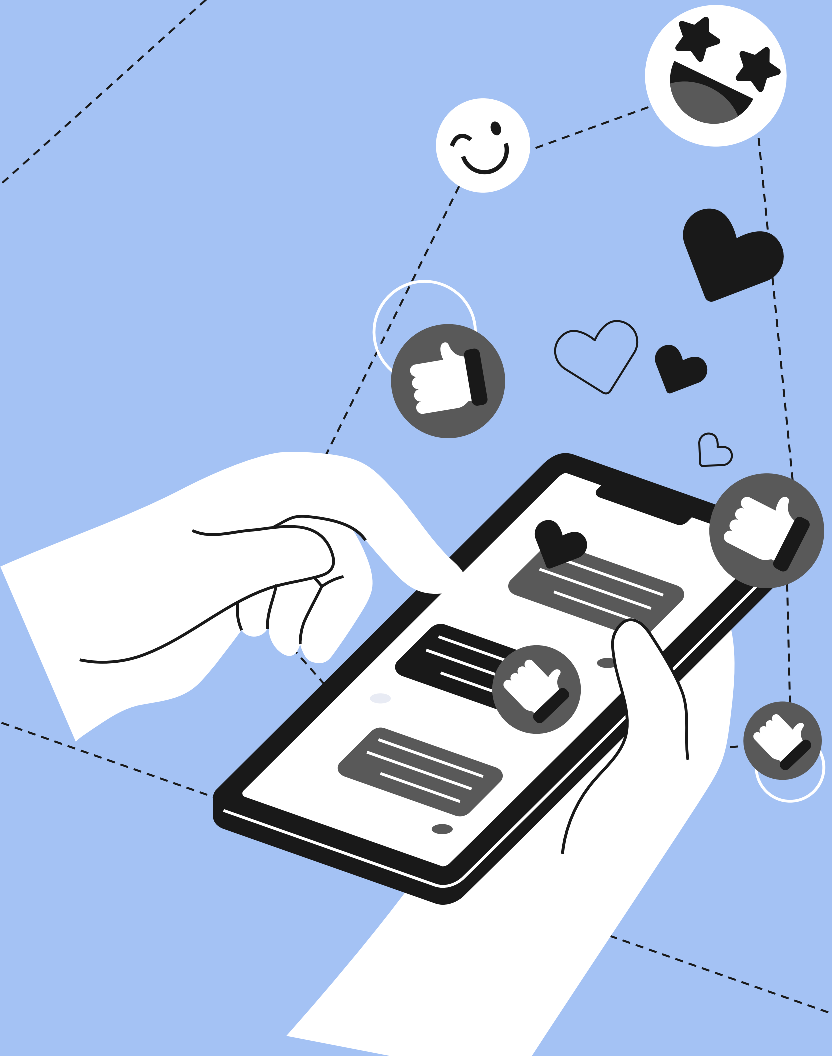


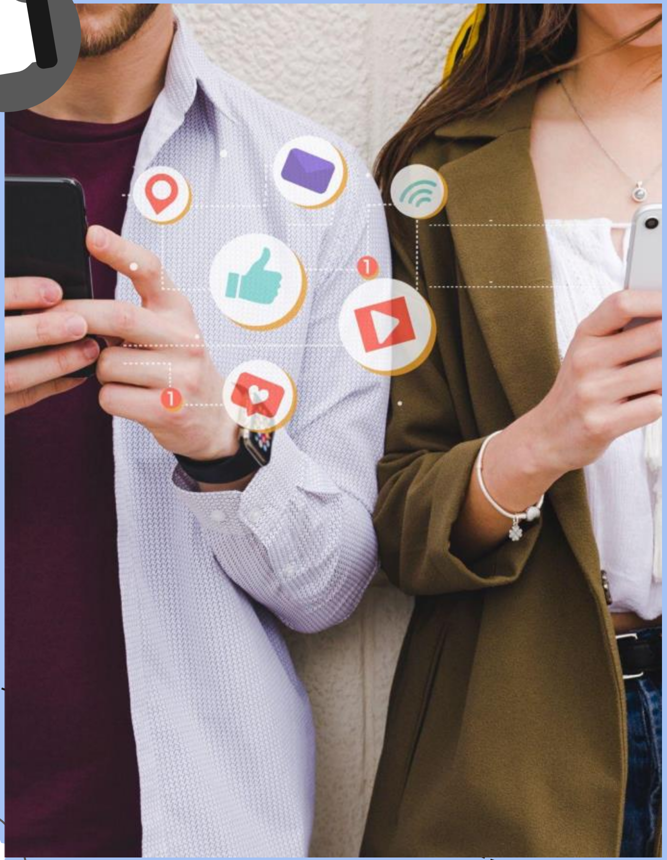
SYRIATEL CUSTOMER CHURN CLASSIFICATION PROJECT

Table of contents

Project Overview

This project aims to predict customer churn for Syriatel, a telecommunications company.

The primary goal is to build a classification model that identifies customers who are likely to churn (“soon” stop doing business with Syriatel) allowing the business to take proactive measures to retain them and improve customer satisfaction.



01

BUSINESS UNDERSTANDING

Objective:

Predict whether a customer will soon stop doing business with SyriaTel

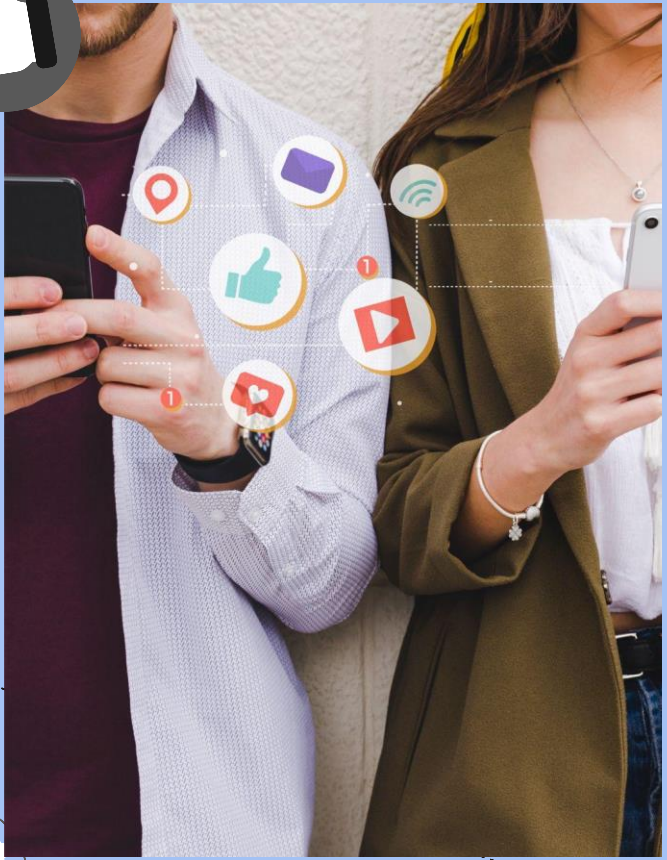
Importance:

By identifying patterns and factors that lead to churn, SyriaTel will be able to minimize revenue loss and enhance customer loyalty. It will implement retention strategies to improve overall customer satisfaction and reduce the financial impact of losing customers.

