SyriaTel Customer Churn Prediction Presentation

Name: Jeffrey Gathigi
Instructor: Brian Chacha
Program: Data Science (Part-Time)

Predicting Customer Churn at SyriaTel

A Data-Driven Approach to Customer Retention

OVERVIEW - THE CUSTOMER RETENTION CHALLENGE

Key Points

The Problem: We're losing customers (churn)

Why It Matters: Acquiring new customers costs 5-25x more than keeping existing ones

Our Solution: Predict who will leave BEFORE they go

Expected Impact: Save millions in revenue through targeted retention

Every time a customer leaves SyriaTel, we lose not just their monthly revenue, but potentially thousands of dollars in lifetime value. Studies show that acquiring a new customer costs between 5 to 25 times more than retaining an existing one. This project uses data to predict which customers are at risk of leaving, giving us time to intervene and keep them.

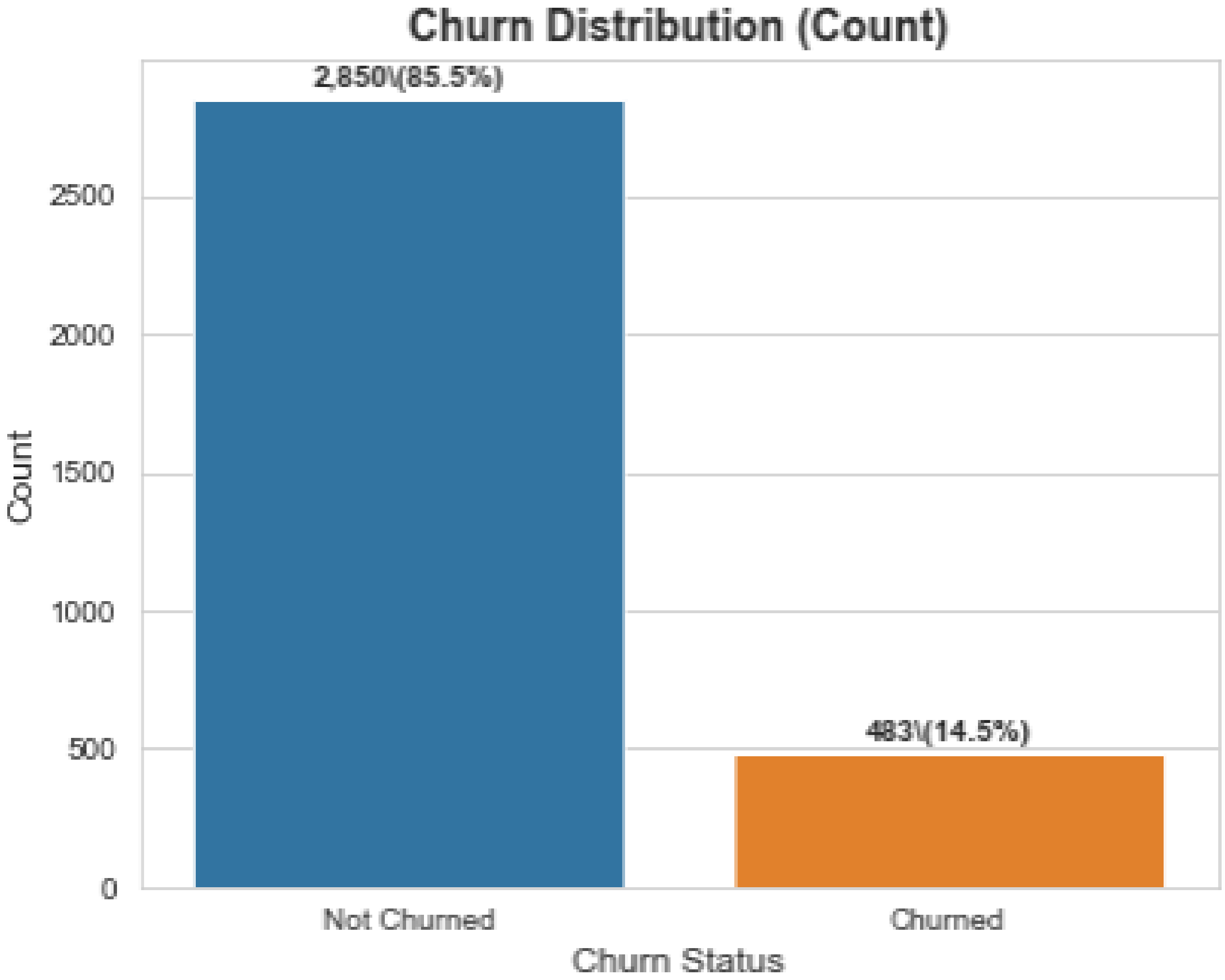
BUSINESS UNDERSTANDING - OUR GOAL

Key Points:
Objective: Build an early warning system for customer churn
Success Metric: Identify at least 70% of customers who will leave
Why 70%?: Better to reach out to some loyal customers than miss those who will leave
Business Impact: Enable proactive retention campaigns

