

Predicting Customer Churn for SyriaTel

Developing a machine learning classifier to identify customers likely to churn,
enabling proactive retention strategies and mitigating revenue loss.



Business Imperative: Retain, Don't Acquire



Profitability

Customer retention is more profitable than acquisition in a saturated telecom market.



Churn Threat

High churn rates directly threaten SyriaTel's long-term profitability.



Early Warning System

Leveraging historical data to build a predictive system for customer departures.

Addressing the Churn Problem

Problem Statement

SyriaTel loses customers without identifying specific triggers until it's too late. Lacks data-driven methods to identify high-risk customers, prioritize retention, and understand pain points.

Primary Objective

Develop a binary classification model to predict customer churn (True/False).

Success Metrics & Actionable Insights

Primary Metric: Recall >80%

Maximize "capture rate" due to high cost of customer loss vs. promotional incentives.



Secondary Metric: Precision

Ensure marketing budget ROI by not wasting efforts on loyal customers.

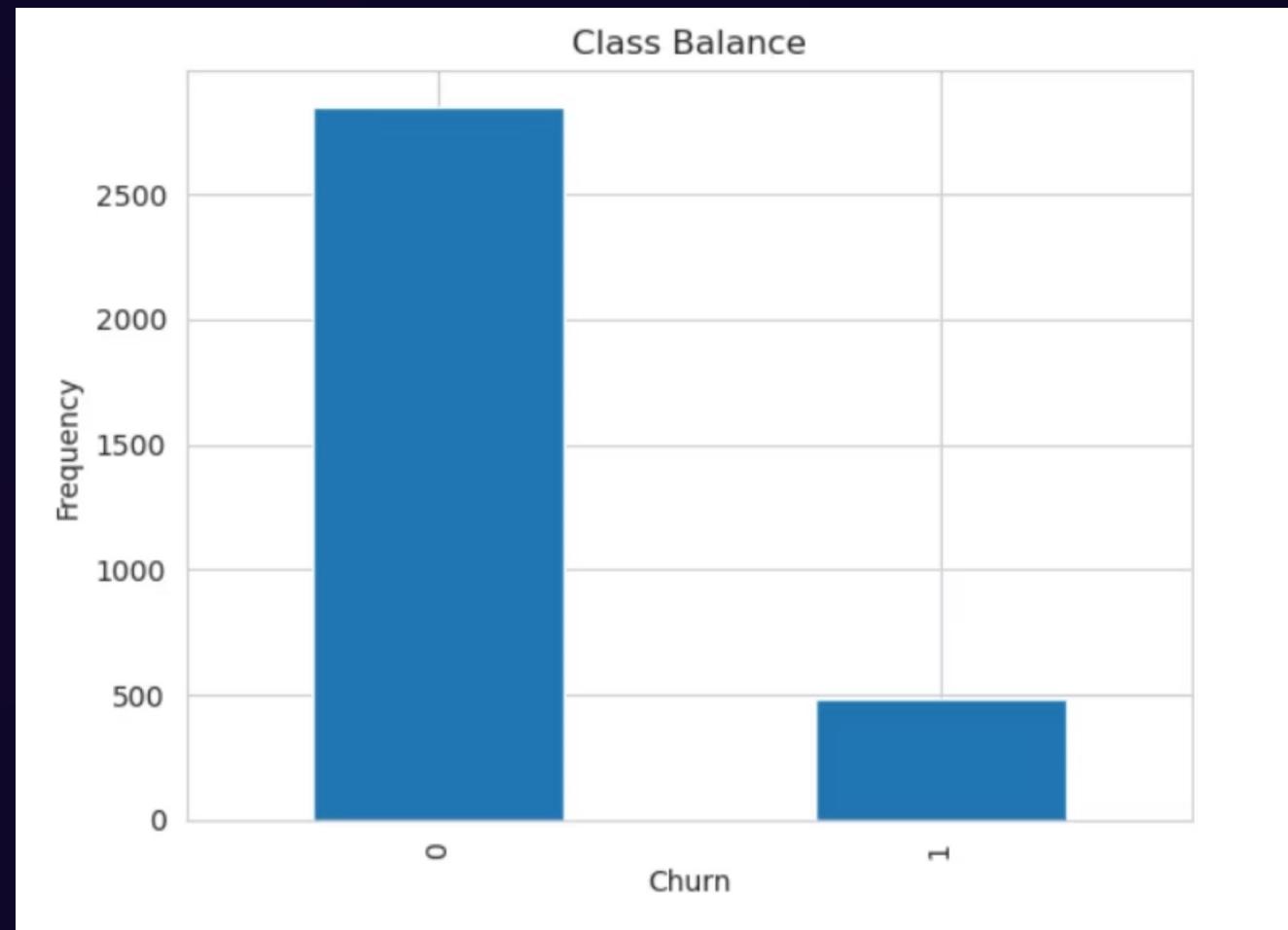
Actionable Insight

Identify top 3-5 features predicting churn to address root causes (e.g., rates, customer service).

