Customer Churn Prediction at SyriaTel
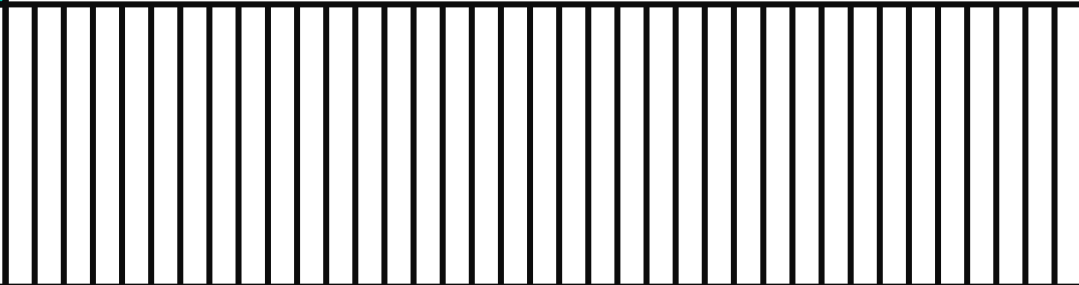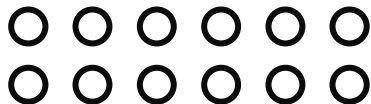
Project Overview

- This project aims to develop a classification model to predict customer churn and recommend strategies to reduce churn rates.

- Addressing customer churn helps syriaTel minimize financial losses and improve customer retention.

Business Understanding

📞 Syriatel Telecommunications is a company dedicated to connecting people through seamless communication, it is struggling to retain its customer base, as customer churn continues to rise.

📢 This churn not only affects the company's revenue but also raises the cost of acquiring new customers in order to retain market share.

🚩 Identifying the factors that contribute to customer churn and predicting which customers are at risk allows SyriaTel to implement targeted retention strategies, thereby increasing customer loyalty and decreasing turnover.

This analysis looks into the patterns and predictors of customer churn at SyriaTel
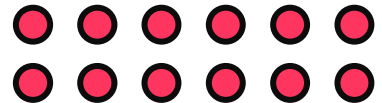
Objective & Performance Metrics

The main aim of this project is to create a model that would accurately predict whether customer would churn out or not.

As such the project will focus on optimizing the following metrics:

Recall, Precision & F1_score

AUC score from the ROC curve.

The main objective is to optimize recall.

Data Understanding

Variables

• The Dependent variable is a binary. 1: Churn. 0: Not Churn
• The independent variable are features such as talk time, call rate etc

Shape: The dataset contains 21 columns and 3333 rows. This means there are 21 different feature variables each with 3333 records.
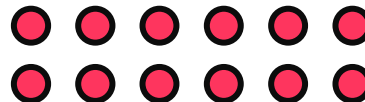
Null & Duplicate values

• There are no missing values.
• There are no duplicated rows in this dataset.

**Consumer Data from Syria Telecommunication's company**

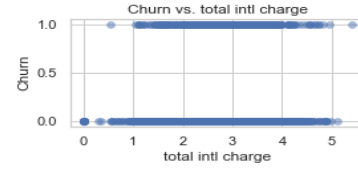# Exploratory Data Analysis

Univariate, Bivariate and Multivariate analysis were adopted in the analysis and helped in identifying significant churn predictors, such as contract type, internet service quality, and charges.

