Customer Churn for SyriaTel

Telecommunications Limited

OVERVIEW

OVERVIEW

BUSINESS AND DATA UNDERSTANDING

MODELLING

EVALUATION

RECOMMENDATIONS AND NEXT STEPS

OVERVIEW

Syriatel is a provider of mobile telecommunication and data services based in Damascus, Syria.

The company offers services including calls, news, messages, global systems for mobile communications (GSM) and internet services, thereby making the lives of customers easier with reasonable prices.

It is one of the only two providers in Syria, the other being MTN Syria.

In 2022 the Syrian telecommunications authority awarded the third telecom license to Wafa Telecom and is keen to use data to ensure it retains its clients.

BUSINESS PROBLEM

SyriaTel would like to retain their market share in the Telecommunications Industry by ensuring high levels of customer satisfaction. By predicting churn (i.e loss of customers),SyriaTel will be able to predict factors that influence customer loss so as to put in place measures for retaining clients and position as market leaders in the Telecommunication Industry.

The key objectives are to determine:

1.Develop a predictive model to identify customers who are at risk of churning from SyriaTel's services for early intervention and improve on customer sartisfaction.

2.Identify key features that predict Churn

3.Strategic Decision-Making:Provide management with insights for decision-making on resource allocation.

DATA UNDERSTANDING

This project uses data from SyriaTel Telecommunications Company for 3333 users across 51 states on calls made at different times of the day , the duration of the calls, the charges as well as their international and voice plans.

Columns Descriptions:

Churn: Indicates if the customer has stopped doing business with SyriaTel. (False = No churn, True = Churned)

State: The U.S. State of the customer. (Requires one-hot encoding; not ordinal)

Account Length: A smaller number signifies an older account. (Indicative of Customer Lifetime Value)

Area Code: Area code of the customer's phone number.

Phone Number: The customer's phone number.

International Plan: Whether the customer has an international plan. ('yes' or 'no'; binary and thus effectively one-hot encoded)

Voice Mail Plan: Whether the customer subscribes to a voice mail plan. ('yes' or 'no'; binary and thus effectively one-hot encoded)

Number of Voice Mail Messages: Total number of voice mail messages left by the customer.

Total Day Minutes: Aggregate of daytime minutes used.

Total Day Calls: Total number of calls made during the day.

Total Day Charge: Total charges incurred for daytime calls.

Total Eve Minutes: Total minutes spent on calls in the evening.

Total Eve Calls: Number of calls made during the evening.

Total Eve Charge: Charges for evening calls.

Total Night Minutes: Total minutes for nighttime calls.

Total Night Calls: Number of calls made at night.

Total Night Charge: Nighttime call charges.

Total Intl Minutes: Cumulative international minutes (covering day, evening, and night).

