

SyriaTel Customer Churn Modeling Project

By Brian Amani

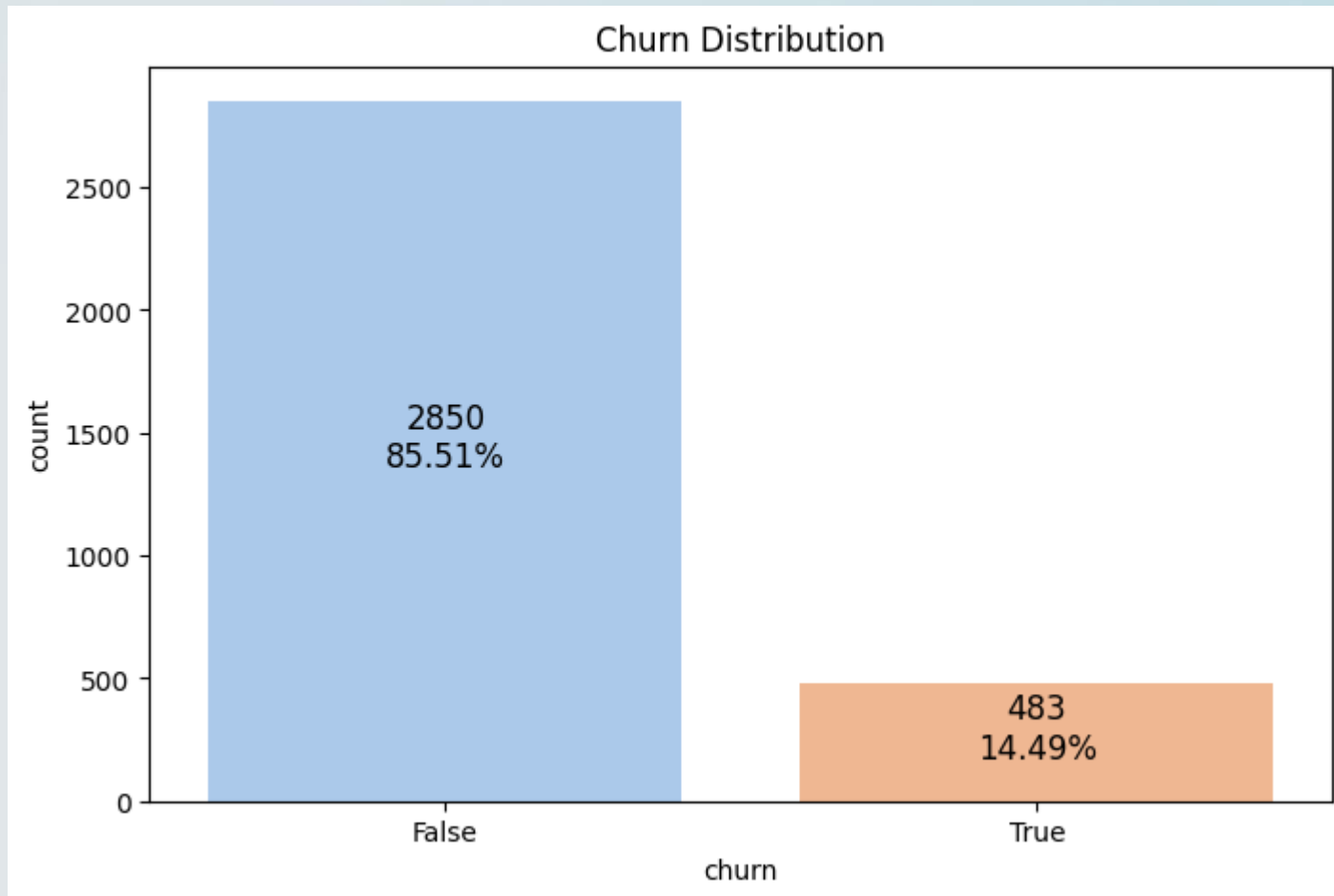
Project Outline

- Overview & Objectives
- Data Understanding
- Modelling
- Evaluation & Conclusion

