

Customer Churn EDA & Data Preparation Report

Customer Churn - Data Preparation and EDA Report

1. Data Sets Selected & Rationale

Customer_Demographics Captures age, gender, marital status, and income level; useful for understanding customer profile patterns related to churn.

Transaction_History Includes transaction dates and amounts; essential for measuring spending behavior.

Customer_Service Tracks interaction types and resolution statuses; potential indicator of dissatisfaction.

Online_Activity Login frequency and platform usage; reflects engagement level.

Churn_Status Target variable for model training (0 = Active, 1 = Churned).

2. EDA & Insights

Churn distribution is imbalanced with more active users than churned ones.

Churned users tend to spend less and have lower login frequencies.

Absence of service interactions correlates positively with churn.

3. Data Cleaning & Preprocessing

Missing service interaction values were filled with "No Interaction" or "None".

Categorical variables were one-hot encoded (e.g., Gender, IncomeLevel, ServiceUsage).

Numerical features (Age, LoginFrequency, AmountSpent) were normalized using StandardScaler.

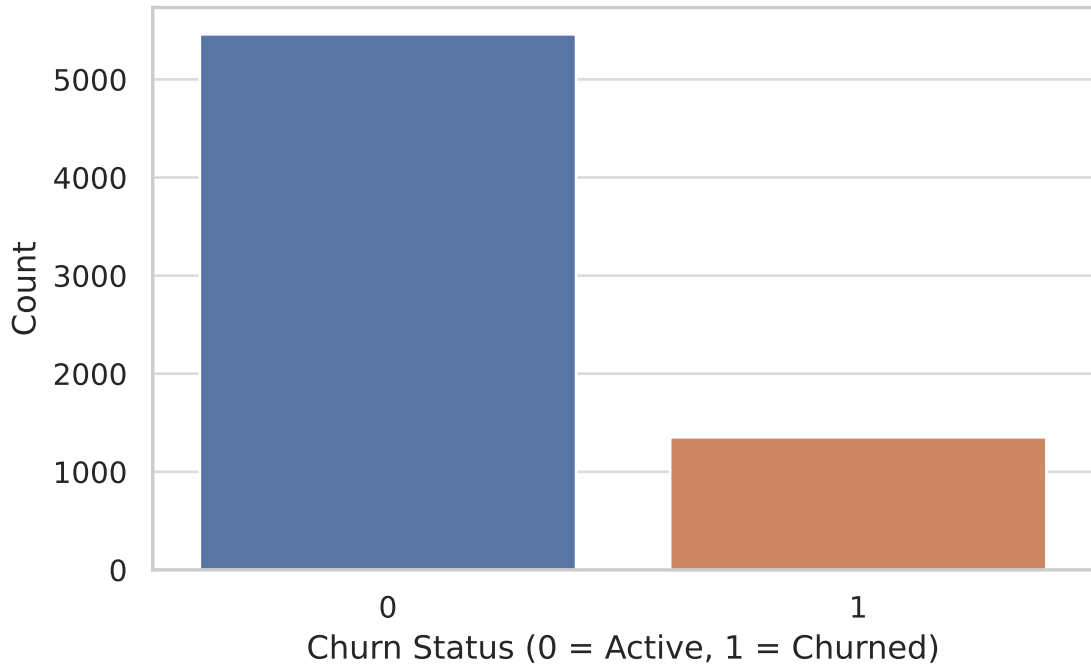
Non-predictive ID and date fields were removed before model development.

