Project Summary: Bank Customer Churn Prediction

Developed By: Shivansh Sharma

Objective:

To predict whether a customer will leave the bank (churn) using machine learning models, enabling the bank to take proactive retention measures.

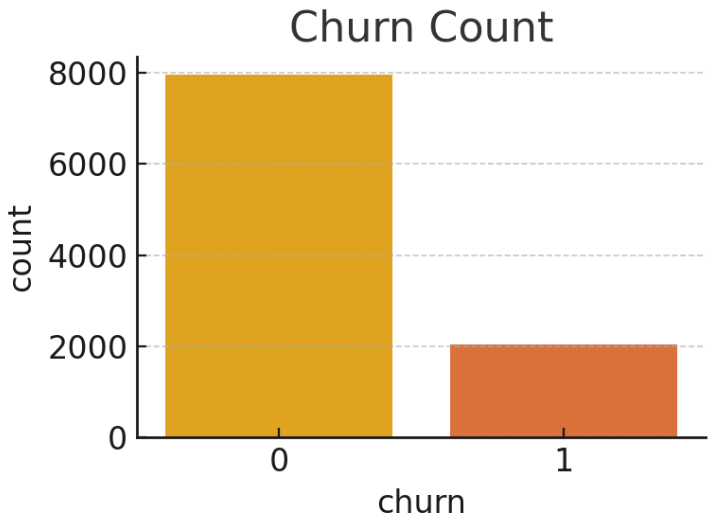
Dataset Overview

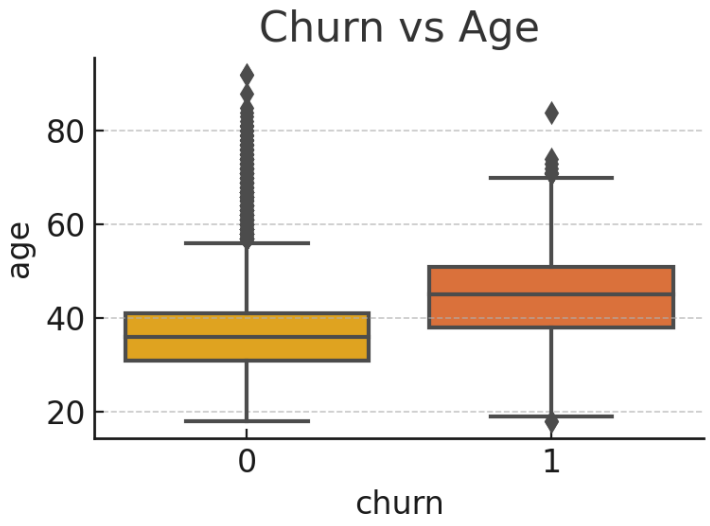
- Source: Bank_Customer_Churn_Prediction.csv

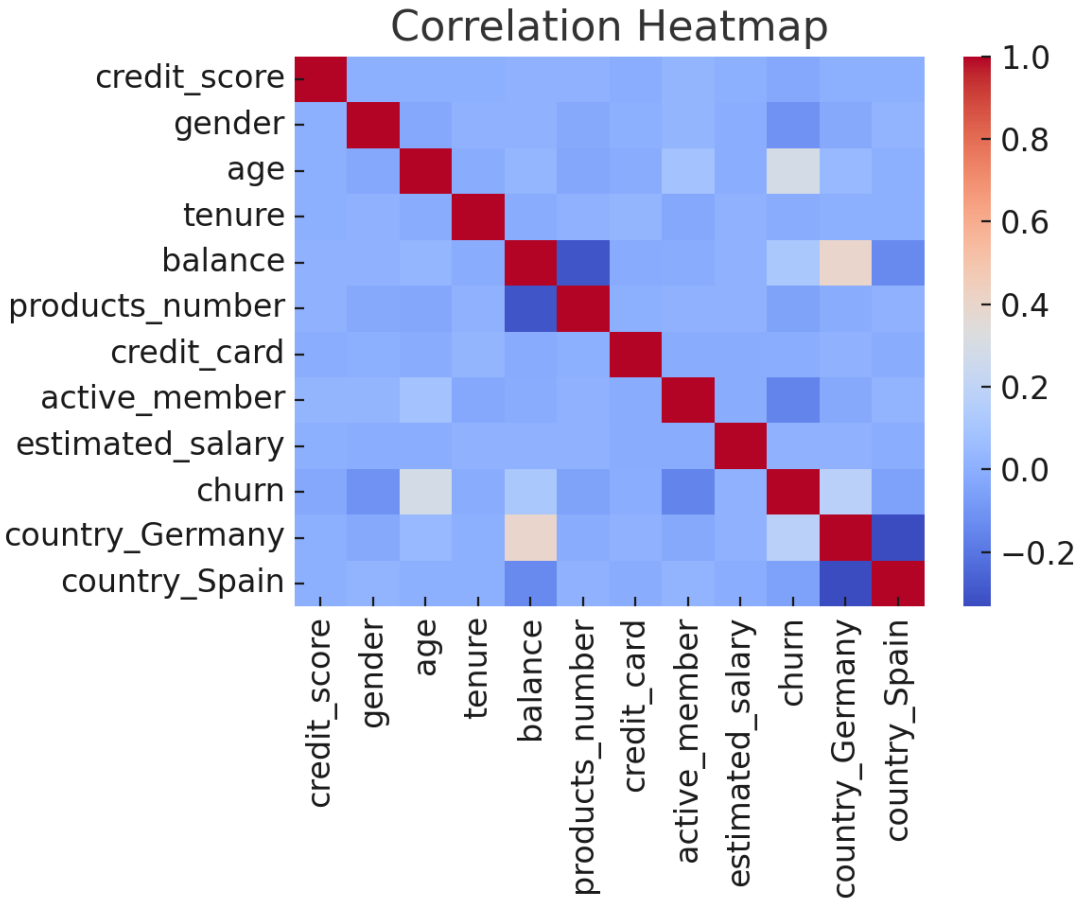
- Features: Age, Gender, Country, Balance, Products, etc.

- Target Variable: churn (1 = left bank, 0 = stayed)

Visualizations







Data Preprocessing

- Dropped column: customer_id

- Encoded gender (0: Female, 1: Male)

- One-hot encoded 'country'

- Feature scaling using StandardScaler

Model Building

Logistic Regression

- Accuracy: ~81%

- Good for interpretability

Random Forest Classifier

- Accuracy: ~86%

- Chosen for final predictions due to better performance

- ROC-AUC curve used for evaluation

Feature Importance

Top predictors from Random Forest model:

- balance, products_number, age, is_active, etc.

Key Takeaways

- Data-driven insights help banks prioritize high-risk customers.

- Ensemble models improve churn prediction accuracy.

- Feature importance aids in better understanding customer behavior.