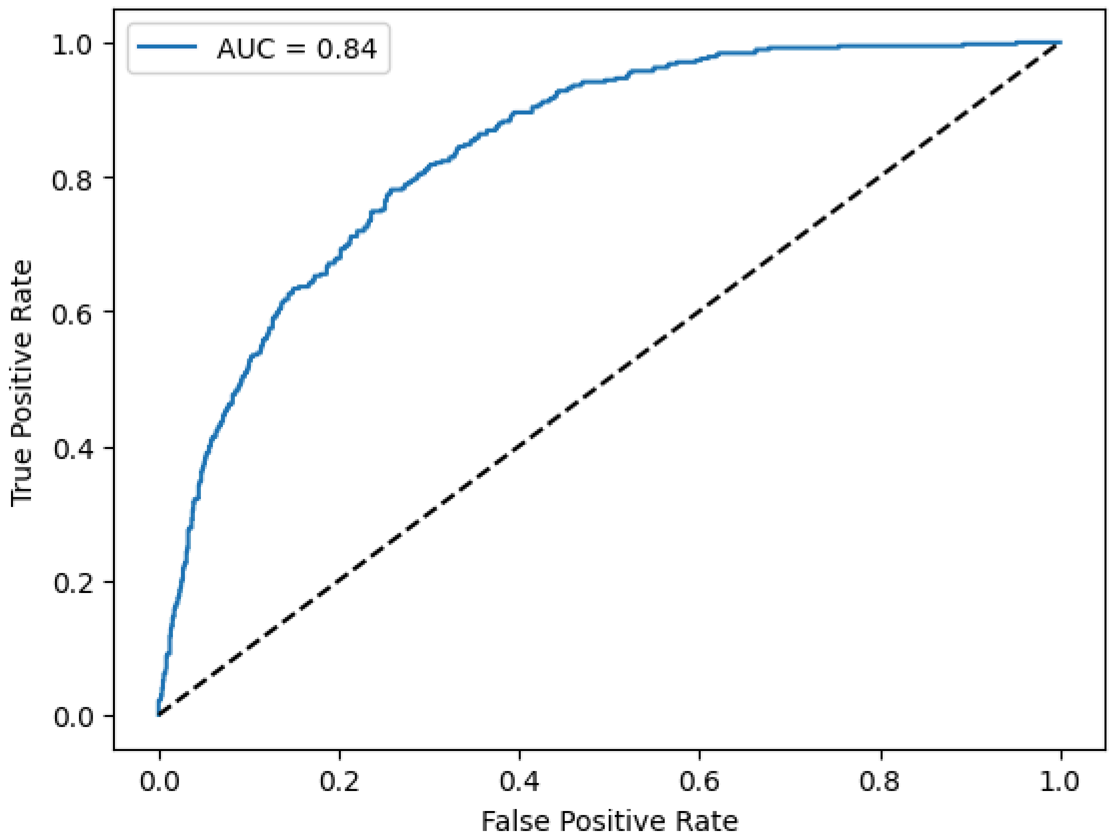
Customer Churn Prediction

What we did

We used the Telco Customer Churn dataset and trained a Logistic Regression model (80/20 split, one-hot encoding for categories). The goal was to predict whether a customer would churn.

How the model performed (test set)



- Accuracy: 0.795
- Precision (churn): 0.634
- Recall (churn): 0.537
- F1: 0.582
- ROC-AUC: ~0.84 (see ROC plot)
- Confusion matrix:
    - True negatives: 919
    - False positives: 116
    - False negatives: 173

    - True positives: 201

Observation: The model is good at avoiding false alarms (few non-churners flagged) but misses some actual churners. If catching more churners is the priority, we should lower the decision threshold slightly.

What seems to drive churn (high level)

- Higher risk: Month-to-month contracts, higher monthly charges, short tenure, electronic check payments, fiber internet, and not having support/security add-ons.
- Lower risk: One- or two-year contracts, longer tenure, automatic payments, and having support/security add-ons.

Mistakes the model makes

- False positives (116): Some loyal customers get flagged—cost is mostly outreach time/discounts.
- False negatives (173): Some churners are missed—this is usually more costly for the business.

Possible fix that we can try: Move the decision threshold below 0.50 to catch more churners, accepting a few more false positives.

Recommendations to reduce churn

1. Contracts & pricing: Offer incentives to move month-to-month → annual; provide bundle or loyalty discounts for high monthly bills.
2. Onboarding & early care: Focus on new, short-tenure customers with check-ins in the first 1–3 months.
3. Payments: Nudge electronic-check users to switch to autopay (card/bank) with small rewards.
4. Service bundles: Promote Tech Support and Security/Backup add-ons to at-risk segments.
5. Fiber customers: Provide value messaging or tailored bundles to offset higher perceived cost.

Quick modeling improvements (next pass)

- Drop customerID and convert TotalCharges to numeric (and fill any missing).
- Try class_weight='balanced' or a lower threshold to boost recall.
- Compare with Random Forest or XGBoost and pick the best on validation.

Conclusion

With AUC ~0.84, the model already ranks risk well. A small threshold tweak and the quick fixes above should help catch more churners and guide targeted retention offers.