Capstone Project

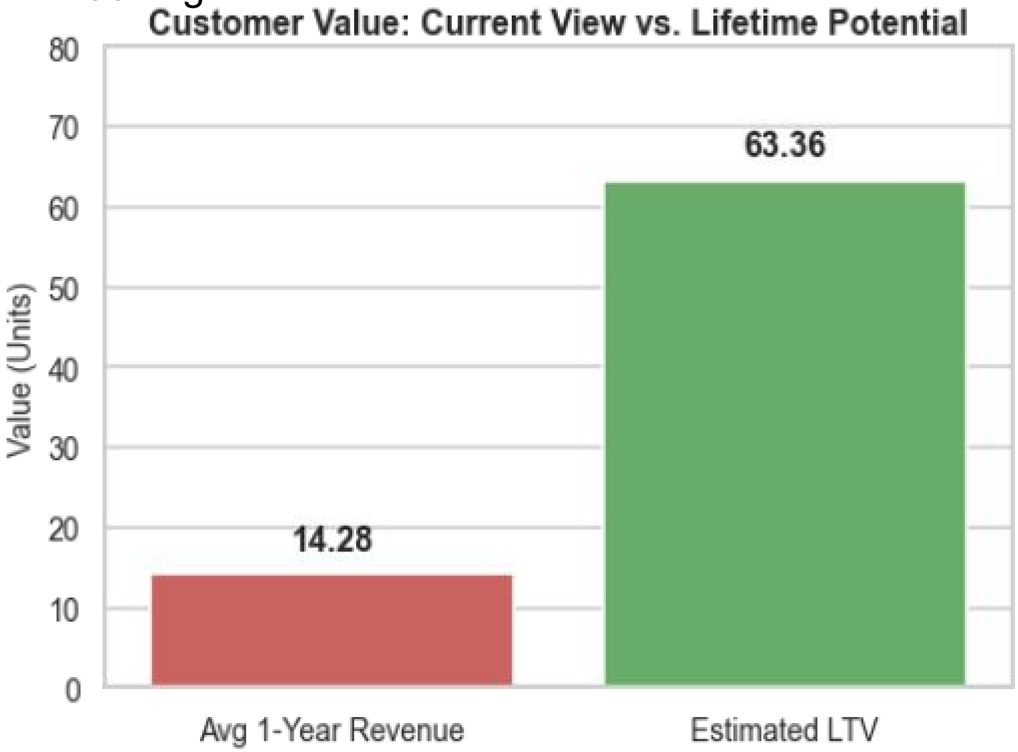
Final Presentation

PART 01

WHAT'S THE PROBLEM

The Blind Spot in SaaS Practices

Paylocity's current view of customer value is backward-looking.



- SaaS companies often know **how much they made**, not **how much value clients still hold**
- Even when using LTV, most rely on **averages** — ARPU × retention
- Paylocity makes most decisions based on average **1-year revenue**
- That **downplays retention, growth, and time value of money**

What's the Problem

Misallocation Without Forward Value

Same revenue, different lifetime value.

- There's no way to **tell who's high-value vs short-term**
- Resource allocation becomes **guesswork**
- Retention, upsell, and acquisition aren't targeted
- The current model **misses growth signals and customer heterogeneity**

What’ s the Problem

Data Sources Behind the Model

What powers our analysis: 5 years of billing, behavior, and customer profile data

- 📊 **1. Billing Data (NetSuite)**
  - 5 fiscal years of client-level billing
  - Reflects **cash flow**, not booked revenue
  - *Note: Excludes child-parent mapping (reseller/partner level)*
- 👥 **2. Client Profiles (Salesforce)**
  - Segment, Employee Count, Industry, Region
  - Used for segmentation and CAC estimation
- 🚥 **3. Account Health (Salesforce)**
  - 5 years of status data (Green / Yellow / Red / Grey / Black)
  - Used to infer retention risk and model churn patterns
- 📁 **4. CAC Estimates (Provided)**
  - Segment-level CAC based on **lead → close rates**
  - *Benchmark only — not actual spend*

PART 02

HOW WE APPROACHED IT

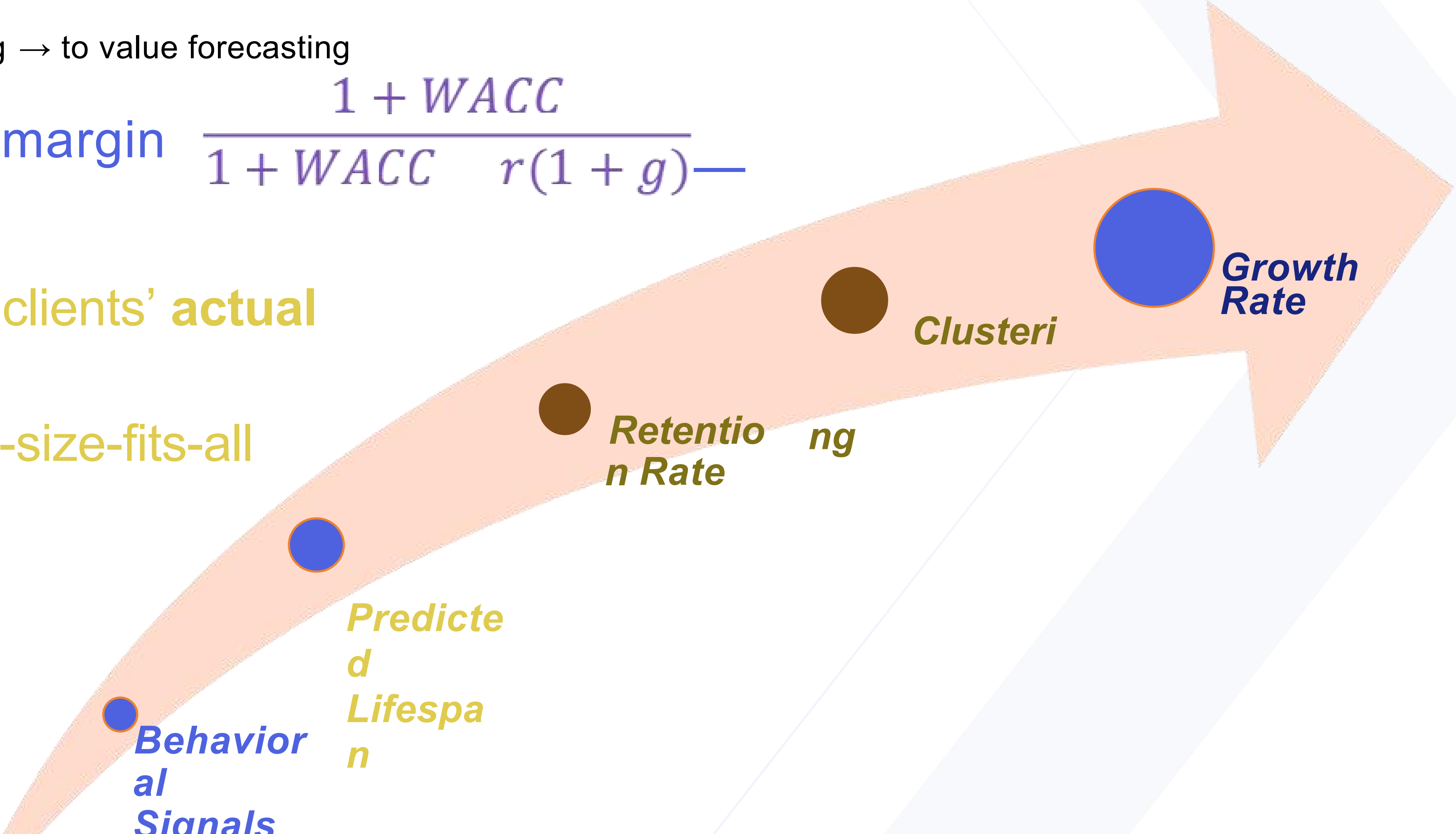
What’s the Problem → How We Approached It → What We Discovered → How It Drives Strategy

