Project Introduction – Predicting Customer Churn at SyriaTel

This project focuses on helping SyriaTel reduce customer churn.

Using customer service data, call patterns, and plan details, we trained models to identify customers likely to leave.

The end goal is to help the business intervene before valuable customers churn.

Student: Shem Nderitu

Blog Post URL: https://github.com/Shemnderitu/SyriaTel-Customer-Churn-Project

Business Context & Stakeholders

Business Need: SyriaTel wants to identify customers at risk of leaving the service in order to reduce revenue loss and maintain market share.

Key Stakeholders:

Retention & Loyalty Team

to act on churn signals

Executive Team

to align churn reduction with KPIs

Success Indicators:

High accuracy

in identifying churners

Insights

that guide strategic action

A model

ready to integrate into business systems

Understanding the Data

The dataset includes 3,333 customer records. Features include:

Customer profiles

state, account length

Service usage

daytime/evening/night minutes and calls

Subscription status

international and voicemail plans

Support history

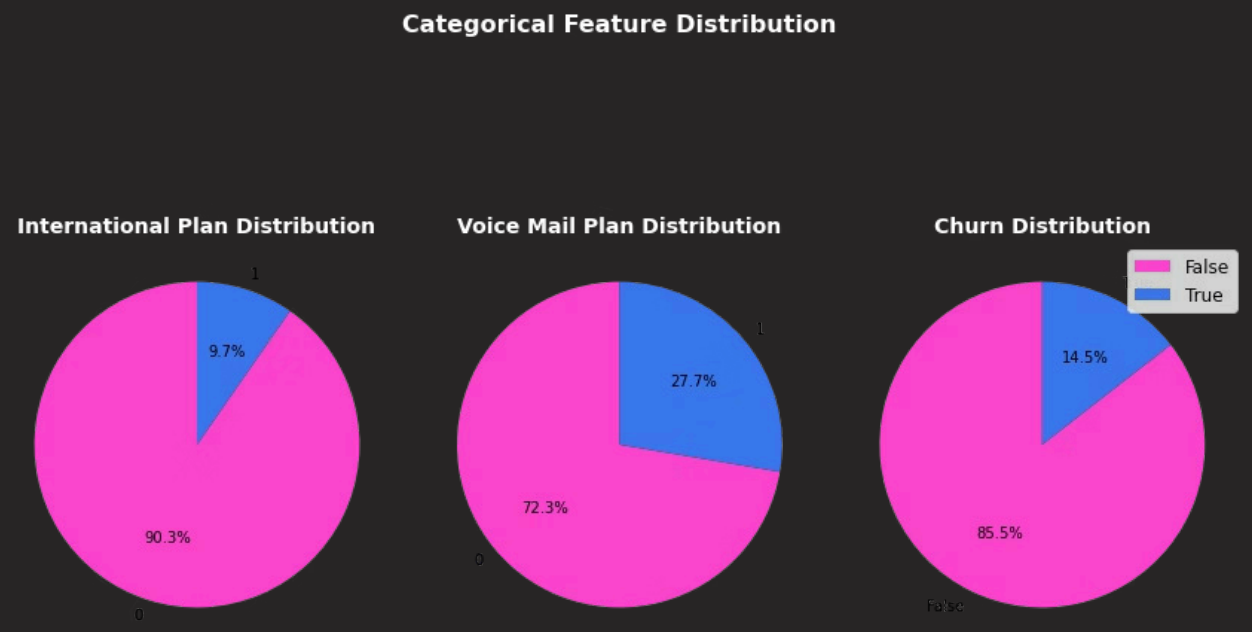
number of service calls

Target

whether the customer churned or not

