TELECOM CHURN ANALYSIS:A SYRIA TEL PERSPECTIVE

BUSINESS UNDESTANDING

- Stakeholder: SyriaTel, a prominent player in the telecommunications sector, is committed to comprehending and addressing the challenge of customer churn to optimize revenue streams and elevate overall customer satisfaction.

- Business Challenge: SyriaTel grapples with a substantial rate of customer churn, where subscribers terminate their services, posing a financial risk. This underscores the necessity of recognizing patterns and elements that contribute to the phenomenon of churn.

- Project Goals:

- Anticipate whether a customer is inclined to churn in the immediate future. Reveal insights actionable for devising targeted retention strategies. Importance: Mitigating customer churn is pivotal for ensuring a resilient and profitable customer base. Proactively addressing the factors influencing churn empowers SyriaTel to introduce customer-centric initiatives, enhance service quality, and cultivate enduring customer loyalty.

- Key Inquiries:

- What are the primary drivers of customer churn within the telecom industry? Can we effectively predict customers with the highest likelihood of churning? How can the findings from the analysis guide the implementation of personalized retention strategies? Performance Metrics:

- Accuracy in predicting churn/non-churn. Precision and recall metrics to strike a balance between false positives and false negatives. Evaluation of feature importance to pinpoint crucial factors contributing to churn.

Steps followed

- Data analysis
- Model building and evaluation
- Hyper parameter tuning
- Conclusion
- Recommendation

Data Features

- state: state where a customer lives

- account length: the number of days the customer has had an account

- area code: the area code of the customer

- phone number: the phone number of the customer

- international plan: true if the customer has the international plan, otherwise false

- voice mail plan: true if the customer has the voice mail plan, otherwise false

- number vmail messages: the number of voicemails the customer has sent

- total day minutes: total number of minutes the customer has been in calls during the day

- total day calls: total number of calls the user has done during the day

- total day charge: total amount of money the customer was charged by the Telecom company for calls during the day

- total eve minutes: total number of minutes the customer has been in calls during the evening

- total eve calls: total number of calls the customer has done during the evening

- total eve charge: total amount of money the customer was charged by the Telecom company for calls during the evening

- total night minutes: total number of minutes the customer has been in calls during the night

- total night calls: total number of calls the customer has done during the night

- total night charge: total amount of money the customer was charged by the Telecom company for calls during the night

- total intl minutes: total number of minutes the user has been in international calls

- total intl calls: total number of international calls the customer has done

- total intl charge: total amount of money the customer was charged by the Telecom company for international calls

- customer service calls: number of calls the customer has made to customer service

