

How many times is too many, and how long an audience stays worth targeting - a brief for your AI assistant

Hand this whole file to your assistant along with the notebook from <https://www.stochasticstrata.com/blog/how-many-times-is-too-many>. It tells the assistant what to do, what to refuse, and what to say before anyone changes a frequency cap or a membership window on the strength of the number.

For the assistant: what you are being asked to do

The person you are helping wants two numbers: how many times one person should see their ad before the next impression stops paying for itself, and how long the audience stays worth keeping as its members get older. They may not write code. Do the technical parts yourself and walk them through one step at a time. Ask for one thing, wait, then ask for the next.

- Find out which of the two files they can produce (both, ideally). The frequency file needs a per-person log, which most ad dashboards do not export - say so early, and steer them to an owned channel (email, push) or a log-level export if they have one.
- Ask what an impression costs them (cost per thousand) and what one sale is worth to them in margin. The cap is read off those two numbers; without them there is no cap.
- Map their columns onto the ones the notebook expects, set the file paths in the marked cells, and run it for them if you can.
- Read the refusals and the warnings out loud. They are the most useful output in the file.
- Before they change a cap or a window, say the caveats at the end of this brief. Every one of them is measured.

The two files

THE FREQUENCY FILE: one row per frequency bucket, three columns - 'frequency' (times a person saw the ad in the window; a top bucket may be written 15+), 'reach' (people in that bucket) and 'conversions' (how many of them converted). The window is theirs to state, and the cap comes out in impressions over that window.

THE AUDIENCE FILE: one row per person, three columns - 'joined' (the date they entered the audience), 'converted_on' (the date they bought, blank if they have not) and 'as_of' (the date of the export, the same on every row).

Or the raw logs: for frequency, one row per impression and one row per conversion, each with a person identifier; for the audience, an events log with a person identifier, an event name and a timestamp. The notebook consolidates both.

Queries to adapt

```
-- One row per frequency bucket: the people who saw the ad f times in the window, and how many
-- of them converted. Needs a PER-PERSON log - one row per impression with a person identifier,
-- and one row per conversion with the same identifier (a log-level ad-server export, or an
-- email/push tool's send and order logs). The standard dashboards do not carry this.
WITH per_person AS (
  SELECT person_id, COUNT(*) AS impressions
  FROM `project.dataset.impressions`                -- one row per impression
  WHERE event_date BETWEEN '2026-05-01' AND '2026-05-28' -- the window the cap will be stated in
  GROUP BY person_id
),
converted AS (
  SELECT DISTINCT person_id
  FROM `project.dataset.conversions`                -- one row per conversion
  WHERE event_date BETWEEN '2026-05-01' AND '2026-05-28'
)
SELECT
  LEAST(p.impressions, 15) AS frequency, -- 15 stands for the '15+' bucket
```

```

COUNT(*) AS reach,
COUNTIF(c.person_id IS NOT NULL) AS conversions
FROM per_person p
LEFT JOIN converted c USING (person_id)
GROUP BY frequency
ORDER BY frequency;

```

```

-- One row per person: the day they first visited (joining the audience), the day they first
-- bought (blank if they have not), and the date of the export. From a GA4 export in BigQuery;
-- adapt the dataset name, the two event names and the date window.
SELECT
  user_pseudo_id AS person_id,
  MIN(IF(event_name = 'first_visit', PARSE_DATE('%Y%m%d', event_date), NULL)) AS joined,
  MIN(IF(event_name = 'purchase', PARSE_DATE('%Y%m%d', event_date), NULL)) AS converted_on,
  CURRENT_DATE() AS as_of
FROM `project.analytics_XXXXXXX.events_*`
WHERE _TABLE_SUFFIX BETWEEN FORMAT_DATE('%Y%m%d', DATE_SUB(CURRENT_DATE(), INTERVAL 90 DAY))
AND FORMAT_DATE('%Y%m%d', CURRENT_DATE())
GROUP BY user_pseudo_id
HAVING joined IS NOT NULL;

```

How to run it

- Local: open the notebook, set the file paths and the two economic numbers in the marked cells, run all.
- Colab: upload the notebook and their files, set the paths, run all. Tell them plainly that this puts their files on Google's servers. If that is not acceptable, use the local route.
- Nothing leaves their machine on the local route. Say so - people ask.

Can you do the arithmetic yourself? The cap, yes - and the check says so

The curve is share-who-bought = $\text{ceiling} \times f^n / (K^n + f^n)$, with f the times seen, K the half-way frequency and n the shape. Fit it like this. For every pair (K, n) on a grid - K from 0.5 to 40 in steps of 0.25, n from 0.5 to 4 in steps of 0.05 - compute the unit curve $g = f^n / (K^n + f^n)$ at each bucket's frequency; the ceiling for that pair is $\text{sum}(w \times g \times y) / \text{sum}(w \times g)$, floored at zero, with w the people in the bucket and y the share who bought; keep the pair with the smallest sum of $w \times (y - \text{ceiling } g)^2$, scanning n outer and K inner with the first minimum winning; then refine twice, rescanning twenty steps either side of the best pair at a tenth of the step.

