



Student Name:

Dennis Kobia

Dspt 04





Project Name:

SyriaTel Customer Churn Prediction



Business and Data Understanding

Objective: The primary objective of the project is to develop a predictive model that can identify customers who are likely to churn in the near future. This will allow SyriaTel to implement targeted strategies to retain these customers.

Impact of Churn: Understanding the financial implications of customer churn for SyriaTel is crucial. This could include lost revenue, acquisition costs for new customers, and the potential impact on the company's reputation and market share.

Customer Segmentation: Recognizing that not all customers have the same likelihood of churning. There may be different segments of customers with varying churn probabilities. Identifying and targeting these segments effectively can be a key strategy.

Business Goals and KPIs: Understanding SyriaTel's specific business goals, such as revenue targets, customer retention targets, and customer satisfaction goals. These KPIs will guide the evaluation of the project's success.

Data Acquisition

I have chosen a dataset from the curated list, below is the link:

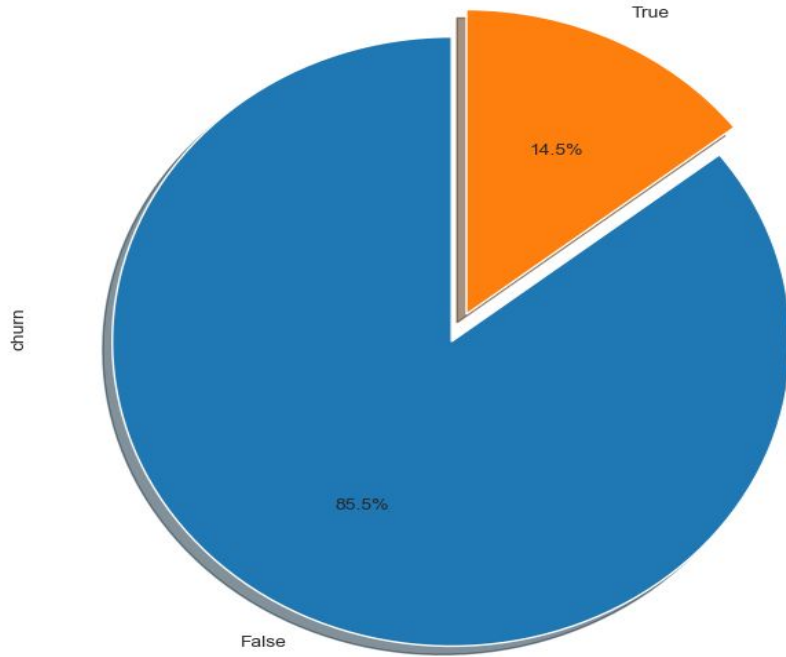
<https://www.kaggle.com/datasets/becksddf/churn-in-telecoms-dataset>