Overview & Objectives

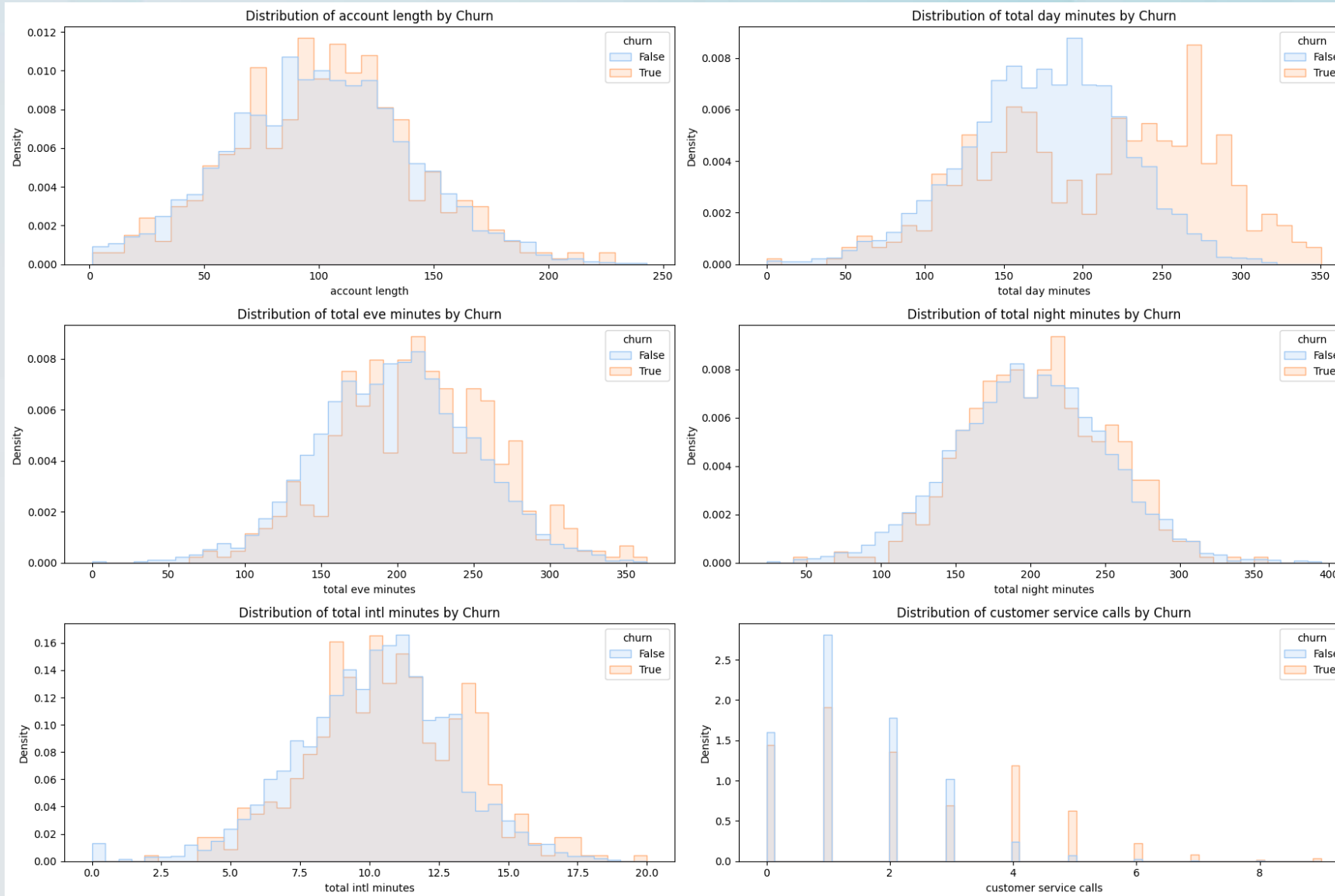
- This project focuses on developing a model to predict whether a customer is likely to churn from SyriaTel, a telecommunications company. The goal is to help SyriaTel reduce churn by identifying high-risk customers and applying targeted retention strategies.
- By analyzing customer data, the model provides actionable insights to stakeholders aiming to minimize churn and prevent revenue loss. The results enable SyriaTel to identify at-risk customers and take proactive steps to retain them, supporting efforts to improve customer loyalty and profitability. retention tactics.

