Predicting Customer Churn Rates in Kenyan Banks Using Time Series Analysis

Abstract:Churn rates in business is defined as the rate at which customers end their relationship with a certain organization within a specified period. In the recent past, there has been a rise in competition especially in the banking sector. This is partly attributed to advancement in technology leading to easy entry of fintech entities in banking industry as new players. Moreover, the rise in Fintech companies generates stiff competition which makes banking customer retention vital to secure profitability levels and protect existing market shares. For the companies to enhance profitability, customer retention oriented strategies need to be put in place. This study explored the application of time series analysis to forecast Kenyan commercial banks' customer churn rates while identifying customer patterns that were likely to drive proactive retention strategies in a bid to curb churn rates. The study developed time series models that will accurately predict churn rates in Kenya commercial banks. The study made use of historical bank data and applied Autoregressive Integrated Moving Average(ARIMA) and exponential smoothing models to predict future bank trends. Time series forecasting models stood out for their simplicity in computation, minimal parameter requirements and their ability to handle different data patterns including trends and seasonality. The analysis included, customer account activity, transaction volume, loan uptake and repayment. External economic indicators along with regulatory changes were included in the study in order to boost the forecasting strength of the models. The research findings indicated how time series analysis effectively detect high risk churn periods which help banks to create targeted retention strategies. Results from the study also indicated that economic downturns together with regulatory changes contributed to customers attrition. A limitation of this study was the use of simulated data that might not entirely reflect what happens in the Kenyan banks in terms of customer churn. These time series models should be validated by real life banking data in the future to ensure a better generalization and evaluate their predictability in the real market environment.

Keywords: Digital Credit Provider, ARIMA, Moving Average, Simulation.

1 Introduction

Customer retention remains a major challenge for commercial banks worldwide, a situation equally reflected in Kenya's banking sector. Currently, the competition in banking sector in Kenya has grown fierce because of mushrooming financial service providers including traditional banks, microfinance institutions, Savings and Credit Cooperative Organizations (SACCOs) as well as digital credit providers(DCP). The adoption of mobile banking solutions such as M-Pesa, Airtel Money together with digital lending apps has reshaped the market competition by providing customers both convenience and accessibility to financial solutions. According to a study by [5], mobile banking drives population level financial inclusion but also face impediments to its development. Moreover, [1] observed that the rapid rise of fintech entities has intensified competition in the banking industry, compelling commercial banks to adopt analytical tools and predictive models to strengthen digital risk management and customer retention strategies. Kenyan banks now face increased pressure to both protect its customer base and find new ones to sustain their positions in the industry.

Kenya's banking sector serves a wide array of customers including elite urban dwellers to less enlightened semi-urban and rural dwellers, each with unique banking needs. Through Central Bank of Kenya (CBK) backing of financial inclusion, various banks have invested heavily in digital platforms and network expansion. The research by [6] showed that customer retention improved through emphasis on service excellence and product quality alongside website usability, competitive pricing, prompt delivery and customization. External economic downturns combined with inflation and regulatory changes substantially worsen the customer retention issue.

Kenyan banks through wide range of services offered can meet customers' needs to enhance customer satisfaction and hence reduce the chance of customer exiting from the organization. This can be done through digitization and proactive customer engagement which will lead to easy access to services. According to [2], Failure to embrace digization can lead to higher churn rate .This has earlier happened to companies such as Kodak.Predicting and understanding customer churn serves as a basic requirement for Kenyan banks to maintain competitive advantage. A time series analysis gives banks the power to understand past patterns and predict future customer churn rates so they can develop tailored preventive strategies. [4], in their study titled "Churn Prediction with Sequential Data and Deep Neural Networks: A Comparative Analysis" which examines churn prediction through recurrent neural networks (RNNs) together with long short-term memory (LSTM) networks in financial service industries. Through RNN implementations scientists found they could boost churn prediction precision which supports developing successful retention approaches. Banks can utilize account activity data and transaction patterns combined with loan performance indicators to find customers at risk for defection so they can develop customized retention strategies.

The study examined the use of time series analysis to forecast churn rates in Kenyan banking institutions. The research intendd to close the predictive analytics-practical decision gap through tools which help banks maintain customers while growing sustainably against industry competition.

2 Literature Review

2.1 Introduction

Statistical researchers have explored factors that influence customer retention in Kenyan banks while uncovering important patterns and trends related to customer attrition.

For instance,[8], investigated on factors affecting customer retention in Barclays Bank of Kenya. She collected data from Barclays bank customers using semi-structured questionnaire and interview done on the staff of Barclays bank. Data was analyzed using SPSS and the findings of the study was accuracy of the transactions, accessibility of the bank location, quality of service and appearance of the bank contribute to customer churn. In addition, what other customers say about the bank was also noted as the key factor that can lead to customer churn.

Another study by [6] analyzed the customer retention strategies implemented by banks in Kenya. The study utilized a descriptive survey design where senior managers and customer service representatives were issued with questionnaires which explored strategies covering loyalty programs along with relationship management and service quality. The author utilized both qualitative and quantitative techniques to analyze the obtained data. The findings from the study showed that relationship marketing was the most commonly adopted strategy. However, loyalty programs and personalized customer service were emphasized as essential tools for retention. It was also noted that banks invested heavily in staff training to enhance customer service. Even though this study provided valuable insights on customer retention strategies it suffered shortcomings. For instance, the study heavily dependent on self-reported data, something likely to introduce bias. Furthermore, the study did not address the long-term effectiveness of these strategies neither did the author compare the strategies with customer retention practices in other industries.

