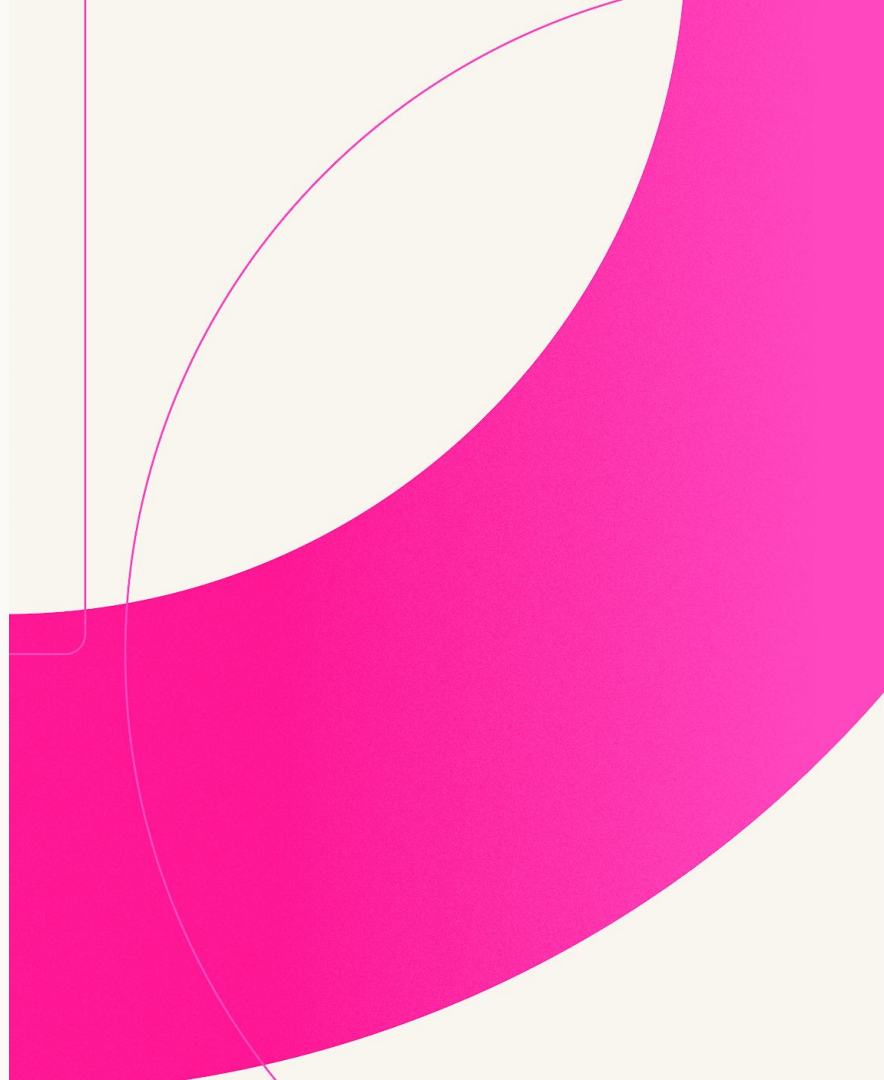Webinar

Launching Geolift: A Game-Changer for Modern Marketers

eppo

Agenda

1. Learning Culture and Searching for Incrementality

2. Fireside chat with Bryce Casavant, Senior Data Scientist at Whoop

3. Eppo Geolift Launch

4. Q&A

Your webinar crew

Elizabeth George
Head of Marketing

eppo

Greg Dale
Marketing Experimentation Lead

eppo

Bryce Casavant
Senior Data Scientist

WHOOP®

eppo

VOLVO
coinbase
perplexity
DRAFTKINGS
descript
zalando
Delivery Hero
The Farmer's Dog
Handshake
PATREON
Airtable

