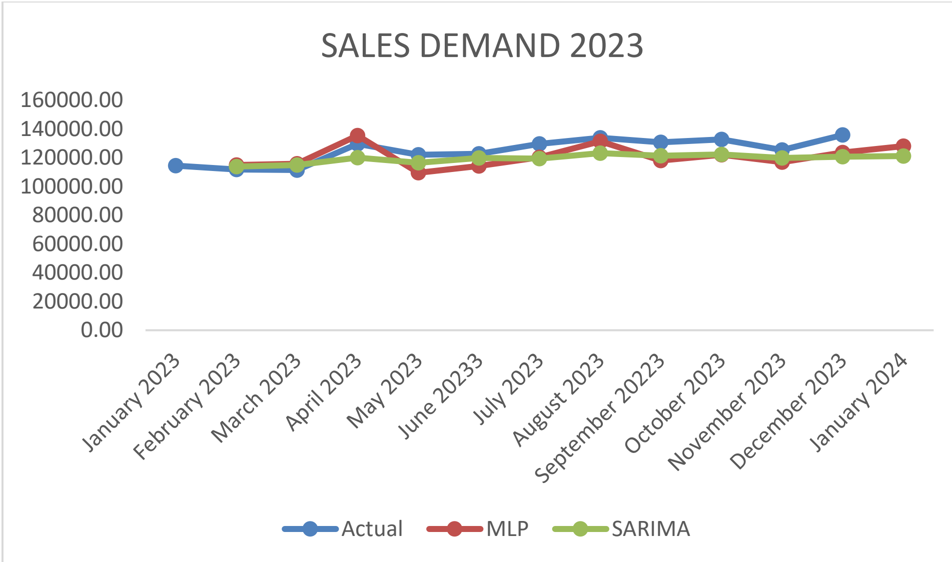

Figure 15: Comparison of SARIMA and MLP

From the data provided, it appears that the MLP (Multilayer Perceptron) method outperforms the SARIMA (Seasonal Autoregressive Integrated Moving Average) method in terms of accuracy when forecasting sales demand. To further elaborate, according to accuracy metrics MAPE and RMSE for both methods. MLP method obtained a MAPE of 0.345 and an RMSE of 357.06, while the SARIMA method obtained a MAPE of 11.043 and an RMSE of 9451.967, respectively. The lower the values of these metrics, the better. Therefore, based on output results from Figure 15 and Table 10, we can conclude that the MLP method provides better accuracy in forecasting sales demand for the given period.

4.8 Findings

According to the findings above it shows that forecasting sales demand using MLP is the best compared to SARIMA. Figure 4.13 below shows the forecasted graph for 2024 where sales show that there will be a rise in sales and Table 4.9 below shows the