- churn: true if the customer terminated their contract, otherwise false

# Continuation of the data features

- total eve minutes: total number of minutes the customer has been in calls during the evening
- total eve calls: total number of calls the customer has done during the evening
- total eve charge: total amount of money the customer was charged by the Telecom company for calls during the evening
- total night minutes: total number of minutes the customer has been in calls during the night
- total night calls: total number of calls the customer has done during the night
- total night charge: total amount of money the customer was charged by the Telecom company for calls during the night
- total intl minutes: total number of minutes the user has been in international calls
- total intl calls: total number of international calls the customer has done
- total intl charge: total amount of money the customer was charged by the Telecom company for international calls
- customer service calls: number of calls the customer has made to customer service
- churn: true if the customer terminated their contract, otherwise false
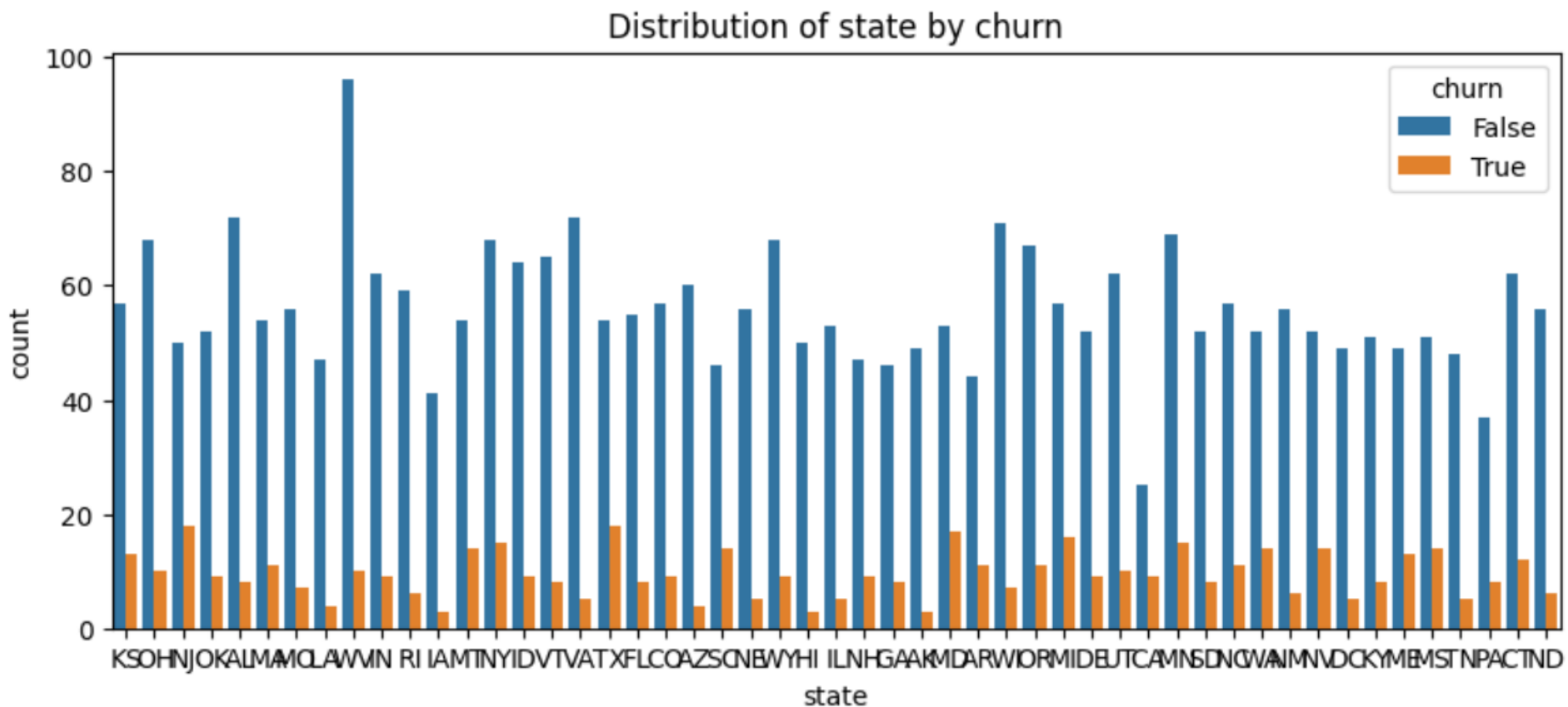
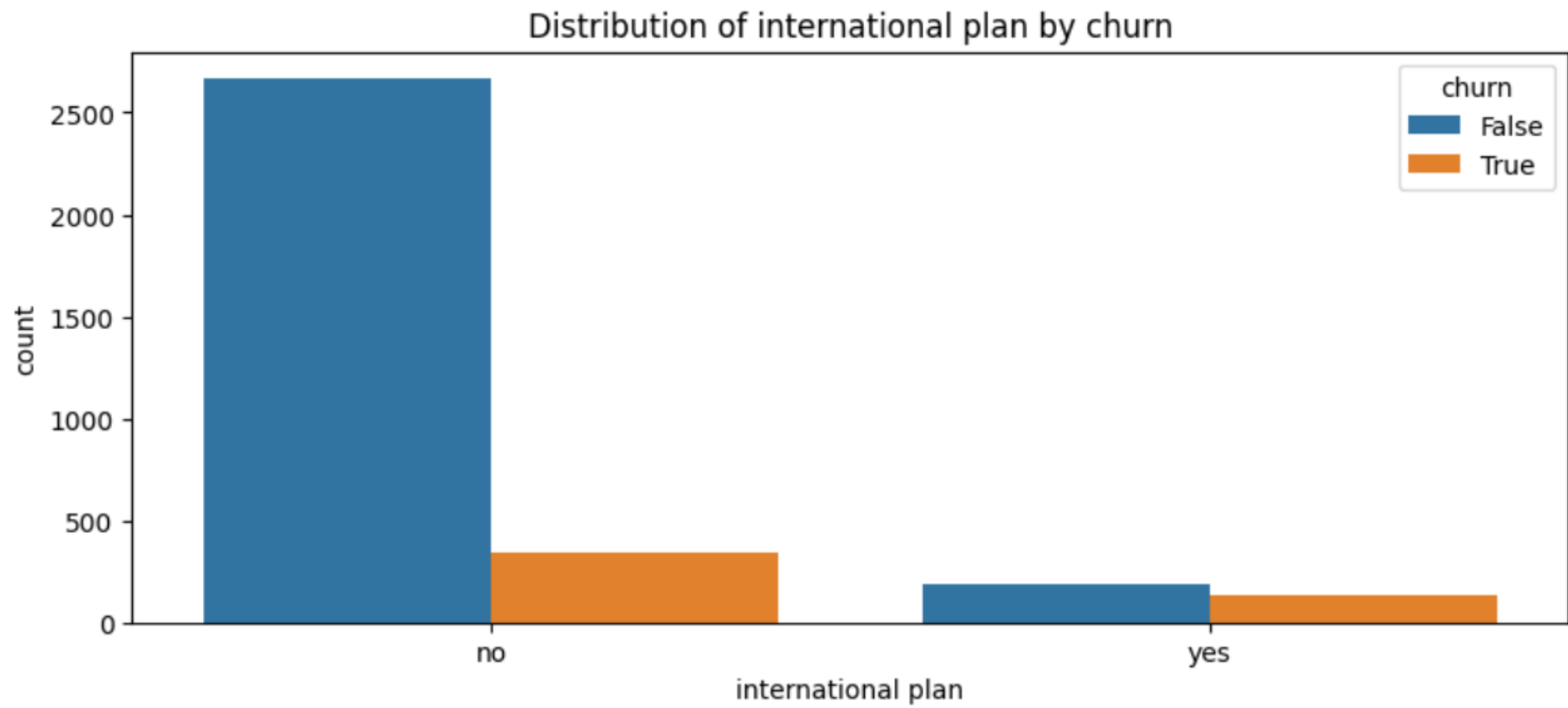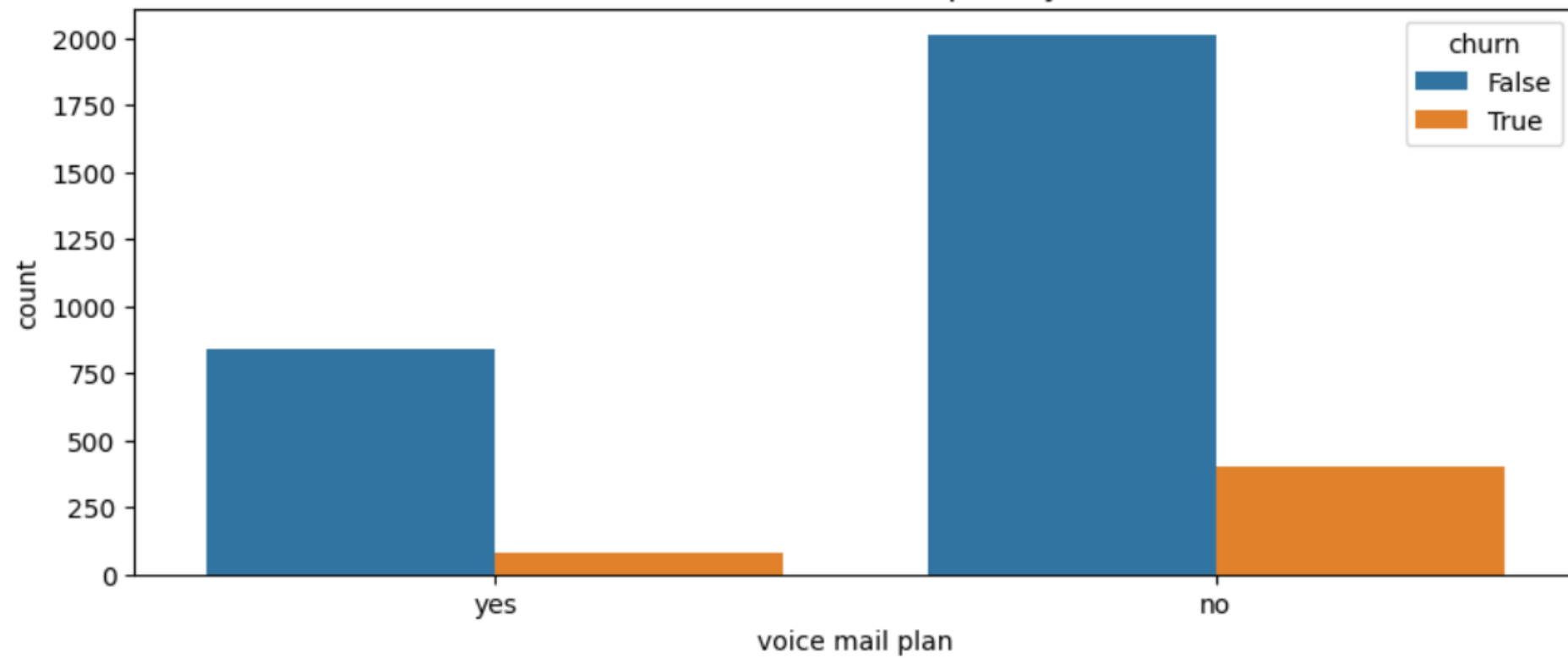# DATA ANALYSIS

