Customer Retention: Telecom Churn Reduction Using Predictive Analytics

BO **by BERYL OKELO**

Business Problem

Customer churn heavily impacts telecom revenue. This project uses predictive models to identify at-risk customers and informs early retention efforts.

Data Understanding and Feature Overview

1

Dataset Source

Kaggle Telecom Churn Dataset with 3333 customer records and 21 features.

2

Target Variable

Churn is binary: True (1) means customer left; False (0) means active.

3

Key Features

- Location: state, area code
- Account info: account length, international plan, voice mail plan
- Usage: day, evening, night calls
- Customer service calls

Exploratory Data Analysis: Churn Patterns

1

Class Imbalance

Only 14% of customers churn, highlighting the need to handle imbalance in modeling.

2

Top Correlated Features

•International plan subscription

•Customer service calls

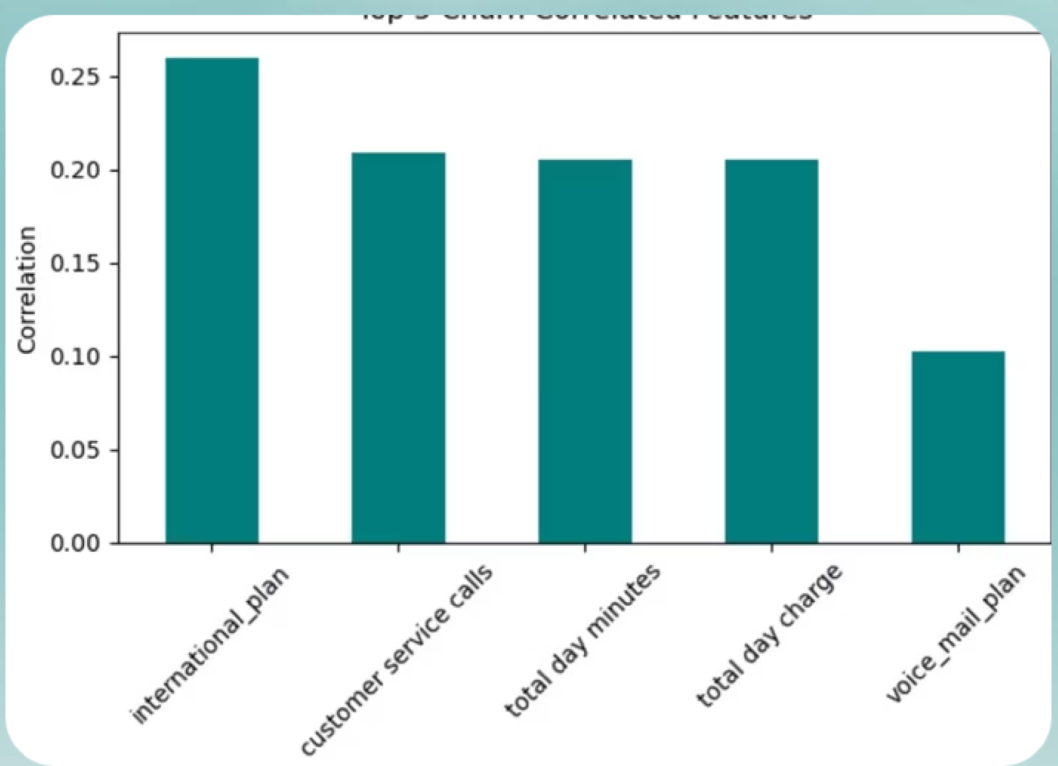
•Total day minutes and charges

•Total evening minutes

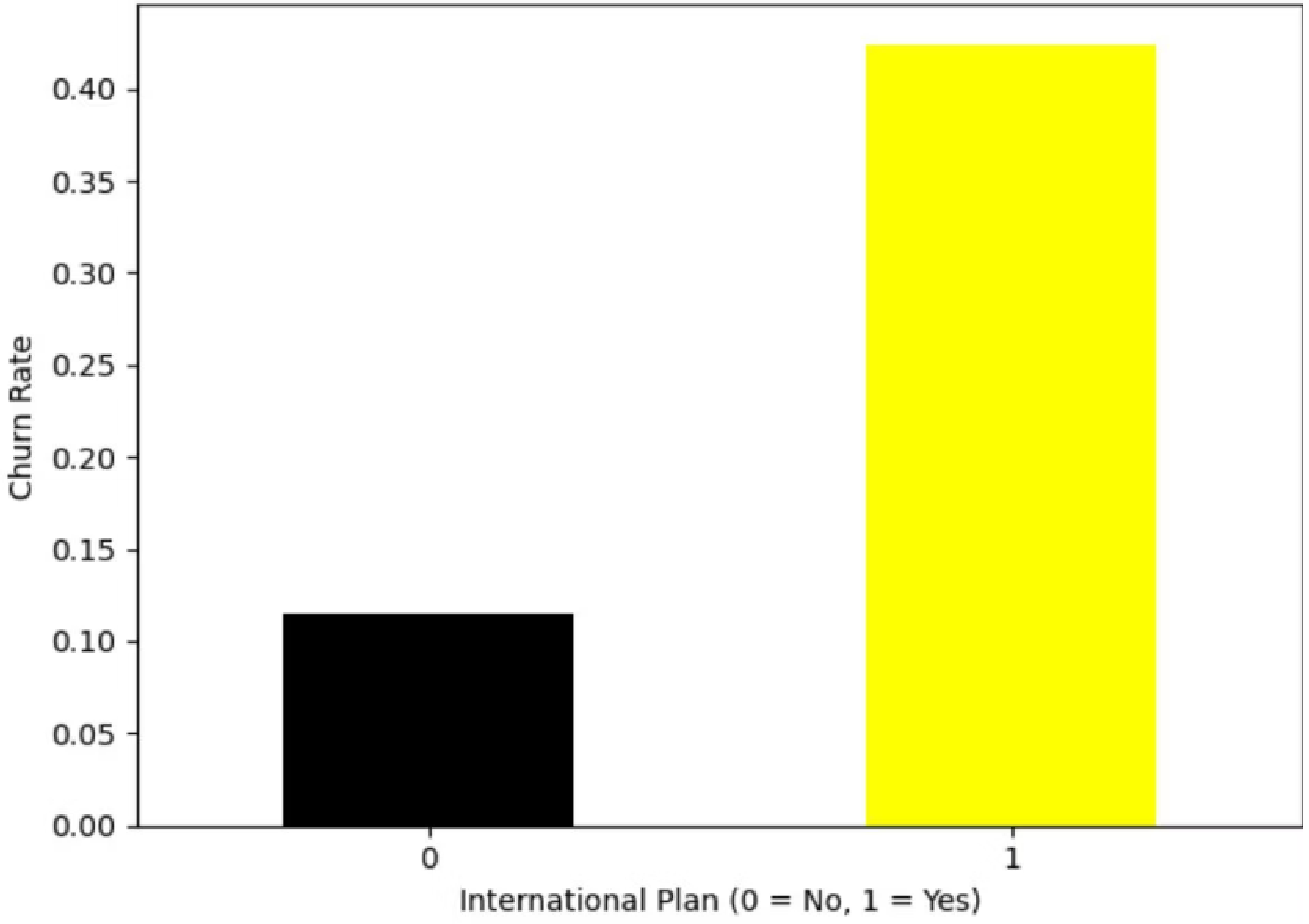
3

Insights