A Behavior-Driven Pipeline

From revenue tracking → to value forecasting

$$\text{LTV} = \text{annual margin} \; \frac{1 + WACC}{1 + WACC \quad r(1+g) -}$$

- We modeled clients’ **actual behavior**
- No more one-size-fits-all



Modeling Retention — Predicting How Long Clients Stay

Retention is derived, not assumed

Δ *The Problem We Faced*

- Multi Collinearity
- Limited features
- Low predictive value
- No access to behavioral logs
- Missing transactional signals
- Lacked engagement metrics

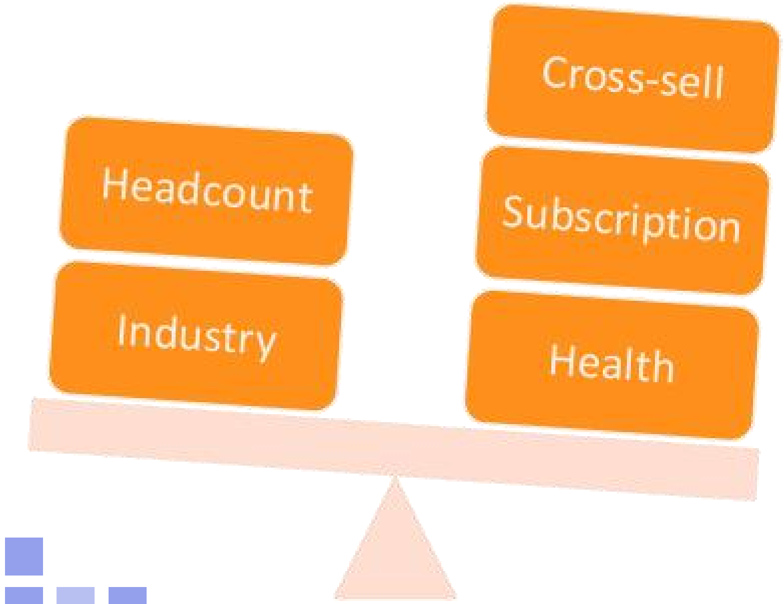
What's the Problem → How We Approached It → What We Discovered → How It Drives Strategy

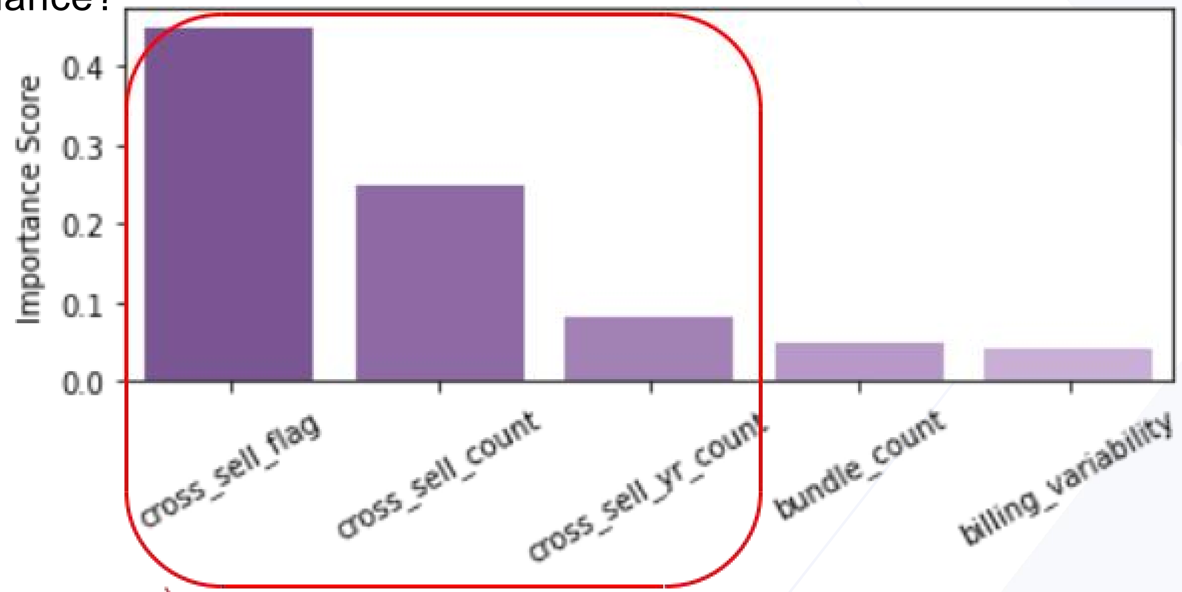
Modeling Retention — Feature Importance (XGBoost) Top 5

What features impacting the model performance?

Predicting Power:

Profile < **Behavior**





💡 *Cross-selling features altogether explain over 75% variance*

Modeling Retention — Results & Performance

How Did Feature Engineering Boost Performance?

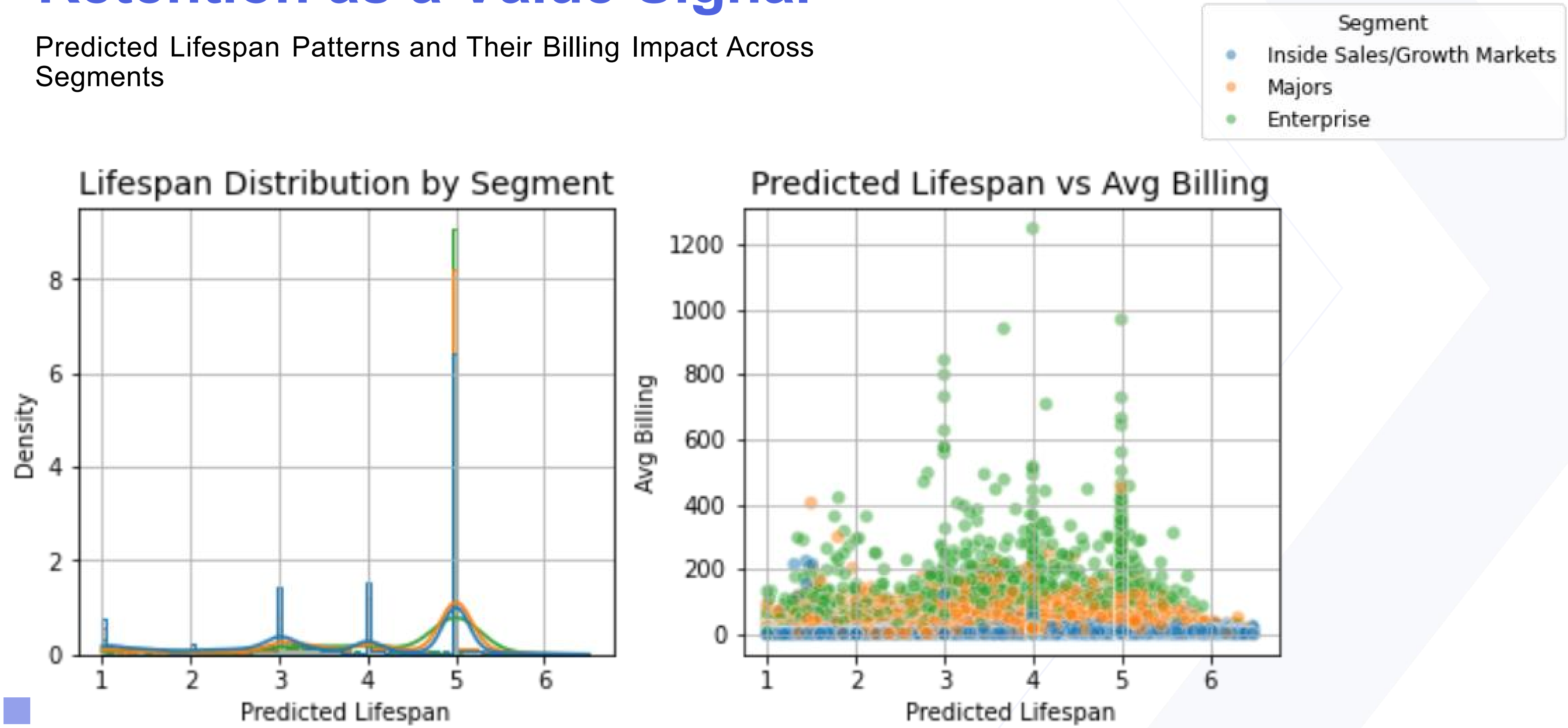
Feature Engineering

Model	R^2 Before FE	R^2 After FE	Impact of FE
OLS	-	0.471	Strong
Ridge	0.0618	0.4711	8x Better
XGBoost	0.2344	0.6812	3x Better

