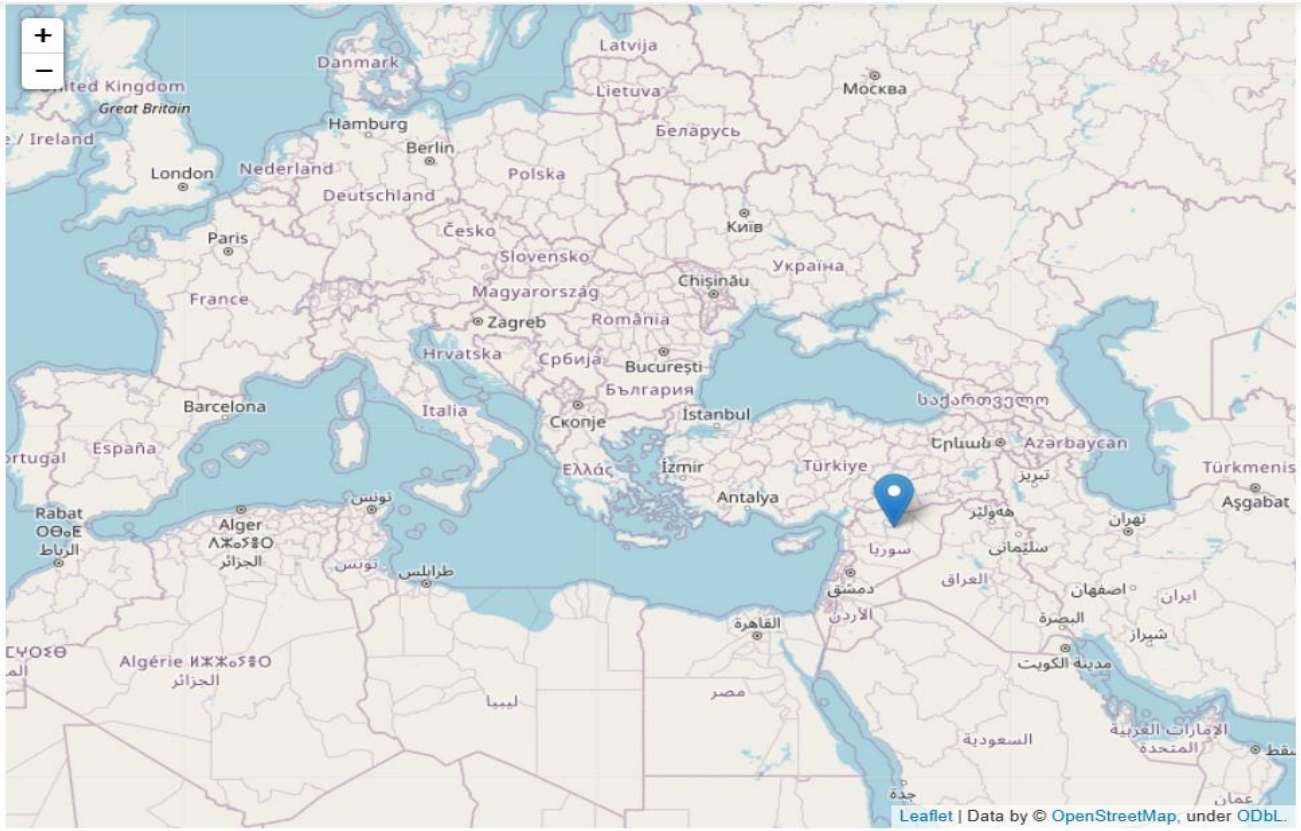
# PREDICTIVE MODEL ON CUSTOMER CHURN RATE FOR SYRIATEL COMPANY

# Location of Syria




## Deductions

- Using the Area Code provided to locate the Location of Syria.
- Officially Syrian Arab Republic.
- Located in West Asia located in the Eastern Mediterranean and the Levant.
- Bounded by the Mediterranean Sea to the west, Turkey to the north, Iraq to the east and southeast, Jordan to the south, and Israel and Lebanon to the southwest.

# Project Overview

**SyriaTel Customer Churn Prediction**

- SyriaTel , a telecommunications company, aims to improve customer retention by predicting customer churn.
- Churn refers to customers who discontinue their service subscription.

**Project Objectives:**

- Analyze customer data to understand patterns and factors contributing to churn.
- Develop and validate predictive models to accurately identify potential churners.
- Provide actionable insights for targeted customer retention strategies.

**Main Stakeholders:**

- **Senior Management:** Interested in overall business impact and strategic insights for decision-making.
- **Marketing Team:** Needs to design targeted retention campaigns based on model predictions.
- **Customer Service Team:** Requires insights to proactively address customer issues and improve service.
- **Data Science Team:** Responsible for developing, validating, and maintaining the predictive model.
- **IT Department:** Supports data infrastructure, model deployment, and integration with existing systems.
- **Sales Team:** Uses insights to enhance customer interaction and retention efforts.

# Business Understanding

- Customer churn, the phenomenon where customers cease doing business with a company, is a critical concern for telecommunications companies like SyriaTel.

- Retaining customers is essential for maintaining revenue and growth in this competitive industry.
- Identifying factors contributing to churn, such as service dissatisfaction or competitive offers, SyriaTel can take targeted actions to mitigate churn and improve customer retention.

# Problem Statement

- SyriaTel, a telecommunications company, faces the challenge of customer churn, where customers discontinue their services.
- This attrition impacts revenue and profitability.
- The business seeks to proactively identify customers at risk of churning and implement effective retention strategies to mitigate revenue loss and maintain customer loyalty.

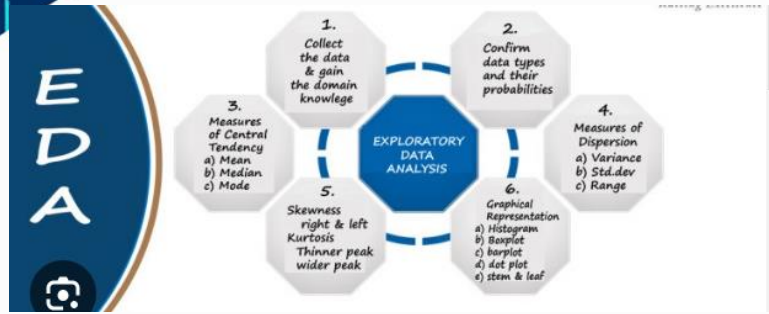
- **Specifically, the project aims to address the following questions:**

- 1. What are the primary factors driving customer churn for SyriaTel?

- 2. Which machine learning modelling technique to apply in accurately predicting Churn so as to take proactive measures?

- 3. What actionable insights can SyriaTel derive from the predictive model to improve customer retention efforts?

- 4. What strategies can SyriaTel put in place to reduce churn rate?




# Exploratory Data Analysis

# DATA Understanding

## SyriaTel

- For purposes of this project, the ***'SyriaTel_df.csv***' dataset was used.
- The dataset had 3,333 rows and 21 columns
- The columns provided had numerical, categorical and string  data types.

