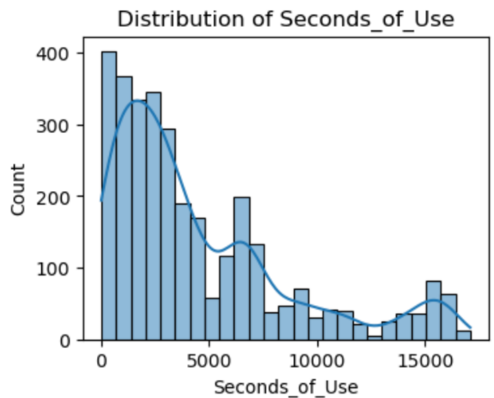
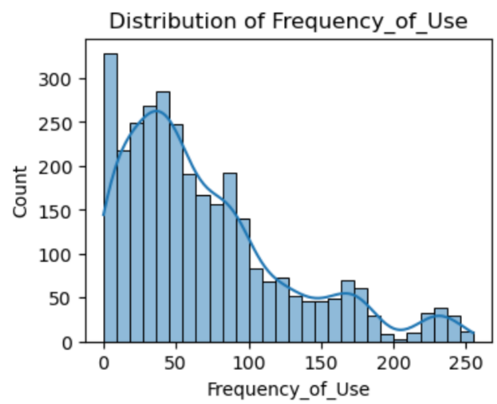
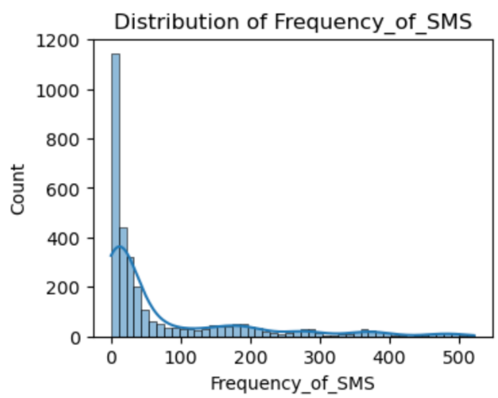
Capstone Contact Center Churn

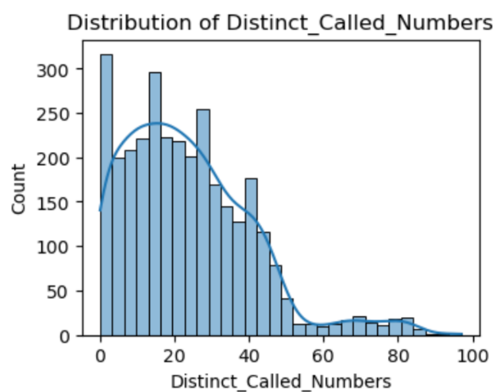
Harish Laxmi Narasimha Venugopal

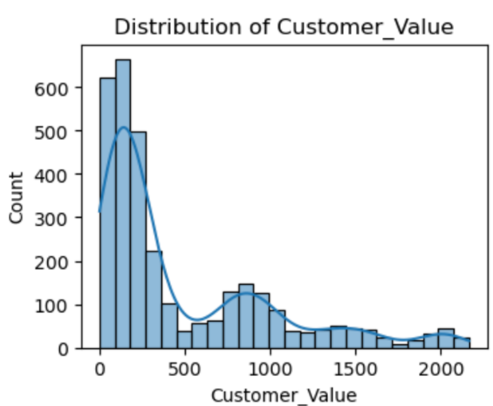
Exploratory Data Analysis (EDA)
Histograms for numerical columns

