

Bank Customer Churn Analysis

&

BI Pipeline

Amirhossein Tonekaboni

SAP Business One Consultant
BI Developer



Bank Customer Churn Analysis

End-to-end data engineering project analyzing 10,000 customer accounts at a European bank to identify churn patterns and risk factors. This project implements a complete **data pipeline** including data cleaning, exploratory analysis, database integration, and interactive visualization through containerized BI tools.



Excel: Data Source

- 10,000 records
- Maven Analytics

Python: EDA

- Explore
- Clean
- Analyze
- Visualize

Database: ETL

- Store data
- Secure connection

Deploy BI tool

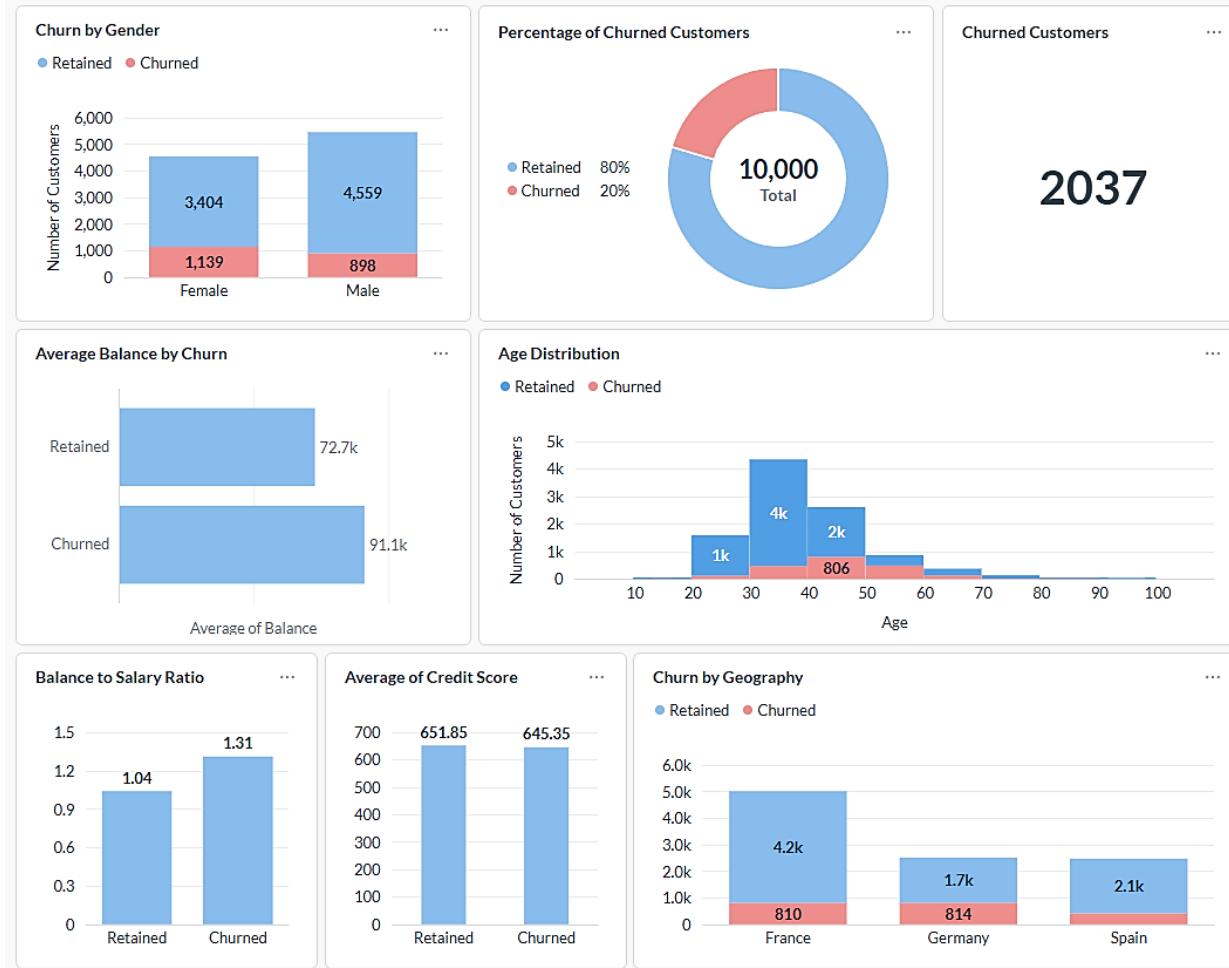
- Container setup
- Port mapping

Interactive Dashboard

- Visualize KPIs
- Filter data
- Business insights

Project Overview

Project Objective: Build a complete **data pipeline** from raw Excel files to an interactive **BI dashboard**, covering data cleaning, database integration, and visualization.



Source: [Bank Customer Dataset](#) from Maven Analytics (Chris Bruehl)

Data Cleaning & Exploratory Analysis

Data Preparation

- Merged related data tables using pandas
- Removed duplicate records
- Standardized data types and currency formats
