

## Summary and Recommendation:- Churn Prediction Analysis

The **Churn Prediction Analysis** project is a comprehensive machine learning solution designed to identify customers who are likely to leave a bank. Using a real-world dataset from a European bank, which includes data on 10,000 customers, the project aims to provide valuable insights that can help businesses improve customer retention and reduce revenue loss due to churn.

The dataset includes demographic information (such as geography, age, and gender), financial details (like balance, credit score, and estimated salary), and customer behavior indicators (such as tenure, number of products, and activity status). The target variable, **Exited**, indicates whether or not a customer has churned.

The project begins with thorough **data preprocessing**, which includes handling missing values, encoding categorical variables, and scaling numerical features. A detailed **exploratory data analysis (EDA)** was conducted to uncover trends and patterns in the data. Visualizations such as histograms, box plots, correlation heatmaps, and churn distribution charts provided meaningful business insights.

To predict churn, three different machine learning models were implemented and compared:

- **Logistic Regression** – A baseline model to understand linear relationships.
- **Random Forest Classifier** – A robust ensemble method capable of handling non-linear patterns.
- **XGBoost Classifier** – An optimized gradient boosting algorithm known for its high performance on classification tasks.

All models were evaluated using accuracy, precision, recall, F1-score, and ROC-AUC metrics. **XGBoost and Random Forest outperformed Logistic Regression**, achieving higher accuracy and better generalization on the test data.

Feature importance analysis from Random Forest and XGBoost models revealed that **Age**, **Balance**, **Geography**, and **Number of Products** were the most influential factors affecting churn. These findings can help banks identify high-risk customers and design targeted strategies to retain them.

This project highlights the application of machine learning in solving a real business problem, covering the full data science pipeline—from raw data to actionable insights. It demonstrates the power of predictive modeling in making informed, data-driven business decisions.

---