SyriaTel Customer Churn: Analysis & Actionable Insights

A Data-Driven Approach to Reducing Customer Attrition

Author: Rose Miriti

Date: July 21, 2025

The Problem: Customer Churn is a Drain on Revenue

Customer churn directly impacts SyriaTel's bottom line.

Reduced Revenue

Every customer who leaves lowers our Average Revenue Per User (ARPU).

Increased Costs

Acquiring new customers is far more expensive than retaining existing ones.

Competitive Market

In a market with low switching barriers, even a small increase in churn significantly hurts profitability.

Goal:

Use data to predict which customers are likely to churn, allowing for proactive and targeted retention efforts.

Our Approach: From Data to Decision

Key Points:

Data Exploration

Analyzed a dataset of 3,333 customers, examining attributes like tenure, service usage, plan details, and location.

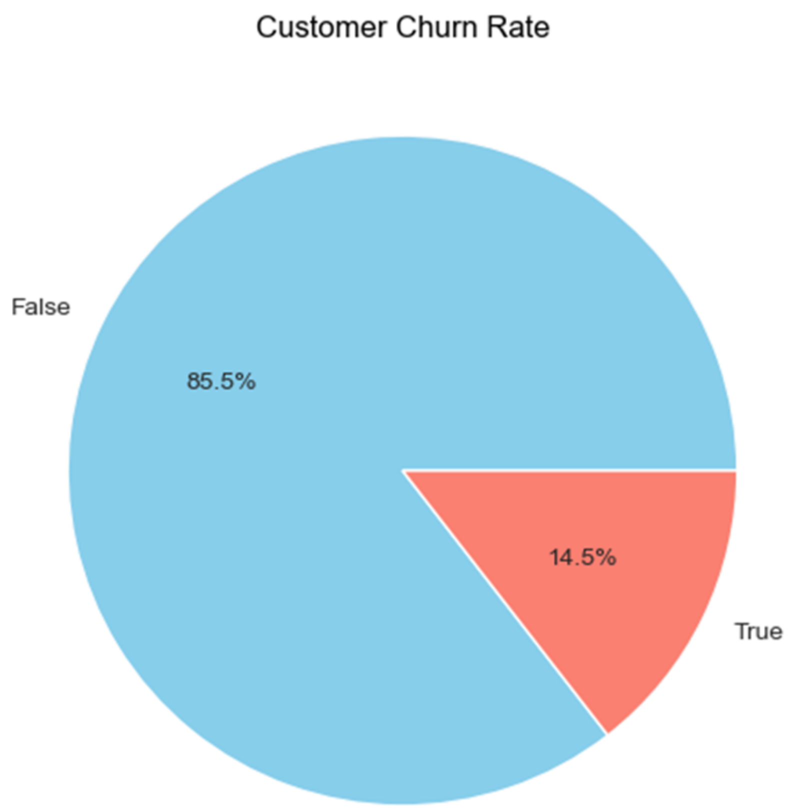
Predictive Modeling

Built and compared several machine learning models (Logistic Regression, Random Forest, Gradient Boosting) to find the most accurate predictor of churn.

Insight Generation

Identified the key drivers of churn to inform actionable business strategies for the Customer Retention, Marketing, and Finance teams.

A 14.5% Churn Rate: An Opportunity for Improvement



14.5%

Churn Rate

While this may seem like a small small percentage, it represents a represents a significant portion portion of our customer base. base.

