High-Tech Association (HTA)

December 11, 2025

Amazon Cart Optimization

by Gina Liao, Roxanne Li, Jacky Lu, James Shao, and Tim Su

Executive Summary

This initiative addresses high Amazon Cart abandonment rates by leveraging comprehensive abandonment data to drive a more precise understanding of customer segmentation.

We are developing a professional, Data Science-driven Dashboard product designed to equip regional managers with the analytical tools necessary to better categorize customer cohorts and develop targeted retention strategies.

By successfully executing the product roadmap, this vital management tool is projected to achieve full operational availability and global deployment within the next 12 months.

Agenda

- Business Problem, Stakeholders and Users
  - Tim Su
- CRISP-DM Data Strategy, Quality, and Ethics
  - Gina Liao
- BUS Matrix + Data Science and Analytics design
  - James Shao
- Prototype, demo Business value summary and expected outcomes
  - Jacky Lu
- Implementation roadmap, dependencies, scalability
  - Roxanne Li

Business Use Case

Use Case: Predictive Cart Abandonment & Recovery

- High-intent customers abandon carts at checkout → direct revenue leakage
- Existing systems focus on post-abandonment recovery (emails, coupons)
- No real-time prediction of who will abandon + why + when to intervene
- Opportunity: Shift from reactive to predictive, behavior-driven recovery

Problem Statement

Core Business Problem

- Amazon loses revenue due to harmful cart abandonment at final funnel stage
- Not all abandonment = lost sale (browsing, comparison, save-for-later)
- Key challenge: Differentiate harmful vs. benign abandonment in real time

Gaps in Current System

Current systems do NOT:

- Identify which customers are most likely to abandon
- Detect early friction signals (price, shipping, trust, sentiment)
- Prioritize high-risk products or categories
- Personalize recovery based on trust or behavior
- Provide a unified executive + regional dashboard
- Automate cross-region interventions

Result: Late reaction, lost revenue, inefficient recovery

Stakeholders & Importance

PRIMARY STAKEHOLDERS

- Data Science & Engineering Teams
  - Build, maintain, and monitor segmentation & prediction models
- Product & UX Teams
  - Convert model outputs into usable dashboard insights
- Regional Managers
  - Use dashboard to act on high-risk segments & friction signals
- Executives (Retail, Ops, Consumer Engagement)
  - Own scaling, funding, and global rollout

SECONDARY STAKEHOLDERS

- Customer Experience Teams
  - Validate friction signals (shipping, packaging, pricing)
- Marketplace Sellers / Amazon Basics
  - Adjust pricing, quality, and bundling based on insights

- Enables consistent adoption and measurable business impact
- Strengthens alignment across data, operations, and marketing teams
- Ensures scalable and responsible deployment of AI-driven interventions

Data Understanding and Preparation Stages

DATA SOURCE REQUIREMENTS

- **Behavioral Session Logs**
  - Track browsing depth, revisit frequency, and comparison behaviors that signal hesitation
- **Cart Activity Events**
  - Time added, time removed, abandonment flags, price at moment of decision
- **Product Attributes**
  - Category, pricing volatility, Amazon Basics status, and item pairings for bundle opportunities
- **Shipping & Delivery Estimates**
  - Shipping delays or high fees that contribute to abandonment risk
- **Review Sentiment & Friction Signals (NLP)**
  - Packaging complaints, quality concerns, and pricing sentiment that drive last-minute hesitation
- **Recommendation Interaction Data**
  - Clicks, trust score patterns, and past response to personalized nudges

Modeling Approach

BEHAVIORAL SEGMENTATION MODEL

Groups customers into distinct behavior profiles (e.g., Review-Reliant, Routine Buyers, Explorers, Price-Sensitive) to understand who is most at risk of abandoning and why

CART ABANDONMENT PREDICTION MODEL

Predicts the likelihood a user will drop off using session data, friction signals, trust score, price sensitivity, and product attributes

SATISFACTION & FRICTION DRIVER MODEL (NLP + FEATURE IMPORTANCE)

Identifies the specific factors causing hesitation:

- Shipping delays
- Packaging complaints
- Quality inconsistencies
- Price competitiveness
- Review sentiment

This helps quantify why customers abandon

INTERVENTION UPLIFT MODEL

Recommends the best action for each segment, such as:

- Price-drop nudges
- Review reassurance prompts
- Bundle recommendations
- Amazon Basics alternatives

Forecasts conversion uplift from personalized interventions

Compliance and Data Quality

LEGAL & ETHICAL FACTORS

- Algorithmic Bias
- Dark-Pattern Nudging
- Data Privacy
- Transparency Standards

Compliance Frameworks: GDPR, CCPA, Explainable AI standards, consumer data-consent rules

Usability

- Regional-manager user testing
- Guided dashboards with interpretation tooltips
- Executive summary views for rapid decision-making

Feasibility

- Cloud-native infrastructure
- Microservice-based deployment
- Model-agnostic, scalable pipelines

