# SYRIATEl CUSTOMER ANALYSIS

EDWIN RURUNG'A

# CONTENT

1. Overview
2. Business Understanding
3. Data Analysis
4. Modelling
5. Conclusion

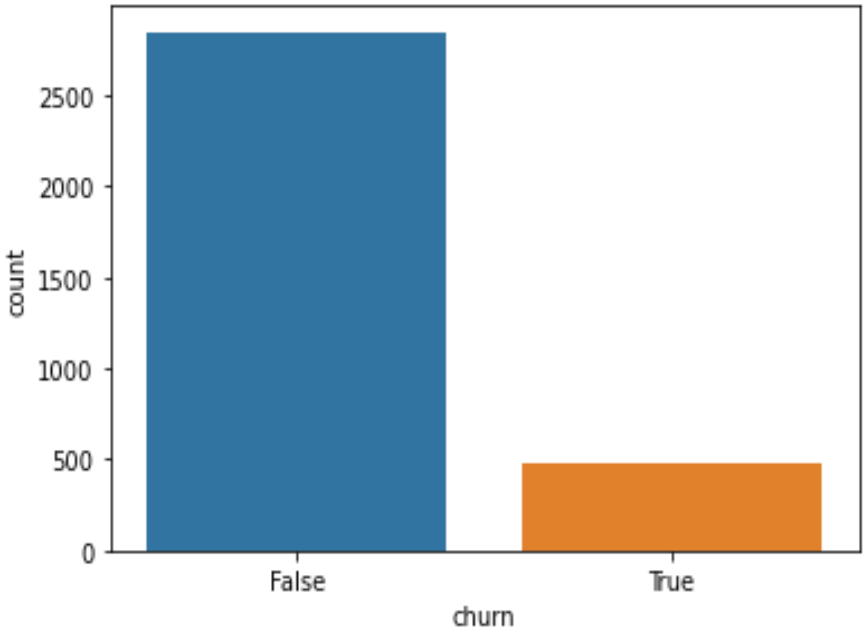
# Overview

- SyriaTel is a telecommunications company in Syria. They have been informed that some of their customers have started to churn, discontinue their service.
- This analysis will determine what features will indicate if a customer will ("soon") discontinue their service
- Concerns regarding the customer turnover rate have arisen due to increased competition in the dynamic telecom market, where customers have the opportunity to switch between multiple providers.

# Business Understanding

- The nature of this issue is framed by the reality that it is less expensive to lose current consumers than it is to acquire new ones. Therefore, it is far preferable to keep your current clients than to find new ones.
- To effectively retain consumers in the face of increased competition in the telecom sector, client turnover must be predicted. SyriaTel is aware of how crucial client retention is to cost effectiveness and competitive parity.
- SyriaTel can foresee customer behavior and take preventative action to lower churn rates thanks to data analysis and predictive analytics.

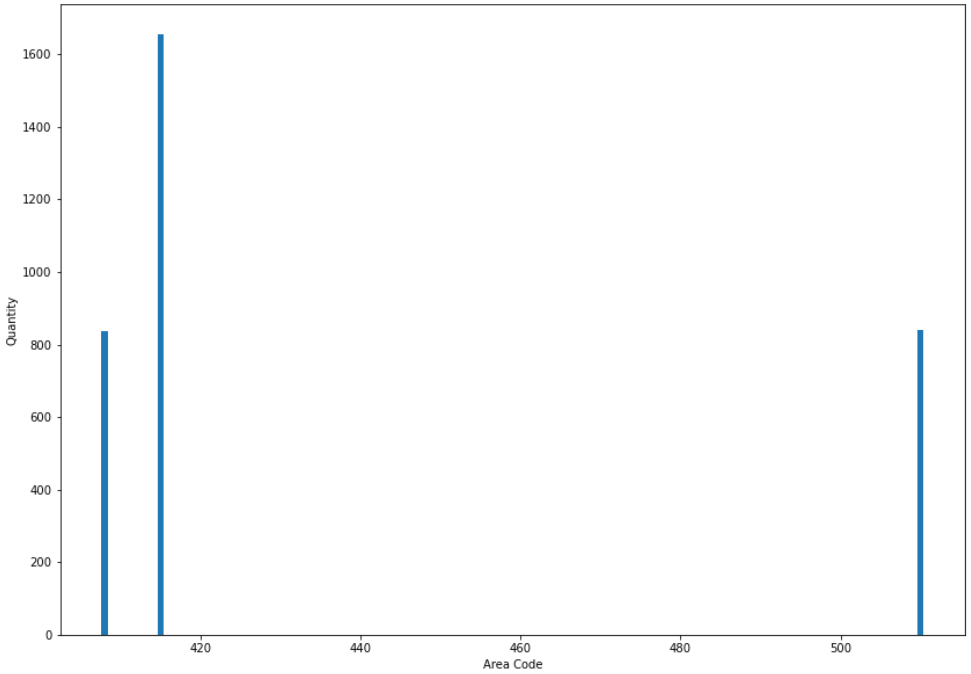
# Data Analysis




Of the 3,333 customers in the dataset, 483 have terminated their contract with SyriaTel. That is 14.5% of customers lost.