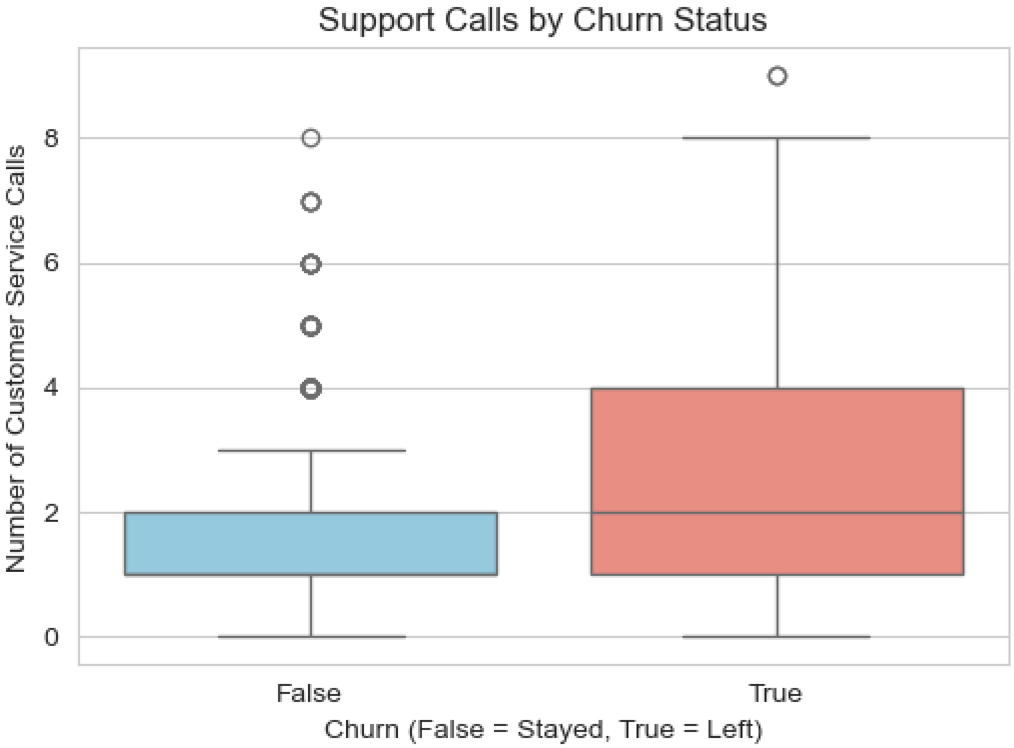
483

Customers Lost

Out of 3,333 customers in our dataset.

This is a substantial number of customers to lose, highlighting a significant opportunity to improve retention and preserve revenue.

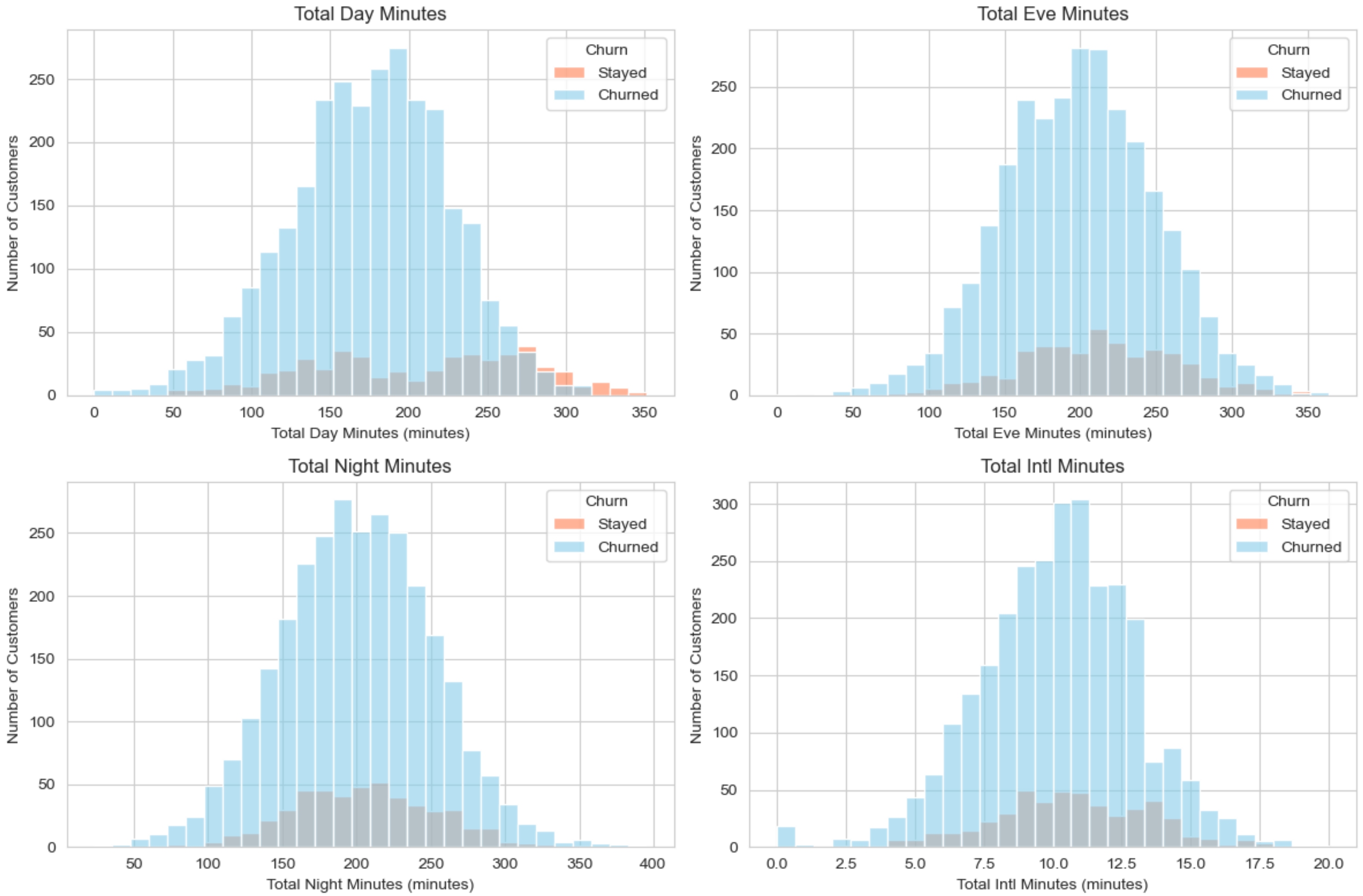
Does calling customer service signal higher churn risk?



