

SAMPLE DELIVERABLE • CHURN / PROPENSITY MODEL

Predicting 60-day churn at Contoso Play

Fictional subscription video service • 10,000 synthetic subscribers • Python / scikit-learn • SHAP explainability

SAMPLE - synthetic data - fictional company

CHURN PROPENSITY MODEL

Who will cancel in the next 60 days, and why?

Contoso Play (fictional subscription service) • 10,000 subscribers • gradient boosting + SHAP

AUC (HOLD-OUT)

0.80

Gini 0.59

TOP-DECILE LIFT

3.5x

vs average churn rate

CHURNERS IN TOP 30%

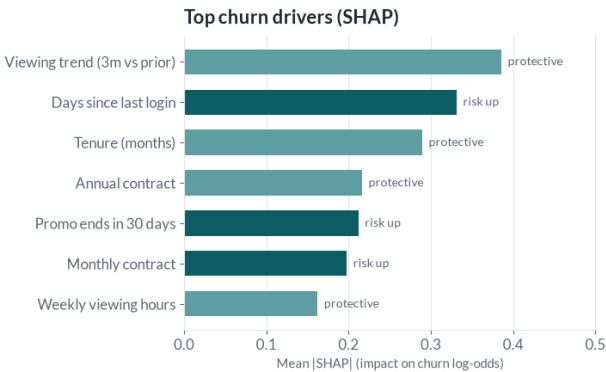
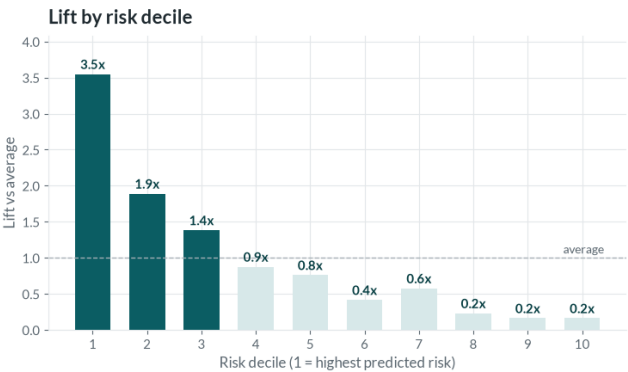
68%

of all hold-out churners

BEST CAMPAIGN TARGET

Top 15%

net INR 28.5 L per 100k subs



SAMPLE - synthetic data, fictional company • Contoso Play is fictional; model trained on simulated data

Shashwat Kumar • Illustrative work sample

Executive summary

Contoso Play loses about **14.5%** of its subscribers every 60 days. We built a gradient-boosting model that ranks every subscriber by their risk of cancelling in the next 60 days. On a held-out 30% sample it scores an **AUC of 0.80**. The riskiest 10% of customers churn at **3.5x** the average rate, and the top 30% contains **68%** of all churners.

- **What drives churn:** viewing trend (3m vs prior), days since last login and tenure (months) carry the most signal, followed by contract type and promo expiry.
- **Where to act:** a retention offer aimed at the **top 15%** of scored customers maximises net value: about **INR 28.5 L** per 100,000 subscribers per cycle at 1.0x ROI, under the stated assumptions (page 5).
- **Explainable:** every score comes with its top reasons, so CRM can match the offer to the reason (payment fix, content nudge, plan switch).

Recommendation: run a 6-week randomised pilot on the top 15%, holding out 20% of them as a control, before rolling out.