Stakeholders:

The Primary Stakeholder: SyriaTel Telecommunications Company interested in reducing customer churn and improving customer retention.
Secondary Stakeholders:

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



02

DATA UNDERSTANDING

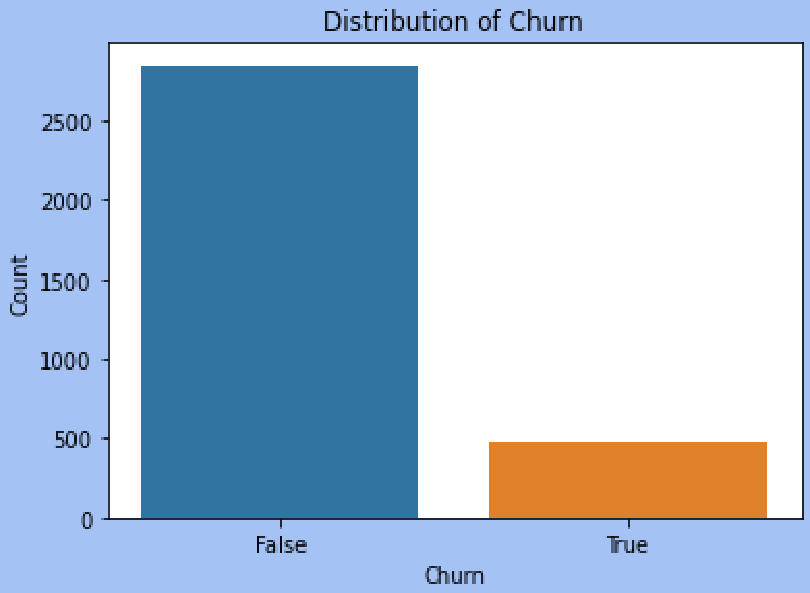
Data Overview: This dataset includes customer information and their usage patterns with SyriaTel Telecommunications Company.

Data Source	SyriaTel Telecommunications Company
Features	State, Account Length, Area Code, Phone Number, International Plan, Voice mail Plan, Total day minutes, Total Day calls, Total Day Charge, Total Eve Minutes,, Total Eve Calls,, Total Eve Charge, Total Night minutes, Total Night calls, Total Night Charge, Total Intl Minutes, Total Intl Calls, Total Intl Charge, Customer Service calls, Churn
Target Variable	Churn: "True"- Indicates the customer churned "False"- Indicates the customer did not churn
Predictors	State, Account Length, Area Code, Phone Number, International Plan, Voice mail Plan, Total day minutes, Total Day calls, Total Day Charge, Total Eve Minutes,, Total Eve Calls,, Total Eve Charge, Total Night minutes, Total Night calls, Total Night Charge, Total Intl Minutes, Total Intl Calls, Total Intl Charge, Customer Service Calls

Distribution of the Target Variable(Churn)

Observations:

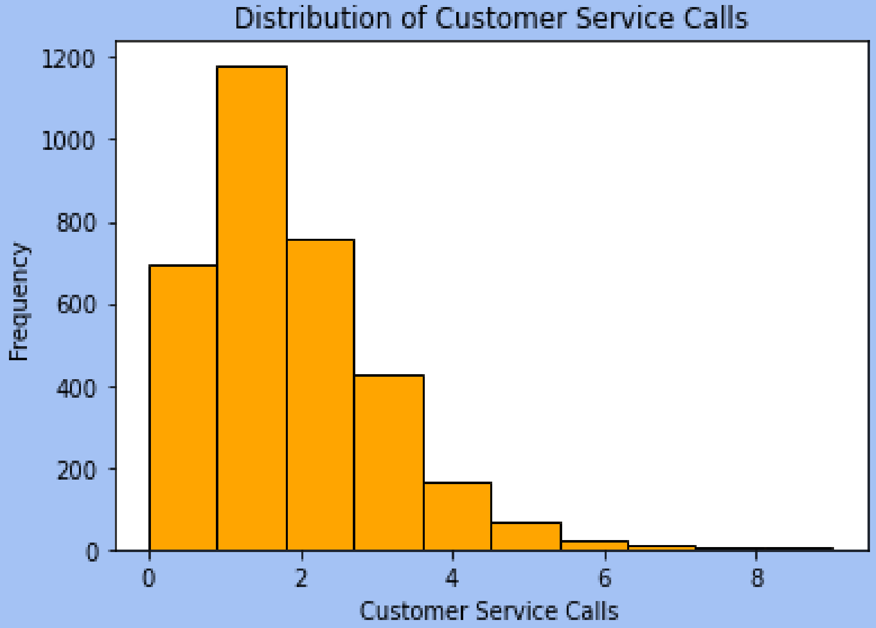
- False(Not Churned):Approximately 2800 customers
- True(Churned): Approximately 500 customers
- Class Imbalance between the two classes



Distribution of Customer Service Calls

Observations:

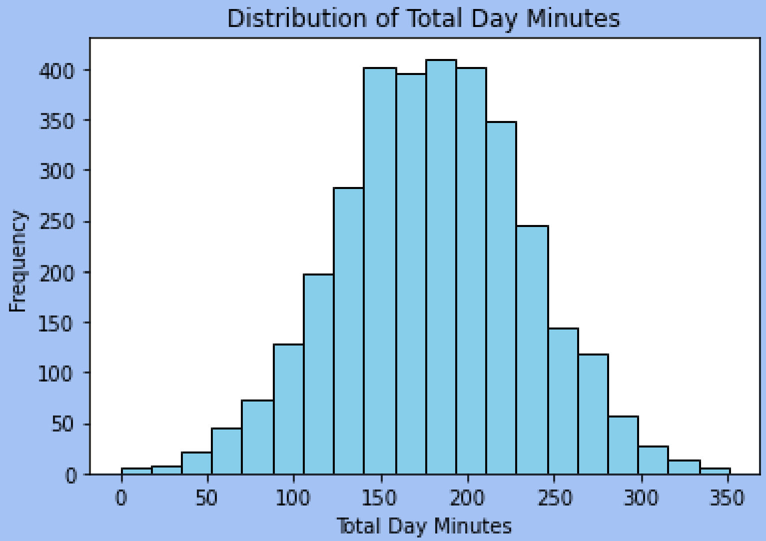
- The distribution is right skewed indicating most customers make a relatively small number of calls while a smaller group makes a larger number of calls.
- The most frequent number of calls is between 2 and 3 suggesting that a significant portion of customers contact 2-3 times



Distribution Of Total Day Minutes

Observations:

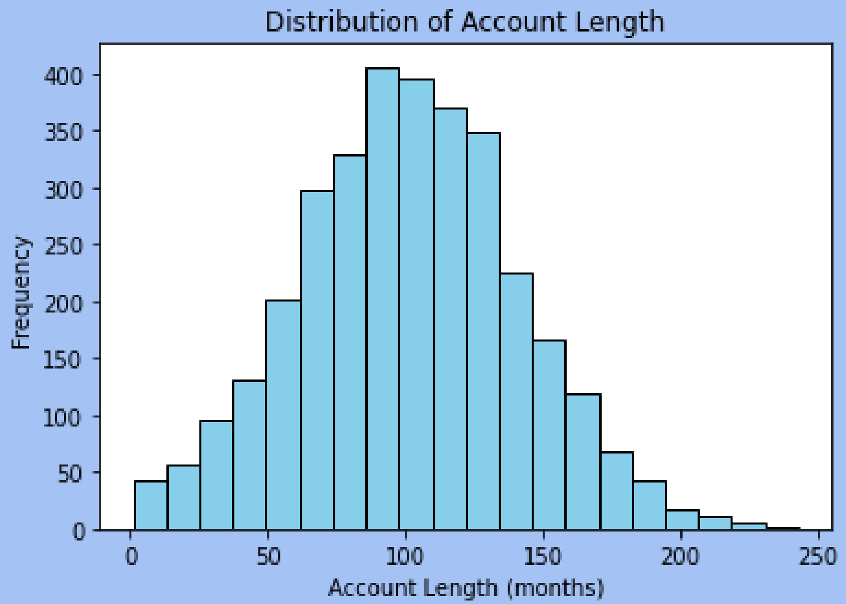
- The distribution resembles a normal distribution indicating that majority of customers use a moderate amount of day minutes with fer using significantly more or less
- The mean total day minutes is around 200



