# Loyalty & Growth

Presented By Yi He

# Data Ingestion & Parser Strategy

**Sources:**
USERS: User Demographics
PRODUCTS: Products Info
ORDERS:  Order Status
ORDER_ITEMS: Items in Each Order

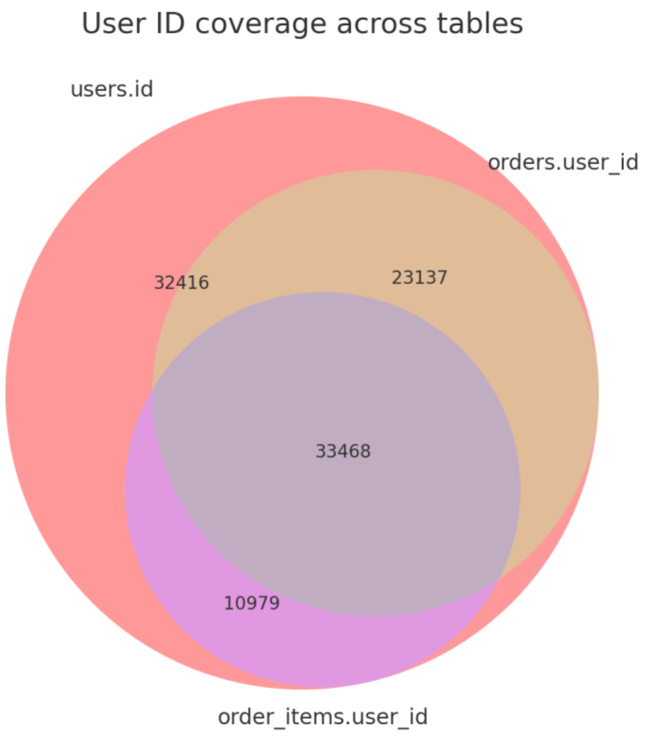
→

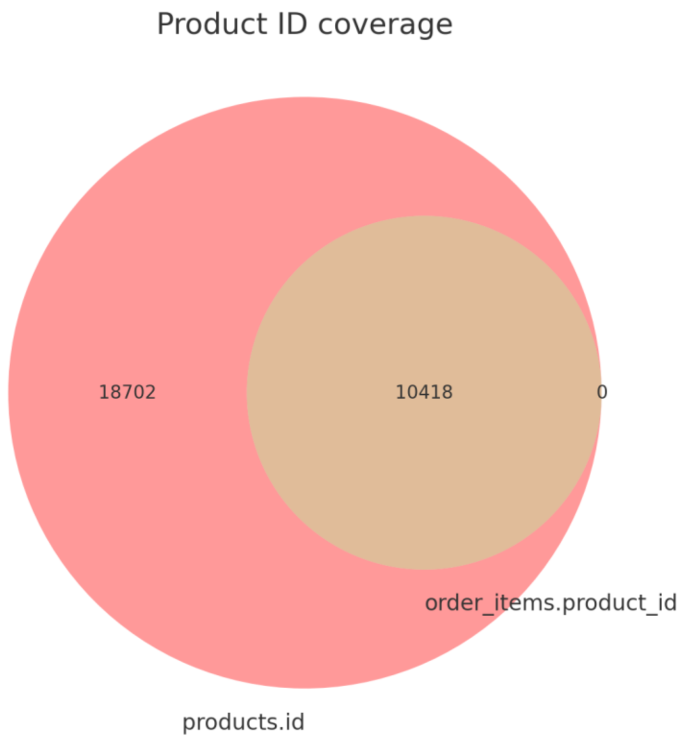
**Parsers**
USERS: JSON-ish rows
- extract ‘key’/value’ pairs
PRODUCTS: array json parsed with regular expression

