

CUSTOMER CHURN ANALYSIS AND SERVICE OPTIMIZATION



WINFRED KINYA

Machine Learning Project

Phase 3

07/06/2024

OVERVIEW
TELECOM COMPANIES FACE
REVENUE LOSSES FROM
CUSTOMER CHURN.

UNDERSTANDING CHURN
FACTORS IS CRUCIAL FOR
PROFITABILITY.

This project's goals:

- ❖ Identify Customer Churn rate
- ❖ Evaluate the current call charges for different plans
- ❖ Choose appropriate models for predicting churn
- ❖ Segment Customers Based on Service Usage Patterns, Satisfaction Levels, and Likelihood of Churn.

DATA UNDERSTANDING



- ❖ state 3333 non-null object
- ❖ account length 3333 non-null int64
- ❖ area code 3333 non-null int64
- ❖ phone number 3333 non-null object
- ❖ international plan 3333 non-null object
- ❖ voice mail plan 3333 non-null object
- ❖ number voicemail messages 3333 non-null int64
- ❖ total day minutes 3333 non-null float64
- ❖ total day calls 3333 non-null int64
- ❖ total day charge 3333 non-null float64
- ❖ total eve minutes 3333 non-null float64
- ❖ total eve calls 3333 non-null int64
- ❖ total eve charge 3333 non-null float64
- ❖ total night minutes 3333 non-null float64
- ❖ total night calls 3333 non-null int64
- ❖ total night charge 3333 non-null float64
- ❖ total intl minutes 3333 non-null float64
- ❖ total intl calls 3333 non-null int64
- ❖ total intl charge 3333 non-null float64
- ❖ customer service calls 3333 non-null int64
- ❖ churn

STATISTICAL SUMMARY

Descriptive statistics.

Account Length: Mean - 101, Min - 1, Max - 243

Voicemail Messages: Mean - 8, Min - 0, Max - 51

Day Minutes: Mean - 179, Min - 0, Max - 350

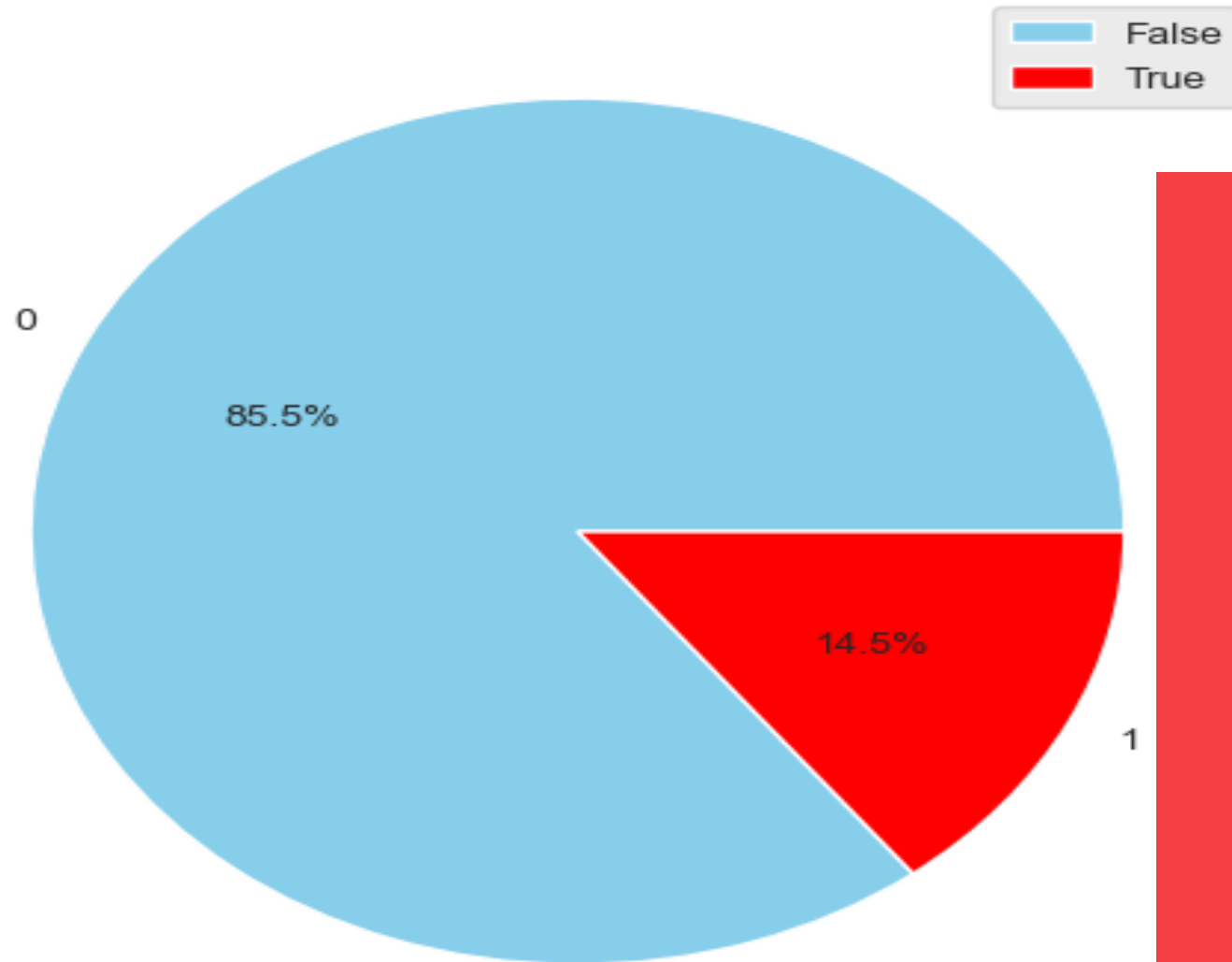
Evening Minutes: Mean - 200, Min - 0, Max - 363

