

# Customer Churn Prediction for SyriaTel: A Machine Learning Approach to Enhance Retention Strategies.



# Background

SyriaTel, operating in a competitive telecommunications market, faces significant challenges with customer churn, which negatively impacts both revenue and brand loyalty. As the cost of acquiring new customers rises, retaining existing ones becomes crucial for sustaining long-term growth and profitability. Understanding the factors that contribute to customer churn is essential for devising effective retention strategies. By leveraging customer behavior and demographic data, this analysis seeks to develop a predictive model that can accurately identify customers at risk of leaving. The ultimate goal is to enable SyriaTel to take proactive measures, reduce churn rates, and mitigate the associated financial consequences.



# Problem Statement

- ▶ SyriaTel is confronted with the challenge of retaining customers in a highly competitive telecommunications market. Customer churn results in revenue loss and negatively impacts the company's reputation and market position. The goal is to create a predictive model that can accurately identify customers at risk of churning, allowing SyriaTel to take proactive steps with targeted retention strategies.



# Objectives

- I. Minimize customer churn
- II. Create a predictive model to curb customer churn
- III. Help the company maximize on profits
- IV. Help the company identify reasons for customer churn so that they can make remedial actions where necessary.

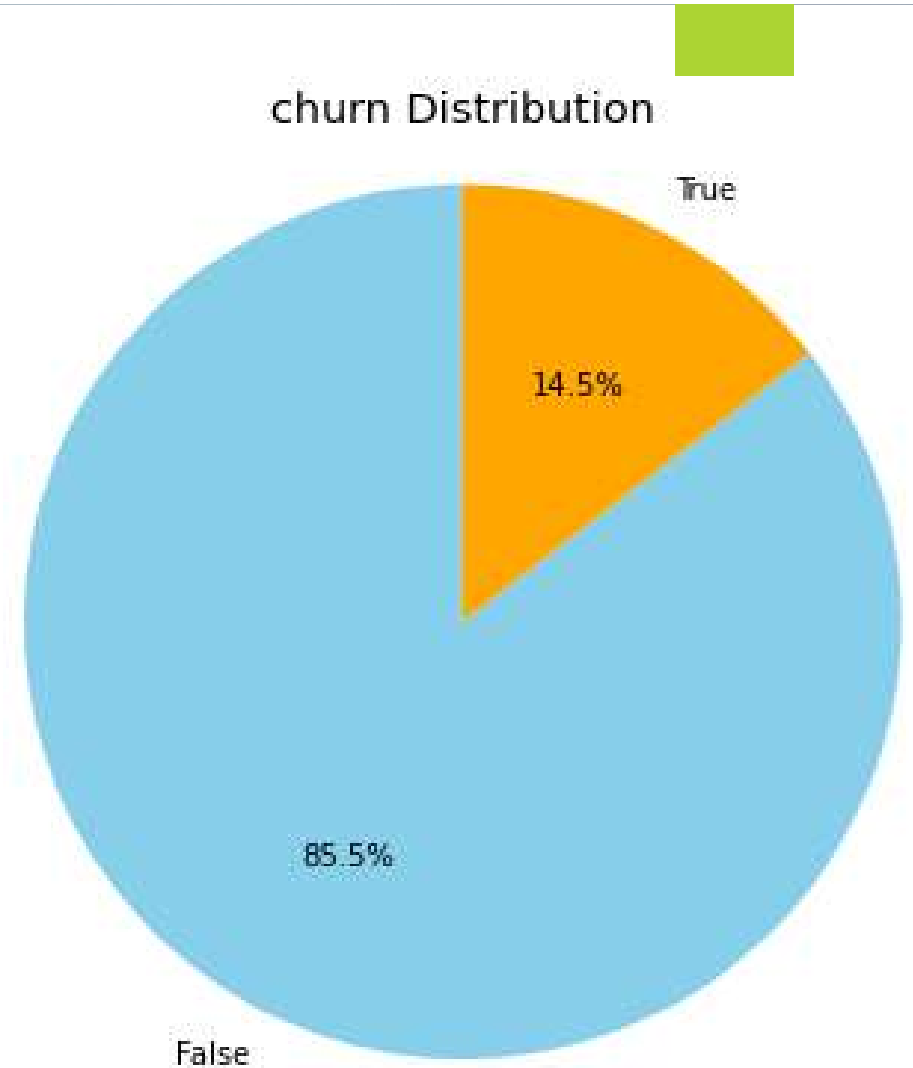


# DATA UNDERSTANDING

- ▶ The data is sourced from <https://www.kaggle.com/datasets/becksddf/churn-in-telecoms-dataset>
- ▶ The dataset has 3333 rows and 21 columns and has no null values or duplicates