Viability

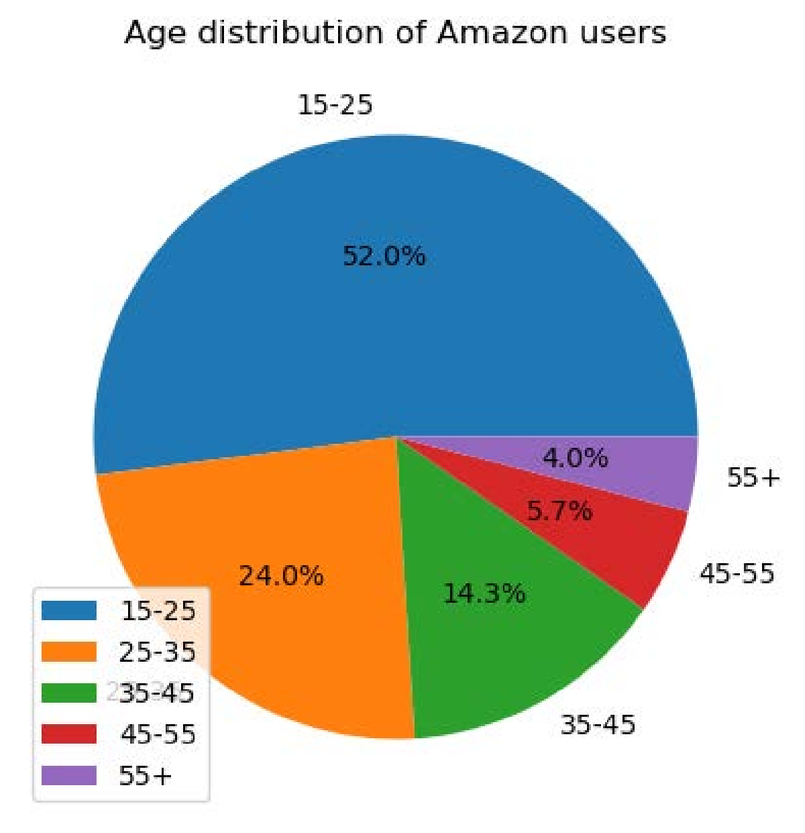
- Revenue lift from better conversion
- Improved marketing efficiency
- Higher retention
- Lower customer-experience operating costs

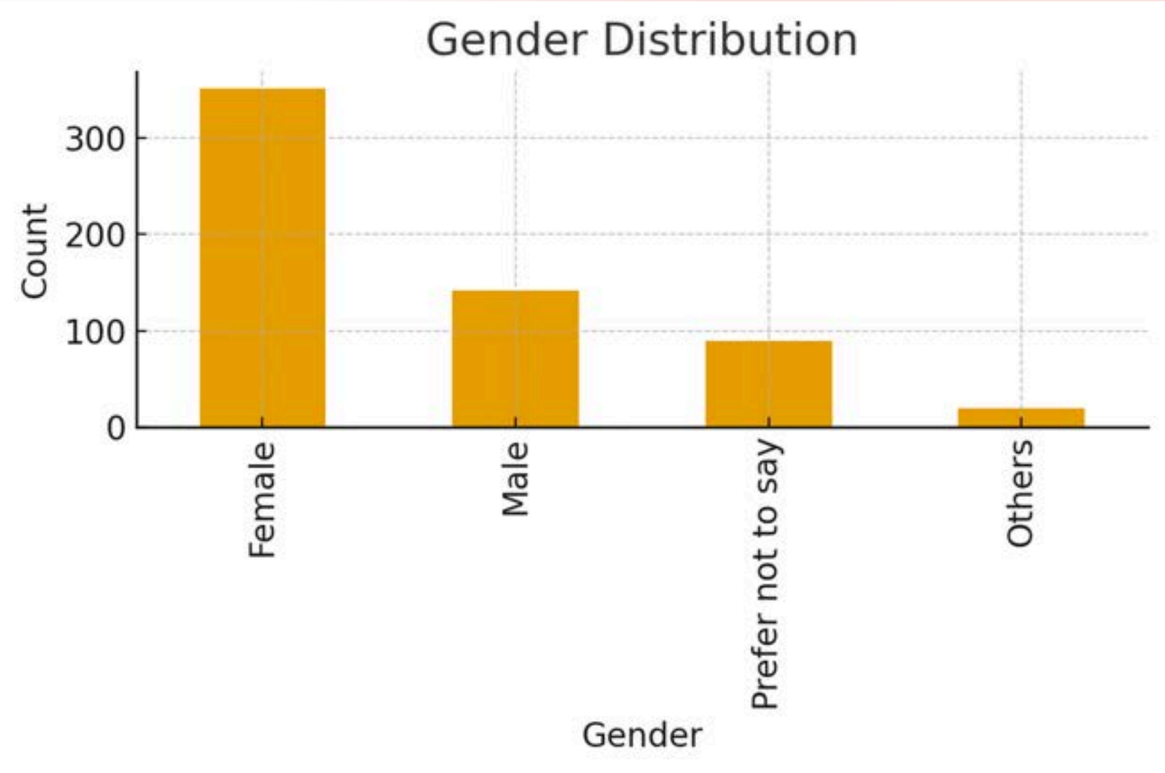
Dimension	Risk	Impact
Accuracy	Misreported fields, noisy clickstream logs	Incorrect predictions, poor segment assignment
Completeness	Missing demographic or session data	Unreliable clustering + biased models
Consistency	Inconsistent categorical labels across sources	Misalignment in segmentation + metrics
Validity	Outliers, bot traffic, corrupted logs	Skewed behavior patterns + false positives
Timeliness	Delayed ingestion or outdated behavior	Model drift and stale personalization
Uniqueness	Duplicate accounts or merged user IDs	Inflated frequency / unstable lifetime value

