Customer Churn Analysis

Leveraging advanced predictive models to identify at-risk customers and drive targeted retention strategies

Team Members: Sarah Joseph, Hashmat Subazali, Mohammed Abdulaziz, & Charles Amadi

Date: 12/3/2025

Agenda

Business Problem & Objective

Defining the challenge and desired outcomes.

Data & Methodology

Collecting and structuring relevant data.

EDA Insights

Exploring data to uncover patterns and trends.

Feature Engineering

Transforming data for model readiness.

Modeling & Evaluation

Training and Evaluating machine learning models.

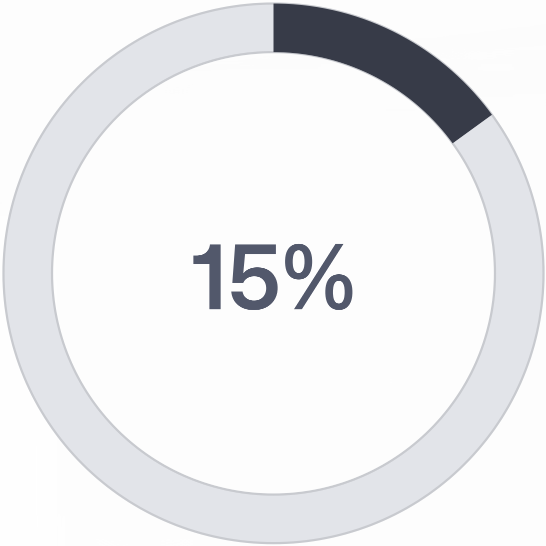
Business Impact & Recommendations

Translating findings into actionable strategies.

Challenge: Churn Threatens Revenue Growth

Business Problem & Objective

- LLOYDS Banking Group was experiencing high churn among personal and small business customers and don't know who was at risk of leaving and why. Our mission is to move them from a reactive 'goodbye' to a proactive 'how can we help you stay?
- **Objective:** Build a reliable classification model to identify customers at-risk churning and recommend interventions to reduce attrition.



Profit Reduction

Impact of churn on overall company profitability.

