Customer Churn

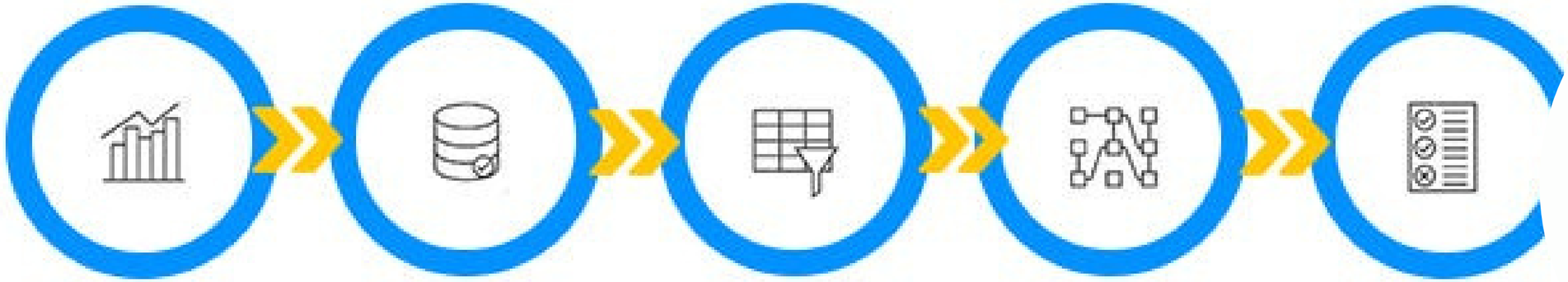
Author: Stephanie Mbithe

ProjectPro

INTRODUCTION

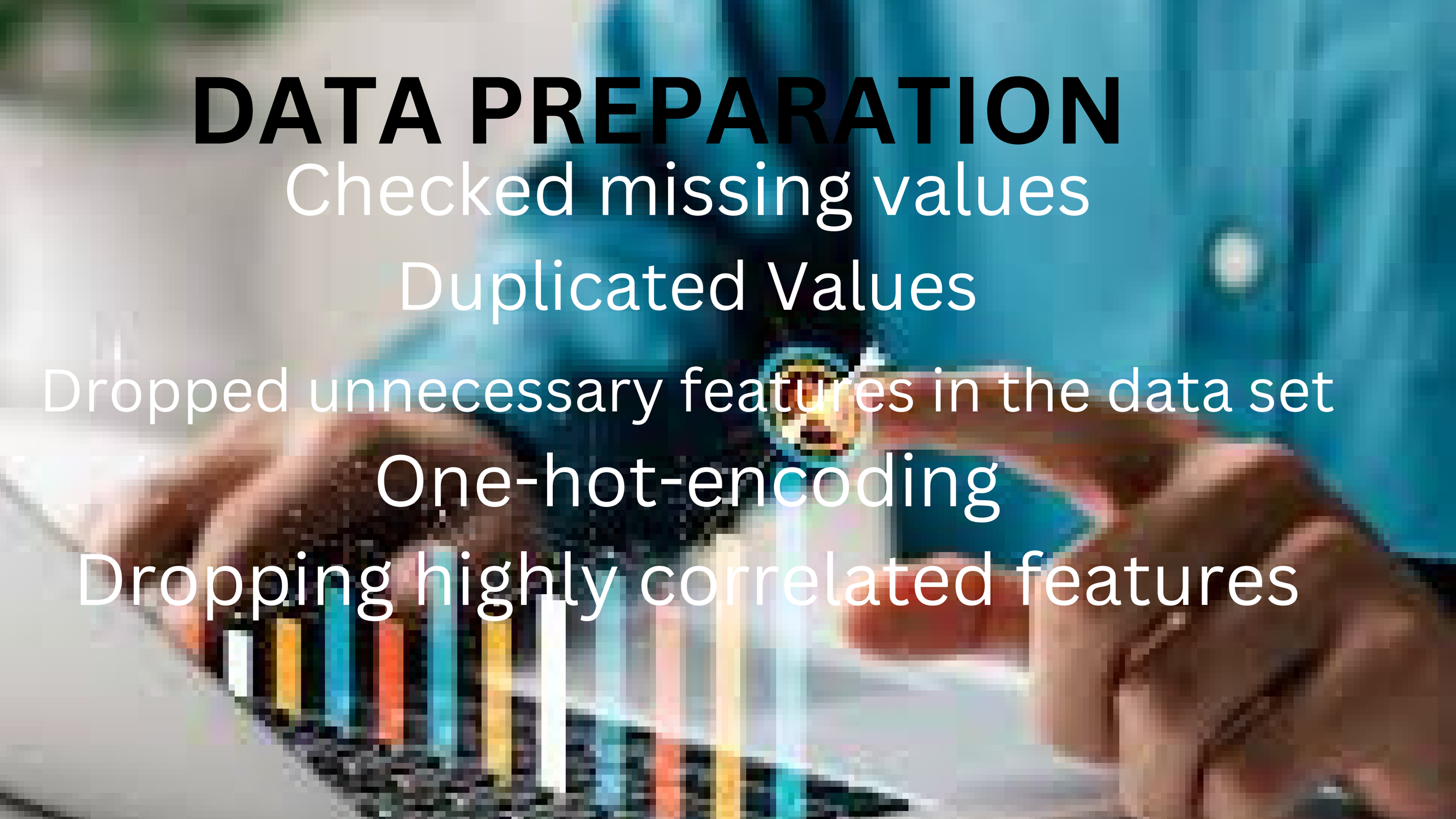
DATA UNDERSTANDING
DATA PREPARATION
MODELLING
EVALUATION
CONCLUSION
RECCOMENDATIONS

Data Science Process



BUSINESS UNDERSTANDING

SyriaTel is a telecommunications company that is interested in reducing the financial losses caused by customers who churn, i.e., customers who terminate their business relationship with the company. To address this problem, we will build a binary classifier to predict whether a customer will soon churn or not. By identifying predictable patterns that are indicative of customer churn, SyriaTel can take proactive measures to retain customers and minimize revenue loss.



