Cohort Analysis + LTV:CAC (Marketing Analytics Case Study)

by Konrad Kuleta

The following analysis uses a synthetic e-commerce dataset to answer:

- **Retention:** What share of customers return in months 1–12 after their first purchase?
- **LTV:** How does cumulative revenue per customer evolve over the first 12 months?
- **Unit economics:** How does **LTV:CAC** compare across acquisition channels?

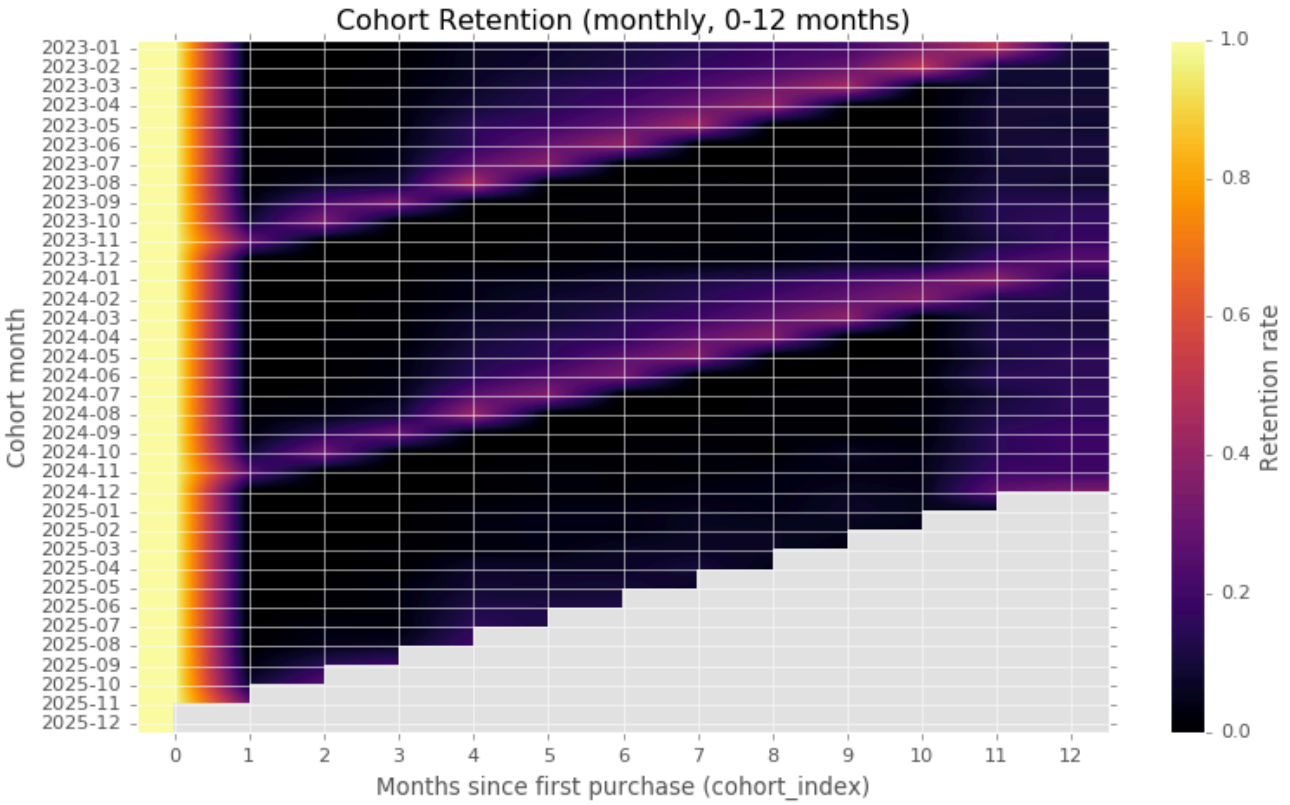
Key assumptions

- **LTV proxy:** LTV is approximated using a constant **45% contribution margin** applied to net revenue.
- **Organic:** Organic has **0 recorded spend** in this dataset, so **CAC is not applicable** for Organic.