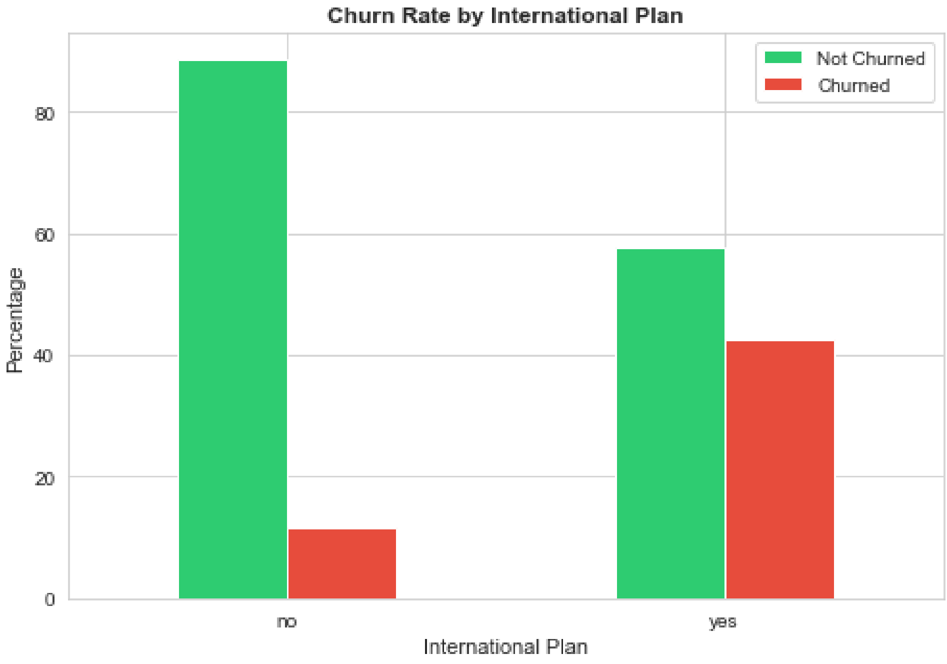
Our goal is straightforward: we want to know which customers are thinking about leaving before they actually do. We set a target to correctly identify at least 70% of customers who will churn. You might ask, 'why not 100%?' Well, in the real world, we'd rather reach out to some customers who weren't actually planning to leave, than miss the ones who are. The cost of a retention offer is much less than losing a customer entirely.

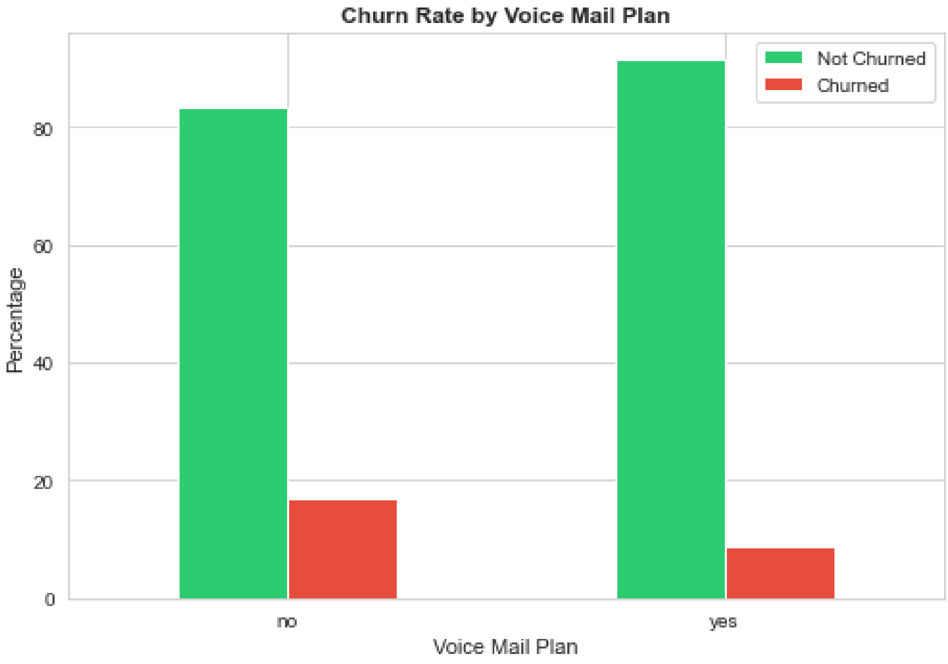
Data Understanding - What We're Working With

