PREDICTIVE ANALYTICS MODEL FOR SMALL AND MEDIUM ENTERPRISES IN KENYA, FORECASTING ON SUPPLY AND DEMAND

BY

BY MBURU JOHN MUREITHI

A research thesis submitted in partial fulfillment of the requirements of the degree of Master of Science Computer Information Systems (M Sc. CIS).

KENYA METHODIST UNIVERSITY

October 2022

DECLARATION

This research project is my original work and has not been submitted to any other university for an academic award.

Sign ……………………………………… Date ………01/09/2022……………….

MBURU JOHN MUREITHI
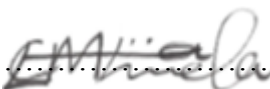
CIS-3-0561-1/2018

We confirm that the work reported in this thesis was carried out by the candidate under our supervision.

OMWANDO NYAKONI

KENYA METHODIST UNIVERSITY

SUPERVISOR

Sign ……………………………………… Date ………01/09/2022……………….

DR. LAWRENCE MURIIRA

KENYA METHODIST UNIVERSITY

SECONDARY SUPERVISOR

Sign ……………………………………… Date ………01/09/2022………….

COPYRIGHT

DEDICATION

I would like to dedicate this research thesis to JABALI and TAMASHA.

You have within you the ability to create that which you constantly seek!

EVERYTHING that you're searching for is already within you!

ACKNOWLEDGMENT

ABSTRACT

Predicting business operations is a critical task for small and medium enterprises (SMEs). With increased unpredictability in the business environment, small enterprises find themselves in the receiving end simply because they do not have the tools in decision-making like their counterparts who have established business decision-making tools. With increased use of ICT, SMEs can now tap into the power of data to support decision-making. Transactional data like sales, purchases, payments, service requests, invoices, purchase orders, delivery notes, repairs, and maintenance notes are collected by SMEs and are readily available. Currently, SMEs in Kenya have limited on non-existent quantitative forecasting systems in place, predictive analytics is a preserve of well-funded, large, well-established or international companies. SME owners can now benefit from the power of predictive analytics in areas like sales, purchases, business lead generation, recommendations systems, and risk prediction. With predictive analytics, SME owners can have added confidence in decision-making to help propel their business to successful ventures. Predictive analytics algorithms like clustering algorithms, linear regression, and classification algorithms can be used to aid SME owners and managers gain insights into their business by identifying relationships and associations between the various variables in their business or identification of trends. SMEs in Kenya, have in the past run out of business due to wrong decisions associated with lack of information regarding their customers, their business, or the business sector as a general. SMEs' contribution to Kenya's GDP growth is vital and the use of technology and ICT could mitigate the challenge of access to information that SMEs have. The use of technology to predict business operations and performance is the next frontier in ensuring business sustainability, job security, and a good business environment. This research aimed to solve this information gap by designing a demand and supply forecasting model that equips SME owners and managers with insightful information about the demand and supply of the products they are selling, helping them make informed decisions based on their sales and purchase data. The study was carried out on five (5) SMEs in Nairobi and Muranga who had access to ICT infrastructure and had some knowledge of digital book-keeping methods but had no forecasting or prediction systems in place. A fully functional web passed demand and supply forecasting model accessible via a browser was designed. A beta test was conducted by the five respondents with positive feedback. The majority of the respondents, as the study found out, derived great value from forecasting their demand and supply and were able to stock right and meet their clients' needs better, the forecasting model was developed using the agile prototyping method of software development with cascading style sheets (CSS), hyper-text markup language (HTML), react graphical user interface and an R based forecasting engine. The designed system allowed the users to interact with a user-friendly graphical user interface on either a mobile device or a computer allowing more freedom and flexibility in accessing the platform. It allowed SMEs to upload their sales and purchase data in a predefined format that the forecasting model consumes and provide accurate demand and supply forecasts and provided forecasting accuracies in the 95% quartiles for all the SMEs sampled. The model was evaluated by the same SMEs with 100% of them indicating that they would be relying on the forecasting model for all future forecasts. The designed model allows SMEs to benefit from the demand and supply forecasts irrespective of its sector and the type of enterprise engages in.

TABLE OF CONTENTS

List of Equations

List of Tables

List of Figures

Abbreviations and Acronyms

Abbreviations	Description
API	Application Programming Interface
ARIMA	Autoregressive Integrated Moving Averages
BI	Business intelligence
CRM	Customer Relationship Management
CSS	Cascading Style Sheets
EM	Expectation–maximization (EM) algorithm
GDP	Gross domestic product
ICT	Information communication technology
MAE	Mean Absolute Error
MAP	Maximum a Posteriori assessments
MAPE	Mean Absolute Percentage Error
MASE	Mean Absolute Scaled Error
NN	Neural Networks
OOP	Object-oriented programming
PII	Personally Identifiable Information
RF	Random Forest
RWD	Responsive Web Design
SMEs	Small and medium enterprises
SAAS	Software as a service
SDLC	Software development life cycle
UAT	User acceptance testing

CHAPTER ONE

1.1.   INTRODUCTION

1.2.   Background of the Study

Predictive analytics consists of several statistical and analytical processes that are used to create models that forecast future outcomes (Nyce, 2007). The forms taken by a predictive model depend on what they intend to predict. The emergence of knowledge discovery as a new technological frontier is credited to the rapid advancement and broad application of information technology. According to Zhang and Zhang (2002) knowledge discovery is an iterative process of four steps which are, problem definition, pre-processing, data extraction, and post-processing. Predictive modeling tools forecast future outcomes that enable businesses to make proactive, data-based decisions (Pousttchi et al., 2015). Predictive analytics modeling is a collection of computational mechanisms and methods that enables an organization to detect data patterns. Organizations use predictive analytics to achieve their desired goals and increase their profits. Predictive analytics can enable small and medium enterprises to improve their existing processes, understand customer behavior, isolate new opportunities, and foresee problems before they harm their business resulting in a greater return on investment (Goul et al., 2005). A predictive analytics model will allow SMEs to understand their business verticals like Sales, Purchases, Training, Marketing, and in turn, getting insights into future trends and opportunities. Such information is vital in decision making that could keep SMEs afloat. Hamiche et al., identified Demand forecasting as an important aspect in supply chain decision making. As forecasting provided a basis step for activity planning in response to customer demand (Hamiche et al., 2018). They go ahead and explain that having the exact amount of

product is the ideal case for any organization as It ensures that customer orders are fulfilled and that there are no dead stock which frees cash to be used in other areas for the company. There were 22.86 million internet users in Kenya in the year 2020, this was an increase of 3.2 million users from the previous year account to a 16% increase, this account to 43% internet penetration (Kemp, 2021). In the same report, mobile penetration in Kenya was estimated to be 98%, this clearly indicates the reliance on mobile devices for the access of internet connection in Kenya. It is therefore important to have this in mind when developing forecasting applications as the main mode of access to the platform would be through small-screen devices (Kenya National bureau of statistics [KNBS], 2016).

The use of predictive analytics is not new in the world of business, large organizations have access to skilled data engineering teams that develop machine learning and forecasting tools that such organizations can use and provide a competitive edge. Findings discovered that the adoption of information systems by SMEs were more influenced by technological and organizational factors than environmental factors, this means that organizations that are willing to experiment, management support, organization readiness and the size of the organization are better adopters of Information Systems. (Ramdani et al., 2009). With this in mind, it is important to find a forecasting solution for SMEs. With the high closures of SMEs in Kenya attributed to financial and market information (Mugo, 2016), this research explores the use of predictive analytics in forecasting demand and supply. This is because SMEs face uncertain market demands and needs that lead to their inability to meet their customers' needs. Giving these SMEs the ability to forecast their demand and supply could be a possible avenue for enhanced revenue and reduced losses (Bowen, et al., 2009).

This Study will demonstrate unique opportunities available for SMES in Kenya to begin the adoption of predictive analytics. The study focuses on the useability of forecasting models and the flexibility to allow SMES from different sectors to be able to accurately forecast demand and supply. This study also recognizes that not all SMEs are technology savvy, the use of common spreadsheets can be used in the collection of historical data for the purposes of forecasting. In the end, any SME that requires forecasting solutions can easily access the forecasting model either on premise or in the cloud using either personal computer systems or mobile devices.

Currently, there is a lack of access to locally made open-source applications targeting demand and supply forecasting for SMEs in Kenya, where SME owners and managers can easily forecast their organizations' demand and supply. This is surprising because, based on a survey by the Central Bank of Kenya (Central Bank of Kenya [CBK], 2018) on the National economy reported that small and medium enterprises constitute 98% of all enterprises in Kenya, with a 4.7% increase in GDP, 3% was attributed to SMEs. However, a report by the National Bureau of statistics, small and medium enterprises in Kenya contributed about 40% of Kenya's GDP as of 2008 but 62% of these SMEs closed in less than 4 years. 2.2 million SMEs closed shop between 2011 and 2016 alone. These closures were attributed to the SMEs lack of access to enough market information, unlike the well-established businesses. This information gap leads to unpredictable markets, which in turn hinders the SMEs ability to plan their sales. (Bowen et al., 2009). A Demand and supply forecasting model will provide the SME owners and managers vital information regarding their business and business operations. This lack of insights into market movements greatly hinders the growth and sustainability of SMEs. SMEs need a data acquisition model that allows them to mine data from their transactional data sources into formats that are suitable for forecasting sales, once this data is extracted and prepared, meaningful insights can be derived after forecasting, however, there is a disconnect

between the data acquisition and forecasting as data that the SMEs store is not in a format that can be easily used in forecasting, this includes aspects like the format used to store transactional data, the databases used to store the data. This means that for each SME to fully utilize forecasting, individual customization is required in the process of ensuring that the data is ready for forecasting.

1.3.   Statement of the Problem

The lack of access to demand and supply forecasting tools for SMEs in Kenya leaves SMEs unprepared to meet their obligations to customers and suppliers.

1.4.   Objectives of the Study

The SME space is currently ripe for predictive analytics due to the increase in the use of technology. The purpose of this study is to design a supply and demand forecasting model for SMEs. The framework will report the current state of the business upon analysis of its sales and purchase transactional data. The Framework will consume this transactional data, transform it as required, perform analysis on the data, and provide demand and supply forecasts helping owners make informed decisions. The study will focus on identifying easily interpretable data stories, deducing relationships on the provided data set, and attempting to forecast the behavior of a new dataset; this reflects the main objective of data mining and knowledge systems (Velickov & Solomatine, 2000). The broad objective of the study will thus be, to design a free to use, demand and supply forecasting model, that can be used by small and medium enterprises in Kenya. The model will be able to identify patterns and visualize trends and forecasts. Due to the target users of the model being SME owners who might not have access to powerful infrastructure, the designed model will be easily installed on-premises or accessed in the cloud as software as a service (SAAS) on a hosted public domain.

4

1.4.1.　　Broad Objectives

I) To identify and analyze supply predictors captured by SMEs by analyzing orders made to suppliers.

II) To identify and analyze demand predictors captured by SMEs by analyzing delivery notes made to customers.

III) To identify and build a reusable and scalable predictive analytics model to forecast supply and demand based on the data collected in I and II

IV) To Evaluate the Effectiveness of the proposed model

1.4.2.　　Specific Objectives

I) Compare existing models and implement the best model for predicting demand and supply.

II) Build a data extraction user interface

III) Build a presentation layer to display the supply and demand forecasts

IV) To evaluate the effectiveness of the designed forecasting model

1.5.　　Research Questions

I) What predictors are captured by SMEs in Kenya that relate to Demand and Supply?

II) What is the best forecasting model for SMEs in Kenya?

III) How effective is the proposed model in forecasting demand and supply for SMEs in Kenya?

1.6.　　Justification and significance of the Study

Small and medium enterprises in Kenya and in a unique position, there is a great opportunity for predictive analytics adoption to address specific issues that hinder their growth and sustainability while competing with large established corporations. While there are have been

challenges in the adoption of information technology due to limitations of funds and lack of the expertise to design inhouse customizable forecasting models, it is still possible to overcome these challenges by use of open source, free to use predictive analytics model that can help these SMEs gain tremendous insights on their businesses that were not previously available. This research is significant in terms of its practical contribution to the owners and managers of small and medium enterprises in the Republic of Kenya and can be used as a basis and reference for other sub-Saharan economies.

1.6.1. Management of SMEs

The research will influence how SME managers in Kenya make business decisions. According to the Communication Authority of Kenya Report (Communication authority of Kenya [CAK], 2018), there were a total of 45.7 million data and broadband internet subscribers in Kenya, this high penetration of internet will increase the adoption of the framework. Managers that adopt the framework should expect personalized prediction of business supply and demand giving them the ability to make informed decisions for their business.

1.6.2. Employees of SMEs

Employees in SMEs that adopt the framework will have a better understanding of the business that they work in. this will allow them to direct their efforts towards the most important and rewarding activities to the SMEs. Employees in direct contact with the customer will benefit from having a better understanding of their customers and hence provide better services.

1.6.3. Policy Makers

Having a holistic view of market trends and projections, policymakers will be able to legislate and provide guidance to the public on nationwide trends.

1.6.4. Future Researchers

This research will form a solid foundation for researchers interested in SMEs in Kenya and Africa in general. Predictive analytics continues to take shape around the globe meaning more and more researchers will continue studying this research area.

1.7. Scope and Limitations of the Study

Lack of data is one of the most common barriers to predictive analytics. Making predictions requires the availability of good data (Davenport, 2014). However, the use of technologies like mobile money to make transactions in prominent within Kenya's SMEs. Due to this limitation, the scope of the study will focus on SMEs that have formal business book-keeping applications and have access to the internet. It is paramount that a customer's data is respected. Predictive models collect and use personally identifiable information (PII) about private citizens to forecast future behavior. In the past privacy-aware individuals have found such data usage as intrusive and dangerous. There is the anticipation of resistance by SMEs to share their business data. Business operations are sometimes a guarded secret that managers and owners might not be willing to share this information. As will be described in the methodology, the designed framework will not store the SMEs data, data will be used to make predictions and then discarded. The results of the prediction can, however, be stored to establish industry trends.

This research will target SMEs that have record keeping systems in place, without data, predictive analytics and forecasting is not possible, SMEs will need historical data from their business, this historical data will be extracted from their existing bookkeeping systems into a predefined format that the forecasting model will consume.

1.8. Definition of general terms

There are general concepts that one needs to understand to be able to dive in into the forecasting of Demand and Supply, this section defines these terms.

1.8.1. Demand and supply

As per the definition in microeconomics (Dilts, 2004), supply and demand is an economic model used to derive the price. This law implies that with everything held constant, the price for a particular product will vary until it settles at a point where the quantity demanded will be the same as the product quantity supplied.

1.8.2. Forecasting model development

Forecasting model development can be considered as any other software development project and can be implemented in the same manner a software project is implemented.

1.8.3. Evaluation of Software systems

Just like the evaluation of any software system, the evaluation of the forecasting model should be verified by the end users who are the source of the requirement or by a person representing the final end users. This can be through beta testing or User acceptance testing (UAT).

1.9. Thesis Outline

1.9.1. Literature review

Chapter two provides an overview of demand and supply forecasting, predictive analytics, business intelligence theoretical review of forecasting methods, it also compares the various techniques used by previous researchers, the chapter concludes with a summary section that identifies the gaps and opportunities identified by previous researchers and possible solutions.

1.9.2.   Methodology

This chapter highlights the three-stage approach that was used in this research, requirements gathering and analysis, coding and evaluation. Further, the chapter describes the respondents and the data collected, the research process used to answer the research questions is described. The summary section in the chapter provides a summary of the research procedure used to address the research questions.

1.9.3.   Results and discussions

In this chapter, the final prediction model is presented, the advantages of the model are discussed, the forecast results from the model are discussed and presented. The chapter also discusses the respondent who took part in the study. This chapter discusses the results in a three-stage study, the requirement analysis, the coding and the evaluation. The chapter finally concludes with a summary of the findings.

1.9.4.   Summary Conclusions and Recommendations

The chapter summarizes the thesis research questions, and presents the conclusion based on the results. An analysis of how the result answers the research question is laid out. The conclusions from the research are described in the section with recommendation for future research being laid out in the last part of the section.

CHAPTER TWO

2.0. LITERATURE REVIEW

2.1. Introduction

Small and Medium Enterprises in sub-Saharan Africa face a myriad of challenges in their attempt to make their enterprises profitable and successful. SMEs in developing countries like Kenya are the backbone of economic development. SMEs provide daily life sustenance and economic empowerment (Ayanda & Laraba, 2011). Since small and medium enterprises are largely owned or run by sole owners/managers, it is expensive for them to access relevant information systems as a base for structured decision-making. The adoption of forecasting tools and financial literacy can largely improve the small and medium enterprises' decision-making process. These forecasting models can predict data values which adopt similar outcome found from contrasting datasets. In the past, forecasting frameworks have been used in many diverse areas such as retail outlets, healthcare, financial institutions, and insurance (Smita & Sharma, 2014). Predictive analytics provide information that helps in making an informed decision of their businesses and enterprises.

2.3. Previous Studies and Approaches to forecasting

An example of successful forecasting was when Yan et al. (2015) took data from retail stores renowned for selling women's clothing and identified one hundred (100) respondents. They recommended an algorithm that provides consumers with valuable information at the most crucial time which resulted in a pleasant shopping experience for their customers. Another research conducted by Sharma, Goyal and Mittal, studied customer relationship management (CRM) applications and the benefits of analyzing collected data, by Analyzing CRM sales

using data analytics techniques he concluded that it is important for businesses to use data analytics technology, by integrating customer information and other relevant information and providing personalized services to customers (Sharma et al., 2018). These studies show that customization of experience is required as each customer has different needs. In this same light, SMEs need customized insights into their businesses. After studying data from Singapore based small and medium enterprises, Thong identified a positive correlation between the role of small and medium enterprises performance and a CEO's characteristics like knowledge in technology and innovation. He also identified that innovation characteristics and organizational characteristics were significant factors in the adoption and use of technology. The SME industry generates large data with the potential to gain insights with the use of predictive analytics (Thong, 1995).

In an attempt to construct a prediction model for mining clinical data, Zhou and his colleagues described eight Stages that can be used in the process of constructing a prediction model, the first step involved the identification of the clinical issues, here they focused on identifying the issues that they intended to investigate (Zhou, et al., 2019).

In our context, this involved the the identification of attributes that SME owners and managers use for the forecasting of demand and supply. Any SME can benefit from the forecasting of its business operations, however this research focused on demand and supply. The next step involved the identification of the research strategies that would yield the best and most accurate outcome, this involves the review of available data, the type of data collected by the SMEs that relates to demand and supply would significantly change our approach as to which strategy to use in our forecasting model. Upon reviewing the available data collected by the SMEs, the next step would be the identification of predictors, to be precise, what attributes are collected by the SMEs that would aid in the forecasting of the Issues Identified in the first step, for our case, what attributes can we use as predictors for demand and supply? As discussed in the

coming section, the availability of this data is really the key to a successful forecast. Since our intention was the forecasting of demand and supply, it was paramount that data relating to demand, and supply would be collected to ensure accurate forecasting. Zhou and his colleagues (Zhou, et al., 2019) indicated that Determination of the outcome should then be done, here an analysis of the expected outcome was determined, for our case, having sales and purchase data from SMES might lead to having, sales per day or sales per month and then translating this as low or high demand, the more sales you have for a given month the higher the demand for that month and vice versa. Once this was established, the construction of the prediction model could begin. In this stage, the selection of tools to be used for forecasting was done, the selection of the tool was based on the research strategy identified, the predictors and the expected outcome of the prediction. Other factors like skills of the researcher in a specific programming language, the cost of the programming platform and the availability of a vibrant community also affected the selection of the prediction model. Once the model was designed then the evaluation of the model was done.

Figure 2.1

Research process of prediction models
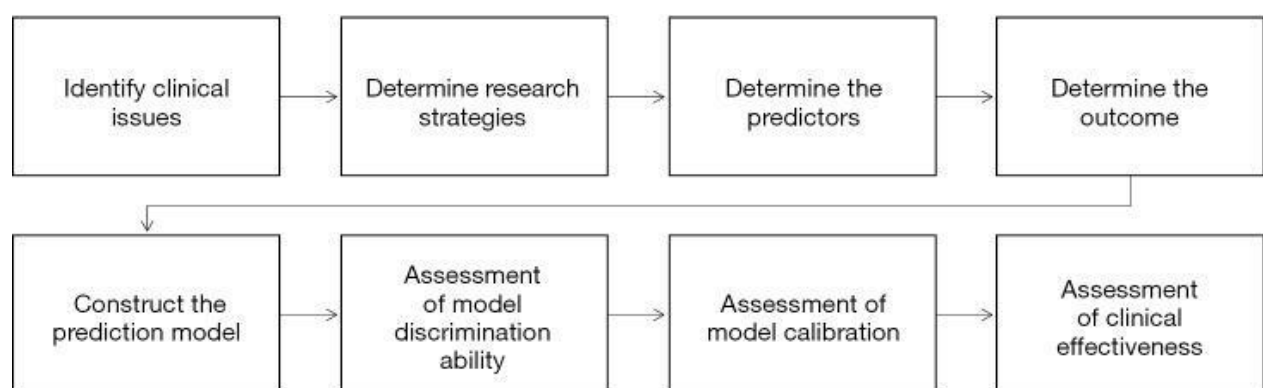
Quantitative forecasting are techniques that employ the use of historical data in time series in the identification, modeling and extrapolating trends found in these historical data. A general

method is followed as defined by Montgomery Jennings and Murat. (Montgomery et al., 2015) was as follows

I) Problem definition

II) Data collection.

III) Data analysis

IV) Model selection and fitting.

V) Model validation

VI) Deployment

VII) Monitoring

From the outset, it is evident that the method described by Montgomery resembles the classical software development life cycle (SDLC) with an additional step of data collection. And as we will discuss more about the design process of software systems in chapter 3, it is important to outline the relationship between software development and predictive analytics model design. The design of a predictive analytics model is a software development process. Like any software development project, the predictive analytics model design will follow a chosen model. The choice of the model depends on the requirements, end user responsibility, cost, team composition and skill amongst others.