Customers who contact customer service more frequently are **significantly more likely to churn**.

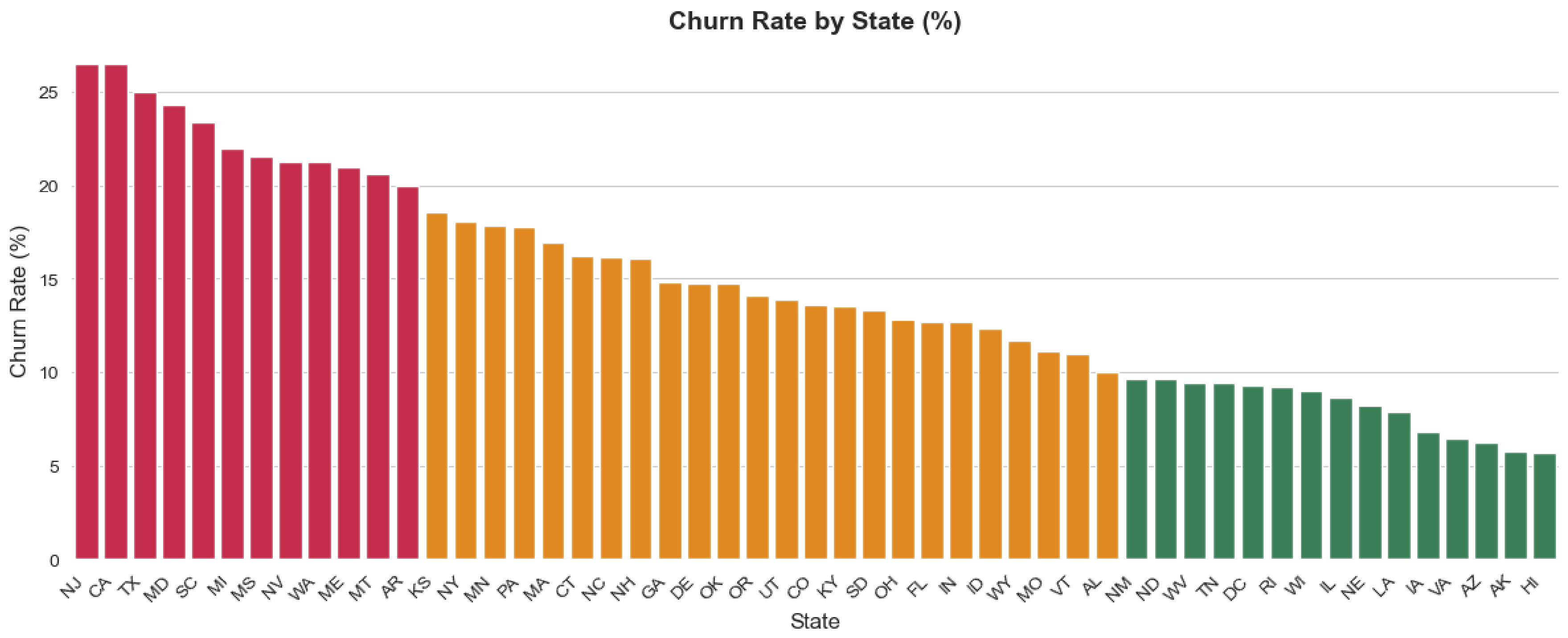
Frequent support interactions are a clear early warning signal of customer dissatisfaction.