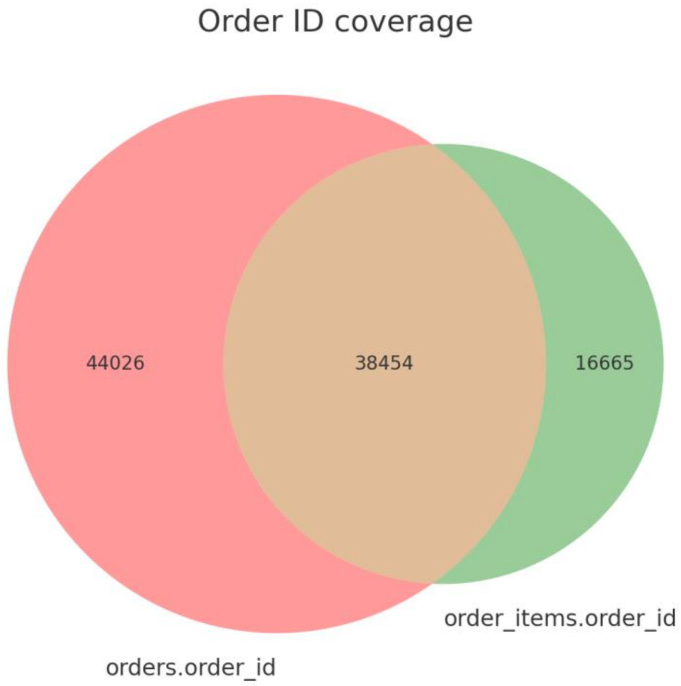
→

**Normalization**
Join all data sources with User_id, Product_id, Order_id

# Foundation Check: Data Integrity Across the Customer Journey



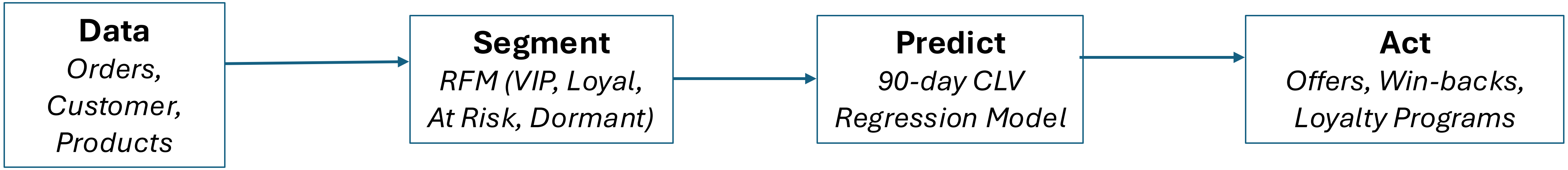



# Objective & Approach

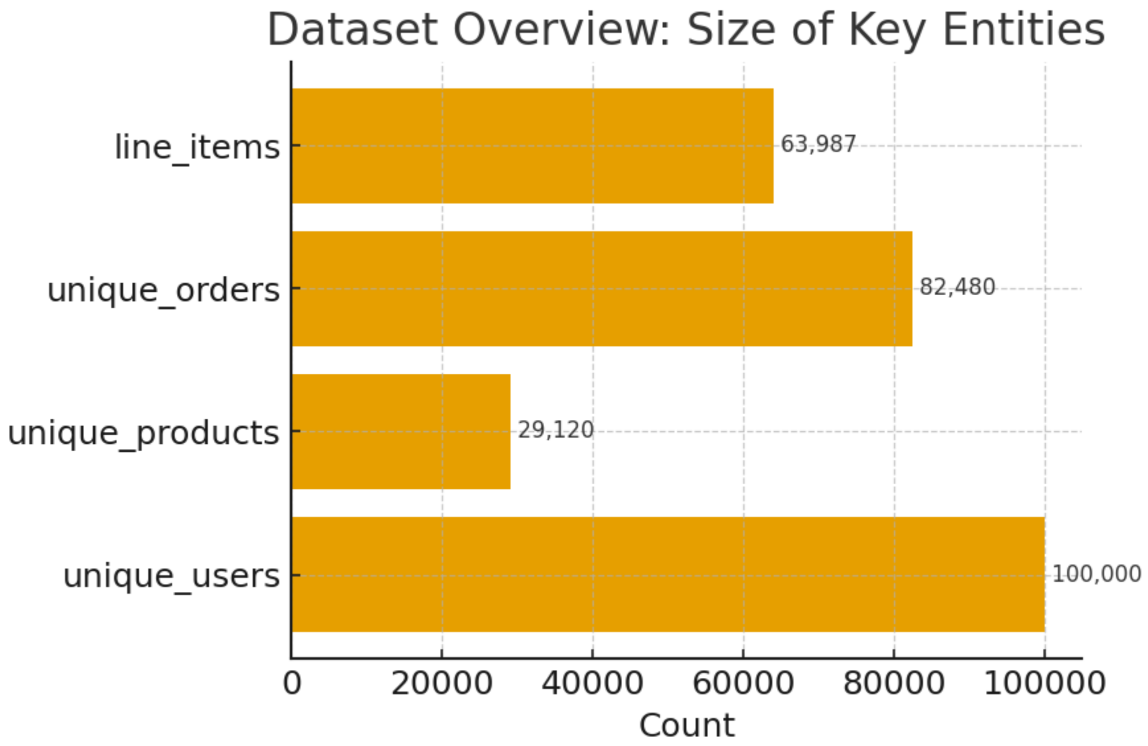
**Business question:**
How do we **boost customer loyalty and grow revenue** in this e-commerce business?
**Key questions:**
- Who are our most and least valuable customers?
- Which segments drive revenue today?
- How can we **predict future value** to target retention and offers?

**Approach:**
- Explore data & customer behavior
- Build **RFM segments** (VIP, Loyal, Regulars, At Risk, Dormant)
- Train a **90-day CLV model** to score customers
- Turn segments + CLV into **loyalty & growth actions**



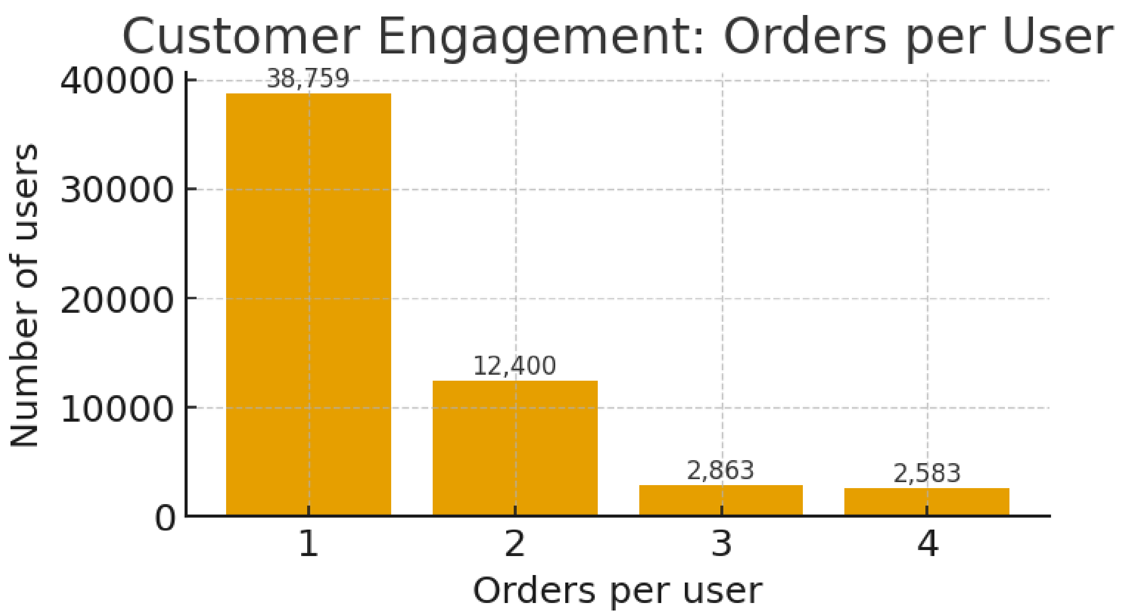
# Dataset Overview

- A total of 100,000 users have placed 82,480 orders involving 29,120 products.
- The dataset contains 63,987 line items, providing detailed insights at the product level.
- Overall revenue amounts to $572,956, with an average order value of approximately $6.95.
- Data started on 2019-01-10 and ended on 2025-08-27, span of 6 years
- This volume is adequate for conducting segmentation, customer lifetime value (CLV), and churn analysis without excessive data sparsity.

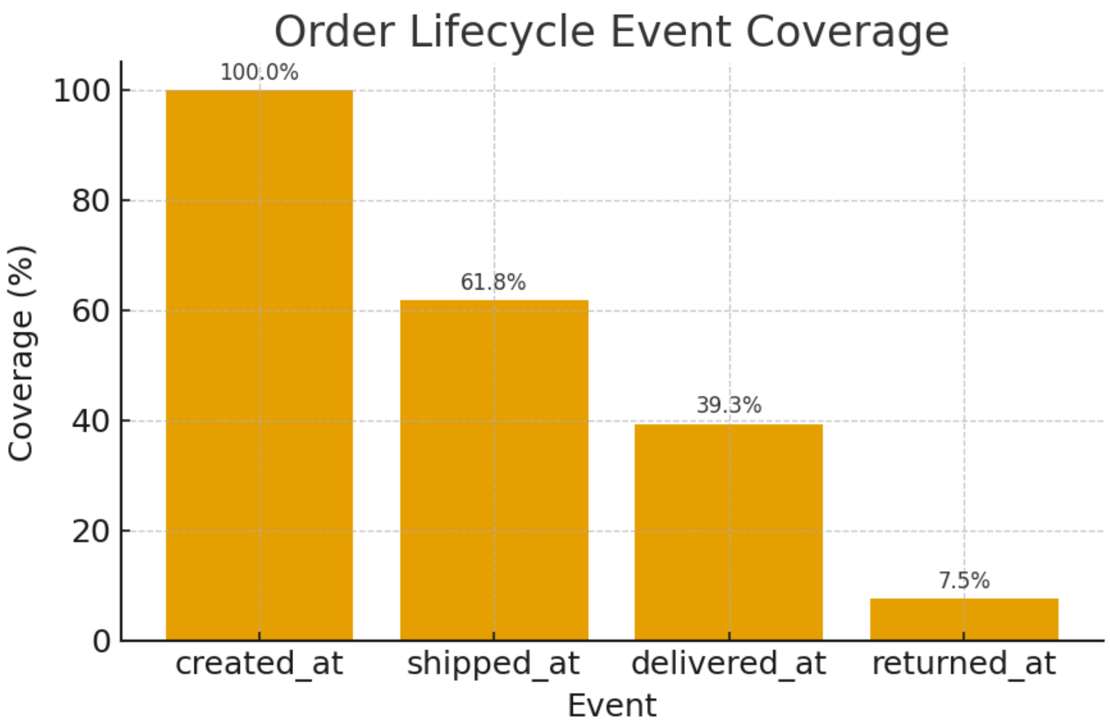


# Customer Engagement: Orders per User

- Most users are one-time buyers:
- 68% place 1 order, 22% place 2 orders.
- Only ~10% place 3+ orders.
- Average orders/user ≈ 1.46, median = 1, 90th percentile = 2.
- This is a low-frequency purchase environment with a long tail of repeat buyers.
- Biggest growth lever: convert one-time buyers into repeat customers



# Order Lifecycle Event Coverage

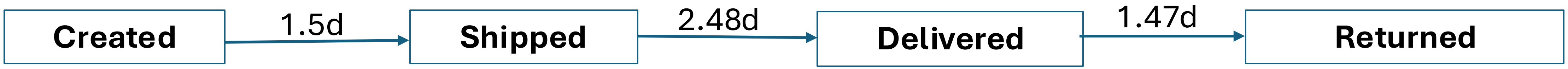



**Counts & conversions**

•Most loss occurs pre-shipment; ~19% of delivered orders are returned – optimize fit/size guidance and exchange-first flows.