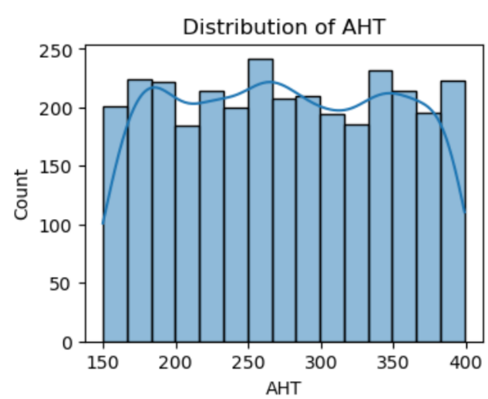


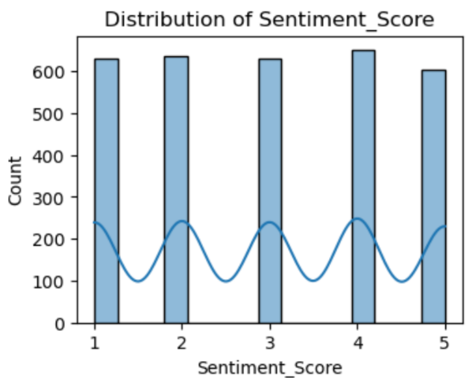


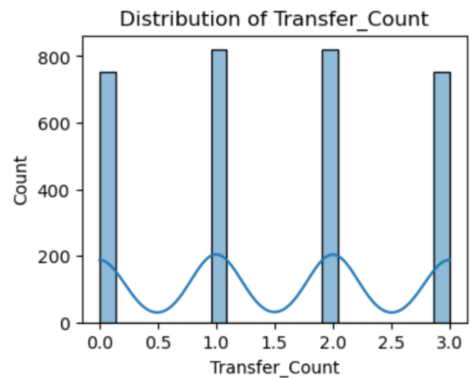


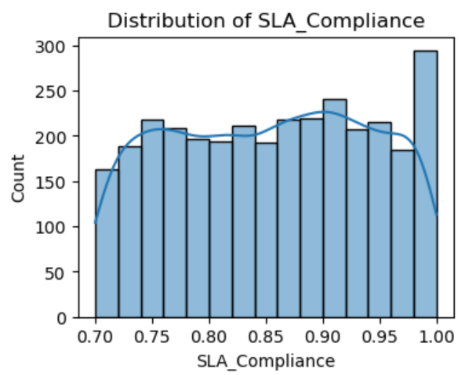


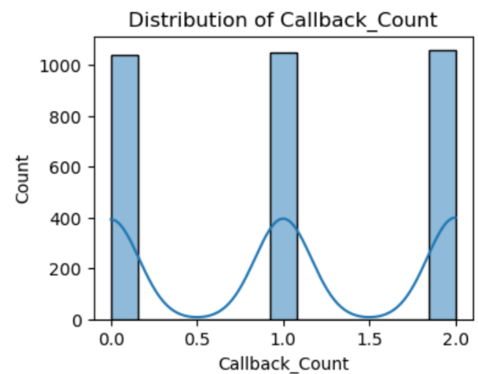












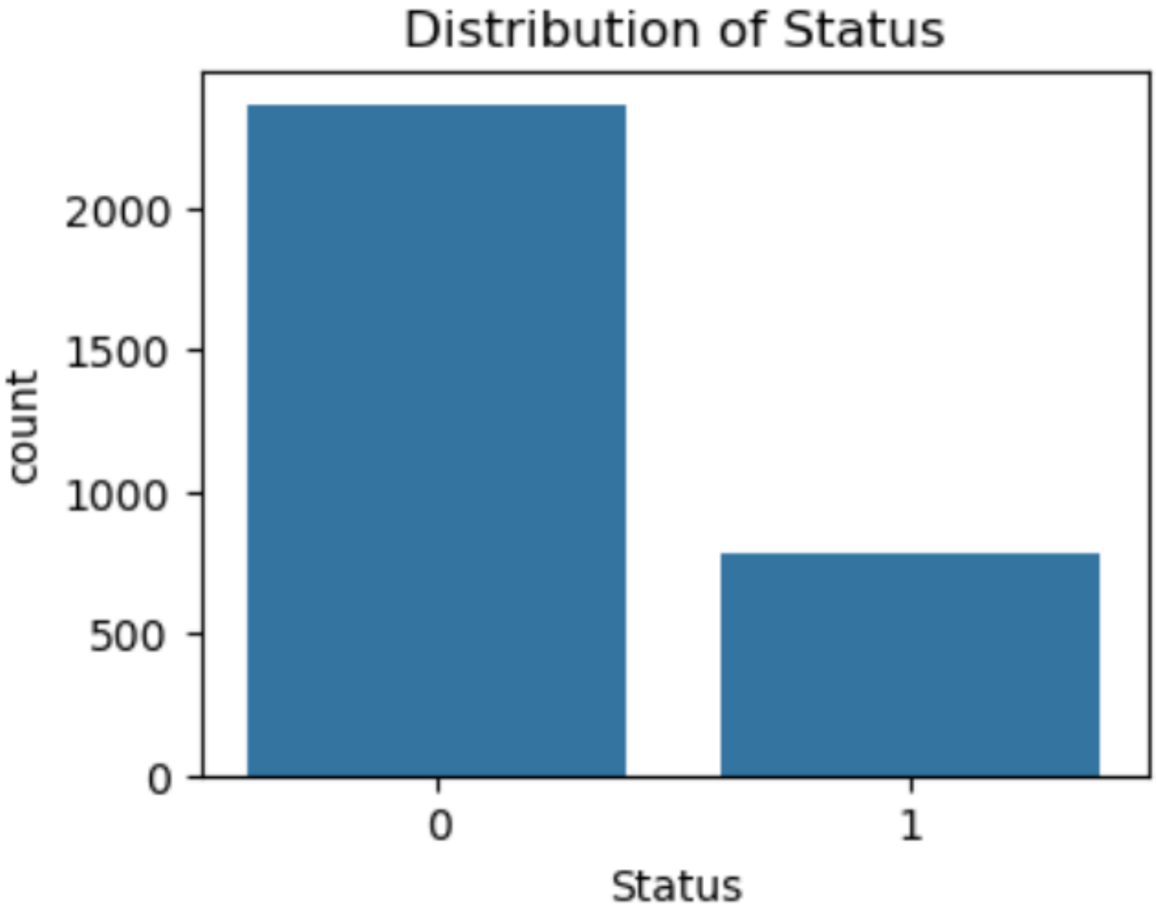
Leveraged the above plots to gain valuable insights into customer behavior and identify areas for improvement in your contact center operations

- **Seconds of use** A large number of users have very low usage (close to zero seconds). This suggests a segment of users who might be inactive, have just signed up, or are not heavily utilizing the contact center services.
- There's a gradual decrease in the number of users as the **frequency of use** increases, with some small spikes along the way
- **Frequency of SMS** Heavily right-skewed, indicating that most customers send a low number of SMS messages.
- **Distribution of Distinct called numbers** - This may be a baseline for normal calling behavior
- **Distribution of customer value** most customers have relatively low value, while a smaller group of high-value customers contributes disproportionately to revenue. Prioritize retention efforts on the high-value segment. Investigate the characteristics of the high-value segment.
- **Distribution of Average Handling Time** - AHT is fairly consistent across the customer base., meaning contact center agents are doing a great job at handling customer queries within the organizational AHT standards.
- **Distribution of Sentiment score** – Showcase scores ranging from 1 to 5., indicating customer spread with respect to sentiment scores. Will need to analyze

factors driving high and low sentiment scores. Correlate sentiment with other variables (e.g., AHT, call reason) to identify areas for improvement

- **Distribution of transfer count** - Many calls are resolved without a transfer, but a significant number of calls require 1, 2, or even 3 transfers. Minimize transfers by improving agent training, providing better knowledge base tools, and optimizing call routing. High transfer counts often indicate customer frustration., leading to a churn.
- **Distribution of SLA compliance** - Roughly uniform distribution with some clustering around certain compliance scores. Need to identify why there are differences and address the root causes of non-compliance to ensure consistent service levels.
- **Distribution of Callback_Count** - Many calls are resolved without a callback, some resolved with 1 or 2. Improve first-call resolution to reduce the need for callbacks. Analyze the reasons for callbacks. This will be a recommendation to business.