2.2 ARIMA model

A study by [1] investigated how Artificial Neural Networks (ANNs) and ARIMA models predict Customer Lifetime Value (CLV) while assessing their effects on societal and industrial growth. The research used time series data for purchasing behavior of customers. Both ANNs and ARIMA applications were preferred in this study because of their capability to handle nonlinear and linear patterns respectively. A comprehensive evaluation of the predictive accuracy centered around comparing the two methods. The comparison of model predictions was done by use of Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) to evaluate their predictive accuracy. The findings from the study. The results showed that both ARIMA and artificial neural network models can be used to predict customer lifetime value.

3 MATERIAL AND METHOD

3.1 Data

The study used 10,000 simulated data based on information from three banks in Kenya, KCB bank limited, Equity Bank Kenya limited and Cooperative bank of Kenya. The dataset used was defined over period of three years so as to capture seasonal trends and behavioral changes that greatly impact churn rates.

3.2 ARIMA (Auto-Regressive Integrated Moving Average) Model

The widely used time series forecasting method is ARIMA which integrates three key components through its AR (Auto-Regressive) and I (Integrated) and MA (Moving Average) elements. The mathematical expression for the ARIMA model takes the following form:

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (1)$$

Where:

- Y_t is the value of the time series at time t,
- p is the order of the Auto-Regressive (AR) term,
- $\phi_1, \ldots, \phi_p$ are the AR coefficients,
- q is the order of the Moving Average (MA) term,
- $\theta_1, \ldots, \theta_q$ are the MA coefficients,
- ε_t is the white noise error term at time t.

The ARIMA model has three main parameters:
P: The order of the auto-regression (AR), i.e., the number of lag observations included in the model.

d: The degree of differencing (I), used to make the series stationary.

q: The order of the moving average (MA), representing the number of lagged forecast errors in the prediction.

Stationarity check and differencing

Time series data must be stationary before applying ARIMA forecasting method. The first step was therefore differencing when data remains non-stationary. We obtained the differenced data by applying the formular below;

$$Y_t' = Y_t - Y_{t-1} \tag{2}$$

Where Y_t' is the first-order differenced series at time t, Y_t is the original series at time t, and Y_{t-1} is the previous value of the series at time $t-1$.

The process was repeated until the series became stationary. The AR part captures the relationship between the current value and its past values while the MA part accounts for the dependency between the current value and past forecast errors. We then selected optimal values for p, d, and q using AIC (Akaike Information Criterion). The ARIMA model was fitted then used to make forecasts by projecting future churn rates in banks.

3.3 Exponential Smoothing Models

A bank, upon implementing new strategies or interventions for instance promotional offers, a successful marketing campaign etc, recent data will more accurately reflect the effectiveness of these actions. Such a scenario calls for a model that gives emphasis on recent occurrences and Exponential Smoothing Model is a good example. By giving more weight to these recent behaviors, the model can gauge the impact of interventions more accurately, adjusting churn predictions based on the success of these actions. The general form of exponential smoothing is:

$$\hat{Y}_{t+h} = \alpha Y_t + (1-\alpha)Y_{t-1} \tag{3}$$

Where;

- $\hat{Y}_{t+h}$ is the forecast for the h-step ahead, future churn rate in this case.
- α is the smoothing parameter (between 0 and 1), controlling the weight given to the most recent observation.

Since our dataset was comprised of historical data from banks, exhibiting both trend and seasonality, the Exponential Smoothing model can further be expanded to the Holt-Winters Seasonal Model.

4 RESULTS AND DISCUSSION

This chapter outlines and discusses the empirical findings of the implementation of time series modeling in prediction of customer churn in Kenyan banks. The analysis was organized in accordance with the following objectives:

i. To identify key patterns and trends that contribute to customer churn rates in Kenyan banks ; and

ii. To develop time series forecasting models that accurately predict churn rates in Kenyan banks.

Each section has a descriptive analysis followed by model fitting procedures, diagnosis evaluation and interpretation of the forecast results on the basis of the empirical/theoretical literature.

4.1 Identifying Key Patterns and Trends that Contribute to Customer Churn Rates in Kenyan banks

4.1.1 Exploratory Data Analysis

An exploratory data analysis was undertaken to investigate temporal variation in churn behavior within quarterly periods. The analysis revealed hidden patterns of long term trends, seasonal changes and irregular changes that defined customer attrition in banking industry. Figure 1 presents the observed churn rates over the study period, offering an initial visual interpretation of how customer exit behavior evolves with time. Figure 1 illustrates observed churn rates over the period of study. The data clearly shows an upward trend with regular fluctuations illustrative of seasonal variations in banking sector.

