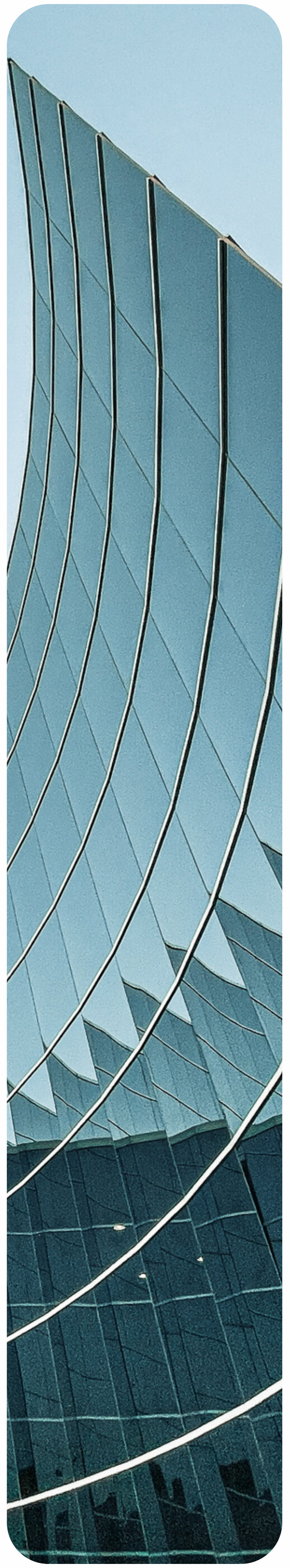



LLOYDS
BANKING GROUP

Customer Churn Prediction:

Data Preparation Report

Date: May 17, 2025

Selected Dataset

The dataset contains 5,204 customer records and includes the following eighteen features:

Demographics	Transaction Date	Service Interaction	Digital Engagement	Target Variable
Age	Transactiondate	InteractionDate	LastLoginDate	ChurnStatus
Gender	AmountSpent	InteractionType	LoginFrequency	
MaritalStatus	DaysSinceInteractionn	ResolutionStatus	ServiceUsage	
IncomeLevel		InteractionLag	DaysSinceLogin	
		DaysSinceInteraction	LoggedInLast30Days	

Rationale:

These features were selected as they provide a balanced view of customer behaviour ,spending , engagement, and satisfaction. These behavioral and engagement metrics are essential for churn prediction

Explanatory Data Analysis(EDA)

Statistical Summaries:

Feature	Mean	Std Dev	Notes
Age	Scaled	Scaled	Standardized age distribution
AmountSpent	Scaled	Scaled	Right-skewed; high variance
Login Frequency	Scaled	Scaled	Varies widely; potential churn link
DaysSinceLogin	Scaled	Scaled	Higher values may indicate churn