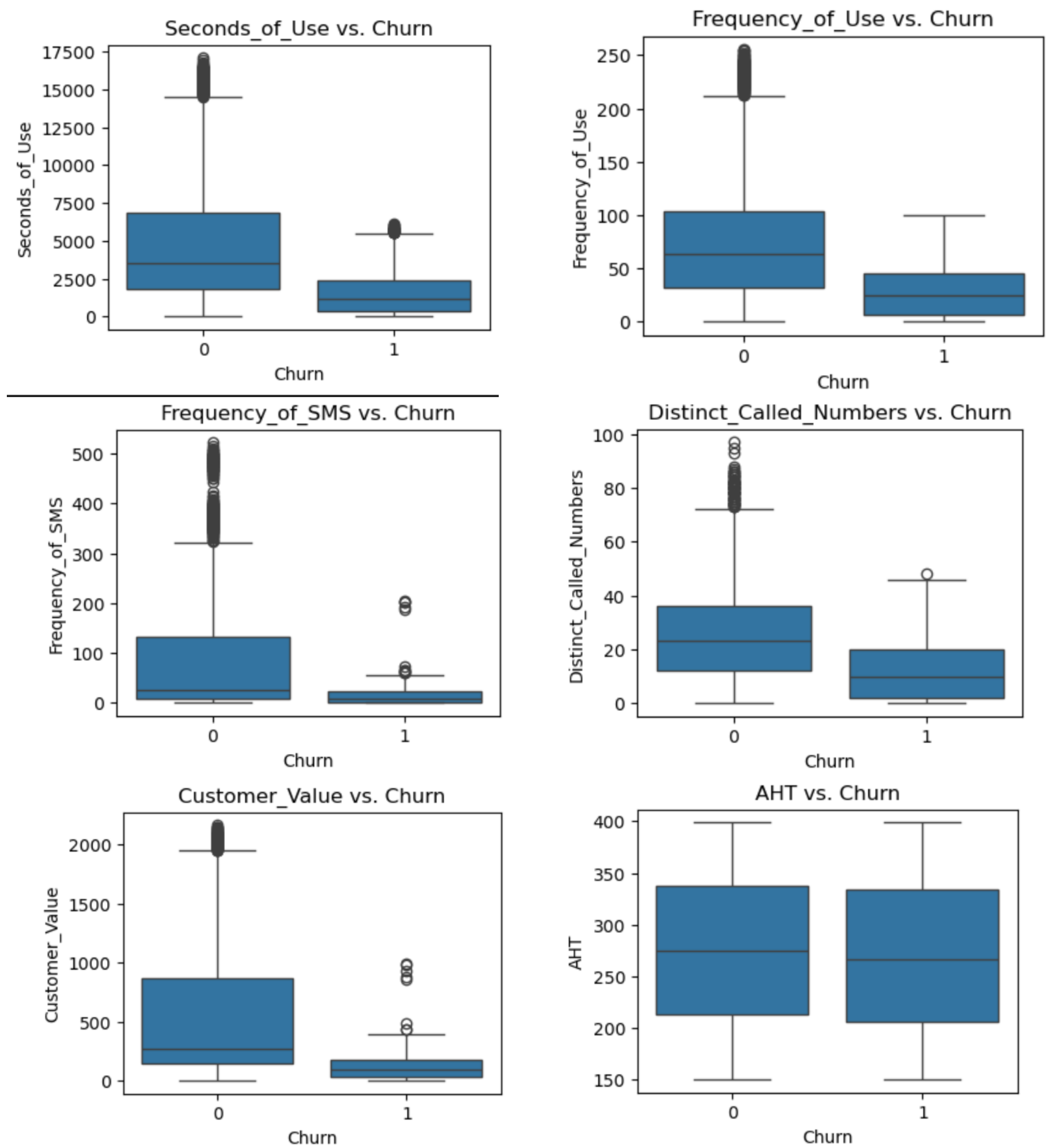
Exploratory Data Analysis (EDA)
Bar charts for Categorical columns – after encoding

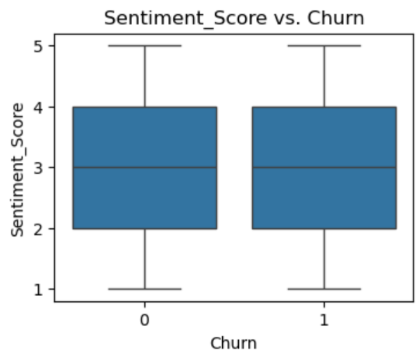


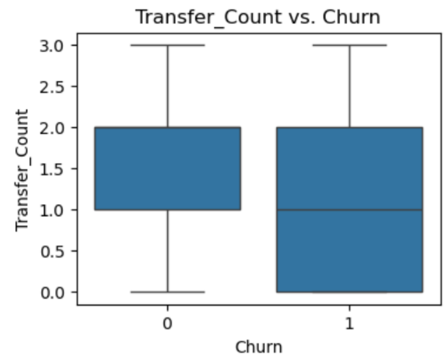
- Status = 0: Approximately 2,300 customers (roughly 75% of the dataset)
- Status = 1: Approximately 800 customers (roughly 25% of the dataset)
- Status = 1 represents customers on basic software subscription plans
- Status = 0 represents customers with enhanced features, or premium

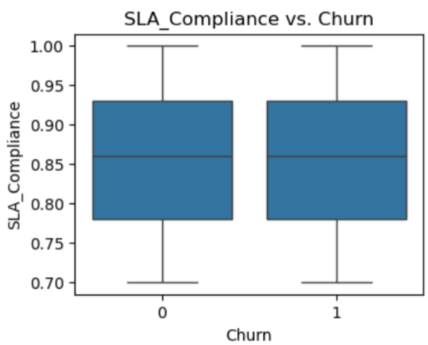
- "Status" is a categorical variable in the dataset.
- The distribution of "Status" is imbalanced, with more customers having Status = 0 than Status = 1.
- **Customers with Status = 1 (basic software subscription) have a higher churn rate compared to customers with Status = 0.**
- **It would be good to target the customers with basic subscription plans with value added feature capabilities to uplift them to premium subscription with value added features.**

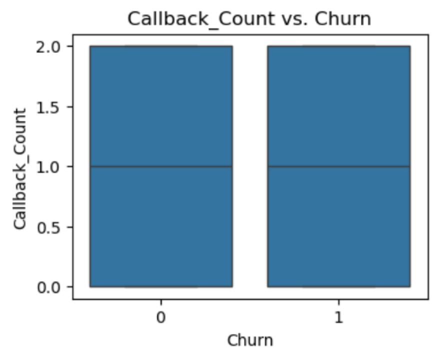
Below box plots help understand **usage-based metrics and customer value** (key indicators of churn risk). By focusing on proactive engagement with at-risk customers, we can potentially reduce churn and improve customer retention











- **Seconds of use vs Churn** - Customers who churned (Churn = 1) generally had significantly lower "Seconds_of_Use" than those who didn't churn (Churn = 0). Low usage is a strong indicator of potential churn. Target these customers with engagement campaigns
- **Frequency of use vs Churn** – Similar to seconds of use vs churn. , customers who had a low frequency of use churned high. Reinforce low usage = high churn %
- **Frequency of SMS usage** – Customers who had a churn didn't use SMS as a communication channel to interact with the contact center either proactively and/or reactively.
- **Distinct Called Numbers vs Churn** – Customers who churned didn't call different distinct called numbers, which means these customers didn't try different product options, explored the product portfolio, features and support to take maximum benefit of the product /subscription.
- C**ustomer value vs Churn** - Customers who churned had a significantly lowe "Customer_Value". Protecting high-value customers should be a priority, and you can identify at-risk customers via the other metrics.

