



SyriaTel Churn Prediction

By Joshua Kiluu

Contents

01 SyriaTel Business Understanding.

02 Data Collection and Preprocessing

03 Exploratory Data Analysis

04 Churn Prediction Modeling

05 Insights and Recommendations

06 Conclusion

01

Introduction to SyriaTel

...

Overview of SyriaTel

Leading Telecom Provider

SyriaTel is the largest telecommunications company in Syria, offering a wide range of services.

Diverse Product Portfolio

SyriaTel provides mobile, fixed-line, and internet services to residential and business customers.

Importance of Churn Prediction

Understand customer behavior

Identify factors leading to customer churn

Optimize retention strategies

Develop targeted campaigns to reduce churn

Improve business performance

Increase revenue and profitability through reduced churn



Objectives of the Study

01

Understand customer churn

Analyze factors contributing to customer attrition at SyriaTel.

02

Develop predictive model

Identify customers likely to churn and enable targeted retention efforts.

03

Recommend strategies

Suggest initiatives to improve customer loyalty and reduce churn.



02

The DataSet

...

Demographics & Account Info:

- State, Account length, Area code

Plans Subscribed:

- International plan (yes/no)
- Voice mail plan (yes/no)

Usage Metrics:

- Voice mail activity: number of vmail messages
- Call usage (minutes, calls, charges) across Day, Evening, Night
- International usage: intl minutes, intl calls, intl charge

Customer Service:

- Number of customer service calls

Target (Churn):

- Whether the customer stayed (False) or left (True)

Customer satisfaction and feedback collected through online and phone surveys.

03

Data Preprocessing

...

Data Cleaning and Transformation

- Dataset had no nulls and duplicates
- Dropped 'state' and 'area code' columns.
- Handled multicollinearity by dropping the charges columns.
- Handled class imbalance using SMOTE.

Feature Engineering

Identified Relevant Features

Analyze customer data to determine the most predictive variables for churn.

Encoded Categorical Variables

Convert categorical data into numerical format for model input. These are international plan and voice mail plan.

Scaling

Bring features into a similar scale.

