

# Data Dictionary — Module 3: CRO Terminal

Dokumen ini menjelaskan seluruh kolom data yang digunakan di Module 3: Data Steering (CRO Terminal). Platform: Shopify · Data Sources: GA4 + Cloudflare + Shopify API

## 1. `generate_traffic_data()` — Raw & Filtered Traffic

Data harian yang menggabungkan informasi dari GA4, Cloudflare, dan Shopify. Fungsi utamanya adalah membedakan traffic manusia asli dari bot.

Kolom	Tipe	Range (Dummy)	Penjelasan	Sumber di Production
<code>date</code>	datetime	30 hari terakhir	Tanggal data	—
<code>total_sessions</code>	int	3,500 – 8,125	Total session mentah termasuk bot. Weekend di-boost +25%.	GA4: <code>sessions</code>
<code>bot_sessions</code>	int	15–30% dari total	Jumlah session yang teridentifikasi sebagai bot.	Cloudflare Bot Score < 30
<code>human_sessions</code>	int	total - bot	Session manusia asli setelah filtering. Ini yang disebut " <b>True Traffic</b> ".	Cloudflare + GA4
<code>bot_sub_1s</code>	int	40–60% dari bot	Bot yang bounce dalam waktu kurang dari 1 detik. Ciri khas automated scraper.	Cloudflare / Server Logs
<code>bot_known_ips</code>	int	20–35% dari bot	Bot dari IP address yang sudah masuk database bot (Googlebot, Bingbot, SEMrush, dll).	Cloudflare Known Bots DB
<code>bot_no_js</code>	int	sisa bot	Bot yang tidak mengeksekusi JavaScript. Browser asli pasti jalankan JS, bot sering skip.	Cloudflare JS Challenge
<code>bot_pct</code>	float	15.0 – 30.0	Persentase bot dari total traffic. Metric "seberapa kotor" traffic kita.	Dihitung
<code>mobile_sessions</code>	int	62–72% dari human	Session dari perangkat mobile (smartphone).	GA4: <code>deviceCategory</code>
<code>desktop_sessions</code>	int	22–30% dari human	Session dari perangkat desktop/laptop.	GA4: <code>deviceCategory</code>

Kolom	Tipe	Range (Dummy)	Penjelasan	Sumber di Production
tablet_sessions	int	sisa	Session dari tablet (iPad, dll).	GA4: deviceCategory
lcp_mobile	float	2.0 – 4.5s (spike: 6.5s)	<b>Largest Contentful Paint</b> di mobile. Seberapa cepat elemen terbesar di halaman tampil. Target: <2.5s.	Google PageSpeed API / CrUX
lcp_desktop	float	1.2 – 2.8s	LCP di desktop. Biasanya lebih cepat dari mobile.	Google PageSpeed API / CrUX
src_google_ads	int	20–28% dari human	Session dari Google Ads (Search + Shopping).	GA4: sessionSource / sessionMedium
src_meta_ads	int	18–25% dari human	Session dari Meta Ads (Facebook + Instagram Ads).	GA4: UTM parameters
src_tiktok_ads	int	8–15% dari human	Session dari TikTok Ads.	GA4: UTM parameters
src_organic_search	int	12–18% dari human	Session dari pencarian organik Google/Bing (bukan iklan).	GA4: sessionMedium = organic
src_direct	int	10–15% dari human	Session dari akses langsung (ketik URL / bookmark).	GA4: sessionMedium = (none)
src_email	int	5–10% dari human	Session dari email marketing (Klaviyo, Mailchimp, dll).	GA4: UTM medium = email
src_social_organic	int	sisa (~8%)	Session dari social media organik (bukan ads).	GA4: sessionMedium = social

## 2. generate\_funnel\_data() — Shopify Conversion Funnel

Data harian perjalanan pelanggan dari landing sampai beli. Setiap step menunjukkan berapa orang yang "lolos" ke tahap berikutnya.

