MAY 1, 2025

CUSTOMER RETENTION ENHANCEMENT THROUGH PREDICTIVE ANALYTICS REPORT

ATOMA, ERIKA
LLOYDS BANKING GROUP
UK

Table of Contents

Abstract

This report presents the exploratory data analysis and data preparation steps undertaken to predict customer churn using various data sets. This includes data collection, visualization and cleaning processes, and providing a foundation for developing a predictive model.

Data

The following data sets that were identified as relevant for predicting customer churn:

1. Customer Demographics
2. Transaction History
3. Customer Service Interactions
4. Online Activity
5. Churn Status

These data sets provide comprehensive insights into customer behavior and engagement.

Situation

Over the past few months, there is a worrying trend of increased customer churn, particularly among young professionals, and small business owners. This poses a substantial threat to Lloyds Banking Group's market position and long-term profitability.

The client, SmartBank, a subsidiary of Lloyds, has reported that a substantial portion of their customer base is at risk of moving to competitors offering more personalized banking solutions.

SmartBank has tasked our team with developing a predictive model to identify at-risk customers and propose targeted interventions to retain them.

Complication

Identify problem(s) or opportunity

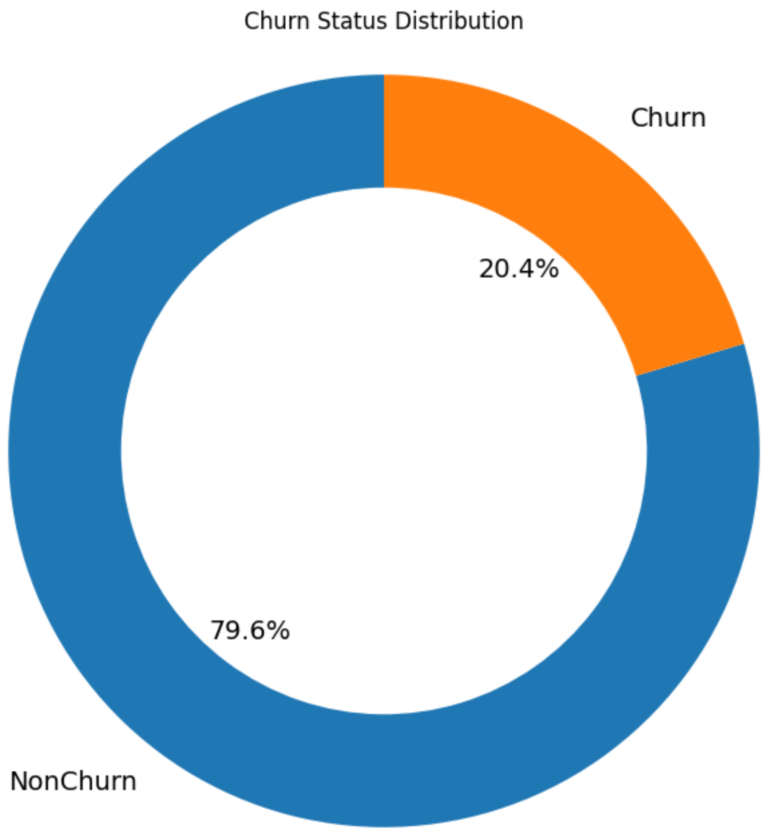

Figure 1: Churn Status Distribution

The churn rate is 20.4% across all the users in 2022. Therefore, this poses a problem that there is low retention as customers move to other competitors.

