Predicting Customer Churn to Reduce Revenue Loss

Telecom Customer Churn Classification Project
Stakeholder: Customer Retention & Marketing Team
Objective: Identify customers likely to leave
Goal: Reduce churn and protect revenue

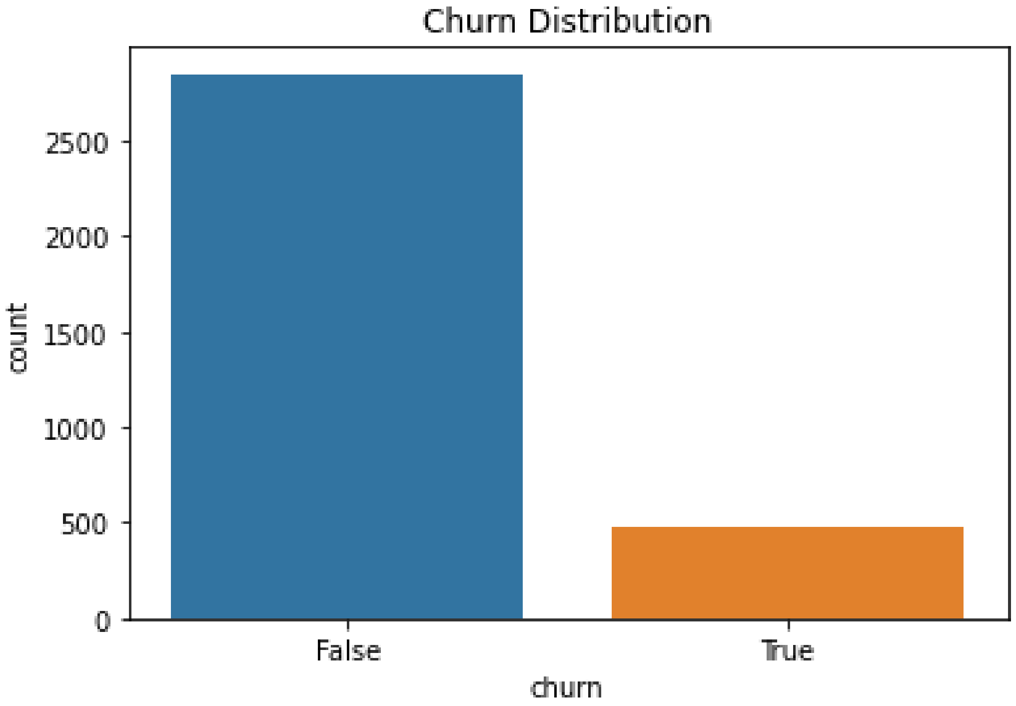
Project Overview

- **Business problem:** Customer churn leads to lost revenue and increased costs associated with acquiring new customers. By identifying customers who are likely to churn before they leave, the business can prioritize outreach, improve customer service experiences, and design targeted retention offers.
- **Solution:** Predict churn before it happens
- **Approach:** Classification modeling
- **Outcome:** Actionable retention targeting

Business Understanding

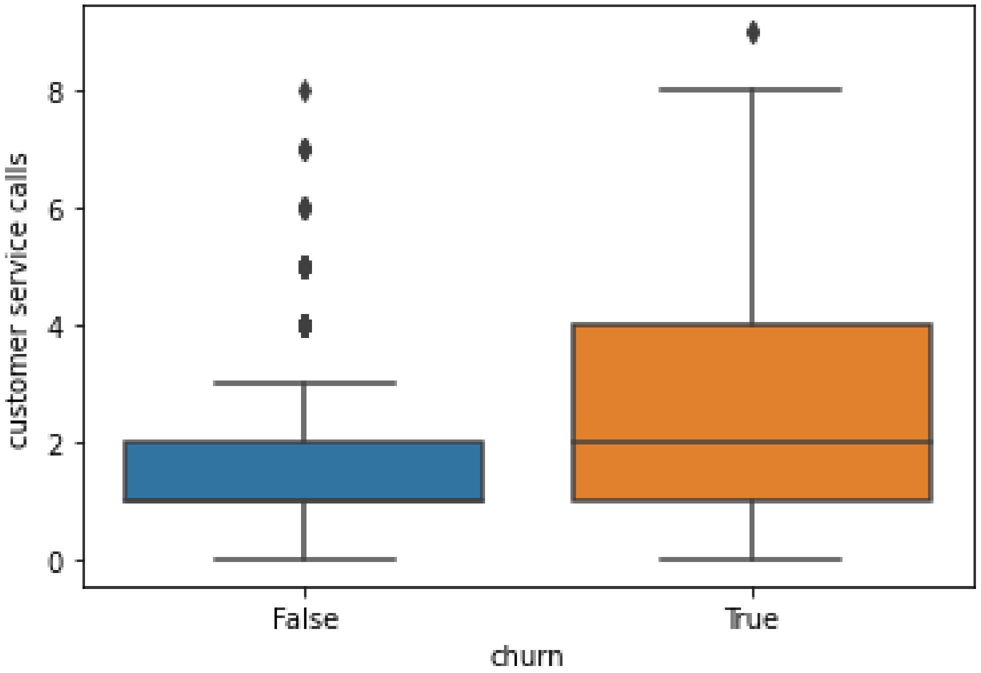
- Losing customers reduces lifetime revenue
- Retention efforts are costly
- Need to prioritize high-risk customers
- Goal: Improve intervention efficiency

Data Understanding



The dataset shows significantly more customers staying than churning. Although churners represent a smaller portion of the dataset, they account for a disproportionate share of revenue risk. This highlights the importance of focusing on churn detection rather than relying solely on overall accuracy.

