E-COMMERCE & DTC INTELLIGENCE

E-commerce and DTC store intelligence

A continuously tracked universe of ~531,000 Shopify and direct-to-consumer stores: catalog, pricing, paid acquisition and technology stack, at store level, point-in-time.

~531k stores	~41.5M products	~47.9k creatives	API or bulk
Continuously tracked	SKU-level catalog	Meta ad tracking	JSON / CSV / Parquet / S3

PRODUCT

What this is

A store-level view of the direct-to-consumer web. Shopify and DTC storefronts are discovered, then their catalog, pricing, paid-acquisition creatives and technology stack are checked on a recurring schedule and diffed over time. The result is a single research-ready feed per store: profile and lifecycle status, SKU-level catalog with images and prices, Meta ad creatives with landing pages and target countries, detected technology and tracking-pixel stack, and a multi-store operator grouping for entity resolution. See data overview for the broader catalog.

DATASET

Coverage

Numbers refresh continuously. Snapshot below was generated from the live database for this document.

~531,000

Tracked stores

~41.5M

Store products (SKUs)

~4.0M

Product images

~47,900

Tracked ad creatives

~8,200

Multi-store organizations

Store-level

Granularity

Domain-keyed store ID | URL-unique products (no dupes) | Platform, country, currency, category | Openings & closures tracked over time

SCHEMA

Per-store fields

Store profile

One row per store, domain-unique.

FIELD	DESCRIPTION
`store_id`	Stable unique store ID
`domain`	Canonical storefront domain (unique key)
`store_name`	Store / brand name
`platform`	E.g. `shopify`, `woocommerce`, `custom`
`status`	`active` / `inactive` / `closed`
`country` / `currency`	Detected market and store currency
`category`	Merchandise category classification
`product_count`	Live catalog size
`organization_id`	Multi-store operator grouping
`first_seen` / `last_checked`	Discovery and last-refresh timestamps

Products & pricing

One row per product, store x product-unique.

FIELD	DESCRIPTION
`store_id`	Joins to store profile
`product_id` / `handle`	Store product identifier
`title`	Product title
`sku` / `variant`	SKU and variant where present
`price` / `compare_at`	Current and reference price
`currency`	Listed currency
`availability`	In stock / out of stock
`image_url`	Primary product image
`product_category`	Category / product type
`first_seen` / `last_seen`	Catalog lifecycle timestamps

PAID ACQUISITION

Ad creatives and landing pages

~47,900 Meta ad creatives tracked and linked back to the store that runs them, with landing pages and target countries. This is the substrate for "is this brand scaling paid acquisition, and where?" and for reading ad cadence as a demand proxy on DTC brands not visible in filings.

FIELD	DESCRIPTION
`store_id`	Joins to store profile
`creative_id`	Unique ad-creative identifier
`network`	Ad network (currently `meta`)
`creative_url` / `media`	Creative asset URL
`landing_page`	Destination URL on the store
`target_countries`	Declared audience geographies
`first_seen` / `last_seen`	Creative run window (point-in-time)

TECHNOLOGY

Stack and tracking pixels

Detected technology and tracking-pixel stack per store, plus the multi-store operator grouping (~8,200 organizations) that links storefronts run by the same operator for entity resolution.

Technology stack

One row per store x detected technology.

FIELD	DESCRIPTION
`store_id`	Joins to store profile
`technology`	Detected app, theme or service
`category`	Tech category (analytics, payments, …)
`pixel_id`	Tracking-pixel identifier where present
`detected_at`	First-detected timestamp

Organizations

One row per multi-store operator.

FIELD	DESCRIPTION
`organization_id`	Operator grouping key
`store_count`	Stores linked to the operator
`signals`	Shared pixels / payment / DNS signals used to group

METHODOLOGY & DELIVERY

How it's built and shipped

- **First-party collection**: storefronts, catalogs, ad creatives and technology stacks are discovered and parsed by infrastructure we operate, not resold third-party feeds
- **Domain-keyed stores**: every store has a canonical domain; products, creatives and tech join via `store_id`, so the universe is clean to merge internally
- **Store lifecycle tracked**: openings, closures and activity are timestamped, so a closure is a flagged, point-in-time signal
- **SKU-level history**: catalog and pricing carry `first_seen` / `last_seen` per product; price and assortment change is reconstructable over time
- **Multi-store operator grouping**: shared pixels, payment and DNS signals resolve storefronts run by the same operator
- **No consumer PII**: store, catalog, creative and technology data only; no shopper-level data

Delivery

- **Bulk extract**: CSV / JSON / Parquet via S3 or SFTP, full history at the cadence the client wants
- **Sample delivery**: a scoped sample (single category, or single operator universe) delivered for validation before any commercial conversation
- **Production API**: per-client keys, store / product / creative endpoints, on engagement
- **Incremental delivery**: daily or weekly deltas on the same channel; backfill on request
- **Universe mapping**: stores and operators mapped to the client's watchlist or ticker universe on delivery

USE CASES

What a research desk does with this

- DEMAND **Private-brand and category read**: store openings and closures, SKU and price velocity, and Meta ad cadence as demand on DTC brands not visible in filings
- PRICING **Pricing power and assortment**: SKU-level pricing and catalog change across a ~531,000-store universe for category share and margin work
- ACQUISITION **Paid-acquisition intensity**: ad-creative counts, landing pages and target countries as a read on whether a brand is scaling spend, and into which markets
- CATEGORY **Category base rates**: store formation, closure and pricing across categories for cross-sectional demand work
- ENTITY **Operator resolution**: multi-store grouping to size the true footprint of an operator running many storefronts
- MAPPING **Universe ready to join**: domain-keyed, deduped, mergeable; delivered point-in-time to the client universe

ACCESS

Getting the data

Delivered by API or bulk export (CSV, JSON, Parquet, S3), with history and refresh cadence agreed per engagement.