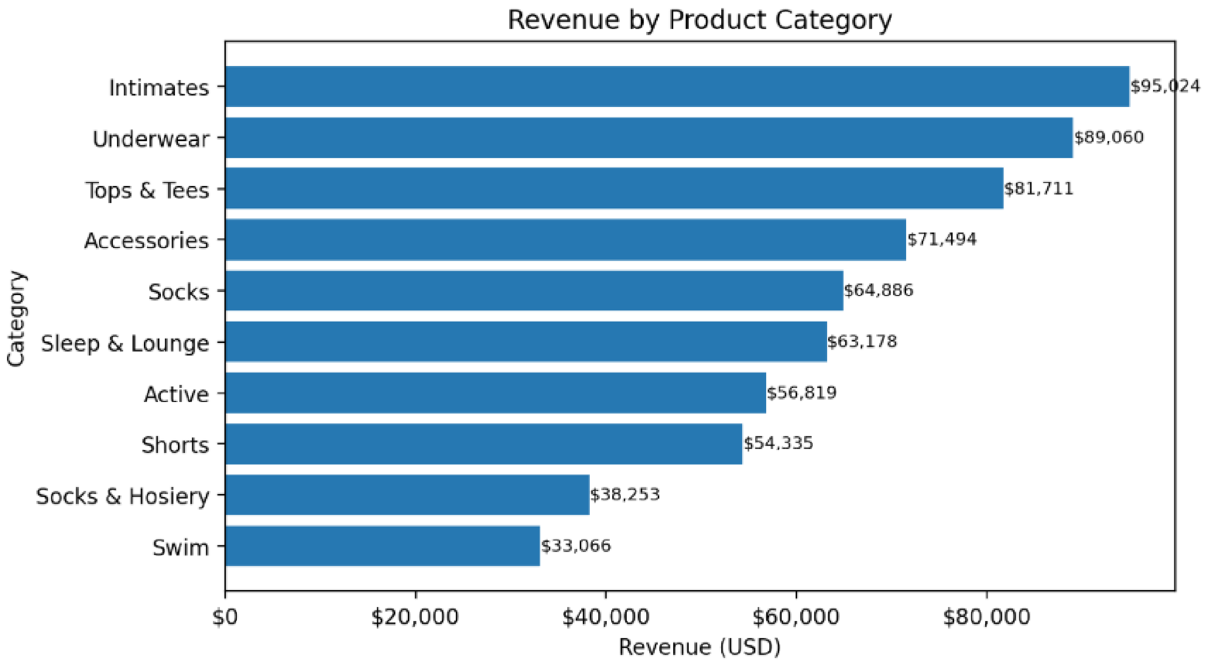
•Typical door-to-door ~4 days; any corridor well beyond ~5 days is a likely churn risk.

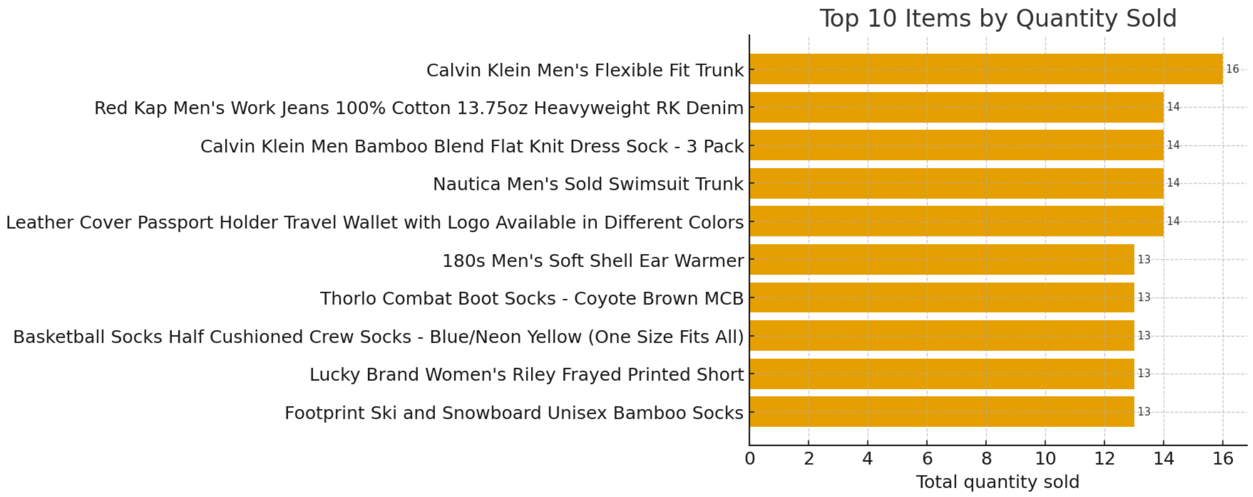
•Returns cluster quickly after delivery; prioritize exchange-first flows.