Handling Class Imbalance

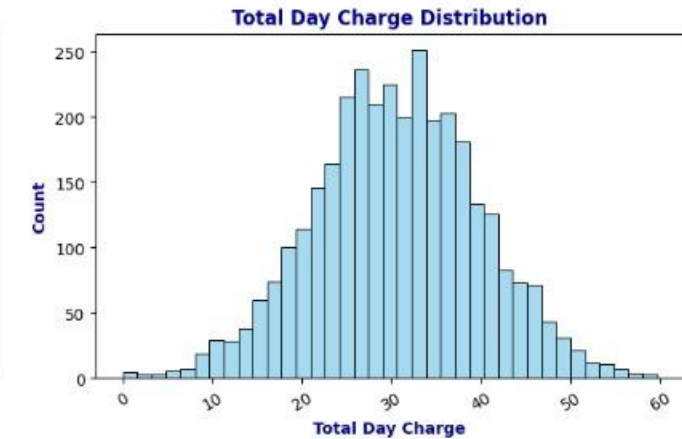
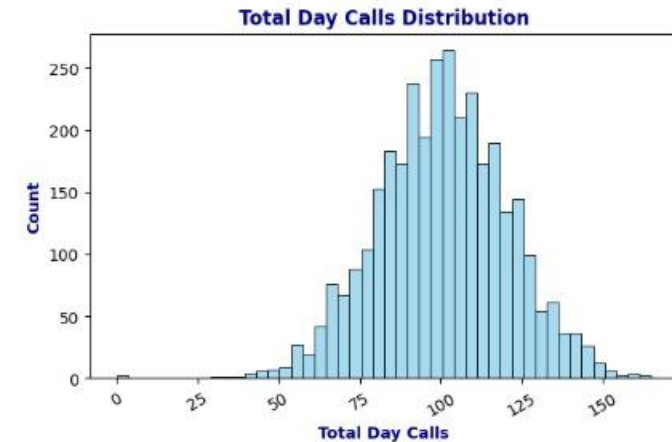
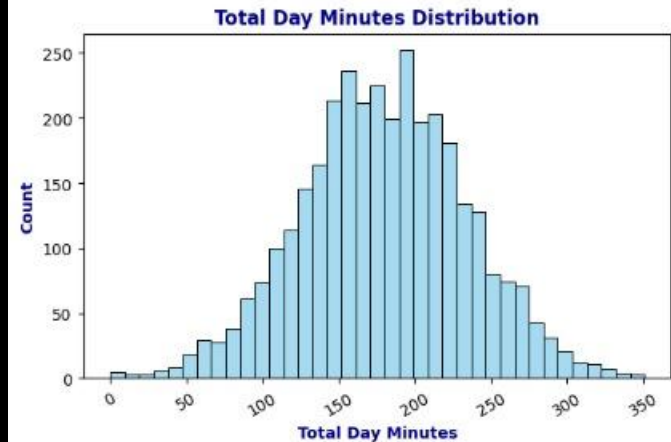
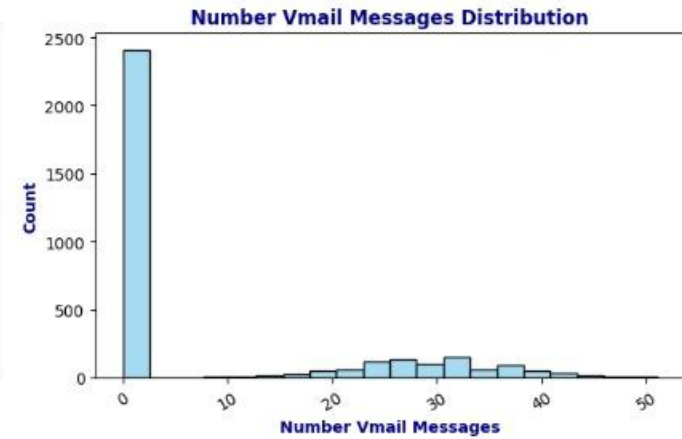
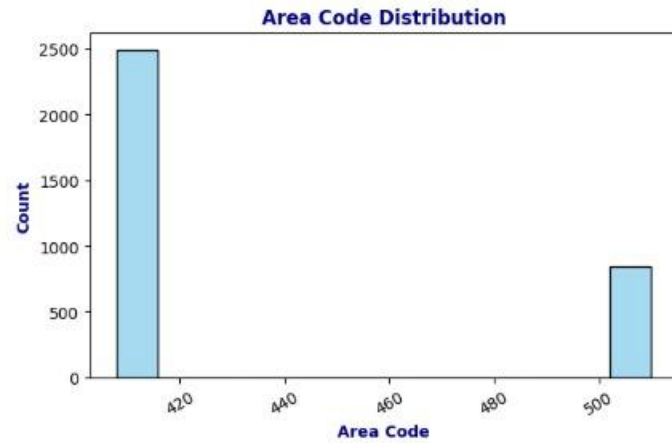
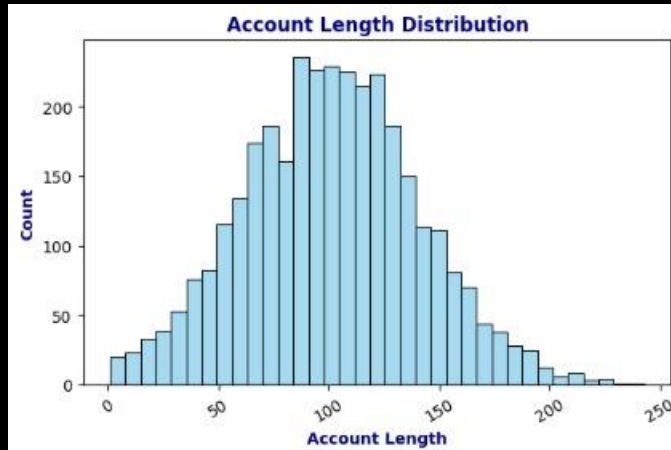
Adjusts the model so that the data does not bias towards the majority class.

