Bank Churn Prediction and Analysis Project

# Project Description

In the banking industry, customer churn—when customers leave the bank and take their business elsewhere—is a significant challenge. Understanding the reasons behind customer churn can help businesses design targeted strategies to improve customer retention, optimize their services, and reduce the risk of losing high-value customers.   
   
 This project focuses on analyzing and modeling customer data to identify the factors that influence why customers leave (churn) or stay (non-churn) with the bank. By using predictive modeling techniques, we aim to classify customers who are likely to churn and investigate the key drivers behind their decision to leave or remain with the bank.  
   
 Through this analysis, we aim to not only predict which customers are likely to leave, but also gain insights into the underlying patterns that contribute to churn and retention. This knowledge will guide the bank in making data-driven decisions to enhance customer satisfaction, design more personalized services, and ultimately increase customer retention.

# Possible Impact of Your Analysis

- Improved Customer Retention: By identifying the most influential factors contributing to customer churn, the bank can take proactive measures to retain valuable customers, such as offering personalized services, improving engagement, or tailoring retention campaigns.  
   
 - Targeted Marketing Campaigns: The analysis can help segment customers based on their likelihood of churning, allowing the bank to focus its marketing efforts on the customers most at risk of leaving.  
   
 - Enhanced Customer Experience: Understanding churn patterns based on factors like age, tenure, and product ownership can help the bank optimize its services to cater to specific customer needs, thus improving overall satisfaction.  
   
 - Increased Profitability: Reducing churn leads to higher customer lifetime value, and retaining customers is often more cost-effective than acquiring new ones. This analysis will provide insights into maximizing customer retention, ultimately boosting profitability.  
   
 By exploring the patterns that distinguish churned customers from non-churned ones, this analysis will assist the bank in making data-driven decisions for better service delivery and customer relationship management.

# Dataset(s)

The dataset for this project is available on Kaggle under the title Bank Churn Prediction. You can access the dataset here:   
 [Bank Churn Prediction Dataset](https://www.kaggle.com/datasets/bonginstates/bank-churn-prediction/data)  
   
 Dataset Description:  
 The dataset contains synthetic data about customers and their banking behaviors, including demographic information, account details, and churn status. It includes the following columns:  
   
 - CustomerId: Unique identifier for each customer.  
 - Surname: Last name of the customer.  
 - CreditScore: A measure of the customer's credit history.  
 - Geography: The country or region of the customer.  
 - Gender: The gender of the customer.  
 - Age: Age of the customer.  
 - Tenure: Number of years the customer has been with the bank.  
 - NumOfProducts: The number of products the customer holds with the bank.  
 - Balance: Account balance of the customer.  
 - HasCrCard: Whether the customer has a credit card.  
 - EstimatedSalary: The estimated salary of the customer.  
 - isActiveMember: Whether the customer is actively using the bank's services.  
 - Exited: Whether the customer has churned (Exited = 1) or stayed (Exited = 0).  
   
 This dataset will be used to analyze and model customer churn and non-churn behavior, focusing on understanding which features most strongly correlate with a customer's decision to leave or stay with the bank.

# Business Problem:

Customer churn—the phenomenon of customers leaving the bank for competitors—is a significant challenge for financial institutions. Churn can result in lost revenue, increased customer acquisition costs, and an overall decline in customer loyalty. Understanding the reasons behind why customers leave (churn) and why they stay (non-churn) is essential for developing effective strategies to reduce churn rates and increase customer retention.

This project focuses on identifying the factors contributing to customer churn and using predictive modeling techniques to classify customers at risk of leaving the bank. By analyzing customer demographics, account details, and behaviors, we aim to gain valuable insights that will guide the bank in developing targeted retention strategies and improving customer satisfaction.

# Business Impact:

The analysis of customer churn and predicate can provide valuable insights that help the bank retain its customers. By identifying key factors influencing churn, the bank can implement targeted strategies to improve customer satisfaction and engagement. This could result in a lower churn rate, increased customer loyalty, and enhanced overall profitability. Additionally, understanding why customers leave can help the bank better allocate resources and optimize its service offerings.

# Dataset(s):

Bank Churn Prediction Dataset - Kaggle

## Strengths:

The dataset contains synthetic data on customer characteristics, including demographics, account details, subscription types, and churn status. This data can be used for predictive modeling and exploratory data analysis to identify the factors contributing to churn.

## Weaknesses:

# The dataset only includes customers from a specific region or demographic, which may not be fully representative of the entire customer base. Additionally, the dataset lacks temporal data such as account activity over time, making it difficult to assess long-term trends and seasonal patterns in customer churn

# Methods:

• Identify characteristics of variables (python)  
• Clean data and handle missing values(python)

• Finding statistics and information about the data (python)

• Create visualizations and identify relationships between variables (python)

• Create the predicative model (python)  
• Create a dashboard to summarize findings (power bi) Creating a Datafolio

• Creating a Datafolio  
• Write a final report summarizing the analysis and recommendations.

# Dashboard:

The dashboard will display key insights into customer churn, including the most significant factors influencing churn, overall churn rate, and trends based on customer attributes. It will also provide comparisons between customers who stayed and those who left.

# Milestones:

# Transform the dataset into a Pandas DataFrame

# Use Pandas functions to discover the DataFrames rows, columns, and values

# Identify null values and deal with them (either drop the entire column, fill it with a pre-defined value, or fill it with the mean/median/mode)

# Check data types and convert any if needed

# Get statistics

# Create Visualizations in Python

# Create a Dashboard

# Design a Portfolio

# Document all the steps done in a detailed final report

# Timeline:

|  |  |
| --- | --- |
| Week | Tasks |
| Week 1 | Data Sourcing & Data Curation |
| Week 2 | Exploratory Data Analysis and modleing |
| Week 3 | Dashboard & Datafolio & Final Report |