Average handle time (AHT) vs Churn – AHT may not be a key feature/driver for churn. No significant difference between customers who churned vs who didn't.

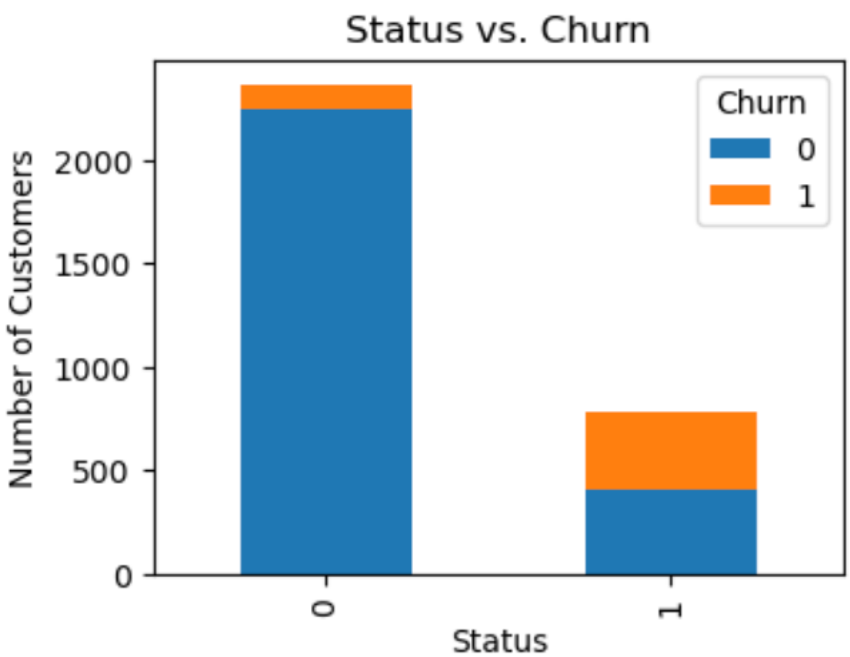
Sentiment score vs churn - overall sentiment may not directly predict churn. Investigate changes in sentiment over time for individual customers. A sudden drop in sentiment could be a red flag.

Transfer count vs churn, SLA compliance vs Churn, Callback count vs Churn - Although the current dataset shows there is little to no difference between customers who churned vs who didn’t. We need to look at this in depth, high % transfers, SLA noncompliance and callback counts on a given customer interaction causes customers to have a bad experience and may move them from positive customers to negative customers leading to a churn. Need additional dataset to evaluate this accurately.

