Telecommunication Customer Churn

A case study of SyriaTel

Moringa School Phase 3 Project

Problem Statement and definition

- SyriaTel is a telecommunications company with **at least more than 3000 subscribers**. The company offers a variety of services which include normal local calls, international calls and voicemail. However, the market conditions seem to make blows to the company quite frequently with **a noted customer churn**. This poses a threat to SyriaTel as it would mean **low turnover and ultimate business decline.**

- In this regard, SyriaTel have shared their customer dataset that would help in **understanding the different patterns portrayed.** Further, the company is **interested in reducing how much money is lost because of customers who don't stick around very long.**

- This project uses **binary classification** to create and predict models that help define the patterns and suggest a resolution in how money lost can be reduced in SyriaTel company.

Case Study objectives

- **Identify Key Factors Influencing Customer Churn:** Determine the most significant factors that contribute to customer churn.
- **Build a Predictive Model for Customer Churn:** Develop and validate a machine learning model to predict whether a customer will churn.
- **Develop Customer Retention Strategies:** Formulate actionable strategies to retain customers identified as high risk for churn.

Data understanding

SyriaTel Customer Churn dataset has:

- 3333 rows
- 21 columns

Data variables

- **account length:** The number of days the account has been active.
- **area code:** The area code of the customer's phone number.
- **phone number:** The customer's phone number.
- **international plan:** Whether the customer has an international plan.
- **voice mail plan:** Whether the customer has a voice mail plan.
- **number vmail messages:** Number of voice mail messages.
- **total day minutes, total day calls, total day charge:** Usage metrics during the day.
- **total eve minutes, total eve calls, total eve charge:** Usage metrics during the evening.
- **total night minutes, total night calls, total night charge:** Usage metrics during the night.
- **total intl minutes, total intl calls, total intl charge:** International usage metrics.
- **customer service calls:** Number of calls to customer service.
- **churn:** Whether the customer has churned or not (target variable).

Methodology

Exploratory Data Analysis(EDA)

The dataset was explored through different aspects encompassing shape, statistical description, missing data, outliers, visualization and polished for data preprocessing.

Data Preprocessing

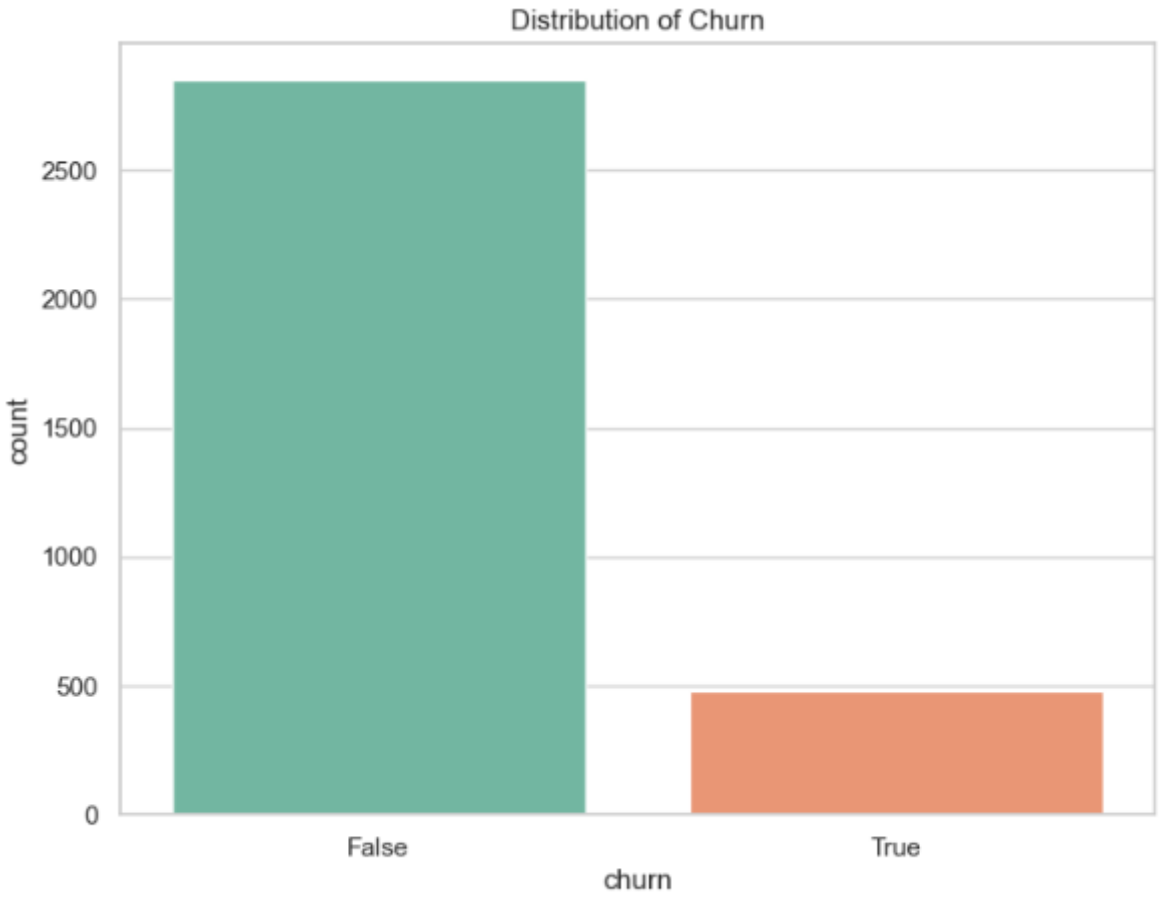
The cleaned data was then preprocessed through encoding of numeric and categorical features, scaling and train-test split ready for model evaluation