Kolom	Tipe	Range (Dummy)	Penjelasan	Sumber di Production
-------	------	---------------	------------	----------------------

Kolom	Tipe	Range (Dummy)	Penjelasan	Sumber di Production
date	datetime	30 hari terakhir	Tanggal data	—
landing_page	int	= human_sessions	Jumlah orang yang masuk ke website (= True Traffic). Step 1 dari funnel.	GA4: <code>page_view</code> event
product_page	int	55–70% dari landing	Jumlah orang yang melihat halaman produk. Drop-off di sini = orang tidak tertarik browsing.	GA4: <code>view_item</code> event
add_to_cart	int	25–40% dari product	Jumlah orang yang menambahkan produk ke keranjang. Drop-off = harga tidak cocok atau produk kurang meyakinkan.	Shopify: <code>add_to_cart</code> event
checkout	int	50–70% dari cart	Jumlah orang yang memulai proses checkout. Drop-off = shipping cost surprise atau harus buat akun.	Shopify: <code>begin_checkout</code> event
purchase	int	45–65% dari checkout	Jumlah orang yang menyelesaikan pembelian. Drop-off = payment gagal atau berubah pikiran.	Shopify: <code>purchase</code> event / Orders API
bounce_rate	float	~30–45%	Persentase pengunjung yang langsung pergi tanpa lihat halaman lain. Rumus: $(\text{landing} - \text{product}) / \text{landing} \times 100$ . Target: <40%.	Dihitung dari GA4
cart_abandonment	float	~50–75%	Persentase orang yang sudah add to cart tapi tidak jadi beli. Rumus: $(\text{cart} - \text{purchase}) / \text{cart} \times 100$ . Target: <70%.	Dihitung dari Shopify

Kolom	Tipe	Range (Dummy)	Penjelasan	Sumber di Production
true_cvr	float	~2–5%	<b>Conversion rate sesungguhnya.</b> Rumus: $\frac{\text{purchase}}{\text{human_sessions}} \times 100.$ Pakai True Traffic, bukan raw traffic. Target: >2.5%.	Dihitung
aov	float	\$45 – \$95	<b>Average Order Value.</b> Rata-rata nilai per transaksi.	Shopify: <code>order.total_price</code>
revenue	float	purchase × aov	Total revenue hari itu.	Shopify Orders API
lcp_mobile	float	2.0 – 4.5s	Page load speed mobile (sama dengan traffic data). Penting karena +1s = -20% CVR.	Google PageSpeed API
lcp_desktop	float	1.2 – 2.8s	Page load speed desktop.	Google PageSpeed API
human_sessions	int	= landing_page	Referensi balik ke True Traffic.	—

### 3. `generate_funnel_by_device()` — Funnel per Device

Breakdown funnel yang sama tapi dipecah per tipe perangkat. Penting karena Mobile dan Desktop punya perilaku sangat berbeda.

Kolom	Tipe	Penjelasan
date	datetime	Tanggal data
device	str	Tipe perangkat: <code>Mobile</code> , <code>Desktop</code> , atau <code>Tablet</code>
sessions	int	Jumlah session untuk device ini
landing_page	int	= sessions (entry point)
product_page	int	Mobile: 45–60%, Desktop: 60–78% dari landing. Desktop lebih tinggi karena lebih nyaman browsing.
add_to_cart	int	Mobile: 20–32%, Desktop: 30–45%. Desktop lebih tinggi karena bisa buka banyak tab untuk compare.
checkout	int	Mobile: 45–60%, Desktop: 55–72%. Desktop lebih tinggi karena form checkout lebih mudah di keyboard.
purchase	int	Mobile: 35–55%, Desktop: 50–70%. Desktop jauh lebih tinggi — ini insight kunci untuk klien.

Kolom	Tipe	Penjelasan
bounce_rate	float	Mobile: 38–52%, Desktop: 25–38%. Mobile bounce lebih tinggi karena loading lambat + layar kecil.
cart_abandonment	float	Persentase cart yang tidak jadi purchase. Mobile biasanya lebih tinggi.
cvr	float	Conversion rate per device. $\text{purchase} / \text{sessions} \times 100$ . Biasanya Desktop 2x lipat Mobile.

**Insight kunci untuk klien:** Kalau CVR Mobile jauh di bawah Desktop, artinya ada masalah UX/kecepatan di mobile site.

---

## 4. `generate_funnel_by_source()` — Funnel per Traffic Source

Breakdown funnel per asal traffic. Setiap source punya "kualitas" pengunjung yang berbeda.

Kolom	Tipe	Penjelasan
date	datetime	Tanggal data
source	str	Asal traffic: Google Ads, Meta Ads, TikTok Ads, Organic Search, Direct, Email, Social Organic
sessions	int	Jumlah session dari source ini
landing_page	int	= sessions
product_page	int	Bervariasi per source. Organic Search paling tinggi (65–80%), TikTok Ads paling rendah (35–50%).
add_to_cart	int	Organic Search dan Direct paling tinggi (niat beli sudah ada). TikTok dan Social paling rendah (iseng klik).
checkout	int	Pola sama — traffic high-intent convert lebih baik.
purchase	int	Organic Search dan Direct jadi juara. Ini bukti bahwa traffic organik punya ROI tertinggi.
cart_abandonment	float	Paid traffic biasanya cart abandonment lebih tinggi — "window shopper" dari iklan.
cvr	float	Conversion rate per source. Organic Search bisa 3–5%, TikTok Ads bisa cuma 0.5–1.5%.

