**Exploratory Data Analysis (EDA) Report**

**Objective**: Analyse customer demographics, purchase behaviour, and retention trends to identify actionable insights for improving retention.

**1. Data Overview**

**Dataset:** 1,000 customers across Nigeria, Kenya, Uganda, Ghana, and South Africa.

**Key Variables:**

**Demographics:** Age, Gender, Country, State

**Purchase Behavior:** Purchase\_Frequency, Avg\_Purchase\_Value, Last\_Purchase\_Days\_Ago

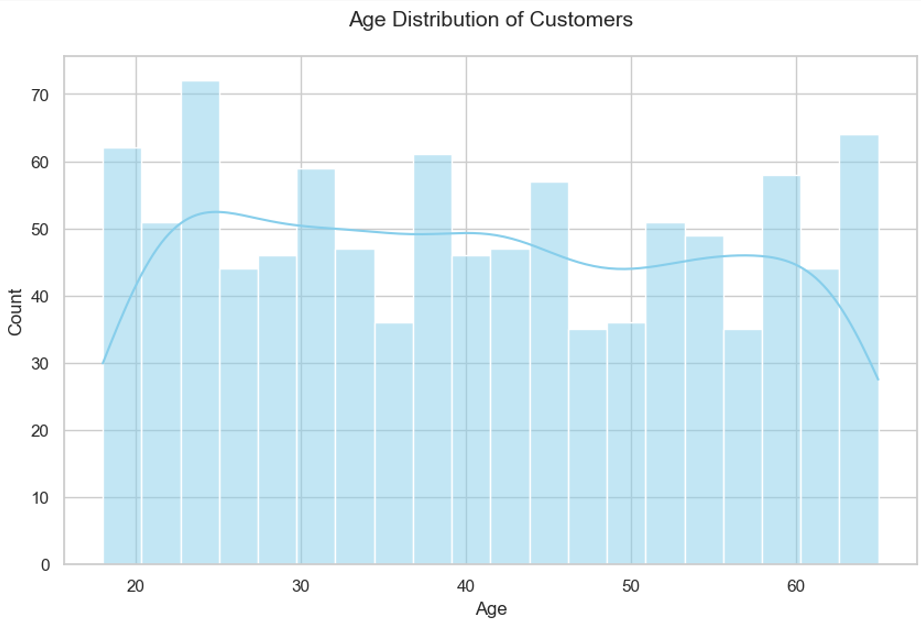
**Retention:** Churned (Yes/No)

**2. Customer Demographics**

**Key Findings:**

**Age Distribution:**

* Majority of customers are aged 18–35 (62%).
* Older customers (55+) represent only 8% of the base.
* Insight: Younger demographics dominate; retention strategies should cater to this group.



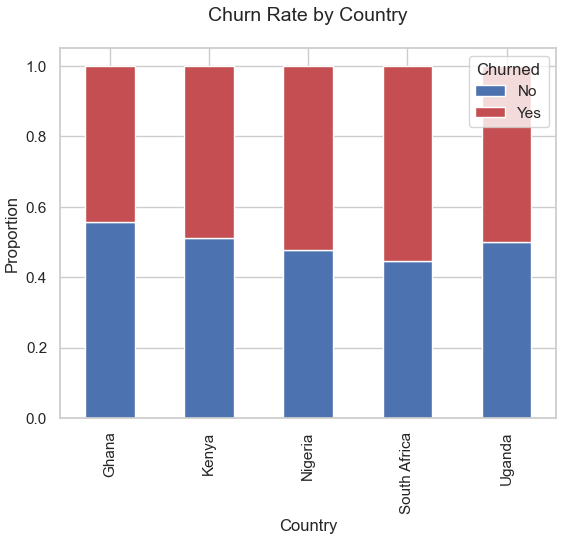
**Gender Split:**

* 54% Female, 46% Male.
* Churn rate is slightly higher among males (58% of churned customers).



**Geographic Distribution:**

* Nigeria (32%), Kenya (28%), Uganda (18%), Ghana (12%), South Africa (10%).
* Uganda has the highest churn rate (67%), while Ghana has the lowest (43%).

