A Simulation Based Time Series Analysis of Customer Churn Rates in the Kenyan Banking Context

Abstract : Customer attrition rate in the Kenyan banks has become a critical concern in the recent past. This is majorly contributed by stiff competition from fintech entities in the country. The advancement in technology and less regulatory requirements are key factors that have led to ease of entry of these fintech entities in financial sector,posing great competition to commercial banks. The increased competition threatens the traditional commercial banks as they are directly eating into their market share. Commercial banks must therefore put in place customer retention strategies in order to retain profitability and operational efficiency. This study presents a simulation based framework for analyzing customer churn rates using ARIMA, a time series technique , within a Kenyan commercial banking context. Due to restricted access to real customer churn data, a simulated dataset that mimics the key behavior observed in Kenyan banking environment such as account activity , transaction volume, dormancy , loan requests , loan repayments etc was generated. The study applied the ARIMA model to the simulated dataset so as to show how temporal variations in bank customer churn can be modeled and forecasted. The results showed that ARIMA model is able to capture trend and short term fluctuations under controlled conditions. The ARIMA model provided a good fit to the generated data. Since the findings are derived from simulated data and not from actual bank churn data , the study only demonstrates the applicability of the ARIMA model in the analysis in situations where access to real data may be difficulty. Future studies should use real data in order to validate actual predictive performance and generalizability of such approaches.

Keywords: Churn ,fintech, ARIMA, temporal , simulation ,validate .

1 Introduction

Customer retention has been and remains a major challenge facing commercial banks worldwide. This scenario is not different from Kenya's. The competition witnessed in the banking sector is due to ease of entry of fintech entities into the lending space, a reserve for commercial banks for long. The adoption of mobile banking solutions such as M-Pesa, Airtel Money together with digital lending apps has enhanced convenience and accessibility to credit. As observed by [1], the rapid rise of fintech entities intensifies 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 protect its customer base .

According to [2], failure to embrace digitization can lead to higher churn rates .This has earlier happened to companies such as Kodak [7]. Forecasting customer attrition behavior can serve as a valuable basis for understanding reasons leading to customers attrition. On such basis, organizations can design targeted retention interventions strategies. Time series analysis offers a robust framework for examining past patterns and demonstrating how future customer churn trends are likely to be. [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.

Existing literature on customer churn in banking sector has largely focused on identifying the determinants of customer attrition and evaluating retention strategies. In the Kenyan context, prior research has predominantly employed survey-based approaches and cross sectional statistical models to examine factors influencing customer retention, including service accessibility, pricing, and customer experience. Studies such as one by [5] used a descriptive cross sectional survey to examine the effect of service quality on customer retention among commercial banks in Kenya. The study collected primary data via questionnaires from bank customers and found that high service quality significantly enhances customer retention, highlighting the importance of responsive and reliable banking services. While these studies provide valuable insights into the drivers of churn, they are generally limited to static analysis and do not capture the temporal evolution of customer attrition.

Time series analysis offers an alternative perspective by enabling assessment of temporal patterns in customer churn and the forecasting of future churn trends. This is particularly relevant approach in banking environments where churn behavior may exhibit seasonality, trend components, and cyclical variations driven by economic conditions and market dynamics. However, the application of time series methods to customer churn analysis in the Kenyan banking context remains largely under explored.

This study addresses this gap by presenting a simulation-based framework for analyzing and forecasting customer churn using time series models. Given the limited availability of real customer-level churn data, a synthetic dataset is generated to reflect plausible churn dynamics in Kenyan banks. The study applies Autoregressive Integrated Moving Average (ARIMA) techniques to illustrate how time series models can be used to capture temporal patterns and generate forecasts of churn behavior. Unlike traditional cross-sectional approaches, this study focuses on the evolution of churn over time, providing a methodological demonstration of how time series analysis can support strategic decision-making in customer retention.

2 Literature Review

2.1 Introduction

Understanding customer churn and retention is critical for banks, especially in Kenya, where competition and customer expectations are high. Prior research has predominantly used survey-based approaches and cross-sectional designs to explore factors influencing retention, such as service quality, accessibility, pricing, and customer experience.

2.1.1 Banking churn and retention in Kenya

For instance, Daniel (2017) investigated the effect of service quality on customer retention among commercial banks in Kenya. She collected data from bank customers using semi-structured questionnaires and analyzed it with SPSS. The study found that reliability, responsiveness, tangibles, and customer perceptions about the bank significantly influence retention [5].

