Distilo

RELATÓRIO DE ANÁLISE

Análise de Churn

A worked example of patient recall (fictional data).

Gerado em	16/08/2026
Cliente	Sample report · synthetic data
runId	demo-churn

Who to call back: 1,274 patients and R$ 825,644 in paid doses nobody has given

I found **1,274 patients in an actionable situation, 39.6% of your base of 3,221**, and within that **R$ 825,644 in doses that have already been paid for and never given**. That is the most uncomfortable number in the report, because it is money that has already come in and service that has not yet gone out.

What the base shows

You have 3,221 patients on record, 19,247 doses sold and R$ 10,748,740 in recorded revenue between 26/12/2023 and 30/04/2026. The growth is real: purchases went from 201 doses in December 2024 to 2,446 in January 2026.

The operation is also organised in a way that makes the analysis easier. In 99.9% of cases the patient buys the whole course at once: whoever buys hepatitis B buys the three doses on the same day, whoever buys Qdenga buys the two. That means each purchase is a closed commitment, and each dose not given is an open promise. I call each of those purchases a “course” in the rest of the text.

Reference date for everything here: 30/04/2026, the last day that appears in the data. No calculation uses today’s date.

The problem

Of the 19,247 doses sold, **1,467 are marked as not given: 7.6% of everything you sold**. They split into two quite different problems, and the difference matters for the front desk’s script.

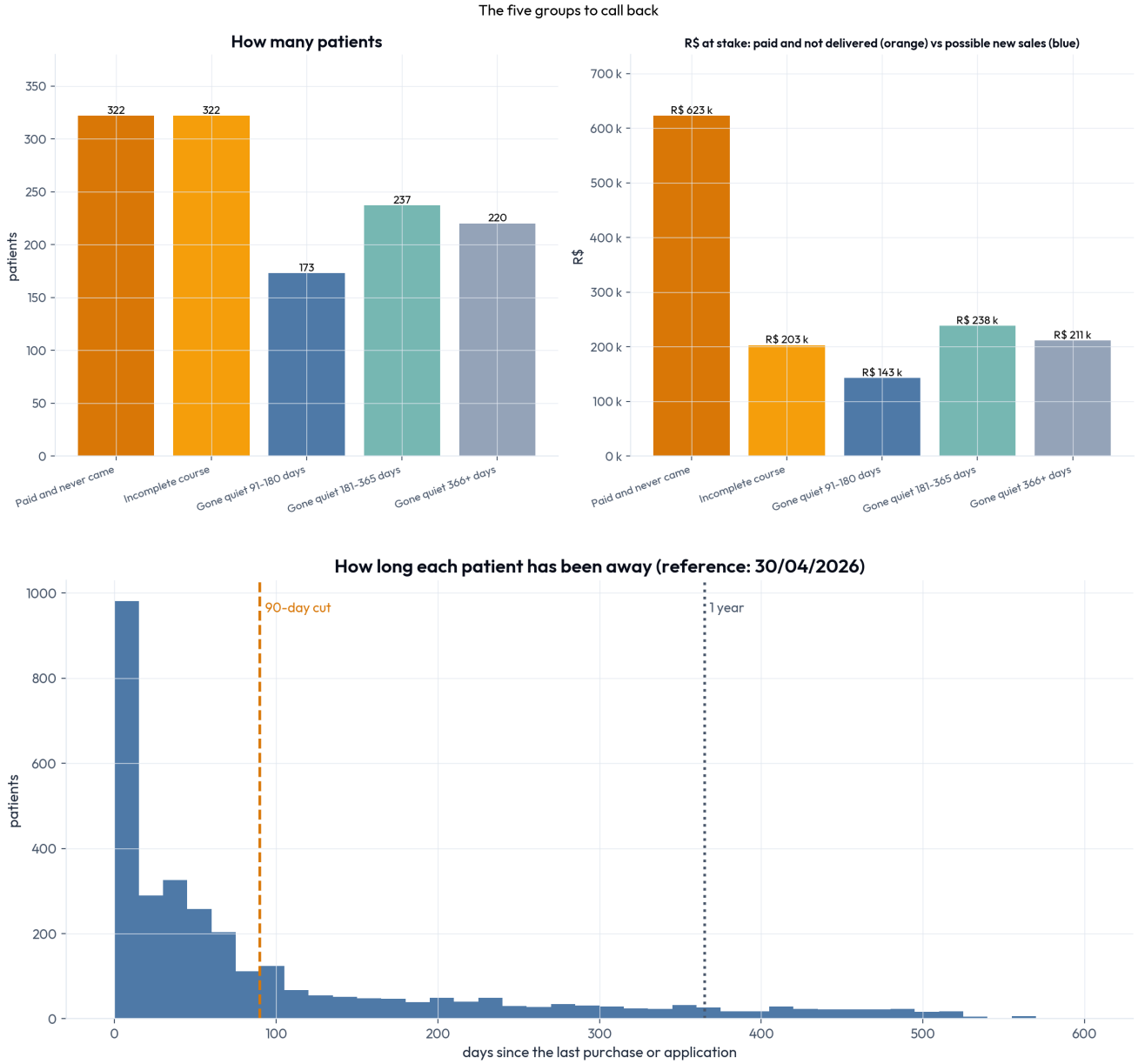
322 patients paid for the full course and never showed up, not even for the first dose. That is 656 courses, 1,107 doses and R$ 622,914. The average wait on those courses is already 200 days, and 203 of those 322 patients have been waiting for more than 180 days. The most extreme case is a patient with 6 doses pending and R$ 6,628 paid.