✅ ***Key success:***
Feature engineering increased model performance 8x for Ridge and 3x for XGBoost, proving that creative data engineering can overcome behavioral data limitations and deliver actionable business intelligence.

Retention as a Value Signal

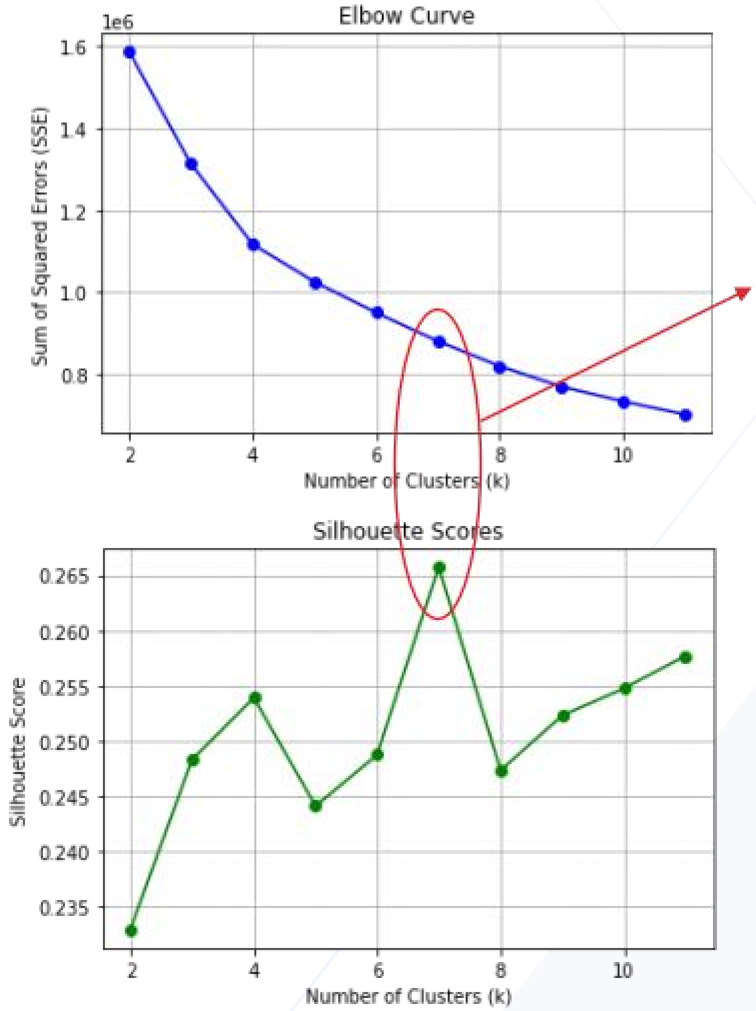
Predicted Lifespan Patterns and Their Billing Impact Across Segments



Growth Assignment — Based on Cluster Behavior, Not Heuristic

Growth isn't adjusted based on heuristic rules — it adapts to cluster performance

- All clients start with baseline growth rate 4.4% (inflation)
- We used **KMeans++ to cluster clients (k = 7)** based on behavioral and billing features
    - Clustering quality was validated using **Elbow** and **Silhouette** methods



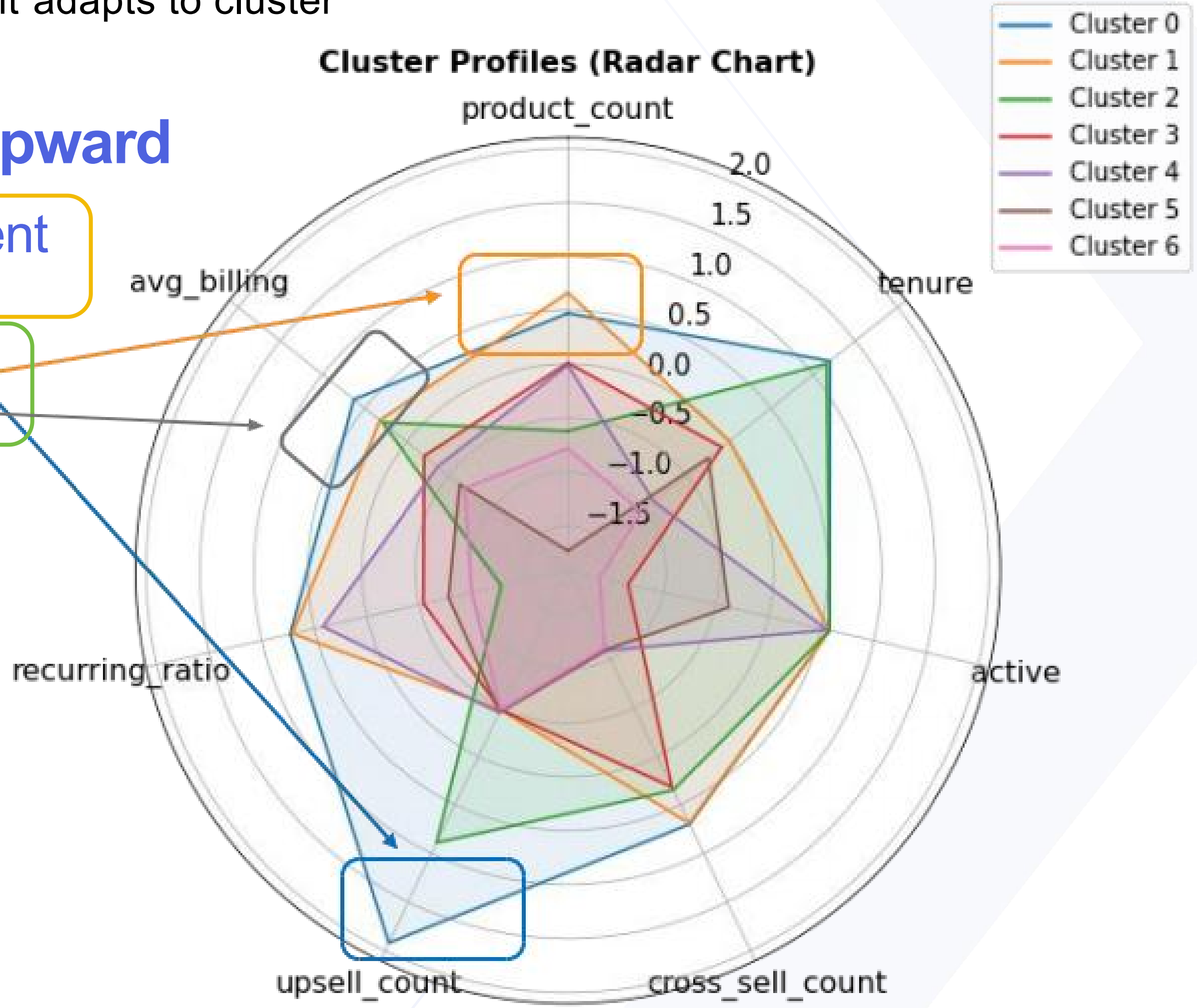
7 clusters offered the best balance of compactness and separation.