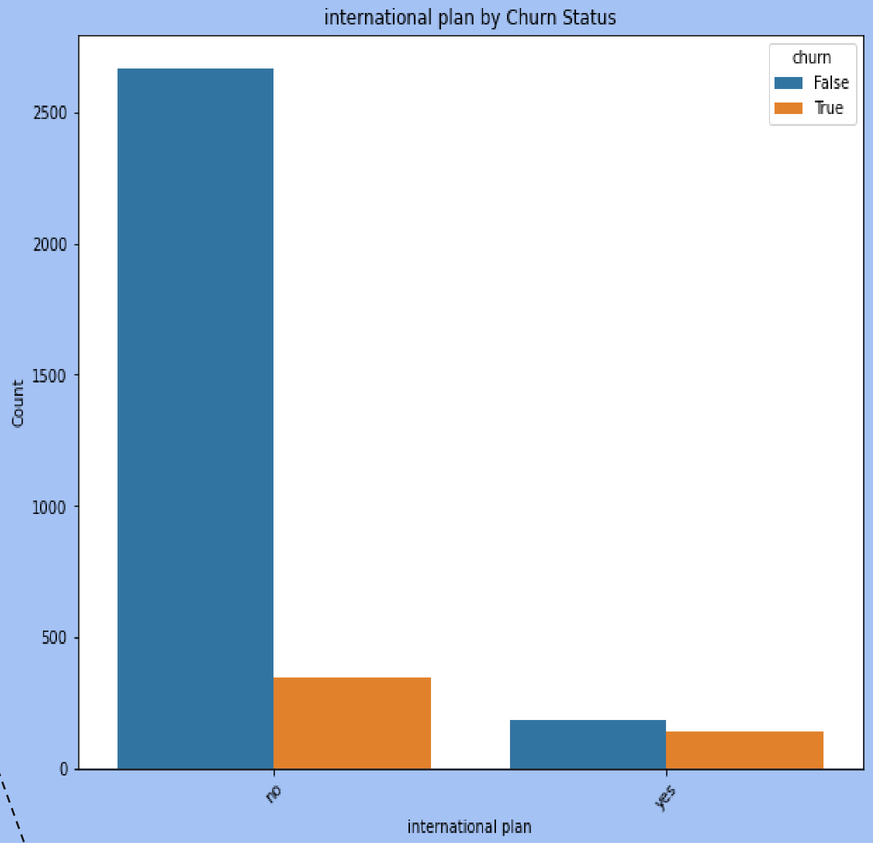
Distribution of Account Length

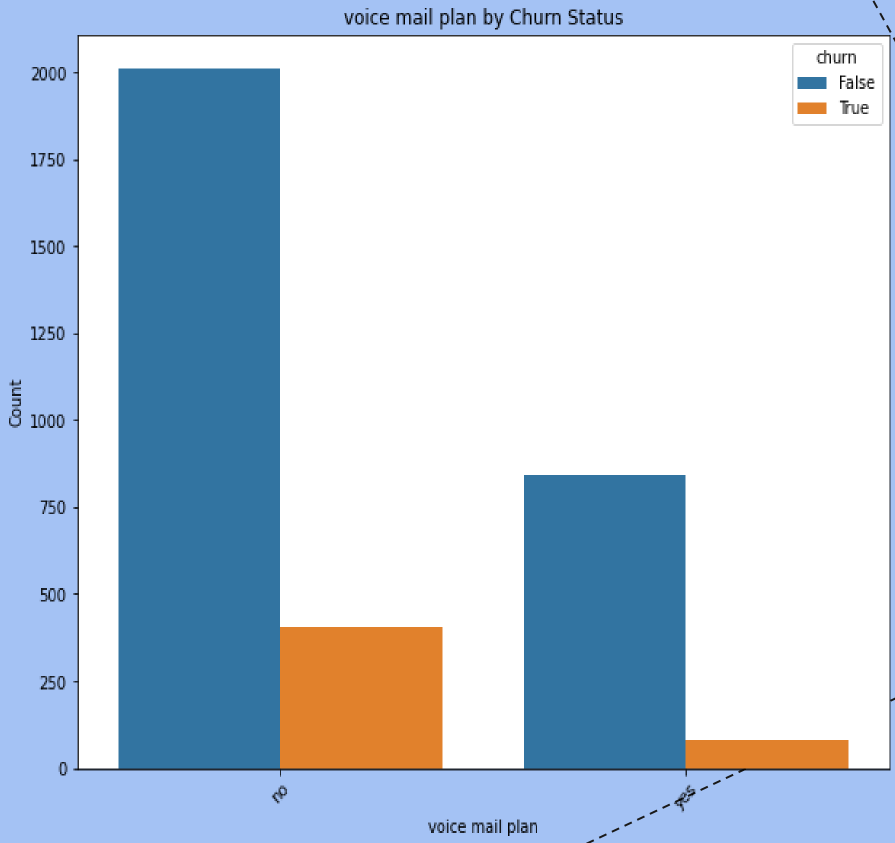
Observations:

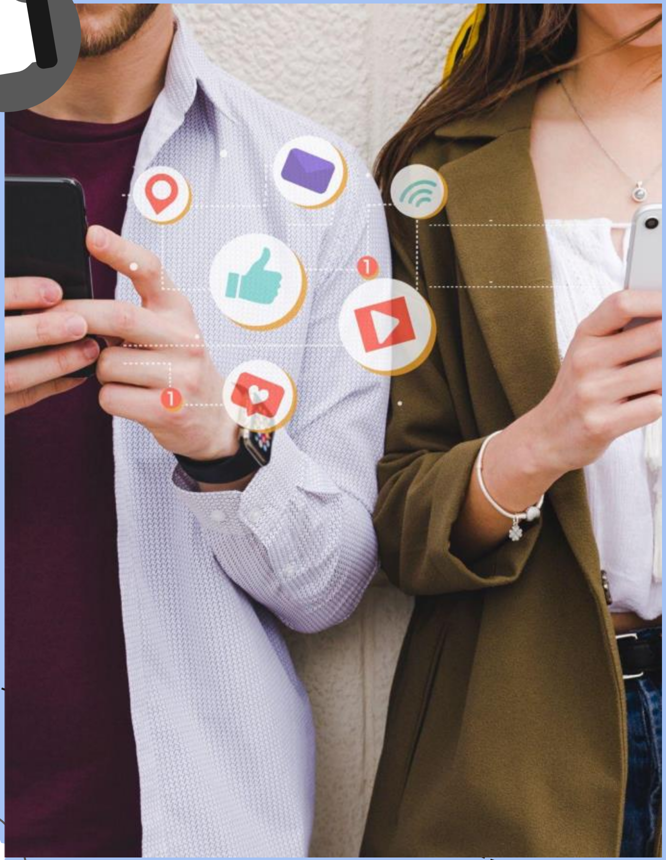
- The distribution appears to be normal suggesting majority of customers have account lengths close to the average with fewer having longer or shorter lengths.
- The mean account length is around 100 months.