Another 322 patients took the first dose and did not come back to finish. That is 360 doses and R$ 202,730. These are the easiest to bring back: they have already walked into the clinic, already trusted you, already taken a needle. The normal gap between the first and the second dose in your base is 77 days, so anyone past that is genuinely late.

The two groups do not overlap: they are 644 distinct patients.

Another 630 patients have nothing pending, but have gone quiet. They bought, took everything and did not come back. I split them into three bands because the conversation changes in each one: 173 have been gone for 91 to 180 days, 237 for 181 to 365 days and 220 for more than a year. Together they left R$

1,800,096 in the till over time, and 494 of them bought two or more different vaccines, so they were not one-time customers.



The first bar in the chart above is inflated, and the next block explains why.

Before you dial: two things to check in the system

These two checks are gates. Until they are done, part of the lists may be pointing at the wrong patient, and nothing annoys a customer more than being chased for something they have already done. I estimate 2 to 3 hours of work for the two.

Gate 1: the 851 administrations stamped on the last day of the data

On 30/04/2026 the data records **851 administrations**. On a normal day in the two previous months there are 40. That is twenty-one times the normal traffic, which would already call for an explanation, but two details close the case.

First: **not one of those 851 is a first dose**. They are all second and third doses. On any other day in the data, 62% of administrations are first doses. A real mutirão (a one-day mass-vaccination drive) would bring new people along with it.

Second: the gap between purchase and administration collapses. Across the whole base, the second dose is never given earlier than 60 days after purchase and the third never earlier than 120 days, which makes clinical sense. On this specific day there are second doses at 30 days and third doses at 36 days.

The two possible readings are: **(a) an extraction artefact**, where doses that were scheduled or still pending were stamped with the date the file was generated, or **(b) a real mutirão** of second and third doses that had been piling up. It cannot be settled with this file alone, and I will not settle it for you.

The test that separates the two: **open the appointment book for 30/04/2026 and count how many people actually went through the vaccination room**. If it is 30 or 40, it is an artefact. If it is 851, it was a mutirão. The full list is in `aplicacoes_ultimo_dia_conferir.csv`, ordered by the purchase-to-administration gap, so the least plausible cases come first.

What is at stake: if it is an artefact, those 851 doses (R$ 446,514, 736 patients) are currently counted as delivered and are in fact pending. The “paid and never came” group and the “incomplete course” group would grow, and the R$ 825,644 figure would rise. Note that most of those 736 patients bought recently, so they are probably not overdue cases but doses still within the window. That is why I did not include them in the call lists: including them now would be chasing people who owe nothing.

Gate 2: the 1,324 repeated courses

I found **1,324 courses where the same patient bought again a vaccine they had already bought before**, involving 1,073 patients and R$ 1,530,840, or 14.2% of recorded revenue. The median gap between the earlier purchase and the repeat is **92 days**.