Night Minutes: Mean - 200, Min - 23, Max - 395

International Minutes: Mean - 10, Min - 0, Max - 20

Customer Service Calls: Mean - 1.6, Min - 0, Max - 9

Proportion of Churn



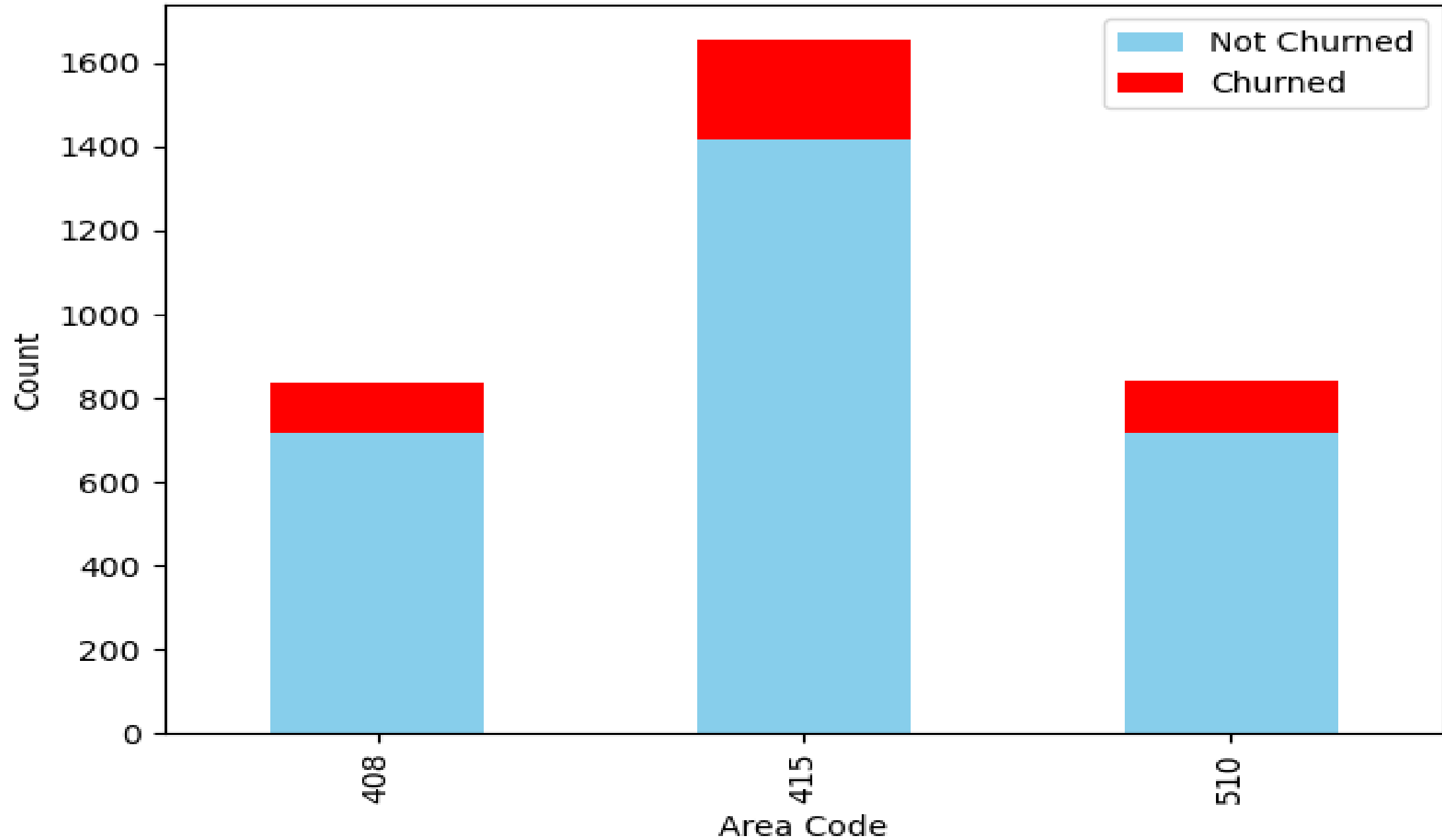
NON-CHURNED CUSTOMERS:

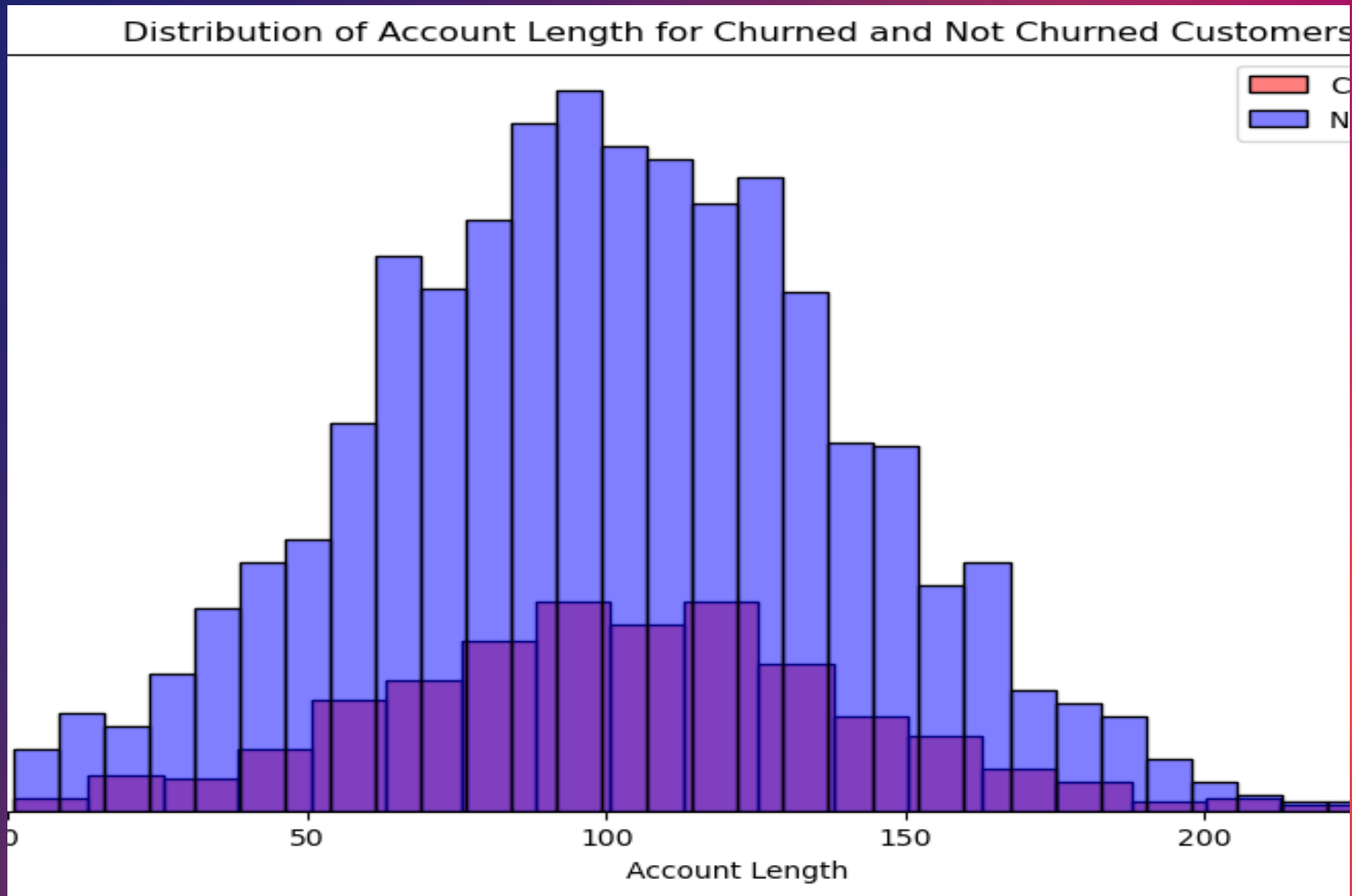
85.5% OF THE CUSTOMERS HAVE NOT CHURNED, INDICATING THAT THE MAJORITY OF THE CUSTOMER BASE REMAINS WITH THE COMPANY.