Class Imbalance Analysis of Categorical Variables







03

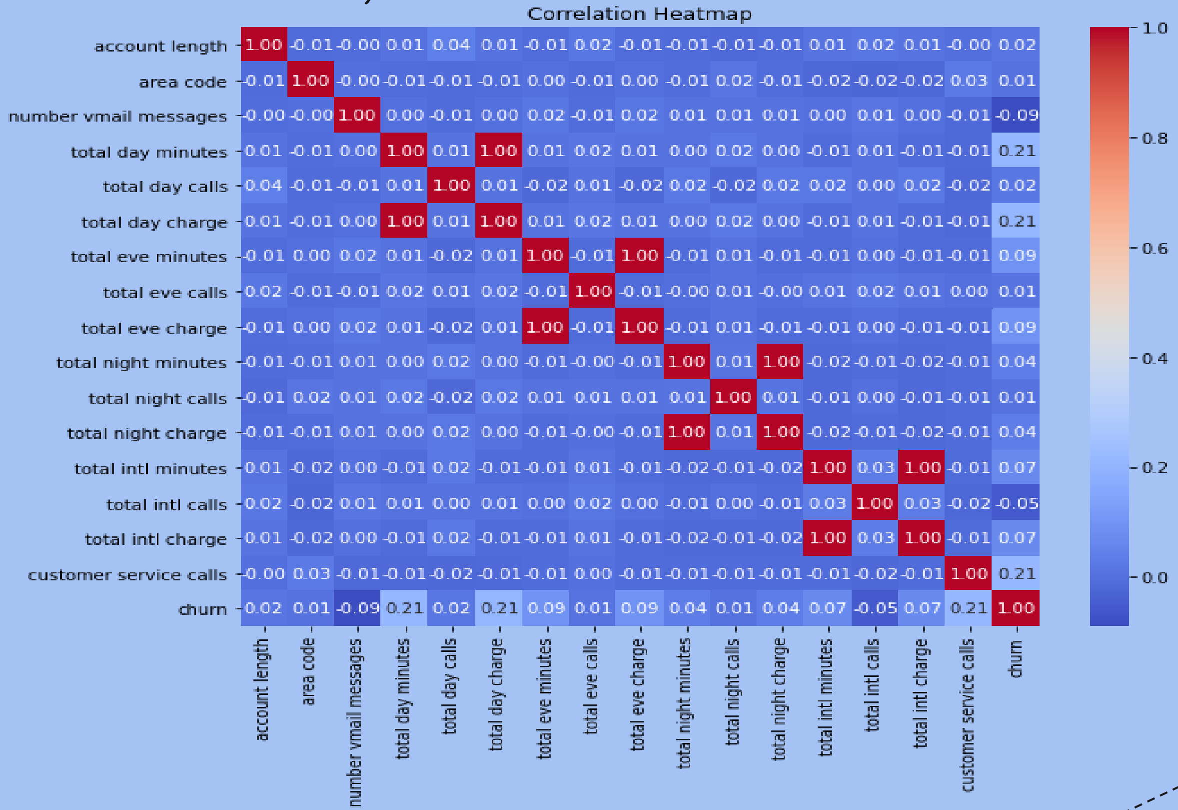
DATA PREPARATION

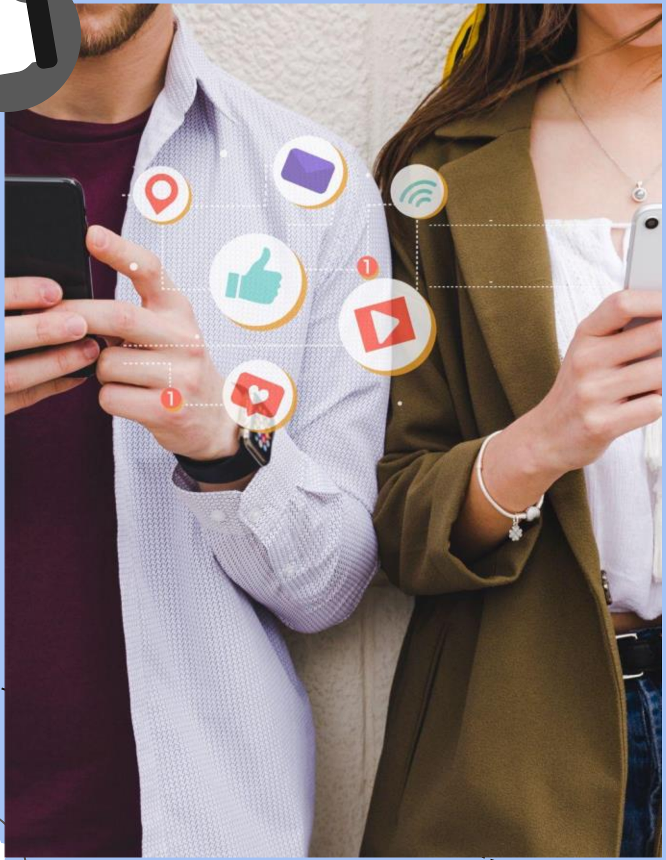
Data Preprocessing

- Data Cleaning- Since there were no missing values and duplicates in the dataset, no imputation was necessary.
- Encoding Categorical variables-We used one-hot encoding to convert categorical variables (international plan', 'voice mail plan') into numerical format.
- Addressing Multicollinearity-Features that exhibited high multicollinearity were removed using VIF analysis, reducing the dataset from 21 columns to 14 (including target variable).
- Feature Scaling- Ensuring numerical features are on the same scale using 'MinMaxScaler' to standardize the range.
- Handling Class Imbalance- SMOTE was applied to balance classes in the target variable
- Feature Selection- Selecting a subset of relevant features (variables, predictors) for use in the model construction.

Correlation Heat Map

- The heatmap is symmetrical indicating that the correlation between any two variables is the same regardless of the order in which they are considered.





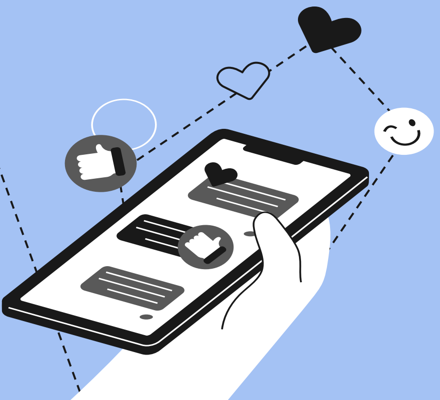
04

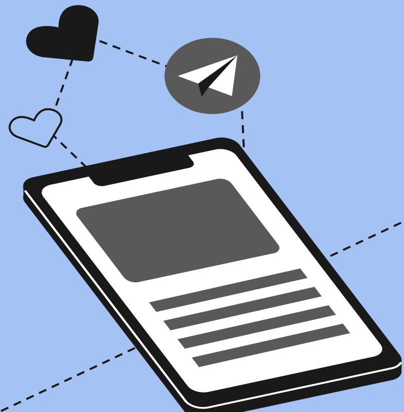
MODELING

Two models were built and evaluated:
Train-Test Split: 70.0% - 30.0%

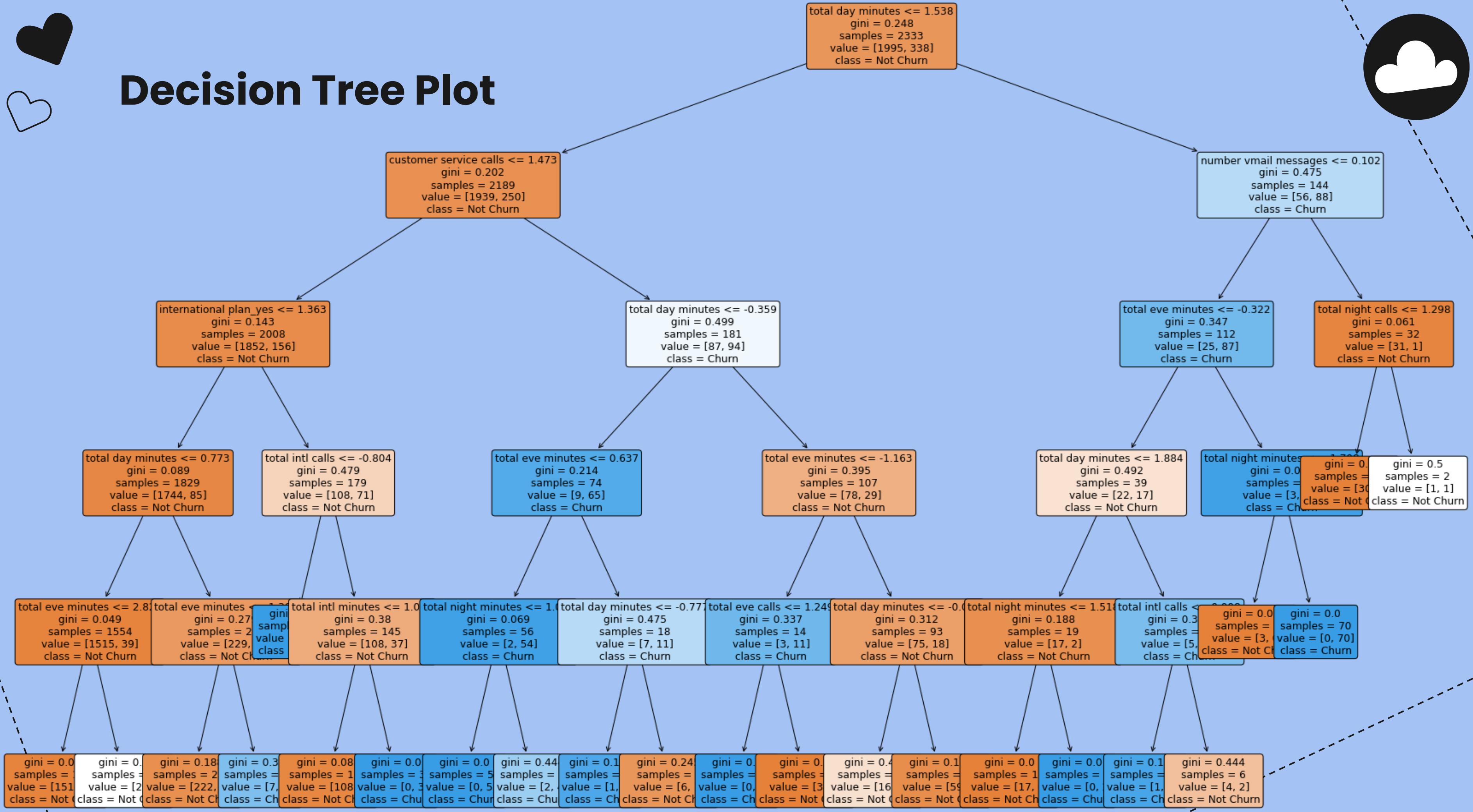
1. Logistic Regression:
- Baseline and tuned models were built.
- The model performed well for the “non-churn” class but struggled with recall for the churn class