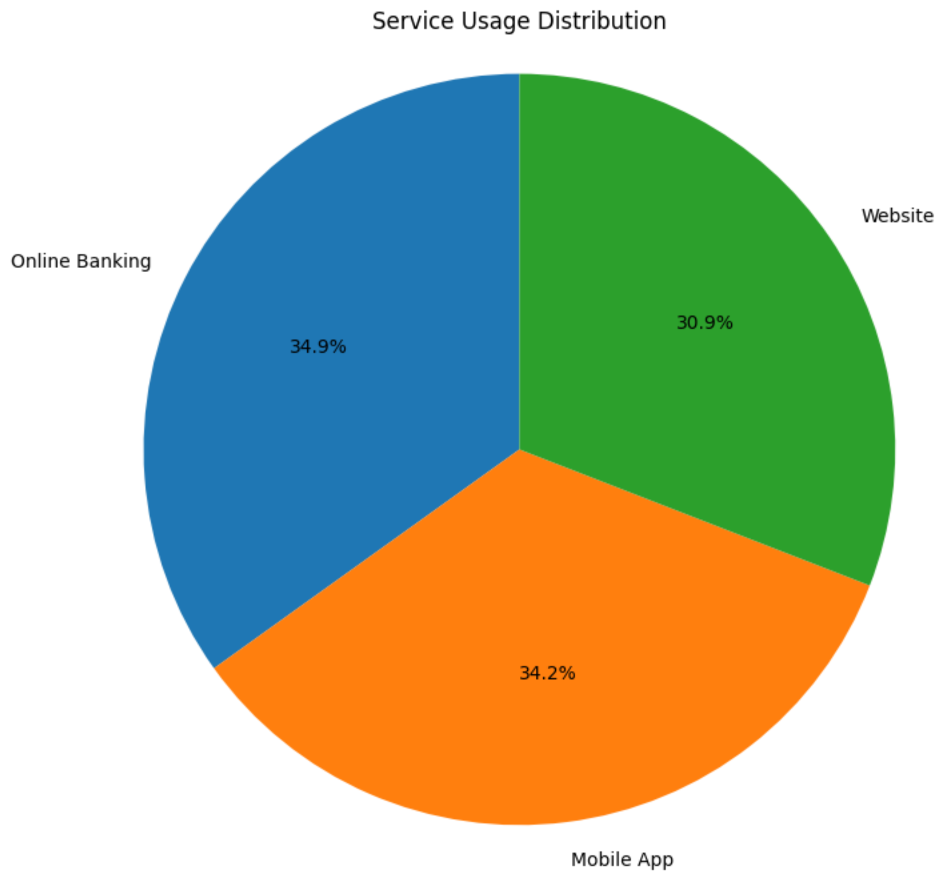

Figure 2: Service Usage Distribution

The distribution of how customer access Smart Bank's services, includes Online Banking (34.9%), Mobile App (34.2%), and Website (30.9%). The digital platforms are actively used, with Online Banking leading the top service usage. However, the close split suggests there is no dominant platform, which presents an opportunity to deliver consistent and personalized experiences across all channels.

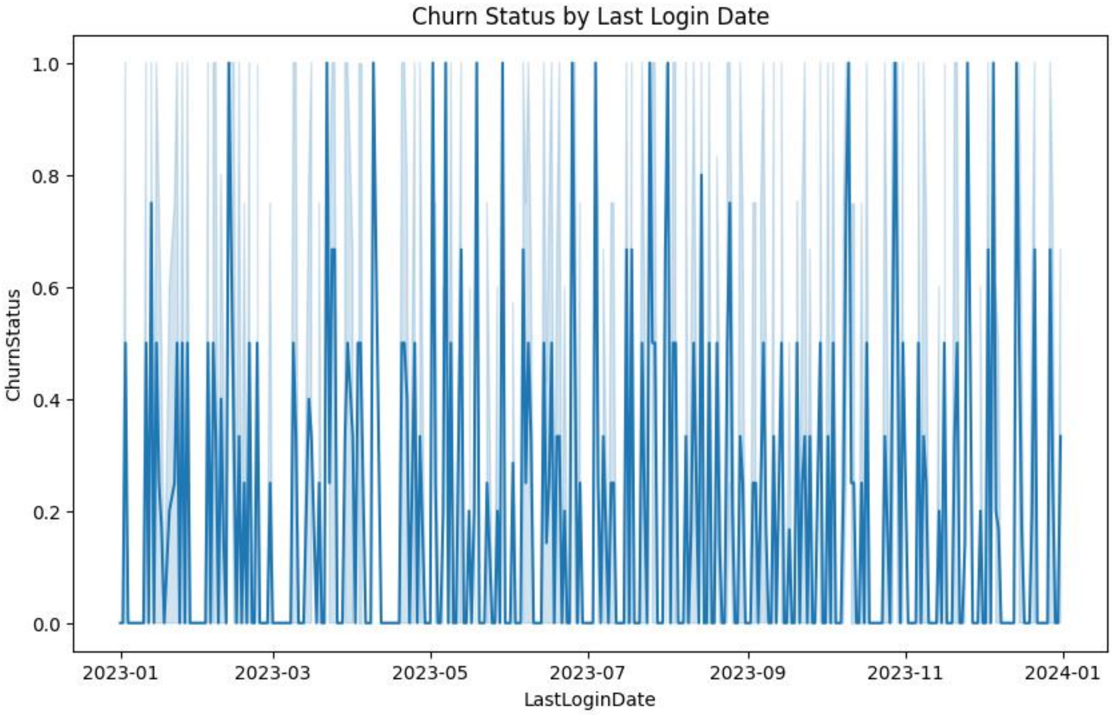

Figure 3: Churn Status by Last Login Data

The times series shows the occurrences where the customers churning over the course of 2023 based on their last login dates. There is no visible decline in churn towards the end of the year. This explains that customers may churn regardless of season, and therefore continued engagement is critical throughout the year.