Distribution of churn
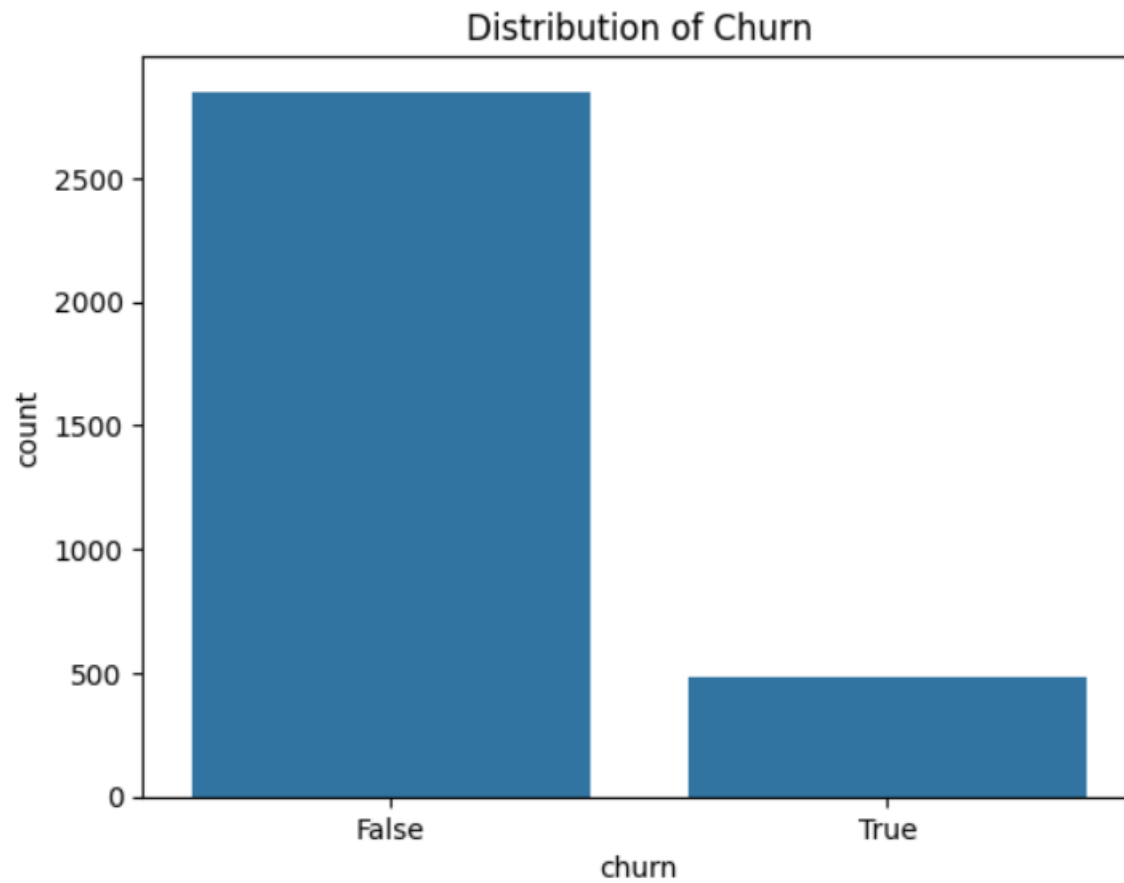
Analysis on "area code"
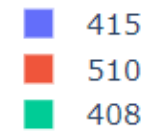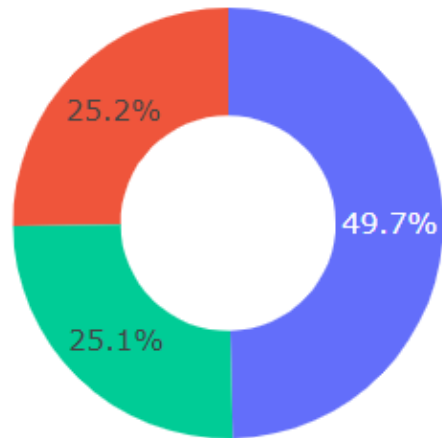
Analysis of the feature "churn"

