

Presented by Norman Mwapea

# SYRIATEL CUSTOMER CHURN PREDICTION



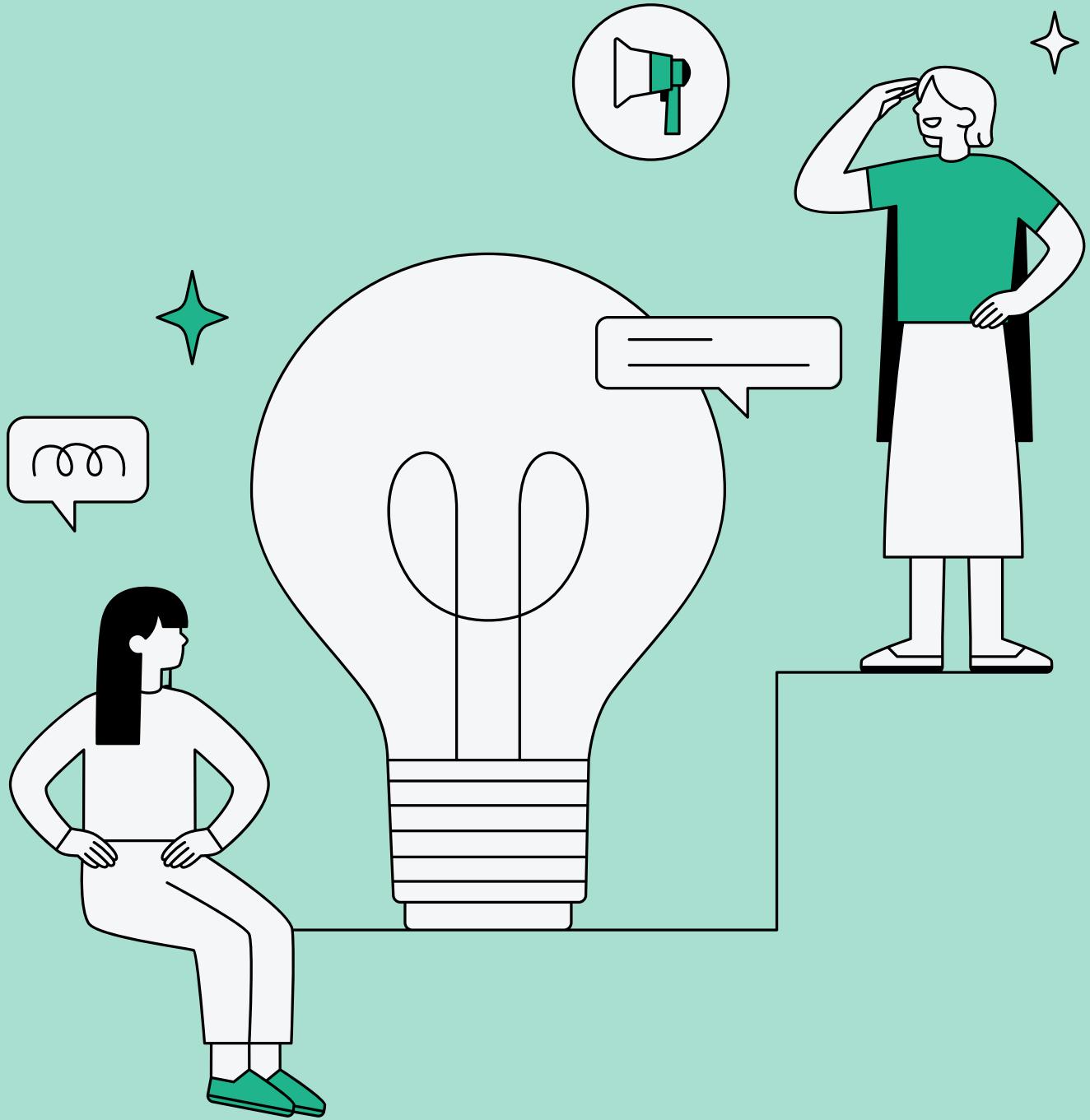
Powered by Ahjin Analytics.

# Problem Statement

Customer churn remains one of the most critical challenges in the telecommunications sector. Losing a customer directly impacts recurring revenue and increases acquisition pressure to replace churned subscribers. However, churn is driven by complex behavioral, service-related, and pricing factors that cannot be fully understood without a structured data-driven approach.