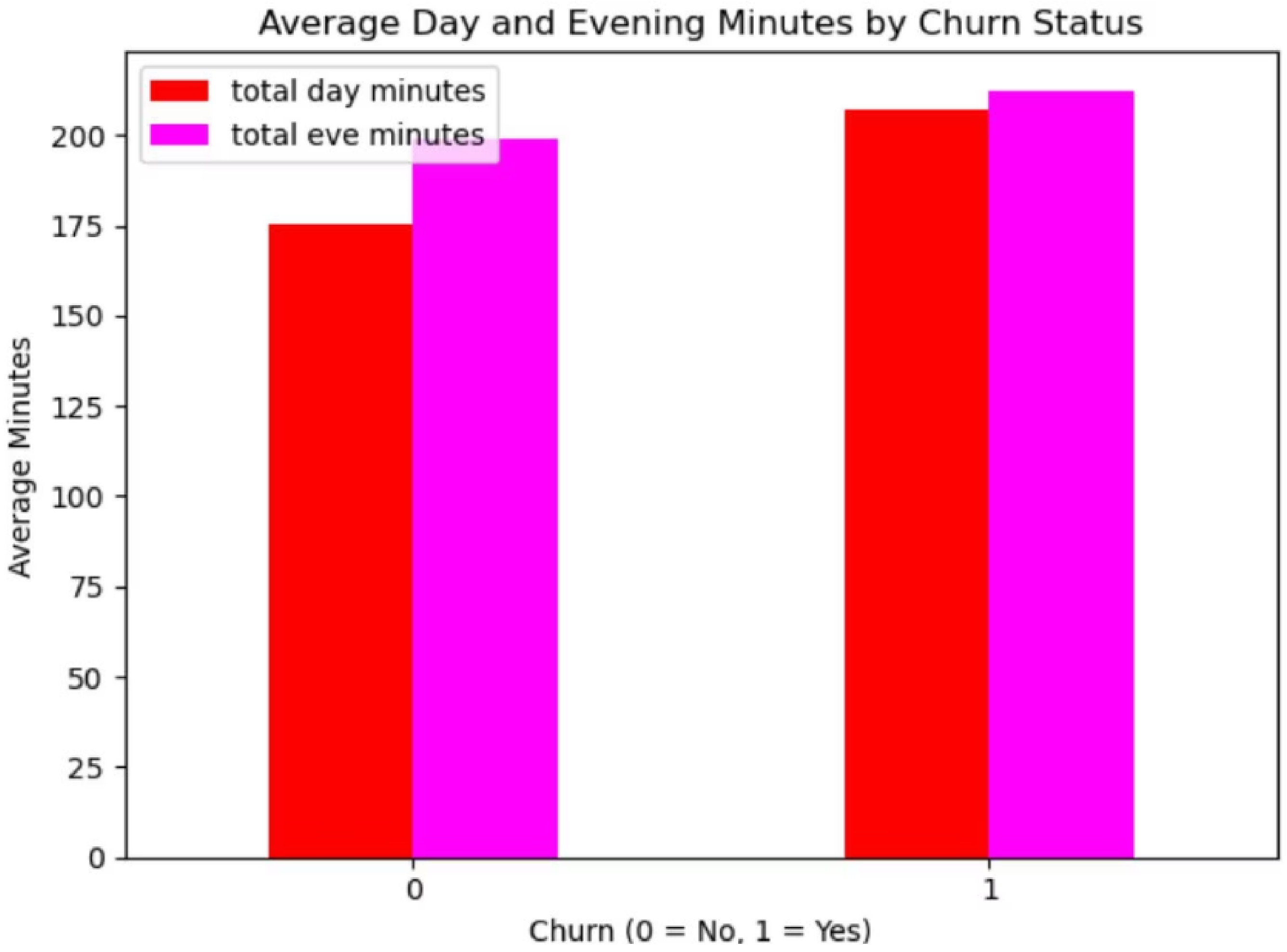
Customers with international plans and frequent service calls are more likely to churn due to high charges or dissatisfaction.



Churn Rates by International Plan and Time of Day Usage



Customers with international plans have a 42% churn rate versus 11% for those without, likely due to higher call costs or roaming issues.



Higher call usage during day and evening correlates with increased churn. Evening minutes are generally higher, possibly due to customer availability outside work hours.