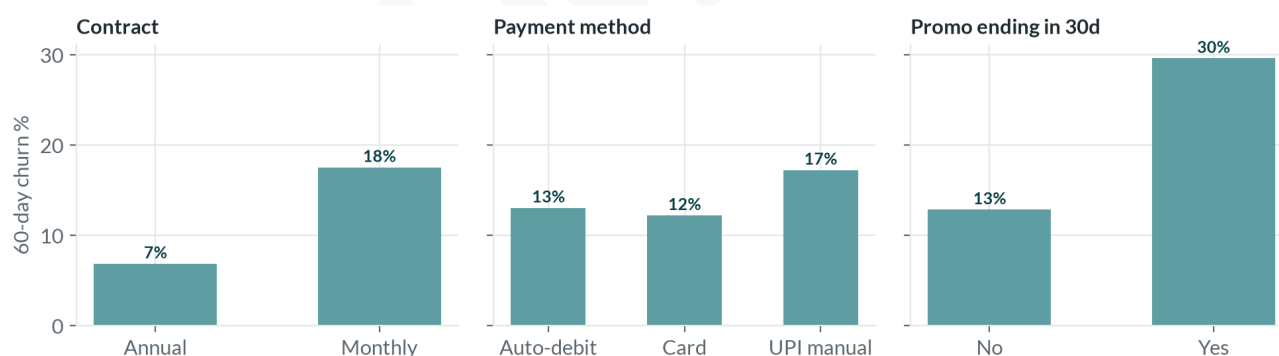
DATA & FEATURES

One row per subscriber, one snapshot date

We took a snapshot of each active subscriber with their behaviour up to that date, then labelled whether they churned in the following 60 days. Features use only information available at the snapshot, so nothing leaks from the future. The data is synthetic but follows patterns typical of subscription businesses.

Feature	Description	Type
tenure_months	Months since first subscription	Numeric
contract	Monthly or annual plan	Category
plan / monthly_fee	Basic / Standard / Premium and price paid (INR)	Category / numeric
payment_method	Auto-debit, manual UPI or card	Category
weekly_hours	Average viewing hours per week (last 4 weeks)	Numeric
viewing_trend_pct	% change in viewing, last 3 months vs prior 3	Numeric
days_since_login	Days since the last app or web session	Numeric
support_tickets_90d	Support tickets raised in 90 days	Count
failed_payments_6m	Failed or bounced renewals in 6 months	Count
devices / profiles	Linked devices and viewer profiles	Count
discount_active / promo_ends_30d	On a promo price; promo price ends within 30 days	Flag
region / age_band	Metro / Tier-2 / Tier-3; age band	Category
churn_60d (target)	Cancelled or lapsed within 60 days of the snapshot	Binary

Churn rate by key segment



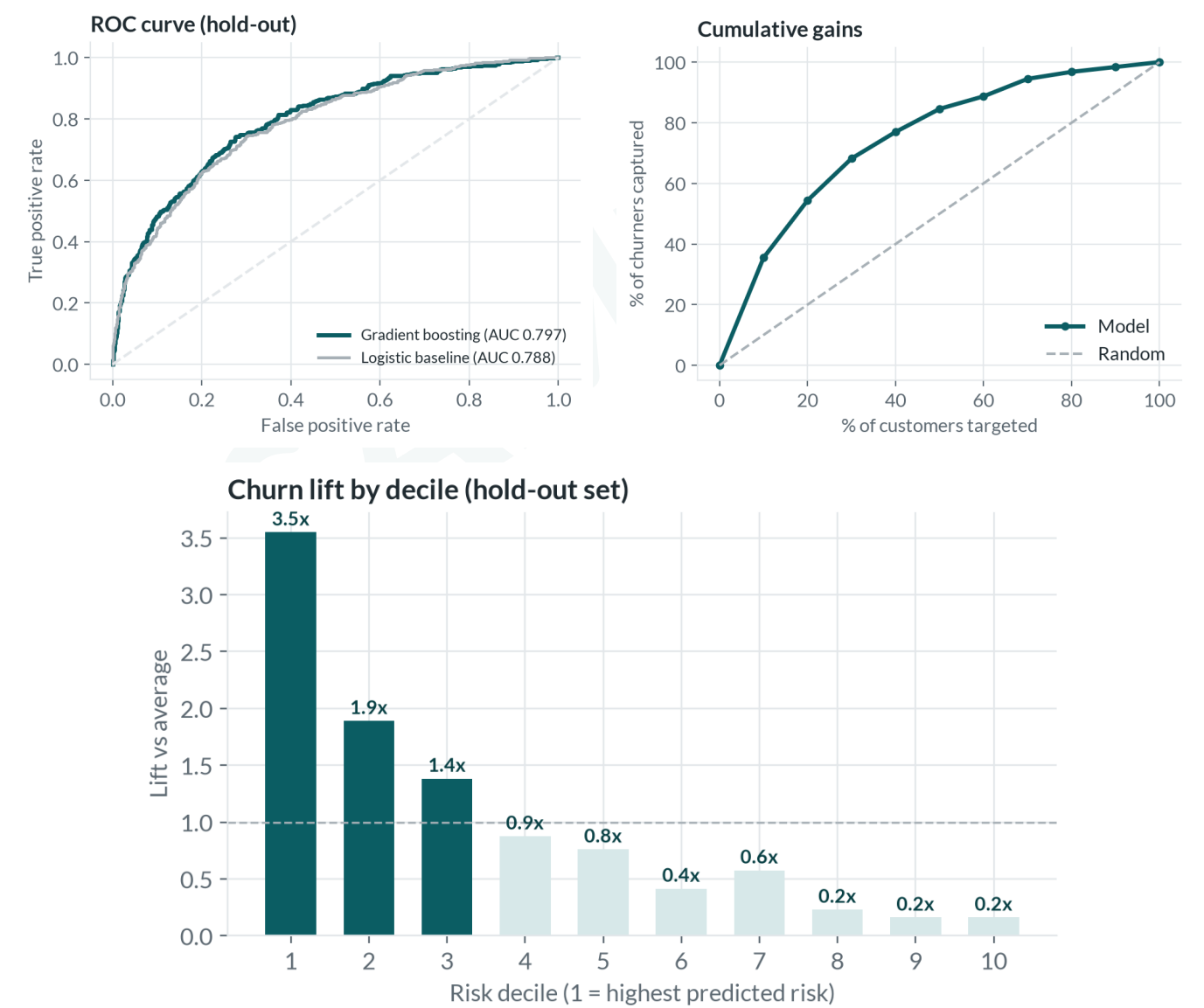
Monthly contracts, manual UPI payers and subscribers whose promo price is about to end churn far more often. These single-variable cuts are a starting point. The model combines them with behavioural signals to rank each person.