- **Data frame  understanding involved investigating:**

✓ Missing values
✓ Head and tail
✓ Duplicates
✓ Unique values
✓ Univariant, Bivariant and Multivariate Analysis

**Data Preparation involved**

✓ Outlier identification and handling.
✓ One hot and Cording Categorical Columns.

- After Data preparation, a dataframe of 3,333 rows and 67 columns were adopted for further analysis.
- Main columns colums considered for the analysis included
- Numerical Colums :'Account Length', 'Area Code', 'Number Vmail Messages', '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‘
- Categorical Columns: “State”, “International Plan, “Voice Mail Plan”,”Churn”

# Descriptive Analysis

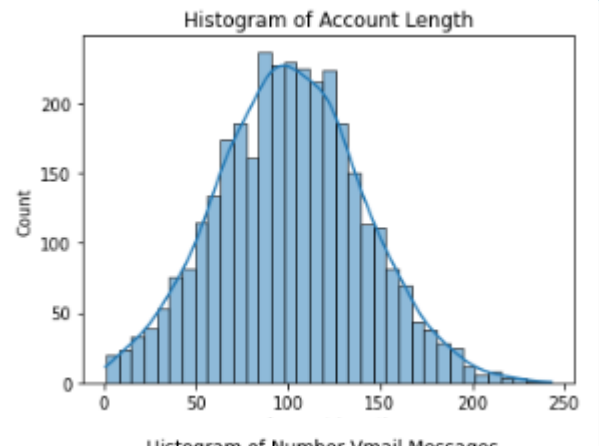
| | Account Length | Area Code | Number Vmail Messages | Total Day Minutes | Total Day Calls | Total Day Charge | Total Eve Minutes | Total Eve Calls | Total Eve Charge | Total Night Minutes | Total Night Calls | Total Night Charge |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| count | 3333.0 | 3333.0 | 3333.0 | 3333.0 | 3333.0 | 3333.0 | 3333.0 | 3333.0 | 3333.0 | 3333.0 | 3333.0 | 3333.0 |
| mean | 101.06480648064805905 | 437.18241824182416622 | 8.0990099009900991 | 179.7750975097509354 | 100.4356435643564396 | 30.562307230723075 | 200.9803480348034839 | 100.11431143114310771 | 17.08354035403540294 | 200.8720372037203674 | 100.1077107710771088 | 9.0393249324932494 |
| std | 39.8221059285956045 | 42.3712904856066146 | 13.6883653720385983 | 54.46738920237137194 | 20.0690842073008966 | 9.2594345539305003 | 50.7138444258119989 | 19.9226252939431028 | 4.31066764311034056 | 50.5738470136583587 | 19.5686093460585582 | 2.275872837660029 |
| min | 1.0 | 408.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 23.1999999999999993 | 33.0 | 1.040000000000000036 |
| 25% | 74.0 | 408.0 | 0.0 | 143.6999999999999886 | 87.0 | 24.4299999999999997 | 166.5999999999999943 | 87.0 | 14.160000000000000014 | 167.0 | 87.0 | 7.5199999999999996 |
| 50% | 101.0 | 415.0 | 0.0 | 179.4000000000000057 | 101.0 | 30.5 | 201.4000000000000057 | 100.0 | 17.120000000000001 | 201.19999999999998863 | 100.0 | 9.05000000000000071 |
| 75% | 127.0 | 510.0 | 20.0 | 216.4000000000000057 | 114.0 | 36.7899999999999991 | 235.30000000000001137 | 114.0 | 20.0 | 235.30000000000001137 | 113.0 | 10.5899999999999999 |
| max | 243.0 | 510.0 | 51.0 | 350.8000000000000114 | 165.0 | 59.6400000000000006 | 363.6999999999999886 | 170.0 | 30.9100000000000001 | 395.0 | 175.0 | 17.7699999999999996 |

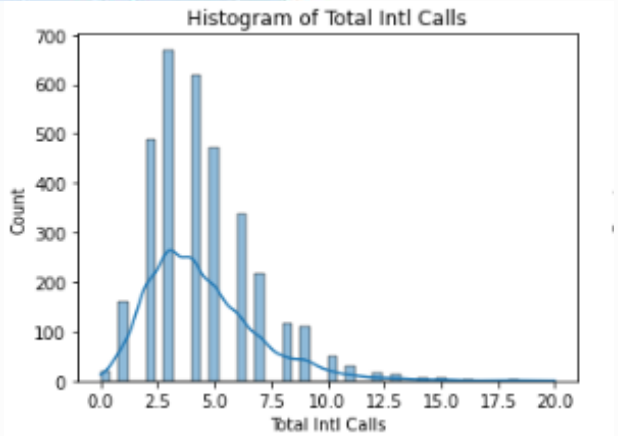
| | total night charge | total intl minutes | total intl calls | total intl charge | customer service calls |
|---|---|---|---|---|---|
| count | | | | | |
| mean | 12.5600000000000005 | 9.9000000000000004 | 6 | 2.6699999999999999 | 2 |
| std | 8.6099999999999994 | 9.5999999999999996 | 4 | 2.5899999999999999 | 3 |
| min | 8.6400000000000006 | 14.099999999999999964 | 6 | 3.81 | 2 |
| 25% | 6.2599999999999998 | 5.0 | 10 | 1.3500000000000001 | 2 |
| 50% | 10.8599999999999994 | 13.6999999999999993 | 4 | 3.7000000000000002 | 0 |
| 75% | | | | | |
| max | | | | | |

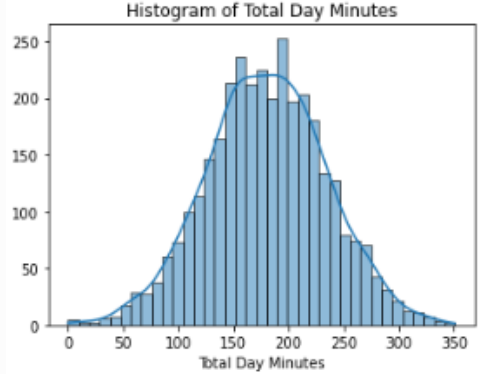
## Deductions

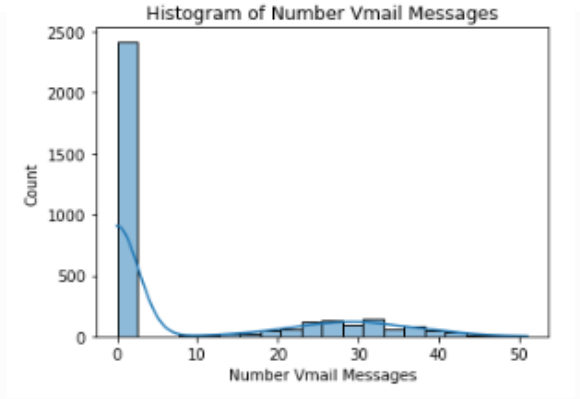
- Total number of customer is 3,333.
- Mean account length 101.1.
- Max account length 243.
- Mean Total Day Calls is approximately 100 calls.
- Max Total Night Calls is 175.
- Std for Total Day Change is 9.3.
- Max Customer Service Calls is 9.