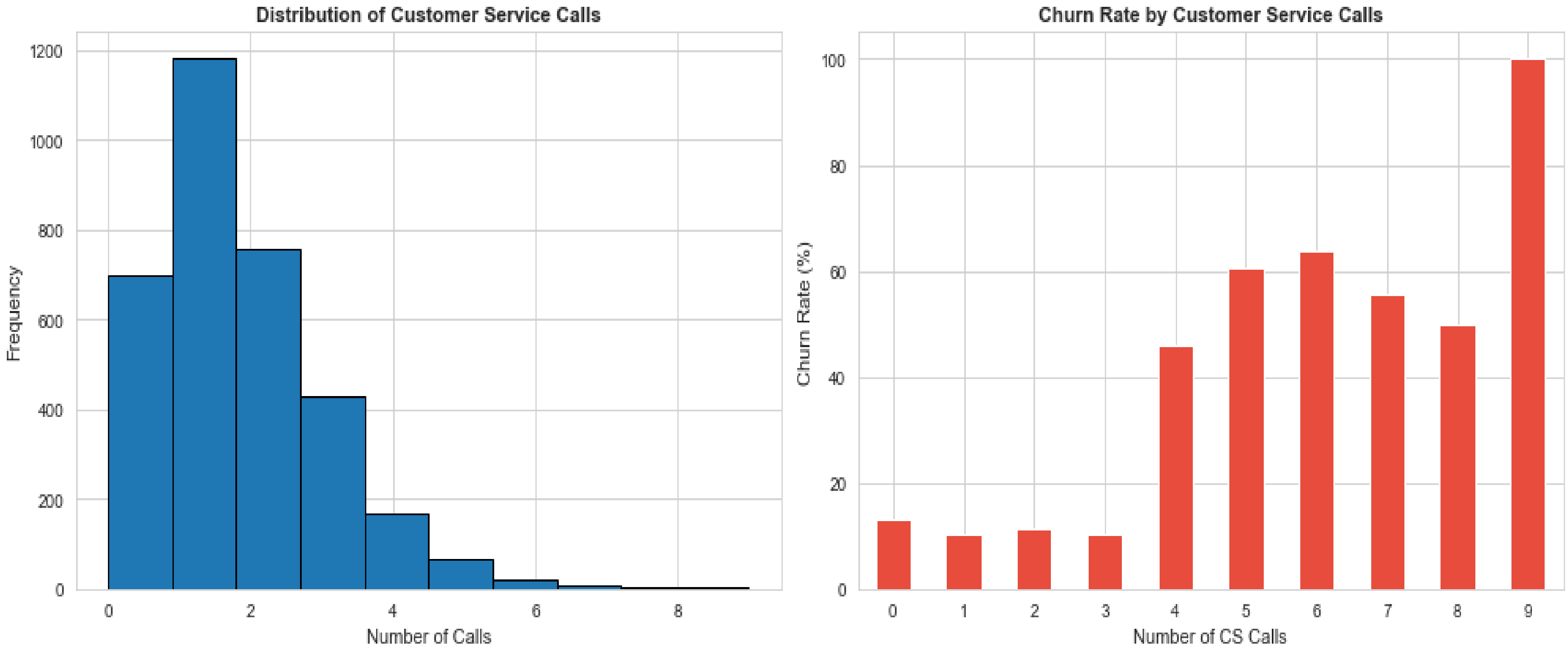
Key Statistics:
3,333 Customers analyzed
Account Information: Length of relationship, location
Usage Patterns: Call minutes, charges, service plans
Service History: Customer service interactions
Current Churn Rate: 14.5% (483 customers left)
We analyzed data from 3,333 customers. For each customer, we have information about their relationship with SyriaTel: how long they've been with us, which services they use, their calling patterns, and importantly, whether they've contacted customer service. We also know which customers ultimately left—that's our churn status. Currently, about 14.5% of our customers are churning, which represents significant lost revenue.