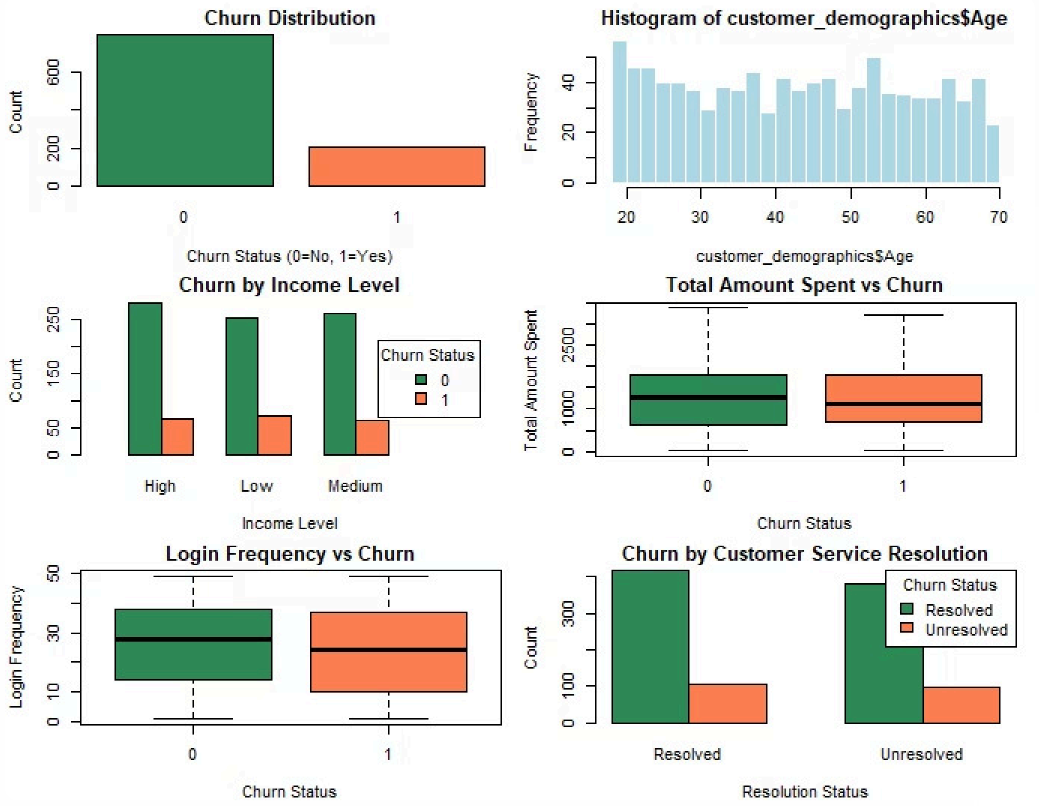
Data & Methodology

Data Sources

5 data sources consolidated into one master dataset

- Demographics
- Transactions
- Service
- Activity
- Churn Labels

EDA Insights



Key Findings from Exploratory Data Analysis

- **Churn Distribution:** Dataset is **imbalanced** (more non-churners), which will require balancing strategies (SMOTE, class weights, etc.) in modeling.
- **Customer Demographics:** Customers are fairly evenly distributed across ages 18 – 69, with no obvious skew, suggesting demographic-specific engagement strategies.
- **Income & Churn:** Churn is present across all income levels. Slightly higher churn counts in Low income groups (Which could suggest income sensitivity plays a role in churn).
- **Spending Habits:** Total amount spent does not serve as a strong differentiator between churned and retained customers, implying other factors are more influential.
- **Login Frequency:** Churners show lower median login frequency compared to retained customers. It suggests engagement is a strong predictor of retention. etc.

We created features to capture churn behavior

CustomerValueScore

High-value customers deserve different retention strategies

EngagementScore

Recent and frequent engagement indicates loyalty

SupportEfficiency

Poor support experiences often lead to churn

HasUnresolvedIssues

Binary flag for pending problems. Unresolved issues are strong churn predictors

TransactionFrequency

Measuring intensity of platform usage

ProductDiversityScore

Customers using multiple products are stickier

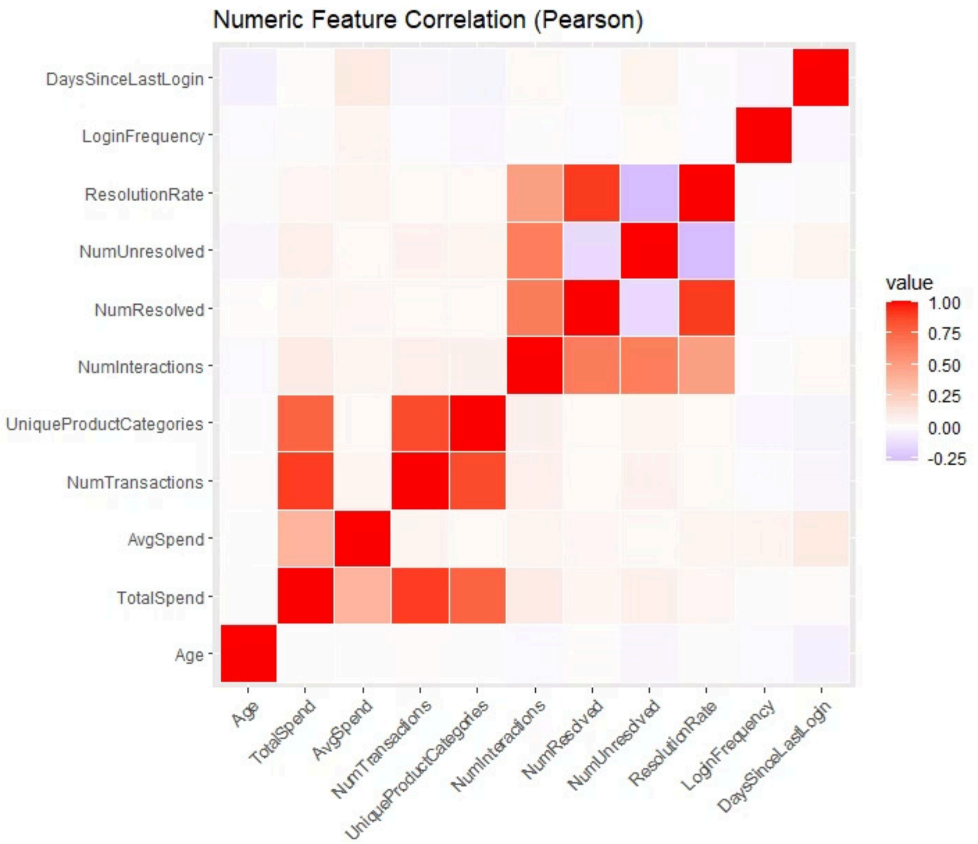
Activity Level

categorical recency segmentation. Different interventions for different activity levels