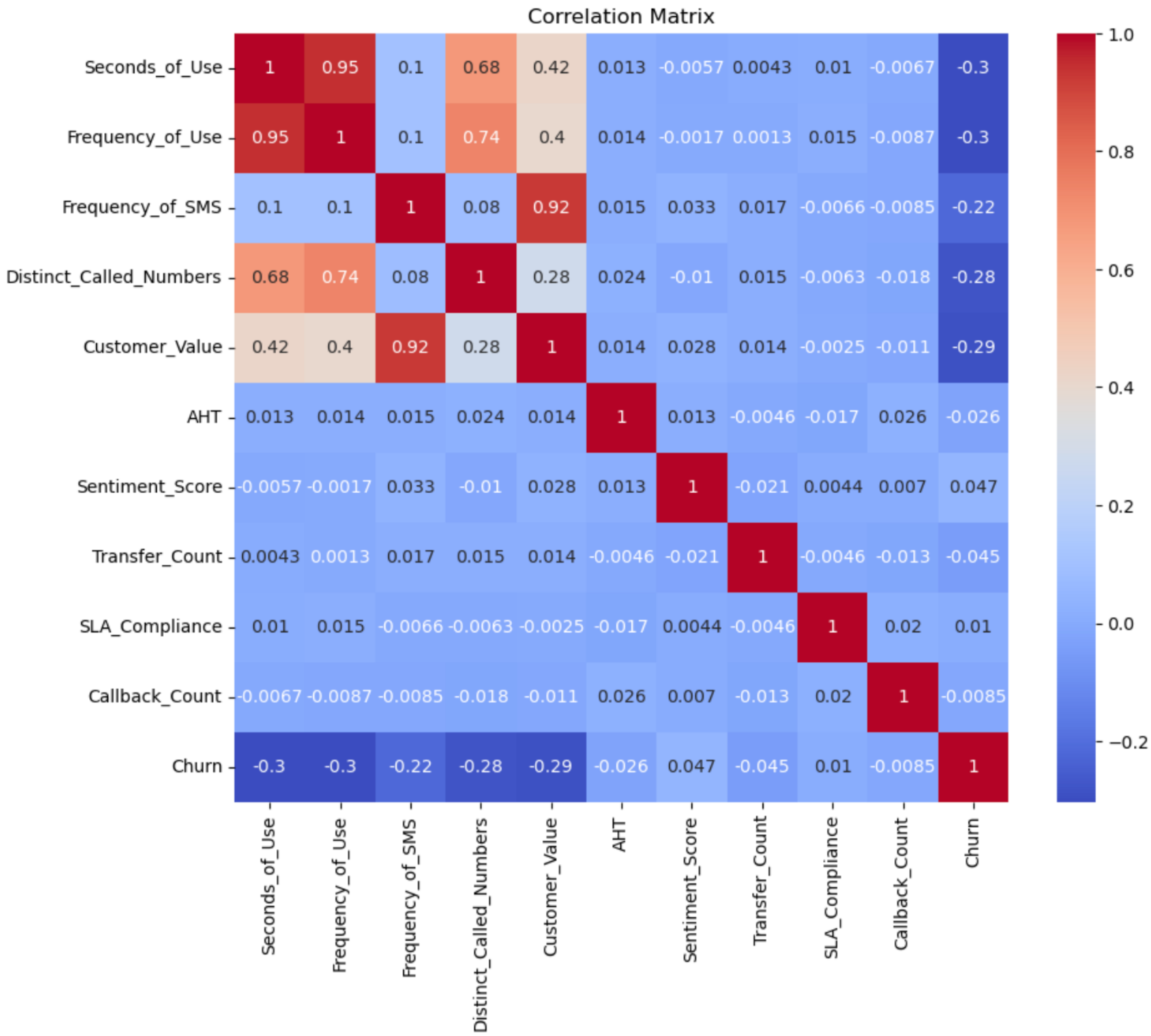
In summary -

- The strongest predictors of churn in these plots are related to usage ("Seconds_of_Use", "Frequency_of_Use", "Frequency_of_SMS", "Distinct_Called_Numbers") and "Customer_Value".
- Target customers with low usage and value with proactive engagement strategies (e.g., personalized offers, onboarding support, highlighting valuable features).
- While important for overall customer service, "AHT", "Sentiment_Score", "Transfer_Count" and "SLA_Compliance" don't appear to be strong indicators of churn in this dataset on their own. Consider investigating trends in sentiment (sudden decreases) and combining these metrics with usage data for a more nuanced analysis.

Churn Risk Assessment – Helps identify business strategies for customer retention



- The stacked bar chart shows two categories of "Status" (0 and 1) and their respective churn behavior (Churn = 0 or 1)
- Customers with **Status** = **1** have a significantly higher proportion of churn (orange segment) compared to customers with **Status** = **0**, where churn is much lower
- **There more customers with "Basic Subscription (base plan)" who would cancel a subscription service than engaged customers with added features who actively use a program**
- The rate of cancelations within the **"Basic Subscription (base plan)"** with "Status=1" is higher.



Correlation matrix shows the correlation coefficients between different variables in the dataset, including Churn. Red indicates positive correlation, blue indicates negative correlation, and the intensity of the color indicates the strength of the relationship

Positively correlated features (higher churn risk):

- Complains (0.53) → Customers with complaints are more likely to churn.
- Status (0.50) → Certain customer statuses may indicate higher churn probability.
- Tariff_1 (0.11) → Some tariff plans may lead to higher churn.

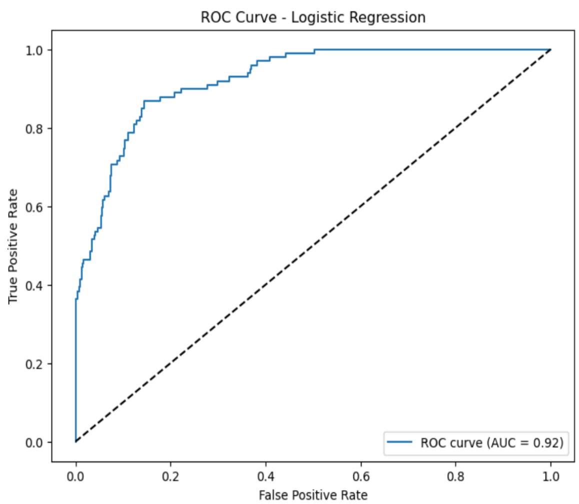
Negatively correlated features (less likely to churn):

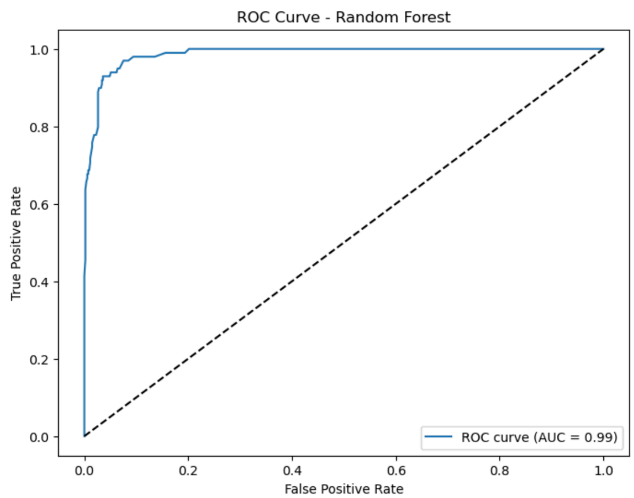
- Customer_Value (-0.29) → High-value customers are less likely to leave.
- Seconds_of_Use (-0.30) → More call usage is associated with retention.

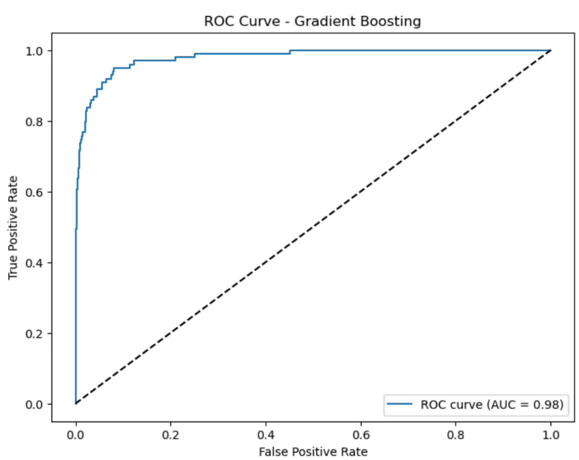
- Frequency_of_SMS (-0.22) → Higher SMS usage means lower churn.
- Subscription_Length (-0.03) → Longer subscription durations correlate with lower churn

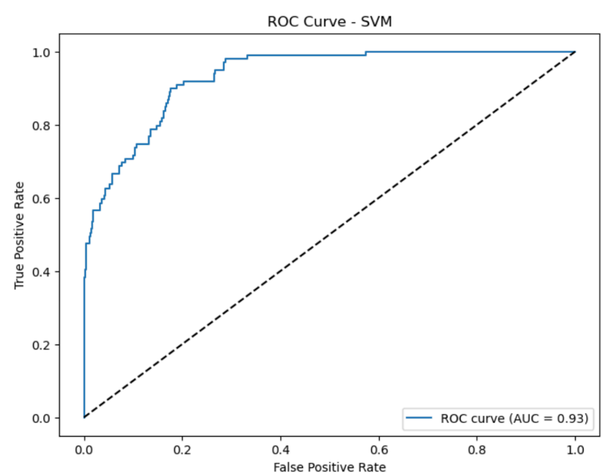
<u>Train and evaluation of each classifier w/ ROC curve plots</u>