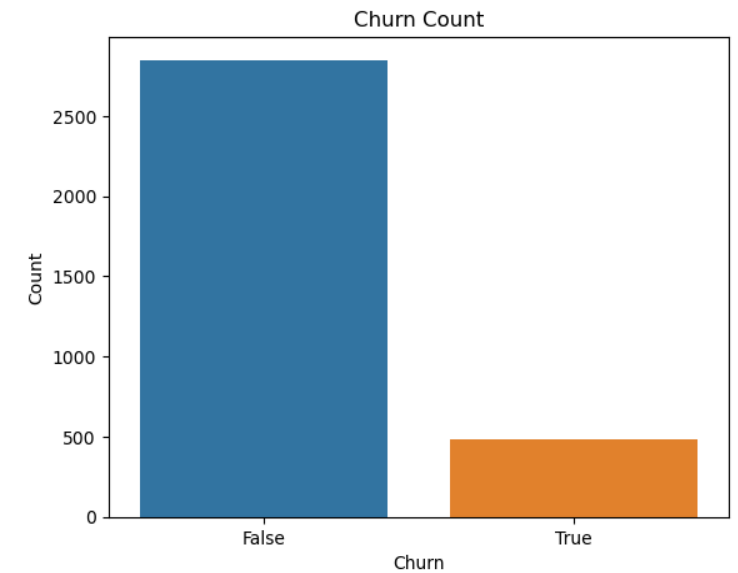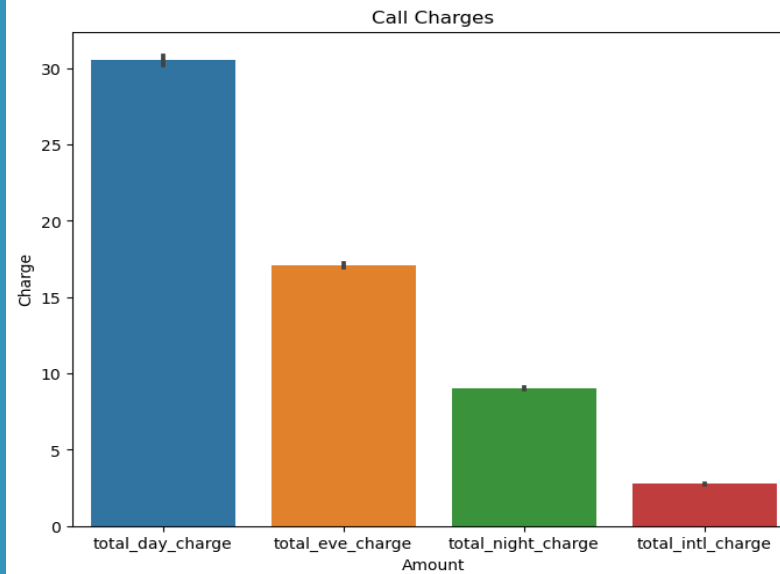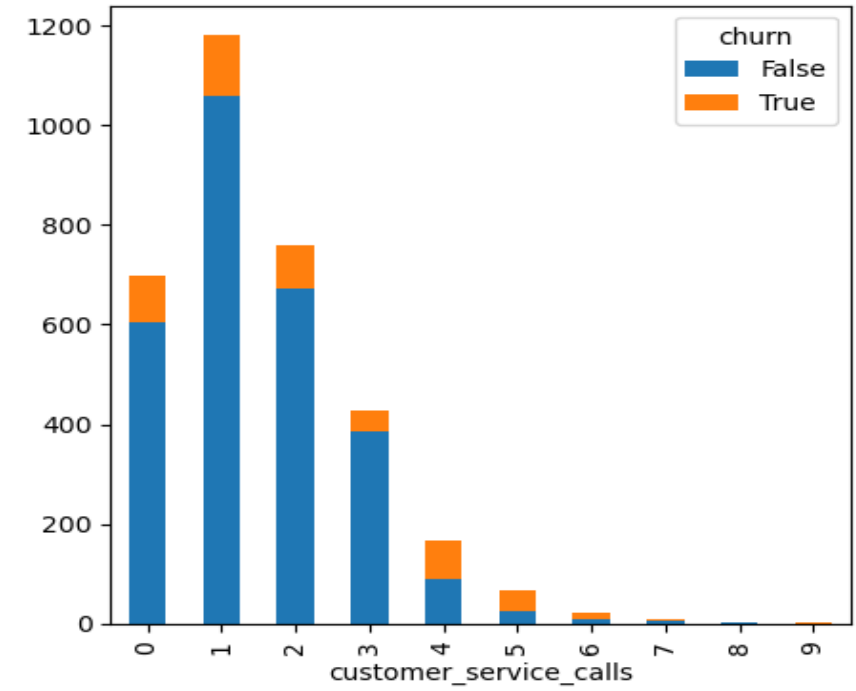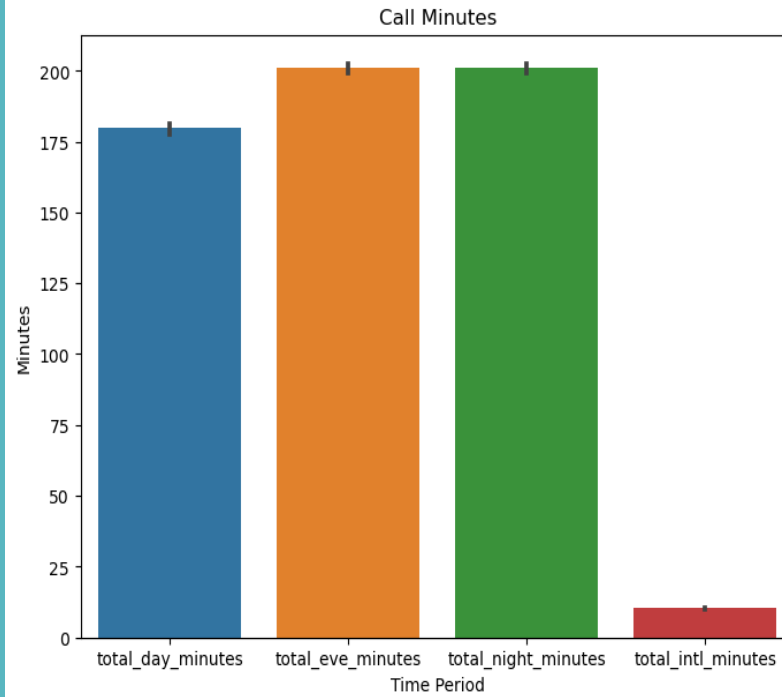Total Intl Calls: Total number of international calls (across all time periods).

