# SYRIATEL CUSTOMER CHURN PREDICTION

PRESENTED BY: LILIAN WANGUI

DATE: 23/07/2025

# INTRODUCTION AND OVERVIEW

- **GOAL:** To build a robust classification model to predict customer churn for SyriaTel.
- **PROBLEM:** Customers discontinuing services (churn) lead to significant revenue loss.
- **IMPORTANCE:** Proactive identification of at-risk customers enables targeted retention strategies, improving customer satisfaction and minimizing financial impact.
- **METHODOLOGY:** Analyzing historical customer data to uncover patterns and factors indicative of churn.

# BUSINESS PROBLEM

- **Core Problem: Predicting whether a customer will churn.**
- **Impact:**
  1. Direct revenue loss from lost subscriptions.
  2. Increased acquisition costs to replace churned customers.
  3. Negative impact on brand reputation and market share.

- **Strategic Need:** Implement effective retention strategies to enhance customer loyalty and reduce financial impact.

# PROJECT OBJECTIVES

- **Our Aims: Driving Retention with Data**

1. **Develop a Churn Prediction Model -** Build an effective classification model for SyriaTel.
2. **Evaluate Model Performance –** Rigorously assess effectiveness using appropriate metrics for imbalanced data(Precision, Accuracy, F1 score, ROC AUC)
3. **Identify Key Churn Drivers –** Determine which customer attributes and behaviors are most significant in predicting churn.
4. **Propose Actionable Retention Strategies –** Formulate data-driven recommendations to reduce customer attrition.

# KEY BUSINESS QUESTIONS

- **Answering Critical Business Needs**

1. **Who is most likely to churn?**

   -Identifying high risk customer segments.

2. **What differentiates churners from non–churners?**

   -Understanding the characteristics and behaviors that lead to churn.

3. **How effectively can we predict churn?**

   -Assessing model accuracy and trade-offs of predictions.

4. **How can insights inform retention strategies?**

   -Translating model findings into actionable business recommendations.

# BUSINESS STAKEHOLDERS

**The primary Stakeholder:** SyriaTel Telecommunications Company is interested in reducing customer churn and improving customer retention.

Secondary Stakeholders:

1. **Customer Care Service Teams:** To develop strategies and interventions for high churn risk customers.
2. **Marketing Teams**: To create targeted campaigns and offers aimed at churn-risk customers.
3. **Financial Analysts:** To evaluate financial impact of customer churn.

# DATA UNDERSTANDING

- **Data Source** – Syria Tel  Telecommunication Company.
- **Size** – Over 3,000 customer records.
- **Structure** – Each row represents a new customer.
- **Key Attributes** – Demographics, service subscriptions, usage patterns, customer support interactions.
- **Target Variable** – Churn (Binary: TRUE for churned, FALSE for retained)
- **Initial Steps** – Handling missing values, correcting data types, and encoding categorical variables.

# DATA OVERVIEW

- Understanding Customer Attributes.

**1. Usage Patterns:**

-total day minutes, total evening minutes, total night minutes, total international minutes(and corresponding calls/charges)

**Relevance:** Quantify service consumption; high usage/charges could indicate satisfaction or, conversely, dissatisfaction due to high bills.

**2. Customer Service Interactions:** customer service calls

**Relevance:** A high number of calls often indicates unresolved issues or dissatisfaction, serving as a strong indicator of churn.

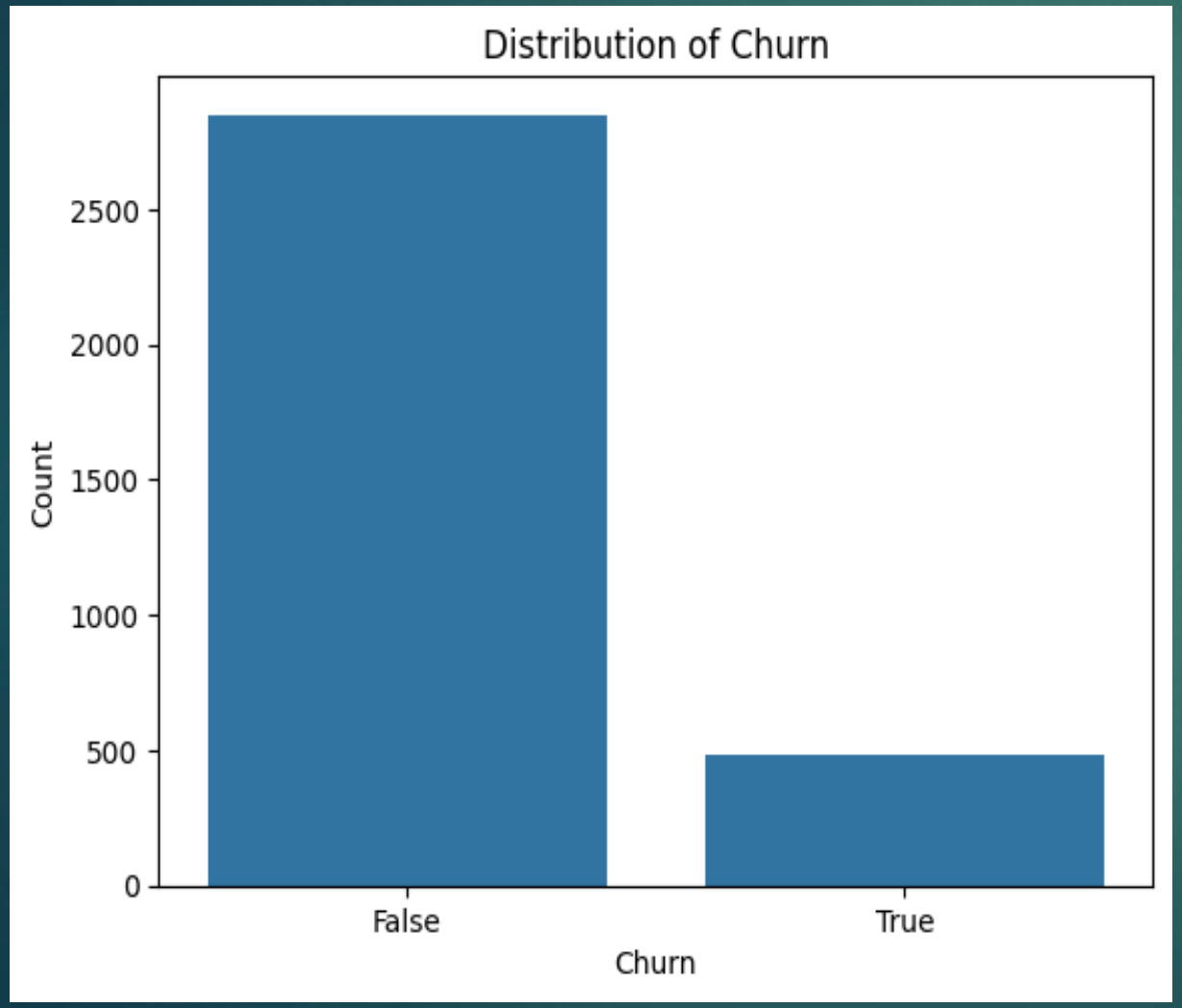
**3. Plan Subscriptions:** international and voicemail plan.