Data Understanding .contd



Customers who churn tend to make more customer service calls compared to those who remain. This suggests that increased service interactions may reflect dissatisfaction or unresolved issues, making it a strong early indicator of churn risk.

Data Preparation

Data was prepared carefully to avoid leakage and ensure a fair evaluation. To improve the detection of churners, I also handled the class imbalance cause churn cases were fewer than the non-churn

- Removed identifiers – phone number, area code, state
- Encoded categorical variables
- Balanced class handling
- Train / Validation / Test split

Modeling

➢I began with simple baseline models to establish a performance benchmark. I built a Logistic Regression model to provide an interpretable, probability-based churn prediction.

➢I then introduced a Decision Tree model to capture potential nonlinear patterns in customer behavior.

➢Finally, tuned the models particularly focusing on improving recall to ensure we detected more at-risk customers. Each iteration was driven by the goal of increasing business value, not just improving accuracy.

Model Evaluation

```
Final Tuned Model
              precision    recall  f1-score   support

       False       0.95      0.76      0.84       570
        True       0.34      0.74      0.47        97

    accuracy                           0.76       667
   macro avg       0.64      0.75      0.66       667
weighted avg       0.86      0.76      0.79       667

Final ROC-AUC: 0.8149032374751312
```

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

       False       0.88      0.96      0.92       570
        True       0.55      0.25      0.34        97

    accuracy                           0.86       667
   macro avg       0.71      0.61      0.63       667
weighted avg       0.83      0.86      0.84       667

Baseline ROC-AUC: 0.8178875022608065
```

Baseline Logistic Regression compared to Tuned Logistic Regression

Key Findings

Baseline accuracy: **86%**, but only detected **25% of churners**

Final model detects **74% of churners**

ROC-AUC remains strong at **~0.81**

Improved churn detection outweighs slight drop in accuracy

Model Evaluation .contd

Decision Tree Classification Report

	precision	recall	f1-score	support
False	0.94	0.95	0.95	570
True	0.69	0.63	0.66	97
accuracy			0.91	667
macro avg	0.82	0.79	0.80	667
weighted avg	0.90	0.91	0.90	667

Decision Tree ROC-AUC: 0.7907487791644059

Tuned Decision Tree Classification Report

	precision	recall	f1-score	support
False	0.94	0.96	0.95	570
True	0.76	0.64	0.69	97
accuracy			0.92	667
macro avg	0.85	0.80	0.82	667
weighted avg	0.91	0.92	0.91	667

Tuned Decision Tree ROC-AUC: 0.803734852595406

After tuning the Decision Tree, we achieved improvements in both accuracy and precision, while maintaining strong churn detection. The ROC-AUC also increased slightly, indicating better overall predictive discrimination.

Recall (Churn detection) improved from **63% → 64%**

Precision improved from **69% → 76%** higher precision reduces unnecessary retention effort

Overall accuracy increased from **82% → 85%**

ROC-AUC improved from **0.79 → 0.80**

Tuned model provides more balanced performance

Final Model Selection

Final Tuned Model

	precision	recall	f1-score	support
False	0.95	0.76	0.84	570
True	0.34	0.74	0.47	97
accuracy			0.76	667
macro avg	0.64	0.75	0.66	667
weighted avg	0.86	0.76	0.79	667

Final ROC-AUC: 0.8149032374751312

Tuned Decision Tree Classification Report

	precision	recall	f1-score	support
False	0.94	0.96	0.95	570
True	0.76	0.64	0.69	97
accuracy			0.92	667
macro avg	0.85	0.80	0.82	667
weighted avg	0.91	0.92	0.91	667

Tuned Decision Tree ROC-AUC: 0.803734852595406

Final Model Selection: Tuned Logistic Regression

Detects **74% of churners** (highest recall)

Strong predictive stability (ROC-AUC ~0.81)

Better aligned with revenue protection goals

Provides probability-based risk scoring

More suitable for targeted retention strategy

Business Interpretation

Although the Decision Tree achieved higher overall accuracy, it missed more customers who were about to churn. The Logistic Regression model identifies a greater proportion of at-risk customers, making it more effective for proactive retention efforts. Since missing a churner directly impacts revenue, the Logistic model provides stronger business value.

In churn prediction, missing a customer who is about to leave is more costly than contacting a customer who would have stayed. Therefore, recall is more critical than overall accuracy.