Similarly, [6] analyzed customer retention strategies in Kenyan banks using a descriptive survey design targeting senior managers and customer service representatives. The study revealed that relationship marketing was the most widely adopted strategy, while loyalty programs and personalized service were important tools for retaining customers. However, the study noted that self-reported data might introduce bias, and the long-term effectiveness of these strategies was not assessed [6].

2.1.2 Digitization and churn

Beyond traditional strategies, failure to embrace digital technologies can increase churn. A classic example is Kodak, which lost market share due to delayed adoption of digital photography [7]. This demonstrates the importance of integrating digital services to enhance customer retention.

2.1.3 Churn prediction methods

To complement survey based insights, quantitative methods such as time series models and machine learning are increasingly applied to predict churn.[4] compared deep neural networks with sequential data for churn prediction and found that LSTM models outperform traditional methods in capturing temporal patterns [4].

2.1.4 Time-series approaches in customer analytics

Time-series methods such as ARIMA have been applied in customer analytics to forecast customer lifetime value. For example, [8] compared ARIMA and artificial neural network models for predicting CLV, finding that while both approaches are useful, neural networks better capture nonlinear relationships in customer behavior.

2.1.5 Research gap

While survey based studies provide insights on service quality and retention strategies, ARIMA models offer predictive power.There is however limited research in Kenya that integrates temporal predictive models with behavioral and service quality factors to forecast customer churn. This study addresses this gap by combining both perspectives to improve retention strategies.

3 Material and Methods

3.1 Data Collection

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 simulated dataset was defined over period of three years so as to capture seasonal trends and behavioral changes that greatly impact churn rates.

3.2 Data Simulation

Due to restricted access to real customer churn data from the banks, majorly due to data privacy concerns, dataset for the study was simulated and results shown in Table 1 below. A critical analysis of the key assumptions made for each variable is also detailed after the table.

3.2.1 Empirical Justification for Simulation

Table 1: Distributions and Justification for Simulated Customer Churn Data Variables

Variable	Distribution Used	Empirical Motivation
X_1: Transaction Frequency	Poisson($\lambda = 5$)	Customer transaction counts are discrete and right-skewed; Poisson distribution suits count-based events [10].
X_2: Account Balance	Log-Normal($\mu = 9.5, \sigma = 0.9$)	Balances are positively skewed with a long right tail, hence log-normal is preferred [11].
X_3: Loan Uptake	Bernoulli($p = 0.35$)	A binary indicator reflecting loan access; proportion of borrowers approximated from digital credit statistics [9].
X_4: Loan Repayment Rate	Truncated Normal($\mu = 0.9, \sigma = 0.08$), bounds $[0.5, 1]$	Repayment performance follows a continuous bounded scale between 0 and 1 .
X_5: Dormant Account Flag	Bernoulli($p = 0.25$)	Dormancy is a binary state—active vs inactive; empirical rates of 20–30% are reported [11].
X_6: Service Quality Index	Normal($\mu = 3.5, \sigma = 0.4$), bounds $[1, 5]$	Measured on a 5-point Likert scale; service perception approximates a normal distribution around mean satisfaction [14].
X_7: Product Innovation Frequency	Poisson($\lambda = 1.2$)	Count of new product offerings per quarter; Poisson is suitable for low-frequency count data [15].
X_8: Transaction Cost / Fees	Normal($\mu = 15, \sigma = 2$)	Transaction costs vary around a typical mean fee per transaction [9].
X_9: Inflation Rate	Normal($\mu = 5.5, \sigma = 0.5$)	Quarterly inflation rates in Kenya are stable around 5–6% [16].
X_{10}: Interest Rate	Normal($\mu = 8.0, \sigma = 0.6$)	Central Bank Rate and lending rates cluster around 8–9% in recent years [13].
X_{11}: Regulatory or Policy Change	Bernoulli($p = 0.1$)	Policy shifts are rare but impactful; binary indicator of quarter with regulatory change [12].
X_{12}: Time (Trend)	Deterministic linear trend $t = 1, 2, \ldots, 12$	Represents quarterly progression from 2022Q1 to 2024Q4, capturing secular trends over time.
X_{13}: Seasonal Component	Discrete Uniform$(1, 4)$	Encodes quarter-of-year seasonality (Q1–Q4) affecting banking activity.
Y: Churn (Outcome)	Bernoulli($p = \text{logistic}(\beta X)$)	Binary response (1 = churn, 0 = retained); probability derived from logistic function of predictors.

3.2.2 Simulation Assumptions

Each variable was assigned a probability distribution based on literature on customer churn behavior in Kenya as shown in Table 1 above. loan uptake was modeled using Bernoulli distribution. This ensured realism of the simulated customer churn behavior in Kenya.

