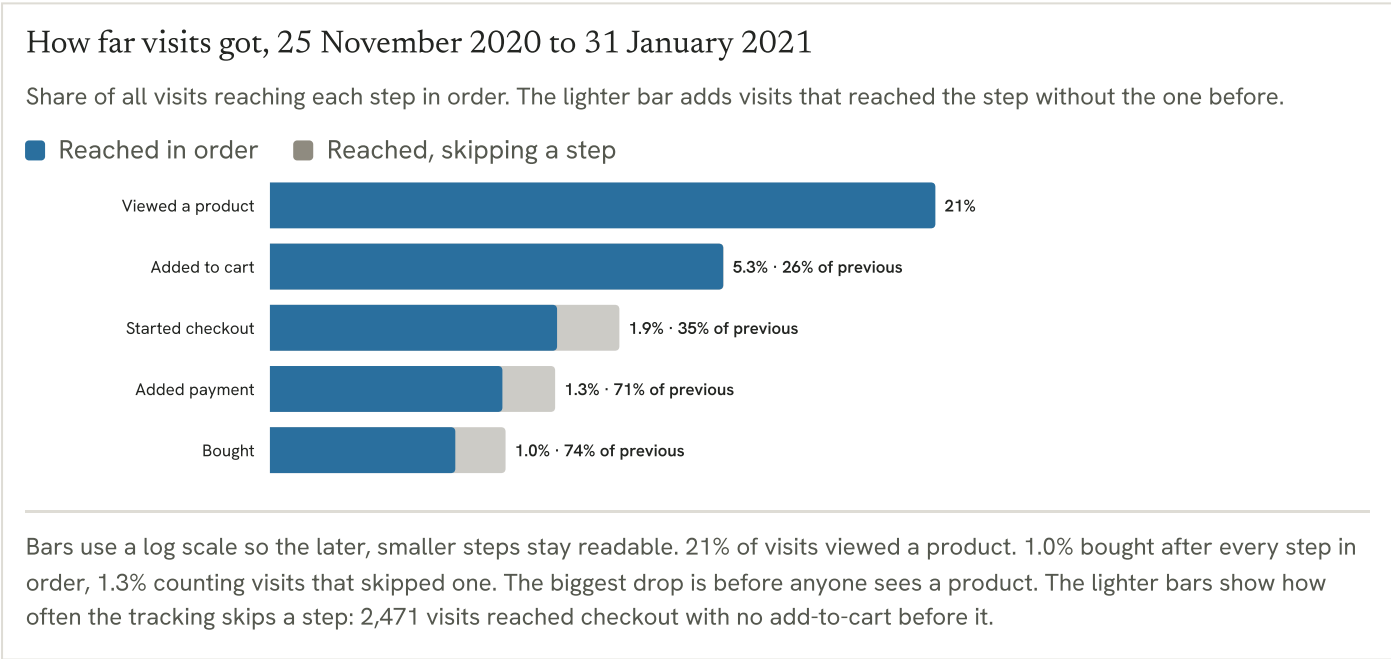


One purchase in four had no add-to-cart event. Fix the tracking first, then test a shorter checkout.

4,295,584 tracked events from 270,154 visitors to the Google Merchandise Store, November 2020 to January 2021. We checked the tracking before trusting it, rebuilt each visit, measured where people drop out, tested a model that spots likely buyers at their first product view, and turned the evidence into a ranked list of tests.

Independent sample using public data, to show what a Dunwell Conversion Check delivers. Google obfuscated this data, so revenue is a sample value, not the shop's takings. Every figure on this page comes from one results file.