Data Overview & Preparation

The SyriaTel dataset contains 3,333 clean records with no missing values. Features include account profiles, usage metrics (minutes, charges), and customer service calls. The target variable, churn, is highly imbalanced, with only 14.5% of customers having churned.

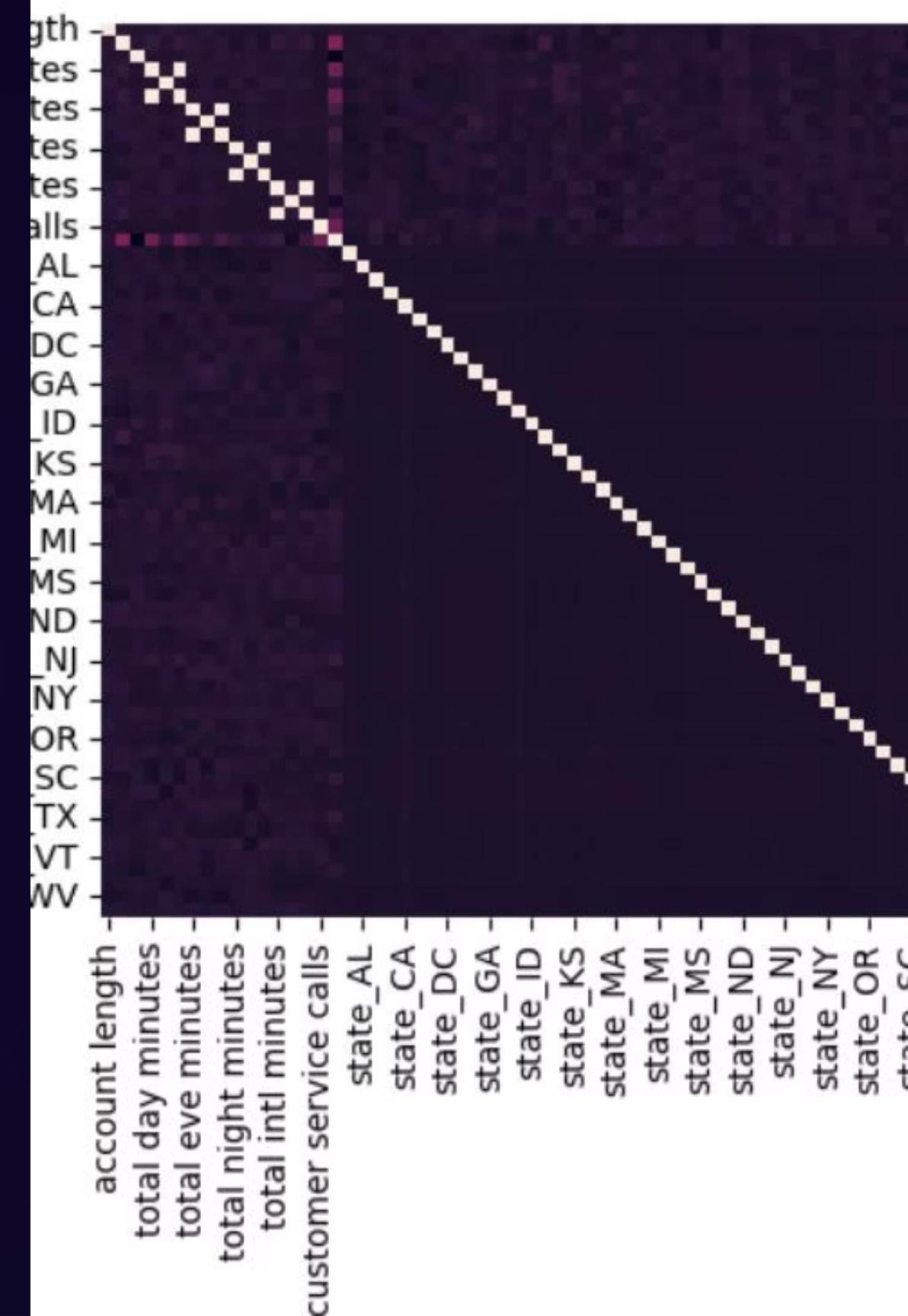
Data preparation involves dropping redundant features (e.g., phone number), removing highly correlated variables, scaling numerical data, and encoding categorical variables to optimize the dataset for modeling.