SyriaTel currently lacks clear visibility into which customer attributes, usage patterns, or service experiences contribute most to attrition. Without this understanding, the organization cannot effectively identify at-risk customers or design targeted retention strategies.

This project addresses the gap by analyzing usage behavior, plan characteristics, and customer service interactions to uncover the key drivers behind churn and provide a reliable predictive foundation for early intervention.



# Business Understanding and Objectives

The main objective is to support SyriaTel in reducing customer churn through analytical insight and predictive modeling. To achieve this, the project focuses on four strategic aims:

- Identify the variables most strongly associated with churn, including service interactions, usage intensity, and plan features.
- Enable early detection of customers likely to leave by developing a modeling-ready dataset suitable for predictive algorithms.
- Guide effective retention initiatives by translating analytical findings into practical actions such as plan adjustments, improved service handling, and targeted communication.
- Protect long-term revenue by reducing unnecessary customer loss and prioritizing retention resources on the highest-risk profiles.
- This business understanding provides the foundation for all subsequent analytical and modeling decisions.



# Data Understanding



The dataset consists of 3,333 customer records and includes usage behavior, service interaction metrics, and plan subscription indicators. Key fields include:

- Usage metrics: day, evening, night, and international minutes and calls
- Plan information: international plan, voicemail plan
- Customer interaction activity: customer service calls
- Engineered features: usage ratios, total usage minutes, service intensity (calls per month)
- The target variable is churn, indicating whether a customer left the service.
- While the dataset provides a strong behavioral foundation, it lacks demographics, billing history, and temporal patterns, which limits certain types of analysis but still enables robust exploratory and predictive modeling.

# Data Analysis: Key EDA Findings

01.

Customer service calls show the strongest relationship with churn; frequent callers are significantly more likely to leave.

02.

Daytime usage and charges are higher among churners, suggesting dissatisfaction driven by cost sensitivity or poor plan fit.

03.

International plan users churn more frequently than those without the plan.

04.

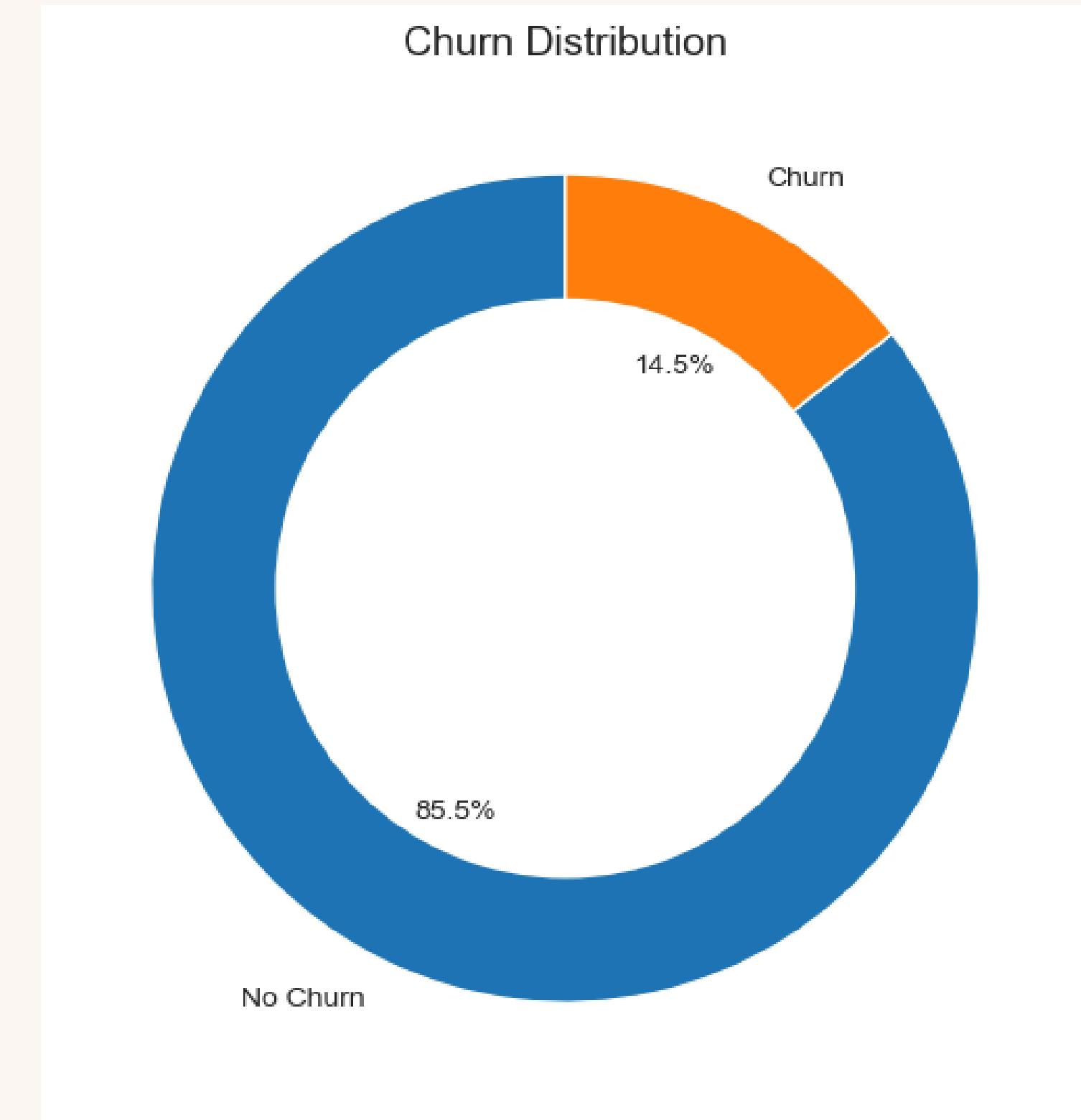
Voicemail plan subscribers churn less, indicating improved engagement or value perception.

