Problem Statement

Background:

A major bank is experiencing customer churn in its credit card division. Understanding the factors that lead to customer attrition is critical for designing retention strategies and improving customer satisfaction.

Objective:

You are provided with a synthetic dataset containing customer information, transaction history, and demographic details. Your task is to build a predictive model that identifies customers likely to close their credit card accounts.

Task Instructions

1. Data Understanding & Cleaning

- Explore the dataset and summarize key statistics.
- Identify and handle:
  - ✓ Missing values
  - ✓ Outliers
  - ✓ Duplicate records
  - ✓ High cardinality of categorical features
  - ✓ Imbalanced target variable (AttritionFlag)
  - ✓ High dimensionality

2. Feature Engineering

- Create meaningful features that may improve model performance.
- Consider dimensionality reduction techniques if needed.

3. Model Development

- Train at least two classification models (e.g., Logistic Regression, Random Forest, XGBoost).

- Use appropriate techniques to handle class imbalance (e.g., SMOTE, class weights).
- Perform hyperparameter tuning and cross-validation.

4. Model Evaluation

- Evaluate models using:
  - ✓ Accuracy
  - ✓ Precision, Recall, F1-score
  - ✓ ROC-AUC
  - ✓ Confusion Matrix
- Justify your choice of the final model.

5. Insights & Recommendations

- Identify key drivers of attrition.
- Provide actionable recommendations for the bank.

6. Bonus (Optional)

- Build a simple dashboard or app (e.g., Streamlit) to visualize attrition risk.
- Deploy the model and provide a link or demo.

Scoring Rubric (100 Points Total)

Category	Criteria	Points
Data Cleaning & Preprocessing	Handling missing values, outliers, duplicates, cardinality, imbalance	20
EDA & Feature Engineering	Quality of insights, creativity in feature creation	15
Modeling	Choice of models, handling imbalance, tuning	20
Evaluation	Use of appropriate metrics, clarity of comparison	15
Interpretability	Identification of key features, business relevance	10
Presentation	Code readability, documentation/PPT, visualizations	10
Bonus	Dashboard/app deployment, innovative approaches	10
Total		100