Distribution of international plan by churn

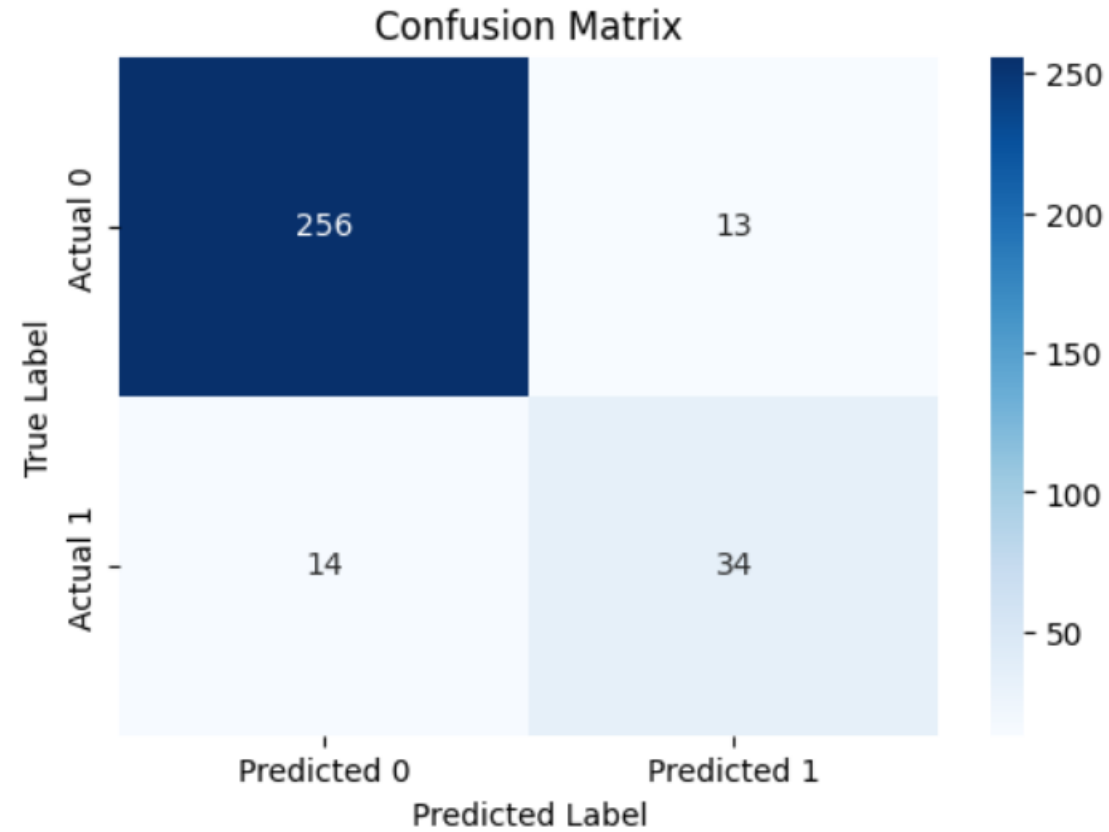Distribution of voice mail plan by churn

MODEL BUILDING & EVALUATION

- dt: Decision Tree Classifier
- knn: K-neighbors Classifier
- rf: Random Forest Classifier

