SYRIATEL CUSTOMER CHURN ANALYSIS

PRESENTATION BY:

VIVIAN MOSOMI

INTRODUCTION

BUSINESS PROBLEM

- SyriaTel - A Telecommunication Company, is facing customer attrition. Understanding and predicting customer churn is crucial for the company's sustainability and growth in the telecommunications industry. The company is interested in reducing how much money is lost because of customers who don't stick around very long by being able to use a predictive model that can identify customers who are likely to churn based on various factors.

OBJECTIVE

- To be able to predict and identify factors causing customer churn, this will help the company can take measures to retain customers and prevent financial loss

EXPLORATORY DATA ANALYSIS

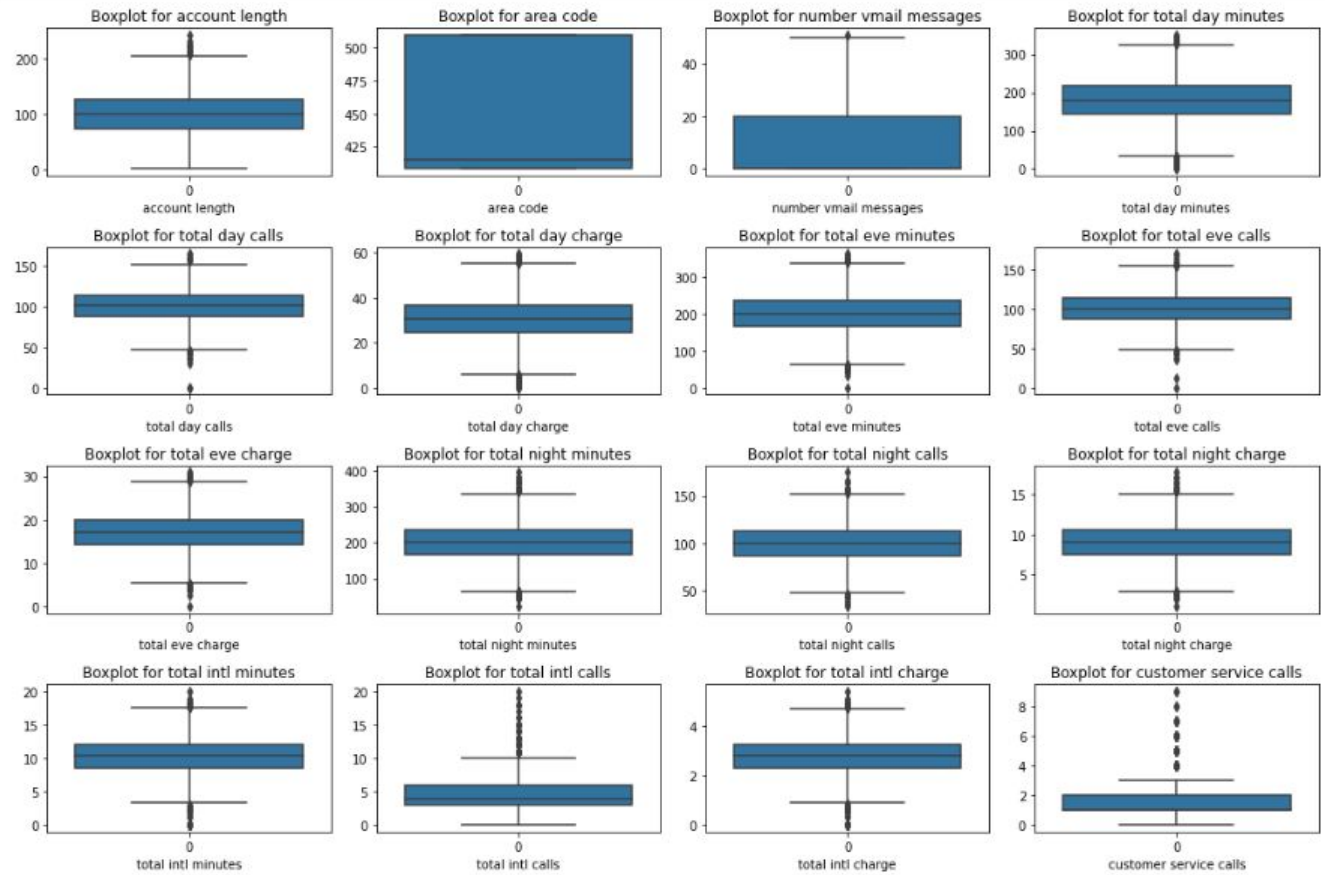
DATA UNDERSTANDING

- The dataset consists of 3333 rows, 21 columns

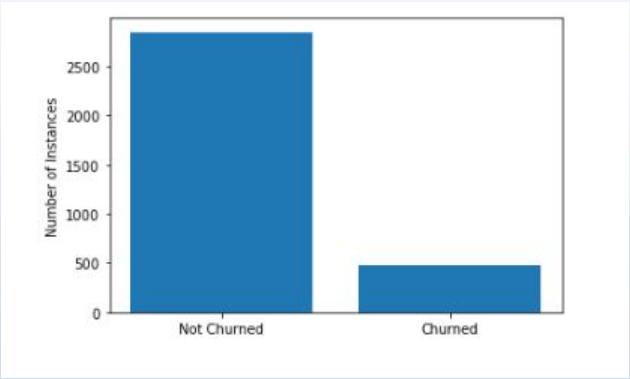
FEATURE COLUMNS

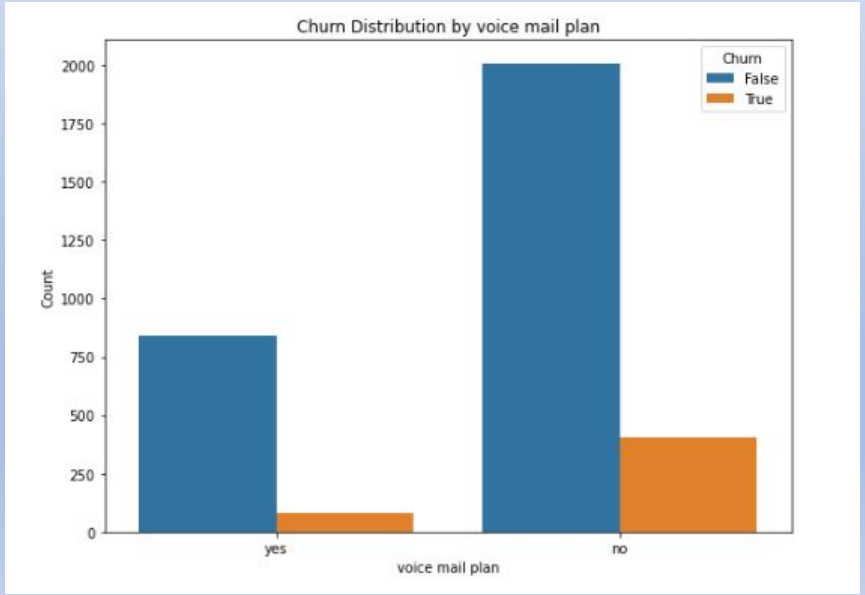
state, account length, area code, phone number, international plan, voice mail plan, number vmail messages, total day minutes, total day calls, total day charge, total eve minutes, total eve calls, total eve charge, total night minutes, total night calls, total night charge,
total intl minutes, total intl calls, total intl charge, customer service calls,
churn

VISUALIZATION



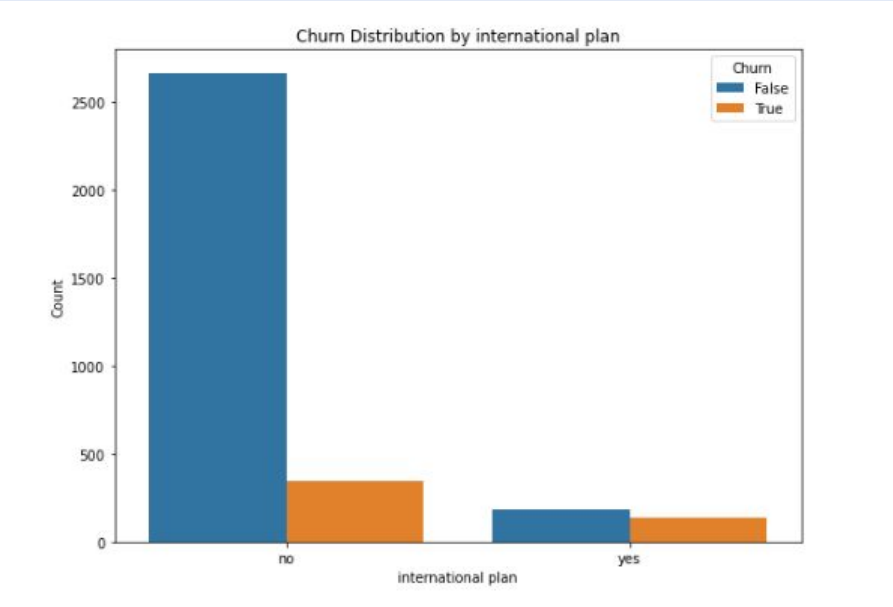
- analyzing the outliers in every numeric column