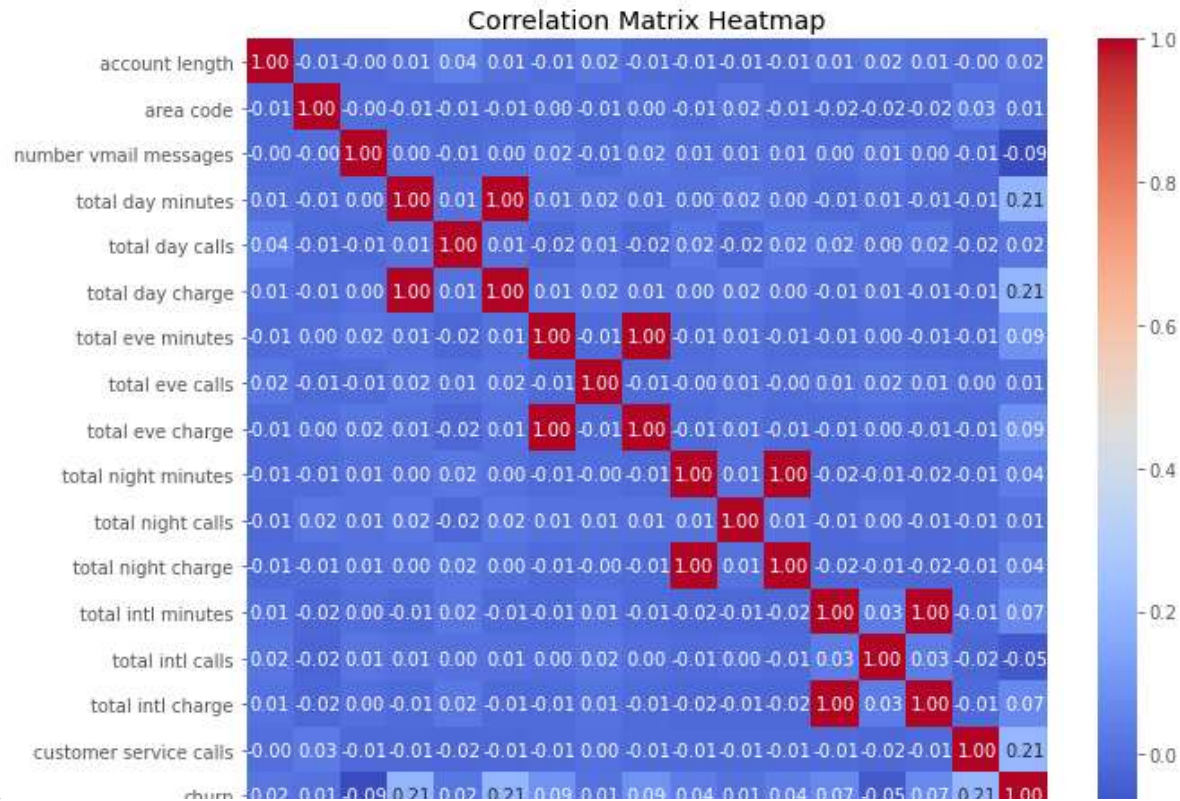
# Data visualization

This pie chart shows the distribution of churn in the dataset. The majority class (85.5%) represents customers who did not churn, while the minority class (14.5%) represents customers who did churn.





# Correlation



# Correlation explained

- ▶ This correlation matrix heat map visualizes the relationships between various features in the dataset, including the target variable churn.
- ▶ Churn Relationships: The churn variable shows a moderate positive correlation with features like total day minutes, total day charge, and customer service calls, indicating these features may be significant predictors of churn.



# Modeling

- ▶ This project has used 3 models to make predictions. The project has used linear regression, decision trees and random forest model.
- ▶ The aim is to establish which is the best predictive model amongst the 3.

# 1. Logistic regression model

- ▶ The performance metrics for your Logistic Regression model indicate excellent predictive performance:
  - ▶ Accuracy (96.67%): This high accuracy suggests the model correctly predicts churn and non-churn cases most of the time.
  - ▶ Precision (97.00%): A precision score of 97% indicates that when the model predicts a customer will churn, it is correct 97% of the time.
  - ▶ Recall (96.67%): This score implies that the model correctly identifies 96.67% of the actual churn cases, showing it is effective in minimizing false negatives.
  - ▶ F1-score (96.68%): The F1-score balances precision and recall, confirming that the model achieves a strong trade-off between these metrics.
- AUC (1.0): The perfect Area Under the Curve (AUC) score suggests the model is excellent at distinguishing between churn and non-churn customers.

## 2. Decision tree

- ▶ Precision (97.00%): The model is highly precise, meaning it correctly identifies 97% of the customers it predicts will churn.
- ▶ Recall (96.67%): The model successfully captures 96.67% of the actual churn cases, ensuring most at-risk customers are flagged for retention efforts. High recall is critical for minimizing the number of customers who churn undetected.
- ▶ Accuracy (96.67%): The model's overall performance is strong, with nearly 97% of predictions being correct. This reflects a well-balanced model that performs effectively.

### 3. Random forest

- ▶ Accuracy (100%): The model correctly predicted all instances, achieving perfect accuracy.
- ▶ Precision (100%): Every customer flagged as likely to churn was indeed a churner, with no false positives.
- ▶ Recall (100%): The model captured all actual churners, leaving no false negatives.
- ▶ F1-score (100%): A perfect balance between precision and recall, indicating exceptional predictive performance.
- ▶ ROC AUC Score (100%): The model perfectly distinguishes between churners and non-churners, showcasing outstanding reparability.

# Conclusion

- ▶ From the above models, random forest model is performing better than the logistics and decision tree model.
- ▶ Recommendations:
  - ▶ 1.Introduce loyalty programs that reward long-term customers with benefits such as free upgrades, exclusive services, or special discounts.
  - ▶ 2.Customers with frequent customer service calls may indicate unresolved issues. Streamline complaint resolution processes to address their concerns promptly.
  - ▶ 3.Design personalized offers such as discounts, loyalty rewards, or improved service plans to retain these customers