2. Decision Trees Classifier:
- Baseline and tuned models were built.
- The tuned model achieved better performance overall, particularly in improving recall for the “churn” class.





Decision Tree Plot

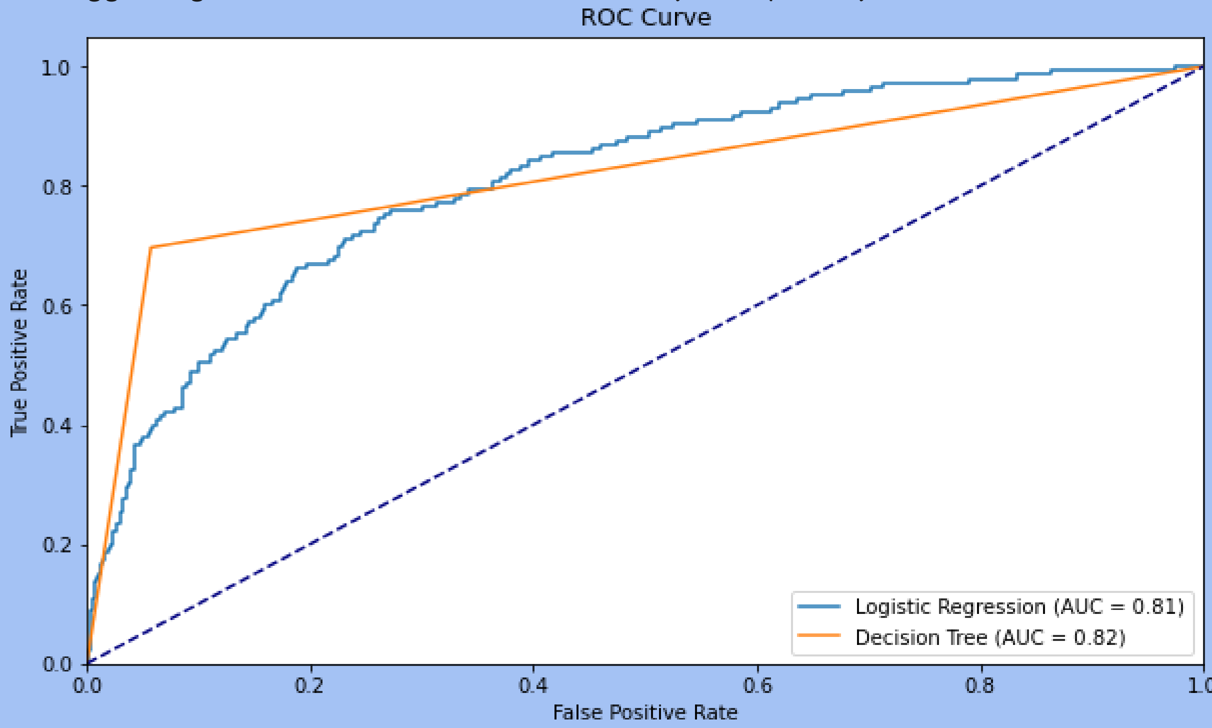


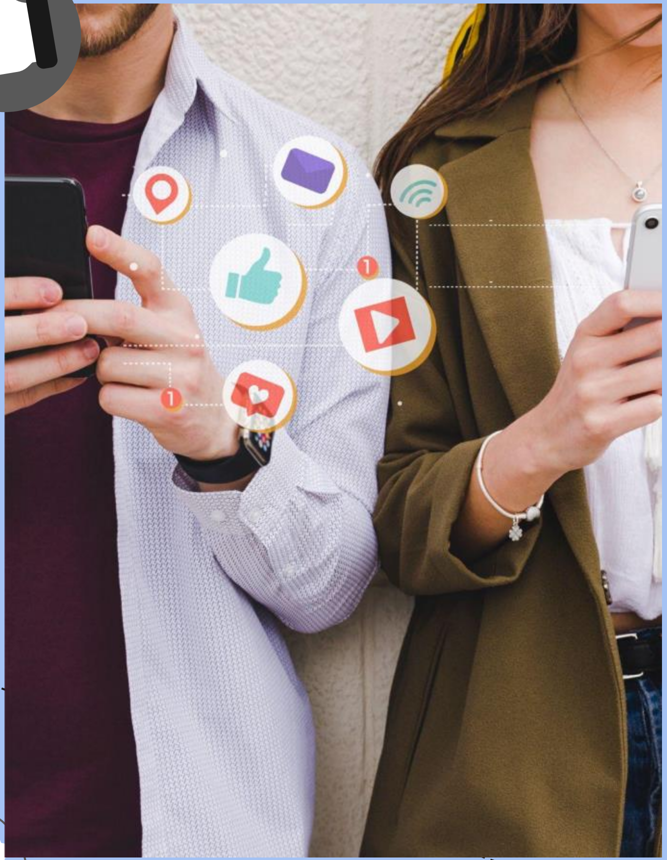
Observations:

- The decision tree is moderately deep and complex with multiple levels and branches indicating it considered multiple factors to make predictions.
- Total Day Minutes feature is used as the root node ,suggesting that it is the most important predictor of the target variable (churn).
- Customer Service Calls feature is also used at a higher level indicating it's importance in decision making process.
- Other features like international_plan_yes, total_ eve_minutes and total_intl_minutes are also used at various levels suggesting their relevance in predicting churn.
- The leaf nodes contain the predicted class labels(Churn or Not Churn).
- The number of sample at each leaf node indicates the number of instances that fall into that category.

ROC Curve for Logistic and Decision Tree

- Both models demonstrate good classification performance as the ROC curves are significantly above the random guess line(diagonal line) although the decision tree seems to be slightly closer to the top left corner suggesting better balance between sensitivity and specifity.





05

EVALUATION

Model Metrics

Both models were evaluated using the following metrics:

- Accuracy: The percentage of correct predictions made by the model.
- Precision: The proportion of true positive predictions relative to the total number of positive predictions.
- Recall- The proportion of true positive predictions relative to the total number of actual positives.
- F1- Score: The harmonic mean of precision and recall.
- ROC-AUC : The area under the Receiver Operating Characteristic Curve, measuring the model's ability to discriminate between classes.

