# **Customer Churn Analysis Report**

## **1. Introduction**

This report summarizes the exploratory data analysis, data cleaning and preprocessing, and key insights derived from our customer churn dataset. The goal is to provide a clear, concise foundation for predictive modeling and guide future analytical efforts to reduce churn.

## **2. Data Sources & Feature Selection**

* **Data Loading**: Combined customer demographics, transaction history, login records, and service usage logs.
* **Feature Selection Rationale**: Chose features with known impact on churn (Age, AmountSpent, LoginFrequency, IncomeLevel, ServiceUsage, ProductCategory, Gender, MaritalStatus).

## **3. Exploratory Data Analysis (EDA)**

* **Age Distribution**: Age is broadly distributed; a slight concentration in the 25–45 range, with churn rates higher among younger customers.
* **Spending Patterns**: Most customers spend modest amounts; high spenders exhibit lower churn.
* **Login Frequency**: Higher login frequency correlates with retention; low-activity users show elevated churn.
* **Service & Product Usage**: Certain service packages and product categories are associated with distinct churn behaviors.
* **Demographic Segments**: Churn varies by gender, marital status, and income level—informing targeted interventions.

## **4. Data Cleaning & Preprocessing**

* **Handling Missing Values**: Imputed numeric fields with median; categorical fields with the most frequent value.
* **Type Corrections**: Converted ChurnStatus to binary; ensured numeric types for continuous features.
* **Encoding**: Label-encoded binary categories; one-hot encoded multi-class variables.
* **Outliers & Skew**: Capped AmountSpent at the 99th percentile; log-transformed LoginFrequency.
* **Feature Engineering**: Binned Age into age groups; created a log-scaled login frequency.
* **Scaling**: Standardized numeric features for consistent modeling.

## **5. Visualization Highlights**

* **Histograms**: Illustrated distributions of age, spending, and login frequency by churn status.
* **Boxplots**: Revealed distributional differences across churn for age, income, and login behavior.
* **Scatter Plot**: Showed the relationship between Age and AmountSpent, highlighting churn clusters.
* **Countplots**: Compared churn counts across gender, marital status, service usage, income levels, and product categories.

## **6. Conclusions & Recommendations**

* **Key Insights**:
  1. Younger, low-activity, and low-spending customers churn at higher rates.
  2. Income level and service usage category significantly influence retention.
* **Actionable Recommendations**:
  1. **Targeted Engagement**: Design loyalty programs for younger and low-frequency users.
  2. **Personalized Offers**: Develop incentives for high-risk segments based on service usage.
  3. **Monitoring & Alerts**: Implement real-time churn-risk scoring to trigger retention actions.

## **7. Next Steps**

1. **Model Development**: Train and validate classification models (e.g., logistic regression, random forest, XGBoost).
2. **Evaluation**: Use precision, recall, AUC, and business KPIs to select the optimal model.
3. **Deployment**: Integrate churn prediction into CRM workflows for proactive outreach.
4. **Continuous Improvement**: Monitor model performance, retrain regularly, and incorporate new data (e.g., customer feedback).