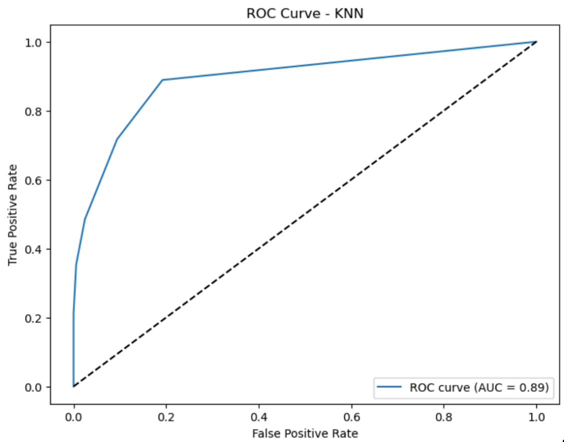
<u>Precision, Recall, F1-score present in the README and Jupyter notebook</u>











- **Random Forest**: Exhibits the highest AUC (0.99), indicating the best performance among the models in distinguishing between customers who will churn and those who will not.
- **Gradient Boosting**: Also shows excellent performance, with an AUC of (0.98).
- **SVM**: Demonstrates good performance as well, achieving an AUC of (0.93)
- **Logistic Regression**: Performs well, with an AUC of 0.92, showing a decent ability to discriminate.
- **KNN**: Shows fairly good performance with an AUC of 0.8.

Based on the AUC values, Random Forest is the best-performing model, followed closely by Gradient Boosting. The Logistic Regression, SVM, and KNN models also exhibit good performance.

While Random Forest has the highest AUC, other factors should be considered when choosing the final model:

- **Interpretability**: Logistic Regression is more interpretable than Random Forest or Gradient Boosting. If understanding the factors driving churn is crucial, Logistic Regression may be preferred.
- **Computational Cost**: Complex models like Random Forest and Gradient Boosting can be more computationally expensive to train and deploy than simpler models like Logistic Regression.

- **Overfitting**: Random Forest and Gradient Boosting are prone to overfitting, especially with high-dimensional data. Regularization techniques and careful hyperparameter tuning are necessary to mitigate overfitting.

Recommendations:

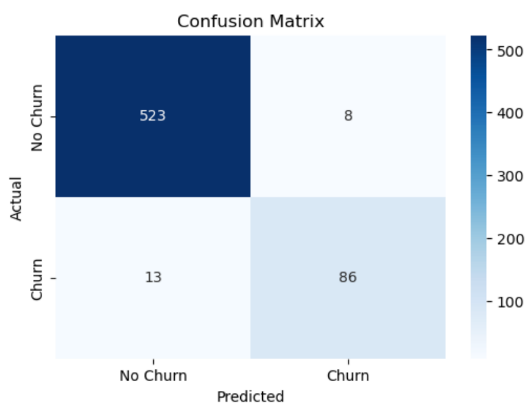
1. Random Forest or Gradient Boosting: As top-performing models, consider using Random Forest or Gradient Boosting as your primary churn prediction model.
2. Evaluate Logistic Regression: If interpretability is a key requirement, evaluate Logistic Regression and consider using it if its performance is acceptable.
3. Hyperparameter Tuning: Optimize the hyperparameters of all models using techniques like cross-validation to achieve the best possible performance on your specific dataset.
4. Model Validation: Thoroughly validate the selected model on a held-out test set to ensure its generalization performance.

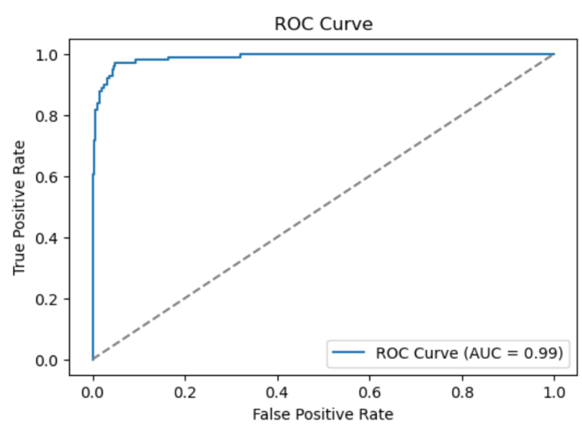
In summary, the ROC curves indicate that all the models have good predictive power for churn, with Random Forest and Gradient Boosting showing the best performance. As an additional recommendation it would be ideal to choose the final model based on a balance of performance, interpretability, computational cost, and risk of overfitting.

Additional Findings with XGBoost

In addition to the above, performed Extreme Gradient Boosting classifier based analysis. Listed below are the XGBoost classifier before and after hyperparameter tuning.

XGBoost model performed exceptionally well with high accuracy (97%), precision (91%), and an outstanding AUC score (0.99). It effectively identifies most churners while keeping false alarms low. However, addressing false negatives could further enhance its ability to detect all potential churners, making it even more robust for real-world applications like customer retention strategies.





1. **Performance Metrics**

- Accuracy: 0.97 (97%) The model correctly classifies 97% of all predictions, indicating high overall performance.
- Precision: 0.91 (91%) Among all instances predicted as "Churn," 91% are actual churners. This shows the model has a low false positive rate.
- Recall: 0.87 (87%) The model identifies 87% of actual churners, meaning it performs well in detecting churn but misses some cases (false negatives).
- F1 Score: 0.89 The F1 score balances precision and recall, showing the model is well-rounded in handling both false positives and false negatives.
- ROC AUC Score: 0.99 A near-perfect score of 0.99 indicates excellent separability between the "Churn" and "No Churn" classes.

2. **Confusion Matrix**

The confusion matrix provides a breakdown of predictions:

- True Negatives (523): The model correctly predicts "No Churn" for 523 customers.
- False Positives (8): Only 8 customers were incorrectly predicted as "Churn" when they did not churn.
- False Negatives (13): The model misses 13 actual churners, predicting them as "No Churn."
- True Positives (86): The model correctly identifies 86 customers who actually churned.

Insights:

- The model performs exceptionally well in predicting "No Churn" cases, with only a small number of false positives.
- While recall is strong at 87%, the false negatives (13) suggest there is room for improvement in capturing all churners.

3. **ROC Curve**

The ROC curve evaluates the trade-off between the true positive rate (recall) and the false positive rate:

- The curve is very close to the top-left corner, indicating excellent performance.
- AUC = 0.99 confirms that the model has near-perfect discrimination between the two classes.