Modeling: Logistic Regression Baseline

1

Performance Metrics

Accuracy: 78%, Precision: 39%, Recall: 77%, ROC-AUC: 83%

2

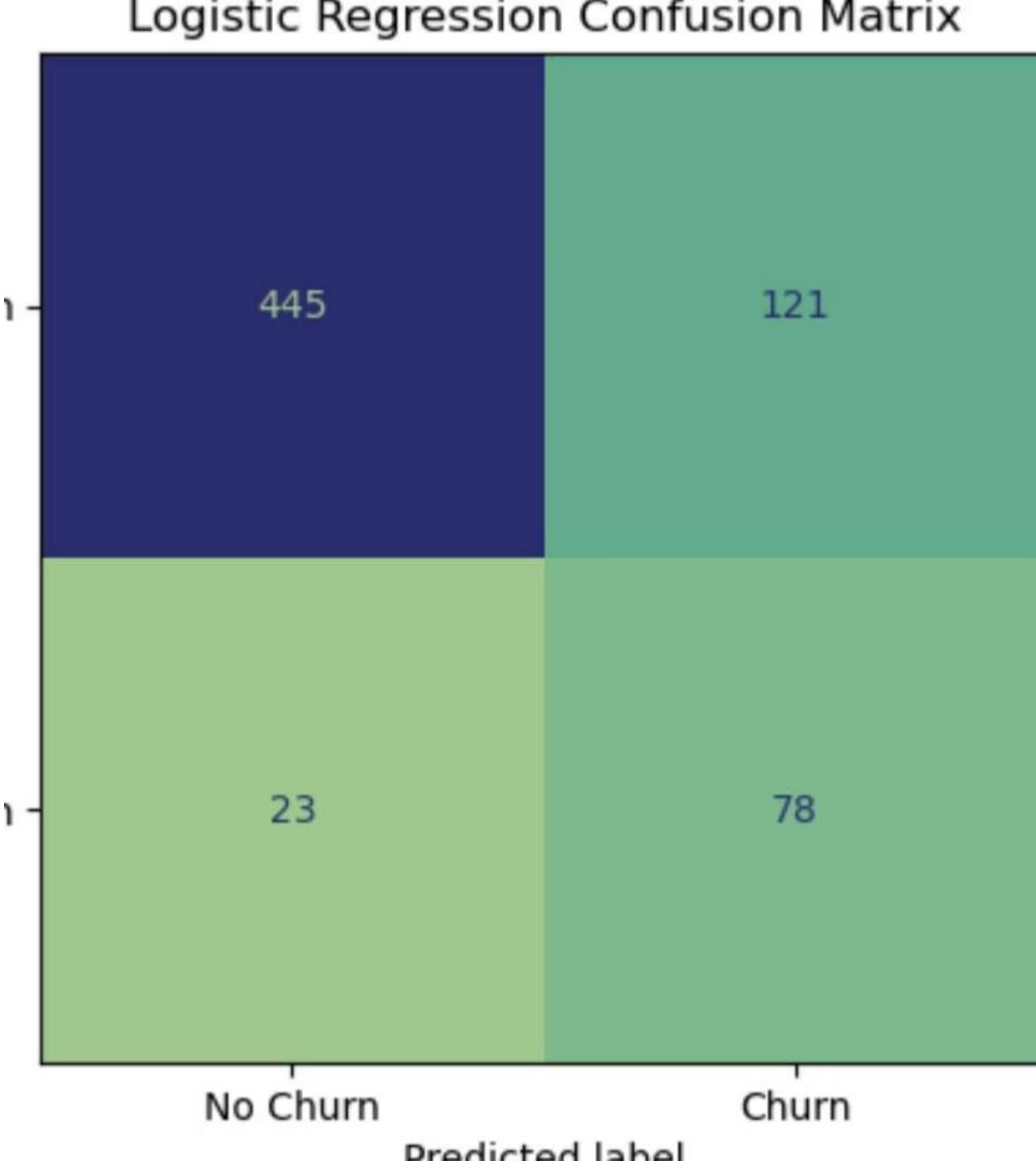
Interpretation

The model identifies 79% of non-churners and 77% of churners but has low precision, meaning many predicted churners do not actually churn.

3

Confusion Matrix Highlights

- True Positives: 78
- False Positives: 121
- False Negatives: 23
- True Negatives: 445