Churn Rate

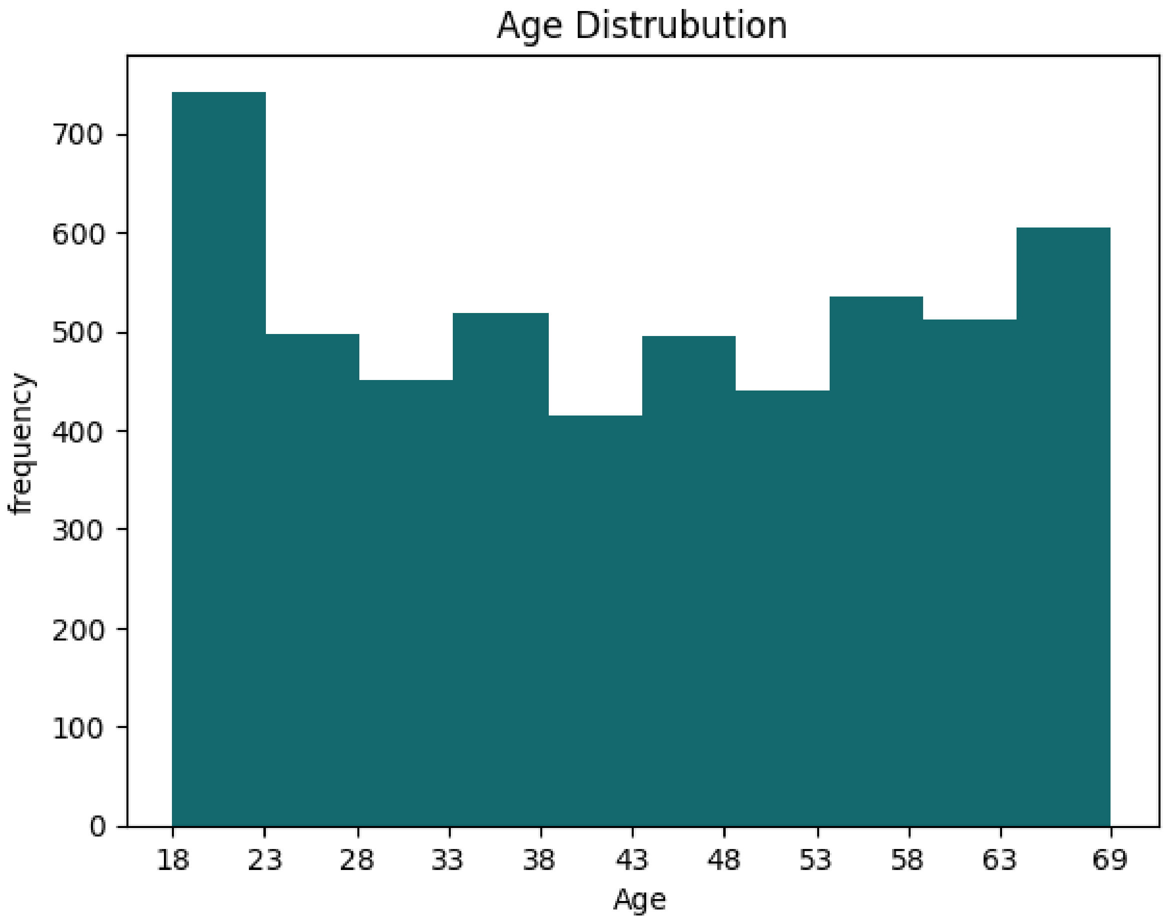
Approximately 60% churn, indicating classs imbalance