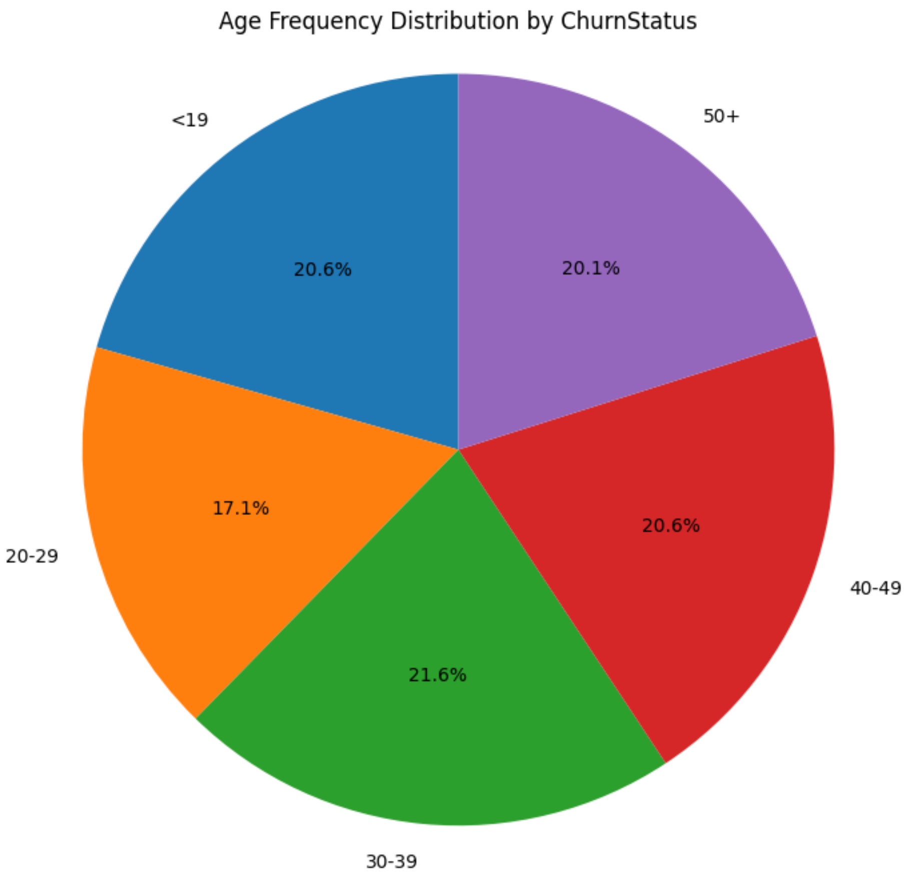

Figure 4: Age Frequency Distribution by Churn Status

Churn is relatively evenly spread across age groups, with a slightly concentration among working-age adults (30 to 39 years). The highest proportion of churn is 21.6% from 40-39 age group. Whereas the lowest proportion of churn is from the 20-29 age group with 17%.

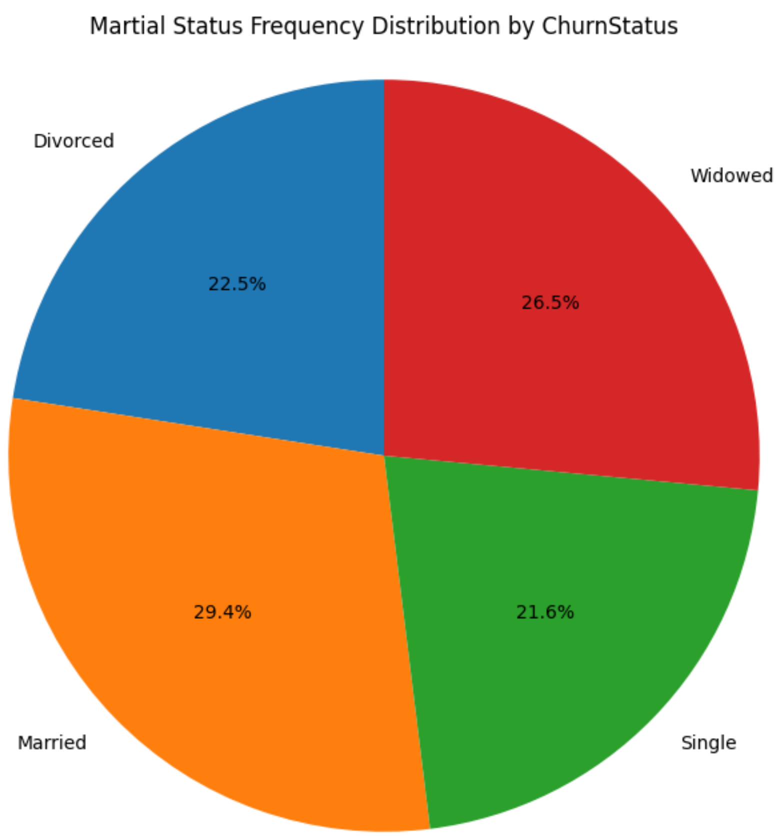

Figure 5: Martial Status Frequency Distribution by Churn Status

Married individuals have the highest churn rate (29.4%), and the single customers has the lowest churn rate (21.6%). This suggests that life-stage factors (e.g., family responsibilities, financial stress) may influence dissatisfaction for married couples with SmartBank's services. Therefore, tailored communication and family-centric benefits might improve retention for this customer segment.