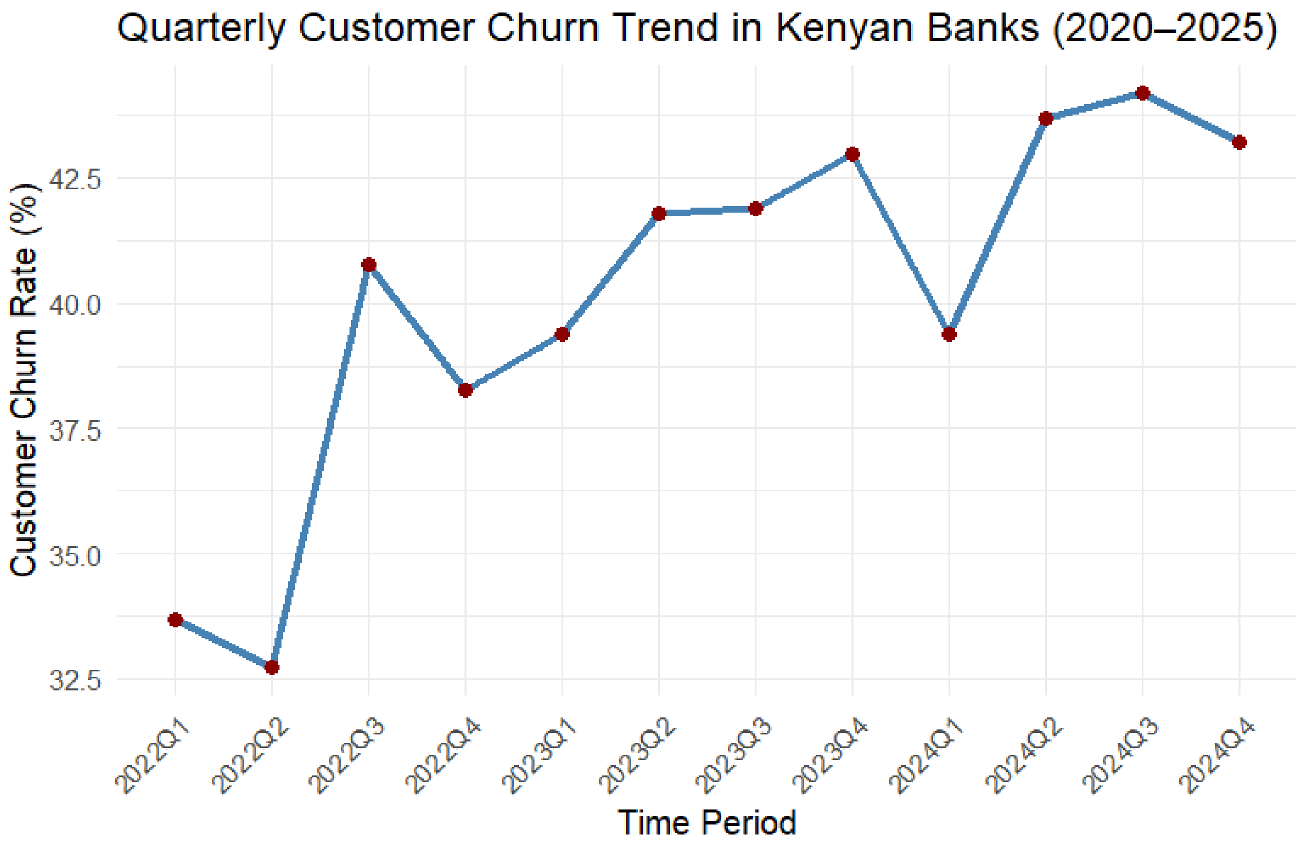

Figure 1: Quarterly customer churn trend in Kenyan banks.

The plot showed cyclical peaks in the last quarter of each year.This coincided with heightened competition from digital lenders and income mobility among customers. The observed recurrence confirmed our choice of time series models, its capability to capture not only stochastic but also predictable features such as trend, seasonality and variations.

To complement the line graph in Figure 1, Table 1 gives a numerical summary of quarterly churn rates in the period of study. The data in the table, gives a clearer comparison of seasonal fluctuations and average churn rates over time. This strengthens the observed temporal patterns and informs the subsequent tasks of predictive time series models.

The results in Table 1 shows an increasing trend in customer churn rates over the period of study. However, minor fluctuations between successive quarters were observed. Notably, the churn rates rose from averagely 33.69% in 2022Q1 to more than 43% by the end of 2024. This showed a gradual increase of customer attrition in Kenyan banks. This pattern further revealed regular highs during the last two quarters of every year.This suitably aligns with seasonal dynamics observed during exploratory data analysis. Such repetitive variations imply that customer changing traits could be influenced by cyclical factors such as year end financial realignments, successful market campaigns and changing credit demands.

Table 1: Quarterly Customer Churn Rates in Kenyan Banks (2020–2025)

Time Period	Seasonal Component	Avg. Churn Rate
2022Q1	1	33.69
2022Q2	2	32.73
2022Q3	3	40.77
2022Q4	4	38.25
2023Q1	1	39.38
2023Q2	2	41.78
2023Q3	3	41.90
2023Q4	4	42.98
2024Q1	1	39.38
2024Q2	2	43.70
2024Q3	3	44.18
2024Q4	4	43.22

We also computed descriptive statistics in order to provide an overview of the data's central tendency and variability. These statistics provided a clearer understanding of the overall distribution of churn behavior among Kenyan banks. Table 2 shows the results of this descriptive analysis.