**Relevance:** The presence or absence of these plans can affect perceived value and churn rates.

**4. Account Details:** account length, state

**Relevance:** Longer acc correlates with lower churn. Geographic factors also play a role.

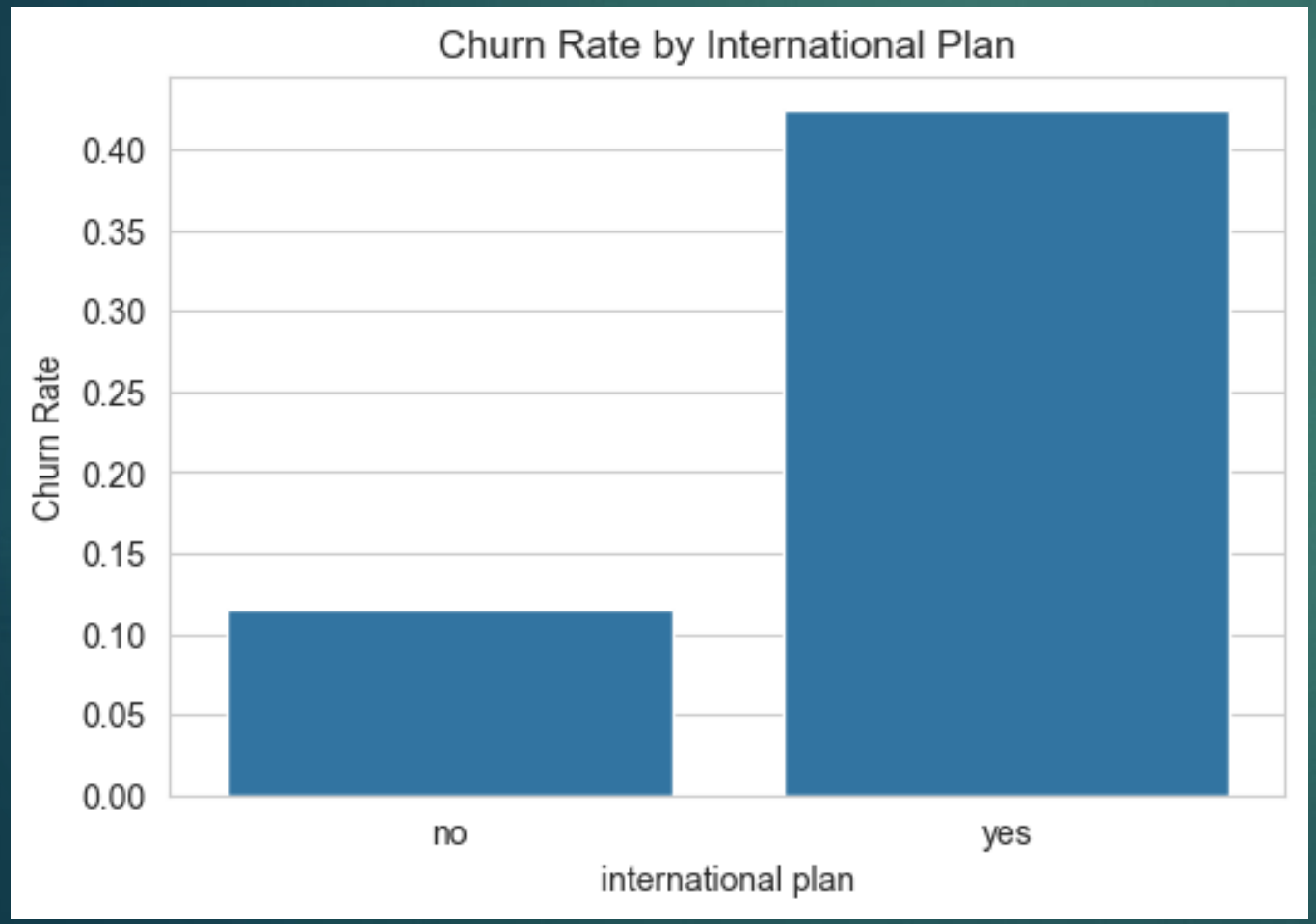
# DISTRIBUTION OF TARGET VARIABLE




Observations:

- There is a significant imbalance between the two classes. The majority of customers have remained with the service (churn = False), while a much smaller proportion have churned (churn = True).
- False (Not Churned): Approximately 2,800 customers
- True (Churned): Approximately 500 customers

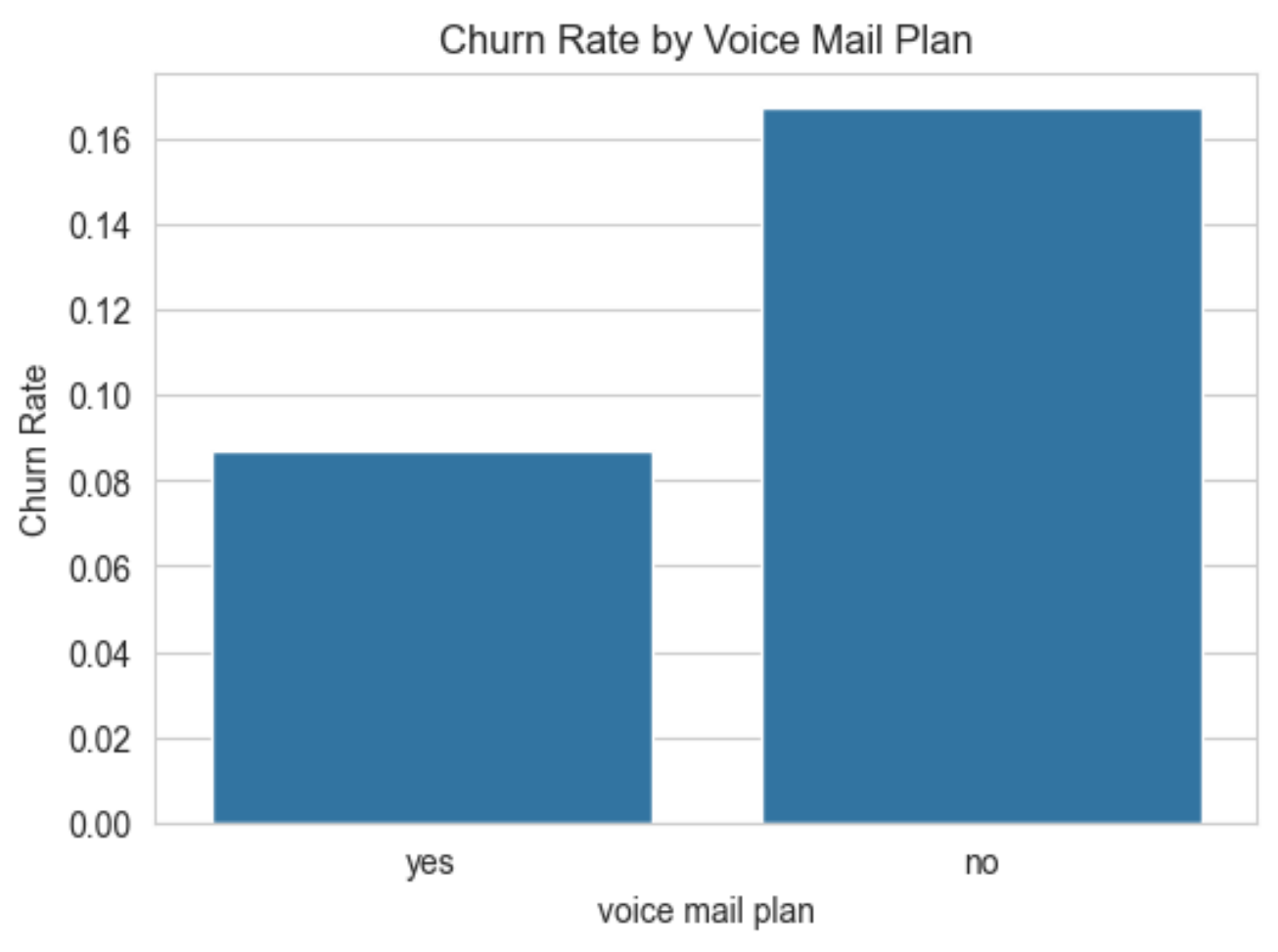
# CHURN RATE BY INTERNATIONAL PLAN




Observations:

International Plan
- Customers with an international plan have a higher churn rate, 42.4%
- Customers without an international plan have a much lower churn rate, 11.5%
- ➤ This indicates that subscribing to the international plan is a strong churn signal.