That is clinically odd. I left quadrivalent flu out of the count precisely because it is the only periodic booster in your price list. What remains is hepatitis B repeated, HPV-9 repeated, shingles repeated, Qdenga repeated: courses you take once in a lifetime, bought again three months later. The repeats show up across all vaccines in roughly the same proportion (334 in Qdenga, 180 in shingles, 153 in hepatitis B, 150 in HPV-9), which weakens the idea that this is the behaviour of one specific product.

The two readings: **(a) duplicate record**, where the same course was entered twice in the system, or **(b) a real sale**, with different people using the same record (a mother booking for her child under her own name, for instance) or revaccination after a lost vaccination card.

The test that separates them: **take 10 cases from the top of `cursos_repetidos_conferir.csv` and check in the prontuário (the patient record) whether the two purchases have separate invoices and different vaccine batch numbers**. Two invoices with distinct batches is a real sale; a single invoice is a duplicate.

What is at stake for the call-back: **52 of the pending items I listed sit in repeated courses** (42 in the "paid and never came" group and 10 in "incomplete course"), adding up to R$ 45,906. If they are duplicates, calling those patients is inviting them to take a dose they have already taken. Every call list carries the `repeticao_a_conferir` column marked true in those cases: **filter on it and leave those patients for the end of the queue until gate 2 is answered**.

The contact-recovery plan

With the two gates answered, this is the queue. The front desk makes 40 contacts a day, about 5 minutes each, which is 3h20 of daily work and costs R$ 100 a day at R$ 30 an hour. The whole operation, with the four touches described below applied to the 1,274 patients, comes to around R$ 12,740 and 425 hours of front-desk time. The cost in money is irrelevant next to what is at stake; the limiting factor is the front desk's diary, and that is why the order below matters.

What each group returns per real spent, at 30% conversion and with the four touches included in the cost:

Group	Contacts	Cost	Return at 30%	Ratio
Paid and never came	322	R$ 3,220	R$ 186,874	58x
Incomplete course	322	R$ 3,220	R$ 60,819	19x
Gone quiet 91-180 days	173	R$ 1,730	R$ 42,785	25x
Gone quiet 181-365 days	237	R$ 2,370	R$ 71,548	30x
Gone quiet 366 days or more	220	R$ 2,200	R$ 63,390	29x
Flu overdue	146	R$ 1,460	R$ 8,322	6x

Before reading the table, a warning about adding up: the flu row brings no new people, the 146 are already counted inside the three gone-quiet rows, which is why the contacts column adds up to 1,420 for 1,274 distinct people. Add the columns as a ceiling, never as a floor. Two readings of this table. First, **the return of the first two groups is not new money**: those R$ 186,874 and R$ 60,819 are already in your till and what the call does is turn a service debt into service delivered, taking that amount out of the risk of becoming a refund request. Only the gone-quiet and flu rows are new sales. Second, **no group falls below 2x**, so none needs to be dropped on cost. What decides the order is front-desk hours, not money. All the return rows carry the same fragility: the 30% is a market benchmark, not your number (low confidence). And in the new-sale rows (gone quiet and flu) the return is at full price: what the dose costs the clinic to buy is not in the data, so the real ratio in those four rows is lower than the table shows, though there is enough room for none of them to fall below 2x.

The four-touch sequence

It applies to every group, changing only the text and the opening channel.

On **day 0** goes a light message, no offer and no chasing, just a reminder that there is a dose reserved in the person's name. On **day 3**, whoever has not replied gets a reminder with a concrete deadline, something like "I can hold this slot until Friday". On **day 7** comes a real incentive: a priority off-peak slot, or home administration if the clinic offers it, or a small gift. On **day 14** it is a human phone call, even if the three earlier touches went by WhatsApp, because this is the last card to play. Whoever does not respond to the fourth touch leaves the active queue and only comes back in the following quarter's mass campaign. Do not push: going beyond that burns the brand and does not convert.

Today: the 50 largest in the "paid and never came" group

Open `pagou_e_nao_veio.csv`, which already comes ordered from the largest to the smallest pending amount, and take the first 50 rows. **That is R$ 206,166 paid and not delivered, concentrated in 50 people.** This batch is not for the front desk: the caller is you or management, because the subject is serious money and the conversation can turn into a refund request.