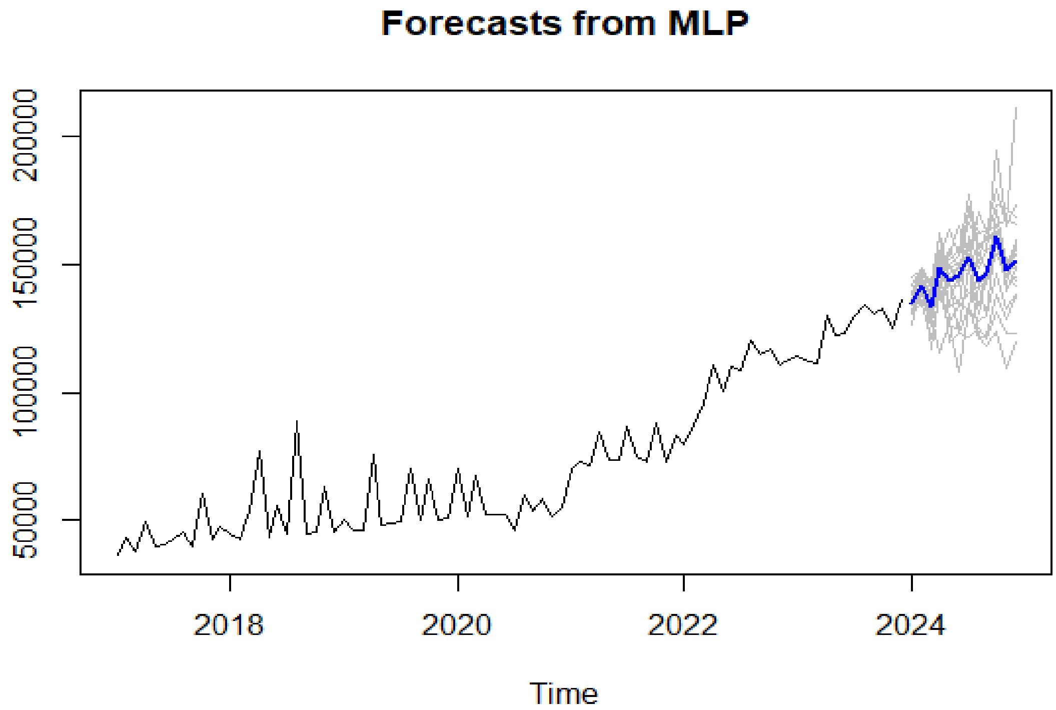

Figure 16: Forecasted sales demand for 2024

Table 11: Forecasted values for 2024

Forecast:	
Jan-24	135443.7
Feb-24	139408.8
Mar-24	134321.3
Apr-24	146127.3
May-24	140727.9
Jun-24	144254.7
Jul-24	146625.4
Aug-24	149529.2
Sep-24	145606.5
Oct-24	162068
Nov-24	144432.7
Dec-24	147253

4.9 Conclusion

Using the methodology described in the previous chapter, this chapter describes how to represent and analyze data. According to the formulated time series model, Corked Spin Investments is projected to experience a consistent increase throughout 2024. New knowledge on the performance of foresting sales and how it will fill the gaps that have been identified has also been gained in this chapter. Conclusions and recommendations shall be discussed in the next chapter.

CHAPTER 5:SUMMARY ,CONCLUSION AND RECOMMENDATIONS

5.0 Introduction

The chapter will summarize the findings and discuss the implications and limitations of the study. We will also provide recommendations for further research and improvements to the models used in the study. Overall, this chapter will provide a comprehensive analysis of the results obtained using time series analysis for forecasting sales demand, which can be valuable to organizations in making informed decisions regarding their sales strategy and resource allocation

5.1 Summary

The researcher's main focus was on the time series analysis of the sales data that was taken from Corked spin investment from the year 2017 up to 2023. The researcher also viewed the literature review of the study and highlighted the research gaps from the previous studies. Ethical consideration was also shown to show the security measures that were taken and the privacy of the data that was used. The study was quantitative research. Research questions and objectives were answered during the study. In fulfillment of the primary objectives, the study delves into the fundamental principles of time series analysis, emphasizing their relevance to predicting future sales within a manufacturing context. Comparing the strengths and weaknesses of different models, including SARIMA and MLP, the study evaluates their ability to analyze and forecast sales demand in the manufacturing industry. The study focuses on a direct comparison of SARIMA and MLP performance, highlighting their advantages and limitations in accurately predicting sales demand. Utilizing the chosen model, the study generates specific sales forecasts for the next 12 months in the year 2024.

After examining the sales history for Corked Spin Investments, both SARIMA and MLP models were developed. The MLP model used multiple hidden layers to analyze patterns in the past demand values. The SARIMA model incorporated seasonal, trend, and error components. Decomposition revealed a seasonal trend in the sales demand data. Autocorrelation analysis indicated significant autocorrelation at lags of 12, indicating seasonality. Seasonality testing confirmed a strong seasonal component in the data. The SARIMA model captured the seasonality and autocorrelation well, but its forecasting accuracy was limited. Consistently, the MLP model demonstrated superior performance compared to the SARIMA model in terms of forecast accuracy, as assessed through metrics such as Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). The MLP model was used to forecast sales demand for Corked Spin Investments for twelve months in 2024. The forecasted demand showed a slight upward trend, with seasonal fluctuations. This study supports previous findings that the MLP model is generally more effective than the SARIMA model for forecasting sales demand in manufacturing Choi and Oh (2019)

