PRESENTATION #2

MAXIMIZING LIFETIME RETURN THROUGH DATA-DRIVEN INSIGHTS

- **Team Name:** Vector Visionaries
- **Sponsor:** Vector Controls and Automation Group
- **Group Members:** Ziyu Zhang, Shihui Chen, Yiyao Zhang, Chenyan Wu, Bihan Wang, Ennan Zhao, Annie Guo

Vector Visionaries

AGENDA

- **Executive Summary**
- **Modeling Approach 1: Clustering**
  - Customer & Industry Segmentation
  - Playbook Templates
- **Modeling Approach 2: Forecasting**
  - Industry Seasonality
  - Customer Growth Potential
- **Operationalization & Scalability**
  - Industry Seasonality
  - Customer Growth Potential
  - Success Criteria

EXECUTIVE SUMMARY

- **Recap:**
  - **Data Treatment & EDA:** Cleaned and evaluated the data; surfaced early patterns in revenue, customer activity, and churn signals.
  - **Business Understanding:** Vector's revenue is highly concentrated, with a small loyal segment driving most sales, while the majority show irregular engagement and limited visibility into churn risk and timing. With clearer signals, sales time can be allocated far more efficiently.
  - **Modeling Initial Stage:** Established the roadmap for a scalable retention framework powered by clustering, churn modeling, and seasonality analysis.

- **Overarching Goal:** Build a repeatable, data-driven system that guides the Vector sales team on who to prioritize and when to engage.

MODELING APPROACH 1: CLUSTERING

Vector Visionaries

OBJECTIVE AND WORKFLOW

K-means Clustering

- Grouped customers (K=4) and industries (K=6) using behavioral metrics.

Identified key customer and industry patterns

Churn Modeling

- Built a monthly panel with features capturing recency, frequency, monetary value, and volatility
- Compared Logistic Regression vs. HistGradientBoosting

Flag-at-Risk Table

- Segmented customers into four churn-risk tiers.
- Analyzed each tier's key metrics.
- Linked each risk tier to customer and industry clusters by customer_id

Predicted account risk

Playbook Integration

- Structured a three-layer playbook
- Mapped customers to each playbook

Translated analytics into action

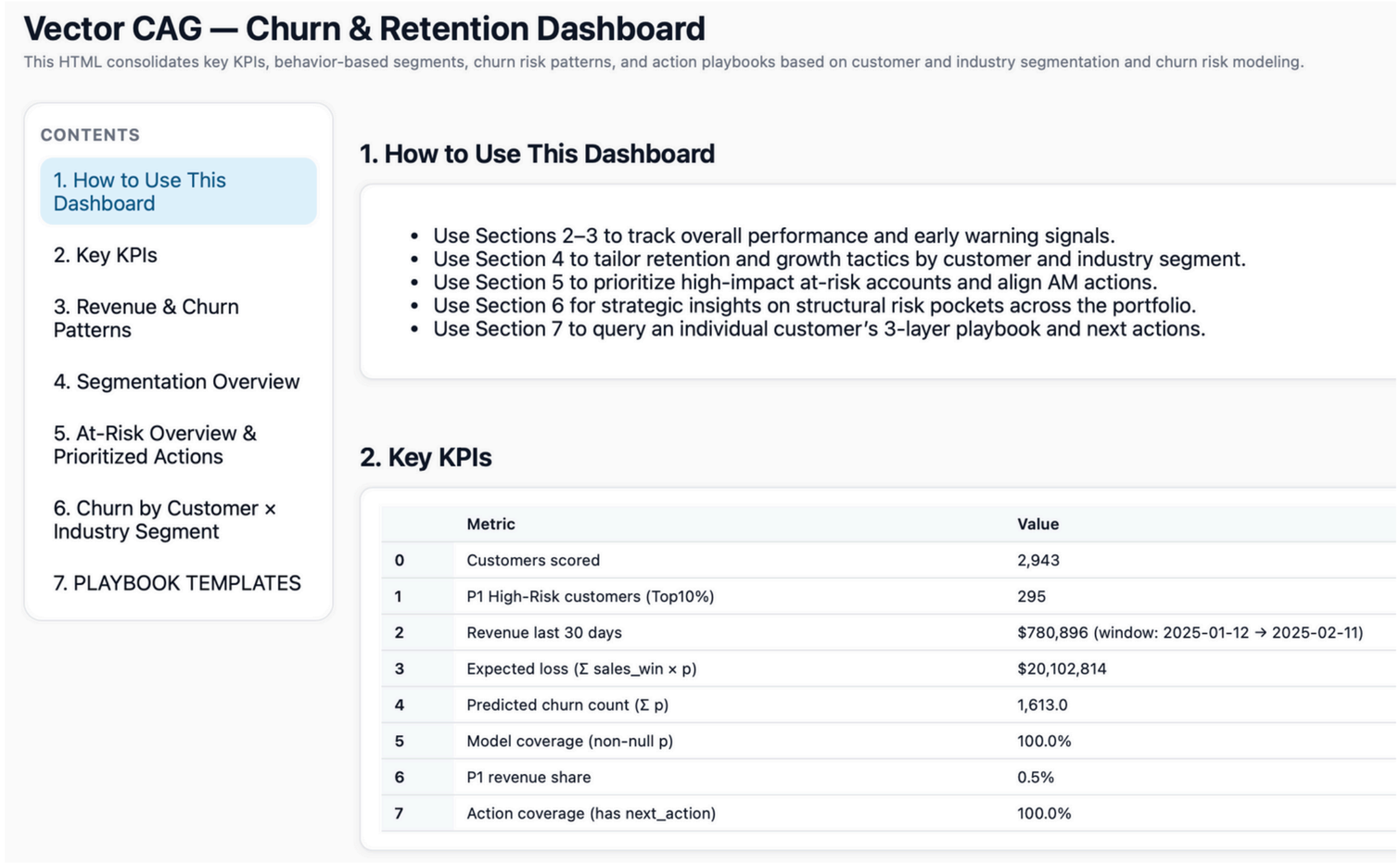
Dashboard

- Consolidated model outputs into the dashboard.
- Showed churn scores, segments, and playbook actions.

Customer Cluster Strategy & Industry Cluster Strategy + **Churn Risk Tier Strategy** = **Tailored Retention Strategy**

SAMPLE DASHBOARD

This dashboard summarizes all analytical components we will introduce next: revenue and churn patterns, segmentation, at-risk prioritization, and playbook generation. In the following sections, we walk through each component in detail: how it works, what insights it reveals, and how it connects to the overall strategy.



Vector CAG — Churn & Retention Dashboard

This HTML consolidates key KPIs, behavior-based segments, churn risk patterns, and action playbooks based on customer and industry segmentation and churn risk modeling.

CONTENTS

1. How to Use This Dashboard

- Use Sections 2–3 to track overall performance and early warning signals.
- Use Section 4 to tailor retention and growth tactics by customer and industry segment.
- Use Section 5 to prioritize high-impact at-risk accounts and align AM actions.
- Use Section 6 for strategic insights on structural risk pockets across the portfolio.
- Use Section 7 to query an individual customer's 3-layer playbook and next actions.

