Maré Alta case: data dictionary

Tables and columns of the extract

1. Structure of the extract

The extract has five tables that describe 1,150 points of sale of Maré Alta Beverages over 36 months, from October 2023 to September 2026. All the data is synthetic and was generated for the programme.

The fiscal year runs from October to September, so FY24 covers October 2023 to September 2024, FY25 the following twelve months and FY26 the twelve months up to September 2026. Volumes are expressed in hectolitres (1 HL = 100 litres) and monetary values in Brazilian reais. Months are stored as the first day of the month, in the format `YYYY-MM-DD`.

Table	One row is	Key	Rows (raw / curated)
`pos`	a point of sale	`pos_id`	1,150 / 1,150
`products`	a product line	`product_id`	5 / 5
`sales_monthly`	a point of sale, a month and a product	`pos_id`, `month`, `product_id`	99,897 / 99,324
`trade_spend`	a point of sale, a month and a type of spend	`pos_id`, `month`, `spend_type`	11,328 / 11,328
`churn_scores`	a point of sale scored by the churn model	`pos_id`	1,150 / 1,150

The tables `sales_monthly`, `trade_spend` and `churn_scores` relate to `pos` through `pos_id`, and `sales_monthly` relates to `products` through `product_id`.

The two releases

The extract is distributed in two releases with the same structure. The raw release is the extract as it left the source systems and is the one examined on the first day. The curated release is the version that the data team delivered after the working group reviewed the raw one, and it is used on the second and third days. The data team corrected in it only what could be corrected without a business decision, and a note delivered with it lists what was changed and what was left as it was.

2. Table pos

Column	Type	Description
`pos_id`	text	Identifier of the point of sale.
`region`	text	Commercial region: Southeast, South, Northeast or Central-West.
`state`	text	Brazilian state, two-letter code.
`channel`	text	Type of outlet: Bar, Restaurant, Mini-market, Supermarket or Convenience store. Bars and restaurants form the on-trade, in which the product is consumed on the premises; the other three form the off-trade.
`tier`	text	Potential tier attributed by the sales team: A (highest), B or C. The classification was made in 2023 and has not been reviewed since.
`income_index`	number	Income index of the neighbourhood of the outlet, from 0 to 100, bought from an external provider.

Column	Type	Description
`sales_rep_id`	text	Sales representative responsible for the outlet.
`source_system`	text	System in which the outlet is managed: `ERP_CORE`, the corporate system, or `LEGACY_NE`, the system of the distributor acquired in the Northeast in 2023.
`cooler_programme`	text	Yes when the outlet joined the Cooler Programme, No otherwise.
`programme_start_month`	date	Month in which the cooler was installed. Empty for outlets outside the programme.

3. Table products

Column	Type	Description
`product_id`	text	Identifier of the product line, P01 to P05.
`product_name`	text	Commercial name and pack size.
`family`	text	Core lager, Premium or Non-alcoholic.
`pack_type`	text	Returnable bottle, One-way bottle or Can.
`list_price_brl_per_hl_fy24`	number	List price per hectolitre in FY24, net of taxes. List prices were raised by 4% at the start of FY25 and again at the start of FY26.
`cogs_brl_per_hl_fy24`	number	Cost of goods sold per hectolitre in FY24.

4. Table sales_monthly

Column	Type	Description
`month`	date	First day of the month of the invoices.
`pos_id`	text	Point of sale.
`product_id`	text	Product line.
`volume_hl`	number	Volume invoiced in the month, in hectolitres.
`gross_revenue_brl`	number	Volume multiplied by the list price in force.
`discount_brl`	number	Discounts granted on the invoice.
`net_revenue_brl`	number	Gross revenue minus invoice discounts.
`cogs_brl`	number	Cost of goods sold of the volume invoiced.

Gross margin is net revenue minus cost of goods sold, and it does not include trade spend. A point of sale has no row in a month in which it did not buy a given product.

5. Table trade_spend

Column	Type	Description
`month`	date	First day of the month in which the spend was booked.
`pos_id`	text	Point of sale that received the investment.
`spend_type`	text	Cooler installation (one-off, at the start of the programme in the outlet), Cooler maintenance (monthly), POS activation (promotional actions at the outlet), Merchandising materials or Listing fee (quarterly payments to supermarkets).
`amount_brl`	number	Amount booked.

Invoice discounts are recorded in `sales_monthly` and are not repeated in this table, so the commercial investment of a period is the sum of the two.

6. Table churn_scores

Column	Type	Description
`pos_id`	text	Point of sale scored.
`snapshot_date`	date	Date on which the model was run, 30 June 2026, using data up to that day.
`model_version`	text	Version of the model.
`churn_score`	number	Score between 0 and 1 produced by the model. Higher values indicate that the model considers the outlet more likely to stop buying.
`predicted_churn`	0 or 1	1 when the score is 0.5 or higher, the threshold adopted by the analytics team.
`actual_churn_90d`	0 or 1	1 when the outlet placed no order in the 90 days after the snapshot (July to September 2026).