- Handled missing values with appropriate strategies (categorical labeling, median imputation)
- Identified and treated outliers through statistical analysis
- Normalized categorical label variations

	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	EstimatedSalary	Balance	NumOfProducts	HasCrCard	IsActiveMember	Exited
0	15634602	Hargrave	619	France	Female	42.0	2	101348.88	0.00	1	Yes	Yes	1
2	15647311	Hill	608	Spain	Female	41.0	1	112542.58	83807.86	1	Yes	Yes	0
3	15619304	Onio	502	France	Female	42.0	8	113931.57	159660.80	3	No	No	1
4	15701354	Boni	699	France	Female	39.0	1	93826.63	0.00	2	No	No	0
5	15737888	Mitchell	850	Spain	Female	43.0	2	79084.10	125510.82	1	Yes	Yes	0

Exploratory Data Analysis

Visualization & Analysis

- Compared numeric features (Balance, Age, Credit Score) by churn status
- Distribution plots and outlier treatment
- Correlation analysis between features

Key exploratory findings

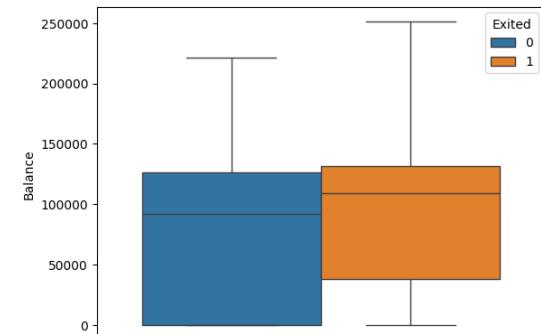
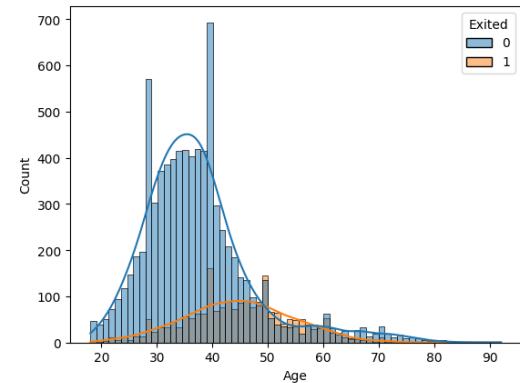
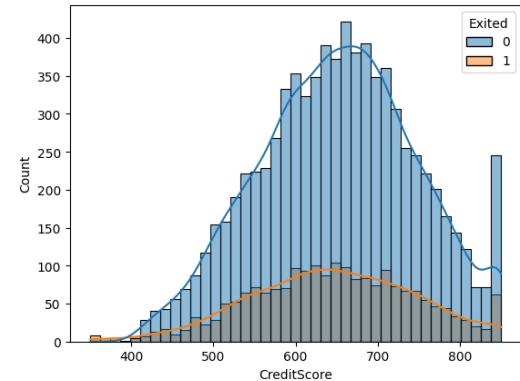
- Analyzed churn distribution across customer base
- Examined relationships between demographic features and churn behavior
- Compared financial metrics between churned and retained customers