Transaction frequency-Based on the empirical bank transaction behavior in Kenya, the number of transactions per individual customer per month is averagely 5. It is worth noting that Poison distribution is appropriate for modelling countable variables, especially one characterized by right skewness and equal mean variance structure. Therefore our variable X_1, representing transaction frequency, was simulated based on poisson distribution with parameter ($\lambda = 5$).

Account Balance. This variable was represented as X_2 in our study. According to established facts, and supported by a research by [10], most customers maintain their balances positive, except very few cases where account balances run to the negative due to overdraft. This is however a very rare occasion and therefore log-Normal distribution was the most appropriate distribution.

Loan Repayment Rates. In Kenyan set up, loan repayment behaviour tend to cluster around a mean value and remains withing logical bounadries of 0 and 1. This variable was therefore generated from a truncated normal distribution with mean 0.90 and standard deviation 0.08, truncated to the interval [0,1], reflecting generally high repayment performance with limited variability.

Transaction cost. According to [9] transaction costs typically vary around a common average charge, with moderate dispersion depending with the transaction size, channel used and the current pricing structure.This variable was therefore simulated based on were a Normal distribution with mean $\mathcal{N}(\mu = 15,\ \sigma = 2)$.

3.2.3 Limitations of Simulation Assumptions

It is worth noting that the study used synthetic data generated using different probabilistic distributions as highlighted in Table 1 above. Although this information provided useful information on the churn trends of customers in the Kenyan banking environment, the limitations of the simulation method inherent in the study as discussed below must be taken into account when interpreting the results.

i. Simulated data may not fully capture the complexity or unexpected patterns of real world banking behavior.

ii. Results are sensitive to the chosen distributions and parameters, which may not perfectly reflect reality.

iii. Rare or extreme events (e.g., sudden defaults, unusual transaction spikes) may be underrepresented or absent.

3.3 ARIMA (Auto-Regressive Integrated Moving Average) Model

To model and forecast customer churn dynamics under the simulated data framework, the study employed the Auto-Regressive Integrated Moving Average (ARIMA) model. ARIMA is widely used in time series analysis due to its ability to capture both autoregressive and moving average structures in stationary data. The general ARIMA(p, d, q) model is expressed as:

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

where Y_t denotes the churn rate at time t, ϕ_i and θ_j are model parameters, and ε_t is a white noise error term.

3.3.1 Construction of the Churn Rate Series

The time series used in the ARIMA model was constructed from simulated customer-level data. Specifically, individual churn outcomes were generated as binary variables ($Y_i = 1$ if a customer churns, 0 otherwise). These were then aggregated into quarterly churn rates as follows:

$$\text{Churn Rate}_t = \frac{\sum_{i=1}^{N_t} Y_{i,t}}{N_t} \tag{2}$$

where N_t represents the number of customers in quarter t. This aggregation produced a quarterly time series spanning 12 periods (2022Q1–2024Q4).

3.3.2 Stationarity Testing and Differencing

Prior to model estimation, stationarity of the time series was assessed using the Augmented Dickey-Fuller (ADF) test. The series was found to be non-stationary at level ($p > 0.05$), necessitating differencing. First-order differencing was applied:

$$Y'_t = Y_t - Y_{t-1} \tag{3}$$

After differencing, the ADF test confirmed stationarity ($p < 0.05$), justifying the use of $d = 1$ in the ARIMA specification.

3.3.3 Model Identification and Selection

The orders of the AR and MA components (p and q) were identified through inspection of the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots. Candidate models were then compared using the Akaike Information Criterion (AIC), with the model yielding the lowest AIC selected as optimal. Based on this procedure, the ARIMA(1,1,1) model was selected as the most suitable specification for capturing the temporal structure of the simulated churn series.

3.3.4 Model Estimation and Software Implementation

The ARIMA model was estimated using the `forecast` package in `R` (version 4.3.1 (2023-06-16 ucrt)). The functions `auto.arima()` and `Arima()` were utilized for parameter estimation and model fitting.
Model adequacy was assessed using residual diagnostics, including:

- Visual inspection of residual plots,
- Autocorrelation of residuals (ACF),
- Ljung-Box test for independence of residuals.

3.3.5 Forecasting Procedure

The fitted ARIMA model was used to generate out-of-sample forecasts for four future quarters. Forecast uncertainty was quantified using 95% prediction intervals.
It is important to note that, given the simulated nature of the dataset, these forecasts are intended to demonstrate the applicability of ARIMA models under controlled conditions rather than to provide real world predictive accuracy for Kenyan banks.