Table 2: Descriptive Statistics of Quarterly Customer Churn Rates in Kenyan Banks

Statistic	Value
Mean	40.16
Median	40.77
Minimum	32.73
Maximum	44.18
Standard Deviation (SD)	3.58
Coefficient of Variation (%)	8.92

The results in Table 2 showed that the mean quarterly churn rate across Kenyan banks was 40.16%. This closely aligns with the approximated median of 40.77%. The closeness of the two values shows a relatively balanced distribution of churn rates over time. The observed range from 32.73% to 44.18% implied a reasonable variation, while the standard deviation of 3.58% and coefficient of variation of 8.92% indicated that churn levels remained consistent across the periods(quarters). In summary, the descriptive statistics confirmed that customer attrition followed a stable but progressively increasing trend. This therefore called for further exploration in order to unearth deeply underlying temporal features.

Further, we derived seasonal indices in order to quantify the relative effect of each period on overall customer churn. These indices revealed periods of either above or below average churn rates. This offered deeper insight into the cyclical nature of customer churn behavior in Kenyan banks. Table 3 summarizes the computed indices and the average churn rates for each quarter. From Table 3 above, it is clear that cyclical variations in customer churn rates significantly differed across the four quarters. The last two quarters had indices above 1 (1.05 and 1.03, respectively). These are without doubt periods of higher than average customer churn rates. On the other hand, the first two quarters, recorded indices below one, reflecting lower churn rates.

Table 3: Seasonal Indices for Quarterly Customer Churn

Seasonal Component	Average Churn (%)	Seasonal Index
1	37.48	0.93
2	39.40	0.98
3	42.28	1.05
4	41.48	1.03

This pattern suggests that customer exits tend to increase during the latter quarters of the year. Factors such as high competition, year end financial adjustments and heightened income mobility among customers are responsible for this occurrence. Such dynamics underscore the importance of incorporating seasonality component.

By carrying out exploratory data analysis, descriptive statistics and seasonal decomposition, we were able to fulfill the first objective of the study which aimed at identifying key patterns and trends that contribute to customer churn rates in Kenyan banks. The findings showed a gradual but steadily upward trend in churn rates, moderate variability across periods and distinct seasonal highs during the latter part of each year. These revelations confirm the presence of both long term and cyclical variations in customer churn, providing a solid justification for the development of time series forecasting models under objective two of the study.

In summary, the observed seasonal and upward trend in Kenyan banking churn confirms the broader hypothesis and supports the significance of dynamic, time sensitive modeling approaches in predicting customer attrition.

4.2 Developing Time Series Forecasting Models To Predict Churn Rates in Kenyan Banks

Following the results of the exploratory analysis of Section 4.1, which successfully revealed the major patterns and temporal trends of customer churn behavior, this section is dedicated to fulfilling the second specific objective of the study that aimed at creating time series forecasting models that would accurately forecast the rates of customer churn within Kenyan banks. The analysis acknowledges the trend and seasonal characteristics of churn dynamics, and thus the use of forecasting methods, such as the Autoregressive Integrated Moving Average (ARIMA), Holt-Winters, and Autoregressive Integrated Moving Average with external Variables (ARIMAX) models.Since the aspects of dynamics of churn follow both a trend and a seasonal pattern, the analysis uses forecasting methods, such as the Autoregressive Integrated Moving Average (ARIMA), the Holt Waver, and the Autoregressive Integrated Moving Average with external Variables (ARIMAX) model. These models were specifically chosen due to their strength in the capturing stochastic, deterministic, and externally driven aspects of time varying data. The analysis of the same below gives specification of the model, estimation, diagnostic analysis and comparison of forecasting performance to be able to decide which method was the most reliable in predicting churn rates with time.

4.2.1 ARIMA Model Selection Process

Identifying a suitable ARIMA model was key to accurate time series prediction of customer churn rates. It included determining the order of Autoregressive (AR) terms, differencing (I) needed to obtain stationarity, and moving average (MA) elements which best represented the structure of the data. Proper model selection balances goodness-of-fit with parsimony, avoiding over fitting while preserving the essential temporal dynamics. Statistical tests used to select the ARIMA model in this study included the Augmented Dickey Fuller (ADF) test of stationarity, and inspection of autocorrelation (ACF) and partial autocorrelation (PACF) plots to identify the best combination of model parameters that best predicts customer churn in Kenyan banks.Figure 2 below shows the step by step ARIMA model selection procedure used in this analysis.