2. Key KPIs

	Metric	Value
0	Customers scored	2,943
1	P1 High-Risk customers (Top10%)	295
2	Revenue last 30 days	$780,896 (window: 2025-01-12 → 2025-02-11)
3	Expected loss (Σ sales_win × p)	$20,102,814
4	Predicted churn count (Σ p)	1,613.0
5	Model coverage (non-null p)	100.0%
6	P1 revenue share	0.5%
7	Action coverage (has next_action)	100.0%

- **Key KPIs**: Monitor overall activity.
- **Revenue & Churn Patterns:** Identify long-term trends and understand churn risk distribution.
- **Segmentation Overview:** Review customer and industry clusters to understand behavioral differences and tailor engagement strategies.
- **At-Risk Overview**: Focus retention effort on the highest-impact accounts.
- **Churn by Customer × Industry Segment:** Spot risk across segments to inform resource allocation and monitoring.
- **Playbook Templates - Interactive**: View the personalized 3-layer playbook for any account and track their recent order activity.

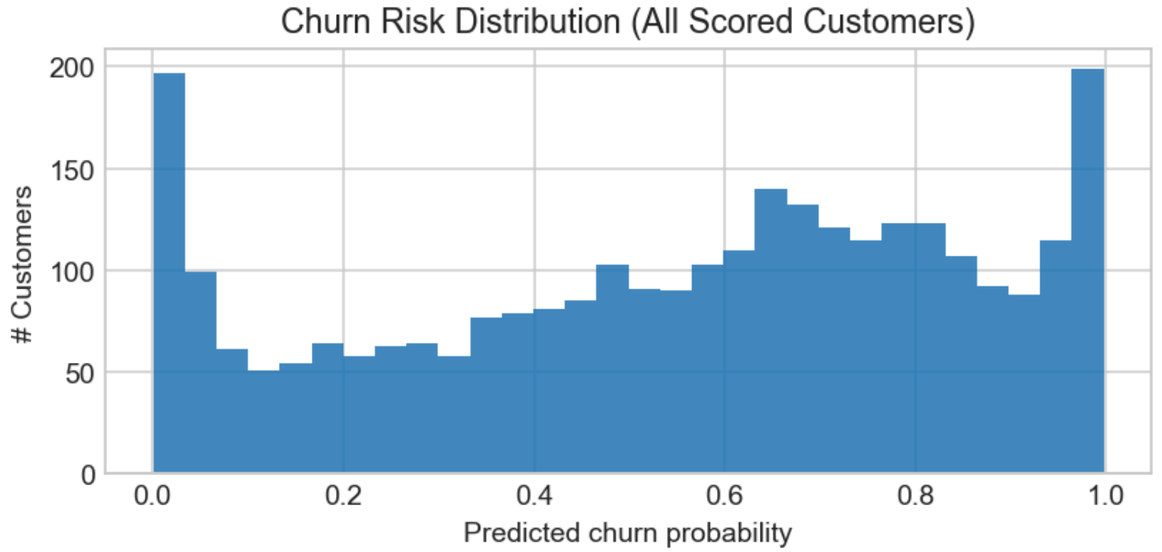
REVENUE TREND & CHURN RISK PATTERN

Revenue Trend

- Revenue shows a steady upward trend with mild seasonality.
- Indicates healthy baseline demand — focus on maintaining stability rather than recovery.

Churn Risk Distribution

- The churn risk distribution is U-shaped, with many customers at the very low- and very high-risk ends.
- This indicates polarized behavior and highlights the need for targeted segmentation instead of broad marketing campaigns.



CUSTOMER SEGMENTATION OVERVIEW & STRATEGY

The customer segmentation built upon K-Means (K=4) on many standardized features, revealing distinct relationship stages across 9,000+ customers.

	customer_cluster_id	cluster	customers	med_orders	med_revenue	med_avg_order	med_recency	med_tenure	strategy
0	0	C0 — Large, Low-Value & Inactive	3,921	3.0	$5,619	$1,485	1322	134	Win-back: reactivation + replenishment
1	1	C1 — High-Value Loyal	2,775	26.0	$66,639	$2,605	119	1713	Retain & grow: dedicated AM call + premium service
2	2	C2 — Moderate-Value Newer	2,395	1.0	$1,010	$524	410	0	Onboard & build: usage cadence + cross-sell starter bundle
3	3	C3 — High-Return / Unstable	191	4.0	$647	$149	1264	105	Stabilize: root-cause review + replacement SKU

Behavioral Insights

- 70 % of customers fall into low-to-mid value groups; the top 30 % drive major revenue.
- Retention priority shifts from reactivation (C0) → nurture (C2) → loyalty protection (C1).
  - C0 – Inactive & Low-Value
  - C1 – High-Value Loyal
  - C2 – Moderate-Value Newer
  - C3 – High-Return / Unstable
- Combining these segments with churn risk creates clear “who and how to engage” playbooks, which eventually becomes the first layer of the playbook.

INDUSTRY SEGMENTATION OVERVIEW & STRATEGY

The industry segmentation was built using K-Means (K=6) on standardized features, revealing six structural market segments with distinct revenue scales and stability patterns.

	industry_cluster_id	customers	med_orders	med_revenue	med_avg_order	med_recency	med_tenure	strategy
0	0	583	27.0	$53,268.25	$1,989.72	118	1571	Baseline stable: standard cadence + EOQ reminders + standing orders
1	1	493	42.0	$130,189.01	$3,133.98	111	1810	Project pipeline sync: pre-bid/bid reminders + bulk pricing + delivery scheduling
2	2	96	21.0	$54,551.34	$2,231.16	106	1410	Fragmented/small: micro-offers + lower MOQs + simplified SKUs + self-serve guides
3	3	428	2.0	$1,748.42	$709.62	174	0	Unstable/irregular: risk review + credit/lead-time checks + small test orders
4	4	1,014	27.0	$66,953.93	$2,361.85	120	1746	Steady high-volume: stock alerts + VMI/blanket PO + auto-replenishment
5	5	329	14.0	$28,667.35	$1,999.56	137	1276	Emerging/niche: trial bundles + sampling/case study + feedback loop

Behavioral Insights

- **Revenue concentration**: Over 75% of total industry revenue comes from mid-scale (C0), project-based (C1), and high-volume (C4) sectors.
- **Engagement stability**: Project-heavy industries (C1, C3) are more volatile, while core sectors (C0, C4) provide steady recurring orders.
- **Growth potential**: Emerging and niche clusters (C5) show rising but unpredictable demand—ideal for pilot programs or sampling outreach.
- This will eventually become the second layer of the playbook.

AT-RISK OVERVIEW

BAND SUMMARY & RISK PLAYOOK COVERAGE

priority_band	customers	avg_churn	total_expected_loss	risk_playbook
P1 Top10%	295	97.44%	$8,986,503	High-touch
P2 10-30%	588	83.73%	$6,636,152	Low-touch
P3 30-60%	883	62.97%	$3,914,560	Mid-touch
P4 60-100%	1177	23.17%	$367,818	Monitor only

KPI SUMMARY

	Metric	Value
1	Customers scored	2,943
2	Revenue last 30 days	$780,896 (2025-01-12 → 2025-02-11)
3	Expected loss (Σ sales_win × p)	19,905,033
4	Model coverage (non-null p)	100%
5	Action coverage (has next_action)	100%

Key Insights & Takeaways

- Churn probabilities were binned into four risk tiers (P1–P4).
  - **P1 (Top 10%)** represents the most critical accounts—high churn probability (~0.97) but small volume share (~10%).
  - **P4 (Bottom 40%)** are stable customers requiring only monitoring, not outreach.
- This prioritization allows sales teams to focus on 20–30% of accounts driving 70% of expected churn loss.
- Combined with playbook mapping (risk × customer × industry), each band now has a clear engagement path, which eventually becomes the third layer of the playbook.