Insights:

- The high AUC value reflects that the model is highly effective at distinguishing between churners and non-churners across different thresholds.

Key Strengths

1. High Accuracy: The model achieves an impressive accuracy of 97%, making it reliable for most predictions.
2. Strong Precision: With a precision of 91%, it minimizes false positives, which is crucial for avoiding unnecessary interventions for non-churning customers.
3. Excellent AUC: The high AUC score demonstrates that the model effectively separates churners from non-churners.

Areas for Improvement

1. False Negatives: While recall is strong, reducing the number of false negatives (13) would further improve the model's ability to capture all churners.
    1. Possible actions: Adjusting decision thresholds or using techniques like oversampling/undersampling to handle class imbalance if present.
2. Class Imbalance Check: If churn cases are significantly fewer than non-churn cases, consider rebalancing the dataset to ensure better recall without sacrificing precision.

Training XGBoost Model

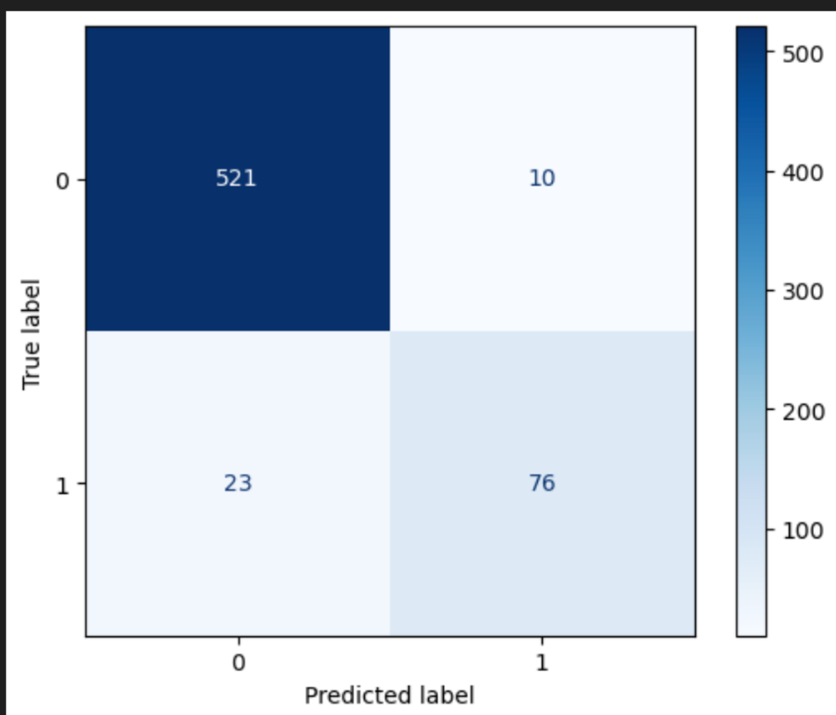
```
Training XGBoost Model...

XGBoost Classifier:
Accuracy: 0.9476
AUC: 0.9793
              precision    recall  f1-score   support

           0       0.96      0.98      0.97       531
           1       0.88      0.77      0.82        99

    accuracy                           0.95       630
   macro avg       0.92      0.87      0.90       630
weighted avg       0.95      0.95      0.95       630

Confusion Matrix:
 [[521  10]
 [ 23  76]]
```



1. Initial Model Performance

- Accuracy: 94.76% (indicates the overall correctness of predictions).
- AUC (Area Under the Curve): 0.9793 (shows high discrimination ability between classes).
- Precision, Recall, F1-Score:
  - Class 0 (majority class): High precision (0.96), recall (0.98), and F1-score (0.97).
  - Class 1 (minority class): Lower precision (0.88), recall (0.77), and F1-score (0.82), indicating some difficulty in correctly identifying minority class instances.
- Confusion Matrix:
  - True Negatives (TN): 521
  - False Positives (FP): 10
  - False Negatives (FN): 23
  - True Positives (TP): 76
  - The model misclassifies some instances, especially for Class 1.

Tuning XGBoost

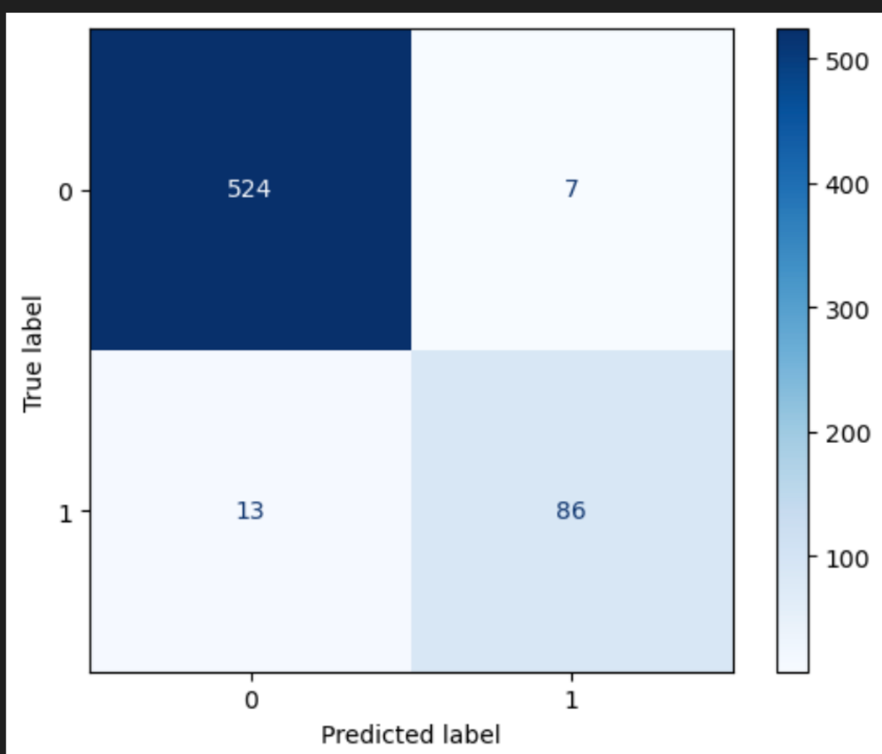
```
Tuning XGBoost Model...
Best Parameters: {'learning_rate': 0.1, 'max_depth': 7, 'n_estimators': 200, 'subsample': 0.8}

Tuned XGBoost Classifier:
Accuracy: 0.9683
AUC: 0.9866
              precision    recall  f1-score   support

           0       0.98      0.99      0.98       531
           1       0.92      0.87      0.90        99

    accuracy                           0.97       630
   macro avg       0.95      0.93      0.94       630
weighted avg       0.97      0.97      0.97       630

Confusion Matrix:
 [[524   7]
 [ 13  86]]
```



2. Tuned Model Performance

- After hyperparameter tuning with parameters like learning_rate=0.1, max_depth=7, n_estimators=200, and subsample=0.8:

- **Accuracy**: Improved to 96.83%.
- **AUC**: Increased to 0.9866, indicating better class separation.
- **Precision, Recall, F1-Score**:
    - Class 0: Precision, recall, and F1-score improved slightly.
    - Class 1: Significant improvement in precision (0.92), recall (0.87), and F1-score (0.90), reflecting better handling of the minority class.