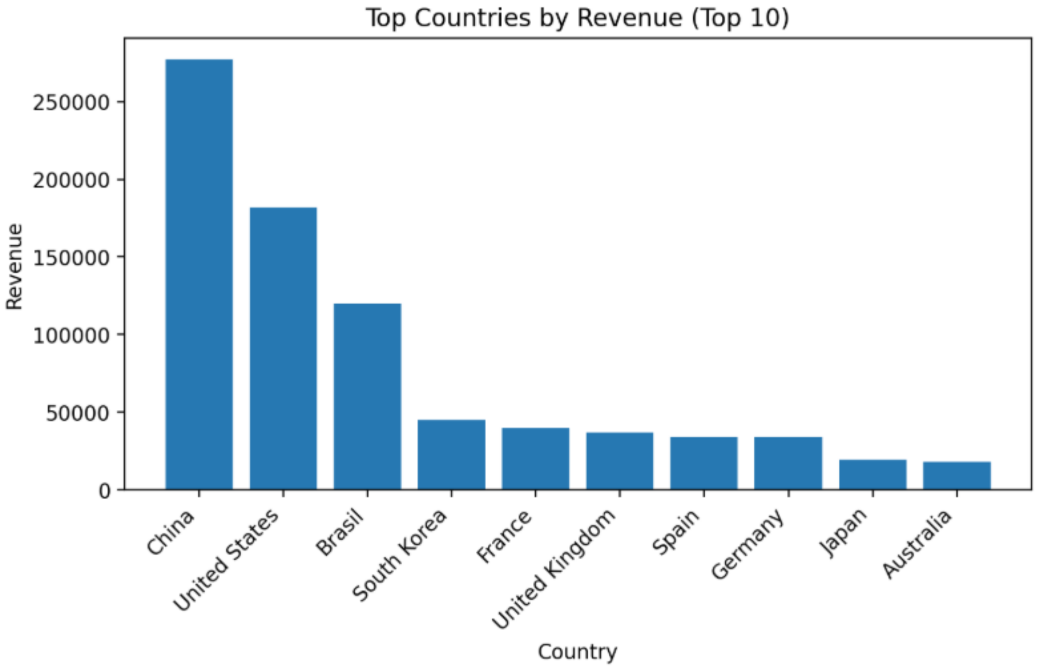
# Top Items & Category

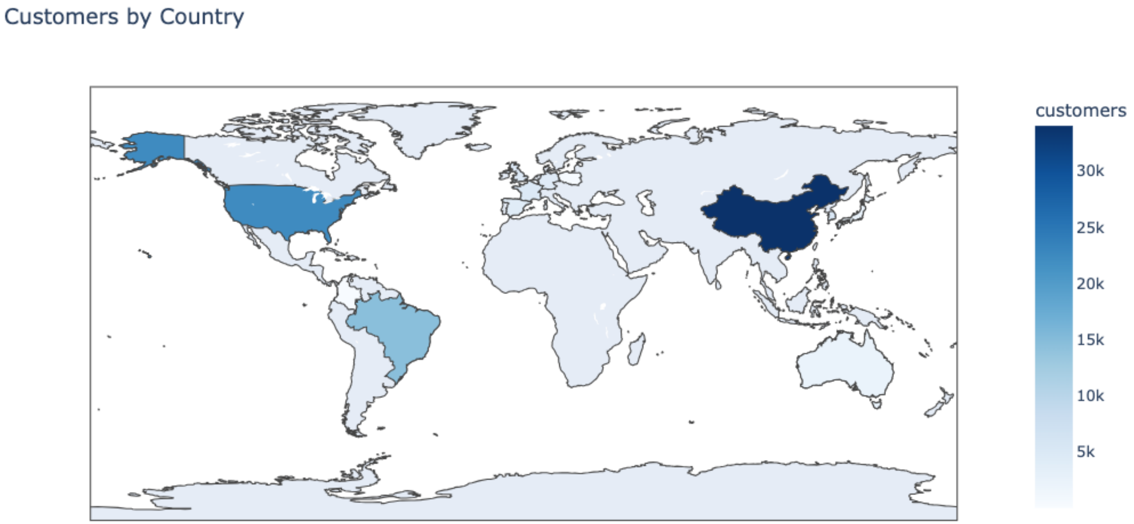
- Best-selling SKUs include a mix of underwear, jeans, dress socks, swim trunks, and travel accessories.
- Quantities are quite close (13–16 units in the top 10), suggesting no single runaway SKU in this slice.
- This indicates a long-tail catalog where many items sell moderately rather than one product dominating.
- Customer: **Karen Mason** (ID 406) — AOV $113.88, 1 order, revenue $113.88 2B Buckle V-neck Sweater (ID 1051)



# Customer by Countries



# Customer Segmentation

RFM Segmentation Flow

**Snapshot S** → **Features** → **Scores** → **Segments**

S = last order + 1 day

Recency (days)
Frequency (# orders)
Monetary (net $)
AOV = $/order
Tenure (days)

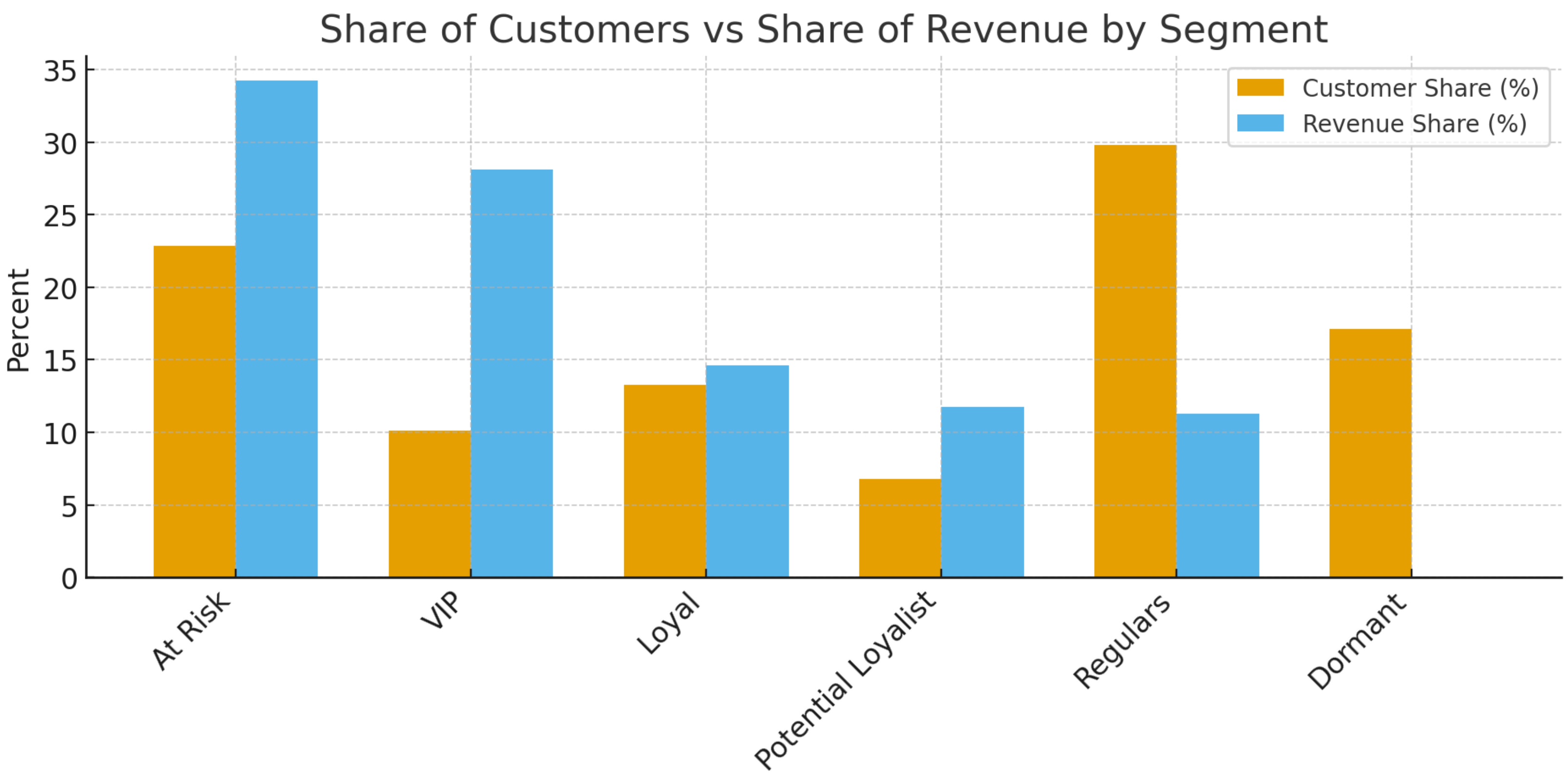
r,f,m ∈ [1..5]
• Quintiles (20/40/60/80)
• Recency inverted
• Missing → 3