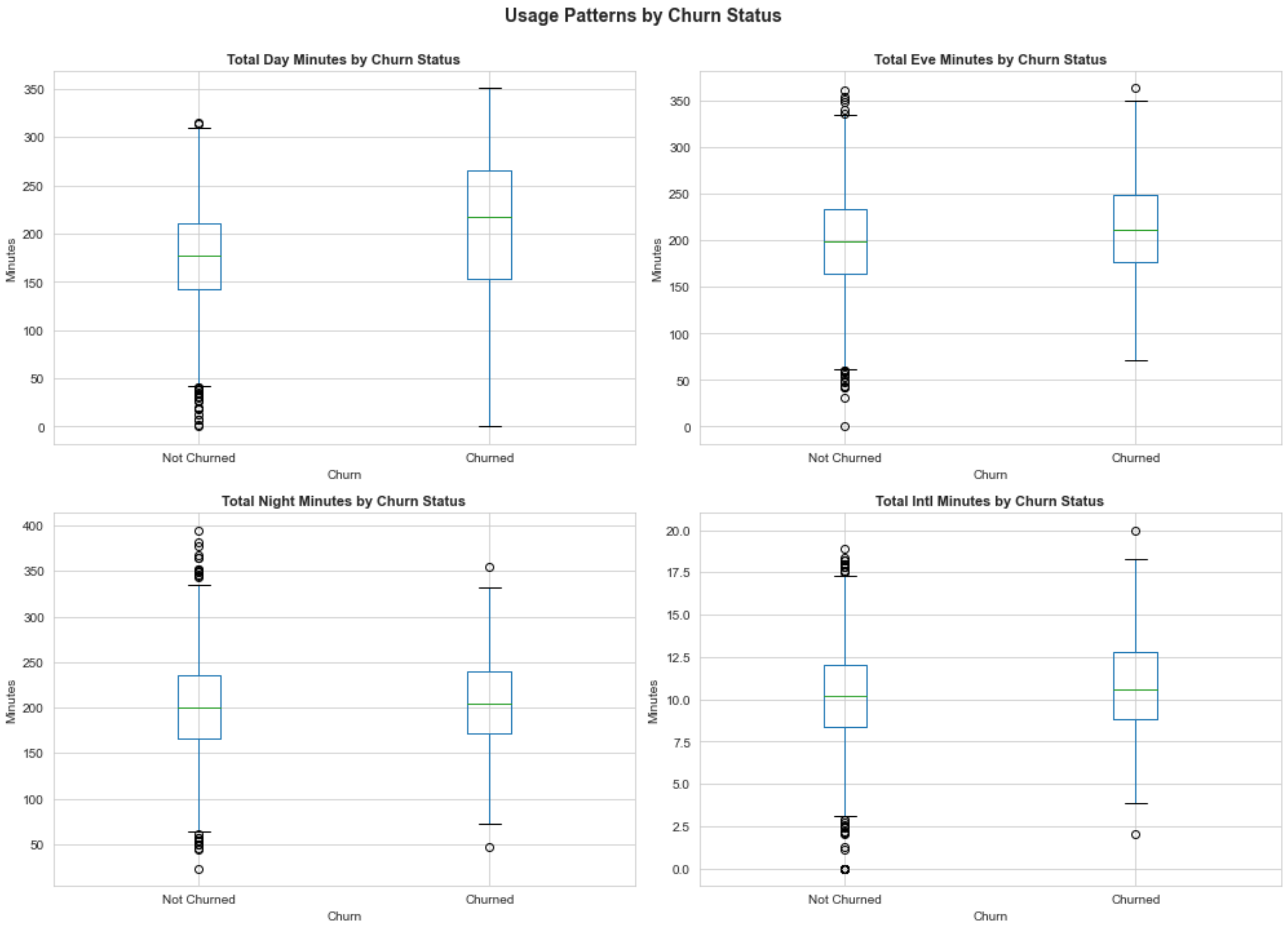

Churn Distribution (Count)
2,850\(85.5%)
483\(14.5%)
2500
2000
1500
1000
500
0
Count
Not Churned
Churned
Churn Status

Key Insights - What the Data Reveals






Distribution of Customer Service Calls
Frequency
Number of Calls
0
200
400
600
800
1000
1200
0
2
4
6
8
Churn Rate by Customer Service Calls
Churn Rate (%)
Number of CS Calls
0
20
40
60
80
100
0
1
2
3
4
5
6
7
8
9