05.

Evening and night usage exhibit minimal differences between churners and non-churners.

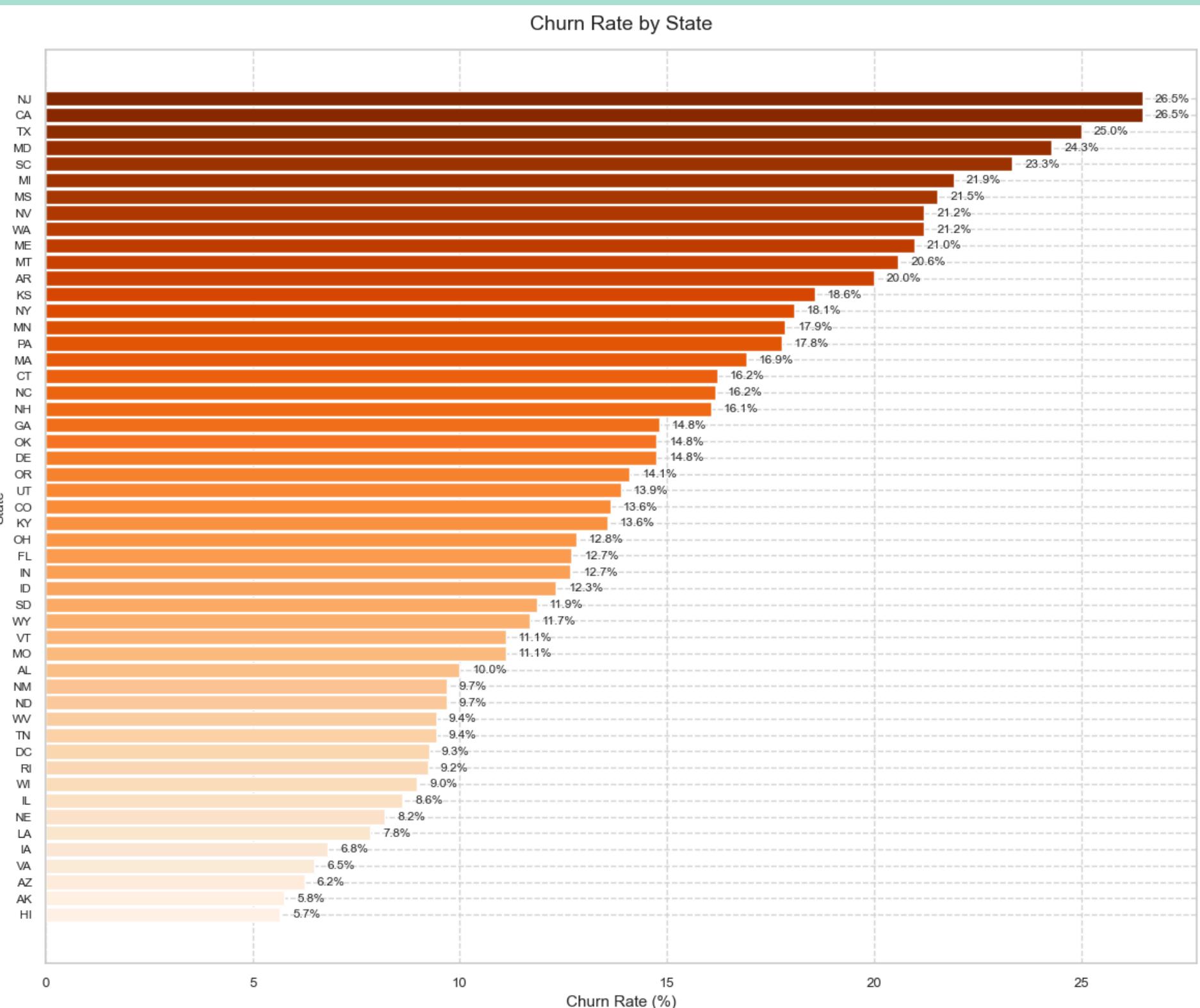
# Churn Distribution

The churn distribution shows that a majority of customers remain with the company, while approximately 14.46% have churned.



# Churn By State

1. Churn rates vary across states, with higher churn observed in regions such as New Jersey, California, Texas, and Maryland. These states tend to be more competitive or urban markets.