Business: online shop for Google-branded clothing and gifts
Visits: 360,129 from 1 Nov 2020 to 31 Jan 2021 Test month for the model: January 2021
Prepared by Dunwell Insights, October 2026



The tracking misses carts, checkout loses half its starters, and most "weak" traffic is just new visitors

1.3% Of visits ended in a purchase 274,090 visits, 25 Nov to 31 Jan	25% Of purchases had no add-to-cart event a tracking gap, not a customer choice	51% Of checkouts didn't end in a purchase 34% never reached payment	38% Of buyers in the model's top 10% test month, against 20% for a simple group rate
--	--	--	---

What we found

- The cart tracking can't be trusted yet.**
Add to cart wasn't recorded at all on 18 days in November while checkouts and purchases carried on. Even after it came back, 25% of purchases had no add-to-cart event in the visit. Any cart-based rate from this setup understates what customers really do, so we kept the cart step out of the model and out of the first test.
- Checkout is where the money leaks.**
Of visits that started checkout, 66% reached payment and 49% bought. These people have chosen products and said they want to pay. Losing half of them is the most expensive drop in the shop.
- Search ads look like the worst traffic. Like for like, they're as good as any.**
As recorded, 0.98% of visits from search ads bought, the lowest of any source. But 91% of them were first visits, and new visitors buy at a fifth of the rate of returning ones (0.67% against 3.29%). With the new-returning mix held equal, search ads come out at 1.48%, and every other source at 1.32% to 1.37%.
- Visits that land on clothing pages buy far less, and not only because they're new.**
A quarter of visits start on a clothing page. As recorded they bought at 0.67% against 1.79% for the home page. About 40% of that gap is visitor mix: like for like it's 1.23% against 1.91%. The rest is real: fewer view a product (16% against 20%), and of those who do, fewer buy (4.1% against 9.0%).
- Mobile is not the problem here.**
Mobile visits bought at 1.39%, desktop at 1.31%. The gap stays tiny with traffic source and visitor type held equal. Device and country barely change any rate in this data, which may be partly Google's obfuscation. We don't recommend mobile work on this evidence.
- January's drop was the season, not the visitors.**
The share of visits that bought fell from 1.57% in December to 0.94% in January. Only 10% of that came from a change in who visited. The rest happened inside every group, which fits the end of the holiday rush.
- A model can pick out likely buyers at their first product view.**
Using only what was known at that moment, LightGBM cut prediction error (log loss) by 10% against a simple group rate on the test month, and its top 10% of visits held 38% of buyers. A plain logistic regression did almost as well, and the difference between them wasn't clear-cut.

What we'd do

- First, fix the add-to-cart tracking.** Check the tag on every way of adding to cart, then confirm for two weeks that cart events line up with orders. It's quick, and every later test depends on it.

Then test a shorter checkout. Guest checkout first, fewer fields, delivery cost before payment. At January's traffic, about 433 visitors a week start checkout. A 4-week test could detect a 12% relative lift in buying; 8 weeks could detect 8%.

Judge traffic like for like. Compare sources and pages with new and returning visitors held equal before cutting any budget. Paid search brings new visitors, and new visitors buy less wherever they come from.

1 · TRACKING FIRST

Add to cart went missing for 18 days, and still misses about one purchase in four

Before building a funnel we checked what the tracking actually recorded. There were no missing visitor or visit ids and no exact duplicate events, which is good. Three problems stood out. Cart events stopped for most of November. On some days from 12 November, checkout events were sent twice per checkout. And the order ids were patchy: 883 purchase events carried "(not set)" instead of an id.



Tracking checks		
Check	Result	What we did
Missing visitor or visit id	0	Nothing needed
Exact duplicate events	0	Nothing needed
Days with no add-to-cart events	18 in November	Funnel starts 25 November
Purchases with no add-to-cart in the visit	25%	Cart step kept out of the model and the first test
Shipping step recorded with checkout	11,105 of 11,106	Dropped as a separate step

Check	Result	What we did
Purchase events with "(not set)" id	883	Counted by visit, not by id
Order ids on more than one purchase event	326	Counted once
Order revenue differs from item revenue	838 ids	Reported, not patched
Visits with no page view at all	18,342	99% are returning visitors; shown as "Unknown" first page
Product price on the product-view event	never filled in	Left out of the model