Growth Assignment — Based on Cluster Behavior, Not Heuristic

Growth isn't adjusted based on heuristic rules — it adapts to cluster performance

For select clusters, we **adjust upward**

- **Cluster 0:** High Billing + High Engagement
- **Cluster 1:** High Billing + Broad Usage
- **Cluster 4:** New & Active



What's the Problem → *How We Approached It* → *What We Discovered* → *How It Drives Strategy*

Growth Assignment — Based on Cluster Behavior, Not Heuristic

Growth isn't adjusted based on heuristic rules — it adapts to cluster performance

For select clusters, we **adjust upward**

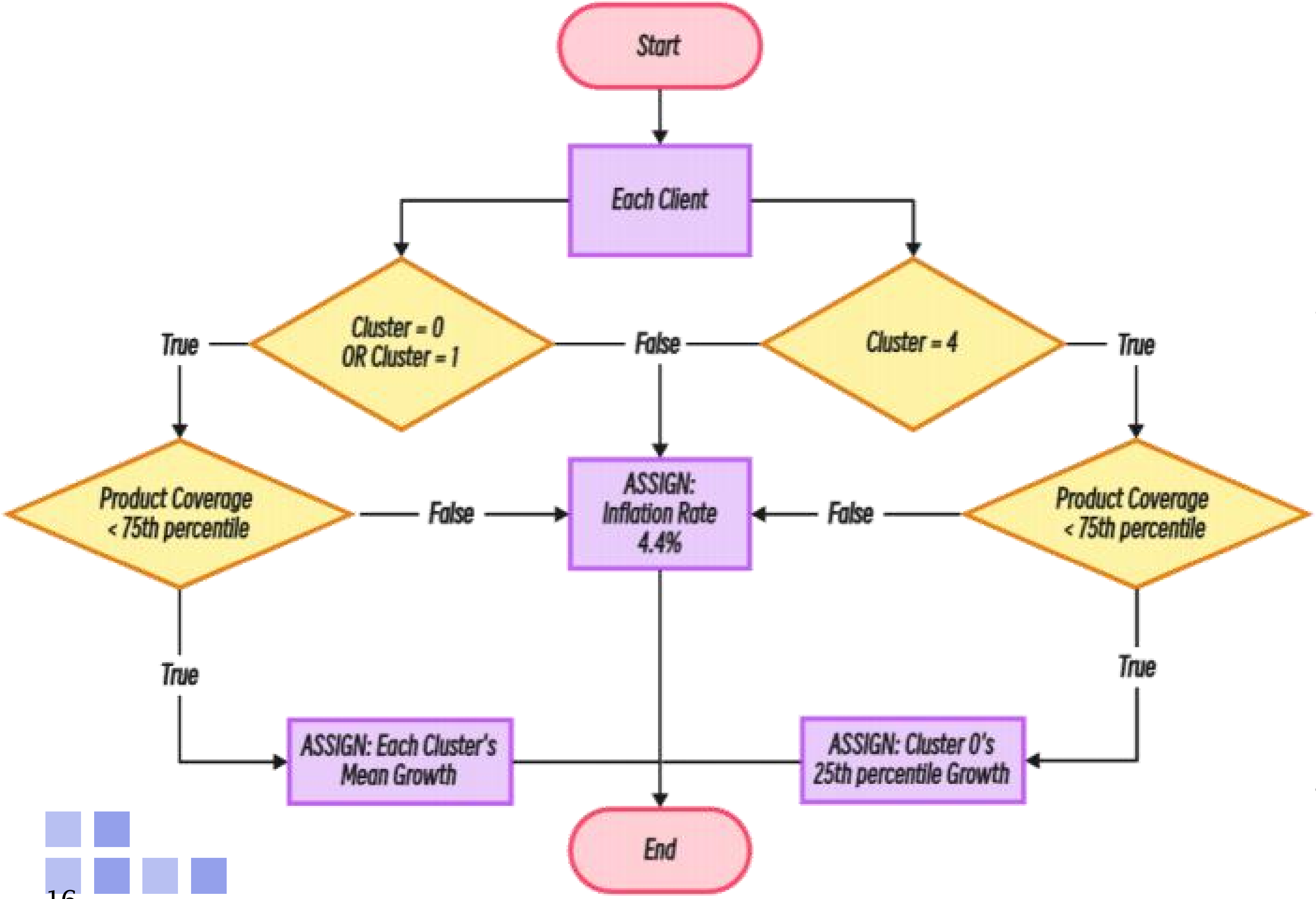
- **Cluster 0**: High Billing + High Engagement
- **Cluster 1**: High Billing + Broad Usage
- **Cluster 4**: New & Active

cluster	0	1	4	2	3	5	6
market_size_mil	1.98	3.18	1.56	1.25	1.38	0.33	1.10
avg_billing	23.31	15.92	6.45	13.98	7.67	6.18	3.88
tenure	5.00	3.18	1.79	4.93	3.03	2.80	1.57
active	1.00	1.00	0.99	1.00	0.16	0.58	0.04
growth	0.20	-0.01	-0.00	0.08	-0.03	-0.00	-0.00
recurring_ratio	0.79	0.78	0.67	0.00	0.29	0.20	0.11
cross_sell_count	3.55	3.47	0.03	2.42	2.10	0.02	0.00
risk_avg	0.84	0.71	0.51	0.69	2.77	0.96	2.68

What’ s the Problem → How We Approached It → What We Discovered → How It Drives Strategy

Growth Assignment — Based on Cluster Behavior, Not Heuristic

Growth isn’t adjusted based on heuristic rules — it adapts to cluster performance



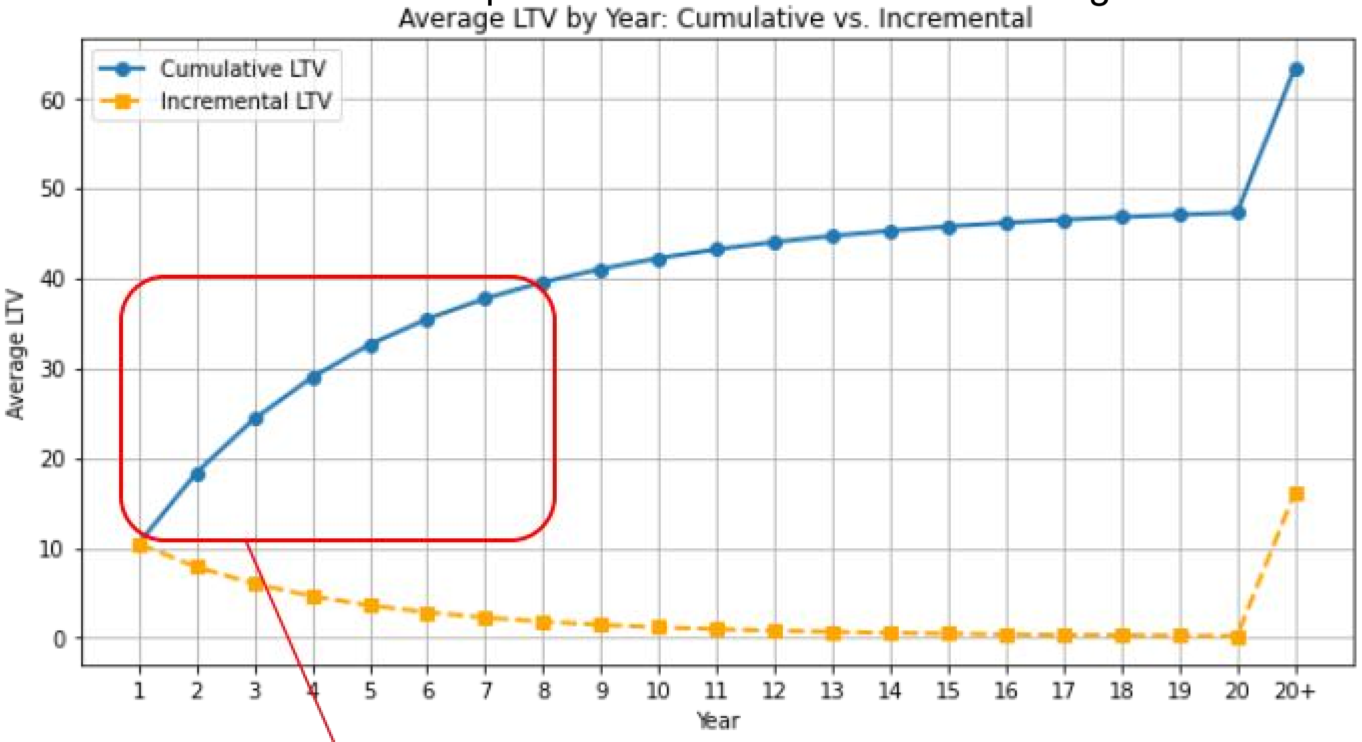
- **Adjust ↑ for clients with room to expand 、**
  Defined as below the 75th percentile in product coverage
- **If in Cluster 0 or 1**
  Assigned the cluster’s historical ***mean*** *growth*
- **If in Cluster 4 (new, active)**
  Mapped to the ***25th percentile*** *growth of Cluster 0*
  Why? We treat them as early-stage versions of Cluster 0 — with caution
- **All others default to 4.4% baseline (inflation rate)**