Data Modeling

Different data models were performed and fine-tuned to define the best model that would suit SyriaTel's problem solution.

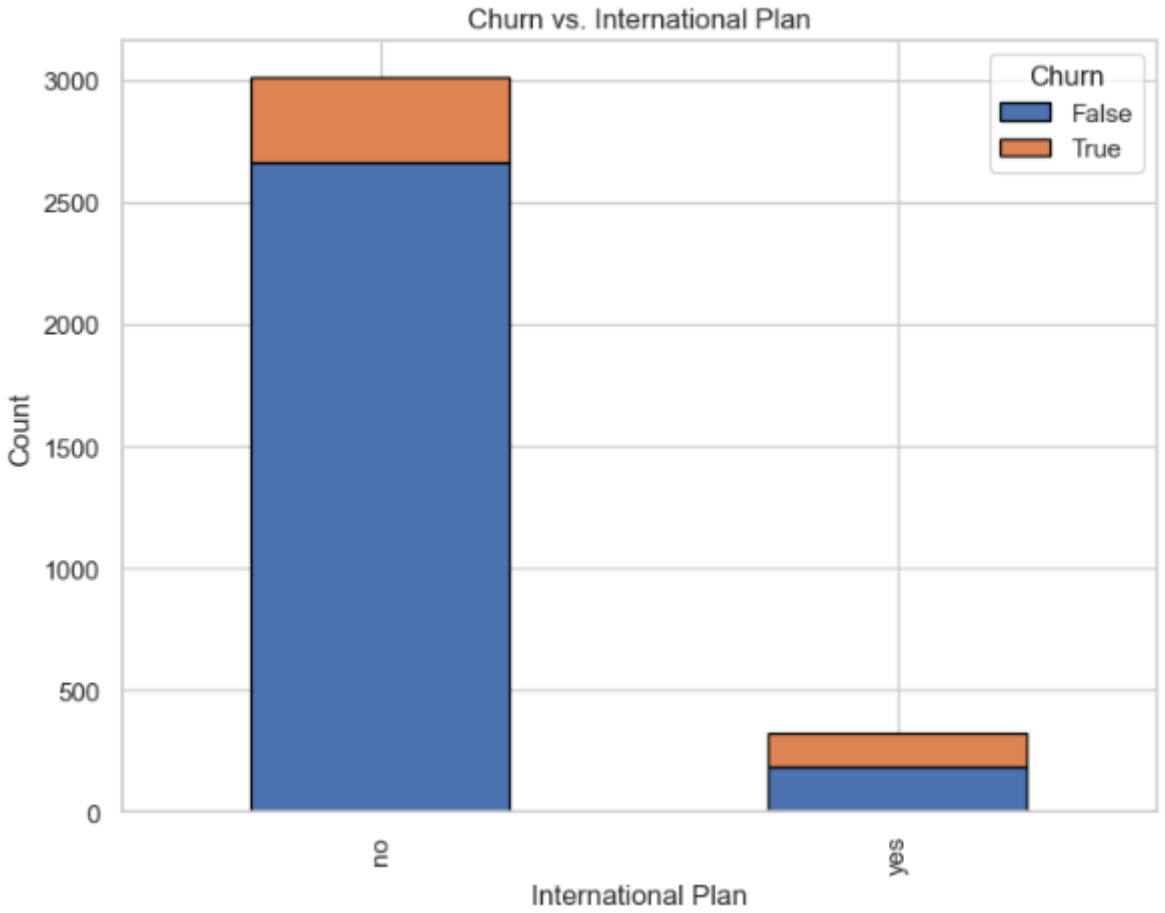
Observations