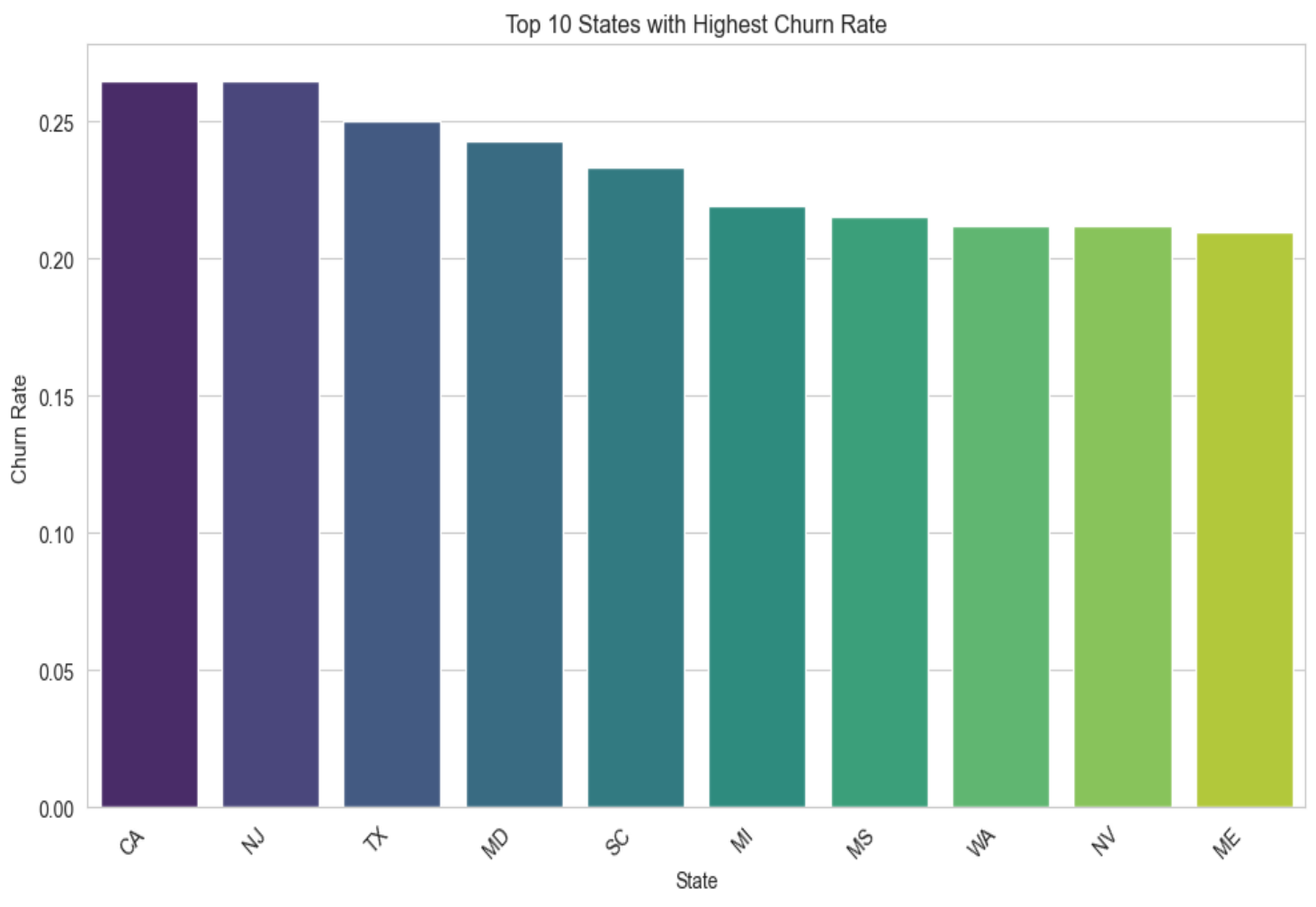
# CHURN RATE BY VOICE MAIL PLAN




Observations:

Voice Mail Plan
- Customers with a voice mail plan churn less, 8.7%
- Customers without a voice mail plan churn more 16.7%
- ➤ This suggests that voice mail features may be linked to higher customer satisfaction or retention.