Worth knowing what this cut by money does to the queue: 50 names chosen by amount are not a portrait of the group. Bela Vista gets in with 22.6% of its patients in the group and Riacho Doce with 5.3%; the over-60s get in at 21.3% and children up to 12 at 8.7%. That is not your choice or mine, it is the vaccine price calling the shots, because expensive courses concentrate in those taking shingles and HPV-9. Since the other 272 in the same group get the same script within eight working days, the imbalance lasts a week and does not become policy. If week 1 slips, redo the batch taking the 5 largest from each neighbourhood instead of the 50 largest overall.

The script is about service, not sales: the person paid, the clinic owes them, you called to book. The target is **30 of the 50 with a date booked by Friday**. The number you check on Friday is how many of those 50 names have a slot in the system. If the answer is below 20, the problem is not the list, it is the script, and it is worth changing the opening before scaling to the other 272.

If you do not do this: those 203 patients who have already waited more than 180 days keep their money sitting in your hands, and every month that passes raises the chance they become a refund request instead of a booking.

Week 1: the other 272 in the same group, with a script test

The remaining 272 go by message, 40 a day, eight working days. Here it is worth **splitting the list in half, 136 and 136, and using two different scripts**: one focused on the money ("you have R$ X in paid doses waiting") and the other on protection ("your hepatitis B course is incomplete and protection only closes on the third dose"). Measure how many bookings each half generated and use the winner in every following wave. Target for the week: **80 bookings, around 30% conversion** (low confidence, it is a market benchmark and not a number from your base).

Whoever does not respond to the fourth touch goes onto a waiting list and gets one last attempt together with the seasonal campaign.

Week 2: the 322 with an incomplete course

`esquema_incompleto.csv`, 40 a day, also by message. This group is cheaper to convert because the person has already come once and already has a dose in their arm. The argument is clinical and short: half

a course does not protect. Target: **100 bookings in the week**, and the Friday number is how many second and third doses were actually given, not how many messages went out.

The largest slice of the group is hepatitis B, with 84 of the 322, followed by Qdenga with 69, shingles with 52, HPV-9 and meningococcal B with 40 each and meningococcal ACWY with 37. Since hepatitis B is the only three-dose vaccine, it deserves a script of its own: whoever stopped at the second dose has a different clinical argument from whoever stopped at the first.

Week 3: the recent and medium gone-quiet

`sumidos_sem_pendencia.csv`, filtering the "recente" (recent: 173 people, gone for 91 to 180 days) and "medio" (medium: 237 people, 181 to 365 days) bands. There is no debt here at all: it is new sales. The courses sold at your clinic range from R$ 190 to R$ 4,468, so the two bands cannot be projected at the same price: I used the basket each one actually bought, **R$ 824 per course in the recent band and R$ 1,006 in the medium one** (the average of the group's own historical courses, not the catalogue). At 30% conversion, the recent band is worth **R$ 42,785** and the medium band **R$ 71,548** (low confidence on both, the conversion rate is a market benchmark; the price is high confidence, it comes from the real basket each band bought).

The order of attack is exactly this: people who went quiet 4 months ago convert better than people who went quiet 10 months ago, and you want to learn from the easy band before spending front-desk hours on the hard one.

Week 4 onwards: the cold ones and the flu campaign

The 220 who have been gone for more than a year do not deserve four individual touches. **Do a single mass send** with a concrete offer and let whoever replies join the normal queue.

Alongside it goes `gripe_anual_vencida.csv`, with **146 patients whose last quadrivalent flu purchase is more than 365 days old**. An important warning: **all 146 are already inside the gone-quiet list**, they are not new people. And the premise that the dose is annual comes from the package insert, not from your data: in your base, the median gap between two quadrivalent flu purchases is **81 days**, which is far too short to be an annual booster and is probably one more symptom of the gate 2 problem. Treat this list as an alternative script for the gone-quiet, not as a segment of its own. At R$ 190 a dose and 30% conversion, that is R$ 8,322 (low confidence, premise not confirmed in the data).