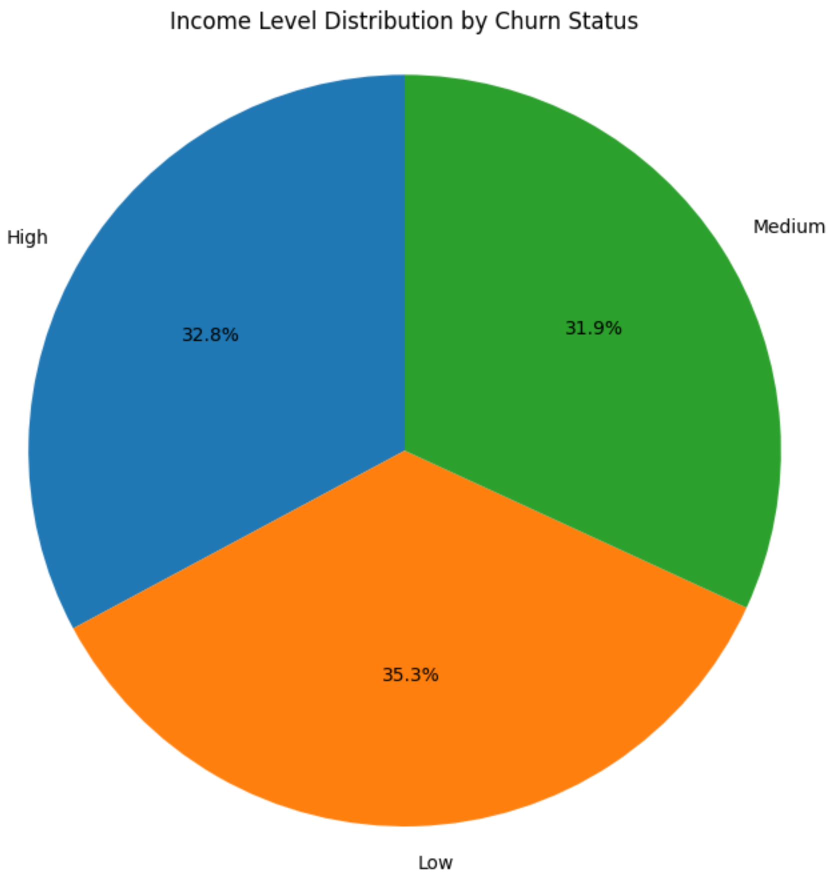

Figure 6: Income Level Distribution by Churn Status

The low-income earners have the highest churn rate of 35.3%, high-income earners had a churn rate of 32.8%, and the medium income of 31.9%. The lower income earners had the highest churn, possibly due to unmet financial needs. Therefore, affordability-focused incentives or fee waivers could be effective retention strategies.

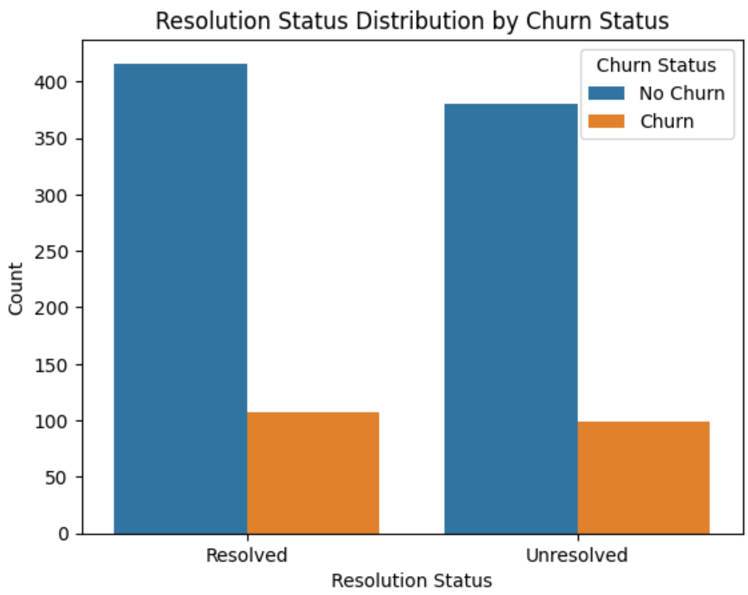

Figure 7: Resolution Status Distribution by Churn Status

Customers seems to churn more when issues remained unresolved. However, there is some churn when the issues are resolved. Therefore, there is some correlation between unresolved customer complaints and higher churn. Improving complaint resolution time and quality is crucial for customer retention.

