SYRIATEL TELECOMMUNICATION DATA REPORT

Title: Improving Customer Retention Strategies at SyriaTel Telecom Ltd.

Authors: *Hassan Ali, Rose Matoke, Kenneth Kiplagat, Celine Sitina, Felix Musau.*

Date: September 13, 2025.

Introduction

This report looks into why customers are leaving Syriatel and what can be done to keep them. Churn isn't random. It's linked to key features such as customer experience and usage habits. Identifying these features is a big step towards improving customer retention and predicting our company's performance over time. In this analysis, we apply model evaluation to identify the most effective model for understanding churn and guiding future business decisions.

Problem Statement

Syriatel is experiencing customer churn that directly threatens its revenues and market position. While it has historically maintained a strong market share, rising competition, service quality issues, evolving customer demands, and broader political-economic instability increase the likelihood of customer attrition. Without effective churn prediction and retention strategies, Syriatel risks losing valuable customers, resulting in revenue loss, reduced market share, and diminished competitiveness.

Business objectives

Main objective:

To develop a machine learning classifier that predicts whether a Syriatel customer is likely to churn, enabling data-driven strategies for proactive retention.

Specific objectives:

1. To explore customer demographics and usage behavior influencing churn.

2. To determine how charges influence customer churn

3. To develop and evaluate machine learning models that classify whether a customer is likely to churn.

4. Optimize the models for best performance.

5. To provide actionable insights that support Syriatel in designing targeted retention campaigns (e.g., loyalty programs, personalized offers).

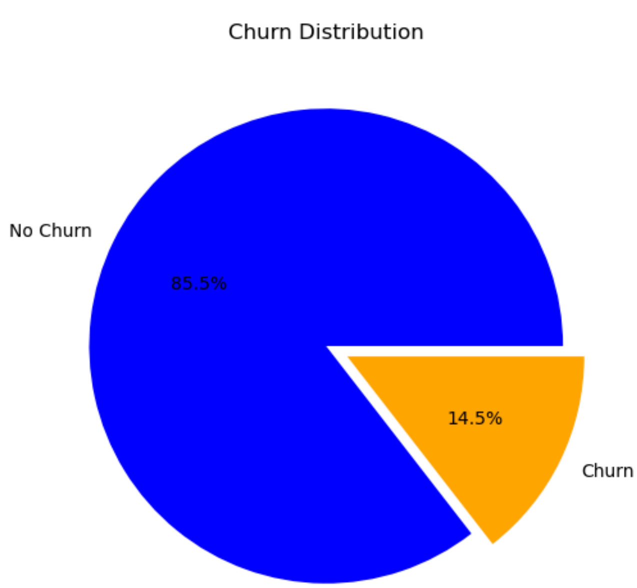
Success Criteria

1. Model Performance: Achieve at least 85% accuracy and a high AUC score (>0.85) in predicting churn.

2. Business Impact: Provide insights that reduce churn rates by enabling proactive retention strategies, targeting high-risk customers before they leave.

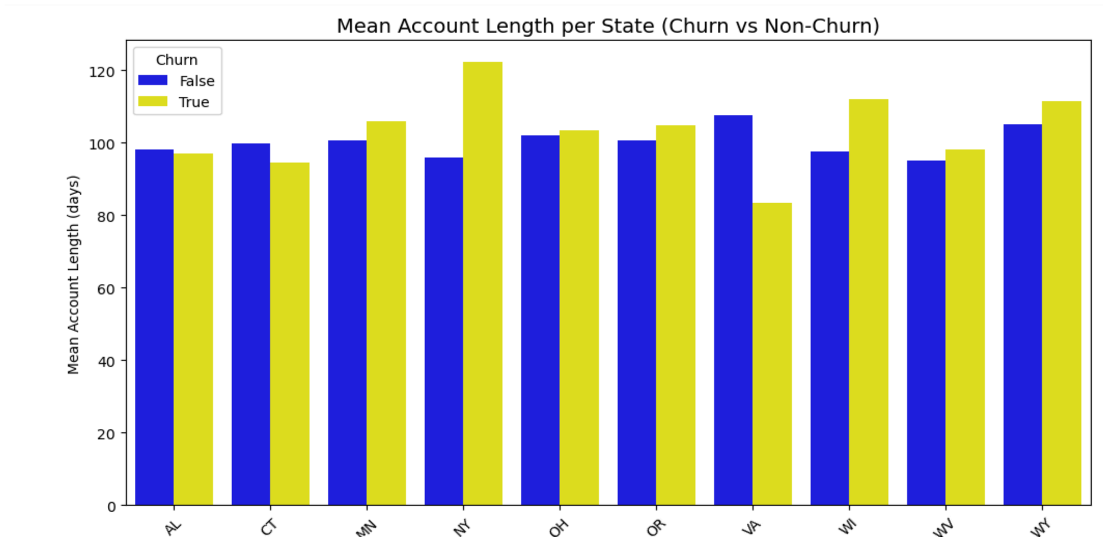
Findings & Analysis

1. Churn vs Non-Churn



As seen from the pie chart, 85.5% of the customers stayed while 14.5 churned.