DATA PREPARATION

Checked missing values

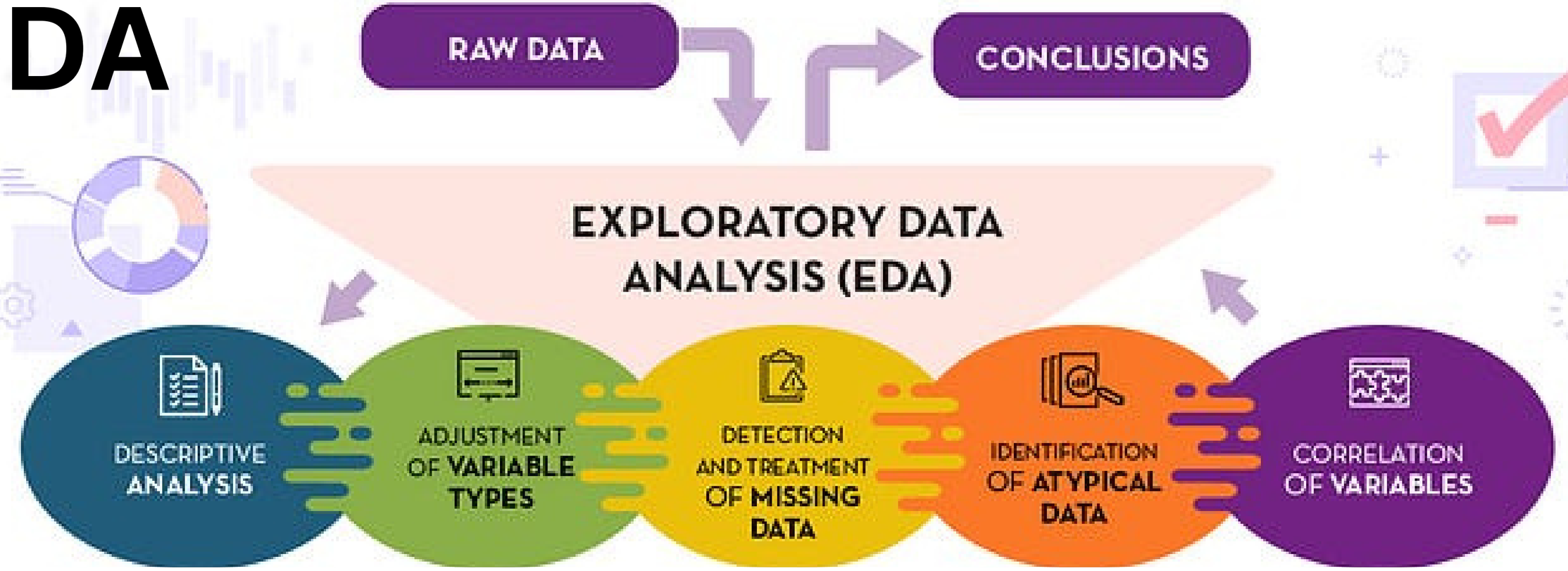
Duplicated Values

Dropped unnecessary features in the data set

One-hot-encoding

Dropping highly correlated features

EDA



Checked for unique features

Performed a categorical and numerical feature analysis

Dealt with outliers

Correlation Heatmap

Pairplots