How does plan usage relate to churn?



- ➢ Customers who use their plan's included minutes less are more likely to leave.
- ➢ This suggests that customers who don't feel they are getting value from their plan are at a higher risk of churning.

Are certain regions more at risk?



- **High Churn "Hotspots":** New Jersey and California have the highest churn rates, both around **26%**. Texas, Maryland, and South Carolina also show churn rates above 23%.
- **Low Churn States:** Hawaii and Alaska have the lowest churn rates, around 6%.
- This indicates a clear need for localized retention strategies.

Our Baseline Model: Logistic Logistic Regression

We started with a simple, interpretable model Logistic Regression to establish a performance performance baseline.

It achieved a Recall of 70%, meaning it found 7 out of 10 customers who were about to churn. to churn.

However, its Precision was only 33%. This means that for every 3 customers it flagged as a churn risk, 2 were actually loyal actually loyal customers (false positives), which would lead to inefficient sp

70%

Recall

We can correctly identify 70% of customers who are about to churn.

33%

Precision

When our model predicts predicts a customer will will churn, it is correct 33% of the time.

0.80

ROC-AUC

This score indicates **good separability**. It means there's an 80% chance the model will rank a random churner higher than a random stayer.

Where the Baseline Model Succeeds and Fails

This chart breaks down the model's predictions, showing us exactly where it was right and where it was wrong.

What it got right:

•Correctly identified 70% of customers who actually churned (True Positives).

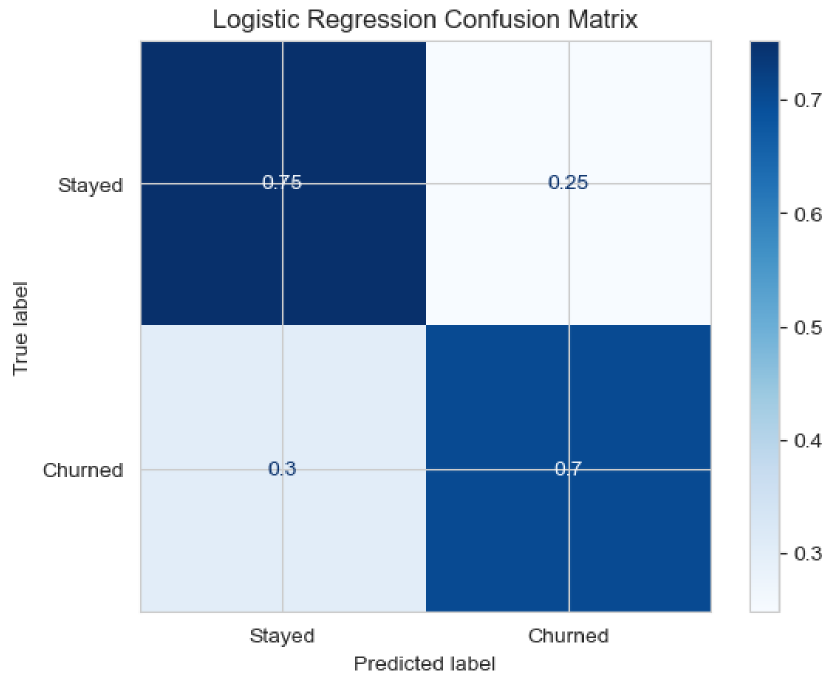
•Correctly identified 75% of loyal customers who stayed (True Negatives).

Where it struggled:

•Missed Opportunities (30%): Failed to identify 30% of customers about to leave (False Negatives), representing significant lost revenue.

•Wasted Effort (25%): Incorrectly flagged 25% of loyal customers as at risk (False Positives), leading to inefficient spending.