Observations 1/4: Churn distribution



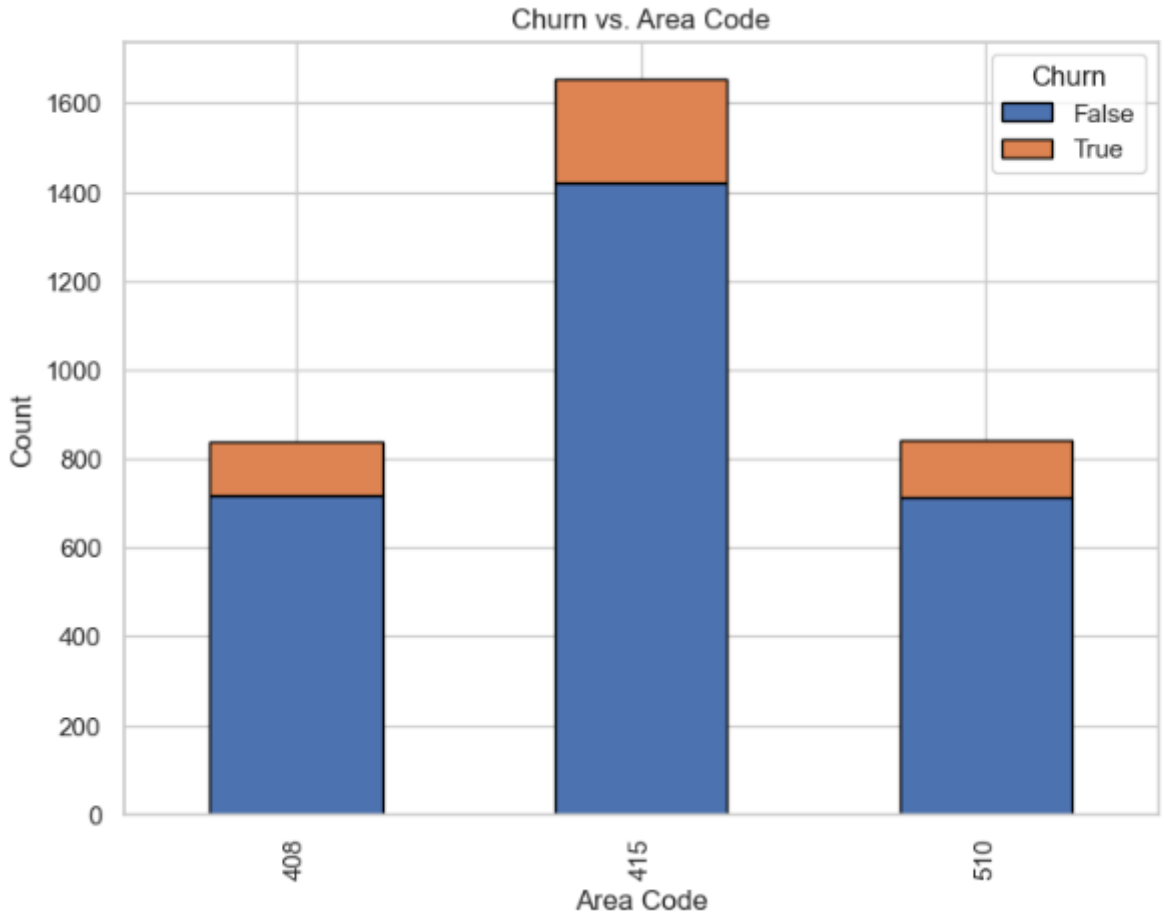
According to the churn distribution among the subscribers, **about 500** have exited SyriaTel which is a worry to the company.