AgeGroup

The Age feature was binned into categories (18-24, 25-34, 35-44, 45-54, 55-64, 65+) to capture non-linear relationships

Correlation Analysis With Key Variables Insight



Key Correlation Insights & Multicollinearity

Highly correlated features can cause model instability & make interpretation difficult

- **TotalSpend**, **AvgSpend** and **NumTransactions** etc. To avoid redundancy, TotalSpend was dropped in favor of the more granular AvgSpend & NumTransactions features etc.
- **Customers** on long-term contracts typically exhibit lower churn rates, highlighting the importance of tailored contract incentives.
- **Technical Support Interactions:** Frequent technical support calls, particularly unresolved issues, show a clear link to increased churn risk.

Our Approach: Five Models Trained & Evaluated

We developed & tested five distinct machine learning algorithms to find the most effective predictor of customer churn. Each model was rigorously trained, validated, and tested on unseen data.

Logistic Regression

Classic statistical approach providing baseline performance

Random Forest

Ensemble method combining multiple decision trees

Gradient Boosting

Advanced boosting technique for sequential learning

XGBoost

Optimized gradient boosting with regularization

Neural Network

Deep learning model capturing complex patterns

Model Performance Comparison

After comprehensive evaluation across multiple metrics, we identified clear performance differences. The Neural Network emerged as our champion model due to its superior recall - the ability to correctly identify customers who will actually churn.

Model	Accuracy	Precision	Recall	Specificity	F1	ROC-AUC
Logistic Regression	0.513	0.223	0.575	0.497	0.322	0.492
Random Forest	0.779	0.250	0.050	0.962	0.083	0.552
GBM	0.794	0.400	0.050	0.981	0.089	0.489
XGBoost	0.749	0.222	0.100	0.912	0.138	0.502
Neural Network	0.432	0.208	0.650	0.377	0.315	0.522