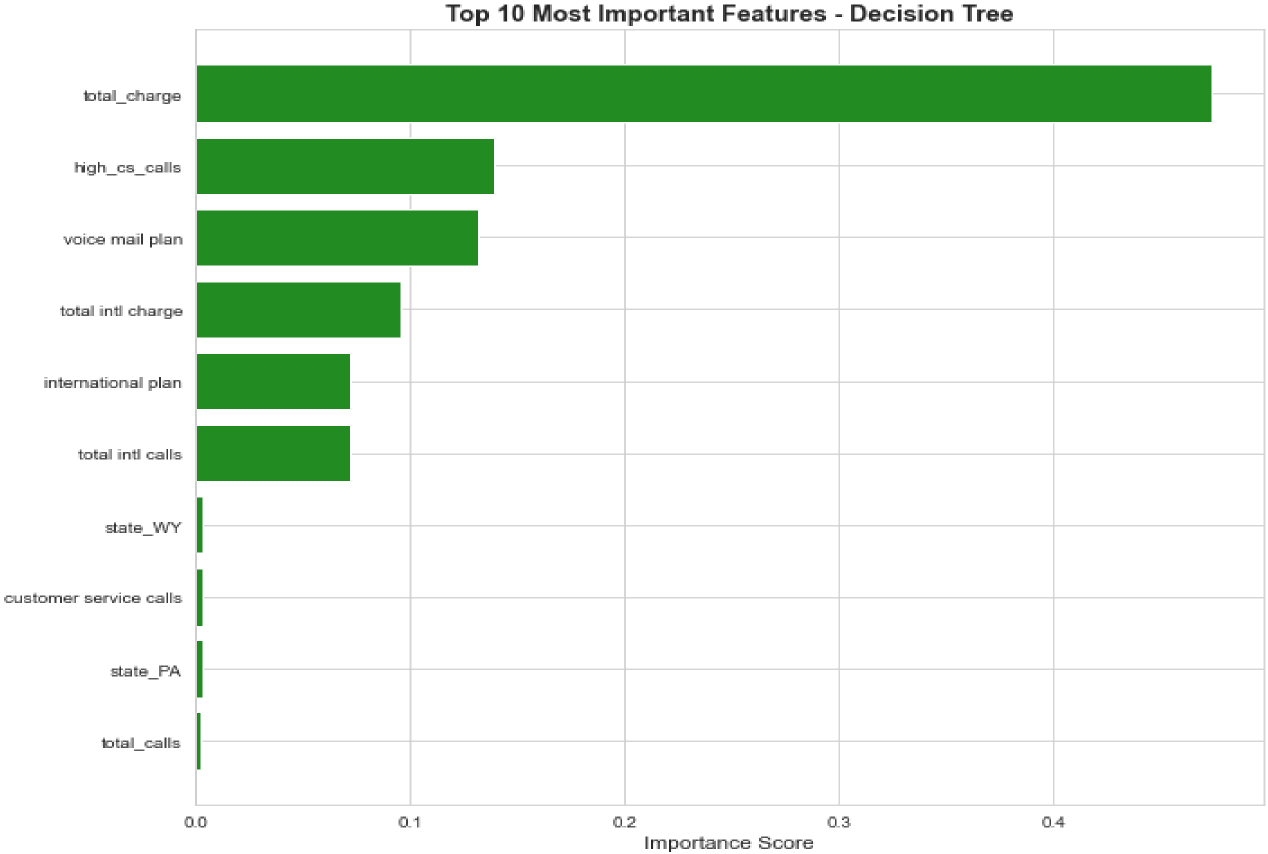

Usage Patterns by Churn Status
Total Day Minutes by Churn Status
Total Eve Minutes by Churn Status
Total Night Minutes by Churn Status
Total Intl Minutes by Churn Status
Minutes
Churn
Not Churned
Churned
0
50
100
150
200
250
300
350
400
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0

Key Insights - What the Data Reveals

Before building our prediction model, we explored the data to understand what drives churn. Four major patterns emerged: First, customers who call our support line 4 or more times are at extreme risk—repeated calls signal unresolved problems. Second, customers with international plans leave at three times the normal rate, suggesting pricing or value issues. Third, customers with our voice mail plan are more loyal—value-added services create stickiness. Finally, customers with very high charges are more likely to leave, possibly due to bill shock or better competitor pricing.