Still on flu, there is a segment that appears in none of the lists above: **1,896 of your 3,221 patients (58.9%) have never bought quadrivalent flu**, and **1,027 of them are still active**, with nothing pending and without having gone quiet. The other 869 are already in the call lists, so do not add the two groups together. Those 1,027 are in `sem_gripe_ativos.csv`, have spent R$ 4,050 at the clinic on average and have never taken the only vaccine in your catalogue that repeats. The front desk sends them a single message at the opening of the next seasonal campaign, 100 names a day over two weeks, offering a flu slot. At R$ 190 a dose and 15% conversion, that is **R$ 29,270** (low confidence: the rate is a seasonal-campaign benchmark, it does not come from your base, and the annual-booster premise still depends on gate 2). The number you check on Friday is how many flu doses were given, not how many messages went out; if the first week comes in below 50 doses given, stop the send and test another offer on 100 names before spending the

rest of the list. If you do not do this, that slice of the base keeps buying once in a lifetime and never enters an annual cycle.

The high-value list, and why it gained 31 names

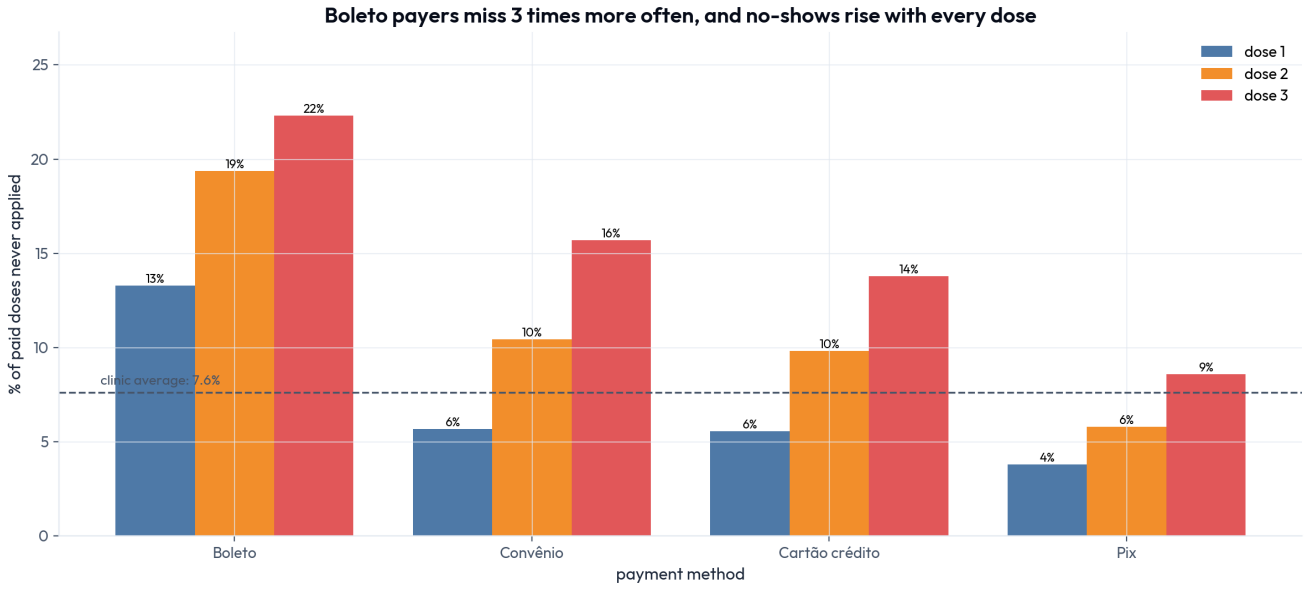
`alto_valor_em_risco.csv` holds **159 patients**: the 128 in the top decile of historical spend (the 10% who spent the most) among all the actionable patients (cut-off at R$ 4,702.80, average of R$ 5,788 per patient) plus 31 that I added on purpose. It is not a separate group: it is a cut across the three previous groups, so **do not add its R$ to the others or you will count twice**.

The reason for the 31: when you order a queue by money spent, the money carries along whoever it favours. I measured that queue's selection rate by payment method, acquisition channel, neighbourhood, sex and age band. The biggest deviation is by **age band**: adolescents aged 13 to 18 get onto the list at 14.0%, while adults aged 46 to 60 get on at 8.7%. The ratio between the lowest and the highest rate is 0.62, below the 0.80 level usually used as the acceptable floor. It makes sense: HPV-9 costs R$ 1,117 and is an adolescent vaccine, so age rode into the queue on the back of the product price. Except that fixing age on its own does not fix the queue: when you pull the adult bands in, the deviation moves to neighbourhood, which goes from 0.67 to 0.61. That is why the quota runs in rounds, always on the most unbalanced cut of the moment, until none of the five sits below 0.80.