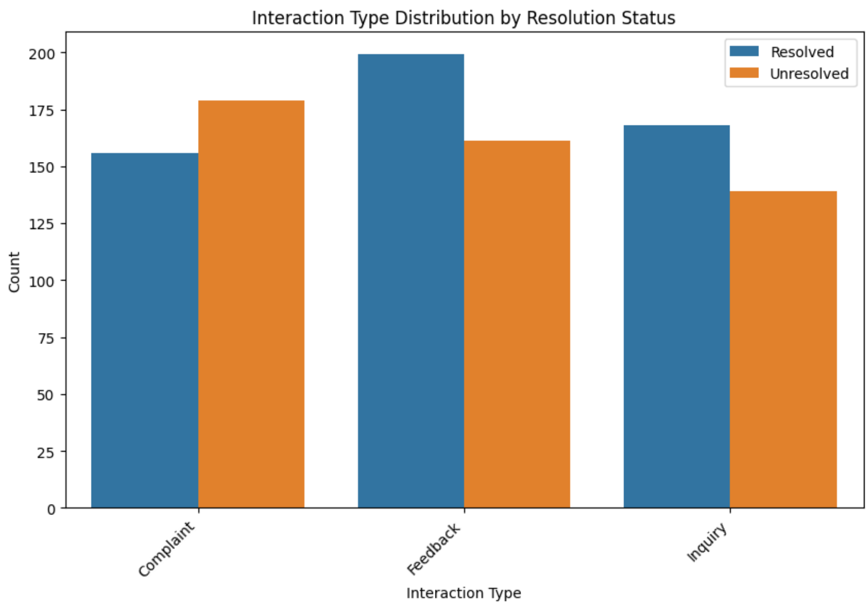

Figure 8: Interaction Type Distribution by Resolution Status

It appears that complaints and feebacks are often unresolved. There are 179 complaints and 161 feeback that are unresolved. We can see that the unresolved complaints outweighs the resolved complaints. This is problematic since they represent dissatification. These interactive types should be prioritized in customer support workflows.

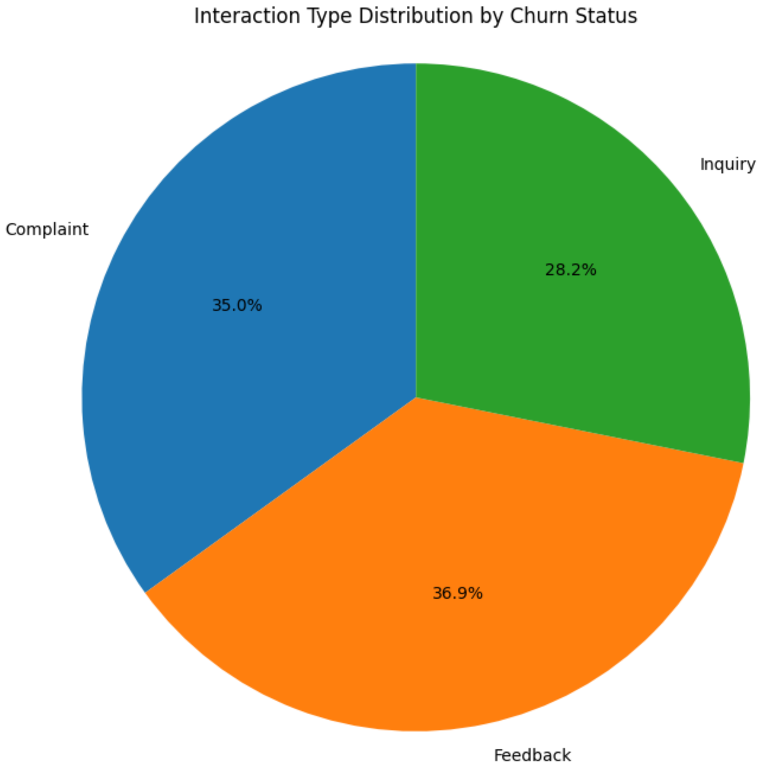

Figure 9: Interaction Type Distribution by Churn Status

Following to the resolution status, we can see that feedback and complaints are the main drivers when it comes to churning at 36.9% and 35% respectively. Whereas inquiries are less associated with churn. The customers who raise complaints or give feedback, which means that Smart Bank needs to act quickly and appropriately to reduce the likelihood of customers churning.