Customer Behavior Breakdown

Categorical Feature Distribution Pie Chart

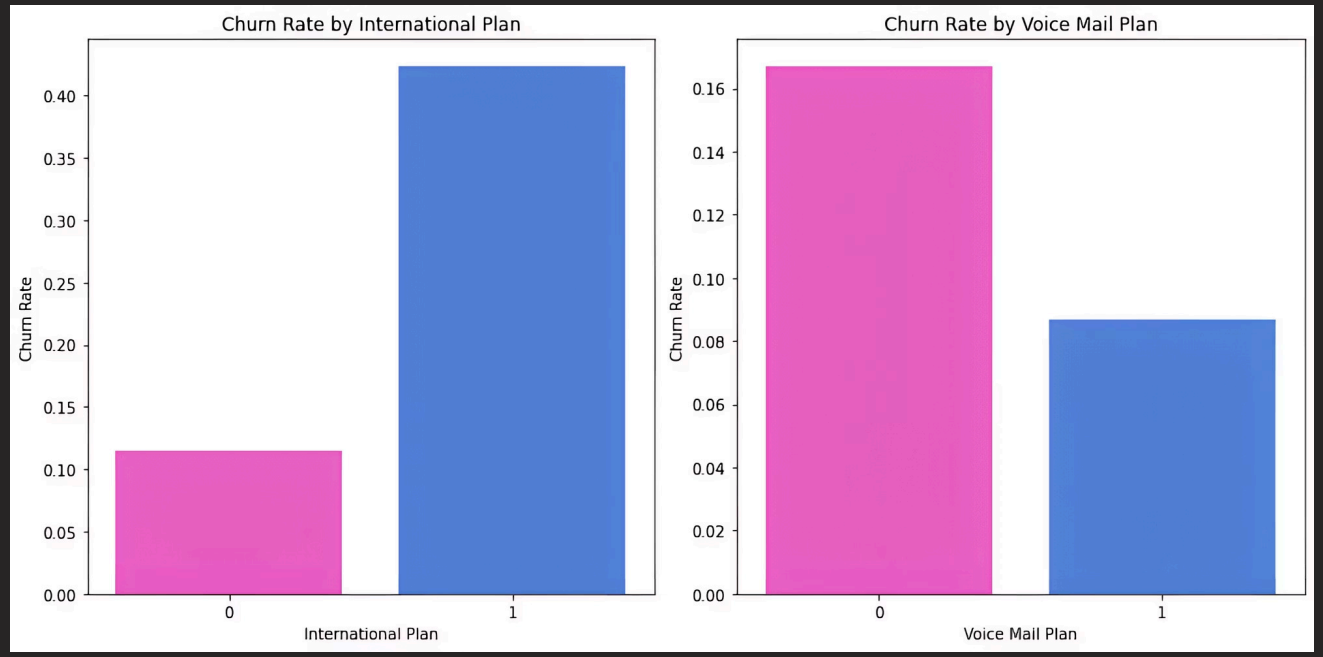


Insights:

- Only **9.7%** use an international plan; these users churn more.
- **27.7%** have voicemail – they tend to churn less.
- Most customers stay, but around **14.5%** have churned.

Churn Patterns by Plan Subscription

Plan Subscription vs. Churn Bar Chart



Takeaways:

- Churn is much higher among customers with an international plan.
- Voicemail plan users show lower churn.
- These patterns help us segment risk groups early.

Modeling: How We Predicted Churn

We tested 3 machine learning models:

Model	Description
Logistic Regression	A simple, fast model used for binary decisions
Decision Tree	Splits customer behavior into easy-to-follow decision paths
Random Forest	Combines many decision trees for more accurate predictions

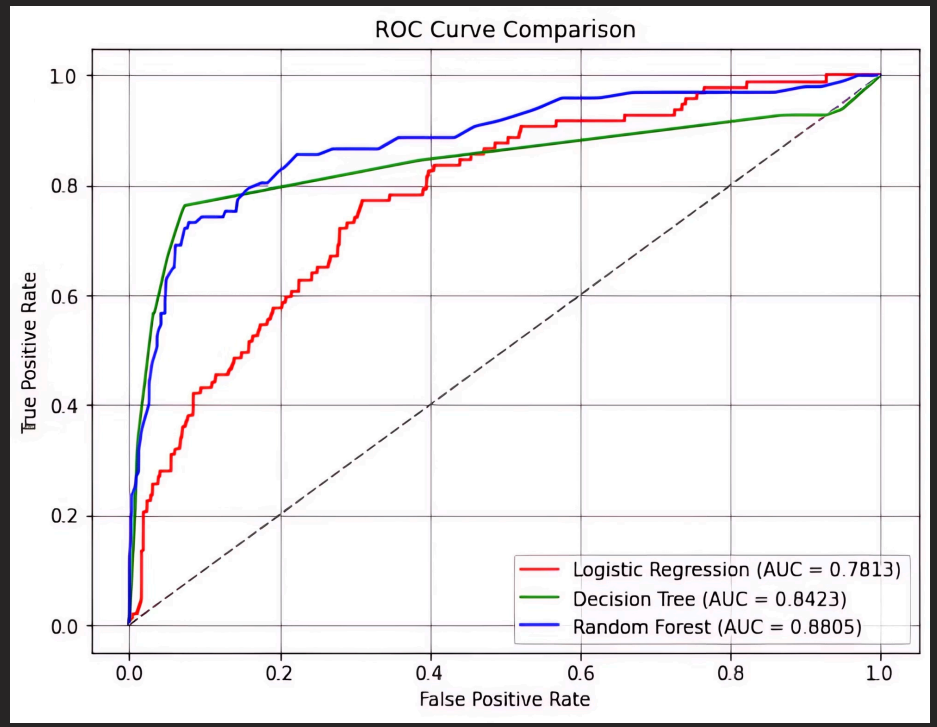
All models were trained using balanced and scaled data.

Performance Metrics – How Each Model Performed

Metric	Logistic Regression	Decision Tree	Random Forest
Accuracy	70%	90%	**90%**
Precision (Churn)	0.30	0.65	**0.63**
Recall (Churn)	**0.76**	0.75	0.69
ROC-AUC	0.7813	0.8423	0.8805

How Confident Are the Models?

