Documentation for Bank Customer Churn Prediction Project

➢ Overview

- This project focuses on predicting customer churn in a banking environment using machine learning. The goal is to identify customers likely to leave the bank, enabling proactive measures to retain them.

➢ Dataset Description

- **Target Variable**:
  - Exited: Indicates whether the customer has churned (1) or stayed (0).
- **Features**:
  - RowNumber: Row number in the dataset.
  - CustomerId: Unique identifier for each customer.
  - Surname: Customer's last name.
  - CreditScore: Customer's credit score.
  - Geography: Country of residence.
  - Gender: Gender of the customer.
  - Age: Customer's age.
  - Tenure: Number of years the customer has been with the bank.
  - Balance: Customer's account balance.
  - NumOfProducts: Number of products used by the customer.
  - HasCrCard: Whether the customer owns a credit card (1: Yes, 0: No).
  - IsActiveMember: Indicates if the customer is an active member (1: Yes, 0: No).
  - EstimatedSalary: Estimated annual salary of the customer.

➢ Libraries and Tools Used

- **Libraries**:
  - NumPy, Pandas: For data manipulation and analysis.
  - Matplotlib, Seaborn: For visualization.
  - Scikit-learn: For preprocessing and modeling.

➢ Steps in the Notebook

1. **Data Loading**:
   - The dataset is loaded into a Pandas DataFrame from a CSV file.
2. **Exploratory Data Analysis (EDA)**:
   - Includes visualizations and statistical summaries of key features.
3. **Preprocessing**:
   - Encoding categorical variables, handling missing values, and scaling numeric features.
4. **Model Building**:
   - Selection and training of machine learning models.
5. **Evaluation**:
   - Performance metrics to evaluate the predictive power of the model.

➢ Visualizations

- **Analysis of the relationship between the credit score (CreditScore) and the exited**
  - **Note**: customers with low credit scores (less than 600) leave with a higher percentage compared to those in the higher category (700+)
  - **Resolution**: focus on improving the experience of customers with medium and low credit scores, such as providing financial advice or incentive programs.
- **Analysis of the relationship between age and exited.**
  - **Note**: customers with a high balance tend to leave the bank more often than customers with a low balance
  - **Resolution**: improve investment offers or offer higher benefits to clients with high balances
- **Analysis of the relationship between the number of products (NumOfProducts) and exited**
  - **Note**: customers who own one product are most likely to leave.
  - **Resolution**: encourage customers to increase the use of products by offering incentives when using more than one product.
- **Analysis of the relationship between the activity (IsActiveMember) and exited**
  - **Note**: inactive clients leave at a higher percentage compared to active ones.
  - **Resolution**: implementation of interactive programs to motivate inactive customers to participate more with the bank.
- **Analysis of the relationship between estimated salary (EstimatedSalary) and exited**
  - **Note**: there is no obvious effect of the estimated salary level on departure, as the departure percentages are distributed almost equally at all salary levels.
  - **Resolution**: focus on other characteristics, such as activity and number of products, when developing customer retention strategies.
- **Analysis of the impact of Geography on Exited**
  - **Note**:
    - **In France:** most customers stayed with the bank, the percentage of departures is low compared to other regions.
    - **In Spain:** the percentage of customers leaving is average, but it is less pronounced compared to Germany.
    - **In Germany:** the percentage of departures is the highest compared to other regions, which indicates a possible problem with services or interaction with customers.
  - **Resolution**:
    - **In France:** maintaining the quality of services and the continuity of Special Offers to maintain customer loyalty.
    - **In Spain:** improving customer interaction through targeted support programs to prevent departures.
    - **In Germany:** improving the services and offers offered to customers to reduce departure rates, such as offering competitive advantages or loyalty programs.

- **The relationship between gender and the client's exit status**
  - **Note**:
    - The percentage of departures is higher among female customers compared to male ones.
    - Male clients show a higher retention rate with the bank.
  - **Resolution**:
    - Improve interaction with female clients by customizing services and programs that suit their needs.
    - Study the reasons that lead to females leaving more and work on addressing them through questionnaires or direct interaction.
    - Promote retention strategies for female clients by offering offers or incentives specifically targeted at them.

➢ Instructions for Running the Notebook

1. **Dependencies**:
   - Ensure Python 3.x is installed.
   - Install required libraries using pip install -r requirements.txt.
2. **Dataset**:
   - Place the Churn_Modelling.csv file in the same directory as the notebook.
3. **Execution**:
   - Run each cell sequentially to process the data, build the model, and visualize results.

➢ Key Insights

- The importance of age, credit score, and activity status in predicting churn.
- Customers with higher engagement tend to stay longer.