Distribution of Numerical Features
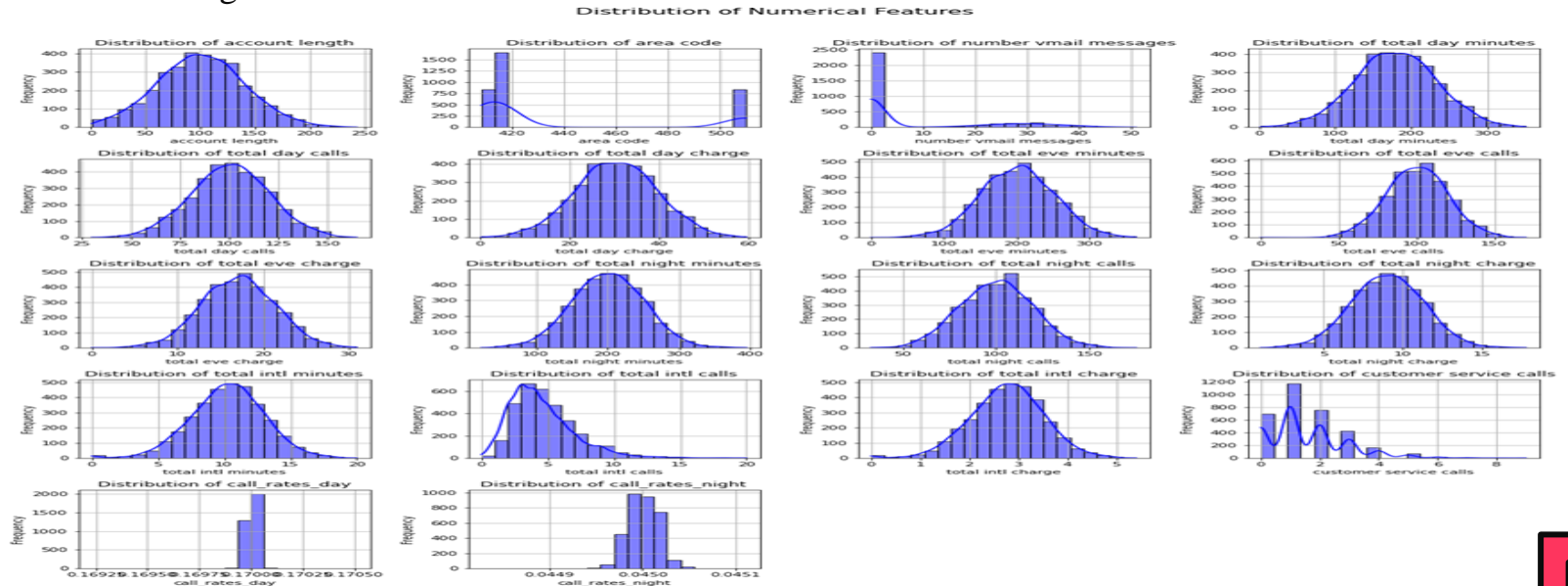
Scatter Plots of Churn vs. Numerical Variables

Correlation Heatmap
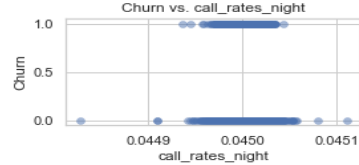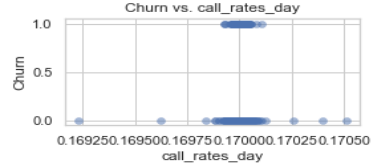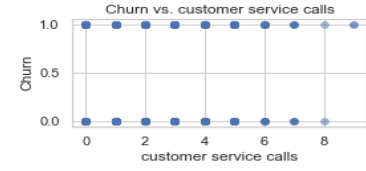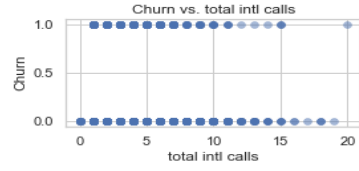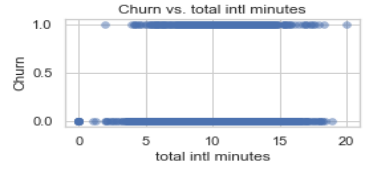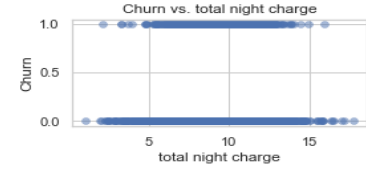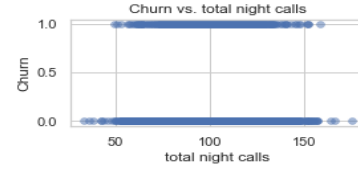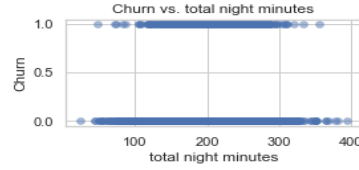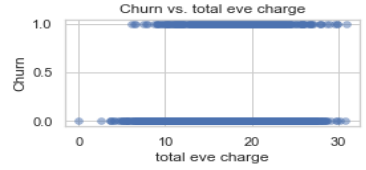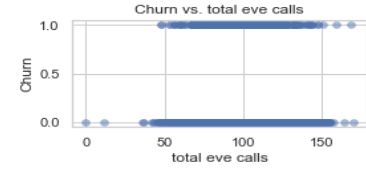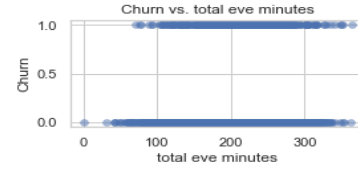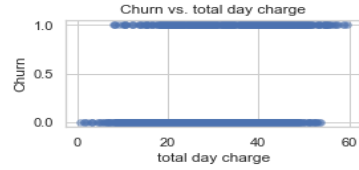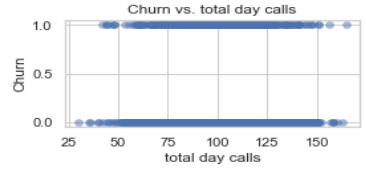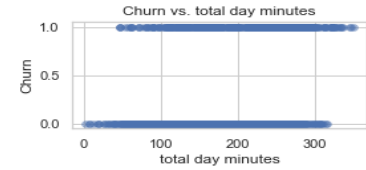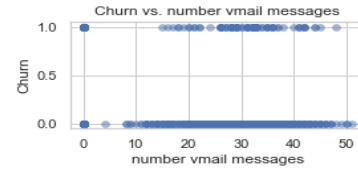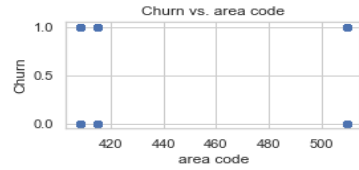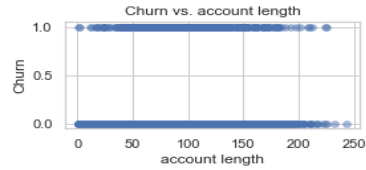
From the Heat Map as shown, we have four pairs that have a very high correlation of 0.99 and above. In our model we dropped one of the features in each of the pairs