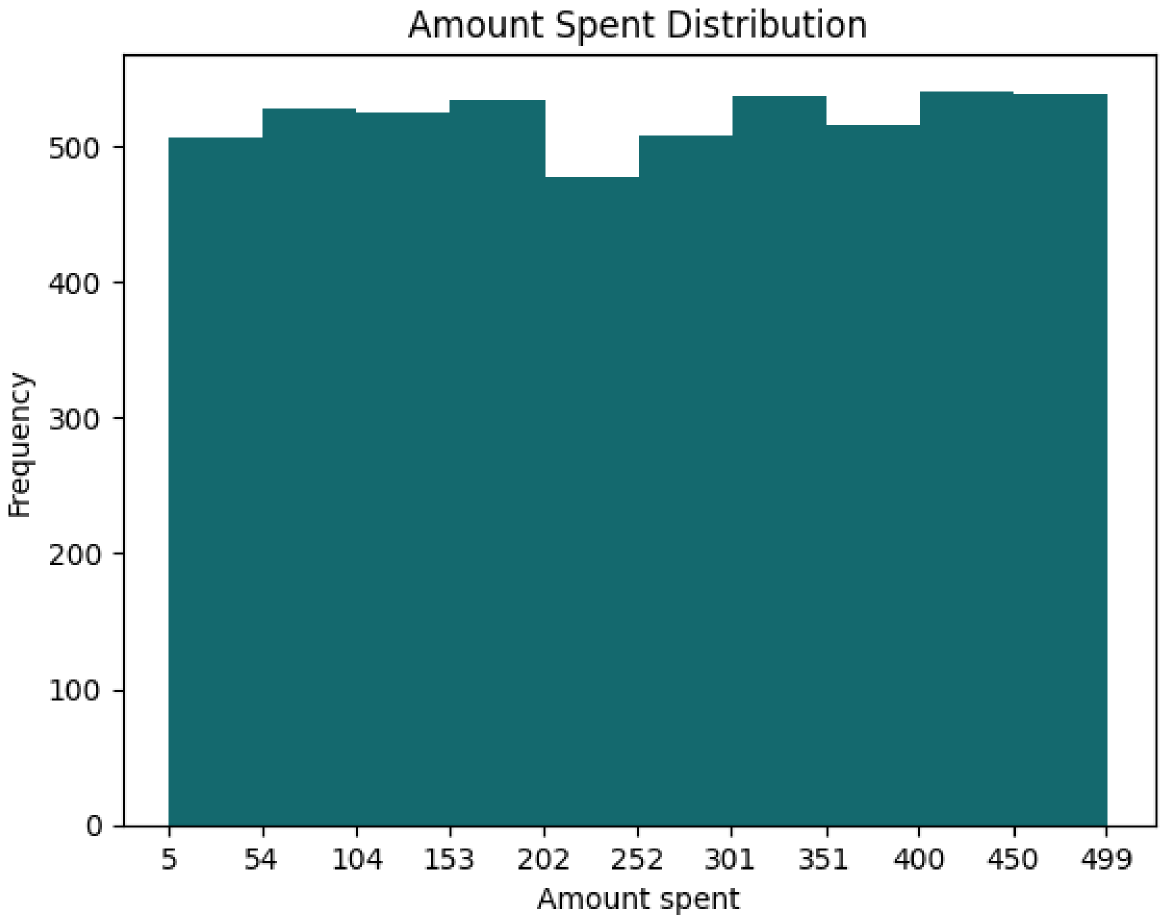
Interaction Lag

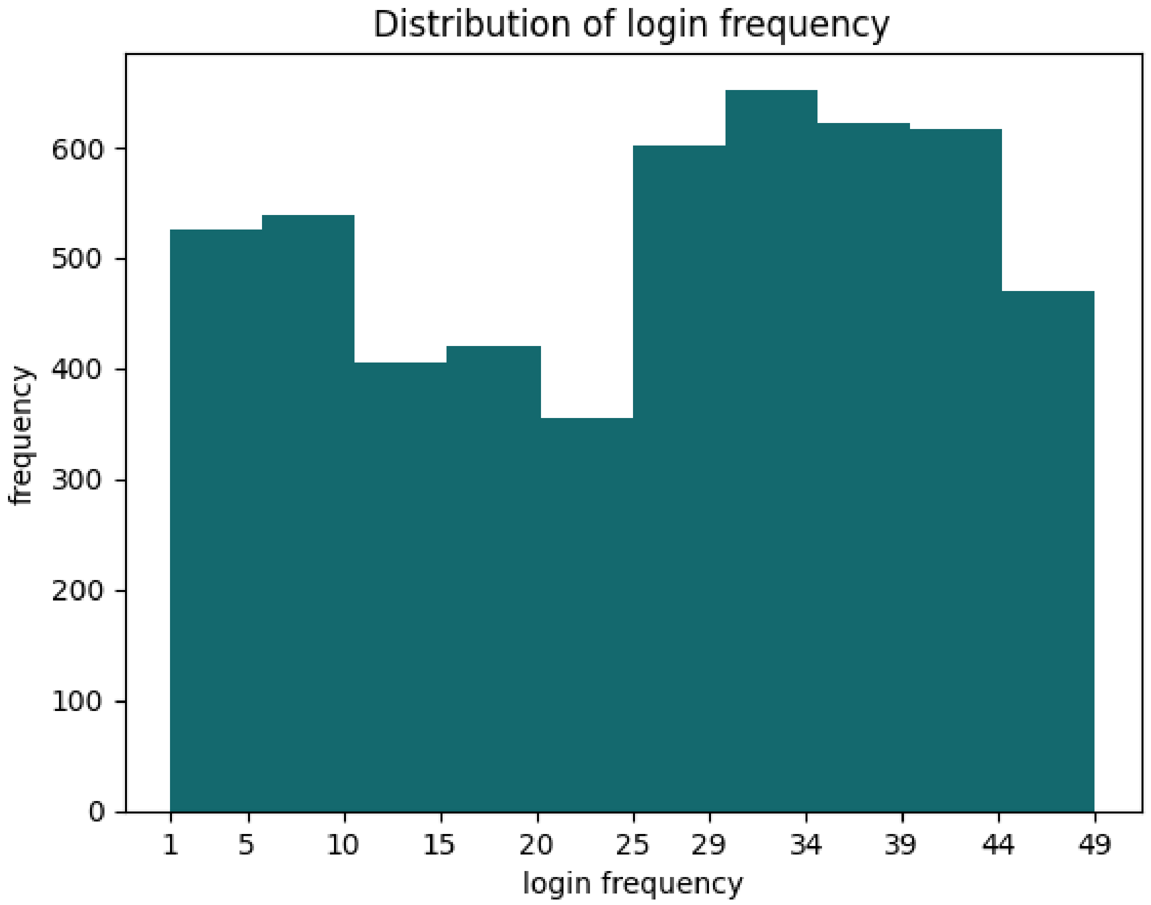
Longer lag between spending and interaction may signal dissatisfaction