Data Preparation and Feature Engineering

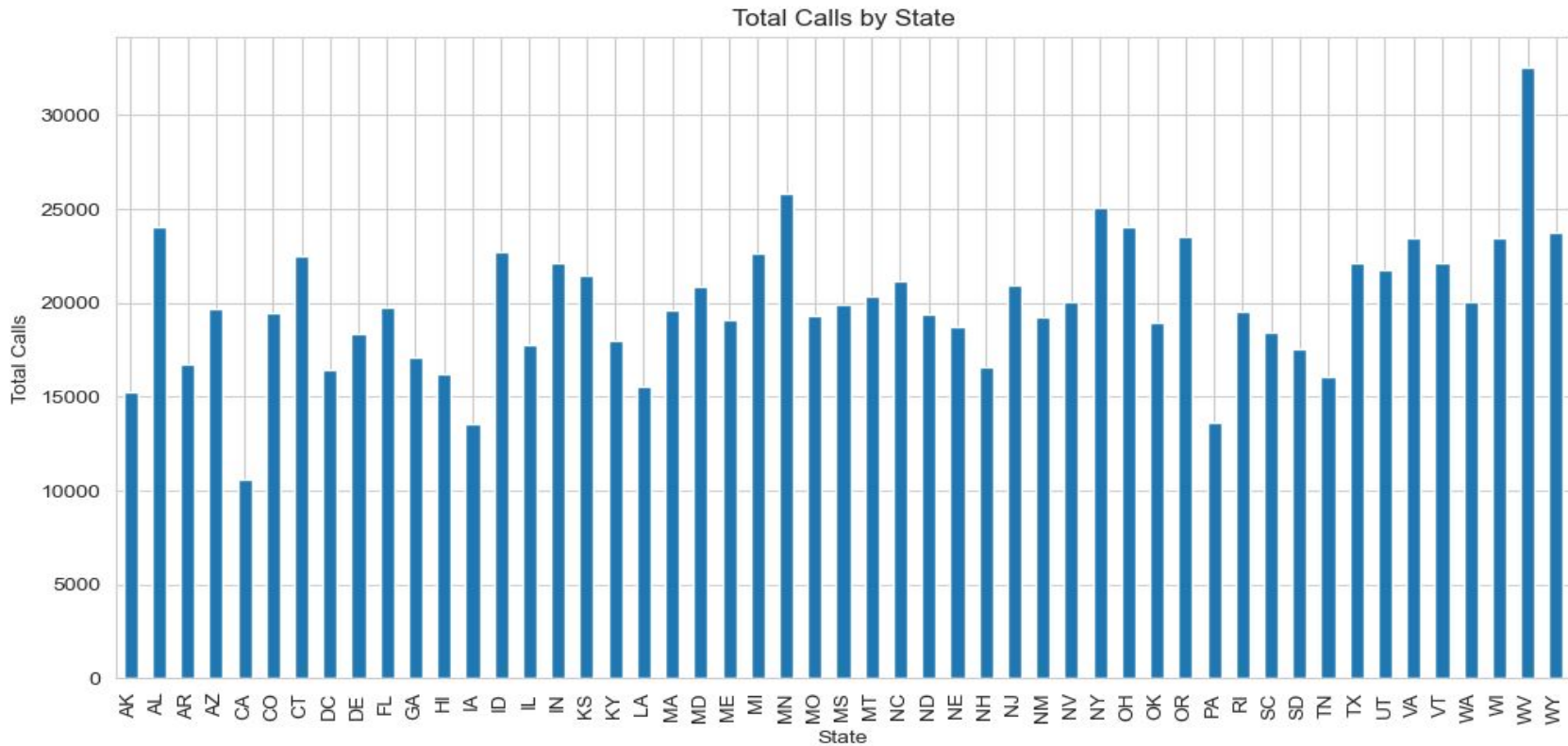
There was neither missing values nor duplicates in the SyrialTel data. I went ahead and checked for outliers which I found and removed.

I also separated categorical columns and the numerical columns for easier analysis. I dropped columns which were of no use to the achievement of the project objectives. I used the groupby method to visualize the pre-existing relations in columns . I used matplotlib and seaborn for visualization. In order to select the best features for modelling I used heatmap to check the correlations between the Predator variables. I went ahead to encode categorical variables and performed feature scaling.

