MEASUREMENT PARTNER CASE STUDY · SYNTHETIC DATA · YOOSEOP SEONG

FIZLY Zero: measuring a store-led drink launch

A fictional Korean drink brand launches a zero-sugar line with heavy short-form video. Three quarters of sales are in stores, where no pixel looks. Every number below is graded against the hidden truth of the simulated world.

One question, seven instruments: what does the launch video return per won?

INSTRUMENT	WINDOW	ESTIMATE [INTERVAL]	TRUTH
Platform report	weeks 83-88	**1.51**	0.73
Conversion lift (user holdout)	weeks 83-88	**0.70** [0.54, 0.85]	0.73
Geo test, online part	weeks 83-88	**0.68** [0.59, 0.78]	0.75
Geo test, stores + online	weeks 83-88	**1.68** [1.32, 2.04]	1.92
Geo test with 2-week cooldown	weeks 83-90	**2.27** [1.82, 2.72]	2.62
MMM calibrated to the geo test	weeks 79-104	**5.97** [5.19, 6.77]	4.48
MMM calibrated, both video channels together	weeks 79-104	**4.50** [3.93, 5.12]	4.14

Truth is the simulated answer to each instrument's own question: online rows against the true online return, the geo rows against the regions that went dark, the model rows against the half-year return. Short video alone is the one row whose interval misses; the readout reports the two video channels together.

What the work found

- The platform credits **2.2x** the online orders video adds; conversion lift (people) and the geo test's online part (regions) agree on the real number.
- Online alone, video looks like a loss. With store sales and carry-over, both video channels return **4.5** won of shelf sales per won over the half year, against a break-even of 3.3 at 30% contribution.
- Sustain video sits near break-even and the model runs high on it, so the readout asks for a short-video-only step-down test (Q1) instead of a cut. Search and retail media look best in platform reports and fall short of break-even even at the top of their intervals.
- When pixel matching falls from 75% to 30%, reported ROAS drops from 1.52 to 0.60 while the true return stays at 0.73: a privacy change reads as a performance collapse.

What a measurement partner did

1. Agreed the questions, read dates and rules with advertiser and agency before launch
2. Sized every study from the advertiser's own history (holdouts, regions, sales at risk)
3. Ran brand lift, conversion lift, a store-level geo test and a vendor tracker in one flight
4. Reconciled seven answers to one question and tied each to the decision it should drive
5. Handed over a bilingual readout, a launch playbook, a program view by advertiser tier, and a product memo

How it was checked

9 of 11 headline intervals contain the simulated truth (two misses reported). Repeated runs: 40 per study. The unchanged pipeline on the lab's first world (media at 15% of sales) writes the opposite verdict: launch video has not paid back (1.9 vs truth 2.0).

Demo: https://seongyooseop.github.io/fizly-launch-measurement/site/ · Code and work log: https://github.com/seongyooseop/fizly-launch-measurement

All brands, regions and numbers are simulated. Methods are platform-neutral; the README maps them to named lift tools.