The distribution of the binary classes shows a data imbalance. This needs to be addressed before modeling as an unbalanced feature can cause the model to make false predictions

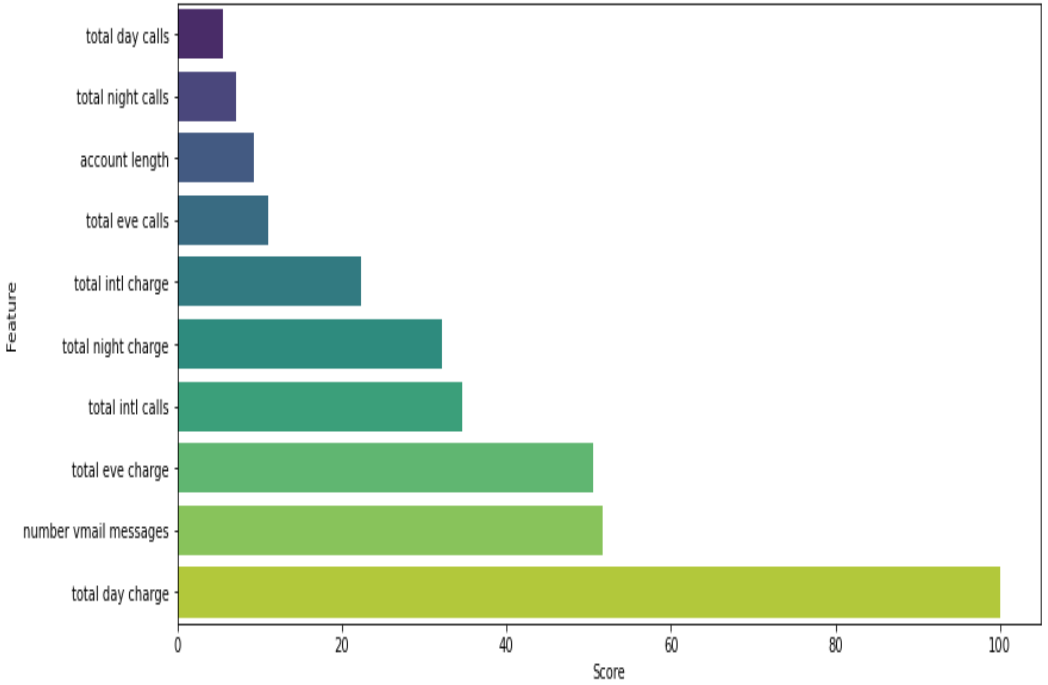
# Area Code Analysis




Half of the customers have the area code 415.

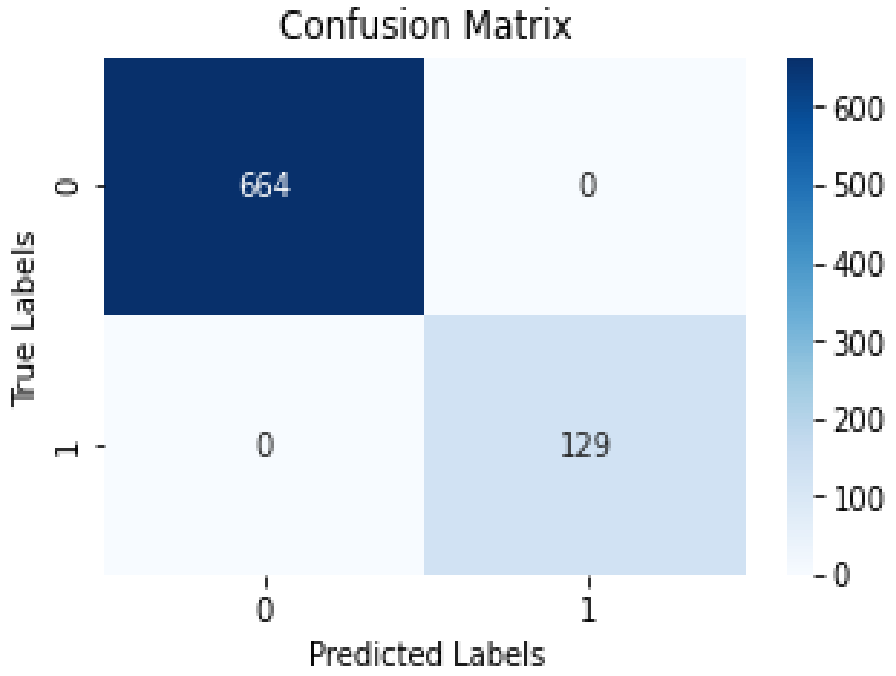
One fourth of customers have the area code 510 and another fourth have the area code 408.