Feature Engineering

- Removed identifier columns not suitable for modeling
- One-hot encoded categorical variables with `pd.get_dummies`
- Created derived feature: balance-to-income ratio for wealth assessment

Technologies

- Python, Pandas, NumPy, Matplotlib, Seaborn, Jupyter Notebook



Database Integration

SQLA

PostgreSQL Setup

- Created PostgreSQL database using DataGrip IDE
- Established secure connection using SQLAlchemy with environment variables
- Automated data transfer from pandas dataframe to database table

Technologies

- PostgreSQL, SQLAlchemy, DataGrip, psycopg2

```
from sqlalchemy import create_engine
from sqlalchemy.engine import URL
import os
```

```
target_db_name = "pgbank"
target_host = "localhost"
target_port = 5432
```

```
pg_user = os.getenv('PG_USER')
pg_password = os.getenv('PG_PASSWORD')

if pg_user and pg_password:
    print("User Pass checked.")
else:
    print("Check CMD")
```

```
connection_url = URL.create(
    "postgresql+psycopg2",
    username=pg_user,
    password=pg_password,
    host="localhost",
    database="pgbank",
    port=5432
)
```

```
engine = create_engine(connection_url)
with engine.connect():
    print("connected")
```

Business Intelligence Dashboard



Deployment

- Deployed Metabase as Docker container
- Connected Metabase to PostgreSQL database
- Built custom interactive dashboard with multiple visualizations

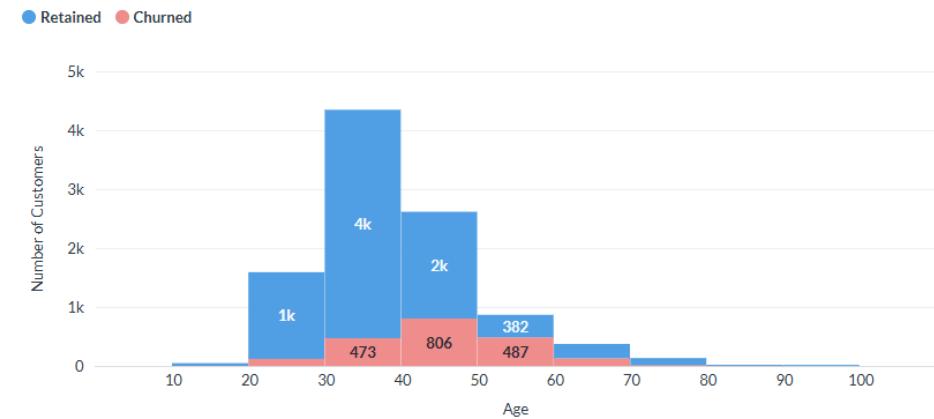
Key Metrics

- Total churned customers: 2,037 (20.4% churn rate)
- Churn distribution by demographics and geography
- Financial behavior comparison between customer segments

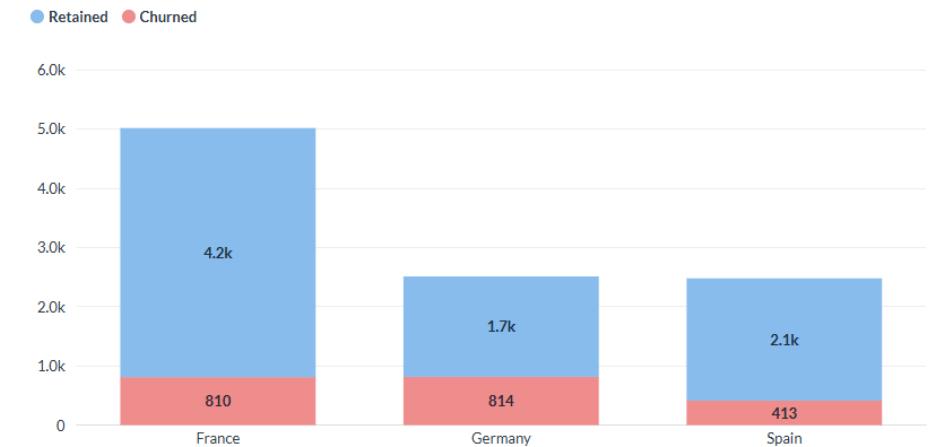
Technologies:

- Metabase, Docker

Age Distribution



Churn by Geography

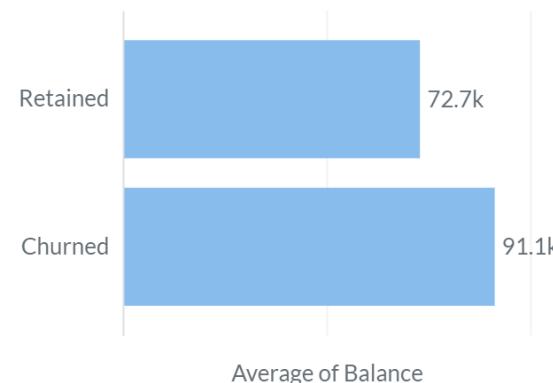


Key Insights

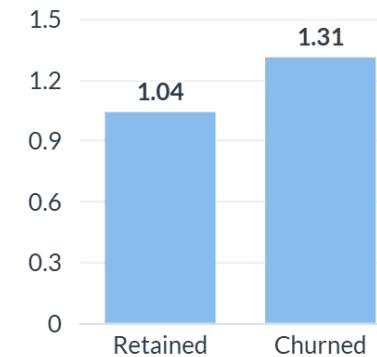
Key findings from the analysis:

- High-Value Customer Loss:** Churned customers maintained significantly higher average balances (€91.1k vs €72.7k), indicating the bank is losing its most valuable customers.
- Wealth Concentration Risk:** Customers who left had 26% higher balance-to-income ratios (1.31 vs 1.04), suggesting dissatisfaction among customers with larger deposits relative to their income.
- Overall Churn Rate:** 20.4% of customers (2,037 out of 10,000) have churned, representing significant revenue loss.
- Geographic Variation:** Germany leads with 32.4% churn rate, others around 16%.
- Age Factor:** Middle-aged customers (40-50 range) show distinct churn patterns compared to other age groups.
- Gender Factor:** Female customers showed significantly higher churn rates (25.1%) compared to males (16.5%), suggesting females are 1.5x more likely to leave.

Average Balance by Churn



Balance to Salary Ratio



Churned Customers

2037

Project Highlights

Complete Data Flow

- **ETL Pipeline:** Automated data extraction, transformation, and loading workflow
- **Database Integration:** Loaded cleaned data into PostgreSQL using secure connections (SQLAlchemy)
- **Containerization:** Docker deployment with isolated environment setup (Metabase)
- **Data Quality:** Statistical outlier detection and treatment (Pandas)
- **Business Intelligence:** Interactive dashboards with filtering capabilities
- **Feature Engineering:** Created analytical features (balance-to-income ratio)

Excel Data → Python / Pandas → PostgreSQL → Docker / Metabase → Interactive Dashboard

Future Enhancements

- Develop machine learning model for churn probability prediction
- Implement automated data refresh pipeline for real-time updates
- Add customer lifetime value (CLV) analysis
- Create customer segmentation using clustering algorithms
- Integrate additional data sources (transaction history, customer service interactions)

Technologies Stack

Category and Technologies

- **Languages:** Python, SQL
- **Libraries:** Pandas, NumPy, Matplotlib, Seaborn, SQLAlchemy, psycopg2
- **Database:** PostgreSQL
- **Tools:** Jupyter Notebook, DataGrip, Docker Desktop, Metabase

Links

- [GitHub Repository](#)
- [Maven Analytics Dataset](#)

Amirhossein Tonekaboni

SAP Business One Consultant
BI Developer