In 2014, Amarasingham and his colleagues researched on the implementation of an electronic health care predictive analytics system (e- HPA), in their research, they discuss the challenges of implementing such a system, most notable were stakeholder management, privacy, consent policies and data quality assurance (Amarasingham et al., 2014). These challenges do exist in the implementation of predictive analytics for SMEs as well. This poses the question; how do we engage stakeholders and ensure that they fully participate in the design of the forecasting model? How do we ensure privacy in a cloud-based system? How do we ensure data quality? As will be discussed later, Software development techniques like Agile and Agile prototyping

allow for closer cooperation between the development and the users are better suited to solve some of this problem identified in the implementation e-HPA.

2.4. Demand and Supply Planning

When Saubar and Smith compared demand forecasting methods used by the supply chain department (SCM) at Bilports, a company selling building products in the United States to open-source forecasting methods they described Demand and Supply planning as one of the toughest aspects of supply chain management because it had the greatest planned future spending but had the largest disparity between performance and satisfaction. They further explain that traditional data analysis processes could no longer fulfill the needs of volatile markets, they propose a robust demand forecasting model as the ability to respond was no longer as crucial as the ability to forecast. (Shubbar & Smith, 2019).

2.5. Forecasting methods

Forecasting can be done by use of empirical qualitative analysis, qualitative methods, or by using mathematical quantitative analysis. There is no forecasting method that can be described as the best model. Each model has advantages and disadvantages and it's up to the individual performing the forecast to select the best model depending on the needs and the data available (Fildes & Lusk, 1984).

Table 2.1

Comparison between Qualitative and Quantitative Methods

	Qualitative Methods	Quantitative Methods
Characteristics	Based on human judgement and opinions. It is subjective and non-mathematical	Based on Computation and quantifies options mathematically
Advantages	Can leverage latest changes in environment and privileged information	Consistent and Objective and can leverage large quantities of data at once
Disadvantages	Can be biased	Quantifiable data is often unavailable. The forecasts are only as good as the data that is used in forecasting

Qualitative forecasting methods are based on the judgment and opinions of individuals to forecast future outcomes. Since they have a strong reliance on an individual's experience it is commonly riddled with inaccuracies and is considered an unreliable forecasting approach. Surprisingly, SME owners still favor judgmental methods over other more complex quantitative methods because they use wisdom, experience, and foreknowledge.

Figure 2.2

Qualitative Forecasting Models
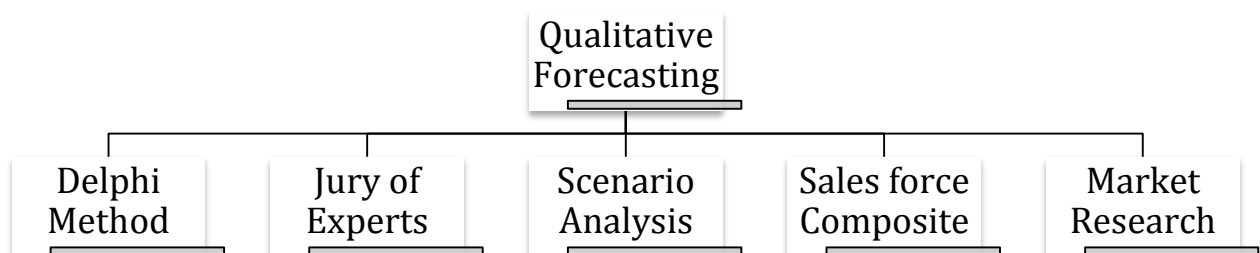
Quantitative forecasting models are used to generate a future forecast using prior data. Mathematics and statistics are used in the historical data to generate predictions. Some of the Models used in such forecasting are time series models such as moving averages and exponential smoothing and causal such as regression and econometrics. Quantitative

forecasting provides an objective view based on the strength of past sales data. It can be difficult for SMEs to get an unbiased peek at how their business is doing. Qualitative forecasting provides an unbiased forecast based on hard numerical data from quantitative forecasting, this provides the necessary information that allows SME owners and managers to get a true feel of the organization's projections enabling them to make informed decisions for the future. Quantitative forecasting models are further classified into Time series models ad Causal models.

Causal forecasting models assume that the forecast variable has a cause-effect association with one or more independent variables. Causal forecasting models habitually take into consideration all potential factors that can affect the dependent variable. The data needed for the before-mentioned forecasting can vary from private transactional data to obvious information like a countries macroeconomics indicators and social chatter, causal forecasting models require constant revision and updates to take into consideration new information. Regression models, econometric model and leading indicator models are some of the common causal forecasting models today.

Time series models chronologically order historical datasets over time. This ordering of data is one of the differences between causal models and time series models. In time series models, it is assumed that forecast components like trends seasonality, and cycles will occur again in the future. Line charts are most commonly used to represent time-series forecasts. The time-series model help SMEs identify cyclical patterns, trends, growth rates, and any deviation or shift in the series of data. Some of the common time series models are moving averages, exponential smoothing and trend Projection algorithms

Figure 2.3

Quantitative Forecasting Model

```
                    Quantitative
                    Forecasting
         ┌──────────────┴──────────────────────┐
     Time Series                             Causal
       Models                                Models
   ┌──────┼────────────┐                       │
Moving   Exponential  Trend                   Line
Average   Smoothing   Projection            Regression
```
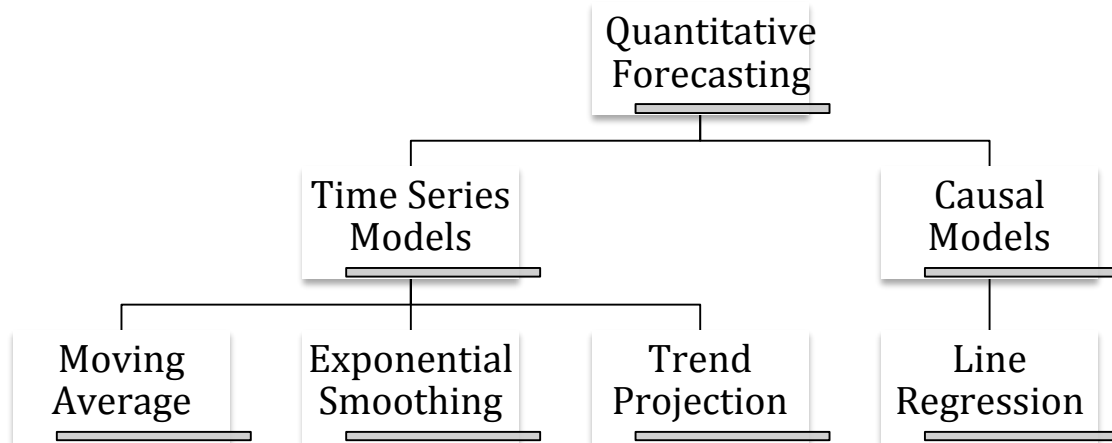
2.6. Forecasting Algorithms

There are four classifications of Predictive analytics; Descriptive analytics, in descriptive analytics, a comprehensive analysis of live data to depict the current state of the business. Diagnostic analytics, in diagnostic analytics, there is an isolation of confounding information, a drill-down on current data to isolate the root cause as to why things are the way they are in the business. Predictive analytics, where, automated decisions using algorithms from historical patterns predict future outcomes. Prescriptive analytics, in this type of analytics, application of advanced analytics to make specific recommendations based on champion - challenger testing strategy outcome (Patil & Bhosale, 2018). Predictive analytics utilizes several methods to analyze current and previous data sets to infer future events. These methods are further categorized into three which are predictive models, descriptive models, and decision models (Halili & Rustemi, 2016). Forecasting models target specific links, and patterns that are known to result in predictable outcomes. By identifying the explanatory variables, the prediction of dependent variables can be done. Descriptive frameworks attempt to create segmentations that

are usually used to categorize customers. Predictive models largely focus on specific events or behaviors, while descriptive models identify the different relationships between variables of interest.

2.6.1. Classification algorithms

Classification algorithms forecast the value of an attribute based on the values of other attributes (Bhardwaj & Pal, 2011). One such classification algorithm is the Bayesian classification, in this classification algorithm, a conditional probability is used to guess a value based on a set of data as input. Bhardwaj & Pal (2011) defined the Bayes classification is as follows;

Equation 2.1

Naive Bayes Classification Algorithm

$$P(c \mid x) = \frac{P(c \mid x)P(c)}{P(x)}$$

$$P(c \mid X) - P(x_1 \mid c) \times P(x_2 \mid c) \times \ldots \times P(x_n \mid c) \times P(c)$$

This algorithm is termed naive because it considers the independence of the attributes. This classification algorithm is considered both a descriptive and predictive algorithm. Naive Bayes algorithm is simple to understand and use. Other classification algorithms require a single scan on the data (Pandey & Pal, 2011). The main demerit of using the naive Bayes classifier algorithm is that it needs minimal training data to establish the parameters required for classification. Independent variables are considered as only the variances of the variables for each class that need to be established and not the whole covariance matrix. Despite their naive design and assumptions, naive Bayes classifiers have worked well in many complex real-world situations. The data classification process entails learning and classification. During the learning process, the training data is evaluated by the classification algorithm. During the classification method, the test data is used to appraise the correctness of the classification rules.

If the correctness is adequate the rules can be deployed to new data tuples (Halili & Rustemi, 2016).

K-nearest neighbors' technique is another classification method the samples which are of the same type will be in each other's vicinity. K-nearest neighbor classifier searches for patterns in space for k-objects that are closest to it and will commission the class by identifying the class label which is mostly used. For the value of k=1, the samples which are closest to unknown samples are given pattern space (Patra & Singh, 2013). The main advantage of k-nearest is that it achieves good results on applications that have a sample with many class labels. While learning is easy, it is more efficient when the training data more. The main detriment of k-nearest techniques is that it is slower than other classification techniques and that it allocates the same weight to each attribute which might confuse when there are many irrelevant attributes (Anuradha & Divakar, 2013).

Decision trees, in a decision tree the "root" is considered as the main node which does not have any incoming edges from any other node. Other nodes in the tree have one incoming edge. Each node is split into two or more nodes. The main advantage of decision trees is that it has a fast learning and classification speed (Senthilnath, et al., 2016). The main disadvantage of decision trees is that a small amount of change in the data can cause varying looking trees and that construction of a decision tree may affect badly irrelevant attributes (Senthilnath, et al., 2016). Senthilnath, et al. (2016) concludes that all the classification algorithm has a set of advantages and disadvantages. He also concludes that each algorithm has a specialized area of implementation and that each benchmark cannot be accomplished by one classification algorithm. In this regard, two or more classifiers can be used to achieve the implementation and performance needed.

2.6.2. Regression algorithm

Regression analysis provides a relationship between a dependent variable and a set of predictors. Regression analysis is considered as a type of supervised learning which encompasses the segregation of the data set into training data and validation data (Halili & Rustemi, 2016; Hetzel, 1988). Regression algorithms showcase relationships between the target output and input features to forecast the value for new data. Regression algorithms can be used in a myriad of scenarios like in finance, trend analysis, and the stock market. Some of the popular regression algorithms include linear regression algorithm, regression trees algorithm, lasso regression algorithm, and multivariate regression algorithm. A simple linear regression algorithm enables users to compile and study the relationships between continuous variables. There are several types of regression algorithms;

Linear regression is a model that provides a linear relationship between the input variables (x) and the single output variable (y). The variables can be computed from a linear consolidation of the input variables (x). When there is one input variable (x), the algorithms are termed as a "simple" linear regression algorithm. Least Absolute Selection Shrinkage Operator (LASSO) is a regression algorithm where shrinkage is considered as a constraining parameter. The main goal of Least Absolute Selection Shrinkage Operator regression is to obtain the subset of predictors that provide the least prediction error for a quantitative variable. Variables with a coefficient that is equivalent to zero (0) post the shrinkage process are removed from the model. Variables with non-zero regression coefficients are strongly correlated with the return variable.

2.6.3. Seasonal Naïve Forecasting

A forecast that is equal to the latest observation in a time series is known as a no change or "naïve 1" forecast. Due to its simplicity, naïve forecasts are usually used as a benchmark to determine if the costs and effort of applying more sophisticated methods are justified by their increased accuracy (Halili & Rustemi, 2016; Hetzel, 1988). Seasonal naïve forecasting is

desirable in data that is seasonal. The forecast is made to be similar to the last value from the same season of the year (e.g., the same week of the previous month). Generally, the forecast for time **T+hT+h** is denoted as follows: -

Equation 2.2

Seasonal Naïve forecasting

$$\gamma T + h|T = \gamma T + h - p(n + 1)$$

Where $p = the\ seasonal\ period,$

And, $n = the\ integer\ part\ of\ (h-1)/m$

Considering daily forecasting, the forecast for a similar day in the future is equivalent to the last observed day value. Considering a scenario of monthly data, the forecast of all similar months in the future (e.g. January) values is equal to the last observed January value. The same logic applies for other months and quarters, and other periods of the forecast.

Table 1.2

Seasonal Naive Forecast

Month (t)	Sales (A)	Forecast (F)
Jan 2020	31	-
Feb 2020	37	-
Jan 2021	40	31
Feb 2021	30	37

As shown in the table above, the seasonal naïve forecast for Jan 2021 is the actual sales for Jan 2020, this is similar for Feb 2021 which is the actuals sales for 2020. The prediction confidence levels of this model can be evaluated by eventually checking the actuals vs the forecast value. For our case for example, the prediction confidence for Jan 2021 would be (Sales A – Forecast F)/Sales A. which translates to 100 - {((40-31)/40) *100} = 77.5%. in R Programming, this can

be achieved by the use of the *snaive ()* function. This function accepts two parameters as *snaive (y, h)* where y is the predictor and h is the forecast period.

2.6.4. Naïve Forecasting

I naive model, the forecasts are equal to the last observed value. Naïve forecasting is denoted by the formula:-

Equation

2.3 Naïve forecasting

$$yT + h|T \ = \ yT$$

Table 2.2

Naive Mode forecasting.

Date (t)	Sales (A)	Forecast (F)
1	21	21
2	27	21
3	30	27
4	25	30

In the Naïve Model, previous observed values are used as forecast for the next period, in this forecast, the assumption is that the data has no seasonality. As displayed In the table above. The forecast for the 1st day is same as the actuals because there is no previously observed value that can be used to perform the forecast. The forecast for the 2nd day is the actual value from the 1st day and the forecast for the 4th day is the actual value from the 3rd day.

2.6.5. Simple Exponential Smoothing Algorithm

Exponential smoothing utilizes a weighted average procedure with measures declining exponentially as data age. The forecast for the following period (t) is computed using the following formula.

Equation 2.4

Exponential smoothing forecasting

$$Ft = ft - 1 + (At - 1 - ft - 1)$$

Another way to calculate this is

$$Ft = At - 1 + (1-)Ft - 1$$

Where is the smoothing coefficient with a range of between 0 and 1. A is the observation and

F is the forecast value. A popular feature of the exponential smoothing method is that forecasts

made include a portion of every piece of historical demand. There are contrasting weights fixed

on the historical values, with older data given lower importance.

Table *2.4*

Exponential smoothing forecasting.

Date (t)	Sales (A)	Forecast (F)
1	31	30
2	37	30.1
3	40	37.1
4		40.3

The calculation for the forecast for day 2 is as follows $31+ 0.1(31-30) = 30.1$

$where$ = 0.1. Note that the Forecast for Day 1 is a guess because there is no historical data

for the calculation. In R Programming this can be achieved by the function **ses(y, h).** where y

is the predictor and h is the forecast period.

2.6.6. Autoregressive Integrated Moving Average (ARIMA) Model

Auto-Regressive Integrated Moving Average refers to a class of models that try to elaborate a

given time series depending on previous values (Hyndman & Khandakar, 2008). In the auto-

regression model, a variable forecast basis is the compounding of previous values of the variable. The ARIMA model of order **n** is denoted as

Equation 2.5

Autoregressive forecasting

$$y_t = C + \emptyset_1 \, y_{t-1} + \emptyset_2 \, y_{t-2} + \cdots + \emptyset_n \, y_{t-n} + \varepsilon_t$$

Where ε_t = white noise.

The autoregressive forecasting algorithm resembles multiple regression albeit with lagged values of y_t as predictors. This is model is considered to be an autoregressive model of order n AR(n).

The Moving average forecasting model utilizes previous forecast errors. It is denoted by the formula:-

Equation 2.6

Integrated forecasting

$$y_t = C + \varepsilon_1 + \emptyset_1 \, \varepsilon_{t-1} + \emptyset_2 \, \varepsilon_{t-2} + \cdots + \emptyset_p \, \varepsilon_{t-2}$$

Where ε_t = white noise.

This is referred to as a moving average model of order, q MA(q).

A combination of the two forecasting algorithm is jointly considered as a non-seasonal ARIMA forecasting model.

Equation 2.7

Autoregressive Integrated Moving Average (ARIMA) forecasting

$$y'_t + c + \emptyset_1 y'_{t-1} + \cdots + \emptyset_p y'_{t-p} + \emptyset_1 \, \varepsilon_{t-1} + \emptyset_q \, \varepsilon_{t-q} + \varepsilon_t$$

Where y'_t = differenced series. This is an ARIMA(p,d,q) model where

p = autoregressive.

d = differencing;

q = moving average.

2.6.7. Clustering algorithm

Clustering techniques identify distribution patterns and relationships among datasets. Predictive analytics clustering is a framework that combines clustering and forecasting (Blockeel et al., 1998). K-means clustering is a segregation based algorithm. It segregates 'n' data sets into K sets of. Cluster resemblance is computed by the average value of the objects in a collection, which can be identified as the center of the cluster. This follows an initial randomly selected K the objects as the average of the cluster. Objects are assigned to the K clusters which have the lowest distance between objects and center of the cluster. The average is updated until all the objects are assigned a mean value this process iterates until the assignments are stable (Bala, et al., 2014). Hierarchical Clustering technique either combines or divides similar data elements by developing a ranking of clusters called dendrograms (Rajalingam & Ranjini, 2011). This method creates collections progressively. It is categorized into two, Agglomerative algorithm and Divisive algorithm. (Rajalingam & Ranjini, 2011)

Expectation–maximization (EM) algorithm is a recursive technique that helps in for identification of the likelihood or Maximum a Posteriori (MAP) assessments of specifications in an analytical system, where the system largely relies on latent variables. The EM recurs between performing an expectation step that creates a function for the expectation of the probability computed using the current estimate of the parameters, and a maximization step, which in turn computes parameters with a view of increasing the expected probability found on the E step (Dobbin & Simon, 2011)

Another clustering algorithm is the density-based algorithm that is used when clusters are irregular. It identifies the main objects i.e. objects that have dense zones. It then attempts to connect core objects and their zones to create denser clusters.

2.6.8. Machine learning algorithms

Machine learning is used to train computer systems on how to act on datasets better (Richert, et al., 2013). Neural networks (NN) assign all inputs to one of the identified output classes sorely based on an indicator value. After an input value runs through a sequence of batches of activation systems, it is considered to have been processed (Hatem et al., 2018).

Figure 2.4

Artificial Neural Network
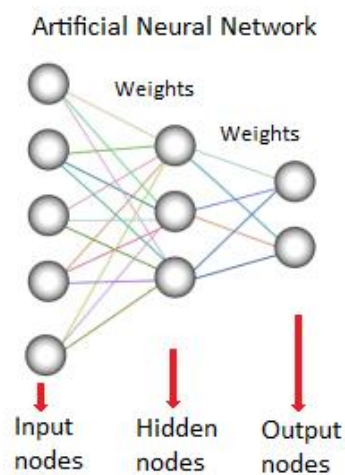
Figure 2.0-1 Artificial Neural Networks

In Random Forest (RF) Algorithm, the predicted result is a distinct value (Breiman, 2001), Studies on random forest algorithm have affirmed the argument that the random forest algorithm has a huge prediction efficiency with superb resilience for abnormal values and noise. It trains all trees separately using an indiscriminate sample of the data (Biau & Scornet, 2015) that result in the high efficiency. Some key aspect of machine learning is supervised learning, unsupervised learning, and reinforced learning. Supervised learning uses a training dataset to teach models to yield the desired output. in this learning methodology, the main goal is to map an output Y=f(X) from input variables (X) and output variables and using an algorithm. Unsupervised learning provides the model to independently identify and uncover patterns and information usually in unlabeled data. in this model, the input data(X) and has no output variable with an aim of learning and modeling the underlying composition.

Reinforcement learning provides both qualitative and quantitative frameworks (Dayan & Niv, 2008) .Reinforced learning concerned with how intelligent systems ought to make decisions in an environment.

2.6.9. Evaluation of Prediction Accuracy

The Evaluation of predictive performance can be assessed by use of several common methods; these methods rely on the test data set which provide as more objective basis than the training period in the assessment of the prediction accuracy. Before evaluating a forecasts accuracy, a distinction between bias and precision must be made. Bias represents the historical average error. In short forecasts will on average be higher or lower than the actual while precision measures the spread between forecasts and actual value. To evaluate a forecast's accuracy, a comparison of the forecast to the actual outcome is needed. This is always a backwards looking forecast which unfortunately, does not provide insight into the accuracy of the forecast in the future and it is impossible to establish this accuracy. The common forecasting accuracy measures are as follows;

Table 2.5

Predictive Analytics Accuracy

Accuracy Measure	Calculation
Mean Absolute Error	$MAE = average(\lvert e_t \rvert)$
Mean Squared Error	$MAE = average(\lvert e_t^2 \rvert)$
Mean Absolute Percentage Error	$MAPE = 100 \times average(\lvert \frac{e_t}{y_t} \rvert)$
Mean Absolute Scaled Error	$MAPE = MAE/Q$

where Observation is y_t, forecast is $\widehat{y}_t$, then forecast error is $e_t = y_t - \widehat{y}_t$,

I)   Mean Absolute Error

Mean absolute error (MAE), is the degree of errors amongst paired observations, it is considered as the simplest measure of accuracy. MAE indicates the degree of error we can expect from the forecast on average. MAE is not scaled to the mean; this makes it unsuitable for measuring accuracy.