- **Confusion Matrix**:
    - TN: 524
    - FP: 7
    - FN: 13
    - TP: 86

The number of misclassifications decreased for both classes.

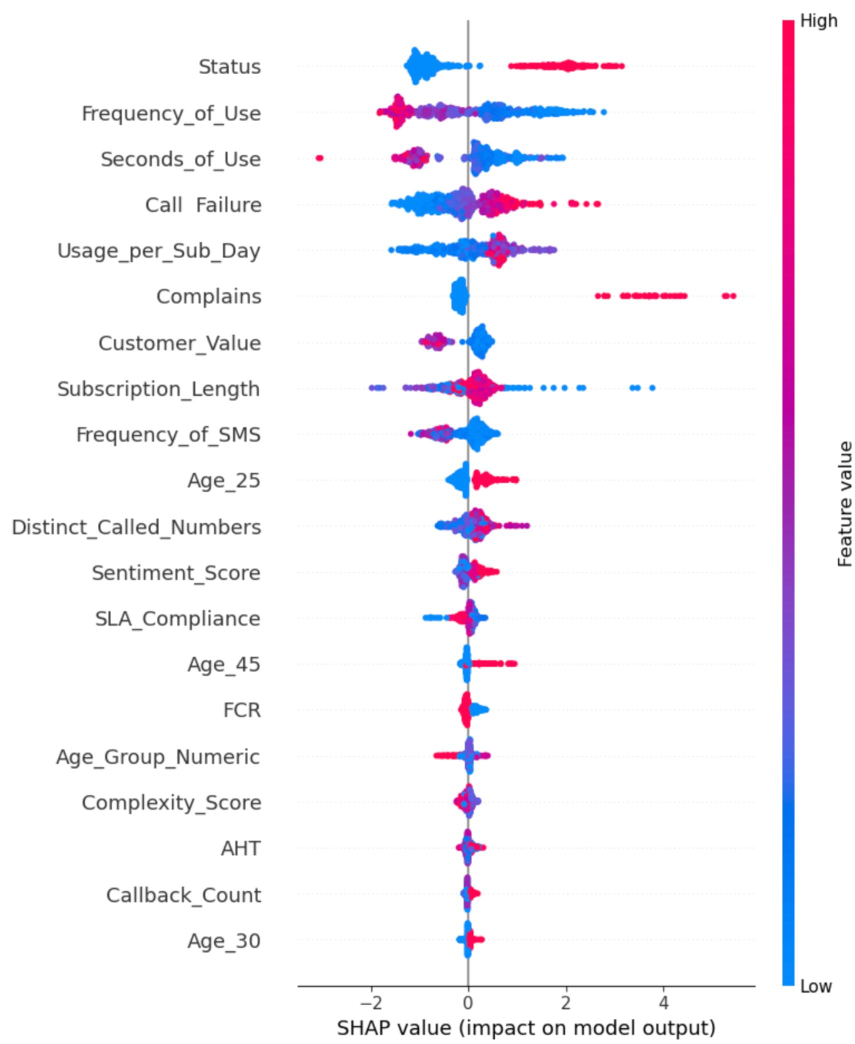
<u>Key Takeaways</u>

- The XGBoost model performs well initially but shows bias toward the majority class.
- Hyperparameter tuning significantly improves the model's performance, particularly for the minority class, as seen in better precision, recall, and fewer misclassifications.
- The tuned model is more balanced and effective at distinguishing between the two classes while maintaining high overall accuracy and AUC.

SHAP Interpretation

I tried to perform SHAP based interpretation to gain deeper insights into the decision-making process of XGBoost model, helping validate its reliability and interpretability.

The SHAP plot reveals that customer usage metrics (Status, Frequency_of_Use) and service quality indicators (Call Failure, Complains) are the most influential features in the model's predictions. These insights can guide targeted interventions, such as improving service reliability or addressing complaints to enhance predictive accuracy and customer satisfaction.



This SHAP (SHapley Additive Explanations) summary plot shows how each feature impacts the model's predictions.

- x-axis (Impact on Model Output)

- Negative values decrease the likelihood of churn.

- Positive values increase the likelihood of churn.

- y-axis (Model Features)

- The order of the features on the Y-axis is based on overall importance.

- Color (Feature Value Intensity)

- Blue indicates lower feature values.

- Red indicates higher feature values.

SHAP Key Insights

- Features at the top of the plot have the highest impact on the model's predictions, while those at the bottom contribute less.

- Top Impactful features

- Status: This feature has the most significant influence, with a wide range of SHAP values indicating strong predictive power.

- Frequency_of_Use and Seconds_of_Use: These usage-related metrics also play critical roles in determining predictions.

- Call Failure and Usage_per_Sub_Day: Indicators of service quality and usage patterns are highly relevant.

- Lesser Impactful features

- Features like Age_30 and Callback_Count, located near the bottom, have minimal influence on predictions. These might be less relevant for decision-making or could be candidates for removal during feature selection.

- In Status, higher values (red) are associated with positive impacts on predictions, while lower values (blue) contribute negatively.

- In Complains, higher values (red) negatively impact predictions, suggesting that more complaints correlate with unfavorable outcomes.

- Wider distributions of SHAP values for features like Status and Frequency_of_Use indicate variability in their influence across different samples.

- Narrower distributions for features like Age_Group_Numeric suggest consistent but less impactful contributions.

- Features related to customer usage patterns (Frequency_of_Use, Seconds_of_Use) are critical for understanding customer behavior.

- Service quality metrics (Call Failure, Complains) are key drivers for predicting outcomes, emphasizing their importance in improving customer satisfaction.