Analytics BUS Matrix

Business Process/Objective	Develop dynamic customer segmentation to personalize experiences, predict behavior, and improve conversion + retention.
Key Metrics (Facts)	CLV, AOV, purchase frequency, recency, category affinity scores, abandonment probability, retention rate, MoM behavioral drift.
Dimensions	Age, gender, region, acquisition source, device type, category preference, time period, session attributes, discount sensitivity, lifecycle stage.
Data Sources	User profiles (PII-compliant), transaction logs, clickstream/session data, customer feedback, historical orders, browsing behavior, time-series activity logs.
Owner/Stakeholders	Data Science, CRM/Growth, Product, Pricing, Analytics, Compliance.
DS/AI Opportunities	Clustering, time-series segmentation, predictive scoring models, next-best-action engines, price elasticity modeling, behavioral drift detection.
Ethics & Compliance	PII protection, explainability, non-discriminatory pricing and recommendations, transparent data usage, strict opt-in policies for personalization.

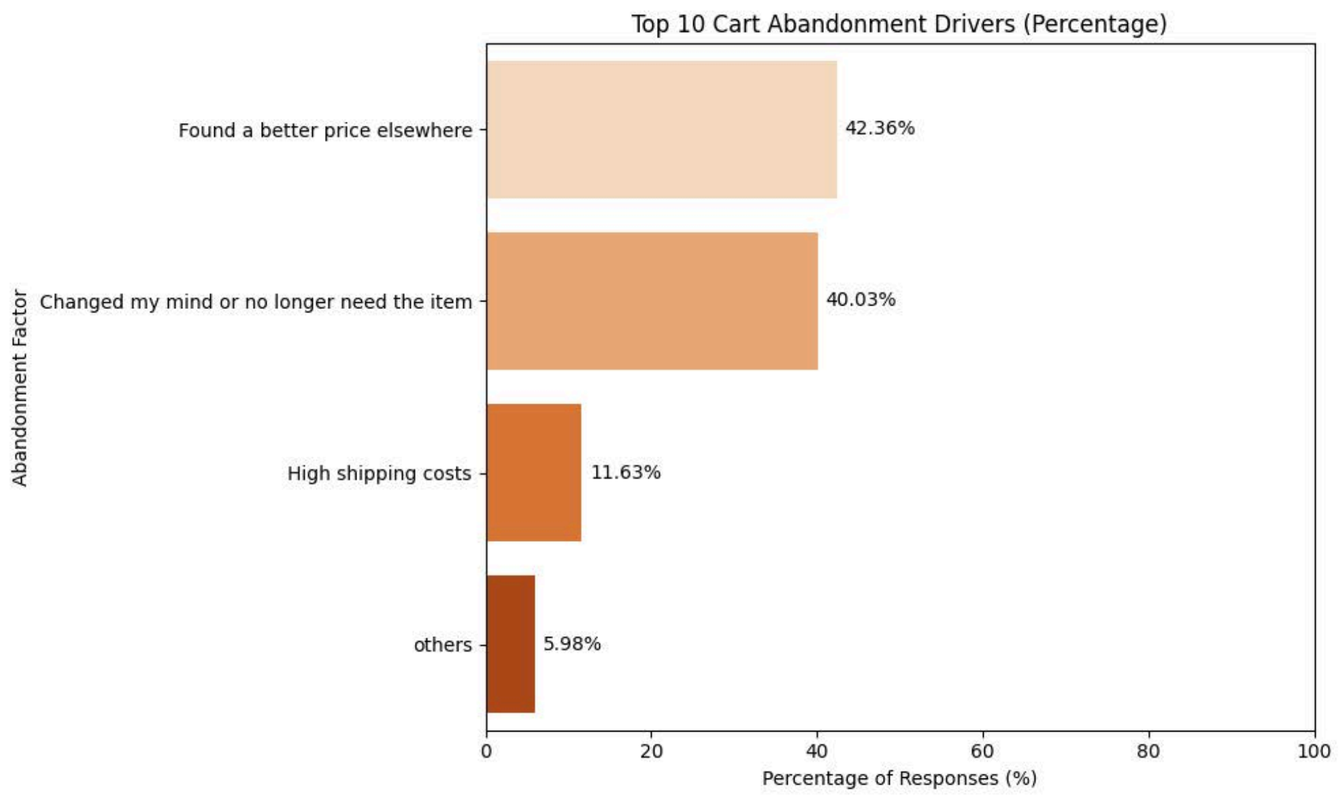
Who Are Our Users?