CHURNED CUSTOMERS:

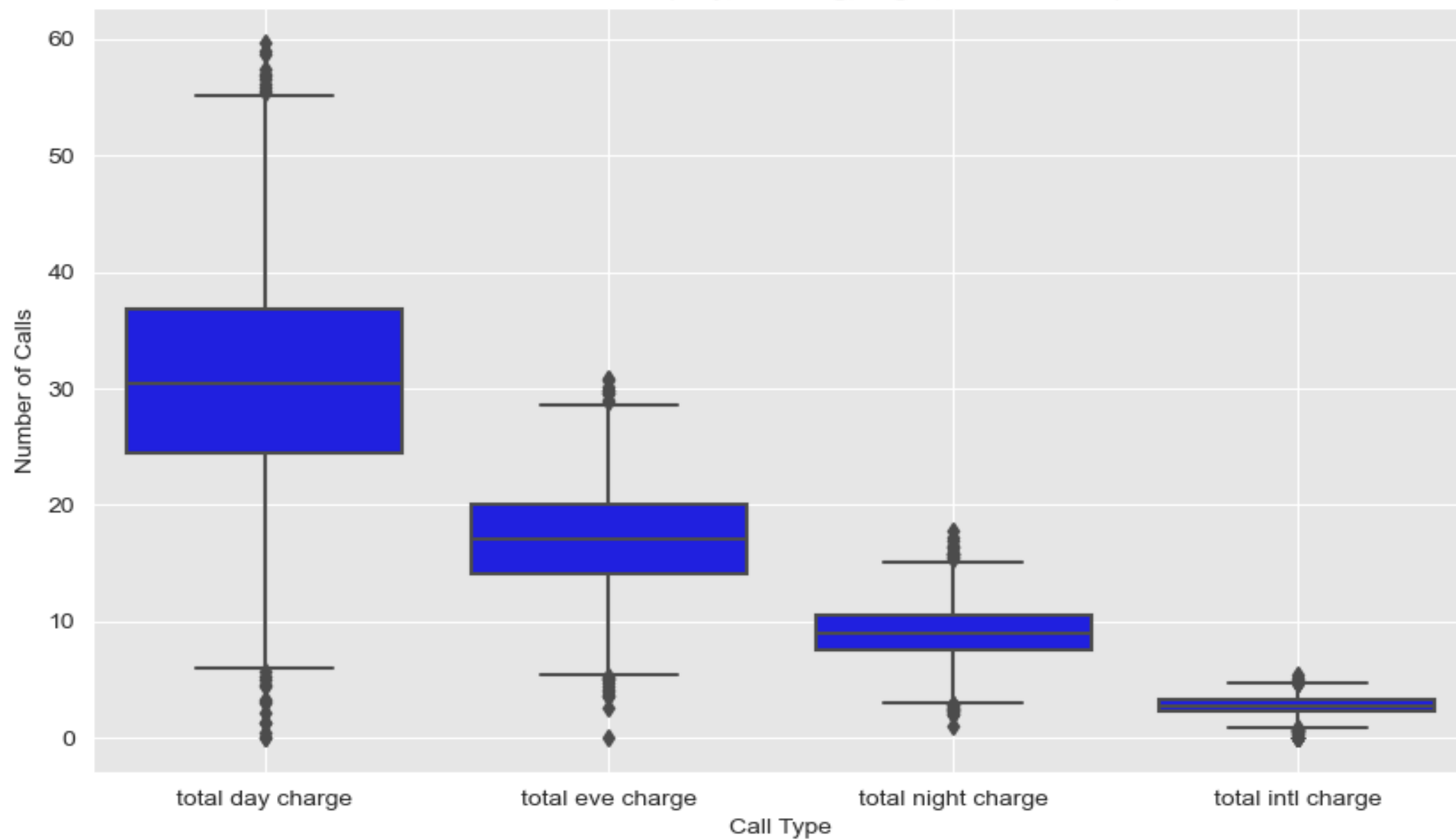
14.5% OF THE CUSTOMERS HAVE CHURNED, WHICH IS A SIGNIFICANT MINORITY THAT NEEDS ATTENTION.