5.2 Recommendations

The findings show that there will be an increase in sales demand, and it shows months where there will be sharp increases and sharp decreases in demand according to seasons. Accurate forecasting is important for manufacturers to plan production levels, procurement, hiring needs, and inventory management. Better predictions can help optimize these operational processes. Comparing techniques provides valuable insight into which approaches are best suited for different contexts. This company now knows neural networks work well for their industry based on past evidence

Corked Spin Investments should consider adopting an MLP neural network model to forecast their sales demand 12 months ahead into 2024. This approach proved much more accurate than alternative methods like SARIMA in similar studies and market contexts. They will need clean, historical time series data on their monthly or quarterly sales amounts for the past 5+ years to train and validate the MLP model. This data should be thoroughly checked for completeness and anomalies beforehand. An expert in neural networks and time series analysis would need to develop the MLP model architecture, identify appropriate parameters, train it on the historical data, and select the final model configuration. This requires specialized statistical software. The forecasts produced by the MLP model should not be considered perfect predictions, as there is still some uncertainty. A range or confidence intervals around the projections would help in planning. The model should be re-trained periodically as new actual sales data becomes available, such as every 6-12 months, to keep it up-to-date with the latest trends and patterns. While an improvement, automated forecasts still require human judgment. The results should be reviewed by experienced sales or demand planners before being set in stone. Unusual forecasts may require further investigation.

Despite its superiority, the study acknowledges that the MLP model is more complex and computationally intensive than SARIMA. This complexity may necessitate more advanced technical expertise and resources for implementation. The study concludes that the MLP model offers a more accurate and reliable solution for forecasting sales demand compared to SARIMA. While requiring more technical expertise and resources, it provides invaluable insights into future sales trends, enabling informed decision-making for Corked Spin Investments and other manufacturing companies.

5.3 Areas of Further Research

Future research could explore the integration of additional variables, such as macroeconomic indicators or competitor activity, into the MLP model to potentially further enhance its predictive power. This would contribute to a more comprehensive and robust forecasting framework for the manufacturing industry.

REFERENCES

Adeyinka, D.A. and Muhajarine, N., 2020. Time series prediction of under-five mortality rates for Nigeria: comparative analysis of artificial neural networks, Holt-Winters exponential smoothing and autoregressive integrated moving average models. *BMC medical research methodology*, *20*, pp.1-11.

Appelbaum, D., 2016. Securing big data provenance for auditors: The big data provenance black box as reliable evidence. *Journal of emerging technologies in accounting*, *13*(1), pp.17-36.

Bowersox, D.J., Stank, T.P. and Daugherty, P.J., 1999. Lean launch: managing product introduction risk through response-based logistics. *Journal of Product Innovation Management: AN International Publication of The Product Development & Management Association*, *16*(6), pp.557-568.

Chase, C.W., 2013. *Demand-driven forecasting: a structured approach to forecasting*. John Wiley & Sons.

Chen, C., Wang, Y., Huang, G. and Xiong, H., 2019, December. Hierarchical demand forecasting for factory production of perishable goods. In *2019 IEEE International Conference on Big Data (Big Data)* (pp. 188-193).

Chong, A.Y.L. and Zhou, L., 2014. Demand chain management: Relationships between external antecedents, web-based integration, and service innovation performance. *International journal of production economics*, *154*, pp.48-58.

Coveney, M. and Cokins, G., 2017. *Budgeting, Forecasting, and Planning In Uncertain Times*. John Wiley & Sons.

Dagum, E.B. and Bianconcini, S., 2016. *Seasonal adjustment methods and real time trend-cycle estimation*. Berlin/Heidelberg, Germany: Springer International Publishing.

Davis, D.F. and Mentzer, J.T., 2007. Organizational factors in sales forecasting management. *International Journal of Forecasting*, *23*(3), pp.475-495.