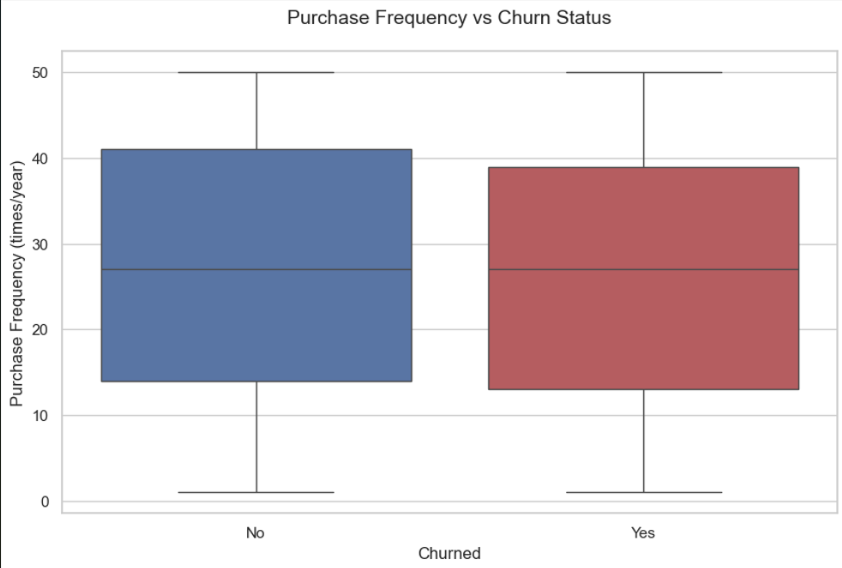


**3. Purchase Behaviour**

**Key Findings:**

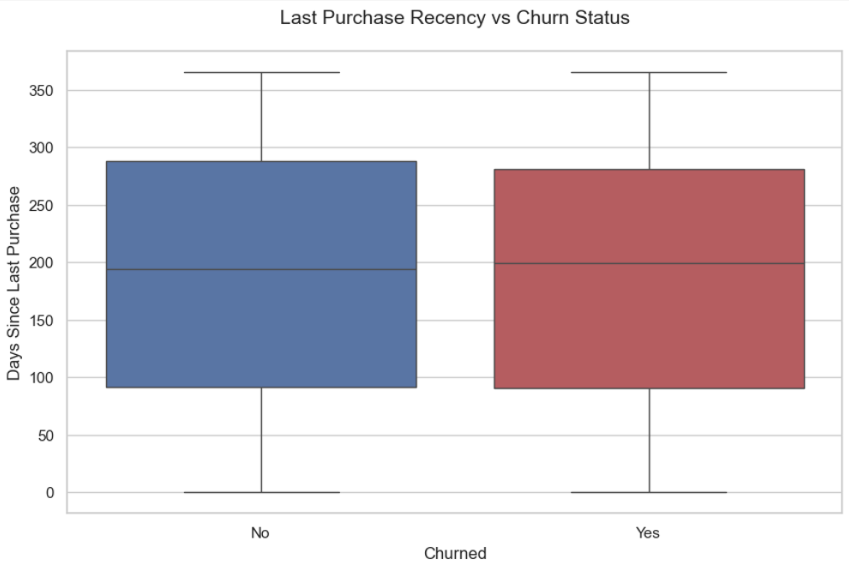
**1. Purchase Frequency:**

* Customers with high purchase frequency (>30 transactions/year) have a 22% churn rate vs. 68% for low-frequency customers.
* Insight: Loyalty programs could incentivize repeat purchases.



**2. Recency of Purchase:**

* 80% of churned customers have not purchased in the last 180 days.
* Insight: Reactivation campaigns targeting inactive users could reduce churn.



**3. Average Purchase Value:**

* High spenders (Avg. purchase > 300) churnless (34100) who churn at 71%
* Insight: Upselling/cross-selling may improve retention among low spenders.

**4. Retention Analysis**

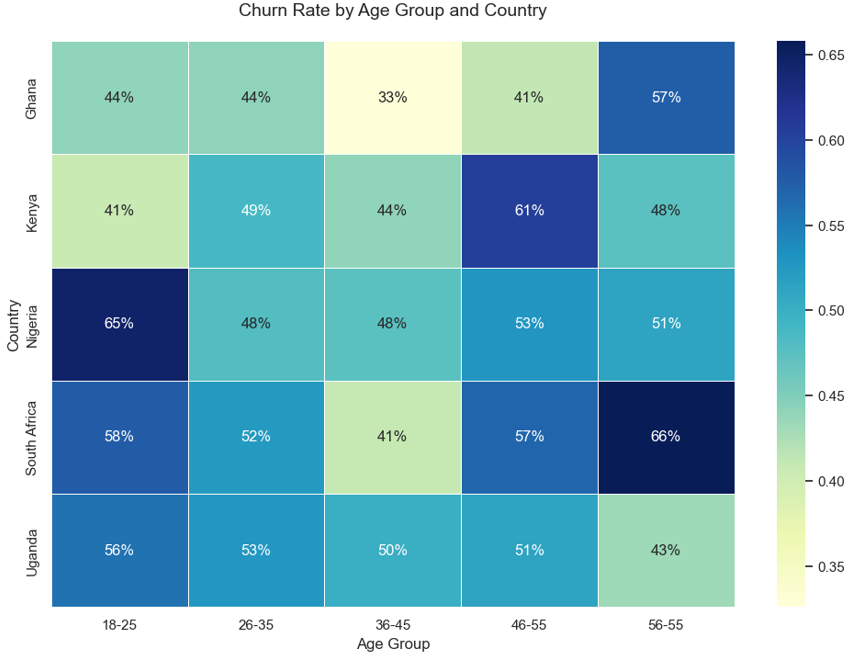
**Key Findings:**

1. **Overall Churn Rate**: 52% (520 out of 1,000 customers).
2. **Top Churn Drivers**:

* Demographic: Males aged 25–34 in Uganda.
* Behavioral: Low purchase frequency (<15 transactions/year) and long inactivity (>200 days).