The file you will use already comes balanced: **I added the 31 highest-spending patients from each under-represented group**, 20 by age band (7 from the 46-60 band, 4 from 31-45, 3 from 0-12, 3 from 19-30 and 3 from 60+) and 11 by neighbourhood (4 from Alto da Serra, 3 from Boa Vista, 2 from Vila Nova, 1 from Morada do Sol and 1 from São Cristóvão), until each group reached 80% of the rate of the most favoured group. It costs **R$ 77.50 of front-desk time** and those 31 have an average value of R$ 4,570, so this is not charity, it is expensive people that ordering by decile would have left out. In the delivered file none of the five cuts sits below 0.80: the tightest is still age band, now at 0.80, and neighbourhood rises to 0.81. The `entrada` column in the file says whether each row came in by the value decile or by the balancing quota.

What makes a patient miss: the payment method

Your overall no-show rate is 7.6% of doses sold. It is not spread evenly, and the cut that separates most is the payment method.



Whoever pays by Boleto (a bank slip paid offline) misses 16.0% of the time. Whoever pays by Pix (instant bank transfer) misses 4.8%. That is 3.3 times more. Boleto accounts for 14.2% of your doses sold and for **28.6% of all the pending value, R$ 236,418**. The effect is not an illusion of mixed audiences: it shows up inside each acquisition channel separately, from 22.8% on Boleto from Google down to 2.5% on Pix from referrals.

The channel also separates well: **whoever came from Google misses 13.0%, whoever came by referral misses 4.3%**, and Google on its own carries R$ 339,371 in pending doses. A patient who arrives through search has less of a bond and slips away more easily.

And no-shows grow with each dose of the course: 5.9% on the first, 9.5% on the second, 12.7% on the third. One caveat in reading that: **the third dose only exists for hepatitis B in your price list**, so that last step mixes “it is the third dose” with “it is hepatitis B”. The rise from the first to the second dose, that one does show up across six different vaccines and is solid.

That turns into two process changes, and they are what stop the list from reappearing six months from now:

The first is **booking every dose of the course at the moment of purchase**, with a date in the system, especially for those paying by Boleto and those who came from Google. Today the person pays for the whole package and leaves with no date for the second dose; 77 days later nobody remembers. Responsible: the front desk, from next Monday, for 100% of new sales.

The second is **an automatic alert on the dashboard**: a paid course with no administration within 30 days joins the contact queue on its own, and a dose in a started course that passes 100 days from the previous one raises a flag. That way the front desk works 5 to 10 names a day instead of discovering 1,467 stalled doses two years from now. Responsible: whoever looks after the system, built by the end of the month.

How to measure whether it worked

On the Friday of each week, calculate per group: how many were contacted, how many replied, how many booked and how many showed up. The number that rules is the last one. Divide the bookings by the front-desk hours spent on that group and compare the groups against each other.

On the following Monday, attack first the group with the best result per hour and push down the queue whatever came in weak. If a group delivers less than 10% conversion two weeks in a row, take it out of the individual queue and move it to a mass send. My bet is that “incomplete course” beats “paid and never came” per hour spent, even with less money at stake, because the conversation is shorter and the patient already knows the clinic. If the opposite happens, swap the order of weeks 1 and 2 in the next rounds.

Limitations you need to know

The data ends on 30/04/2026 and today is 16/08/2026. That is 108 days that are not in the file. Everything I call “gone for 100 days” may have come back in the meantime. I did not build those 108 days into any number, so cross the lists against the current system before dialling.

Purchases stop before administrations. The last recorded purchase is on 10/04/2026, but the flow already thins out on 03/04: between 3 and 10 April there are 42 purchases in total, against an average of 32.3 a day in March. In practice the sales record was cut off on 02/04/2026. That does not affect the pending lists, but it means anything I say about “last month’s sales” is incomplete, and that is why I do not say it.