Scaling

Categorical Features Analysis

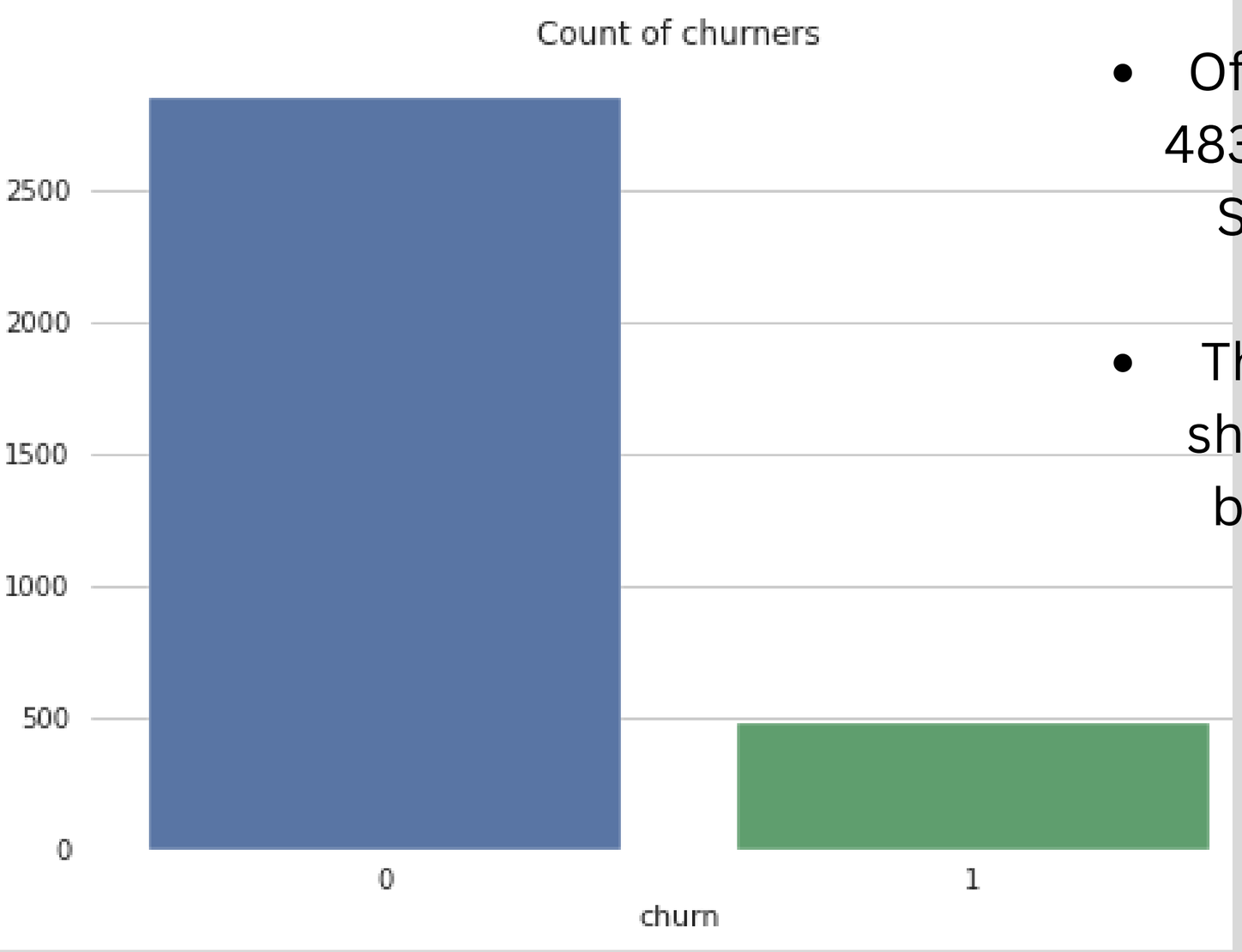
state

area code

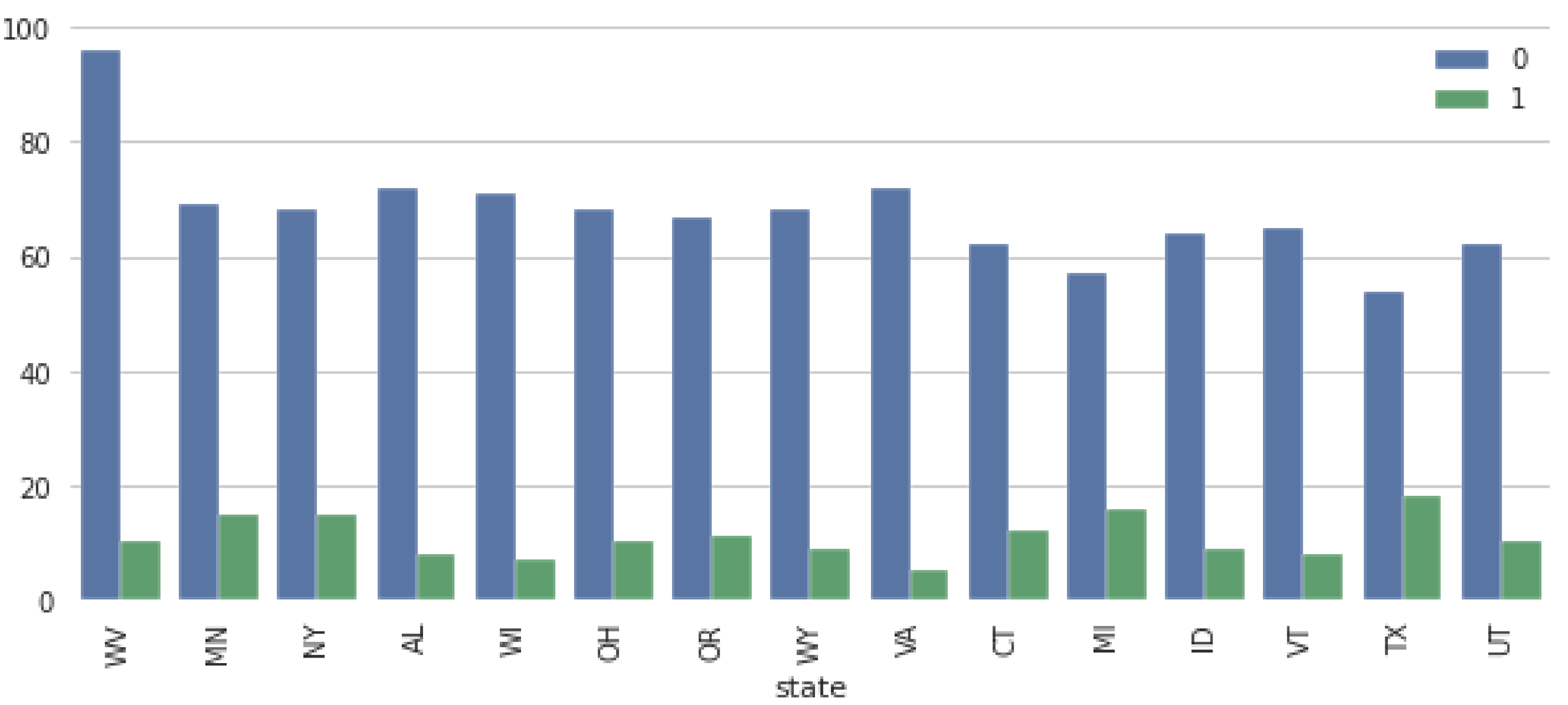
international plan

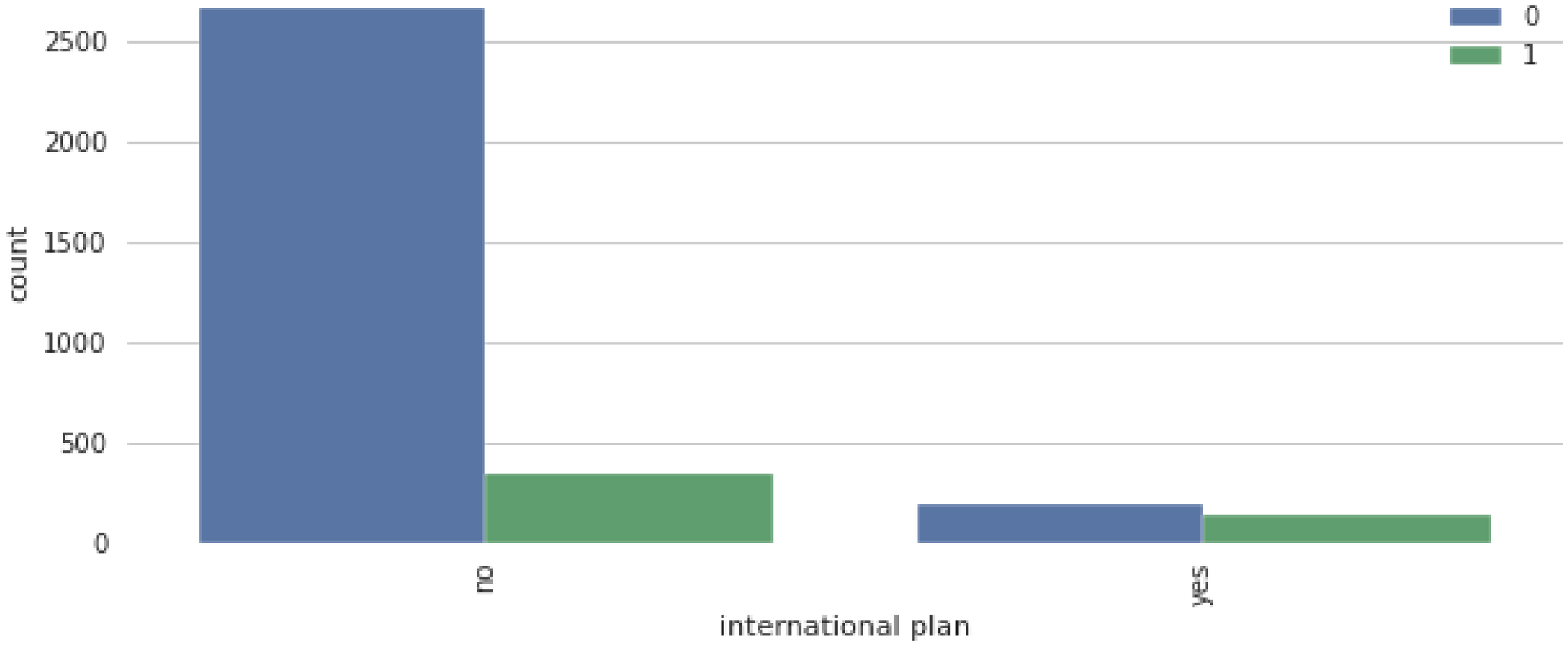
voice mail plan

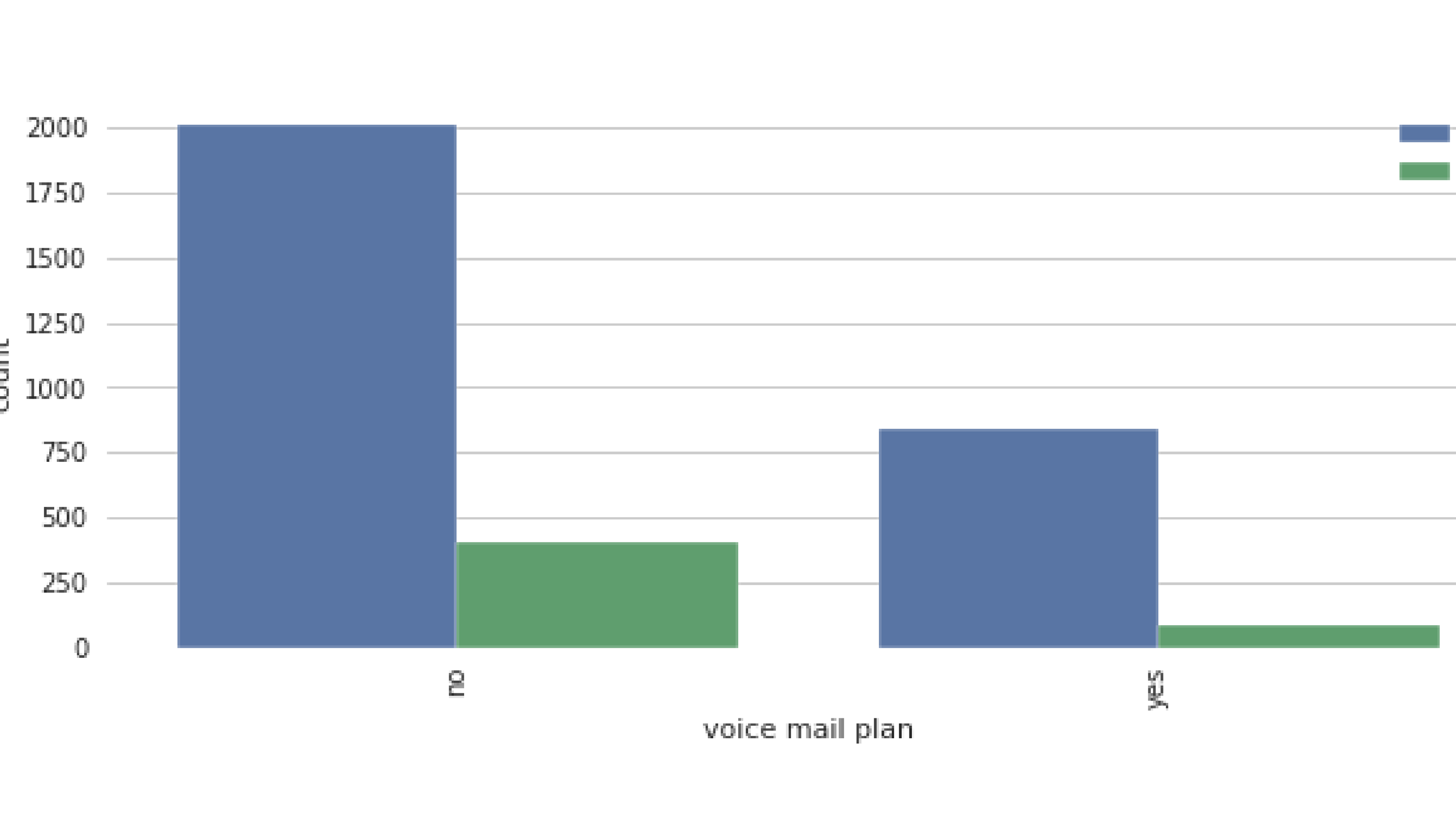
CHURN is our dependent variable we will compare with the categorical feutures.