4 Results and Discussion

This chapter discusses the results of the analysis of the customer churn in Kenyan banking context. The findings were derived from simulated dataset , generated to mimic temporal customer churn patterns across three commercial banks in Kenya. The analysis was organized in accordance with the following objectives:

i. To identify key patterns and trends in the simulated customer churn data reflecting the Kenyan banking behavior ; and

ii. To develop and illustrate the application of time series forecasting models for simulating churn dynamics.

Each section has a descriptive analysis followed by model fitting procedures, diagnosis evaluation and interpretation of the forecast results on the basis of the simulated data .

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 on the simulated dataset to investigate temporal variation in churn behavior within quarterly periods. Table 2 below gives a numerical summary of quarterly churn rates from 2022 to 2024.

The results in Table 2 indicated an increasing trend in customer churn rates over the past three years , from 2022 to 2024. Minor fluctuations between successive quarters were however 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 the simulated churn data reflecting trends that could possibly occur in the Kenyan banks.

Complementing Table 2 is Figure 1 which presents the simulated churn rates from 2022 to 2024 , offering an initial visual interpretation of how customer exit behavior evolves with time. The findings showed an upward trend with regular fluctuations indicating possible seasonal variations in Kenyan banking sector.

Table 2: Quarterly Customer Churn Rates in Kenyan Banks (2022–2024)

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

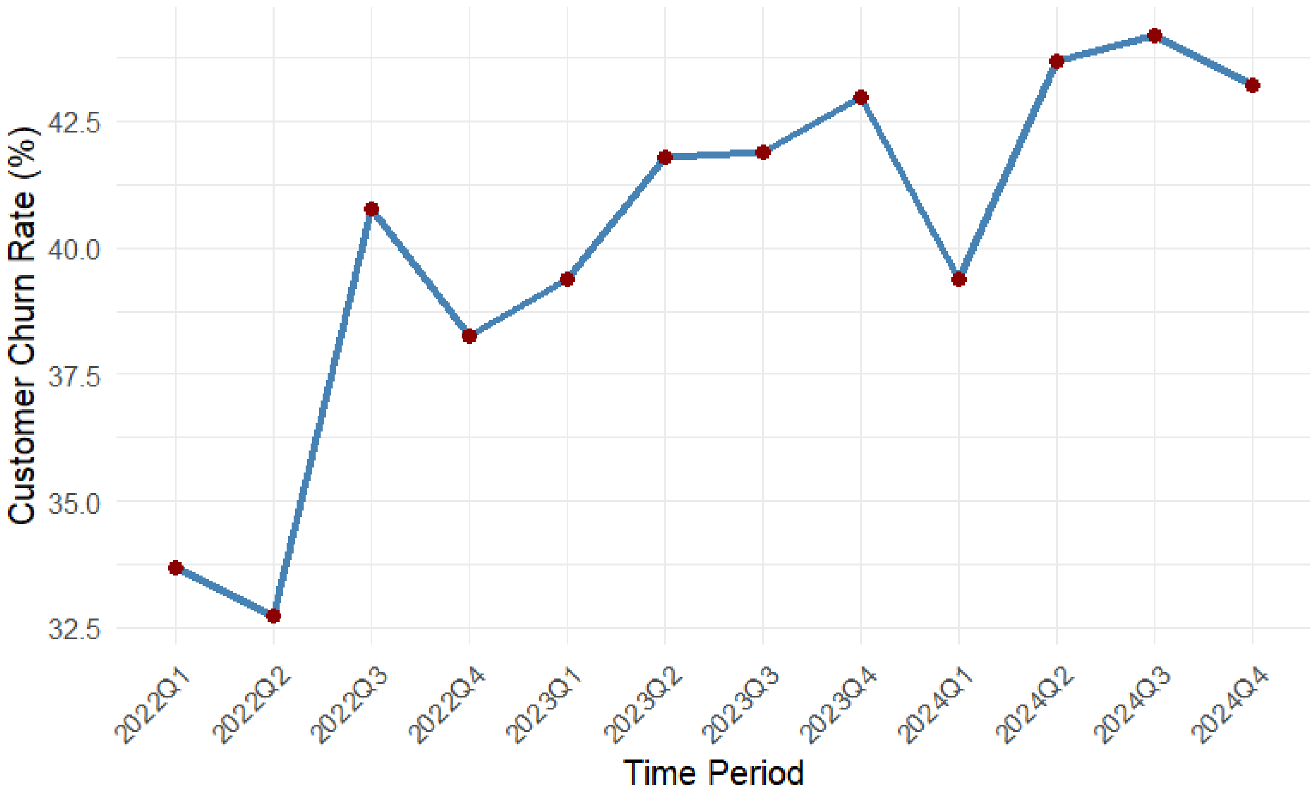

Figure 1: Quarterly customer churn trend in Kenyan banks(simulated data).