### Karakteristik per source:

Source	Tipe Pengunjung	CVR Range	Kenapa
Organic Search	Cari produk spesifik	3–5%	Niat beli sudah tinggi
Direct	Pelanggan lama / referral	3–5%	Sudah kenal brand
Email	Subscriber yang engaged	2.5–4%	Sudah punya hubungan

Source	Tipe Pengunjung	CVR Range	Kenapa
Google Ads	Cari solusi/produk	2–3.5%	Intent medium-high
Meta Ads	Scroll lalu tertarik	1–2.5%	Impulse, butuh nurturing
Social Organic	Lihat konten lalu penasaran	0.8–2%	Eksplorasi, belum ready beli
TikTok Ads	Nonton video lalu klik	0.5–1.5%	Paling impulsif, "just checking"

## 5. `generate_page_speed_data()` — Core Web Vitals

Data performa website harian. Sangat penting karena kecepatan langsung mempengaruhi conversion.

Kolom	Tipe	Range (Dummy)	Penjelasan	Target
<code>date</code>	datetime	30 hari terakhir	Tanggal data	—
<code>lcp_mobile</code>	float	2.2 – 4.2s (spike: 7.0s)	<b>Largest Contentful Paint (Mobile).</b> Waktu sampai elemen visual terbesar di halaman selesai dirender. Ini metric kecepatan #1.	< 2.5s 
<code>lcp_desktop</code>	float	1.0 – 2.5s	LCP untuk Desktop. Biasanya lebih cepat.	< 2.5s 
<code>fid_mobile</code>	float	80 – 250ms (spike: 500ms)	<b>First Input Delay (Mobile).</b> Waktu dari pertama kali user tap/klik sampai browser merespons. Mengukur interaktivitas.	< 100ms 
<code>fid_desktop</code>	float	30 – 120ms	FID untuk Desktop.	< 100ms 
<code>cls_mobile</code>	float	0.05 – 0.25	<b>Cumulative Layout Shift (Mobile).</b> Mengukur seberapa banyak elemen halaman "bergeser" saat loading. Angka tinggi = tombol bergeser saat mau diklik.	< 0.1 
<code>cls_desktop</code>	float	0.02 – 0.12	CLS untuk Desktop.	< 0.1 

### Benchmark Core Web Vitals (Google):

Metric	Good 	Needs Improvement 	Poor 
LCP	< 2.5s	2.5s – 4.0s	> 4.0s
FID	< 100ms	100ms – 300ms	> 300ms
CLS	< 0.1	0.1 – 0.25	> 0.25

**Kenapa ada speed spike?** Di dummy data, ada 7% kemungkinan mobile LCP naik ke 4.5–7.0s. Ini mensimulasikan kejadian nyata seperti: install Shopify app baru yang berat, update theme, third-party script yang lambat, atau CDN bermasalah.

---

## Hubungan Antar Data

```

generate_traffic_data()
  |
  +-- human_sessions --> generate_funnel_data()
    |
    +-- landing_page = human_sessions
    +-- product_page → add_to_cart → checkout →
purchase
  |
  +-- mobile/desktop/tablet_sessions --> generate_funnel_by_device()
    |
    +-- Funnel per device type
  |
  +-- src_google_ads ... src_social_organic --> generate_funnel_by_source()
    |
    +-- Funnel per traffic
source
  |
  +-- lcp_mobile/desktop --> generate_page_speed_data()
    |
    +-- + FID + CLS (Core Web Vitals lengkap)
  
```

---

## Catatan untuk Integrasi Production

Dummy Function	Production Data Source	API / Method
generate_traffic_data()	GA4 + Cloudflare	GA4 Data API v1 + Cloudflare Analytics API
generate_funnel_data()	GA4 + Shopify	GA4 Funnel Report + Shopify Orders API
generate_funnel_by_device()	GA4	GA4 Data API dengan dimension <code>deviceCategory</code>
generate_funnel_by_source()	GA4	GA4 Data API dengan dimension <code>sessionSource</code>
generate_page_speed_data()	CrUX / PageSpeed API	Chrome UX Report API atau PageSpeed Insights API