PART 03

WHAT WE DISCOVERED

Retention Builds the Base, Acquisition Extends the Frontier

Retention creates exponential value — but can't fuel growth forever



- Retention delivers steep value gain in early years
- By year 4, half of LTV has already materialized
- After year 8, marginal value flattens
- New clients are needed to maintain revenue momentum

💡 ***More than 60%** of LTV realized within the first 8 years*

Retention creates value. Acquisition sustains momentum.

You need both!

Climbers & Churning: Nurture or Exit

Emerging upside — and the warning signs

- ◆ Churning *(C6)*
  - Extremely low activity and poor retention
- High churn scores, often post-onboarding drop-offs
  - Strategic moves:
    - Exit flow design — offboarding, feedback capture
    - Or experiment with last-touch recovery campaigns
    - Consider reducing CS effort / automating deactivation

strat_seg	4-Churning	6	
market_size_mil	1.10	1.10	*million ppl (staff)*
avg_billing	3.88	3.88	*units*
tenure	1.57	1.57	*years*
active	0.04	0.04	*%*
growth	-0.00	-0.00	*%*
recurring_ratio	0.11	0.11	*%*
cross_sell_count	0.00	0.00	*#*
risk_avg	2.68	2.68	*x/4 score*

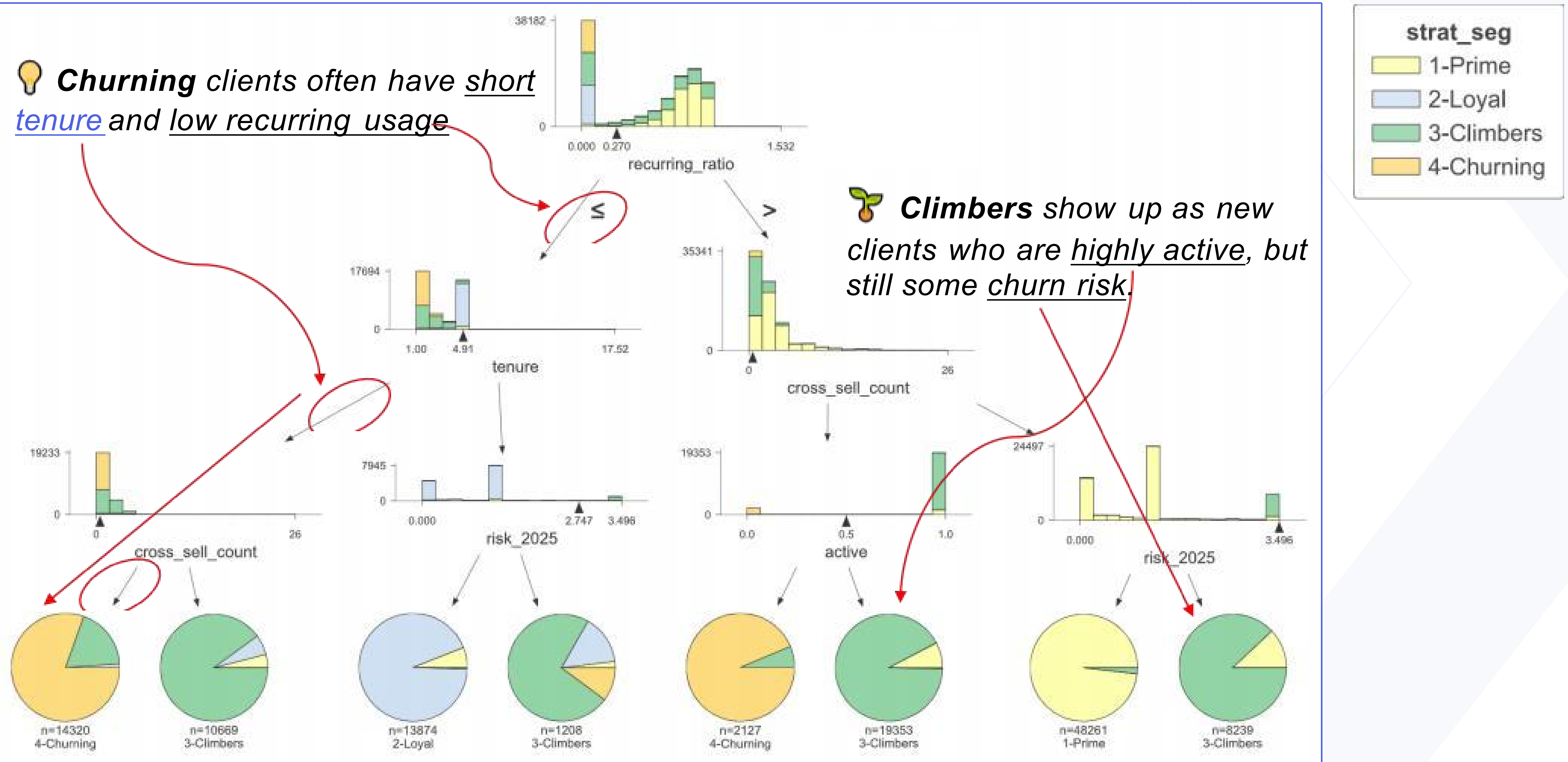
⚠ *Note*: This merge favors clarity. **For actual execution, we recommend operating at the 7-cluster level** to retain granularity.

What's the Problem → How We Approached It → What We Discovered → How It Drives Strategy

Validation with Decision Tree

💡 ***Churning** clients often have short tenure and low recurring usage*

🌱 ***Climbers** show up as new clients who are highly active, but still some churn risk.*



What We Discovered

ROI Snapshot: Where Do We Win?

LTV-to-CAC ratios tell us which segments drive profitable growth. (Based on Paylocity's internal segmentation and CAC assumptions.)