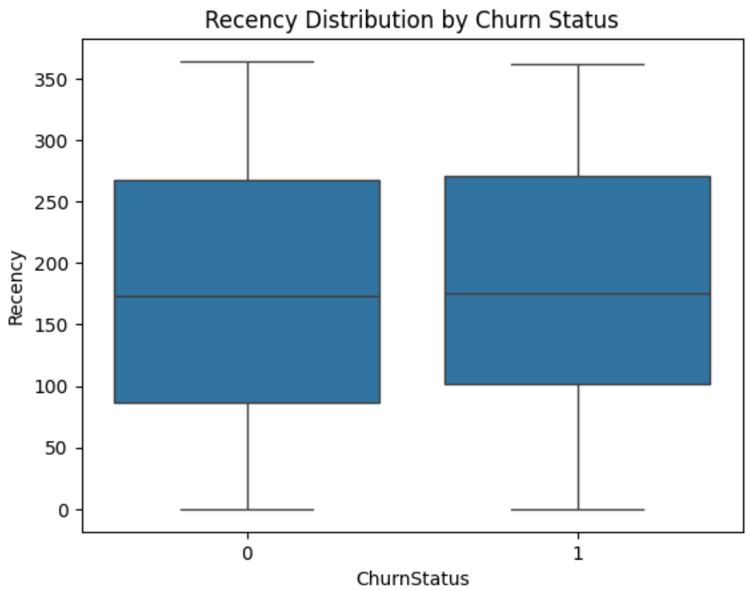

Figure 10: Recency Distribution by Churn Status

Churners have higher recency gaps (avg. 179.98 days) than non-churners (avg. 177.62). As shown in the boxplot, the median recency is 173.5, the interquartile recency is 176.25.

Question

1. Does a higher number of complaints increase the likelihood of customer churn?
    a. The null hypothesis: There is no relationship between the number of complaints and customer churn.
    b. The alternative hypothesis: A higher number of complaints is associated with a higher likelihood of churn.
2. Do unresolved customer issues increase the likelihood of customer churn?
    a. The null hypothesis: There is no difference between customer with resolved and unresolved interactions.
    b. The alternative hypothesis: Customers with unresolved issues are more likely to churn than those with resolved ones.
3. Do the types of customer interaction increase the likelihood of customer churn?
    a. The null hypothesis: Churn is equally likely across all types of customer interaction.
    b. The alternative hypothesis: Customers who submit complaints or feedback are more likely to churn than those who submit inquiries.
4. Does recency of the last login date predict churn?
    a. The null hypothesis: There is no relationship between the last login date and churn.
    b. The alternative hypothesis: Customers who haven't logged in recently are more likely to churn.

Methods

The methods that will be used are:

- Logistic regression or chi-squared test is used to check whether complaint volume significantly predict churn status.
- Decision tree, Random Forest, XGBoost, and Neural Networks for machine learning analysis.
- Convert login dates to a recency measure (days since last login) and login frequency, then model churn using logistic regression or survival analysis.
- Transaction_History variable is aggregated by CustomerID to obtain new variables: total spent, average spent, number of transactions.
- Customer_Service is aggregated by CustomerID to obtain new variables: number of interactions, resolution rate, last interaction.

Data Cleaning and Preparation

Data cleaning involved handling missing values, detecting and addressing outliers, and normalising/standardising numerical features. Categorical features were encoded using one-hot encoding to prepare the data for machine learning algorithms.

There are no missing values and outliers are shown in the data.

Model Selection and Building

The machine supervised algorithms that were considered were decision tree, random forest, XGBoost and neural network. After considering all the algorithms, the XGBoost model was selected for its balance of accuracy and interpretability.

Model Evaluation

The model's performance was evaluated using precision, recall, F1 score, and ROC-AUC metrics, after applying SMOTE and scaling to address data imbalance.

The ROC curve above shows that XGBoost achieved the highest AUC score (0.57) among all tested models, followed by the Neural Network (0.56) and Random Forest (0.55). The Logistic Regression and Decision Tree models performed near or below random guessing (AUC ≈ 0.50), which indicates their limited ability to separate churners from non-churners in this context.

While the AUC values are relatively low across all models—likely due to class imbalance and limited signal in the features—the results suggest that XGBoost has a slight edge and is better suited for detecting churn patterns in this dataset. However, further optimization is recommended to improve sensitivity to churners.

The confusion matrix for the best-performing model (XGBoost with SMOTE and scaling) reveals: 92 customers correctly identified as non-churners, 15 non-churners incorrectly flagged as churners, 23 churners missed by the model and only 4 churners correctly identified. These results show that while the model is effective at detecting non-churners (high specificity), it struggles to identify actual churners (low recall). Only 15% of churners were detected (4 out of 27), which limits its usefulness in retention-focused strategies unless further refined.

The feature importance analysis from the XGBoost model highlighted that the number of interactions, service usage, and resolution rate are the strongest predictors of churn.

