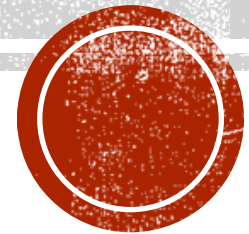


CUSTOMER CHURN PREDICTION PROJECT

BY: Rose Syomiti Muthini

Date: 22/07/2025



OVERVIEW

- Loss of existing customers is a major challenge for telecommunications companies like SyriaTel. Acquiring new customers is often more costly than retaining current ones, and high churn rates negatively impact revenue, profitability, and market share.
- This project aims to leverage historical customer data to predict churn, enabling SyriaTel to take proactive measures that improve customer retention, satisfaction, and competitive positioning.



BUSINESS PROBLEM

- SyriaTel is experiencing significant customer churn, leading to revenue loss and declining market share.
- Without a clear understanding of churn patterns and high-risk customer segments, the company struggles to develop effective retention strategies.
- There is need to predict key drivers to customer churn and which customers are at high risk of leaving.



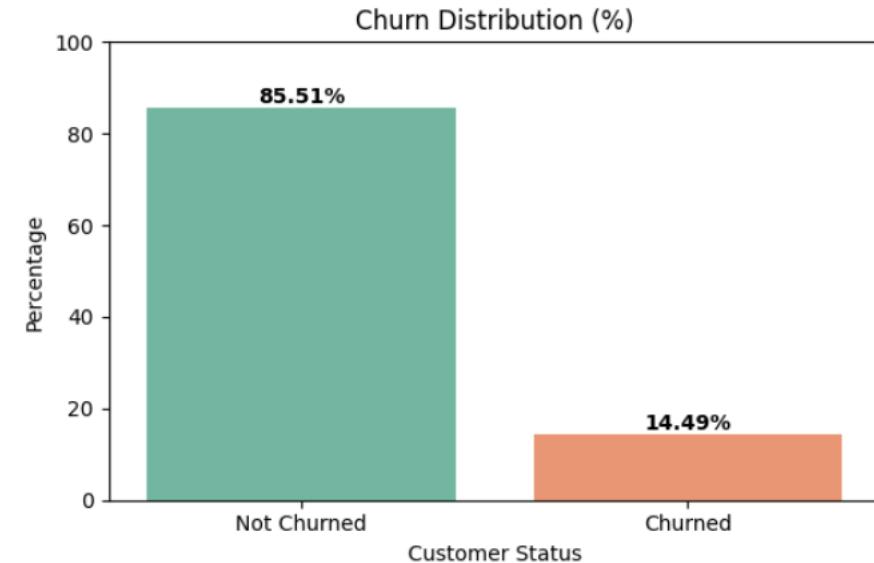
OBJECTIVE

- Predict which customers are likely to churn.
- Identify the most important factors driving churn.
- Deliver actionable insights to support targeted retention campaigns and improve overall customer loyalty.



DATA UNDERSTANDING

- The data contain 3,333 records with no missing values.
- It contain information on customer behavior and usage metrics to predict customer churn.
- Churn is the binary variable indicating whether the customer has left the company.