Observations 2/4: Churn vs international plan



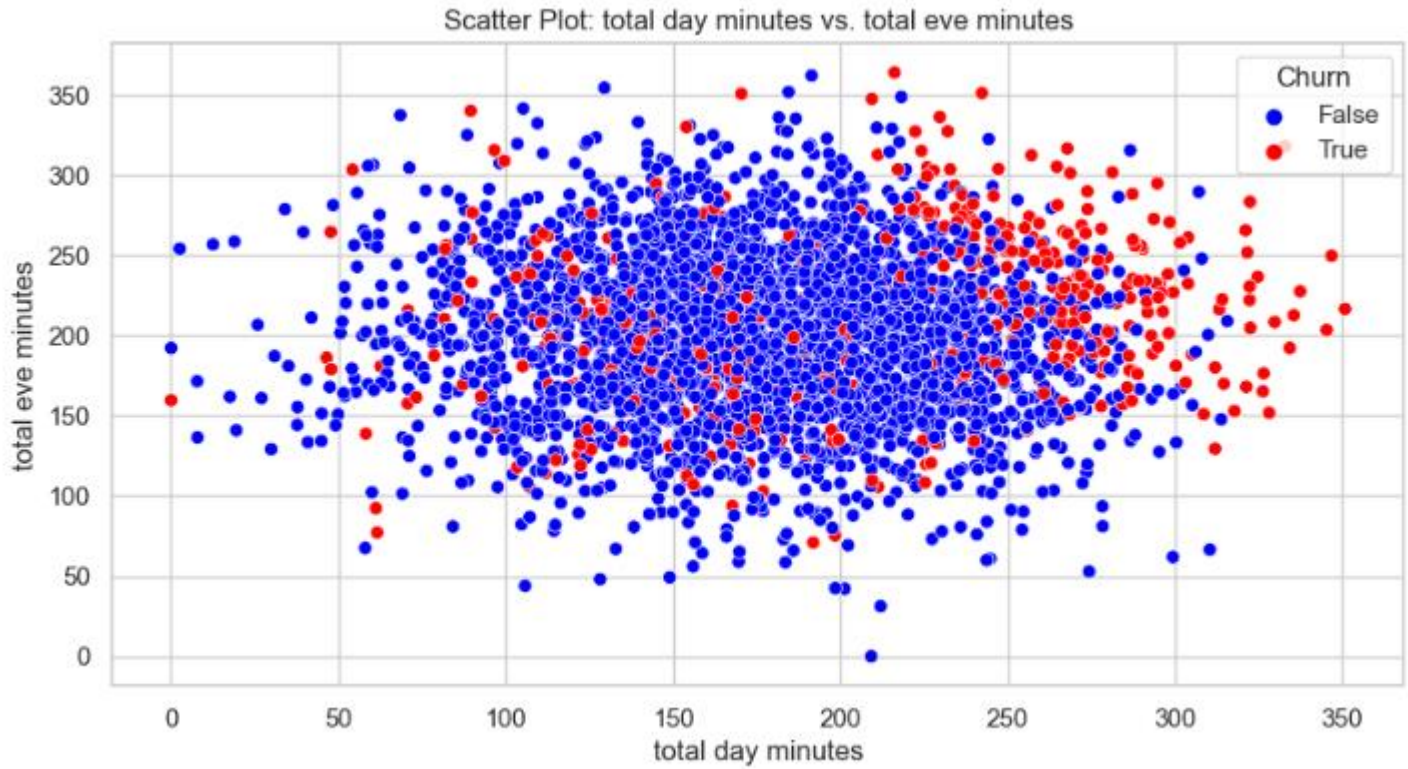
- A large proportion of the churn customers have no international plan.
- However, nearly half of those with the international plan have churned, which is a risk to the business in its international plan entity.

Observations ¾: Churn vs area code



- Area code 415 has the most subscribers and most exited (churn) customers.

