Exploratory Data Analysis (EDA) Report

# 1. Univariate Analysis

Univariate analysis explores each feature individually:  
  
- Numerical features like `credit\_score`, `age`, `balance`, and `estimated\_salary` show varied distributions.  
- Categorical features:  
 • Gender is almost evenly distributed.  
 • Most customers hold 1 or 2 products.  
 • Around 20% of customers churned, indicating an imbalanced dataset.

# 2. Bivariate Analysis

Bivariate analysis checks the relationship between two variables, particularly with the target variable `churn`:  
  
- Gender: Female customers have a slightly higher churn rate.  
- Active Member: Inactive members churn more.  
- Product Number: Customers with more products churn less.  
- Numerical vs Churn:  
 • Age and Balance show higher values among churned customers.  
 • Credit Score is slightly lower for churned customers.  
- Correlation Heatmap:  
 • No strong multicollinearity.  
 • Features are mildly correlated with `churn` (e.g., `balance`, `active\_member`).

# 3. Multivariate Analysis

Multivariate analysis looks at interactions between multiple variables:  
  
- Pairplot: Distinct patterns observed between `balance`, `estimated\_salary`, and `churn` clusters.  
- Barplots:  
 • Churn varies by gender and country.  
 • Customers who are active and have fewer products churn less frequently.

# Summary of Insights

- About 20% of the customers have churned.  
- Age and balance show visible difference in distribution between churned and non-churned customers.  
- Inactive members and customers with fewer bank products are more likely to churn.  
- Gender and geography show moderate influence on churn behavior.  
- The dataset is clean with no missing values after preprocessing.