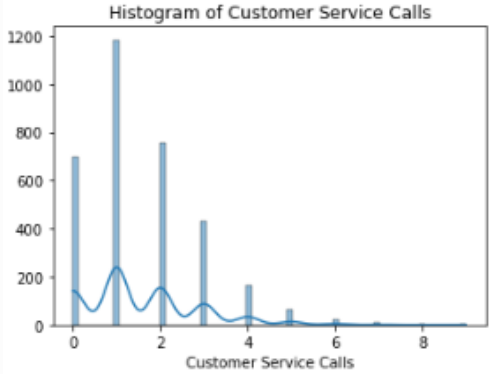
# Univariant Analysis



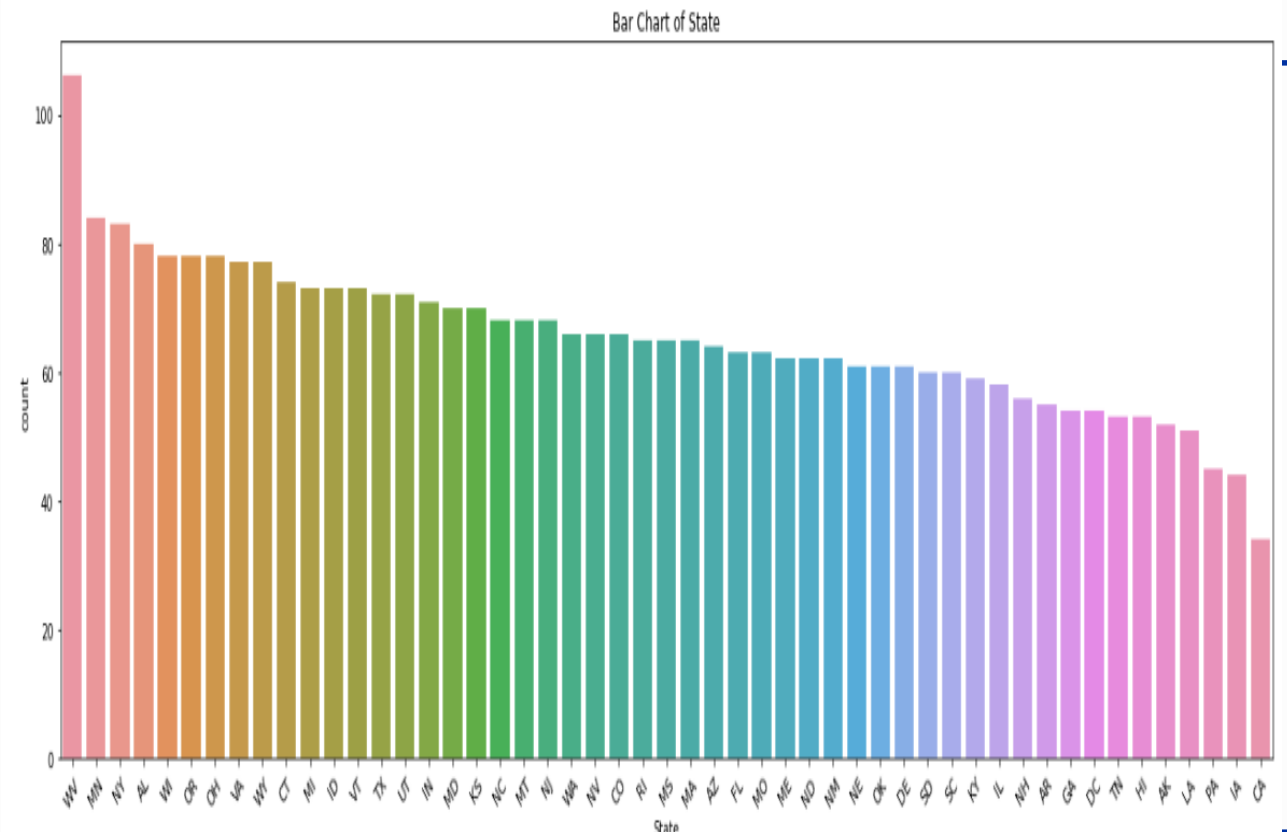




## Deductions

- Account length is positively skewed.
- Total Intl Calls is positively skewed.
- Total Day Minutes is nearly uniformly distributed.
- Majority of customers do not use the Voice Mail Messages.

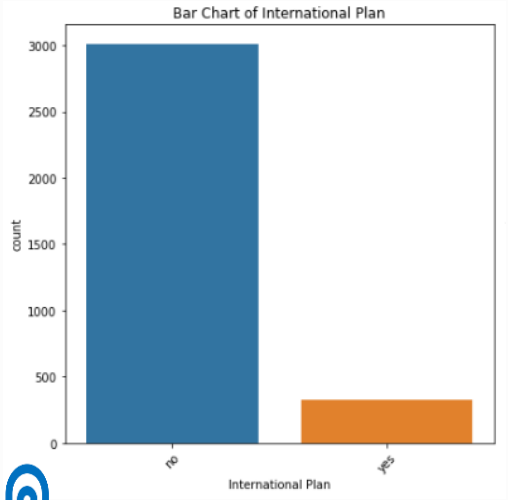
# Univariant Analysis

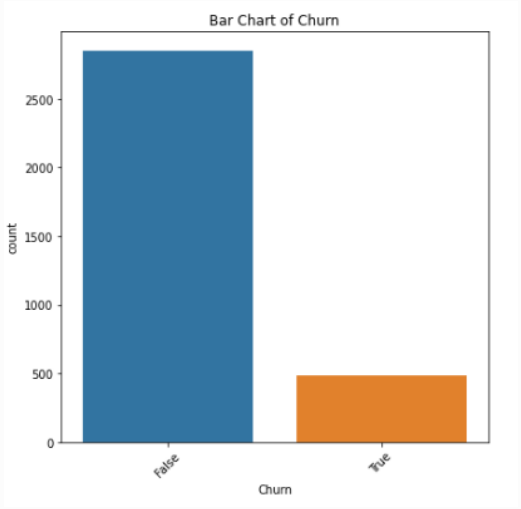


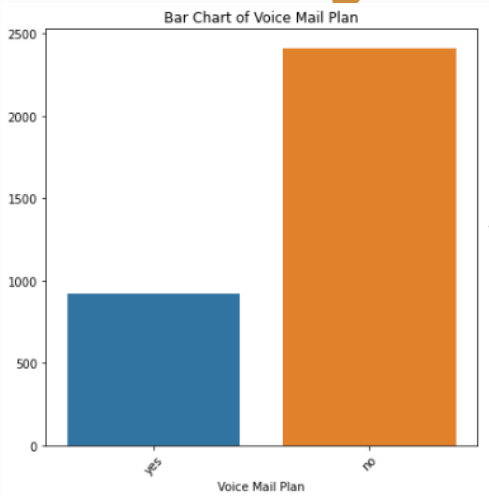

## Deductions

- WV state has the highest number of customers.
- CA state has the lowest number of customers.
- Other states with considerable number of customers include:
- MN, NY,AL,WI,CR,CH among others.