The baseline model is a decent start, but it misses too many at-risk customers and creates too many false alarms, highlighting the need for a more powerful and precise model.



An Advanced Approach - The Random Forest Model

We then tested a Random Forest model to see if we could improve the precision of our predictions and reduce false alarms.

64%

Recall

We can correctly identify 64% of customers who are about to churn.

73%

Precision

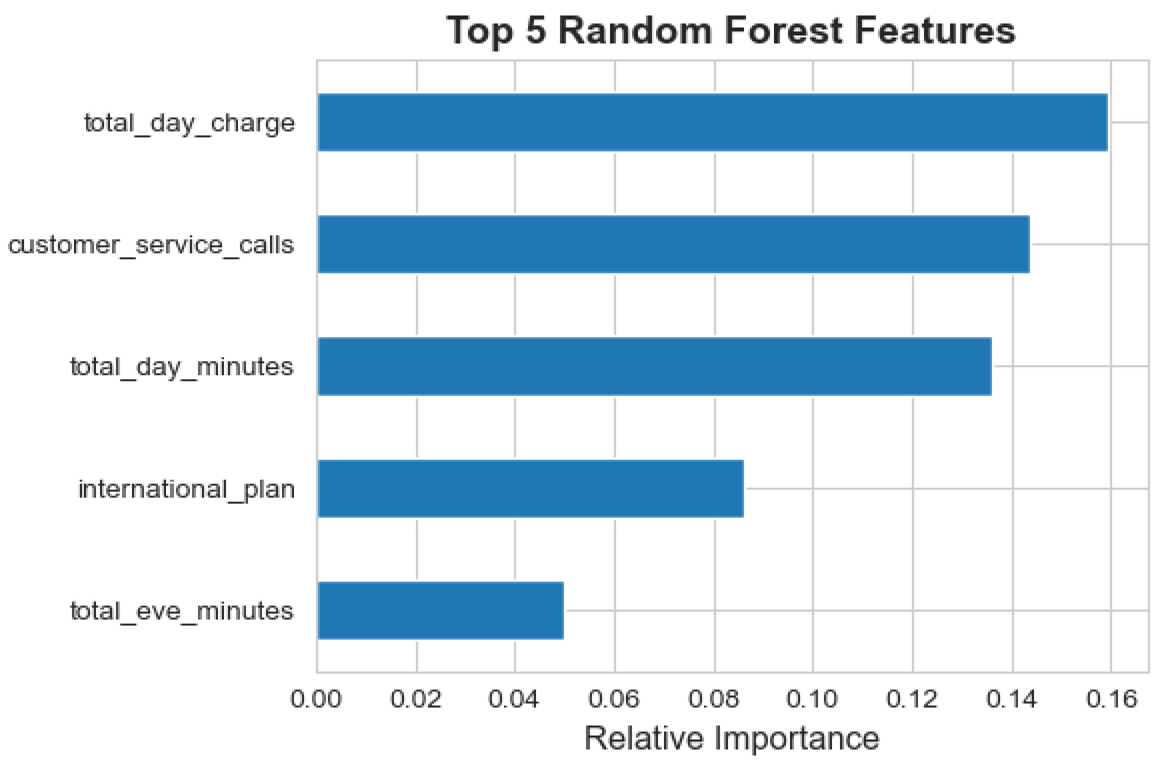
When our model predicts a customer will churn, it is correct 73% of the time.

0.89

ROC-AUC

The model is highly effective at distinguishing between churners and stayers.

What the Random Forest Model Focuses On



1. **Total Day Charge:** The single most important factor. This confirms that billing and cost are primary drivers of a customer's decision to leave.
2. **Customer Service Calls:** Reinforces our earlier finding that frequent support interactions are a major red flag.
3. **Total Day Minutes:** Similar to day charges, this shows that daytime usage patterns are a critical indicator of customer engagement and churn risk.
4. **International Plan:** The model learned that simply being enrolled in this plan is a significant risk factor.
5. **Total Eve Minutes:** The model also considers evening usage, indicating that a customer's overall engagement pattern is important.

Our Predictive Tool: The Gradient Boosting Model

We selected the **Gradient Boosting model** as our final choice due to its superior superior performance in identifying potential churners.

78%

Recall

We can correctly identify identify 78% of customers who are about about to churn.