Churn Pie_Chart



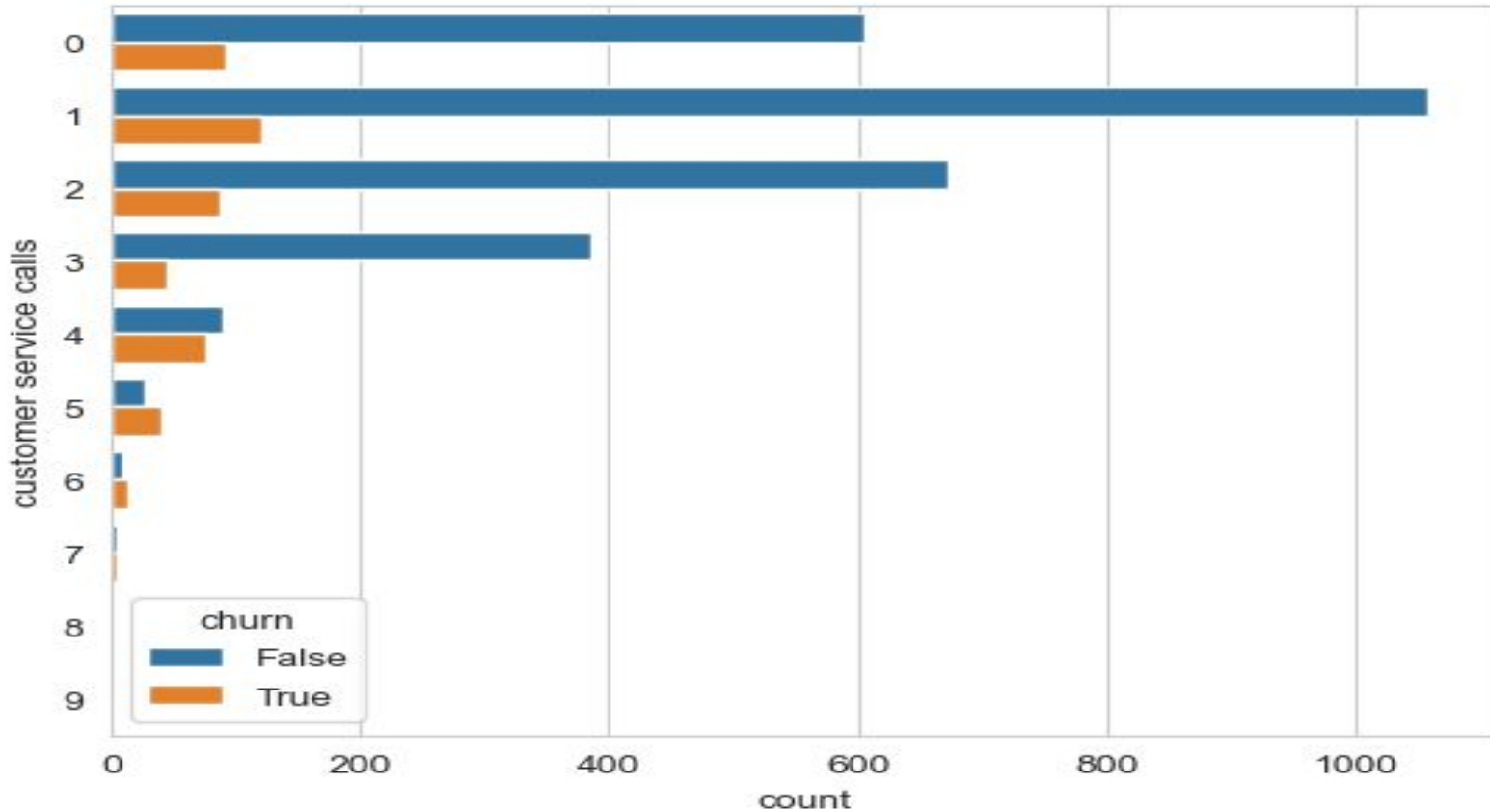
In the above Pie Chart, 14.5% shows customers who left Syrialtel and 85.5% shows customers retained during the period under review