2. Conversely, states such as Hawaii, Alaska, Virginia, and Iowa exhibit lower churn levels.



# Modeling Approach

01.

Cleaned and transformed the dataset, removing redundant features and applying feature engineering.

02.

Split data into 80% training and 20% testing with stratification to retain churn distribution.

03.

Addressed class imbalance using class weighting and imbalance-tolerant algorithms.

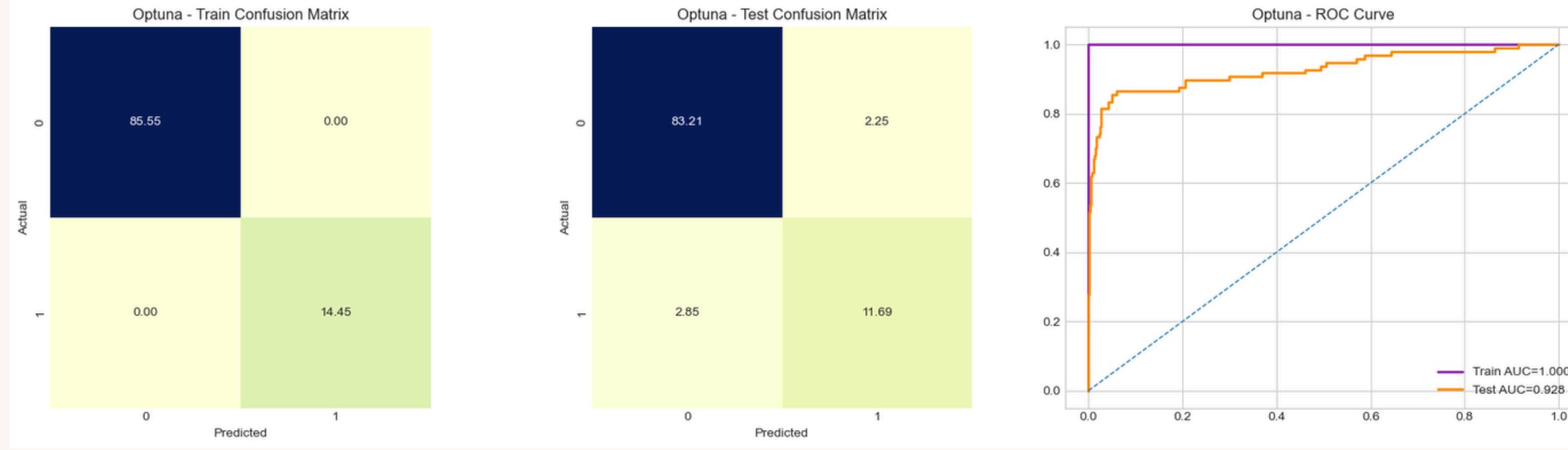
04.