# Univariant Analysis



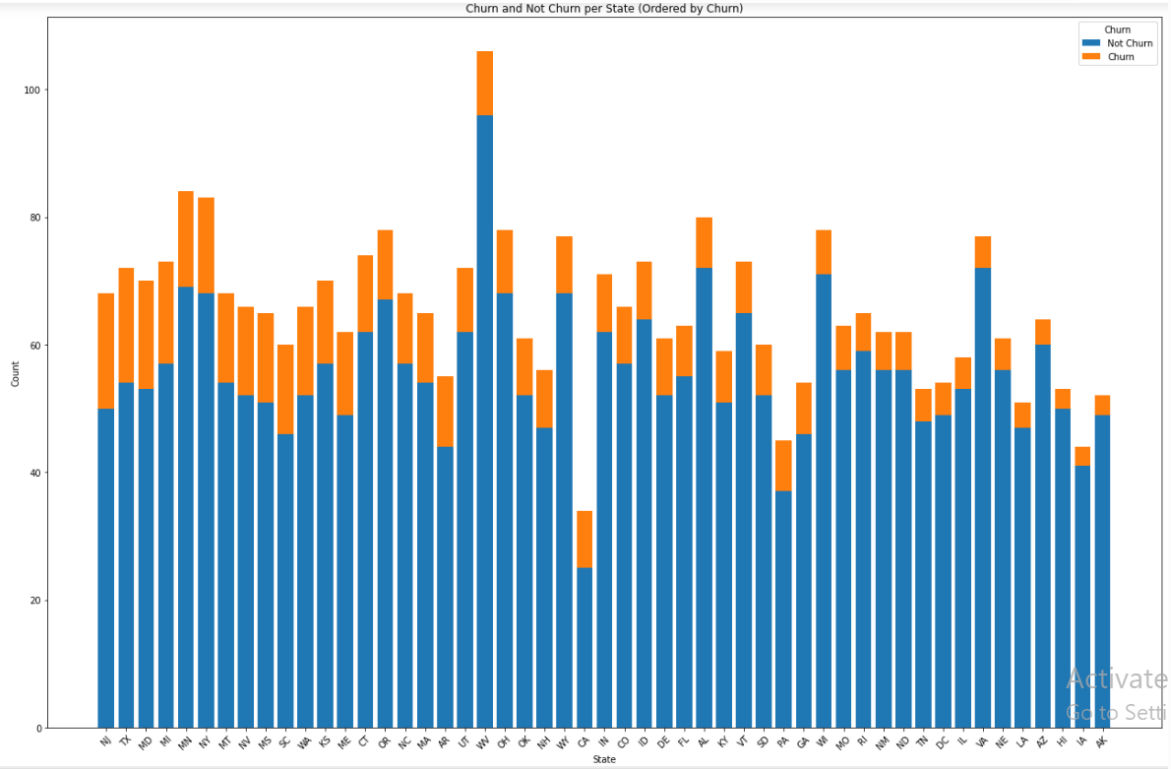




## Deductions

- About 3000 customers are not enrolled for the International Plan
- Approximately 500 customers are churn.
- About 2400 customers have not enrolled for the Voice Mail Plan.

# Bivariant Analysis



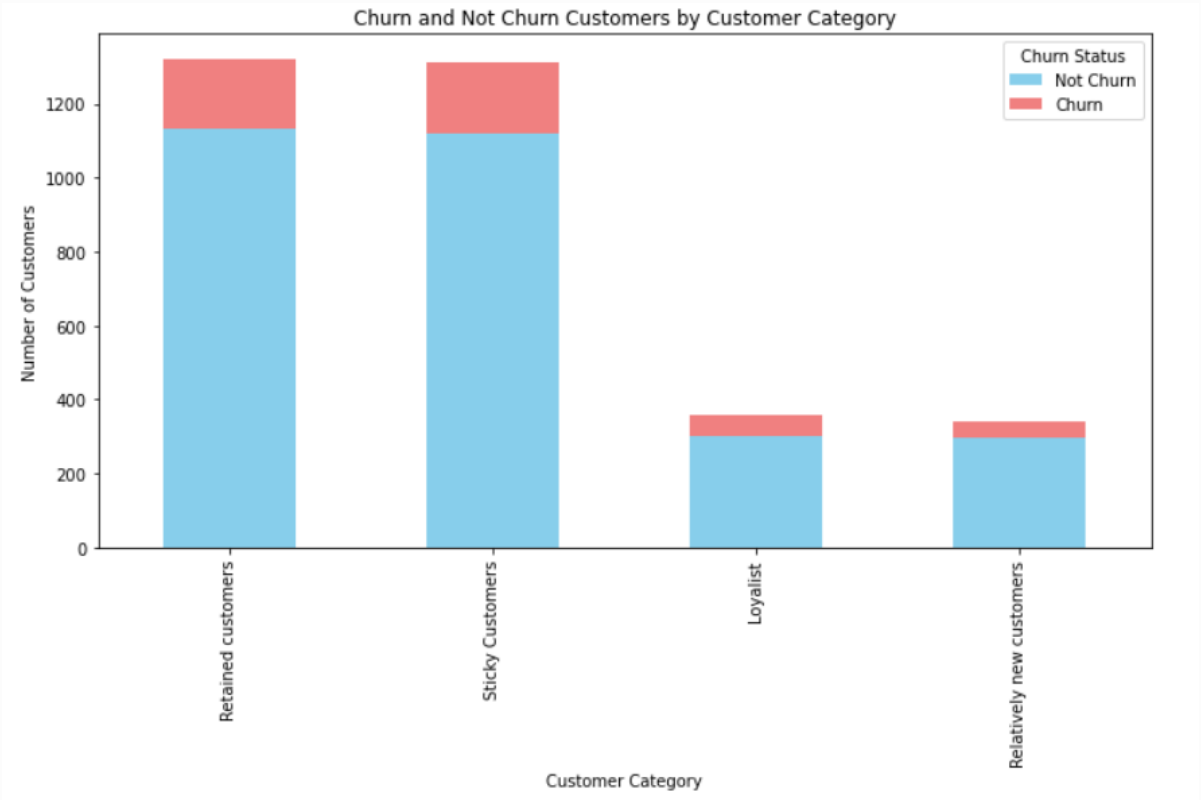

Deductions

- NJ state has the highest churn.
- AK state has the least churn.