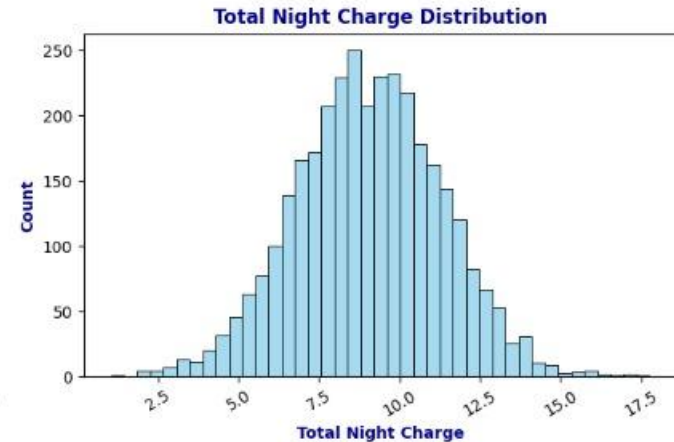
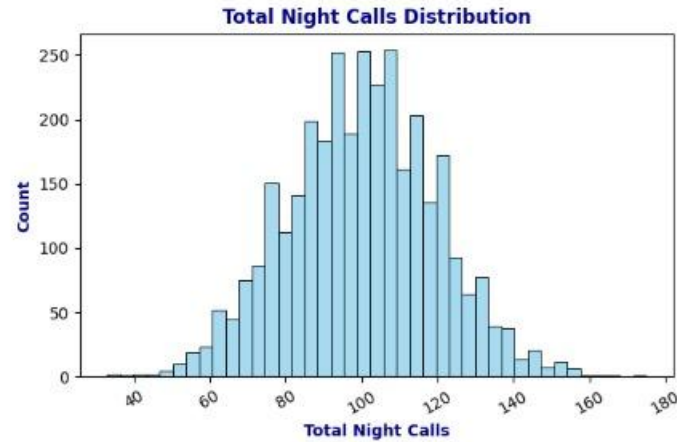
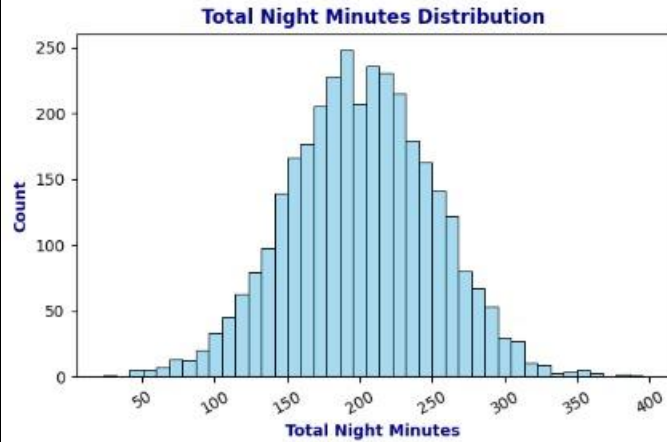
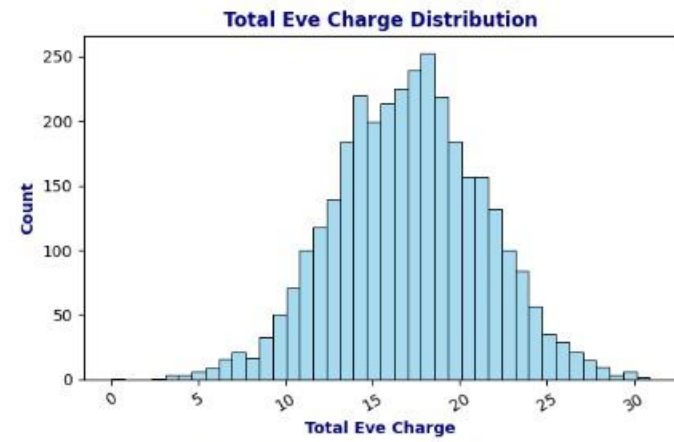
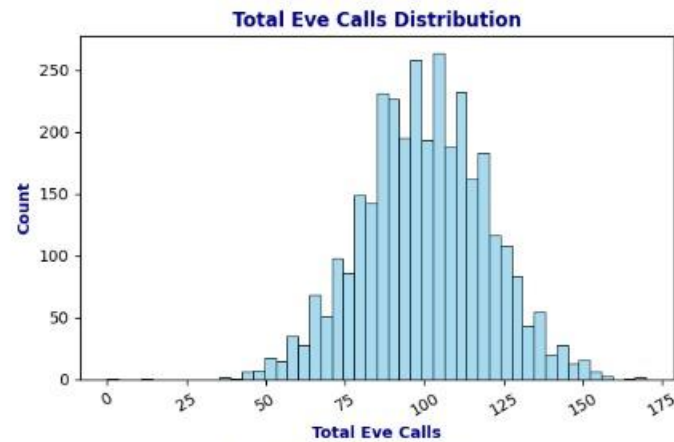
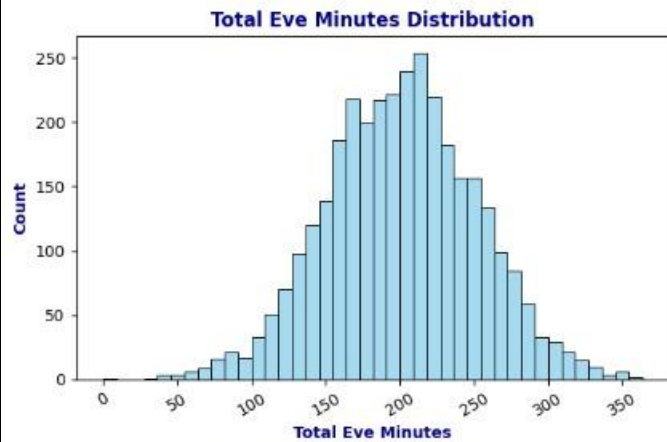
04

Exploratory Data Analysis

...