Doganis, P., Alexandridis, A., Patrinos, P. and Sarimveis, H., 2006. Time series sales forecasting for short shelf-life food products based on artificial neural networks and evolutionary computing. *Journal of food engineering*, *75*(2), pp.196-204.

Edmundson, B., Lawrence, M. and O'Connor, M., 1988. The use of non-time series information in sales forecasting: A case study. *Journal of Forecasting*, *7*(3), pp.201-211. (Edmundson, 1988)

Esling, P. and Agon, C., 2012. Time-series data mining. *ACM Computing Surveys (CSUR)*, *45*(1), pp.1-34.

Firoz, A.S., 2014. Long term perspectives for Indian steel industry. *Ministry of Steel, Economic Research Unit mimeo, New Delhi, May*.

Gonçalves, J.N., Cortez, P., Carvalho, M.S. and Frazao, N.M., 2021. A multivariate approach for multi-step demand forecasting in assembly industries: Empirical evidence from an automotive supply chain. *Decision Support Systems*, *142*, p.113452.

Guðmundsson, J.I., 2022. *A systematic categorization process facilitating the selection of demand forecasting methods* (Doctoral dissertation).

Güven, İ. and Şimşir, F., 2020. Demand forecasting with color parameter in retail apparel industry using artificial neural networks (ANN) and support vector machines (SVM) methods. *Computers & Industrial Engineering*, *147*, p.106678.

Hodson, T.O., 2022. Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geoscientific Model Development Discussions*, *2022*, pp.1-10.

Hyndman, R.J. and Athanasopoulos, G., 2018. *Forecasting: principles and practice*. OTexts.

Keleperas, D., 2023. Exploring the importance of demand forecasting in B2B: Enhancing the overall customer experience and operational efficiency by utilizing forecasting and collaboration at a manufacturing company.

Kerkkänen, A., 2010. *Improving demand forecasting practices in the industrial context*. Lappeenranta University of Technology.

Khosroshahi, H., Husseini, S.M. and Marjani, M.R., 2016. The bullwhip effect in a 3-stage supply chain considering multiple retailers using a moving average method for demand forecasting. *Applied Mathematical Modelling*, *40*(21-22), pp.8934-8951.

Kilimci, Z.H., Akyuz, A.O., Uysal, M., Akyokus, S., Uysal, M.O., Atak Bulbul, B. and Ekmis, M.A., 2019. An improved demand forecasting model using a deep learning approach and proposed decision integration strategy for the supply chain. *Complexity*, *2019*.

Lalou, P., Ponis, S.T. and Efthymiou, O.K., 2020. Demand forecasting of retail sales using data analytics and statistical programming. *Management & Marketing*, *15*(2), pp.186-202.

Powers, W.J., Rabinstein, A.A., Ackerson, T., Adeoye, O.M., Bambakidis, N.C., Becker, K., Biller, J., Brown, M., Demaerschalk, B.M., Hoh, B. and Jauch, E.C., 2019. Guidelines for the early management of patients with acute ischemic stroke: 2019 update to the 2018 guidelines for the early management of acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*, *50*(12), pp.e344-e418.

Preis, S.J., 2021. *Analysis and Evaluation of the Impacts of Predictive Analytics on Production System Performances in the Semiconductor Industry* (Doctoral dissertation, University of Gloucestershire).

Sankaran, S., 2014. Demand forecasting of fresh vegetable product by seasonal ARIMA model. *International Journal of Operational Research*, *20*(3), pp.315-330.

Småros, J., Lehtonen, J.M., Appelqvist, P. and Holmström, J., 2003. The impact of increasing demand visibility on production and inventory control efficiency. *International journal of physical distribution & logistics management*, *33*(4), pp.336-354.

Smith Jr, A.A., 1993. Estimating nonlinear time-series models using simulated vector autoregressions. *Journal of Applied Econometrics*, *8*(S1), pp.S63-S84.

Uzzaman, I., Rahman, M., Alam, M.S. and Alam, S., 2016. Simulation of cement manufacturing process and demand forecasting of cement industry. *Global Journal of Researches in Engineering: G Industrial Engineering*, *16*(2).