- Other states with considerable churn include:

TX
MD
MI
AM
NY
MT
MV
MS
SC

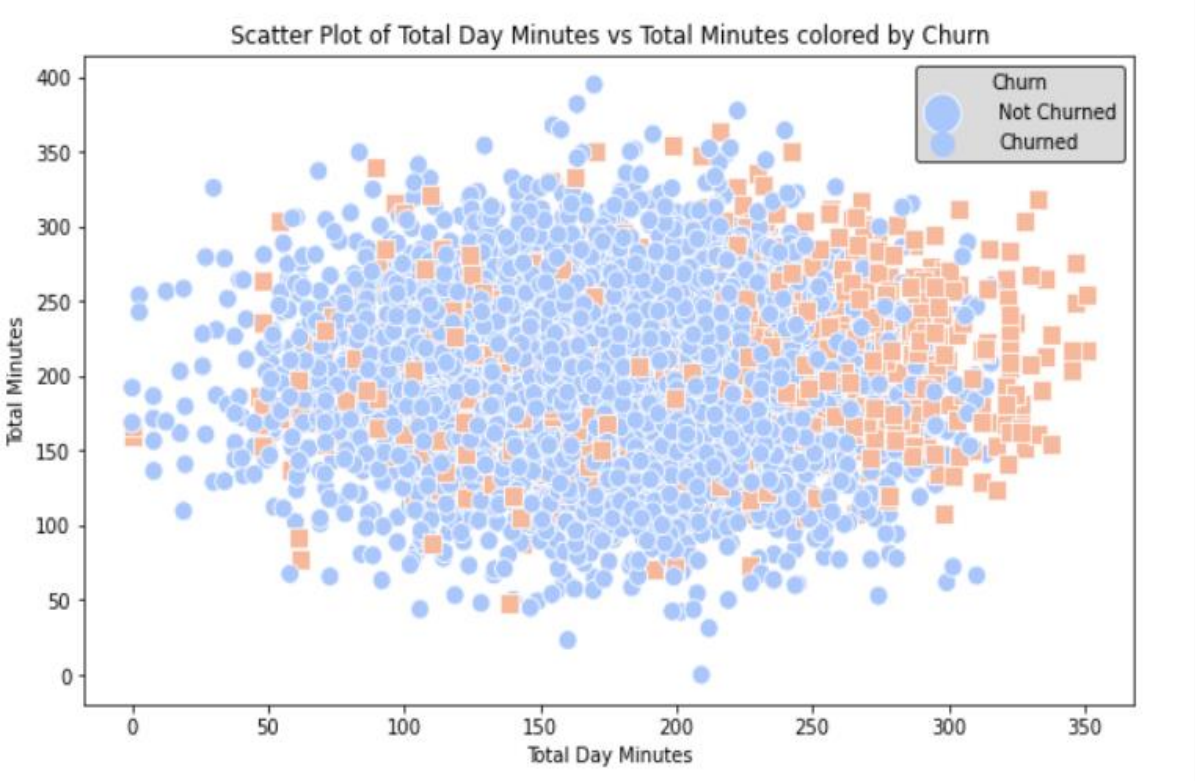
# Bivariant Analysis




## Deductions

- Relatively new customers are those with account length about 50.
- Retained customers have account about 100.
- Sticky customers have account length above 100 but less 150
- Loyalists have account length above 150.
- Retained and Sticky customer categories have majority of the churn.
- Loyalists and relatively new customers have least least churn.
- Majority of the total customers are under the Retained and Sticky customer categories.

# Multivariant Analysis




Deductions

There are few churn within the Total Minutes and Total Day Minutes features. This insight can be investigated for a better understanding.

# Normality and Spread of the Cleaned Dataset.

| | Skewness | Kurtosis |
|---|---|---|
| Account Length | 0.09656281161489656 | -0.1094739184341575 |
| Number Vmail Messages | 1.2642543349768245 | -0.0528515105905245 |
| Total Day Minutes | -0.0290639795181198 | -0.0217101179240888 |
| Total Day Calls | -0.1117363237307519 | 0.24101722895174227 |
| Total Day Charge | -0.0290701779270378 | -0.0215817191450336 |
| Total Eve Minutes | -0.0238667088046375 | 0.0237916804447047 |
| Total Eve Calls | -0.0555381300016192 | 0.20404769217448226 |
| Total Eve Charge | -0.023847250496277 | 0.02364954586272594 |
| Total Night Minutes | 0.008917275580987895 | 0.08388775499253365 |
| Total Night Calls | 0.03248494205404463 | -0.0737112242125884 |
| Total Night Charge | 0.008882237062694412 | 0.08373508611499814 |
| Total Intl Minutes | -0.2450256034866443 | 0.606471635404318 |
| Total Intl Calls | 1.3208833668164015 | 3.07716543898885142 |
| Total Intl Charge | -0.2451761045009844 | 0.6068966666527675 |

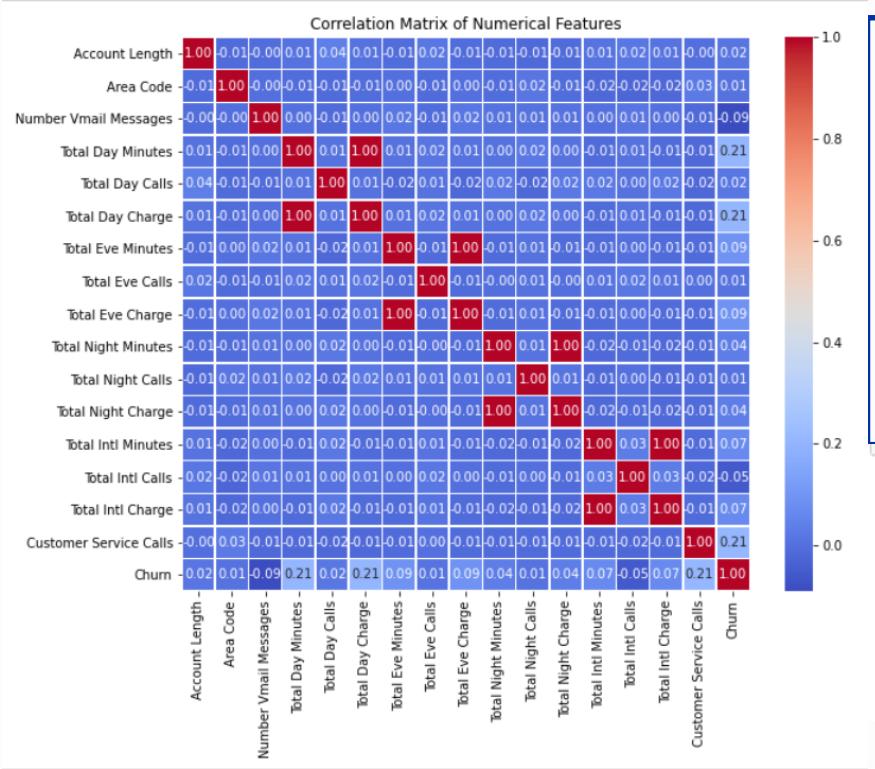
| | Std Dev |
|---|---|
| Account Length | 39.8221059285956045 |
| Number Vmail Messages | 13.6883653720385983 |
| Total Day Minutes | 54.46738920237137194 |
| Total Day Calls | 20.0690842073008966 |
| Total Day Charge | 9.2594345539305003 |
| Total Eve Minutes | 50.7138444258119989 |
| Total Eve Calls | 19.9226252939431028 |
| Total Eve Charge | 4.31066764311034056 |
| Total Night Minutes | 50.5738470136583587 |
| Total Night Calls | 19.5686093460585582 |
| Total Night Charge | 2.275872837660029 |
| Total Intl Minutes | 2.791839548408416 |
| Total Intl Calls | 2.461214270546094 |
| Total Intl Charge | 0.753772612663046 |
| Customer Service Calls | 1.3154910448664767 |

| | Mean | Median |
|---|---|---|
| Account Length | 101.06480648064805905 | 101.0 |
| Number Vmail Messages | 8.0990099009900991 | 0.0 |
| Total Day Minutes | 179.7750975097509354 | 179.4000000000000057 |
| Total Day Calls | 100.4356435643564396 | 101.0 |
| Total Day Charge | 30.562307230723075 | 30.5 |
| Total Eve Minutes | 200.9803480348034839 | 201.4000000000000057 |
| Total Eve Calls | 100.11431143114310771 | 100.0 |
| Total Eve Charge | 17.08354035403540294 | 17.120000000000001 |
| Total Night Minutes | 200.8720372037203674 | 201.19999999999998863 |
| Total Night Calls | 100.1077107710771088 | 100.0 |
| Total Night Charge | 9.03932493249324942 | 9.05000000000000071 |
| Total Intl Minutes | 10.23729372937293824 | 10.3000000000000007 |
| Total Intl Calls | 4.4794479447944795 | 4.0 |
| Total Intl Charge | 2.7645814581458144 | 2.7799999999999998 |
| Customer Service Calls | 1.5628562856285628 | 1.0 |