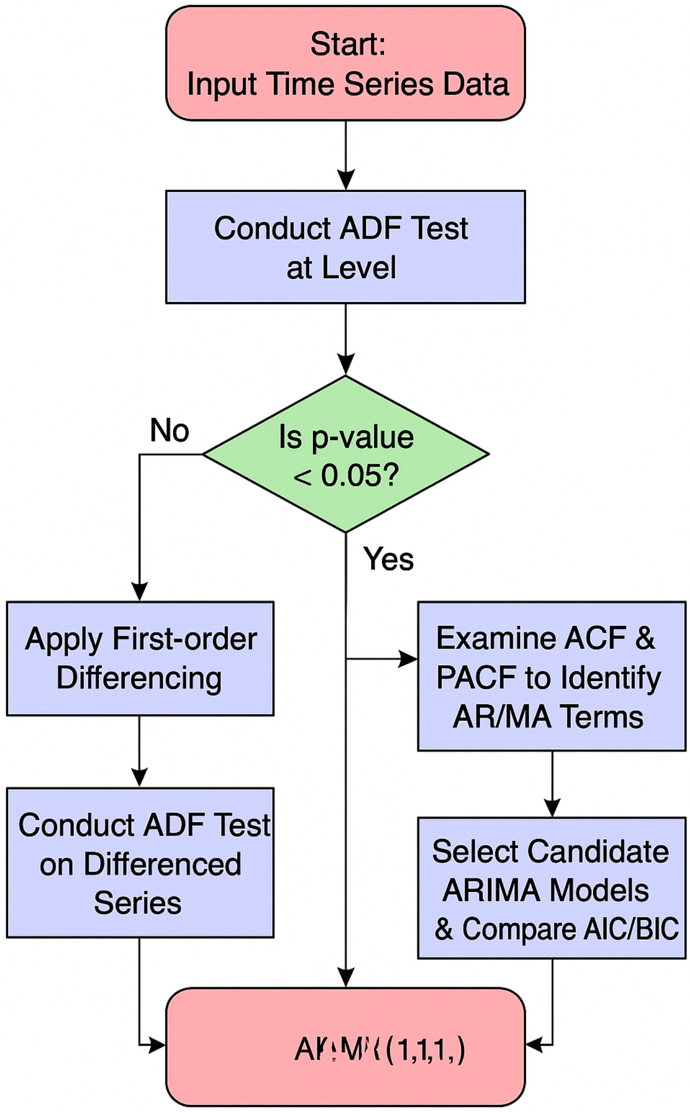

Figure 2: ARIMA Process Flow

4.2.2 ARIMA Model Estimation

The initial forecasting model adopted was the Autoregressive Integrated Moving Average (ARIMA) model which was chosen due to its strength in estimating stationary stochastic processes. The series was tested using the Augmented DickeyFuller (ADF) test, which proved non-stationarity at the level form ($p > 0.05$). Order one differencing was then applied, and the series became stationary ($p < 0.05$). A inspection of the autocorrelation (ACF) and partial autocorrelation (PACF) plots was used to inform the choice of an(1,1,1) specification.

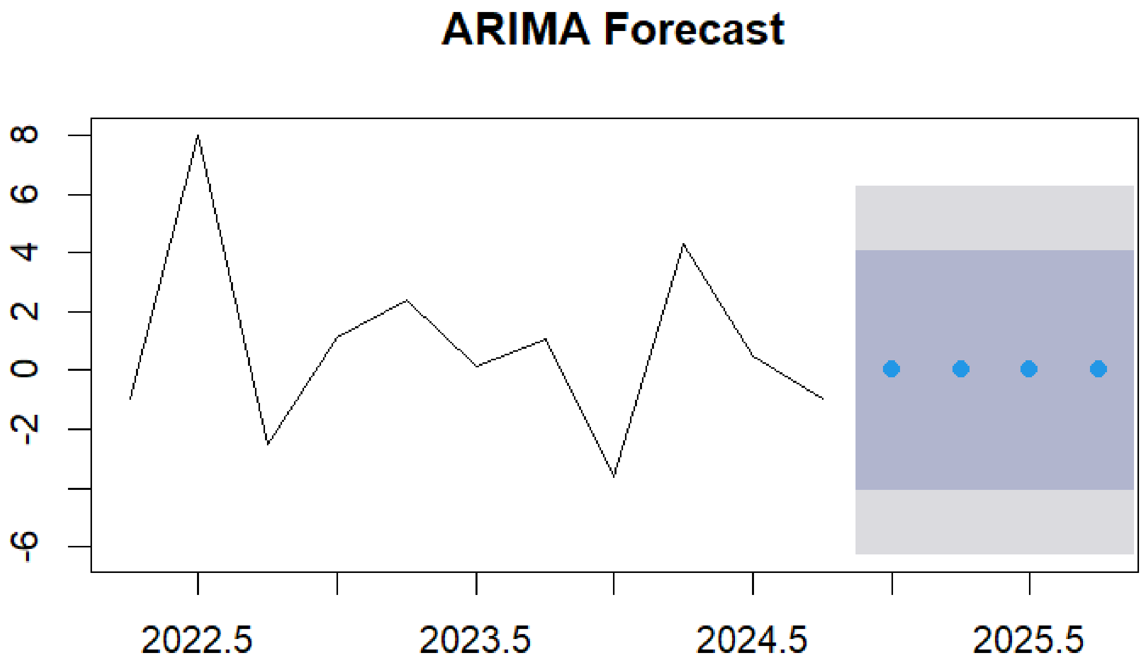

Figure 3: ARIMA(1,1,1) forecast of customer churn rates.

As presented in Figure 3, The ARIMA model was able to capture the underlying trend in customer churn as well as the short term variation. This model predicted a medium term though consistent increase in the churn rates during the next four quarters, indicating that the customer retention in the industry was becoming more unstable.