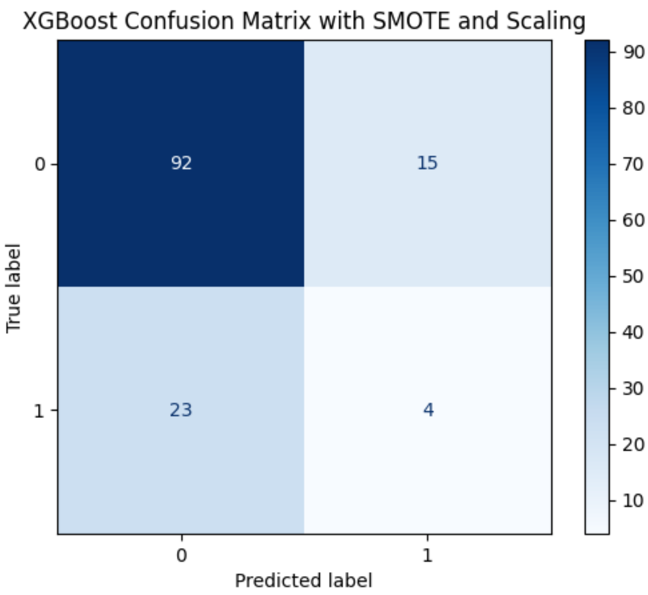

Figure 11: XGBoost Confusion Matrix

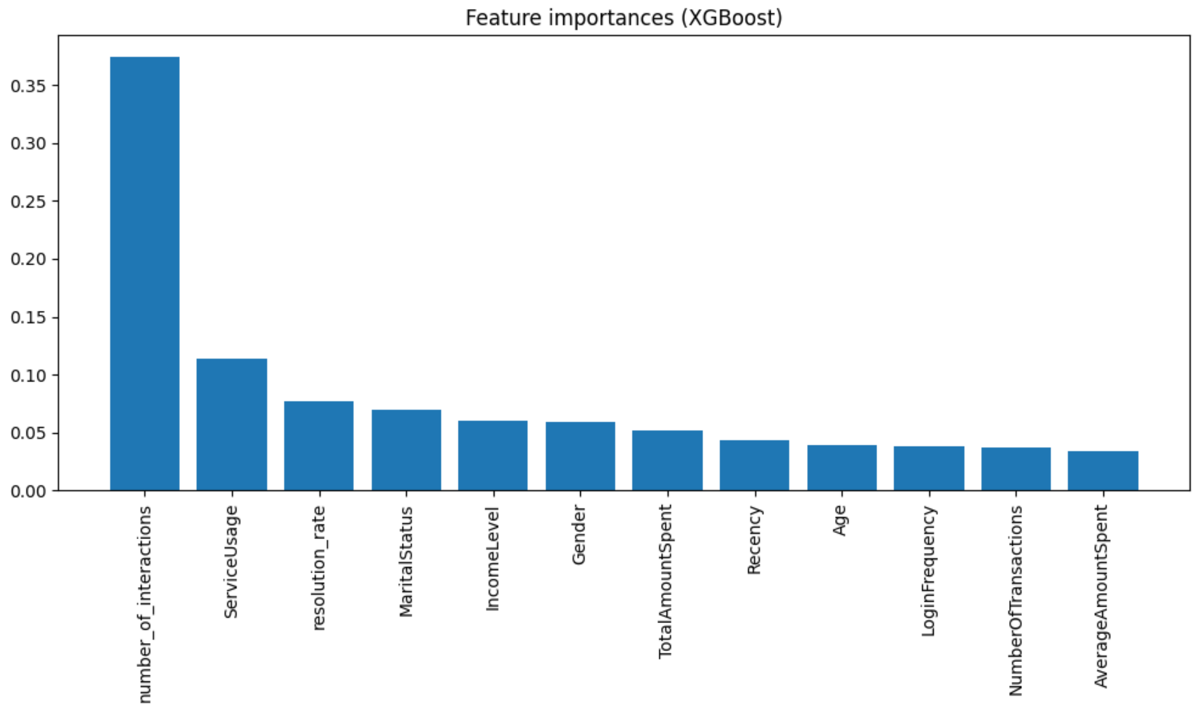

Figure 12: XGBoost Feature Importances

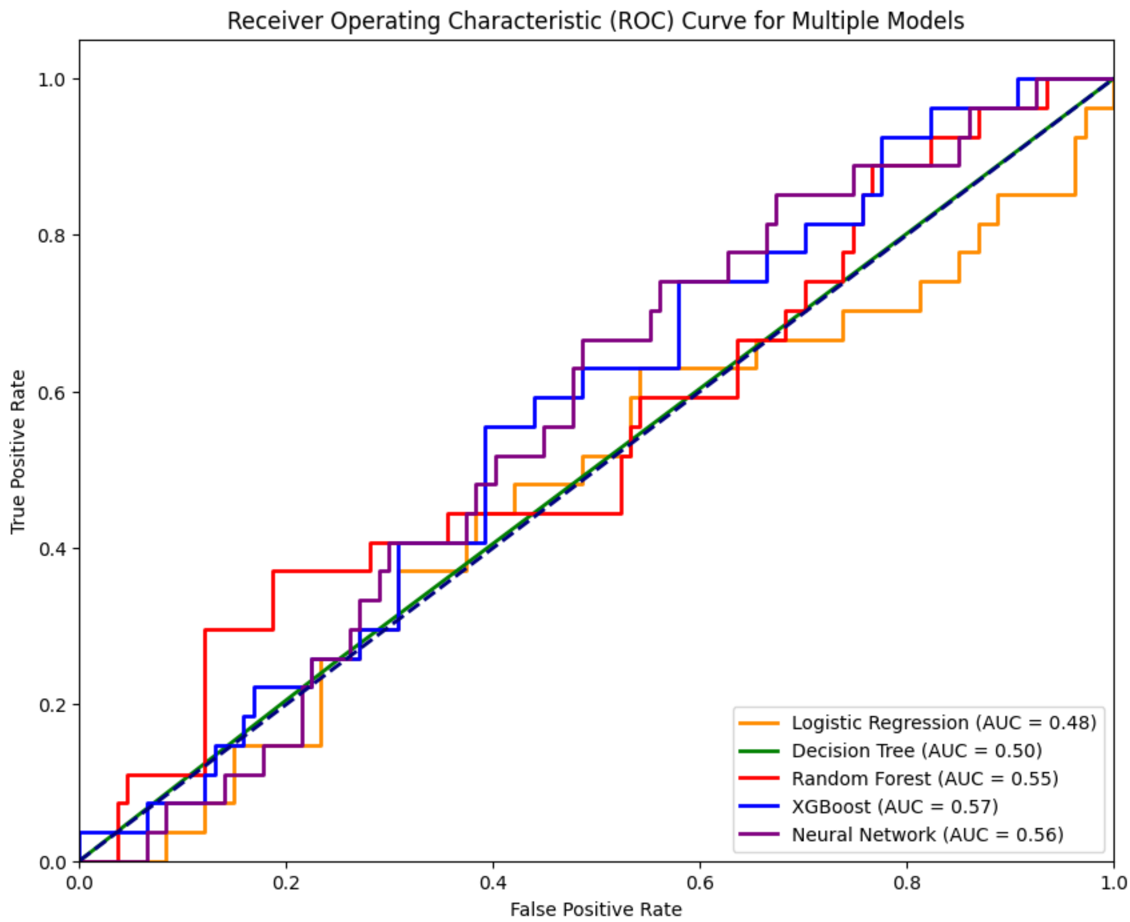

Figure 13: AUC-ROC for Multiple Models

Recommendations and Business Implications

This report's findings provide a data-driven foundation for implementing targeted, high-impact churn reduction strategies. The analysis revealed that churn is particularly prevalent among young professionals and small business owners, underscoring the importance of personalized retention strategies for these segments. For instance, SmartBank could introduce tailored financial tools such as startup loan options, tax advisory content, or digital-first budgeting tools designed for freelancers and business owners. For working-age adults, particularly those who are married, family-centric product bundles—like joint savings accounts or childcare-oriented financial incentives—could address churn linked to life-stage pressures.

The usage distribution across SmartBank's digital channels—online banking (34.9%), mobile app (34.2%), and website (30.9%)—suggests customers are accessing services across multiple platforms with no single dominant channel. This presents an opportunity to invest in a unified and consistent user experience across all interfaces. By integrating predictive analytics with customer behavior data, the bank could deliver platform-agnostic personalization, such as financial insights and personalized offers that follow the user regardless of their chosen touchpoint.

To directly address churn risk, SmartBank should deploy real-time churn alerts generated from the predictive model. These alerts can be integrated into CRM systems to trigger targeted interventions: personal follow-up from relationship managers, in-app nudges, or personalized retention campaigns. Early identification and outreach to at-risk customers can significantly reduce the probability of churn, especially when combined with incentives such as loyalty rewards or financial wellness consultations.