Cohort Heatmap

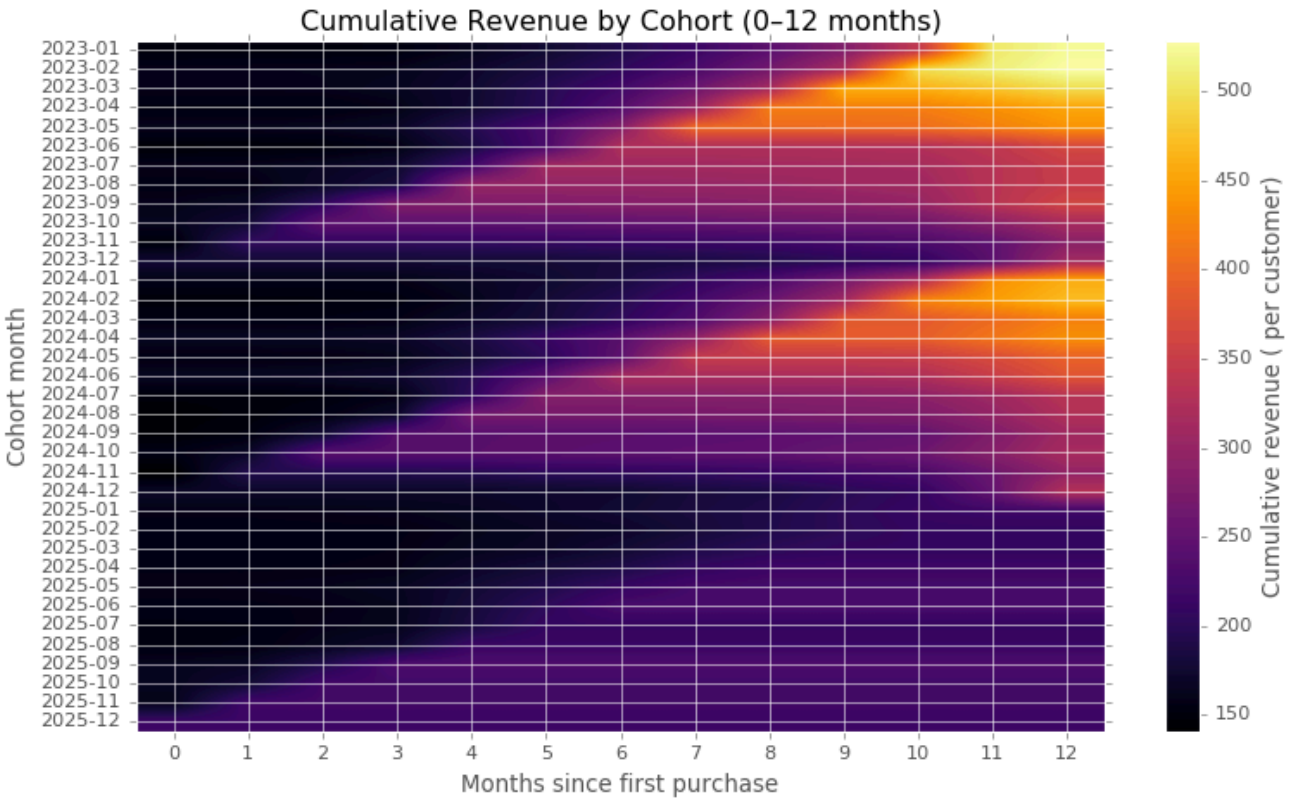
Data is loaded from pre-cleaned CSV files and transformed into cohort features (cohort_month, cohort_index). Definitions:

- **cohort_month**: month of a customer's **first** purchase
- **order_month**: month of a given order
- **cohort_index**: number of months since first purchase (0, 1, 2, ...)