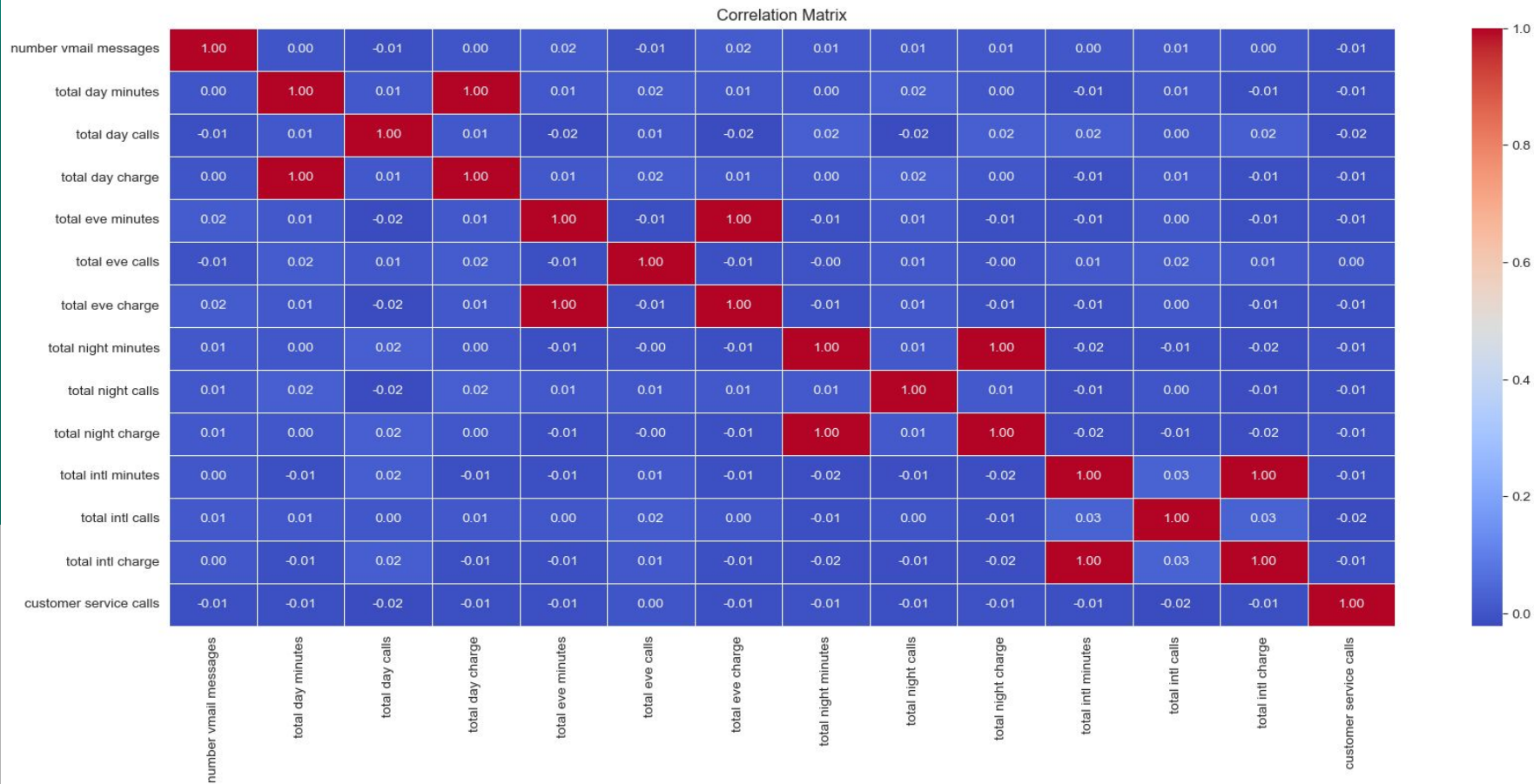
The figure above shows total calls per state. West Virginia has the highest call. California has the least calls