PCTY Segment	Avg LTV	CAC	LTV/CAC
1-Growth	30.48	6.00	5.08
2-Major	167.63	5.01	33.46
3-Enterprise	572.82	14.28	40.11
Total	**63.36**	**6.04**	**10.49**

- **CAC**: Estimated, not modeled — provided by client as rough benchmarks
- Enterprise ROI is exceptional (40x), but accounts for only **16% of billing**
- Growth segment has decent return (5x) — but potential may be capped
- **Takeaway**: Helps **prioritize investment** across segments based on potential payoff

PART 04

HOW IT DRIVES STRATEGY

Turning Segment Insights into Strategy

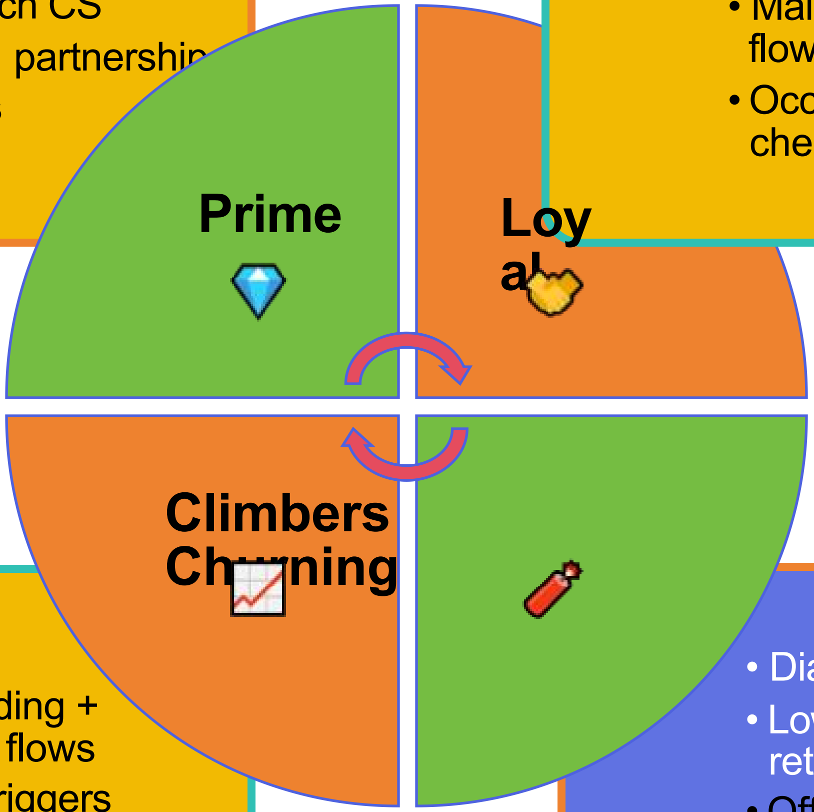
Tailored actions by … aximize value

- High-touch CS
- Strategic partnership
- Referrals

LTV = 88.41

- Maintenance flows
- Occasional check-ins

LTV = 53.61



LTV = 17.65

- Onboarding + Nurture flows
- Upsell triggers

LTV = 11.07

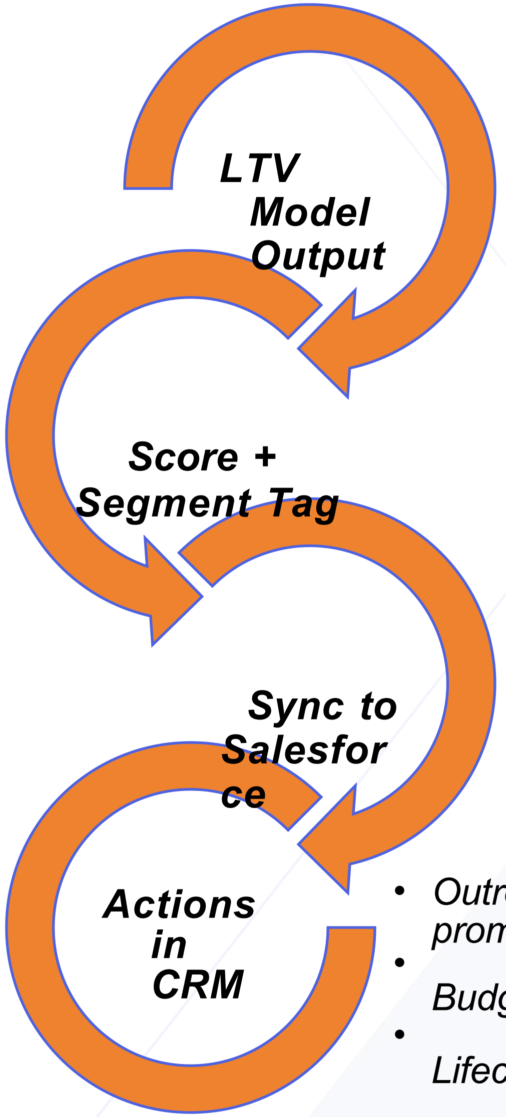
- Diagnostic flags
- Low-cost retention
- Offboarding flow

What' s the Problem

Scoring and Prioritization in Salesforce

Using LTV + Retention to guide budget and outreach in real-time

- **Why this matters**
  - Your CRM (e.g., Salesforce) is where reps make day-to-day decisions
  - Embedding LTV + Segment scores = smarter resource allocation
  - Let sales/CS see: *who to focus on, who's at risk, who to grow*
- **What we recommend**
  - Integrate LTV scores and segment labels into Salesforce client records
  - Create workflows that:
    - Flag high-potential climbers for onboarding push
    - Set CS priorities by segment (e.g., Prime = high touch)
    - Trigger retention playbooks for Churning clients