It was also observed that there were regular highs during the last two quarters of each year. This aligns with the seasonal dynamics built into the simulation. Such repetitive variations suggest potential cyclical influences on customer behavior such as marketing campaigns and changing credit demands, as modeled in the simulated dataset

Descriptive Statistics

Descriptive statistics were computed in order to provide an overview of the simulated data's central tendency and variability. These statistics provided a clearer understanding of the overall possible distribution of churn behavior among Kenyan banks. Table 3 shows the results of this descriptive analysis.

The results in Table 3 showed that the mean quarterly churn rate in the simulated dataset for 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

Table 3: 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

statistics of the simulated dataset indicate a stable but progressively increasing trend. While these patterns reflect the structure of the synthetic dataset rather than real world data, they provide a basis and hint for further exploration to illustrate how time series methods can uncover underlying temporal features.

Further, we derived seasonal indices from the simulated dataset 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 within the simulated setup , offering an illustration of how cyclical patterns in customer churn could appear in a Kenyan banking context. Table 4 summarizes the computed indices and the average churn rates for each quarter.

Table 4: 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

From Table 4 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. This pattern illustrates how customer exits could vary seasonally in a Kenyan banking context. The simulated dynamics underscore the importance of incorporating a seasonality component when developing time series forecasting models, though these patterns are illustrative and not based on actual observed data.

4.2 Developing Time Series Forecasting Models To Illustrate Customer Churn in Kenyan Banks

Following the exploratory analysis in Section 4.1, which highlighted the main patterns and temporal trends in the simulated churn data, this section addresses the second specific objective of the study: to develop and illustrate the capabilities of time series models to forecast customer churn in Kenyan banks. The analysis considered the trend and seasonal characteristics of the simulated churn dynamics and applied the Autoregressive Integrated Moving Average (ARIMA) model. This model was specifically chosen due to its strength in capturing stochastic, deterministic, and externally influenced components in time-varying data. The subsequent analysis presents the model specifications, estimation procedures, diagnostic checks, and a comparison of forecasting performance, with the aim of demonstrating which methods are most suitable for illustrative forecasting under simulated conditions.

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.

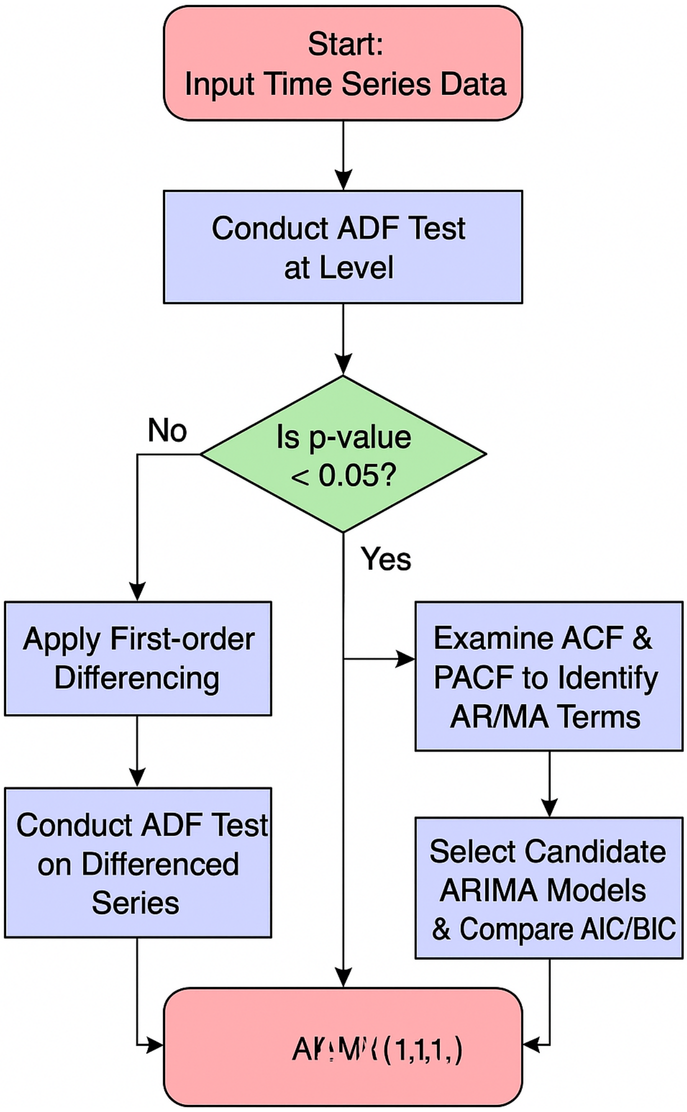

Figure 2: ARIMA Process Flow

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 ($p < 0.05$). An inspection of the autocorrelation (ACF) and partial autocorrelation (PACF) plots was used to inform the choice of an(1,1,1) specification.