AT-RISK OVERVIEW

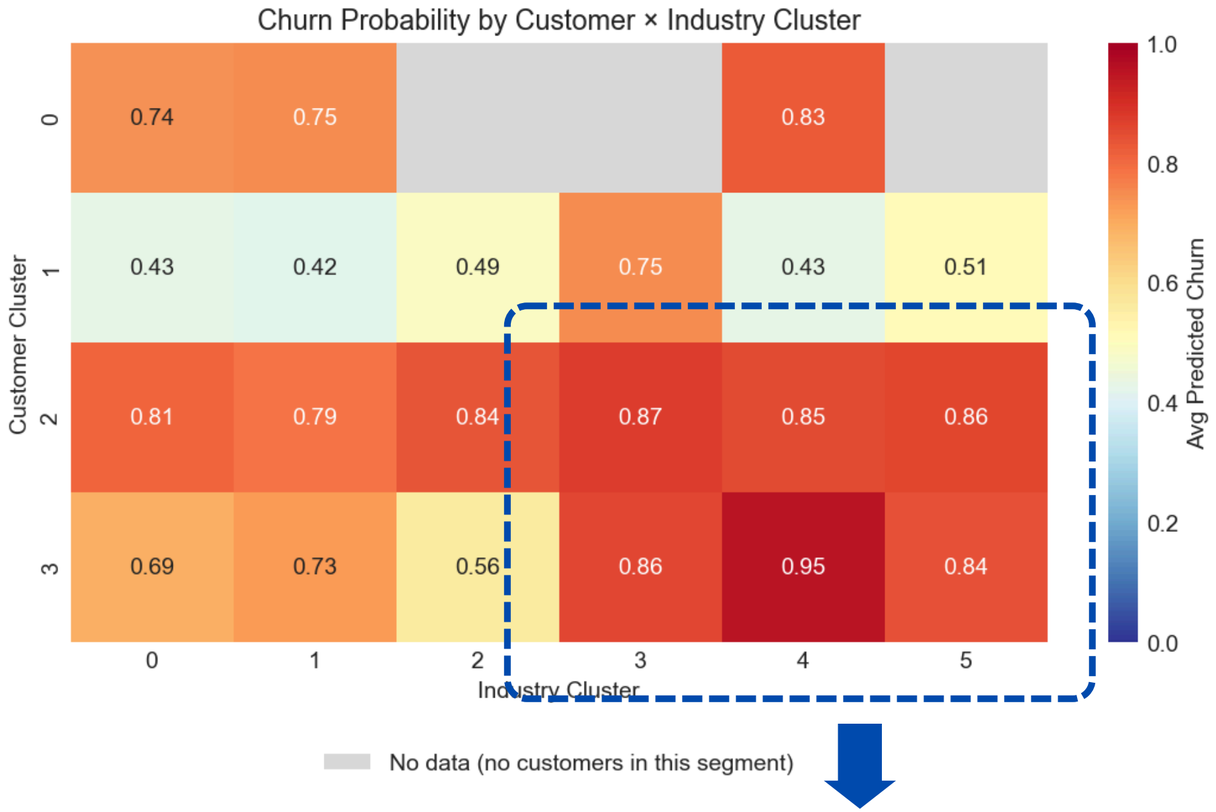
TOP P1 CUSTOMERS BY EXPECTED LOSS

	customer_id	sales_win	churn_prob	expected_loss	customer _cluster_id	industry _cluster_id	risk_playbook	cust_playbook	ind_playbook
266	23275	$40,163	94.40%	$37,903	1	5	High-touch	Retain & grow	Emerging/niche
278	44355	$26,205	93.80%	$24,586	1	4	High-touch	Retain & grow	Steady high-volume

Key Insights & Takeaways

- Identified Top 10 customers in P1 (Top 10 % risk) ranked by **expected loss = sales_win × churn_prob**
  - P1: High-touch (AM call + personal follow-up)
  - P2: Mid-touch (targeted email + personalized offer)
  - P3: Low-touch (automated nurture or periodic check-in)
  - P4: Monitor-only (passive tracking via dashboard)
- Expected loss analysis directly feeds into the Flag-at-Risk dashboard, ranking accounts by potential revenue impact.

CHURN RISK HOTSPOTS: CUSTOMER × INDUSTRY SEGMENTS



C2/C3 customers in Industry Clusters 3–5 likely face project-driven budgets, usage volatility, or replacement cycles, making them

Why It Matters

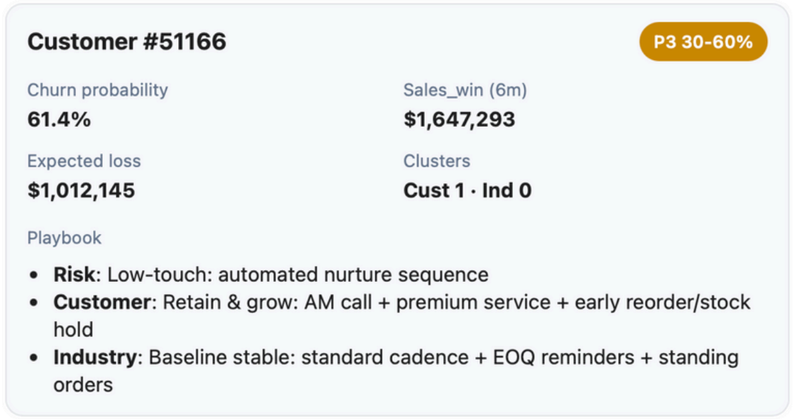
- High-risk hotspots appear where unstable customer segments overlap with volatile industries, pointing to where proactive outreach will have the greatest impact.

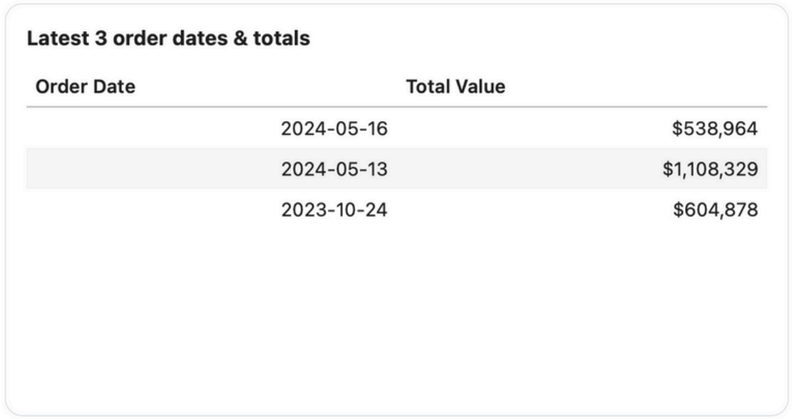
Business Insight

- Some segments are high-risk due to both customer behavior and industry factors.
- Retention strategies should focus on high-risk intersections and engage customers with timing aligned to industry cycles.

PLAYBOOK TEMPLATE

pattern	customer_id	expected_loss	industry_cluster_id
Monitor only + Retain & grow + Baseline stable	#51166	$785,767	0
Monitor only + Retain & grow + Steady high-volume	#25148	$263,199	4
Low-touch + Retain & grow + Baseline stable	#10730	$248,024	0
Low-touch + Retain & grow + Steady high-volume	#10601	$240,270	4
Monitor only + Retain & grow + Project pipeline sync	#20223	$219,341	1





How to Use Template

- Match customers to the closest pattern and apply the corresponding outreach level and messaging.
- This approach scales retention efficiently

Interactive Customer Lookup

- Allows users to look up customer ID and instantly view their churn risk, segment profile, recommended playbook, and recent order patterns.
- It enables Account Managers to quickly understand a customer's situation and take the next best action.