Univariate Analysis(Usage Patterns)



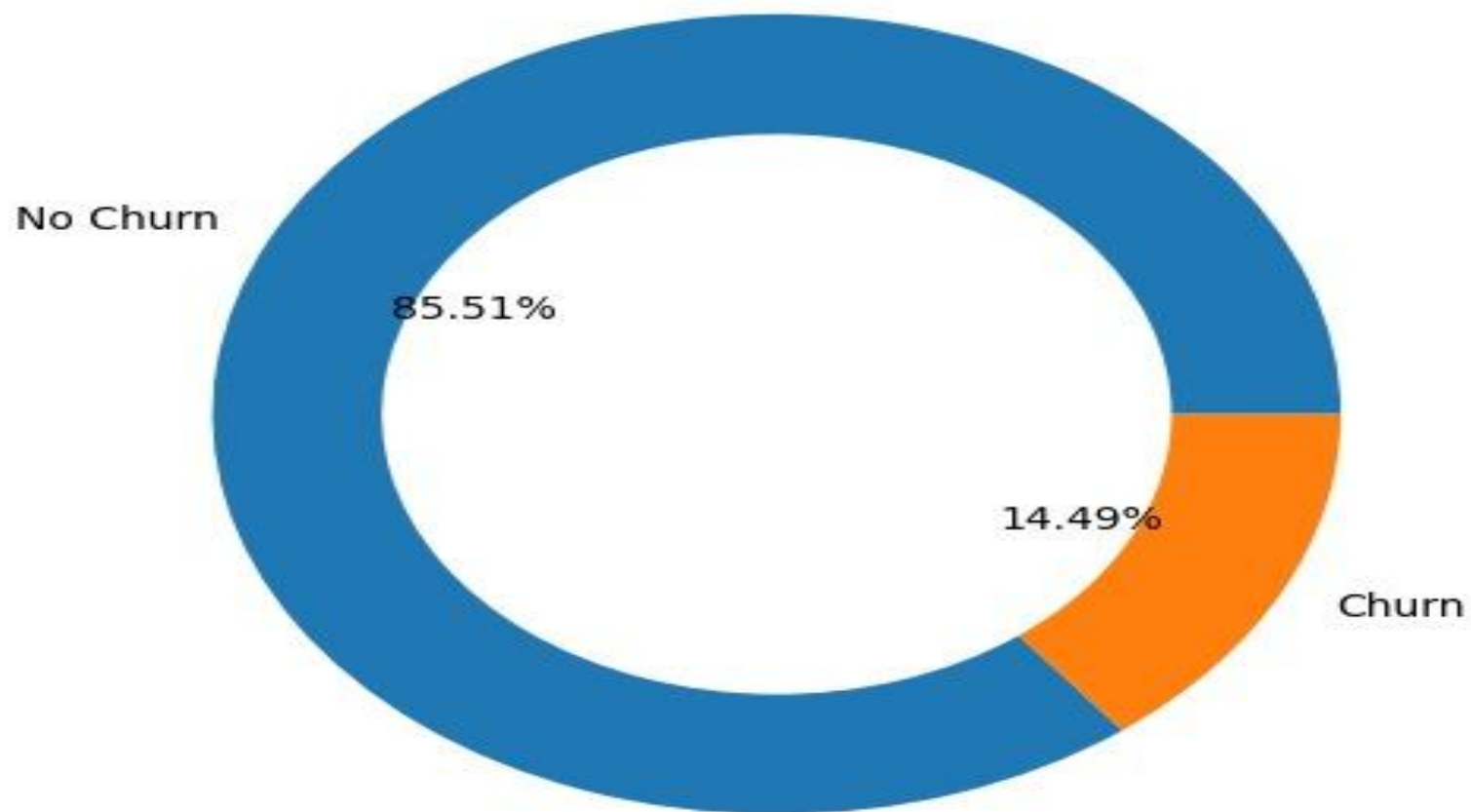


Key Observations from Distributions.