80%

Precision

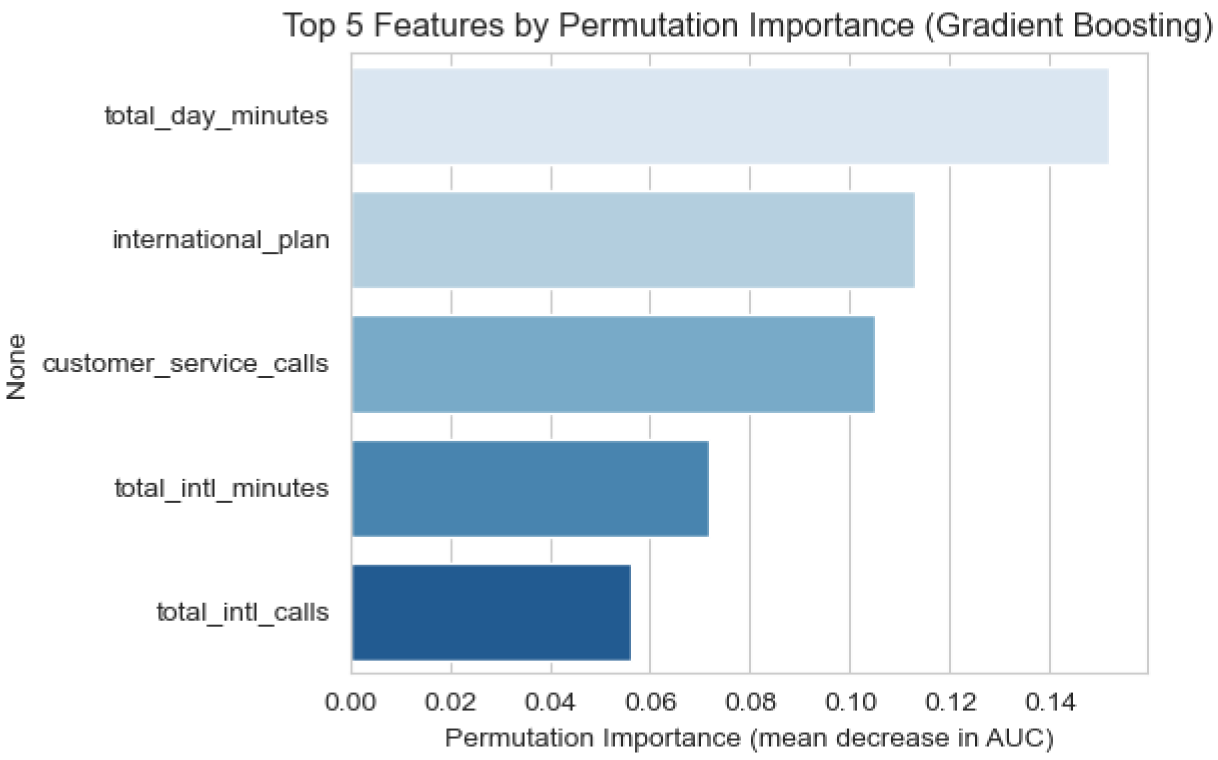
When our model predicts a customer will churn, it is correct 80% of the time.

0.90

ROC-AUC

The model is highly effective at distinguishing distinguishing between between churners and

What Matters Most? Top 5 Churn Predictors



1. **Total Day Minutes** (How much customers use their phone during the day is the strongest signal).

2. **International Plan** (Customers with this plan are more likely to churn).

3. **Customer Service Calls** (The number of calls to support is a major indicator).

4. **Total International Minutes** (The amount of time spent on international calls).

5. **Total International Calls** (The number of international calls made).