Data Understanding



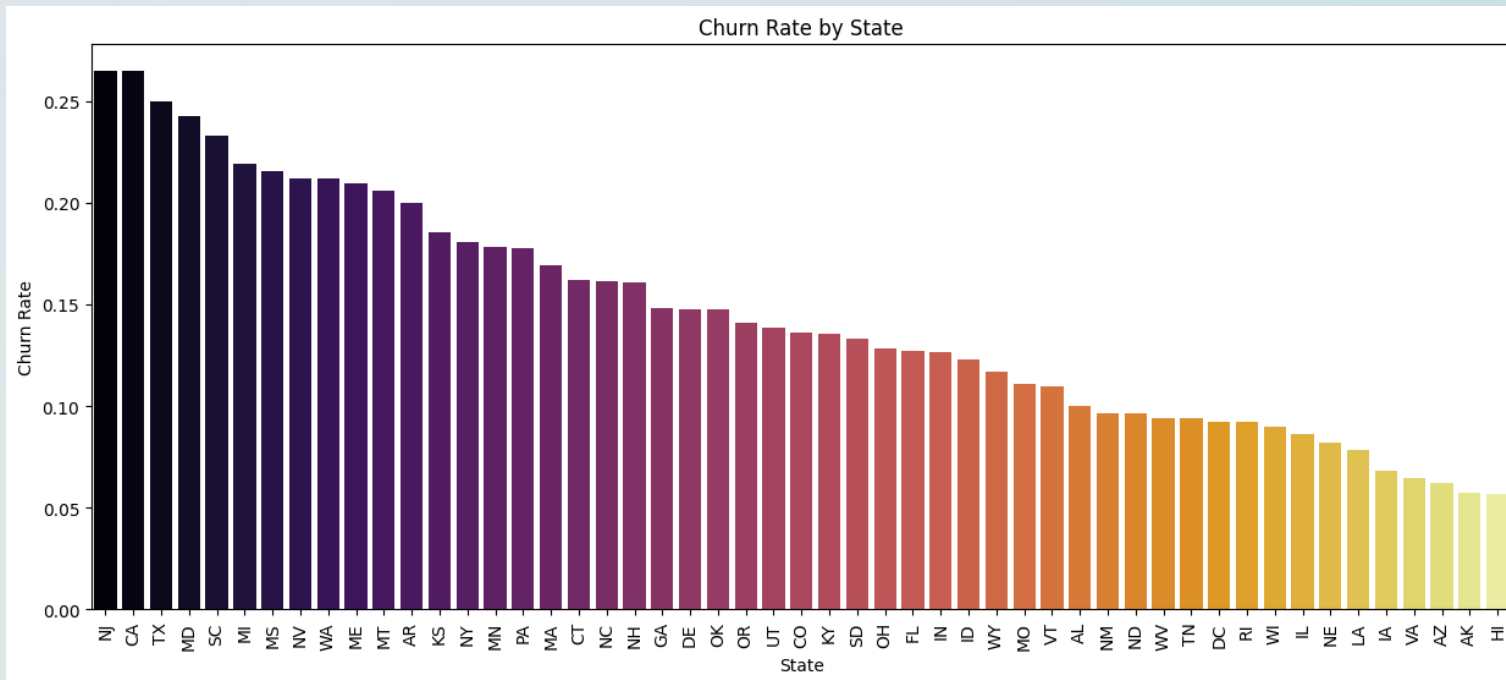
This figure shows an imbalance in churn class. 14.49% churned vs 85.51% unchurned