STRAVA
zapier
twitch
ShareChat
Rakuten
cameo

airbnb
Uber
STITCH FIX
Linked in
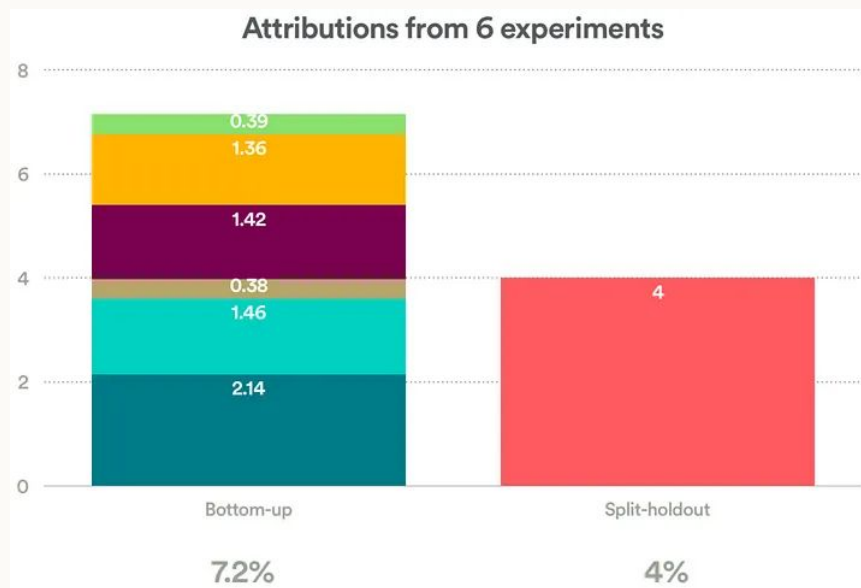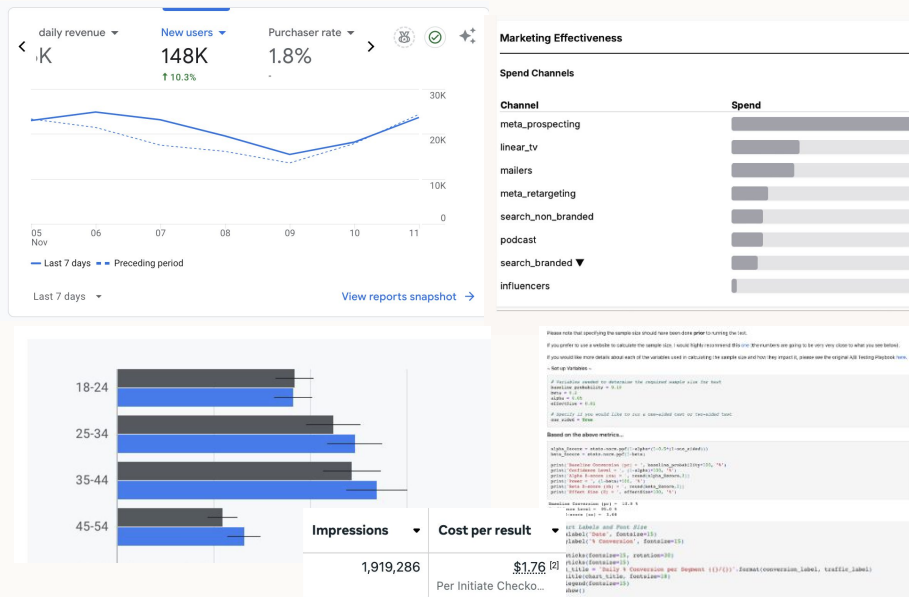ClickUp
gojek
GOLD BELLY
miro
GoodRx
ID.me