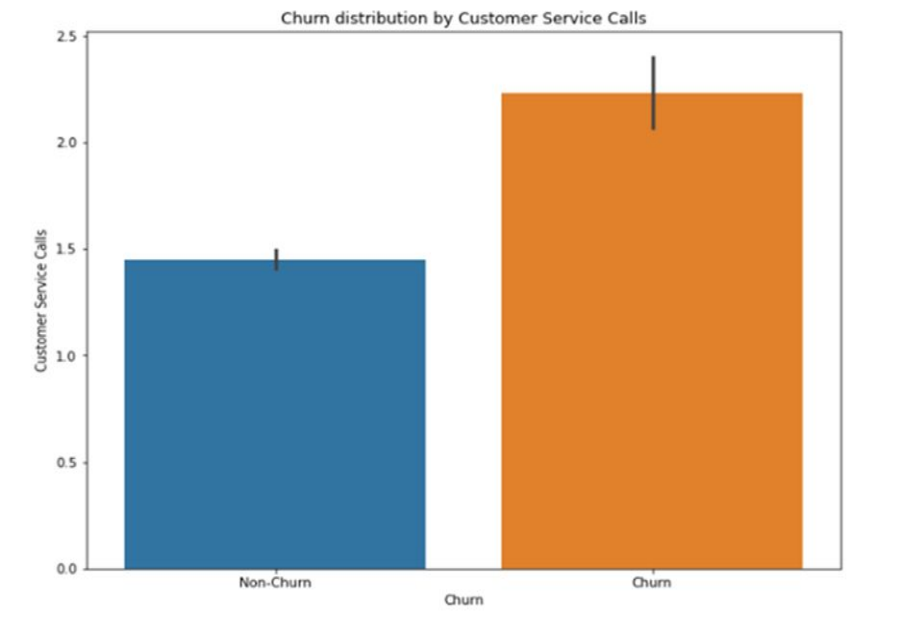
- Visualizing the relationship between number of churn and not churned instances

- Visualizing the churn distribution by voice mail plan



- Visualizing the churn distribution by international plan

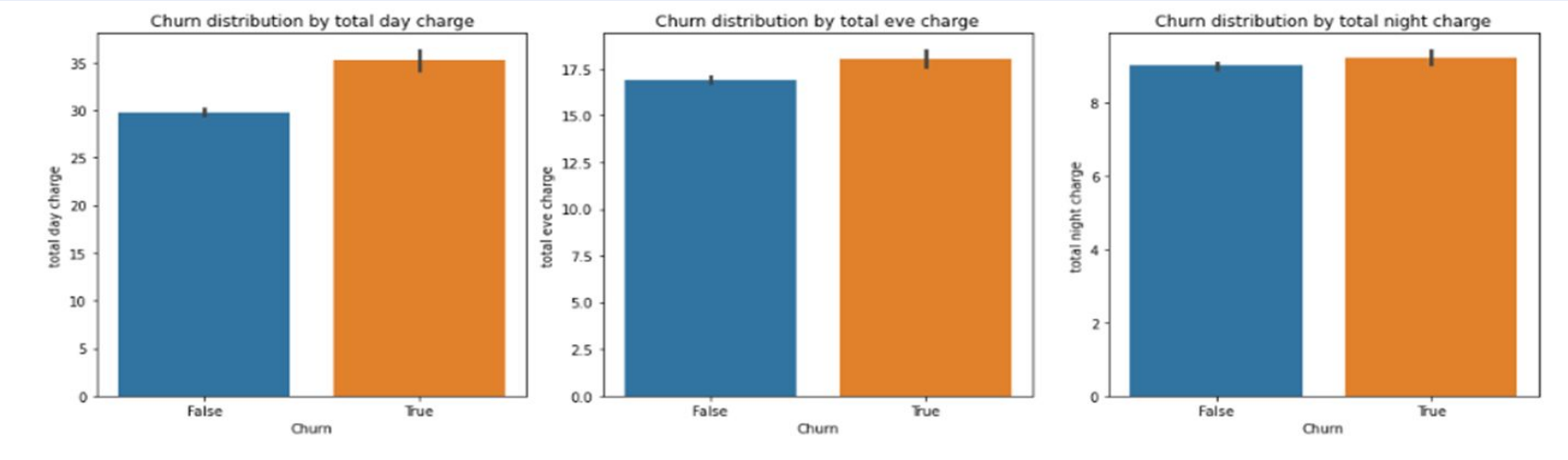
- From the above plots of voice mail plan and international plan, we can see that the count of customers with no voice mail and international plan have a high churn count.