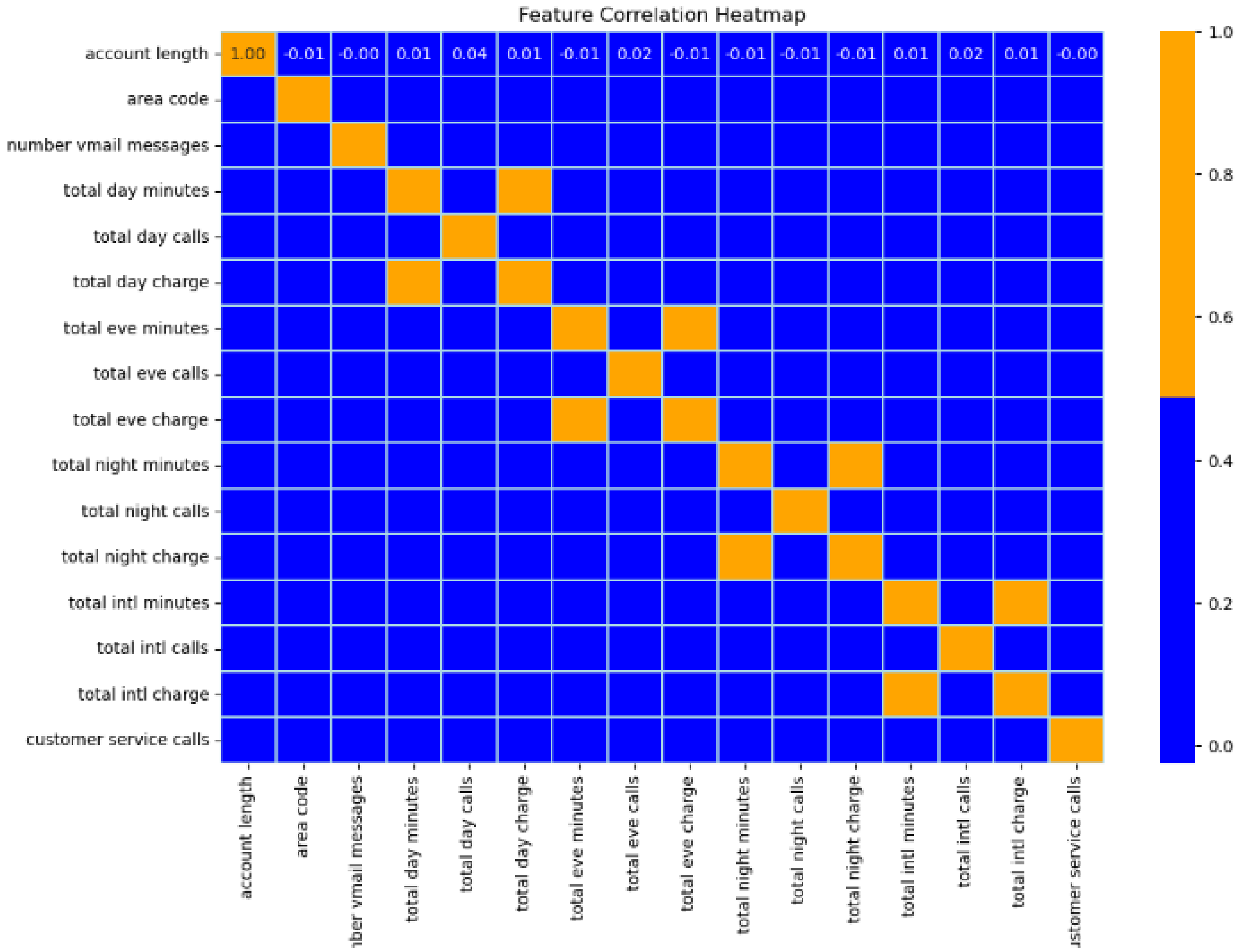
2. Mean Account length per state



The above visualization illustrates well how account length varies with churn. The states with the top mean account length are also indicated above as Alabama, Connecticut, Minnesota, New York and Ohio.

3. Correlation Heatmap

The following heatmap shows correlation between different features in the dataset



Modelling & Evaluation

Creating the baseline model with the selected features as international plan, customer service calls, total day charge, total eve charge and the target variable is churn

The following comparison models were created:

1. Logistic regression (with multiple features)
2. Decision Tree
3. Random Forest

We conducted hyperparameter tuning for all of them to optimize model performance, and concluded the following:

	Baseline Model	Logistic regression	Decision Tree	Random Forest
Accuracy	75%	76%	81%	89%
Precision	94%	95%	94%	85%
Recall	76%	76%	95%	92%
F1 Score	84%	84%	94%	76%

Recommendations

1. Adopt Random Forest as the Primary Model

- Random Forest achieved the highest ROC_AUC (0.896) and Average Precision (0.706), making it the most reliable choice for predicting churn.
- **Business impact**: It can accurately flag customers most at risk, by an accuracy of 93%

2. Use Logistic Regression When Interpretability Matters

- Logistic Regression performed weaker with an ROC_AUC of 0.815 and an Average Precision of 0.452, but it's easy to interpret
- **Recommendation**: Use it alongside Random Forest in scenarios where transparency and stakeholder trust are more important than raw performance.

3. Treat Decision Trees as a Supporting Model

- Decision Tree had and ROC_AUC of 0.893 and Average Precision of 0.797. These were solid, but still below Random Forest.

- **Recommendation**: They can be useful as a simple, explainable baseline, but not as the main production model. Monitor and retrain regularly using updated datasets to ensure the model adapts to evolving customer behavior.