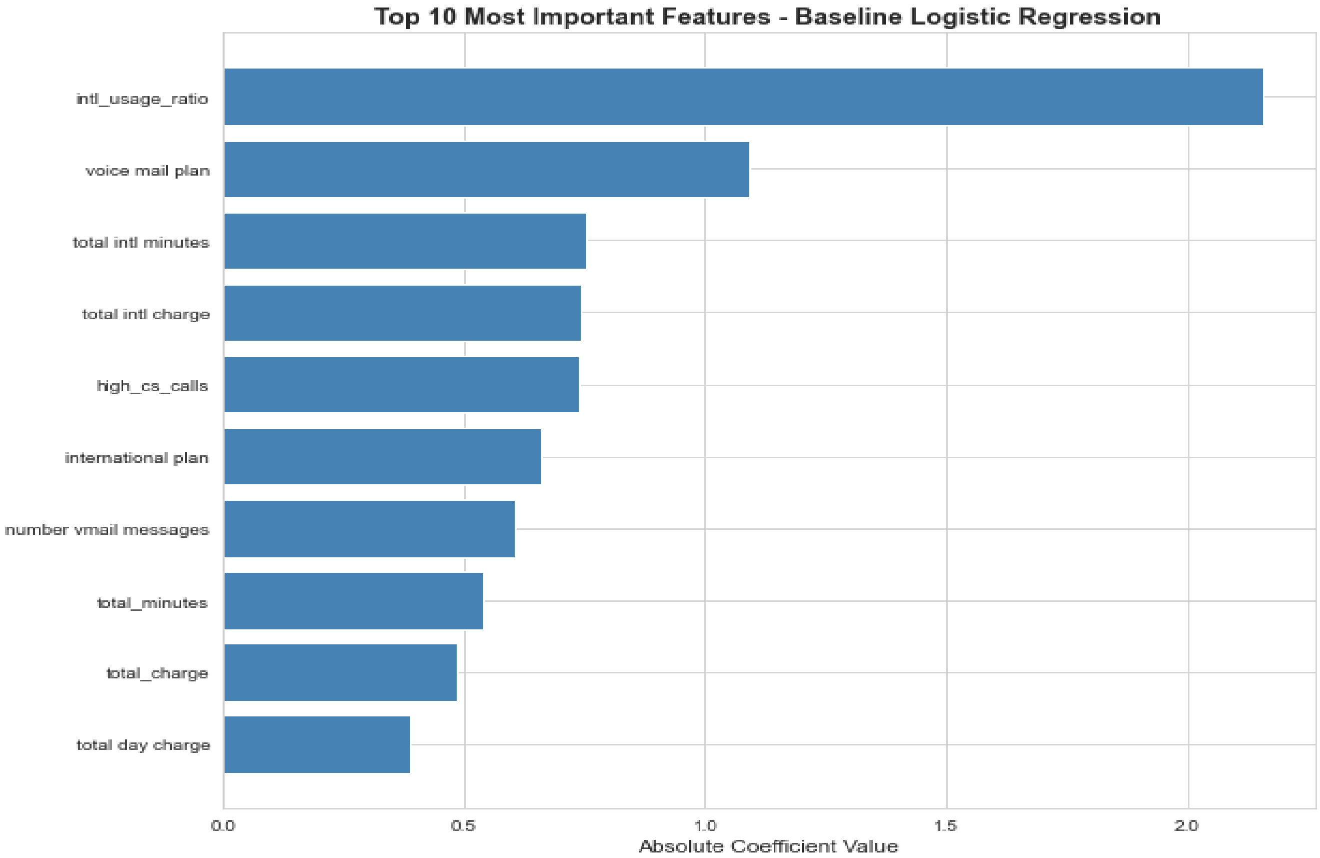
Understanding Classification Models

Decision tree diagram



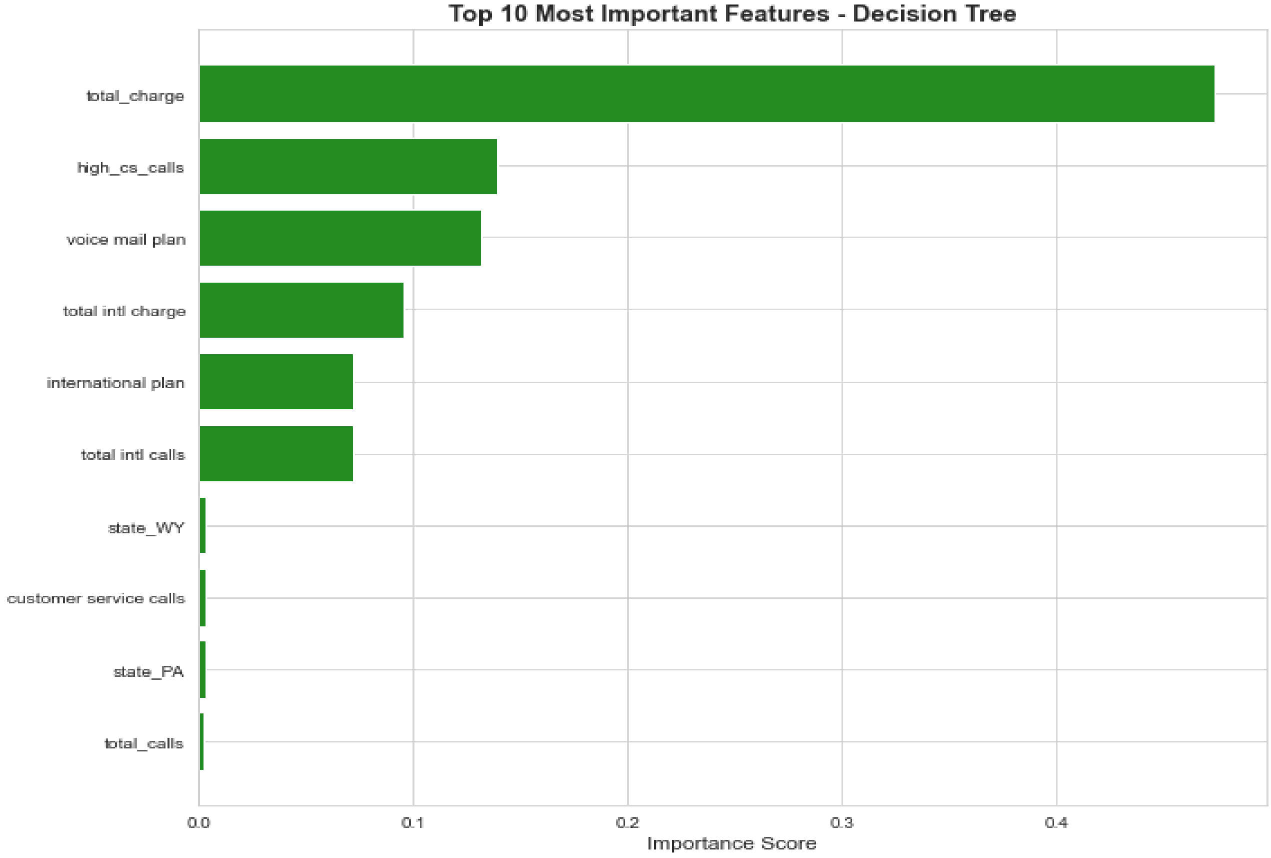
Modeling Approach - Testing Different Methods