II)   Mean Squared Error

Mean Squared Error avoids the pitfalls of positive and negative forecast errors offsetting each other. This is achieved by calculating the square of the forecast error and then using the mean of the squared forecast error.

III)   Mean Absolute Percentage Error

Mean Absolute Percentage Error (MAPE) is the sum of the individual absolute errors divided by the actual forecasts. It is the mean of percentage errors.

IV)   Mean Absolute Scaled Error

The Mean Absolute Scaled Error (MASE) is a measurement of forecast accuracy, the scaling of the errors is derived from the in-sample MAE of the naïve forecasting method. in this method, one-period-ahead forecasts is generated from each data-point in the sample.

 2.7. Challenges of software systems implementation in SMEs

As described in previous sections, predictive analytics and forecasting methodologies are highly technical subjects and may not be easily understood by the end users. This poses a great risk to the success of any software development project. End users having never used forecasting models before might also have unrealistic expectations of the application being developed. Zowghi et al. (2015) further, In their paper on problems and challenges of user

involvement in software development, explains the difficulty of involving users especially those who are not experienced in software development projects in the development of complex systems is a major challenge as the contribution by these users is limited and might not be helpful. They go ahead to describe effective communication of goals and the intent of developing an application is of high importance. They explain expectation management as key to the success of the software project and smooth user acceptance.

In Kenya, access to the internet is mainly through mobile devices, with 75% of users accessing the internet via small devices (Statcounter, 2021), the development of a forecasting model that is intended for use by Kenyan SMEs need to take into account that the users will largely be using small screen devices. Development in cascading style sheets allow for the development of responsive web applications that allow end users to use software applications from various devices and browsers, this allows for a greater reach in audience as opposed to native mobile applications that might have problems with different operating systems of the devices in use. web Applications are easier to design and implement and less platform-specific as compared to mobile applications (Turner-McGrievy , et al., 2017). Responsive Web Design (RWD) is a software design strategy that enables web applications to render well in multiple environments and device screens. This approach has taken off in recent times because more and more people continue to access the internet through mobile devices and, there has been a tremendous need to have a design approach where one web application can be run in multiple environments and is adopted appropriately for easy use in those environments.

The effective communication of the goals and strategic intent of developing new software or making major changes to an existing one sets the overall benchmark for what the users should expect from the delivered system. Furthermore, looking at users' expectations from a finer-grained point of view, requirements specification document will set the expectations of the

users of the end product and are used by them during UAT The effective communication of the goals and strategic intent of developing new software or making major changes to an existing one sets the overall benchmark for what the users should expect from the delivered system. Furthermore, looking at users' expectations from a finer-grained point of view, requirements specification document will set the expectations of the users of the end product and are used by them during UAT

2.8.Summary, gaps and opportunities from previous works

After initial analysis of the available data and the objective of the study, regression algorithms would best be used to achieve the goals of the research, forecasting of demand and supply has a time element attached to it, this means that the forecasting data will be converted into time series data for easy forecasting of future values. This is evident in previous time series data forecasting done previously by researchers like Thong and Yap (1995) and Yan et al, (2018) who were forecasting future time series values. The designed model should be designed in a way that allows the majority of the users in Kenya to access it through their mobile devices, Responsive Web Design (RWD) approach would best suit this need as the design model can be accessed through multiple devices with different screen sizes.

CHAPTER THREE

3.0. METHODOLOGY

3.1. Introduction

This chapter discusses the appropriate research methodologies that are appropriate in achieving the objective of the study. (Mouton, 1996) emphasizes that the overriding principle in the choice of the various techniques is that they must be appropriate for the task at hand. This chapter will analyze the modeling technique that will be used to develop the forecasting model and some software development techniques that could be used in the delivery of the final model.

3.2. Predictive analytics modeling and software development processes

According to (Eckerson, 2006), Predictive can be developed in six basic steps as shown in the figure below.

Figure 3.1.

Predictive Modelling Process
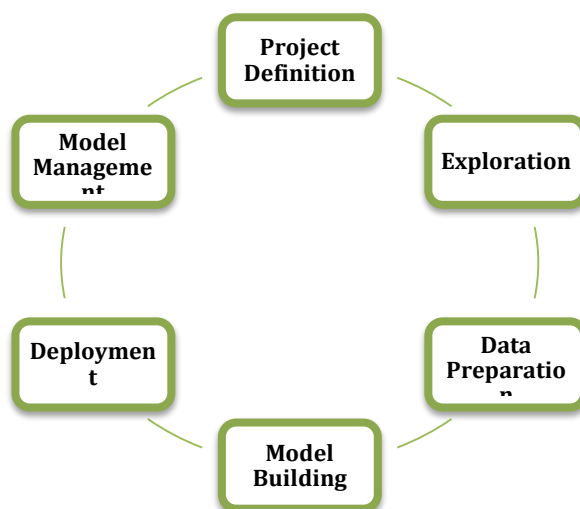
I)   Project Identification and Definition

Here the purpose and chosen outcomes for the project are outlined and translated into predictive analytic targets and tasks. The researcher needs specific targets and outcomes at this stage that can be evaluated later to gauge the success of the model developed.

II)   Data Exploration

The availability of data will determine how the model will be developed. An analysis of the available source data to identify the most relevant data to be used in developing the model

III)   Data Preparation

In this step, data extraction from the source system is carried out. An extraction transformation and Loading (ETL) process is usually followed to mine data from source systems. The transformation process transforms data into its final format as per the forecasting objective.

IV)   Model Development

When creating a forecasting model, an approach of development, testing, validation, and evaluation is generally followed. The objectives identified in the project definition are evaluated at this point to check if they have met the desired expectations.

V)   Model Deployment

This is the final step in the model design process where the model is finally delivered in an environment that allows the model to be used by users end users to derive business decisions or processes. This environment is usually referred to as a production environment.

3.2.1. Ad hoc Software development

This is a software development technique that does not have any structured planning. In this methodology, there are no guidelines or processes. The advantage of this technique is that it is easy to learn, however, the disadvantages are that there are no clear start and end of tasks,

there is no task prioritization leading to ignoring of a more important task, unable to scale the application, not easy to review the application.

3.2.2. Waterfall model

There are two steps common to the development of computer systems regardless of the size or complexity, the first step is the analysis and the second is coding. This method assumes that highly complex systems can be built in a single step without revisiting requirements. It does not take into account any changing business requirements or technology. The main advantage of the waterfall model is that at the end of every phase there is a detailed evaluation of the project, a thorough evaluation of the phase can be done.

Figure 3.2

Waterfall software development model
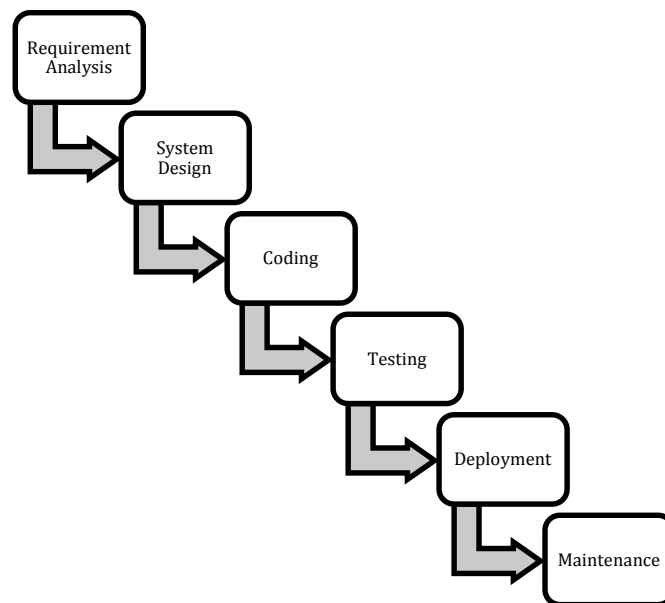
One of the disadvantages of a waterfall project is that there is no room to accommodate frequent customer changes and that the product can only be seen in the later phases of the project.

3.2.3. The V-shaped software development model

The V-model is a software development life cycle technique that requires the execution of processes to occur in a sequential V-Shaped manner known as verification and validation process. It is based on the association of a testing phase for each corresponding development stage (McHugh, et al., 2013).

Figure 3.3

V-shaped Software Development Model
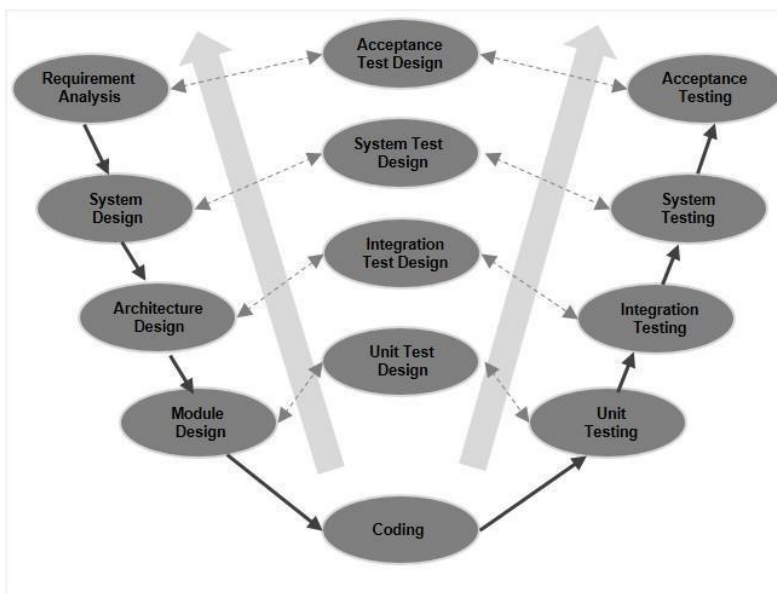
The main advantage of this model is that it is easy to comprehend, manage, and use. The drawback of the model is not flexible to changes.

3.2.4. Prototyping model

The Prototyping Model is a software development life cycle where an initial proof of concept of the final solution or product is developed, tested, and then remodeled based on end-user inputs until a successful prototype is finally developed from which a complete solution is finally developed from. This model works best when the requirements are not initially known or clear.

Figure 3.4

Prototyping Software Development Model
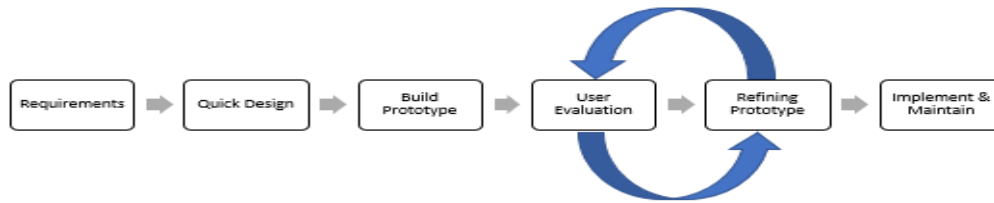
3.2.5. Agile model

According to the agile manifesto, the greatest priority is to satisfy the end-users by having their inputs early and incorporates these inputs continuously throughout the software development process. Agile software development methodology prioritizes a high end-user engagement throughout the software development process. This is usually done through regular walkthroughs and demonstrations of the current state and release of small parts of the software to production giving customers a chance to start using the solution and provide feedback before the solution is even fully completed (Szalvay, 2004).

Figure 3.5

Agile Software Development Model
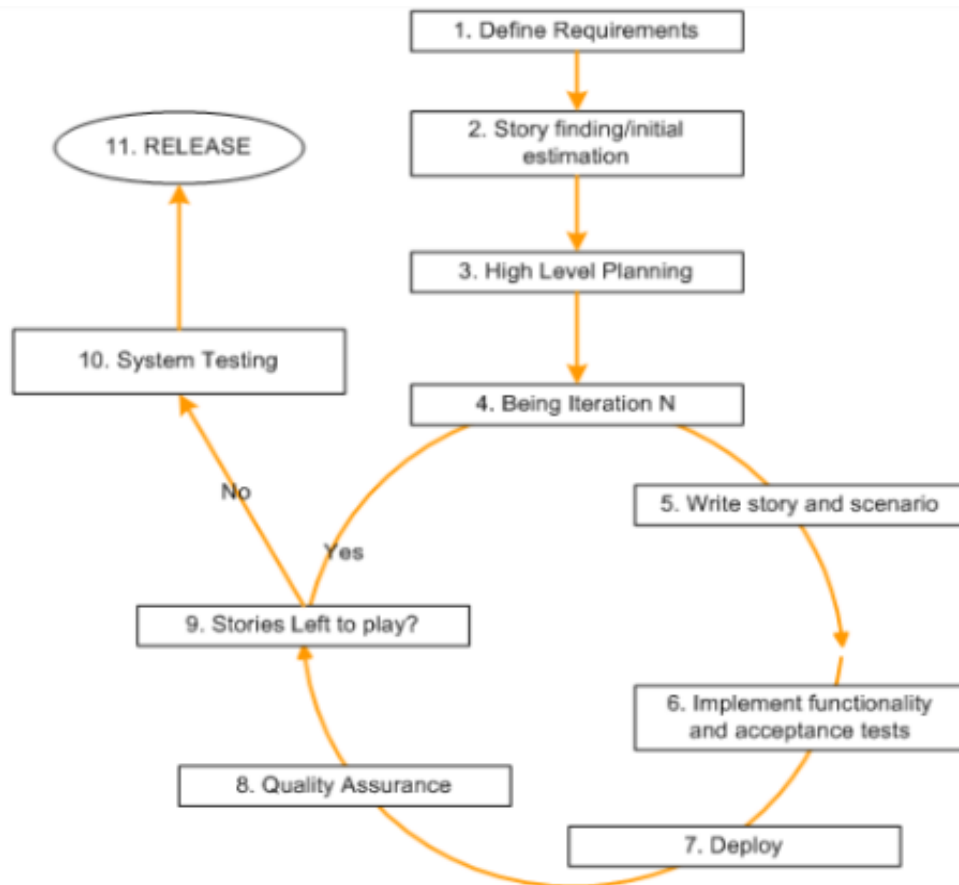
The choice of which software development methodology used will generally depend on the type of system being designed, the size of the project, the delivery mechanism amongst others. Each methodology has its own advantages and disadvantages as discussed in the table below.

Table 3.1

Advantages and Disadvantages of various Software development methodologies.

Model	Advantages	Disadvantages
Waterfall	Has clear structure and what needs to be done in the phase	Change implementation is difficult and costly
	End Goals are determined early	Excludes end user and clients during until the end of the project
	Information in well managed	

		Delays testing until the end of the project.
V shaped	Simple to use	Rigid and not flexible to change
	Testing activities can begin before coding	Software is developed during implementation so there is no early visibility of the end product
	Proactive defect tracking and prevention of defects from downward flow	
	Good for small projects with well understood requirements	Changes in requirements results in changes of all related documents and tests
Prototyping	Active involvement of end users	Scope creep is usually a problem
	Users have a better understanding of what is being developed and have ownership of the end product	Incomplete or inadequate problem analysis
	Errors are detected early	
	Missing functionality can be identified early	
	Good for scenarios where the requirements are not fully understood or where the end user is unsure of what is needed.	
Agile	Improved quality	Requires dedicated cross functional teams
	Focus on business value	
	Stakeholder engagement	Requires agile management tools for management
	Transparency Early and Predictable delivery	
	Easily accommodates change	

These methodologies can be used individually or as a combination of two or more depending on the end product being developed and designed.

3.3. Selected Methodology

With the view of the methodologies discussed, the methodology used in developing a predictive analytics model was a variant of prototyping and agile. Agile prototyping is an

efficient technique that allows customers an early preview of the design process, these previews are repeated in a design, test repeat cycles that helps in discovering the user's needs. It helps in focusing on the end-users' needs and priorities and at the same time helps in evaluating the idea being tested. This Methodology was used because it provided an opportunity for quick evaluation of the final forecasting model early enough. Prototypes can be modularized into functions allowing for the identification of errors and misunderstood requirements. This methodology is not subjected to rigid pre-planning of phases and maturity of the software development process but provides the early confirmations of requirements or questions. The research was designed into three distinct phases.

Figure 3.6

Three phase development process
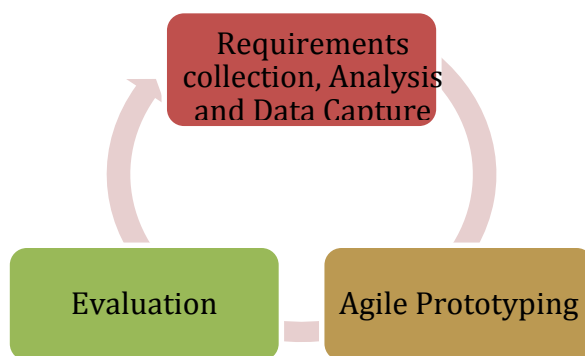
I)   Requirement collection, Analysis and Data Capture

Requirements gathering and analysis are one of the main activities in software development that has to be executed correctly and all the requirements are understood by both the development team and the end-users. If misunderstood, the end product will not meet the end user's needs, while time and effort would have been used in the development of the product. It is therefore important to capture the user's requirements correctly. However, when designing software solutions for users who have not used a similar system in the past

presents a myriad of challenges for example when users pose unrealistic expectations or unclear requirements (Zowghi, et al., 2015).In such cases, agile prototyping becomes a good methodology as the communication between the end-users and the development team is increased leading to early clarification of expectation and stimulation of interest and engagement from the user throughout the development of the product.

II)   Coding (Agile Prototyping)

As discussed earlier, agile software development allows for the development of software systems in small chunks and getting feedback from the client. Prototyping is in a way related to agile software development in that a prototype can be designed quickly and presented to the client for feedback. The difference is that in Agile development, requirements are well understood by both the client and the developer while in prototyping, one party is not fully aware of the requirements and the prototype provides a talking point and potential use case. This is the approach used for this research as the end-users targeted by the research were SME owners and managers whose organizations had never used forecasting applications before and Thurs were not aware of the possibilities and the for the model would take. in this case, a prototype would be designed and presented to the end-users for feedback and the feedback would be implemented and be reviewed in the next demo session.

III)   Evaluation

Pan explains that it is not possible to test quality directly, certain quality related factors can be measured to evaluate quality of a software system (Hetzel, 1988). Hetzel goes ahead to list some of the software quality factors that can be evaluated to determine software quality (Hetzel, 1988).

Table 3.2

Typical Software Quality Factors

Functionality	Engineering	Adaptability
Correctness	Efficiency	Flexibility / Scalability
Reliability	Testability	Reusability
Usearbility	Documentation	Maintenability
Integrity	Structure	

In the evaluation phase, Reliability, Usability, Documentation, and Scalability were taken into consideration for the evaluation of the forecasting model. The main purpose of this phase is to verify with the user that the designed solution meets expectations. If the expectations are met then, based on the type of the software development project, then project closure is conducted. For this research, Evaluation was conducted as a survey to the user group, the results of the survey would determine if the requirements have been met. There are two variations of user acceptance, intermediate user acceptance testing which is done at the end of each software development stage, or final user acceptance testing that is done at the stage of project completion (Jeong, et al., 2015). Only the final user acceptance testing was carried out for this research and its main objective was to confirm the operational readiness of the model, the accuracy of the model and the useability of the model. On completion of the development of the prototype, users were given access to the platform where they used the forecasting model on their business, during this pilot phase another survey was sent to the participating SMEs to evaluate the functionality and the efficiency of the forecasting model. An acceptance criterion for the forecasting model was designed for the end users.

Table 3.3

User Acceptance Checklist.

ID	Description	(Pass/Fail)	Comments
1	As a user, I should be able to Access the Forecasting model on a mobile device's browser and on a Desktop Computer's browser So that I can make sales forecasts		
2	As a user, I should be able to Upload Transactional data as per the provided template So that, I can see the Daily forecasts And Analyze the uploaded data		
3	As a user, I should be able to Analyze uploaded data by - Transaction type - My custom Category - Location - Service type - Sector		
4	As a user, I should be able to see forecasts uploaded data by - Transaction type - My custom Category - Location - Service type - Sector		
5	As a user, I should be able to select a desired forecasting algorithm from the available list		
6.	As a User, I should be able to select the forecast period using the slider		
7.	As a user, I should be able to navigate to the model documentation site.		

3.3.1. Design of Data Collection Instrument

This data collection was done through online questionnaires technique. The questionnaires where structured and administered to the respondent. Follow up video calls were done with the respondents who required further information or clarifications.

The questionnaire consisted of 4 parts as shown in APPENDIX 2:

I)   Sales and purchase data capture

Sales and purchase transactional data for the month of April 2020 was solicited. This would form the main variable used to forecast demand and supply. The data captured were Date of Sale, number of Sale quantities per day, Number of sales returns per day, Number of fulfilled purchase orders per day and the number of cancelled purchase orders per day.

II)   Location and Size of SME

The Location of the and the Size of SMES was also captured in the questionnaires, this helped to understand the location of the business from a geographic perspective, and also the economy of the area for example rural area or Urban area. The size of the enterprise with regards to the number of employees was also captured as per the definition of SMEs in Kenya described earlier.

III)   Business main business classification

The respondents were asked to classify their enterprise according to the global industry classification standard. (Morgan Stanley Capital International, 2021)

IV)   Business Type

The respondents were also asked to provide details about how thy classify their business and the options were, Merchandising Business, Service Business, Manufacturing Business, Hybrid Business.

3.3.2. Evaluation Questionnaire

An evaluation questionnaire was sent to the user participating in the pilot phase of the app, the objectives of the evaluation questionnaire was to establish the impact of the designed model on and its applicability in performing forecasts for SMEs.

3.4. Research Schedule

Table 3.4

Project Schedule.

Task	Duration (Days)	Output
Proposal Acceptance	1	Approved Project Proposal
Initial requirements analysis	10	Backlog
Backlog Prioritization	1	Prioritized backlog
Development	30	Framework Development
End to End Testing	30	System testing report
Pilot	5	Acceptance Report
Deployment & monitoring	3	Deployment Report
Documentation	30	Full Documentation of the Framework (Continuous process throughout the project)

Upon approval of the proposal to begin research, the initial requirements analysis was carried out as per the schedule above

3.5. Research Budget

Table 3.5

Project Budget.