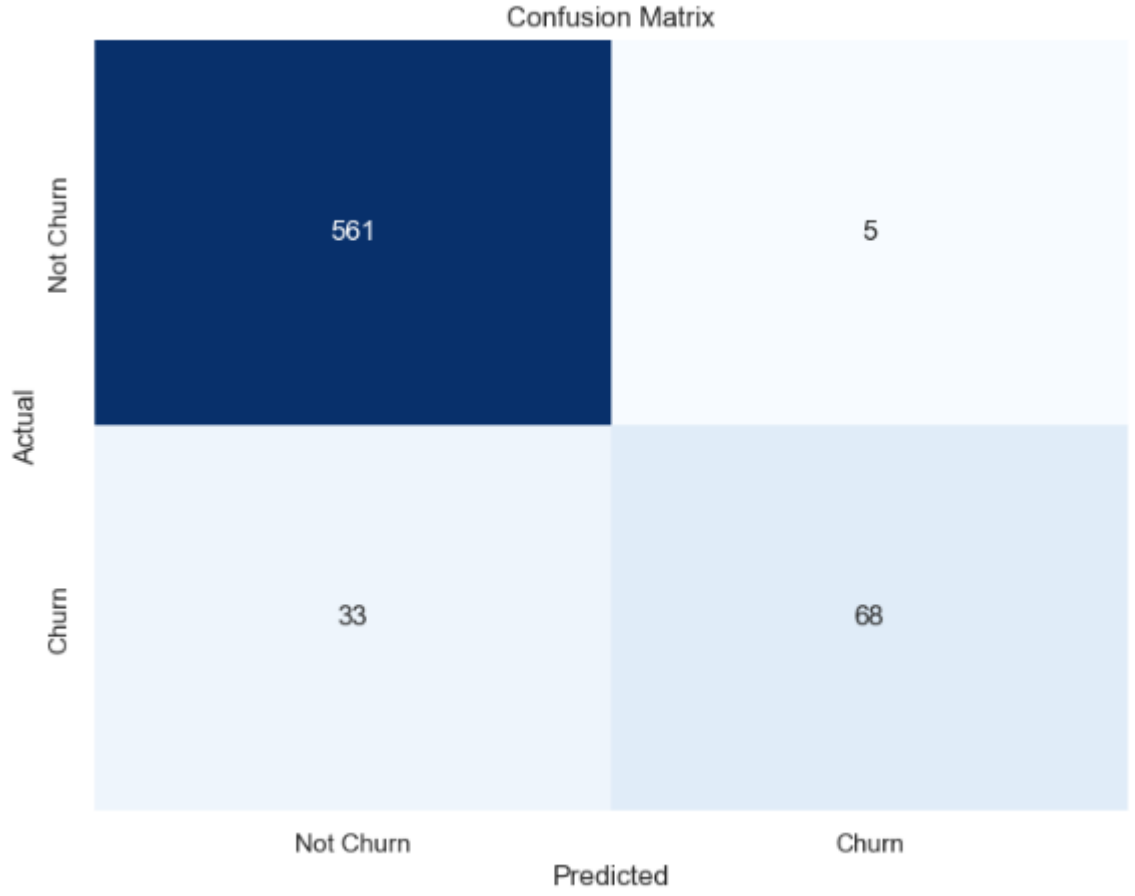
Observations 4/4: Churn distribution vs total day and evening minutes



- Subscribers who have churned tend to have high number of day and evening minutes spent on the SyriaTel platform, another risk to the business.

Model evaluation: Best performing model

Random Forest Model



- The model had the best prediction accuracy of customer churn in relation to the dataset.
- Accuracy: **97.56%** of the model predicted correctly of customer churn or not-churn.

True Negatives (TN): 561 instances where the model correctly predicted not-churn customers.

False Negatives (FN): 33 instances where the model incorrectly predicted not-churn customers.

False Positives (FP): 5 instances where the model incorrectly predicted positive churn customers.

True Positives (TP): 68 instances where the model predicted positive churn customers correctly.