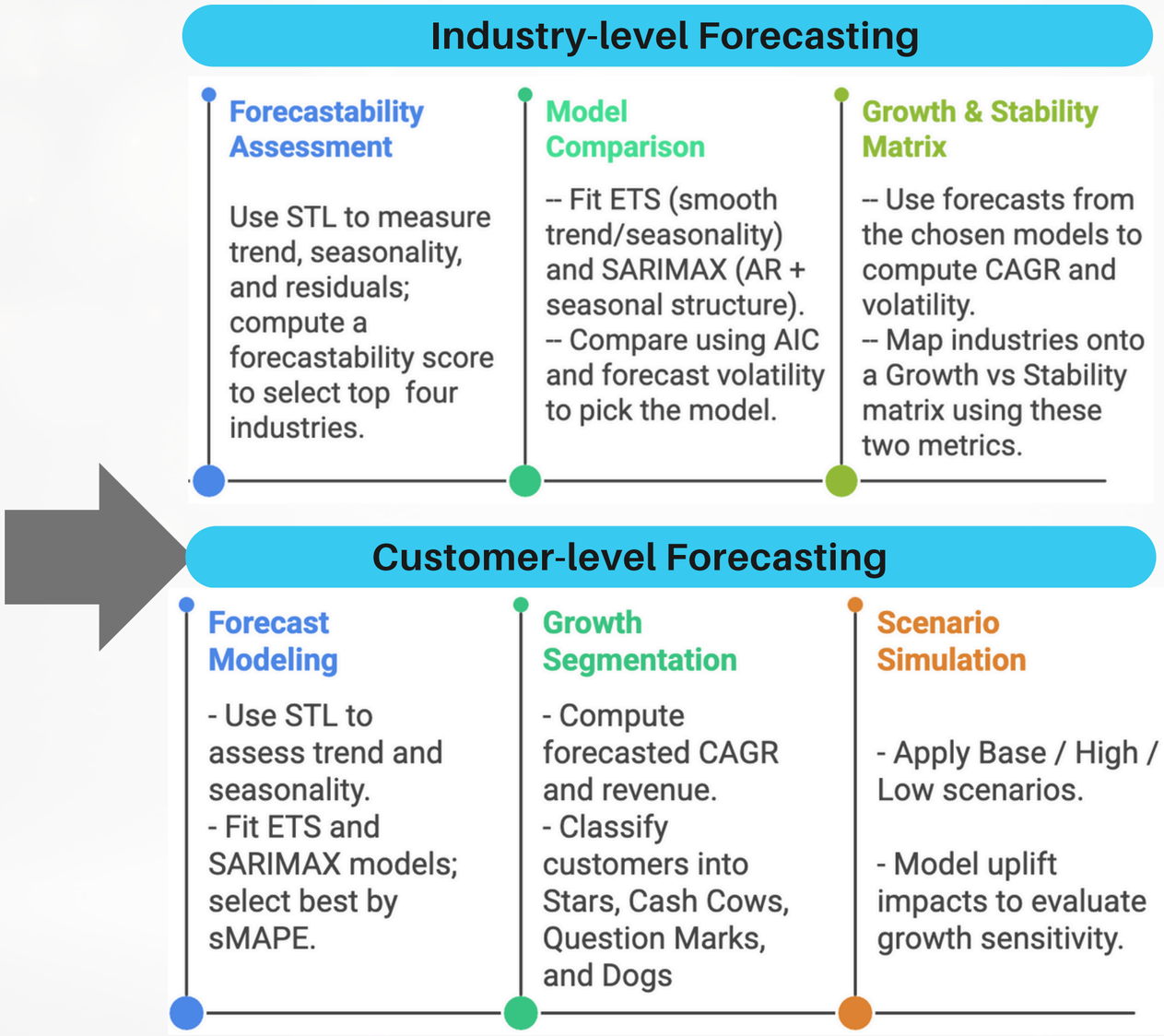
MODELING APPROACH 2: FORECASTING

Vector Visionaries

OBJECTIVE AND WORKFLOW

Data Preparation & STL Decomposition

- Consolidated monthly revenue data by industry and customer.
- Applied STL decomposition to extract trend, seasonality, and residual components.



Integration & Business Application

- Integrated industry and customer forecasts into one revenue view
- Identified growth opportunities and volatility risks
- Informed capacity, investment, and account planning.

STL DECOMPOSITION OVERVIEW

- STL decomposition separates each time series into trend, seasonality, and residuals, allowing structural diagnostics before forecasting.

Interpretation:

- Oil & Gas shows a mostly flat trend with large residual spikes → weak seasonality and higher volatility.
- Chemical has a smooth upward trend and stable seasonality → strong fit for ETS.
- Dealers / Resellers display steady trend growth with moderate seasonality → predictable structure for ETS.
- Water & Waste Water shows gradual trend increase and visible seasonal swings → supports seasonal ETS.
- Food & Beverage has smooth trend and clean residuals → high forecastability with ETS.
- Power & Energy has irregular seasonality and large residual outliers → SARIMAX needed to capture volatility.

INDUSTRY FORECASTABILITY ASSESSMENT

Industry Forecastability Scores

industry	n_points	seasonality_strength	trend_strength	resid_ratio	forecastable	score
Chemical	72	0.442	0.682	0.245	TRUE	0.601
Resellers, Reps, and Integrators	72	0.399	0.713	0.229	TRUE	0.599
Water and Waste	72	0.319	0.729	0.232	TRUE	0.573
Food & Beverage	72	0.291	0.739	0.243	TRUE	0.563
Misc	72	0.279	0.653	0.307	TRUE	0.511
Uncategorized	72	0.265	0.659	0.329	TRUE	0.504
Oil & Gas	72	0.282	0.257	0.578	FALSE	0.3
Automotive	72	-0.092	0.424	0.626	FALSE	0.208
Semiconductors	72	0.201	0.124	0.715	FALSE	0.187
Renewables	72	-0.025	0.29	0.714	FALSE	0.163
Power and Energy	72	0.027	0.209	0.813	FALSE	0.132
Life Science	72	0.105	0.091	0.783	FALSE	0.122

- **Seasonality strength** shows how much seasonal structure exists.
- **Trend strength** measures the long-term direction.
- **Residual ratio** captures noise in the data (lower = cleaner).
- **Forecastability score** blends these three signals to measure predictability.
- **Forecastable industries** are those with trend or seasonality > 0.3 and residual ratio < 0.6