The Winning Model: Neural Network

65%

Recall Rate

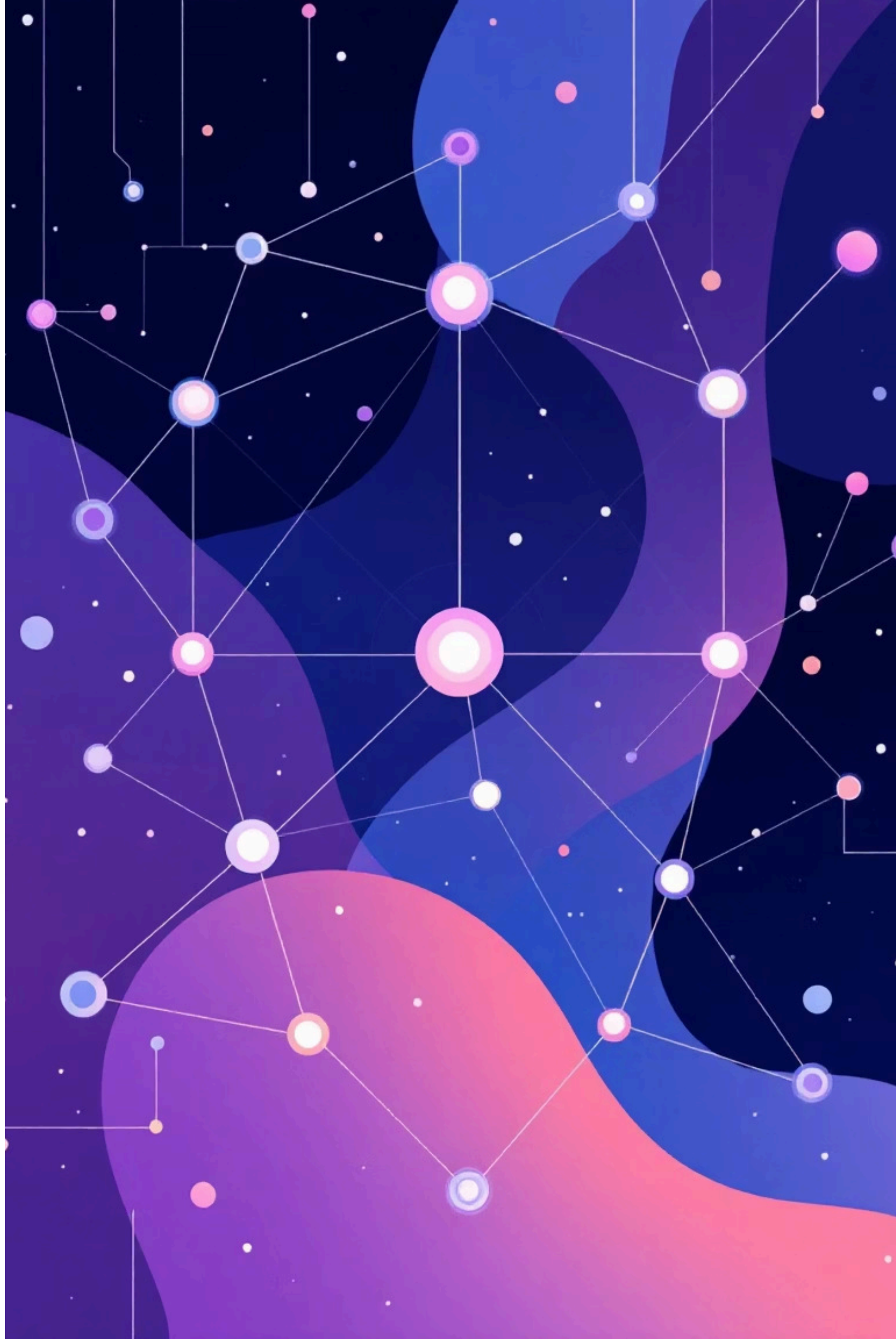
Successfully identifies 65 out of 100 customers who will churn

0.52

ROC-AUC Score

Demonstrates model's predictive power or ability to distinguish churners from non-churners

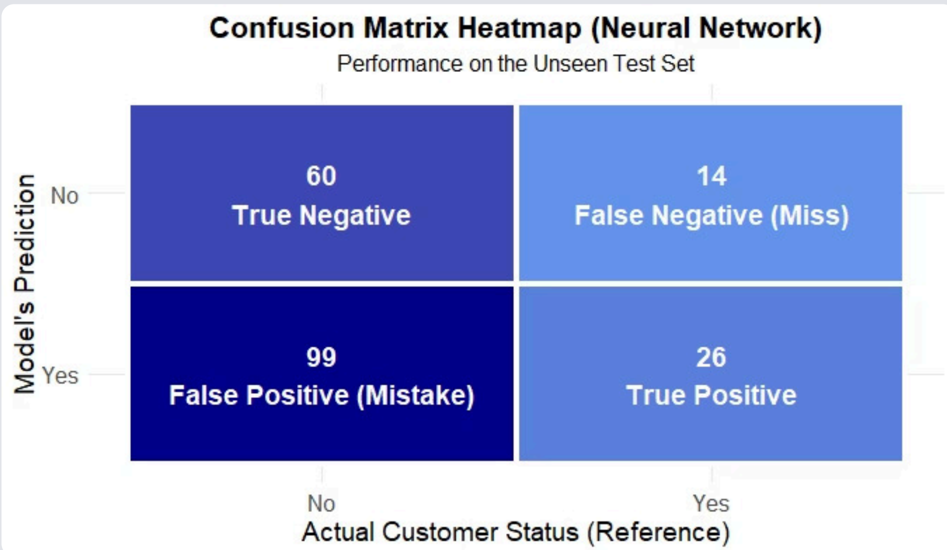
While the ROC-AUC score indicates room for improvement, the high recall rate means we can proactively reach the majority of at-risk customers before they leave.