Item Description	Qty	Amount (KES)	Total (KES)
Development Laptop	1	120,000.0	120,000.0
Printing & Photocopy	1	5,000.0	5,000.0
Transport & Logistics	1	20,000.0	20,000.0
Web hosting	1	3,000.0	3,000.0
Domain name	1	2,500.0	2,500.0
Publishing	2	5,000.0	10,000
		Total :	**160,500.00**

The budget for the for the research was fully through self-funding. There were no attempts or solicitation of for project funding.

3.6.Technology and Tooling

SMEs owners and managers will need to interact with the forecasting model designed conveniently through web interfaces. Web interfaces are not limited to computers and can be accessed through mobile devices through web browsers.

3.6.1. Client Server Architecture

Client-server architecture is also known as the Client-Server Network or Network computing Model this is because in client-server design all services and requests are spread over the network. Client-server works like a distributed computing system. All components perform their tasks independently from each other. There are three main tiers in client server architecture.

I) 1 Tier (Single tier) Architecture

In the 1-tier client-server architecture, all the client-server user interface environment, configuration setting, and data logic are on the same system. Since both the client and the server exist on the same machine, these systems are not commonly used because they are not easily scalable.

II) 2 Tier Architecture

In the 2-tier client-server architecture, the database is stored in the server machine and separate from the client-side. Here, the database is more secured as it is not exposed to the user in the client machine. several clients can make requests to the server in this architecture.

III) 3 Tier Architecture

The 3-tier client-server architecture is an extension of the 2-tier architecture. There are three layers, the presentation layer, the application layer, and the database layer. The main goal of a 3-tier architecture is, to segregate the application layer from the database layer so that there are program-data independence and support varied views of data.

Figure 3.7

Three tier Client/Server Architecture
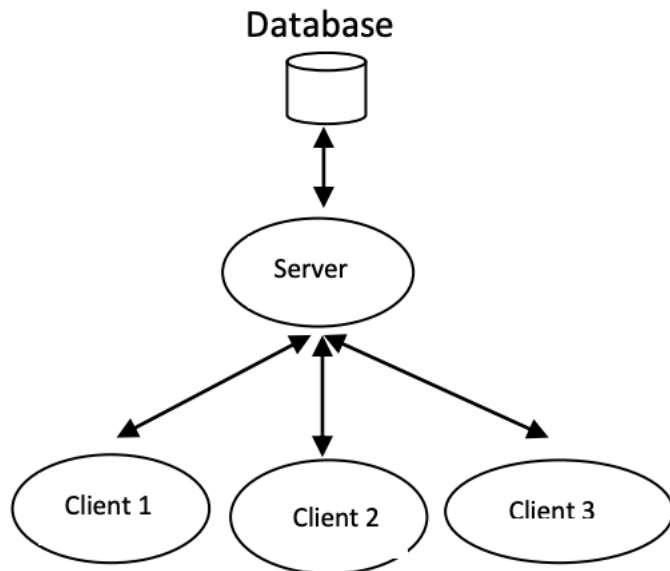
With the emergence of cloud computing, software as a service (SaaS) has taken the mainstream. SaaS is one of the most popular forms of cloud computing. Customers can access offerings hosted by a service provider over the Internet without the need for local installation of the application. The infrastructure and hosting logistics, and overheads are removed from the client. The client can use the application from anywhere as long as there is a connection to the cloud infrastructure.

For this research, the 3-tier approach was used, this allows the hosting of the model and having SMEs connect to the software from anywhere in the world. SME Owners can opt to host the application locally in their premise of use the publicly hosted version.

3.6.2. Programming language

This Research was implemented in R and Rshiny (R studio, 2021), R is a powerful language for analytical computing that combines lazy functional features and object-oriented programming (OOP) (R Foundation, 2021). According to the R Foundation, R compiles and

runs on multiple platforms including, UNIX platforms, Windows and macOS operating systems. R can is an implementation of S. While there are some differences, code written in S can run in R. It is an integrated suite of software modules use for data manipulation, calculation, and graphical display. It contains an effective data handling and storage capability, calculation suites, data analysis tools, and graphical tools amongst other features (R Foundation, 2021). R is a as an environment in which statistical techniques are deployed, implemented and can be extended greatly using packages from the CRAN that covers a variety of modern statistics and statistical solutions (R Foundation, 2021).

Shiny is an R package that makes it easy to build interactive web apps straight from R. Shiny provides the capability to host the designed model on shiny's cloud platform, shinyapps.io, or on-premises with open-source Shiny Server, which offers features such as multiple apps on a single server and deployment of apps (R studio, 2021). In this research, the designed model was hosted on shiny's cloud platform.

R shiny has user reactive expressions that allows that give control as to which parts of your user interface need to change whenever you have large computation. Cascading Style Sheets (CSS) was used in the implementation of the presentation layer, this was specifically important since it allowed the adaption of the presentation layer for small devices. Mobile devices are widely used by SME owners and it was important that this group of users be considered in the development of the user interface.

3.7. Summary of selected methodology and technology used

Forecasting model is undertaken just like a software development project, for this research a hybrid of agile software development and prototyping known as agile prototyping was used.

This allows for both the benefits of having a sample of the end product quickly while at the same time having the end users as a core part of the review process and the acceptance of the model. This helps when the users are unclear about the requirements of the system either because they have not used a similar system before or in cases where users are unsure about what functionality they need. R, Rshiny, Reactive R and Cascading style sheets were used in the implementation of the model with the final model being deployed in Rshiny's Cloud infrastructure.

CHAPTER FOUR

4.0. RESULTS AND DISCUSSIONS

4.1. Introduction

This chapter discusses the designed demand and supply forecasting model designed based on the collected training data. This chapter will discuss the training data that was used and the actual forecast results from the designed model.

4.2. Respondents

For this study, the data that was collected were sales and purchases data for the month of April 2020 from 5 small and medium enterprises in Nairobi and Muran'ga counties in Kenya. This data was used to model the training and the testing data. The data was used to help the firms forecast the demand and supply for the first two weeks of May 2020. For clarity, demand was modelled from the SMEs sales, the more sales an SME makes, the more the demand for the product. The supply was modelled from the SMEs purchases from its suppliers, the more the purchase fulfilments the more the supply. The SMEs were classified into 3 locations types, Rural areas, where the SMES operated fully in a rural set up, Urban Area where the SME operations were fully in an urban set up and both rural and Urban centre where the smes had operations in both rural and urban areas.

Table 4.1

Respondents Location distribution.

SME Location	Number	percentage%
Urban	3	60%
Rural Area	1	20%
Both Rural and Urban	1	20%

The SMEs were categorized into sectors based on the global industry classification standards (Morgan Stanley Capital International, 2021). This categorization helped in identification of any sector specific trends and behavior that could be modeled into the framework.

Table 4.2

Respondents SME Category.

SME Category	Number	percentage%
Energy	1	20%
Materials	2	40%
Industrials	1	20%
Consumer Discretionary	1	20%
Consumer Staples	0	0%
Health Care	0	0%
Information Technology	0	0%

The SMEs were further categorised based on the end product or service and how this end product was achieved. Merchandising businesses sell goods that are bought on wholesale and then sell them at a higher retail price. Manufacturing Business are enterprises that buy raw materials and process these raw materials to come up with a final product. Usually, the processing changes the state of the raw material or combines several raw materials to come up with a final product. Service business are business that provide services to its client while goods and service business combine both the provision of a service and the sale of goods. Hybrid business functions with multiple means to reach its clients.

Table 4.3

Respondents SME Product Type.

SME end product	Number	percentage%
Merchandising Business	1	20%
Service Business	1	20%
Manufacturing Business	0	0%
Hybrid Business	1	20%
Goods and Services	2	40%

4.3. Training Data Preparation and Analysis

The following data was identified as necessary for the forecasting of Supply and demand irrespective of the forecasting method in use.

Table 4.4

Prepared Training Data

Custom Category	Date	Sales	Sales Refund	Purchases	Purchase Cancelation	Location	Service	SME Size	Sector
Store A	1-Apr-20	0	1	0	0	Rural & Urban	Goods & Services	50	Energy
Store A	2-Apr-20	0	0	0	0	Rural & Urban	Goods & Services	50	Energy
Store A	3-Apr-20	2	0	0	0	Rural & Urban	Goods & Services	50	Energy
Store A	4-Apr-20	33	3	0	0	Rural & Urban	Goods & Services	50	Energy

The research respondent was able to download this template directly from the designed forecasting platform and edit it locally in with their own data. The file was then be uploaded back into the platform for forecasting.

i. **Custom Category** – This can be merchants own categorization, it can be the categorization of the business or products. Examples of possible values are [Product Names, Product Codes, Store Names, Salesperson Names]. This option provides flexibility for classification.

Table 4.5

Custom Category used to identify a store

Custom Category	Date	Sales	Sales Refund	Purchases	Purchase Cancelation
Store A	1-Apr-20	0	1	0	0
Store B	1-Apr-20	0	0	0	0
Store B	3-Apr-20	2	0	0	0
Store A	4-Apr-20	33	3	0	0

Table 4.6

Custom Category used to identify a product.

Custom Category	Date	Sales	Sales Refund	Purchases	Purchase Cancelation
Product A	1-Apr-20	0	1	0	0
Product B	1-Apr-20	0	0	0	0
Product B	3-Apr-20	2	0	0	0
Product A	4-Apr-20	33	3	0	0

Table 4.7

Custom Category used to identify a Salesperson.

Custom Category	Date	Sales	Sales Refund	Purchases	Purchase Cancelation
Salesperson A	1-Apr-20	0	1	0	0
Salesperson B	1-Apr-20	0	0	0	0
Salesperson B	3-Apr-20	2	0	0	0
Salesperson A	4-Apr-20	33	3	0	0

The Custom Category is unique for each day and there cannot be duplicate custom categories for the same day. The quantities indicated for sales and purchases are thus considered the sum for custom category item for a specific date. E.g. Total Sales for Store A on1st April or Total Sales of Product A on 1st April. The Custom Category field can only take one variable and cannot be used to track multiple attributes.

ii.   **Date** – This takes the format Day, Month, Year e.g. (1, April 20). This records when the sales, refunds, purchases, and purchase occurred for each custom category attribute. There can be only one day's entry per each custom category attribute.

Table 4.8

Invalid use of date attribute for forecasting.

Custom Category	Date	Sales	Sales Refund	Purchases	Purchase Cancelation
Salesperson A	1-Apr-20	0	1	0	0
Salesperson A	1-Apr-20	1	0	0	0
Salesperson A	3-Apr-20	2	0	0	0
Salesperson A	4-Apr-20	33	3	0	0

The combination of date and custom category attribute was designed to be unique. The above table contains duplicate entries for 1st April 2020 and for this case, data cleaning needs to be done by the SME's.

iii.   **Sales** – Sales captures a complete Sale order and does not take into account the products within the sales order. If users need to capture product-wise sales, then the sales here have to be in conjunction with having the product code or product description in the *custom category* field.

iv.   **Sales Refund** – A sales return is an alteration to sales that happens from a substantiated return by a purchaser of a product that was previously purchased.

v.   **Purchases** – Purchases by the SME from its suppliers. The unit of measurement here is independent of the sale quantity.

vi.   **Purchase Cancelation** – Purchase cancelation refers to when an SME returns purchased products back to its suppliers.

vii.   **Location** – The Location from which business runs its operations can result in different behavior in terms of demand and supply, for this research this considered the following possibilities, Rural Area, Urban Area, and Both Rural and Urban Area.

viii. **Service** – Service refers to the type of undertaking the SME carries out. The possible options here are; Merchandising, Service, Manufacturing and Hybrid Business,

ix. **Company Size** – According to the Micro and Small Enterprise Act of 2012, small and medium enterprises employ 100 employees.

x. **Category** – Categorization here refers to sectors in which an SME can belong to, according to the Global Industry Classification Standard, Sales and sales refunds represents demand forecasting from the perspective of the SME managers and owners. Ideally, a business will strive for high sales and low sales refunds

Purchase and purchase cancelations represent supply. Ideally, a business will strive to have just enough quantities to fulfill its sales commitments. Having a high supply and low demand implies that a business has products that it cannot sell while having high sales and low supply means that a business cannot fulfill its customers' demand (Whelan & Msefer, 2001).

For this research, the training data from each SMEs was grouped into two parts, training data and testing data. In 2011 Dobbin and Simon developed a non-parametric algorithm for determining an optimal splitting proportion that can be used in a dataset. They performed multiple simulations to better understand the factors that determine the best split proportions and to evaluate commonly used splitting strategies of 50% or 1/2 and 66.6% or 2/3 Training Data. They conclude that the optimal proportion depended on the dataset size and classification accuracy targeted. The general approach of allotting 66.6% or 2/3 of cases for training was close to optimal (Dobbin & Simon, 2011). For this study the dataset was split into 60% for training and 40% for testing which can be considered a 2/3.

Table 4.9

Training Dataset Grouping.

Data Set (30 Days)	
Training Data 20 Days (66.6%)	Test data 10 Days (33.3%)

4.4. Forecasting Model Architecture

An option of an on-premise or cloud solution is available for the forecasting model designed. This gives the SME owners the flexibility to choose the kind of infrastructure that they need. A cloud solution is ideal for SMEs that do not have the required infrastructure to host an on-premise solution. The cloud solution allows users to make demand and supply forecasts by accessing the hosted, the cost of hosting and running of the model is not a concern of the SME owner, this gives the SME owners the freedom to focus on their core business and do not need to worry about infrastructure management. The On-premises solution allows SMEs to host the model in their local infrastructure. This provides a sense of ownership and accountability to internal IT teams, the cost of hosting the model and its maintenance is sorely the responsibility of the SMEs. With the model's sources code publicly available, the SMEs owners can make modifications to the model or addition of new features.

Figure 4.1

Designed Predictive analytics Forecasting model
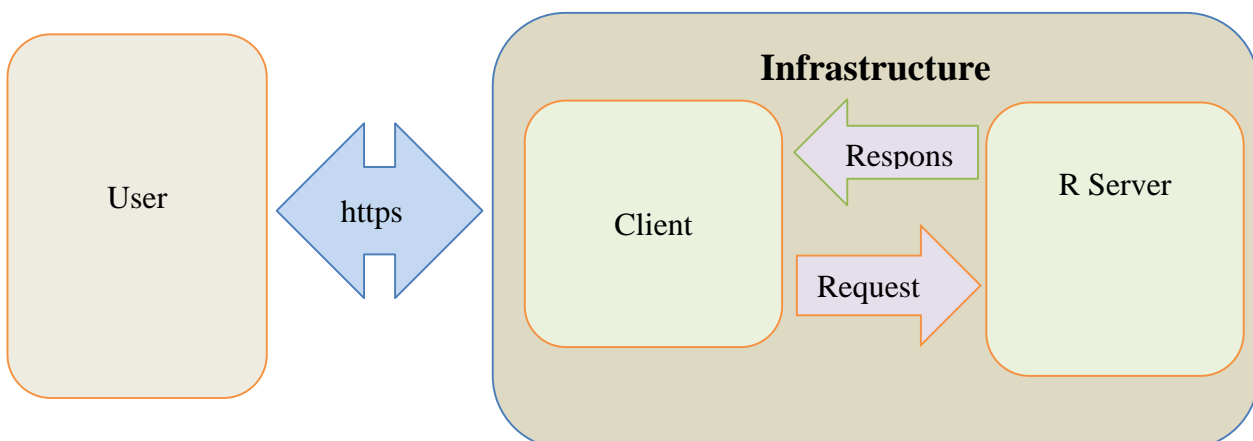
The adopted client-server architecture allows users to access the model through any device with a browser and internet connection. through these browsers, users will be able to access the client-side of the designed model. Through http requests, the user can make requests to the server for specific forecasting requests.

Figure 4.2

Designed User Interface
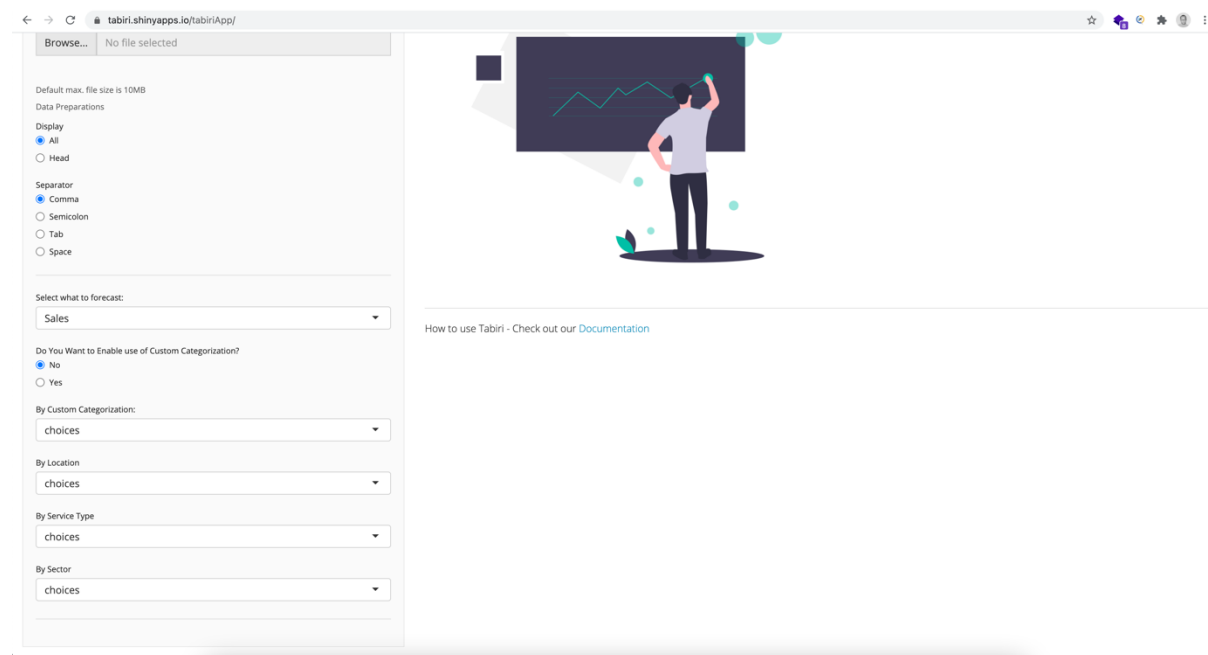
SME owners and managers can upload CSV files in a predefined format into the forecasting model by uploading the file through the user interface. The design framework automatically cleans the data in-case there are null values and makes the forecasts. The user can choose the desired forecasting algorithm from a drop-down menu giving greater flexibility in forecasting.

Figure 4.3

Designed Forecasting model data flow.
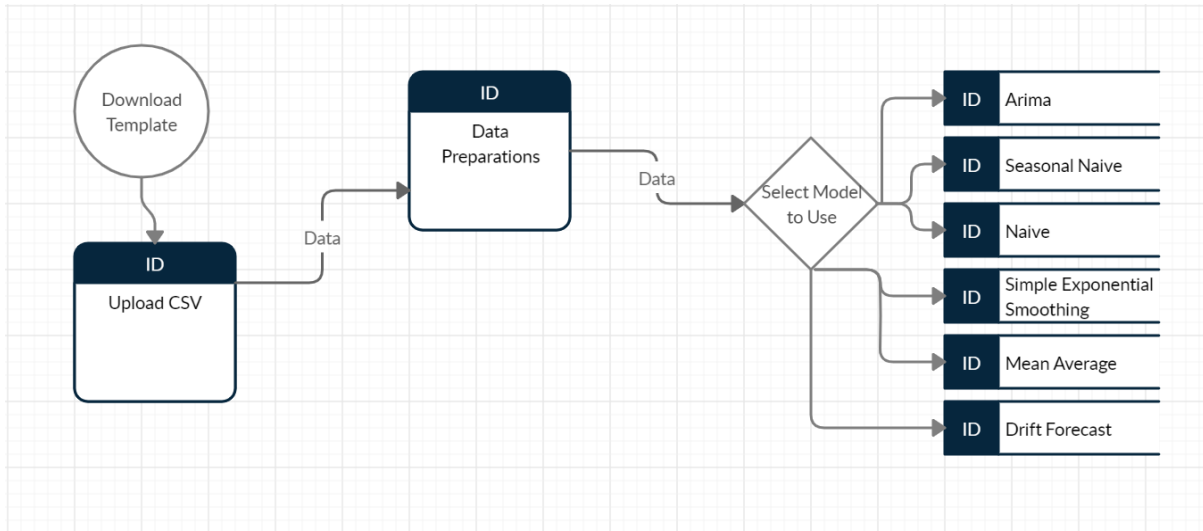
4.4.1. R Packages

R is an open-source software environment for statistical computing and graphics that and runs on a both UNIX OS, Windows OS and Macintosh OS (R Foundation, 2021)

R includes:-

- An efficient data processing and storage,

- A large suite of libraries used for various computational requirements,

- Analysis libraries,

- Graphical facilities user interface

R Packages are small units of reproducible R code. R packages can be installed as follows in R console.

```
install.packages("ggplot2")
```

Where "ggplot2" is the name of the package that is being installed. All packages have to initialized this is achieved by the code

```
library(ggplot2)
```

Where "ggplot2" is the library that is being initiated for use.

4.4.2. Project Structure

```
TabiriApp
        server.R
        ui.R
  Data
    - Template.csv
  www
    - tabiri.png
```

The Overall design of the Model followed a Client-Server architecture. The point of entry is the ui.R this renders the user interface from which users can make requests to the server. Server.R is the server-side code that fulfills requests from the user interface. The server performs all the computations as per the requests from the client-side and returns the results. The data folder contains a pre-defined .csv template that users can download and populate the purchase and sales data. The designed model can accept data in this prescribed format only. The www folder contains all the assets that will be used by the user interface. This can be .css, .png, .jpg or any other asset that is needed by the user interface.

