Unlocking Insights: Decoding Churn Analysis

Leveraging Data Insights to Retain
Your Valued Customers

PRUDENCE COREDO

Summary

- This project entails of a churn analysis on Syriatel, a telecommunications company ,providing a range of services.

- A model was built to predict which customers are likely to discontinue their services with the company in order to find methods that helped retain their customers and attract new ones .

- This analysis can also be used to help reduce the amount of money lost when customers leave.

- This was done by building 3 different models to see which one was most accurate.

- Different features were analyzed to see which one had the biggest impact on churn or if there was any relation at all .

Agenda

- **Business Problem**

- **Data**

- **EDA**

- **Results**

- **Conclusions**

Business Problem

Syriatel is experiencing high churn rates making them loss customers and in turn reduced revenue.

Stakeholder : Syriatel Executives.

Data

- Outliers were dropped.

- The dataset contained 3333 rows and 21 columns

- One-hot encoding performed on categorical features.

- Numerical features were scaled.

- Synthetic Minority Oversampling Technique ("SMOTE") applied to resolve class imbalance.

- Hyperparameter tuning was done to improve he model

EXPLORATORY DATA ANALYSIS
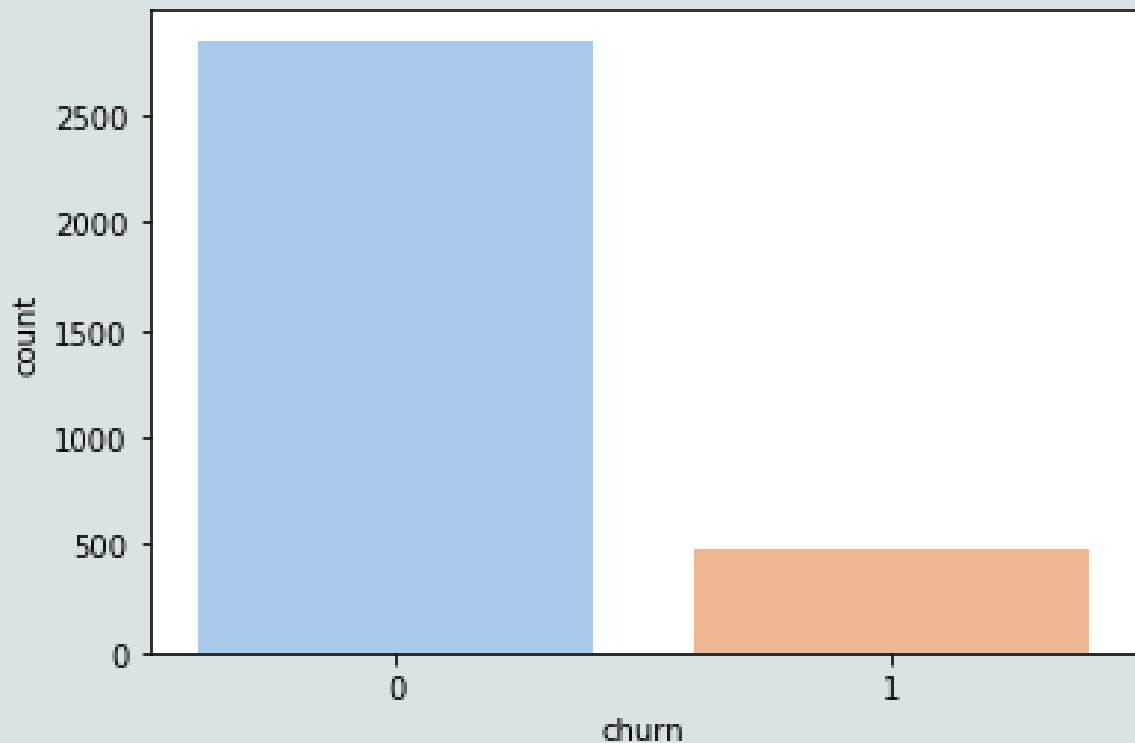
- ## Churn Analysis

- **483 out of 3,333 have terminated their contract with Syriatel and 2850 still have an account.**

