PlotStudio

Churn Prediction, Risk Profiling, and Lift Chart Evaluation

Apr 5, 2026 · 4m 37s

QUESTION

How can we predict which customers will churn in the next 30 days, create churn risk profiles, and evaluate the model with a lift chart?

Churn Prediction Model Evaluation

Overview

The churn model is strong enough to support targeted retention action: it materially outperformed the naive baseline and ranked likely churners well, with a test ROC-AUC of `0.842`. In business terms, the model is not just better at classifying churn overall — it is good at pushing real churners toward the top of the outreach list, which is exactly what matters for a lift-based retention program.

Key Patterns & Observations

The naive benchmark was operationally useless despite decent-looking accuracy. Because churn is only about a quarter of the customer base, a majority-class baseline reached `0.735` accuracy simply by predicting retention for everyone, but its precision and recall were both `0.000`, and ROC-AUC was `0.500`.

Model	Accuracy	Precision	Recall	F1	ROC-AUC
Naive baseline	0.735	0.000	0.000	—	0.500
Churn model	0.807	0.658	0.567	0.609	0.842

The trained model added real detection power, not just cosmetic metric improvement. Precision of `0.658` means about two out of three customers flagged as likely churners actually churned in the test set, while recall of `0.567` means the model captured more than half of all true churners. That is a meaningful jump from the baseline's complete failure to identify churners.

Lift is strong enough to make prioritization worthwhile. The cumulative gains curve **Figure 1**