Test and learn across **Marketing, Product, AI**

Real Business Metrics

Highly **Trusted Statistics**
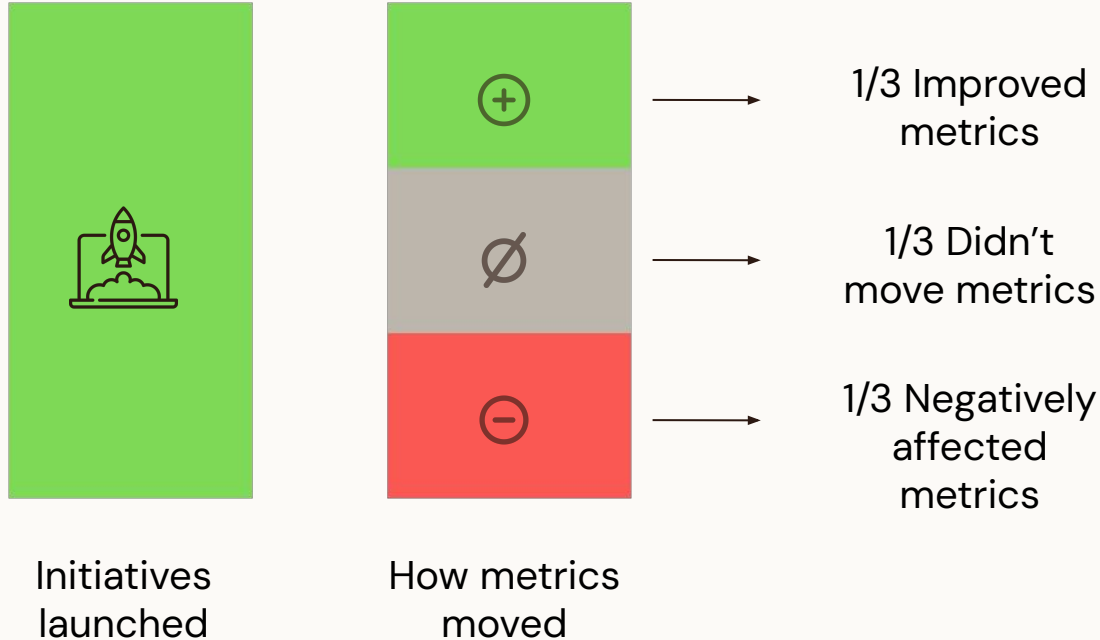
Learning Culture and Searching for Incrementality

So much data, yet it never adds up

Attributions from 6 experiments

Attribution, MMM, surveys, in-platform data, custom analyses...

But where's my revenue lift?

You never know what initiatives will drive impact across a business

Initiatives launched

How metrics moved

1/3 Improved metrics

1/3 Didn't move metrics

1/3 Negatively affected metrics

You never know what initiatives will drive impact across a business

- What gets celebrated?
- What gets more investment?
- What institutional knowledge is built?

Experimentation means businesses celebrate, repeat, remember **the right actions**

Different experimentation methods drive marketing incrementality studies across channels

Web	Email, SMS, Push	CRM Advertising	Offline Media	Digital Media
Content, UX, LPs	Messaging, Timing, Segments	Direct Mail, Custom Audiences, Retargeting	Regional TV, Radio, Out-Of-Home	YouTube, branded search, TikTok, etc.

A/B Tests

Geolift Tests

Within advertising, two key questions we ask: Where should I advertise and how much should I spend?

Sit-back Inventory

Branded Search

New Channels

Let's say I want to test TikTok

- **My design:** I launch in **Chicago** and decide I'll "match markets" and pick **Milwaukee** to compare results to

- **This works**! US Midwest, similar climate, culture.... not bad

- **But**: Chicago is *much* larger, with a higher average household income, cost of living, larger cultural reach...

