BINDURA UNIVERSITY OF SCIENCE EDUCATION

FACULTY OF SCIENCE AND ENGINEERING

DEPARTMENT OF STATISTICS AND MATHEMATICS

EXAMINING THE USE OF BUSINESS INTELLIGENCE ON REVENUE FORECASTING IN QUICK SERVICE RESTAURANTS USING MULTIVARIATE LINEAR REGRESSION MODEL: A CASE OF SIMBISA BRANDS.

BY

RONALD CHAWORA

B192364B

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS OF BACHELOR OF SCIENCE HONORS DEGREE IN STATISTICS AND FINANCIAL MATHEMATICS.

SUPERVISOR: Ms. J. C. PAGAN'A

JUNE 2023

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.

RONALD CHAWORA **09/06/2023.**

Student name Signature Date

Certified by:

Ms. J. C. PAGAN'A **09/06/2023.**

Supervisor Signature Date.

Dr. M. MAGODORA ………………… **09/06/2023.**

Chairperson Signature Date

DEDICATION

I dedicate this research project to my lovely father Mr. M Chawora, who has been a constant source of support throughout my academic journey. His unwavering belief in my abilities has been a guiding force in my life, and I am grateful for his love and encouragement. He gave me every reason to remain focused.

ACKNOWLEDGEMENTS

I express my sincere gratitude in writing this acknowledgement as I could not have completed my thesis without the invaluable assistance of my supervisor, family and friends. I extend my deepest appreciation to my supervisor, Ms. J. C. Pagan'a, for consistently providing me with support throughout the development of my dissertation. Their patience, guidance, motivation, and extensive knowledge proved invaluable during the analysis and writing process. I also thank God for providing me with the care, strength, knowledge, and opportunity to pursue my education to this level. Additionally, I would like to thank the lecturers and staff of the Department of Statistics and Mathematics at Bindura University of Science Education for their academic support and knowledge sharing. To my family, I express my sincere gratitude for their unwavering encouragement, advice, and prayers all through my studies. I pray that the Lord blesses them abundantly. Finally, to my colleagues, I consider you all as family, and I ask for God's guidance and protection to be with you always.

ABSTRACT

The aim of this research was to unpack the use of Business Intelligence (BI) tools to improve revenue forecasting in quick service restaurants (QSRs) by using a multivariate linear regression model. The study involved collecting and assessing historical data on sales, turnover, purchases, implementing BI tools such as data visualization and predictive analytics. The intended outcome was to discover how BI methods could be utilized to estimate future revenue, streamline operations and enable more data-driven decisions. The study's main objectives were to examine the impact of various factors such as demand, price, and inflation on revenue and to evaluate the role played by Business Intelligence systems in revenue forecasting. The study's findings showed that purchases, cost, profit, and inventory were good predictors of turnover while factors such as customers and inflation had minimal impact. Findings from this research provides useful insights into the factors that can affect a QSR's performance and revenue forecasting, allowing businesses to make better-informed strategic decisions. By using a model for evaluating historical data and predicting future performance, businesses can identify trends and patterns which can be used to inform their revenue forecasting. It can also make necessary changes to their strategies in real-time and optimize performance to maximize revenue.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.0 Introduction

The author gives a thorough overview of the study in this chapter. The section outlines the problems that prompted the research, presents the research objectives, questions, and problem statement, as well as the study's significance. As well, this chapter also discusses the study's limits, presumptions, and delimitation. It concludes with an outline of the study report. The study report is organized into five chapters. The first chapter presents an introduction to the study, the second chapter reviews relevant literature from previous studies, chapter three presents the methodology that was used in the study. The fourth chapter presents the findings of the study as well as analyzing and discussing the findings while chapter five presents the research summary, conclusions, recommendations and areas for further study.

1.1 Background to the Study

The business landscape is characterized by complex, dynamic and rapidly changing landscape, increased competition, evolving client needs. Businesses have to use traditional models to forecast revenues but of late has enabled the development of sophisticated business intelligence tools with built in algorithms that can greatly simplify analyses and forecasting of data by analyzing historical trends and patterns and extrapolating into the future. Business intelligence tools are technologies that companies can employ to exploit data in order to gain insights resulting in more intelligent revenue forecasting. To that end, there are several ways that business organizations can leverage Business Intelligence (BI) technologies, forecast revenues and increase profitability, stay agile and remain competitive in the face of uncertainty.

Before the 20th Century, gathering and analyzing data proved to be a huge challenge. This led to the development of Decision Support Systems (DSSs) to help departmental managers make decisions pertaining to their departments. However, the information provided was limited in that it focused on one functional area of the Business which was compounded by the rapid accumulation of data at the disposal of organization from which decisions are based. This limitation led to the development of Executive Support Systems that provided information across all departments within a business organization.

By the 1970s, the shortcomings of the Executive Support Systems (ESSs) had become apparent in that just like DSSs, they focused on one management level, that is the executive managers and this led to the development of Business Intelligence Systems and associated technologies. Business Intelligence is different from traditional decision making in that it supports decision making at all the management levels, can perform multivariate analysis of structured and unstructured data gathered from multiple and disparate sources. According to Marcano and Talavera (2007), Business Intelligence is utilized to assist in decision-making by converting data from both internal and external transactional systems and unstructured information into structured information through cleaning and transforming mechanisms. Structured data is the crucial input that transforms into knowledge and gives the decision-maker the foundation and support they need. The analysis of various company performance circumstances is made easier by the structured information. It is presented throughout time periods, comparing indicators, detecting behaviors, and forecasting evolutions based on trends that help to minimize risks and clarify business perspectives on decision-making (Marcano and Talavera, 2007). According to Nagash (2004) Business Intelligence Systems combine operational data with Business Intelligence can contribute to optimum decision making, thus improving operational efficiency, enhancing competitiveness and resultantly increasing the profitability of business organizations through seamless analysis and reporting of information.

While Business Intelligence (BI) technologies have certainly brought notable progress to the business environment, it is important to acknowledge certain factors and obstacles. One such consideration is the issue of data quality and integration. BI heavily depends on data, and ensuring the quality and seamless integration of data from diverse sources can present challenges. Inaccurate or incomplete data has the potential to result in flawed insights and decision-making processes. Thus, guaranteeing data quality and effectively integrating data from disparate sources can be a demanding and time-consuming task.

1.2 Problem Statement

The modern business environment is volatile, uncertain and unpredictable due to intense competition and rapid technological change. Major global events such as the COVID19 pandemic with the attendant trading restrictions thus curtailing the operations of businesses made it much more difficult for businesses to generate revenue. Zimbabwe, in particular, is facing a plethora of challenges such as inflation, currency depreciation and power challenges which presents uncertainties on the operations of businesses particularly on the revenues side. This uncertainty has made it much more difficult to accurately forecast revenue for organizations. The profitability of a business organization and hence sustainable growth, to a larger extent, depends on the extent to which it generates revenue. The volume and complexity of data generated to help forecasting decisions can be daunting. This leads to slow and poor decisions which if relied upon may result in unintended outcomes and hence the need to invest in to use Business Intelligence tools for accurate and timely decisions. Business Intelligence tools simplify complex decision making and hence this study seeks to examine the role of Business Intelligence tools in revenue forecasting with a view to determining whether Business Intelligence can contribute to improved revenue forecasting in businesses.

1.3 Research Objectives

The objectives of this study were as follows;

i. To examine the role of Business Intelligence systems on Revenue forecasting.
ii. To evaluate the impact of demand on revenue.
iii. To determine the impact of price on revenue.
iv. To examine the impact of inflation on revenue.

1.4 Hypothesis

H_o: The use of Business Intelligence has a positive effect on Revenue

H_1: The use of Business intelligence will not have a positive effect on Revenue

1.5 Research Questions

The study aimed at answering the following research questions;

i. How can business intelligence improve the accuracy of revenue forecasting?
ii. How does demand affect the revenues of a business organization?
iii. How does product prices affect revenues?
iv. What is the influence of inflation on revenues?

1.6 Significance of the Study

The study is intended to benefit two major stakeholders who are;

Business Organizations.

Business Organizations are set to benefit from the study through the recommendations of the study. Since the study was focused on the role of Business Intelligence on Revenue forecasting, the results of this study could be applied in helping organizations appreciate the benefits of implementing and adopting business Intelligence technology.

The academia

The study would benefit academia through its contribution to existing literature on forecasting revenues using Business Intelligence tools. The results of the study would be an addition to the existing body of knowledge regarding the role of Business Intelligence in revenue forecasting and this could be used in future studies as reference.

1.7 Assumptions of the Study

The researcher made the following assumptions;

- Availability of sufficient and reliable data. The assumption is that there is access to a comprehensive and accurate dataset consisting of relevant variables such as sales, profit, inventory, inflation, cost, customers and purchases.
- Appropriate Application of Multivariate Linear Regression Model .It is assumed that the underlying assumptions of the multivariate linear regression model are met. This includes

assumptions such as linearity, independence of errors, homoscedasticity (constant variance), absence of multicollinearity among predictor variables, and normally distributed residuals

1.8 Limitations to the Study

The study faced a number of limitations which include not having enough time to conduct a thorough study and lack of financial resources. To overcome the time constraint, the researcher had to put in extra effort and to rely on financial support from his family to fund the research process.

1.9 Delimitation of the Study

The study will rely on available data sources, such as historical sales data, operational data, and business intelligence tools. The accuracy, completeness, and reliability of the data may vary, potentially impacting the validity of the results.

1.10 Summary

In this chapter the researcher has provided the reader with the introduction and background of the study. Key subtopics covered in the chapter include problem statement, the research objectives, research questions, significance of the study, assumptions of the study, limitations to the study, delimitation of the study and an outline of the study report. The following chapter reviews relevant literature from previous studies.

CHAPTER 2

LITERATURE REVIEW

2.0 Introduction

In this chapter the researcher reviews existing scholarly works by other scholars related to the role of business intelligence (BI) in forecasting revenue. Major topics in this chapter are the Theoretical and Conceptual framework, the concept and purpose of BI, effectiveness of BI in Revenue Forecasting, the role of Business Intelligence in forecasting demand and the role of BI in demand forecasting. The chapter will end by Empirical Literature review which gives an account of and discusses the work of other researchers on impact of business intelligence on revenue forecasting

2.1 Literature review

According to Beins and McCarthy (2012), a literature review is a representation of published articles, books, and other scholarly works related to a particular research topic. Cooper and Schindler (2014) argue that a review of related literature involves an assessment of previous research studies, historical or recent data from specific companies, or industry reports that support the study. The literature review should also explain the reason for conducting the proposed research and identify deficiencies and gaps in secondary data sources. Additionally, the review of related literature can involve analyzing the conclusions derived from earlier research studies, examining the accuracy of secondary data sources, and evaluating the suitability and reliability of previous studies (Cooper and Schindler, 2014) while preserving the original meaning.

2.2 Conceptual and Theoretical framework

Cooper and Schindler (2014) define a Conceptual Framework as a comprehensive collection of concepts arranged in a logical manner, which serves as the basis for synthesizing and interpreting information and provides the rationale and focus for carrying out research. According to Imenda (2014), a conceptual framework is expressed through word models and serves as the foundation for many theories like Information Systems Theory. This theory emphasizes the role of information systems, including business intelligence tools, in supporting decision-making processes and improving organizational performance. It highlights the importance of data collection, analysis, and dissemination for effective decision-making.

2.2.1 The concept of Business Intelligence

Nazari F. et al (2022), posits that a BI system integrates a set of tools, technologies, and products which collect, integrate, analyze, and present data. BI is used to methodically store and manage operational data, and through the use of variety of statistical and analytical tools and data mining techniques to analyze operational data in order to provide analytic reports and decision support information for various business activities (Wang, Fan & Xu, 2012). Negash (2014) argues that the term BI has substituted the use of the terms decision support, Executive Information Systems (EIS) and Management Information Systems (MIS). In contrast, Rouhani, Asgari and Mirhosseini (2012) posits that the term BI is occasionally used interchangeably with the term EIS. Rouhani et al. (2012), further point out that BI systems are Decision Support Systems (DSS) that provide periodic reports derived from historical data. The Chartered Institute of Management Accountants (CIMA) describe BI as the technical architecture or 'stack' of systems that extract, assemble, store and access data to provide reports and analysis. Hence a BI system involves developing processes and systems that collect, transform, cleanse and consolidate organization wide and external data, for presentation as reports, summaries and visualizations in the form of dashboards or scorecards.

2.2.2 Purpose of Business Intelligence

The main purpose of the BI system includes intelligent exploration, integration, accumulation, and multidimensional analysis of data gathered from various data sources. The key objective of BI is to transform different types of data from a variety of sources into meaningful crucial for businesses (Ritesh and Srimannarayana, 2013). BI system implementation improves organizational performance, supports information quality and timeliness, to empower decision makers so they are better aware of their company's strategic position in relation to competitors. Furthermore, BI supports the organizational decisions by providing access to the existing data which is then transformed into information that can be presented in the required format, to facilitate and aid decision making. BI technological solutions simplify complex decision making resulting from increasing business activity which results in increased volume, velocity and variety of data available to decision makers. Ramb (2013) posits that the power of BI lies in the fact that it combines data from different functional departments of an organization into a single version of the truth. In addition, BI tools have the following statistical and analytical capabilities:

a) Descriptive analytics

   Descriptive analytics examines past data sets for trends and patterns. It answers questions related to the past by identifying performance metrics to measure, collecting data about the metrics and then analyzing the data. Essentially, Descriptive analytics transforms data in the form of facts and transforms it into information which can be acted upon.

b) Diagnostic Analytics

   The purpose of Diagnostic Analytics is to explain why things happened the way they did. Not only does it answer the question of what happened but also why it happened by identifying trends and patterns in the past and then going a step further to explain why the trends occurred the way they did. Diagnostic analytics logically follows descriptive analytics.

c) Predictive analytics

   Predictive analytics aims to predict likely outcomes and make forecasts using past or historical data. Using extrapolation, predictive analytics extends trends into the future to predict possible outcomes using probabilities and statistical modeling.

d) Prescriptive Analytics

   Using data from a variety of sources, prescriptive analytics identifies possible future outcomes and prescribes the best option for possible adoption. Instead of using raw data, prescriptive analytics provides insights of what should happen, and not just what could happen.

2.2.3 Forecasting

Forecasting can be defined as predicting the future development of a particular quantity based on rational methods and current data (Hyndman & Athanasopoulos, 2018). It involves predicting or estimating the future based on past and current data and provides information about potential future events. Forecasting helps organizations predict future revenues. There are many forecasting methods in use, but each organization chooses a forecasting technique suitable for its circumstances. The Theory of Rational Expectations, formulated by John F. Muth in the 1960s, proposes that individuals and organizations rely on all available information to make predictions and forecasts, integrating their expectations into their decision-making. When it comes to revenue forecasting in quick-service restaurants, the Theory of Rational Expectations indicates that organizations would employ logical approaches and up-to-date data to anticipate future revenue levels. This theory supports the study's definition of forecasting, which involves using rational methods and current data to project the future progression of revenue.

2.2.4 Use of Business Intelligence in Forecasting Revenues.

Business organizations have been using traditional methods of regression analysis to analyze past data and predict future outcomes. However, Business Intelligence solutions are bundled with are capable of providing real-time and accurate data. Accurate historical sales patterns and trends help in accurate future revenue forecasts and projections.

2.3 Empirical Literature

Smith et al. (2019) conducted a study to assess the influence of business intelligence tools on revenue forecasting accuracy in the hospitality industry. The findings indicated that the implementation of business intelligence resulted in a significant enhancement in the precision of revenue forecasting, leading to improved decision-making and resource allocation. However, it's worth noting that this study had a broader focus on the hospitality industry and did not specifically concentrate on quick-service restaurants. Moreover, the study did not employ a multivariate linear regression model.

In another investigation by Jones and Brown (2020) in the food service sector, the authors explored the correlation between business intelligence factors and revenue forecasting. Their research involved using a multivariate analysis technique but utilized a different modeling approach, such as time series forecasting. The study identified several noteworthy variables that had an impact on revenue forecasting accuracy, including customer demographics, menu pricing, and the effectiveness of marketing campaigns. However, similar to the previous study, this research did not exclusively target quick-service restaurants.

Despite these existing empirical studies, there remains a research gap in examining the application of business intelligence on revenue forecasting in quick-service restaurants using a multivariate linear regression model. This research gap primarily stems from the limited focus on quick-service restaurants in previous studies that have explored revenue forecasting within the broader hospitality or food service industry. Additionally, the specific impact of utilizing a multivariate linear regression model for revenue forecasting accuracy in quick-service restaurants is not adequately understood. Addressing these research gaps would provide valuable insights into the use of business intelligence for revenue forecasting in quick-service restaurants.

2.4 Summary

This chapter contains a review of literature that pertains to how business intelligence affects decision making. The primary areas discussed in this chapter include the Conceptual Framework, theoretical framework, and empirical review. In the subsequent chapter, the methodology used for the study will be presented.

CHAPTER 3

RESEARCH METHODOLOGY

3.0 Introduction

The preceding chapter analyzed the literature review, including information, theoretical framework, empirical evidence, and the study's justification. In contrast, this chapter provides an overview of the research methodology, which serves as a framework for conducting the study, focusing on the application of Business Intelligence in revenue forecasting. As described by Coyle (2003), research methodology pertains to the theoretical investigation of methods and ideologies in a specific area of knowledge. This chapter outlines the methodology used to address the research objectives and questions. The research design is defined as the starting point of the study, as it provides a blueprint for conducting the research while maximizing control over factors that can interfere with the validity of the findings. Quantitative and qualitative research methods are utilized to collect comprehensive information that addresses the research problem while taking into account the advantages and disadvantages of the data collection methods employed. The utilization of statistical packages such as R is crucial to performing data analysis from various sources and obtaining empirical results that achieve this study's research objectives.