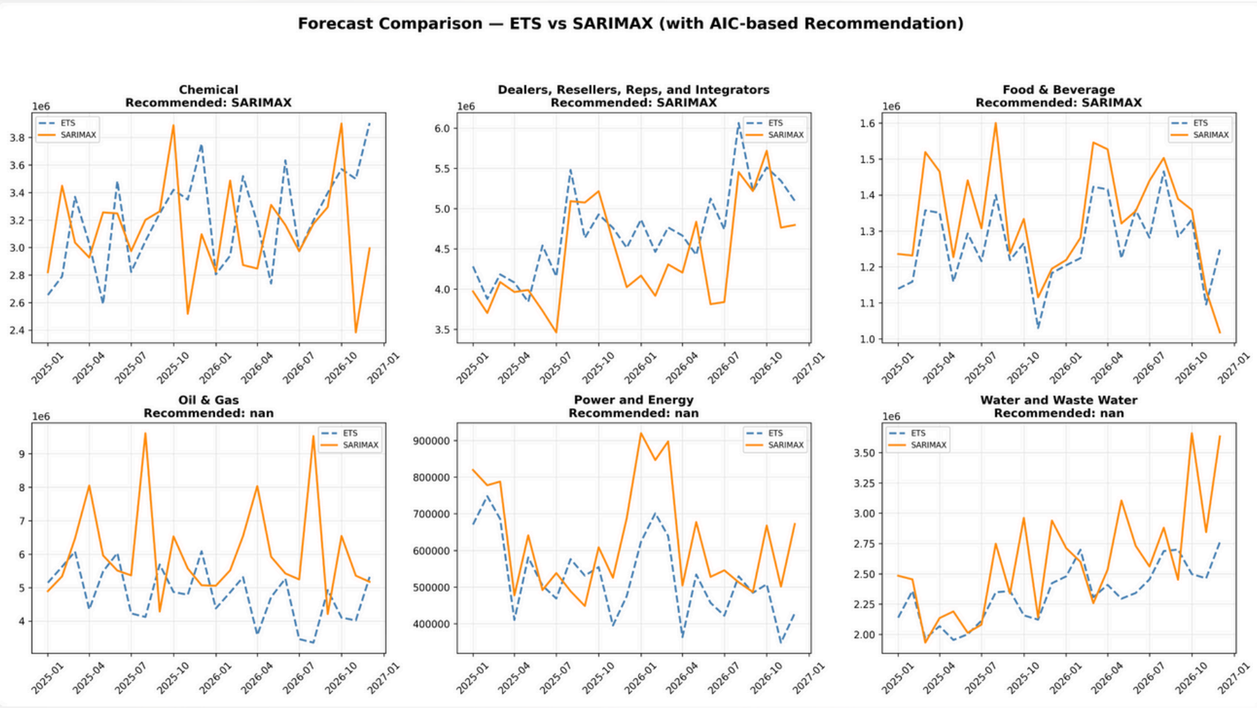
Forecastable industries

- Chemical
- Dealers / Resellers / Reps / Integrators
- Water & Waste Water
- Food & Beverage
- Oil & Gas
- Power & Energy

Why these Six?

- The first four have the strongest forecastability.
- Oil & Gas and Power & Energy are high-priority business sectors..

MODEL RECOMMENDATION

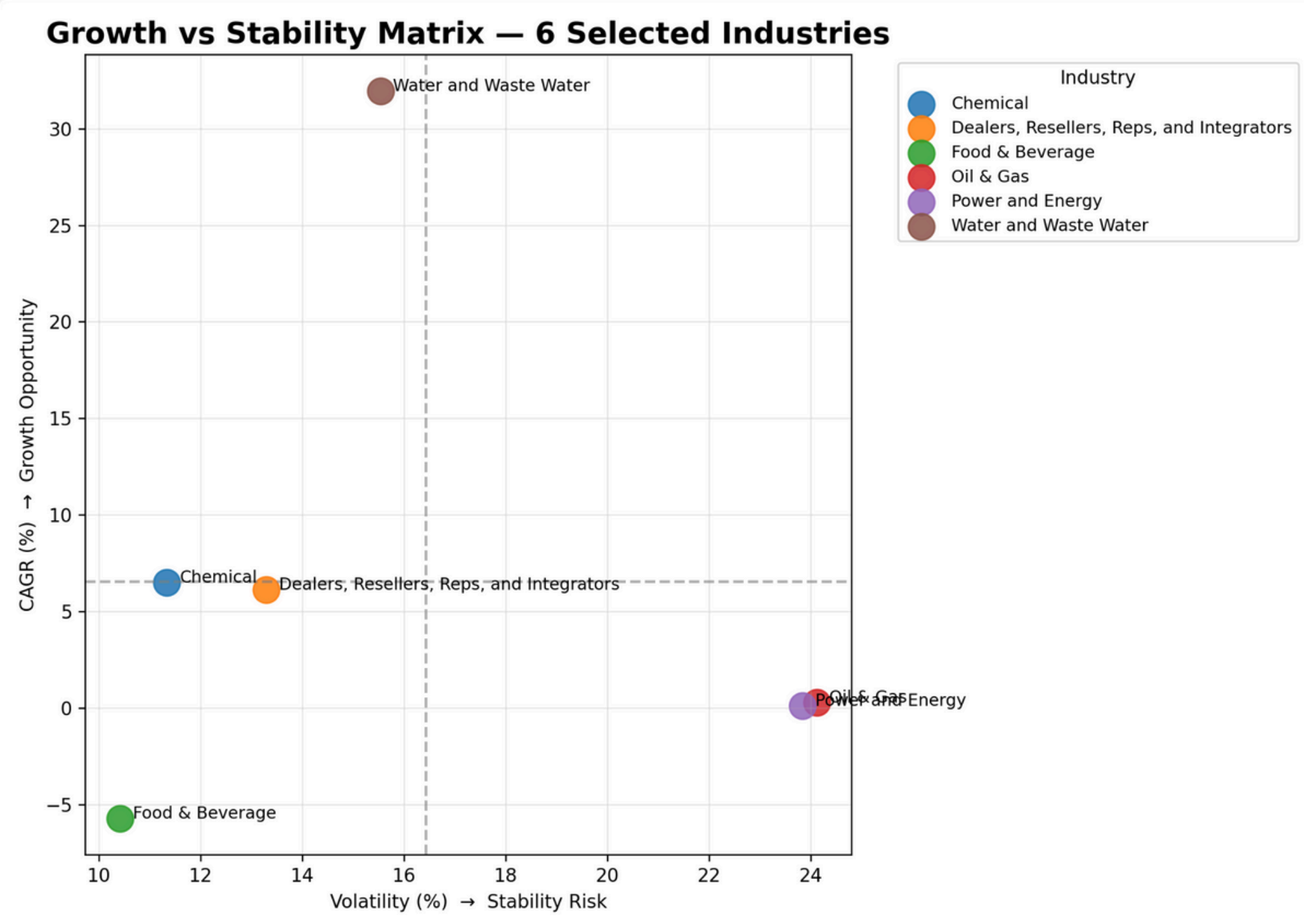


industry	model	mean_forecast	std_forecast	max_forecast	min_forecast	best_aic	volatility_ratio
Chemical	ETS	3204458	365056.5	3905038	2587851	1924.842715	0.113921
Chemical	SARIMAX	3120744	354993.3	3901263	2382754	-2.045585	0.113753
Dealers, Resellers,	ETS	4731781	549337.4	6063341	3841329	1948.744958	0.116095
Dealers, Resellers,	SARIMAX	4414467	633067.9	5718095	3461779	19.180031	0.143408
Food & Beverage	ETS	1263759	109817.3	1465758	1030102	1771.336382	0.086897
Food & Beverage	SARIMAX	1333445	151035	1600318	1017895	-2.508048	0.113267
Oil & Gas	ETS	4824519	815343.5	6087430	3354069	2106.58030	0.169
Oil & Gas	SARIMAX	6050272	1437837	9615724	4208186	53.921899	0.237648
Power and Energy	ETS	526752.5	109646.9	747881.5	348586.1	1835.730726	0.208156
Power and Energy	SARIMAX	627166.3	146254.5	919396.2	448529	81.386748	0.233199
Water and Waste	ETS	2337076	237540	2762650	1953306	1856.736306	0.10164
Water and Waste	SARIMAX	2599035	456483	3660387	1931092	21.411702	0.175636

Recommended models:

- **Chemical** → SARIMAX – lower AIC, better fit to irregular fluctuations.
- **Dealers / Resellers / Reps / Integrators** → ETS – smoother trend, less variance.
- **Food & Beverage** → SARIMAX – captures short-term spikes and seasonal shifts.
- **Water & Waste Water** → ETS – stable seasonality and low volatility.
- **Oil & Gas** → ETS – handles high volatility and large residual swings.
- **Power & Energy** → ETS – better suited for irregular patterns and noise.

