Display campaign: who should get the ad?

From: Data Science | To: Marketing leadership | Basis: randomized experiment, 14.0M users | All figures out-of-sample

> **Recommendation:** stop buying impressions for everyone. Targeting the **30% of users our model ranks most persuadable keeps ~77% of the campaign's incremental conversions at 30% of the impression cost** — roughly 2.6× the conversions per dollar of the treat-everyone campaign. An aggressive 10% variant keeps 64% of the lift at ~6.4× efficiency. Pilot the 30% policy.

FIRST: THE CAMPAIGN DOES WORK

Before deciding who to target, we verified the experiment itself (assignment ratio exactly as designed; feature balance audited) and measured the overall effect with covariate adjustment: the ads lift conversion by **+0.10 percentage points** against a 0.19% baseline — a ~50% relative lift, measured to a precision of ±0.006pp. Notable: the unadjusted readout said +0.115pp; a quarter of that was imbalance between the groups, not the ad. The adjusted number is the one to plan on.

BUT THE LIFT IS NOWHERE NEAR EVENLY SPREAD

We trained a model to predict each user's *individual* response to being shown the ad, and validated its ranking on 4.2M held-out users who were genuinely randomized — so the numbers below are measured, not modeled:

Policy	Impressions bought	Incremental conversions kept	Efficiency vs. treat-all
Treat everyone (today)	100%	100%	1.0×
Target top 30% (recommended)	**30%**	**~77%**	**2.6×**
Target top 10% (aggressive)	10%	~64%	6.4×

"Incremental conversions" = conversions caused by the ad, measured by treatment-vs-control comparison within each targeting tier of the held-out sample.

WHAT COULD GO WRONG

- **Don't use the model to blacklist "harmed" users.** The model flags its bottom tier as people the ad would repel; the held-out data says the opposite — they respond mildly positively. This is a known artifact of the modeling approach at the extremes. The top-of-ranking is validated; the bottom-of-ranking story is not.
- These are intent-to-treat numbers: only ~3.6% of "treated" users were actually reached by an ad. Per *reached* user, effects are much larger — reach is the bottleneck worth separate attention.
- The model ranks well but over-states magnitudes at the top (predicted ~2× the measured lift). Use it to rank, not to promise a forecast of absolute lift.

SUGGESTED NEXT STEP

Run the 30% policy head-to-head against treat-all for one budget cycle with a 90/10 holdout, and confirm the 2.6× efficiency in production before making it default.