# Modelling




- The total day charges has the most impact in terms of churn when it comes to the logistic regression model
- Whereas the total day call has the least impact

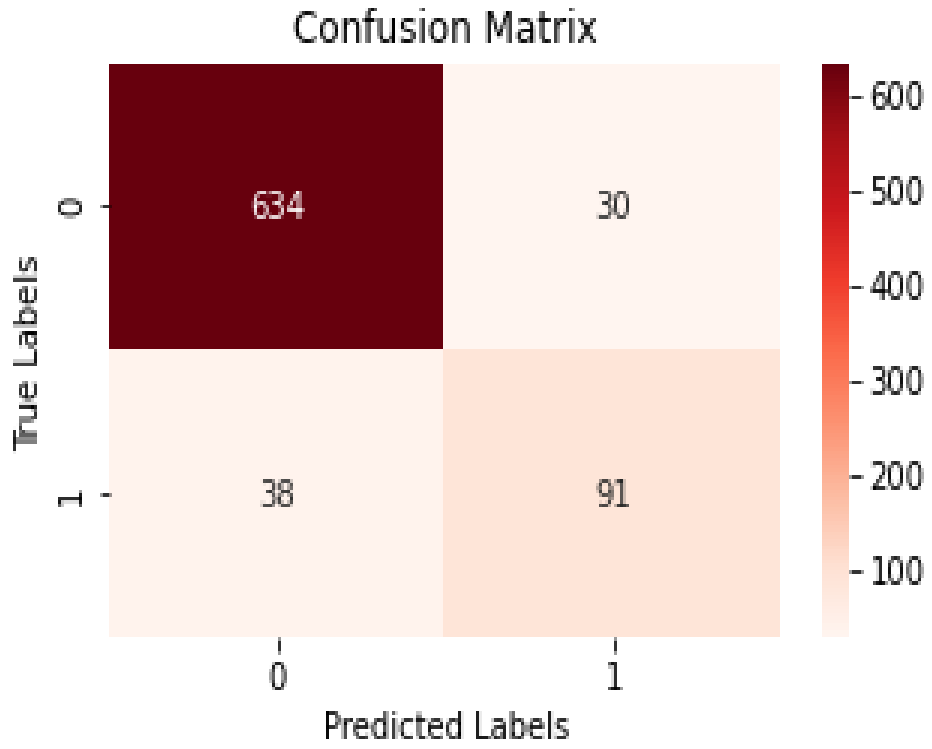
# Logistic Regression




According to the logistic regression classifier model, total day charge, number of voicemail messages and total evening charge are the top three important features.

Model accuracy is 76.5%, which isn't bad. F1 score is only 50.2% which means the test will only be accurate half the times it is ran

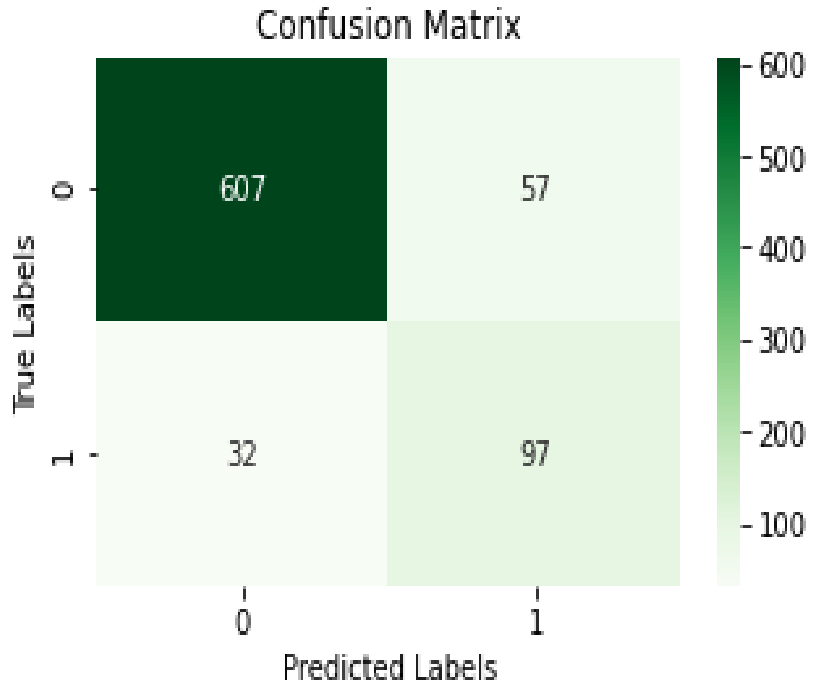
# Random Forest Classifier




According to the random forest classifier, total day charge, customer service calles and "international plan is yes" features have the highest impact on the model.

Accuracy and F1 score are much higher for this model, which is good news.

# Decision Tree Classifier




According to the decision tree classifier, customer service calls total day charge and total evening charge are the three most important for the model.

The accuracy and F1 score for this model is not as great as model 2.

# Recommendations

Explore alternative algorithms: Consider trying different machine learning algorithms to improve churn prediction accuracy.

Gather more relevant data: Collect additional data to enhance the accuracy of the churn prediction models. Continuously evaluate and refine models

Regularly assess model performance, fine-tune parameters, and retrain models to ensure accuracy and effectiveness in predicting customer churn which will help in ensuring the model is able to track real time data and provide accurate information.