Complaint resolution emerged as a critical churn driver. Customers with unresolved complaints were significantly more likely to leave the bank. This calls for the launch of a dedicated complaint-to-loyalty pipeline—an initiative that ensures complaints are resolved swiftly and followed by customer-centric recovery strategies. For example, customers who submitted complaints could receive resolution summaries, apology notes, and small-value compensations (e.g., waived service fees or cashback) to rebuild trust and loyalty.

Another key finding was the impact of digital recency. Churners had longer recency days since their last login, with a median of 173.5 days. SmartBank should act proactively by implementing recency-based reactivation campaigns. Customers inactive for 90 days or more could receive personalized messages reminding them of product benefits, limited time offers to re-engage, or tutorials highlighting underused features of their accounts.

In addition, the data shows that lower-income earners experience the highest churn rates. This indicates a need to design affordability-sensitive offerings. These could include simplified, fee-free accounts, personalized debt management tools, and access to budget coaching—all of which signal empathy toward financial vulnerability and build long-term loyalty among this segment.

Finally, relationship marketing should become a central pillar of SmartBank's customer engagement strategy. Regular, value-driven communication through email and app notifications should reinforce the product's daily relevance. Examples include financial health summaries, personalized savings milestones, or success stories from customers with similar profiles. These nudges build emotional loyalty by embedding SmartBank into customers' financial journeys.

To ensure accountability and continuous improvement, the bank should track performance metrics such as churn reduction by segment, customer satisfaction scores, complaint resolution times, and retention rates of previously at-risk users. Through these integrated strategies—grounded in behavioral data and advanced analytics—SmartBank can significantly reduce churn while enhancing customer lifetime value and brand trust.