Bank Churn Analysis

By: Reagan Odhiambo Otieno
Date: 10/22/2025

Table of Contents

Introduction

Customer retention is a crucial aspect of long-term business success, especially in the competitive banking sector where customer acquisition costs are high. This analysis aims to understand the key drivers behind customer churn, the rate at which customers discontinue their banking relationship. By integrating Python, Excel, and Power BI, this project systematically merges, cleans, and analyzes customer data to uncover patterns and factors influencing churn. The insights generated are visualized through an interactive Power BI dashboard, enabling data-driven decision-making to enhance customer loyalty and retention.

The primary objectives of this project were to consolidate data from multiple sources, perform feature engineering for deeper insights, and visualize key performance indicators such as churn rate, retention rate, and customer distribution across demographic and behavioral segments. The analysis not only identifies who is leaving but also provides actionable recommendations on how the bank can proactively retain its customers.

1. Data Merging (Python)

Objective: Combine multiple sheets from the Excel dataset into one master file.

Steps:

1. Imported necessary Python libraries (pandas).
2. Read all Excel sheets and merged them using the common key CustomerId.
3. Ensured all rows matched correctly by using CustomerId as the merge key.
4. Exported the clean merged file for further processing in Excel.

```
import pandas as pd
import re

# Path to your Excel file
file_path = "Bank_Churn_Messy.xlsx"

# Read all sheets
sheets = pd.read_excel(file_path, sheet_name=None)

# Function to find the Customer ID column in each sheet
def find_customer_id_column(columns):
    for col in columns:
        clean_col = col.strip().lower().replace(" ", "").replace("_", "")
        if re.search(r'cust|customer.*id|id', clean_col):
            return col
    return None

# Standardize the key column name in all sheets
aligned_dfs = []
for name, df in sheets.items():
    id_col = find_customer_id_column(df.columns)
    if id_col:

        df = df.rename(columns={id_col: "Customer_ID"})
        aligned_dfs.append(df)
    else:
        print(f"⚠ No Customer ID column found in sheet: {name}")

# Merge all sheets using Customer_ID
merged_df = aligned_dfs[0]
for df in aligned_dfs[1:]:
    merged_df = pd.merge(merged_df, df, on="Customer_ID", how="outer")

# Save the merged result
merged_df.to_excel("Bank_Churn_Merged_By_CustomerID.xlsx", index=False)

print("✅ All sheets merged successfully by 'Customer_ID' into 'Bank_Churn_Merged_By_CustomerID.xlsx'")
```

Output: Bank_Churn_Merged_By_CustomerID.xlsx

2. Data Cleaning (in Excel)

Objective: Prepare the dataset for analysis by correcting inconsistencies and ensuring accuracy.

Steps:
- Removed duplicates based on CustomerId.
- Handled missing values appropriately.
- Standardized categorical data formats.
- Converted Exited column: 1 = Yes, 0 = No.

```
=IF(B2=1, "Yes", "No")
```

- Changed data types for numeric and text columns.

Output: Clean, analysis-ready Excel dataset.

3. Feature Engineering (in Excel)

Objective: Create new derived columns for better analysis and segmentation.

Steps:
- Created Age Groups using IF formulas.

```
=IF(C2<=25,"Young Adults", IF(C2<=35,"Early Career", IF(C2<=45,"Mid Career",
IF(C2<=55,"Experienced", IF(C2<=65,"Pre-Retirement","Retired")))))
```

- Created Tenure Groups using IF formulas.

```
=IF(D2<=2,"New Customer", IF(D2<=5,"Established", IF(D2<=8,"Loyal","Very Loyal")))
```

- Verified logic and consistency.

Output: Added Age Group and Tenure Group columns.

4. Importing and Modeling in Power BI

Objective: Load the cleaned dataset and create analytical measures.

Steps:
- Imported cleaned Excel file into Power BI.
- Verified data types in Model view.
- Created DAX measures including Total Customers, Churned Customers, Churn Rate (%), Retention Rate (%), and Average Balance.
- Used DIVIDE() function for accurate percentage calculations.