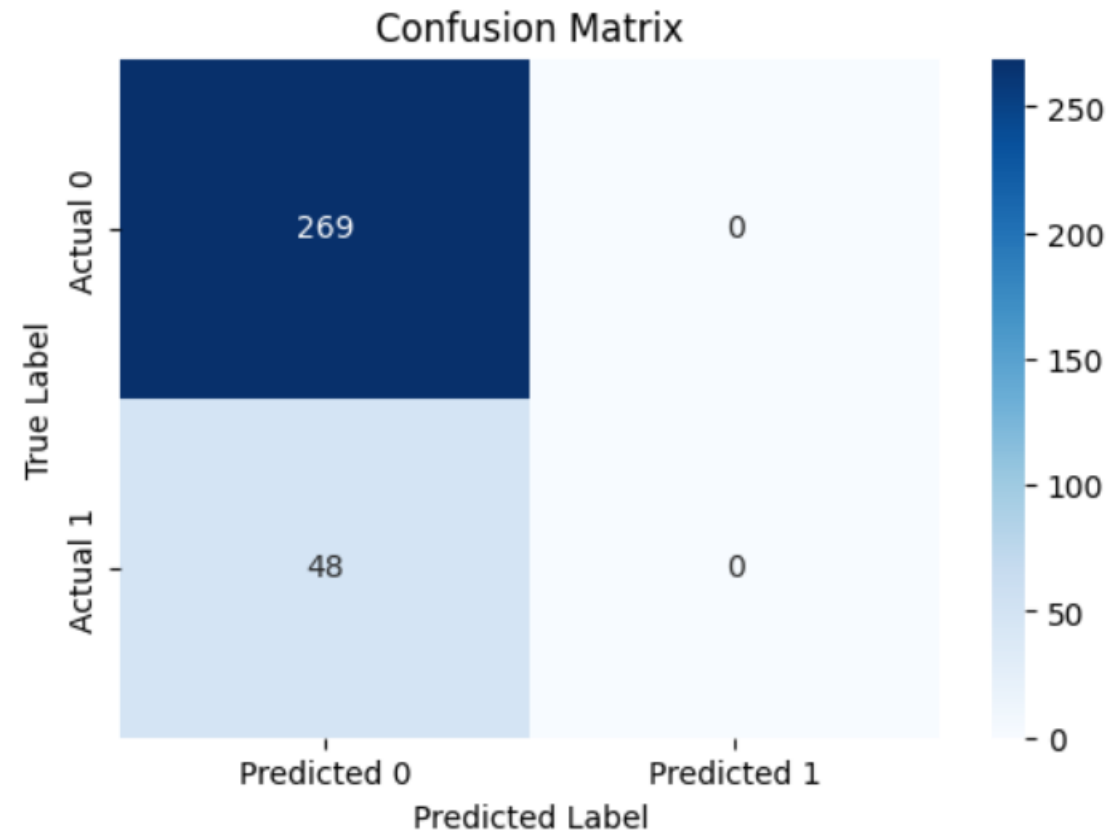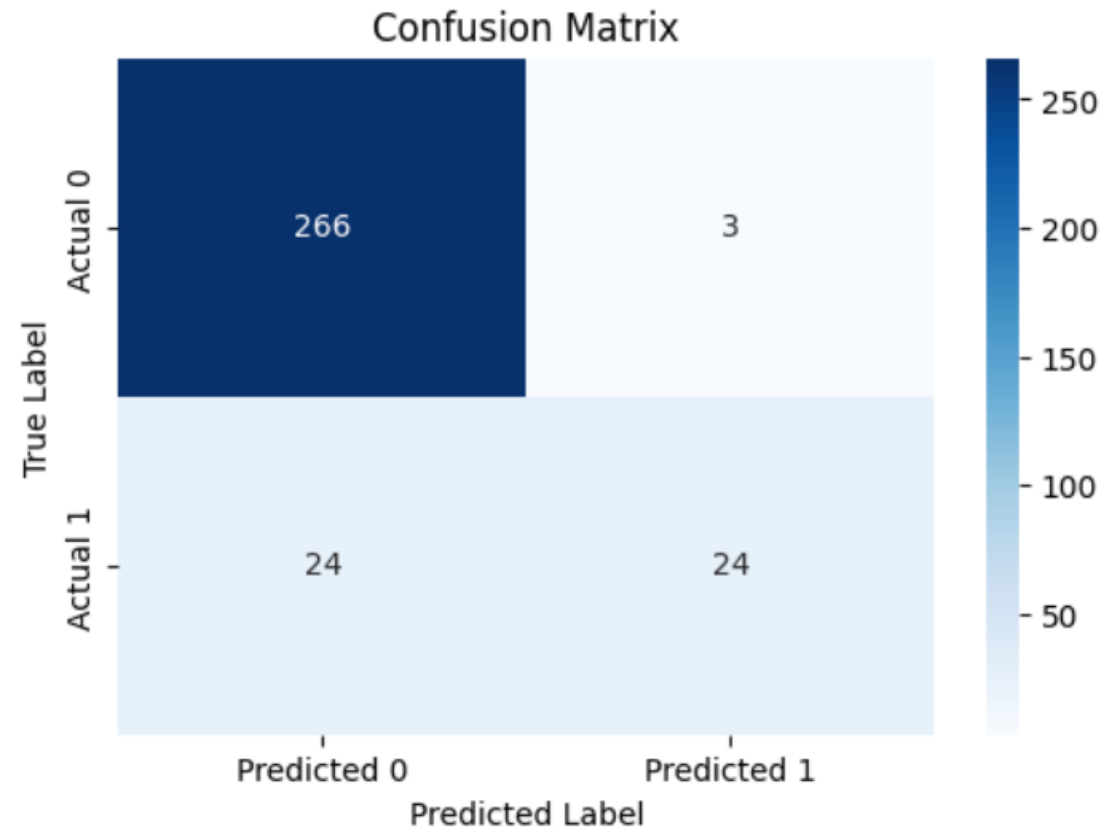# Confusion matrix for dt