4. Output

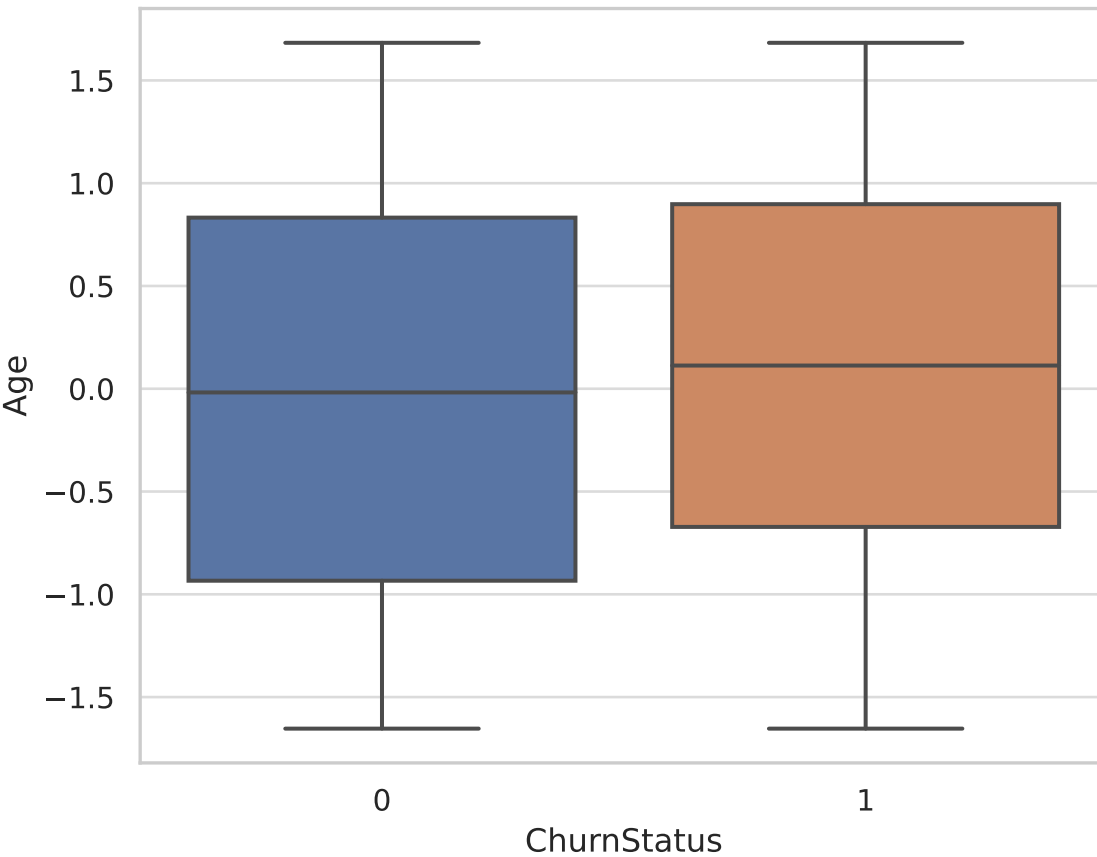
Cleaned and preprocessed data frame has 6,812 rows and 20 features + 1 target.

Prepared dataset is now ready for predictive modeling.

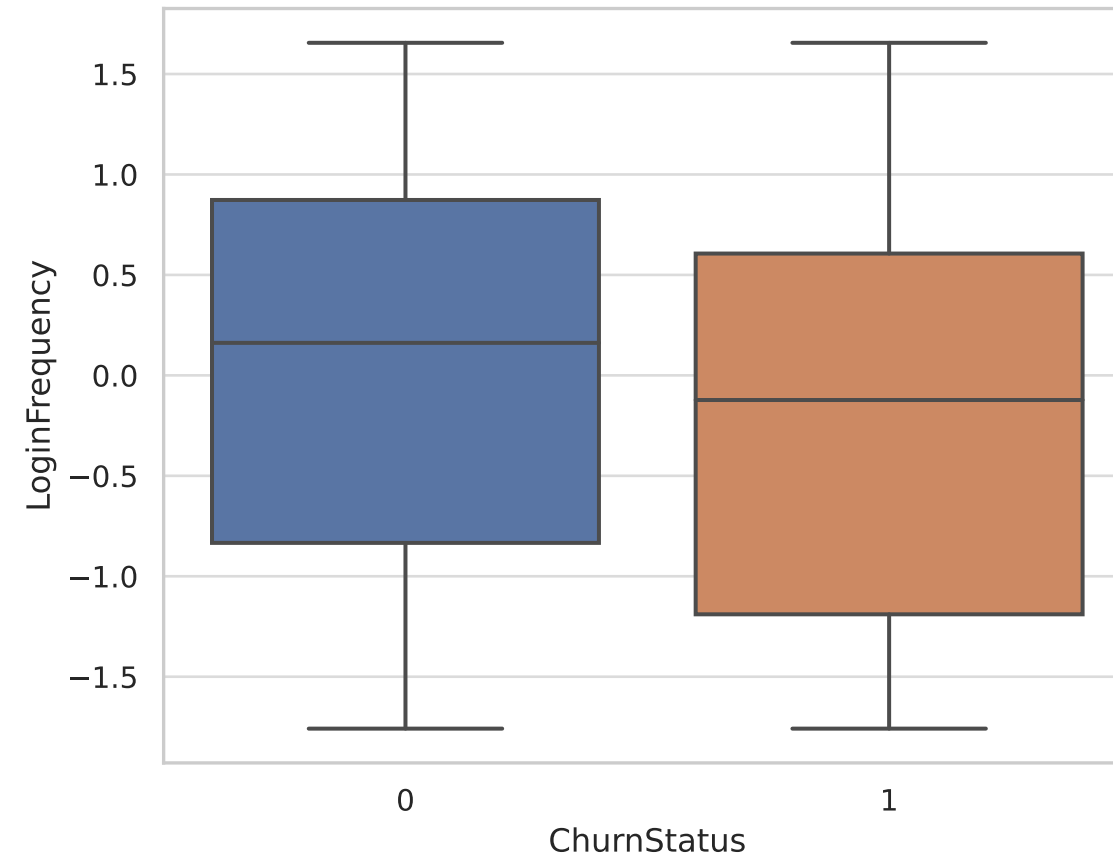
Churn Distribution



Age vs Churn



Login Frequency vs Churn



Amount Spent vs Churn