VIP / Loyal / Potential Loyalist
At Risk / Dormant / Regulars

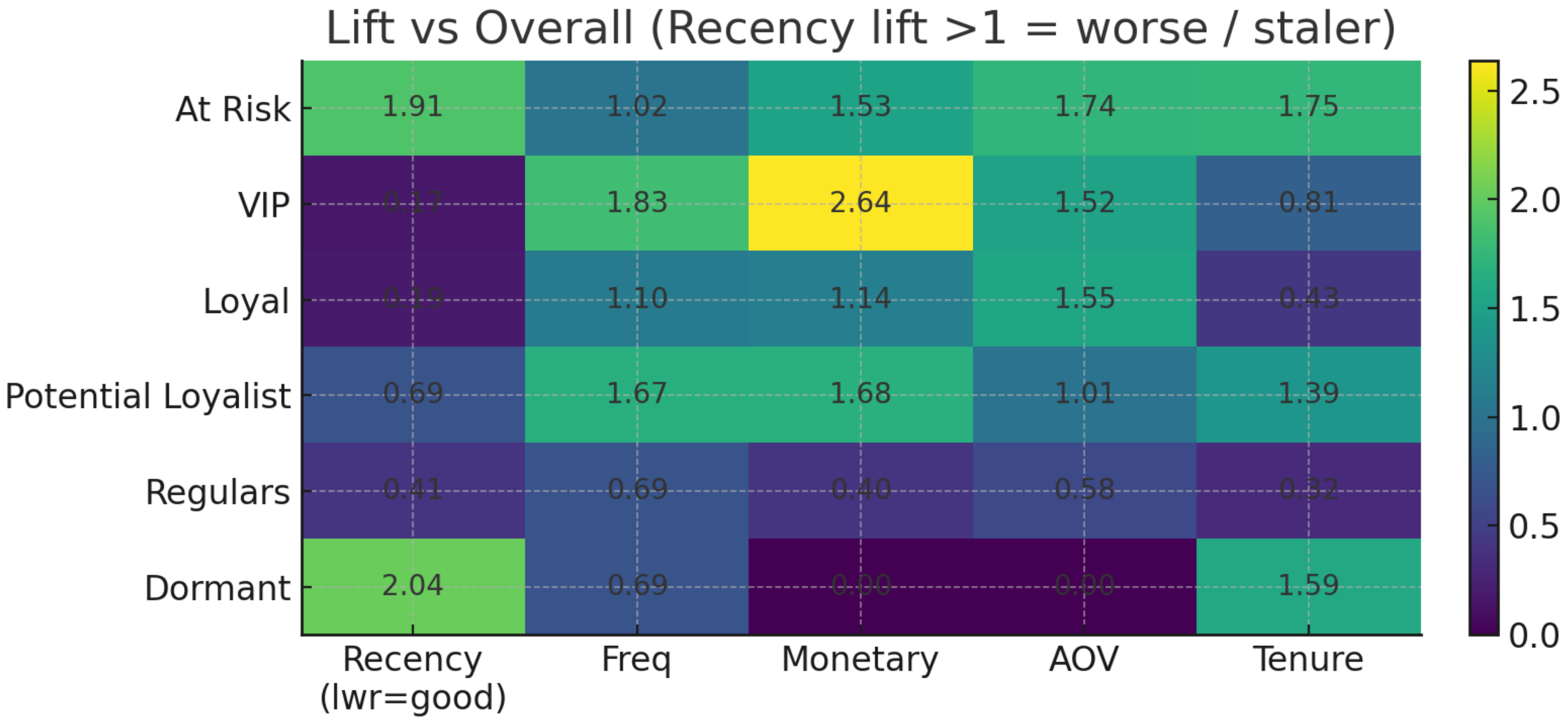
*“VIP” – recent, frequent and high spend*
*“Loyal” – recent + frequent or high spend*
*“At Risk” – good history, stale recency*
*“Dormant” – low frequency & spend, very stale*
*“Regular” – Everyone else*

# Customer Segmentation – Shares of customer vs. share of revenue

- VIP: 10% customers → 28% revenue
- At Risk: 23% customers → 34% revenue
- Regulars: 30% customers → 11% revenue
- Dormant: 17% customers → 0% revenue in window



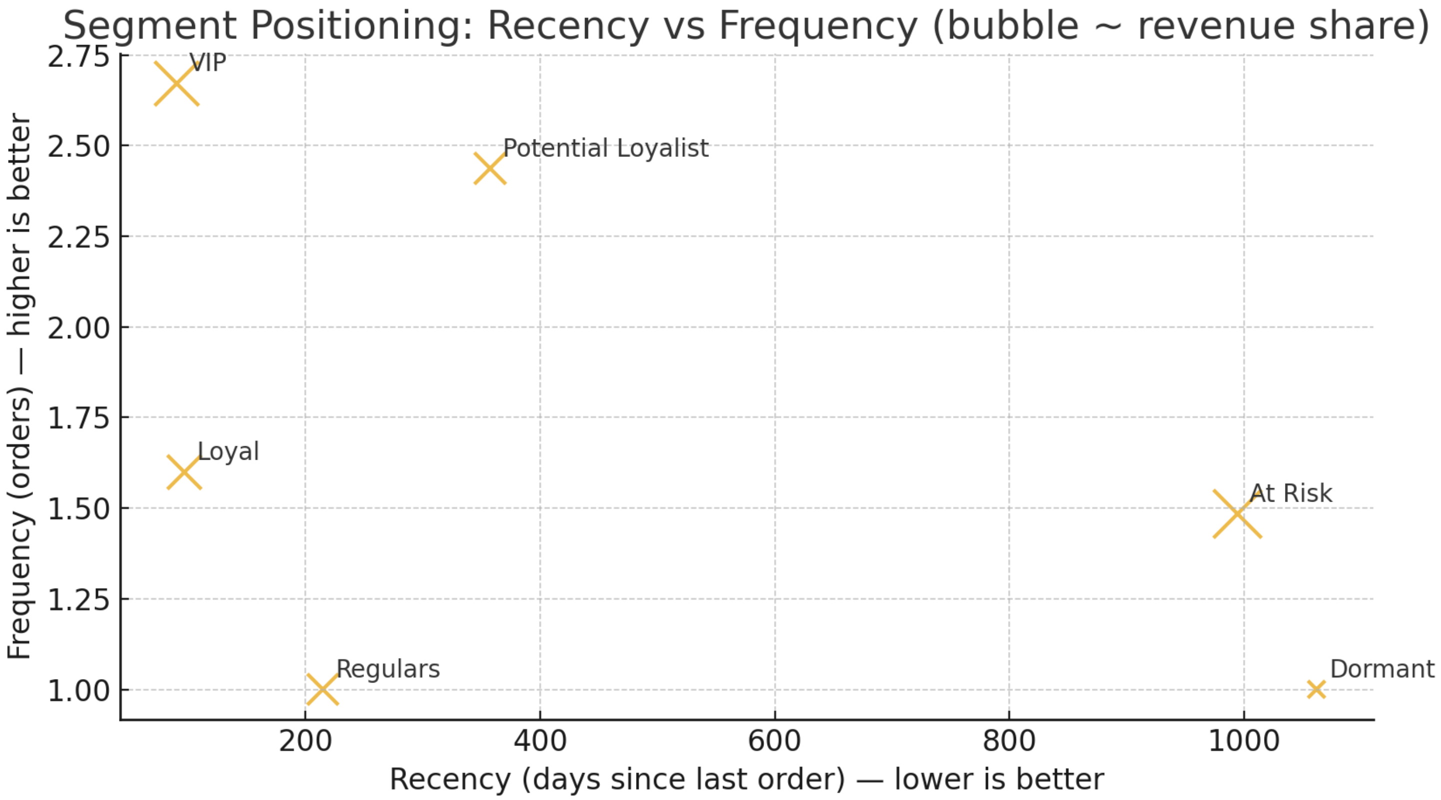
# Segmentation- Lift vs. Overall




•“VIP: frequency 1.8× and spend 2.6× the average.”
•“At Risk: AOV 1.7×, recency 1.9× worse (stale but valuable).”