Confusion matrix for knn

Confusion matrix for rf

```
Evaluation for dt:
Accuracy: 0.9148
ROC-AUC: 0.8300


Evaluation for knn:
Accuracy: 0.8486
ROC-AUC: 0.5000


Evaluation for rf:
Accuracy: 0.9148
ROC-AUC: 0.7444
```

Interpretation:

- Decision Tree (dt):The model achieved a high accuracy of 91.80%, indicating that it correctly classified a large portion of the instances.The ROC-AUC score of 83.19% suggests good discrimination performance.K-Nearest Neighbors (knn):The accuracy is 84.86%, indicating reasonable performance, but it's lower than the Decision Tree.The ROC-AUC score is 50.00%, suggesting that the model's discrimination performance is close to random chance. This might indicate an issue with the chosen k value or the suitability of the KNN algorithm for your dataset.Random Forest (rf):The model achieved a high accuracy of 93.06%, which is the highest among the three models.The ROC-AUC score of 79.65% indicates good discrimination performance but is slightly lower than the Decision Tree.

HYPERPARAMETER TUNING

```
Classification Report:
              precision    recall  f1-score   support

           0       0.93      0.99      0.96       269
           1       0.93      0.58      0.72        48

    accuracy                           0.93       317
   macro avg       0.93      0.79      0.84       317
weighted avg       0.93      0.93      0.92       317
```

Interpretation

- The classification report shows that the model performs well on both classes, with high precision, recall, and F1-score for class 0. The performance for class 1 is also reasonable.

- The overall accuracy of 93% on the test set is a positive outcome, indicating the model's effectiveness in making correct predictions.