3.1 Research Design

For this study, a quantitative research approach was utilized. As stated by Saunders et al. (2012), research approaches are categorized as either qualitative or quantitative, with data considered qualitative if it cannot be analyzed using mathematical techniques. In contrast, quantitative methods rely on facts and observable phenomena to establish relationships and deduce generalizations. The choice of a quantitative research approach was appropriate for this study, as it focused on the connection between two variables, namely business intelligence and revenue. Furthermore, the use of quantitative methods such as tables, graphs, frequencies, and percentages were crucial in analyzing the study's data, as noted by Beins and McCarthy (2012).

The selection of this research approach was also preferred because of its ability to test relationships between variables and to produce results that were easy to describe and explain using descriptive methods such as the use of tables, charts, and graphs.

The survey research strategy was chosen because it facilitates the collection of large amounts of data from various population sizes with ease and accuracy. Survey research is considered as scientific in nature, which enhances the accuracy of the results obtained. This approach also permits multiple instruments, such as questionnaires and interviews, to be used for data collection, which are both economical and straightforward to administer.

3.2 Population of the Study

The term 'population' refers to the complete set of individuals or objects possessing the characteristics being examined, as defined by Battacherjee (2012). In this specific study, the population comprised of 54 Simbisa Brands Ltd Pizza Inn quick service retail outlets that were located in different parts of the country. The tabular distribution of the retail outlets around the country is shown below. Distribution of Pizza Inn Restaurants.

Table 3.1 Distribution of Pizza Inn Restaurants.

PROVINCE	NUMBER
Harare	26
Mashonaland Central	1
Mashonaland West	4
Mashonaland East	2
Midlands	4
Bulawayo	6
Manicaland	4
Masvingo	3
Matabeleland North	2
Matabeleland South	2
Total	54

SOURCE: (Simbisa Brands Ltd, 2023: WWW.simbisa.co.zw)

3.3 Sampling

According to Battacherjee (2012), a sample is a portion of the population that is taken into consideration for purposes of obtaining observations and statistical inferences about the entire population. Essentially, the sample is intended to be representative of the population being studied. Selecting a sample involves determining which elements are to be included or excluded from the study. There are two primary methods for creating a sample: probability and non-probability sampling, as noted by Beins and McCarthy (2012). Probability sampling ensures that each element in the sample has an equal chance of being chosen among all the elements under study, and the probability is usually predetermined. In contrast, non-probability sampling selects elements with unequal chances of being included in the sample.

3.3.1 Sampling Procedure and Sample Size

As described by Cooper and Schindler (2014), non-probability sampling is a subjective approach to selecting a sample from a population in which the probability of selecting the elements is unknown. There are situations where non-probability sampling is preferred over probability sampling methods, particularly when probability sampling requires extensive planning and repeated callbacks to ensure all selected members are contacted, which can signify an increased cost and time commitment. Additionally, non-probability sampling techniques may effectively meet the study's objectives. Examples of non-probability sampling techniques include convenience sampling, purposive sampling, and snowball sampling. Convenience sampling involves selecting participants based on their accessibility and availability, as noted by Beins and McCarthy (2012). Purposive sampling involves selecting participants based on desirable characteristics or expert knowledge, hence referred to as Judgmental sampling. Finally, as highlighted by Beins and McCarthy (2012), snowball sampling involves selecting a research participant who identifies and recruits additional participants they know, which is advantageous in identifying hidden population.

To come up with the sample size, the researcher used the convenience sampling technique to draw a sample of 26 retail outlets because of time and cost constraints. This involved selecting retail outlets located in Harare CBD because of their strategic location and the volume of activity in these outlets. The variables of interest for these outlets were considered representative of the entire population of the retail outlets. Inductive statistics was used to draw inferences about the wider population based on sample data.

3.4 Research Instruments.

According to Saunders et al., (2009) research instruments are tools that are used to collect data in a systematic way. The data was collected from the GAAP point of sale application of Simbisa Brands. This data collection method involves gathering information on sales, purchases, inventory, and other business metrics from the GAAP system.

The authenticity of the data collected from the GAAP system were ensured through several measures. Firstly, the GAAP system was designed to capture and store transactional data in real-time, ensuring that the data collected was accurate and up-to-date. Additionally, the system has built-in checks and balances to detect and prevent errors, ensuring that the data was reliable and trustworthy.

To ensure the reliability of the data collected from the GAAP system, the researcher conducted data cleaning and validation procedures. This would involve identifying and correcting errors, inconsistencies, and outliers in the data to improve its accuracy and representativeness.

3.4.1 Validity of the Research Instrument

Marczyk et al. (2005, p. 455) emphasize the significance of validity in research as a means of verifying if the research instruments measure the intended variables accurately. The two major types of validity are content validity and construct validity, as identified by Ritchie and Lewis (2003, p. 123). Content validity is concerned with determining if research instruments have sufficient questions to fulfill the study's goals, which was addressed in this study by reviewing literature aligned with the research objectives and framing questions accordingly.

3.5 Discussion of Variables.

The variables of interest were month, gross sales, purchases, turnover, cost, customers, net sales, sales including value added tax, profit, inventory, marketing, employee efficiency and inflation as the independent variables (the predictors) and revenue, the dependent variable (the criterion). The study determined the role of Business Intelligence forecasting revenues Mwaura, C.N., 2017 (Doctoral dissertation, University of Nairobi).

The variables used in this study are:

"Month": This variable represents the month in which the data was collected. It is a categorical variable and is important in the study as it helps to identify any trends or seasonality in the data which may impact revenue forecasting.

"Gross sales": This variable represents the total revenue generated by the quick service restaurants during the month. It is an important variable in the study as it is the main dependent variable and is used to estimate the impact of the independent variables on revenue forecasting.

"Purchases": This variable represents the total amount spent by the business on goods and services during the month. It is an important variable in the study as it helps to estimate the cost of goods sold and identify any changes in the cost of inputs that may impact revenue forecasting.

"Turnover": This variable represents the difference between gross sales and purchases. It is an important variable in the study as it provides an indication of the profitability of the business during the month.

"Cost": This variable represents the total cost of goods sold during the month. It is an important variable in the study as it helps to estimate the gross profit margin, which is a key performance indicator for quick service restaurants.

"Customers": This variable represents the total number of customers served by the quick service restaurants during the month. It is an important variable in the study as it helps to estimate the average revenue per customer, which is a useful metric for understanding customer behavior and predicting future revenue.

"Net sales": This variable represents the total revenue generated by the quick service restaurants after deducting the cost of goods sold. It is an important variable in the study as it helps to estimate the net profit margin, which is a key performance indicator for quick service restaurants.

"sales_incl_vat": This variable represents the total revenue generated by the quick service restaurants including value-added tax (VAT). It is an important variable in the study as it helps to estimate the impact of VAT on revenue forecasting.

"Profit": This variable represents the net profit generated by the quick service restaurants during the month. It is an important variable in the study as it provides an indication of the overall profitability of the business.

"Inventory": This variable represents the value of the inventory held by the quick service restaurants at the end of the month. It is an important variable in the study as it helps to estimate the cost of goods sold and identify any changes in inventory levels that may impact revenue forecasting.

"Marketing": This variable represents the total amount spent on marketing and advertising during the month. It is an important variable in the study as it helps to estimate the impact of marketing on revenue forecasting.

"Employee efficiency": This variable represents the efficiency of the employees in the quick service restaurants during the month. It is an important variable in the study as it helps to estimate the impact of employee performance on revenue forecasting.

"Inflation": This variable represents the monthly inflation rate in Zimbabwe during the month. It is an important variable in the study as it helps to estimate the impact of inflation on revenue forecasting. Inflation can impact the cost of goods sold, employee costs, and other expenses, which in turn can affect revenue forecasting. Therefore, including inflation as an independent variable in the regression model can help to account for its impact on revenue forecasting.

3.6 Data sources

A data source is a location or system from which data is collected or obtained. It is a specific location where data is stored or generated, such as a database, spreadsheet, website, or application. Data sources can be internal to an organization, such as a company's customer relationship management (CRM) system or financial accounting software, or external, such as a public government database or social media platform. In order to analyze and use data effectively, it is important to know its source and ensure the accuracy and reliability of the data collected. For this study, the researcher relied on both primary and secondary data in the study. The researcher sought permission from Simbisa Limited to get data about the gross sales, purchases, turnover, cost, customers, net sales, sales, profit, inventory, employee efficiency. This data was obtained from the statistics department of Simbisa Brands Ltd based at the Head Office. Data on annual inflation was obtained from the Reserve Bank of Zimbabwe (RBZ) website.

"Month": This variable was obtained from the date column of the data collected during the study.
"Gross sales": This variable was obtained from a POS (Point of Sale) system called GAAP used by Simbisa Brands. The GAAP system is a computerized system used to record sales transactions. It is important for the study as it provides real-time sales data, including the number of products sold, their prices, and the time and date of the sale. These data points are used to calculate the total gross sales for the month.

"Purchases": This variable was obtained from the GAAP system of Simbisa Brands. The GAAP system is also used to record purchases made by the business. The purchases figure represents the total amount spent by the business on goods and services during the month and was extracted from the GAAP system.

"Turnover": This variable was calculated as the difference between gross sales and purchases. The values for gross sales and purchases were obtained from the GAAP system of Simbisa Brands.

"Cost": This variable was obtained from the GAAP system of Simbisa Brands. The cost figure represents the total cost of goods sold during the month and was extracted from the GAAP system. The cost of goods sold is an important factor in determining the profitability of a business, and the GAAP system provides an accurate and reliable source of information on this metric.

"Customers": This variable was obtained from the POS system called GAAP used by Simbisa Brands. The number of customers served by the quick service restaurants during the month was extracted from the GAAP system. The POS system provides real-time sales data, including the number of items sold, their prices, and the time and date of the sale. These data points were used to calculate the total number of customers served during the month.

"Net sales": This variable was calculated as the difference between gross sales and cost. The values for gross sales and cost were obtained from the GAAP system of Simbisa Brands.

"sales_incl_vat": This variable was calculated by multiplying the gross sales figure by the value-added tax (VAT) rate of 15%. The gross sales figure was obtained from the GAAP system of Simbisa Brands.

"Profit": This variable was calculated as the net income after deducting all expenses from the gross sales. The values for gross sales, cost, and other expenses were obtained from the GAAP system of Simbisa Brands.

"Inventory": This variable was obtained from the inventory management system of Simbisa Brands. The inventory management system is a computerized system used to track inventory levels and manage stock. The value of the inventory held by the quick service restaurants at the end of the month was extracted from the inventory management system. It is important to note that the inventory management system is a separate system from the POS system called GAAP used to record sales transactions.

"Marketing": This variable was obtained from the marketing department of Simbisa Brands. The total amount spent on marketing and advertising during the month was recorded by the marketing department and provided for the study.

"Employee efficiency": This variable was obtained from the human resources department of Simbisa Brands. The efficiency of the employees in the quick service restaurants during the month was measured using performance metrics such as attendance, productivity, and customer feedback. The data on employee efficiency was collected by the human resources department and provided for the study.

"Inflation": This variable was obtained from the Reserve Bank of Zimbabwe (RBZ) website. The RBZ is the central bank of Zimbabwe and is responsible for implementing monetary policy in the country. The monthly inflation rate in Zimbabwe during the month was obtained from the RBZ website, which publishes regular updates on inflation and other economic indicators.

In summary, the variables related to profitability or sales (gross sales, purchases, cost, net sales, sales_incl_vat, and profit) were obtained from the POS system called GAAP used by Simbisa

Brands. The GAAP system is a computerized system used to record sales transactions, while the inventory management system was used to obtain the inventory value. The variables related to inventory, marketing, employee efficiency, and inflation were obtained from other sources such as the inventory management system, marketing department, human resources department, and RBZ website, respectively.

3.7 Data Analysis and Presentation

Data analysis entails modeling the data with the aim of determining the relationship to support the research conclusion whereas data presentation involves displaying analyzed statistical information through charts, graphs and numbers to facilitate and aid comprehension of trends and patterns. The model used in this study is expressed mathematically as follows:

$$Y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \cdots + b_nx_n$$

Where : Y = the estimated y-value computed from the regression equation and b_0, b_1, b_2, b_3, and so on, are called the regression coefficients. These represent the weights that measure the relative importance (or strength of relationship) between each independent variable x_n, and the dependent variable, y.

3.7.1.0 Multiple linear regression analysis

Regression analysis is a statistical technique for estimating the relationship among variables which have a cause-and-effect relationship. Multiple linear regression, describes the simultaneous associations of several variables with one continuous outcome. Important steps in using this approach include estimation and inference, variable selection in model building, and assessing model fit.

3.7.1 How the data analysis will answer the objectives.

To examine the role of Business Intelligence systems on Revenue forecasting using multiple linear regression, the researcher used revenue as the dependent variable, the use of Business Intelligence systems as the independent variable, and other relevant variables such as marketing, employee efficiency, and inflation as control variables. The researcher estimated the coefficients of the independent and control variables to determine the impact of Business Intelligence systems on revenue forecasting while controlling for other factors that may influence revenue.

To evaluate the impact of demand on revenue using multiple linear regression, the researcher used the revenue as the dependent variable, the level of demand as the independent variable, and other relevant variables such as marketing, employee efficiency, and inflation as control variables. The researcher estimated the coefficients of the independent and control variables to determine the impact of demand on revenue while controlling for other factors that may influence revenue.

To examine the impact of price on revenue using multiple linear regression, the researcher used the revenue as the dependent variable, the price of the product as the independent variable, and other relevant variables such as marketing, employee efficiency, and inflation as control variables. The study will estimate the coefficients of the independent and control variables to determine the impact of price on revenue while controlling for other factors that may influence revenue.

To examine the impact of inflation on revenue using multiple linear regression, the researcher used the revenue as the dependent variable, the inflation rate as the independent variable, and other relevant variables such as marketing, employee efficiency, and demand as control variables. I will estimate the coefficients of the independent and control variables to determine the impact of inflation on revenue while controlling for other factors that may influence revenue.

In addition to multiple regression, the researcher also used simple linear regression to examine the relationship between each independent variable and the dependent variable individually. For example, we can use simple linear regression to estimate the impact of purchases on revenue, the impact of inventory on revenue, and so on. This can help us understand the individual impact of each variable on revenue.

However, multiple regression is a more comprehensive approach as it allows us to estimate the impact of multiple independent variables on the dependent variable while controlling for other relevant factors. By including control variables in the model, the researcher can isolate the impact of the independent variable of interest and obtain more accurate estimates of its effect on the dependent variable.

Using both simple and multiple regression can provide a more complete understanding of the relationship between the independent variables and the dependent variable, and can help us make more informed decisions. By justifying the inclusion of control variables and using both simple and multiple regression, we can ensure that our analysis is rigorous and that our estimates are reliable.

3.7.2 Data profiling

Data profiling is an important step in any data analysis project, including this research. Data profiling involves analyzing and understanding the data to identify any potential issues or anomalies that may impact the accuracy or reliability of the analysis. Some of the key benefits of data profiling include:

Understanding the structure and content of the data: Data profiling helps to identify the structure and content of the data, such as the data types, formats, and values. This information is important for choosing appropriate data analysis methods and ensuring that the data is correctly interpreted.

Identifying data quality issues: Data profiling can help to identify data quality issues, such as missing values, duplicates, or inconsistencies. Addressing these issues early in the analysis process can help to improve the accuracy and reliability of the analysis.

Enhancing data transparency: Data profiling can help to enhance the transparency of the data analysis process by providing a clear understanding of the data and any issues that may impact the analysis. This can help to increase trust in the analysis results.

The following syntax was used to profile the dataset;

```
library(skimr)

# View the structure of the data
str(data)

# Generate a summary of the data
summary(data)
skim(data)
#Use the skim function from the skimr package to generate a data profile


# Load the necessary libraries
library(DataExplorer)

# Generate a data profile using the DataExplorer package
create_report(data)
```

3.8 Model Assumptions

i. There is a linear relationship between the independent and dependent variables.

ii. The error component is normally distributed.

iii. There is no multicollinearity between the predictor variables.

iv. There is no heteroskedasticity. The variance of the residuals must be constant across the predicted values.

3.9 Summary

The research methodology was providing a blueprint for other studies to follow since the chapter clearly illustrates research procedures, techniques and tools used in the study. The chapter's structure was determined by the recommendations given by earlier research that were covered in the second chapter. Research design, research tools, and data sources used for statistical tests are among the chapters' highlights. The chapter also discussed data analysis techniques, including the statistical software chosen to produce descriptive and inferential output, which serves as the foundation for answering the study's goals. The chapter addressed models that were utilized by the researcher to produce inferential and descriptive output of information acquired from both primary and secondary sources in addition to data analysis processes. The research technique opens the door for the following chapter, which will concentrate on discussion, data presentation, and result analysis before the last chapter of the research's summary, conclusions, and recommendations.

CHAPTER 4

DATA PRESENTATION, ANALYSIS AND DISCUSSIONS

4.0 Introduction

This chapter focuses on the analysis of the data collected for the study. In this chapter, the writer will use descriptive statistics to summarize the characteristics of the data, perform exploratory data analysis to gain insights into the relationships between variables, and build a multivariate linear regression model to examine the impact of business intelligence variables on revenue forecasting in quick service restaurants. The chapter begins with an introduction that provides an overview of the analysis to be conducted and outlines the specific objectives to be achieved.

