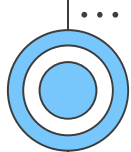


Customer churn prediction analysis.

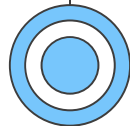
...





Business understanding

Syria Tel is facing the challenge of customer churn. Minimizing churn is crucial for sustained profitability. Acquiring new customers is costly, and retaining existing ones is a key strategy to maintain market share and revenue stability. Therefore, this analysis plays a vital role in predicting this thus improving on customer service and make more informed decisions.



Data understanding

The dataset used for this analysis comprises historical customer data, including features such as call duration, call charges, account length, customer service interactions, and more. The target variable is binary, indicating whether a customer has churned or not.



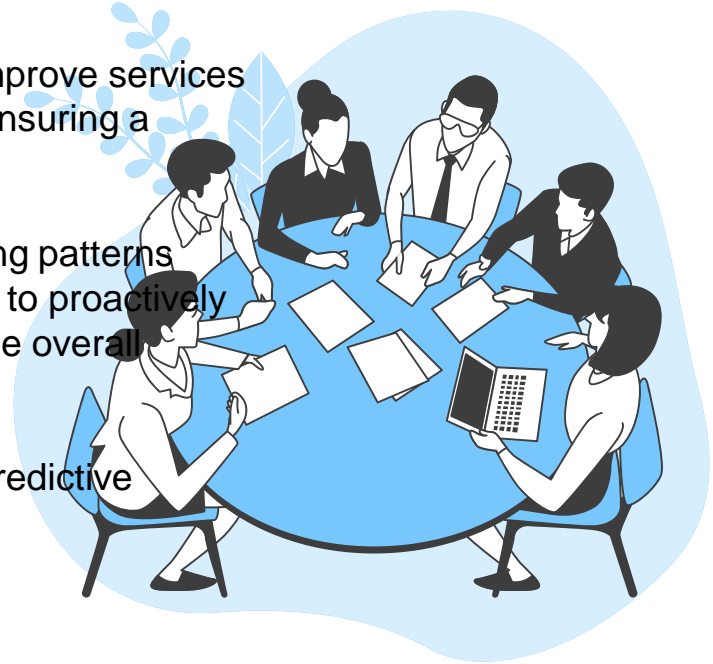
Target audience

a) Syria Tel company at large :This mainly involves the company's executives, shareholders and investors.

b)Customers who are indirectly impacted. The goal is to improve services and customer satisfaction, potentially reducing churn and ensuring a positive experience for existing customers.

c)Customer Service Teams who benefit from understanding patterns indicative of potential churn. This information enables them to proactively engage with customers, address concerns, and enhance the overall customer experience

d)Anyone interested in gaining insights into this Syria Tel predictive customer churn analysis.



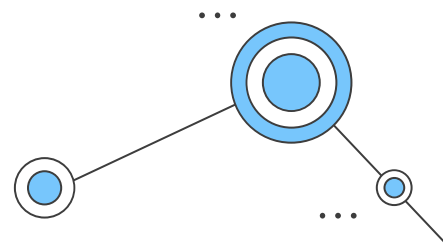
Objectives

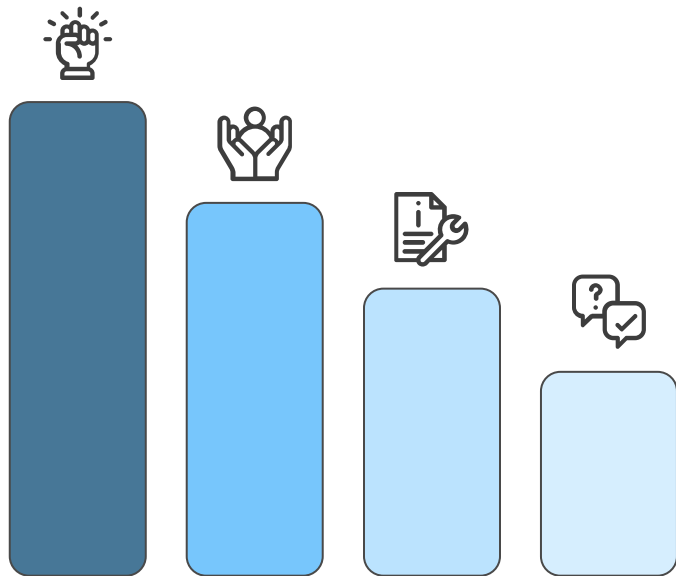
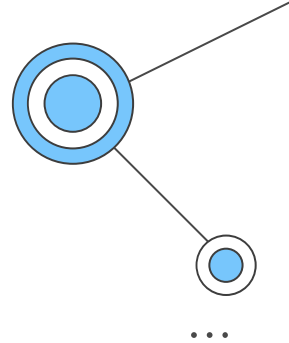
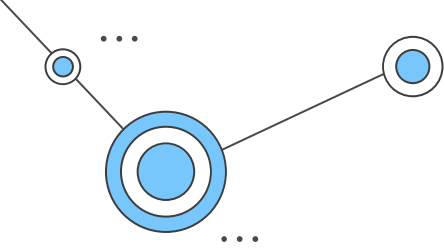