4.4.3. Download Template

For the user to be able to download the preformatted template, a button is provided on the user interface. This is achieved by the following line of code on the ui.R file.

```
downloadButton("downloadData", "Download Dataset Template")
```

When a user clicks on the download button on the user interface, the "downloadData" function in the server.R file is invoked. The data is downloaded with the predefined name "template.csv"

```
output$downloadData <- downloadHandler(
  filename = function() {
    paste(input$template, "template.csv", sep = "")
  },
  content = function(file) {
    write.csv(template(), file, row.names = FALSE)
  }
)
```

Users can use this template to populate their purchases and sales data

4.4.4. Upload CSV

Using the predefined template users can upload their daily sales and purchase data for forecasting. On the user interface the application provides functionality for users to upload a file, this is achieved by the following code on the ui.R file.

```
fileInput('file1', Select a file to be uploaded,
    accept = c(
      '.csv'
    ))
```

The server-side keeps listening to the client-side through its reactive property. If a user uploads a file, the server automatically reads the file and renders the results on the user interface. The data is read using the following function on the server-side,

```
mydata <- reactive({
  inFile <- input$file1
  if (is.null(inFile))
    return(NULL)
  data <-read.csv(inFile$datapath, sep=input$sep)
  data

  #outputing the head or all data only
  if (input$disp == "head") {
    return(head(data))
  }
  else {
    return(data)
  }
```

```
  if(is.null(mydata()))
 })
```

If no data uploaded, the user interface will not display any of the forecasting graphs. Based on the user selection. The server will read the head of the file if the "header" checkbox is selected, this allows for a quick view of the uploaded file before the user reading the whole uploaded data. If the data is uploaded, the graphs are automatically rendered by the function *renderUI* rendering the user interface on the server-side. This *renderUI* function will check for if there is data uploaded to the forecasting model.

```
  if (is.null(mydata()))
```

If this function is null, then the user has not uploaded any file and thus there is no need to attempt to forecast. If the Function is not null the forecast of the uploaded data is performed.

4.4.5. Data Analysis and preparation in R

Data Preparation should be conducted by the user on the predefined data before uploading the data. This means that the user should endeavor to have the correct data in the correct columns as the model is based on the template. The server side will perform some basic data preparation. This is done based on the selections on the user interface. This means a subset of the data will be used for forecasting depending on the selections on the user interface e.g Sales Quantities forecasting, Sales Returns forecasting, Purchases forecasting, Order Cancellation forecasting. The data is prepared by the following function on the server-side.

```
  data <- reactive({
   if (input$drill == 'yes'){
      itemdata_sr <- subset(mydata(),Custom_Category==input$item)
      itemdata_sr_loc <- subset(itemdata_sr,Location==input$location)
      itemdata_sr_loc_ser <- subset(itemdata_sr_loc,Service==input$service)
      itemdata <- subset(itemdata_sr_loc_ser,Sector==input$sector)
      itemdata$Date <- as.Date(paste(itemdata$Date, sep = ""), format = "%d-%b-%y")
```

```
    #Remove N/A Dates
    itemdata <- itemdata %>% filter(!is.na(Date))
    return(itemdata)
    itemdata
  } else {
    itemdata <- subset(mydata())
    itemdata$Date <- as.Date(paste(itemdata$Date, sep = ""), format = "%d-%b-%y")
    return(itemdata)
    itemdata
  }
})
```

Once correct data has been prepared based on the user interface filters, the data is converted into a time series data. This is because the sales and purchase data is indexed in time order. This is achieved by the following function on the server-side. Each forecast element is separately created as a time series and used separately during the forecast.

```
#Sales
ts_Sales <- reactive({
  itemdata <- ts(data()[,c(3)],start=c(2020,4),frequency=365)
  return(itemdata)
})
#Sales_Refunds
ts_Sales_Refund <- reactive({
  itemdata <- ts(data()[,c(4)],start=c(2020,4),frequency=365)
  return(itemdata)
})
#Purchases
ts_Purchases <- reactive({
  itemdata <- ts(data()[,c(5)],start=c(2020,4),frequency=365)
  return(itemdata)
})
#Purchase_cancelation
ts_Purchase_Cancelation <- reactive({
  itemdata <- ts(data()[,c(6)],start=c(2020,4),frequency=365)
  return(itemdata)
})
```

4.4.6. Autoregressive Integrated Moving Average (ARIMA) Model

A combination of integration with autoregression and a moving average model result in a non-seasonal ARIMA model.

I)   Autoregressive model

An autoregression model forecasts the variable of interest by using a linear combination of previous values. This makes auto regressive model a regression of the variable against itself. (Rob & George, 2018).

an autoregressive model of order p can be written as

$$y_t = c + \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \cdots \emptyset_p y_{t-p} + e_t$$

where e_t is white noise. (Rob & George, 2018)

II)   Moving Average Model

Instead of using historical values of the forecast variable in the regression model, moving average model utilises past forecast errors to forecast future values.

$$y_t = c + \varepsilon_t + \emptyset_1 \varepsilon_{t-1} + \emptyset_2 \varepsilon_{t-2} + \cdots \emptyset_q \varepsilon_{t-p}$$

where ε_t is white noise. This is referred to as a moving average model of order q.

Moving average model are used in forecasting future values, unlike the moving average smoothing model which is used for estimating the seasonality of past values. These models are linear since the future values are cramped to be linear functions of past data. While other time series forecasting models attempt to describe the trend and seasonality of the data points. ARIMA aims to explain the autocorrelation between the data points.

The auto.arima() function in R can be implemented as follows on time-series data

```
else if(input$fplot == "Armia"){
  aadata <- auto.arima(PURCHASE_ORDER_CANCELATION_QUANTITIES)
  aaforecast <- forecast(aadata(), h = input$forecast_days)
  return(aaforecast)
  }
```

Figure 4.4

Autoregressive Integrated Moving Average (ARIMA) Forecast.
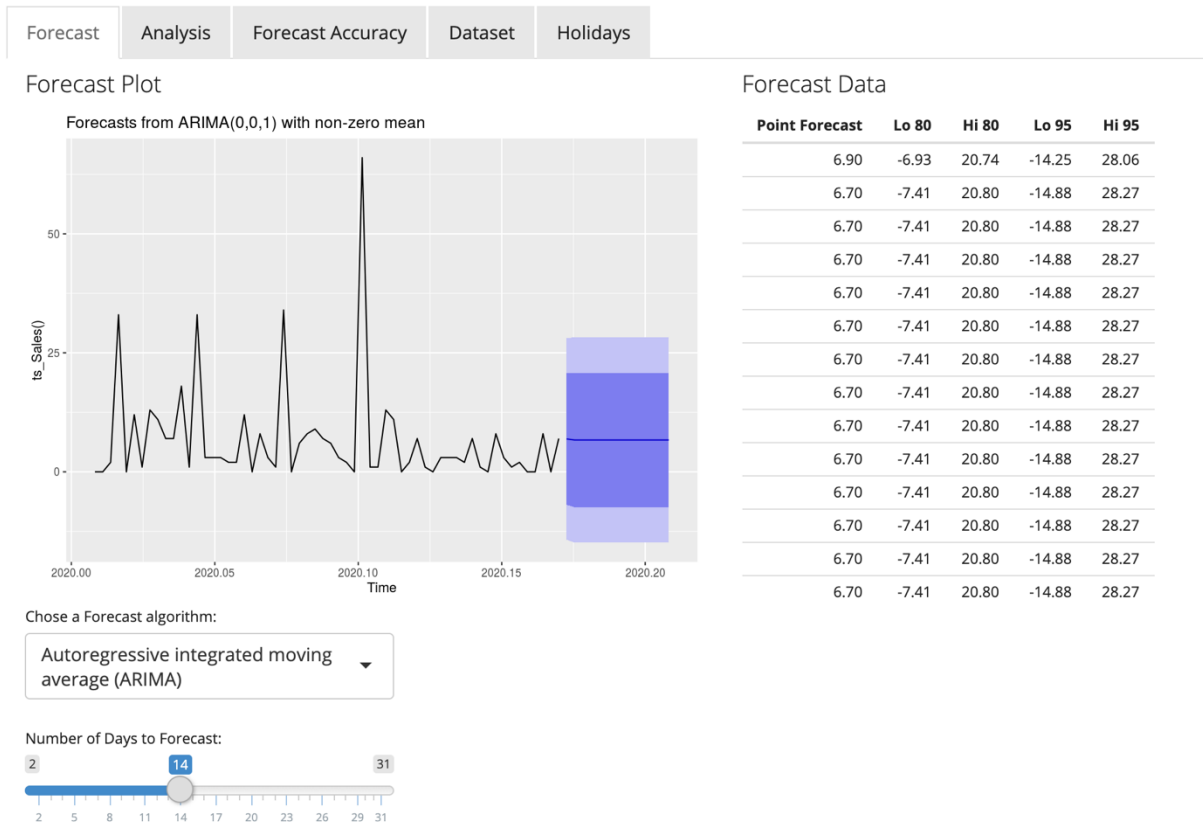
The values of p and q that are appropriate for the available data can be determined by use of ACF plot, and the closely related PACF plot.

Figure 4.5

Decomposition test.

62
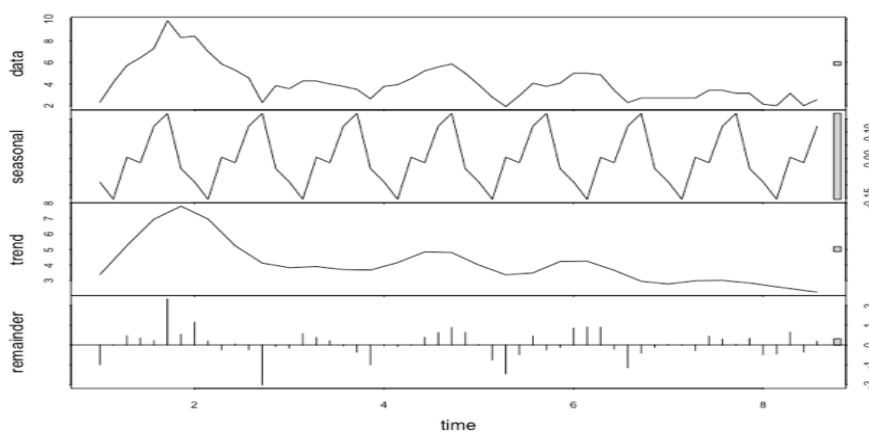
Figure 4.6

Augmenter Dickey-Fuller test.

```
> adf.test(cnt_week_sales, alternative = "stationary")

        Augmented Dickey-Fuller Test

data:  cnt_week_sales
Dickey-Fuller = -4.5843, Lag order = 3, p-value = 0.01
alternative hypothesis: stationary
```

The More negative you are with the Dickey-Fuller Value the more correct you are. This was true for the data as indicated by our sales decomposition test.

Figure 4.7

Decomposition plot Before ACF.
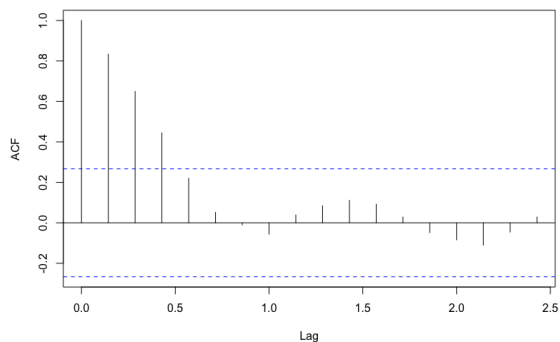
It is evident that there were still numerous outliers in the data, this data was reduced after decomposition by the autocorrelation function that was used

Figure 4.8

Autocorrelation function plot (ACF)

ACF for Differenced Series
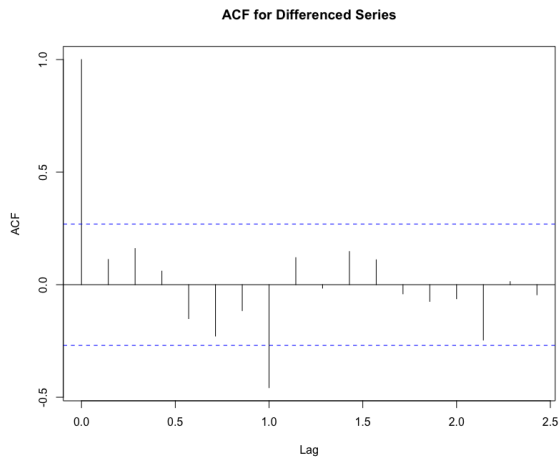
Upon running the dickey fuller test, the best model for ARIMA was of the order (1,1,1) and others of the order (0,0,1).

Figure 4.9

ARIMA (1,1,1) model residuals.

(1,1,1) Model Residuals
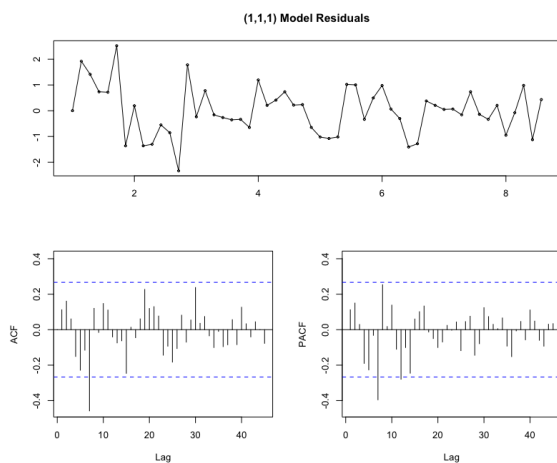
Low RMSE values indicates that the model has good accuracy

4.4.7. Seasonal Naive Model

Pre-processed time series data cab forecasted using the function

```
SALES_REFUND_QUANTITIES <- ts_Sales_Refund()
if (input$fplot == 'Seasonal Naive'){
  fcsnv <- snaive(SALES_REFUND_QUANTITIES, h = input$forecast_days)
```

```
    autoplot(fcsnv)
  }
```

Ts_Sales_data is the preprocessed time series data for Sales Refunds, each Seasonal Naïve

forecast for each attribute can be forecasted by interchanging the time series data. this function

is dependent on a variable fplot that is available on the user interface as

```
selectInput(
  "fplot",
  "Chose a Forecast algorithm:",
  c("Seasonal Naive","Naive","Simple Exponential Smoothing","Mean Average Forecast","D
rift Forecast")),
```

The user can select the desired algorithm to forecast the data. When the user selects the

Seasonal Naïve option from the drop-down menu, the seasonal naïve algorithm is executed and

displayed on the graph.

Figure 4.10

Seasonal Naïve forecast.
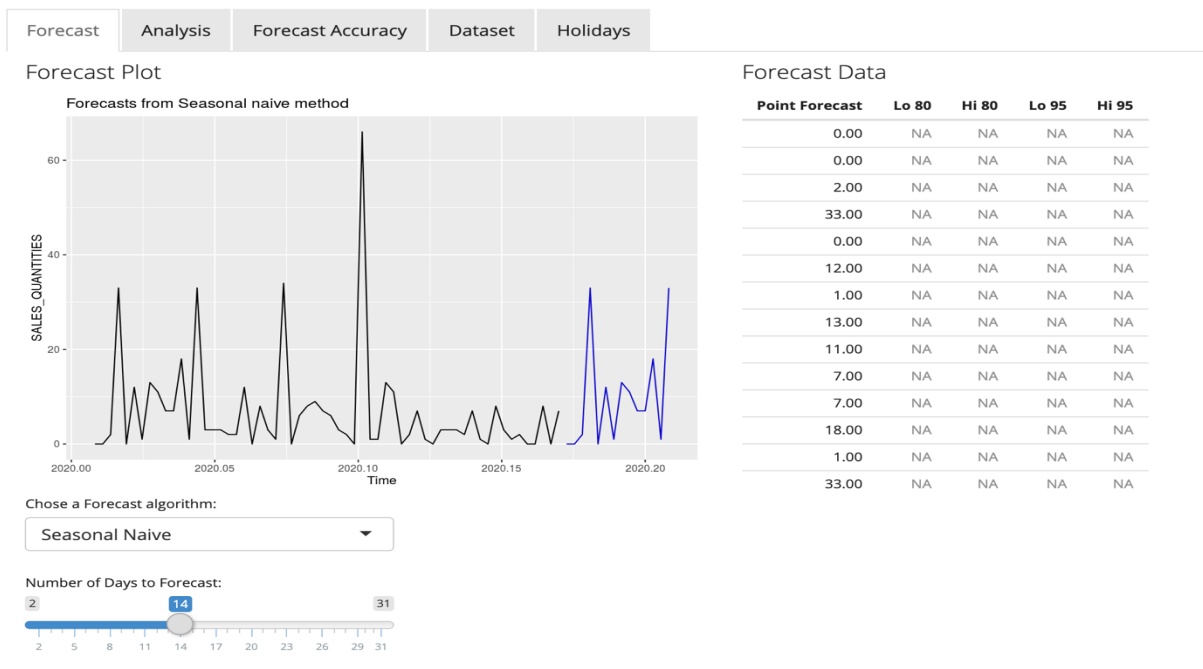
4.4.8. Naive Model

```
else if(input$fplot=='Naive'){
    fcnv <- naive(SALES_REFUND_QUANTITIES, h = input$forecast_days)
    autoplot(fcnv)
```

Just like the seasonal naïve, when a user selects the naïve forecasting method, then Naïve forecasting is done using the above function.

Figure 4.11

Naïve forecast.
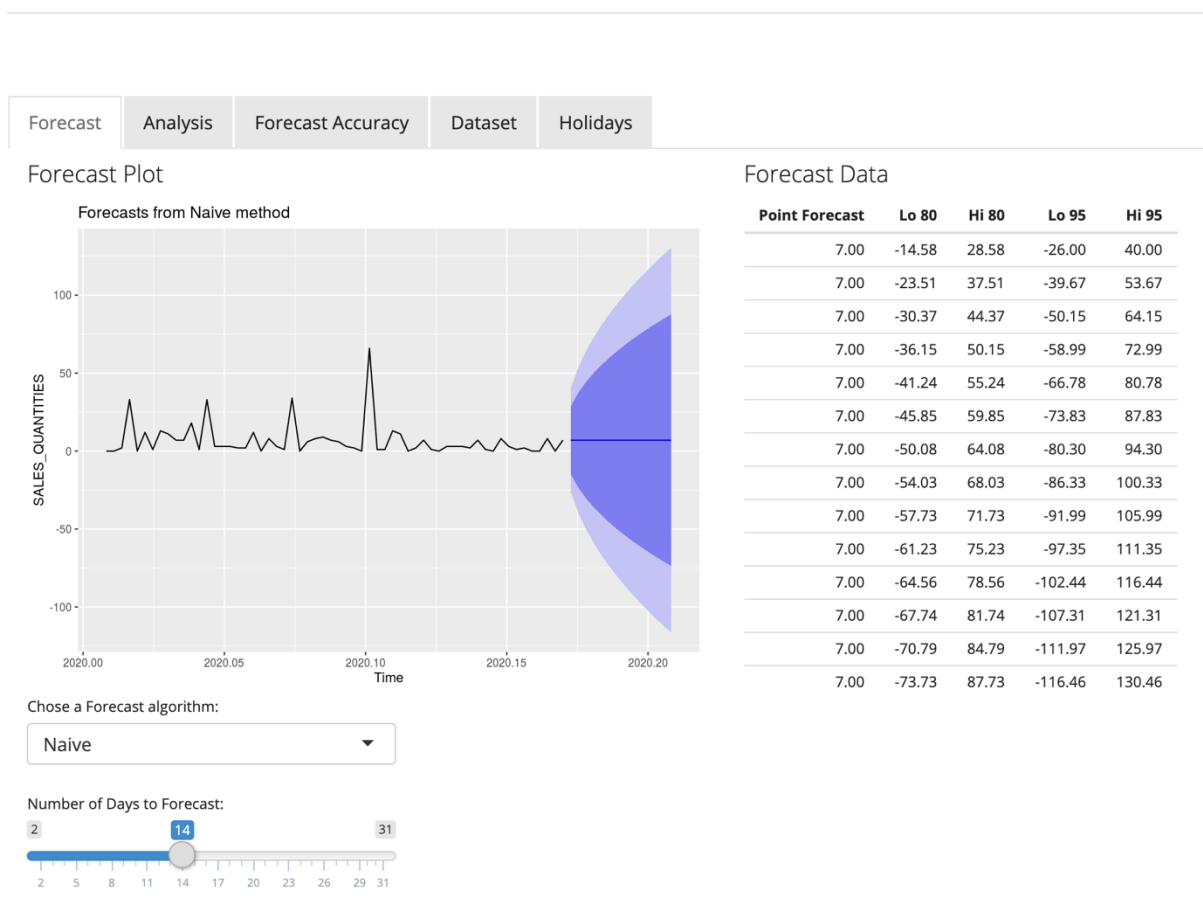
4.4.9. Simple Exponential Smoothing forecasting

Simple exponential smoothing is achieved by the following function when the Option is selected from the user interface by the user.

```
else if(input$fplot == "Simple Exponential Smoothing"){
```

```
    fcses <- ses(PURCHASE_ORDER_CANCELATION_QUANTITIES, h = input$forecast
_days)
    return(fcses)
    }
```

Figure 4.12

Simple Exponential Smoothing

4.4.10. Mean Average Model

The Mean Average model is achieved by the following function

```
 else if(input$fplot == "Mean Average Forecast"){
  fcmean <- meanf(PURCHASE_ORDER_CANCELATION_QUANTITIES, h = input$forec
ast_days)
    return(fcmean)
    }
```

Figure 4.13

Mean Averages Forecast.

4.4.11. Drift Model

The Drift model is achieved using the following function

```
else if(input$fplot == "Drift Forecast"){
    fcdrf <- rwf(PURCHASE_ORDER_CANCELATION_QUANTITIES, h = input$forecast
_days)
      return(fcdrf)
    }
```

4.5. Evaluation

The respondents were given access to the designed model for evaluation. They were asked to use the model to forecast the first two weeks of May 2020. Actual Demand and supply for the first two weeks of May 2020 were compared with the designed model forecasts. The light blue line represents the forecast for the two weeks of May 2020, while the grey line represents the actual sale. The forecast falls withing the 95% accuracy quadrant.