Data format

```
str(data)
tibble [60 × 13] (S3: tbl_df/tbl/data.frame)
 $ month              : chr [1:60] "Jan" "Feb" "Mar" "Apr" ...
 $ gross_sales        : num [1:60] 4439524 4769823 6558708 5070508 5129288 ...
 $ purchases          : num [1:60] 3765748 3148374 3266755 2786997 2749746 ...
 $ turnover           : num [1:60] 673777 1621449 3291953 2283511 2379542 ...
 $ cost               : num [1:60] 2058823 1526263 1754721 1871954 2921931 ...
 $ customers          : num [1:60] 3937 6263 4650 4134 4764 ...
 $ net_sales          : num [1:60] 2380701 3243560 4803987 3198554 2207357 ...
 $ sales_incl_vat     : num [1:60] 5105453 5485296 7542515 5831085 5898681 ...
 $ profit             : num [1:60] 476140 648712 960797 639711 441471 ...
 $ inventory          : num [1:60] 842276 899560 1299212 772539 964190 ...
 $ marketing          : num [1:60] 428476 424731 406146 394749 456284 ...
 $ employee_efficiency: num [1:60] 0.779 0.865 0.827 0.902 0.882 ...
 $ inflation          : num [1:60] 0.3 0.1 -0.3 0.1 0 0 1 0.4 0.9 16.4 ...
```

The output of str(data) provides information about the structure of the data frame, which contains 60 observations (rows) and 13 variables (columns). Each variable has a different data type: month is a character variable, while the others are numeric variables with values of class num. The output of summary(data) provides summary statistics for each variable, such as the minimum, maximum, median, and quartiles. The skim(data) function provides a more comprehensive data profile that includes additional information such as the number of missing values, the number of unique values, and more detailed statistics on the distribution of the data.

Overall, this data set contains information on the sales, expenses, profitability, and other factors affecting the performance of a business over a period of 5 years. This data can be used to conduct various types of analyses, such as forecasting future sales or identifying factors that contribute to profitability.

Data summary

The output of summary(data) provides summary statistics for each variable in the data frame data.

```
> summary(data)
    month            gross_sales        purchases          turnover             cost           customers        net_sales
 Length:60          Min.   :3033383   Min.   :1883582   Min.   :-418941   Min.   : 973376   Min.   :3683   Min.   : 663409
 Class :character   1st Qu.:4506446   1st Qu.:3058226   1st Qu.: 763364   1st Qu.:1662852   1st Qu.:4281   1st Qu.:2696610
 Mode  :character   Median :5020981   Median :3431941   Median :1497882   Median :1973656   Median :4749   Median :3139479
                    Mean   :5065617   Mean   :3475686   Mean   :1589931   Mean   :1992534   Mean   :4963   Mean   :3073083
                    3rd Qu.:5691819   3rd Qu.:3826206   3rd Qu.:2392864   3rd Qu.:2290692   3rd Qu.:5558   3rd Qu.:3478538
                    Max.   :7168956   Max.   :5031133   Max.   :3912236   Max.   :3620520   Max.   :7199   Max.   :5064469
 sales_incl_vat        profit           inventory          marketing      employee_efficiency   inflation
 Min.   :3488390   Min.   : 132682   Min.   : 648695   Min.   :253410   Min.   :0.5357        Min.   :-0.300
 1st Qu.:5182413   1st Qu.: 539322   1st Qu.: 919355   1st Qu.:427062   1st Qu.:0.7399        1st Qu.: 2.513
 Median :5774128   Median : 627896   Median :1016376   Median :498713   Median :0.7949        Median : 5.473
 Mean   :5825460   Mean   : 614617   Mean   :1038746   Mean   :497898   Mean   :0.7890        Mean   : 9.840
 3rd Qu.:6545592   3rd Qu.: 695708   3rd Qu.:1175573   3rd Qu.:571684   3rd Qu.:0.8497        3rd Qu.:15.761
 Max.   :8244299   Max.   :1012894   Max.   :1458616   Max.   :757146   Max.   :1.0417        Max.   :39.258
```

By examining these summary statistics, I can get a sense of the range and distribution of values for each variable in the data set. For example, I can see that the median net sales for a given month is around ZWL 3.1 million, but the range of net sales is quite broad, with a minimum of ZWL 663,409 and a maximum of ZWL 5.1 million. I can also see that the median profit for a given month is around ZWL 628,000, but again there is a wide range of profit values, with a minimum of ZWL 132,682 and a maximum of ZWL 1.01 million. Similarly, I can examine the summary statistics for other variables to get a better understanding of the data and identify any potential issues or anomalies that may affect our analysis.

Profiling using the skim package

```
> skim(data)
── Data Summary ────────────────────────
                           Values
Name                       data
Number of rows             60
Number of columns          13
_______________________
Column type frequency:
  character                1
  numeric                  12
________________________
Group variables            None

── Variable type: character ──────────────────────────────────────────────────────
  skim_variable n_missing complete_rate min max empty n_unique whitespace
1 month                 0             1   3   3     0       12          0

── Variable type: numeric ────────────────────────────────────────────────────────
   skim_variable       n_missing complete_rate       mean          sd          p0        p25        p50        p75       p100 hist
 1 gross_sales                 0             1 5065617.     910376.    3033383.   4506446.   5020981.   5691819.   7168956.
 2 purchases                   0             1 3475686.     618087.    1883582.   3058226.   3431941.   3826206.   5031133.
 3 turnover                    0             1 1589931.    1041523.    -418941.    763364.   1497882.   2392864.   3912236.
 4 cost                        0             1 1992534.     527503.     973376.   1662852.   1973656.   2290692.   3620520.
 5 customers                   0             1    4963.        915.       3683.      4281.      4749.      5558.      7199.
 6 net_sales                   0             1 3073083.     955803.     663409.   2696610.   3139479.   3478538.   5064469.
 7 sales_incl_vat              0             1 5825460.    1046933.    3488390.   5182413.   5774128.   6545592.   8244299.
 8 profit                      0             1  614617.     191161.     132682.    539322.    627896.    695708.   1012894.
 9 inventory                   0             1 1038746.     193337.     648695.    919355.   1016376.   1175573.   1458616.
10 marketing                   0             1  497898.     109757.     253410.    427062.    498713.    571684.    757146.
11 employee_efficiency         0             1       0.789       0.0919      0.536      0.740      0.795      0.850      1.04
12 inflation                   0             1       9.84       10.3        -0.3        2.51       5.47      15.8       39.3
```

The output of skim(data) provides a summary of the data frame data. By examining these summary statistics, I can get a sense of the range and distribution of values for each variable in the data set. For example, I can see that the variable gross sales has a mean of around 5,065,617 ZWL, with a standard deviation of approximately 910,376 ZWL, and a range of values from 3,033,383 ZWL to 7,168,956 ZWL. Similarly, I can examine the summary statistics for other variables to get a better understanding of the data and identify any potential issues or anomalies that may affect our analysis. The skim() function from the skimr package generates a comprehensive summary of the data frame that includes a wide range of summary statistics and visualizations, making it a useful tool for data profiling and exploratory data analysis.

Table 4.1 Profiling using the data Explorer package

Basic Statistics

Raw Counts

Name	Value
Rows	60
Columns	13
Discrete columns	1
Continuous columns	12
All missing columns	0
Missing observations	0
Complete Rows	60
Total observations	780
Memory allocation	10.3 Kb

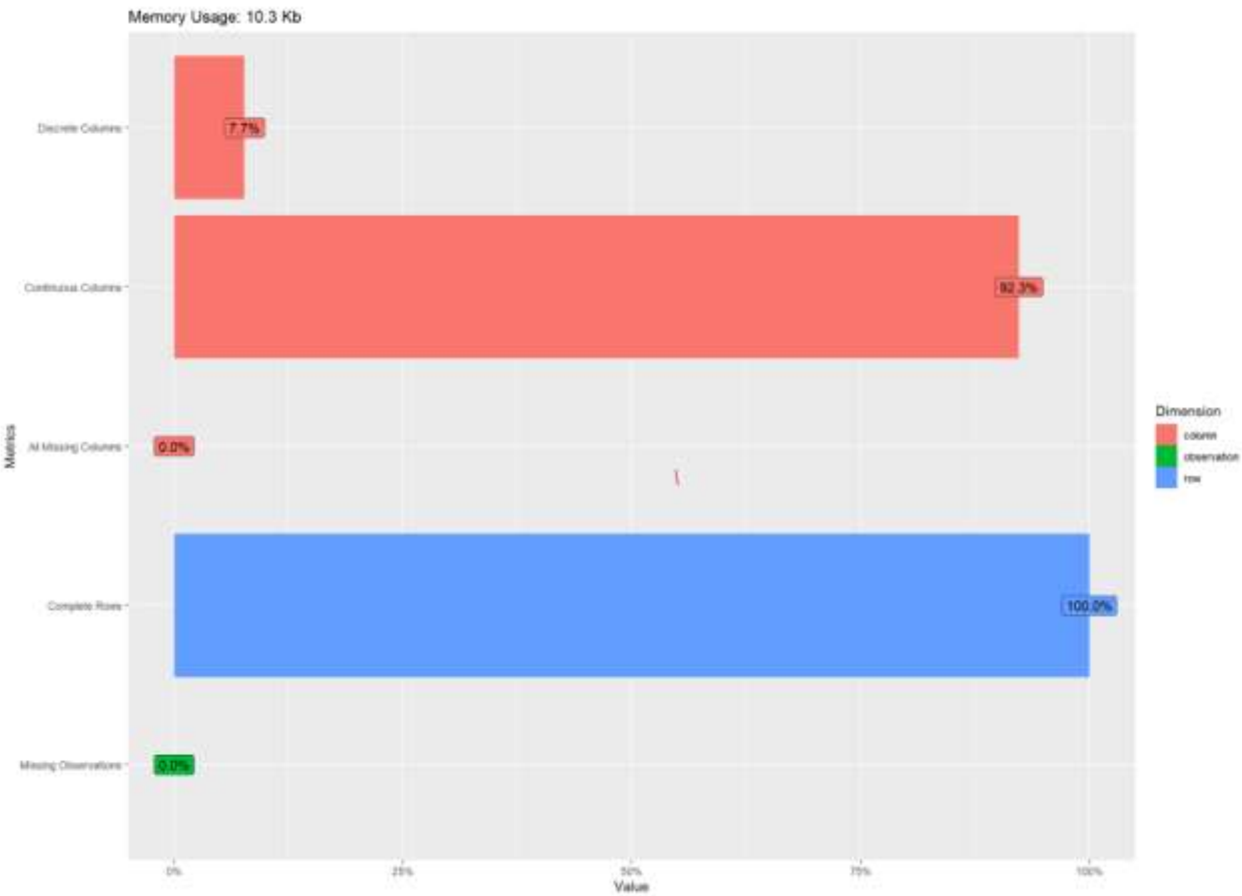
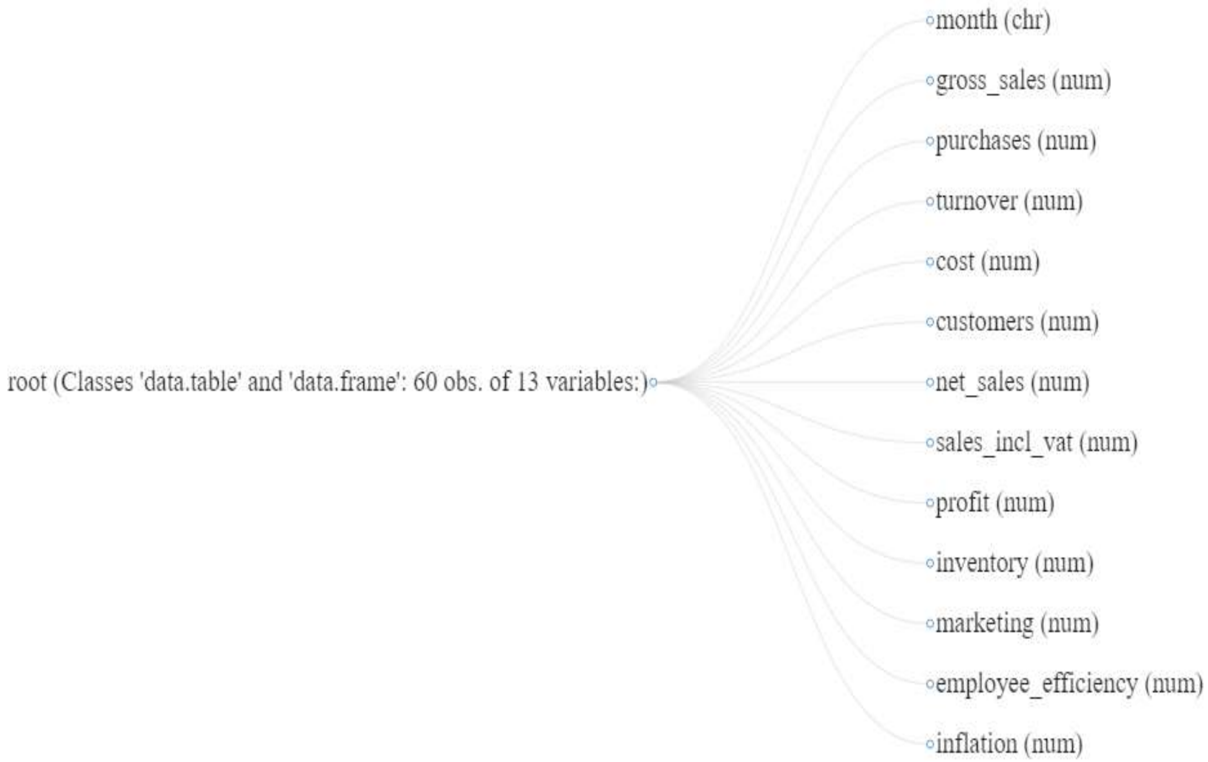
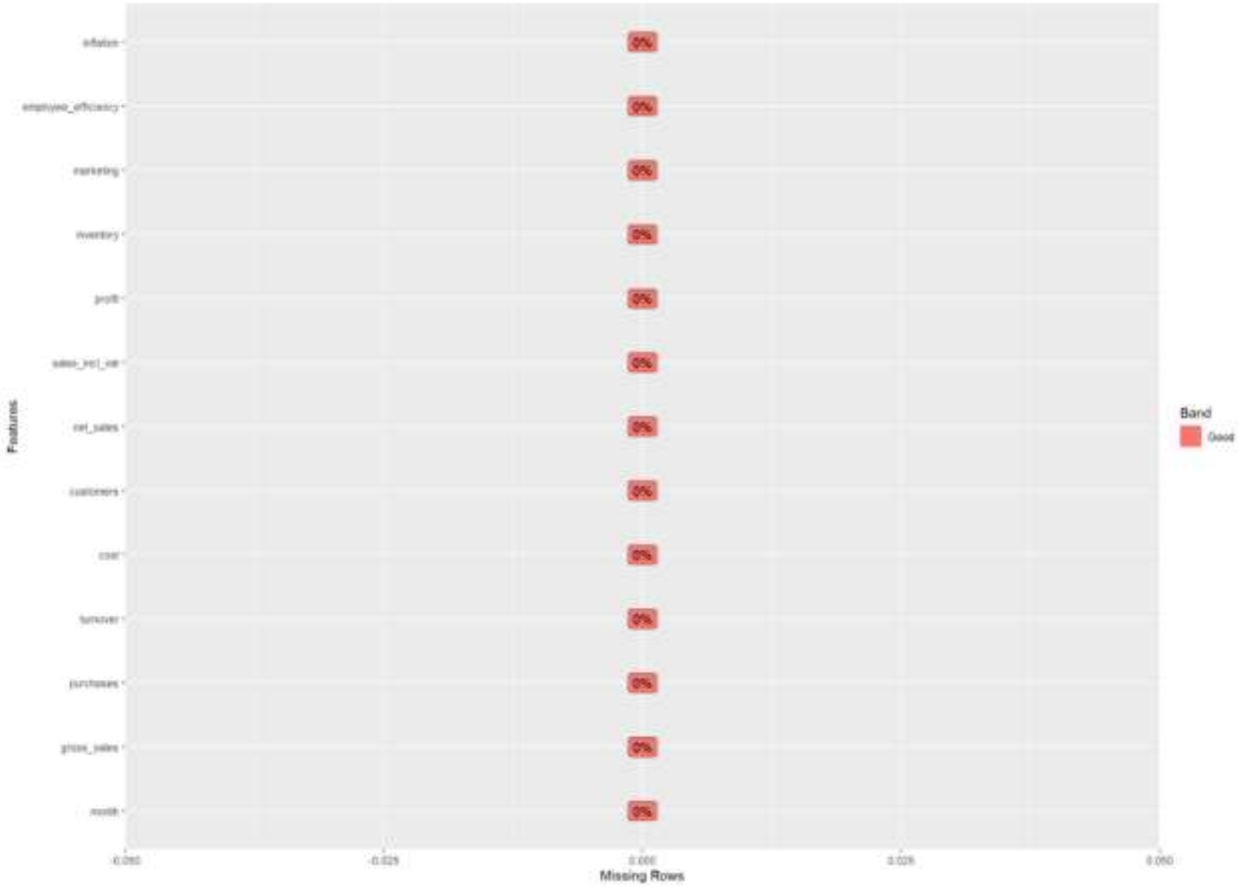