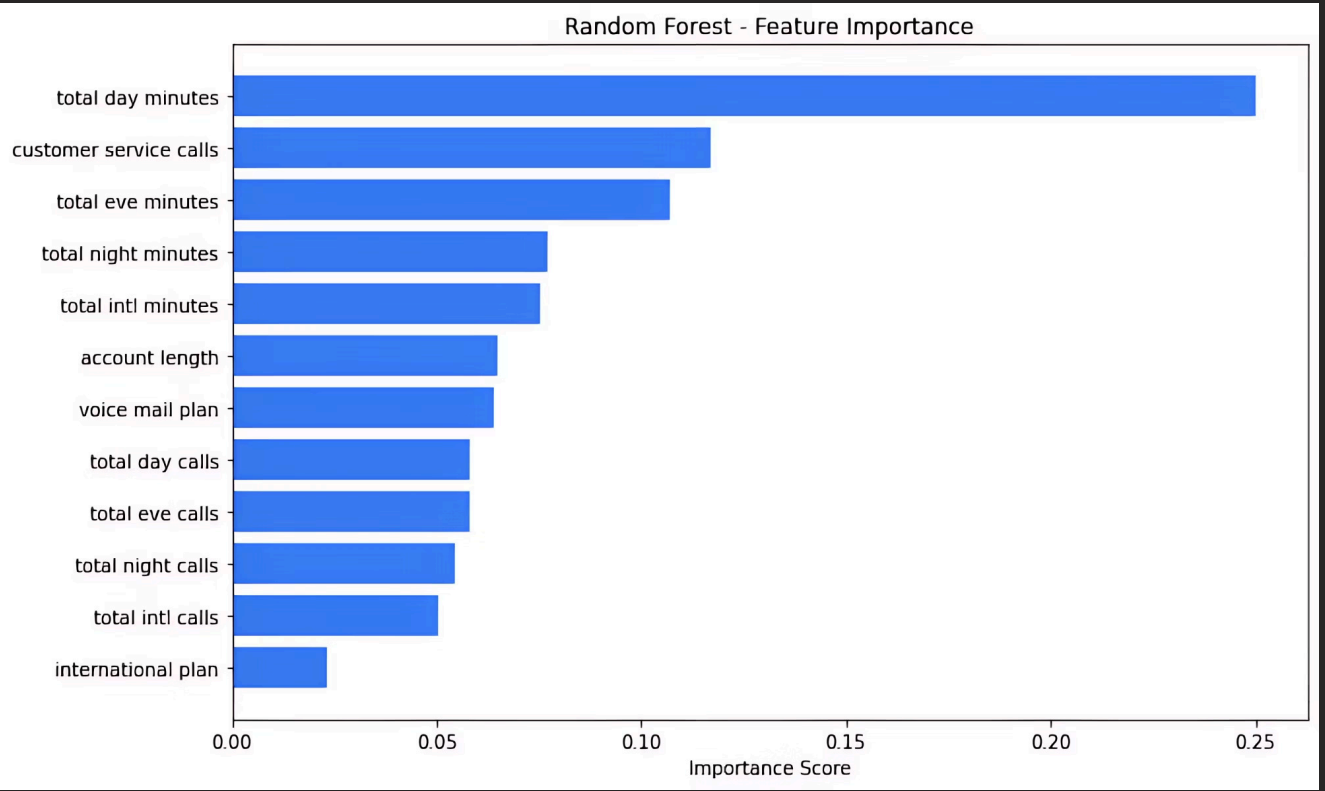
ROC Curve Comparison



- The **Random Forest curve** hugs the top-left the closest — meaning it balances false alarms and missed churns well.
- **Logistic Regression** has the highest recall but may raise too many false alarms.
- **Decision Tree** is a middle ground: high accuracy and clear rules.

What Matters Most for Predicting Churn?

Random Forest Feature Importance Bar Chart



Top Signals of Churn:

- **Having an international plan**
- **High international minutes**
- **Large daytime usage**

Lower Importance:

Voicemail plan and night call volume had little influence.

Key Conclusions

1

Logistic Regression

Good for fast, simple insights with high recall.

2

Decision Trees

Offer clear rules for customer segmentation.

3

Random Forest

Most reliable and well-rounded, with the best balance of performance.

Business Recommendations

🎯 Focus retention efforts on:

- Customers with international plans
- Those calling support frequently

💬 Boost customer experience by:

- Monitoring and addressing frequent complaints

📢 Market voicemail plan usage to at-risk users

🧮 Integrate the **Random Forest model** into CRM for live churn alerts

Next Steps

Deploy model in production

Set up dashboards for churn predictions

Retrain model quarterly to keep predictions sharp

Test additional features

(billing complaints, network issues) to improve predictions

Thank You

Questions?

Let's work together to keep SyriaTel customers happy and loyal.

Linkedin: https://www.linkedin.com/in/shem-nderitu-165b97184/