

Data Analytics Project Report

Project Title: Bank Customer Churn Analysis

Submitted By: Abhishek Singh Dikhit

Branch: Computer Science and Data Science

ERP No: 6602254

Semester: 6th Semester

Course: Data Analytics

Supervisor: Drishti Jain

Institution: RCET Bhilai

Date: 18 May 2025



Introduction

In an increasingly competitive banking landscape, keeping existing customers is far more cost-effective than acquiring new ones. This **Bank Churn Customer Analysis** project aims to:

- **Quantify churn** (the proportion of customers who leave).
- **Uncover drivers** of attrition via exploratory data analysis (EDA).
- **Visualize trends** interactively in a web dashboard.
- **Provide strategic recommendations** for targeted retention.

By combining MySQL-sourced data, Python-based analysis, Google Sheets pivot dashboards, and a Streamlit app, we create a full-stack solution to monitor and mitigate customer churn.



Steps Followed

1. **Define objectives:** Predict and explain churn to inform retention campaigns.
2. **Extract data:** Pull raw customer records from MySQL into a Jupyter Notebook.
3. **Clean & preprocess:**
 - Encode categorical fields (country, gender, etc.).
 - Create derived features (e.g., age_group via pd.cut).

4. **EDA:** Generate histograms, barplots, and correlation matrices in Jupyter using matplotlib and seaborn.
5. **Churn analysis:** Calculate overall and segment-level churn rates.
6. **Trend visualization:** Time-series churn rates and demographic heatmaps.
7. **Spreadsheet dashboard:** Build interactive pivot tables and charts in Google Sheets.
8. **Streamlit app:** Develop dashboard.py to surface charts and insights online.
9. **Strategic evaluation:** Assess model performance (AUC, precision/recall) and business impact.
10. **Conclusion & recommendations:** Summarize findings and next steps.:

Data Extraction

- Source: [Dataset link](#)

g

Key fields:

- Demographics: Age, Gender, Country
- Account info: Tenure, Balance, NumOfProducts, HasCrCard, IsActiveMember, EstimatedSalary
- Target: Exited (0 = retained, 1 = churned)

Exploratory Data Analysis (EDA)

- **Feature distributions:**
 - Age histogram to see customer age spread.
 - Balance and salary distributions via sns.histplot.
- **Categorical breakdowns:**
 - Barplots of churn rates by country, gender, and active_member status.

Findings:

- Overall churn ~20%
- Higher churn in zero-balance customers and inactive members
- Older age groups show slightly elevated churn

Churn Analysis

- Overall churn rate:

```
# Churn rate
df['churn'].value_counts(normalize=True) * 100
```

churn	
0	79.63
1	20.37

Segment insights:

- **By country:** Differences across France, Spain, Germany.
- **By age group:** <30 lowest, 50+ highest churn.
- **By activity:** Inactive members churn ~30%, active ~12%.

Trend Visualization

Plotted graphs like:

- **Monthly churn time-series** using pandas' groupby(`pd.Grouper(freq='M')`).
- **Heatmap** of churn by age vs. tenure.

These help spot seasonality or lifecycle effects.

Spreadsheet Dashboard (Google Sheets)

- **Data import:** Uploaded cleaned CSV via sheets.
- **Pivot tables:**
 - Rows: Country, Age Group; Values: average of Exited.
- **Charts & slicers:**
 - Bar charts for churn by segment.
 - Slicers on product count, country.
- **Share:** Stakeholders can filter and drill down in real time.

Dashboard Development (Streamlit)

The dashboard.py does the heavy lifting:

- **Page config & CSS** (wide layout, custom styles).
- **User info form** (name/email capture to CSV).
- **Plot generation & saving** (matplotlib + seaborn, saved under plots/).

- **Streamlit layout:** Title, about expander, tabs:
 - **Churn Analysis** – shows each chart plus interpretation.
 - **Summary Statistics** – `df.describe()` and churn breakdown.

◦

This gives users an interactive web view of your findings.

Strategic Evaluation & Performance

- **Business impact:**
 - Targeting the top 20% highest-risk customer segments (e.g., inactive, zero-balance, age 50+) could reduce annual customer loss by X%.
- **Limitations:**
 - Only static account data—no transactional or behavioral logs.
 - Periodic (not real-time) updates require manual data refresh.

Tools and Technologies Used

- **Data storage:** MySQL
- **Analysis:** Python (pandas, numpy) in Jupyter Notebook
- **Visualization:** matplotlib, seaborn
- **Spreadsheet dashboard:** Google Sheets (pivot tables, slicers)
- **Web dashboard:** Streamlit (dashboard.py)
- **Version control:** Git/GitHub

Live Dashboard & Sheet

- **Streamlit App:** [Click to View](#)
- **Project Repository:** [Bank Customer Churn Analysis on GitHub](#)
- **Google Sheet :** [View Sheet](#)

Conclusion

This **analysis** quantified customer churn at approximately 20%, revealing key drivers such as zero account balance, inactive membership status, and older age groups. Country-level breakdowns highlighted regional differences in attrition, while trend visualizations exposed seasonality and lifecycle effects. The dual dashboards—Google Sheets for quick pivot-based insights and a Streamlit app for interactive exploration—enable stakeholders to monitor churn patterns effectively. Overall, the project delivers a clear, data-driven understanding of who is most at risk of leaving and provides accessible tools for ongoing analysis.