2 · THE FUNNEL

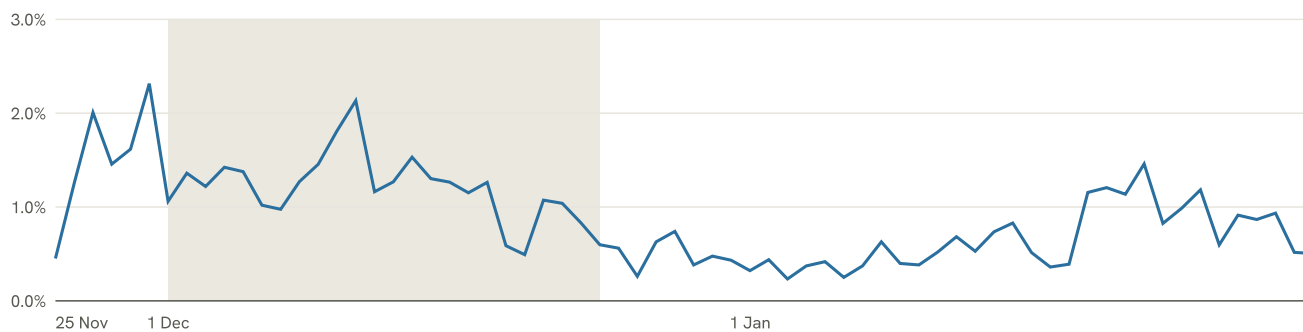
79% of visits never viewed a product. Of those who started checkout, half didn't buy.

A visit counts at a step only if it also reached the step before, at or before that time. Visits that reached a step some other way, such as straight to checkout from a basket filled earlier, are counted separately so nothing is hidden. We show visits and visitors side by side. Visitor rates are higher because people get more than one try.

Each step, by visit and by visitor					
Step	Visits	From the step before	Skipped a step	Visitors	Share of visitors
Viewed a product	56,233 20.52% of visits	-	0	45,653	21.53%
Added to cart	14,618 5.33% of visits	26% 26% to 26%	38	12,067	5.69%
Started checkout	5,091 1.86% of visits	35% 34% to 36%	2,471	4,588	2.16%
Added payment	3,598 1.31% of visits	71% 69% to 72%	1,431	3,205	1.51%
Bought	2,668 0.97% of visits	74% 73% to 76%	1,003	2,520	1.19%
Ranges resample whole visitors 1,000 times. Buyers took a median of 14 minutes from first product view to purchase. 3,671 visits bought; 379 of them had no usable order id.					

Share of visits that bought, each day

Tracked period. Grey band: the 2020 holiday rush.



Buying peaked in early December and fell away after Christmas. Check any test against the same weeks, not against the holiday peak.

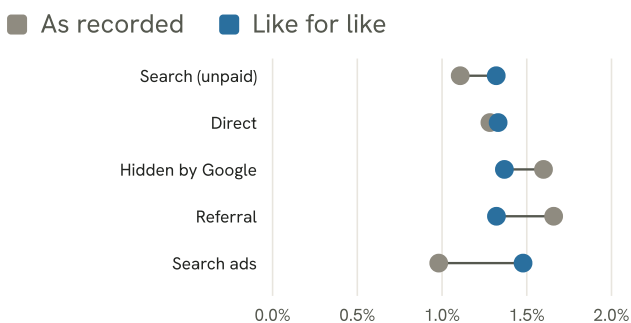
3 · WHO BUYS

New visitors buy at a fifth of the rate of returning ones, and that explains most "weak" traffic

We compared visits by device, traffic source, first page, new or returning, and country, over the tracked period. Groups with fewer than 1,000 visits are folded into "other". Every rate is also shrunk toward the average, so a small group can't top the list by luck. The like-for-like rate holds the mix of new and returning visitors fixed, because that one difference is bigger than any other in the data.