MODEL PERFORMANCE

Ranks risk well and stays honest on unseen data

Model: scikit-learn GradientBoostingClassifier (350 trees, depth 3, learning rate 0.05, 80% subsampling), trained on 70% of the data with a stratified split and evaluated on the remaining 30%. A logistic regression is the transparent baseline.

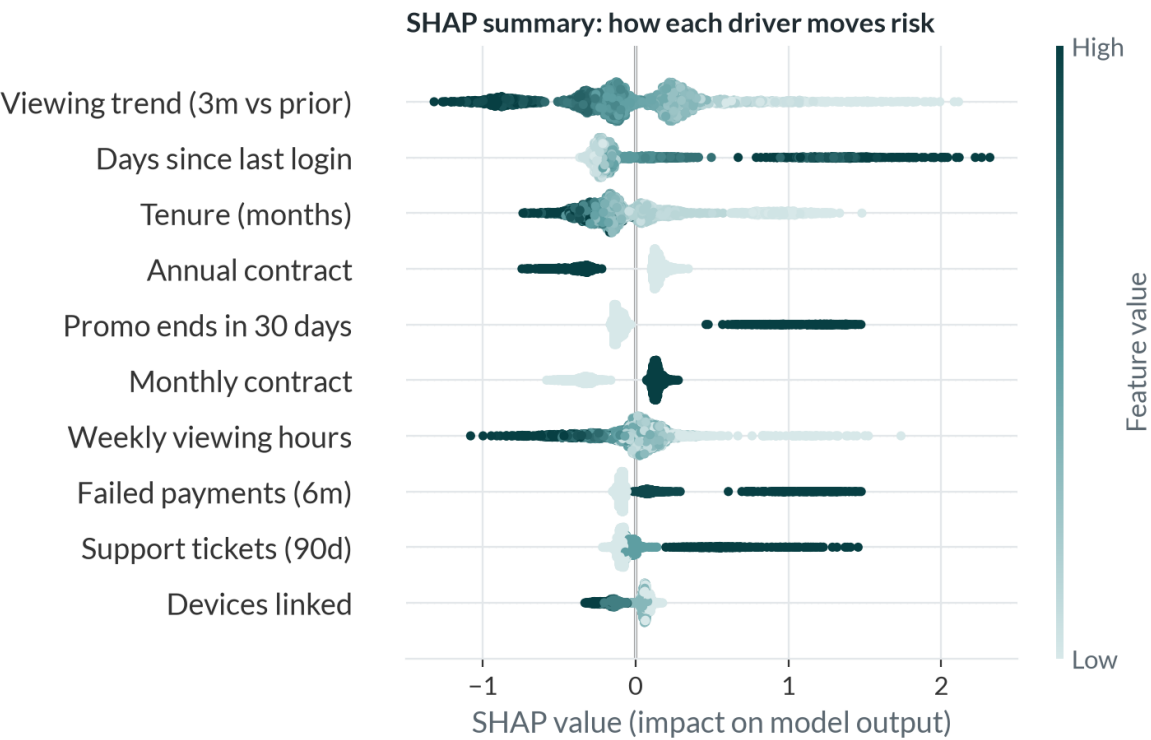
Metric	Gradient boosting	Logistic baseline	What it means
AUC (hold-out)	0.797	0.788	Chance that a random churner is ranked above a random stayer
Gini	0.594	0.576	$2 \times \text{AUC} - 1$, a common credit-risk convention
PR-AUC	0.447	-	vs 0.145 for random ranking
Brier score	0.101	-	Accuracy of the probabilities (lower is better)
Train AUC	0.859	-	Modest gap to hold-out, so overfitting is limited
Top-decile lift	3.55x	-	Churn rate in the riskiest 10% vs average