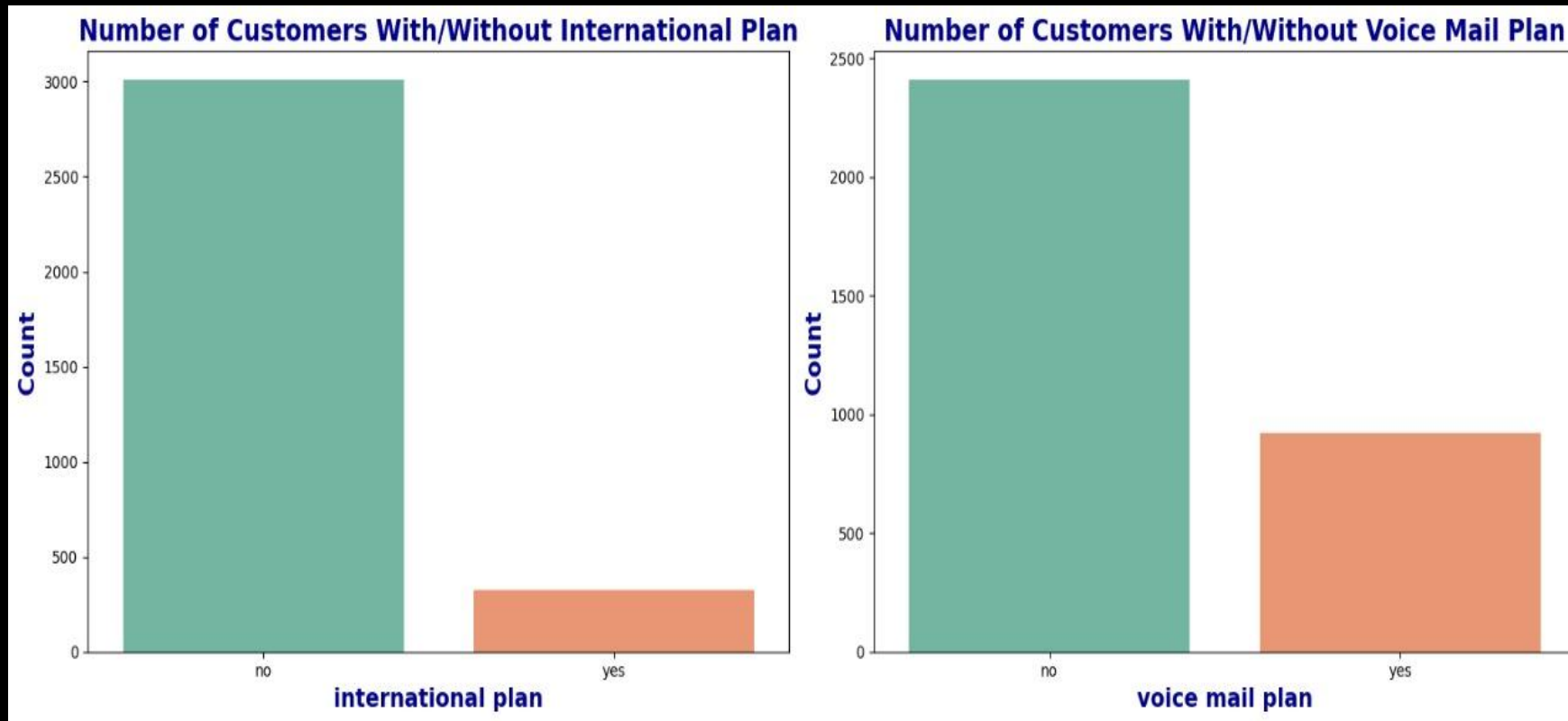
- Most customers have zero voicemail messages.
- Day/Evening/Night Minutes and Charges have a normal distributions. This indicates balanced usage across customers.
- Most customers rarely use international calling.
- A few customers make many service calls.