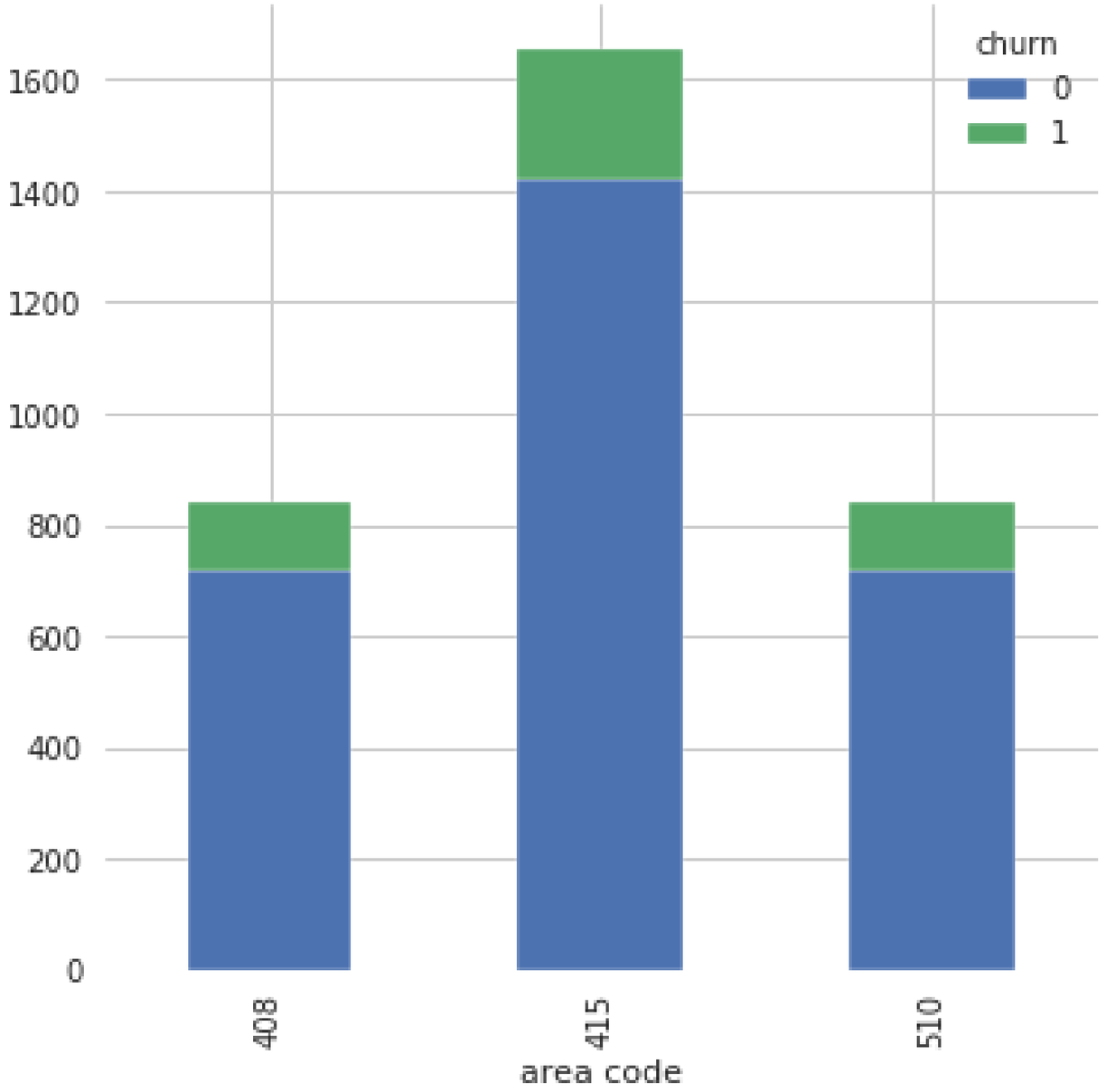


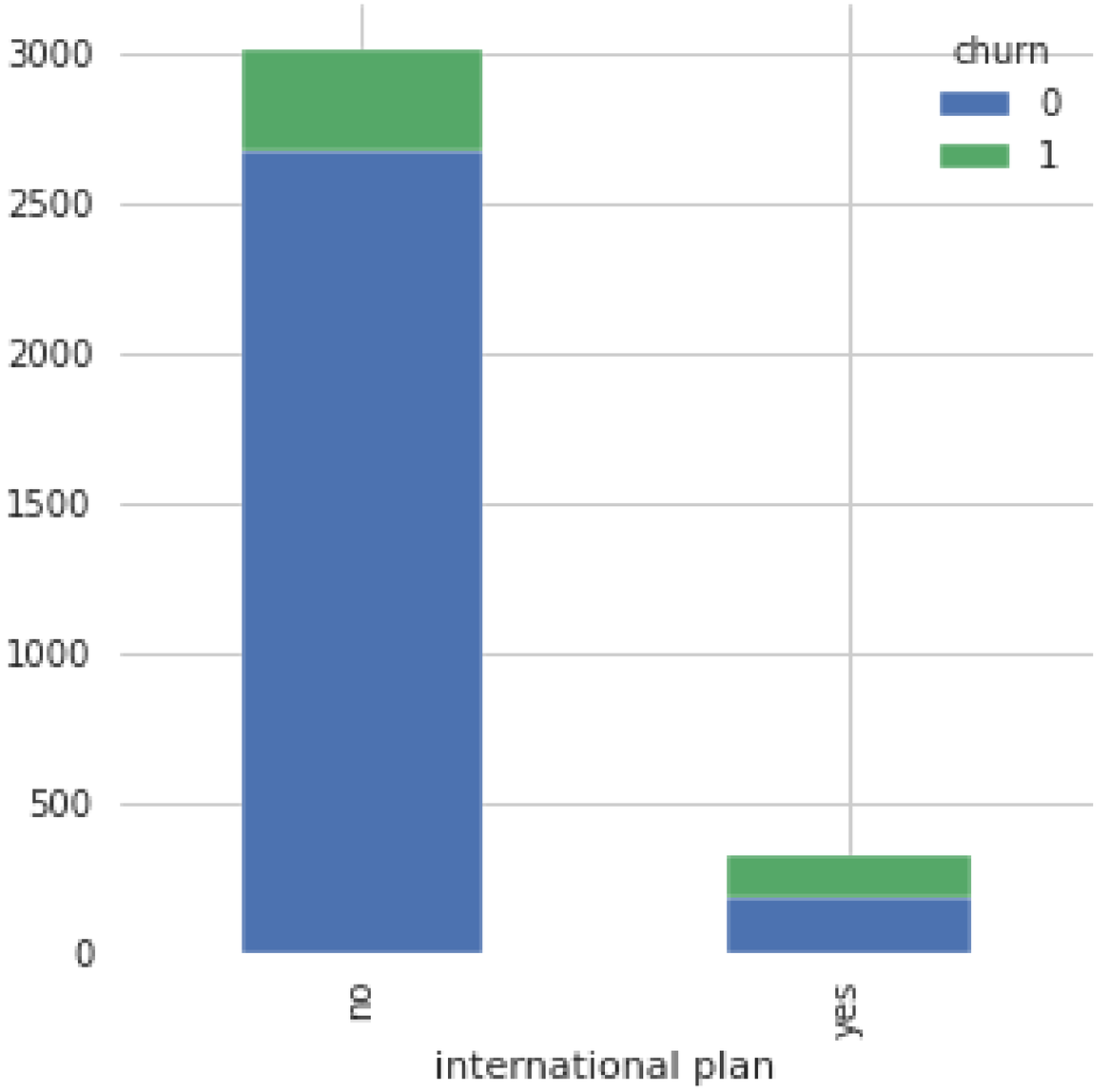
- Of the 3,333 customers in the dataset, 483 have terminated their contract with SyriaTel. That is 14.5% of customers lost.
- The distribution of the binary classes shows a data imbalance. This needs to be addressed before modeling as an unbalanced feature can cause the model to make false predictions.