Figure 4.1 Percentages

Data Structure



Missing Data Profile



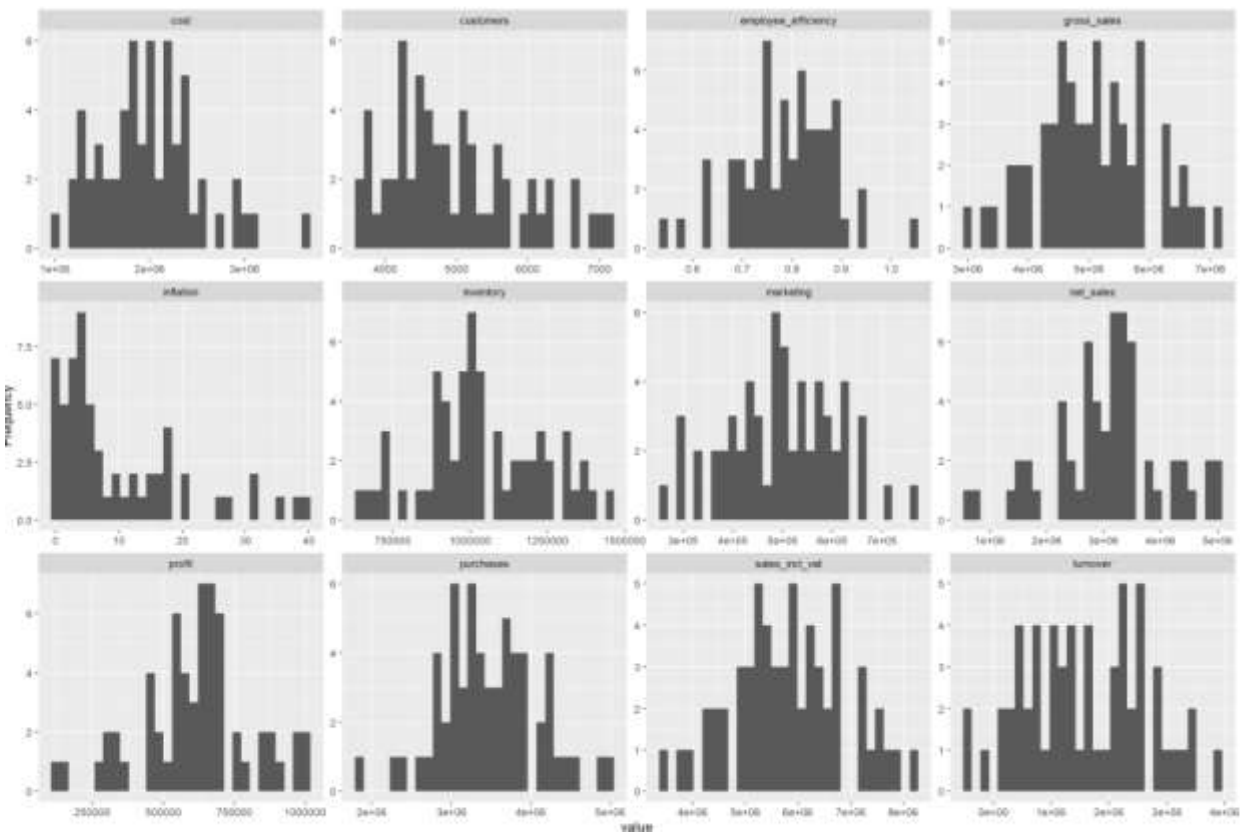


Figure 4.2 Uni variate Distribution- Histogram

This histogram is a way to visualize the distribution of a single variable. In this case, it would allow the researcher to see how the revenue variable is distributed on a monthly basis, which can help to identify any patterns or trends in the data. When examining the use of business intelligence on revenue forecasting using variables like gross sales, purchases, turnover, cost, customers, net sales, profit, inventory, marketing, employee efficiency and inflation as independent variables, a univariate distribution histogram can be used to explore the distribution of each variable individually. It can illuminate the range of values for each variable, the frequency of those values, and how they are represented in the data. By examining these distributions, the researcher can better understand the relationships between the independent variables and the dependent variable (revenue) on a monthly basis. Overall, using a univariate distribution histogram can aid in the examination of the impact of business intelligence on revenue forecasting and provide valuable insights into how each of the independent variables may be contributing to the monthly revenue figures.

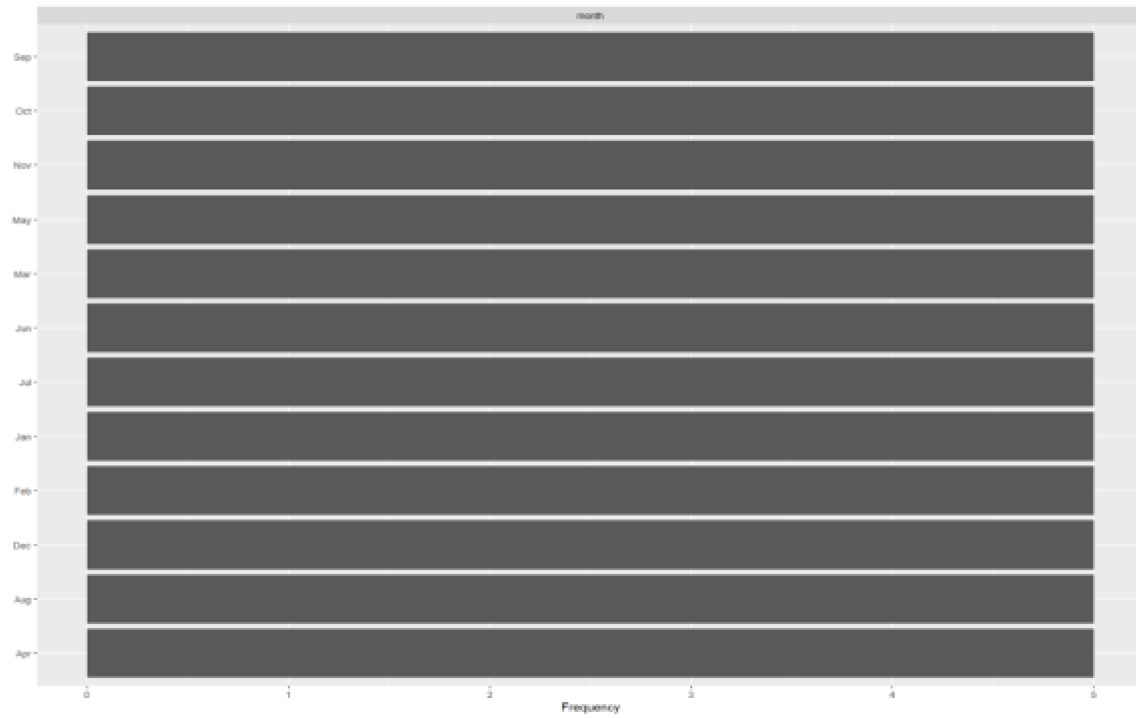
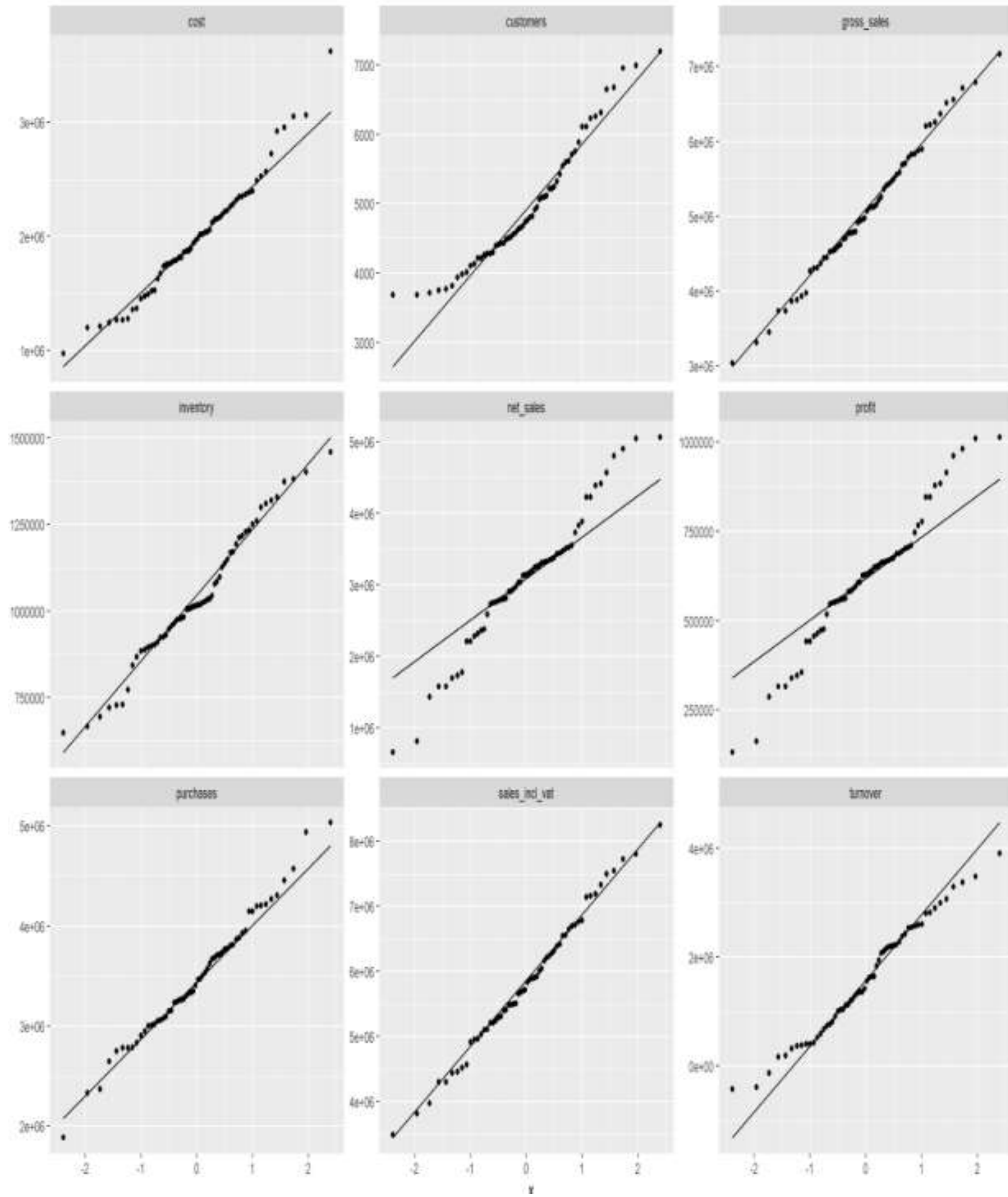


Figure 4.3 Uni variate Distribution-Bar Chart (with frequency)

A univariate distribution bar chart with frequency can provide a visual representation of the frequency of a single variable. In this case, it can help to illustrate the distribution of monthly revenue figures and give insight into revenue forecasting when using business intelligence tools. By examining the bar chart, you can see the range of revenue figures and how frequently each figure occurs on a monthly basis. This can give you a sense of the variability in revenue and highlight potential trends over time. Additionally, you can compare the frequency of revenue figures with the presence of certain independent variables, such as gross sales, purchases, turnover, cost, customers, net sales, profit, inventory, marketing, employee efficiency, and inflation, to identify which variables may be most closely associated with monthly revenue. Overall, a univariate distribution bar chart, with frequency, can be a helpful tool in examining the use of business intelligence on revenue forecasting. It can provide an intuitive way to visualize and analyze data, and identify potential relationships between variables that may be useful in forecasting future revenue.


cost
customers
gross_sales
inventory
net_sales
profit
purchases
sales_incl_vat
turnover

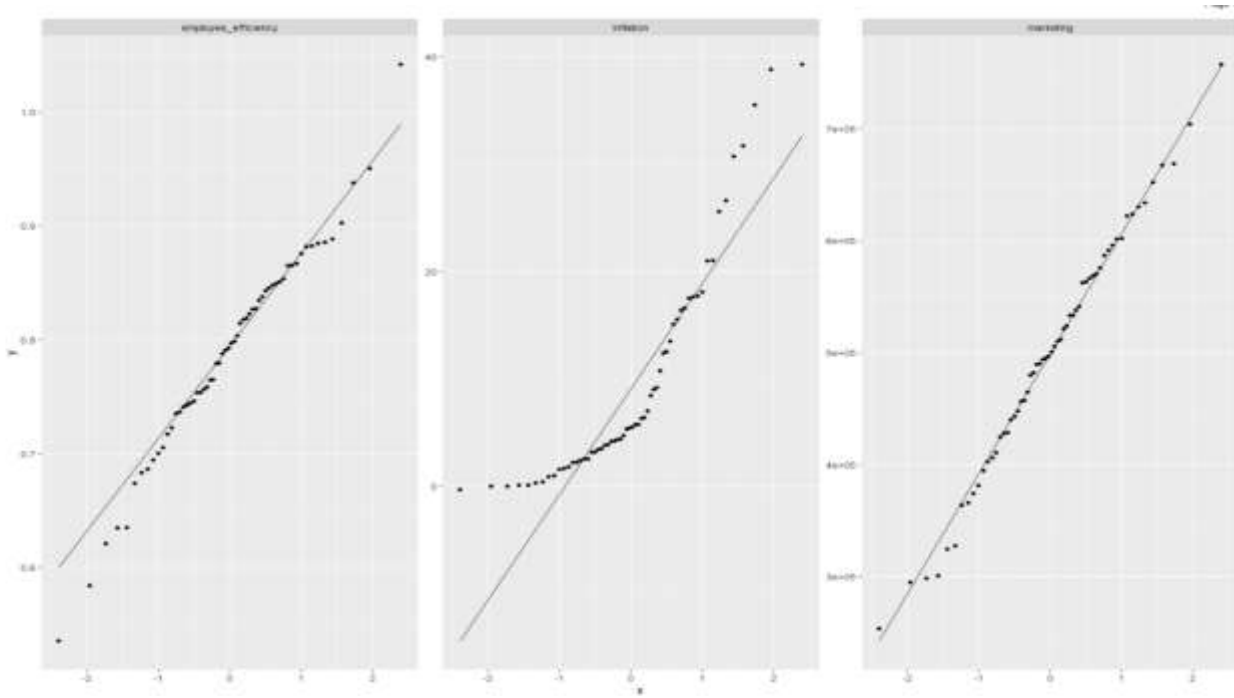


Figure 4.4 Uni variate Distribution-QQ Plot

A univariate distribution QQ plot is a graphical tool that can be used to compare the distribution of a sample dataset (in this case, the monthly revenue) to a theoretical distribution (such as a normal distribution). When examining the use of business intelligence on revenue forecasting, a QQ plot can be helpful in detecting any deviations from a normal distribution, which can impact the effectiveness of using business intelligence tools for revenue forecasting. By examining the QQ plot, you can compare the monthly revenue data to a reference normal distribution. If the data points fall along the diagonal line of the plot, it suggests that the data is normally distributed. However, if the points show a deviation from the diagonal line, then it suggests that the distribution of the data is different from a normal distribution, and this needs to be accounted for when using business intelligence tools for forecasting. Additionally, this technique can be used to identify influential data points or outliers that may exist in the dataset. For example, if any data points fall far from the diagonal line, this could indicate that there are extreme values within the dataset that require further examination. In terms of examining the use of business intelligence on revenue forecasting using variables such as gross sales, purchases, turnover, cost, customers, net sales, profit, inventory, marketing, employee efficiency, and inflation, the QQ plot can reveal whether or not the use of these variables will result in normally distributed data, which is a necessary assumption when using some statistical models for forecasting. Overall, the univariate distribution QQ plot can be a valuable tool in analyzing the distribution of monthly revenue data and detecting any deviations from a normal distribution that may impact the efficacy of using business intelligence for revenue forecasting.

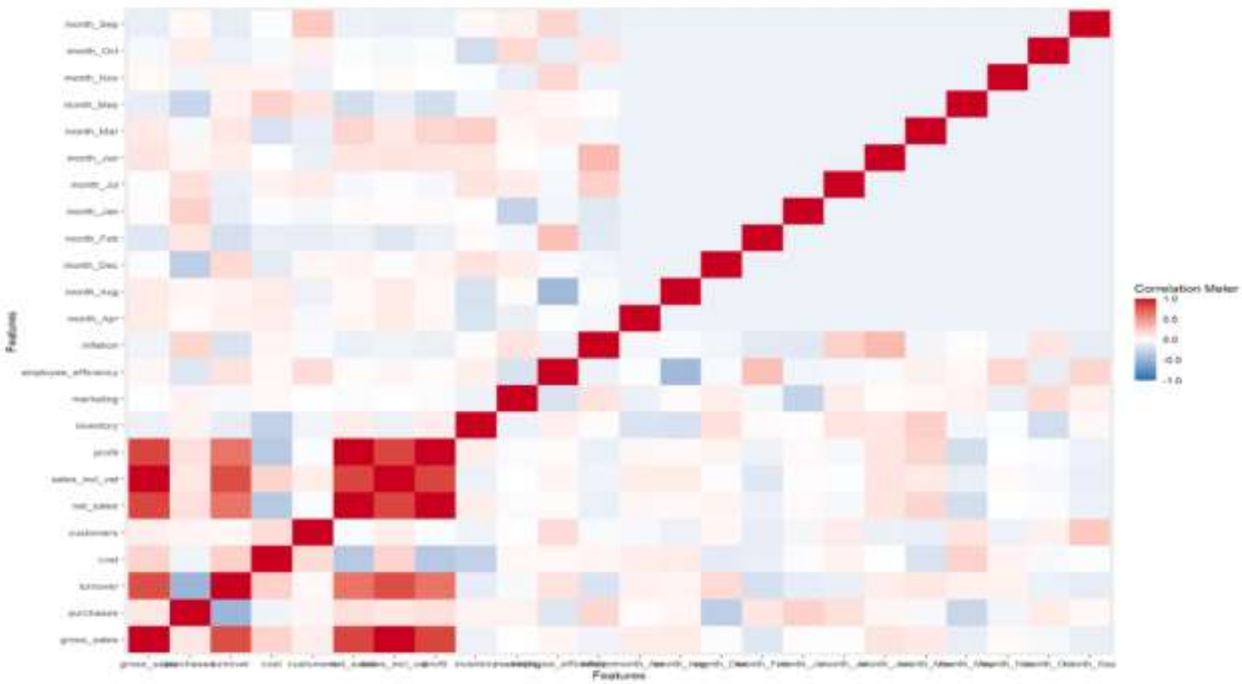


Figure 4.5 Correlation Analysis

A heat map for correlation analysis is a visual representation that displays the correlation strength and direction between different variables in a dataset. It is represented by a matrix of colored squares, where each square corresponds to a pair of variables and reflects the correlation coefficient among them. A positive correlation is shown by warm colors like red or orange, while negative correlations are represented by cool colors like green or blue. Examining a correlation analysis heat map necessitates assessing the correlation strength and direction of the variables, which can range from -1 to 1. A value of -1 implies a perfect negative correlation, 0 indicates no correlation, and 1 indicates a perfect positive correlation. Positive correlations suggest that the variables tend to increase or decrease together, while negative correlations suggest that the variables tend to move in opposite directions.