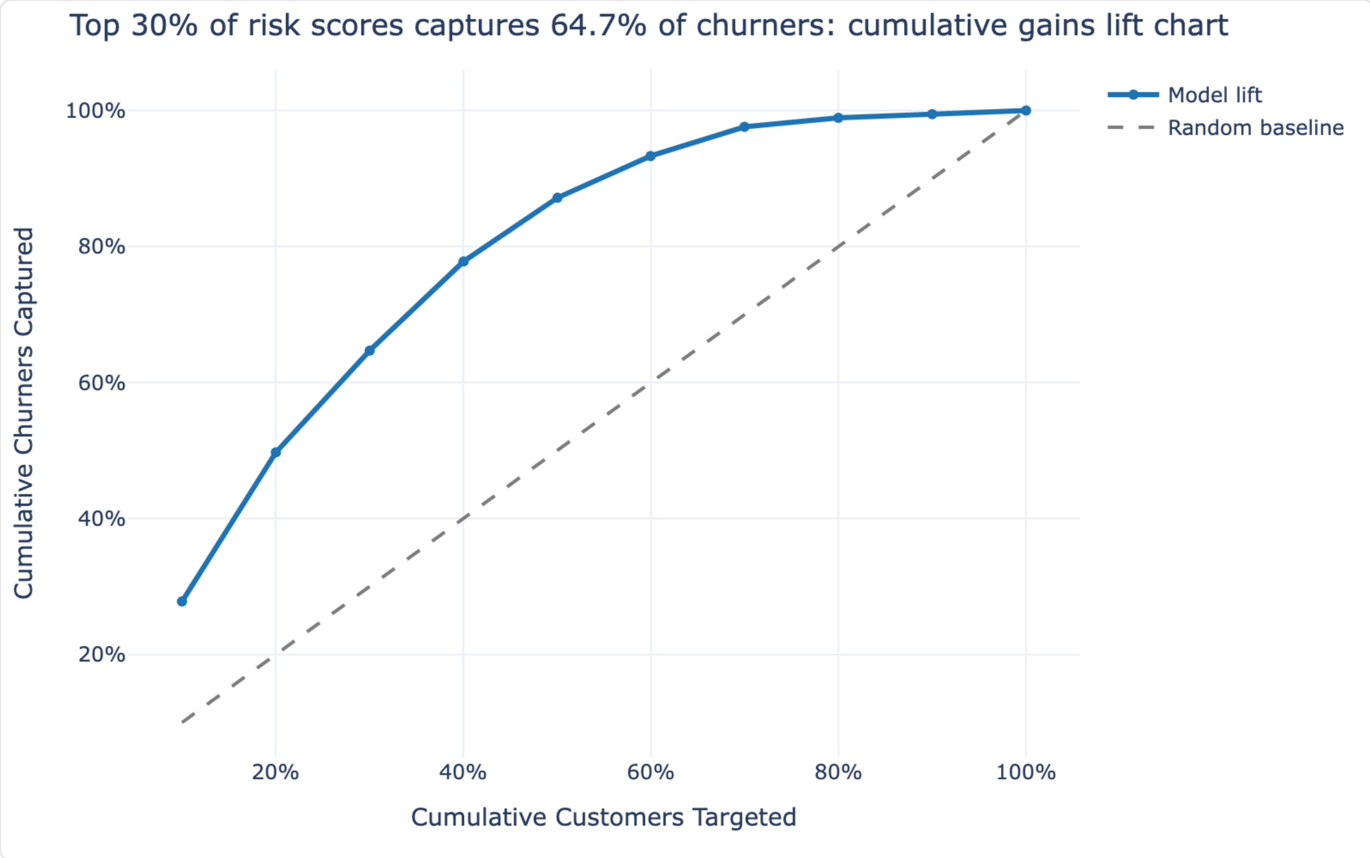

Figure 1

shows that targeting the top `30%` of customers by predicted risk captures `64.7%` of actual churners in the test sample.

Lift metric	Value
Test customers	1409
True churners in test set	374
Churners captured in top 30% of risk scores	64.7%

That means the model is concentrating churn risk more than twice as efficiently as random targeting, which makes it suitable for ranked retention campaigns when budget or contact capacity is limited.

Statistical Insights

The classification model was evaluated on a stratified holdout sample, so the reported results reflect out-of-sample performance rather than in-sample fit. The most decision-relevant ranking metric is:

$$\text{ROC-AUC} = 0.842$$

This means a randomly chosen churner is assigned a higher predicted risk than a randomly chosen non-churner about `84.2%` of the time — strong ranking performance for a business-facing churn model.

The strongest model drivers were:

Feature	Direction of effect	Interpretation
tenure	Lowers churn risk	**Dominant stabilizer** — longer relationships strongly reduce churn risk, reinforcing the early-lifecycle retention problem
MonthlyCharges	Lowers churn risk in multivariate model	Likely reflects overlap with other pricing/service variables; not a simple standalone “higher price = higher churn” story once bundle mix is controlled
InternetService_Fiber optic	Raises churn risk	Fiber customers appear more churn-prone, consistent with higher-price or expectation-sensitive segments
Contract_Two year	Lowers churn risk	**Strong retention anchor** — long commitments materially reduce churn
TotalCharges	Raises churn risk	Higher cumulative spend is associated with churn after controlling for tenure and monthly charges, suggesting interaction with pricing mix or customer value tension
Contract_One year	Lowers churn risk	Fixed-term structure reduces attrition versus month-to-month
StreamingMovies_Yes	Raises churn risk	Entertainment add-ons are associated with elevated churn once other services are held constant
StreamingTV_Yes	Raises churn risk	Similar signal: higher-risk bundle mix rather than broad loyalty
MultipleLines_Yes	Raises churn risk	More complex service setups may correlate with higher bills or more competitive switching behavior
PaperlessBilling_Yes	Raises churn risk	A modest but persistent indicator of higher churn propensity

---

Interpretation & Implications

Tenure and contract structure dominate the churn story. The model’s strongest protective signals are longer tenure and fixed-term contracts, which means the retention opportunity is concentrated among newer, less-committed customers rather than evenly spread across the base. That argues for proactive onboarding, early-tenure support, and month-to-month conversion offers as the first line of defense.

High-risk commercial patterns are clustered, not random. Fiber optic service, paperless billing, streaming add-ons, and multiple lines all point to a segment that may be more price-sensitive or more willing to switch despite higher product engagement. This is important because it suggests churn is not just about low-value customers leaving; some richer service bundles may still be fragile if perceived value does not match the bill.

The lift result makes the model actionable. If the business can only intervene on a fraction of accounts each month, focusing on the top-risk tranche should recover a disproportionate share of likely churners. In practice, that means the next step should be risk tiering and revenue-at-risk analysis, so retention offers can be aligned to both likelihood of churn and financial exposure.