- **Younger users dominate the platform:** The 15–25 age group accounts for 52% of all respondents, making them the most active Amazon demographic in our dataset
- **Female users represent the majority**, indicating a gender-skewed engagement pattern on the platform

Where Do We Lose Them? (Abandoment Drivers)



- **Price Sensitivity is the top barrier -** the most common reason for cart abandonment is finding a better price elsewhere
- **Shoppers often change their minds -** deciding the item is no longer needed after adding it to the cart

Solution Design

Behavioral Segmentation model

Segmentation:
K-Means, Gaussian Mixture Models, Hierarchical Clustering

- flexible, probabilistic customer segments that evolve

Satisfaction & Friction Driver Model (NLP + Feature Importance)

NLP, PCA, SHAP,

- real-time personalization with exploration

Cart Abandonment Prediction Model

Predictive Modeling:
LightGBM, CatBoost, Logistic Regression

- strong with tabular behavioral data

Key Metrics for Models

- Customer Lifetime Value, RFM (Recency, Frequency, Monetary) scores, purchase frequency
- Category affinity, price sensitivity
- Session depth, recency, time-between-purchases
- Lifecycle stage, churn/abandonment probability
- Segmentation filters: age, region, device, time period, category

Modeling Precaution

- Acknowledge demographic bias in modeling because dataset is skewed
- Privacy/PII constraints (GDPR/CCPA) on personal data usage
- Model drift if behavior changes over time
- Interpretability challenges with deep models (LSTM, Transformers)
- Data leakage in time-series modeling if not carefully split

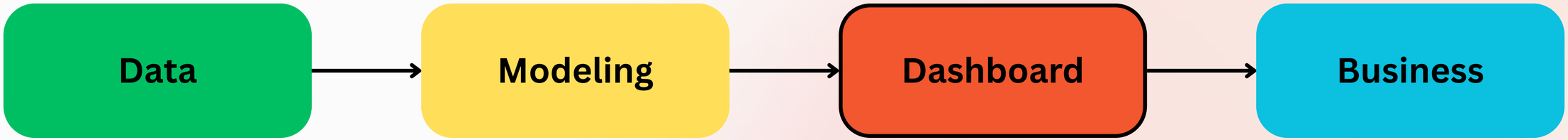
Implementation Roadmap

	Phase 1: Month 0-4	Phase 2: Months 5–8	Phase 3: Months 9–12
Product Focus	Behavioral segmentation model	Operational dashboard	Real-time, automated global system
Primary Users	Data Scientists, Data Analysts, Business Analysts, Pilot Users(Selected managers)	Regional Managers, Category Managers, Operations & Marketing Teams	Global Executives, Enterprise Growth Teams, Global Category Strategy Teams
Key Objectives	• Build the first version • Identify key patterns • Validate	• Translate analytical outputs into usable business insights • Enable day-to-day decision-making	• Fully automate model • Enable real-time intelligence • Support executive-level decisions
Key Activities	• Data ingestion & feature engineering • Initial clustering model development • Rolling model performance testing • Segment stability & interpretability checks	• Dashboard UX design • KPI visualization • QA, UAT • Regional pilot deployment	• Model serving API deployment • Real-time data pipelines • Global A/B testing of interventions • Enterprise security & access controls
Risk Level	High	Low–Medium	High
Considerations	• Continuous rolling validation on new data • Back-testing on historical quarters • Early feedback from pilot users • Adjusting segment definitions	• Usability and clarity of insights • Stakeholder trust in model outputs • Data freshness and refresh frequency • Clear interpretation guides	• Phased rollout by region • Fail-safe override mechanisms • Strict monitoring • Heavy governance

Dependency & Scalability

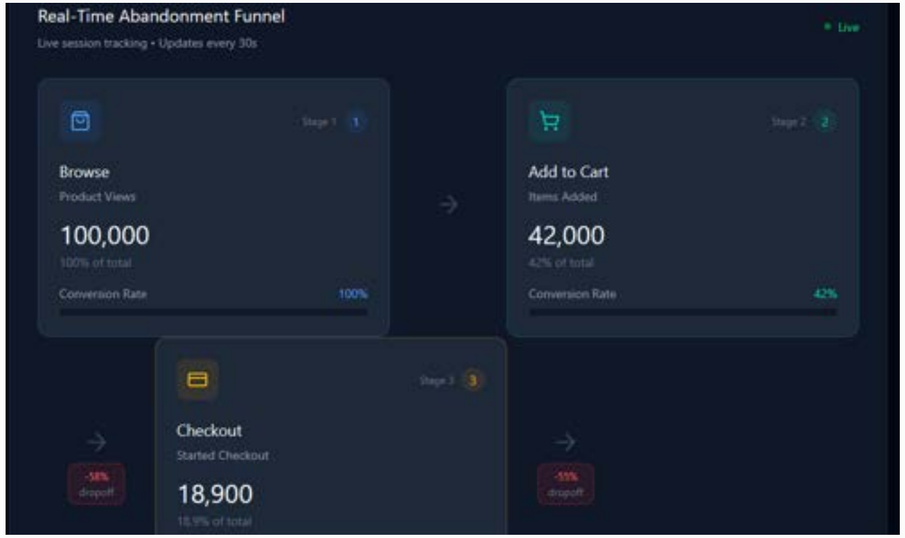
- **Dependency**
  - Phase 1 → Phase 2
    - Stable behavioral segments
    - Validated feature engineering pipeline
    - Model validation and pattern discovery
    - Pilot feedback
  - System-level dependencies
    - Consistent pipeline orchestration (ETL/ELT)
    - Feature store alignment across teams
    - Cross-functional readiness (product, DS, engineering)
  - Phase 2 → Phase 3
    - Proven usability and business adoption
    - Refined KPIs and decision rules
    - Successful regional pilots
    - Clear error-handling and interpretation guides
- **Scalability**
  - Works on high-volume behavioral data → big impact at scale
  - Modular, model-agnostic architecture
  - Reusable across product categories
  - Supports both operational and strategic decisions
  - Integrates with growth, recommendation, and CX systems