Understanding Model Predictions

Confusion Matrix Analysis

Our model's predictions on the test set reveal both strengths and limitations:

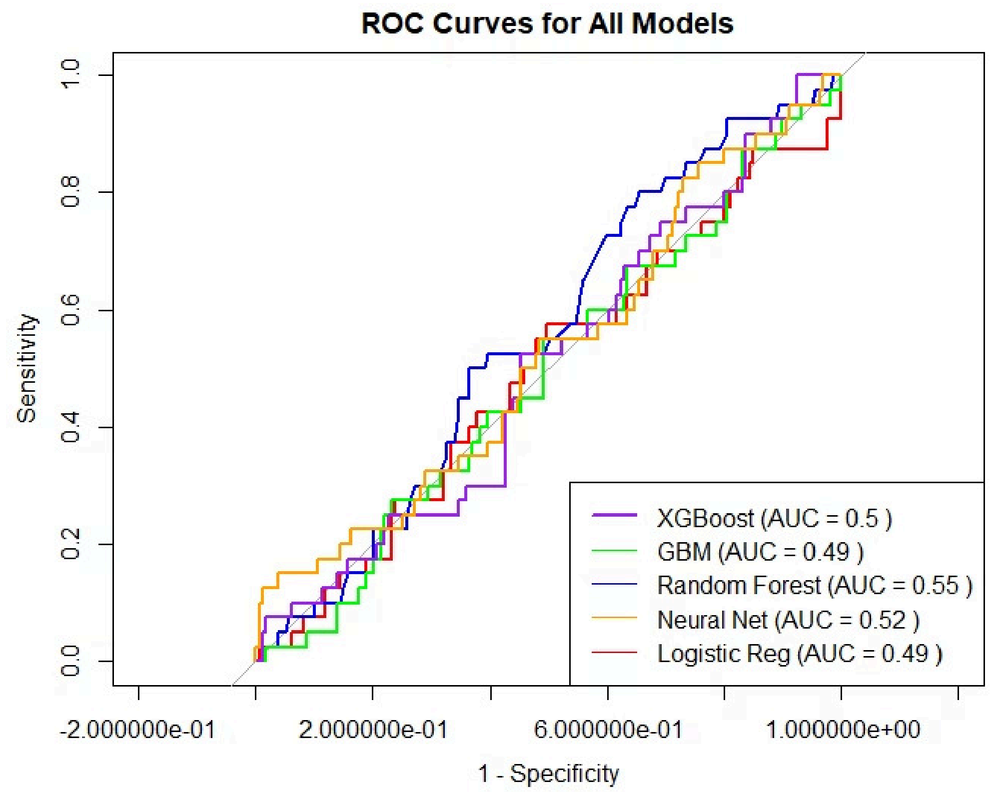


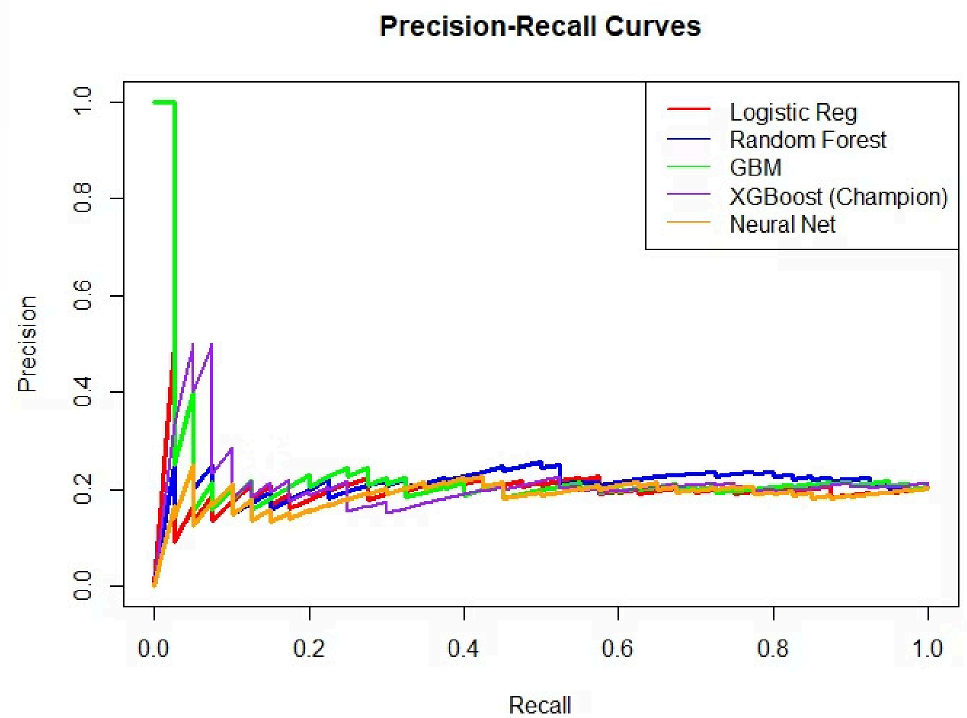
What This Means

- The model correctly catches 26 churning customers while generating 99 false alarms. This trade-off may be acceptable depending on retention cost. The 99 **False positives** are less risky, but still impact retention budget and targeting efficiency