1. Develop a binary classification model that accurately predicts whether a SyriaTel customer is likely to churn in the near future.
2. Explore and utilize historical customer data to identify patterns and behaviors indicative of potential churn.
3. Evaluate and compare the performance of various classification algorithms to select the most effective model for SyriaTel's specific context.
4. Provide actionable insights to SyriaTel's business operations by interpreting the results and highlighting key factors contributing to customer churn

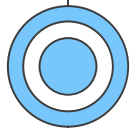
Problem Statement

In the dynamic landscape of the telecommunications industry, SyriaTel faces the challenge of customer churn, which directly impacts its revenue and market share. The company recognizes the need for a proactive approach to identify and retain customers who are likely to discontinue their services in the near future. The objective is to develop a robust predictive model that accurately anticipates customer churn based on historical data and behavioral patterns.





DATA ANALYSIS

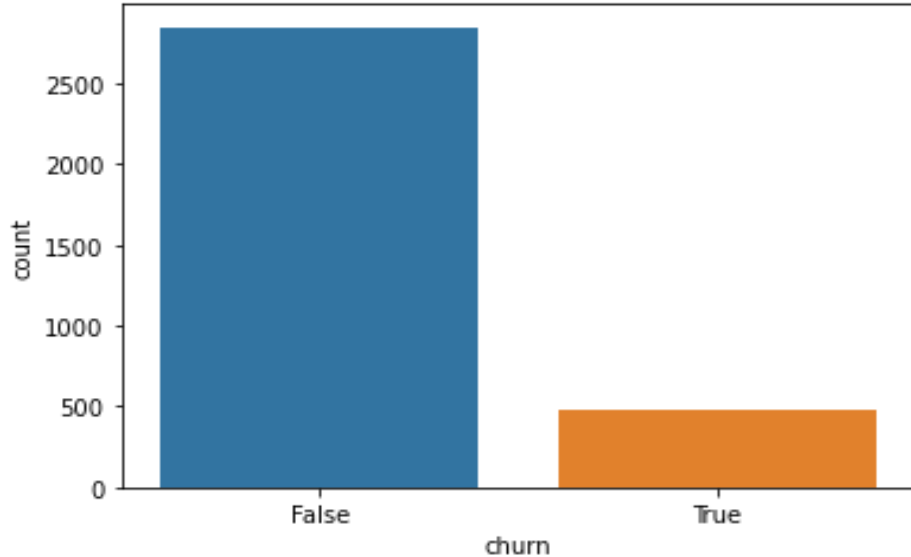


Explanatory data analysis.

The EDA in this project involved checking dataset shape and info, exploring descriptive statistics ,visualizing churn distribution .Categorical data was encoded and dropped just one column which is the phone contact number of customers since it was not helpful in this case.



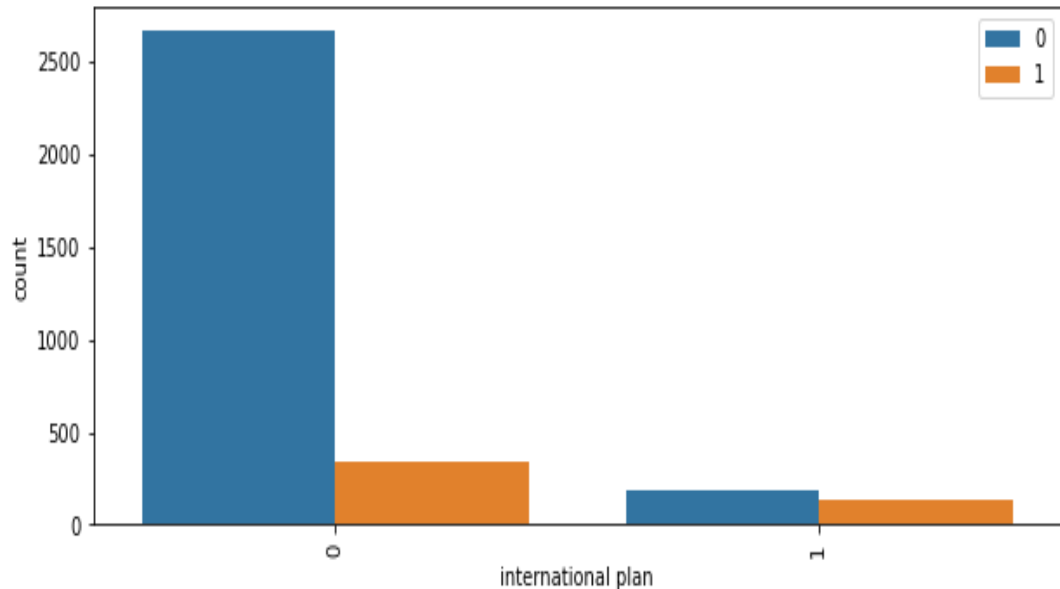
Analysis of rate of customer churning.