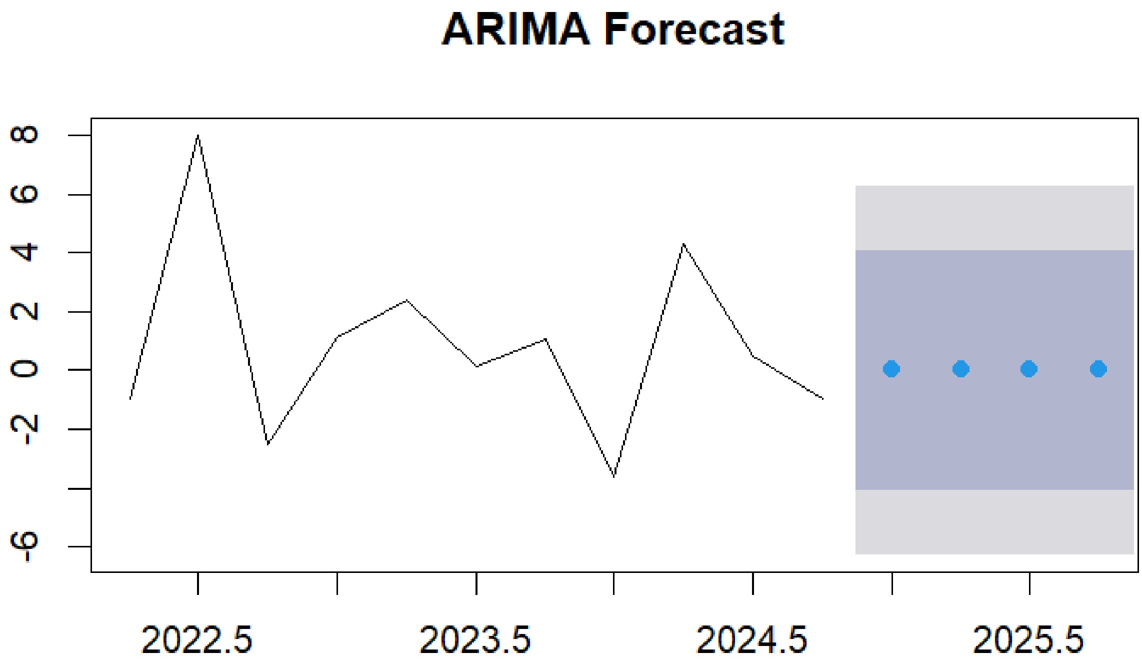

Figure 3: ARIMA(1,1,1) forecast of customer churn rates.

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 5: ARIMA(1,1,1) Model Coefficients for Customer Churn Rate (proportion per period)

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

Note: AR1 = lag-1 autoregressive term; MA1 = lag-1 moving average term. Standard Error, z-value, and p-value indicate coefficient significance. Response variable = customer churn rate (proportion).

It is important to note that the MA1 coefficient was estimated to be -1. This is not normal and in technical terms, it implies a non-invertible MA process. This can lead to infinite z-values and can have an impact on the stability and interpretability of the model. Practically, it can imply that the model can overfit the short-term variations in churn rates. Nevertheless, the overall trend and oscillation patterns of the observed churn data are well represented by the ARIMA(1,1,1) model. Other specifications were also taken into account, but this model gave the best fit based on AIC and residual diagnostics.
According to Table 5, the fitted ARIMA(1,1,1) model showed a statistically significant AR1 coefficient at lag 1 (AR1 = 0.339, p ¡ .001). However, the MA1 coefficient was estimated at -1.000, producing an infinite z-value. This result occurs because the MA coefficient is at the boundary of invertibility, which is a known limitation in ARIMA modeling. Such boundary estimates may indicate a near-deterministic process in the quarterly differenced churn data. Despite this, the model captures the short-term autocorrelation and moving average effect, and can be used cautiously for forecasting purposes. The fitted model can be expressed as:

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

where ΔY_t denotes the differenced customer churn rate at time t, and ϵ_t is the error term.