Churn Distribution

Customer Churn Distribution

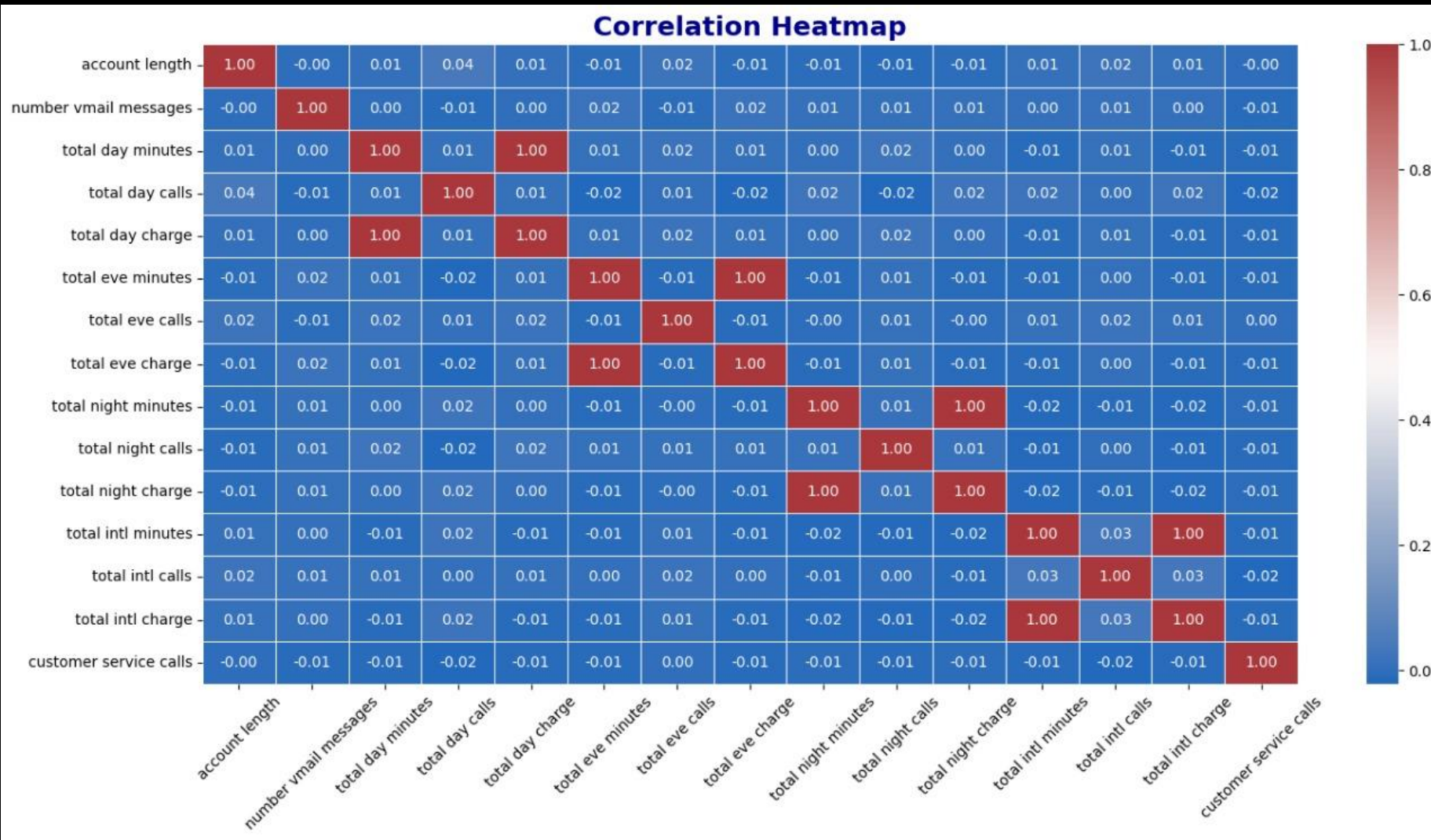


14.45% customers churn while 85.5% do not churn indicating a class imbalance.



Most customers have neither an international plan nor a voice mail plan.

Correlation



Most variables have very weak correlations with each other. There is a perfect correlation between usage minutes and their corresponding charges.

05

Churn Prediction and Modeling

...

Machine Learning Algorithms

Logistic Regression

Baseline model.

Decision Trees

Captures non-linear relationships well

Random Forest

Reduces overfitting compared to a single tree.

XGBoost

Outperforms other models in churn prediction tasks.

Model Evaluation Metrics(Recall, ROC_AUC Score)

