

Dataset 5 – Customer-Level Loyalty & Behavior Panel

Problem Statement: Understand customer behavior, loyalty, churn risk and potential CLV for Heineken’s consumer base across cities and channels.

Grain (each row represents): One row per customer snapshot (RFM-style panel as of end-2024).

Notes: Contains missing values in some behavioral fields, outliers in spend and CLV, and duplicated customer rows to support advanced EDA and modeling exercises.

Data Dictionary:

Column Name	Type	Example	Description
customer_id	String	CUST_01234	Unique identifier for a loyalty program customer.
gender	Category	Male	Customer gender: Male, Female, Other.
age_group	Category	25-34	Age band: 18-24, 25-34, 35-44, 45-54, 55+.
city	Category	Bengaluru	City where the customer is primarily located.
preferred_channel	Category	On-Trade	Most frequently used purchase channel: On-Trade, Off-Trade, E-commerce. Contains missing values.
signup_date	Date	2021-09-10	Date when the customer joined the loyalty program.
last_order_date	Date	2024-11-20	Date of the customer's most recent recorded

			order.
total_orders_last_12m	Numeric (integer)	10	Number of orders placed in the last 12 months.
total_volume_liters_last_12m	Numeric (float)	55.0	Total volume (in liters) purchased by the customer in the last 12 months.
total_spend_last_12m_inr	Numeric (float)	32000.0	Total spend in INR by the customer in the last 12 months. Includes some very high outliers.
avg_discount_pct	Numeric (float)	0.14	Average discount percentage received across the customer's orders. Contains missing values.
recency_days	Numeric (integer)	25	Days since last order as of the reference date (2024-12-31).
frequency_orders_per_month	Numeric (float)	0.8	Average number of orders per month over the last 12 months.
avg_order_value_inr	Numeric (float)	3200.0	Average order value in INR (total spend divided by number of orders).
loyalty_tier	Category	Gold	Loyalty tier: Bronze, Silver, Gold, Platinum, based on total spend.

churn_flag	Numeric (0/1)	0	1 if the customer is considered churned (recency > 90 days), else 0.
last_campaign_responded	Category	Sports_Promo	Last campaign type the customer responded to: Summer_Fest, Sports_Promo, Digital_Coupon, None. Contains missing values.
clv_12m_estimate_inr	Numeric (float)	45000.0	Estimated 12-month customer lifetime value based on spend and churn status. Includes high-value outliers.