- The 14 missed churners (false negative) represent our biggest opportunity for improvement in future model iterations. These are customers who churned but weren't flagged.

ROC & PR Curves

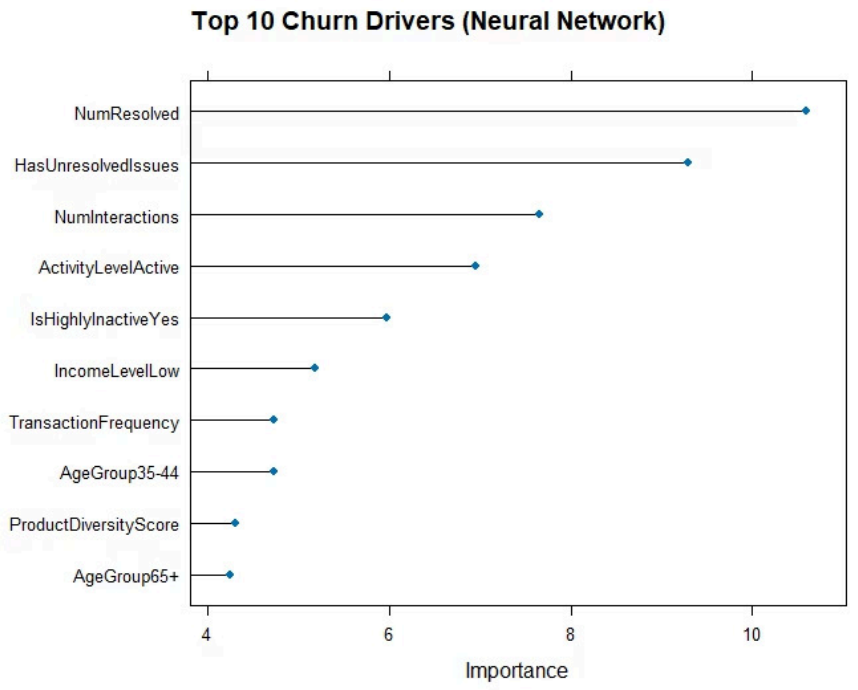
These curves illustrates the model's ability to distinguish between churning and non-churning customers across all probability thresholds.