Purchase rate by traffic source

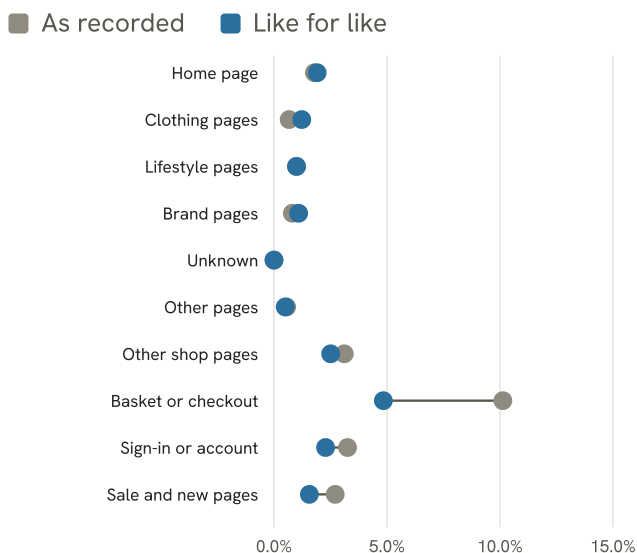
Share of visits that bought. Like for like holds the mix of new and returning visitors fixed.



As recorded, rates run from 0.98% to 1.66%. Like for like, every source sits between 1.32% and 1.48%. "Hidden by Google" covers values Google removed or grouped.

Purchase rate by first page

Groups with at least 1,000 visits. Same two measures.



Basket and sign-in pages convert highest because people arriving there were already buying. "Unknown" is visits with no page view recorded, almost all returning visitors. None of them bought.

New and returning visitors

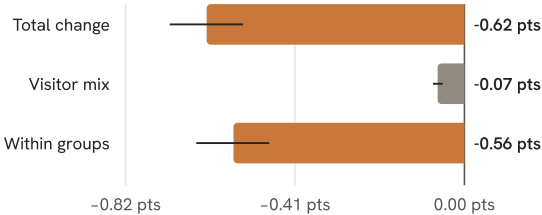
Share of visits reaching each point

	New	Returning
Visits	203,949	70,141
Viewed a product	19%	25%
Bought, if they viewed a product	3.5%	12.4%
Bought, if they started checkout	36%	61%
Bought, all visits	0.67%	3.29%

Returning visitors view more, add more and finish checkout more often. They were 26% of visits and made 63% of purchases.

Why January sold less than December

Change in the share of visits that bought, split into visitor mix and change within groups



Visitor mix means device, traffic source and new or returning. Ranges resample whole visitors. Mix explains 10% of the fall (0.07 of 0.62 points, range 0.05 to 0.08); the rest is lower buying within the same kinds of visit.

These are comparisons, not experiments. Visitors chose how they arrived and where they landed, so a gap between pages doesn't prove a page change would close it. That's why the last section turns each gap into a test.

4 · PREDICTING A PURCHASE AT THE FIRST PRODUCT VIEW

At the first product view, the model cut prediction error by 10% against a simple group rate

The moment we predict from is a visit's first product view, 77,020 visits in all. The question is whether the visit ends in a purchase. The model only sees what was known then: how the visitor arrived, the first page, what they did before the view, the product's category, the time, and their earlier visits. Nothing from later in the visit. Models were trained on 1 November to 15 December, tuned and calibrated on 16 to 31 December, frozen on 11 Oct 2026 at 18:32 UTC, then scored once on January.