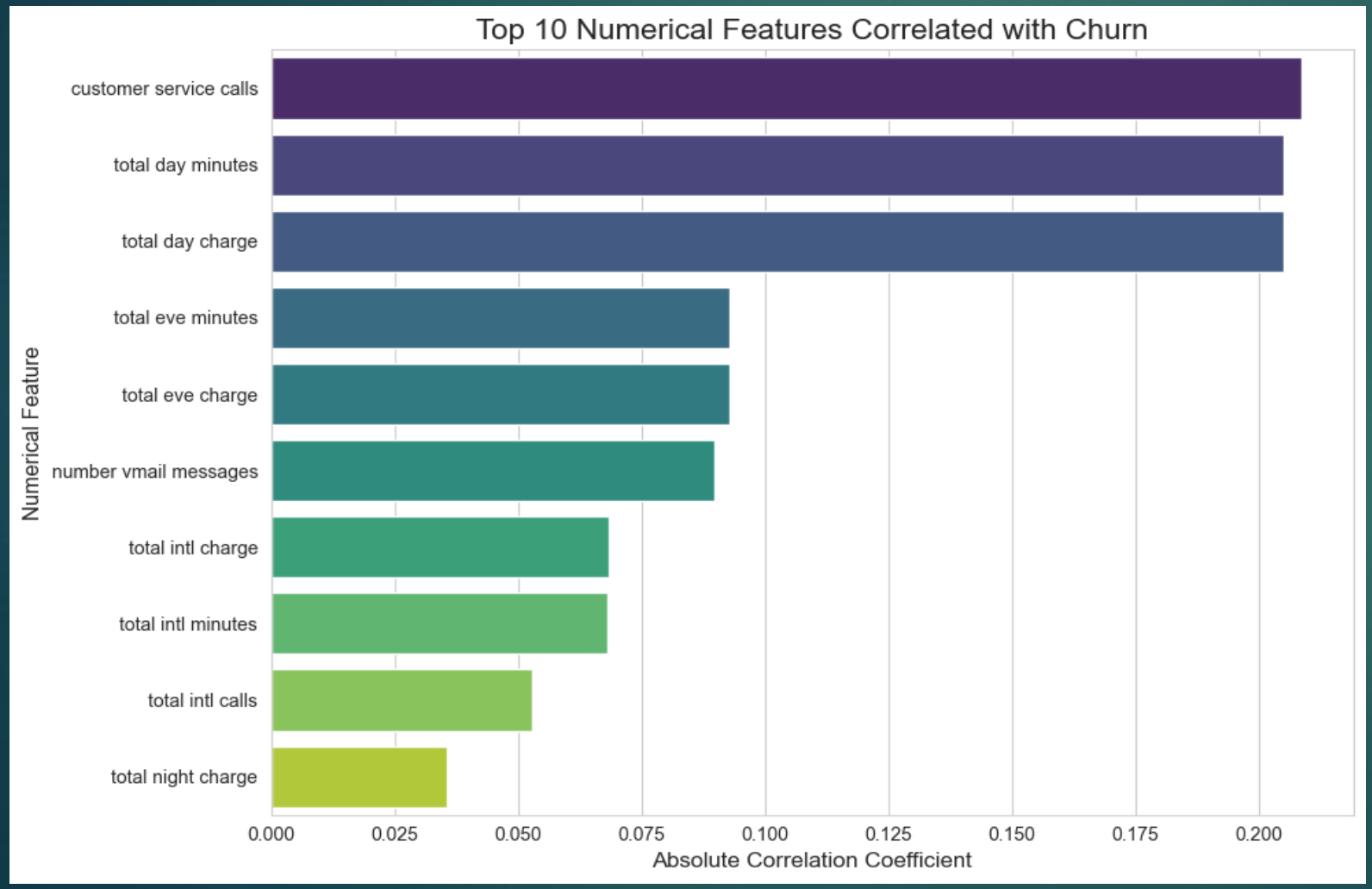
# CHURN RATE BY TOP 10 STATES




Observations:

States like **New Jersey, California, and Texas** have more customers leaving SyriaTel. This means there could be **local problems** like poor service, tough competition, or unhappy customers in those areas.

# TOP FEATURES BY CHURN RATE




Observations:

- Customers with **4+ service calls** are most likely to churn — a sign of unresolved issues.
- Churned customers show **higher day and evening call usage**.
- Heavy usage may indicate **billing frustration** or **poor service quality**.
- Key churn indicators point to **dissatisfaction and unmet expectations**.

# DATA PREPARATION

**Approach to Churn Prediction**

**1. Data Preprocessing:**

- Feature Engineering: Creating new features from existing ones.
- Encoding: Converting categorical variables to numerical.
- Scaling: Normalizing numerical features.

**2. Handling Imbalanced Data:**

- Churn datasets are often imbalanced (fewer churners).
- **SMOTE (Synthetic Minority Over-sampling Technique)** was used to balance the dataset, preventing the model from being biased towards the majority class.

**3. Model Selection & Training:**

- Explored various classification models (e.g., Logistic Regression, Decision Tree, Random Forest).
- **Random Forest** emerged as the top performer.
- **Hyperparameter Tuning:**
- **GridSearchCV** was used to optimize model parameters for best performance, specifically targeting metrics relevant for imbalanced data.

# MODELLING

**Three models were used to predict customer churn:**

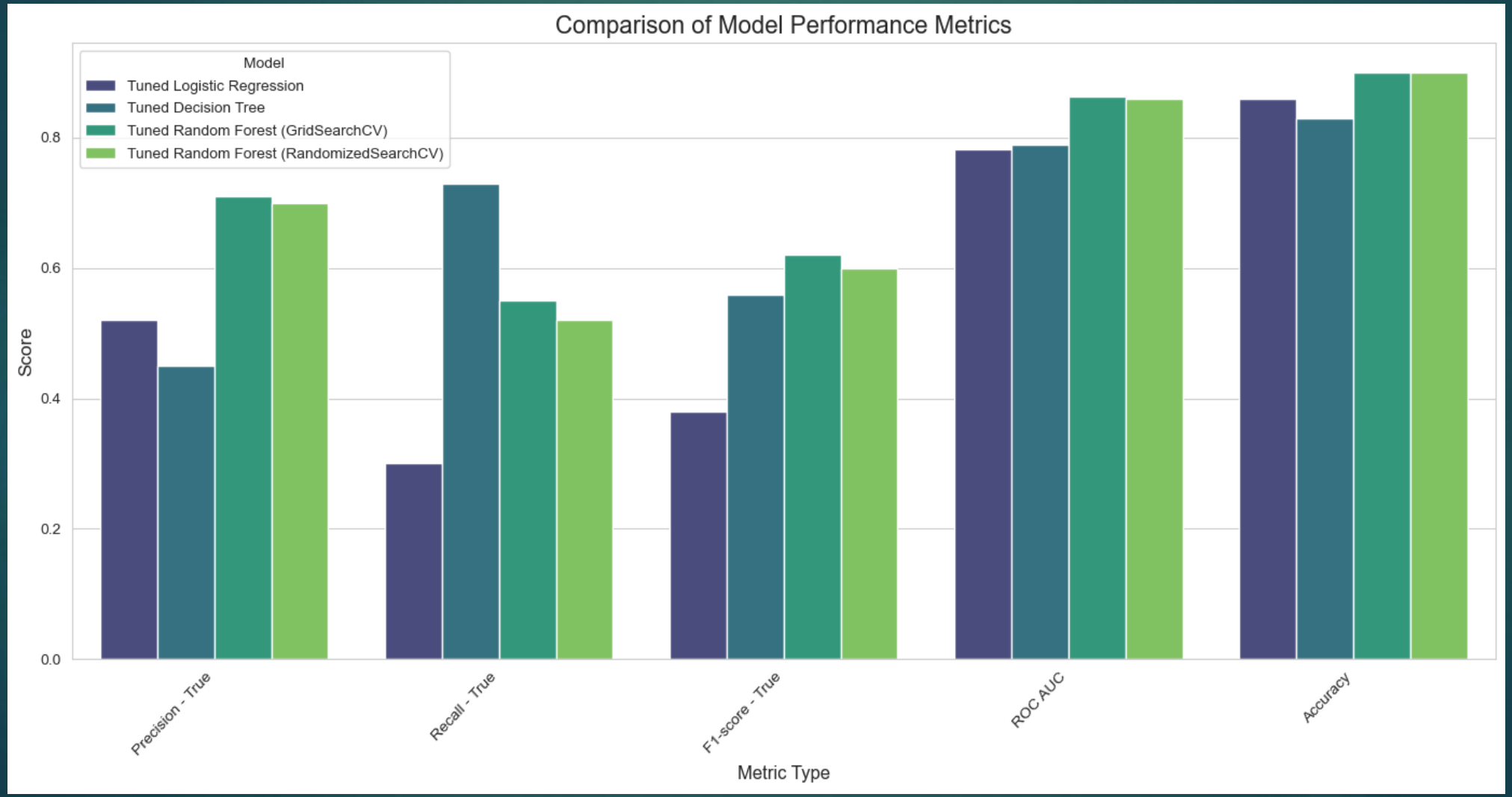
1. **Logistic Regression** – A simple model to establish a baseline.
2. **Decision Tree** – Easy to interpret, but less accurate.
3. **Random Forest** – An advanced model that combines many decision trees for better accuracy.