Model Performance

METRIC	LOGISTIC REGRESSION	DECISION TREE
Overall Accuracy	87%	94%
Precision(Non-Churn)	0.88	0.94
Recall (Non-Churn)	0.99	0.99
F1-Score(Non-Churn)	0.93	0.97
Precision(Churn)	0.68	0.91
Recall (Churn)	0.18	0.65
F1-Score(Churn)	0.28	0.76
ROC-AUC Score	0.812	0.842
Handling Imbalance Data	Struggles	Better

Interpretation:

- The decision Tree has a significantly higher overall accuracy.
- Both models perform well, but the Decision Tree has a higher precision.
- Both models have a high recall for non-churn.
- The Decision Tree slightly outperforms Logistic Regression in F1-Score.
- The Decision Tree makes fewer false positives for churn.
- The Decision Tress Identified more actual churn cases.
- The Decision Tree shows a better between precision and recall
- The Decision Tree has a slightly higher ROC-AUC score.
- The Decision Tree handles Imbalanced data better.

Conclusion

- **Logistic Regression:**

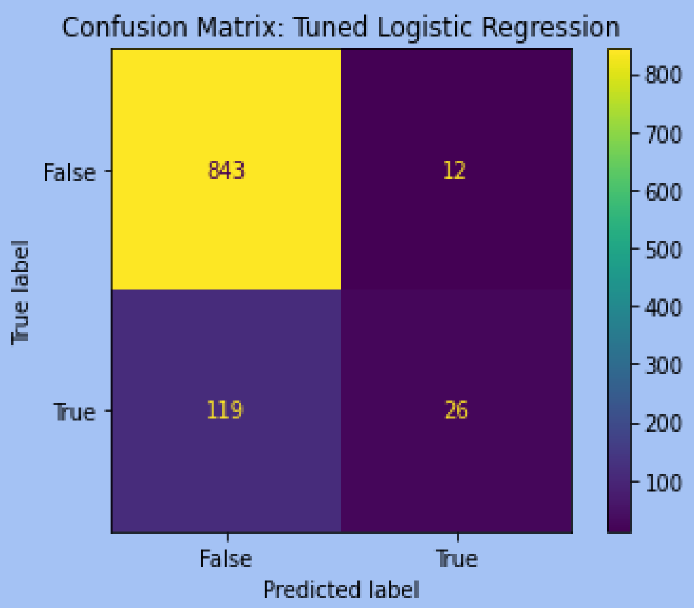
The logistic regression model performs well in identifying non-churn customers but struggles with identifying churn cases (low recall and F1-score for the churn class). Despite tuning, the model shows minimal improvement in capturing churn. This model could be useful in situations where it's crucial to avoid false alarms for churn (e.g., in maintaining positive customer relations by avoiding unnecessary interventions).

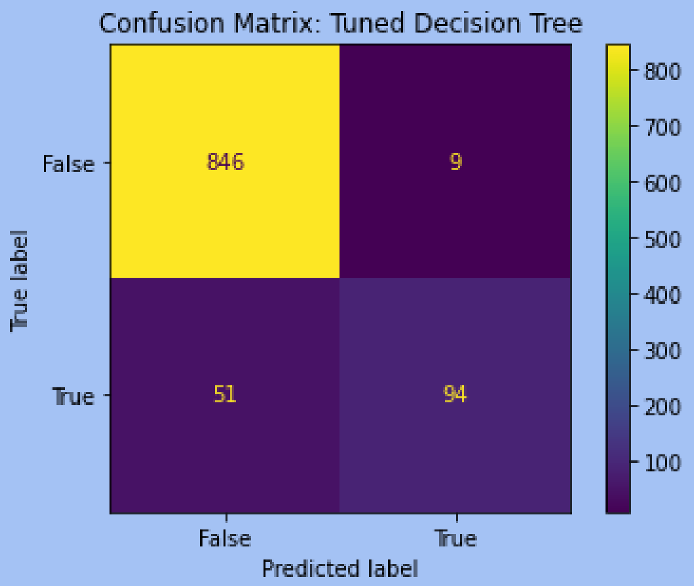
- **Decision Tree:**

The decision tree model significantly outperforms logistic regression in both overall accuracy and identifying churn cases. The tuned model offers substantial improvements in precision and F1-score for the churn class. This model is more balanced, providing better predictions for both classes. It could be useful in scenarios where correctly identifying churn is critical, such as targeting customers for retention efforts

- **The Decision Tree model is the better model because it not only achieves higher overall accuracy but also provides superior performance in identifying churn customers, which is the objective.**

Confusion Matrix





Observations:

- Logistic Regression:

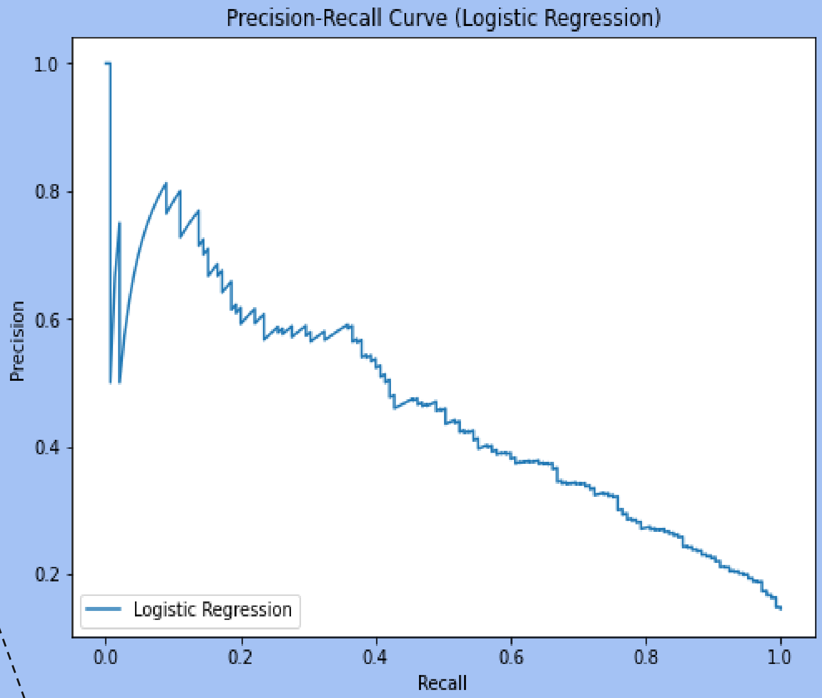
True Negatives(TN): The model correctly identifies a large number of negative instances **(843**)
True Positives(TP): The model correctly identifies a significant number of positive instances **(26)**
False Negatives(FN): There is a relatively small number of false negatives**(119)** showing that the model is not missing many positive instances.
False Positives(FP): The number of false positives is also low**(12)** suggesting the model is not incorrectly classifying negative instances as positive.

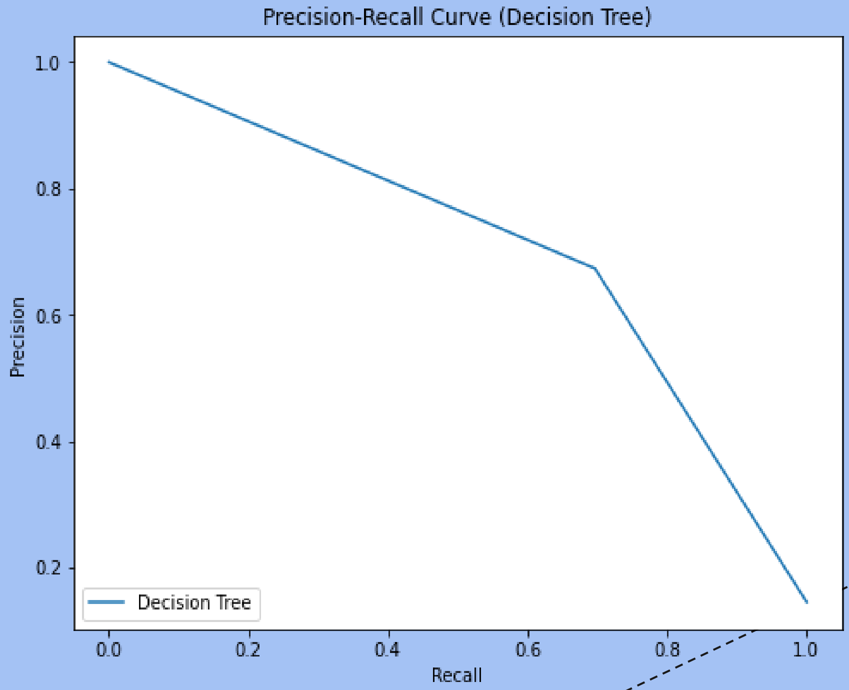
- Decision Tree:

True Negatives(TN): The model correctly identifies a large number of negative instances**(846)**
True Positives(TP): The model correctly identifies a significant number of positive instances**(94)**
False Negatives(FN): There is a relatively small number of false negatives**(51)** showing that the model is not missing many positive instances.
False Positives(FP): The number of false positives is also low**(9)** suggesting the model is not incorrectly classifying negative instances as positive.