- Visualizing the relationship between customer service with churn and non-churner

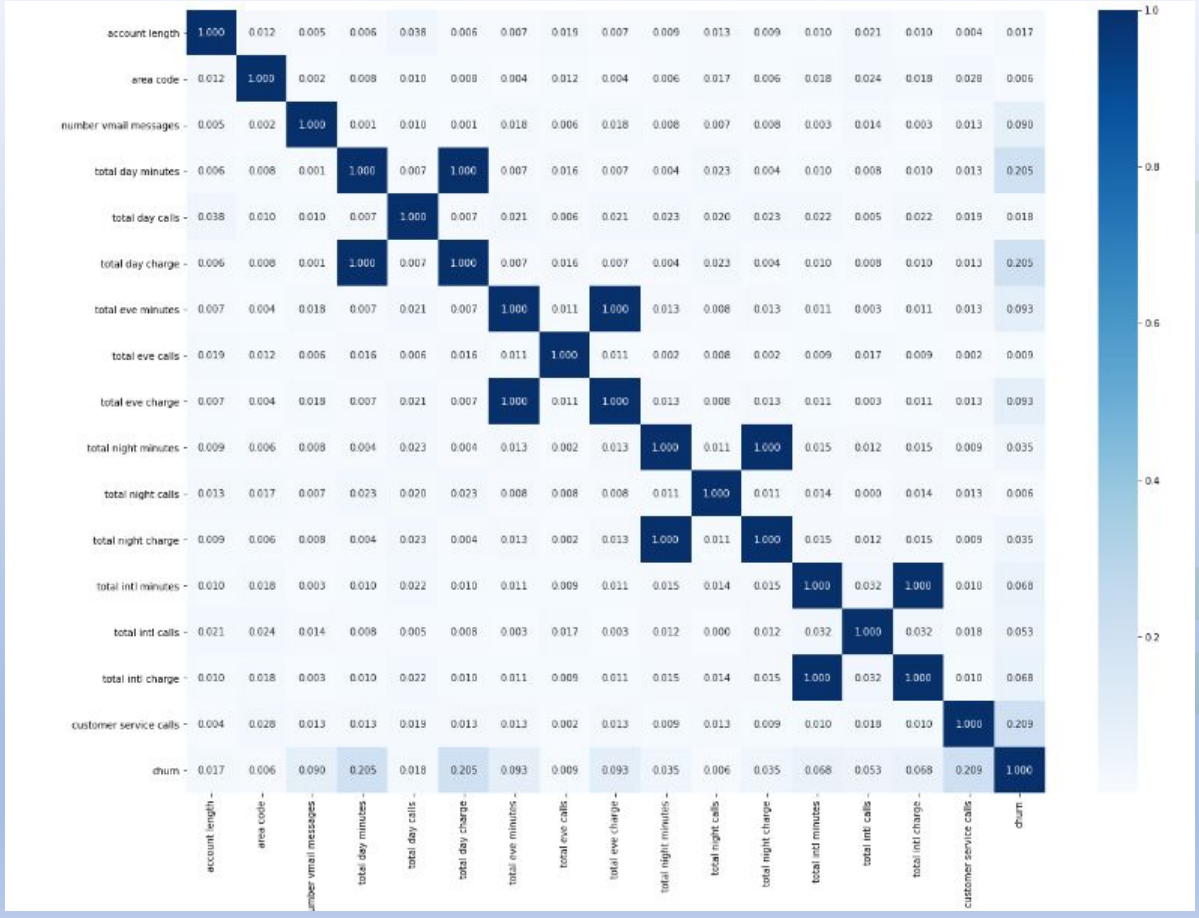
- From the above plot for customer service calls with churn rate, we can see that there's high rate of customer service calls

for churn customers.