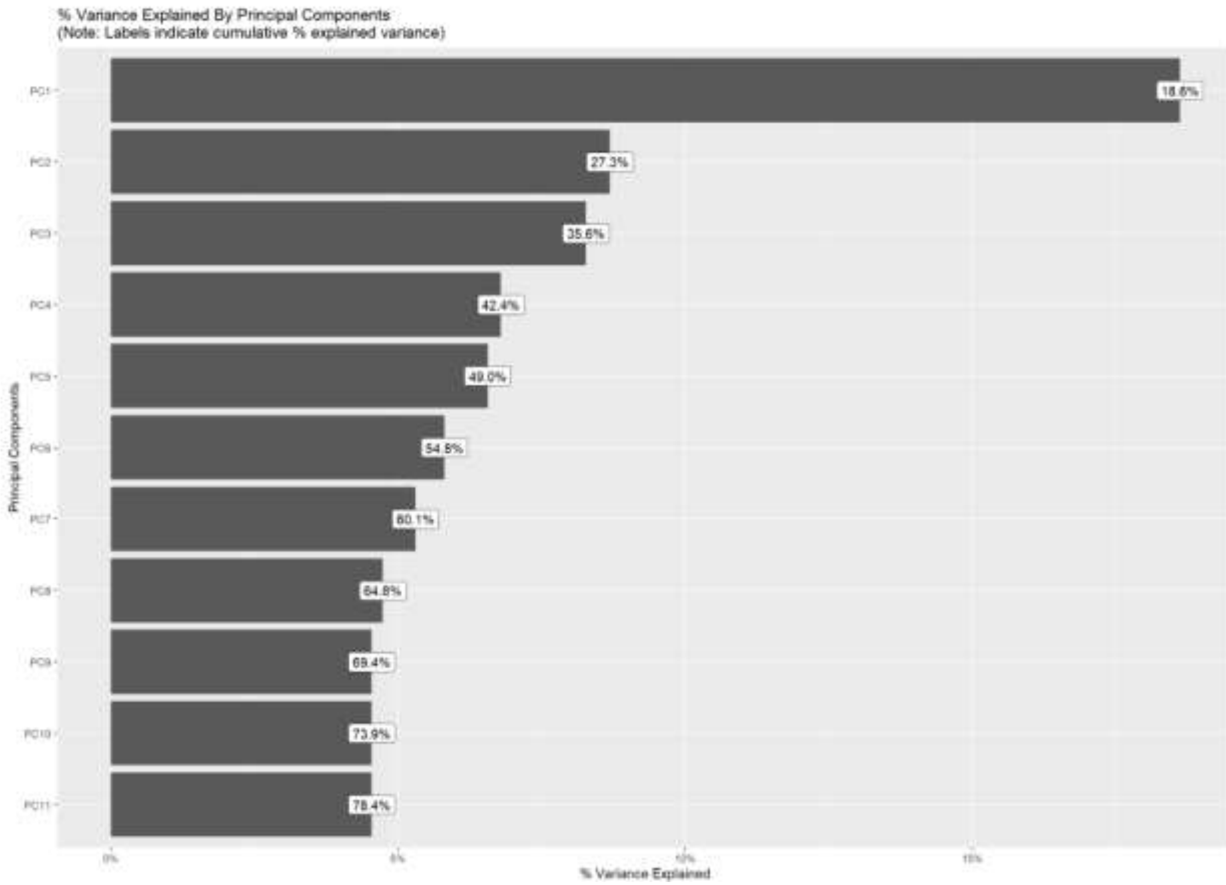
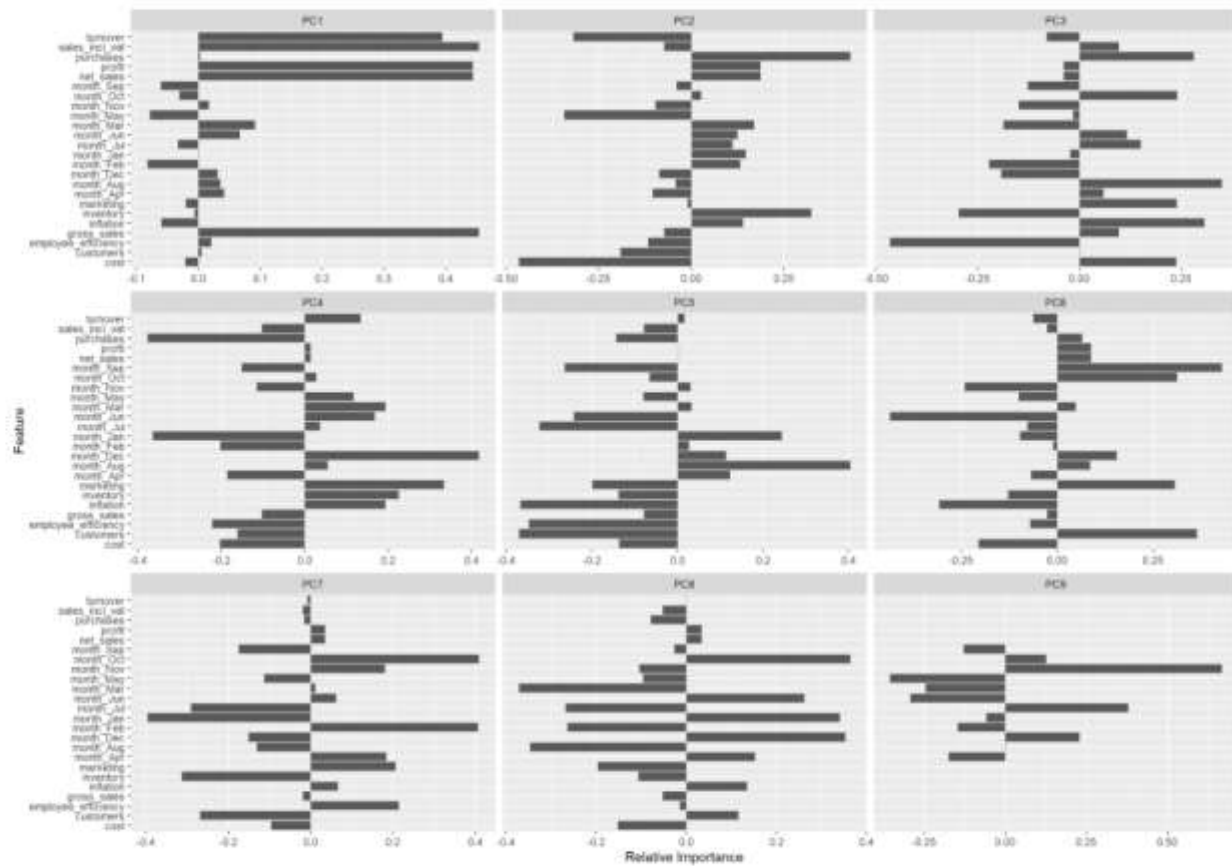
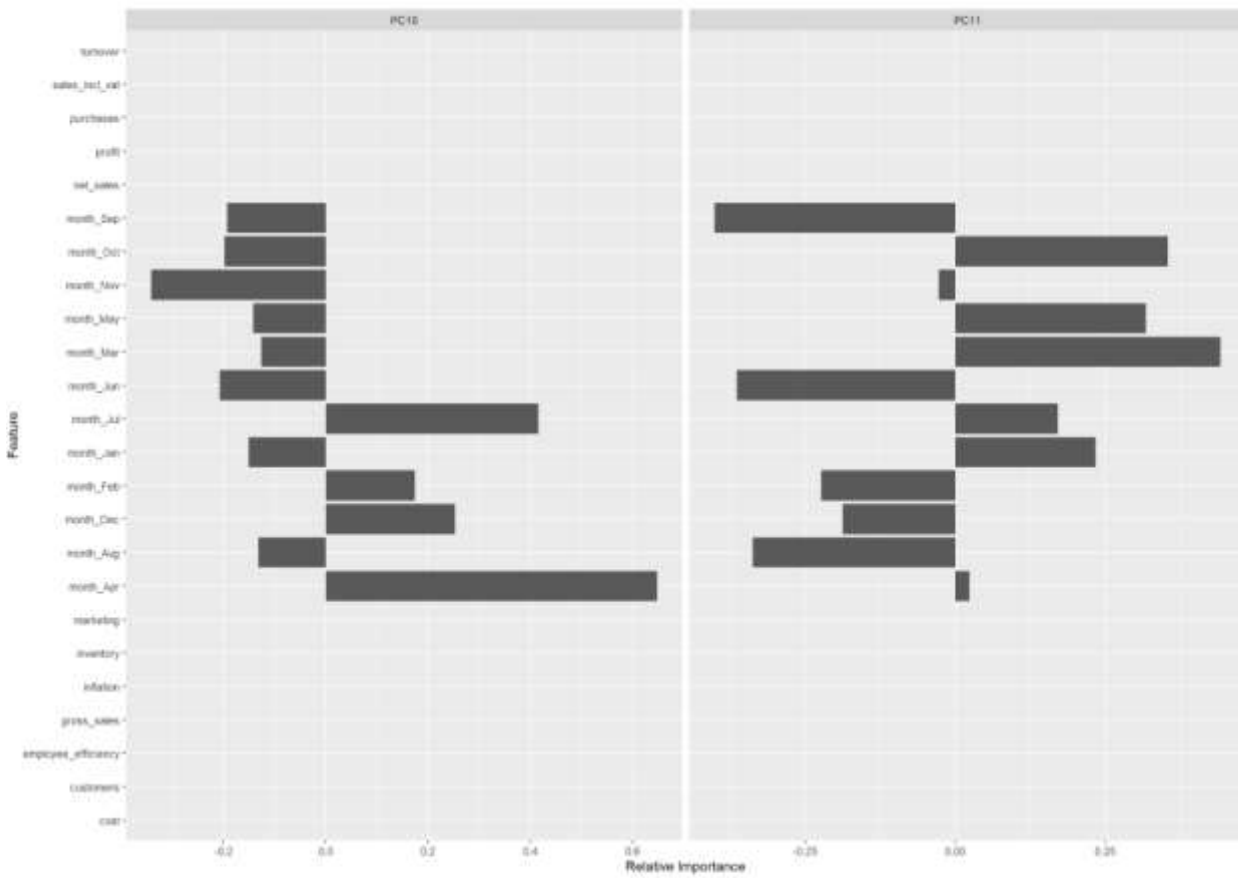


Figure 4.6 Principal Component Analysis

Principal Component Analysis (PCA) refers to a technique for reducing the dimensions of high-dimensional datasets in a bid to uncover patterns. Its objective is to detect principal components, which constitute a fresh collection of variables that capture the vast portion of the original data's variability while simultaneously diminishing the number of dimensions. PCA was used as a form of data preprocessing to reduce the dimensionality of high-dimensional datasets before applying other machine learning algorithms. It can also be used for exploratory data analysis to identify underlying patterns or relationships in the data.





4.2.0 Summary of data profiling

The summary of the data profiling from the DataExplorer package indicates that:

i. There are no missing values in the dataset.

ii. The Q-Q plots for the numeric variables show a normal distribution, suggesting that the data is normally distributed.

iii. The correlation heat map shows evidence of high correlation between some of the variables, indicating that multicollinearity may be an issue in the analysis.

Overall, the lack of missing values and normal distribution of the variables are positive indicators for the quality of the data. However, the presence of high correlation between some of the variables suggests that multicollinearity may be an issue in the analysis. Multicollinearity can cause problems in interpreting the impact of individual predictors on the outcome variable, and it may result in unstable and inaccurate estimates of the regression coefficients. As such, it is necessary to be mindful of the risks posed by multicollinearity and take measures to address it during analysis. These steps might involve removing one of the highly correlated variables or utilizing approaches like ridge regression or lasso regression to regulate it.

4.2.0.0 Data wrangling and manipulation

When dealing with multicollinearity in the data, there are several data wrangling and manipulation techniques that can be used to address the issue:

Remove one of the highly correlated variables: An option that can be utilized is to eliminate one of the variables with significant correlation from the analysis. This can be done based on domain knowledge or statistical significance tests. For example, if two variables are highly correlated but one is deemed less important for the analysis, it can be removed to reduce the multicollinearity.

Combine the highly correlated variables: A viable alternative is to merge the correlated variables into one by establishing a new variable that captures the underlying construct associated with the correlated variables. This can be achieved through calculating the average or sum of the variables, or by developing a brand-new variable that embodies the construct in question.

Regularization techniques: Regularization methods such as ridge regression or lasso regression can be employed to decrease the influence of multicollinearity on the regression coefficients. These techniques add a penalty term to the regression model that shrinks the coefficients towards zero, effectively reducing the impact of the correlated variables on the outcome variable.

Principal Component Analysis (PCA) is a technique used for reducing the number of dimensions in a data set. It can change correlated variables into a smaller set of uncorrelated variables referred to as principal components. These new components can be employed as predictors in the regression analysis, resulting in a decrease in the influence of multicollinearity on the regression coefficients.

In this research I am going to eliminate one of the highly correlated variables which were shown by the correlation heat map. In order for us to do so we need to first know the extend on the correlation so we have the correlation plot with figures. I have used the following code to obtain that:

```
# Check for multi-collinearity
cor(data[, c("gross_sales", "purchases", "turnover", "cost", "net_sales",
             "sales_incl_vat", "profit", "inventory", "marketing",
             "inflation","employee_efficiency","inflation")])
# Calculate correlation matrix
cor_matrix <- cor(data[, c("gross_sales", "purchases", "turnover", "cost",
                           "net_sales", "sales_incl_vat", "profit","inflation",
                           "inventory", "marketing", "employee_efficiency")])
library(ggcorrplot)
# Create correlation plot
ggcorrplot(cor_matrix,
           type = "full",
           colors = c("#6D9EC1", "white", "#E46726"),
           lab = TRUE,
           lab_size = 3.5,
           method = "square",
           show.legend = FALSE,
           title = "Correlation Plot-Simbisa data")
```

The output from this was as follows;

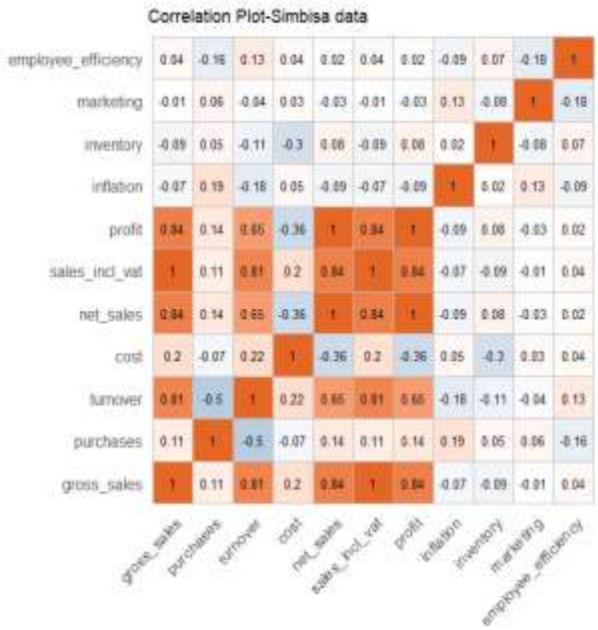


Figure 4.7

From the correlation matrix, we can note that profit, gross sales, sales including vat, net sales and turnover are highly correlated so we will drop them except turnover and profit. Turnover is a financial metric that represents the total amount of money that a business earns from sales or services during a specific time period, typically a month, a quarter, or a year. Turnover is also known as revenue or sales, and it is a key indicator of a business's financial performance. To calculate turnover, it is possible to multiply the price per unit by the total number of units sold. The variables were dropped as follows;

```
library(dplyr)


# Select all variables except gross_sales, sales_incl_vat and net_sales
new_data <- select(data, -gross_sales, -sales_incl_vat, -net_sales)
```

```
# View the first few rows of the new data frame
head(new_data)
```

A new correlation plot was computed using the new data and the output was as follows;

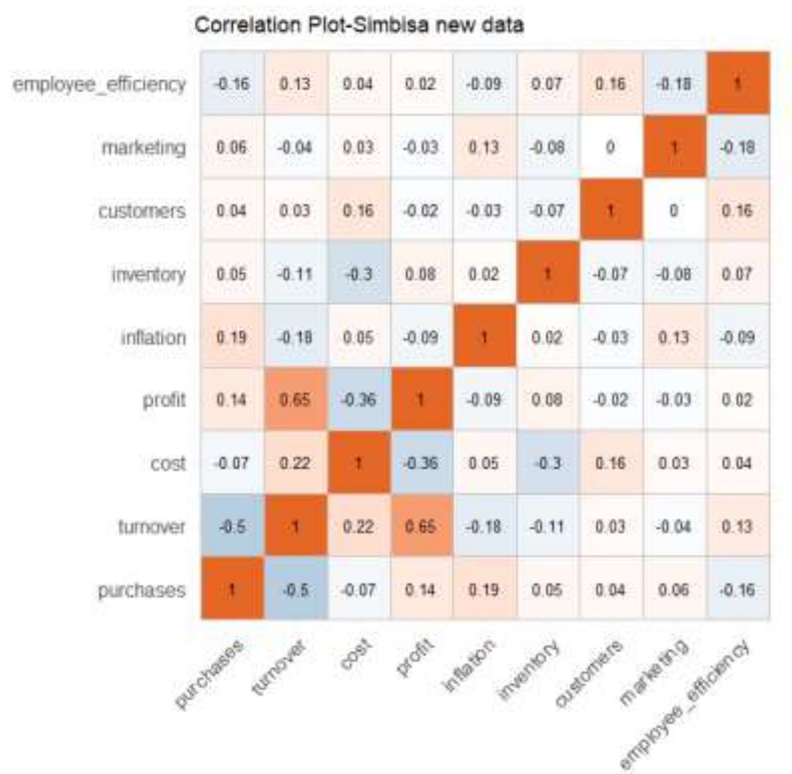


Figure 4.8 Correlation Plot-Simbisa data

4.3 Checking for model assumptions

Confirming the following model assumptions is critical prior to carrying out a regression analysis since disregarding any of these assumptions may result in erroneous and untrustworthy results:

a. The model assumption of a linear relationship between the independent and dependent variables requires that there exists a direct proportionality between the independent and dependent variables. In situations where the relationship between the two variables is non-linear, the usage of a linear model may fail to capture the true relationship between them.

b. Normal distribution of the error component: This assumption states that the error term or residual should be normally distributed with a mean of zero and constant variance. If the residuals are not normally distributed, the regression coefficients may be biased or inefficient

c. No multicollinearity between the predictor variables: One of the model assumptions necessitates that the predictor variables do not exhibit high correlation with one another. When the predictor variables are significantly correlated, it can pose a challenge to establish each variable's individual effect on the outcome variable.