- *Outreach prompts*
- *Budget decisions*
- *Lifecycle playbooks*

What's the Problem

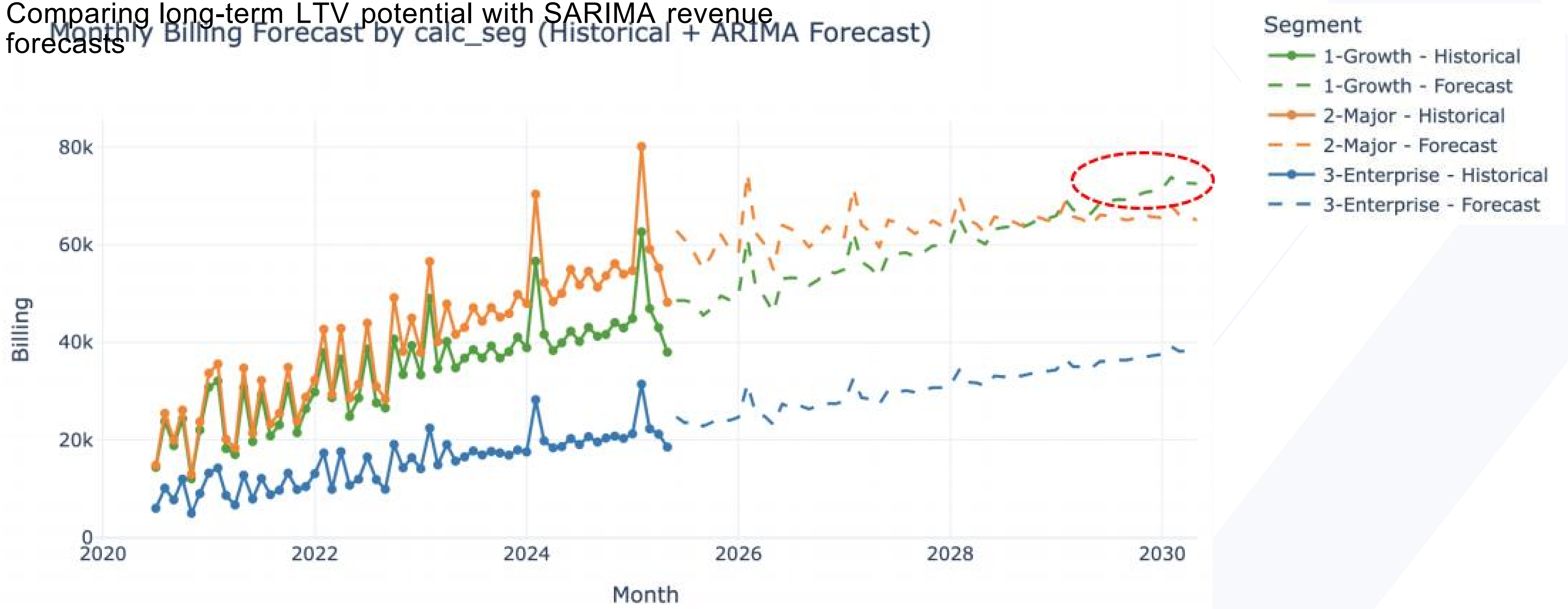
How We Approached It

What We Discovered

How It Drives Strategy

Bridging Potential and Reality

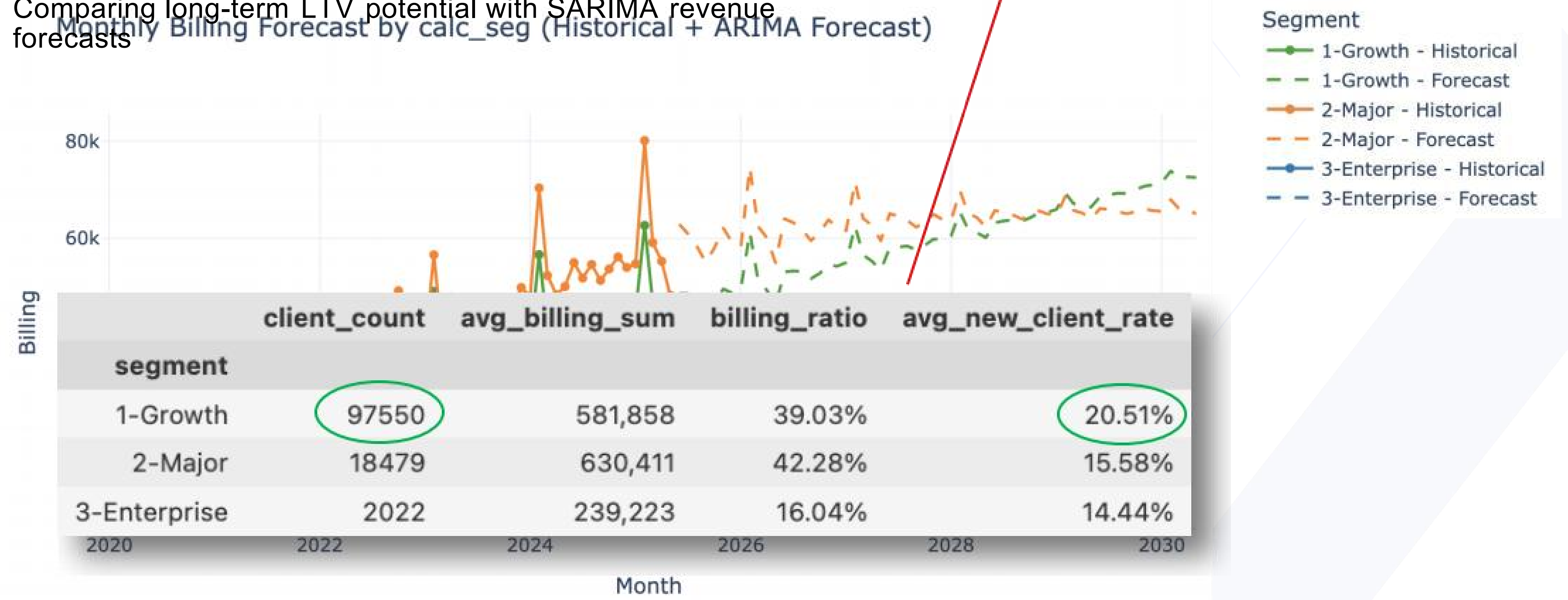
Comparing long-term LTV potential with SARIMA revenue forecasts



Bridging Potential and Reality

Comparing long-term LTV potential with SARIMA revenue forecasts

New client acquisition —
***naturally beyond** LTV's scope*



segment	client_count	avg_billing_sum	billing_ratio	avg_new_client_rate
1-Growth	97550	581,858	39.03%	20.51%
2-Major	18479	630,411	42.28%	15.58%
3-Enterprise	2022	239,223	16.04%	14.44%

PART 05

QUICK WRAP-UP

Quick Wrap-Up

Final Takeaways — Turning Insights into Action

Retention protects. Growth extends. Acquisition expands.

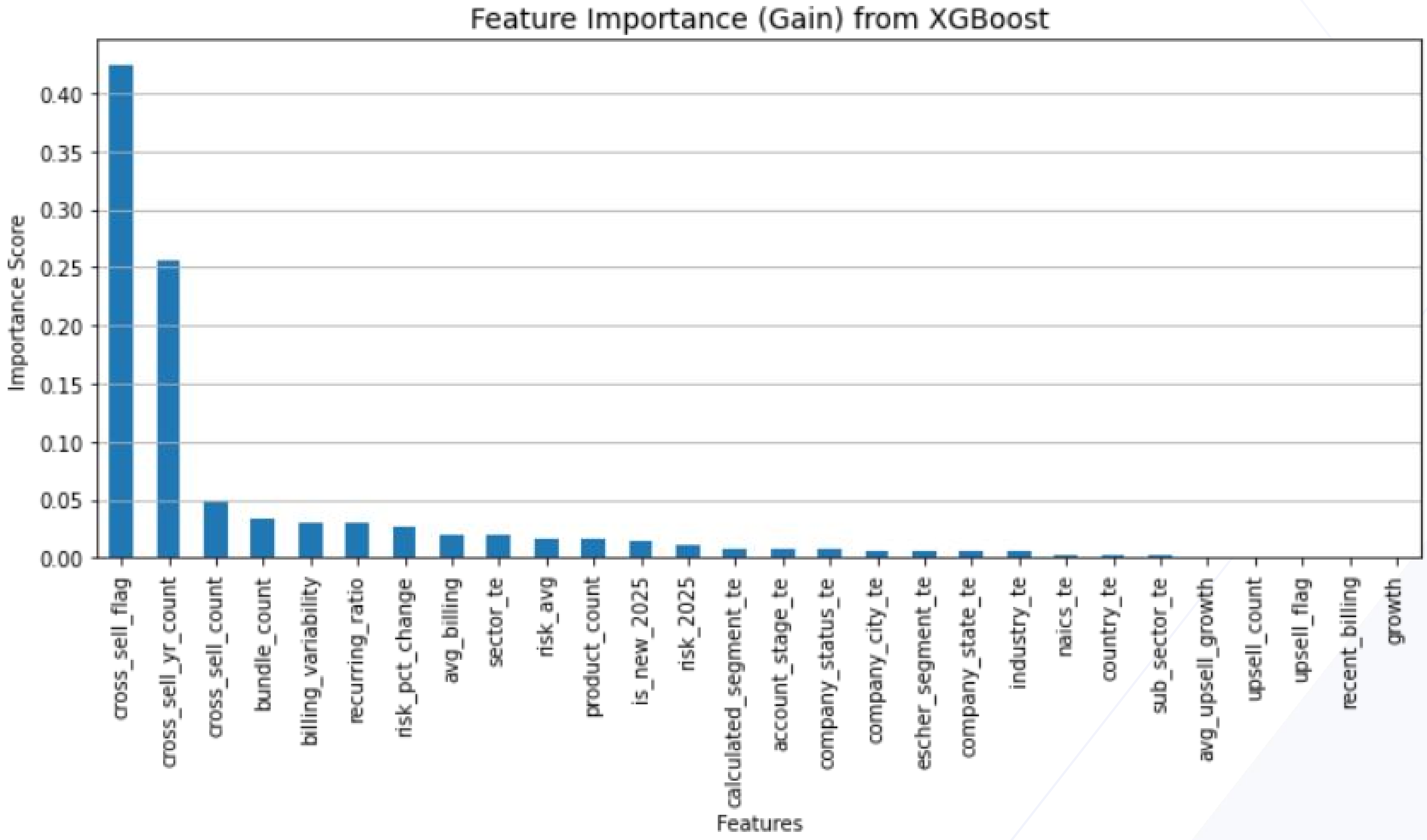
$$LTV = annual\,margin \cdot \frac{1 + \mathrm{i}}{1 + i - r \cdot (1 + g)}$$