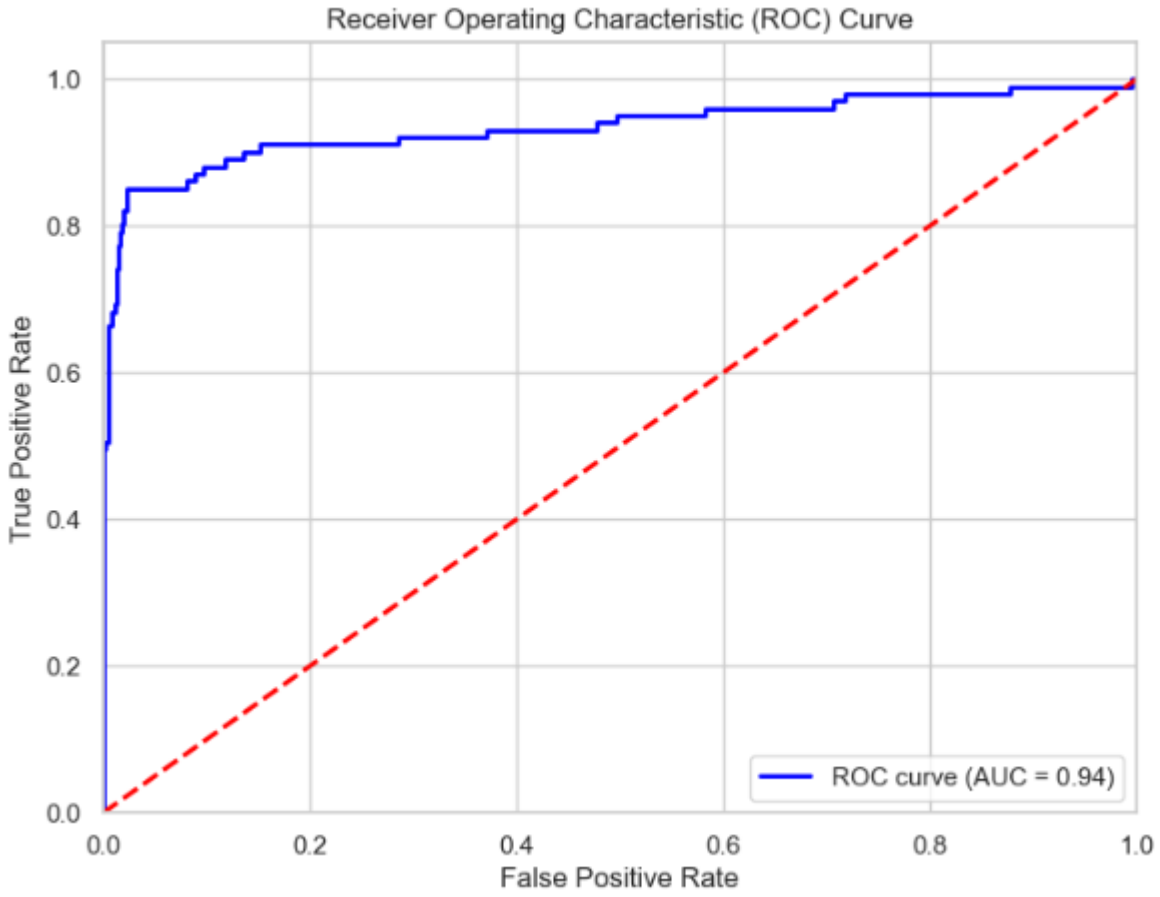
Classification report

Precision:
Non-Churn (False): 0.94 - Out of all the customers predicted as non-churn, 94% actually did not churn. Churn (True): 0.93 - Out of all the customers predicted as churn, 93% actually churned.

Recall:
Non-Churn (False): 0.99 - Out of all the customers who did not churn, the model correctly identified 99% of them. Churn (True): 0.67 - Out of all the customers who churned, the model correctly identified 67% of them.

