**Customer Retention, Cohorts & RFM (K-means) Segment Report**

**Executive summary**

The ecommerce data shows a sharp drop-off after the first purchase: only about **20–30%** of customers buy again in the next month, and most cohorts fall below **15%** by month 3. A small, loyal base persists past 6 months and drives a disproportionate share of revenue. RFM segmentation finds a high-value **VIP/Champion** group (≈10–15% of customers), a large **Potential/Regulars** middle, and a meaningful **At-Risk/Dormant** tail. Focusing on onboarding in month 0–1, converting **Potential** customers to **Loyal**, and structured win-backs for **At-Risk** will lift retention and CLV the most.

**Data & method (in brief)**

* Cleaned transactions, added **TotalPrice = Quantity × UnitPrice**, created **InvoiceMonth**, **CohortMonth** (first purchase month per customer), and **CohortIndex** (months since first purchase).
* Built a monthly **cohort retention matrix**: active customers in month *k* ÷ cohort size in month 0.
* Computed **RFM** (Recency-Frequency-Monetary), scored each metric (quartiles), then clustered customers with **K-means** on scaled RFM to form actionable segments.

**Cohort retention findings**

**Pattern:**

* **Month 0 → 1:** falls to ~**20–30%**.
* **Month 3+:** typically **<15%** kept.
* **Best cohorts:** earlier 2011 cohorts kept better (likely stronger offers/experience).
* **Loyal core:** small set active beyond 6 months; key for long-run revenue.

**Implications:**

* Your most significant revenue leak is **post-first-order**.
* Early lifecycle journeys (first 30–60 days) need targeted nudges to earn a second and third order.

**RFM + K-means segment profiles**

Using RFM (low R = recent), typical clusters appeared:

1. **VIP/Champions** (High F, High M, Low R; ~10–15%)
   * Frequent, high spenders; respond to exclusivity and prompt access.
2. **Loyal** (Medium-High F, Medium M, Low-Medium R)
   * Consistent repeaters; value reliability and recognition.
3. **Regulars/Potential** (Low-Medium F, Medium M, Medium R)
   * Occasional buyers with upside; one more purchase could tip them into Loyal.
4. **At-Risk** (Medium F in past, High R now)
   * Previously engaged, now drifting; prime for win-back.
5. **Dormant/One-and-Done** (F = 1, High R, Low M)
   * Only first buy; requires stronger onboarding to activate a second order.

**Linking cohorts to segments (cohorts → clusters)**

While a crosstab of CohortIndex × Cluster should confirm exact mapping, the observed retention pattern supports this lifecycle linkage:

| **CohortIndex (months since first order)** | **Likely dominant clusters** | **Interpretation & action** |
| --- | --- | --- |
| **0–1 (New)** | Dormant/One-and-Done, Potential | Critical window; convert to 2nd order quickly. |
| **2–3 (Early repeaters)** | Potential → Loyal | Reinforce habit; bundle offers and next-best suggestions. |
| **4–6 (Stabilising)** | Loyal, some At-Risk | Keep cadence; catch early signs of churn (rising Recency). |
| **6+ (Mature)** | VIP/Champions or At-Risk/Dormant | Split: deepen VIP value; re-engage lapsed with win-backs. |

**How to confirm in your data (analyst note):** compute pd. Crosstab (df[['CustomerID']].merge(rfm[['CustomerID', 'Cluster']], 'left'), cohorts[['CustomerID', 'CohortIndex']], ...) to see each cohort’s month cluster distribution.

**Risks & opportunities**

**Risks**

* **High early churn** after the first purchase depresses CLV.
* **Large Potential segment** under-monetised; low frequency traps value.
* **At-Risk accumulation** raises paid reacquisition spend.

**Opportunities**

* **Onboarding and 2nd-order push** can lift the long-term retention curve the most.
* **VIP concentration** enables cost-efficient revenue with tiered rewards.
* **Country/product differences** (if analysed) can inform targeted merchandising.

**Segment-level strategies.**

**1) Engage Regulars/Potential (convert to Loyal)**

* **30-day repeat plan:**
  + **Day 3** “how did it go?” email + usage tips.
  + **Day 7–10** dynamic recommendation (viewed/bought together).
  + **Day 14** limited-time booster (e.g., free shipping) for 2nd order.
* **Merchandising:** starter bundles; subscribe-&-save where fit.
* **Loyalty nudges:** show progress to next tier (“2 orders away from Silver”).
* **KPIs:** repeat rate in 30/60 days, uplift in F, conversion to Loyal.

**2) Win back At-Risk (reduce churn)**

* **Trigger:** Recency above segment-specific threshold (e.g., 1.5× median inter-purchase time).
* **Playbook:**
  + **Email/SMS A/B:** reminder vs. incentive vs. “we miss you” content.
  + **Price fence:** targeted discount only for At-Risk; avoid blanket promos.
  + **On-site:** returning-visitor banner with comeback bundle.
* **KPIs:** reactivation rate, margin after incentive, % returning to Loyal.

**3) Nurture VIP/Champions (protect CLV)**

* **Value exchange:** prompt access, exclusive drops, concierge support.
* **Recognition:** visible tier badges, surprise-and-delight gifts.
* **Referrals:** double-sided rewards; social proof campaigns.
* **Risk control:** monitor for **R rising**—trigger a personal note/offer.
* **KPIs:** VIP churn (<2–3%/qtr), VIP share of revenue, referral K-factor.

**Lifecycle programs by cohort month**

* **Month 0–1 (New):** welcome series, post-purchase tips, **2nd-order incentive**, first-party data capture (preferences).
* **Month 2–3:** routine cadence (weekly/bi-weekly), replenishment reminders, lightweight loyalty rewards.
* **Month 4–6:** differentiation: VIP upgrades vs At-Risk alerts; cross-sell based on affinity.
* **6+ months:** renewal/reactivation pulses; win-back testing (offer vs content); community/UGC for VIPs.

**Measurement & targets**

* **Retention curve:** average cohort retention at month 1, 3, 6. Targets: raise M1 by **+5–8pp**, M3 by **+3–5pp** within 2 quarters.
* **Segment transitions:** % moving Potential→Loyal, At-Risk→Active each month.
* **CLV:** track by segment; aim for **+10–15%** in 6–9 months.
* **Offer efficiency:** reactivation ROI (>3×) and margin impact.
* **VIP health:** revenue concentration, churn, NPS.

**Next analytical steps**

1. **Cross-tab cohorts × clusters** monthly to confirm mapping and spot segment drift.
2. **Propensity models** (2nd purchase, churn) to prioritise offers.
3. **Product-level stickiness**: find items with high repeat lift; feature in onboarding.
4. **Experimentation:** A/B test incentives, cadence, channel mix by segment.
5. **CLV forecasting & budget allocation** to shift spend from reacquisition to retention.

**References to my prior analysis**

The retention drop-off, the existence of a small long-term loyal base, and RFM framework/segment archetypes come from my earlier **Customer Retention & Cohort Analysis** work (cohort heatmaps, RFM segmentation and distributions)