Card 1: Logistic Regression



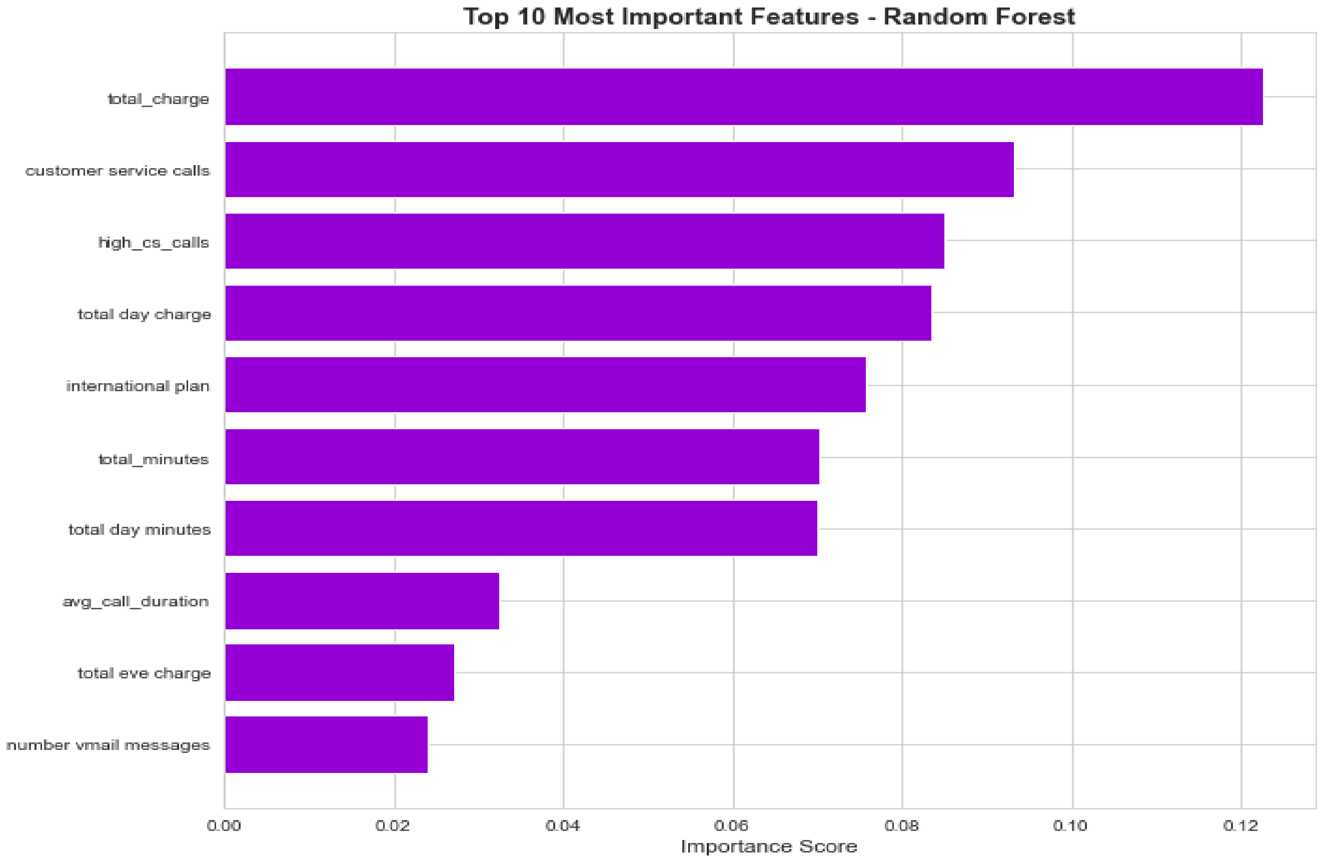
Modeling Approach - Testing Different Methods