4.2.3 Fitting the ARIMA Model

After identifying the most appropriate ARIMA parameters via the stationarity testing and discussion of the autocorrelation structures, the model was then fitted to the observed churn data. This step entailed estimation of the model coefficients and coming up with forecasts to determine the appropriateness of the ARIMA specification to capture the underlying trends and short term oscillations in customer churn rates. In this sub section, the ARIMA(1,1,1) model was implemented, its parameters estimated and the time series illustrated accordingly.

Table 4: ARIMA(1,1,1) Model Coefficients for Customer Churn Rate

Term	Estimate	Std. Error	z-value	p-value
AR1	0.339	0.009	37.667	$<$ .001
MA1	-1.000	0.000	-Inf	$<$ .001

According to the findings shown in Table 4, the fitted ARIMA(1,1,1) model of the customer churn rates showed statistically significant coefficients of Autoregressive at lag 1 (AR1 = 0.339, p = .001) and coefficient of Moving average at lag 1 (MA1 = -1.000, p = .001). These outcomes indicated that the change in the churn rates of the previous quarter and the forecasting errors of the past period played a key role in the current churn processes. On this basis, the ARIMA(1,1,1) model was discovered to be:

$$\Delta Y_t = 0.339\,\Delta Y_{t-1} + \epsilon_t - 1.000\,\epsilon_{t-1}$$

where ΔY_t denotes the differenced customer churn rate at time t, and ϵ_t the error term at time t. This formulation captured both the short term autocorrelation and the moving average effect in the quarterly churn data

.

4.2.4 Predicted Churn Trends Over Four Quarters

In order to determine the anticipated trends during the upcoming year, we used the fitted ARIMA model to make predictions of the next four quarters. The projected values gave a clue to the future trend which could be used to make proactive plans and make informed decisions by the Kenyan banks. These forecasted results are represented in Figure 4 below:

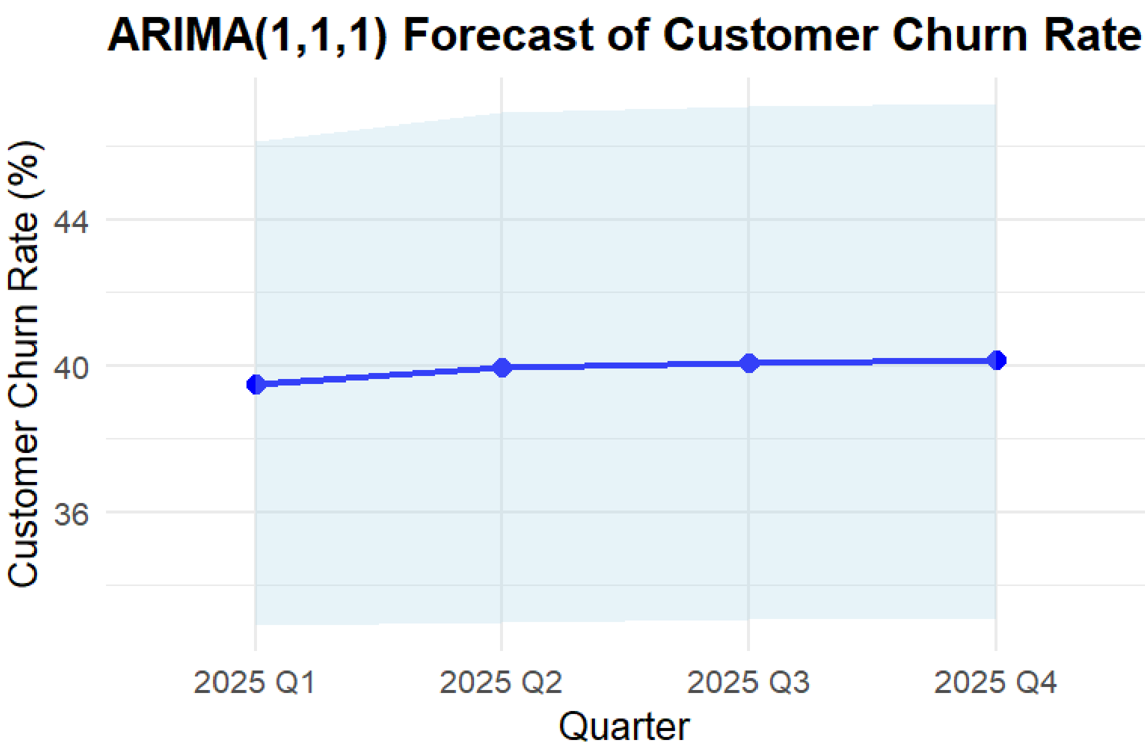

Figure 4: Forecasted ARIMA results for 4 quarters

The four-quarter forecast presented in Figure 4 was a projection of the trends given by the fitted ARIMA model. These estimated values helped in accomplishing the second specific objective , Creating time series forecasting models that predict churn rates accurately. The findings identified the intervals that may experience increased or decreased churn, thus giving a hint that could inform timely intervention and strategic decision making.

4.2.5 ARIMA Model Validation and Residual Analysis

Residual diagnostics were performed to determine the adequacy of the model and independence of the errors.The findings of the residual diagnostics are presented in Figure 5.

