SyriaTel Customer Churn Prediction

The core objective of this study is to predict churn in advance and pinpoint the primary factors that may influence customers to migrate to other telecom providers. The major questions this project seeks to answer include:

- What are the factors that are contributing to customer churning?
- What attributes do the customers who churn have?
- How can SyriaTel increase customer retention?

Project Overview

- 1 Business Problem
- 2 Data Loading and Understanding
- 3 Data Preparation
- 4 Modeling
- 5 Evaluation
- 6 Model of Choice: Decision Trees
- Recommendations and Future Investigations
- Conclusion

Project objective

The main objective for this project is to build predictive model with 85% accuracy that will help to:

* 1.To understand the factors that contribute to customer churn.'

* 2.To understand the attributes of the customer who churn

* 3.To enhance overall customer retention.

1. Business Problem

- This project aims to create a predictive model for SyriaTel, enabling the identification of customers prone to churning. By analyzing diverse customer data encompassing demographics, usage patterns, complaints, and billing records, we seek to unveil correlations and factors influencing churn.

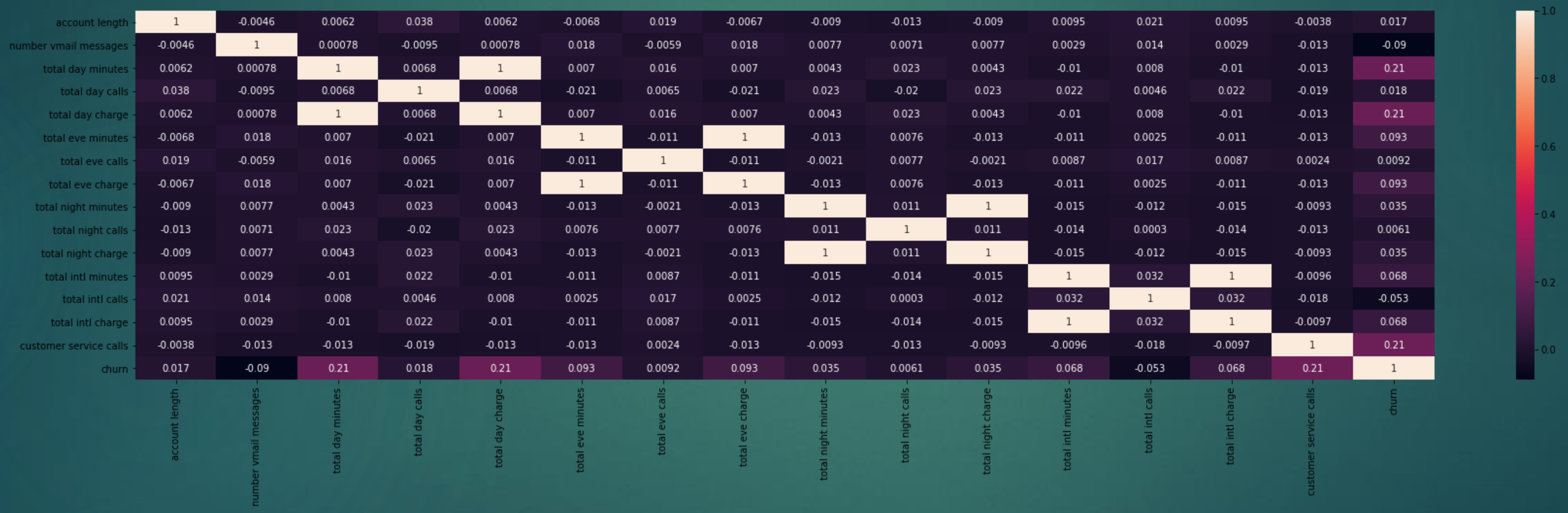
2. Data Loading and Understanding

During our EDA, there are some findings that have come up and need to be addressed before getting into modeling. Below are the findings:

