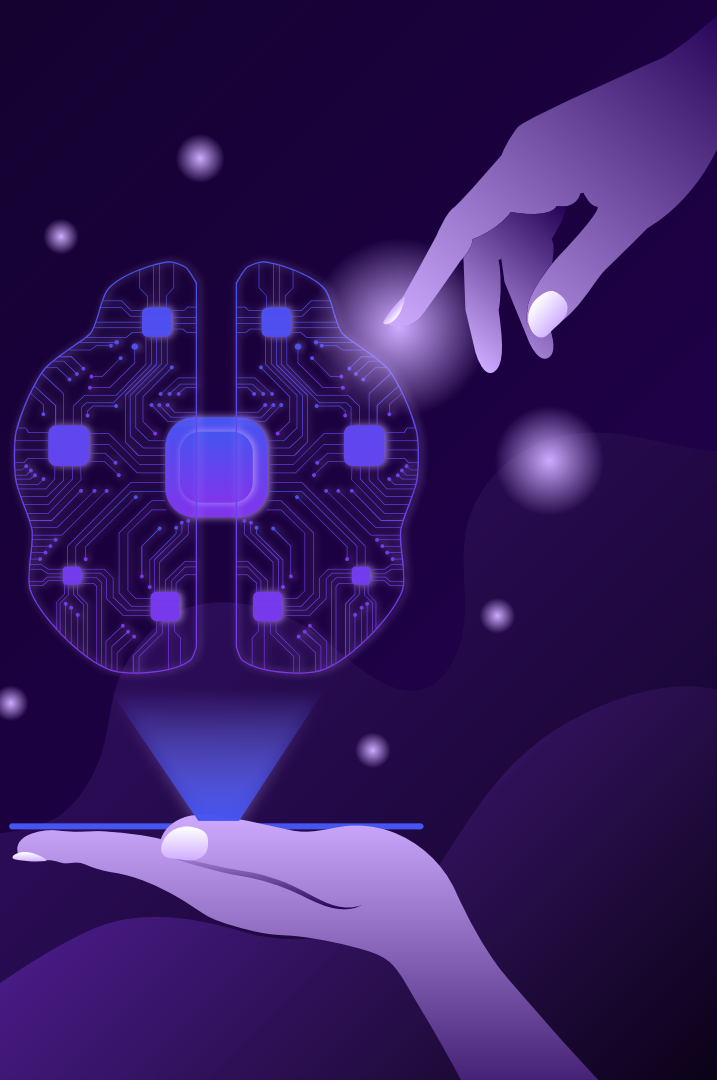
SyriaTel Customer Churn Prediction

By Muthoni Kahuko



Project Overview

SyriaTel, a telecommunications company, is facing a high churn rate as many customers are stopping to do business with them. The company wants to address this issue and figure out how to reduce money lost because of customers who don't stick around very long.

The company has requested the data analysts to develop a model that can predict the customer churning. SyriaTel aims to gain insights into factors associated with churn in order to reduce the rate, increase customer retention and overall growing profit.

Churn - the rate at which customers stop doing business with a company over a given period of time.

Project Workflow

01 Business Understanding

02 Data Understanding

03 Data Preparation

04 Modeling

05 Model Evaluation

06 Recommendations

Business Problem

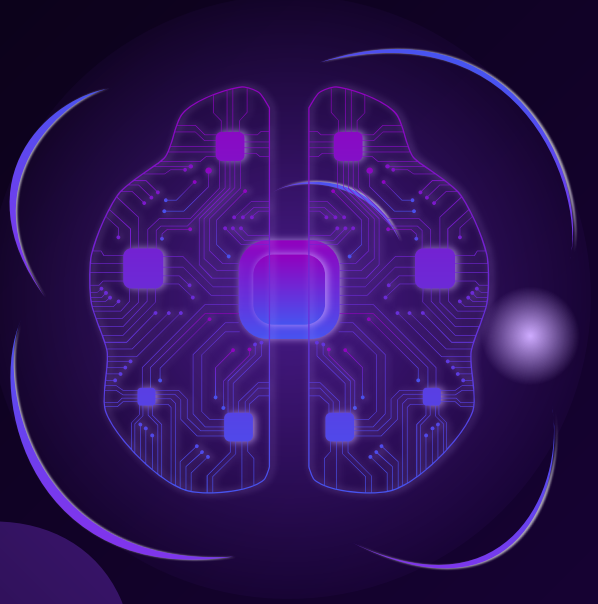
Syriatel, a telecommunications company, is facing a high churn rate with many of their customers discontinuing their subscription services and switching to different telecommunication companies.