- **Best Model: Tuned Random Forest Classifier**

- It gave us the **highest performance**, especially on:

- **F1-Score** – Balances catching churners (recall) without too many false alarms (precision).
- **ROC AUC** – Measures how well the model separates churners from non-churners.
- **Why it matters**: The Random Forest model **identified churners more accurately** than the others, making it our best tool for targeting customers at risk of leaving.

# COMPARISON OF THE MODELS



# COMPARISON SUMMARY

We tested and compared several models to find the most effective way to predict customer churn at SyriaTel:

- **Best Overall Model**: The **Tuned Random Forest** (via GridSearch) gave the strongest, most balanced results. It was best at identifying customers likely to churn **while minimizing wrong predictions**.
- **Decision Tree**: This model was **great at catching churners** but came with more false alarms (mistargeting loyal customers).
- **Logistic Regression**: Simple and easy to interpret, but **less accurate** than tree-based models in this case.
- **Key Metrics**:
    - **F1-Score**: Tuned Random Forest had the highest score — meaning it best balanced precision and recall.
    - **ROC AUC**: It also had the best ability to separate churners from non-churners.
    - **Precision**: Helped avoid mistargeting loyal customers.
    - **Recall**: Captured true churners effectively.

**Conclusion**: The **Tuned Random Forest** model is our top choice. It strikes the best balance between identifying churners and avoiding costly errors — making it the most practical and reliable tool for reducing churn at SyriaTel.

# MODELS PERFORMANCE

| Model | Accuracy | Precision (True) | Recall (True) | F1-Score (True) | ROC AUC Score |
|---|---|---|---|---|---|
| Logistic Regression | 85% | 49% | 36% | 41% | 76.25% |
| Tuned Logistic Regression | 86% | 52% | 30% | 38% | 78.26% |
| Decision Tree | 83% | 45% | 73% | 56% | 78.97% |
| Tuned Decision Tree | 83% | 45% | 73% | 56% | 78.97% |
| Random Forest | 90% | 68% | 54% | 60% | 85.81% |
| Tuned Random Forest | 90% | 71% | 55% | 62% | 86.25% |
| Tuned RF (RandomSearchCV) | 90% | 70% | 52% | 60% | 85.90% |

# MODELS PERFORMANCE: INSIGHTS

- **Best Performing Model**: The **Tuned Random Forest** model outperformed others with:
  - The highest **ROC AUC score** (0.8625)
  - Strong balance between **accuracy** (90%) and **recall** (55%) for churned customers
- **Why Random Forest Wins**:
  - Unlike Logistic Regression, which struggles with recall (36%), Random Forest better captures true churners (Recall: 55%)
  - High accuracy ensures overall reliability, while improved F1-score (0.62) signals better consistency for minority class
- **Business Impact**:
  - Identifying 55% of potential churners can help SyriaTel proactively intervene (e.g., offer incentives, personalized plans)
  - Reducing churn directly saves revenue and improves customer lifetime value
- **Next Steps**:
  - Deploy the Tuned Random Forest model into production
  - Regularly retrain with updated customer data
  - Use model insights to segment high-risk customers for retention strategies