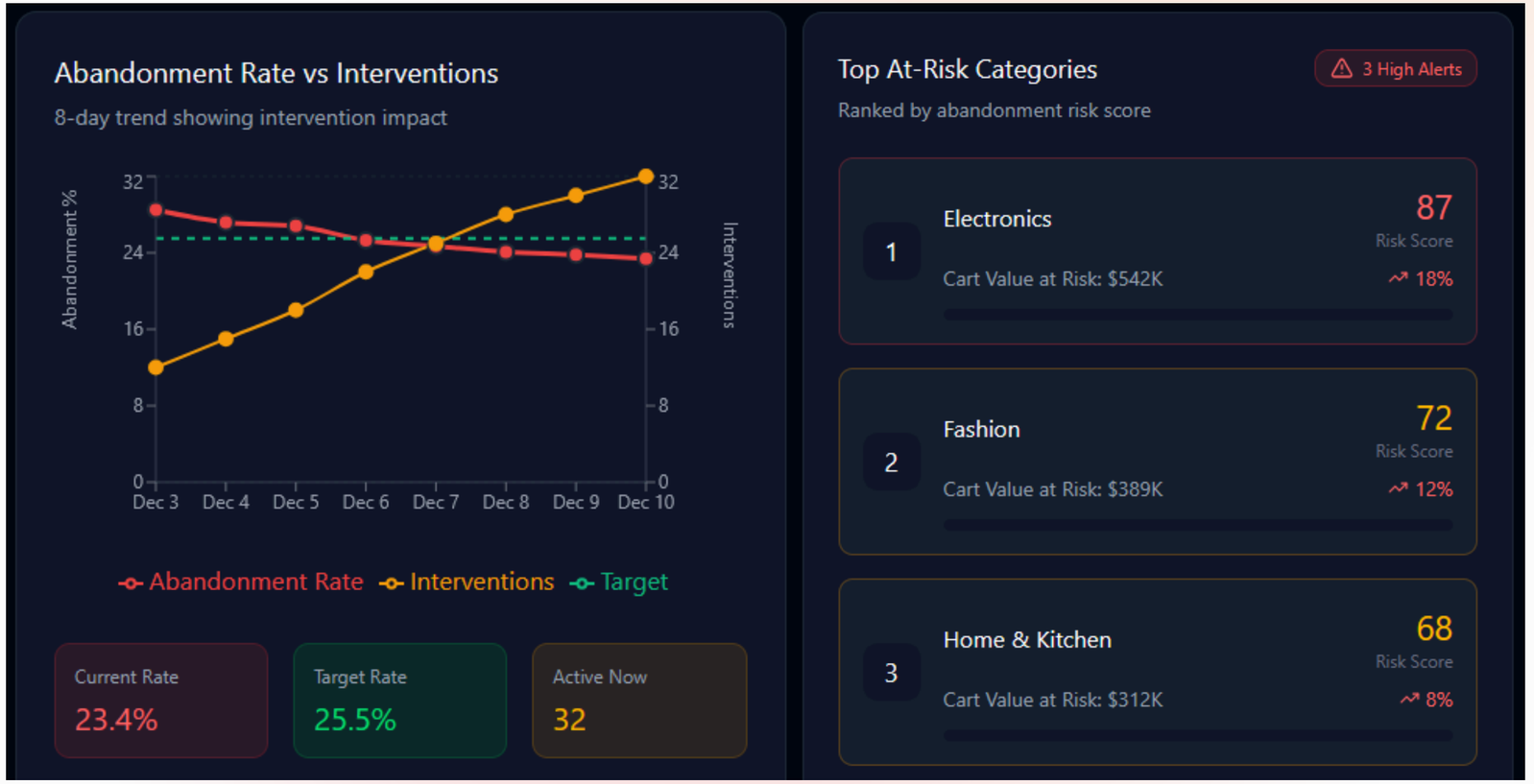
Demo



Cart Abandonment Tab

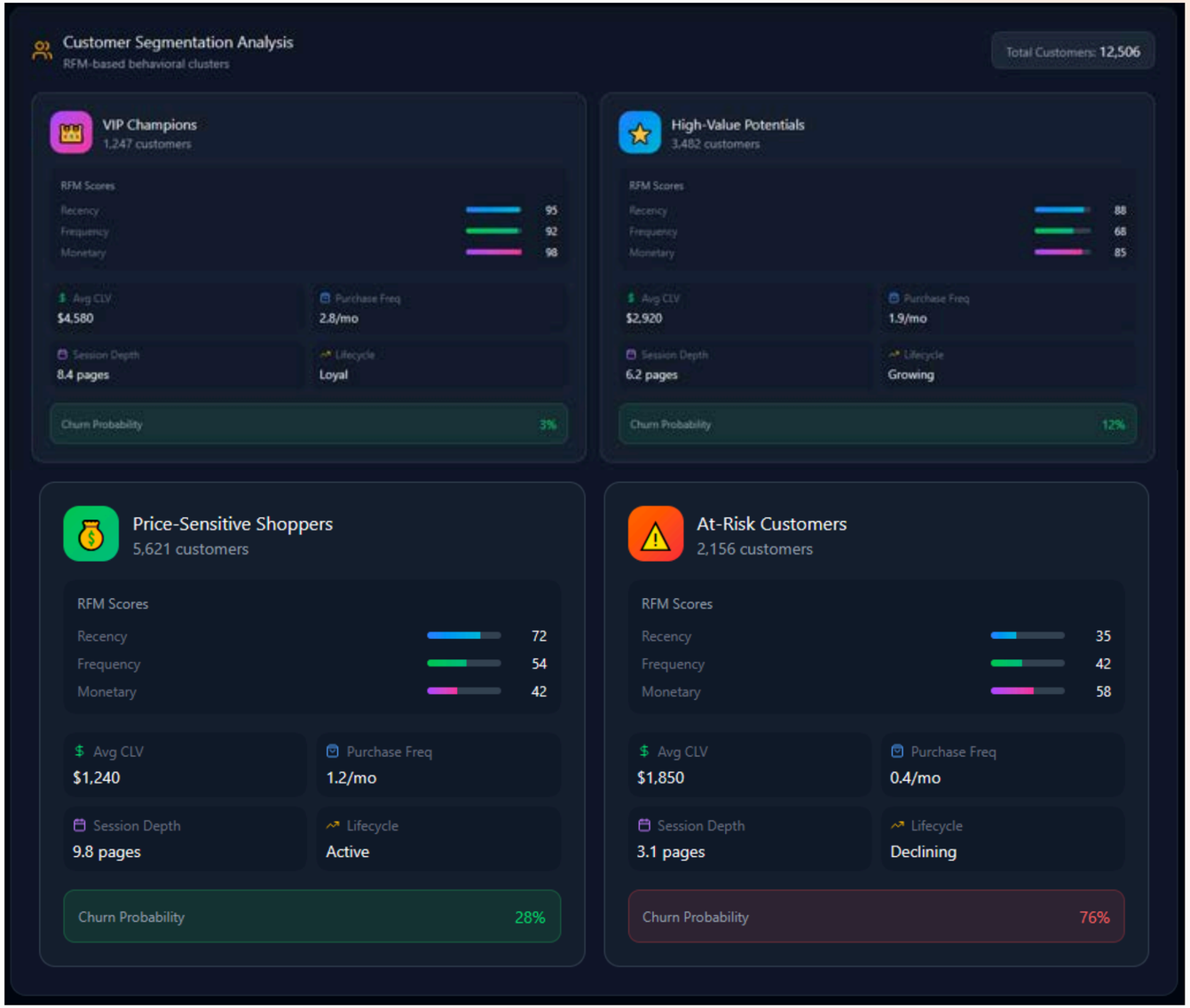
- **Who is most likely to abandon their cart?**
- **What segments require retention or incentives?**
- **Where in the funnel are customers dropping off?**
- **How can we proactively intervene (alerts, offers, reminders)?**