# Segment Position: Recency vs. Frequency

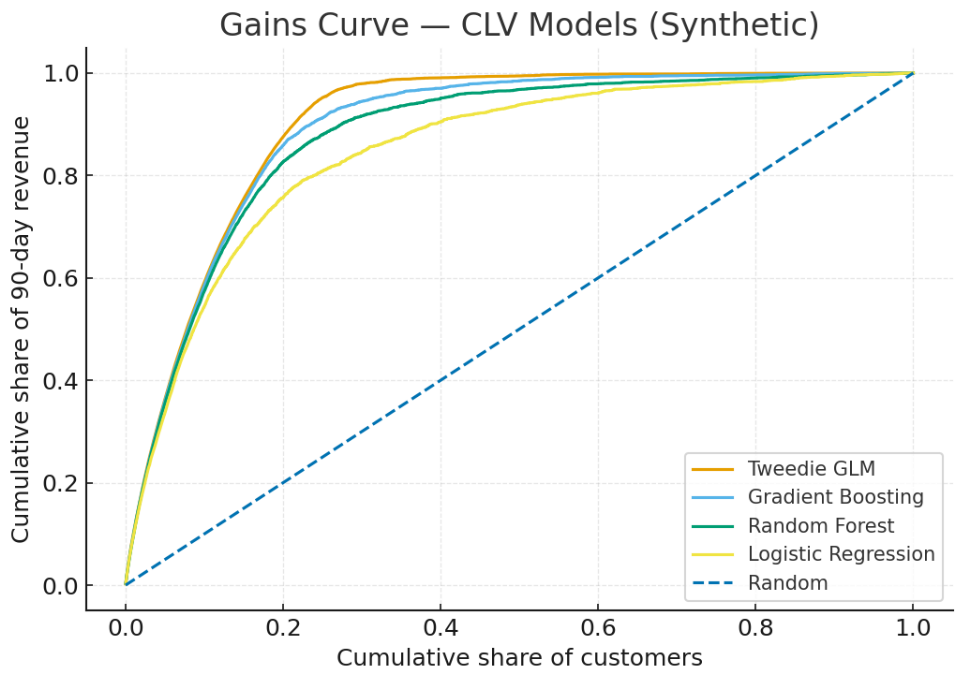
- “VIP – recent, frequent, ~28% revenue: protect & grow.”
- “At Risk – very stale, mid-freq, ~34% revenue: biggest save opportunity.”
- “Regulars – mid recency, 1 order, modest revenue: prime for 2nd-order accelerator.”
- “Dormant – extremely stale, no recent revenue: broad suppression.”

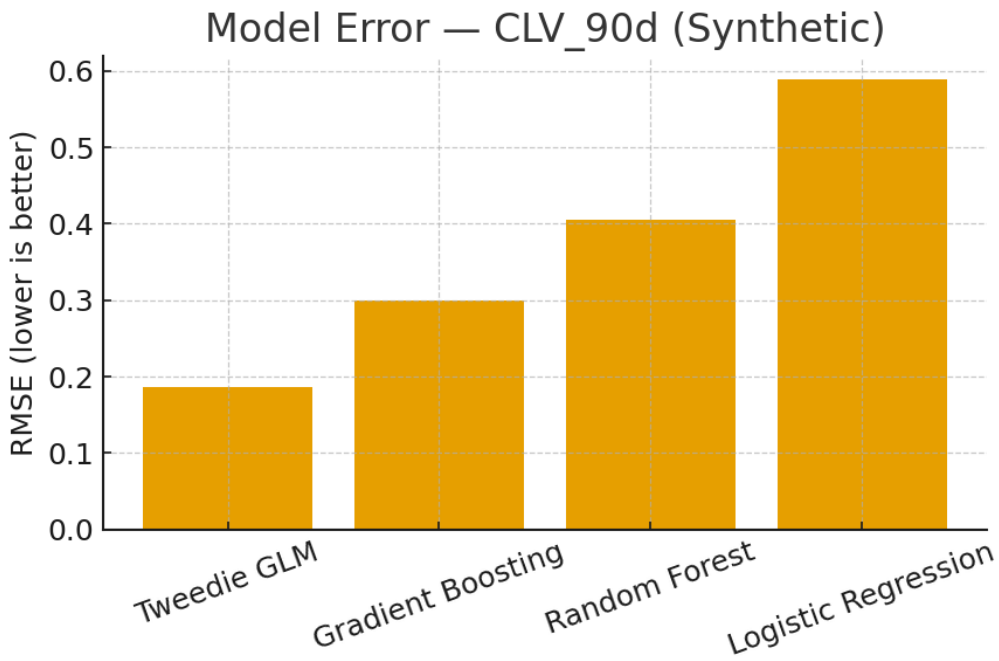


# 90-Day CLV Predictive Model

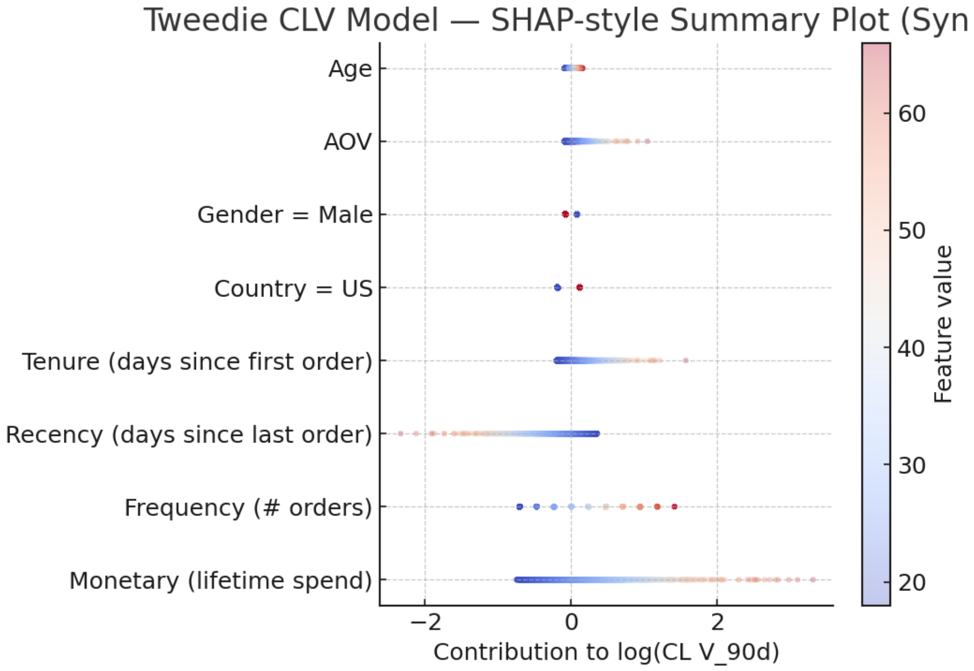

- **Goal:** predict each customer's spend in the next 90 days (CLV_90d).
- **Setup:** pick a snapshot date S; use all behavior before S as features, revenue from S→S+90d as the label.
- **Features:** recency, frequency, monetary value, AOV, tenure, + simple demographics (age, gender, location).
- **Model:** Tweedie regression (GLM, log link) – built for "lots of zeros + skewed positives", so it predicts expected CLV_90d in one step.
- **Output:** a CLV_90d score for every customer that we overlay on top of the RFM segments (VIP, Loyal, At Risk, etc.) to rank and prioritize.