Card 2: Decision Tree



Modeling Approach - Testing Different Methods

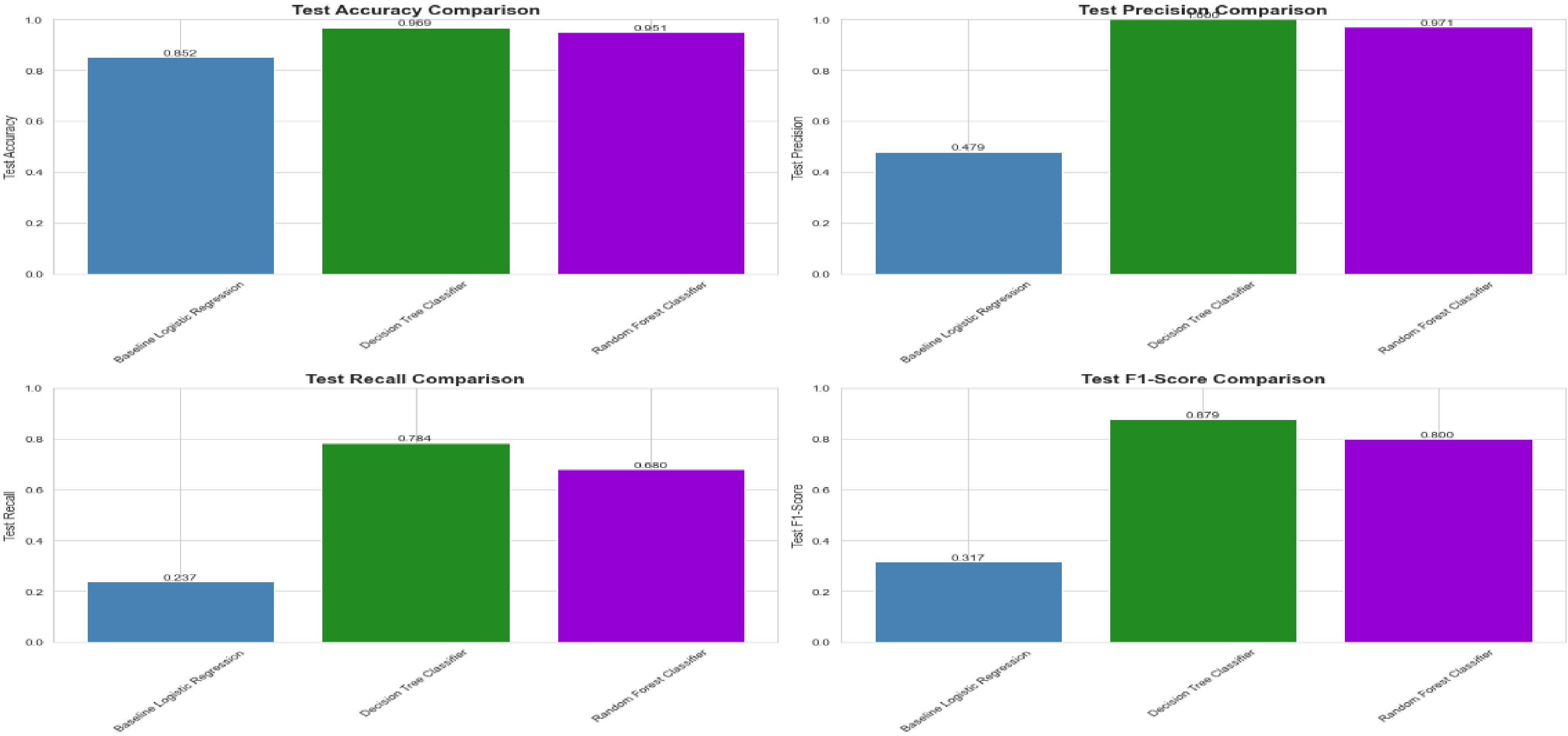
Card 3: Random Forest



Modeling Approach - Testing Different Methods

We tested three different approaches to find the best predictor.The first, Logistic Regression, is simple and transparent but only catches 24% of customers who will churn—not good enough. The third, Random Forest, is very sophisticated and accurate overall, catching 68% of churners. But the second approach, Decision Tree, is our winner: it catches 78% of customers who will leave, and importantly, it never raises false alarms. It's also easy to explain and understand, which matters for implementation.

Model Evaluation - How Well Does It Work?



Model Evaluation - How Well Does It Work?

Let me translate the technical metrics into business terms. Our most important measure is called 'recall'—it means: out of every 100 customers who will actually leave, we correctly identify 78 of them. That gives us 78 opportunities to intervene and save those relationships. Yes, we miss 22, but 78% is excellent in the industry. Our overall accuracy is 97%, meaning we make the right call almost every time. And here's something special: our 'precision' is 100%, meaning when our model flags someone as at-risk, it's always right—we have zero false alarms. This means every alert we get is worth investigating.

What Drives Churn? - Most Important Factors

Our model tells us exactly which factors matter most for predicting churn. The number one predictor is customer service calls—specifically, when someone calls four or more times, they're at extreme risk. This tells us they have unresolved problems. Second, total charges matter: high bills make customers shop around. Third, having an international plan is surprisingly risky, suggesting we need to review our pricing and value proposition for these customers. On the positive side, customers with our voice mail plan are much more loyal—value-added services create stickiness. These insights aren't just academic—they tell us exactly where to focus our retention efforts.

Business Recommendations - Taking Action

Based on our findings, I recommend a three-tier action plan. First, immediate interventions for high-risk customers: anyone with 4 or more service calls gets assigned to a dedicated retention specialist, and we need to review our international plan pricing immediately.

Second, medium-term strategies like automated usage monitoring to catch bill shock before it happens, and promoting value-added services like voicemail to increase loyalty.

Third, systematic changes: deploying this model to score all customers monthly and triggering automated retention workflows. This moves us from reactive to proactive retention.

Future Enhancements - Continuous Improvement

This model is just the beginning. Over the next year, we can enhance it significantly: moving from monthly batch predictions to real-time scoring, adding more data sources like competitor pricing and app usage, and building advanced capabilities like trend detection and optimal offer selection. We can also integrate this more deeply into our operations with automated alerts to account managers and executive dashboards. The foundation we're building today will support increasingly sophisticated retention strategies as we learn and improve.

Thank You

Thank you for your time. I'm happy to answer any questions or dive deeper into any aspect of the analysis

- Jeffrey Gathigi

- Data Science Program (Part-Time)

- Instructor: Brian Chacha

- Email: jeffrey.gathigi@student.moringaschool.com

- LinkedIn: https://www.linkedin.com/in/jeffrey-gathigi-38a373363/

- GitHub: https://github.com/GATHIGIMUREITHI/phase3project.git