- **Retention protects the core** — most value is realized early
  - Strengthen loyalty *(Prime & Loyal)* and reduce churn through segment-based playbooks
- **Growth matters — once retention is strong**
  - Expand future stars *(Climbers)* with upsell nudges and onboarding reinforcement
- **Acquisition expands the frontier** — it’s how scale happens
  - Target strategically with *real-time clustering*, not just demographic profiles

Appendix 1.1

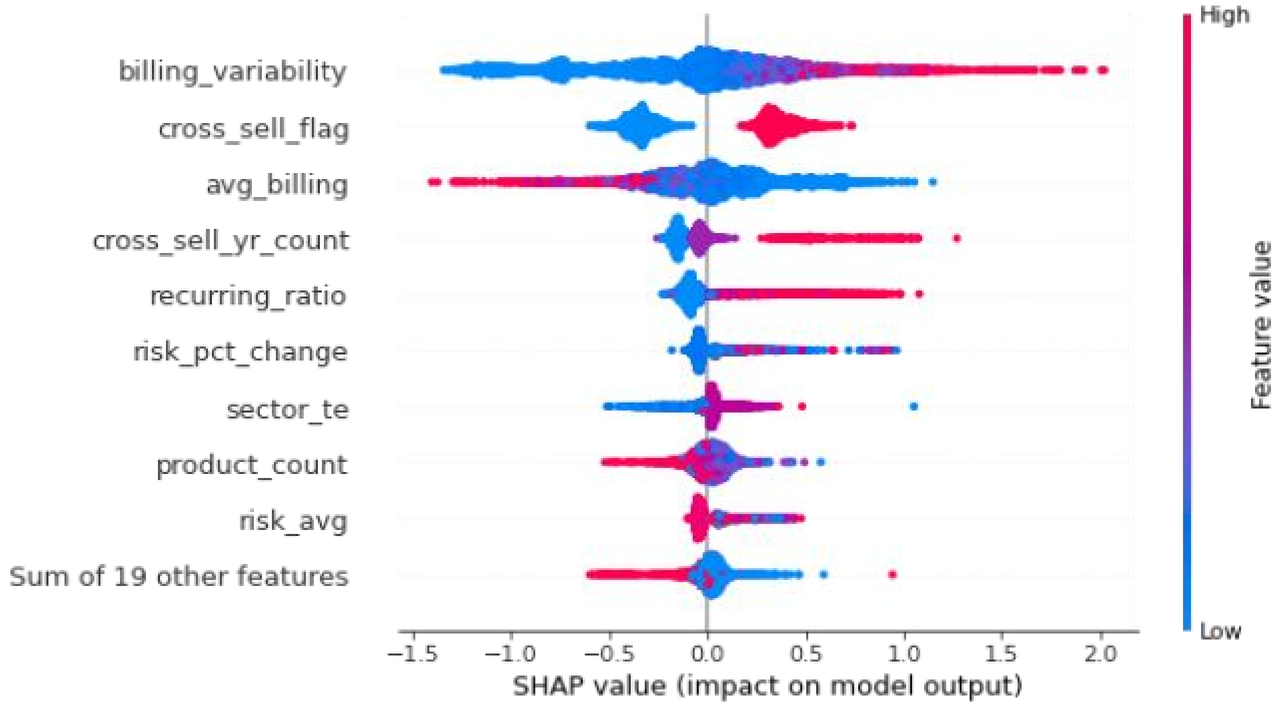
Modeling Retention — Feature Importance (XGBoost)

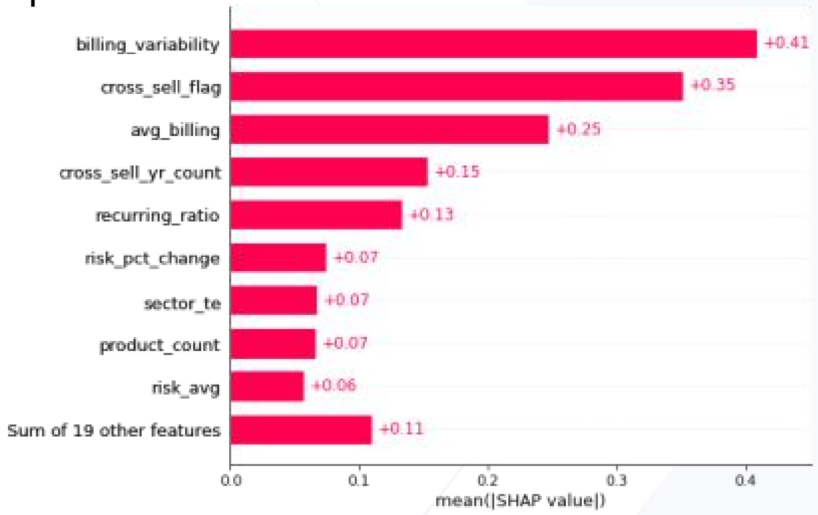
Top drivers of customer lifespan prediction, ranked by model contribution

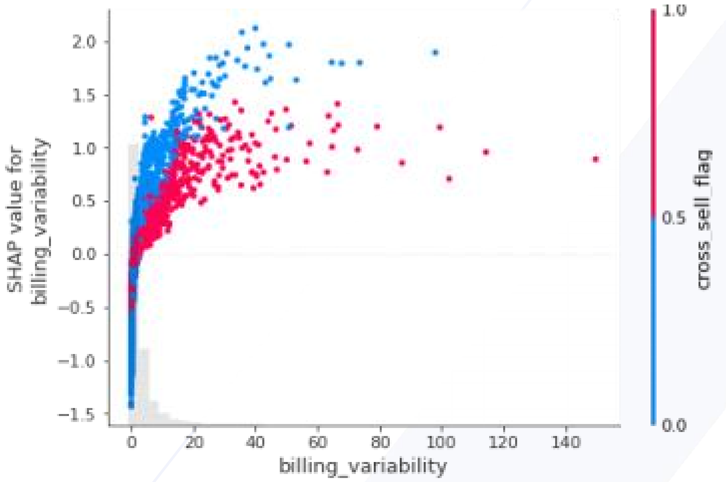


Modeling Retention — SHAP Analysis

Explaining the model: how each feature pushes retention predictions up or down







Appendix 2.1
SARIMA Model Selection Summary by Segment

Auto-selected via AIC optimization using 60 months of segment-level billing data

Segment	ARIMA Order	Seasonal Order	AIC	Notes
1-Growth	(3,1,1)	(1,0,0)[12]	1167.56	Acceptable, modest variance
2-Major	(4,1,0)	(1,0,0)[12]	1199.23	Slightly higher, still stable
3-Enterprise	(4,1,1)	(1,0,0)[12]	1082.56	Lowest AIC, most stable model

- We selected the SARIMA configuration for each segment based on the lowest AIC value found via *auto_arima*.
- All models have moderate complexity and include yearly seasonality.
- AIC values range from 1082 to 1199, indicating reasonable fit given the data scale (~60 monthly observations).

Appendix 2.2
SARIMA Residual Diagnostics

Residuals across all segments behave as white noise — no signal left unexplained.