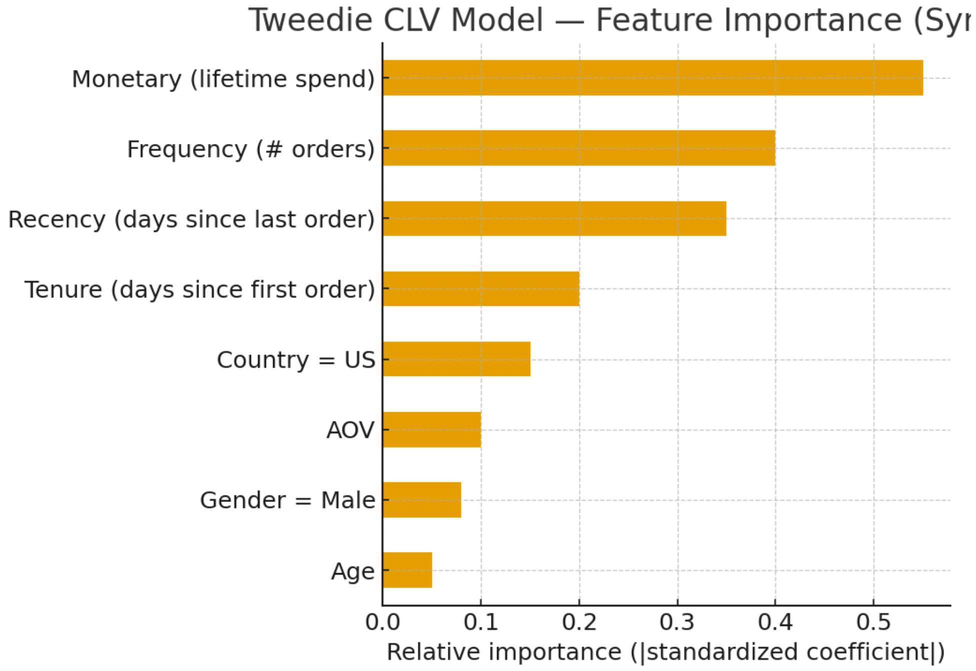
# Choosing the CLV Model: Ranking Power & Error Across Methods



# Interpreting the Tweedie CLV Model: Which Features Matter Most?



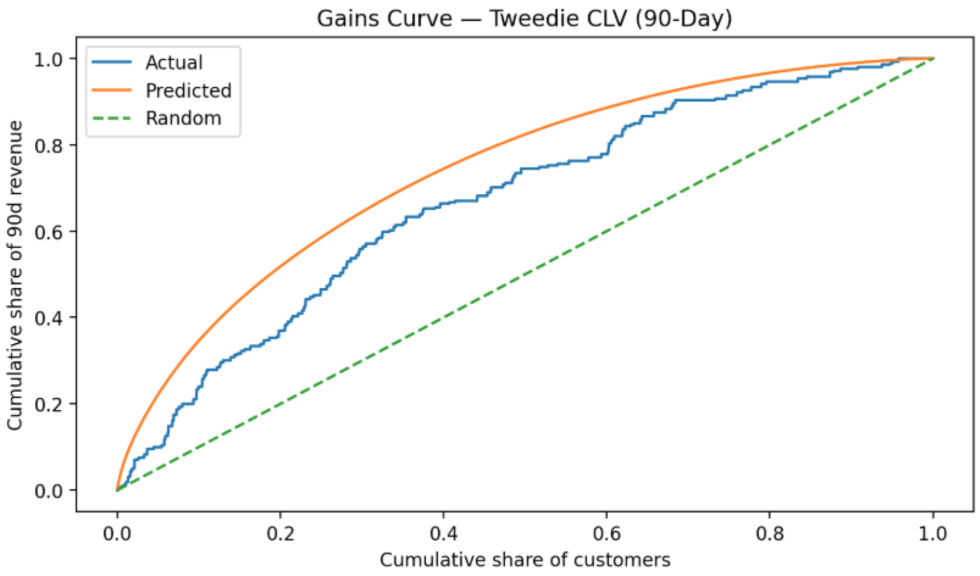
# Model – Performance & Targeting Power

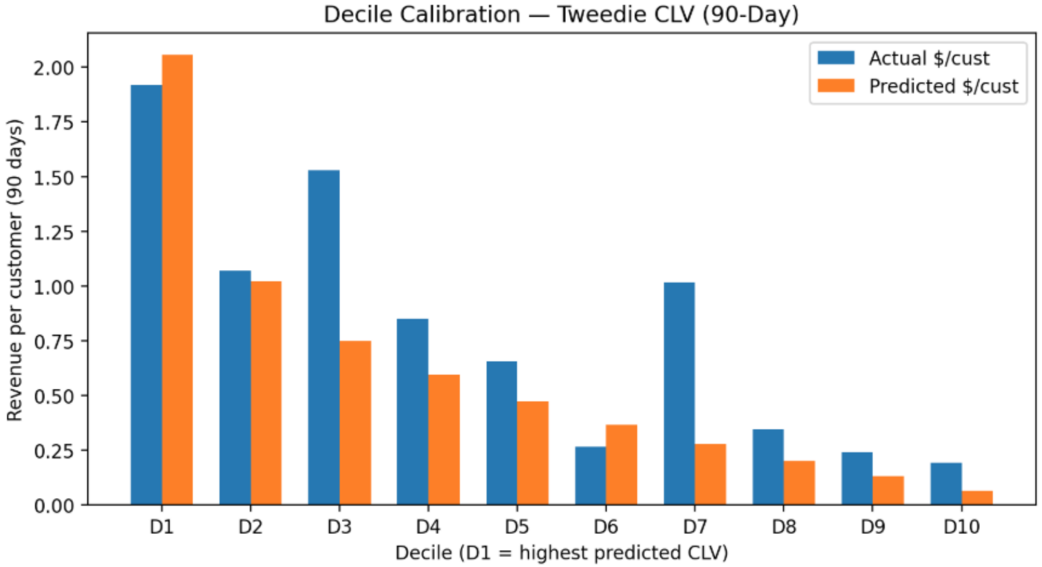
**Ranking / targeting (test set)**

- Avg 90-day revenue per customer: \$0.81
- Top 10% by predicted CLV_90d
  - \$1.92 per customer (≈ 2.4× average)
  - 24% of 90-day revenue vs 10% at random
- Top 20% by predicted CLV_90d
  - 37% of revenue vs 20% at random

**Takeaway**

- Model clearly separates high- vs low-value customers
- Concentrates future revenue in the highest-scored group → ideal for targeting offers, discounts, and premium service



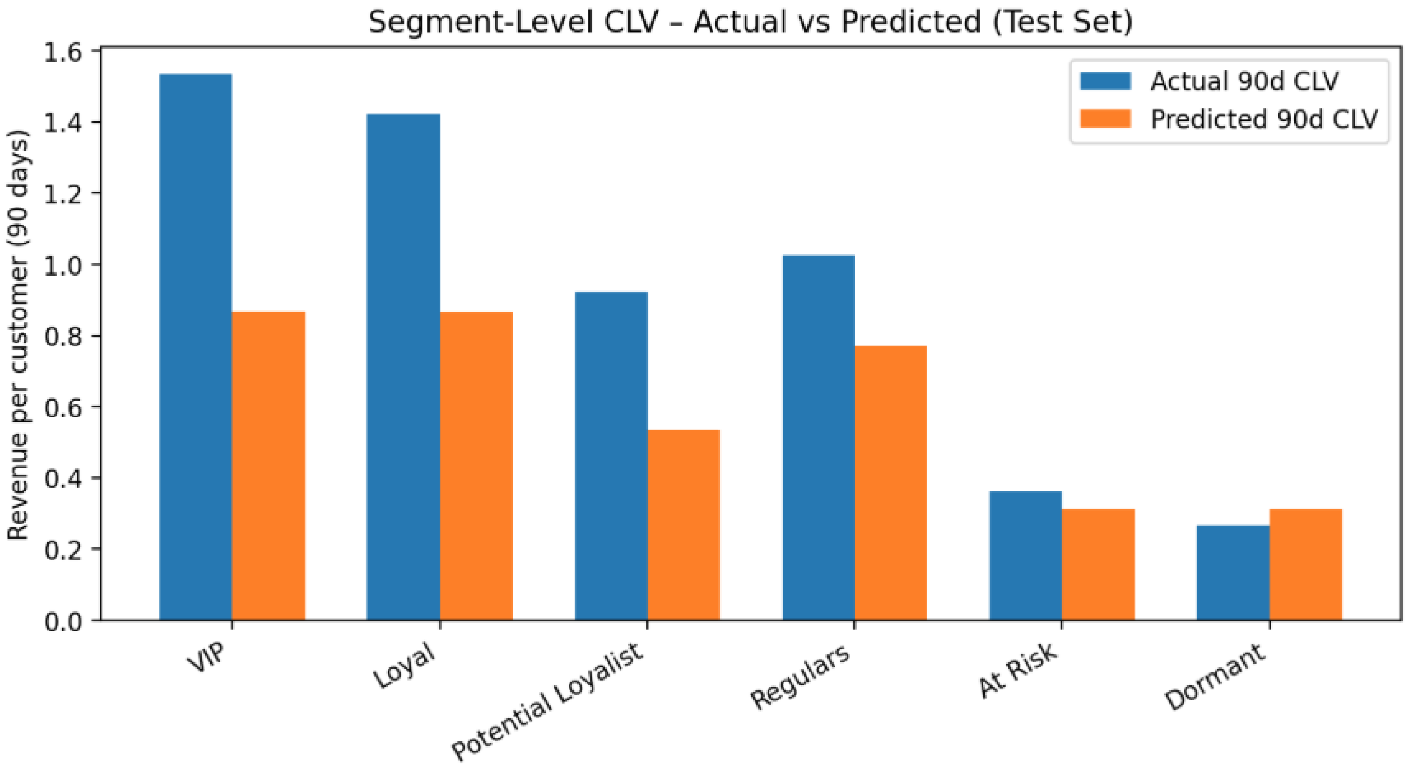
# What the Model Learns (and How It Aligns with RFM Segments)