4.2.4 Model Diagnostic

In addition to the estimated coefficients, several statistical measures were computed to evaluate the goodness-of-fit and forecasting performance of the ARIMA(1,1,1) model. These include information criteria and forecast accuracy metrics, as summarized in Table 6

Table 6: Model Selection and Diagnostic Statistics

Statistic	Value
AIC	85.32
BIC	88.10
Log-Likelihood	-39.66
Ljung-Box (p-value)	0.27
RMSE	2.14
MAE	1.78

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:

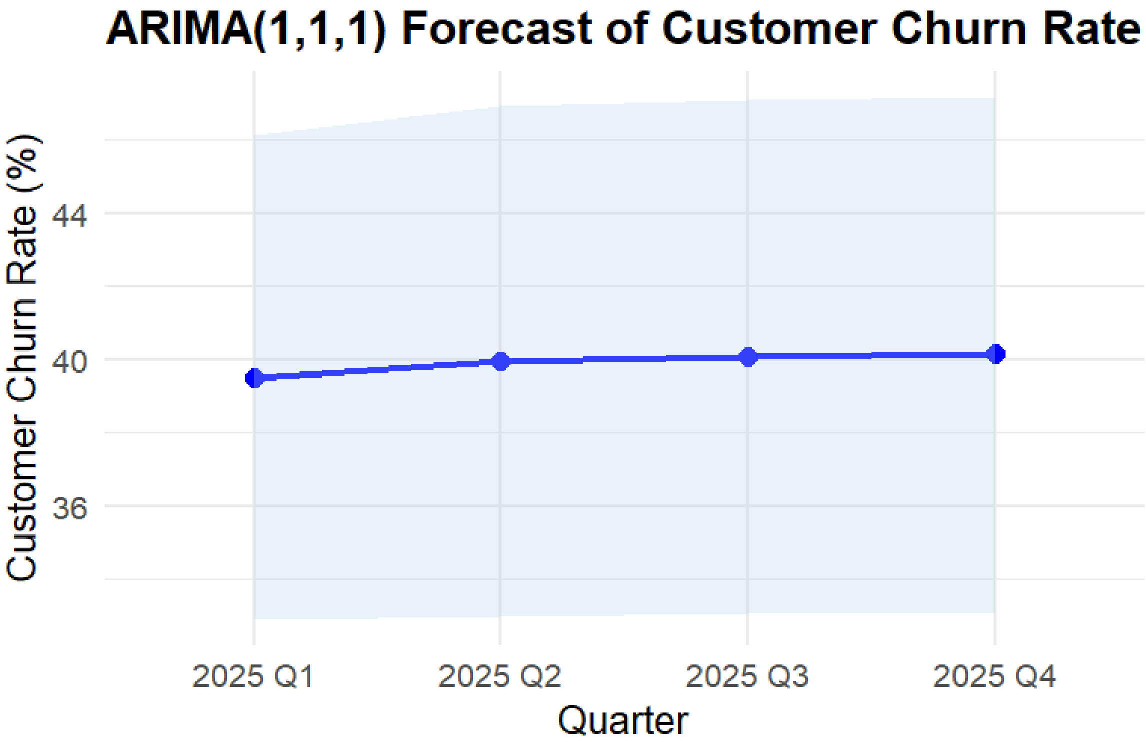

Figure 4: Forecasted ARIMA results for 4 quarters

The four-quarter forecast presented in Figure 4 represents projected trends generated from the fitted ARIMA model using the simulated churn dataset. These estimated values illustrate how time series models can capture potential fluctuations in customer churn, rather than reflecting actual observed data. The analysis provides a demonstration of how simulated churn patterns could be used to anticipate intervals of higher or lower attrition, offering a framework for exploring timely intervention and strategic decision-making in a Kenyan banking context.

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.