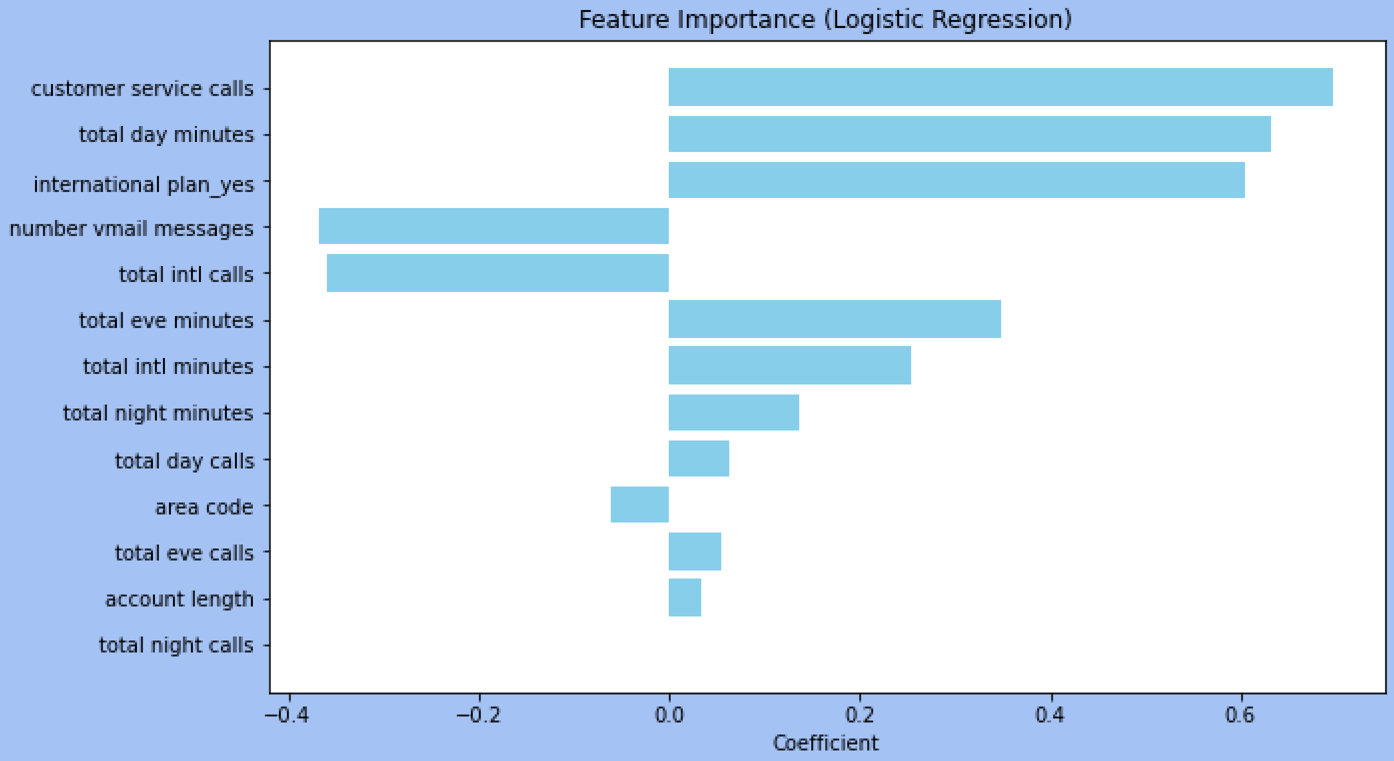
Precision-Recall Curve

- The curves generally show a downward trend indicating that as the model recalls more positive instances (higher recall), it tends to sacrifice precision(the proportion of positive predictions that are actually correct) This is a common trade-off in classification problems.