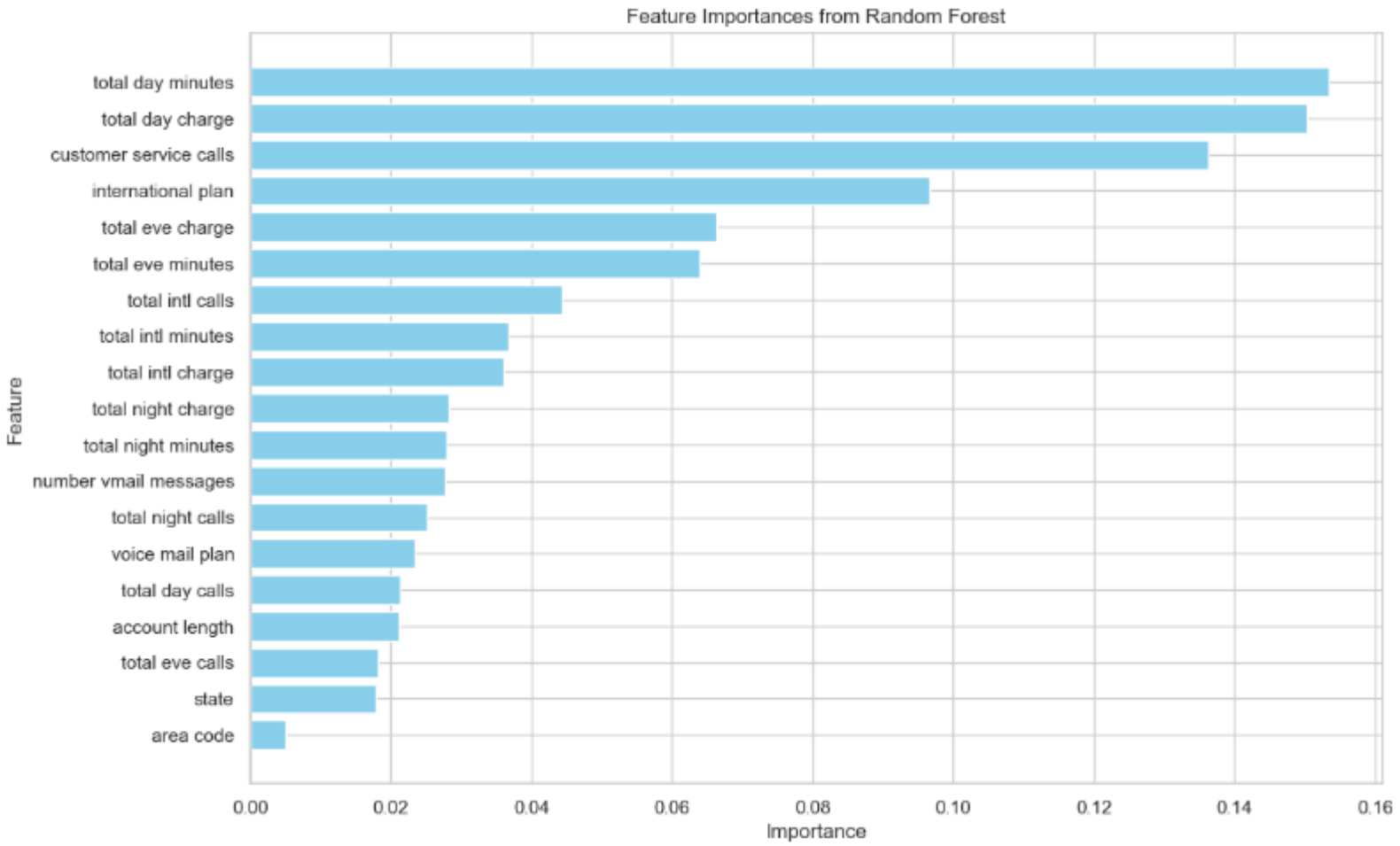
F1-Score:
Non-Churn (False): 0.97 - The F1-score is the harmonic mean of precision and recall for non-churn, indicating high accuracy. Churn (True): 0.78 - The F1-score for churn indicates a relatively good balance between precision and recall.

Model ROC curve



From the ROC curve above, the model performance had a good prediction of customer churning since the area under the curve is 0.94, which is close to 1.

Features of importance



Total day minutes and charge, customer service calls and international plan are the main contributing factors to customer churning.