Key Drivers of Customer Churn & Recommended Intervention Strategy

Key Drivers of Customer Churn



Focus on High-Risk Customers

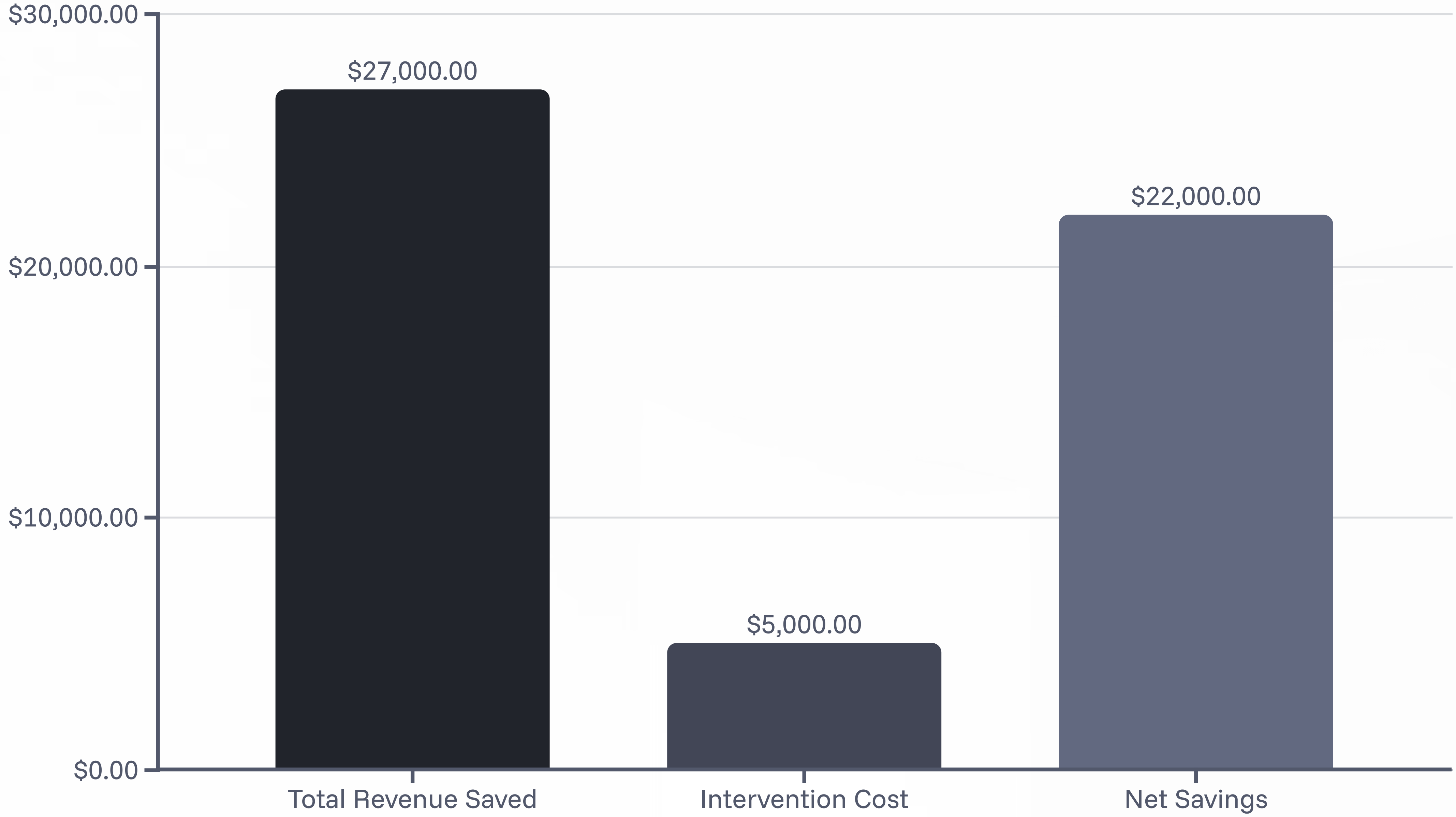
We recommend implementing targeted retention campaigns for the **top 10% highest-risk customers** identified by our model.

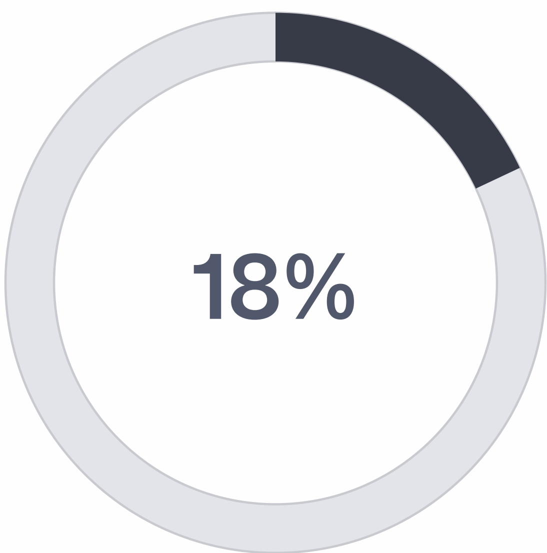
Interventions should address the primary churn drivers:

- Prioritized customer service outreach for unresolved issues
- Personalized engagement campaigns to reignite activity
- Proactive account reviews and value demonstrations
- Exclusive offers or incentives tailored to individual needs

Projected Financial Impact

A conservative ROI analysis demonstrates the business value of our predictive model. By investing in targeted interventions, we can achieve substantial returns through retained customer revenue.





Customer Save Rate

18 out of 100 targeted customers retained

Return on Investment

Every dollar invested returns $4.40 in retained revenue