Data Understanding

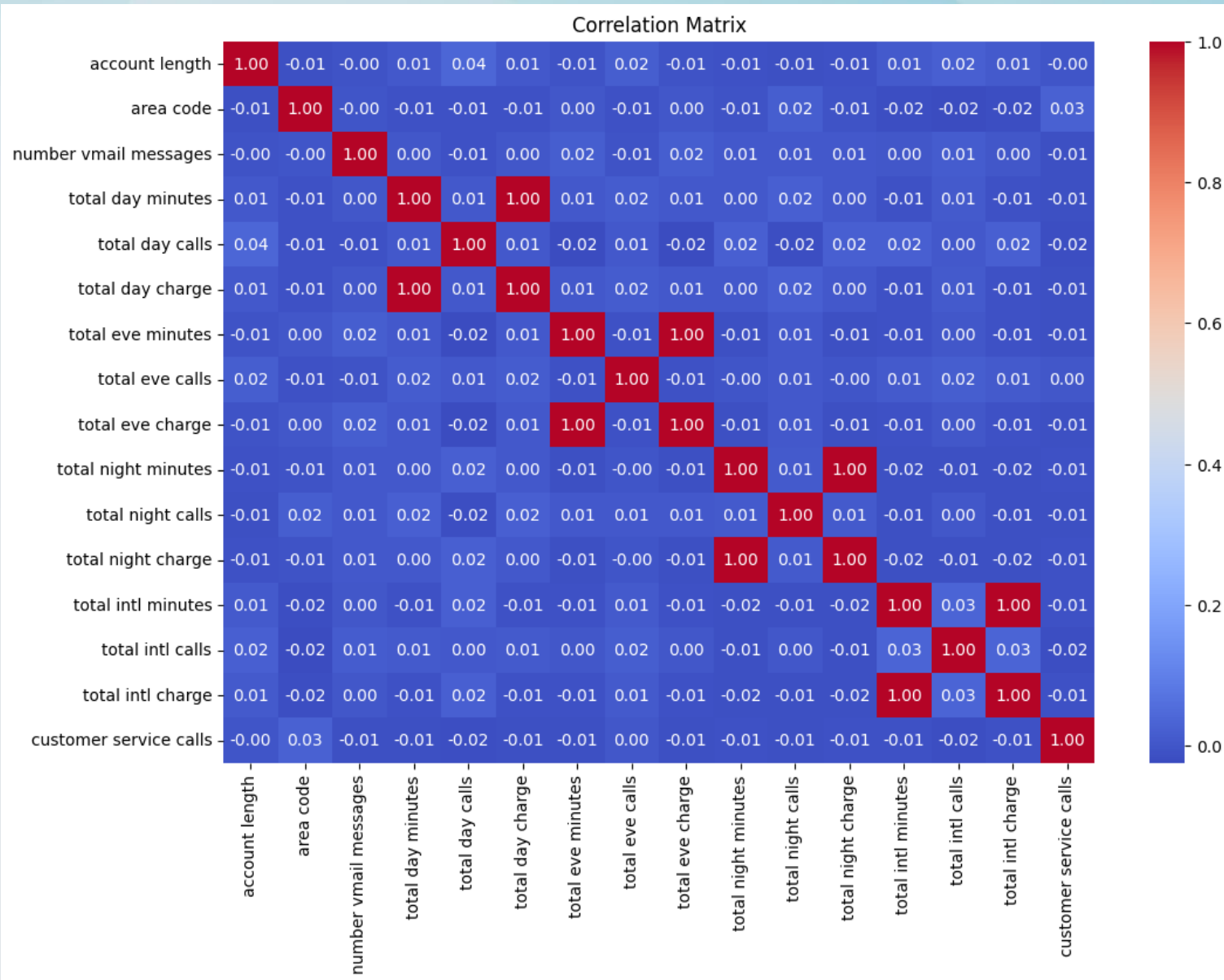


The distribution plots show that customers with higher total day minutes and more frequent customer service calls are more likely to churn, suggesting possible dissatisfaction among heavy daytime users and those with ongoing service issues.