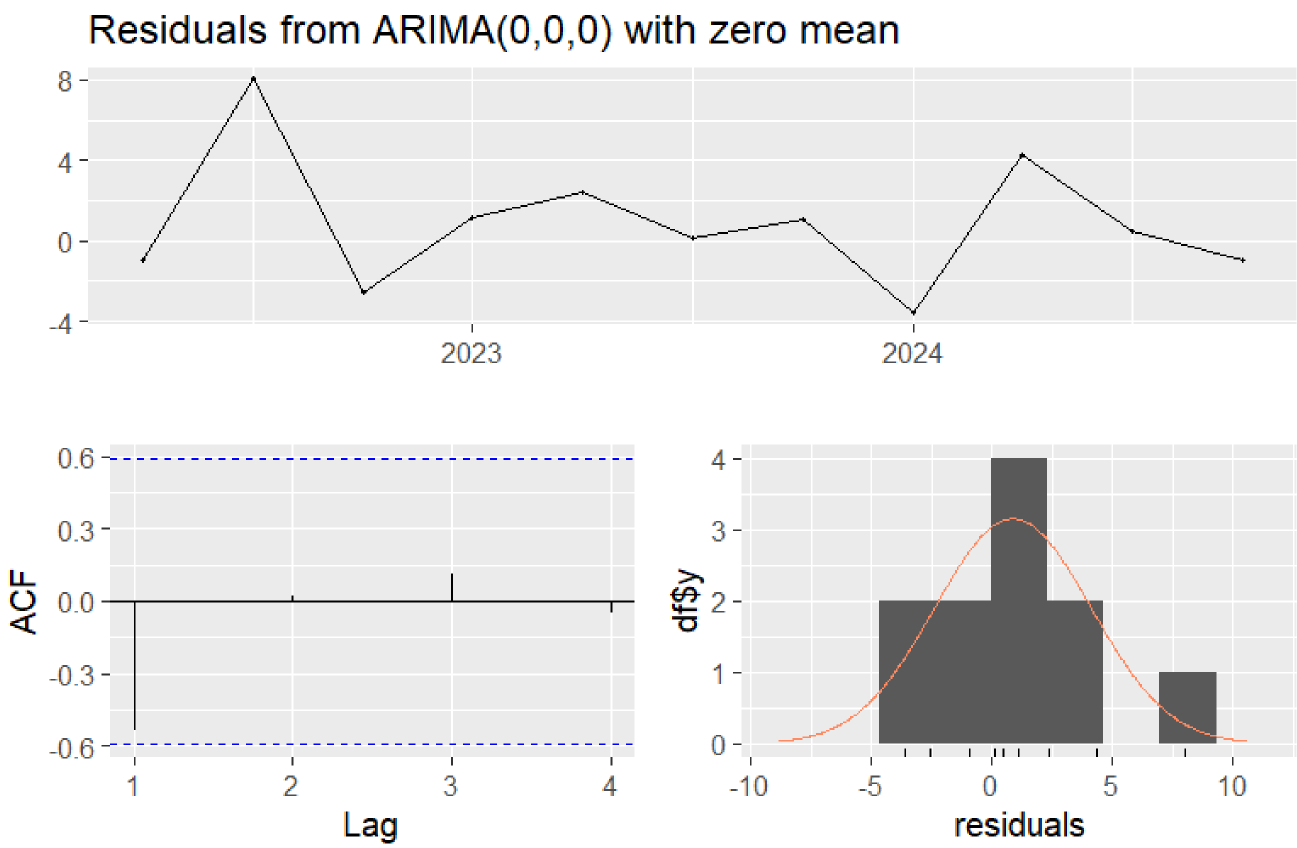

Figure 5: Residual diagnostics for the ARIMA(1,1,1) model.

As shown in Figure 5 above, it was evident that the values of the residual were symmetrically distributed about zero with no noticeable evidence of autocorrelation and therefore the ARIMA(1,1,1) model was an adequate representation of the underlying data generating process. The white noise Ljung Box test had a non-significant value ($p > 0.05$) ,which supports the adequacy of the model.

4.2.6 Residual Diagnostics from Regression Analysis

Residual analysis was performed to assess the randomness and distributional assumptions of the regression-based models.

