

Customer Churn - Data Preparation Report

1. Dataset Summary

Source file: Customer_Churn_Data_Large.xlsx

Rows: 1000, Columns: 7

No missing values detected.

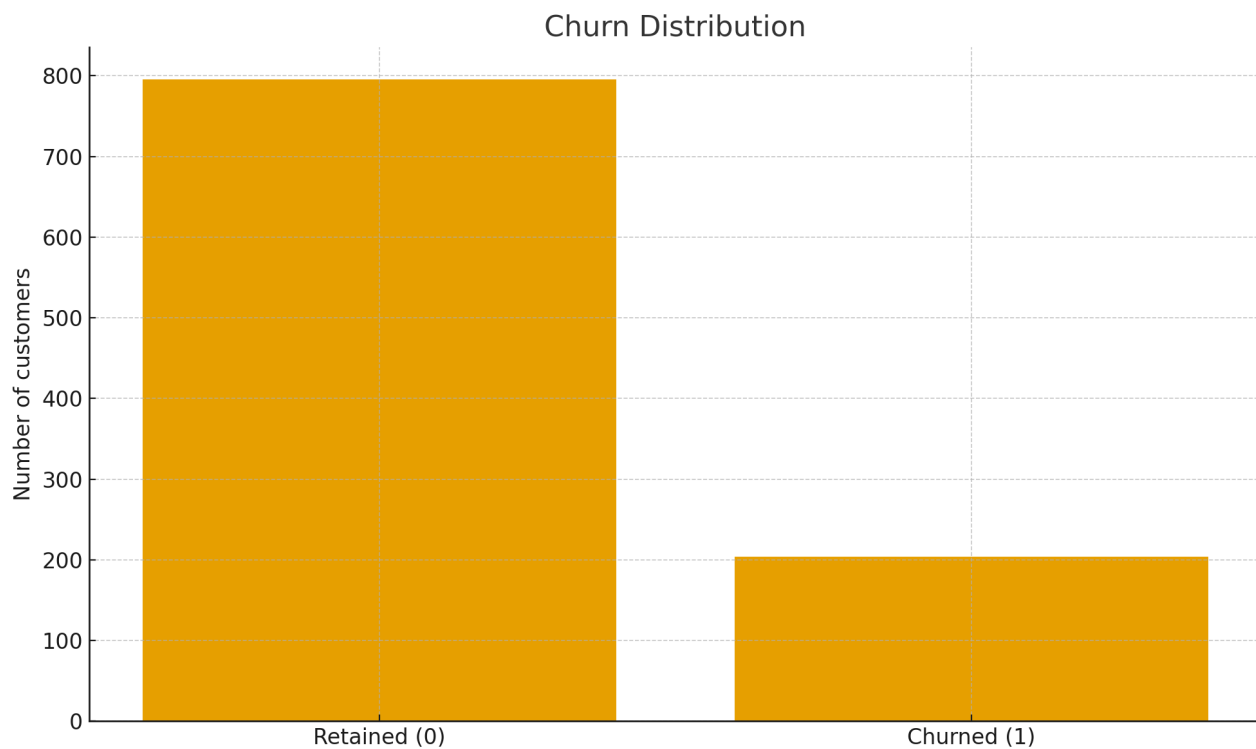
Dropped columns: CustomerID.1 (if existed).