Recommendations

Recommendations

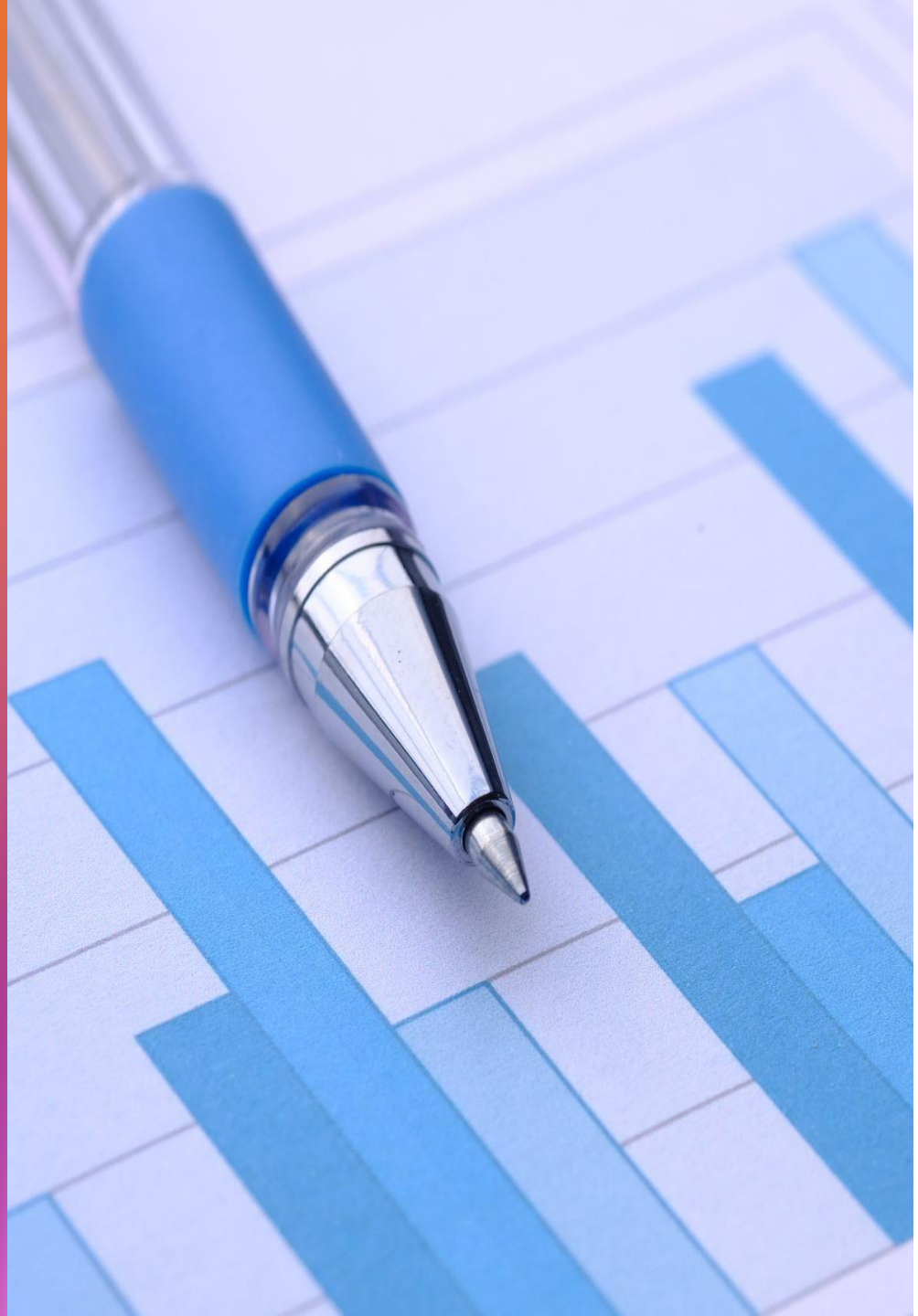
Proactive customer service: Since customer service calls are a significant factor, providing proactive and effective customer support can help address issues or concerns promptly, potentially reducing churn.

Personalized offers or incentives: Identifying customers with international plans and offering personalized discounts or incentives may encourage them to stay with the telecommunications company.

Monitoring usage patterns: Monitoring total day minutes and charges can help identify customers who are using the service extensively, potentially indicating dissatisfaction or a need for alternative plans. Offering tailored solutions or upgrades may help retain these customers.

Targeted communication: Utilizing the insights from the predictive model, targeted communication strategies can be implemented to reach out to customers at high risk of churn. This may involve personalized outreach campaigns, targeted promotions, or loyalty programs aimed at retaining these customers.

Feedback mechanisms: Implementing effective feedback mechanisms to gather insights from churned customers can help identify underlying issues and inform strategies for continuous improvement.



Project Progress Report

1. I first read through the case study and designed a business problem.

2. Further, I defined the business objectives that would help in the solution process.

3. Load the dataset for exploration to understand all included variables. Further determine relevant and irrelevant columns for use in analysis.
4. Prepare the data for preprocessing by checking its feasibility ie. Missing data, duplication and outlier handling.

5. Next, I did data preprocessing through numerical and categorical encoding, data scaling to ensure that all features contribute equally to the model's performance and helps in achieving faster convergence and better performance and performing test-train split for modeling.

6. Further performed different models and fine-tuned each for model performance and did model evaluation. From these, I chose the best model based on the accuracy – Random Forest(97%).

7. I then drew conclusions and recommendations that align with the EDA and objectives relevant for SyriaTel stakeholder.

Thank you

https://github.com/Annegit1/Phase_3_project