- **This means that 14.5% of the customers are lost.**

- **0 – False ; have not churned**

- **1 – True ; Have churned**

International plan and Voice mail plan
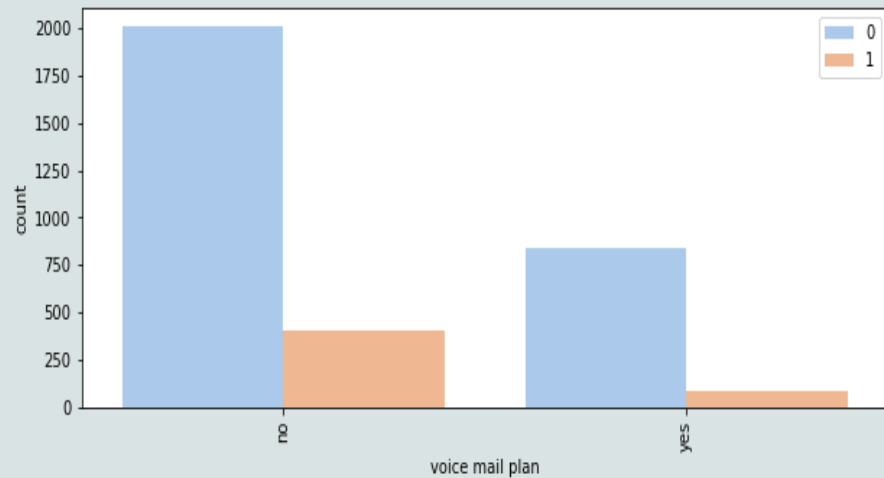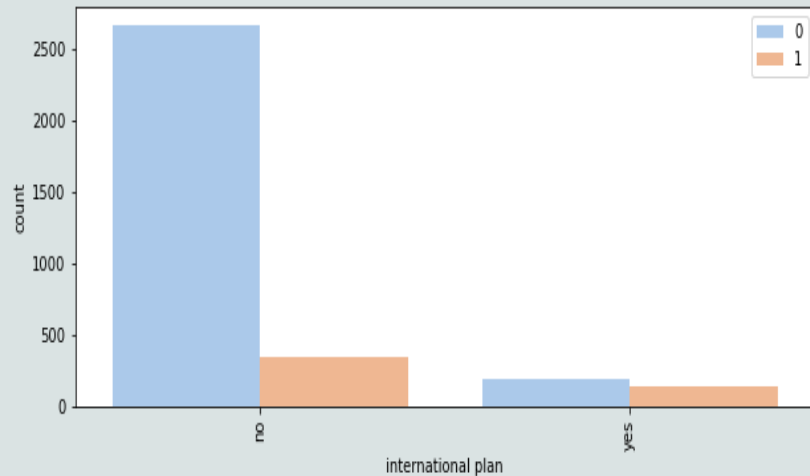
- customers without a voice mail plan and those without an international plan had a higher churn rate compared to customers with a voice mail plan and an international plan.

- The proportion of churn for customers with an international plan is significantly higher compared to those without. This indicates that the presence of an international plan may be a significant factor in customer churn.

Modelling

|	Logistic Regression	Decision Trees	Random Forest	Random Forest Tuned
Accuracy	75%	89%	92%	93%
F1-Score	48%	68%	75%	77%
Recall	72&	76%	74%	79%
Precision	36%	61%	76%	76%

Selecting our Model

Confusion Matrix
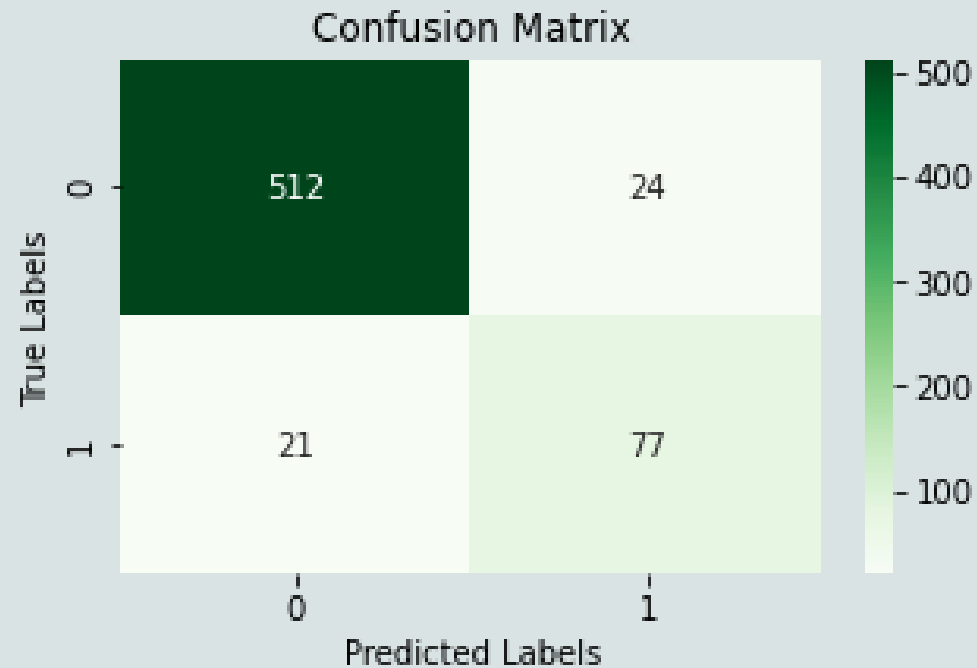
- The chosen model is Random forest classifier after tuning because it has the highest recall score of 79%, this indicates that the model correctly identified around 79% of the actual positive cases,. And an accuracy score of 93%.

- A high f1-score of 78% shows a balanced evaluation of the model's accuracy in predicting both churn and non-churn customers.

Feature Importance

Based on the graph Total day minutes, total evening minutes and total international minutes are the most important features when predicting churn.