Total Intl Charge: Total charges for international calls.

Customer Service Calls: Number of calls made to customer service by the customer.

Target Variable Desription:

Churn: if the customer has churned (true or false)

Exploratory Data Analysis

MODELLING

Methodology

Splitting Data into Training and Validation Sets In machine learning, splitting the dataset into training and validation sets is crucial for model development and evaluation.

Model Training:

The training set is used to train the machine learning model by fitting it to the training data. This involves learning the underlying patterns and relationships between features and target.

Model Evaluation:

The validation set is used to evaluate the performance of the trained model. By assessing the model's performance on unseen data (validation set), we can estimate how well the model generalizes to new, unseen examples.

Preventing Overfitting:

Splitting the data helps in detecting and preventing overfitting, where the model learns to memorize the training data rather than capturing underlying patterns. By evaluating the model on a separate validation set, we can detect overfitting and fine-tune model hyperparameters to improve generalization. Hyperparameter Tuning:

The validation set is also used for hyperparameter tuning, such as optimizing regularization parameters or adjusting the model complexity. This process helps improve the model's performance and generalization ability. Assessing Performance:

Splitting the data allows us to assess the model's performance metrics, such as accuracy, precision, recall, or F1-score, on the validation set. These metrics provide insights into how well the model performs on different aspects of classification or regression tasks.

By splitting the dataset into training and validation sets, we ensure that the machine learning model is trained effectively, evaluated accurately, and optimized for better performance and generalization on unseen data.

Model Evaluation

Our best performing model is gradient boost which has eliminated any false negatives and false positives with an accuracy level of 85%. The machine learning model, evaluated through key metrics such as accuracy, precision, recall, and F1 score, consistently demonstrates superior predictive capabilities.

It has also resulted in eliminating overfitting as well as any false negatives

| Model	Accuracy	Classification Report:			
Logistic Regression	86%	precision	recall	f1-score	support
		0 0.87	0.98	0.92	709
		1 0.57	0.17	0.26	125
		accuracy		0.86	834
		macro avg	0.72	0.57 0.59	834
		weighted avg	0.82	0.86 0.82	834
Cross Validated Logistic Regression	86%				
Decion Tree Classifier	0.936450839				
Mean Cross Validation Score for Random Forest Classifier:	91.61%				
		Classification Report:			
Gradient Boost	91%	precision	recall	f1-score	support
		0 0.97	0.93	0.95	709
		1 0.68	0.82	0.74	125
		accuracy		0.91	834
		macro avg	0.82	0.87 0.84	834
		weighted avg	0.92	0.91 0.92	834

Recommendations

Top three Features that can be used to predict churn are:

- **Total day Minutes**
- **Total day Charges**
- **Customer Service calls**

Top 10 Feature Importances
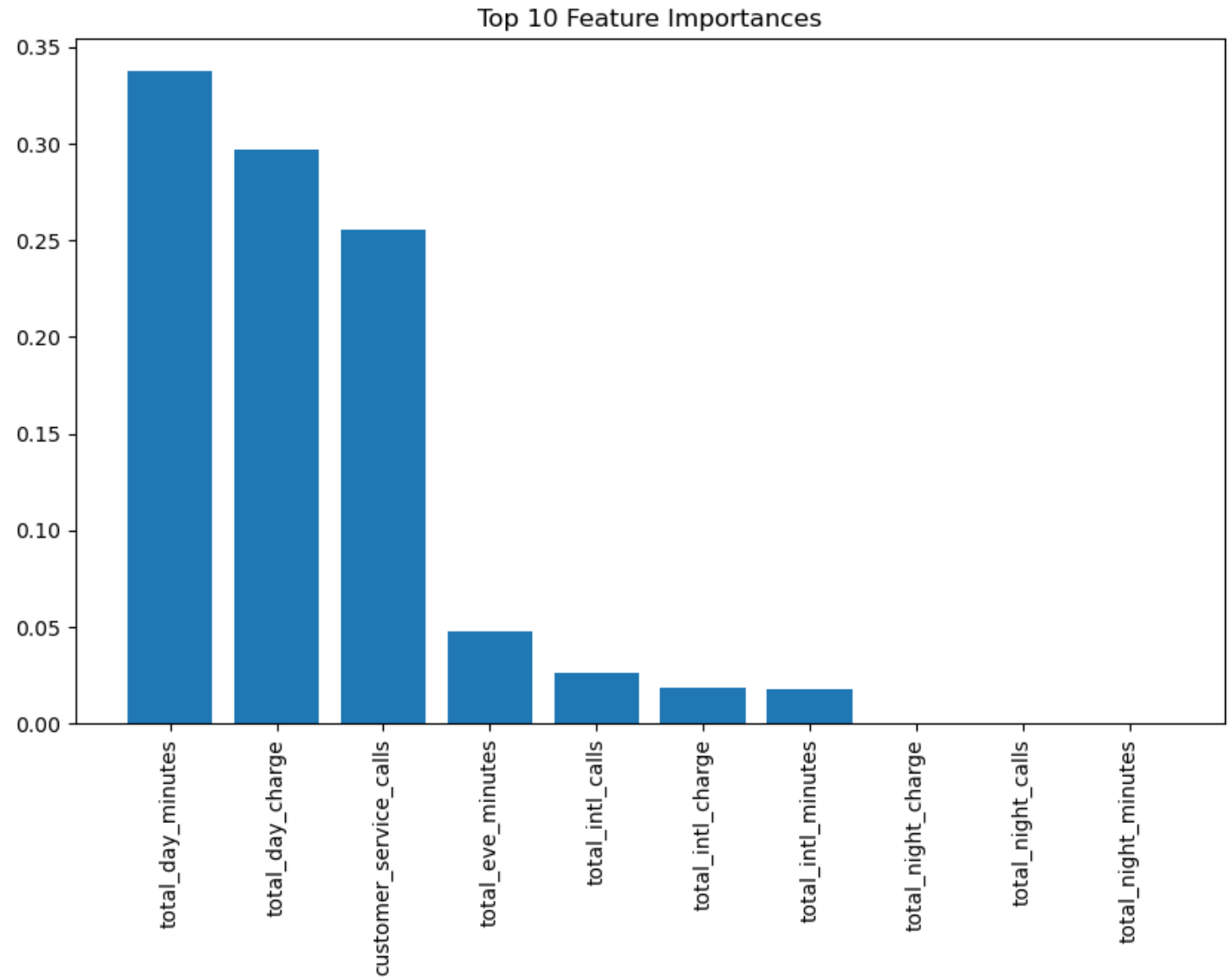
Next Steps

Investigate ways to incentivize customers who have high call minutes and with total day charges

Further investigation should be devoted to looking into the other characteristics of these customers to find out why there was a need to make this many calls to customer service and how the company could better assist these customers.

Thank You!