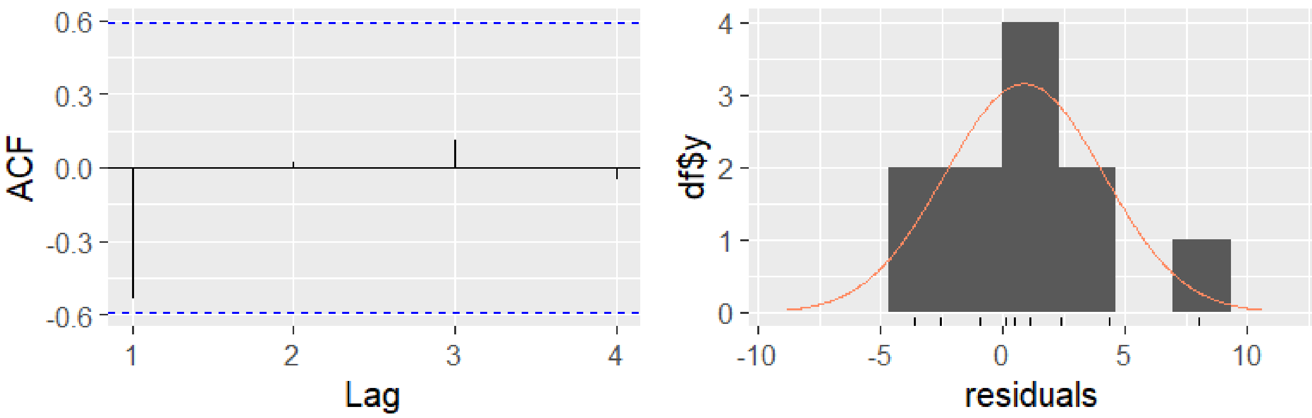

Figure 5: Residual diagnostics for the ARIMA(1,1,1) model.

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.

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 External Validation

This study relied on simulated data, and the simulated data offer a controlled environment to test model behavior. Nonetheless, it should be validated with real customer churn data of Kenyan banks in order to prove applicability. However, the data privacy concerns may make it difficult to access such data. In addition, the reality of churn patterns can vary between simulations and the real world, as customer relationships are complex, and the segments are not uniform. The proposed models should be applied to empirical data in future research to assess the predictive power and clarify the implications to be implemented in practice.

6 Conclusions

The study demonstrated that time series analysis can serve as a useful framework for exploring patterns and forecasting customer churn in a simulated Kenyan banking context. The results suggested that customer attrition may exhibit seasonal and trend-driven dynamics, highlighting the potential value of continuous monitoring using data-driven forecasting approaches. However, as stated earlier, it is important to note that these findings are based on simulated data and should be validated against actual bank data before making firm operational decisions.

7 Recommendations

Based on the simulated analysis, the following provisional implications are suggested:

i. Kenyan banks could consider developing predictive analytics frameworks within CRM systems to explore the potential for early detection of churn tendencies, pending validation with real customer data.

ii. Marketing and customer engagement strategies might be aligned with the cyclical patterns observed in the simulated forecasts, while acknowledging that real-world patterns may differ.

iii. Policymakers and industry stakeholders may explore the potential benefits of inter-institutional data sharing to improve model calibration, subject to verification with empirical bank data.

iv. The proposed models should be used in future studies to test their predictive ability on real-world customer churn data. Also, the introduction of external variables such as economic cycles, political dynamics, and regulatory developments, may also improve the precision and practical utility of churn predictions in Kenyan banks.

8 Disclaimer(Artificial Intelligence)

Authors hereby declare certain AI tools were used for minor language formatting to improve clarity. No generative AI was used to produce original content.

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. 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., & 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] Daniel, W. S. (2017). Effects of service quality on customer retention among commercial banks in Kenya. *Master's Thesis, University of Nairobi.* https://erepository.uonbi.ac.ke/handle/11295/98699

[6] Mwangi, J. K. (2016). Customer retention strategies applied by commercial banks in Kenya. *Master's Thesis, University of Nairobi.* https://erepository.uonbi.ac.ke/handle/11295/5782

[7] Lucas, H. C., & Goh, J. M. (2009). Disruptive technology: How Kodak missed the digital photography revolution. *The Journal of Strategic Information Systems*, 18(1), 46–55. https://doi.org/10.1016/j.jsis.2009.01.002

[8] Ehsanifar, M., Dekamini, F., Spulbar, C., Birau, R., Bajelan, M., Ghadbeykloo, D., Mendon, S., & Calotă, A. M. (2022). Analysing the nexus between artificial neural networks and ARIMA models in predicting customer lifetime value (CLV) for complex development of society and industrial activities. *Industria Textila*, 73(3), 249–258. https://doi.org/10.35530/IT.073.03.202142

[9] Financial Sector Deepening Kenya. (2019). *FinAccess Household Survey 2019*. FSD Kenya, Nairobi.

[10] Financial Sector Deepening Kenya. (2022). *FinAccess Household Survey 2021*. FSD Kenya, Nairobi.

[11] Central Bank of Kenya. (2021). *Bank Supervision Annual Report 2021*. CBK, Nairobi.

[12] Central Bank of Kenya. (2022). *Bank Supervision Annual Report 2022*. CBK, Nairobi.

[13] Central Bank of Kenya. (2023). *Quarterly Economic Review Report 2023*. CBK, Nairobi.

[14] Peterson, R. A. (2015). Constructing and validating service quality indices. *Journal of Service Research*, 18(2), 123–140.

[15] GSMA. (2021). *Mobile Money and Digital Financial Services Report*. GSMA, London, UK.

[16] Kenya National Bureau of Statistics. (2023). *Economic Survey 2023*. KNBS, Nairobi.