- Distribution of churn and charges(day, evening and night charges)

From the charges plot above, we can see that the higher the charges the higher the churn rate. For total evening charge and night charge, the churn rate is slightly higher than those who do not churn



- From the correlation matrix, most of the features do not appear to be perfectly correlated but features like:-

total day charge and total day minutes,

-total evening charge and total evening minutes,

-total night charge and total night minutes,

-total intl charge and total intl minutes

MODELLING AND EVALUATION

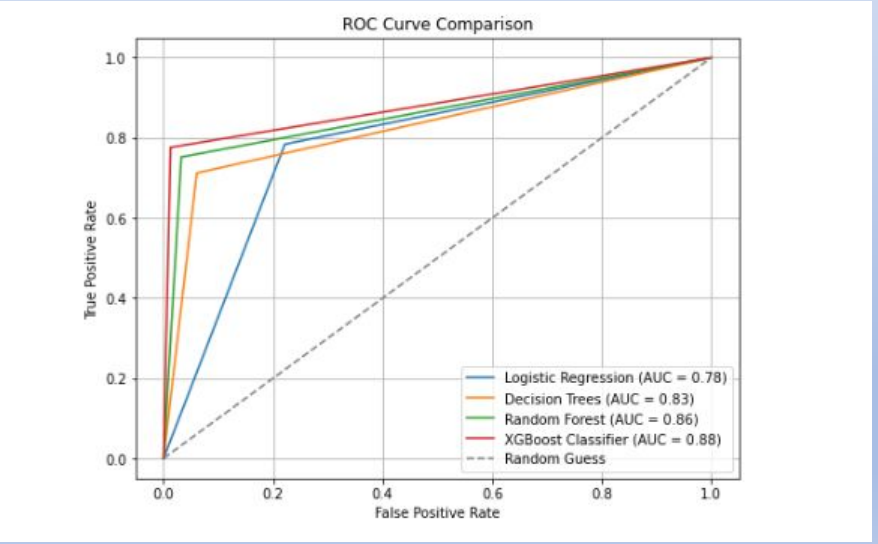
	Logistic Regression	Random Forest Classifier	Decision Tree classifier	XGBoost
precision	0.38	0.80	0.67	0.92
Recall	0.78	0.75	0.71	0.78
F1-score	0.52	0.78	0.69	0.84
Accuracy	0.78	0.94	0.91	0.96

AFTER MODEL TUNING

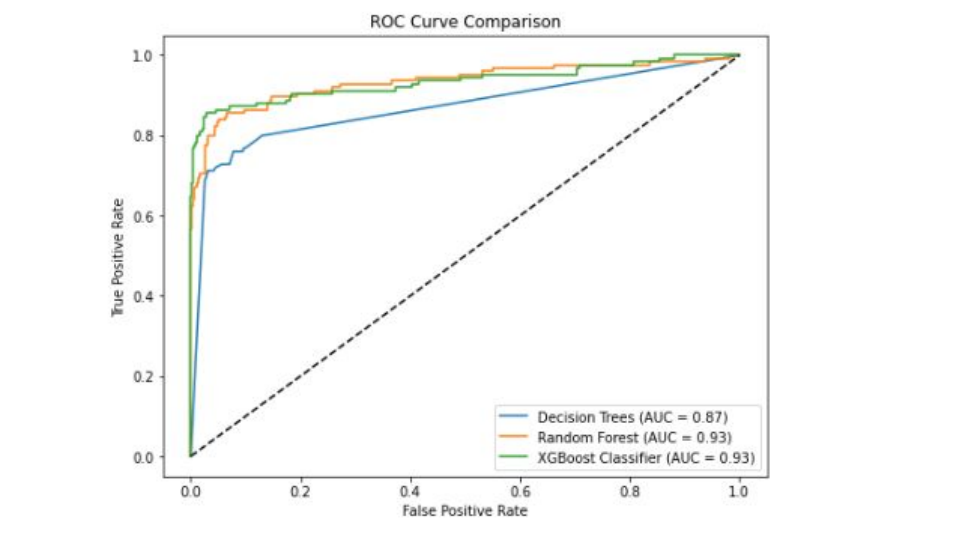
	Baseline Logistic Regression	Random Forest Classifier	Decision Tree classifier	XGBoost
precision	0.38	0.84	0.68	0.94
Recall	0.78	0.77	0.73	0.78
F1-score	0.52	0.80	0.70	0.85
Accuracy	0.78	0.94	0.91	0.96