- The comparison might roughly work, but when you might be making $1m++ decisions, **"rough" is not sufficient**! We need **trust.**

Geolift is a trusted method to identify differences in geographic behavior

Bayesian Synthetic Control Modeling

So, TikTok, in Chicago:

- You can't create a "**parallel universe Chicago**" where the ad didn't run.
- So we create a "**synthetic Chicago**" using data from <u>multiple similar</u> cities without the ad
- Compare sales in **real Chicago** to "**synthetic Chicago**" to see the ad's effect.
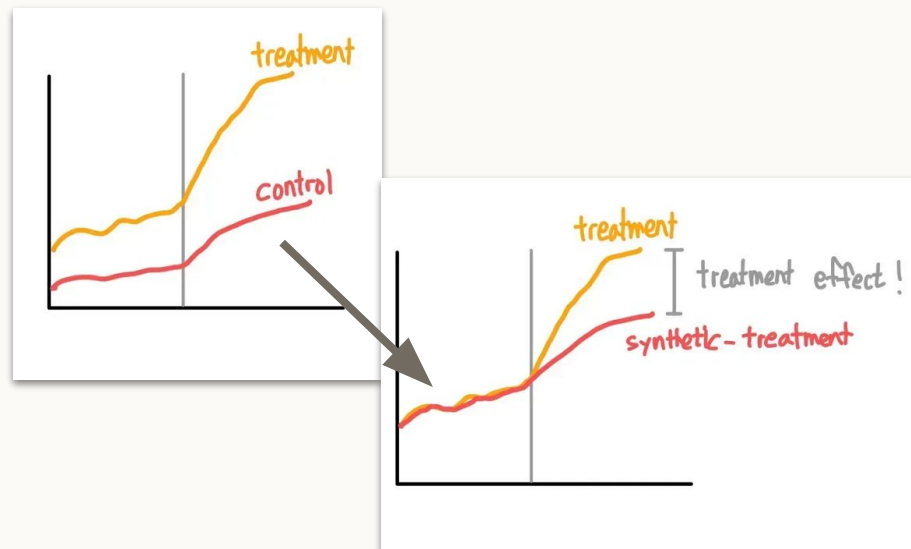
Trust means selecting the right regions for comparison

Identifying similar regions takes rigorous statistics but also an understanding of gotchas

- Weather patterns and events
- Idiosyncratic events (Super Bowl)
- Launches
- Regulatory changes
- Other tests
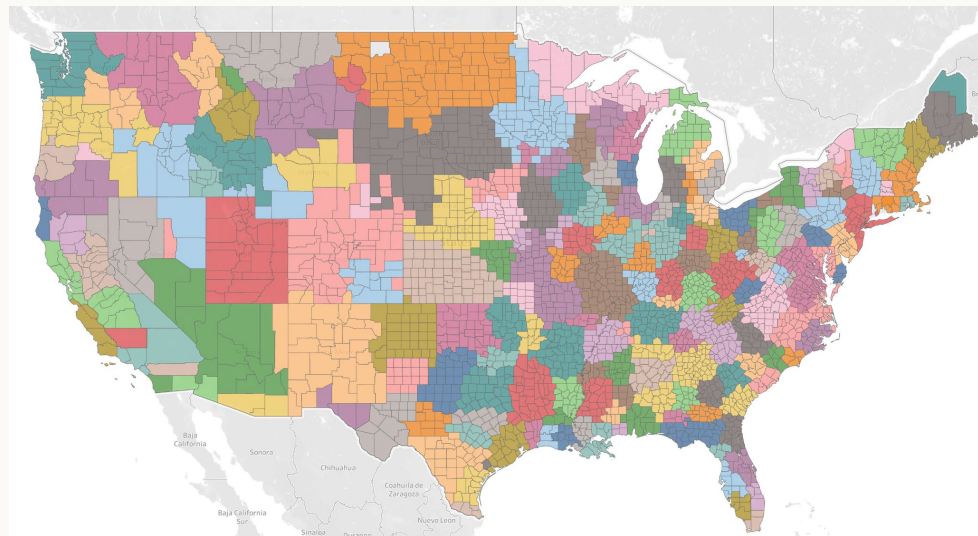
Trust means business metrics, not vendor events