Cross validation

```
Cross-Validation Results (Accuracy): [0.93345009 0.92994746 0.91578947 0.93333333 0.92280702]
Mean Accuracy: 0.9270654745445048
Standard Deviation: 0.006835756465122311
```

Interpretation

The high mean accuracy and low standard deviation are positive indicators that the Random Forest model is performing well and consistently across different subsets of the training data.

The standard deviation provides an idea of how much the model's performance varies between folds. In this case, the low standard deviation suggests a stable and consistent model.

Feature Importance Analysis

Feature Importances in Random Forest Model
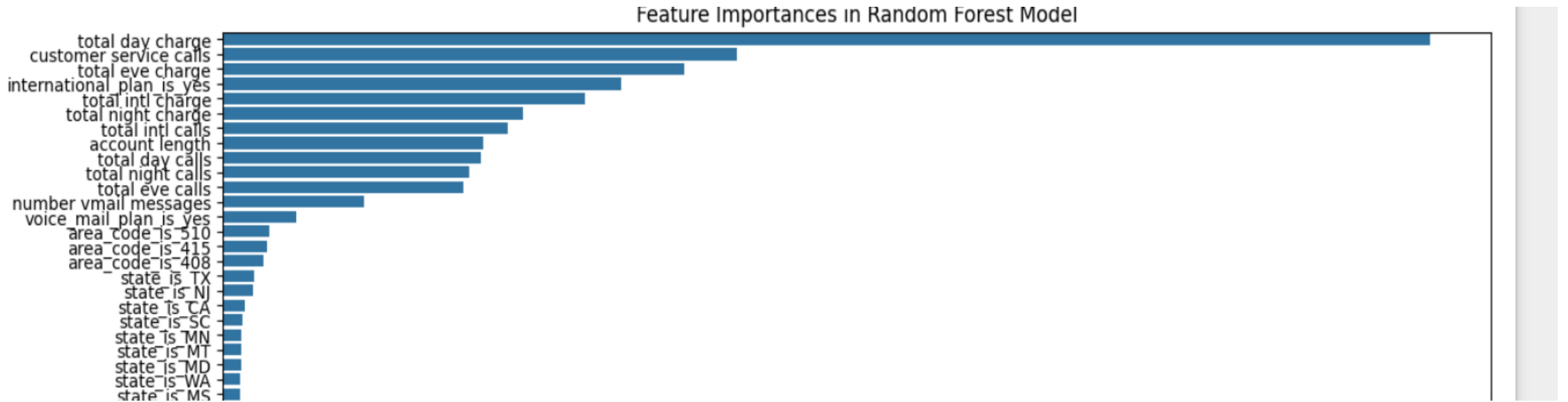
Reassessing the performance of a model after addressing class imbalance

```
Accuracy: 0.973186119873817
Precision: 1.0
Recall: 0.82653061224489
F1-Score: 0.9050279329608939
AUC-ROC: 0.9909952787084984
```

Testing set using Python and scikit-learn

```
Testing Accuracy: 0.9700315457413249
Testing Precision: 0.9876543209876543
Testing Recall: 0.8163265306122449
Testing F1-Score: 0.893854748603352
Testing AUC-ROC: 0.9897578434358819
```

Conclusion

- The Random Forest model appears to excel in terms of accuracy and F1-score when compared to the other two models. It is crucial to consider both precision and recall, especially in accordance with the specific requirements of the business problem at hand. The model demonstrates robust adaptability to fresh, unseen data, evident in its elevated testing accuracy, precision, recall, and AUC-ROC. It maintains a notable precision level, minimizing false positives—a critical aspect for SyriaTel in mitigating the financial risks linked to customer churn. Additionally, the model exhibits strong recall, effectively identifying a substantial portion of genuine churn cases. The F1-Score contributes to a well-balanced perspective, considering both precision and recall. In conclusion, the machine learning model, based on the testing outcomes, appears to adeptly address SyriaTel's business challenge regarding customer churn

Recommendation

- Explore the option of deploying the model in a live environment for immediate predictions. Establish monitoring systems to continuously assess the model's performance over the course of time. Engage with stakeholders to incorporate the model's predictions into focused retention strategies. Persist in refining and enhancing the model based on continuous feedback and evolving trends in the telecommunications industry.