Data Understanding



The plot shows a significant decline in churn rate by state from left to right. This is indicative that the state has an effect on churn rate.



The correlation matrix shows that for the most part, the numeric variables in the data set are independent. Let us now analyse the categorical variables

Modelling

I created 5 models:

1. Simple logistic regression model
2. SMOTE to refine the logistic regression model
3. Logistic regression model with hyperparameter tuning
4. Decision Tree Model
5. Decision tree with hyperparameter tuning

Model Evaluation

	Model	Accuracy	Precision (Churn)	Recall (Churn)	F1-Score (Churn)	AUC
0	Logistic Regression	0.859070	0.606061	0.198020	0.298507	0.830983
1	Logistic Regression (SMOTE)	0.766117	0.368421	0.762376	0.496774	0.821800
2	Logistic Regression (Tuned)	0.779610	0.392857	0.777778	0.518519	0.821818
3	Decision Tree	0.890555	0.623655	0.715843	0.671641	0.806415
4	Decision Tree (Tuned)	0.917541	0.724138	0.743519	0.734694	0.806415

- Logistic Regression (both the standard and tuned versions) performs reasonably well with high accuracy and AUC but struggles to identify churners effectively due to low recall. It is more reliable for general classification but not optimal for predicting churn, where recall is crucial.
- Logistic Regression (SMOTE) demonstrates a good improvement in recall, but at the cost of low precision and accuracy. It is suitable if the goal is to catch as many churners as possible, even if that means misclassifying more non-churn customers.
- Decision Tree (and its tuned version) provides a solid balance of precision, recall, and F1-score. Its accuracy and performance make it a better choice for predicting churn compared to logistic regression, especially when fine-tuned.

Model Evaluation

Accuracy: 0.9175412293853074

Classification Report:

	precision	recall	f1-score	support
False	0.95	0.95	0.95	566
True	0.72	0.74	0.73	101
accuracy			0.92	667
macro avg	0.84	0.85	0.84	667
weighted avg	0.92	0.92	0.92	667

Model 5 performs the best:

This model performs the best as compared to previous models:

- High Accuracy: The model achieves 91.75% accuracy, demonstrating strong performance in classifying both churn and non-churn customers.
- Churn Recall: The churn recall (0.74) is improved, showing better identification of potential churners compared to previous models.
- Precision-Recall Trade-off: Precision for churn (0.72) is decent, but there's room for improvement, as some non-churn customers may be misclassified as churners.
- Class Balance: The model performs well overall, with a macro-average recall of 0.85 and a weighted recall of 0.92, reflecting strong classification of non-churn customers.

Conclusion & Recommendations

Decision Tree (Tuned) provides the best overall performance in predicting customer churn, with a balanced mix of accuracy, recall, precision, and F1-score. SyriaTel should prioritize this model for identifying high-risk customers and implementing retention strategies. This model can help minimize churn while ensuring that targeted interventions are aimed at the most accurate customer segments.

Recommendations

- **Prioritize Decision Tree (Tuned):** Use this model for accurate churn predictions and targeted retention strategies.
- **Implement Proactive Retention:** Use insights from the churn model to offer personalized promotions and loyalty programs to high-risk customers.
- **Leverage SMOTE for Increased Recall:** Consider using Logistic Regression (SMOTE) for a higher recall rate to catch more churners, while managing precision.
- **Segment Customers:** Tailor retention efforts by segmenting customers based on usage patterns, tenure, or demographics for more effective interventions.
- **Regular Model Updates:** Continuously monitor and update models with new data to maintain accuracy and adapt to changing customer behaviours.

Thank You!

GitHub: @papyrusleaf

LinkedIn:

<https://www.linkedin.com/in/brian-amani-63a64987>