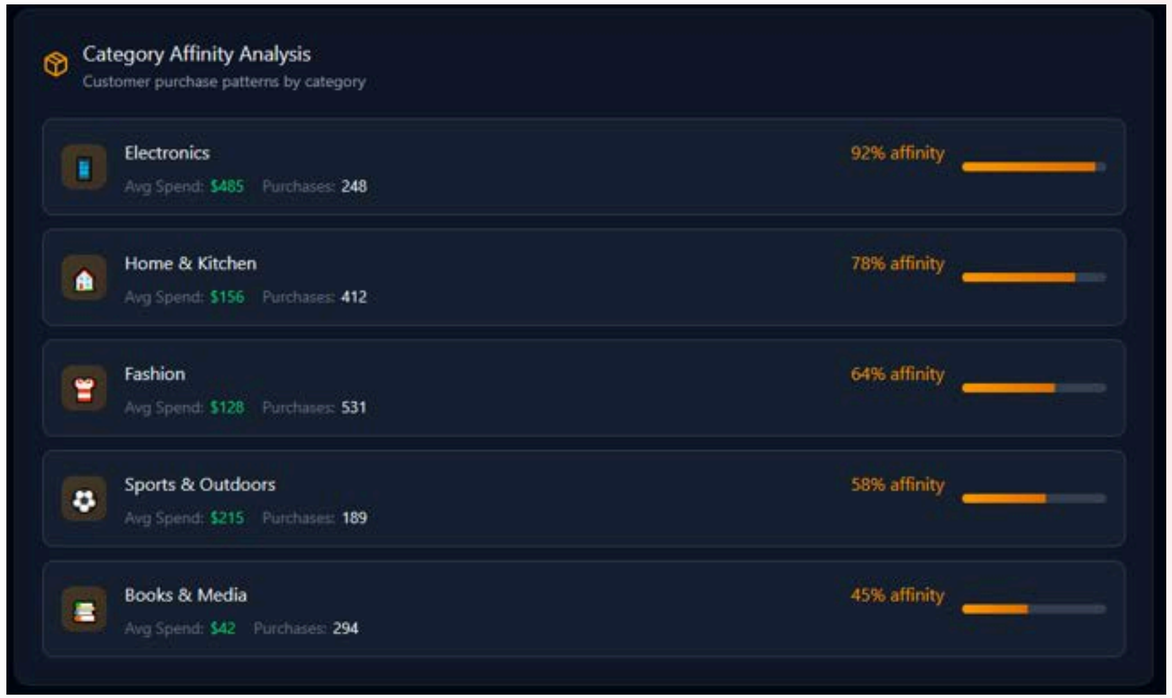
Customer Segmentation Tab

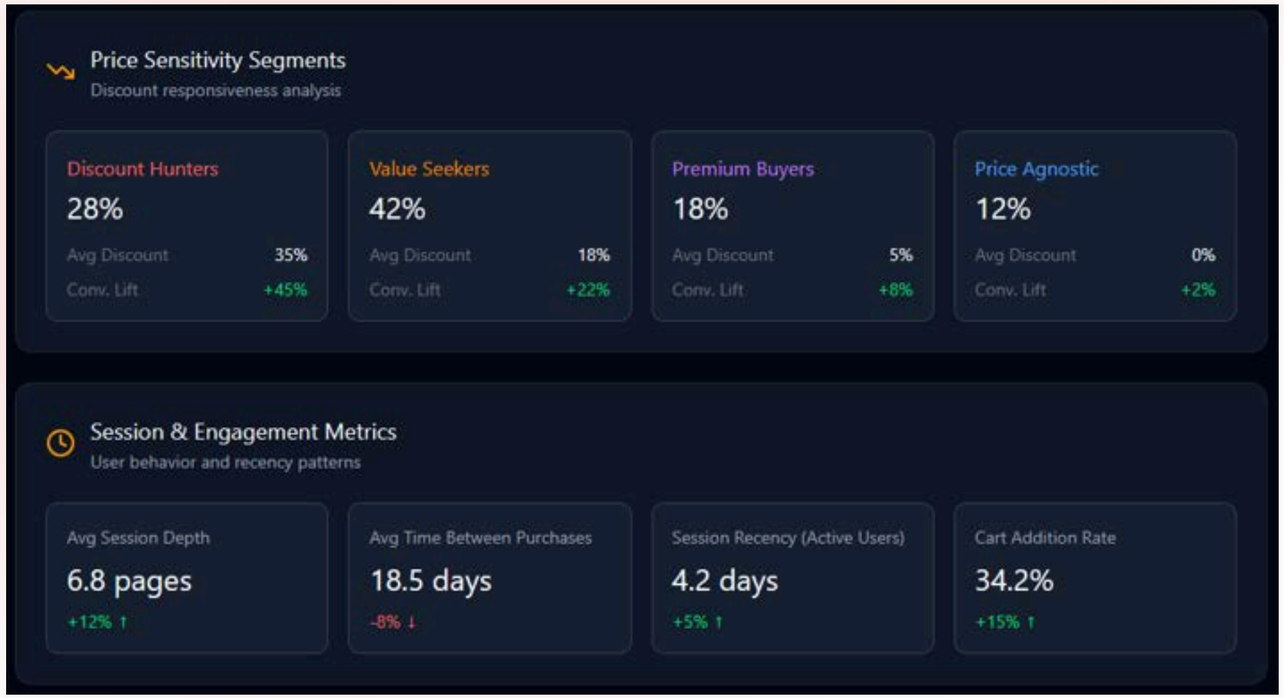
- **Which customers drive the most value?**
- **Which groups are declining or at risk?**
- **How does behavior differ across segments?**
- **Where should marketing focus retention efforts?**



