Customer Churn Analysis Report

This report outlines the process of selecting, building, and evaluating a machine learning model to predict customer churn. The goal is to identify customers at risk of leaving and provide actionable insights to help mitigate this risk.

Data Identification and Collection

The data set includes the following key components:
- Customer Demographics: Age, Gender, Marital Status, Income Level.
- Transaction History: Transaction Dates, Amount Spent, Product Categories.
- Customer Service Interactions: Type and Resolution Status of Interactions.
- Online Activity: Last Login Date, Login Frequency, Service Usage.
- Churn Status: Indicating whether each customer has churned (1) or not (0).

Exploratory Data Analysis

Exploratory data analysis was conducted to uncover patterns and insights that could inform predictive modelling. Key findings include demographic trends, spending patterns, and the impact of customer service interactions on churn.

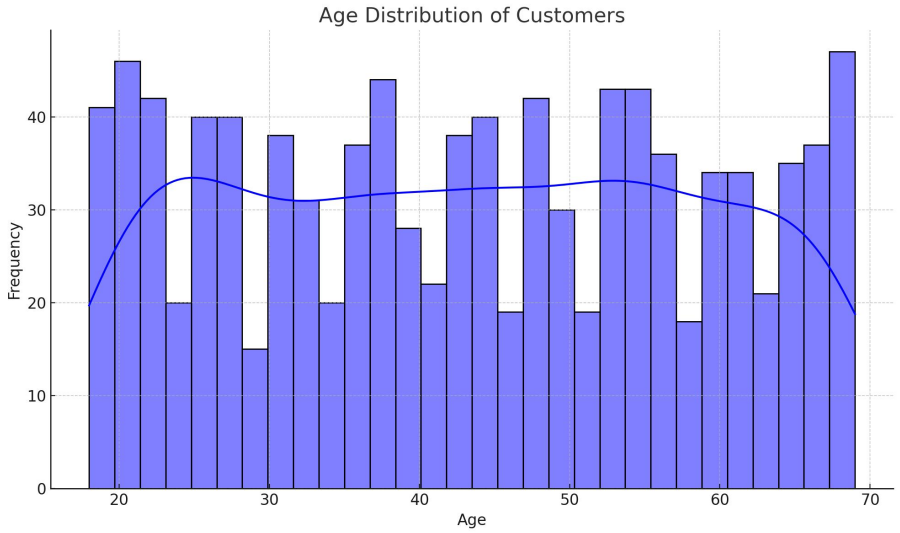

Figure 1: Age Distribution of Customers

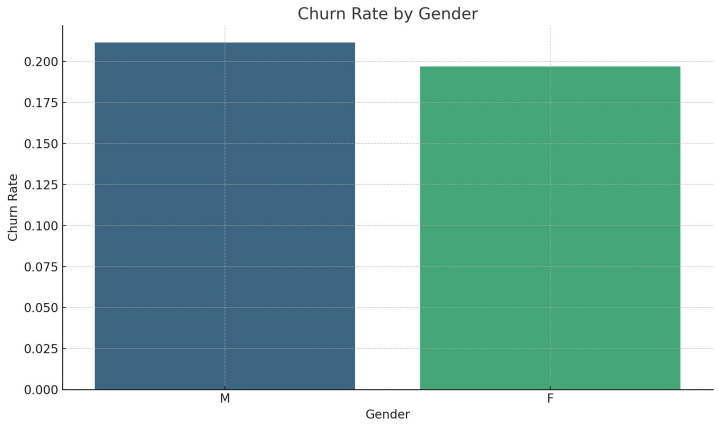

Figure 2: Churn Rate by Gender

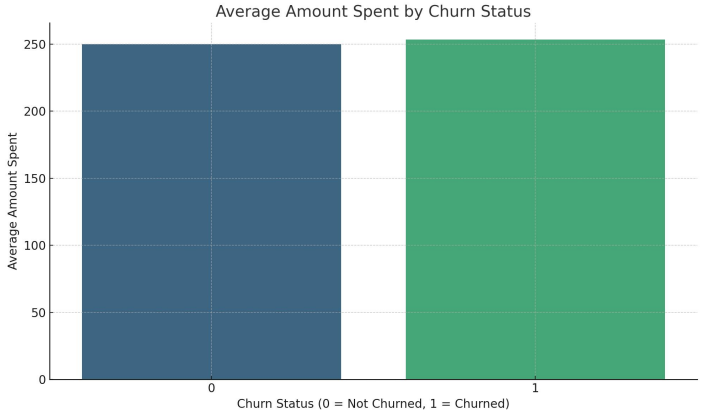

Figure 3: Average Amount Spent by Churn Status

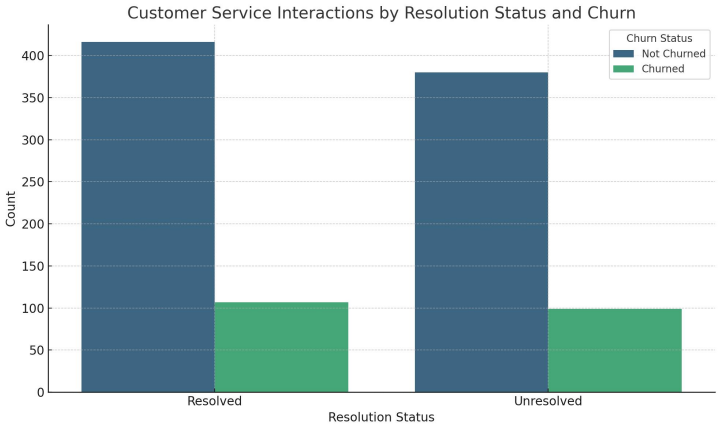

Figure 4: Customer Service Interactions by Resolution Status and Churn

Model Selection and Building

After considering several algorithms, a random forest classifier was selected for its balance of accuracy and interpretability. The model was trained using stratified 5-fold cross-validation, ensuring robustness and generalisation.

Model Evaluation

The model's performance was evaluated using precision, recall, F1 score, and ROC-AUC metrics. Feature importance analysis identified key predictors of churn, such as transaction amount and login frequency.

Recommendations and Business Implications

The model's outputs can be used to develop targeted retention strategies for at-risk customers. Further improvements could include integrating additional data sources and exploring more complex models like XGBoost or Neural Networks.