Channel Metrics

- Clicks, Signups, "Conversions"
- Typically **10–20% inaccurate**
- Cannot capture data **outside of the session**

Business Metrics

- GMV, Activation, Purchases
- What the **CFO reads**
- From **processed, cleaned** sources
- Defined in **SQL**
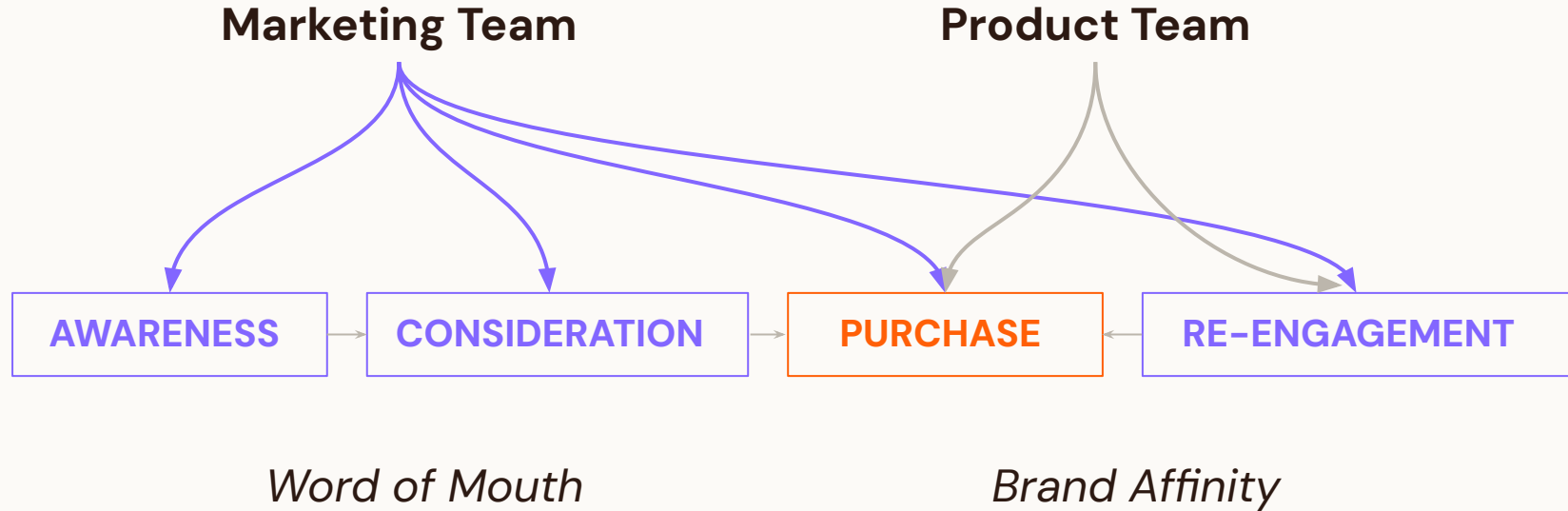
databricks   Google BigQuery   amazon REDSHIFT   snowflake

Trust means a universal language and clarified contributions

Marketing Team

Product Team

| AWARENESS | CONSIDERATION | PURCHASE | RE-ENGAGEMENT |

Word of Mouth

Brand Affinity

Advertising incrementality done right unlocks trust throughout the organization

With the right:

1. Methods
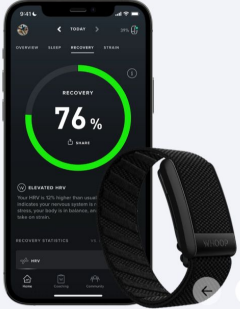2. Region selections
3. Metrics
4. Common language
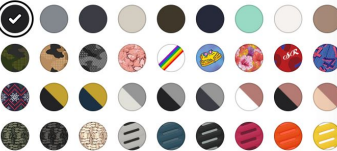
Fireside Chat 🔥