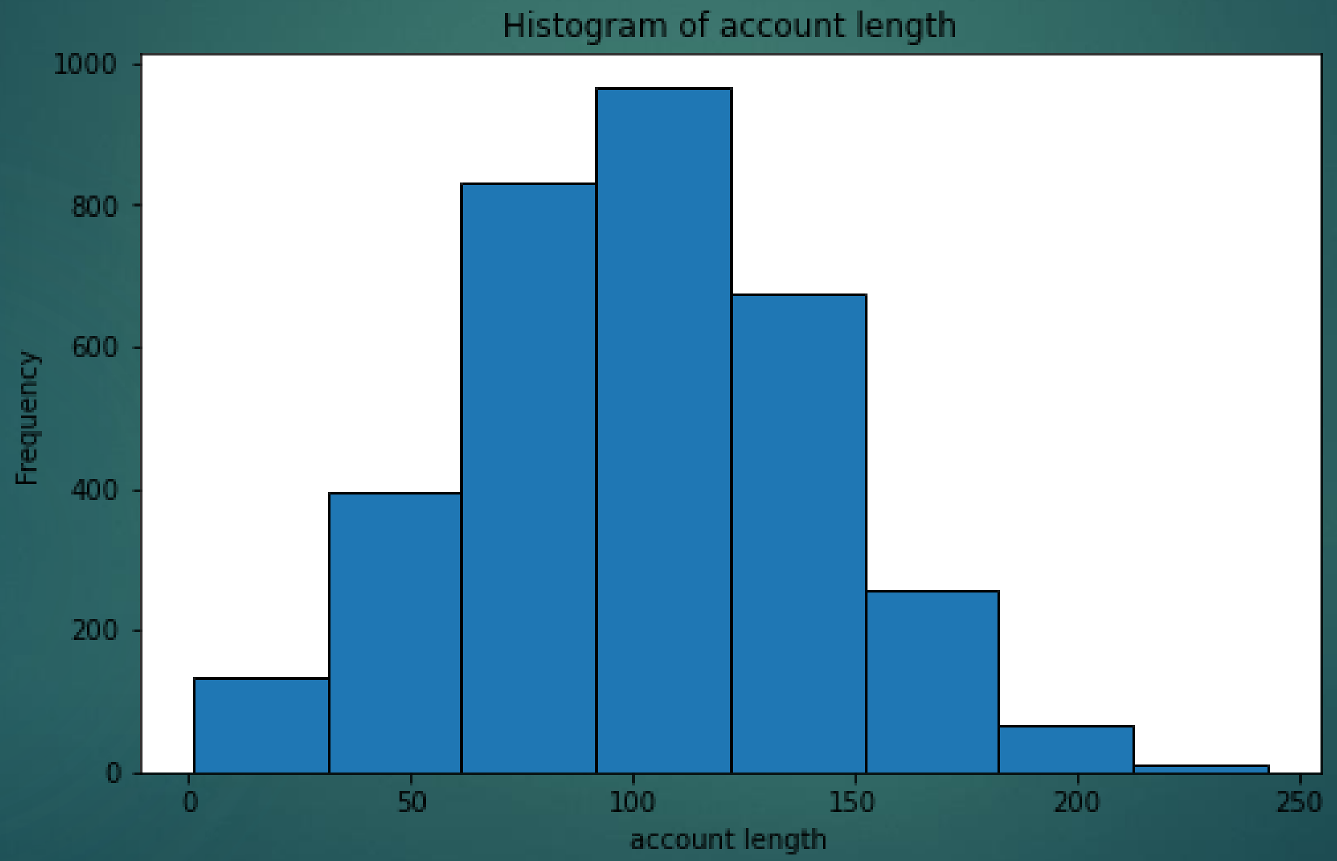
Finding 1: **Converting data type** - an important data type issue related to the 'area code' column. Although it is represented as an integer in the dataset, the values it contains are essentially placeholders or labels, not numerical values that carry mathematical significance.

Finding 2: High correlation - Multicollinearity - Our examination of the heatmap representation of the data revealed that several columns exhibit high levels of correlation with each other. This observation indicates the presence of multicollinearity, a condition where independent variables in our dataset are highly interrelated.