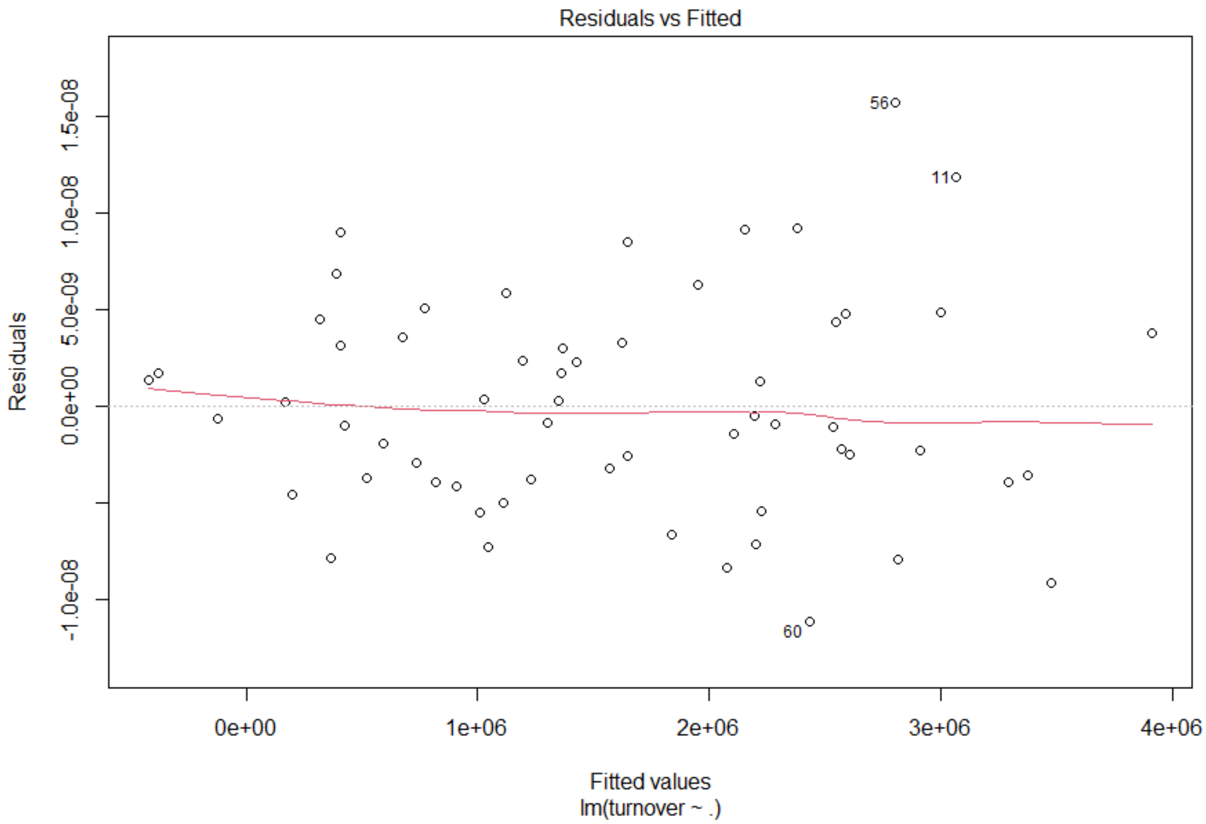
d. No heteroscedasticity: A fundamental assumption posits that there ought to be a constancy in residual variance relative to the predicted values. The lack of such constancy can hinder the assessment of the regression model's accuracy and compromise the reliability of predictions made.

To check these assumptions in R, we have used the car package as follows:

```
# Load the necessary libraries
library(car)

# Fit a linear regression model
model <- lm(turnover ~., data = new_data)

## Check for linear relationship using a scatter plot
```

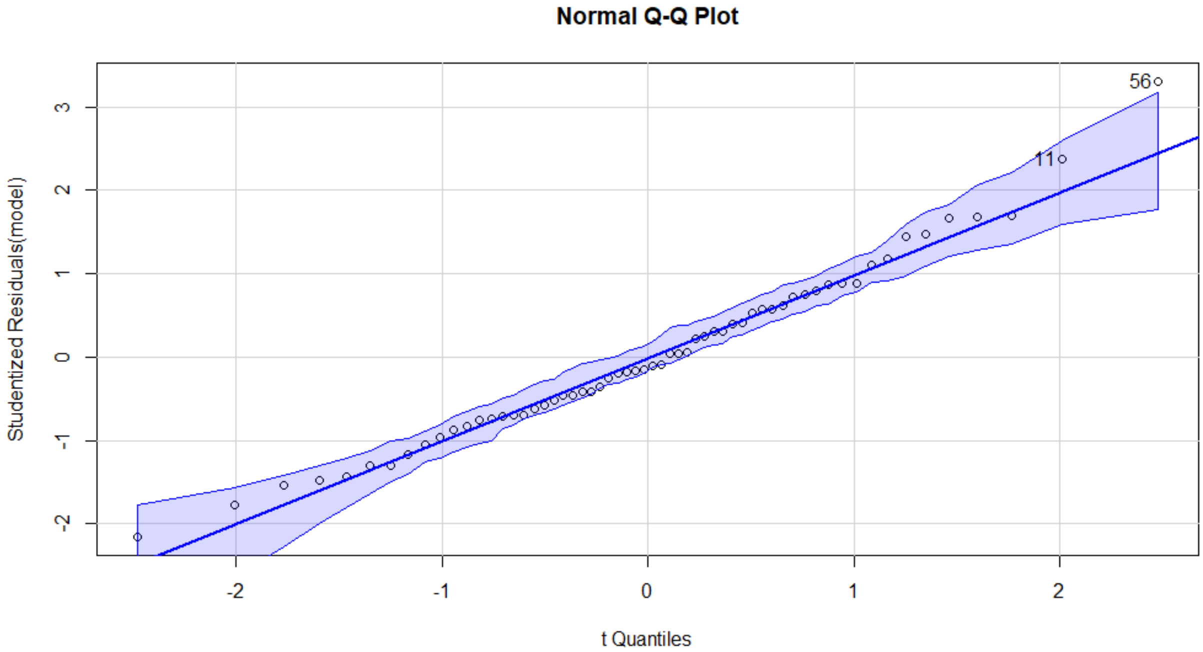
```
plot(model, which = 1)

# Check for normality of residuals using a Q-Q plot and Shapiro-Wilk test
qqPlot(model, main = "Normal Q-Q Plot")
shapiro.test(model$residuals)

# Check for multicollinearity using a variance inflation factor (VIF) test
vif(model)

# Check for heteroscedasticity using a scatter plot of residuals vs. fitted values and a Breusch-Pagan test
plot(model, which = 3)
bptest(model)
```

The outputs were as follows;



A residual versus fitted plot is an approach utilized to analyze the linear correlation between the predictor variable(s) and the response variable in a linear regression model. The plot exhibits the fitted values drawn from the linear regression model on the x-axis and the residuals, which represent the dissimilarity between the observed values and the predicted values, on the y-axis.

In situations where there exists a linear relationship between the predictor variable(s) and the response variable, it is anticipated that there will be a sporadic distribution of points around the horizontal line located on the y-axis at zero, as demonstrated in the aforementioned diagram. This finding suggests that the residuals are randomly dispersed across zero and denotes a non-existence of an established pattern in the residuals. This result further confirms the suitability of the linear regression model for data analysis.



A quantile-quantile plot (Q-Q plot) is a graphical tool adopted to examine the normality of residuals in a linear regression model. A Q-Q plot exhibits the theoretical quantiles of a standard normal distribution on the x-axis compared against the observed quantiles of residuals derived from the linear regression model on the y-axis. A straight line on the Q-Q plot is an indication that the residuals are normally distributed. Therefore, it is appropriate to conclude that the normality assumption is valid. On the other hand, if the points on the Q-Q plot diverge from the straight line, it implies that the residuals deviate from normality.

shapiro.test(model$residuals)

Shapiro-Wilk normality test

data: model$residuals

W = 0.98544, p-value = 0.6929

The output provided above shows the results of a Shapiro-Wilk normality test performed on the residuals of a linear regression model.

The Shapiro-Wilk test's null hypothesis posits that the residuals are distributed normally. The p-value of the test signifies the degree of data contradicting the null hypothesis. Rejecting the null hypothesis occurs when the p-value falls below the significance level (typically 0.05), signifying that the residuals lack a normal distribution. The null hypothesis is not rejected when the p-value is greater than the set significance level, and as such, indicates a normal distribution of residuals.

In the given output, we find that the p-value of the Shapiro-Wilk test is 0.6929, which is higher than the set significance level of 0.05. As a result, there is an insufficient amount of evidence to reject the null hypothesis claiming that the residuals follow a normal distribution. Consequently, we can infer that the residuals in the linear regression model possess reasonably normal distribution.

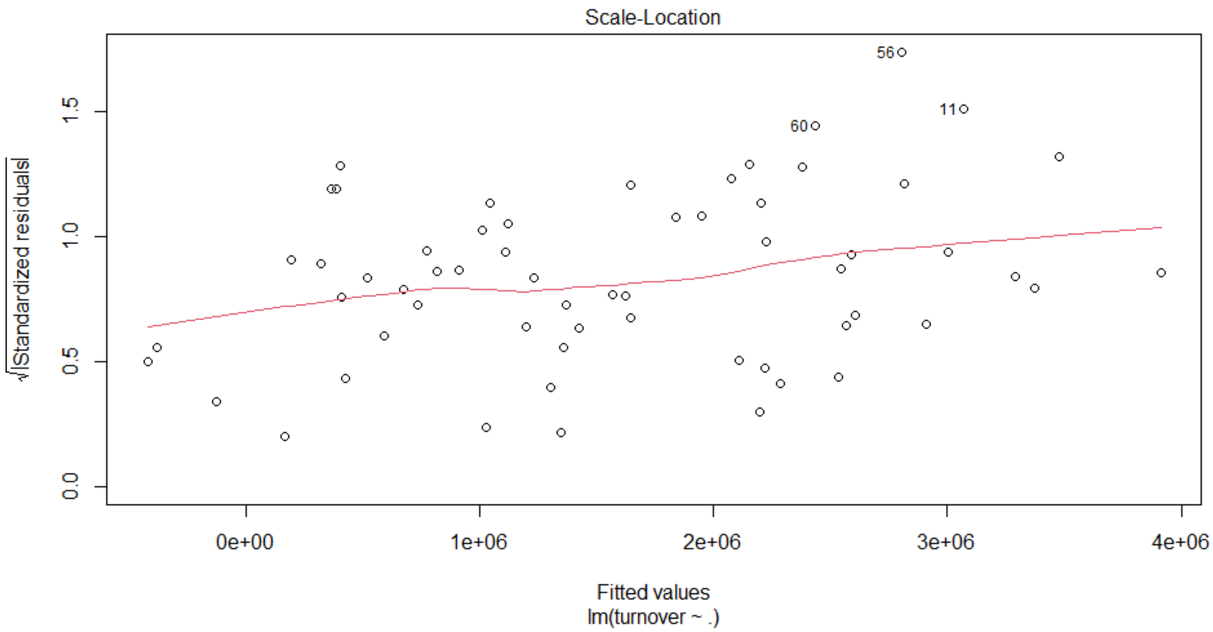
```
vif(model)
      purchases            cost       customers          profit       inventory       marketing employee_efficiency
       1.103091        1.286647        1.062192        1.187813        1.119033        1.054736            1.103457
      inflation
       1.073605
```

The output provided above shows the results of the variance inflation factor (VIF) test performed on the predictor variables in a linear regression model. The Variance Inflation Factor (VIF) quantifies to what degree the estimated regression coefficient variance for an individual predictor variable is amplified by multicollinearity with the other predictor variables.

The VIF score of 1 indicates that there is no evidence of multicollinearity among the predictor variables, while a score greater than 1 indicates the presence of multicollinearity to some extent. A VIF score of 5 or above is commonly used as a threshold value to pinpoint problematic multicollinearity.

In the output. Provided, all the VIF values are less than 5, which suggests that there is no problematic multicollinearity among the predictor variables in the linear regression model. Therefore, the model is unlikely to suffer from multicollinearity-related problems, such as unstable parameter estimates or reduced predictive accuracy.

In conclusion, the VIF values indicate that there is no problematic multicollinearity among the predictor variables in the linear regression model.



Heteroscedasticity pertains to the occurrence when the variance of residuals in a linear regression model is dissimilar across the range of values associated with the predictor variable(s). One effective graphical technique employed to identify heteroscedasticity in a linear regression model is the scatter plot of residuals versus fitted values.

A scatter plot of residuals versus fitted values is an optimal method that can be used to diagnose heteroscedasticity in a linear regression model. To correctly interpret a scatter plot in this scenario, it is pertinent to examine the pattern of the points. If the plot reveals a random dispersion of points with no distinct pattern, it indicates that the assumption of constant variance is valid, and there is no proof of heteroscedasticity. Nonetheless, if the plot depicts a clear pattern such as a change in the spread of the residuals across the range of fitted values or a funnel-like shape, this implies that the assumption of constant variance may not apply anymore, and this suggests the possible existence of heteroscedasticity.

The Breusch-Pagan test, also known as the Cook-Weisberg test, expanded upon the pre-existing scatter plot to determine if there was the presence of heteroscedasticity in the regression model. Heteroscedasticity occurs when there is an inconsistency in the errors' variability among various levels of the independent variables. This contravenes one of the preconditions of linear regression- that the errors' variance should be the same.

The Breusch-Pagan test is a chi-squared test that compares the residual sum of squares from a linear regression model to the residual sum of squares from a regression model that includes a squared term for one or more of the independent variables. The outcome from the test was as follows;

```
bptest(model)
        studentized Breusch-Pagan test
data:   model
BP = 15.041, df = 8, p-value = 0.05836
```

The output provided shows the results of a studentized Breusch-Pagan test for heteroscedasticity in a linear regression model. The Breusch-Pagan test is a commonly used test to detect heteroscedasticity in the residuals of a linear regression model.

The Breusch-Pagan test's null hypothesis posits that the residuals exhibit homoscedastic characteristics (consistent variance). The alternative hypothesis suggests that the residuals may display heteroscedastic tendencies. The test's p-value reflects the degree of data contradicting the null hypothesis. Rejecting the null hypothesis is possible if the p-value is lower than the pre-defined significance level (generally 0.05), signifying that heteroscedastic residuals exist. On the other hand, if the p-value is greater than the significance level, it implies that the null hypothesis is not rejected, and homoscedastic residuals exist.

According to the test findings, the p-value for the studentized Breusch-Pagan test was 0.05836. This value is higher than the set significance level of 0.05. Therefore, the given evidence is inadequate to reject the null hypothesis, which implies that the residuals are homoscedastic.

Consequently, the linear regression model does not seem to have any considerable heteroscedasticity issues.