100
80
60
40
20
0
0
1
WV
MN
NY
AL
WI
OH
OR
WY
VA
CT
MI
ID
VT
TX
UT
state


0
1
2500
2000
1500
1000
500
0
count
no
yes
international plan

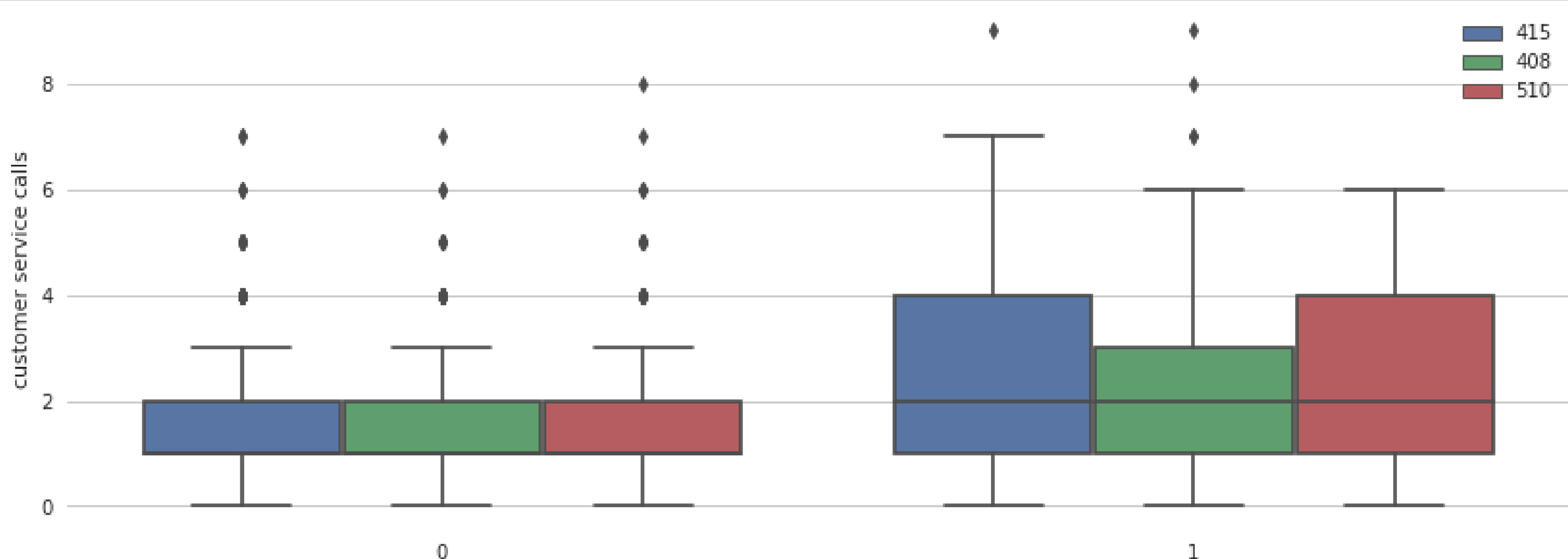

2000
1750
1500
1250
1000
750
500
250
0
count
no
yes
voice mail plan


churn
0
1
1600
1400
1200
1000
800
600
400
200
0
408
415
510
area code

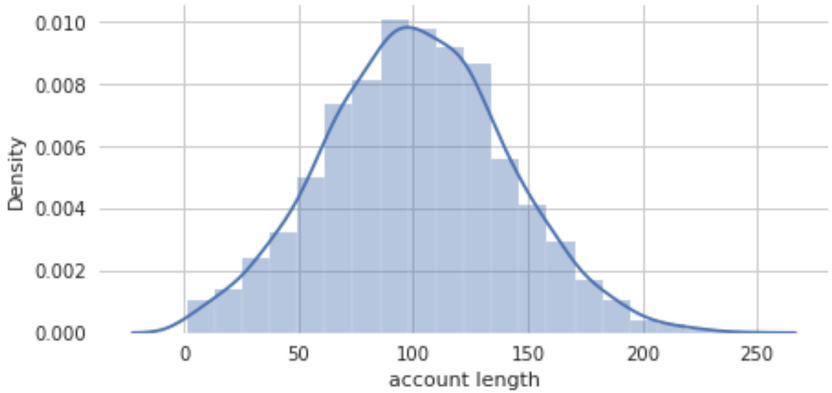

churn
0
1
3000
2500
2000
1500
1000
500
0
no
yes
international plan

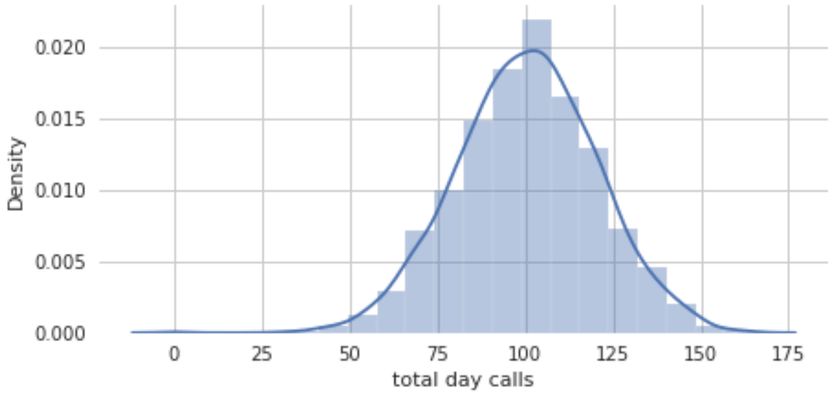