2. Exploratory Data Analysis

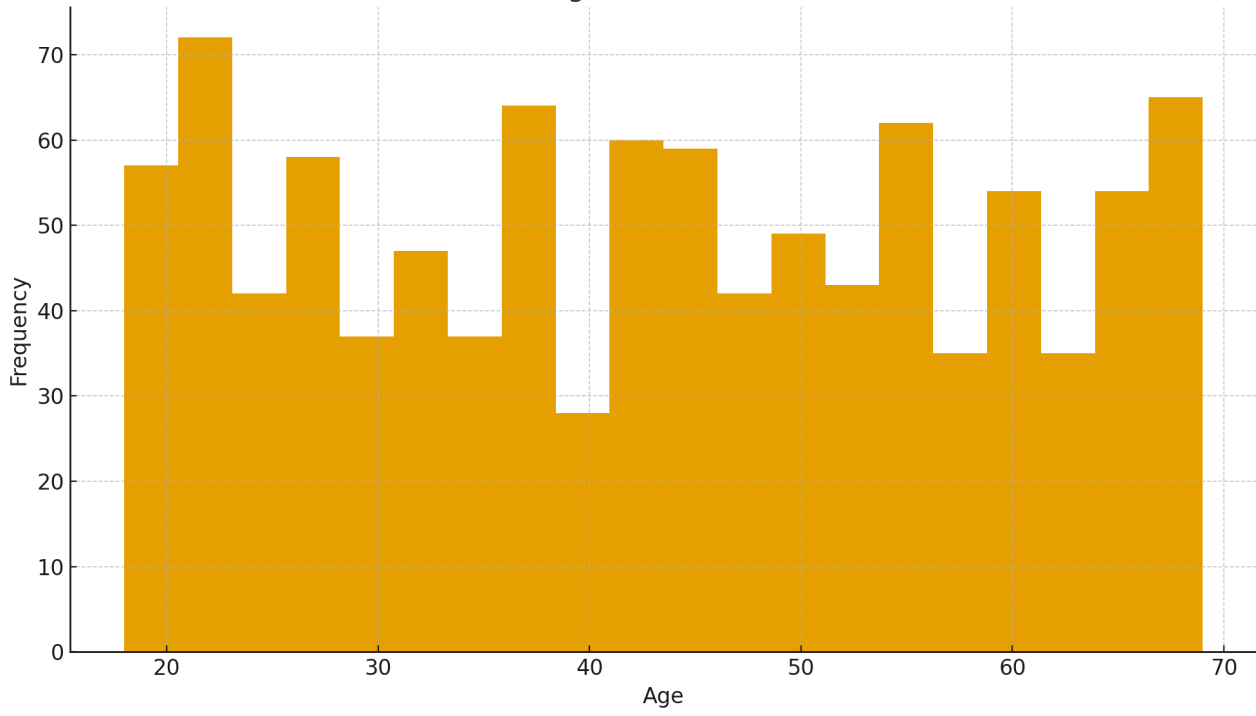
Overall churn rate: 0.204 (204 churned / 1000 total)

Age statistics (before scaling):

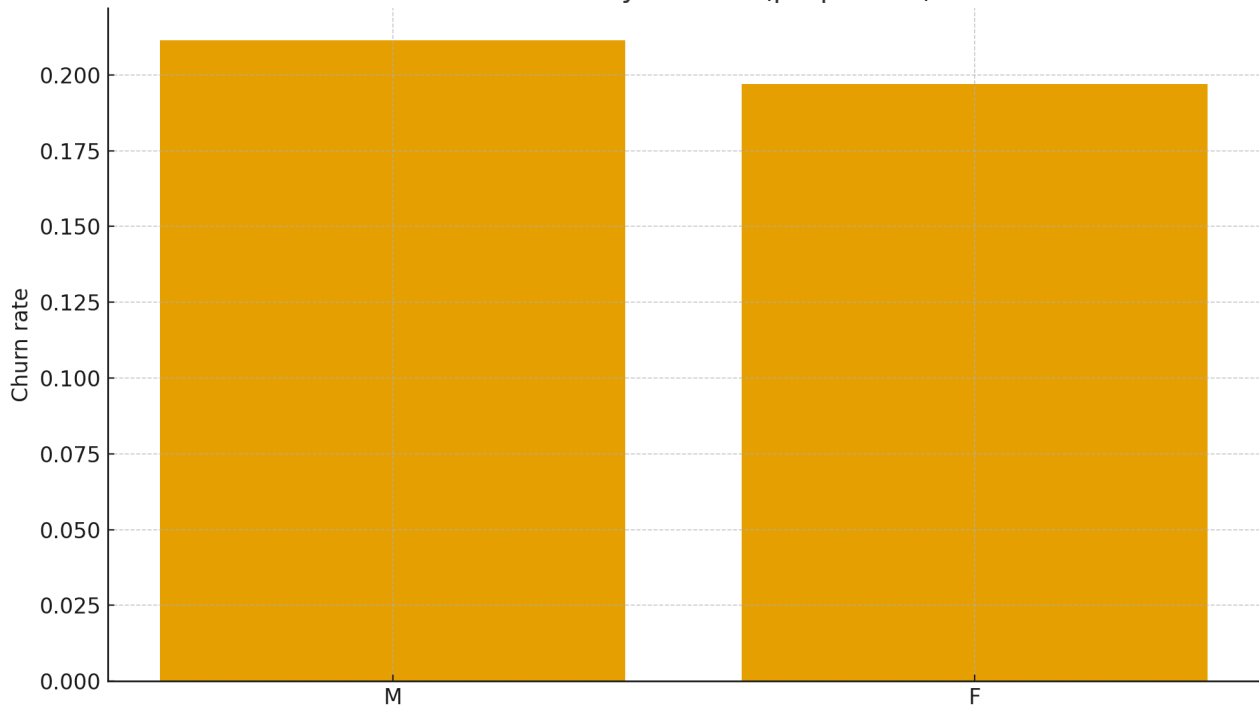
- count: 1000
- mean: 43.27
- std: 15.24
- min: 18
- 25%: 30
- 50%: 43
- 75%: 56
- max: 69