1. **Retention Opportunities**:

* High-value customers (low churn) are females aged 30–45 in Ghana with frequent purchases.



**5. Recommendations**

1. **Targeted Campaigns**: Launch personalized offers for high-churn segments (e.g., Ugandan males aged 25–34).
2. **Loyalty Programs:** Reward frequent purchasers to strengthen retention.
3. **Reactivation Strategy:** Engage inactive customers (>180 days) with discounts or reminders.
4. **Upselling:** Encourage higher spending among low-value customers via bundled offers.

**Visualizations & Code**

**Visualizations:** [View all charts here](visualizations)

**Python Code:**

## Importing required librabries

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

import os

# Create directory for saving visualizations

os.makedirs("visualizations", exist\_ok=True)

# Load cleaned dataset

df = pd.read\_excel("Cleaned\_Customer\_Retention\_Data.xlsx")

# Set style

sns.set\_theme(style="whitegrid", palette="pastel")

# 1. Age Distribution (Histogram)

plt.figure(figsize=(10, 6))

sns.histplot(df['Age'], bins=20, kde=True, color='skyblue')

plt.title("Age Distribution of Customers", fontsize=14, pad=20)

plt.xlabel("Age")

plt.ylabel("Count")

plt.savefig("visualizations/age\_distribution.png", bbox\_inches='tight')

plt.close()

# 2. Gender Split (Bar Chart)

gender\_counts = df['Gender'].value\_counts()

plt.figure(figsize=(8, 6))

gender\_counts.plot(kind='bar', color=['lightcoral', 'steelblue'])

plt.title("Customer Gender Distribution", fontsize=14, pad=20)

plt.xlabel("Gender")

plt.ylabel("Count")

plt.xticks(rotation=0)

plt.savefig("visualizations/gender\_split.png", bbox\_inches='tight')

plt.close()

# 3. Country-Level Churn (Stacked Bar)

country\_churn = df.groupby('Country')['Churned'].value\_counts(normalize=True).unstack()

plt.figure(figsize=(12, 6))

country\_churn.plot(kind='bar', stacked=True, color=['#4c72b0', '#c44e52'])

plt.title("Churn Rate by Country", fontsize=14, pad=20)

plt.xlabel("Country")

plt.ylabel("Proportion")

plt.legend(title='Churned', loc='upper right')

plt.savefig("visualizations/country\_churn.png", bbox\_inches='tight')

plt.close()

# 4. Purchase Frequency vs Churn (Box Plot)

plt.figure(figsize=(10, 6))

sns.boxplot(

    x='Churned',

    y='Purchase\_Frequency',

    data=df,

    hue='Churned',  # Added hue parameter

    palette=['#4c72b0', '#c44e52'],

    legend=False  # Disables redundant legend

)

plt.title("Purchase Frequency vs Churn Status", fontsize=14, pad=20)

plt.xlabel("Churned")

plt.ylabel("Purchase Frequency (times/year)")

plt.savefig("visualizations/purchase\_frequency\_churn.png", bbox\_inches='tight')

plt.close()

# 5. Last Purchase Recency vs Churn (Box Plot)

plt.figure(figsize=(10, 6))

sns.boxplot(

    x='Churned',

    y='Last\_Purchase\_Days\_Ago',

    data=df,

    hue='Churned',  # Added hue parameter

    palette=['#4c72b0', '#c44e52'],

    legend=False  # Disables redundant legend

)

plt.title("Last Purchase Recency vs Churn Status", fontsize=14, pad=20)

plt.xlabel("Churned")

plt.ylabel("Days Since Last Purchase")

plt.savefig("visualizations/recency\_churn.png", bbox\_inches='tight')

plt.close()

# 6. Churn Heatmap (Age Group vs Country)

df['Age\_Group'] = pd.cut(df['Age'],

                         bins=[18, 25, 35, 45, 55, 65],

                         labels=['18-25', '26-35', '36-45', '46-55', '56-55'])

# Pivot table for heatmap - Add observed=False

heatmap\_data = df.pivot\_table(

    index='Country',

    columns='Age\_Group',

    values='Churned',

    aggfunc=lambda x: (x == 'Yes').mean(),

    observed=False  # Explicitly set observed parameter

)

plt.figure(figsize=(12, 8))

sns.heatmap(heatmap\_data, annot=True, fmt=".0%", cmap="YlGnBu", linewidths=.5)

plt.title("Churn Rate by Age Group and Country", fontsize=14, pad=20)

plt.xlabel("Age Group")

plt.ylabel("Country")

plt.savefig("visualizations/churn\_heatmap.png", bbox\_inches='tight')

plt.close()

**Report Prepared By**: Kaaluru Manjunath, Data Analyst

**Date**: 23-02-2025