GROWTH VS STABILITY MATRIX



- **Vertical axis (CAGR):** long-term growth potential.
- **Horizontal axis (Volatility):** forecast stability (higher = more uncertainty).
- The matrix uses these two metrics to position industries based on both growth prospects and stability.

Key Insights:

Water & Waste Water
- Highest long-term growth potential
- Moderate volatility → needs proactive planning

Chemical
- Stable, predictable growth
- Acts as a low-risk base in the portfolio

Dealers / Resellers / Integrators
- Solid growth with lower stability
- Cyclical demand → requires flexible planning

Food & Beverage
- Flat to declining growth
- Mature segment focused on margin efficiency

Oil & Gas
- Very low growth with high volatility
- Requires cautious forecasting and risk buffering

Power & Energy
- Minimal growth and elevated volatility
- Functions as a critical but unstable strategic sector

INDUSTRY-LEVEL BUSINESS IMPLICATIONS

- **Portfolio Strategy:**

Prioritize investment in Chemical and Water & Waste Water, as both show dependable growth with stable, low-variance forecasts. Allocate cautiously to Oil & Gas and Power & Energy due to higher volatility and uncertain outlooks.

- **Risk Management:**

Use volatility and forecast-stability metrics to guide exposure limits, especially for Oil & Gas and Power & Energy. Scenario-based planning and more frequent model refreshes are recommended for these sectors.

- **Operational Planning:**

Adjust inventory, staffing, and production schedules by industry profile—e.g., Dealers/Resellers require flexible operations to handle recurring short-term variability, while Food & Beverage can leverage predictable seasonal peaks for promotion and cashflow planning.

- **Strategic Foresight:**

Use trend and forecast outputs to inform expansion timing, budgeting decisions, and channel or partnership prioritization. Future enhancements may incorporate price elasticity, macro indicators, CRM segmentation, and supply-chain constraints.

CUSTOMER SEASONALITY INSIGHTS & MODELING IMPLICATIONS

Seasonality Distribution

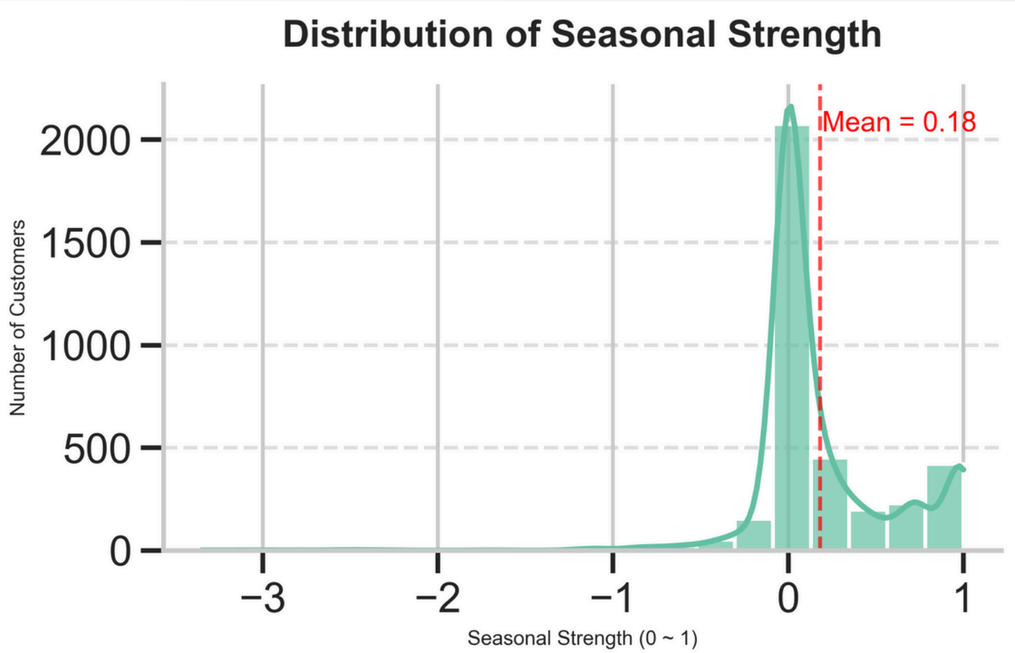
- Most customers show weak or irregular seasonality
- A small subset displays strong, recurring patterns

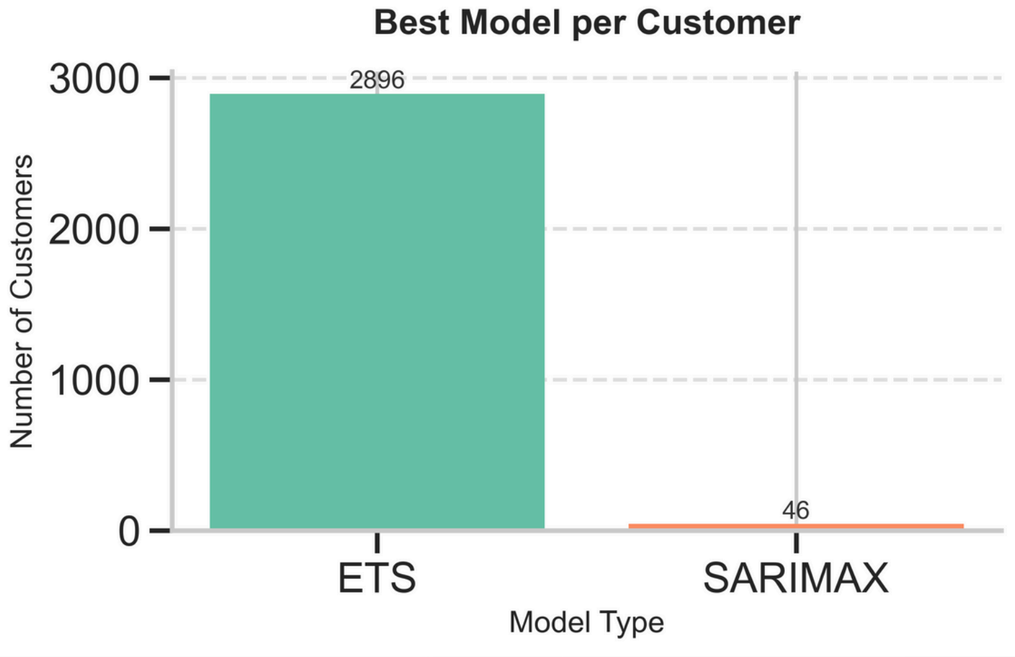
Model Distribution

- Tailored models were applied based on seasonality strength:
  - ETS for stable or trend-driven customers.
  - SARIMAX for high-seasonality customers.
- This segmentation ensures forecasts are both realistic and interpretable

Business Takeaway

- Recognizing when demand is cyclic vs. irregular helps sales and inventory teams plan more efficiently.
- Seasonal customers may need timely engagement or promotional alignment around their recurring demand cycles.