However, the other features do still play a role and need to be taken into consideration

Feature Importances
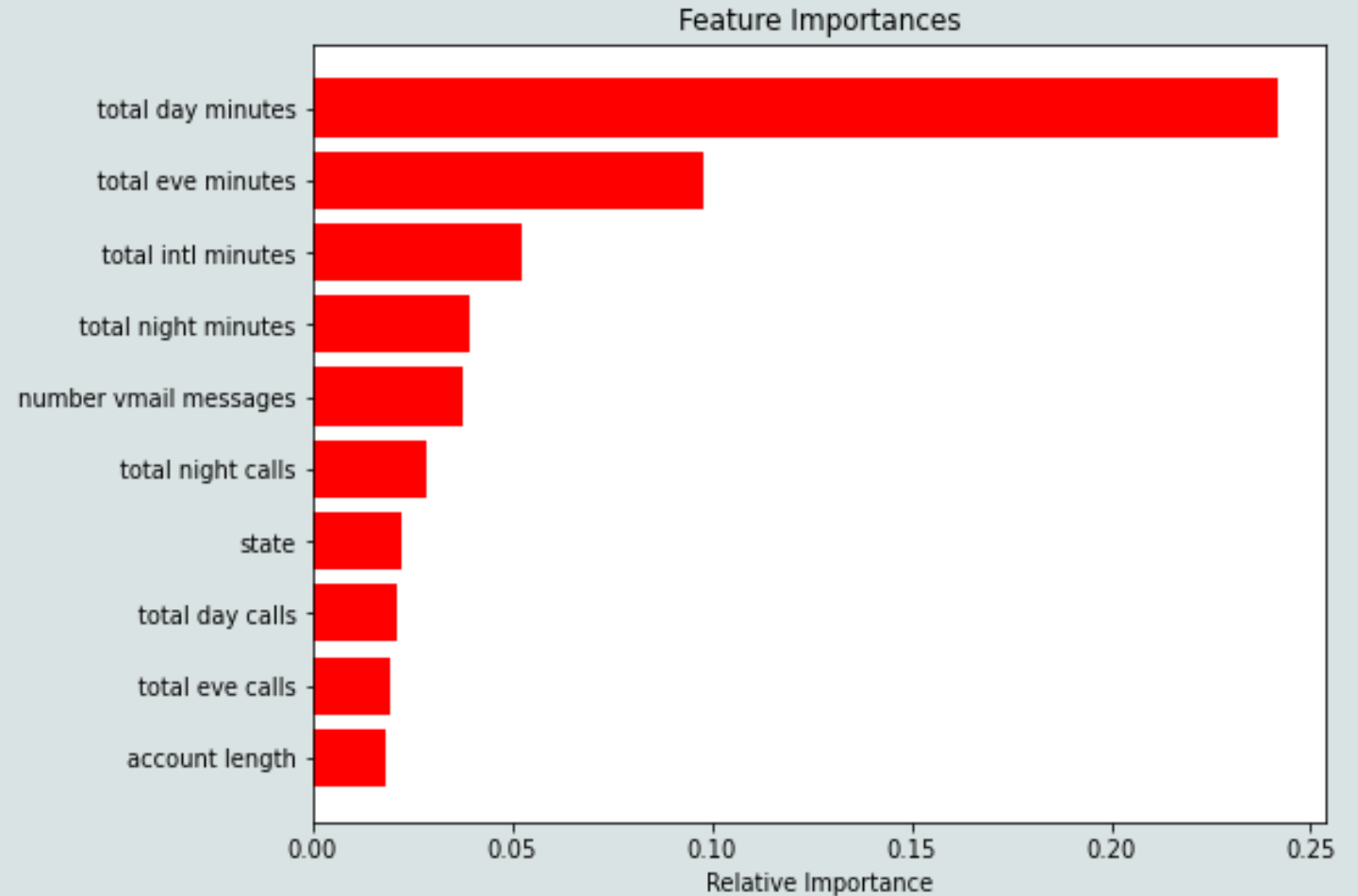
Recommendations

- Implement a strategy that could include giving discounts to international plans and voice mail subscriptions as we've seen those without these plans tend to churn. Many of them may see these plans as too expensive to purchase hence opting not to . This could help attract these niche customer base.

- People with large total call minutes must be charged large amounts of money and may be dissatisfied with the charges hence decide to leave. To counteract this, we could introduce package deals to those who call for long periods of time.

- Loyalty points / cards and deals for customers who have had their account for long periods of time.

- Improve Customer feedback - This could include conducting surveys of customers who have churned to understand why they left and the reasoning behind it. This could also include improving customer service calls as they play a role in churn. Find out the most common issues that customers call about.

Conclusions

NEXT STEPS

- Further investigate on what trends might be causing high churn in certain states such as texas and how we can prevent this.

- There could be othr features that affect churn other than the ones privided in our dataset.

- Explore other model classifiers that may provide better results than random forest classifier.

SUMMARY

- Texas has the highest churn rate.

- Based on the graph Total day minutes, total eveing minutes and total international minutes are the most important features when predicting churn.

- .

Thank you