Choosing the Right Tool: Model Performance

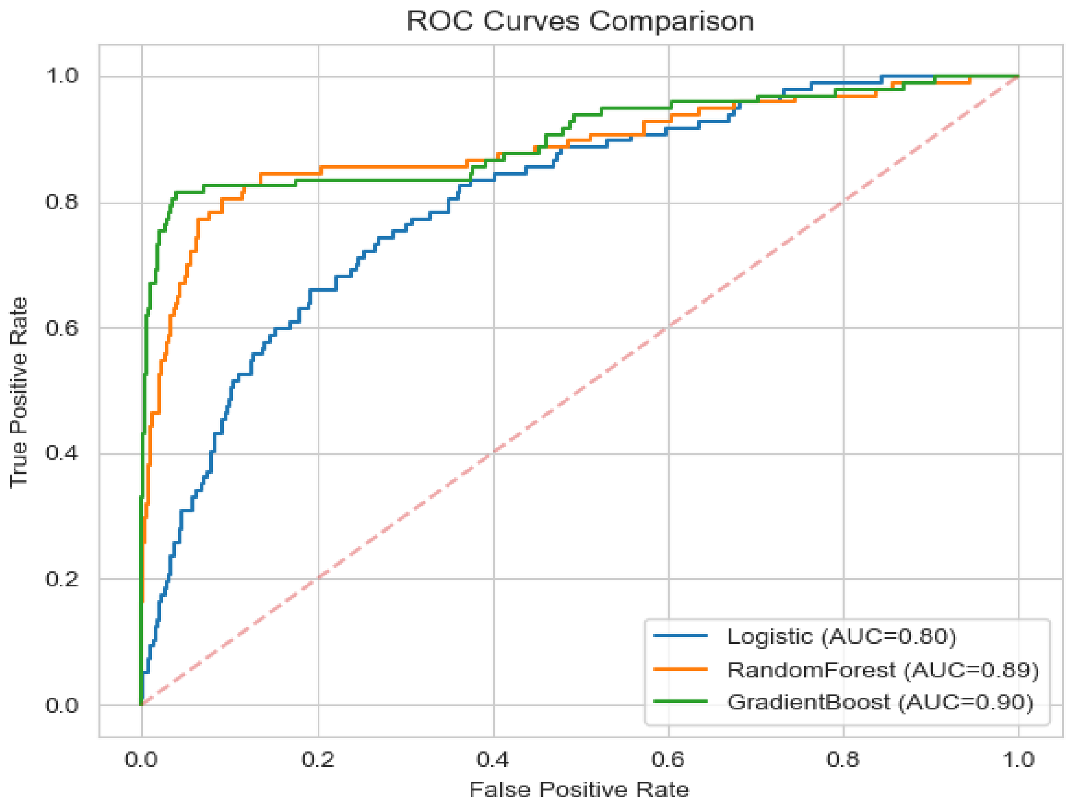
- When compared directly, the **Gradient Boosting** model was the clear winner.
- It provided the best balance of identifying the most churners (highest Recall) while being the most accurate in its predictions (highest Precision).

Model	Recall	Precision	F1-Score	ROC-AUC
Logistic Regression	70%	33%	0.44	0.80
Random Forest	64%	73%	0.68	0.89
Gradient Boosting	78%	80%	0.79	0.90

Visualizing Performance: ROC Curves

- The ROC curve shows how well each model distinguishes between customers who will churn and those who will stay. A model is better if its curve is closer to the top-left corner. compared directly, the **Gradient Boosting** model was the clear winner.

- **Key Takeaway:** The **Gradient Boosting (green line)** curve is highest, confirming it is the most accurate and reliable model for our goal.



Our Strategic Recommendations

Based on our analysis insights, we propose four key actionable strategies to enhance customer retention.

Flag High-Touch Customers

Implement automated alerts for the Retention Team when a customer logs their second service call, enabling proactive support and intervention.

Engage Under-Utilizing Users

Launch a "plan right-sizing" campaign for customers with the lowest usage to increase their perceived value and prevent dissatisfaction.

Focus on Geographic Hotspots

Allocate marketing budget for targeted retention campaigns in high-churn states like New Jersey, California, and Texas.

Review International Plan

The Marketing Department should immediately reassess the pricing and value proposition of the international plan, as it is a primary driver of churn.

Next Steps & Future Work

While our current model is highly effective, continuous improvement is key to sustained success in customer retention.

Continuous Monitoring & Maintenance

Track model performance quarterly and and retrain as needed to adapt to changing customer behaviors and market market dynamics.

Advanced Feature Engineering Engineering

Create more powerful predictive features features (like usage ratios and interaction interaction terms) to uncover new high-high-risk segments and improve model model accuracy.

Explore State-of-the-Art Models Models

In future iterations, test specialized libraries like XGBoost and LightGBM to potentially achieve even higher accuracy and predictive power.