Whoop: a data-driven eCommerce advertiser
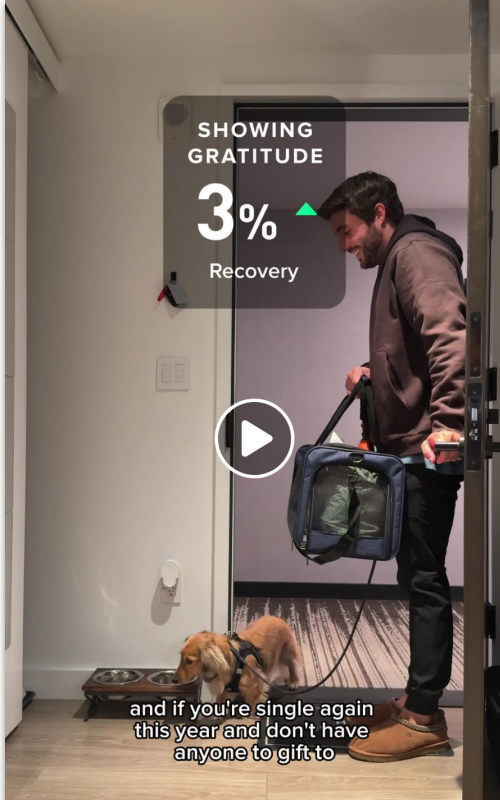
Launching Eppo Geolift

Unifying channels under business-centric experimentation
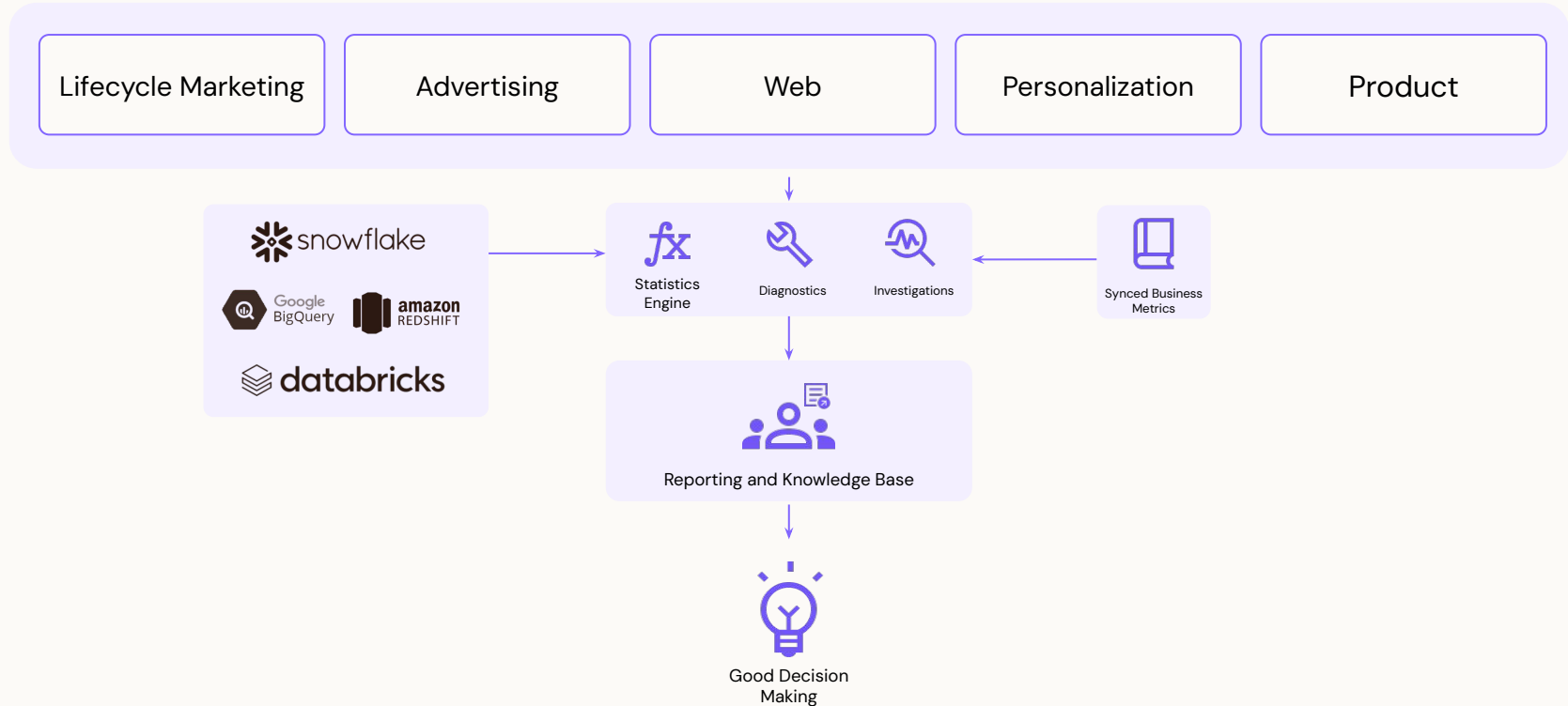
Geolift identifies the real contribution of everything from top-of-funnel to direct response

