**Project Requirements**

**Improving Customer Retention: Analyzing Bank Churn Data**

**Problem Statement:**

1. Banks make continuous efforts to attract and retain customers, they face persistent challenge in the form of customer churn leading to financial losses.
2. It would be advantageous for the banks to understand why customers are leaving.

**Goal:**

1. Figure out why customers are leaving their bank. (Customer Churn Rate)
2. Train a Machine Learning models to find the most important reasons why customers leave.
3. Create a system that predicts which customers are likely to leave, so the bank can try to keep them.

**Steps:**

1. **Data Loading:** Load the provided dataset into a DataFrame for exploration and preprocessing.
2. **ETL Process (Extract, Transform, Load)**:
   1. Data cleaning.
   2. Handling missing values.
   3. Handling inconsistency in the data.
   4. Encode categorical variables.
   5. Scale numerical features.
3. **Exploratory Data Analysis (EDA):**
4. Analyze the distribution of customer churn.
5. Explore relationships between features and churn using visualization techniques.
6. Identify key trends in the data.
7. **Feature Engineering:**
8. Select the most relevant features for the model.
9. Optionally create new derived features based on existing ones.
10. **Model Selection:**
11. Train **Random Forest** models.
12. Compare their performance based on metrics like accuracy, precision, recall, F1-score, and AUC-ROC.
13. **Visualization:**
14. Use **Streamlit** to build a dashboard to showcase insights and model results.
15. Include visualizations like customer churn distribution, feature importance, and churn probabilities.
16. **Prediction and Actionable Insights:**
17. Provide predictions with churn probabilities.
18. Highlight actionable insights for bank decision-makers to retain customers.

# Project Solution Details

Data Source  
  
The project utilizes a dataset containing information about bank customers, including demographics, account details, and churn status. The dataset is loaded from a CSV file named "Bank\_Churn\_Modelling.csv".  
  
Data Preprocessing  
  
- Insignificant columns such as "RowNumber", "CustomerId", and "Surname" are dropped.  
- The target variable "Exited" is renamed to "Churned".  
- Feature engineering is performed to create new features from existing ones:  
    - "Total\_Products" is created from "NumOfProducts" to group customers based on the number of products they use.  
    - "Account\_Balance" is created from "Balance" to categorize customers with zero balance.  
- Categorical features ("Geography", "Gender", "Total\_Products", "Account\_Balance") are one-hot encoded.  
- The "Age" feature is log-transformed to improve model performance.  
- The data is split into training and testing sets using `train\_test\_split`.  
- Class imbalance in the target variable is addressed using SMOTE (Synthetic Minority Over-sampling Technique).  
  
Model Building  
  
- A Random Forest Classifier is used as the prediction model.  
- Hyperparameter tuning is performed using `GridSearchCV` to find the best model parameters.  
- The model is trained on the preprocessed and balanced training data.  
  
Model Evaluation  
  
- The model is evaluated using various metrics, including accuracy, precision, recall, and F1-score.  
- A confusion matrix is generated to visualize the model's performance.  
- Feature importance is analyzed to understand the factors influencing churn prediction.  
- An ROC curve is plotted to assess the model's ability to discriminate between churned and non-churned customers.  
- SHAP (SHapley Additive exPlanations) values are calculated to explain the model's predictions and understand the contribution of each feature.

Visualization with Streamlit  
  
- A Streamlit app is built to provide an interactive interface for visualizing the model's results.  
- The app allows users to select different metrics (Confusion Matrix, Feature Importance, ROC Curve) from a sidebar.  
- The selected visualization is displayed in the main content area of the app.  
  
Conclusion  
  
The Bank Customer Churn Prediction project successfully builds a machine learning model to predict customer churn. The model achieves high accuracy and provides insights into the key factors influencing churn. The Streamlit app enables interactive visualization of the model'sresults, making it easy to understand and interpret the predictions.