Number of Churns vs Not Churned per Area Code





Box Plot of Calls (Day, Evening, Night, International)



MODELS

LOGISTIC LINEAR

A baseline model to establish a benchmark for other models.

KNN MODEL

providing additional insights into customer behavior and further validating the model's effectiveness

DECISION TREE

Used for their interpretability and ability to capture non-linear relationships

RANDOM FOREST

An ensemble method that improves the performance and accuracy over single decision trees.

MODEL RESULTS

KEY METRICS

	Precision	Recall	F1 Score	Accuracy
Logistic Regression	0.95	0.72	0.82	0.74
	0.23	0.69	0.35	
KNN Model	0.96	0.79	0.87	0.78
	0.31	0.74	0.43	
Decision Tree Tuned	0.97	0.89	0.93	0.88
	0.49	0.81	0.61	
Random Forest	0.98	0.96	0.97	0.94
	0.70	0.82	0.76	

MODEL EVALUATION.

For class label 0, all models perform relatively similarly across precision, recall, and F1-score, with scores close to or at 1.0, indicating very high performance.

For class label 1, there is more variation in performance. The Random Forest and Decision Tree models appear to have the highest precision, recall, and F1-score for this class, while the Untuned Decision Tree and Logistic Regression models have the lowest scores.

The KNN model has moderate performance for class label 1, with scores that are lower than the Random Forest and Decision Tree but higher than the Untuned Decision Tree and Logistic Regression

OBSERVATIONS

- Overall, the Random Forest and Decision Tree models seem to be the best performers for both classes, with the Untuned Decision Tree and Logistic Regression models being the least effective, particularly for class label 1. The KNN model falls in between these two extremes.
- The significant drop in performance for the Untuned Decision Tree compared to the tuned Decision Tree suggests that tuning hyperparameters is important for achieving good performance.
- However between the Random Forest and Decision Tree, Checking the accuracy and ROC curve scores: Random Forest accuracy 94%, Decision tree accuracy 88% , The random Forest performs way better on test unseen or new data. Also, The consistent cross validation score performance across folds suggests that the model complexity is well balanced, allowing it to learn the patterns in the data without being overly complex.
- The best model to go by is Random Forest.

RECOMMENDATIONS

1. Conduct detailed analyses to understand the primary factors driving customer churn, this could include surveys and customer feedback. Since Area Code 415 has the highest number of churned customers, it would be beneficial to conduct a detailed analysis to understand why customers are leaving.
2. Reevaluate the voicemail plan's importance and consider repositioning it as an add-on service rather than a standard offering.
3. Implement loyalty programs to reward long-term customers and accounts above a length of 80. incentivize them to stay.



RECOMMENDATIONS

The charges for day, evening, night, and international usage are relatively stable. Regularly review and adjust pricing to remain competitive. Especially Optimized Call Charges for each plan and time of the call.

Evening charges showed the widest range, suggesting potential for targeted plans to optimize revenue.

.Based on the evaluation metrics ,the best performing model is Random Forest thus use it to make customer churn predictions.



THANK YOU