Modeling: Decision Tree

1

Performance Metrics

Accuracy: 93%, Precision: 76%, Recall: 76%, ROC-AUC: 86%

2

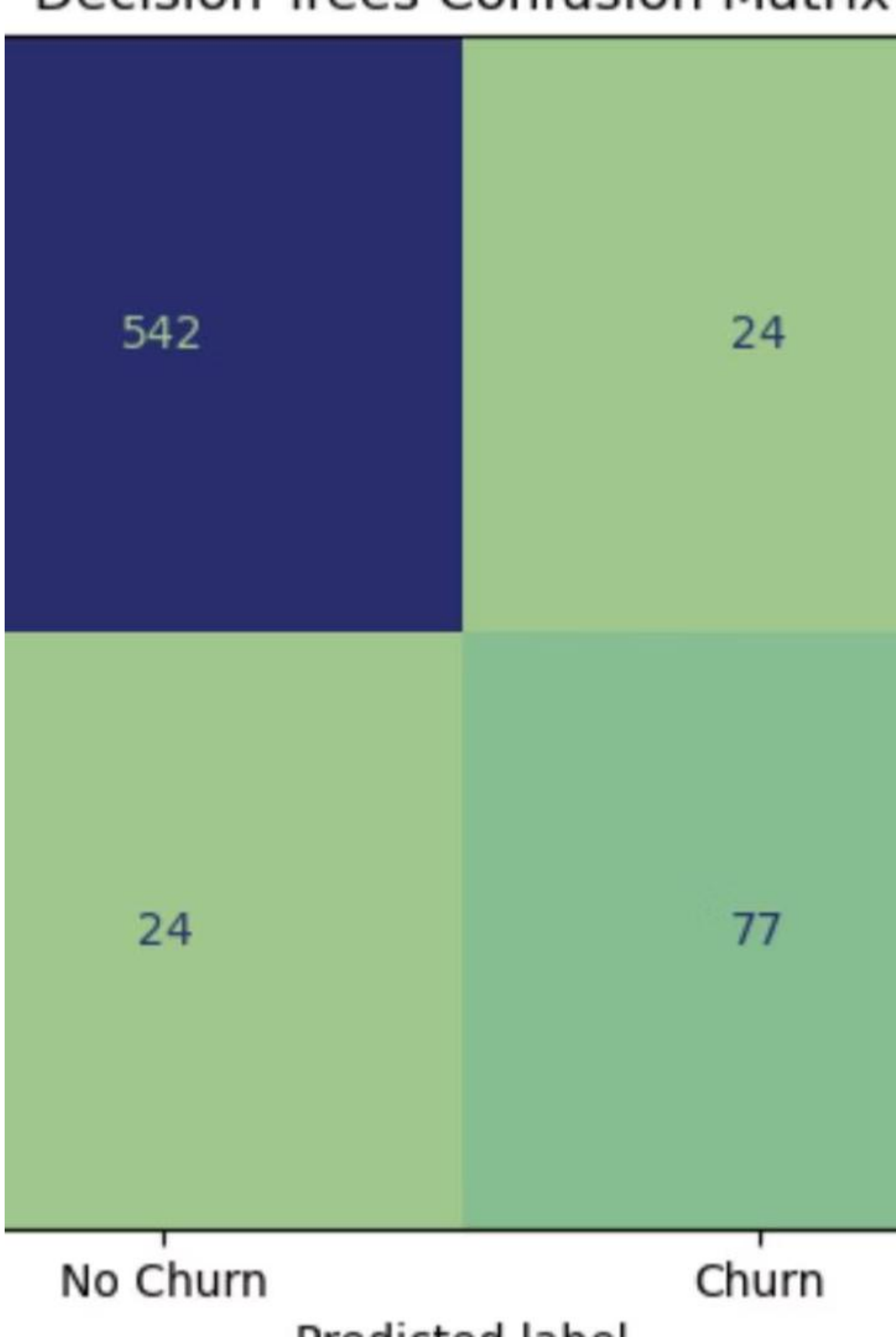
Interpretation

The decision tree model balances precision and recall well, accurately identifying most churners and non-churners.

3

Confusion Matrix Highlights

- True Positives: 77
- False Positives: 24
- False Negatives: 24
- True Negatives: 542



Comparative Analysis and Business Implications

Metric	Logistic Regression	Decision Tree
Accuracy	0.78	0.93
Precision	0.39	0.76
Recall	0.77	0.76
F1 Score	0.52	0.76
ROC-AUC	0.835	0.86

The decision tree outperforms logistic regression across all metrics, providing more reliable churn predictions. This allows the company to allocate retention resources efficiently and target customers truly at risk

Recommendations

1

Customer Segmentation & Personalization

- Use Decision Tree insights to group customers
- Target each segment with tailored offers and campaigns
- Provide personalized plans based on behavior

2

Customer Service Training

- Upskill support teams to handle queries effectively
- Convert support interactions into sales opportunities
- Reduce dissatisfaction through quicker issue resolution

3

Time-Based Offers

- Analyze call time patterns linked to churn
- Introduce:
  - Hourly/Daily bundles
  - Free calls at night
  - Affordable international rates

Conclusion

1

Decision Tree: Best Performing Model

- High accuracy, precision, recall, and F1-score
- Effectively identifies at-risk customers

2

Proactive Strategy Development

- Enables early intervention to reduce churn
- Supports personalized and data-driven actions

3

Business Impact

- Minimize churn-related losses
- Maximize retention and long-term customer value

4

Ongoing Monitoring & Optimization

- Regularly update the model with new churn patterns
- Improve accuracy and adaptability over time