Behavioral Insights Tab

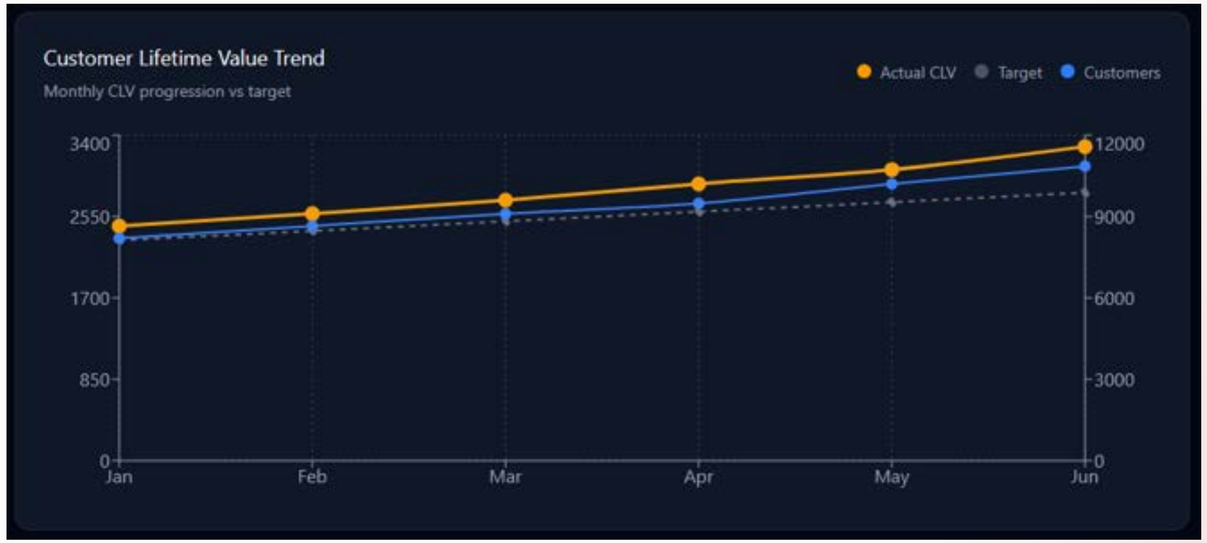
- What categories or behaviors define our customers?
- How sensitive are customers to discounts or pricing changes?
- How deeply do different segments explore the site?
- Are customers buying more frequently over time?

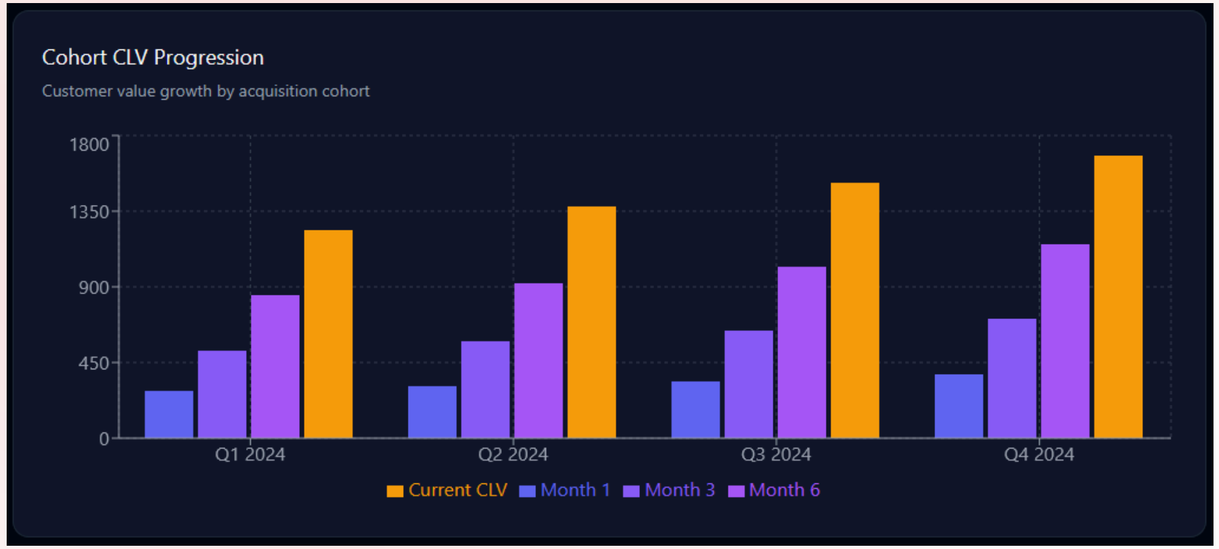




CLV Analytics Tab

- How valuable are our customers now?
- How quickly do we recover acquisition spend?
- Are our cohorts improving or declining in value over time?
- What will customer value look like in the next 12 months?





Business Value Summary

1. Measurable Financial Impact

- Reduction in harmful cart abandonment through predictive identification
- Higher conversion rates via segment-specific nudges and personalized recommendations
- Improved retention by identifying at-risk cohorts early
- Stronger CLV growth driven by timely and relevant engagement
- Marketing efficiency gains from better targeting and reduced blanket discounting

Net effect: Direct revenue lift, optimized spend, and healthier lifetime customer value.

2. Operational & Organizational Impact

- Unified dashboard enabling executives, regional managers, and analysts to act consistently
- Faster, smarter decision-making through real-time insights
- Better regional performance using consistent KPI frameworks
- Improved collaboration across Product, Marketing, Pricing, and Data Science teams
- Reduced operational friction via automated segmentation and prioritization

3. Customer Experience Impact

- More relevant offers based on price sensitivity, behavior, and lifecycle stage
- Fewer interruptions and better timing thanks to predictive risk scoring
- Improved trust and transparency through explainable predictions
- Higher satisfaction driven by better product recommendations and friction reduction

4. Long-Term Strategic Value

- Foundation for AI-driven personalization at enterprise scale
- Scalable segmentation and prediction framework applicable across Amazon verticals
- Real-time intelligence that evolves with customer behavior
- Strengthens Amazon’s competitive moat through superior behavioral understanding

Overall: A high-impact, scalable intelligence system that improves revenue, operational efficiency, and customer satisfaction—while building Amazon’s next-generation predictive commerce infrastructure.