Then the gate: the curve must beat a flat line. Twice the difference in binomial log-likelihood between the fitted curve and the flat overall rate, compared with a chi-square on two degrees of freedom - the tail chance is $\exp(-\text{statistic} / 2)$ - and read a cap only if that chance is under one in twenty.

Then the cap: the whole number of impressions from 0 to 60 that makes a reached person worth the most, which is the value of a sale times the curve minus the cost of one impression times the count. If it lands past the highest frequency in the file, say 'at least the top bucket' rather than the number.

That rebuild was run from this page alone, with no sight of the notebook, on 4 pinned reports, and reproduced the notebook's cap and gate verdict on 4 of them. So the cap you compute by hand IS the notebook's. What the notebook adds is the refusing and the warning - below.

The half-life: the simple way is rough, and the gap is measured

The simple version: build the share-not-yet-bought curve by age (at each age, the share of people watched that long who have not bought - counting a person only while they are still being watched); read the never-buyer share off where the curve flattens, its last value; the buyers' own curve is (share not yet bought - that tail) divided by $(1 - \text{tail})$; straighten it by plotting $\ln(-\ln(\text{buyers' share}))$ against $\ln(\text{age})$ and fitting a line - the slope is the shape, the scale is $\exp(-\text{intercept} / \text{shape})$; the half-life is $\text{scale} \times (\ln 2)^{(1 / \text{shape})}$.

On the article's own audience that simple version lands at 8.3 days against the notebook's 9.4 and a truth of 9.2 - -12 per cent from the notebook. It needs a curve that has flattened, which means an audience watched for about three of its own half-lives; on a younger audience the tail is not the never-buyer share and everything after it is wrong. The notebook fits the never-buyer share and the timing curve together by maximum likelihood and is the number to plan on.

What the notebook does that no rebuild can

- It refuses fewer than four buckets with people in them, fewer than 50 conversions in the report, and fewer than twenty buyers in the audience file, each in plain words. The 50 is measured: below it, the worst tenth of audiences lose more than a tenth of their spend to a fitted cap.
- It warns when the share who bought FALLS with frequency. That is the shape a report takes when buyers are dropped from the audience once they buy, and no cap should be read from it - frequency there is partly a consequence of buying.
- It warns when the cap lands past the top bucket, and when the audience has been watched for fewer than 3 of its own half-lives: a week-old audience returned a half-life of 43.6 days against a truth of 9.2, and said so.

The honest caveats - say these BEFORE they act

- The report's curve is not the person's curve. The platform chose who saw the ad most, and if it chose keen people the report's cap runs late - and often off the end of the data; if it chose people who are merely online more, early. Measured on 499 synthetic audiences. The report cannot say which way its lean runs; only a holdout - cap a random slice, leave the rest - can.
- A cap belongs with its window. Twelve in twenty-eight days is not three in seven. State the window with the number and compare it only with caps over the same window.
- A cap past the top bucket is the shape guessing. Say 'at least the top bucket'.
- An S-shaped curve has a floor as well as a cap: below it extra impressions barely help. A reading of 'no cap' is therefore not a reading that the frequency is right; say so when the audience runs at low frequency.
- How sharply the curve bends carries the brand's size as well as the audience's response - a dominant brand saturates earlier - and no test in the data separates them.
- A promotion or an import that brought a burst of joins mixes two audiences; the half-life across the mixed pool describes neither. Read it off the quiet weeks.
- The half-life describes when the audience's members bought, not what the ads did. The membership window the notebook prices assumes every sale is the ads' doing; if only half are, it shortens - on the article's audience from 38 to 34 days, and to 30 if a quarter are. Only a refresh holdout measures that share.

Numbers to check yourself against

In the article's own synthetic audience - 113,591 people reached over twenty-eight days, at 12 euros per thousand and 25 euros a sale - the fitted curve has a ceiling of 1.85 per cent, a half-way point at 5.2 and a shape of 1.18, and the cap is 10. No cap at all leaves 7.8 per cent of the spend on the table; twelve a month leaves 3.3. The report's buckets are in data/checks.json under the worked example. The analysis places the top bucket at its average frequency; read with the top bucket at its label (15), as an export gives it, the same report gives a cap of 11. If your rebuild reproduces 10 with the average, or 11 with the label, you have it right.

The same audience's membership history gives a half-life of 9.4 days with 12.0 per cent ever buying; by two half-lives 84 per cent of the eventual buyers have bought, by three 96. The economic membership window at these economics is 38 days.

Synthetic data throughout the article and notebook. The method transfers; the numbers do not - which is why the notebook reads the cap and the half-life off the reader's own files rather than shipping them.