Project Objectives

- Identify the factors that lead to customer churn
- Develop a model that can accurately predict which customers will churn
- Take proactive steps to reduce customer churning

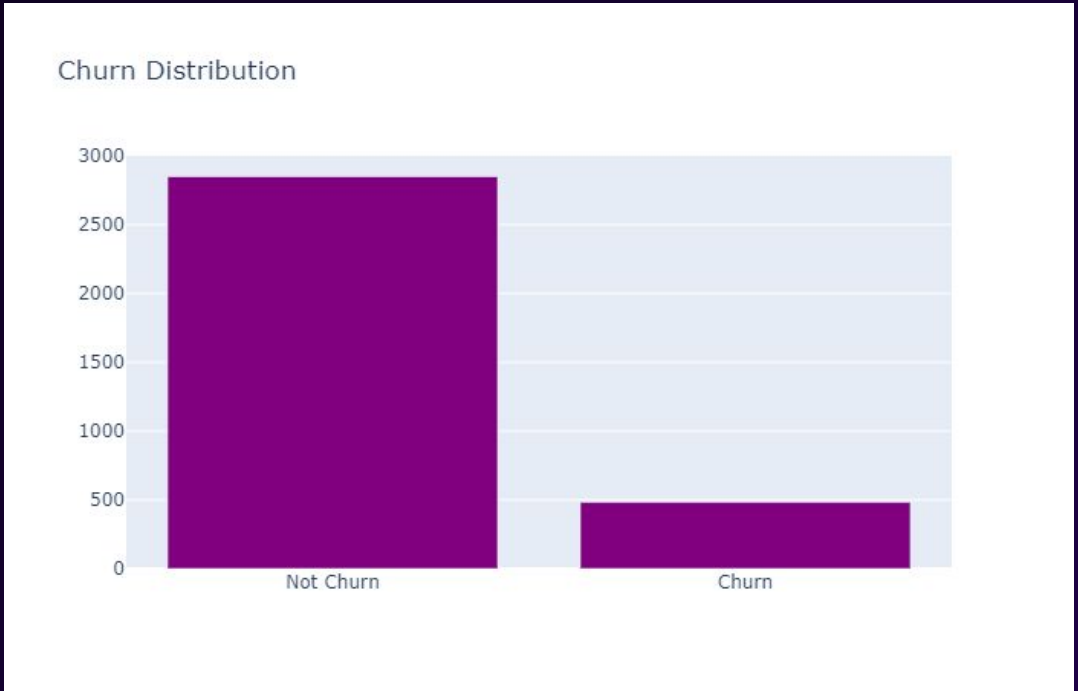
Data Understanding

- We will be working with a dataset from Kaggle website
- It contains 3,333 records and 21 columns of data



Exploratory Data Analysis

Number of customers that churned?