MODEL SELECTION VALIDATION — SARIMAX VS ETS

Strong Seasonality → SARIMAX performs better

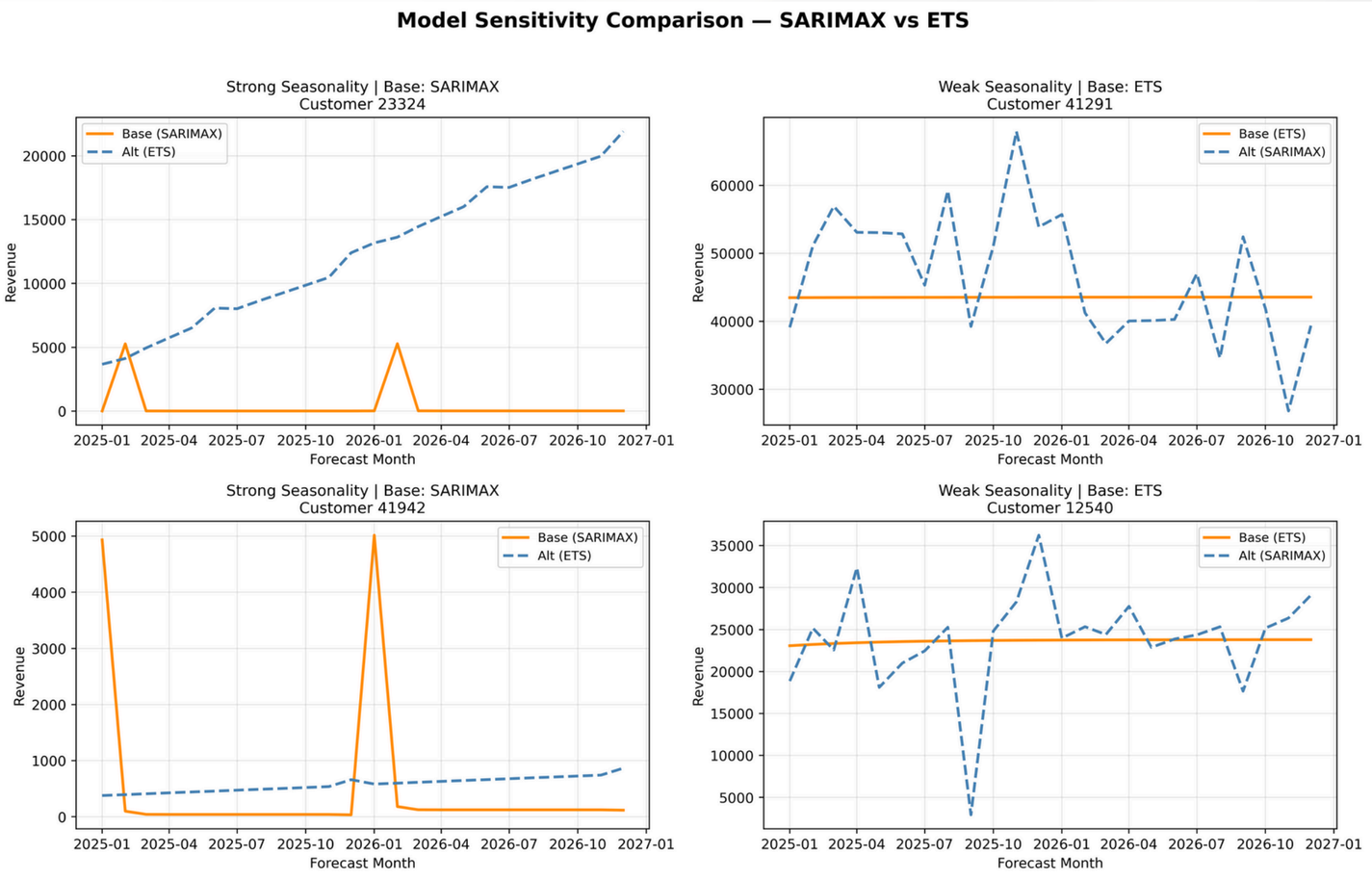
- SARIMAX captures recurring seasonal peaks (e.g., Customer 23324, 41942).
- ETS underestimates amplitude or produces linear growth when strong cyclicality exists.
- SARIMAX's autoregressive seasonal terms adapt to periodic spikes.

Weak Seasonality → ETS shows greater stability

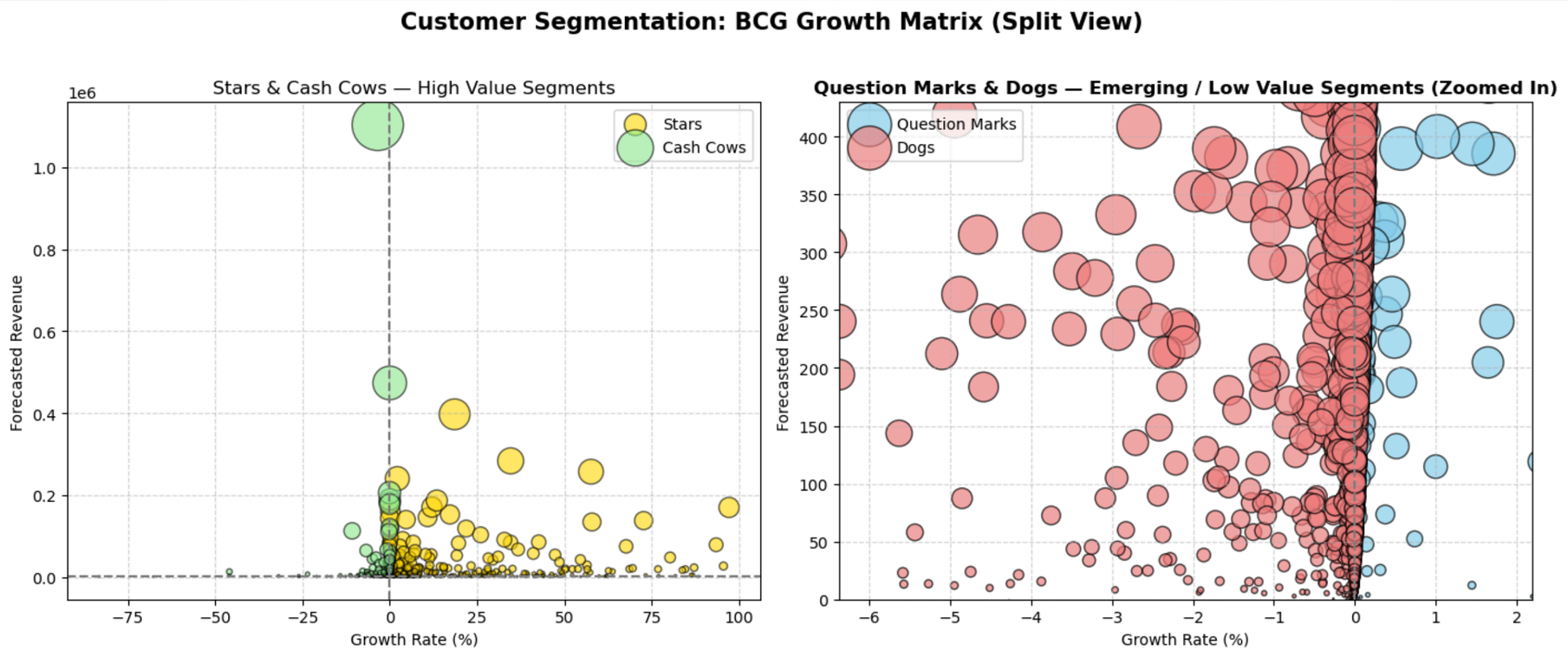
- For flatter, low-variance patterns (e.g., Customer 41291, 12540),
- ETS generates smoother and more realistic long-term projections.
- SARIMAX tends to overreact to noise, producing volatile swings.

Model differentiation ensures tailored accuracy

- ETS for low-variance, high-stability series.
- SARIMAX for high seasonal or cyclic customers.