**Key drivers of predicted CLV**

- **Recency:** more recent purchases → higher CLV_90d
- **Frequency:** more historic orders → higher CLV_90d
- **Monetary / AOV & tenure:** higher spend and longer active relationships → higher CLV_90d

**Alignment with RFM segments**

- **VIP / Loyal:** highest predicted (and actual) CLV_90d → true top-value customers
- **At Risk:** strong past spend but much lower predicted CLV_90d → **value at risk**
- **Dormant:** low historic and predicted CLV_90d → lowest priority for expensive interventions



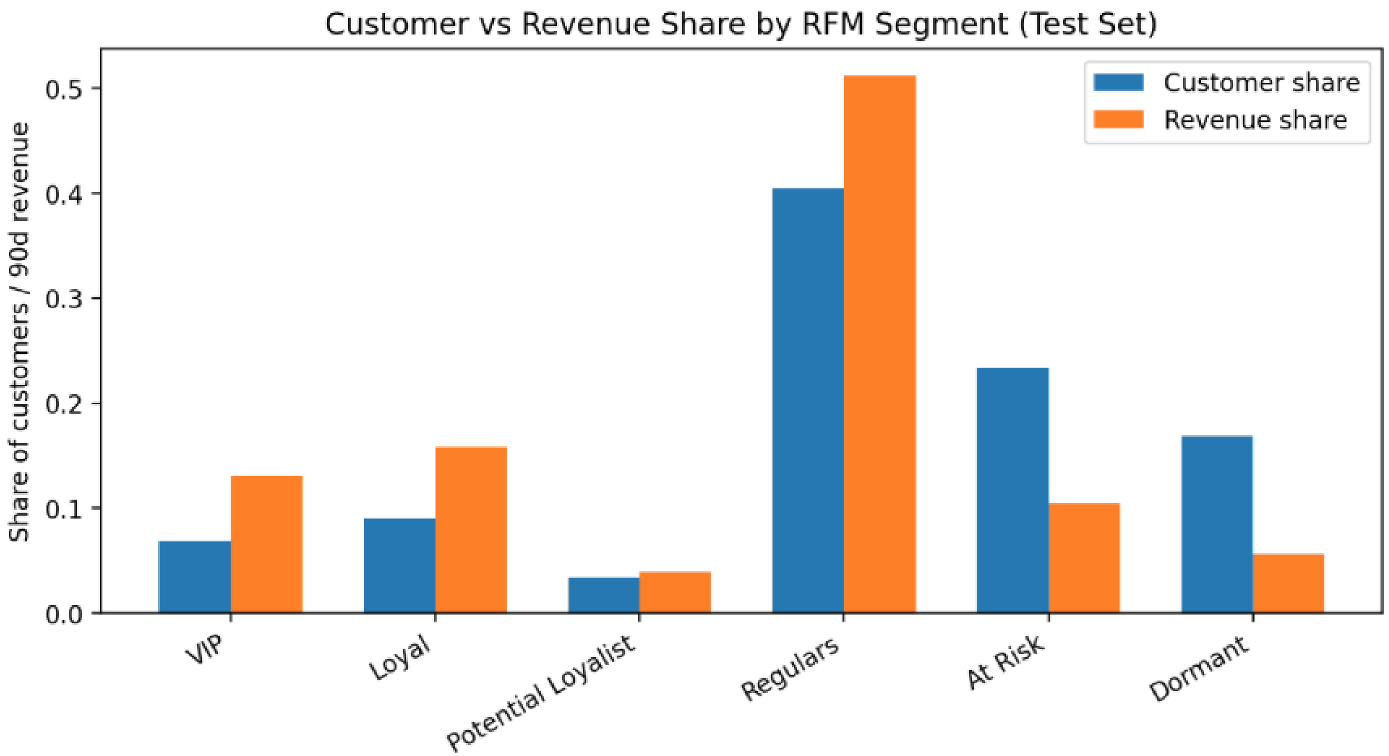
# Strategic Recommendations: Actionable Insights

**Prioritize VIP & Loyal**
~16% of customers, ~29% of 90-day revenue → get best offers, priority service, and early access.

**Strengthen Regulars (core engine)**
~40% of customers, ~51% of revenue → use follow-ups, bundles, and cross-sell to move some into Loyal/VIP.

**Targeted save for At Risk**
23% of customers but only ~10% of revenue → run win-back only for customers with high predicted CLV_90d.

**Low-touch for Dormant**
17% of customers, ~6% of revenue → keep on low-cost, automated comms; avoid heavy discounts.



# Business Impact & Measurement

- VIP & Loyal focus → higher loyalty & revenue
    - KPIs: 90-day repeat rate and CLV_90d uplift vs baseline
- At-Risk save program → prevent revenue loss
    - KPIs: reactivation rate and incremental CLV_90d vs control
- Grow Regulars into Loyal/VIP
    - KPIs: % of Regulars becoming Loyal/VIP and their CLV_90d change
- CLV-based offers → better promo ROI
    - KPIs: promo ROI (profit / promo spend) and avg discount as % of CLV by segment

# Summary & Next Steps

**1. What we learned**

*A small group (VIP + Loyal ~16%) drives ~29% of 90-day revenue.*

*Regulars (~40%) are the core revenue engine (~51% of 90-day revenue).*

*Many customers are At Risk or Dormant with low near-term value.*

**2. What the model adds**

*90-day CLV model ranks customers so the top 10% captures ~24% of future revenue (vs 10% at random).*

*Aligns with RFM segments: highest CLV_90d in VIP/Loyal, lowest in Dormant, and "value at risk" in At Risk.*

**3. What we recommend**

*Double-down on VIP & Loyal, uplift Regulars, and run targeted save programs for high-CLV At Risk.*

*Keep Dormant on low-cost, automated touches and govern offers by CLV to protect margin.*

**4. How we'd measure success**

*VIP/Loyal repeat rate & CLV_90d uplift*

*At Risk reactivation & incremental CLV_90d vs control*

*% of Regulars becoming Loyal/VIP*

*Promo ROI and discount as % of CLV by segment*