Figure 4.14

First 2 Weeks of May 2020 Sales forecast.
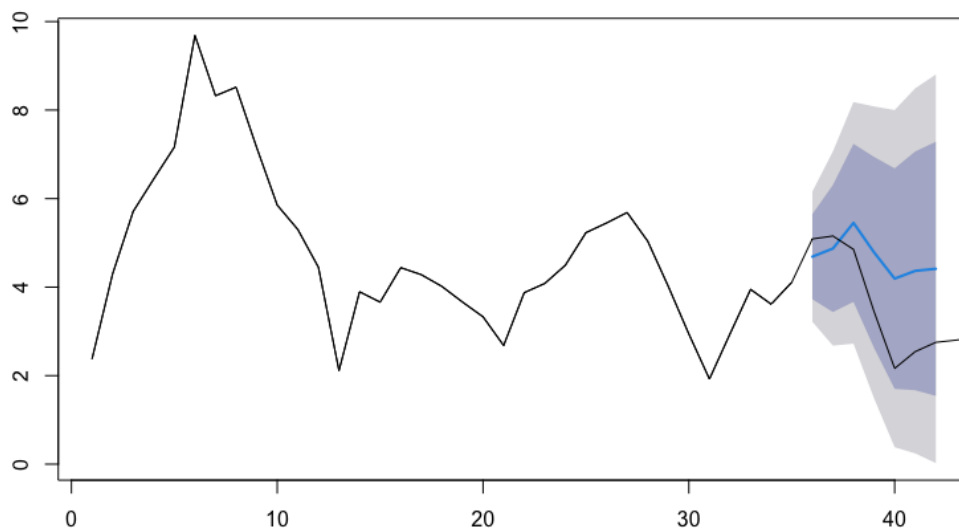
During the evaluation process a questionnaire was sent to the selected respondent whose data was used in the design of the model. The following questions were asked

Question: What Devices are you using to access the forecasting platform

Figure 4.15

Device used to access the forecasting platform

What device are you using to access the forecasting platform?

5 responses

- Laptop — 1 (20%)
- Mobile Phone — 2 (40%)
- Tablet — 2 (40%)
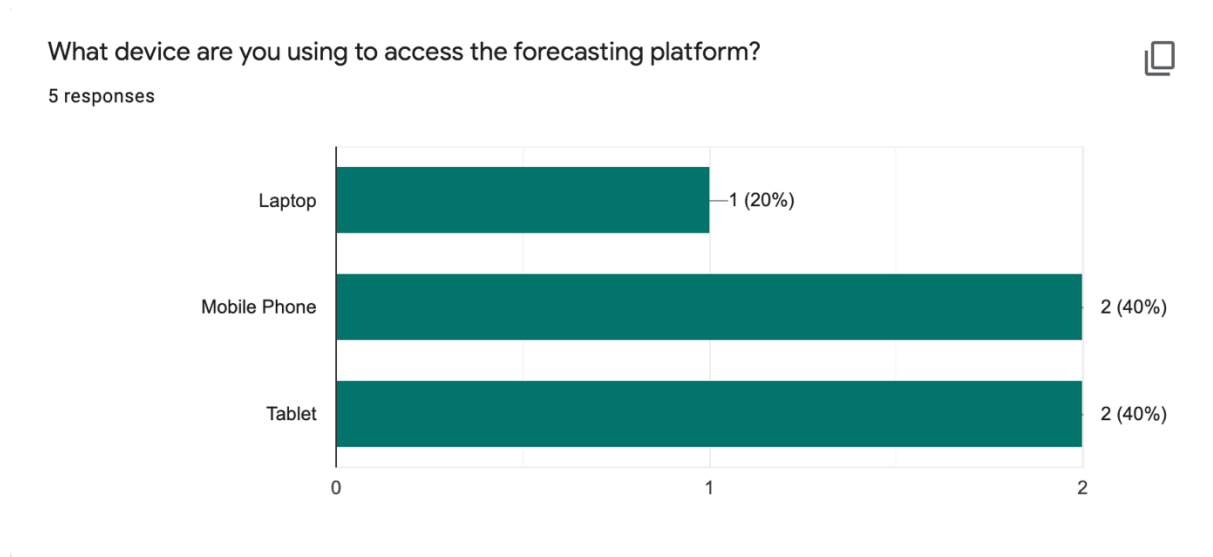
As expected, more users accessed the platform from mobile devices. 80% of the users used a type of mobile device wile 20% used a laptop computer. The mobile design allowed mobile device users to make forecasting without the need of having a laptop computer.

Question: Was the Platform intuitive?

Figure 4.16

Intuitive Presentation Layer Evaluation

Is the platform intuitive?

5 responses

- Yes — 5 (100%)
- No — 0 (0%)
- Some areas are unclear — 0 (0%)
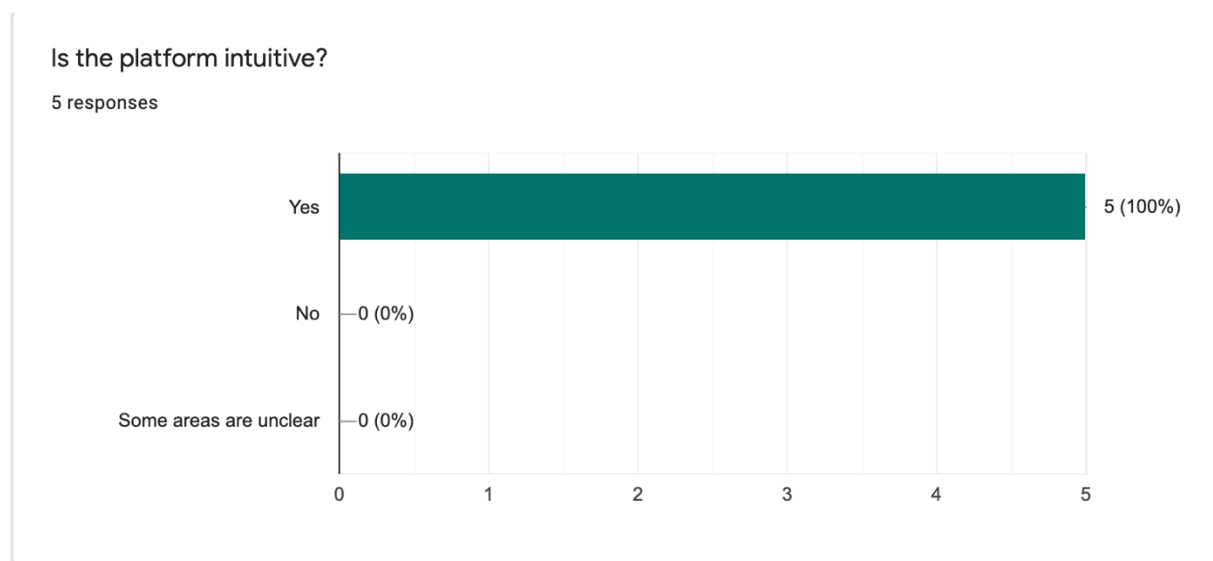
From the response, 100% of the users affirmed that the platform was relatively easy to use and to decipher the forecasts. This is especially important as SMEs might not have skilled IT resources to help in the interpretation of the forecast and the use of the platform.

Question: Did the forecast model help in forecasting Demand and Supply?

Figure 4.17

Evaluation of the Efficiency of the forecasting model
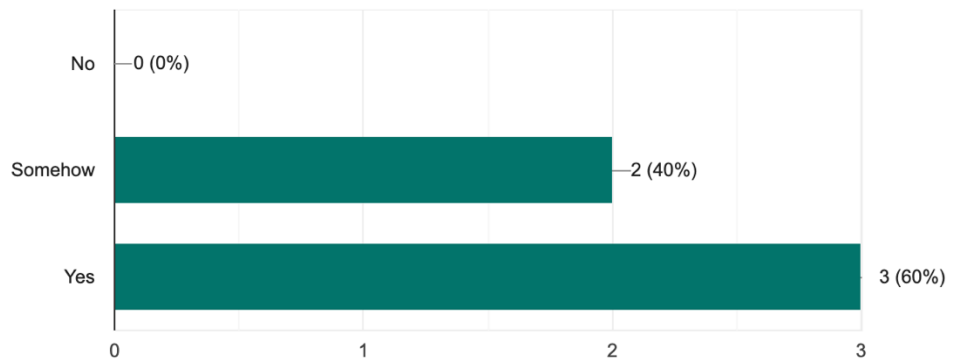
The Main feature of the model was to make forecasts. 60% of the users said that the model helped in making forecasts. Another 40% said that the model somehow helped them make forecast. On further interrogation of the respondents through video interview, the respondent indicated that the model was good enough for them to make forecasts but would like to have additional forecasts. As the research was only focusing on Demand and Supply, other areas of forecasting like lead generation and risk forecast were not in scope.

Question: Will you use the model in future forecasts?

Figure 4.18

Adoption of the forecasting model for future prediction of demand and supply

5 responses
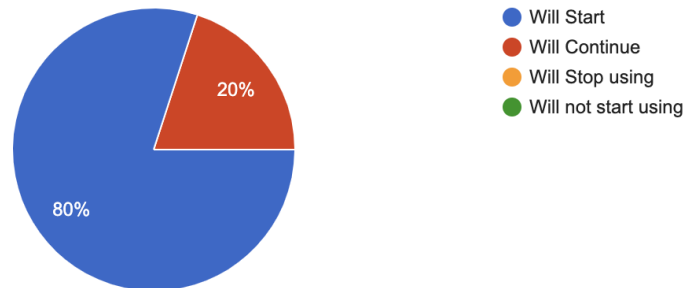
80% of the surveyed users indicated that they will begin using the model in forecasting of demand and supply 20% of the remaining users had already started using the model and indicated they found the model beneficial and will continue to use the mode for future forecasting.

4.6. Strengths of the designed model

During the evaluation of the model, the following several strengths were identified were identified.

4.6.1. Access and Useability

Any SME owner in Kenya can access the hosted version of the model at no cost, upload their transactional data as per the prescribed dataset format, and analysis their demand and supply, and make forecasts. In this architecture, the end-users have the flexibility to access the model on their devices and computer devices. With no registration required, the users can download the template from the platform, fill it with their own transactional data and upload it back to the model for forecasting. With a responsive layout and an intuitive user interface, the application is easy to use and has a help documentation where users can get more details about the forecasting model. A Visual display of the forecast including actual forecast numbers.

4.6.2. Scalability

The Model has a custom field in the data set that allows users to make classified forecasts of demand and supply to a specific matter of interest. For example, an SME owner with multiple outlets might want to make a forecast for all the outlets collectively or might want to have separated forecasts for each outlet. This concept can be applied to other predictors.

Table 4.10

Use of the Custom_Category per respondent

Custom_Category	Respondent Use
Salesperson	1
Outlet (store)	1
Location (city / town center)	1
Product (service)	2

Further, users can select from a list of the forecasting model that they are interested in, there are several forecasting models that can be used to make forecasts. The end-user can also see the forecast accuracy of each model selected. The end-user is also free to select the forecast duration from 2-30 days. In the even that users have uploaded the data and need to make amendments, the changes can directly be made on the uploaded data set, the user is able to view this dataset and search through the dataset. The end users are free to update the holidays that occurred during the forecast period, this helps the model make more accurate forecasts be excluding these days in the training of the model.

4.6.3. Data Security and Privacy

There is no sign-up required for anyone to use the model. Users can directly upload their data and get the forecast presented to them on the screen. The application does not transmit or store any of the uploaded data, the user interaction with the model is managed by sessions. These sessions are unique to each user and when users simultaneously use the model in forecasting

the data is managed separately. The data is fully discarded once the user closes the browser session or when the session expires due to browser inactivity.

4.7. Weakness of the Model

The following weakness was identified in the designed model.

4.7.1. Data Extraction from transactional Systems

While the simple data extraction model adopted is key in allowing users to be able to upload dataset into the model irrespective of the source transactional system being used, the manual extraction of data into a spreadsheet could result in an error during the extraction and subsequently wrong forecasts or errors during forecasts. A solution to this would be to build an Application Programming Interface (API) the SMEs can integrate with their transactional system and the model would extract data directly without the need for manual extraction. This, however, would mean the SMEs must bear the burden of the integration to the Model reducing its adoption and use by SMEs.

4.8. Summary

The methodology chosen in chapter 3 for this research allowed for the design of a practical, accessible, and scalable solution that helps SMEs forecast their Demand and Supply. In the end an Intuitive Web based application was designed to meet this need. SME Owners and managers can now access the forecasting model in any web enabled device through a browser. Of the five SMEs that participated in the research, it was realized that the framework was a much-needed tool to help in the planning of demand and supply.

CHAPTER FIVE

5.0. SUMMARY, LIMITATIONS, AND AREAS FOR FURTHER RESEARCH

5.1. Summary

As more and more SMEs begin to keep transactional records of their business, the opportunities and the possibilities of forecasting are presented. SME owners and managers have a better chance of keeping their business running if they can better forecast product demand and supply. These forecasts help SMEs in planning and ensuring that they meet customer demands and minimize losses. Demand and Supply forecasting is a big problem for SMEs as these businesses usually do not have a lot of financial capital, the strategy is usually one of efficiency in planning and use of available capital to ensure maximum returns. Predictive analytics has not been fully adopted by SMEs which puts them at a disadvantage in competing with well-developed organizations. A model that allows SMEs to forecast demand and supply can greatly help this group in staying afloat. The designed forecasting model is publicly accessible. (Anuradha & Divakar, 2013)

It utilizes 5 forecasting models and users are free to select the models based on the data provided. The model is capable of forecasting Demand and supply attributes with accuracies of more than 80%. The designed model allows users to select a forecasting model from a list of several algorithms. With a high mobile usage penetration in Kenya, a forecasting model capable of running on mobile devices is preferred.

The research aimed at answering three key objectives which were aimed at, comparing existing forecasting model used by SMEs in Kenya, Building a data extraction and presentation layer and to evaluate the effectiveness of the designed model.

Objective 1: Compare existing models and implement the best model for predicting demand and supply.

Literature review of existing models used by SMEs in Kenya revealed that there was a limited used of forecasting tools. SMEs were largely exposed to qualitative forecasting will are not always reliable or transferable since this method heavily relies on the SME owner using experience. Gut feeling and best guessing. The research analysed sales orders and purchase oders used by SMEs to identify predictors that can be used to forecast demand and supply, After and analysis of sales and purchase data that is collected by SMEs, the following predictors or attributes were required for the forecast.

    i.   Sales quantities
    ii.   Sales Refund quantities
    iii.   Purchases quantities
    iv.   Purchase Cancelation quantities
    v.   Location
    vi.   Service
    vii.   Company employee size
    viii.   Category

These predictors were discussed in detail in section 4.3 on training data preparation and analysis. These predictors were present in all the sales orders and purchase orders of the SMES that participated in the research while other predictors like location, type of service provided, company employee size and the business sector were all inferred.

Objective 2: Build a data extraction user interface

The SMEs targeted by this research were SMES that had some sort of formal digital or manual bookkeeping regime in place. A simple digital format easily understood by SMES is the use of spreadsheets to store the required demand and supply predictors. The method chosen to store

these predictors had to allow easy importation of this data into the forecasting engine. A predefined CSV template was designed and made available for users of the forecasting model to use The designed extraction approach allows SME owners to upload their transactional data in a prescribed format in CSV files. Once the users upload transactional data, they were able to see the uploaded data, and make forecasts using the uploaded data . This model allows SME owners with little IT knowledge to still be able to use the model for forecasting.

Figure 5.1

Demand and Supply Uploaded Dataset
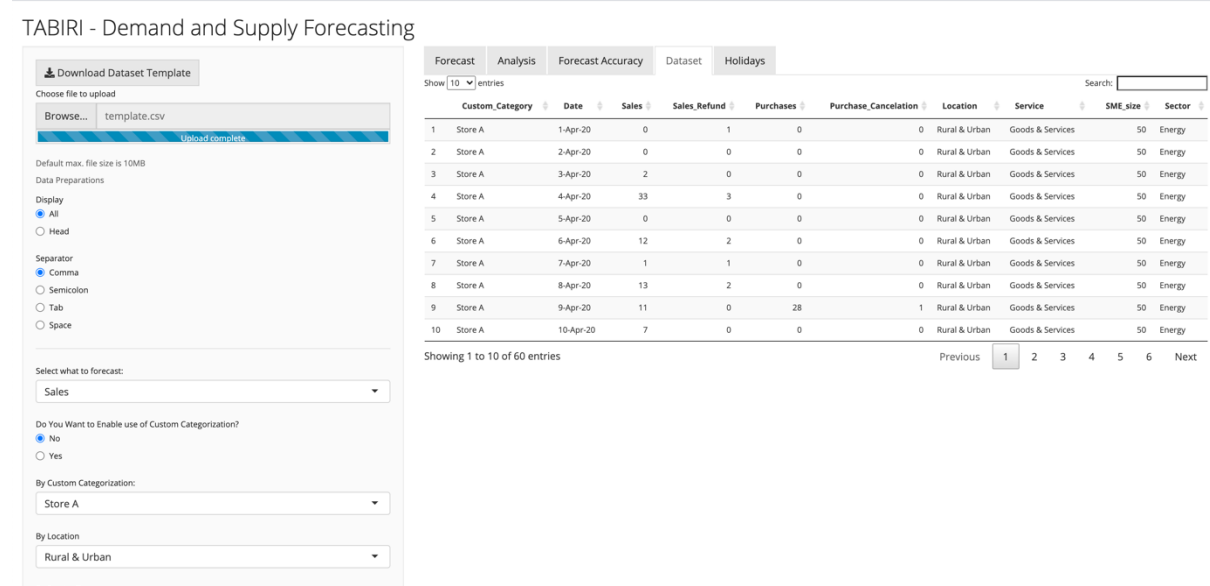
Objective 3: Build a presentation layer to display the supply and demand forecasts

The implementation of Rshiny which utilizes cascading stylesheets (CSS) that delivers the presentation layer to the end-users allows the designed model to be accessed from any device with a browser and a connection to the server hosted on-premise or in the cloud. This means that with a responsive presentation layer, SME owners have a variety of ways that they can access the designed model including their mobile devices.

5.2. Recommendations for further research

The model is limited to forecasting 4 attributes namely sales, sales returns, Purchase Orders, and Purchase cancelations. More variables can be investigated like prevailing political climate, current health pandemic, education of employees, and proprietor that could help improve the accuracy of the model. Further research can be carried out to allow direct integration with the source database so that SME owners do not need to upload the sales and purchase data and this can be read directly by the model from the source tables.

5.3. Outcomes of the research

Sales and Demand forecasting are critical to the survival of small and medium enterprises, it ensures that the SMEs have the right quantities at the right time and maximizes the profits while reducing losses that could have been caused by dead stock. SMEs from any sector can accurately forecast demand and supply. The platform can be accessed on both mobile and desktop devices.

The following are the features of the designed model.

      I)   Cloud or on promise version options

      II) Data privacy.

      III) No signup required

      IV) Multiple forecasting algorithms

      V) Reactive Model

      VI) Responsive Web design

      VII)   Free for use and open source

5.4. Published Papers and Conferences from this Thesis

The following papers were presented in conferences or published as a result of this research.

Publications
Mburu, J M, (July 2020) "*Modelling A Forecasting Platform for Small and Medium Enterprises in Kenya.*" Lukenya University Multidisciplinary Journal (LUMJ), 1 (2), pp. 109-115, 2020, ISSN: 2663-3163.

Conferences
Mburu, J M, (July 2020) "*Modelling A Forecasting Platform for Small and Medium Enterprises in Kenya.*" Lukenya University First Web Conference Series. Nairobi, Kenya.

Mburu, J M, (Sept 2020) "*Performance of Auto-regressive Integrated Moving Average Model (ARIMA) for Forecasting Daily Demand and Supply: A Case of Small & Medium Enterprises in Kenya*" Kabarak University International Research Conference on computing and Information systems: Masters and Doctoral Colloquium.

REFERENCES

Agile Manifesto. (2021). *Manifesto for Agile Software Development:*
https://agilemanifesto.org/

Amarasingham, R., Rachel, E.P., Huesch , M., Nguyen, N.Q., & Xie, B. (2014).
Implementing Electronic Health Care Predictive Analytics: Considerations And
Challenges. *Health Affairs, 33*(7), 1148-1154, DOI: 10.1377/hlthaff.2014.0352.

Anuradha, P., & Divakar, S. (2013). A Survey Report on Text Classification with Different
Term Weighting Methods. *International Journal of Computer Applications, 75*(7), 14-
18 DOI:10.5120/13122-0472.

Ardhapure, O., Patil, G., Udani, D., & Jetha, K. (2016). Comparative Study of Classification
Algorithm for Text Based Categorization. *IJRET: International Journal of Research
in Engineering and Technology,5(2)*, 217-220, DOI:10.15623/IJRET.2016.0502037.

Ayanda, A. M., & Laraba, A. S. (2011). Small and Medium Scale Enterprises as a Survival
Strategy for Employment Generation in Nigeria. *Journal of sustainable development,
4*(1), 200-206, DOI:10.5539/jsd.v4n1p200.

Bala, R., Sikka, S., & Singh, J. (2014). A Comparative Analysis of Clustering Algorithms.
International Journal of Computer Applications, 100(15), 35 - 39;
DOI:10.5120/17603-8293.

Bhardwaj, B. K., & Pal, S. (2011). Data Mining: A Prediction for Performance Improvement
Using Classification. *(IJCSIS) International Journal of Computer Science and
Information Security, 9*(4), 136-140, https://doi.org/10.48550/arXiv.1201.3418.

Biau, G., & Scornet, E. (2015). *A Random Forest Guided Tour. arXiv:1511.05741v1
[math.ST], Test 25*, DOI:10.1007/s11749-016-0481-7.