ROC CURVE ANALYSIS

BEFORE TUNING

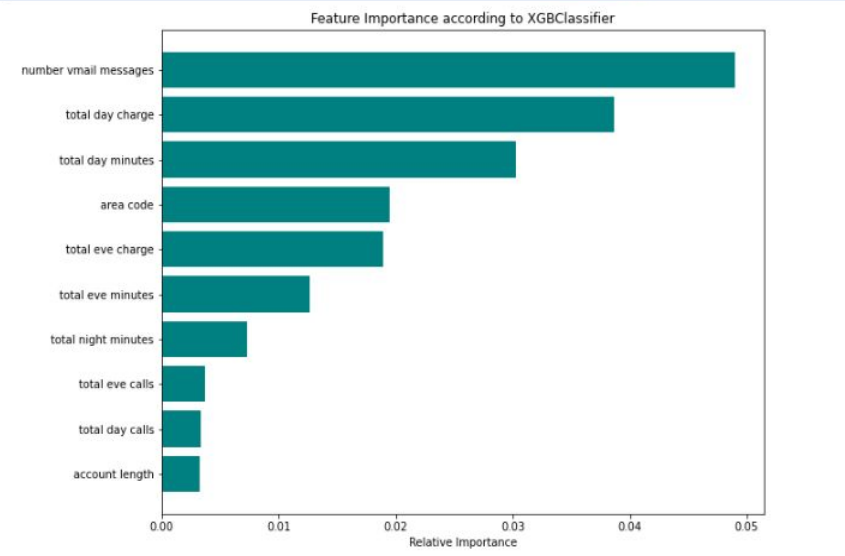


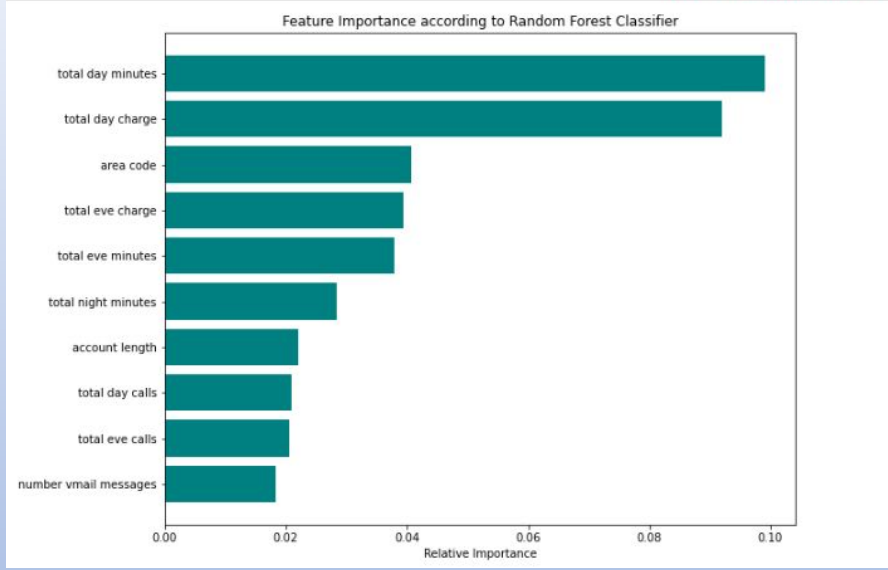
AFTER TUNING



SELECTING OUR MODEL

- The XGBoost and Random Forest models were the most suitable due to its strong performance in terms of recall and F1 score





- 3 top most important features for this model are: number of voicemail messages, total day charge and total day minutes

- 3 top most important features from this model are: total day minutes, total day charge and area code

CONCLUSION

- Using the important features from our top tuned classifiers, that is Random Forest and XGBoost:

-Random Forest has the following important features: - Total day charge - Total day minutes - Area code