Outliers in numeric features

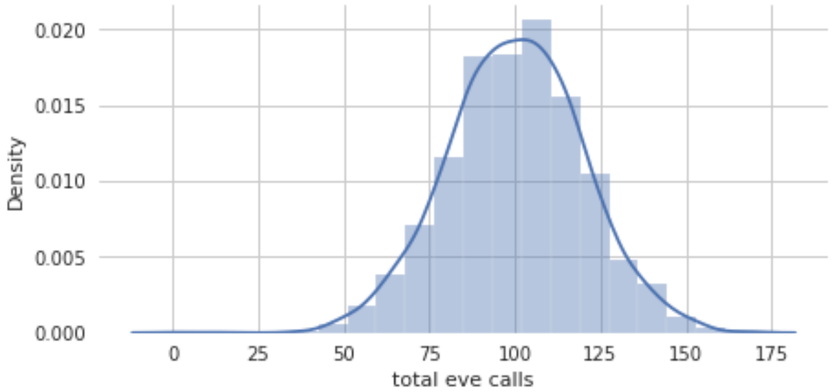
Boxplot of churn vs Area code

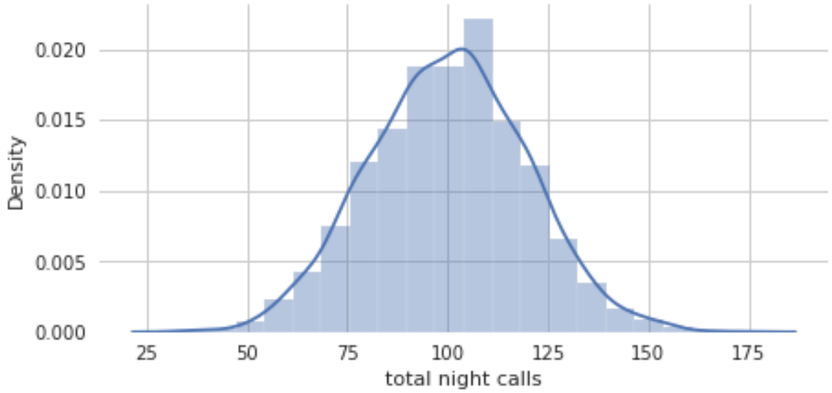


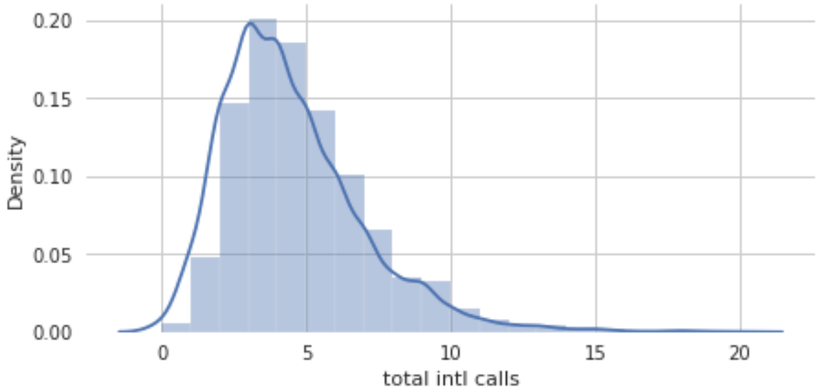
Pairplot distribution for numeric features

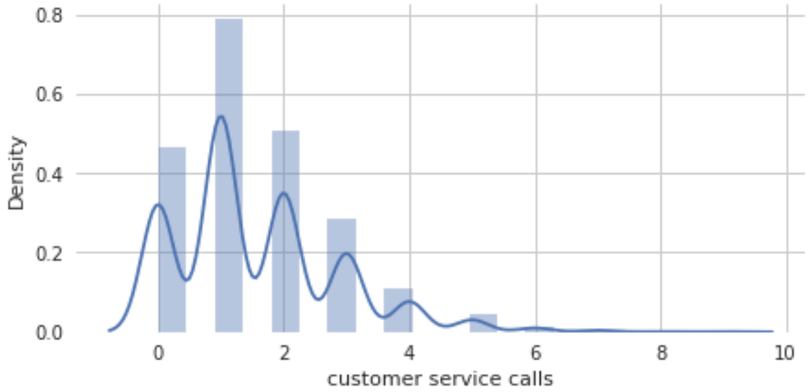






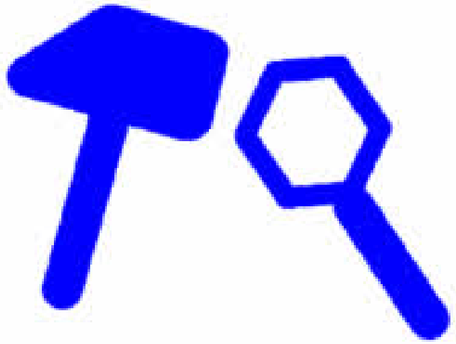


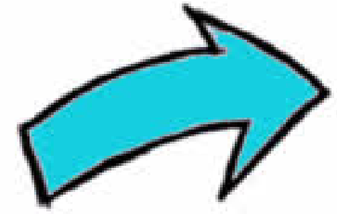


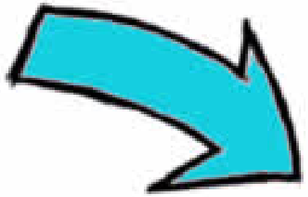


GOOD?

MODEL







Evaluation

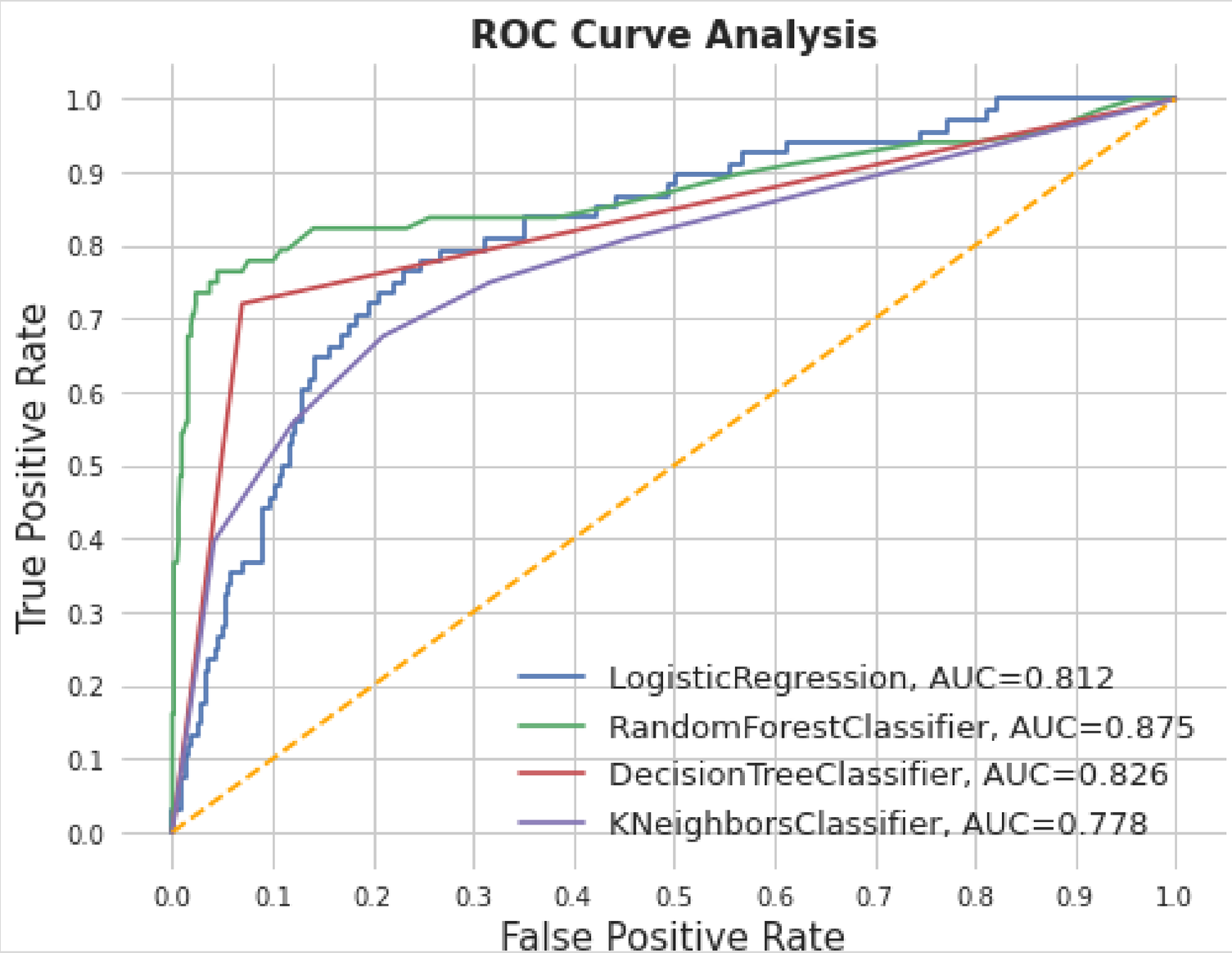
BAD?

