

Datové zdroje pro forecasting

Odkud brát data, jaké mají limity a jak je kombinovat.

PŘEHLED ZDROJŮ

Kde brát data pro forecast

ZDROJ	CO ZÍSKÁTE	LIMITY
GA4	Sessions, CR, Revenue, AOV (historické)	Sampling u velkých objemů; Consent Mode modeluje odhady; revenue ≠ admin (vratky, storna)
BigQuery (GA4)	Nesampulovaná data, vlastní SQL	Vyžaduje propojení GA4 → BQ
Google Ads IS	Podíl zobrazení → potenciál sessions	Pouze Search; zbylé % bude dražší
SimilarWeb	Traffic konkurence, kanálový mix, market sizing	Malé weby (<50k/měs.) nespolehlivé
Google Trends	Sezónní koeficienty (relativní index)	Žádná absolutní čísla; virální buzz zkresluje
Google Market Finder	Objem vyhledávání, CPC benchmarky pro nové trhy	Na úrovni kategorií; CPC orientační
Heureka	Hodnocení, počet recenzí (proxy pro objem)	Nepřímý indikátor; vliv na CR ze srovnávačů
Admin / CRM	Reálné tržby, marže, korekční koeficient	Nutná spolupráce s financemi
Reklamní systémy	CPS, rozpočty, historická data kampaní	Každý systém má jiný atribuční model

SIMILARWEB — PŘESNOST ODHADŮ

Jak moc můžete věřit odhadům

METRIKA	PŘESNOST	POZNÁMKA
Roční traffic	±5 %	Nejspolehlivější
Měsíční traffic	±10-15 %	Sezónní špičky nepřesné
Sezónní peaks	±10-20 %	Podhodnocuje náhlé peaky (Black Friday)
Traffic split (kanály)	±15-20 %	Orientační, ne pro přesné plánování
Malé weby (<50k/měs.)	Nespolehlivé	Nedostatečný vzorek

IMPRESSION SHARE — VÝPOČET POTENCIÁLU

Kolik sessions vám uniká

Potenciální maximum = Aktuální sessions / Impression Share
 Příklad: 10 000 sessions při 70 % IS → max ~14 300 sessions (+43 %)

IS	INTERPRETACE	DOPORUČENÍ
>90 %	Dominujete aukci	Růst = nová poptávka nebo nová klíčová slova
60-90 %	Prostor na růst	Zvýšení rozpočtu/bidů
<60 %	Výrazně omezení	Velký prostor, ale CPS roste s agresivitou

BIGQUERY SQL

Export per kanál a měsíc

```
SELECT
  collected_traffic_source.manual_medium AS medium,
  collected_traffic_source.manual_source AS source,
  FORMAT_DATE('%Y-%m', PARSE_DATE('%Y%m%d', event_date)) AS month,
  COUNT(DISTINCT CONCAT(user_pseudo_id,
    CAST((SELECT value.int_value FROM UNNEST(event_params)
      WHERE key = 'ga_session_id') AS STRING))) AS sessions,
  COUNTIF(event_name = 'purchase') AS transactions,
  SUM(IF(event_name = 'purchase', ecommerce.purchase_revenue, 0)) AS revenue
FROM `project.dataset.events_*`
WHERE _TABLE_SUFFIX BETWEEN '20250101' AND '20251231'
GROUP BY 1, 2, 3
ORDER BY 3, 4 DESC
```

KOREKČNÍ KOEFICIENT (GA4 VS. ADMIN)

Srovnání dat s realitou

Korekční koeficient = Admin revenue / GA4 revenue (typicky 0,85-0,95)
Realistická revenue = Forecastovaná revenue × korekční koeficient

CO ODKUD — RYCHLÝ PŘEHLED

Jakou metriku kde hledat

METRIKA	PRIMÁRNÍ ZDROJ	SEKUNDÁRNÍ ZDROJ
Sessions (historické)	GA4	—
Sessions (potenciál)	Google Ads IS	SimilarWeb
CPS	Reklamní systémy	Google Market Finder
CR	GA4	Admin (korekce)
AOV	GA4	Admin (korekce)
Sezónnost	GA4 (vlastní data)	Google Trends
Konkurence	SimilarWeb	Heureka
Marže	Finance / Admin	—
Market sizing	SimilarWeb	Google Market Finder

GA4 — LIMITY DAT A ZNÁMÉ PROBLÉMY

Na co si dát pozor

Sampling v Explore

Explore reporty samplují nad **10M eventů na dotaz** (free GA4 property). Poznáte podle ikony štítu se žlutým trojúhelníkem. Standardní reporty nesamplují — používejte je pro jednoduché metriky. Pro nesampulovaná detailní data potřebujete BigQuery export.

Export limity v Explore

Explore umožňuje export maximálně **10M buněk** do CSV. Data v Explore jsou dostupná maximálně **14 měsíců** zpětně. Pro YoY srovnání (potřebujete 24+ měsíců) musíte použít BigQuery nebo pravidelný měsíční export do spreadsheetu.

Thresholding (skrývání malých segmentů)

GA4 skrývá data pod cca **40–50 záznamů** v řádku. Souvisí s nastavením Reporting Identity (Observed/Blended) a použitím demografických dimenzí. Řešení: přepněte na „Device-based“ reporting identity, nebo odstraňte demografické dimenze z reportu.

BigQuery vs. UI — klíčové rozdíly

- BQ **neobsahuje modelovaná data** z Consent Mode (tj. BQ revenue < UI revenue).
- BQ batch export trvá až **72 hodin** (ne real-time; streaming export je placený).
- Session rekonstrukce v SQL je netriviální — `ga_session_id` + `user_pseudo_id` + session timeout logika.
- BQ je vždy nesampulovaný — pro přesné forecasty je referenční zdroj.

Attribution scope — pozor na session-scoped dimenze

Session-scoped dimenze (Session source/medium, Session campaign) **VŽDY** používají last non-direct click atribuci, bez ohledu na nastavení Attribution Model v GA4 Admin. To znamená, že assisting kanály (display, social, video) budou systematicky podhodnoceny. Pro kompletní pohled doplňte data z reklamních systémů.