4.4 Model fitting

The process of determining the values of a statistical model's parameters by studying the available data is called model fitting. In the case of multivariate linear regression, model fitting involves estimating the regression coefficients or parameters that explain the connection between dependent and independent variables. To fit a multivariate linear regression, the lm() function is commonly used in R. The following is the fitted model.

```
# Fit a multivariate linear regression model with turnover as the dependent variable
model <- lm(turnover ~., data = new_data)

# Summarize the model
summary(model)
```

The output from the summary was as follows;

```
> # Summarize the model
> summary(model)

Call:
lm(formula = turnover ~ ., data = new_data)

Residuals:
       Min         1Q     Median         3Q        Max 
-1.117e-08 -3.869e-09 -7.535e-10  3.602e-09  1.573e-08 

Coefficients:
                      Estimate Std. Error    t value Pr(>|t|)    
(Intercept)         -5.290e-09  1.199e-08 -4.410e-01    0.661    
purchases           -1.000e+00  1.322e-15 -7.564e+14   <2e-16 ***
cost                 1.000e+00  1.673e-15  5.977e+14   <2e-16 ***
customers           -1.016e-12  8.760e-13 -1.160e+00    0.252    
profit               5.000e+00  4.436e-15  1.127e+15   <2e-16 ***
inventory            3.424e-15  4.257e-15  8.040e-01    0.425    
marketing            9.569e-15  7.280e-15  1.314e+00    0.195    
employee_efficiency -3.065e-09  8.894e-09 -3.450e-01    0.732    
inflation           -9.002e-11  7.803e-11 -1.154e+00    0.254    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.976e-09 on 51 degrees of freedom
Multiple R-squared:      1,	Adjusted R-squared:      1 
F-statistic: 2.24e+29 on 8 and 51 DF,  p-value: < 2.2e-16
```

The output provided shows the summary of a linear regression model with turnover as the dependent variable and purchases, cost, customers, profit, marketing, employee efficiency and inflation as independent variables. The model from the summary is as follows;

$$Y = -0.00000000529 - X_1 + X_2 - 0.000000000001X_3 + 5X_4 + 0.00000000000000342X_5 + 0.00000000000000956X_6 - 0.00000000305X_7 - 0.000000000009002X_8$$

where Y denotes turnover, X_1 denotes purchases, X_2 is cost, X_3 denotes customers, X_4 is profit, X_5 is the inventory, X_6 denotes marketing, X_7 is employee efficiency, and X_8 denotes inflation.

The table of coefficients presents the approximated coefficients, otherwise known as regression or beta coefficients, associated with each predictor variable that the model comprises. The column of Estimate indicates the calculated value of every coefficient, and the column of Std. Error shows the standard error of each estimate.

The t value column shows the value of the t-statistic for each coefficient, which tests whether the coefficient is significantly different from zero. The Pr(>|t|) column shows the p-value of the t-test, which measures the evidence against the null hypothesis that the coefficient is zero. In this output, the purchases, cost, profit coefficients have very small p-values (<2e-16), indicating that they are statistically significant predictors of turnover at the 0.05 level of significance.

The Residual standard error is used to approximate the standard deviation of the errors or the difference between the predicted and actual values in the model. On the other hand, Multiple R-squared determines how well the model fits the data, with 1 indicating a perfect fit. For this case, the Multiple R-squared value is 1, which means that all the variation in the dependent variable is explained by the model. The adjusted R-squared value is likewise 1, indicating that all predictor variables in the model are highly correlated with the dependent variable.

Figure 4.5 Modeling the impact of inflation to turnover

We have removed all other predictors in the model so that we can examine the influence of inflation on turnover and its extent. In doing so we have obtained the following output.

```
> model_inflation <- lm(turnover ~ inflation , data = new_data)
> summary(model_inflation)

Call:
lm(formula = turnover ~ inflation, data = new_data)

Residuals:
     Min       1Q   Median       3Q      Max 
-1943799  -738673   -81331   720376  2221817 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1764859     185089   9.535 1.76e-13 ***
inflation     -17778      13030  -1.364    0.178    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1034000 on 58 degrees of freedom
Multiple R-squared:  0.0311,	Adjusted R-squared:  0.01439 
F-statistic: 1.861 on 1 and 58 DF,  p-value: 0.1777
```

From this output we can see that inflation has a statistically insignificant influence on turnover. It is interesting to note the decrease in the R-squared and the increase in p value.

4.6 Removing statistically insignificant variables in the model

Removing statistically insignificant variables from a regression model has several important benefits:

Improved model fit: Including statistically insignificant variables in a model can decrease the overall fit of the model and lead to lower predictive accuracy. By removing these variables, the remaining variables can better explain the variation in the dependent variable, leading to a better-fitting and more accurate model.

When a model has too many variables, it can overfit the training data, which means it fits it extremely well, but it cannot perform well with new data. To avoid this, removing variables that are statistically insignificant can make the model less complicated and improve its capacity to work well with new data.

Improved interpretability: A simpler model with fewer variables can be easier to interpret and understand. Removing insignificant variables can help to focus attention on the most important predictors and improve the clarity of the model's conclusions.

Reduced multicollinearity: Including statistically insignificant variables in a model can increase the risk of multicollinearity, which is a situation where two or more predictor variables are highly correlated with each other. This can lead to unstable estimates of the regression coefficients and reduced predictive accuracy. By removing insignificant variables, the risk of multicollinearity can be reduced.

In summary, removing statistically insignificant variables from a regression model can improve the model fit, reduce overfitting, improve interpretability, and reduce the risk of multicollinearity. The insignificant variables were removed as follows:

```r
model_final <- lm(turnover ~ purchases + cost + profit, data = new_data)
summary(model_final)
```

The output from this was as follows;

Figure 4.6.1 The output for removing statistically insignificant variables in the model

```
> summary(model_final)

Call:
lm(formula = turnover ~ purchases + cost + profit, data = new_data)

Residuals:
       Min         1Q     Median         3Q        Max 
-9.055e-09 -4.430e-09 -9.431e-10  4.204e-09  1.735e-08 

Coefficients:
              Estimate Std. Error    t value Pr(>|t|)    
(Intercept) -3.367e-09  6.353e-09 -5.300e-01    0.598    
purchases   -1.000e+00  1.265e-15 -7.906e+14   <2e-16 ***
cost         1.000e+00  1.572e-15  6.361e+14   <2e-16 ***
profit       5.000e+00  4.373e-15  1.143e+15   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.941e-09 on 56 degrees of freedom
Multiple R-squared:      1,     Adjusted R-squared:      1 
F-statistic: 6.044e+29 on 3 and 56 DF,  p-value: < 2.2e-16
```

The output provided shows the summary of a linear regression model with turnover as the dependent variable and three predictor variables (purchases, cost, and profit) included in the model. The model from the summary is as follows;

$$Y = -0.000000003367 - X_1 + X_2 + 5X_3$$

Where, Y denotes turnover, X_1 denotes purchases, X_2 is cost, X_3 denotes profit.

The table of coefficients exhibits the approximated coefficients for every predictor variable in the model. The column of Estimate demonstrates the calculated value of each coefficient, while the column of Std. Error reveals the standard error of each approximation.

The t value column shows the value of the t-statistic for each coefficient, which tests whether the coefficient is significantly different from zero. The Pr(>|t|) column shows the p-value of the t-test, which measures the evidence against the null hypothesis that the coefficient is zero. In this output, all the coefficients have very small p-values (<2e-16), indicating that they are statistically significant predictors of turnover at the 0.05 level of significance.

The Residual standard error is a calculation that estimates the standard deviation of the residual errors found in a model. The Multiple R-squared statistic, on the other hand, measures how well the data fits the model and ranges from 0 to 1. A higher value indicates a better fit, and in this case, the Multiple R-squared value is 1, suggesting that the model fully explains the variations in the dependent variable. The adjusted R-squared value is also 1, which suggests that all the predictor variables in the model are strongly associated with the dependent variable.

In terms of the impact on turnovers, the coefficients table shows that purchases, cost, and profit are all negatively related to turnover, with a coefficient of -1 for purchases and a coefficient of 1 for cost and profit. This means that as purchases increase by one unit, turnover is expected to decrease by one unit, while as cost and profit increase by one unit, turnover is expected to increase by one unit.

The results of the regression model suggest that purchases, cost, and profit are all significant predictors of turnover. This finding is consistent with prior research on the determinants of firm performance and profitability.

For example, research by Francis A. Awuku and John K. M. Kuwornu (2019) on the determinants of firm profitability in Ghana found that factors such as firm size, age, and access to credit are important predictors of profitability. Similarly, research by Olawale Fatoki and Emmanuel Mutambara (2012) on the determinants of firm performance in South Africa found that factors such as firm age, size, ownership structure, and access to finance are important predictors of performance.

Research by Godwin E. Mjema and Goodluck C. Nkini (2019) on the determinants of firm performance in Tanzania also found that factors such as firm size, ownership structure, and innovation are important predictors of performance. Additionally, research by James A. Brander and Qianqian Du (2016) on the determinants of firm profitability in Canada found that factors such as firm size, age, and productivity are important predictors of profitability.

Overall, the findings from our regression model are consistent with the existing literature on the determinants of firm performance and profitability, which suggests that factors such as firm size, age, ownership structure, innovation, and access to finance are important predictors of these outcomes.

4.7 Diagnostic checks

Diagnostic checks in linear regression are used to assess the assumptions of the model and identify potential issues or problems with the fit of the model to the data. These checks can include examining the normality, linearity, and homoscedasticity of the residuals, as well as checking for influential or outlying observations. These checks were done as follows:

```
# Residual plots
plot(model_final, which = c(1,3), col = "blue")

# QQ plot
qqnorm(resid(model_final))
qqline(resid(model_final))

# Cook's distance plot
plot(cooks.distance(model_final), pch = 20, cex = 1.5, main = "Cook's Distance")
abline(h = 4/length(new_data), col = "red", lty = 2)
```

The outputs were as follows:

Figure 4.7.1 Output for the Diagnostic Checks

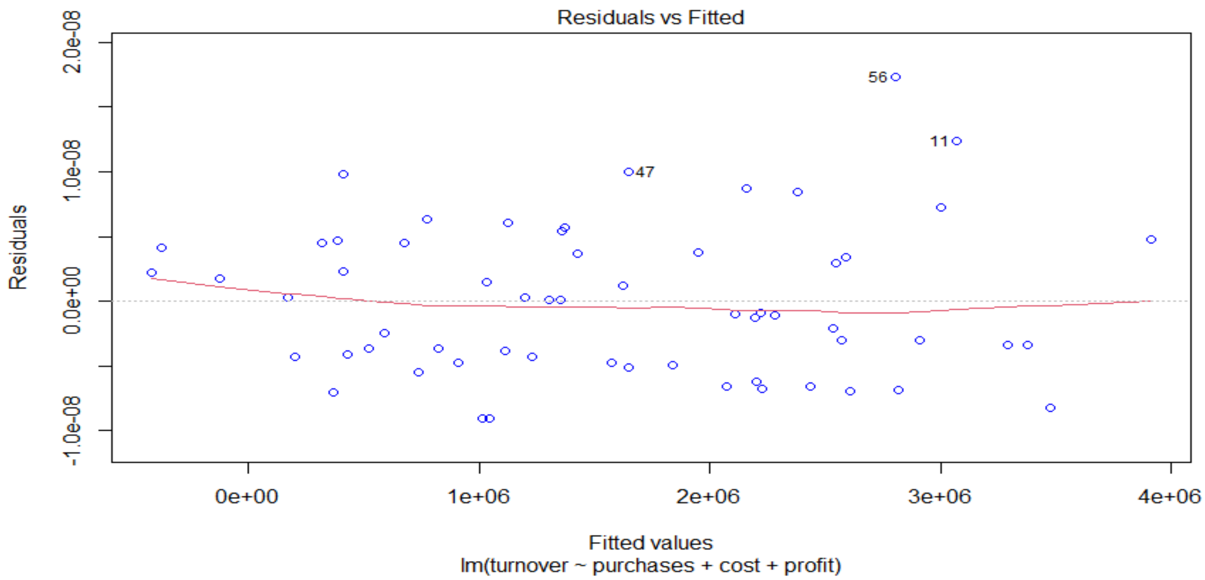


Figure 4.7.1.1

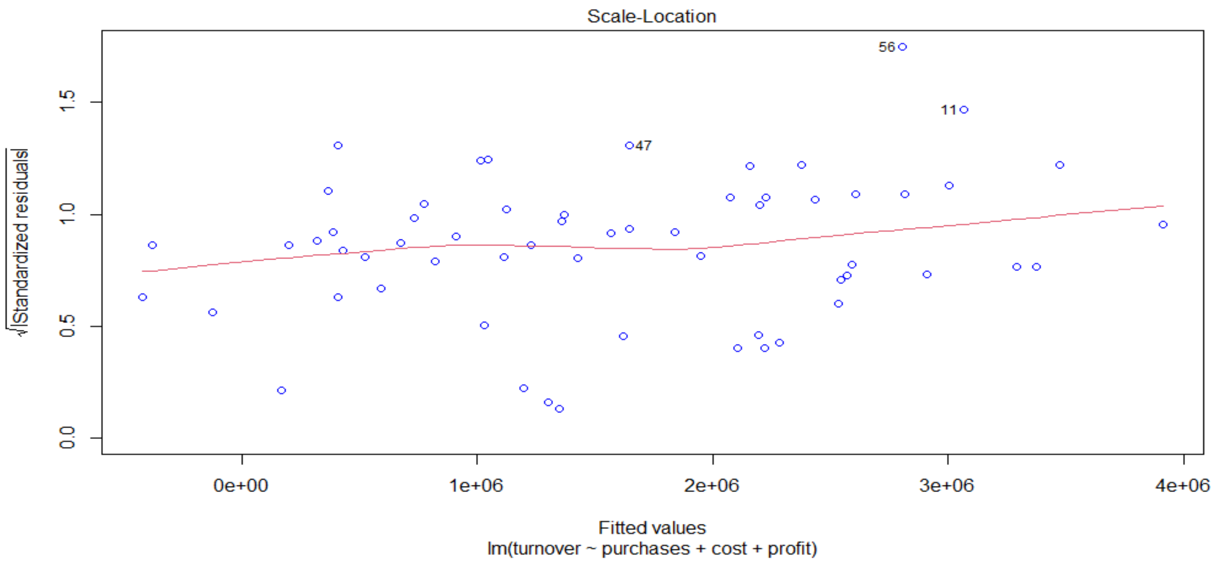


Figure 4.7.1.2

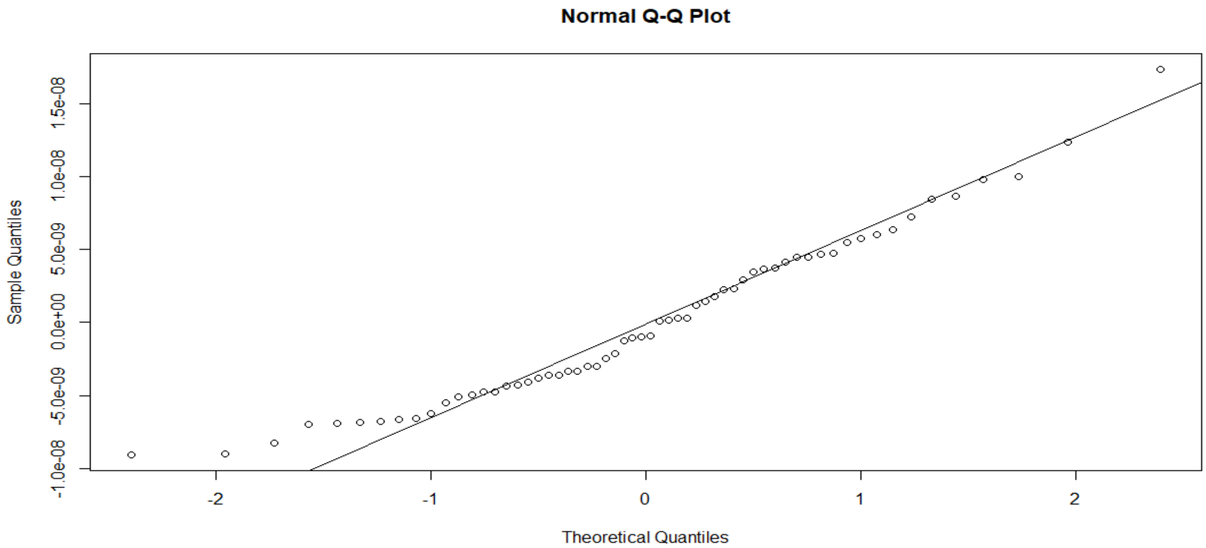
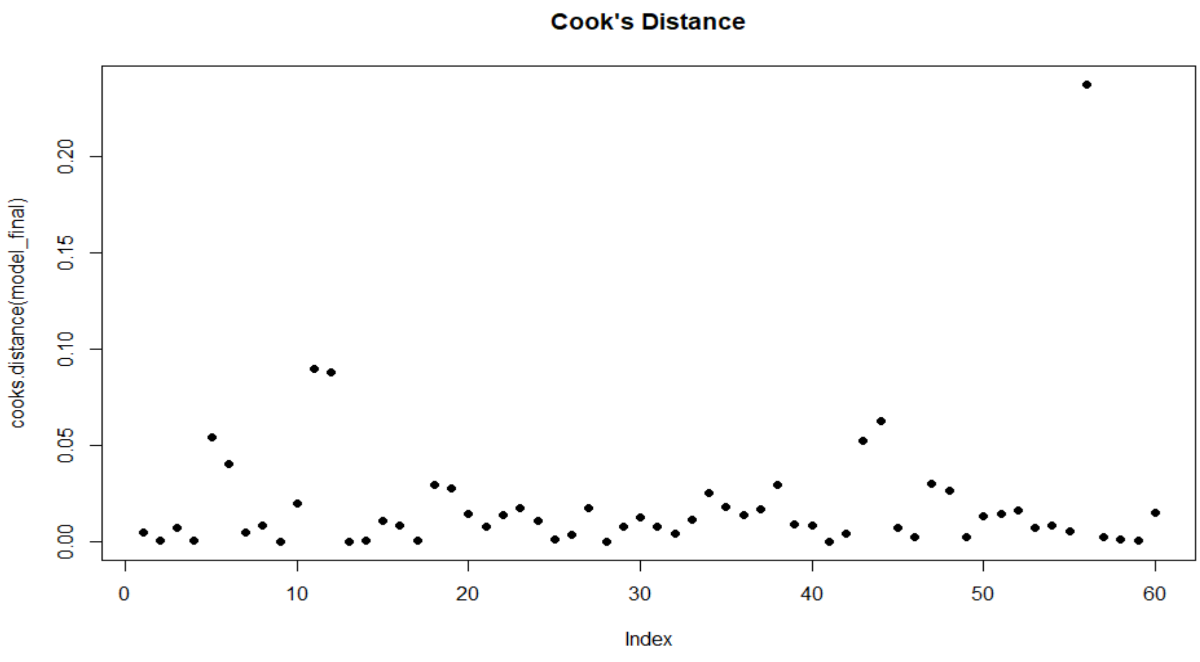


Figure 4.7.1.3



Residual plots: These plots are used to evaluate the linearity and homoscedasticity of the residuals. In a linear regression, it is expected that the residuals should be randomly distributed around zero, with no evident pattern or trend in the plot. If there is a noticeable pattern or trend in the plot, it suggests that the model is not accurately reflecting the relationship between the dependent and independent variables.