Blockeel, H., Raedt, L. D., & Ramon, J. (1998). Top-Down Induction of Clustering Trees. In
J W. Shavlik (eds.)*Proceedings of the Fifteenth International Conference on Machine
Learning*, (pp55-63), Morgan Kaufmann Publishers .doi:10.5555/645527.657456.

Bowen, M., Morara, M., & Mureithi, S. (2009). Management of Business Challenges Among
Small and Micro Enterprises in Nairobi-Kenya. *KCA Journal of Business
Management, 2*(1), 16-31, DOI: 10.4314/kjbm.v2i1.44408.

Breiman, L. (2001). Random Forest. *Machine Learning, 45*, 5-32,
https://link.springer.com/content/pdf/10.1023/A:1010933404324.pdf.

Capital Markets Authority. (2020). *Capital Markets Corporate Governance Survey of Small
and Medium Enterprises (SMEs) in Kenya (An assessment of level of corporate
governance in SMEs).* Kenya National Bureau of Statistics.

Central Bank of Kenya, (2018). *The Year of resilience – Annual Report and Financial
Statements 2017/2018.*
https://www.centralbank.go.ke/uploads/cbk_annual_reports/665458124_Annual%20R
eport%202917%2018%20low%20resolution.pdf.

Communication Authority of Kenya. (2018). *Second quarter sector statistics report for the
financial year 2018/2019.* https://www.ca.go.ke/wp-content/uploads/2019/03/Sector-
Statistics-Report-Q2-2018-19.pdf.

Davenport, T. H. (2014,September 2). A Predictive Analytics Primer. *Harvard Business
Review.* https://hbr.org/2014/09/a-predictive-analytics-primer

Dayan, P., & Niv, Y. (2008). Reinforcement learning: The Good, The Bad and The Ugly.
Science Direct, 18(2), 185-196, DOI 10.1016/j.conb.2008.08.003.

Dhakare, B., Jain, H., VedantSalunke, & Gandla, R. (2018). CRM Application For Analyzing
the Sales Using Data Mining Techniques and Business Intelligence. *Internationa
Journal of Engineering Science and Computing, 8*(3), 16138-16139, https://ijesc.org/.

Dilts, D. A. (2004). *Introduction to Microeconomics E201.* Indiana - Purdue University.

Dobbin, K., & Simon, R. (2011). Optimally splitting cases for training and testing high dimensional classifiers. *US National Library of Medicine National Institutes of Health*, 4(1), 31, DOI: 10.1186/1755-8794-4-31.

Eckerson, W. W. (2006). *Predictive Analytics: Extending the Value of Your Data Warehousing Investment.* Data warehousing Institute.

Fildes, R., & Lusk, E. J. (1984). The Choice of a Forecasting Model. *Omega, 12*, 427-435, DOI:10.1016/0305-0483(84)90042-2.

Goul, M., Balkan, S., & Dolk, D. (2005, January 5-8). *Predictive Analytics-Driven Campaign Management Support Systems* [Paper Presentation]. Hawaii International Conference on System Sciences, Hawaii, United States. DOI: 10.1109/HICSS.2015.568.

Halili, F., & Rustemi, A. (2016). Predictive Modeling: Data Mining Regression Technique Applied in a Prototype. *International Journal of Computer Science and Mobile Computing, 5*(8), 207 - 215, www.ijcsmc.com.

Hamiche, K., Abouaïssa, H., Goncalves, G., & Hsu, T. (2018). A Robust and Easy Approach for Demand Forecasting in Supply Chains. *Science Direct, 51*(11), 1732-1737, https://doi.org/10.1016/j.ifacol.2018.08.206

Hatem, N. M., Shahenda, S. S., & Mohsen, R. .. (2018). Enhancing recurrent neural network-based language models by word tokenization. *Human-centric Computing and Information Sciences, 8*(12), 8-12, https://doi.org/10.1186/s13673-018-0133-x.

Hetzel, W. (1988). *The Complete Guide to Software Testing*, (2nd ed). Denton Island Books

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice .* OTexts.

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software, 27*(3), 1-22, 10.18637/jss.v027.i03.

Ismail, N. A., & King, M. (2006). The alignment of accounting and information systems in SMEs in Malaysia. *Journal of Global Information Technology Management, 9*(3), 24, DOI:10.1080/1097198X.2006.10856425.

Jeong, E. J., Bae, J. H., & Jeong, S. R. (2015). Guidelines Aimed at Reducing the Risks of User Acceptance Delay in the Context of an IT Service Project Management Plan. *International Journal of Electrical and Computer Engineering (IJECE), 5*(4), 832-839, DOI:10.11591/ijece.v5i4.

Kalechofsky, H. (2016). *A Simple Framework for Building Predictive Model.* M Squared Consulting.

Kemp, S. (2021, February). *Digital 2020 Kenya.* Datareportal: https://datareportal.com/reports/digital-2020-kenya

Kenya National bureau of statistics. (2016). *Micro, Small and Medium Establishment (MSME) Survey.* Kenya National bureau of statistics.

Kriesel, D. (2005). *A Brief Introduction to Neural Networks.* Bonn: http://www.dkriesel.com.

Mabula, J. B., & Ping, H. D. (2018). Use of Technology and Financial Literacy on SMEs Practices and Performance in Developing Economies. *International Journal of Advanced Computer Science and Applications(IJACSA), 9*(6), 74-82, (DOI) : 10.14569/IJACSA.2018.090611.

Mahmoud, F. Z. (2017). The Application of Predictive Analytics:Benefits, Challenges and How It Can Be Improved. *International Journal of Scientific and Research Publications, 7*(5), 549 - 566, https://www.ijsrp.org/.

McHugh, M., Cawley, O., McCaffcry, F., Richardson, I., & Wang, X. (2013, May 18-26). *An Agile V-Model for Medical Device Software Development to Overcome the Challenges With Plan-Driven Software Development Lifecycles* [Paper presentation],

5th International Workshop on Software Engineering in Health Care (SEHC), San Francisco, United States. DOI: 10.1109/SEHC.2013.6602471.

Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). *Introduction to Time Series Analysis and Forecasting* (2nd ed). John Wiley & sons.

Morgan Stanley Capital International. (2021). *Global Industry Classification Standard*. Morgan Stanley Capital International: https://www.msci.com/gics

Mouton, J. (1996). *Understanding social research.* Van Schaik.

Mugo, A. N. (2016). *Challenges Facing Kenyan Micro and Small Enterprises in Accessing East African Markets: A Case of Manufacturing MSES in Nairobi.* [Masters thesis, United States International University]. http://erepo.usiu.ac.ke/11732/2581

Nyce, C. (2007, October). Predictive Analytics White Paper. *American Institute for CPCU. Insurance Institute of America*, 9-10, Retrieved from: http://www.theinstitutes.org/doc/predictivemodelingwhitepaper.pdf.

Pandey, U. K., & Pal, S. (2011). A Data Mining View on Classroom Teaching Language. *International Journal of Computer Science Issues, 8*(2), 277-282, https://doi.org/10.48550/arXiv.1104.4164.

Patil, P. M., & Bhosale, A. (2018). Big Data Analytics. *Open Access Journal of Science, 2*(5), 326-335, DOI:10.15406/oajs.2018.02.00095.

Patra, A., & Singh, D. (2013). Neural Network Approach for Text Classification using Relevance Factor as Term Weighing Method. *International Journal of Computer Applications, 68*, 37-41, doi:10.5120/11674-7301.

Pousttchi, K., Tilson, D., Lyytinen, K., & Hufenbach, Y. (2015). Introduction to the Special Issue on Mobile Commerce: Mobile Commerce Research Yesterday, Today, Tomorrow–What Remains to Be Done. *International Journal of Electronic Commerce, 19*(4), 1-20, DOI:10.1080/10864415.2015.1029351.

R Foundation. (2021). *The R Project for Statistical Computing*. Retrieved from R Foundation: https://www.r-project.org/

R studio. (2021). *Shiny*. Retrieved from R studio: https://shiny.rstudio.com/

Rajalingam, N., & Ranjini, K. (2011). Hierarchical Clustering Algorithm - A Comparative Study. *International Journal of Computer Applications (0975 – 8887), 19*(3), 42-46, doi:10.5120/2340-3052 .

Ramdani, B., Kawalek, P., & Lorenzo, O. (2009). Predicting SMEs' adoption of enterprise systems. *Journal of Enterprise Information Management, 22*(1), 10-24, DOI:10.1108/17410390910922796.

Richert, W., & Coelho, L. P. (2013). *Building Machine Learning Systems with Python.* Packt Publishing .

Senthilnath, J., Dokania, A., Kandukuri, M., K.N, R., Anand, G., & S.N.Omkar. (2016). Detection of tomatoes using spectral-spatial methods in remotely sensed RGB images captured by UAV. *Biosystems Engineering, 146*, 16-32; https://doi.org/10.1016/j.biosystemseng.2015.12.003.

Sharma, S., Goyal, D. P., & Mittal, R. (2008). Data mining research for customer relationship management systems: A framework and analysis. *International Journal of Business Information Systems, 3*(5), 549-565, DOI:10.1504/IJBIS.2008.018605.

Shubbar, M., & Smith, J. (2019). Demand Forecasting: An Open-Source Approach. *SMU data science Review, 2*(1), 1 - 18, https://scholar.smu.edu/datasciencereview/vol2/iss1/18.

Smita, S., & Sharma, P. (2014). Use of Data Mining in Various Field: A Survey Paper. *IOSR Journal of Computer Engineering, 16*(3), 18-21, DOI:10.9790/0661-16351821.

Statcounter. (2021). *GlobalStats*. Retrieved from Statcounter GlobalStats: https://gs.statcounter.com/platform-market-share/desktop-mobile-/kenya

Szalvay, V. (2004). *An Introduction to Agile Software Development.* Danube Technologies, Inc.

Thong, J. Y., & Yap, C. S. (1995). CEO Characteristics, Organisational Characteristics and Information Technology Adoption in Small Business. *Omega, 23*(4), 429-442; https://doi.org/10.1016/0305-0483(95)00017-I.

Turner-McGrievy , G. M., Hales , S. B., Schoffman , D. E., Valafar , H., Brazendale , K., Weaver , R. G., . . .& Hester , A. (2017). Choosing between responsive-design websites versus mobile apps for your mobile behavioral intervention: presenting four case studies. *US National Library of Medicine, 7*(2), 224-232, DOI: 10.1007/s13142-016-0448-y.

UC Business Analytics R Programming Guide. (2018). *Benchmark Methods & Forecast Accuracy*. UC Business Analytics R Programming Guide: https://uc-r.github.io/ts_benchmarking.

Velickov, S., & Solomatine, D. (2000, March). Predictive Data Mining: Practical Examples. *Artificial Intelligence in Civil Engineering. Proc. 2nd Joint Workshop*, 1-17, doi=10.1.1.323.8809.

Whelan, J., & Msefer, K. (2001). *Economic Supply & Demand.* MIT.

Yan, J., Gong, M., Sun, C., Huang, J., & Chu, S. M. (2015). Sales Pipeline Win Propensity Prediction: A Regression Approach. *2015 IFIP/IEEE International Symposium on Integrated Network Management (IM)*, 854-857, doi: 10.1109/INM.2015.7140393.

Zahra, F. M., & Mohamadali, N. A. (2019). The Business Intelligence Use In Healthcare And Its Enhancement By Predictive Analytics. *International Journal of Computer Trends and Technology (IJCTT), 67*(7), 26-39, 10.14445/22312803/IJCTT-V67I7P105.

Zhang, C., & Zhang, S. (2008). *Association Rules Mining Models and Algorithms. In Lecture Notes in Artificial Intelligence* (Vol. 2307). Springer.

Zhou, Z.-R., Wang, W.-W., Li, Y., Jin, K.-R., Wang, X.-Y., Wang, Z.-W., . . . & Zhang, T.-S. (2019). In-depth mining of clinical data: the construction of clinical prediction model with R. *National Library of medicine, 7*(23), 796, doi: 10.21037/atm.2019.08.63.

Zowghi, D., Rimini, F. D., & Bano, M. (2015, April 27-29). Problems and Challenges of User involvement in Software Development: an Empirical Study [Paper presentation]. *International Conference on Evaluation and Assessment in Software Engineering,* Nanjing, China*, https://doi.org/10.1145/2745802.2745810.

APPENDICES

APPENDIX 1: DATA COLLECTION RECOMMENDATION LETTER

Kenya Methodist University

P. O. Box 267-60200, Meru, Kenya, Tel: (+254) 724 256 162, 064-3131279/097 Email: vice.chancellor@kemu.ac.ke, Website: www.kemu.ac.ke

September 20, 2022

TO WHOM IT MAY CONCERN

RE: MUREITHI JOHN MBURU   CIS-3-0561-1/2018

This is to confirm that the above named is a student in the Department of Computer Science, in this university, pursuing a Master of Science in Computer Information System.

As a requirement, the student is expected to undertake an independent **primary research** in their area of specialization.

The purpose of this letter is therefore; to introduce the student to you and request you to allow him undertake the research in your organization.

The student has been advised to ensure that all data and information from the organization is treated with utmost confidentiality and only used for academic purposes unless otherwise stated.

Any assistance accorded to him will be highly appreciated.

Yours faithfully,

Prof. Peter Kihara, PhD.
Registrar - Academic Affairs

Nairobi Campus: KeMU Towers, University Way, P. O. Box 45240-00100, Nairobi. Tel: +254-20-2118443/2248172/2247987/0725-751878 • Email: principal.nairobi@kemu.ac.ke
Mombasa Campus: Former Oshwal Academy, P. O. Box 89983, Mombasa. Tel: +254 - 041-2495945 / 8 • Email: director.mombasa@kemu.ac.ke

The Future is Here!

84

APPENDIX 2: NACOSTI LICENSE.

REPUBLIC OF KENYA

NACOSTI

NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Ref No: **750237**

Date of Issue: **04/October/2022**

RESEARCH LICENSE

This is to Certify that Mr.. John Mureithi Mburu of Kenya Methodist University, has been licensed to conduct research in Kirinyaga, Nairobi on the topic: **PREDICTIVE ANALYTICS MODEL FOR SMALL AND MEDIUM ENTERPRISES IN KENYA, FORECASTING ON SUPPLY AND DEMAND for the period ending : 04/October/2023.**

License No: **NACOSTI/P/22/20629**

750237

Applicant Identification Number

Director General
NATIONAL COMMISSION FOR
SCIENCE,TECHNOLOGY &
INNOVATION

Verification QR Code

NOTE: This is a computer generated License. To verify the authenticity of this document,
Scan the QR Code using QR scanner application.

85

APPENDIX 3: COVER LETTER.

Mburu John Mureithi,

Kenya Methodist University,

P.O. Box 7380-30100,

Nairobi.

Dear Respondent,

I am a graduate student at Kenya Methodist University pursuing a master's degree in Computer Information systems. To fulfill my degree requirements; I am undertaking a research project on a Predictive Analytics Framework for SMEs in Kenya.

I have selected you to form part of this study. This is to kindly request you to assist me to collect data by filling out the accompanying questionnaire. The information /data you provide will be exclusively for academic purposes and will be handled with the utmost confidentiality. I look forward to your prompt response.

Thank you in advance,

Yours sincerely,

Mburu John Mureithi

APPENDIX 4: QUESTIONNAIRE.

Question one

Download the Demand & Supply Data Capture excel sheet from
https://docs.google.com/spreadsheets/d/1vEIC9qD4WSSnzOBTlYhnrFW7MDQc2fuX8oc4B7fCTco/edit?usp=sharing

Upload the "Daily Demand & Supply Data Capture" using the excel file downloaded from the link in the description section above

Date (DD-MM-YYYY)	Sales Quantities		Purchase Quantities	
	Orders / Sales Processed per day	Canceled / Refunded Sales per day	Purchase Orders received per day	Unfulfilled orders / Cancelled / Rejected Orders per day
01/04/2020				
02/04/2020				

Question 2

Where is your business located?

Radio group [Rural Area, Urban Area, Both Rural & Urban Area]

Question 3

What sector best categorizes your SME?

Radio group [Energy, Materials, Industrials, Consumer Discretionary, Consumer Staples, Health Care, Financials, Information Technology, Communication Services, Utilities, Real Estate]

Question 4

What best describes what your SME sells?

Radio group [Merchandising Business, Service Business, Manufacturing Business, Hybrid

Business]

Question 5

How many employees work at your business?

Radio group [1-10, 11-50, 51-100, more than 100]

APPENDIX 5: PRESENTATION LAYER SOURCE CODE.

```
# Predictive Analytics Framework For Small and Medium Enterprises
# Sales and Purchases Forecasting
# Kenya Methodist University
# Research By John Mburu
# *******************************************************************************#
# Load Required Libraries
# *******************************************************************************#
library(shiny)
library(shinythemes)


# *******************************************************************************#
# Tabiri App Presentation Layer
# *******************************************************************************#
shinyUI(fluidPage(
   theme = shinytheme("yeti"),
   tags$div(class="Header",style="font-family: Montserrat;",
        tags$h2("TABIRI - Demand and Supply Forecasting")),
   sidebarLayout(
     sidebarPanel(
       downloadButton("downloadData", "Download Dataset Template"),
              fileInput('file1', 'Choose file to upload',
             accept = c(
               '.csv'
             )
       ),
       helpText("Default max. file size is 10MB"),
       h5(helpText("Data Preparations")),
              radioButtons("disp", "Display",
               choices = c(All = "all",
                      Head = "head"),
               selected = "all"),
       radioButtons(inputId = 'sep', label = 'Separator', choices = c(Comma=',',Semicolon=';',Tab='\t', Spa
ce=''), selected = ','),

       tags$hr(),
       # SELECT FORECAST OPTIONS
       selectInput("forecast_item", "Select what to forecast:",
              choices =  c(
                 "Sales" = "sales",
                 "Sales Refund" = "sales_refund",
                 "Purchases" = "purchases",
                 "Cancelled POs" = "purchase_cancelation", selected = NULL)),

       #Allow users to Drill down using custom categorization column [this column can be used to identif
y products or outlets]
       radioButtons("drill", "Do You Want to Enable use of Custom Categorization?",
              choices = c(
                 "No" = "no",
                 "Yes" = "yes",
                 selected = NULL)),

       selectInput(
          "item",
```

```
        "By Custom Categorization:",
        "choices"),
      selectInput(
        "location",
        "By Location",
        "choices"),

      selectInput(
        "service",
        "By Service Type",
        "choices"),
      selectInput(
        "sector",
        "By Sector",
        "choices"),
      tags$hr()

    ),
    mainPanel(
      uiOutput("tb")
    )

  )
))
```