Every model on the test month

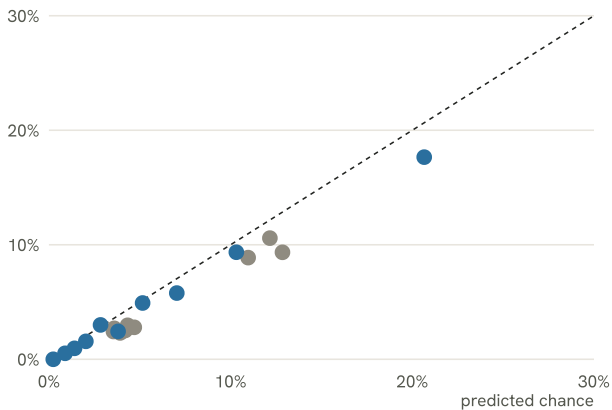
Model	Log loss	Brier score	PR-AUC	ROC-AUC	Average chance
Group rate (device, source, new or returning) late December log loss 0.2092	0.1798	0.0434	8.2% 7.6% to 9.0%	0.671	6.4% actual 4.6%
Logistic regression late December log loss 0.1914	0.1622	0.0415	15.9% 14.4% to 18.1%	0.790	4.9% actual 4.6%
LightGBM (chosen) late December log loss 0.1874	0.1609	0.0412	16.9% 15.2% to 19.2%	0.794	5.4% actual 4.6%

Log loss and Brier score measure how good the chances are; lower is better. PR-AUC measures how well buyers are ranked to the top; a random order would score 4.6%, the base rate. The chosen model's gain over the group rate was 0.0188 (range 0.0166 to 0.0209), resampling whole visitors. Against logistic regression the range crossed zero (-0.0026 to 0.0002).

Predicted against actual

Test month, ten equal groups of visits by predicted chance. On the line means the chances can be taken at face value.

■ Chosen model ■ Group rate

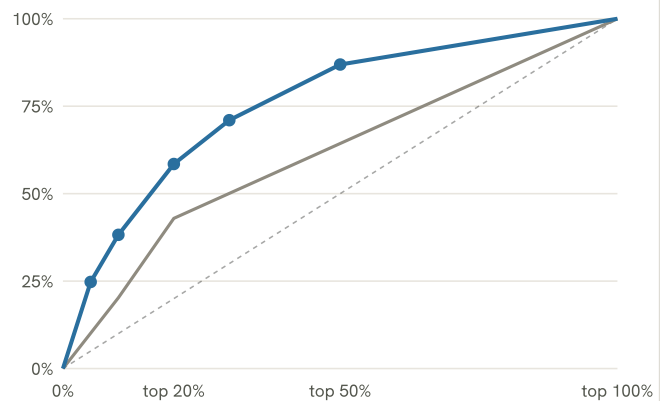


The chosen model's chances were 18% off on average, measured against the base rate, inside the 20% we set in advance. It ran a little high in January (5.4% predicted against 4.6% actual) because it was calibrated on December, when people bought more. The group rate ran higher still.

How many buyers the top of the list holds

Test month. Visits sorted by predicted chance, highest first.

— Chosen model — Group rate
— Random order



The top 20% of visits by predicted chance held 58% of January's buyers, against 43% for the group rate.

What the model leans on

Share of the model's total gain from each input



Earlier visits with a product view, how this visit arrived, how many pages came before the view, and the product's category carry most of the weight. Device and country barely register.

Does it hold up for every group?

Test month, log loss (lower is better)

Group	Group rate	LightGBM	Visits
Desktop	0.174	0.155	13,373
By device			
Mobile	0.188	0.168	9,072
By device			
Tablet	0.199	0.177	553
By device			
Direct	0.181	0.160	5,252
By first source			
Hidden by Google	0.199	0.179	4,832
By first source			
Referral	0.206	0.185	4,114
By first source			
Search (unpaid)	0.157	0.140	7,808
By first source			
Search ads	0.155	0.143	992
By first source			
New in January	0.165	0.146	21,206
By visitor			
Seen before January	0.360	0.335	1,792
By visitor			

92% of January's visits came from people we'd never seen, so the model mostly worked on strangers, and it still beat the group rate for them. The result held when we dropped visits with events recorded at the same moment as the first product view (9,044 visits left). What the model finds is association. A visitor who viewed products on an earlier visit bought 14% of the time, against 3.9% for everyone else, but that doesn't mean getting people to browse more would make them buy.