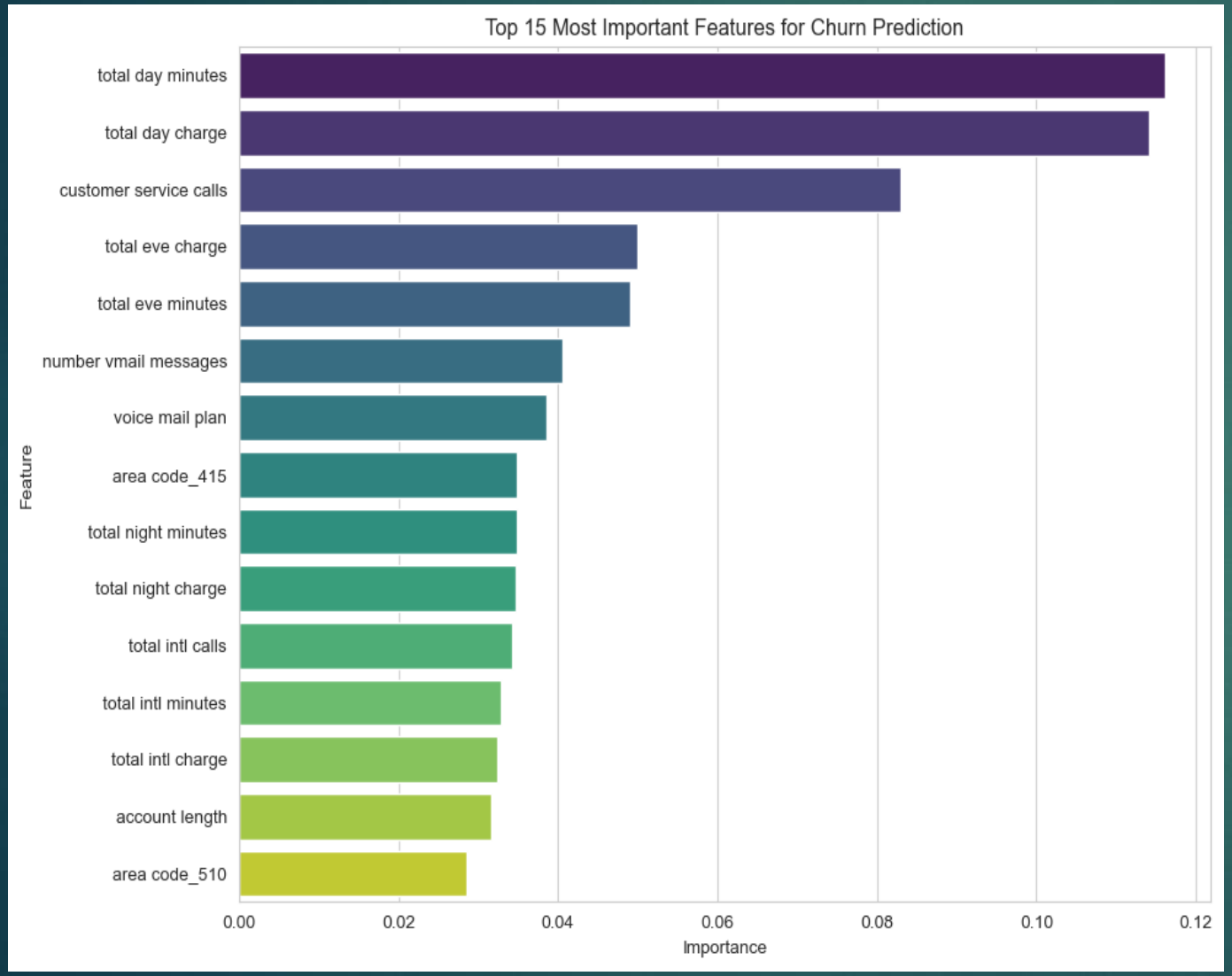
# EVALUATION

In this section, we see how well our model performed and what drives customer churn.

- First, we'll look at the **confusion matrix** to understand how accurately the model predicts churners versus non-churners.
- Then, we'll review the **top features** influencing those predictions — key customer behaviors that signal churn risk.

These insights help us know not just how the model performs, but also **why** customers might be leaving.

# FEATURE IMPORTANCE




**Observations:**

- **Daytime usage** (minutes & charges) is the **strongest churn indicator**.
- **Frequent customer service calls** suggest dissatisfaction → higher churn.
- **Evening usage** has a moderate effect on churn.
- **Voicemail plans** are linked to **lower churn** → a retention opportunity.
- Customers with **international plans** are **more likely to churn**.
- Other moderate factors: **account length**, **area code**, **night & intl. usage**.