The bottom line is that this model has enough discrimination to support prioritized churn prevention, and the strongest levers are early-tenure customers on month-to-month terms with riskier service and billing profiles.

Churn Risk Profiles and Revenue at Risk

Overview

The model can be turned into a practical retention playbook: the top `10%` of scored customers form a clearly defined High-risk segment with a realized churn rate of `73.8%`, far above the `48.9%` Medium tier and `13.4%` Low tier. That means the probability scores are not just mathematically useful—they create actionable customer groups the business can target differently.

The customer-level risk assignment table `df_risk_profiles` shows `141` High-risk, `282` Medium-risk, and `986` Low-risk customers in the holdout sample. High-risk customers are concentrated among **very early-tenure accounts** with median tenure of `4` months, are predominantly **month-to-month**, most often pay by **electronic check**, and are overwhelmingly **fiber optic** subscribers. They also show strikingly low adoption of retention-oriented services such as **OnlineSecurity (`0.7%`)** and **TechSupport (`2.8%`)**, which reinforces the earlier model result that contractual commitment and support-related services are major stabilizers.

Key Patterns & Observations

The High-risk segment is narrow, young, and commercially fragile. Its average *MonthlyCharges* is `82.18`, well above the Low-risk group average of `57.97`, so these are not low-value customers. Instead, they appear to be newer, higher-bill customers whose perceived value may not yet justify their spend.

Risk Tier	Customers	Actual Churn Rate	Average Monthly Revenue
High	141	73.8%	82.18
Medium	282	48.9%	76.44
Low	986	13.4%	57.97

Service mix helps explain why these customers are unstable. In the High-risk tier, `95.7%` have fiber optic internet and `60.3%` have multiple lines, but very few have protection or support add-ons. That pattern suggests customers are buying into relatively expensive core service packages without the stickier support features that often deepen engagement.

High-Risk Profile Metric	Value
Customers	141
Median tenure (months)	4.0
Average MonthlyCharges	82.18
Most common Contract	Month-to-month
Most common PaymentMethod	Electronic check
Fiber optic adoption	95.7%
Multiple lines adoption	60.3%
Online security adoption	0.7%
Tech support adoption	2.8%
Streaming TV adoption	41.8%
Streaming movies adoption	51.1%

Revenue exposure is split between urgency and scale. The revenue summary `df_revenue_at_risk_by_tier` and bar chart **Figure 2**