Vereecke, A., Vanderheyden, K., Baecke, P. and Van Steendam, T., 2018. Mind the gap–Assessing the maturity of demand planning, a cornerstone of S&OP. *International Journal of Operations & Production Management*, *38*(8), pp.1618-1639.

Vergori, A.S., 2017. Patterns of seasonality and tourism demand forecasting. *Tourism Economics*, *23*(5), pp.1011-1027.

Yang, C.H., Chen, B.H., Wu, C.H., Chen, K.C. and Chuang, L.Y., 2022. Deep learning for forecasting electricity demand in Taiwan. *Mathematics*, *10*(14), p.2547.

Zhao, H., Zuo, X. and Lin, P., 2022, February. Sales forecasting for Chemical Products by Using SARIMA Model. In *Proceedings of the 5th International Conference on Big Data and Education* (pp. 419-427).

APPENDICES

```
library(tseries)

library(forecast)

library(ggpubr)

View (YOMILK)

summary (YOMILK)

SM=ts(YOMILK$sales,start = 2017,end = 2023,frequency = 12)

plot(SM,xlab='Time(monthly)',ylab='sales($ USD)',main='SALES FOR CORKED SPIN
INVESTMENTS',type ='o',pch = 20)

adf.test(SM)

SM_Diff1=diff(SM,d=1)

plot(SM_Diff1,main='First Difference(d=1)',xlab='Time(mothly)',ylab='sales($ USD)',type
='o',pch = 20)

acf(SM)

summary(SM)

adf.test(SM_Diff1)


acf(SM_Diff1,main = 'Autocorrelation function for differenced')

pacf(SM_Diff1,main = 'Partial Autocorrelation function for differenced')

library(forecast)


SMmodel=auto.arima(SM)

SMmodel
```

```
summary(SMmodel)

Box.test(Residuals, lag=5, type ="Ljung-Box")

Residuals=residuals(SMmodel)

acf(Residuals,main='Autocorrelation of Residuals')

plot(Residuals,xlab='Time(years)',ylab='Model Residuals')

pacf(Residuals,main='Partial Autocorrelation of Residuals')

hist(Residuals,col = 'skyblue')+ggdensity(Residuals)

hist(Residuals, col = "skyblue", main = "Histogram of Residuals", xlab = "Residuals",
probability = TRUE)

lines(density(Residuals), col = "black", lwd = 2)

library(ggqqplot)

ggqqplot(Residuals,color = ('skyblue'),xlab='Theoritical Quartiles',type ='o',pch=15)

w=forecast(SMmodel,h=12)

 plot(w, xlab='Time(years)',ylab='SM(Sales)')

summary(w)

library(tseries)

SM=ts(YOMILK$sales,start=2017,end = 2023,frequency =12)

SM

plot(SM)

library(nnfor)

library(ggpubr)

#Model fitting and training

mlp.fit=mlp(SM,hd=c(7,6))
```

```
plot(mlp.fit)

print(mlp.fit)

mlp.forecast=forecast(mlp.fit,h=12,include=34)

plot(mlp.forecast)

summary(mlp.forecast)
```

11%	10%	6%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	fastercapital.com Internet Source	2%
2	www.mdpi.com Internet Source	1%
3	ir.buse.ac.zw:8080 Internet Source	1%
4	www.ncbi.nlm.nih.gov Internet Source	<1%
5	Submitted to University of Glasgow Student Paper	<1%
6	vntkumar8.github.io Internet Source	<1%
7	rstudio-pubs-static.s3.amazonaws.com Internet Source	<1%
8	www.science.gov Internet Source	<1%
9	onlineresource.ucsy.edu.mm Internet Source	<1%
10	www.ieworldconference.org Internet Source	<1%
11	Submitted to CSU, San Jose State University Student Paper	<1%
12	articles.cyzerg.com Internet Source	<1%
13	atsa-es.github.io Internet Source	<1%
14	yourstory.com Internet Source	<1%
15	ir.msu.ac.zw:8080 Internet Source	<1%
16	Cristina Fernández. "Streamflow drought time	<1%