Evaluated multiple models, including Logistic Regression, Random Forest, XGBoost, and CatBoost.

05.

Performance metrics prioritized recall, F1-score, and ROC-AUC to ensure strong detection of churners.

	Model	Train Accuracy	Test Accuracy	Train Precision	Test Precision	Train Recall	Test Recall	Train F1	Test F1	Train ROC-AUC	Test ROC-AUC
0	RandomizedSearchCV	99.77	93.40	99.77	93.40	99.77	93.40	997.755	934.033	99.99	92.79
1	Optuna	1.00	94.90	1.00	94.82	1.00	94.90	1.00	94.85	1.00	92.79
2	GridSearchCV	1.00	94.15	1.00	94.12	1.00	94.25	1.00	94.14	1.00	93.11



# Hyperparameter Tuning Results

**Best Model: Catboost tuned using Optuna  
92.79% ROC-AUC On & 94.90% F1 Score on test data**

Three tuning approaches were applied:

- RandomizedSearchCV
- Optuna optimization
- GridSearchCV
- Among the three, Optuna delivered the best performance, achieving:
- 94.90% test accuracy
- Superior recall and F1-score
- Minimal overfitting between train and test sets
- Top predictive features across tuned models include: total daytime minutes, customer service calls, and international plan status.

# Conclusions

1. International Plan is a major churn driver: Customers subscribed to the international calling plan exhibit a substantially higher churn rate. This feature provides one of the clearest separations between churners and non-churners, suggesting dissatisfaction with international pricing or value perception.
2. Voicemail Plan users churn less: Customers with a voicemail plan demonstrate lower churn proportions. This indicates that voicemail subscribers may be more engaged or perceive greater value in their service package.
3. Customer Service Calls strongly correlate with churn: Higher volumes of customer service interactions are closely linked to churn. Frequent support calls likely signal unresolved issues, billing concerns, or dissatisfaction—making this one of the strongest churn indicators.
4. High Daytime Usage is associated with increased churn: Churners generally have higher daytime minutes and charges. Since daytime calls often incur higher rates, heavy usage may lead to billing dissatisfaction or poor plan fit.
5. Evening and Night usage are weak churn predictors: Evening and night usage metrics show minimal differences between churners and non-churners, indicating that these time periods do not significantly influence churn behavior.
6. International usage contributes modestly to churn risk: Total international minutes and calls exhibit moderate differences between churn groups. While not as influential as plan status or service calls, they still add signal related to cost sensitivity.
7. State-level churn differences exist but are not dominant: Certain states show elevated churn, but regional variations are not strong enough to consider state a primary driver. Location plays a secondary role compared to service experience and usage patterns.

# Recommendations

1. Prioritize International Plan customers for retention efforts: Since churn is disproportionately higher among international plan users, targeted communication, improved plan transparency, or redesigned international bundles can help reduce attrition.
2. Strengthen the quality of customer service interactions: High churn among customers with repeated customer service calls underscores the need for better first-call resolution, faster escalation, and more proactive follow-ups.
3. Reassess pricing and plan suitability for heavy daytime users: Provide flexible or personalized plan options, daytime bundles, or protective pricing strategies to mitigate churn among customers who incur high daytime charges.
4. Promote Voicemail Plan adoption: Because voicemail users are less likely to churn, consider offering voicemail promotions, bundling it with other features, or highlighting its value to customers.
5. Implement region-specific monitoring where needed: Although not a primary driver, states with elevated churn may benefit from localized retention campaigns, competitive analysis, or targeted service quality improvements.
6. Flag and intervene with high-risk customer profiles early: Customers who exhibit both high usage and high customer service call volume should be proactively engaged. This group shows the highest likelihood of churn and should receive targeted support or personalized retention offers.

Presented by Norman Mwapea

Thank  
you very  
much!

Powered By Ahjin Analytics