## Deductions

Account length is positively skewed.
Total Intl calls is negatively skewed
Total Day Minutes has the highest std.
Total Intl Charge has the lowest std.
Avg Total Day Charge is 30.5
Avg Total Eve Calls is 100.1
Avg Intl minutes 10.2

# Correlation Matrix




## Deductions

- Features that are highly correlated include: Total Day Charge and Total Day Minutes.
- Total Eve Charge and Total Eve Minutes.
- Total Night Charge and Total Nights Minutes Total Intl charge and Total Minutes.
- This offers insights on opportunities for better packages and loyalty programs.

| | Feature 1 | Feature 2 | Correlation |
|---|---|---|---|
| 13 | Total Day Charge | Total Day Minutes | 0.999999952190397 |
| 34 | Total Eve Charge | Total Eve Minutes | 0.9999997760198517 |
| 64 | Total Night Charge | Total Night Minutes | 0.99999921487588 |
| 103 | Total Intl Charge | Total Intl Minutes | 0.9999927417510258 |

# Hypothesis testing

**Hypothesis Testing**

- **Null Hypothesis (H0):** There is no significant influence of the various factors to churn rate in SyriaTel.

- **Alternate Hypothesis (H1):** There is a significant influence of the various factors to churn rate in SyriaTel.

# Hypothesis results and conclusions

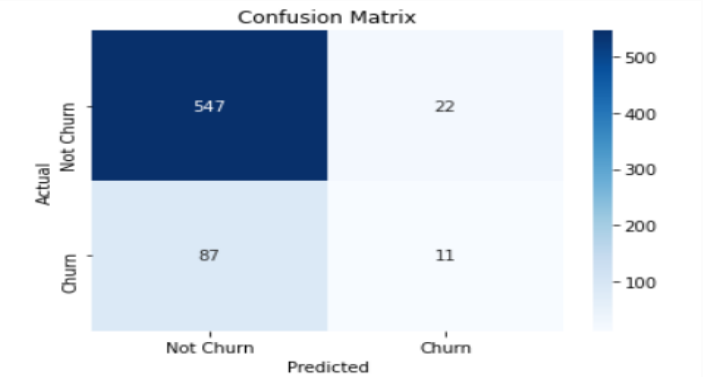
| | Feature | F-Statistic | p-value |
|---|---|---|---|
| 0 | Account Length | 0.9115981986407352 | 0.3397600070569128 |
| 1 | Area Code | 0.12698640858136082 | 0.7215998968016037 |
| 2 | Number Vmail Messages | 27.035911709557691296 | 0.00000021175218402696 |
| 3 | Total Day Minutes | 146.35078521943776764 | 0.00000000000000000000 |
| 4 | Total Day Calls | 1.13541242989728808 | 0.28670102402414055 |
| 5 | Total Day Charge | 146.35065699096048775 | 0.00000000000000000000 |
| 6 | Total Eve Minutes | 28.9325766446506485 | 0.00000000801133856128 |
| 7 | Total Eve Calls | 0.2839943754492388 | 0.5941305829778143 |
| 8 | Total Eve Charge | 28.926443755197127 | 0.00000000803652422777 |
| 9 | Total Night Minutes | 4.20149555022397259 | 0.0404664846378868 |
| 10 | Total Night Calls | 0.12563131916004017 | 0.7230277872159787 |
| 11 | Total Night Charge | 4.2021362787384957 | 0.04045121876901292 |
| 12 | Total Intl Minutes | 15.5834679864501915 | 0.00008005731126549902 |
| 13 | Total Intl Calls | 9.3279453654346529 | 0.002274701409848483 |
| 14 | Total Intl Charge | 15.5925806081700724 | 0.00008001875358306397 |
| 15 | Customer Service Calls | 151.7670126303964366 | 0.00000000000000000000 |

## Conclusions

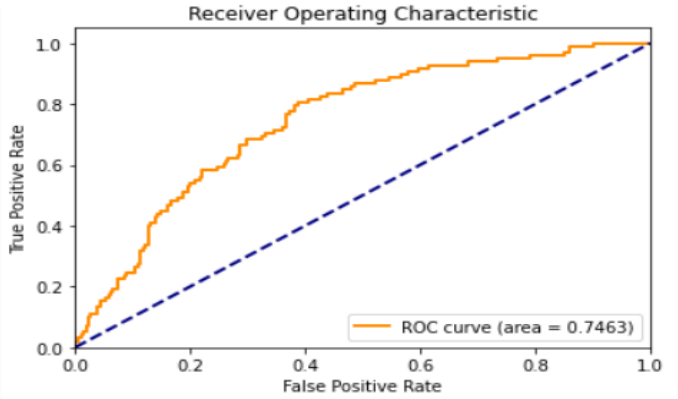
Features such as 'Account Length', 'Area Code', 'Total Day Calls', 'Total Eve Calls', and 'Total Night Calls' have their p-values are greater than the significance level of 0.05.
Therefore, we fail to reject the null hypothesis (H0) for these features. This suggests that there is no significant influence of these factors on the churn rate in SyriaTel.

While features including 'Number Vmail Messages', 'Total Day Minutes', 'Total Day Charge', 'Total Eve Minutes', 'Total Eve Charge', 'Total Night Minutes', 'Total Night Charge', 'Total Intl Minutes', 'Total Intl Calls', 'Total Intl Charge', and 'Customer Service Calls', the p-values are extremely low (close to 0).
Therefore, we reject the null hypothesis (H0) for these features. This indicates that there is a significant influence of these factors on the churn rate in SyriaTel.
In conclusion, there is evidence to suggest that most numerical features have a significant influence on the churn rate in SyriaTel, except for 'Account Length', 'Area Code', 'Total Day Calls', 'Total Eve Calls', and 'Total Night Calls'

# Modelling:Logistic Regression




```
Confusion Matrix:
[[547  22]
 [ 87  11]]
Accuracy: 0.8366
Precision: 0.3333
Recall: 0.1122
F1-Score: 0.1679
ROC-AUC: 0.7463
```