Number of customers;

3, 333

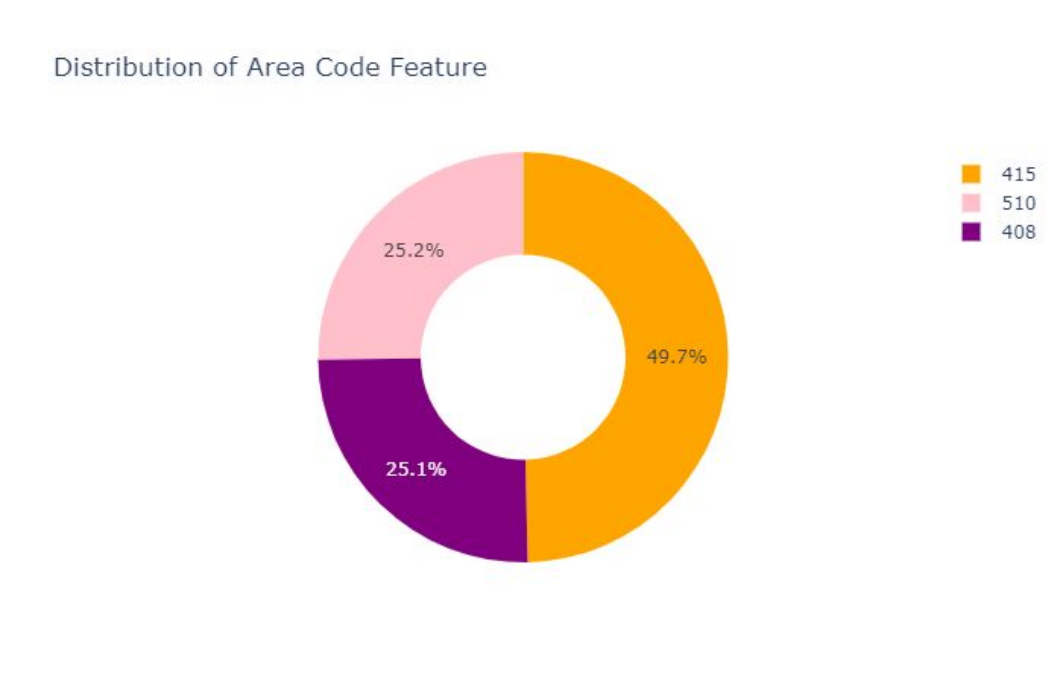
Number of customers that churned;

483

Number of customers that churned;