Heatmap to check how the columns are correlated



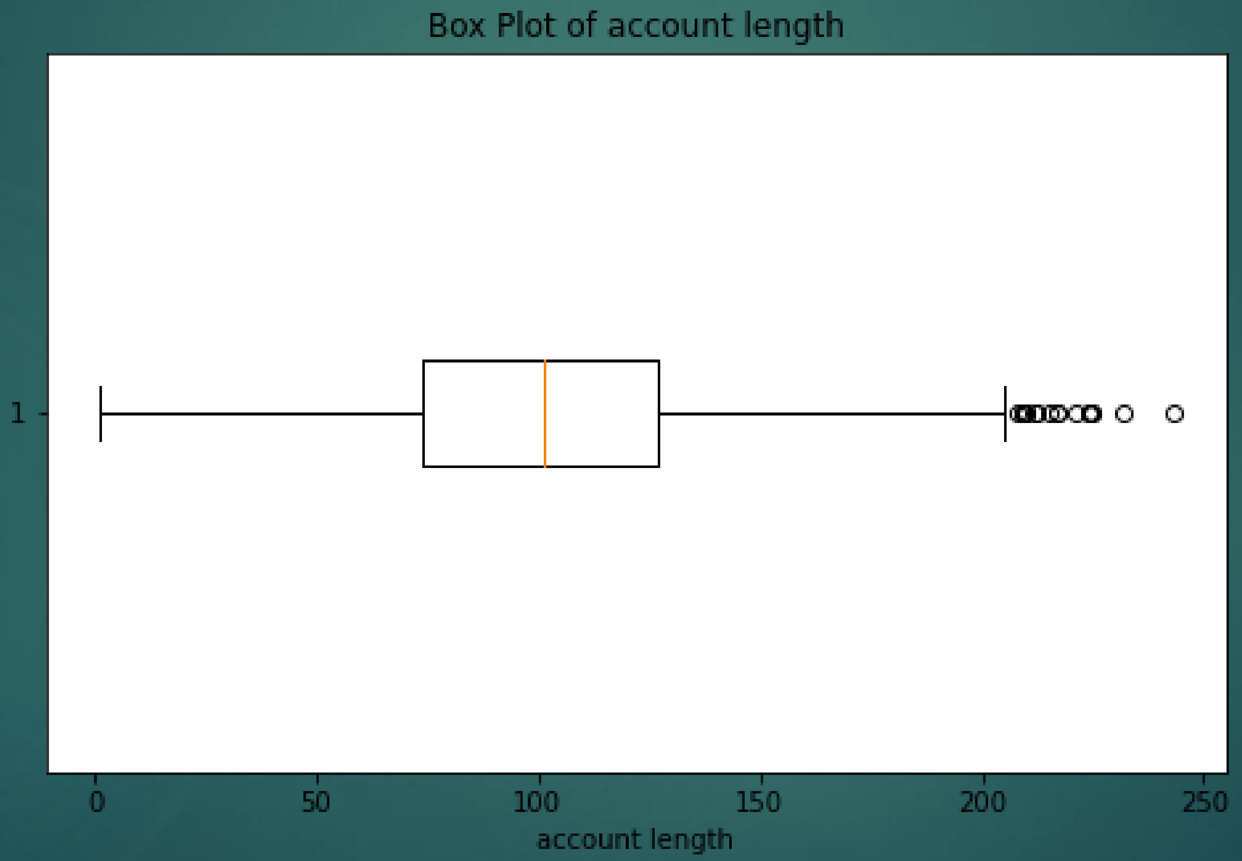
Histograms for numerical features to visualize their distribution



Finding 3: Outliers

we observed the presence of a significant number of outliers in our dataset, as indicated by the boxplots. These outliers have the potential to impact our modeling process. However, it is important to note that, in this case, these outliers are not anomalies that should be removed. Instead, they are a noteworthy aspect of our dataset that we should be aware of during our modeling process.

Box plots to identify outliers and visualize the spread of data



3. Data Preparation

We noted that our dataset does not contain any missing values and duplicates.

We took below steps to prepare our data for modelling.

1. Removing irrelevant columns; we streamline our data by removing columns that are not essential for our analysis or modeling

2. Step 2. Feature Engineering - In our dataset, we have identified that both our target variable and certain feature columns are categorical in nature. To effectively use this data in our modeling process, it is advisable to encode these categorical variables into a numerical format.

Step 3. Choosing the Target and the Features

We've chosen the relevant features and the column churn as our target for our models.

4. Modeling

We developed a predictive model designed to anticipate whether a customer is on the verge of discontinuing their engagement with Syria. The primary objective is to curtail financial losses stemming from customers who have a short-lived association with the entity.

The Model Used:

1. Logistic Regression
2. 2. Decision Trees
3. 3. KNN Classifier Model

5. Evaluation

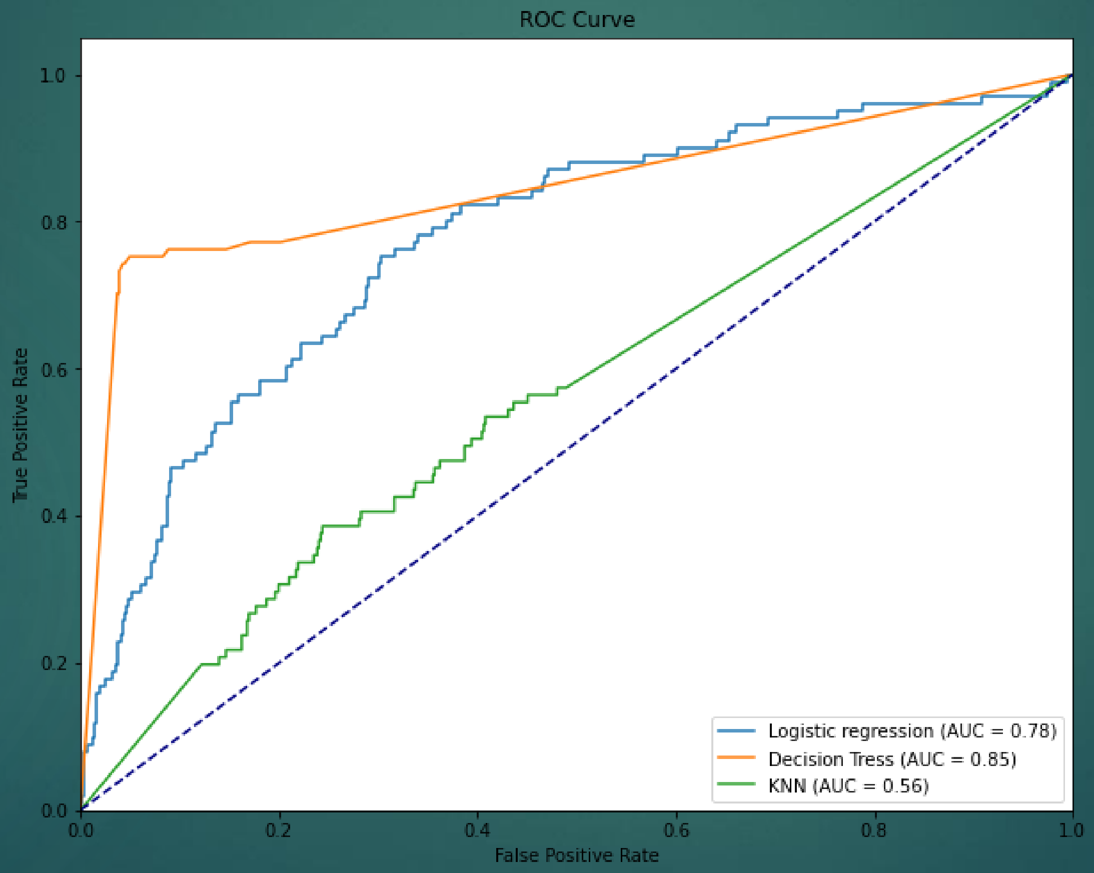
● The logistic regression model shows a balanced performance with reasonably good accuracy, precision, recall, and F1 score. It captures positive cases effectively while maintaining precision. The ROC AUC score is also decent.

● The decision tree model exhibits high accuracy, especially for the majority class. However, it has lower precision, recall, and F1 score for the minority class. This suggests that it may not perform as well on classifying the minority class. The model is well-suited for imbalanced datasets.

● The KNN classifier model achieves decent accuracy and performance for the majority class but struggles with the minority class, similar to the decision tree model.

We determined the best model using the ROC curve to compare the three computed models above

ROC Curve for best model evaluation



6. **Model of Choice: Decision Trees**

- The decision tree was our best model because:
- the model achieved a high recall score of 0.75. In our case, a recall score of 0.75 implies that the decision tree model can capture a significant number of churn cases, correctly identifying 75% of the actual instances of customers who are likely to churn. This high recall score is crucial because it aligns with our objective of identifying and mitigating churn effectively.
- the model demonstrated an accuracy score of 0.91. This indicates that the model performs well in accurately predicting churn, which is essential for making informed business decisions.
- The model's substantial ability to balance precision and recall is reflected in its F1 score of 0.72. A higher F1 score indicates that the model effectively identifies true positives while minimizing false positives and false negatives. In our case, the Decision tree model achieves an F1 score of 0.72, suggesting that it strikes a good balance between precision and recall, resulting in accurate identification of churn cases.

7. Recommendations and Future Investigations

- Customer Service Calls Investigation:
- Dig deeper to understand why some customers need to contact customer service frequently and if there is a relationship between customers who call and their churn rate. This will help in finding ways to better assist them.
- International Plan Churn Investigation:
- Since some of the customers with international plans are leaving, it's essential to explore bundled plan for them in oreder to retain these customers.
- High Churn States Analysis:
- Look into the states where many customers are leaving to identify any patterns or reasons for the high churn rates.
- Incentives for High Bill Customers:
- Find ways to reward customers with high daily charges (over $55) for them to stay with SyriaTel. This might involve offering extra benefits and perks. Currently, all of these high bill customers are leaving, which is a concern.
- Incentives for Customers who stay more than 6 months:
- Find ways to encouragte customers to stay with the company even longer, eg giving them loyalty points offers etc, as this will help in creating a form of loyalty

Conclusion

- It is important to monitor the effectiveness of these measures through ongoing analysis of churn rates, customer feedback, and market trends. Regularly evaluating and adapting these strategies based on customer needs and preferences will help optimize retention efforts and reduce churn effectively. Additionally, providing personalized offers, loyalty programs, and targeted marketing campaigns can further contribute to improving customer satisfaction and loyalty.

Thank you