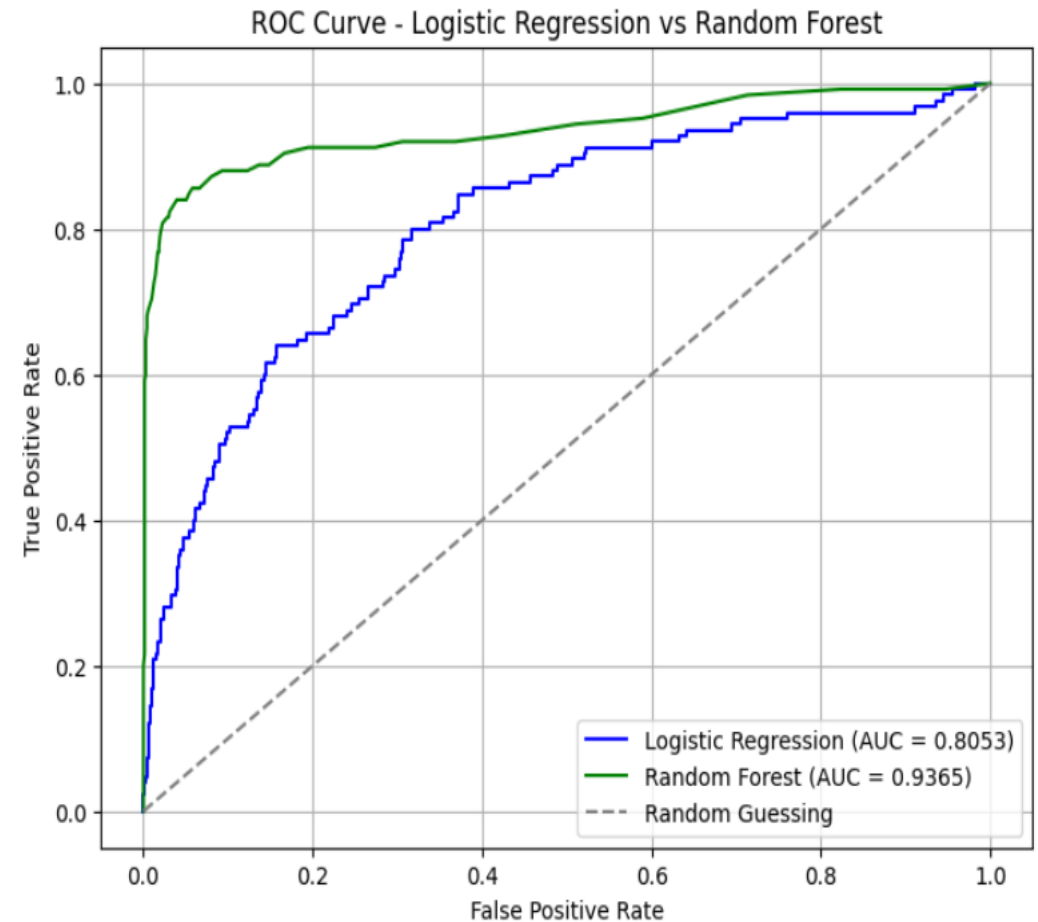
MODELLING

- In this project, I used **two machine learning models** to predict customer churn:
 1. **Logistic Regression**, which is suitable for binary classification tasks like churn prediction and offers good interpretability.
 2. **Random Forest Classifier**, a robust ensemble model that handles nonlinear patterns and class imbalance effectively.



EVALUATION

- **Logistic Regression** performed well on non-churn predictions, with high accuracy overall.
- However, it struggled to identify churners, showing low recall and precision for that class.
- **Random Forest Classifier** demonstrated strong performance across both churn and non-churn classes.
- Achieved higher precision and recall for the churn class compared to logistic regression.



CONCLUSION

- **Logistic Regression** is simple, interpretable, and performs well for non-churn cases, but is less effective at detecting churners.
- **Random Forest** outperformed Logistic Regression, especially in handling class imbalance and correctly identifying churners.



RECOMMENDATIONS

- Based on the evaluation results, I recommend adopting the Random Forest model for predicting customer churn at SyriaTel.
- It delivers stronger performance across both classes, especially in detecting churners and handles class imbalance more effectively than Logistic Regression.



RECOMMENDATIONS

- Based on key features affecting churn:

1. Review International Plan Offering: Users with an **international plan** were more likely to churn

- Action: Reevaluate pricing, usage benefits, or customer targeting for this plan.

2. **Target High Usage Customers Proactively:** Higher **daytime minutes and charges** were associated with churn.

Action: Provide value-added offers or loyalty programs for heavy users to boost retention.

3. Leverage Predictive Modeling: Use the trained model (Random Forest) to **identify at-risk customers in real time**.

Action: Implement a churn risk dashboard and trigger timely interventions (e.g., special offers)



THANK YOU

Author: Rose Syomiti Muthini