Cumulative revenue per cohort

- Cumulative sum across months → **cumulative revenue**

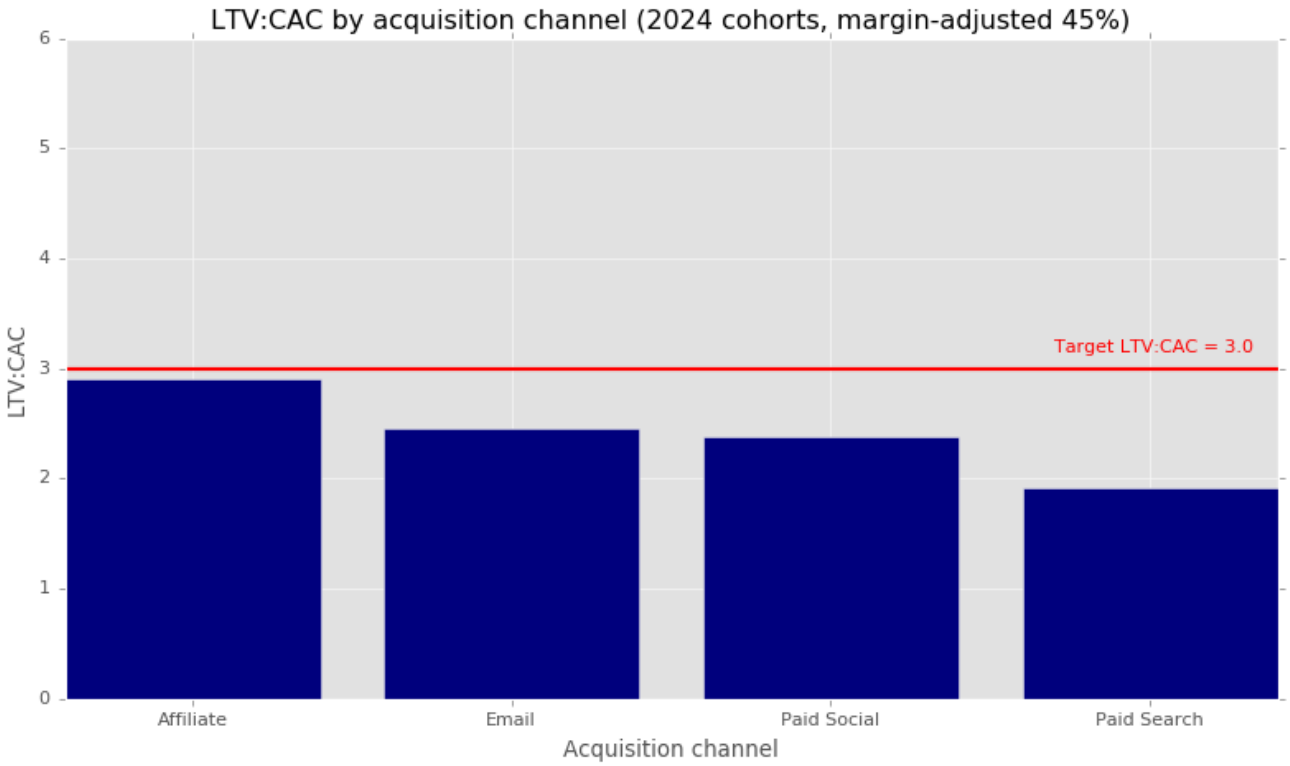


LTV:CAC by acquisition channel

CAC is calculated from **media spend / new customers** (per month & channel). Then, for each channel, 2024 cohorts are summarized and computed:

- **avg CAC** (2024)
- **avg LTV after 12 months** (revenue-based)
- **LTV:CAC** (revenue-based)
- **LTV:CAC** with **margin proxy** (45%)

	channel	avg_CAC_2024	avg_LTV12_2024	LTV:CAC_12m_revenue	LTV:CAC_12m_margin45
0	Affiliate	60.664859	391.367844	6.451311	2.903090
1	Email	68.532846	372.669731	5.437827	2.447022
4	Paid Social	71.923793	381.274949	5.301096	2.385493
3	Paid Search	84.522407	360.202875	4.261626	1.917732
2	Organic	0.000000	371.697546	NaN	NaN



Summary

Key findings:

- **Retention**: Retention drops the most right after the first purchase (months 1–2) and then settles into a lower, relatively stable long tail. In addition, many cohorts show a clear seasonal uplift around December (Q4), which shifts along the “months since first purchase” axis depending on the cohort start month.

- **LTV shape**: Cumulative revenue per customer increases fastest in the first 1–3 months, then grows more gradually — with an additional uplift around Q4/December (seasonality).
- **Unit economics** by channel (12-month, margin-adjusted LTV:CAC): Channel performance is noticeably different. **Affiliate** 2.9 (strongest) > **Email** ~2.45 > **Paid Social** 2.39 > **Paid Search** 1.92 (weakest, below the others) This indicates some channels can tolerate higher CAC due to stronger 12-month revenue per acquired customer, while others struggle to reach the same payback profile.

Recommendations:

- Scale what's closest to target: Prioritize channels with the highest margin-adjusted LTV:CAC (Affiliate / Email / Paid Social) while monitoring diminishing returns.
- Fix the weakest link (Paid Search): Improve unit economics via: a) **CAC reduction**: keyword and query refinement, landing page improvements, ad/creative testing, negative keywords, better intent segmentation. b) **LTV uplift**: bundles, subscriptions, post-purchase cross-sell, lifecycle email flows that target the first 90 days (where retention drops most).

Limitations:

- **Margin proxy**: Contribution margin is assumed constant (45%) to approximate LTV in a margin-adjusted scenario.
- **Organic CAC not observed**: Organic has 0 recorded spend in the dataset, so CAC is not comparable for Organic.
- **Synthetic dataset**: Patterns are simulated and may lack real-world noise and attribution complexity.

Appendix: Sanity checks

```
Orders 2025: 20000
Revenue 2025: 3133059.43
Spend 2025: 380000.02
```