Boosting beats the baseline by a modest margin (0.797 vs 0.788) because it captures threshold effects, such as risk jumping after two weeks without a login. In a real project I would ship the simpler model if the gain did not justify the extra complexity.

CHURN DRIVERS

Why customers leave, in the model's own words



Each dot is one customer. Dots to the right raise churn risk; darker colour means a higher feature value. Falling viewing, long gaps since the last login and short tenure push risk up. Annual contracts, heavy viewing and multiple devices protect against it.

#	Driver	Mean [SHAP]	Perm. importance (AUC drop)	Direction
1	Viewing trend (3m vs prior)	0.386	0.0441	lowers risk when high
2	Days since last login	0.331	0.0489	raises risk when high
3	Tenure (months)	0.290	0.0366	lowers risk when high
4	Annual contract	0.217	0.0151	lowers risk when high
5	Promo ends in 30 days	0.212	0.0304	raises risk when high
6	Monthly contract	0.198	0.0120	raises risk when high
7	Weekly viewing hours	0.163	0.0077	lowers risk when high
8	Failed payments (6m)	0.143	0.0137	raises risk when high

Permutation importance (the AUC lost when a feature is shuffled) agrees with the SHAP ranking. That cross-check tells us the drivers come from the data, not from a quirk of one method.

THRESHOLD & ROI

How many customers to target

The model gives a ranking; the business picks the cut-off. The table below scales hold-out results to a base of 100,000 subscribers for one 60-day cycle.

Target top	Score cut-off	Contacted	Churners reached	Precision	Recall	Saves	Cost	Benefit	Net value	ROI
5%	0.51	5,000	3,167	63%	22%	1,108	INR 9.3 L	INR 27.4 L	INR 18.1 L	1.9x
10%	0.37	10,000	5,133	51%	35%	1,797	INR 18.9 L	INR 43.4 L	INR 24.4 L	1.3x
15%	0.27	15,000	6,833	46%	47%	2,392	INR 28.3 L	INR 56.9 L	INR 28.5 L	1.0x
20%	0.22	20,000	7,867	39%	54%	2,753	INR 38.0 L	INR 65.2 L	INR 27.2 L	0.7x
30%	0.15	30,000	9,867	33%	68%	3,453	INR 56.9 L	INR 81.4 L	INR 24.6 L	0.4x
40%	0.10	40,000	11,133	28%	77%	3,897	INR 77.2 L	INR 91.3 L	INR 14.1 L	0.2x
50%	0.08	50,000	12,233	24%	85%	4,282	INR 96.5 L	INR 100.7 L	INR 4.2 L	0.0x

Assumptions (illustrative): offer = 2 months at 50% off (cost 1 × monthly fee per acceptor); contact cost INR 15; the offer saves 35% of true churners contacted; 50% of non-churners contacted also accept (pure cost); a save is worth 6 months of fee.

Net value peaks at the **top 15%** (score ≥ 0.27). Going wider still reaches more churners, but each extra rupee spent returns less, because more and more offers go to customers who would have stayed anyway.

Deployment & monitoring

- Weekly batch scoring in the warehouse; scores and top-3 reasons pushed to CRM for each subscriber.
- Randomised hold-out (20% of the targeted group) to measure true incremental saves, not just correlation.
- Monitor monthly: AUC and decile lift on fresh labels, population stability index (PSI) on key features, offer take-up and save rate.
- Retrain every quarter, or sooner if PSI goes above 0.2 or AUC drops by more than 0.03.

Handover pack: notebook and script (churn_model.py / .ipynb), feature dictionary, model card, scoring SQL, and a 60-minute walkthrough.

About this sample: Contoso Play is a fictional company and every subscriber record was generated synthetically by the accompanying script. Results show the method and are not benchmarks for any real business. Prepared by Shashwat Kumar as an illustrative work sample.