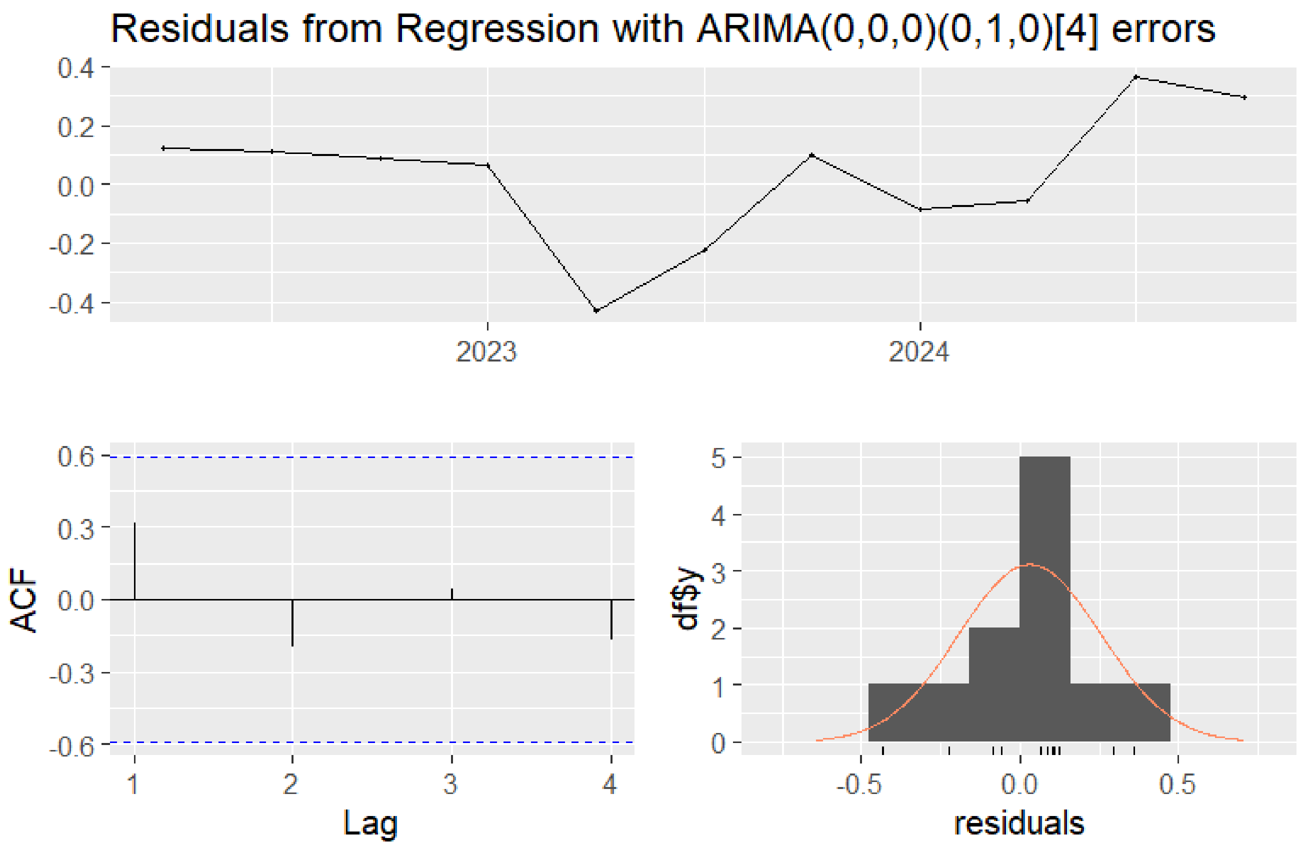

Figure 6: Residual distribution from regression-based forecasting model.

The statistical validity of the fitted models was supported by the fact that the residuals were approximately normally distributed and had homoscedasticity. There was no sign of any systematic bias which indicated that the determinants of customer churn which were important were well represented.

5 Conclusions

The study concluded that time series analysis was a powerful analysis framework for customer churn prediction in Kenya's banking sector. The results also confirmed that customer attrition had seasonal and trend driven dynamics in which customer attrition must be continuously monitored by data-driven forecasting systems.

6 Recommendations

Based on the empirical findings, we made the following recommendations:

i. Kenyan banks should have predictive analytics based models institutionalized within CRM architectures, to detect and preempt churn tendencies.

ii. Kenyan banks should match marketing and customer engagement initiatives with the cyclical churn patterns discovered via time series forecasts.

iii. Policymakers should promote inter-institutional data sharing to better calibrate models and provide more accurate industry-wide churn monitoring.

7 Disclaimer(Artificial Intelligence)

I hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc) and text-to-image generators have been used during writing or editing of manuscripts.

8 Competing Interests

I hereby declare that no competing interests exist.

References

[1] **Barasa, I. N., Wanyonyi, S. W., & Kololi, M. M. (2025).** Application of logistic regression in enhancing digital credit risk management in commercial banks. *Asian Journal of Probability and Statistics*, 27(2), 13–26. Science Domain International (SDI). https://doi.org/10.9734/ajpas/2025/v27i2710

[2] **Brito, J. B. G., Bucco, G. B., & Heldt, R. (2024).** A framework to improve churn prediction performance in retail banking. *Financial Innovation*, 10, Article 17. https://doi.org/10.1186/s40854-023-00558-3

[3] **Shailaja, R. and Kumar, P. (2024).** Service quality and customer retention in the banking industry. *International Journal of Banking Research*, 14(1), 33–48.

[4] Mena, C. G., De Caigny, A., Coussement, K., De Bock, K. W., & Lessmann, S. (2020). Churn prediction with sequential data and deep neural networks: A comparative analysis. *arXiv preprint* arXiv:1909.11114. `https://doi.org/10.48550/arXiv.1909.11114`

[5] Suri, T., & Jack, W. (2016). The long-run poverty and gender impacts of mobile money. *Science*, 354(6317), 1288–1292. https://doi.org/10.1126/science.aah5309

[6] **Zeithaml, V. A., Berry, L. L., & Parasuraman, A. (1996).** The behavioral consequences of service quality. *Journal of Marketing*, 60(2), 31–46. https://doi.org/10.2307/1251929