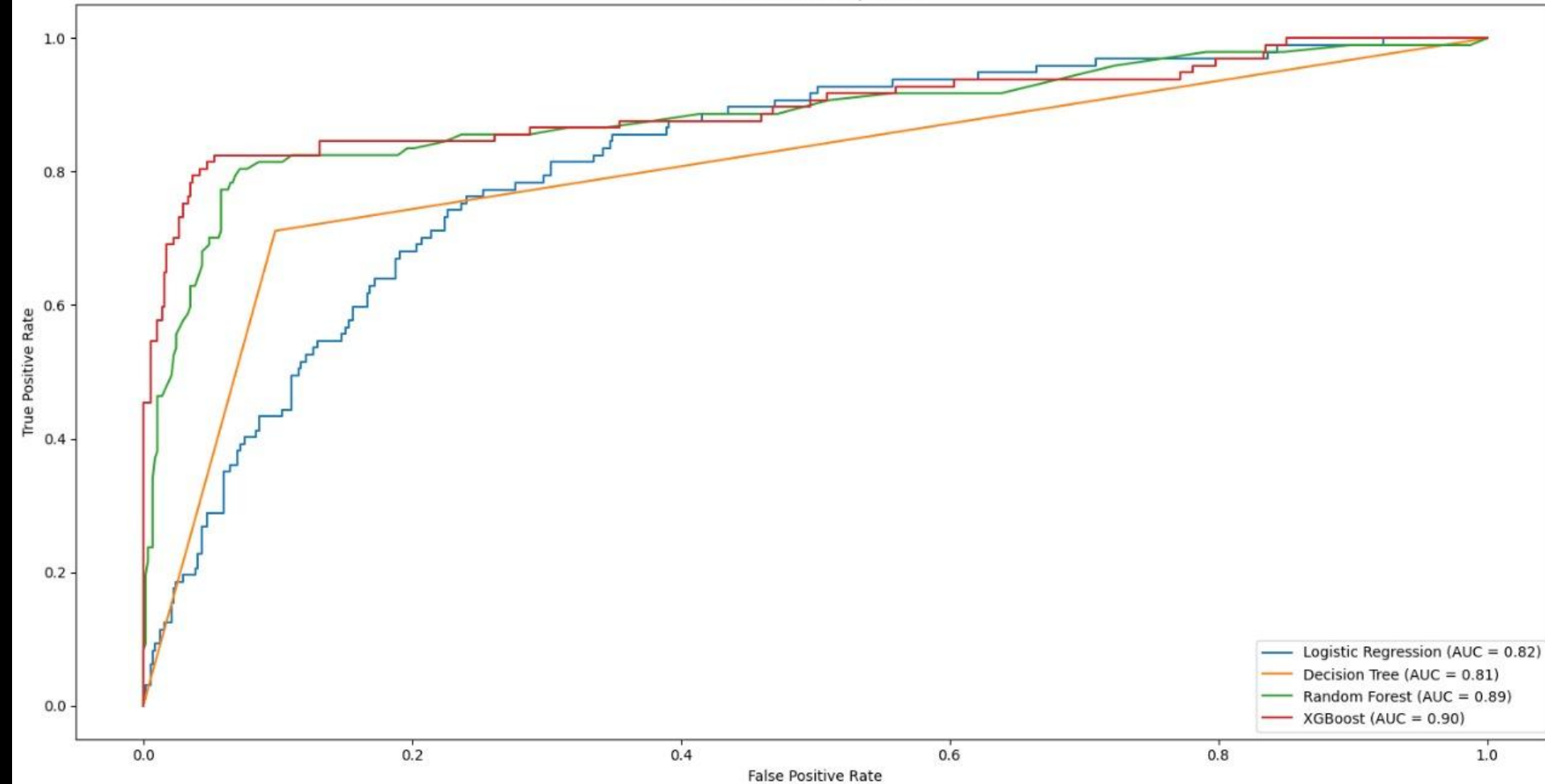
- Recall

Measures how many actual churners were correctly identified by the model helping to minimize false negatives.

- Recall (Logistic Regression): 0.74
- Recall (Decision Tree): 0.71
- Recall (Random Forest): 0.74
- Recall (XGBoost): 0.76

ROC_AUC Comparison

ROC Curves Comparison



The ROC_AUC Score measures the model's overall ability to distinguish between churners and non-churners across all classification thresholds. A score of between 0.81 and 0.90 indicates strong distinguish between churners and non-churners across different thresholds.

Hyperparameter Tuning

Optimized Model Performance

*Adjusting hyperparameters to
improve model accuracy and
generalization.*

XG Boost Tuned Model

- The tuned XGBoost Model had an improvement in the Recall Score from 0.76 to 0.77.