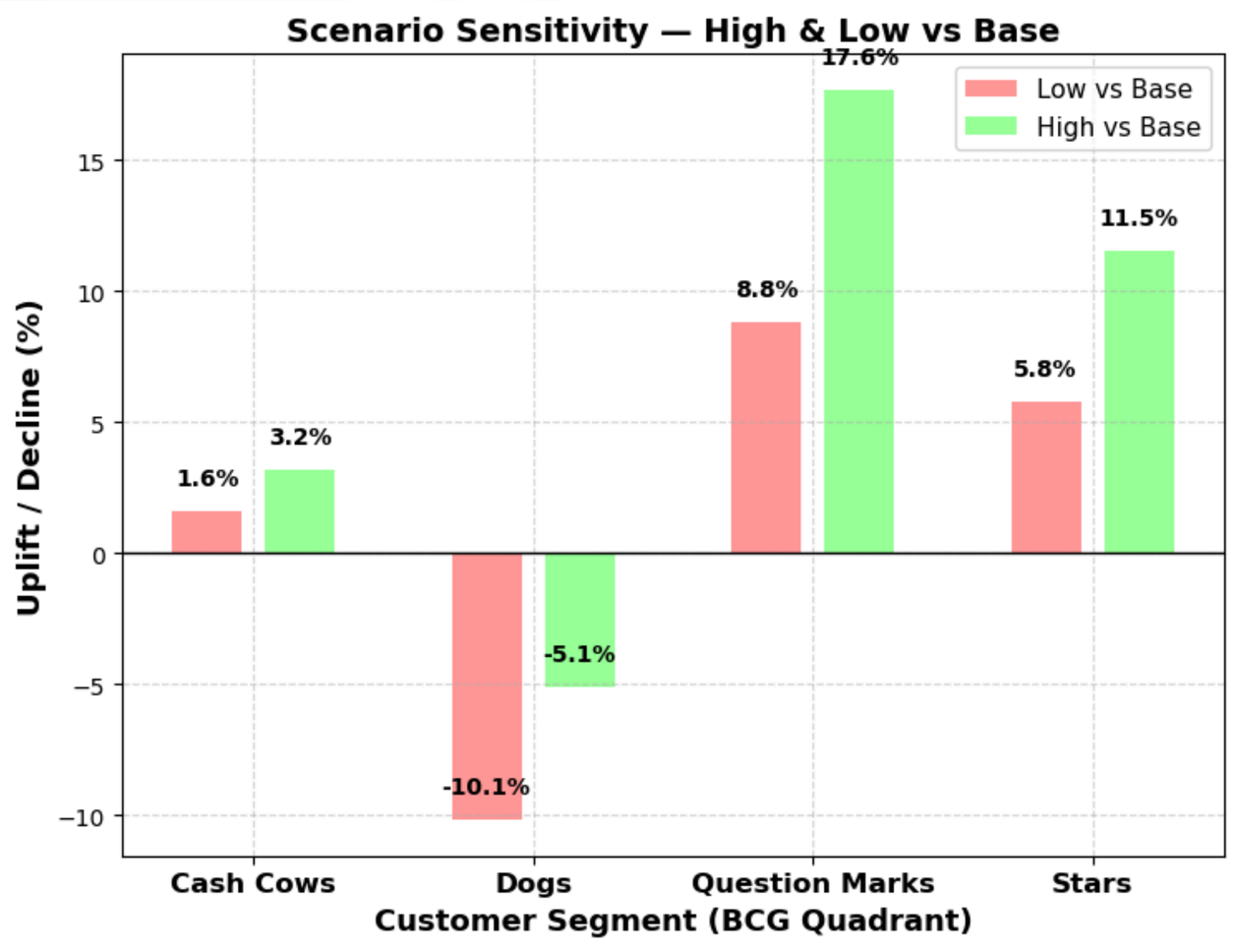
CUSTOMER SEGMENTATION: BCG GROWTH MATRIX



Key Observations

- Most Stars cluster at high growth but relatively low-to-mid revenue
- Cash Cows show high revenue but nearly zero growth, tightly centered around the horizontal axis
- A few large bubble Cash Cows sit near zero growth but extremely high revenue
- Question Marks display low-to-moderate revenue but mixed growth, ranging from slight positive to negative
- Dogs cluster heavily around negative growth and low revenue

SCENARIO FORECAST SIMULATION (BASE / HIGH / LOW)



- **Stars** dominate overall revenue, showing +11–18% potential uplift under high-execution scenarios
- **Question Marks** show the largest relative sensitivity, up to +17% uplift if MDF or PODs execution succeeds
- **Cash Cows** grow steadily (+1–3%), sustaining profitability
- **Dogs** continue to decline unless corrective investment occurs

Segment	Forecast Strategy
Stars	Maintain aggressive growth; align with sales incentives.
Question Marks	Fund execution pilots; track ROI of uplift.
Cash Cows	Protect margins; attach services.
Dogs	Rationalize or reinvest selectively.

GROWTH POTENTIAL CONCLUSION & BUSINESS IMPLICATIONS

Key Findings

- Customer base is highly **heterogeneous** — majority show weak or irregular purchasing patterns, while a smaller subset exhibits predictable, cyclical demand.
- Our **data-driven model assignment** (ETS vs SARIMAX) produced consistent, interpretable forecasts across segments.
- Forecast results reveal clear growth tiers aligned with the **BCG framework**, highlighting where value and potential are concentrated.
- Scenario simulation confirmed **Stars and Question Marks** as the primary growth levers under high-execution conditions.

Strategic Implications

1. **Stars → Sustain Momentum**
   Maintain growth with continued investment and sales incentives; protect strategic accounts.
2. **Question Marks → Accelerate Conversion**
   Allocate marketing and MDF funding; track uplift ROI to validate scalability.
3. **Cash Cows → Optimize Profitability**
   Focus on margin efficiency and attach-rate growth via service offerings.
4. **Dogs → Rationalize Resources**
   Phase out or reinvest selectively based on recovery potential.

OPERATIONALIZATION & SCALABILITY

Vector Visionaries

FINAL DELIVERABLES

WHAT VECTOR VISIONARIES WILL DELIVER

1

Interactive Dashboard

- Churn-Risk Engine + Playbook
  - Risk tiers, early-warning flags, & three-layer retention playbook
- Seasonality & Forecasting
  - Industry timing signals, ETS/SARIMAX forecasts, growth potential

2

Documentation & Code

- Consolidated codes
- Business user guide (refresh process & instructions)
- Research & Action Plan

SUCCESS CRITERIA

Category	Measure	Benchmark	
Churn Model Performance	ROC-AUC	≥ 0.80	✔ 0.84
Forecasting Reproducibility	Automated ETS / SARIMAX scripts	100 % successful reruns	✔
Dashboard Functionality	Interactive visualization (segmentation, churn, seasonality)	100 % functional prototype deployed	In progress and on track to deliver as final deliverables
Code Quality	Analytical toolkit documented for reuse	Executable by sponsor on local setup	✔

OPERATIONALIZING THE FRAMEWORK

HOW VECTOR CAN USE THIS GOING FORWARD

Immediate

- Use customer & industry clusters and apply the retention playbook to steer outreach.
- Leverage seasonality signals (ETS/SARIMAX) to time engagement around industry demand cycles.
- Use the churn-risk dashboard to identify high-impact accounts needing timely follow-up.

Scalable Long-Term

- Integrate clustering labels, seasonality patterns, and churn-risk scores into CRM/Salesforce for daily visibility.
- Automate monthly refresh of segmentation, churn signals, and industry forecasts.
- Develop early-warning triggers using changes in cluster behavior or seasonal deviations.
    - Sudden spike in churn probabiliy and movement between churn tiers are model-based triggers to watch out.

THANK YOU

Vector Visionaries