Modeling
1. K-Nearest Neighbors (KNN)

KNN Classifier:

🎯 **Accuracy**: 0.91

🎯 **Precision**: 0.82

🎯 **Recall:** 0.32

🎯 **F1-score**: 0.46

Modeling
1. Logistic Regression

Baseline Logistic Regression Mode:

🎯 **Accuracy**: 0.88

🎯 **Precision**: 0.58

🎯 **Recall**: 0.22

🎯 **F1-score**: 0.32

Modeling
1. Decision Trees

Decision Tree Classifier:

🎯 **Baseline_dt_Accuracy**: 0.9175412293853074

🎯 **Baseline_dt_Precision**: 0.6744186046511628

🎯 **Baseline_dt_Recall:** 0.6823529411764706

🎯 **Baseline_dt_F1-score**: 0.6783625730994152

Hyper Parameter Tuned Models

KNN Classifier

Accuracy: 0.85

Precision: 0.43

Recall: 0.6

F1-score: 0.49

Logistic Regression Model

Accuracy: 0.79

Precision: 0.35

Recall: 0.764

F1-score: 0.48

Decision Tree Classifier

Accuracy: 0.93
Precision: 0.67
Recall: 0.8
F1-score: 0.73-
Criterion = entropy
max_depth =18
Min_sample_split =2
max_features=15
min_samples_leaf =3

Model Evaluation

XGBOOST CLASSIFIER

🎯 Accuracy = 0.97

🎯 Precision = 0.90

🎯 Recall = 0.82

🎯 F1_score = 0..86

Tuned Boost Classifier achieved a recall score of 0.8235, representing 82.35% of actual churned customers effectively meeting the project's objective with a precision score of 90%

Recommendations
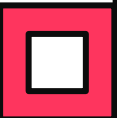

◎ **Customer Retention**: Targeted strategies for at-risk customers, especially those with month-to-month contracts.

◎ **Pricing Optimization**: Competitive and usage-based pricing plans.

◎ **Service Quality**: Focus on improving internet service reliability.

◎ **Customer Support**: Enhance support quality to reduce service calls and improve satisfaction.

Recommendations

- Develop customer retention strategies as recommended

- Segment customers according to their unique characteristics.

- Improve service quality to minimize rate of churn

- Establish a customer Feedback loop for sentiment analysis

- Regularly assess the effectiveness of implemented strategies and initiatives by tracking key metrics and comparing them against established benchmarks.