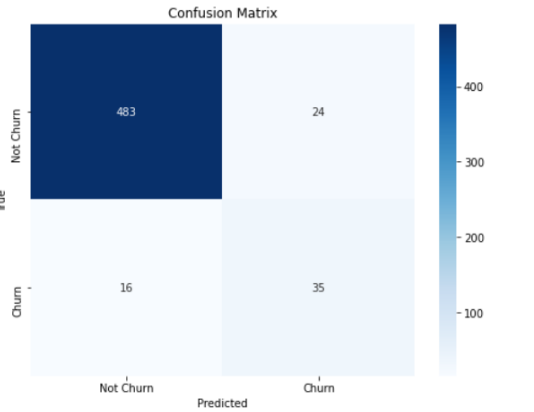



**Criteria for splitting data**:
Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.20, random_state=1, stratify=y)

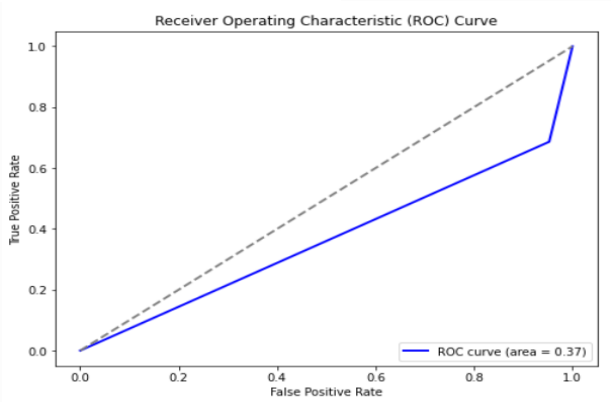
## Deductions

- The model has high accuracy but struggles with precision and recall for the churn class.
- Suggesting that while it correctly predicts the majority of 'no churn' cases,
- It fails to adequately identify 'churn' cases.
- Therefore the need to consider other model techniques.

# Decision Tree Model



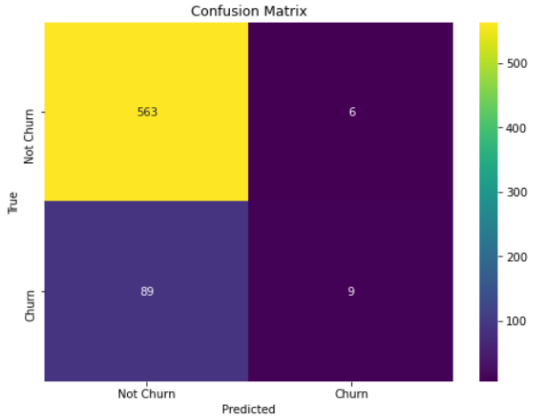

```
Confusion Matrix:
[[483  24]
 [ 16  35]]
Accuracy: 0.9283
Precision: 0.5932
Recall: 0.6863
F1-Score: 0.6364
ROC-AUC: 0.3668
```



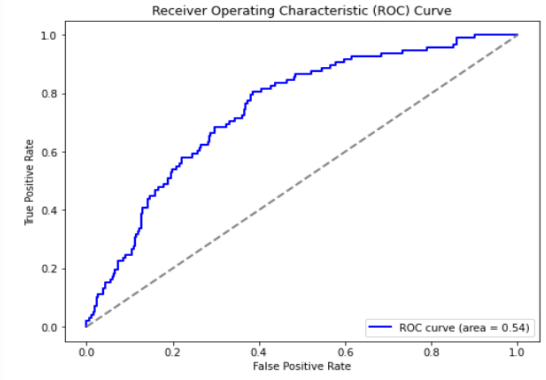

## Deductions

- Model has an Accuracy of 92.8%.
- Precision of 59.3%.
- Recall of 68.6%.
- F1-Score 63.6%
- ROC-AUC of 36.68%.
- While the accuracy is improved, ROC-AUC is still low thus need for another model.

# KNN model



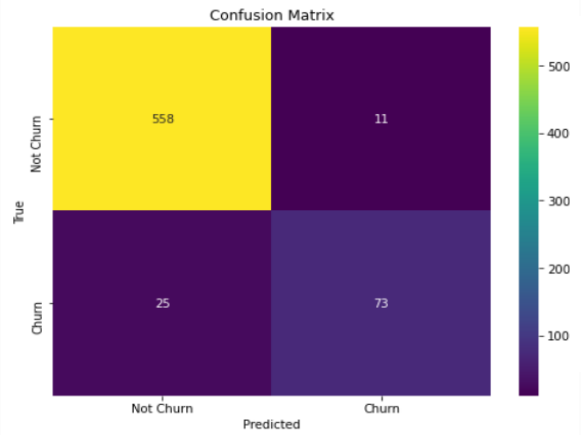

KNN Model Performance:
Accuracy: 0.8575712143928036
Precision: 0.6
Recall: 0.09183673469387756
F1-Score: 0.1592920353982301
ROC-AUC Score: 0.5406459596140741



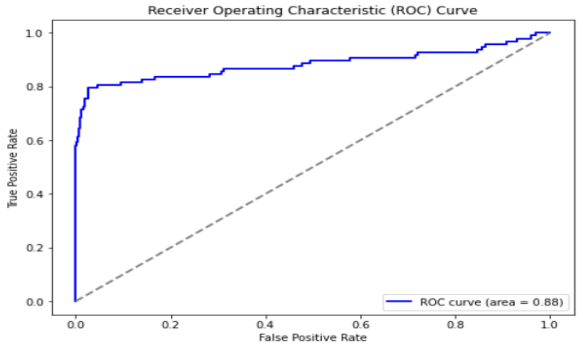

## Deductions

- Model has an Accuracy of 85.7%.
- Precision of 60%.
- Recall of 9.2%.
- F1-Score 15.9%
- ROC-AUC of 54.1%.
- The accuracy is reduced while it has a higher ROC-AUC Score compared with the previous model.

# XGBoost Model




```
XGBoost Model Performance:
Accuracy: 0.9460269865067467
Precision: 0.8690476190476191
Recall: 0.7448979591836735
F1-Score: 0.8021978021978022
ROC-AUC Score: 0.8837380294824432
Confusion Matrix:
[[558  11]
 [ 25  73]]
```




## Deductions

The model has an accuracy of 94.6%.
Precision of 86.9%
Recall of 74.5%
F1 Score of 80.2%.
ROC-AUC Score of 88.3%
By comparison, the model performs better.

# Model Performance

```
Performance Comparison:
          Logistic Regression      Decision Tree                  KNN  \
Accuracy    0.8455772113943029  0.896551724137931  0.8545727136431784
Precision 0.41379310344827586 0.6306306306306306  0.5294117647058824
Recall    0.12244897959183673 0.7142857142857143 0.09183673469387756
F1-Score    0.1889763779527559 0.6698564593301436  0.1565217391304348
ROC-AUC     0.7757074710376242 0.8211147376349486  0.6655876761952585
Average     0.4693006286849591 0.7464878532038736 0.45958612567372625

                    XGBoost
Accuracy  0.9460269865067467
Precision 0.8690476190476191
Recall    0.7448979591836735
F1-Score  0.8021978021978022
ROC-AUC   0.8837380294824432
Average    0.849181679283657

Best Model for each metric:
Accuracy     XGBoost
Precision    XGBoost
Recall       XGBoost
F1-Score     XGBoost
ROC-AUC      XGBoost
dtype: object
```

## Deductions

- XG Boost is the best model by all metrics comparison including.
- Accuracy
- Precision
- Recall
- F1 Score
  ROC-AUC.
- This is the best model to be deployed by SyriaTel to proactively identify and manage churn.