Where True represents the customer churning and False represents the customer not churning.

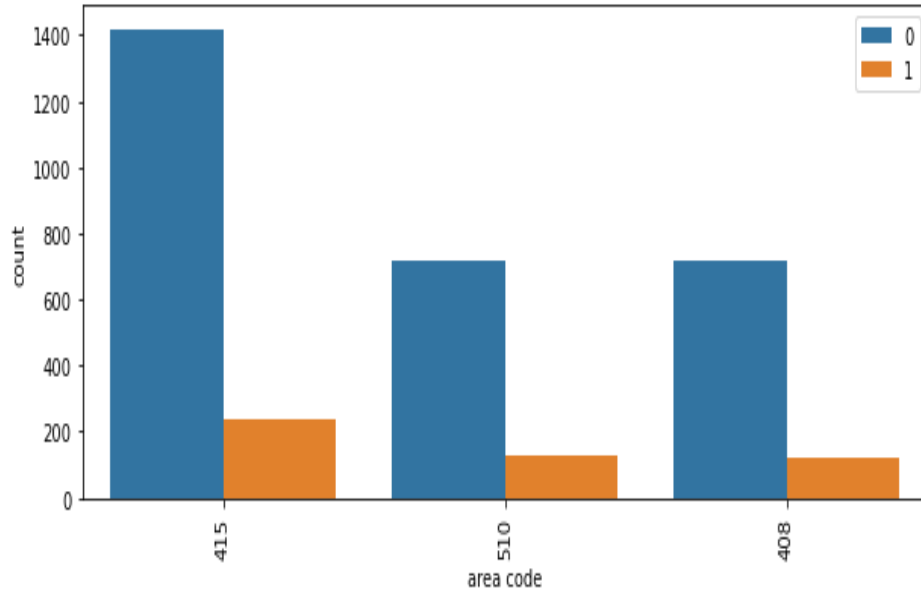
However, this shows that the dataset is imbalanced because the instances in customer not churning (False) is way more than the instances in customer churning .

Churn distribution based on International plan



Customers who have international plan are less likely to churn compared to customers who don't have international plan.

Churn distribution based on area code



There is a high probability that a customer who is going to churn is from 415 area code



Model performances.

LogisticRegression

- The accuracy score of 85.16% in the logistic regression model represents the proportion of correctly classified instances among the total instances

Random Forest

- After tuning the hyperparameters in this model in order to optimize it, it increased its accuracy score to 96% , the model is accurate in predicting the correct class for about 96.10% of the instances.

XGBOOST

- An accuracy score of 95% is pretty good but it did not perform better compared our Random Forest Classifier that used tuned hyper parameters.



Results

Random Forest Classifiers with tuned hyperparameters has proven to be the best model in predicting whether a customer will churn or not with the highest accuracy score of 96.1% compared to logistic regression model and XGBOOST from ensemble methods. This results collectively demonstrate the Random Forest model's stronger performance in predicting the outcome of whether a customer will churn or not.



NEXT STEPS

- 1. Feature Engineering and Selection:** Move on to enhancing the dataset by creating meaningful features that provide valuable insights into customer behavior. Simultaneously, carefully select features that exhibit the most relevance to the prediction of customer churn.
- 2. Model Development and Evaluation:** Advance to the core phase of building predictive models. Experiment with various classification algorithms, such as Logistic Regression, Decision Trees, Random Forest, SVM, and Gradient Boosting, to find the most effective model. Evaluate these models rigorously using performance metrics like accuracy, precision, recall, F1 score, and ROC-AUC.
- 3. Integration and Deployment Considerations:** Transition to discussions around integrating the predictive model into SyriaTel's operational framework. Explore deployment options, ensuring that the model seamlessly integrates into existing systems for real-time prediction. Consider scalability and continuous monitoring aspects for sustained effectiveness.
- 4. Documentation and Communication:** Conclude by documenting the entire process in a comprehensive README file. This document will serve as a reference guide, encapsulating the project's objectives.