5 · WHAT TO TEST

Fix cart tracking first, then test a shorter checkout

Each test comes from a gap in the evidence. For each we measured how many visits a month it would reach and how often those visits buy now, then assumed a lift. Nobody can know the lift before testing, so we show 2%, 5% and 10%. Effort is our judgement: small, medium or large. The list is ranked by extra orders a month per unit of effort. A non-buying visitor isn't counted as lost money.

Fix first: the add-to-cart tracking. Add to cart was not recorded for most of November, and even in January about a quarter of purchases have no add-to-cart event in the session. Check the add-to-cart tag on every way of adding an item (product page, quick add, basket), then confirm cart events match orders for two weeks.

Test backlog

January 2021 traffic. Extra orders assume a 5% relative lift (low case 2%, high case 10%). These lifts are assumptions, not findings.

Test	Visits a month	Buy now	Extra orders a month	Effort
Basket page that moves people to checkout (after the tracking fix) About a third of sessions that add to cart go on to checkout. The cart event is under-recorded, so this rate is less certain than the others. Test a basket page with a single checkout button, delivery cost and returns shown up front.	4,374	18%	40 16 to 80	Small
Shorter checkout before payment (first) Of sessions that start checkout, about a third never add payment details, and about 1 in 4 of those that do still don't buy. Test a shorter checkout: guest checkout first, fewer fields, delivery cost shown before the payment step.	2,083	52%	54 22 to 108	Medium
Clearer route from clothing landing pages A quarter of visits start on a clothing page. They buy at about a third of the rate of visits that start on the home page, and fewer go on to view a product. Test landing pages that lead straight to products: product grid above the fold, sizes and prices visible, one clear next step.	30,436	0.43%	7 3 to 13	Small
Better product pages for lifestyle items Visits that start on lifestyle pages (drinkware, bags and so on) view products at more than twice the usual rate but buy far less often after viewing. Test richer product pages for these items: more photos, size and material details, and use shots. This is where Dunwell Studio work could help.	5,008	1.10%	3 1 to 6	Medium

The basket test ranks highest on paper, but it picks people by the add-to-cart event, which is the broken one. It goes after the tracking fix, so the first test is the checkout. Sample revenue uses January's median order of \$34.

First test: shorter checkout before payment

Who is in it	Sessions that start checkout. About 433 visitors a week at January's traffic.
What is split	Visitors, by user id, so each person sees one version every time.
Main measure	Share of those visitors who buy within 7 days. Now 56%.
Guardrails	Sample revenue per eligible visitor; share of orders refunded or cancelled; page load time; share of visitors who reach checkout.
How long	5% lift: 4,984 visitors per version, about 23 weeks. 10% lift: 1,232 visitors per version, about 6 weeks. 20% lift: 298 visitors per version, under 2 weeks.
What it can detect	4 weeks: 12%, 8 weeks: 8%, 12 weeks: 7%. 80% power, 5% two-sided.
When to stop	Fix the length before starting and don't stop early on a good-looking result. Stop at once only if a guardrail breaks, such as errors at checkout.

A 5% lift would need about 23 weeks at this traffic, too long for one test. Aim for a change big enough to matter, or run it in a busier season. We don't promise a result date the traffic can't support.

6 · HOW WE DID THIS

The plan was locked before any group was compared

The analysis plan was written before any data arrived and locked after the tracking checks, before any group was compared or model fitted. Four changes were made at lock time, all from tracking checks, and they are listed in the plan. Every model choice was frozen with a fingerprint (775404a63422...) before January was scored. The whole study reruns from two query results with one command.