APPENDIX 6: FORECASTING SERVER SIDE SOURCE CODE.

```
# Predictive Analytics Framework For Small and Medium Enterprises
# Sales and Purchases Forecasting
# Kenya Methodist University
# Research By John Mburu
# ******************************************************************************#
# Load Required Libraries
# ******************************************************************************#
```

```
library(shiny)
require(gridExtra)
library(datasets)
library(fpp2)
library(ggplot2)
library(DT)
library(tidyr)
library(dplyr)
library(readxl)
library(zoo)
library(xts)
library(readr)
library(scales)


#arima
library(quantmod)
library(tseries)
library(timeSeries)
library(forecast)
library(plotly)
# ********************************************************************************#
# Server Side Logic
# ********************************************************************************#
#File Size Limitation
options(shiny.maxRequestSize = 9*1024^2)
shinyServer(function(input, output,session) {

  #read in reference CSV for public holidays
  holidays <- reactive({
    myholidays <- read.csv(file = './data/public_holidays.csv')
    return(myholidays)
  })

  #read template
  template <- reactive({
    mytemplate <- read.csv(file = './data/template.csv')
    return(mytemplate)
  })

 # Read uploaded file
  mydata <- reactive({
    inFile <- input$file1
    if (is.null(inFile))
      return(NULL)
    data <-read.csv(inFile$datapath, sep=input$sep)
    data

    #outputing the head or all data only
    if (input$disp == "head") {
      return(head(data))
    }
    else {
      return(data)
    }
```

```r
  })

  #Apply Filters
    data <- reactive({
      if (input$drill == 'yes'){
         itemdata_sr <- subset(mydata(),Custom_Category==input$item)
         itemdata_sr_loc <- subset(itemdata_sr,Location==input$location)
         itemdata_sr_loc_ser <- subset(itemdata_sr_loc,Service==input$service)
         itemdata <- subset(itemdata_sr_loc_ser,Sector==input$sector)
         itemdata$Date <- as.Date(paste(itemdata$Date, sep = ""), format = "%d-%b-%y")

         #Remove N/A Dates
         itemdata <- itemdata %>% filter(!is.na(Date))
         return(itemdata)
         itemdata
      } else {
         itemdata <- subset(mydata())
         itemdata$Date <- as.Date(paste(itemdata$Date, sep = ""), format = "%d-%b-%y")
         return(itemdata)
         itemdata
      }
   })

# Update Inputs with data from the uploaded file
 observe({
  updateSelectInput(session, "item", choices = unique(mydata()[,c(1)])) # first column
  updateSelectInput(session, "location", choices = unique(mydata()[,c(7)])) # 7th column
  updateSelectInput(session, "service", choices = unique(mydata()[,c(8)])) # 8th Column
  updateSelectInput(session, "sector", choices = unique(mydata()[,c(10)])) # 8th Column
 })


  #DECLARE TIME SERIES DATA
  #Sales
  ts_Sales <- reactive({
   itemdata <- ts(data()[,c(3)],start=c(2020,4),frequency=365)
   return(itemdata)
  })
  #Sales_Refunds
   ts_Sales_Refund <- reactive({
   itemdata <- ts(data()[,c(4)],start=c(2020,4),frequency=365)
   return(itemdata)
  })
  #Purchases
   ts_Purchases <- reactive({
   itemdata <- ts(data()[,c(5)],start=c(2020,4),frequency=365)
   return(itemdata)
  })
  #Purchase_cancelation
   ts_Purchase_Cancelation <- reactive({
   itemdata <- ts(data()[,c(6)],start=c(2020,4),frequency=365)
   return(itemdata)
  })

  # ****************************************************************************
*#
  # Download Dataset Template
```

```
   # ***************************************************************************
*#

  # Download CSV template
  # Downloadable csv of selected dataset ----
  output$downloadData <- downloadHandler(
    filename = function() {
     paste(input$template, "template.csv", sep = "")
    },
    content = function(file) {
     write.csv(template(), file, row.names = FALSE)
    }
  )


 # ****************************************************************************
#
 # Text Section Section
 # ****************************************************************************
#

 output$caption <- renderText({

 })




 # ****************************************************************************
#
 # Summary Section
 # Display Forecasting accuracy
 # ****************************************************************************
#
 #"seasonal naive","Naive","Simple Expotential Smoothing","Mean Average Forecast","Drift Forecast"
 output$summary <- renderPrint({
  #data Prep
    f_ts_Sales <- ts_Sales()
    f_ts_Sales_Refund <- ts_Sales_Refund()
    f_ts_Purchases<- ts_Purchases()
    f_ts_Purchase_Cancelation <- ts_Purchase_Cancelation()

    #Seasonal Naive Accuracy
    if (input$f_accuracy=='Seasonal Naive'){
     if (input$accuracy_item=='sales'){
      fcsnv <- snaive(f_ts_Sales, h = input$forecast_days_acc)
      accuracy(fcsnv)


     }
     else if (input$accuracy_item=='sales_refund'){
      fcsnv <- snaive(f_ts_Sales_Refund, h = input$forecast_days_acc)
      accuracy(fcsnv)
     }

     else if (input$accuracy_item=='purchases'){
      fcsnv <- snaive(f_ts_Purchases, h = input$forecast_days_acc)
      accuracy(fcsnv)
```

```
  }
  else if (input$accuracy_item=='purchase_cancelation'){
    fcsnv <- snaive(f_ts_Purchase_Cancelation, h = input$forecast_days_acc)
    accuracy(fcsnv)
  }


}
#Naive Accuracy
else if (input$f_accuracy=='Naive'){
    if (input$accuracy_item=='sales'){
      fcnv <- naive(f_ts_Sales, h = input$forecast_days_acc)
      accuracy(fcnv)

    }
    else if (input$accuracy_item=='sales_refund'){
      fcnv <- naive(f_ts_Sales_Refund, h = input$forecast_days_acc)
      accuracy(fcnv)
    }

    else if (input$accuracy_item=='purchases'){
      fcnv <- naive(f_ts_Purchases, h = input$forecast_days_acc)
      accuracy(fcnv)
    }
    else if (input$accuracy_item=='purchase_cancelation'){
      fcnv <- naive(f_ts_Purchase_Cancelation, h = input$forecast_days_acc)
      accuracy(fcnv)
    }


}

#Exponential smoothing accuracy
else if (input$f_accuracy=='Simple Expotential Smoothing'){

  if (input$accuracy_item=='sales'){
    fcses <- ses(f_ts_Sales, h = input$forecast_days_acc)
    accuracy(fcses)

  }
  else if (input$accuracy_item=='sales_refund'){
    fcses <- ses(f_ts_Sales_Refund, h = input$forecast_days_acc)
    accuracy(fcses)
  }

  else if (input$accuracy_item=='purchases'){
    fcses <- ses(f_ts_Purchases, h = input$forecast_days_acc)
    accuracy(fcses)
  }
  else if (input$accuracy_item=='purchase_cancelation'){
    fcses <- ses(f_ts_Purchase_Cancelation, h = input$forecast_days_acc)
    accuracy(fcses)
  }
```

```r
  }
#Mean Average Accuracy
else if (input$f_accuracy=='Mean Average Forecast'){
  if (input$accuracy_item=='sales'){
    fcmean <- meanf(f_ts_Sales, h = input$forecast_days_acc)
    accuracy(fcmean)

  }
  else if (input$accuracy_item=='sales_refund'){
    fcmean <- meanf(f_ts_Sales_Refund, h = input$forecast_days_acc)
    accuracy(fcmean)
  }

  else if (input$accuracy_item=='purchases'){
    fcmean <- meanf(f_ts_Purchases, h = input$forecast_days_acc)
    accuracy(fcmean)
  }
  else if (input$accuracy_item=='purchase_cancelation'){
    fcmean <- meanf(f_ts_Purchase_Cancelation, h = input$forecast_days_acc)
    accuracy(fcmean)
  }


}

#Drift Accuracy
else if (input$f_accuracy=='Drift Forecast'){

  if (input$accuracy_item=='sales'){
    fcdrf <- rwf(f_ts_Sales, h = input$forecast_days_acc)
    accuracy(fcdrf)

  }
  else if (input$accuracy_item=='sales_refund'){
    fcdrf <- rwf(f_ts_Sales_Refund, h = input$forecast_days_acc)
    accuracy(fcdrf)
  }

  else if (input$accuracy_item=='purchases'){
    fcdrf <- rwf(f_ts_Purchases, h = input$forecast_days_acc)
    accuracy(fcdrf)
  }
  else if (input$accuracy_item=='purchase_cancelation'){
    fcdrf <- rwf(f_ts_Purchase_Cancelation, h = input$forecast_days_acc)
    accuracy(fcdrf)
  }


}
else if (input$f_accuracy=='Summary'){


  if (input$accuracy_item=='sales'){
    summary(f_ts_Sales)

  }
```

```
    else if (input$accuracy_item=='sales_refund'){
      summary(f_ts_Sales_Refund)

    }

    else if (input$accuracy_item=='purchases'){
      summary(f_ts_Purchases)

    }
    else if (input$accuracy_item=='purchase_cancelation'){
      summary(f_ts_Purchase_Cancelation)

    }

   }
 })


 # ***********************************************************************************
#
 # Graphs Section
 # ***********************************************************************************
#
 output$graphs_view <- renderPlot({
  # Sales Plot
  if (input$forecast_item == 'sales'){

    if (input$plot == 'Trend Line') {
     p <- ggplot(data(), aes(x = Date, y = Sales)) +
       geom_line(color = "#00AFBB", size = 1)
     p

     p + stat_smooth(
       color = "#FC4E07", fill = "#FC4E07",
       method = "loess"
     )

    }
    else if(input$plot == 'Area Plot'){
       ggplot(data(), aes(x = Date, y = Sales)) +
       geom_area(aes(color = Custom_Category, fill = Custom_Category),
       alpha = 0.5, position = position_dodge(0.8))

    }
    else if(input$plot == 'Scatter Plot'){
       ggplot(data(), aes(x=Date, y=Sales)) + geom_point() +
       geom_point() +
       geom_smooth(method=lm)

    }
    else if(input$plot == 'Regression Line'){
       ggplot(data(), aes(x = Date, y = Sales)) +
       geom_smooth(aes(color = "Sales"), size = 1, method = "loess") +
       theme_minimal()
    }

  }
```

95

```r
# Sales Refund Plot
else if (input$forecast_item == 'sales_refund'){
  if (input$plot == 'Trend Line') {
    p <- ggplot(data(), aes(x = Date, y = Sales_Refund)) +
      geom_line(color = "#00AFBB", size = 1)
    p

    p + stat_smooth(
      color = "#FC4E07", fill = "#FC4E07",
      method = "loess"
    )

  }
  else if(input$plot == 'Area Plot'){
      ggplot(data(), aes(x = Date, y = Sales_Refund)) +
      geom_area(aes(color = Custom_Category, fill = Custom_Category),
      alpha = 0.5, position = position_dodge(0.8))

  }
  else if(input$plot == 'Scatter Plot'){
      ggplot(data(), aes(x=Date, y=Sales_Refund)) + geom_point() +
      geom_point() +
      geom_smooth(method=lm)

  }
  else if(input$plot == 'Regression Line'){
      ggplot(data(), aes(x = Date, y = Sales_Refund)) +
      geom_smooth(aes(color = "Sales"), size = 1, method = "loess") +
      theme_minimal()
  }

}

# Purchase Plot
else if (input$forecast_item == 'purchases'){

  if (input$plot == 'Trend Line') {
    p <- ggplot(data(), aes(x = Date, y = Purchases)) +
      geom_line(color = "#00AFBB", size = 1)
    p

    p + stat_smooth(
      color = "#FC4E07", fill = "#FC4E07",
      method = "loess"
    )

  }
  else if(input$plot == 'Area Plot'){
      ggplot(data(), aes(x = Date, y = Purchases)) +
      geom_area(aes(color = Custom_Category, fill = Custom_Category),
      alpha = 0.5, position = position_dodge(0.8))

  }
  else if(input$plot == 'Scatter Plot'){
      ggplot(data(), aes(x=Date, y=Sales)) + geom_point() +
      geom_point() +
```

```r
        geom_smooth(method=lm)


      }
    else if(input$plot == 'Regression Line'){
        ggplot(data(), aes(x = Date, y = Purchases)) +
        geom_smooth(aes(color = "Sales"), size = 1, method = "loess") +
        theme_minimal()
    }

  }

  # Purchase Cancellation Plots
  else if (input$forecast_item == 'purchase_cancelation'){
    if (input$plot == 'Trend Line') {
      p <- ggplot(data(), aes(x = Date, y = Purchase_Cancelation)) +
        geom_line(color = "#00AFBB", size = 1)
      p

      p + stat_smooth(
        color = "#FC4E07", fill = "#FC4E07",
        method = "loess"
      )

    }
    else if(input$plot == 'Area Plot'){
        ggplot(data(), aes(x = Date, y = Purchase_Cancelation)) +
        geom_area(aes(color = Custom_Category, fill = Custom_Category),
        alpha = 0.5, position = position_dodge(0.8))


    }
    else if(input$plot == 'Scatter Plot'){
        ggplot(data(), aes(x=Date, y=Purchase_Cancelation)) + geom_point() +
        geom_point() +
        geom_smooth(method=lm)


    }
    else if(input$plot == 'Regression Line'){
        ggplot(data(), aes(x = Date, y = Purchase_Cancelation)) +
        geom_smooth(aes(color = "Sales"), size = 1, method = "loess") +
        theme_minimal()
    }
  }

})
# *******************************************************************************
#
# Forecasting
# Logic for Sales, Sales Refund, Purchases and Purchase Refunds
# Plot Logic
# *******************************************************************************
#

output$forecast_view <- renderPlot({
  # Sales Forecasting
  if (input$forecast_item == 'sales'){
    SALES_QUANTITIES <- ts_Sales()
    if (input$fplot == 'Seasonal Naive'){
```

```
    fcsnv <- snaive(SALES_QUANTITIES, h = input$forecast_days, format = "%d-%b-%y")
    autoplot(fcsnv)
  #return(fcsnv)
  }
  else if(input$fplot=='Naive'){
    fcnv <- naive(SALES_QUANTITIES, h = input$forecast_days)
    autoplot(fcnv)
  }
  else if(input$fplot == "Simple Expotential Smoothing"){
    fcses <- ses(SALES_QUANTITIES, h = input$forecast_days)
    autoplot(fcses)
  }
  else if(input$fplot == "Mean Average Forecast"){
    fcmean <- meanf(SALES_QUANTITIES, h = input$forecast_days)
    autoplot(fcmean)
  }
  else if(input$fplot == "Drift Forecast"){
    fcdrf <- rwf(SALES_QUANTITIES, h = input$forecast_days)
    autoplot(fcdrf)
  }

}

# Sales Refund Forecasting
else if (input$forecast_item == 'sales_refund'){
  SALES_REFUND_QUANTITIES <- ts_Sales_Refund()
  if (input$fplot == 'Seasonal Naive'){
    fcsnv <- snaive(SALES_REFUND_QUANTITIES, h = input$forecast_days)
    autoplot(fcsnv)
  }
  else if(input$fplot=='Naive'){
    fcnv <- naive(SALES_REFUND_QUANTITIES, h = input$forecast_days)
    autoplot(fcnv)
  }
  else if(input$fplot == "Simple Expotential Smoothing"){
    fcses <- ses(SALES_REFUND_QUANTITIES, h = input$forecast_days)
    autoplot(fcses)
  }
  else if(input$fplot == "Mean Average Forecast"){
    fcmean <- meanf(SALES_REFUND_QUANTITIES, h = input$forecast_days)
    autoplot(fcmean)
  }
  else if(input$fplot == "Drift Forecast"){
    fcdrf <- rwf(SALES_REFUND_QUANTITIES, h = input$forecast_days)
    autoplot(fcdrf)
  }
}

# Purchase Forecasting
else if (input$forecast_item == 'purchases'){
  PURCHASE_QUANTITIES <- ts_Purchases()
  if (input$fplot == 'Seasonal Naive'){
    fcsnv <- snaive(PURCHASE_QUANTITIES, h = input$forecast_days)
    autoplot(fcsnv)
  }
  else if(input$fplot=='Naive'){
    fcnv <- naive(PURCHASE_QUANTITIES, h = input$forecast_days)
```

```r
      autoplot(fcnv)
     }
    else if(input$fplot == "Simple Expotential Smoothing"){
      fcses <- ses(PURCHASE_QUANTITIES, h = input$forecast_days)
      autoplot(fcses)
     }
    else if(input$fplot == "Mean Average Forecast"){
      fcmean <- meanf(PURCHASE_QUANTITIES, h = input$forecast_days)
      autoplot(fcmean)
     }
    else if(input$fplot == "Drift Forecast"){
      fcdrf <- rwf(PURCHASE_QUANTITIES, h = input$forecast_days)
      autoplot(fcdrf)
     }

   }
   # Purchase Forecasting
   else if (input$forecast_item == 'purchase_cancelation'){
    PURCHASE_ORDER_CANCELATION_QUANTITIES <- ts_Purchase_Cancelation()
    if (input$fplot == 'Seasonal Naive'){
      fcsnv <- snaive(PURCHASE_ORDER_CANCELATION_QUANTITIES, h = input$forecast_days)
      autoplot(fcsnv)
     }
    else if(input$fplot=='Naive'){
      fcnv <- naive(PURCHASE_ORDER_CANCELATION_QUANTITIES, h = input$forecast_days)
      autoplot(fcnv)
     }
    else if(input$fplot == "Simple Expotential Smoothing"){
      fcses <- ses(PURCHASE_ORDER_CANCELATION_QUANTITIES, h = input$forecast_days)
      autoplot(fcses)
     }
    else if(input$fplot == "Mean Average Forecast"){
      fcmean <- meanf(PURCHASE_ORDER_CANCELATION_QUANTITIES, h = input$forecast_days)
      autoplot(fcmean)
     }
    else if(input$fplot == "Drift Forecast"){
      fcdrf <- rwf(PURCHASE_ORDER_CANCELATION_QUANTITIES, h = input$forecast_days)
      autoplot(fcdrf)
     }

   }
  })
  # ****************************************************************************
#
 # Forecasting
 # Logic for Sales, Sales Refund, Purchases and Purchase Refunds
 # Accuracy Logic
 # ****************************************************************************
#
 output$forecast_accuracy <- renderTable({
   # Sales Forecasting
   if (input$forecast_item == 'sales'){
    SALES_QUANTITIES <- ts_Sales()
    if (input$fplot == 'Seasonal Naive'){
      fcsnv <- snaive(SALES_QUANTITIES, h = input$forecast_days, format = "%d-%b-%y")
      return(fcsnv)
     }
```

99

```r
  else if(input$fplot=='Naive'){
   fcnv <- naive(SALES_QUANTITIES, h = input$forecast_days)
   return(fcnv)
  }
  else if(input$fplot == "Simple Expotential Smoothing"){
   fcses <- ses(SALES_QUANTITIES, h = input$forecast_days)
   return(fcses)
  }
  else if(input$fplot == "Mean Average Forecast"){
   fcmean <- meanf(SALES_QUANTITIES, h = input$forecast_days)
   return(fcmean)
  }
  else if(input$fplot == "Drift Forecast"){
   fcdrf <- rwf(SALES_QUANTITIES, h = input$forecast_days)
   return(fcdrf)
  }

 }

 # Sales Refund Forecasting
 else if (input$forecast_item == 'sales_refund'){
  SALES_REFUND_QUANTITIES <- ts_Sales_Refund()
  if (input$fplot == 'Seasonal Naive'){
   fcsnv <- snaive(SALES_REFUND_QUANTITIES, h = input$forecast_days)
   return(fcsnv)
  }
  else if(input$fplot=='Naive'){
   fcnv <- naive(SALES_REFUND_QUANTITIES, h = input$forecast_days)
   return(fcnv)
  }
  else if(input$fplot == "Simple Expotential Smoothing"){
   fcses <- ses(SALES_REFUND_QUANTITIES, h = input$forecast_days)
   return(fcses)
  }
  else if(input$fplot == "Mean Average Forecast"){
   fcmean <- meanf(SALES_REFUND_QUANTITIES, h = input$forecast_days)
   return(fcmean)
  }
  else if(input$fplot == "Drift Forecast"){
   fcdrf <- rwf(SALES_REFUND_QUANTITIES, h = input$forecast_days)
   return(fcdrf)
  }
 }

 # Purchase Forecasting
 else if (input$forecast_item == 'purchases'){
  PURCHASE_QUANTITIES <- ts_Purchases()
  if (input$fplot == 'Seasonal Naive'){
   fcsnv <- snaive(PURCHASE_QUANTITIES, h = input$forecast_days)
   return(fcsnv)
  }
  else if(input$fplot=='Naive'){
   fcnv <- naive(PURCHASE_QUANTITIES, h = input$forecast_days)
   return(fcnv)
  }
  else if(input$fplot == "Simple Expotential Smoothing"){
   fcses <- ses(PURCHASE_QUANTITIES, h = input$forecast_days)
```

```
      return(fcses)
     }
    else if(input$fplot == "Mean Average Forecast"){
     fcmean <- meanf(PURCHASE_QUANTITIES, h = input$forecast_days)
     return(fcmean)
     }
    else if(input$fplot == "Drift Forecast"){
     fcdrf <- rwf(PURCHASE_QUANTITIES, h = input$forecast_days)
     return(fcdrf)
     }

    }
   # Purchase Forecasting
   else if (input$forecast_item == 'purchase_cancelation'){
    PURCHASE_ORDER_CANCELATION_QUANTITIES <- ts_Purchase_Cancelation()
    if (input$fplot == 'Seasonal Naive'){
     fcsnv <- snaive(PURCHASE_ORDER_CANCELATION_QUANTITIES, h = input$forecast_days)
     return(fcsnv)
     }
    else if(input$fplot=='Naive'){
     fcnv <- naive(PURCHASE_ORDER_CANCELATION_QUANTITIES, h = input$forecast_days)
     return(fcnv)
     }
    else if(input$fplot == "Simple Exponential Smoothing"){
     fcses <- ses(PURCHASE_ORDER_CANCELATION_QUANTITIES, h = input$forecast_days)
     return(fcses)
     }
    else if(input$fplot == "Mean Average Forecast"){
     fcmean <- meanf(PURCHASE_ORDER_CANCELATION_QUANTITIES, h = input$forecast_days)
     return(fcmean)
     }
    else if(input$fplot == "Drift Forecast"){
     fcdrf <- rwf(PURCHASE_ORDER_CANCELATION_QUANTITIES, h = input$forecast_days)
     return(fcdrf)
     }
    else if(input$fplot == "Armia"){
     aadata <- auto.arima(PURCHASE_ORDER_CANCELATION_QUANTITIES)
     aaforecast <- forecast(aadata(), h = input$forecast_days)
     return(aaforecast)
     }

   }
  })



  # *****************************************************************************
#
 # Dataset view
 # View Uploaded Dataset
 # *****************************************************************************
#
 output$data_set <- DT::renderDataTable({
  DT::datatable(mydata(),editable = TRUE)
  })
```

```
# ***********************************************************************
#
# Holidays Plot
# View Holidays
# ***********************************************************************
#
output$holidays <- DT::renderDataTable({
  DT::datatable(holidays(),editable = TRUE)
})




# ***********************************************************************
#
# Dynamically Render the Presentation Layer Tabs After Data Upload
# ***********************************************************************
#
output$tb <- renderUI({
  if(is.null(mydata()))
    h5(tags$img(src='tabiri.png', heigth=900, width=900))
  else
    tabsetPanel(
      #Forecat Tab
      tabPanel(("Forecast"),
          column(6,h4("Forecast Plot"), plotOutput("forecast_view"),
              selectInput(
                "fplot",
                "Chose a Forecast algorithm:",
                c("Seasonal Naive","Naive","Simple Exponential Smoothing","Mean Average Forecast","
Drift Forecast")),
              # SLIDER. NUMBER OF DAYS TO FORECAST
              sliderInput("forecast_days",
                    "Number of Days to Forecast:",
                    value = 14,
                    min = 2,
                    max = 31)
          ),
          column(6, h4("Forecast Accuracy"),tableOutput("forecast_accuracy")
          )
      ),


    # Graph Tab
    tabPanel("Graphs",plotOutput("graphs_view"),
          selectInput(
            "plot",
            "Chose your desired plot type:",
            c("Trend Line","Area Plot","Scatter Plot","Regression Line"))
    ),
    tabPanel("Forecast Accuracy", verbatimTextOutput("summary"),

          selectInput(
            "f_accuracy",
            "Select a Forecast algorithm:",
            c("Naive","Seasonal Naive","Simple Exponential Smoothing","Mean Average Forecast","Drift
Forecast","Summary")),
```

```
        selectInput("accuracy_item", "check accuracy for :",
                choices =  c(
                  "Sales" = "sales",
                  "Sales Returns" = "sales_refund",
                  "Purchases" = "purchases",
                  "Cancelled POs" = "purchase_cancelation", selected = NULL)),

        # SLIDER. NUMBER OF DAYS TO FORECAST
        sliderInput("forecast_days_acc",
                "Number of Days to Forecast:",
                value = 2,
                min = 2,
                max = 31)
        ),
    tabPanel("Dataset",dataTableOutput("data_set")),

    tabPanel("Holidays", dataTableOutput("holidays") )

   )
 })
})
```