EVALUATION Model Comaprison

- Logistic Regression
- Random Forest Classifier
- Decision Tree Classifier
- KNeighboursClassifier

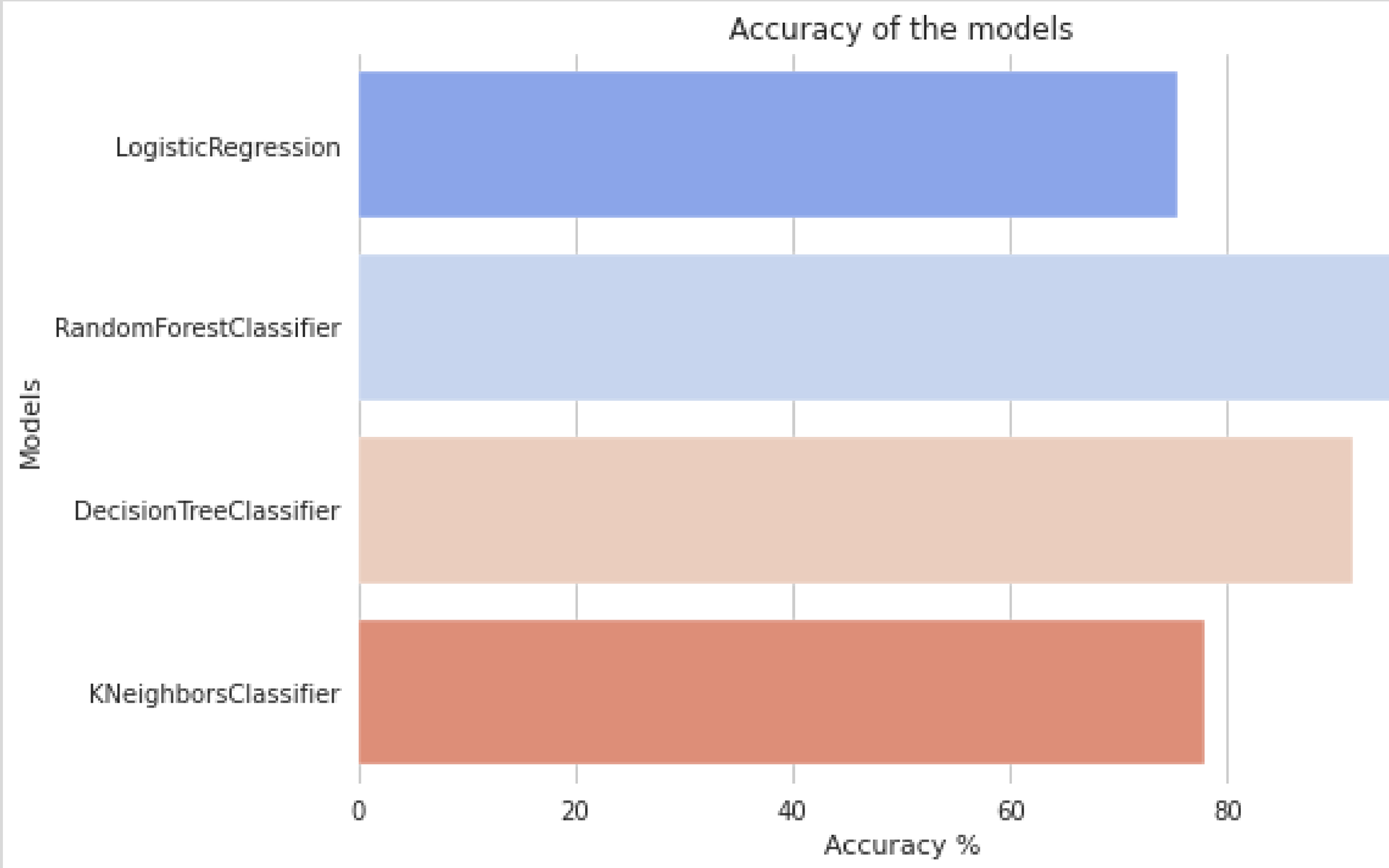
The best is the Random Forest Classifier that hugs the upper left of the graph.

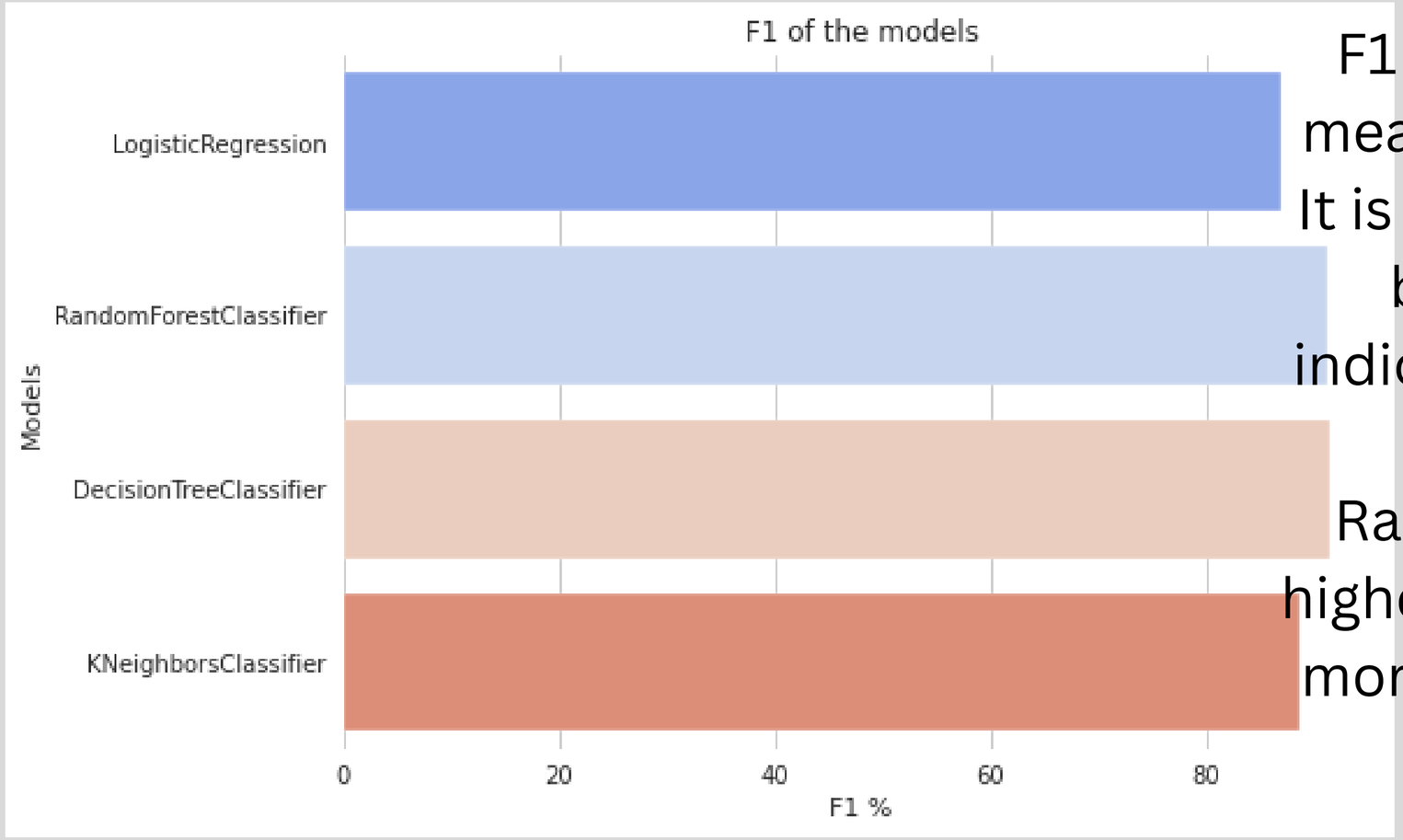
The Roc curve illustrates the true positive against the false positive rate of our classifier.