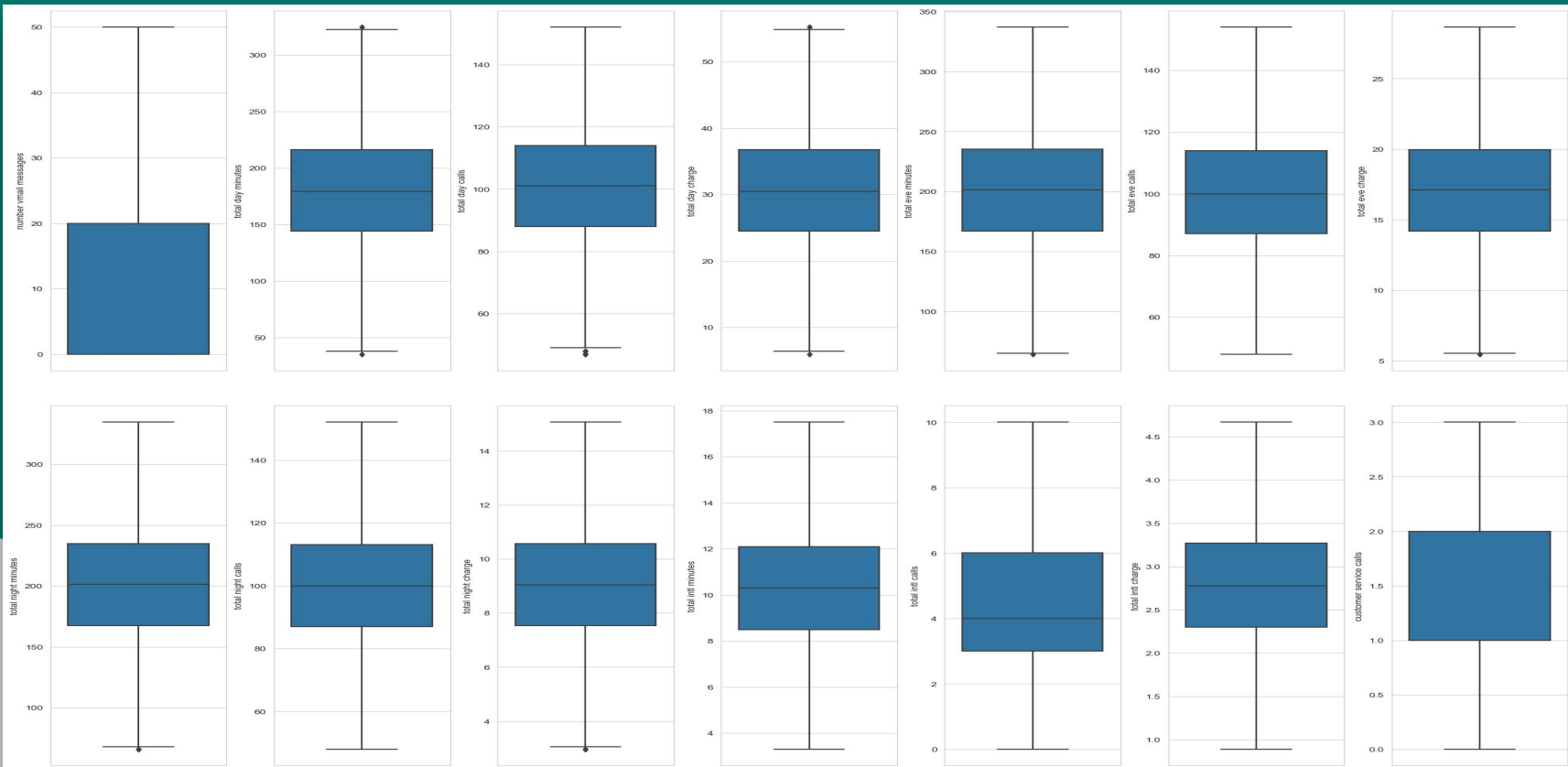
The bar graphs show international calls made by those customers with a plan vs customers without an international plan.



The chart above shows the relationship between customers making calls to the customer care and likelihood of churning



The above Heatmap show the correlation existing between predictor variables



Above are the box and whisker plots to check to show the representation of outliers

Model Development

I have developed and trained 3 classification models on the prepared dataset namely:

- *Logistic regression

- *Decision Trees

- *Random Forest

Model Evaluation

I have evaluated the chosen models using the following metrics

- *Accuracy

- *Precision

- *Recall

- * F1 -score

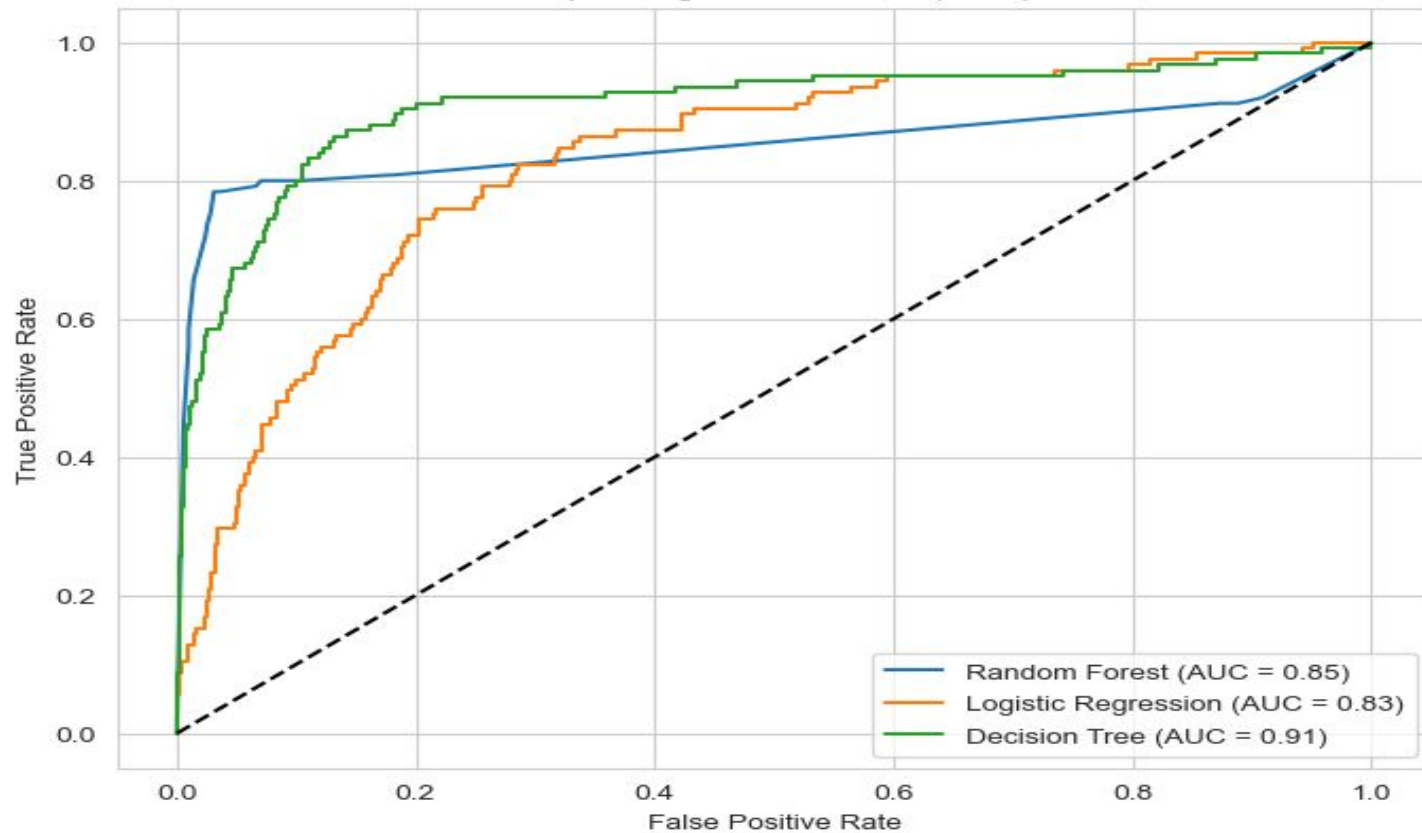
- * ROC - AUC curves

Conclusion of the 3 models based on accuracy

Based on the 3 models the accuracy in predicting customer churn is as follows:

- Logistic Regression performs poorly with 85.1%
- Random Forest classifier showed improvement with 87.1%
- Decision Tree performed the best with 91.5%

Receiver Operating Characteristic (ROC) Curves



Based on the above AUC scores, Decision Tree has the highest score, followed by Random Forest and Logistic Regression has the lowest score



Integration and Deployment

From the 3 adopted models Decision Tree has emerged as the best robust model in predicting SyrialTel customer churn

I will integrate this model into SyrialTel's existing systems for seamless prediction of customer churn.