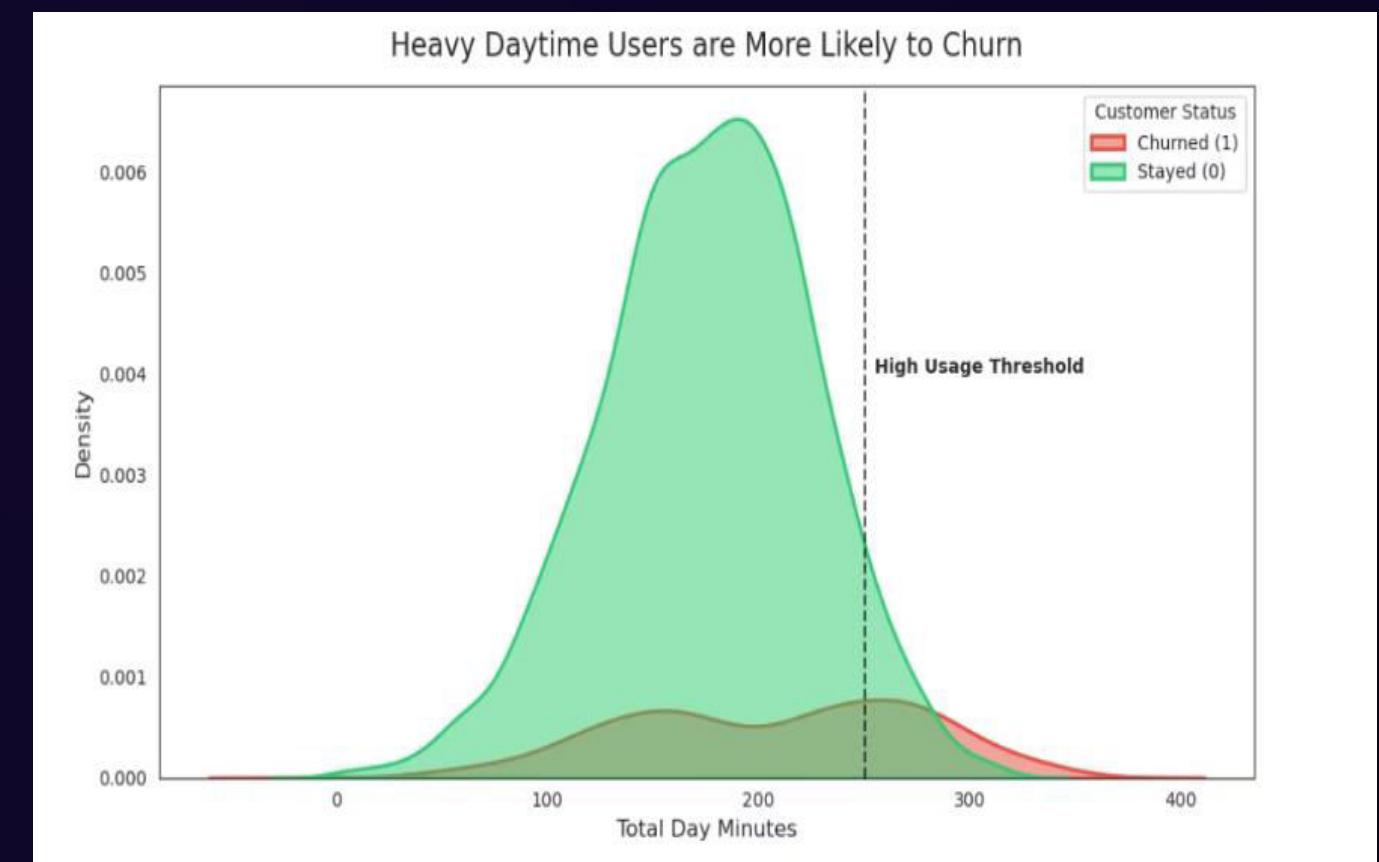
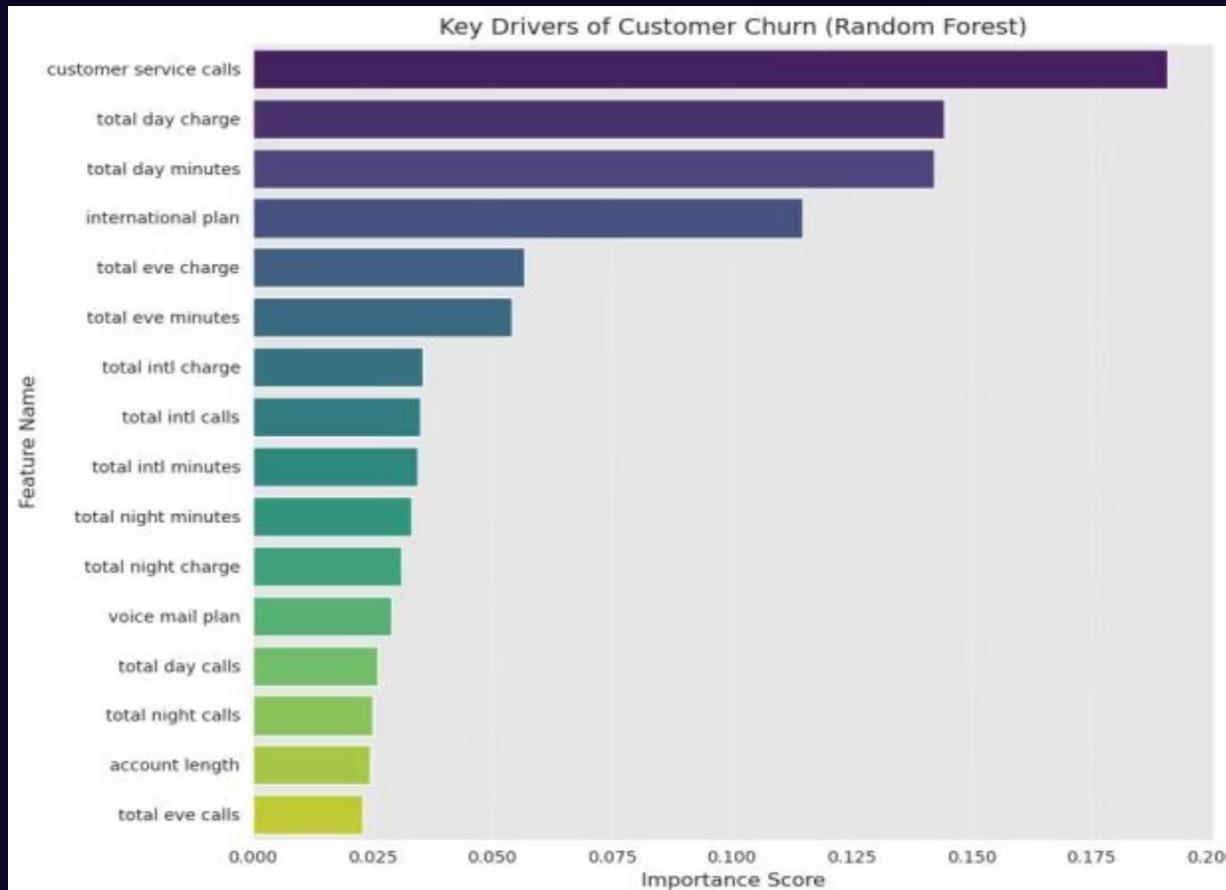
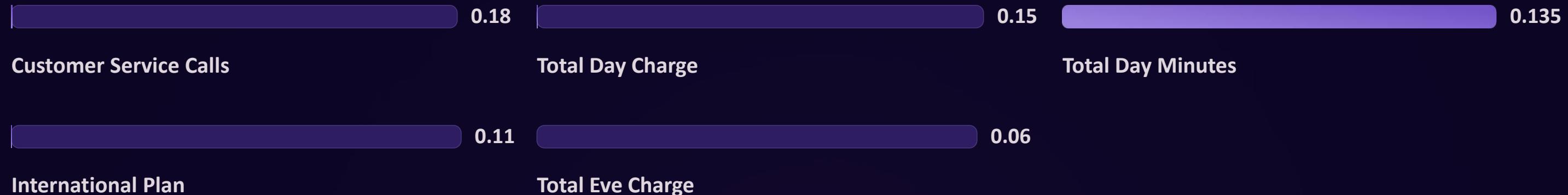
Feature Correlation Analysis