Business Questions

- Existing channel incrementality
- Optimal budget level
- New channel evaluation
- Online impact on offline sales
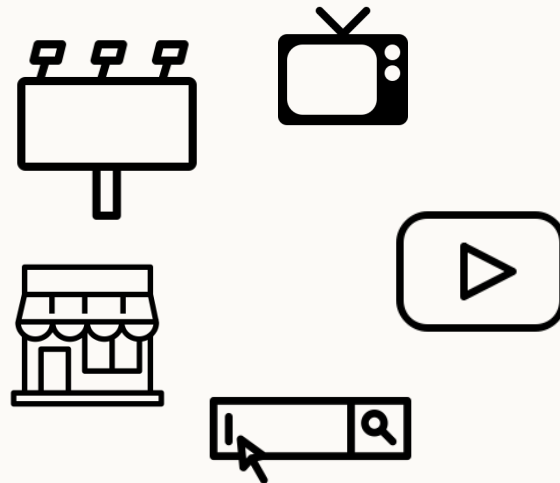
Eppo's approach to incrementality

Always–On Geolift

Automated testing, monitoring and analysis for major digital platforms.

Simple/Powerful Custom Geolift

Smart designs make it easy to find signal in the noise; protocols standardize metrics and best practices.

Real Business Metrics

Measure campaigns on **the same metrics the CFO** uses via your data warehouse.

We *literally* use your data

eppo

Data Warehouse

Metrics

```
SELECT
    user_id
    ,timestamp
    ,cost
FROM orders
WHERE status = 'purchased'
```

snowflake

Google BigQuery   amazon REDSHIFT

databricks

Slice-and-dice Dimensions

```
SELECT
    user_id
    ,persona
    ,marketing_channel
FROM users
```
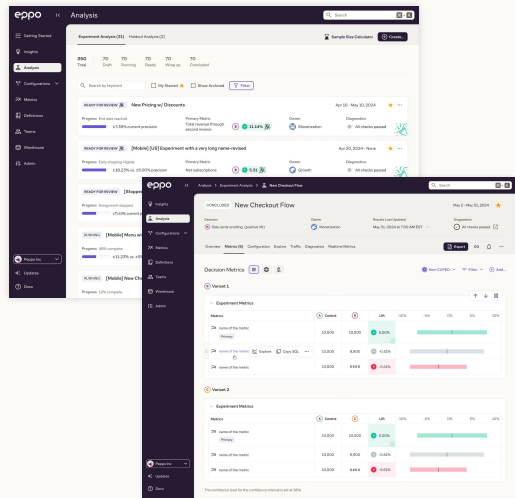
We compute using your warehouse. **Your data never leaves your ecosystem.**

Metrics and dimensions are configured via SQL for maximum flexibility.