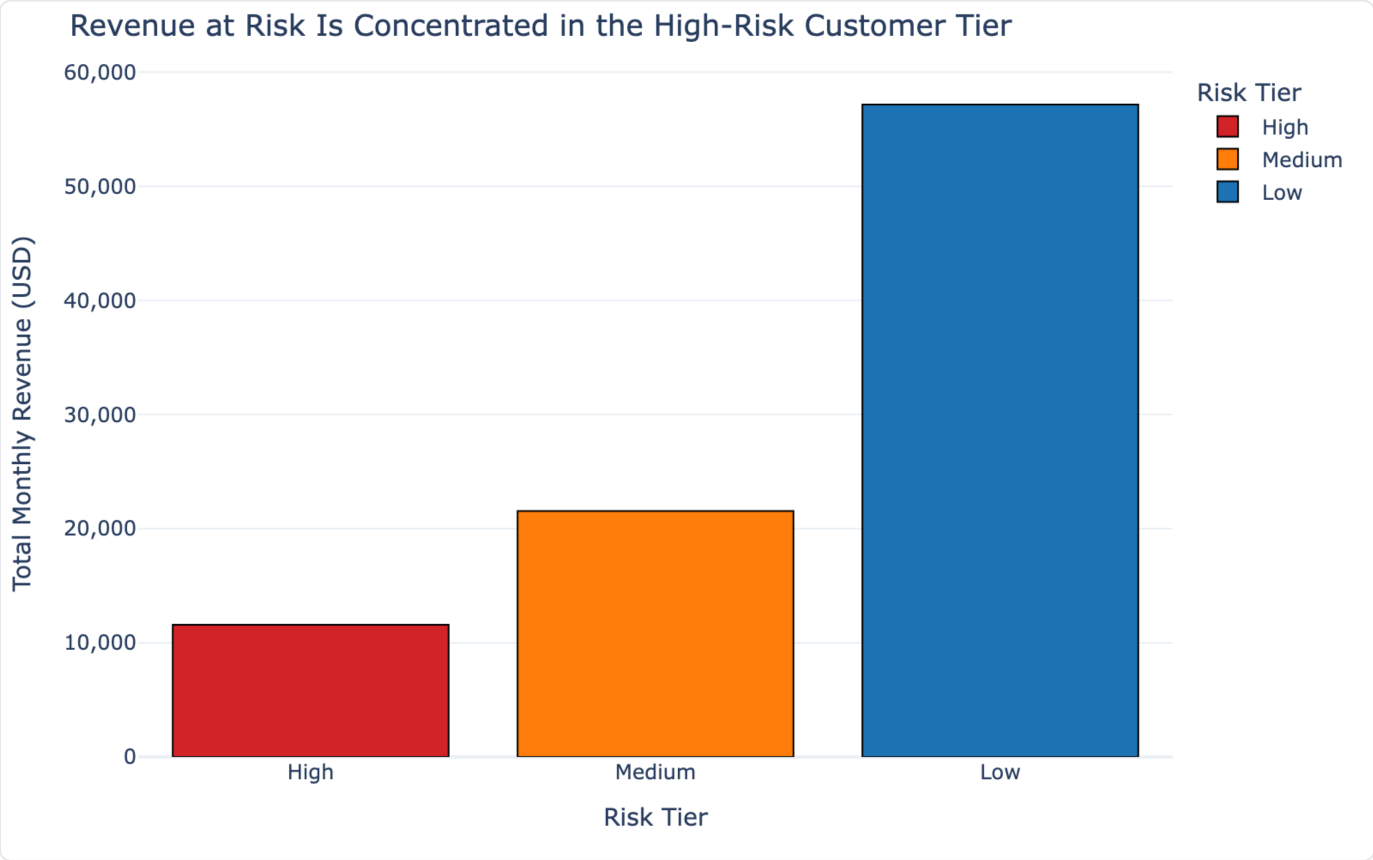

Figure 2

show that the Low-risk tier contains the most total monthly revenue simply because it is much larger, but that should not distract from the High-risk segment's urgency. High-risk customers contribute `11,587.80` in monthly revenue with the highest average revenue per customer, which makes them the most attractive group for immediate intervention on a per-contact basis.

Risk Tier	Customers	Total Monthly Revenue	Average Monthly Revenue	Actual Churn Rate
High	141	11,587.80	82.18	73.8%
Medium	282	21,555.40	76.44	48.9%
Low	986	57,158.00	57.97	13.4%

Statistical Insights

The risk tiers create a strong monotonic separation in realized churn:

$$P(\text{churn} \mid \text{High}) = 73.8\%, \quad P(\text{churn} \mid \text{Medium}) = 48.9\%, \quad P(\text{churn} \mid \text{Low}) = 13.4\%$$

This tier ordering matters because it validates that the predicted probabilities are well calibrated directionally: as model-assigned risk falls, realized churn falls sharply as well. In operational terms, the model is not merely sorting customers randomly into buckets—it is creating meaningful exposure bands with large differences in observed attrition.

Interpretation & Implications

The most important retention opportunity is early-life, high-bill, month-to-month customers. These customers have not yet built tenure, are not protected by longer contracts, and often lack sticky support services. That combination makes them the clearest candidates for onboarding interventions, value-justification messaging, or targeted offers that convert them to more stable contract types.

Medium risk should be treated as the scale play, while High risk is the precision play. The High tier is small but extremely concentrated with likely churners, making it ideal for more expensive interventions such as proactive outreach or personalized offers. The Medium tier is larger and carries more total revenue than High, so it is better suited to lower-cost campaigns such as automated retention journeys, digital nudges, or service bundle promotions.

The bottom line: if the business wants the fastest churn-prevention payoff, it should focus first on new, fiber-heavy, month-to-month customers paying by electronic check and lacking security or tech-support add-ons, because that segment combines the highest churn probability with meaningful monthly revenue exposure.

Data Sources

- df_risk_profiles
- df_revenue_at_risk_by_tier