Identifying the strongest positive and negative correlations helps pinpoint "red flags" for churn and remove redundant features. The heatmap below illustrates these relationships.

Individual states show very low correlation with churn, justifying their removal to avoid high dimensionality and overfitting.

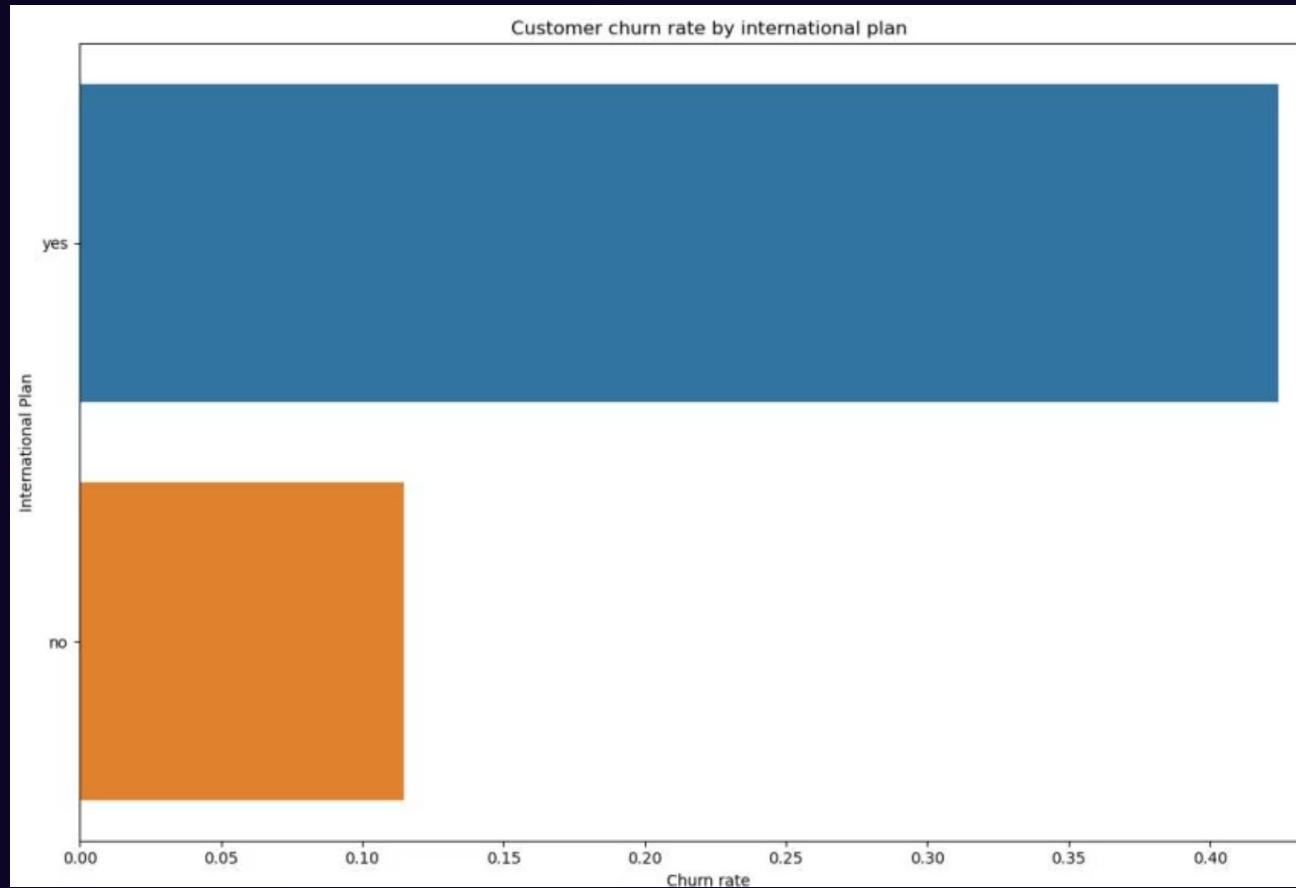


Key Drivers of Customer Churn



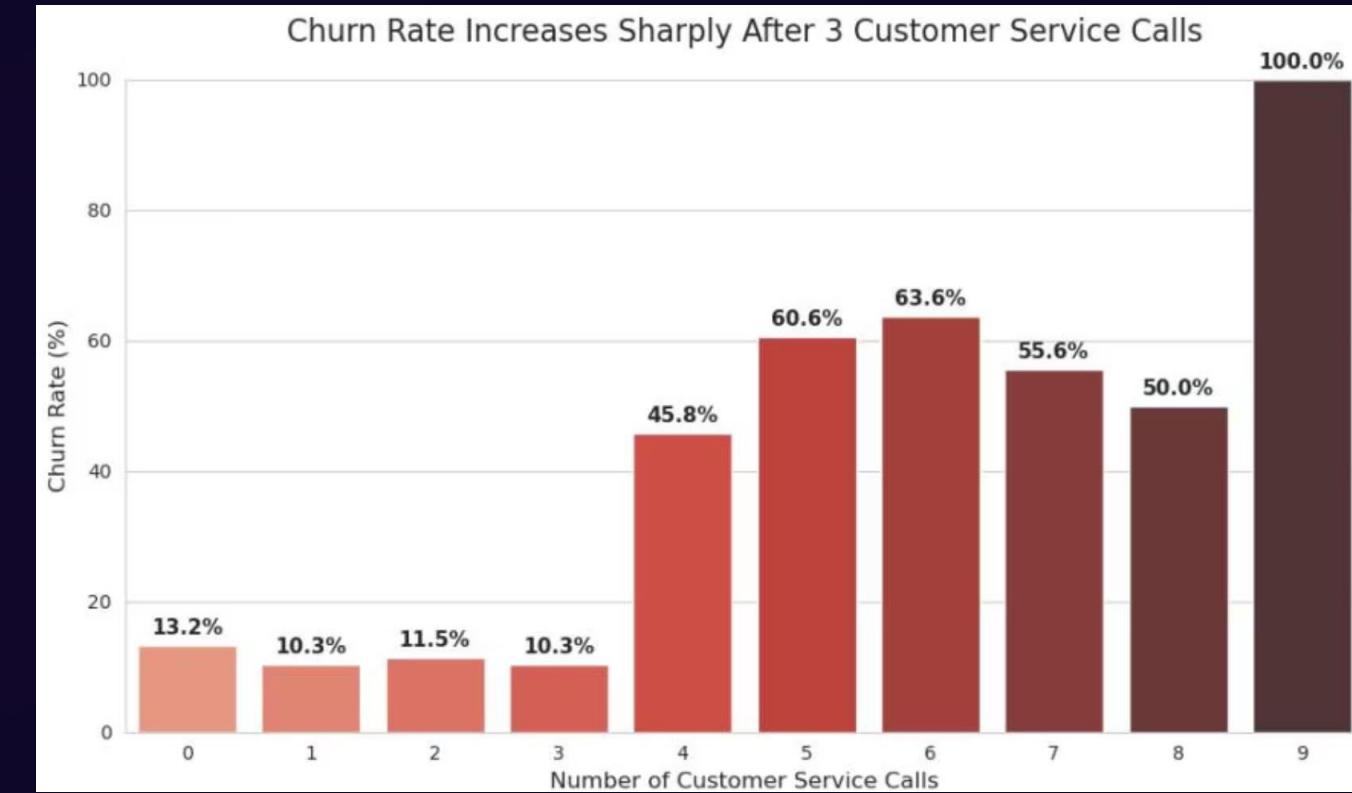
Churn Rate Insights

International Plan Impact



Customers with international plans show a significantly higher churn rate.

Customer Service Calls Threshold



Churn rate increases sharply after 3 customer service calls, reaching 100% at 9 calls.

Model Evaluation: Random Forest Leads

The Random Forest Classifier emerges as the most robust solution, balancing high Recall with strong Precision and ROC-AUC.

Model Configuration	Precision	Recall	F1 Score	AUC	ROC-AUC
LR Default	0.7841	0.3897	0.7525	0.5135	0.8276
LR Balanced	0.7466	0.3468	0.7624	0.4768	0.7874
DT Default	0.9115	0.7019	0.7228	0.7122	0.8340
DT Tuned	0.9115	0.6721	0.8119	0.7354	0.8892
RF Tuned	0.9475	0.8438	0.8020	0.8223	0.9320

Business Recommendations



Usage-Based Incentives

Target high-usage customers with tiered data plans or flat-rate upgrades to mitigate churn from high daily charges.



Support-Triggered Intervention

Implement "priority resolution" for customers reaching three service calls to address persistent issues.



International Plan Optimization

Revise international plan pricing or features to improve perceived value and reduce churn risk.



Proactive Retention Alerts

Deploy automated, personalized discounts to high-risk customers during their predicted departure window.