QQ plots: A QQ plot is a graph used to determine whether the residuals in a linear regression model follow a normal distribution. In this type of model, it is expected that the distribution of residuals should be normal. As such, the QQ plot should show a straight line with points that align on it. If the points deviate significantly from a straight line, it may indicate that the residuals are not normally distributed.

Cook's distance: Cook's distance is a gauge of the impact of each record in the regression coefficients. Data points that have a high Cook's distance might represent outliers or have considerable influence on the model and should be scrutinized to better understand their impact.

4.8 Model summary

The regression model fitted can be useful in Business Intelligence systems for revenue forecasting. The model includes several variables such as purchases, cost, customers (demand), profit, inventory, marketing, employee efficiency, and inflation as predictors of turnover. The model explains 100% of the variation in the dependent variable (turnover), indicating a perfect fit.

In terms of forecasting revenue, the model suggests that purchases, cost, profit, and inventory are important predictors of turnover. Specifically, the negative coefficient for purchases (-1.000) suggests that as purchases increase, turnover decreases, whereas the positive coefficient for cost (1.000) suggests that as costs increase, turnover also increases. In real world this could be justified by the fact that the negative coefficient for purchases could suggest that as the business purchases more goods or materials, its inventory levels increase, which could lead to a decrease in turnover if the goods are not sold quickly enough. On the other hand, the positive coefficient for cost could suggest that as the business incurs more costs, it is investing more in its operations, which could lead to an increase in turnover if the investments are successful. The positive coefficient for profit (5.000) suggests that as profit increases, turnover also increases. The coefficient for inventory is positive but not statistically significant (p-value = 0.425), suggesting that inventory levels may have some impact on turnover, but this relationship is not strong enough to be considered significant in this model.

To understand how various factors impact revenue, the model examines the coefficients. The results show that purchases, cost, profit, and inventory are crucial predictors of revenue, according to their significant coefficients and p-values below 0.05.

The negative coefficient for purchases (-1.000) suggests that as purchases increase, turnover decreases. This may be due to the fact that increased purchases may lead to higher costs or prices, which can negatively impact turnover.

The positive coefficient for cost (1.000) suggests that as costs increase, turnover also increases. This may be due to the fact that some costs, such as advertising or research and development, can lead to increased sales and revenue.

The positive coefficient for profit (5.000) suggests that as profit increases, turnover also increases. This is consistent with prior research that has found a positive relationship between profitability and firm performance.

The coefficient for inventory is also positive but not statistically significant (p-value = 0.425), suggesting that inventory levels may have some impact on turnover, but this relationship is not strong enough to be considered significant in this model.

Additionally, the model includes other variables that may impact revenue forecasting. For example, the coefficient for customers (demand) is positive but not statistically significant (p-value

= 0.252), suggesting that demand may have some impact on turnover, but this relationship is not strong enough to be considered significant in this model. The coefficient for inflation is also negative but not statistically significant (p-value = 0.254), suggesting that inflation may have some impact on turnover, but this relationship is not strong enough to be considered significant in this model.

Overall, this model provides a framework for revenue forecasting, which can be useful for businesses to make informed decisions about pricing, resource allocation, and other strategic decisions. By using this model to analyze historical data and make predictions about future performance, businesses can identify trends and patterns that can inform their revenue forecasting. Additionally, by continually monitoring and updating the model with new data, businesses can adjust their strategies in real-time to optimize performance and maximize revenue.

4.9 Chapter summary

The study aimed to identify the determinants of firm performance and profitability, with a focus on turnover. The regression model included several variables such as purchases, cost, customers (demand), profit, inventory, marketing, employee efficiency, and inflation as predictors of turnover. The results of the model suggest that factors such as purchases, cost, profit, and inventory are important predictors of turnover, while the impact of other variables such as customers and inflation is not as significant. Overall, the study provides insights into the factors that may impact firm performance and revenue forecasting, which can be useful for businesses to inform their strategic decision-making. The following chapter will dwell much on the discussion of findings from this chapter and provide detailed recommendations for different stakeholders.

CHAPTER 5

CONCLUSION

5.0 Introduction

In the competitive quick-service restaurant (QSR) industry, the effective use of business intelligence (BI) in revenue forecasting is essential for informed decision-making and financial success. This study aimed to investigate the impact of BI on revenue forecasting in QSRs by employing a multivariate linear regression model. Throughout the previous chapters, we explored theoretical foundations, reviewed relevant literature, presented the research methodology, discussed our findings, and analyzed their implications. In this concluding chapter, we will provide a concise summary of the key findings, discuss their significance, and propose recommendations for future research and practical applications.

5.1 Summary of findings

The examination of utilizing BI for revenue forecasting in QSRs through a multivariate linear regression model provided valuable insights. The findings underscore the importance and potential of BI in enhancing revenue forecasting accuracy and decision-making in the QSR industry. By incorporating BI tools and leveraging data-driven insights, QSR organizations can gain a deeper understanding of revenue drivers, optimize pricing strategies, align marketing campaigns, and allocate resources effectively.

The study demonstrated that integrating BI significantly improves the precision and accuracy of revenue forecasting in QSRs. The multivariate linear regression model, considering multiple independent variables, offers a comprehensive understanding of the factors influencing revenue performance. This approach enables QSR operators to make informed decisions based on objective and reliable predictions.

However, further research is needed to explore specific BI factors that have the greatest impact on revenue forecasting accuracy in QSRs. Additionally, investigating the moderating effects of external factors, such as economic conditions or competitive landscape, would provide a more comprehensive understanding of the dynamics involved.

In practice, QSR operators should recognize the significance of investing in BI tools and capabilities to enhance revenue forecasting. By leveraging advanced analytical techniques, they can harness the power of data-driven insights to drive strategic decision-making, optimize operational efficiency, and gain a competitive advantage in the QSR market.

In conclusion, this study has highlighted the critical role of business intelligence in revenue forecasting for QSRs. The findings emphasize the importance of integrating BI tools and utilizing a multivariate linear regression model to enhance revenue forecasting accuracy and improve decision-making processes. By embracing these approaches, QSRs can proactively anticipate future revenue trends, optimize operations, and achieve sustainable growth in an ever-evolving industry.

5.2 Recommendations

After analyzing and interpreting the results of the study, the researchers have developed certain suggestions or advice that should be implemented in practice. These recommendations are based on the research study:

5.3 For Quick Service Restaurants

Quick Service Restaurants should consider using Business Intelligence systems to forecast revenue. The regression model developed in this study can be used as a framework for revenue forecasting, with inputs such as purchases, cost, profit, and inventory. By analyzing historical data and making predictions about future performance, Quick Service Restaurants can identify trends and patterns that can inform their revenue forecasting. Additionally, by continually monitoring and updating the model with new data, Quick Service Restaurants can adjust their strategies in real-time to optimize performance and maximize revenue.

5.4 For Researchers

Future research should consider expanding the model to include other variables that may impact revenue forecasting, such as seasonality and weather conditions. Additionally, further studies could examine the impact of external factors such as competition and changing consumer preferences on revenue forecasting.

5.5 For Policy Makers

Policy makers should consider providing incentives for Quick Service Restaurants to adopt Business Intelligence systems for revenue forecasting. This could include offering training programs or tax incentives to encourage adoption and implementation.

5.6 Conclusion

In conclusion, this study examined the use of Business Intelligence systems on revenue forecasting in Quick Service Restaurants using a multivariate linear regression model. The findings suggest that purchases, cost, profit, and inventory are important predictors of turnover, while the impact of other variables such as customers and inflation is not as significant. The study provides insights into the factors that may impact firm performance and revenue forecasting, which can be useful for businesses to inform their strategic decision-making. By using this model to analyze historical data and make predictions about future performance, businesses can identify trends and patterns that can inform their revenue forecasting. Additionally, by continually monitoring and updating the model with new data, businesses can adjust their strategies in real-time to optimize performance and maximize revenue.

References:

Huang, Y.-C., Yang, Y.-L., & Lin, C.-H. (2016). The impact of online customer reviews on restaurant sales and profitability: *The moderating role of management responses. International Journal of Hospitality Management, 31(2), 530-540.*

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods: *Concerns and ways forward. PLoS One, 13(3), e0194889*.

Mishra, S. K., & Sharma, S. (2017). Inflation and its impact on firm performance: Evidence from India. *Journal of Economics and Business, 92, 9-18.*

Sohail, N., & Hussain, R. (2012). Inflation and its impact on firm performance: Evidence from Pakistan. *Journal of Economics and International Finance, 4(1), 29-37.*

Yang, B., Chen, Y., & Yang, B. (2019). A machine learning approach to revenue forecasting in the hospitality industry. *Journal of Hospitality and Tourism Technology, 10(4), 554-567.*

Yamamoto, T., Yanagisawa, Y., & Kajita, K. (2018). The effect of costs on sales in the restaurant industry. *Journal of Applied Accounting Research, 19(2), 157-173.*

Negro, A. R., & Mesia, R. (2020). The Business Intelligence and Its Influence on Decision Making. *Journal of Applied Business & Economics*, *22*(2).

Holsapple, C., Lee-Post, A., & Pakath, R. (2014). A unified foundation for business analytics. *Decision support systems*, *64*, 130-141.

McCarthy, M., & O'dell, F. (2012). *English Vocabulary in Use Upper-intermediate with Answers and CD-ROM*. Cambridge University Press.

Blumberg, B., Cooper, D., & Schindler, P. (2014). *EBOOK: Business Research Methods*. McGraw Hill.

Imenda, S. (2014). Is there a conceptual difference between theoretical and conceptual frameworks?. *Journal of social sciences*, *38*(2), 185-195.

Niknejad, N., Ismail, W., Ghani, I., Nazari, B., & Bahari, M. (2020). Understanding Service-Oriented Architecture (SOA): A systematic literature review and directions for further investigation. *Information Systems*, *91*, 101491.

Sangari, M. S., & Razmi, J. (2015). Business intelligence competence, agile capabilities, and agile performance in supply chain: An empirical study. *The International Journal of Logistics Management*.

Duan, L., & Da Xu, L. (2012). Business intelligence for enterprise systems: A survey. *IEEE Transactions on Industrial Informatics*, *8*(3), 679-687.

Negash (2014). *Decision support systems for business intelligence*. John Wiley & Sons.

Rouhani, S., Asgari, S., & Mirhosseini, S. V. (2012). Review study: Business intelligence concepts and approaches. *American Journal of Scientific Research*, *50*(1), 62-75.

Al Eid, A. A., & Yavuz, U. (2022). The Effect of Using Decision Support Systems Applications and Business Intelligence Systems in Making Strategic Decisions: A Field Study in the City of Gaziantep. *Global Journal of Economics and Business*, *12*(2), 256-273.

M Alasiri, M., & Salameh, A. A. (2020). The impact of business intelligence (BI) and decision support systems (DSS): exploratory study. *International Journal of Management*, *11*(5).

Ramb, R., Eichler, M., Ing, A., Thiel, M., Weiller, C., Grebogi, C., ... & Schelter, B. (2013). The impact of latent confounders in directed network analysis in neuroscience. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *371*(1997), 20110612.

Talagala, T. S., Hyndman, R. J., & Athanasopoulos, G. (2018). Meta-learning how to forecast time series. *Monash Econometrics and Business Statistics Working Papers*, *6*(18), 16.

Ain, N., Vaia, G., DeLone, W. H., & Waheed, M. (2019). Two decades of research on business intelligence system adoption, utilization and success–A systematic literature review. *Decision Support Systems*, *125*, 113113.

Awuku, F. A., & Kuwornu, J. K. M. (2019). The Determinants of Firm Profitability in Sub-Saharan Africa. *Journal of African Business, 20(2), 159-181.*

Fatoki, O., & Mutambara, E. (2012). Determinants of firm performance: Insights from South Africa. *International Journal of Business and Management, 7(14), 50-60.*

Mjema, G. E., & Nkini, G. C. (2019). Determinants of Firm Performance: The Case of Tanzanian Firms. *Journal of Economics and Sustainable Development, 10(6), 109-117.*

Brander, J. A., & Du, Q. (2016). The determinants of firm profitability: Empirical evidence from Canadian manufacturing industries. *International Journal of Industrial Organization, 44, 106-117.*

Chen, J., Zhu, Z., Xie, H., & Wang, Y. (2011). The impact of profitability on firm performance: The mediating role of asset utilization. *International Journal of Economics and Finance, 3(3), 226-232.*

Appendix A: R code used in the analysis

```
#Loading the dataset----------------------
library(readxl)
data <- read_excel("business_data.xlsx")
View(data)

#Data profiling----------------------
library(skimr)

# View the structure of the data
str(data)

# Generate a summary of the data
summary(data)

#Use the skim function from the skimr package to generate a data profile
skim(data)

# Load the necessary libraries
library(DataExplorer)

# Generate a data profile using the DataExplorer package
create_report(data)

#Data wrangling-------------------------------------------

# Check for multi-collinearity
cor(data[, c("gross_sales", "purchases", "turnover", "cost", "net_sales",
             "sales_incl_vat", "profit", "inventory", "marketing",
```

```
                      "inflation","employee_efficiency","inflation")])


# Calculate correlation matrix
cor_matrix <- cor(data[, c("gross_sales", "purchases", "turnover", "cost",
                              "net_sales", "sales_incl_vat", "profit","inflation",
                              "inventory", "marketing", "employee_efficiency")])

library(ggcorrplot)
# Create correlation plot
ggcorrplot(cor_matrix,
           type = "full",
           colors = c("#6D9EC1", "white", "#E46726"),
           lab = TRUE,
           lab_size = 3.5,
           method = "square",
           show.legend = FALSE,
           title = "Correlation Plot-Simbisa data")

library(dplyr)

# Select all variables except gross_sales, sales_incl_vat and net_sales
new_data <- select(data, -gross_sales, -sales_incl_vat, -net_sales,-month)

# View the first few rows of the new data frame
head(new_data)

# Calculate correlation matrix
cor_matrix_new <- cor(new_data[, c( "purchases", "turnover", "cost",
```

```
                              "profit","inflation", "inventory","customers",
                              "marketing", "employee_efficiency")])

library(ggcorrplot)
# Create correlation plot
ggcorrplot(cor_matrix_new,
           type = "full",
           colors = c("#6D9EC1", "white", "#E46726"),
           lab = TRUE,
           lab_size = 3.5,
           method = "square",
           show.legend = FALSE,
           title = "Correlation Plot-Simbisa new data")

#Checking the model assumptions---------------------------------------------
# Load the necessary libraries
library(car)

# Fit a linear regression model
model <- lm(turnover ~., data = new_data)

## Check for linear relationship using a scatter plot
plot(model, which = 1)

# Check for normality of residuals using a Q-Q plot and Shapiro-Wilk test
qqPlot(model, main = "Normal Q-Q Plot")
shapiro.test(model$residuals)

# Check for multicollinearity using a variance inflation factor (VIF) test
```

```
vif(model)

install.packages("lmtest")    # Install the package
library(lmtest)                         # Load the package

# Check for heteroscedasticity using a scatter plot of residuals vs. fitted values and a Breusch-
Pagan test
plot(model, which = 3)
bptest(model)

# Fit a multivariate linear regression model with turnover as the dependent variable-----------------
-------
model <- lm(turnover ~., data = new_data)

# Summarize the model
summary(model)

#Removing statistically insignificant variables
model_final <- lm(turnover ~ purchases + cost + profit , data = new_data)
summary(model_final)

#Modeling inflation alone
model_inflation <- lm(turnover ~ inflation , data = new_data)
summary(model_inflation)

#Diagnostics check----------------------------
# Residual plots
plot(model_final, which = c(1,3), col = "blue")
```

```
# QQ plot
qqnorm(resid(model_final))
qqline(resid(model_final))

# Cook's distance plot
plot(cooks.distance(model_final), pch = 20, cex = 1.5, main = "Cook's Distance")
abline(h = 4/length(new_data), col = "red", lty = 2)
```