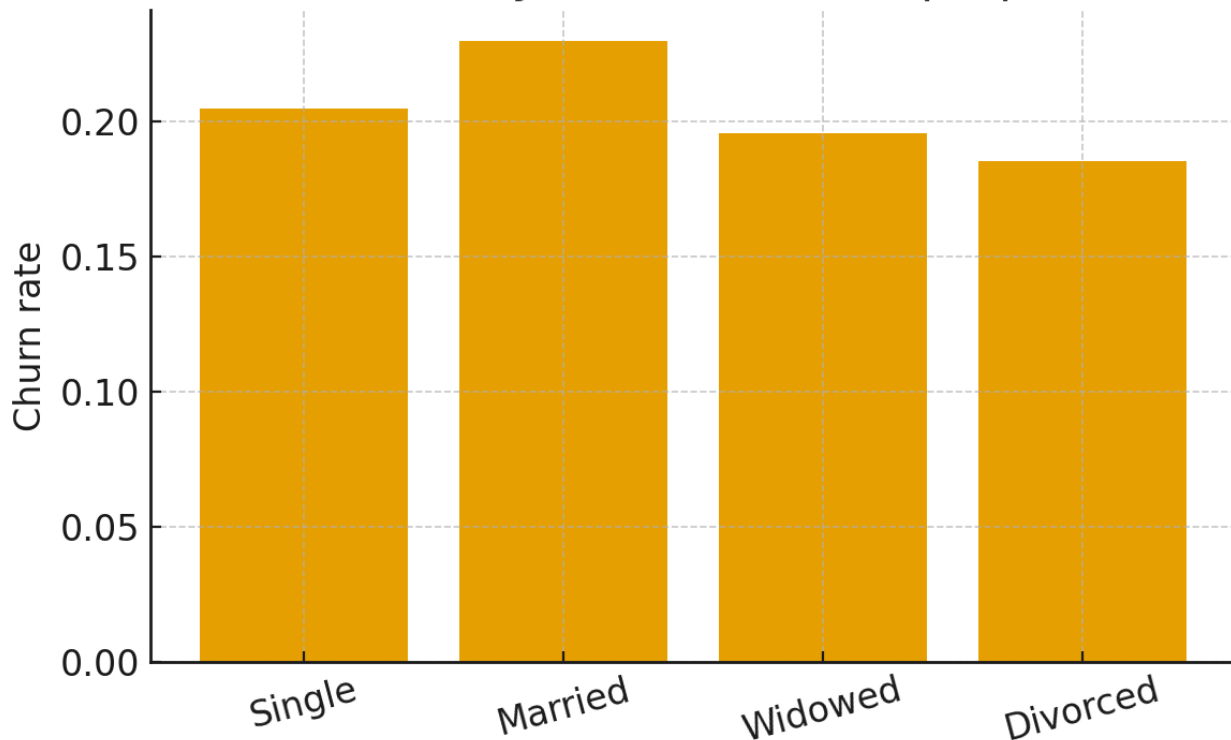
Age Distribution



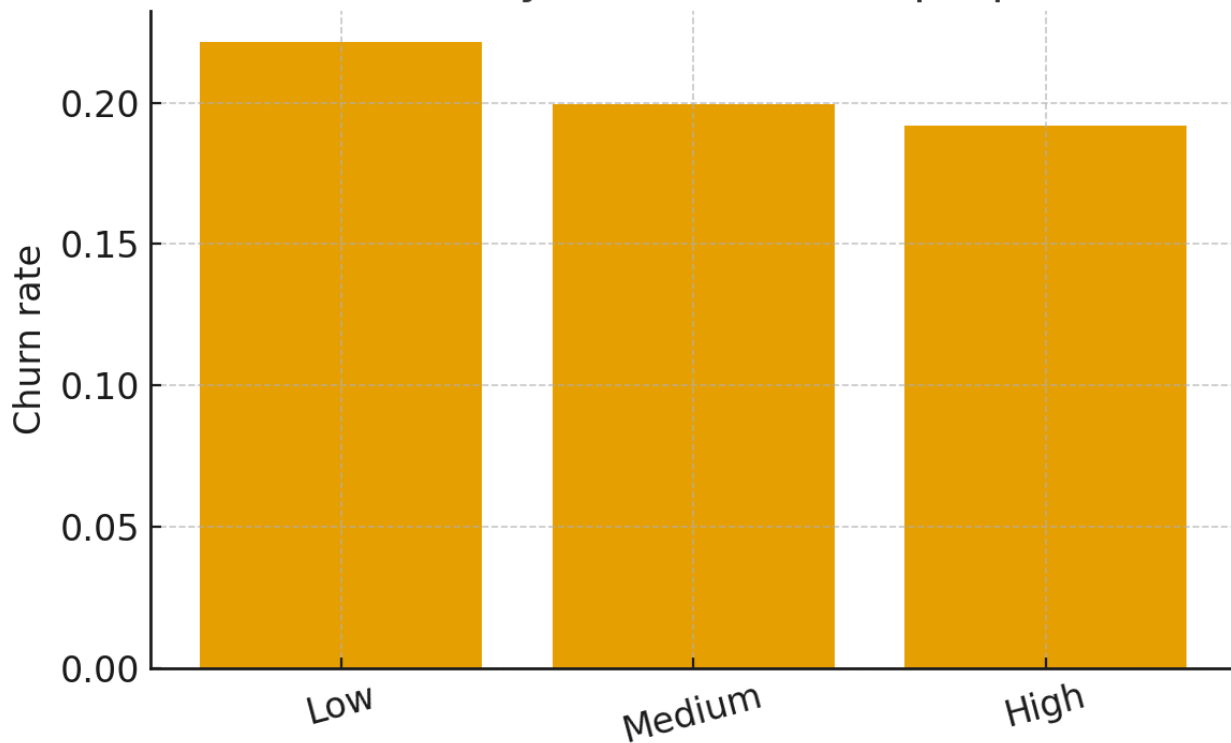
Churn Rate by Gender (proportion)



Churn Rate by Marital Status (proportion)



Churn Rate by Income Level (proportion)



3. Data Cleaning & Preprocessing

- Dropped redundant column 'CustomerID.1'.
- Encoded 'Gender' (M/F) to 0/1.
- One-Hot encoded 'MaritalStatus' and 'IncomeLevel'.
- Standardized 'Age' using StandardScaler.

- Final cleaned dataset saved as 'Customer_Churn_Cleaned.csv'.

4. Suggested Next Steps

- Feature engineering: create tenure-related or usage-derived features if available.
- Split dataset into train/test and try baseline models (Logistic Regression, Random Forest).
- Use cross-validation and evaluate with AUC, accuracy, precision/recall.
- Tune thresholds to balance business priorities (precision vs recall).