World class analysis, reporting, causal inference, and health checks.

Eppo Geolift

Core Test Features:

✓ Experiment Design including:
  - ✓ Region similarity analysis and power analysis
  - ✓ Test/control region selection
  - ✓ Region inclusion/exclusions
✓ State, DMA, Commuting Zone, and Non-Geographic Units
✓ Budget saturation
✓ PhD-level statistical support

Summary

Analysis Last Updated	September 2, 2024
Analysis Start	July 15, 2024
Analysis End	August 21, 2024
Intervention Start	July 15, 2024
Intervention End	August 21, 2024

Overview

Scale 2x

Incremental Purchases
3,250
2,348 - 4,158 @ 95% confidence

Incremental Lift ⓘ
35%
32.53 - 38.86%

$ Per Incremental Purchase ⓘ
$36.80
$32.30 - 43.80

Scale 3x

Incremental Purchases
4,976 ▲ vs Scale 2x
4,585 - 6,158 @ 95% confidence

Incremental Lift ⓘ
28% ▼ 7% vs Scale 2x
25.53 - 32.91%

$ Per Incremental Purchase ⓘ
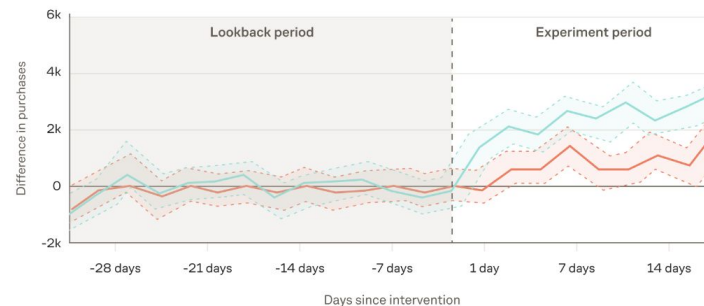**$46.23** ▲ $9.43 vs Scale 2x
$41.86 - 58.28

Difference in Purchases Observed compared to Synthetic Control

☐ Difference between Scale 2x and synthetic control   ☐ Difference between Scale 3x and synthetic control

Lookback period | Experiment period

Difference in purchases

6k / 4k / 2k / 0 / -2k

-28 days | -21 days | -14 days | -7 days | 1 day | 7 days | 14 days

Days since intervention

Eppo's Unique Value Proposition

Alignment at the data and metric level means tests become universally valid throughout the organization

Warehouse-native architecture means data integrity and clarity; no onboarding for Eppo customers

Cutting-edge statistics for test planning and analysis

Single platform for product, CRM, and advertising needs

Q&A

Get your personalized walk–through of Geolift:
[getteppo.com/get–access](getteppo.com/get-access)

Appendix

Eppo's Bayesian Synthetic Control Methodology

Learning relationships between units *before* treatment to provide counterfactual scenarios that estimate causal effects

Test Setup

- Eppo allows unit inclusion/exclusions
- Eppo allows 2- and 3-cell test designs.

Power Analysis

- Eppo provides a list of regions to the user, and can refresh the modeling process if the included regions needs to change or the KPI is not suitable
- Simulate confidence intervals of the Primary KPI to gut-check priors for treatment performance
  a. For example, if prior research efforts find a 5% lift in sales, we'd check that the CIs would be capable of displaying that lift

Analyzing experiment results

Eppo's method can be analogized to an improved CausalImpact. For every treated unit, we have a constant baseline level of the metric being modeled as an intercept. We filter out low-signal small control units, then create synthetic controls based on the the remaining units. For advertising, we then model the ad-effectiveness for each treated unit using a hierarchical model, which includes adstock effects (ad decay) and saturation (or reach) functions.

Results

Once the test is implemented, Eppo will provide interim and final results.