05

Feature Importance

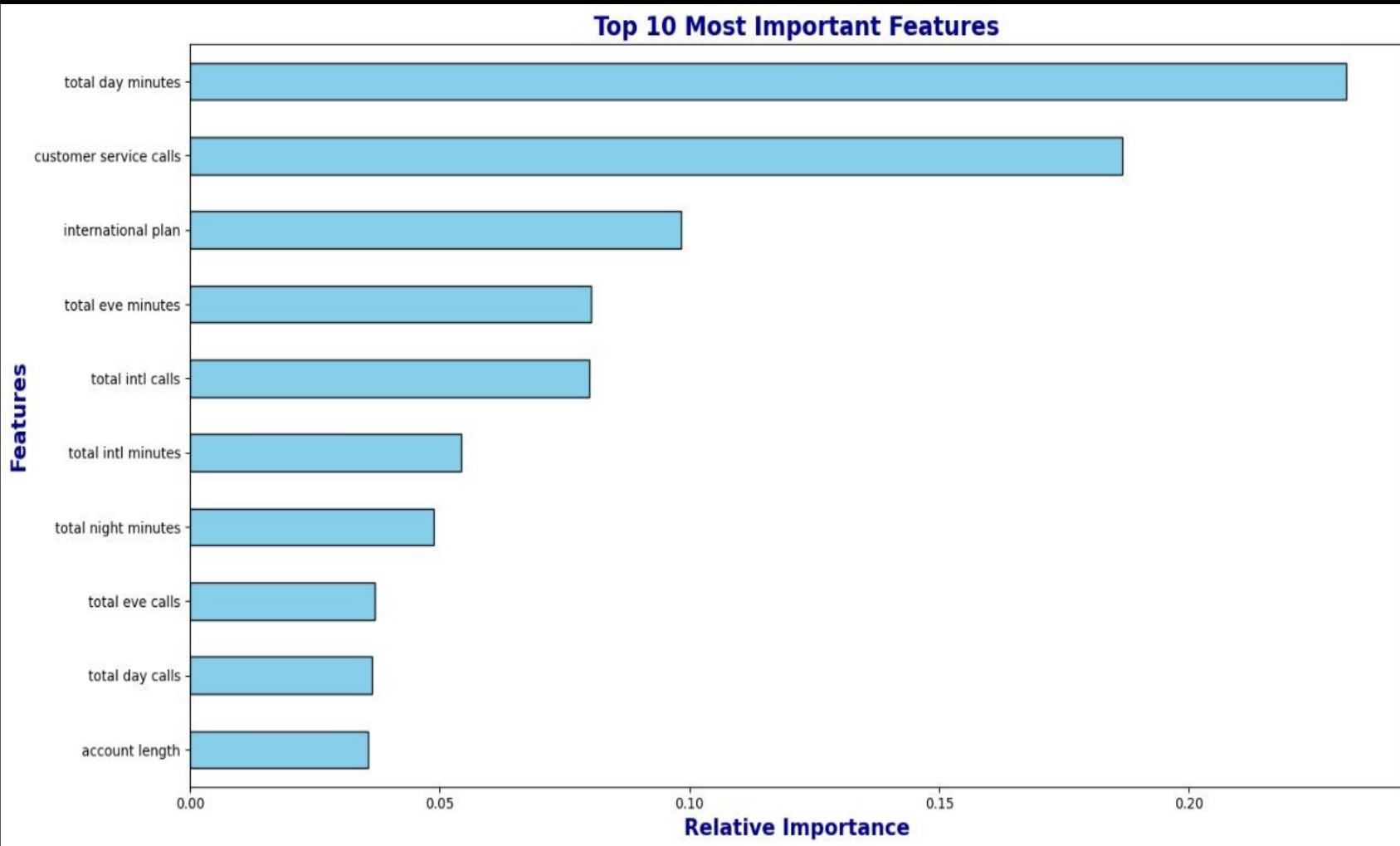
...

01

Feature Importance

...

Top 10 Most Important Features



- **Total Day Minutes**
Higher total day minutes is a strong predictor of churn, suggesting that customers who use more minutes during the day are more likely to switch providers.
- **Customer Service Calls**
Customers who frequently contact support are at higher risk of leaving.
- **International Plan**
Customers with international plans are more likely to discontinue their service.

05

Insights and Recommendations

...

Retention Strategies



Identify at-risk customers

Analyze customer data to pinpoint high-churn risk profiles.



Offer personalized incentives

Provide tailored promotions and discounts to retain valuable customers.



Enhance customer experience

Improve service quality and responsiveness to increase customer satisfaction.

06

Conclusion

...



Summary of Findings

01

Churn rate is 14.5%

Key factors contributing to churn include voice mail plan, international plan and total day minutes.

02

XGBoost is the best model.

It has a recall score of 0.77 after tuning and ROC_AUC Score of 0.89.

Next Steps

Identify Key Drivers

Analyze top factors contributing to customer churn.

Implement Retention Strategies

Develop targeted campaigns to address at-risk customers.

Monitor and Refine

Continuously track churn metrics and adjust strategies.



Thank You