**Action:** Focus on improving daytime service, support quality, and promoting voicemail features.

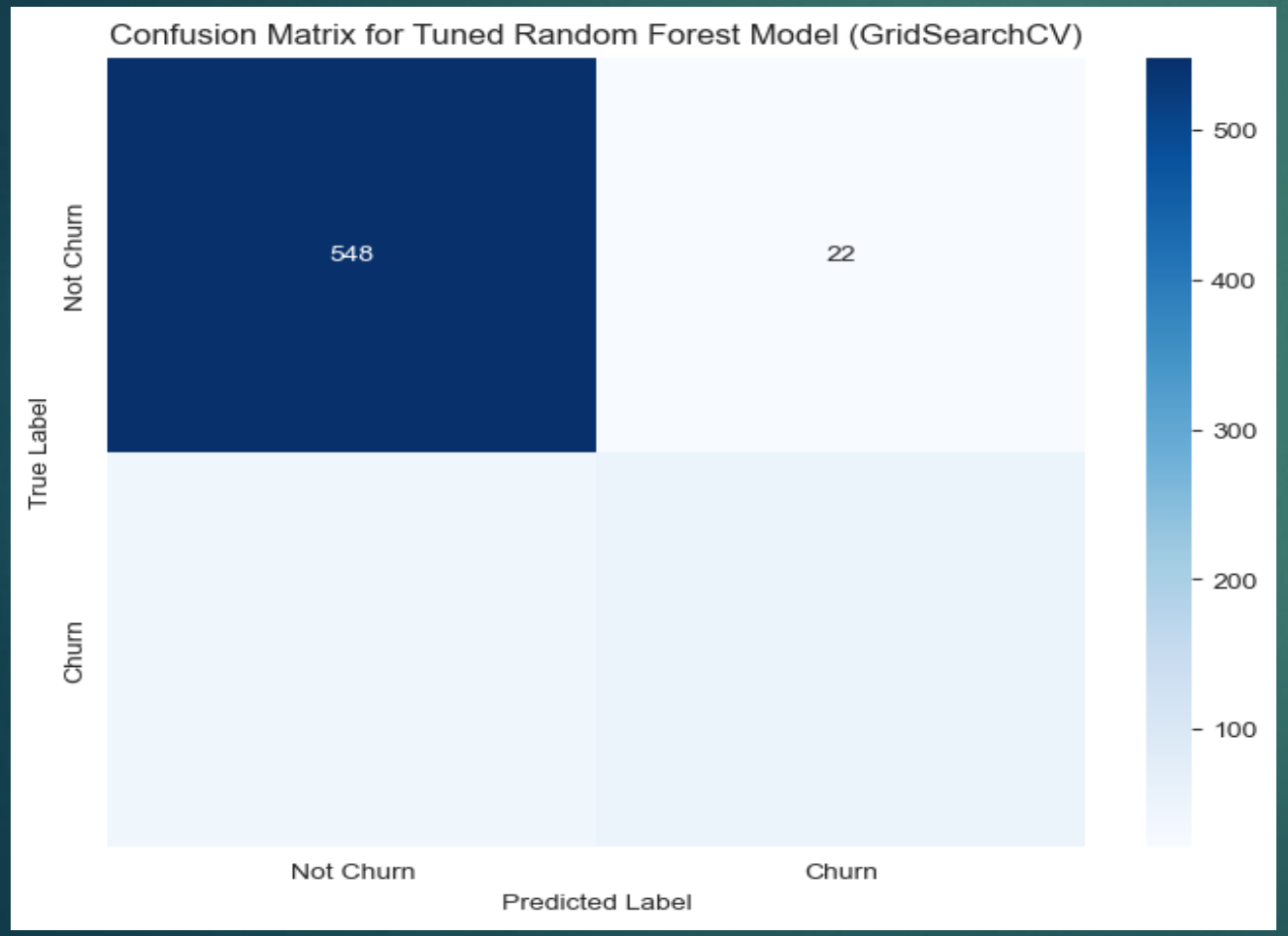
# FEATURE IMPORTANCE

| Feature | Business Meaning | Insight |
|---|---|---|
| **Total Day Minutes / Charge** | Reflects high daytime phone usage | High usage may signal billing dissatisfaction or better offers elsewhere. |
| **Customer Service Calls** | Measures how often a customer contacts support | Frequent calls often indicate service issues — a strong churn signal. |
| **Voicemail Plan & Messages** | Reflects communication preferences | Less voicemail use may mean disengaged or tech-savvy customers. |
| **Evening/Night Usage & Charges** | Indicates off-peak engagement | Customers with lower night/eve usage may be less tied to the service. |
| **International Calls & Charges** | Reflects global connectivity needs | These users may churn if international plans are expensive or unreliable. |
| **Area Code** | Geographic customer grouping | Certain regions (e.g. 415, 510) may have higher churn — possibly due to competition or service quality. |
| **Account Length** | Tenure with SyriaTel | Shorter-tenure customers are typically more likely to churn. |

# BUSINESS TAKEAWAYS

- **Monitor heavy daytime users**: Offer them tailored plans to match their usage.
- **Investigate service call reasons**: High customer service contact is a red flag — consider proactive outreach.
- **Segment by area code**: Identify geographic churn patterns and address regional competition or service issues.
- **Design plans for international users**: Keep high-value international users engaged with competitive offers.

# CONFUSION MATRIX




- **Observations:**

- **Correctly identified loyal customers**: 548
- **Incorrectly flagged loyal customers (false positives)**: 22
- **Missed churners (false negatives)**: 44
- **Correctly identified churners**: 53
  - **Precision: 71%** – Most flagged churners were truly at risk
    **Recall: 55%** – The model caught over half of actual churners

# MATRIX SUMMARY

**What This Means:**

- **Recall (Churn Captured)**: We caught **55% of actual churners**.
- **Precision (Correct Churn Predictions)**: Of those we flagged as churners, **71% were truly at risk**.

**Business Insight**

To reduce customer loss:

- If **every lost customer is expensive**, SyriaTel might want to **focus more on catching more churners (higher recall)** — even if it means more false positives.
- If **resources for retention are limited**, focusing on **more accurate churn predictions (higher precision)** may be better.

# RECOMMENDATIONS

🎯 **Targeted Retention Campaigns**

- Identify high-risk customers based on **daytime usage**, **service calls**, and **international plan** status.
- Offer **personalized incentives** (e.g., discounted plans, loyalty bonuses) to encourage loyalty.

☎️ **Improve Customer Service**

- Analyze common issues behind **frequent customer service calls**.
- Implement **proactive outreach** to resolve recurring problems before customers churn.
- Focus on improving **first-call resolution rates**.

🌍 **Re-evaluate International Offerings**

- Investigate why customers with **international plans** have a higher churn risk.
- Adjust pricing, features, or marketing to better align these plans with customer expectations.

📊 **Monitor Usage Patterns**

- Set alerts for **sudden spikes or drops in total day usage** as potential signals of dissatisfaction or churn risk.

# NEXT STEPS

- **Deploy the Tuned Random Forest Model**

•Integrate it into SyriaTel's CRM system for real-time churn prediction.

- **Set Up Alerts for High-Risk Customers**

•Automatically flag customers with high churn probabilities for proactive engagement.

- **Refine Retention Campaigns**

•Use model insights to tailor personalized retention offers (e.g., loyalty rewards, discounts).

- **Monitor & Re-train**

•Schedule regular model performance checks.
•Retrain model quarterly as new customer behavior data becomes available.

- **Explore Additional Features**

•Consider incorporating **billing history**, **network quality**, or **competitor activity** for deeper insights.

# THANKS

- Do you have any questions?

- Contact info:

lilianwangui266@gmail.com

https://github.com/Wangui10/