Visualizations:

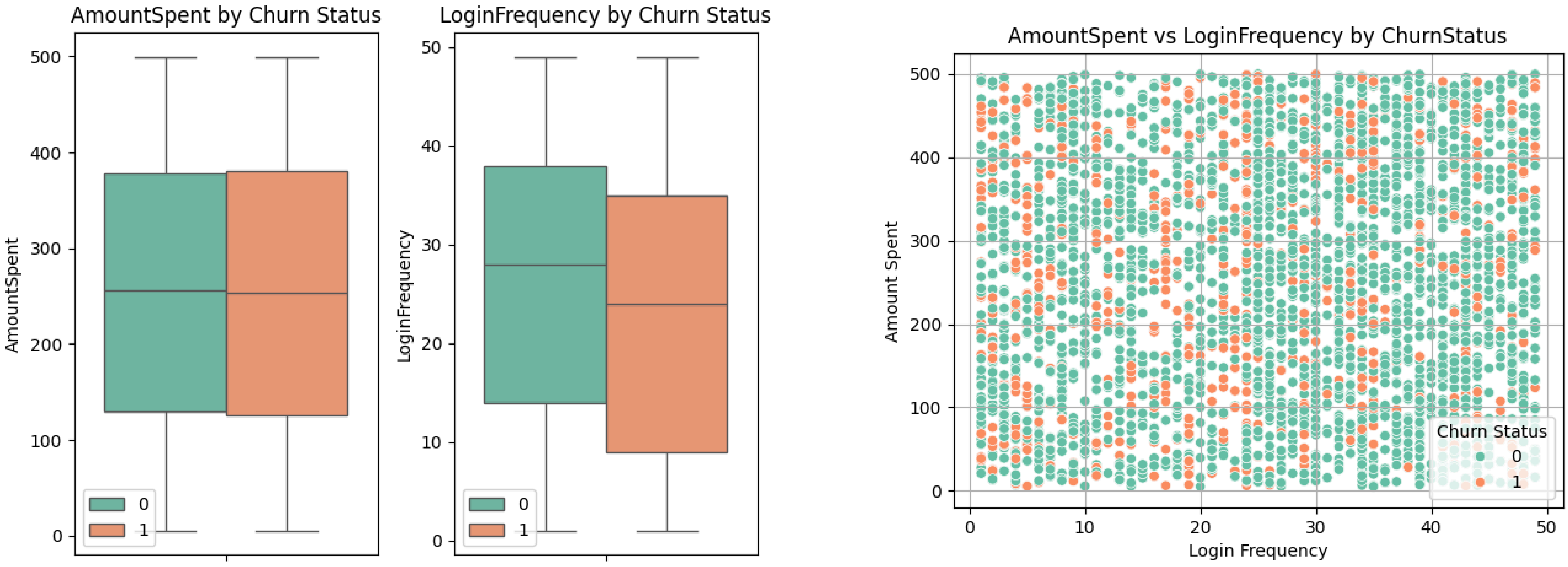
Distribution of Age, Amount spent and Login Frequency:



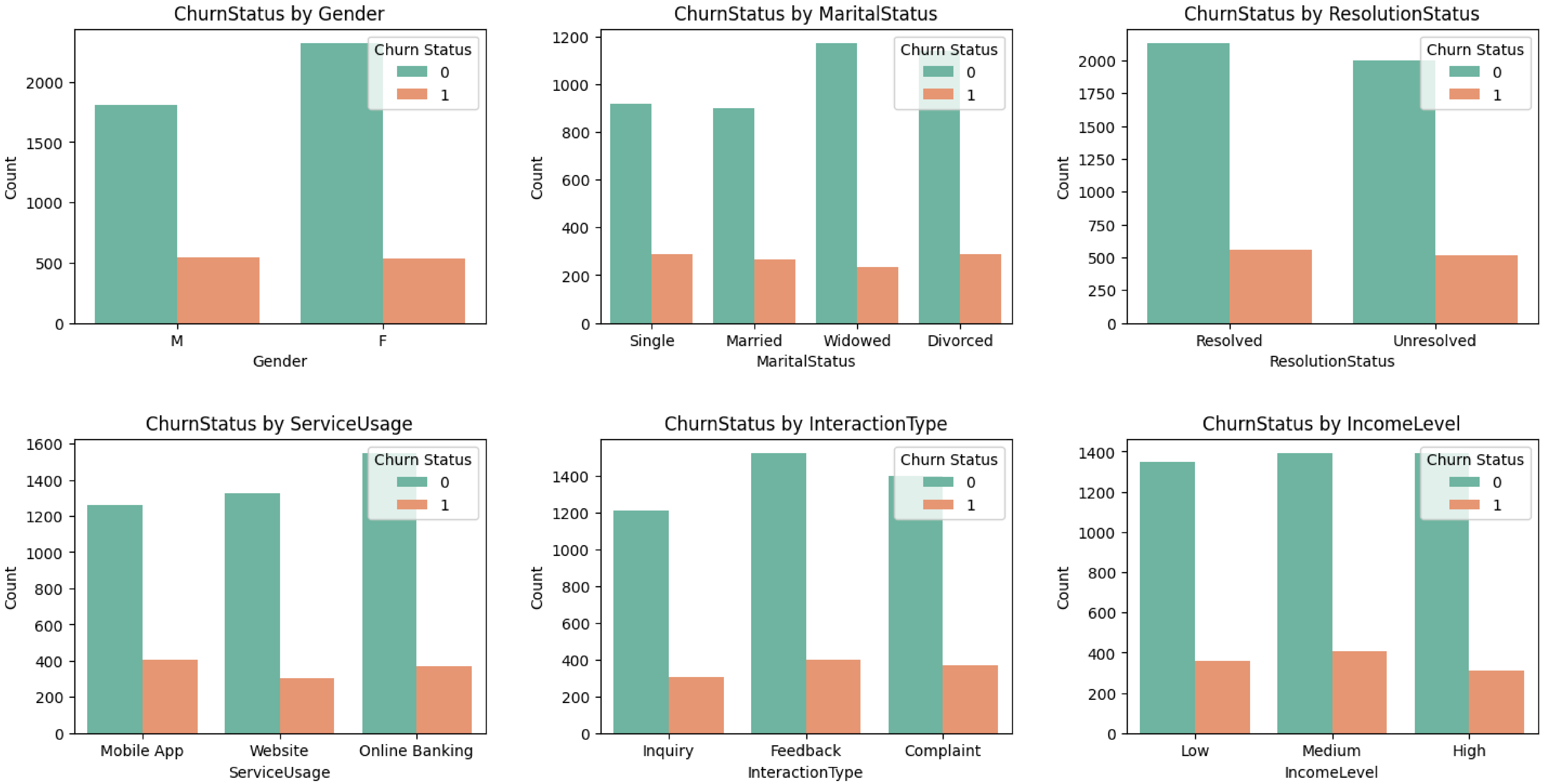




Amount spent and Login Frequency by Churn Status:



Churn Status by Categorical Datas:



Data Cleaning and Preprocessing:

Missing Values Handaling

No missing values present after cleaning (5204/5204 non-null for all columns).

Outlier Detection

- AmountSpent had extreme values
- All numerical features were standardized using StandardScaler.

Feature Engineering from Dates:

Dates were converted into meaningful numerical features:

Engineerd Feaure	Description
DaysSinceTransaction	Recency of last purchase
DaysSinceInteraction	Recency of last service interaction
DaysSinceLogin	Time since last login
InteractionLag	Gap between transaction and interaction
LoggedInLast30Days	Boolean indicator of recent login

Categorical Encoding:

All categorical variables were label encoded to convert them into numerical format:

Variable	Encoding
Gender	Binary (0/1)
MaritalStatus	Ordinal values
IncomeLevel	Ordinal values
ResolutionStatus	Binary (Resolved/ Unresolved)
ServiceUsage	Encoded for Mobile, Web, Online Banking
InteractionType	Inquiry, Complaint, Feedback, etc.

Cleaned Dataset Summary:

- Final shape : (5204 rows × 19 columns)
- Target column : ChurnStatus (0 = Retained, 1 = Churned)
- All features are now numerical and standardized
- Dataset saved as : cleaned_data.csv

Conclusion and Next Step:

This cleaned dataset is now ready for modeling using classification techniques such as:

- Logistic Regression
- Random Forest
- XGBoost
- Neural Networks