14.5%

Distribution of Population

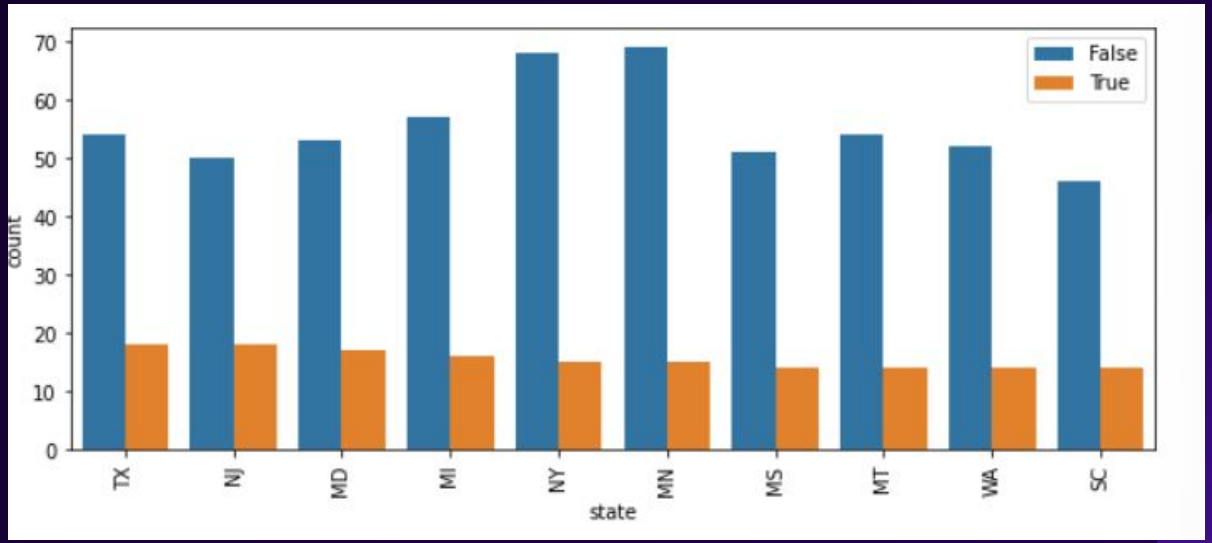


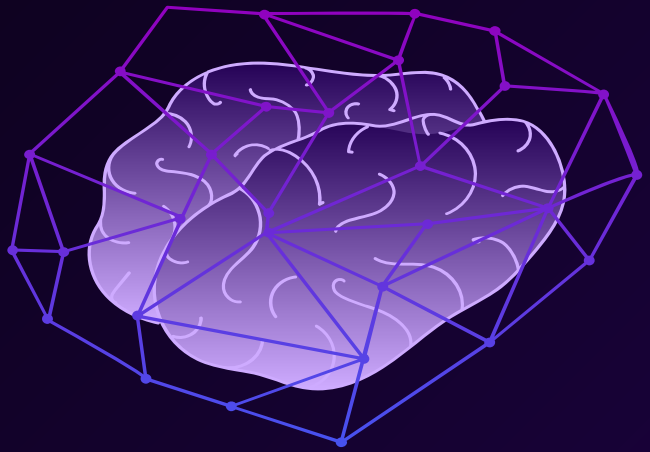
The area code with the highest population is 415 which is 49.7%

Both 510 and 409 have relatively the same size of population with 25.2% and 25.1% respectively.

Distribution of population based on churn rate

Texas, New Jersey are leading with the highest churn rate





Modeling

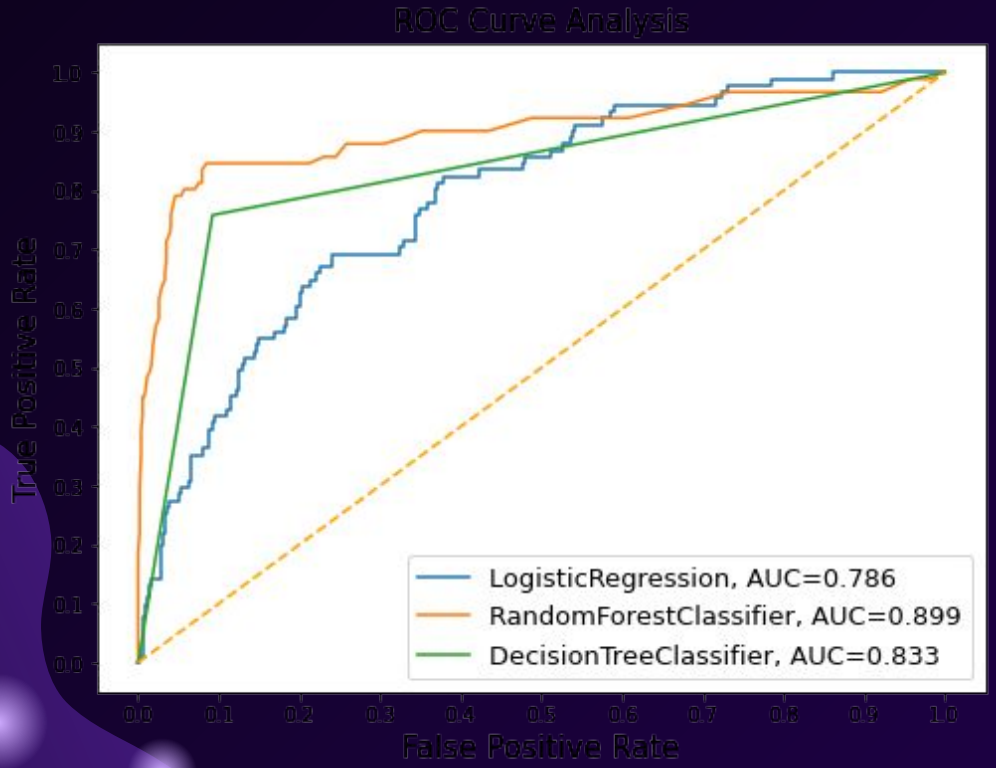
Algorithms used

01 Logistic Regression

02 Decision Tree

03 Random Tree

Model Evaluation using ROC_AUC curve



Results based on ROC_AUC curve:

Random Forest - 0.899

Decision Tree - 0.833

Logistic Regression - 0.786

Model Evaluation using Recall score

classifiers	recall
LogisticRegression	0.659341
RandomForestClassifier	0.791209
DecisionTreeClassifier	0.758242

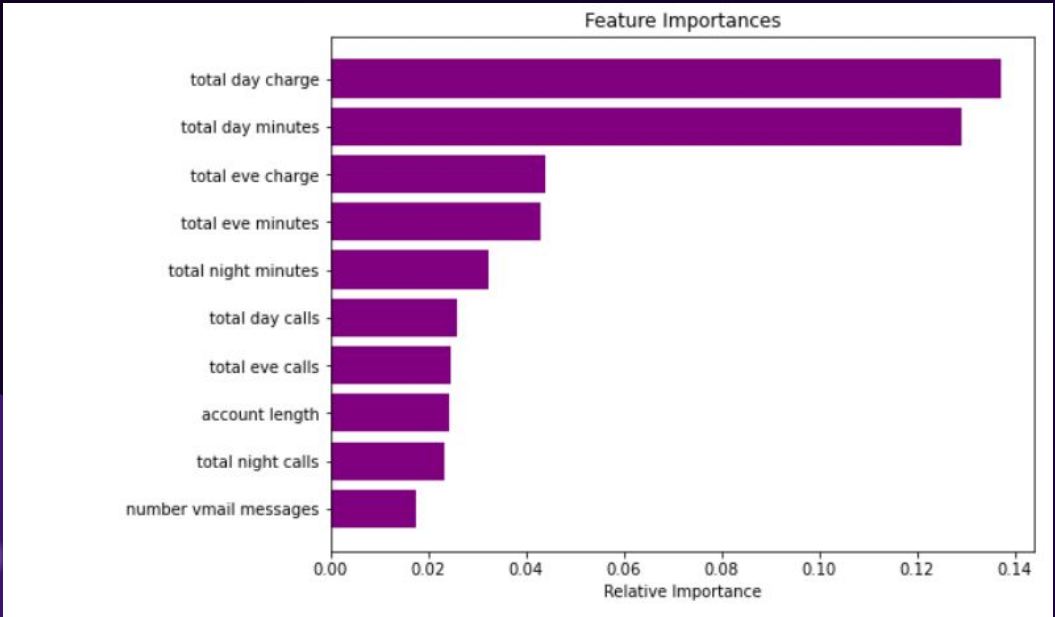
Results based on ROC_AUC curve:

Random Forest - 0.791

Decision Tree - 0.758

Logistic Regression - 0.659

Model Evaluation using Recall score



The most important features as recommended by the model are:

total day charge

total day minutes

total eve charge

Conclusion

The recall score of Random Tree classifier was 0.79. WHile this s a good predictive model we can undertake further engineering to boost the score.

The most important features as recommended by the model are: total day charge, total day minutes and total eve charge

We managed to create a successful predictive model that is able to predict the customer churn rate

Recommendations

Offer discounts and offers to customers in area codes 415 and 510 as these areas have a high churn rate. This can reduce the rate at which customers leave the company

Evaluate the pricing structure for day, evening, and international charges. COnsider adjusting pricing plans, introduce special offers and discounts to address the customers who leave due to higher chargers

Focus on customer retention strategies for states with higher churn rates like Texas, New York and New Jersey. This can involve targeted marketing, discounted rates, and better customer support in those states

Improve customer service quality and reduce the number of customer service calls. Enhance training programs for customer service representatives to ensure prompt and effective resolution of customer issues, leading to higher customer satisfaction and reduced churn.

Thank you

For more information,kindly contact
muthonifk@gmail.com