RANDOM FOREST CLASSIFIER is the best with an accuracy of 91.39%

The worst is Logistic Regression with 75%

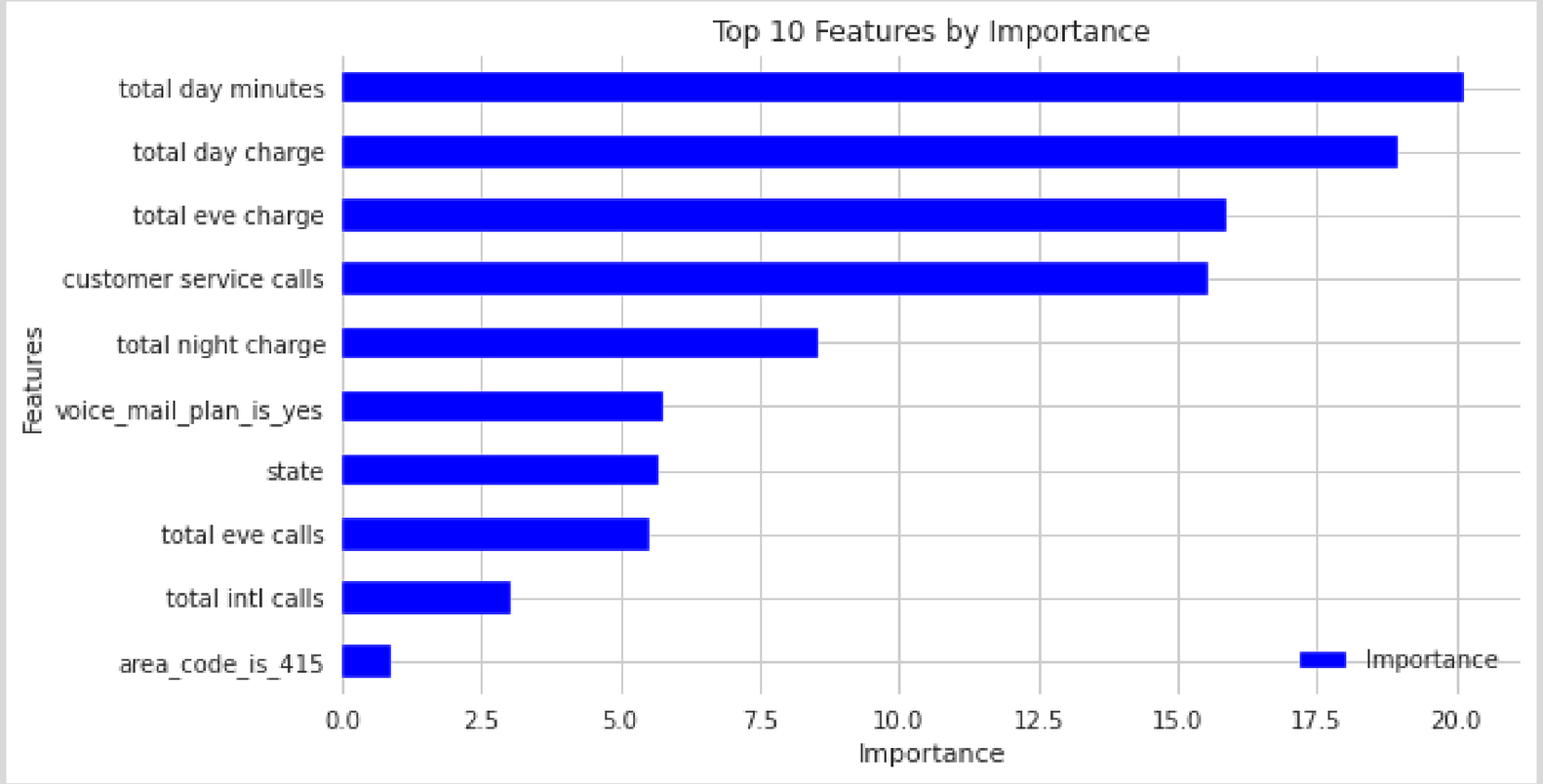




F1 score measures the harmonic mean between precision and recall It is a value between 0 and 1, with 1 being a perfect score and an indication everything was observed correctly.

Random forest classifier had the highest F1 score. false negative have more of a business impact. need to focus on recall

Top 10 features based on importance



Based on the observation we can see the total day minutes is the highest with 20% importance

The least is area code with 0.08%

CONCLUSIONS

- Customers who churn tend to have higher usage during the day and evening, as indicated by the higher values for total day minutes, total day charge, total eve charge, and total eve calls.
- The number of customer service calls is an important indicator of churn, suggesting that customer dissatisfaction or issues may contribute to higher churn rates.
- Offering a voicemail plan may have a positive impact on reducing churn, as customers with a voicemail plan are less likely to churn.
- Geographic location (state) and area code (415 or other) also show some level of influence on churn rates, highlighting the potential impact of regional factors.

RECOMMENDATIONS

1. Firstly, focus on high-value customers by developing tailored plans and promotions to provide cost-effective options based on their usage and charges. Offering personalized incentives can further encourage loyalty and retention among these customers.
2. Secondly, improving customer service is crucial, and Syriatel should enhance processes for prompt issue resolution, reducing the need for customer service calls. Implementing proactive support measures to address customer concerns proactively will also contribute to improved customer satisfaction.
3. Thirdly, analyze factors contributing to high charges during evening and nighttime hours, ensuring that services provided during these periods align with customer expectations and offer value for money.

- Refine the model for improved accuracy and reliability.
- Analyze feature importance to identify key predictors of churn.
- Develop customer segmentation for targeted retention strategies.
- Implement real-time churn monitoring for proactive intervention.
- Monitor customer feedback and satisfaction to address concerns.
- Collaborate with marketing and customer service teams for aligned strategies.
- Evaluate the impact and ROI of churn reduction efforts.

Thank you!

Churn butter,
not customers.

Q&A?:

contact
Stephanie Mbithe
stephmbithe10@gmail.com
https://github.com/Stephie01