# Cross Validation of the Models and making comparisons

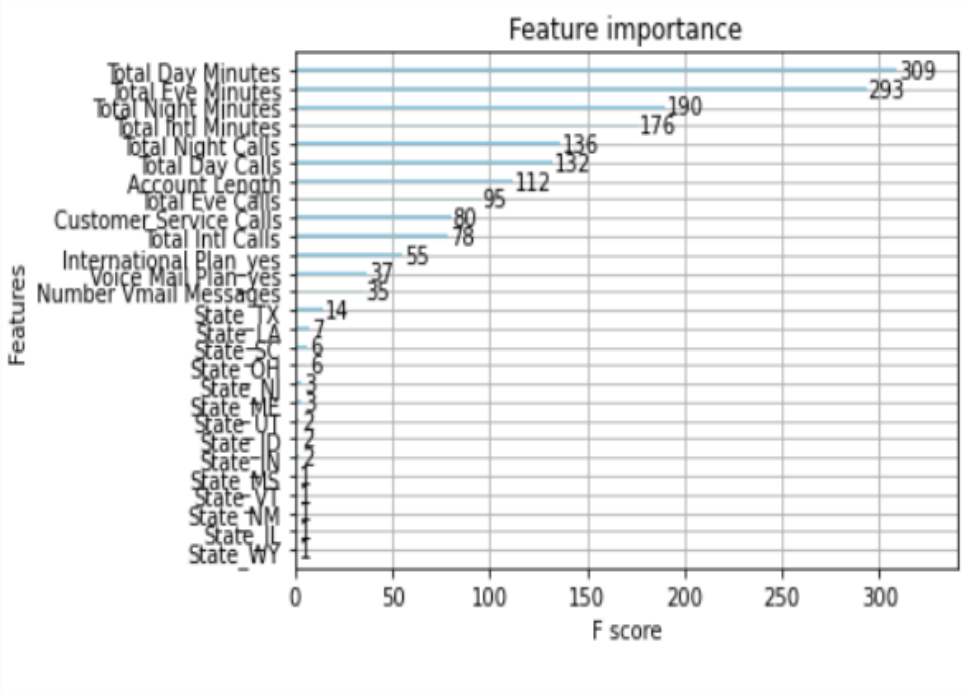
```
Logistic Regression: Average Cross-validation Score = 0.8668
Decision Tree: Average Cross-validation Score = 0.9066
KNN: Average Cross-validation Score = 0.8571
XGBoost: Average Cross-validation Score = 0.9542

Best Model: XGBoost with Average Cross-validation Score = 0.9542
```

## Deductions

By comparison XGBoost has the best Cross Validation Score at 95.4%.

This is enough proof the choosing the model for churn management.

# Variable of Importance




Deductions

- The following variables were identified as Variables of Importance in desceding order include.
- Total Day Minutes
  Total Eve Minutes
- Total Night Minutes
- Total Night Calls
- Account Length
- Total Eve Calls
- Customer Service Calls
- Total Intl Calls
- Intl Pla
- Number of Vmail Messeges
- This provides insights of the variables to consider in developing churn mitigation strategies

# Findings

Deductions

**The primary factors driving customer churn for SyriaTel**.

Based on the feature importance from the XGBoost model:

Total Day Minutes,Total Eve Minutes,Total Intl Minutes, Total Night Calls,Total Day Calls,Account Length,Total Eve Calls,Customer Service Calls, Total Intl Calls, Voice Main Plan and Number of Voice Mail Messages are the main primary factors influencing churn.

2. **Machine learning modelling technique to apply in accurately predicting Churn so as to take proactive measures**.
Based on the comparison of Logistic Regression, Decision Trees, K-Nearest Neighbors (KNN), and XGBoost,
the XGBoost model demonstrated the best performance with the highest accuracy, precision, recall, F1-score, and ROC-AUC.

Therefore, XGBoost is the recommended machine learning technique for accurately predicting customer churn.

3**. Actionable insights that SyriaTel can derive from the predictive model to improve customer retention efforts.**

**Optimize Plans for High Usage**: Introduce or adjust plans to provide better value for customers with high day, evening, international, and night call minutes to reduce cost sensitivity and enhance satisfaction.

**Enhance Call Quality**: Improve network reliability and call quality during peak times to retain high-usage customers.

**Proactive Customer Support:** Monitor and support customers who frequently contact customer service, addressing their issues promptly to prevent churn.

# Findings

## Cont'

Deductions

**Loyalty Programs:** Develop loyalty programs for long-term customers (high account length) and those with high usage across different times and call types.

**Voice Mail Plan**: Review and possibly enhance the voice mail plans to add value for customers who use this service frequently.

5**. Strategies that SyriaTel can put in place to reduce churn rate**
Customized Plans: Offer tailored plans for high-usage customers, especially for those with high day, evening, and international minutes.

**Quality Assurance:** Enhance call quality and reliability during peak usage times

**Proactive Support:** Provide proactive and efficient customer service, especially for those who frequently contact support.

**Loyalty Programs:** Implement loyalty programs targeting long-term and high-usage customers to reward their commitment.

**Voice Mail Enhancements:** Upgrade voice mail plans to add more value for users.

**Usage Alerts:** Send alerts to customers approaching their plan limits to prevent bill shock and encourage plan upgrades.

**Feedback Mechanism**: Regularly gather and act on customer feedback to address pain points swiftly.

**Training and Support:** Provide ongoing training and support for customer service representatives to enhance their problem-solving skills and customer interaction quality.

# Conclusions

## Deductions

- Leveraging the XGBoost model to predict customer churn,
- SyriaTel can identify key factors such as high usage during peak times, frequent international calls, and frequent customer service interactions as primary drivers of churn at SyriaTel.
- By offering customized plans, enhancing call quality, providing proactive support, and implementing loyalty programs, SyriaTel can effectively reduce churn and improve customer retention.

# Recommendations

Deductions

- **Contract Optimization**:
- Offer incentives for customers to switch from month-to-month contracts to longer-term agreements.
- **Target New Customers**:
- Implement retention strategies for new customers to increase their tenure with SyriaTel.
- **Re-evaluate Pricing:**
- Review and adjust pricing structures for high-charge plans to ensure they deliver clear value.
- **Improve Internet Service:**
- Address issues with fiber optic service to enhance reliability and customer satisfaction.
- **Enhance Technical Support:**
- Invest in better technical support to quickly and effectively resolve customer issues.
- **Develop Loyalty Programs**:
- Create loyalty programs to reward long-term customers and reduce churn.
- **Personalize Offers:**
- Use predictive analytics to provide personalized offers and discounts to at-risk customers.
- Proactive Customer Outreach: Regularly engage with customers to identify and address potential issues before they lead to churn.



**Thank you**