- XGBoost Classifier has the following important features: - Number of Voice Mail Messages - Total day charge - Total day minutes

- From the above we can tell that the high total day charge and number of voice mail messages influence churn rate as they are deemed as the important features in the two models.
- From Exploratory Data Analysis, we saw that total evening charge, total night charge and customer service calls also have an influence on the churn rate of customers.
- The states New Jersey, California and Texas have higher churn rates.
- This suggests that addressing service-related issues might mitigate high churn rates. These findings can help direct efforts to retain customers and improve service quality

RECOMMENDATIONS

- Based on the insights gained, it is recommended to focus on enhancing the following:

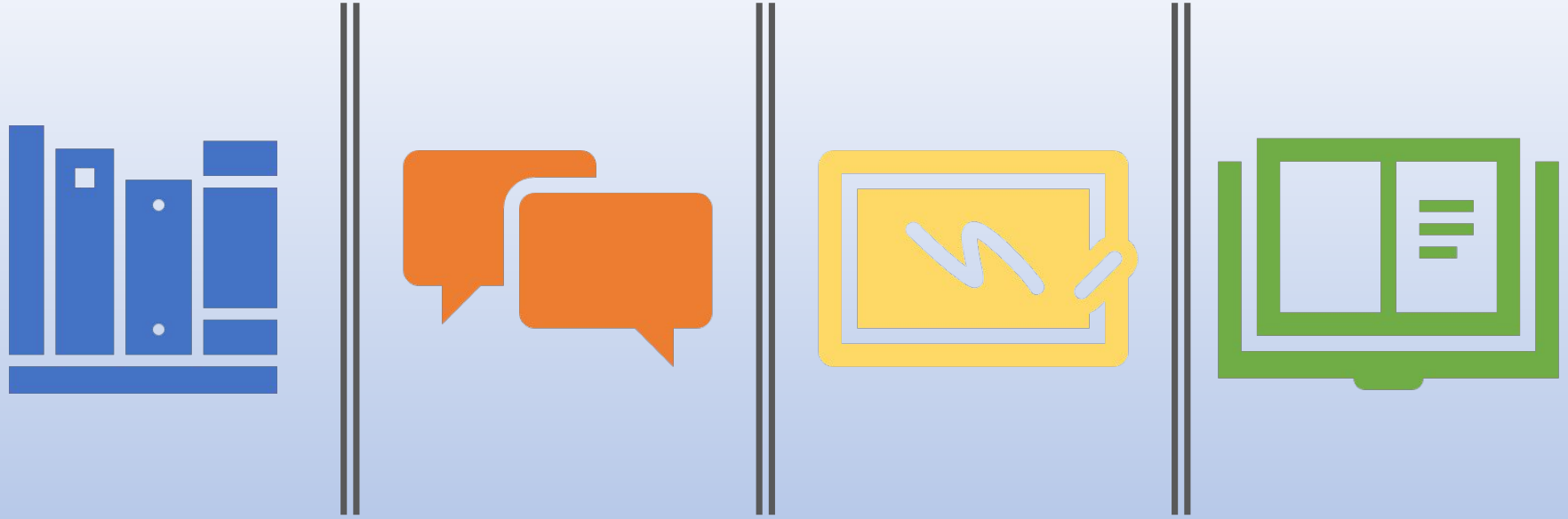
- **Improve customer service quality** so as to reduce the high customer service calls that increase the churn rate. That can be done by first understanding the individual customer needs and trying to maintain high standard of service.

-Syria Tel Company can take measures to **revise pricing strategies for day**, evening and night call charges. The company can negotiate for different plans that offer reduced call charges hence preventing customer attrition.

-**Looking into the cause of high churn rate in New Jersey, California and Texas.** It may be that these states experience poor network coverage or service disruptions hence leading to high churn rate. The Company can also consider marketing the company in those specific states.

-Area Code is also highlighted as one of the important features and hence, the company can look into area codes that have high churn rates and introduce acitivites that will reduce churn rate such as marketing and offering promotions to customers.

- **Limitation:** Syria Tel can consider using the above predictive models to predict customer churn and take measures to enable proactive retention strategies,but should be aware of **computational ability limitations**, especially with models such as random forest and xgboost when using a high number of trees when tuning.



Project Presentation End

THANK YOU!!