1 Check the tracking Ids, duplicates, gaps by day, order ids and revenue	2 Build the funnel Ordered steps by visit and by visitor, with skips counted	3 Compare groups Shrunk rates, like-for- like rates and the mix check	4 Predict, then test Three models frozen on December, scored once on January
---	---	--	---

The models

Group rate	Purchase rate of the visit's device, first source and new or returning group, shrunk toward the average
Logistic regression	L2-regularised, strength picked on late December (C = 10), Platt calibration
LightGBM	15 leaves, learning rate 0.03, 223 rounds, Platt calibration
Ranges	Resampling whole visitors 1,000 times, seed 20261011

Checks we ran

- 4,295,584 events, all placed in a visit
- No missing ids and no exact duplicate events
- Cart gap found by day and the funnel moved to the tracked period
- 17 visits crossing a split date dropped from the model
- 82 visits with a cart or checkout before the first product view kept out of the model
- Encoders, shrinkage and calibration fitted before January
- January scoring refuses to run without the freeze file
- Result checked on new visitors and without tied events

What the model can see

Input	Known at the first product view because
Device, browser, operating system, country	They're set when the visit starts
How the visitor first found the shop, how this visit arrived	Recorded at arrival
First page, hour, day of the week, visit number	Recorded at arrival
Pages, clicks, promotions, searches, scrolls and time before the view	Counted only up to the moment before the first product view
Category of the first product viewed	Part of the view itself
Earlier visits, product views, carts, checkouts and purchases	From the visitor's earlier visits in the data only

Data dictionary

Field	What it is
Visit	One session: a visitor id plus a session id. The same visit can span midnight.
Visitor	The anonymous visitor id Google sets in the browser. One person on two devices counts twice.
Viewed a product	A product-page view event (view_item).
Added to cart, started checkout, added payment	The add_to_cart, begin_checkout and add_payment_info events.
Bought	A purchase event in the visit.
Sample revenue	Purchase revenue in US dollars as obfuscated by Google. Not the shop's takings.
First source	How the visitor first came to the shop, as Google recorded it.
New or returning	Returning means this was not the visitor's first visit.

Decisions log

Decision	Why
Start the funnel on 25 November	Add to cart wasn't recorded on most November days.

Decision	Why
Drop the shipping step	It was recorded with checkout every time, so it isn't a separate customer action.
Predict from the first product view	It's a moment a shop can act on, and only earlier information is used.
Leave product price out of the model	It was never filled in on the product-view event.
Split by date, not at random	A model has to work on future visits, not shuffled ones.
Count visits, not events	Some events fire twice. Visit flags aren't affected.
Rank tests by orders per effort, but skip ones that rely on broken tracking	A test that picks people by a broken event can't be trusted.

What this data can't tell us

- Google obfuscated this data. Some values were replaced with placeholders, and Google warns that it may not be fully consistent. Device and country may have been scrambled, which would explain why they make almost no difference.
- Revenue is a sample value. No figure here is the shop's real takings.
- Visitors are browser ids. Someone who switches device or clears cookies counts as new, so returning visitors are undercounted.
- Nothing here was randomly assigned. Group gaps and model inputs are associations. Only a test can show what a change would do.
- Three months, ending with the holiday season. The test month, January, was quieter than the months the model learned from.
- The lifts in the test backlog are assumptions. The traffic and current rates are measured.

This analysis is based on the data described here and the methods shown. It supports decisions. It is not a promise of any result, and it is not financial, legal or tax advice.

Every job. Done well.

Dunwell Insights · info@dunwellstudio.com

Data: Google Analytics 4 sample ecommerce dataset (obfuscated), Google Merchandise Store, via BigQuery public datasets. Google obfuscated the data, so revenue figures are sample values, not the store's takings. We summarised the data and added estimates. Google does not endorse Dunwell, and this is not a Dunwell client.