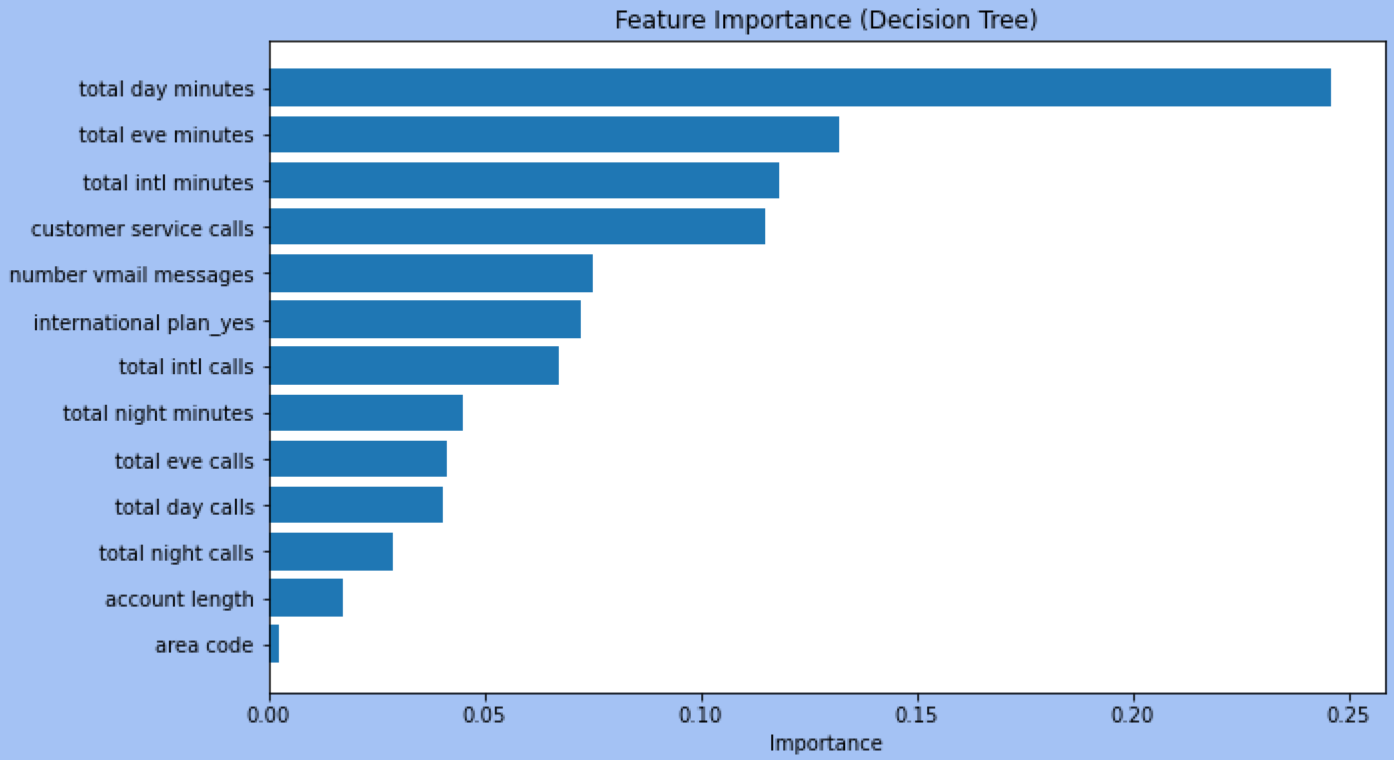


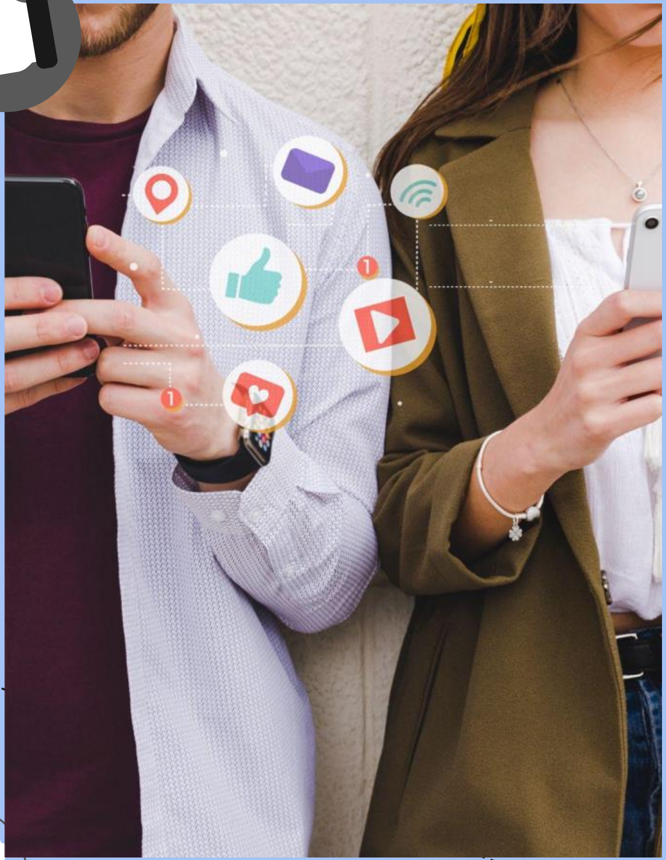


Feature Importance: Logistic Regression



Feature Importance-Decision Tree





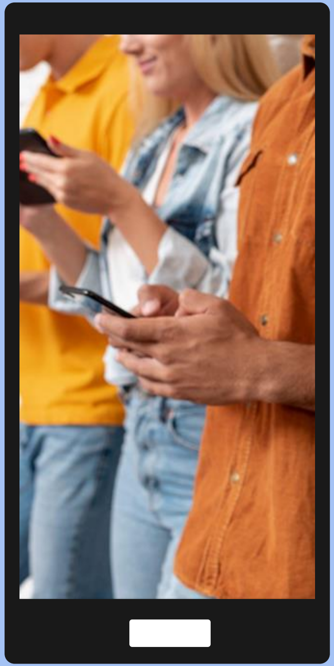
06

RECOMMENDATIONS, NEXT STEPS, THANK YOU

Recommendations

Enhancing Customer Retention

Marketing and Personalization



Monitoring and Evaluation

Operational Improvements

- ## Enhancing Customer Retention

1. **Focus on At-Risk Customers**:

- Focus retention efforts on customers predicted to churn. This could involve offering discounts, improved services, special offers, or loyalty cards to those who are flagged as high churn risk by the model.
- Create personalized retention campaigns for at-risk customers identified by the model, especially those who demonstrate early signs of dissatisfaction.
- Use predictive models to identify customers with a high likelihood of churn.

2. **Improve Customer Service and Experience Quality**:

- Since customer service interactions are a top predictor of churn, improving this experience is critical. Ensure that customer's problems are resolved quickly and effectively to prevent dissatisfaction and churn.
- The high correlation between customer service calls and churn suggests that improving customer service experiences could significantly reduce churn.
- Invest in network infrastructure to reduce call drops or enhance data quality, improve customer support services, and offer better engagement through proactive support.

3. **Incentivize Long-Term Commitment**:

- Introduce loyalty programs or long-term contracts that offer better rates or exclusive benefits to customers who commit for a longer duration.
- Create attractive packages such as bundling services, offering premium features at a discount, or exclusive rewards for staying with Syriatel for an extended period.

4. Optimize Usage Patterns:
- Analyse customer usage patterns and adjust service plans or pricing based on usage metrics. This will increase engagement and retention by aligning services with customer needs and usage behaviour.

5. Identify and Strengthen Relationships with Loyal Customers:
- Use the model to identify loyal customers and deepen their engagement with personalized offers, such as upgrades to premium services or early access to new features.
- Develop a loyalty-based segmentation strategy that rewards high-value and loyal customers with exclusive perks to keep them engaged and reduce potential churn.
- Features like "Total Day Calls" indicate that customer engagement with the service can reduce churn. Therefore, increasing customer interaction with the product through promotions, loyalty rewards, or additional services can strengthen retention.

6. Optimize Customer Onboarding:
- Ensure that new customers are fully supported during the onboarding process to reduce the risk of early churn. This includes providing clear instructions, tutorials, and access to customer support.
- Introduce an automated onboarding system with tutorials, how-to guides, and personalized assistance to reduce confusion and boost customer satisfaction from the start.

- ## Marketing and Personalization

1. **Design Targeted Marketing Campaigns**:
   - Develop marketing campaigns based on insights from the predictive models.
   - Tailor offers and promotions to specific customer segments, improving the effectiveness of marketing efforts and enhancing customer engagement.

2. **Leverage Customer Data**:
   - Utilize customer data insights to create personalized experiences.
   - Increase customer satisfaction and loyalty by providing relevant and personalized interactions.

- ## Monitoring and Evaluation

1. **Regularly Review Model Performance**:
   - Continuously monitor and evaluate the performance of predictive models.
   - Ensure models remain accurate and relevant, making adjustments as needed to maintain effectiveness.

2. **Update Strategies Based on New Insights**:
   - Incorporate new data and insights into business strategies
   - Stay responsive to changing customer behaviours and market conditions, ensuring strategies remain effective and relevant.

- ## Operational Improvements

1. **Allocate Resources Efficiently**:
   - Use model predictions to guide resource allocation in customer service and support.
   - Optimize resource usage to address high-risk customers effectively and improve overall operational efficiency.

2. **Adjust Pricing Strategies**:
   - Analyse pricing structures based on customer usage and churn predictions.
   - Develop pricing strategies that better align with customer value and usage patterns, potentially reducing churn and increasing revenue.
   - If payment-related features (e.g., late payments or prepaid balances) are linked to churn, improve flexibility in payment options or offer better reminders and support around payment deadlines.
   - Features related to specific plans, such as the "International Plan," suggest that pricing and plan features play a crucial role in customer retention. Offering more value in these plans, ensuring they meet customers' needs, or providing personalized plan recommendations could help reduce churn.
   - Implement easier payment plans, send payment reminders through SMS/email, and offer auto-debit options to prevent service disruption.

NEXT STEPS

Model Perfomance Monitoring

- Regularly monitor the performance of both models.
- Ensure that models remain accurate and relevant overtime making adjustments.
- Eliminate the least important features for faster and more interpretable models.

Feature Re-evaluation

- Further refine features to improve predictive power.
- Reasses and potentially reintroduce features that were dropped during feature selection.

Focus on Top Features

- Base business recomendations on top-performing features eg developing initiatives to improve early customer experiences couls have a strong impact.

Experiment with Advanced Techniques

- Explore advanced modeling techniques.
- Enhance predictive accuracy and address any limitations observed in current models.

NEXT STEPS

Continuous Data Improvement

- Invest in collecting more data and improving data quality.
- Provide richer data for model training leading to better predictive performance and more accurate insights.

Stakeholder Engagement

- Regularly engage with stakeolders to understand their evolving needs and how model predictions is applied
- Ensure predictive models align with business goals to provide actionable insights for decision-making.

Update and Refine Business Strategies

- Use insights from model evaluations and performance monitoring to update and refine business strategies.
- Align strategies with predictive insights to enhance customer retention, marketing effectiveness and overall business performance.

Thanks!

Do you have any questions?

angiekale.kale@gmail.com
+254791528138
http://linkedin.com/in/angela-kalelwa-69744a212

CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon**, and infographics & images by **Freepik**