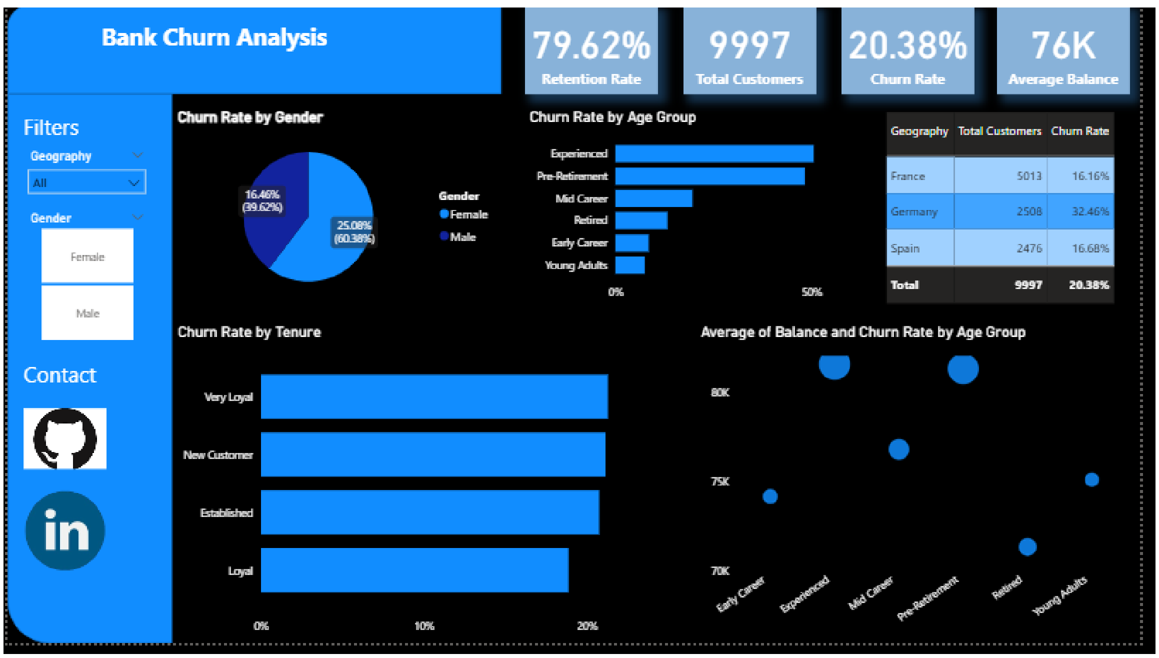
```
Churn Rate =

DIVIDE(

   CALCULATE(COUNTROWS(Bank_Churn), Bank_Churn[Exited] = 1),
```

```
   COUNTROWS(Bank_Churn)
)
```

5. Dashboard Creation in Power BI

Objective: Design an interactive dashboard to visualize customer churn insights.

Steps:
- Created KPI cards for Retention Rate, Total Customers, Churn Rate, and Average Balance.
- Designed charts: Pie (Gender), Bar (Tenure, Age Group), Table (Geography), Bubble (Balance vs Age Group).
- Added filters for Geography and Gender.
- Included LinkedIn and GitHub contact icons.
- Applied consistent formatting and alignment.



Link for live dashboard: https://app.powerbi.com/reportEmbed?reportId=bc11e71a-0e91-490f-b104-32a5d7bac019&autoAuth=true&ctid=c5f7004a-3295-4205-9179-1b0ca576040c

Output: Interactive Power BI Dashboard – Bank Churn Analysis.

6. Insights & Recommendations

Key Insights:
- 20% churn rate indicates moderate attrition.
- Females (25%) churn more than males (16%).
- Germany has the highest churn (32%).

- Experienced and pre-retirement customers show higher churn.
- Low-balance customers are more likely to leave.

Recommendations:
- Launch loyalty and retention programs.
- Improve onboarding for new customers.
- Personalize services for female and older clients.
- Target low-balance customers with engagement offers.
- Track churn trends continuously.

7. Final Deliverables

Output Summary:
- Merged Excel File: Combined all sheets by Customer ID.
- Cleaned Dataset: Standardized and feature-engineered.
- Power BI Dashboard: Interactive visuals with key churn insights.
- Executive Summary: Highlights findings and recommendations.

Conclusion

The Bank Customer Churn Analysis successfully identified significant patterns that influence customer attrition. Findings revealed that churn is higher among female customers, clients from Germany, and those nearing retirement age or maintaining lower account balances. These insights highlight the need for personalized retention strategies, such as loyalty programs and improved engagement for specific customer groups.

Through the integration of Python for data merging, Excel for cleaning and feature engineering, and Power BI for modeling and visualization, the analysis provided a clear and actionable understanding of customer churn dynamics. Implementing the recommended strategies will help the bank strengthen customer relationships, reduce churn, and ultimately improve long-term profitability.