The gone-quiet group depends on nobody having bought outside the system. If the clinic records the administration of a vaccine the patient brought in, or a visit without a sale, those events are not here and some people on the list may be active.

The `data_cadastro` column was discarded. It runs from 05/01/2025 to 21/03/2026, but purchases start on 26/12/2023, and **1,922 of the 3,221 patients have a purchase earlier than their own registration date**. The column seems to record a system migration, not the patient’s arrival. No calculation here uses it.

The sex column is not reliable. It agrees with the ending of the first name in only 52.3% of cases, the same as a coin toss. I did not use sex to prioritise anyone, and the balancing quota of the high-value queue was decided by age band and neighbourhood, which are the cuts that were actually out of balance. The detail is in the appendix.

There is no phone number or email in the files. The lists come out with an identifier and a name; the front desk needs to cross them with the appointment book to get the contact.

The 30% to 40% conversion rates do not come from your base, but from market benchmarks for call-backs. They are the most fragile number in the report. After the first week you will have your own number, and it is worth more than mine.

Appendix

Files produced

- `pagou_e_nao_veio.csv` (322 rows): ordered by pending value descending. Biased towards those who bought an expensive vaccine: whoever is owed an MMR at R$ 210 drops to the end even after waiting longer.
- `esquema_incompleto.csv` (322 rows): same ordering and same bias.
- `sumidos_sem_pendencia.csv` (630 rows): ordered by time band and, within it, by historical spend descending. Under-serves those who spent little and kept coming back.
- `gripe_anual_vencida.csv` (146 rows): ordered by days since the last flu dose, oldest to most recent. A cut of the gone-quiet list, no new names.
- `alto_valor_em_risco.csv` (159 rows): ordered by historical spend descending, with the balancing quota by age band and by neighbourhood at the end. A cut of the three main groups; do not add to them.
- `sem_gripe_ativos.csv` (1,027 rows): ordered by historical spend descending. Active patients who have never bought quadrivalent flu; biased towards those who have already spent a lot, whoever bought one cheap vaccine just drops to the end.
- `cursos_repetidos_conferir.csv` (1,324 rows): ordered by course value descending. Audit list for gate 2.
- `aplicacoes_ultimo_dia_conferir.csv` (851 rows): ordered by the gap between purchase and administration, ascending, so the least plausible cases come first. Audit list for gate 1.

Images: `buckets_acionaveis.png`, `dias_desde_ultima_atividade.png`, `falta_por_pagamento.png`.

How I arrived at the groups

Term	Definition used
Reference date	30/04/2026, the latest date in any date column of the data
Course	Set of rows for the same patient, same vaccine and same purchase date
Paid and never came	Course with zero doses given
Incomplete course	Course with at least one dose given and at least one pending
Gone quiet	Patient with no pending dose and no purchase or administration for more than 90 days
Gone-quiet bands	Recent 91-180 days, medium 181-365 days, cold 366 days or more
Suspected repeat	Second or later purchase of the same vaccine by the same patient, excluding quadrivalent flu
Average course value	R$ 969 across the whole base (11,088 courses sold); in the new-sale projections, the band's own average: R$ 824 recent, R$ 1,006 medium, R$ 960 cold
Contact cost	R$ 2.50, from R$ 30 an hour and 12 contacts an hour

The "incomplete course" rule uses the protocol declared in the data itself, in the `protocolo_doses` column, which is consistent per vaccine (hepatitis B always 3, HPV-9 always 2, quadrivalent flu always 1, and so on). Since the protocol comes from the data and not from a heuristic, confidence in this group is high.

No-show rate by cut

Percentage of doses sold that were not given, and the value sitting still in each cut.

Cut	Highest rate	Lowest rate
Payment method	Boleto 16.0% (R$ 236,418)	Pix 4.8% (R$ 236,084)
Acquisition channel	Google 13.0% (R$ 339,371)	Referral 4.3% (R$ 193,379)
Dose number	Third 12.7%	First 5.9%
Vaccine	Hepatitis B 9.0%	Pneumococcal 20 6.1%
Age band	13-18 years 9.3%	46-60 years 7.0%
Neighbourhood	Jardim das Acácias 8.7%	Morada do Sol 6.4%
Sex	Male 7.9%	Female 7.4%

Two negative readings that also hold: **neighbourhood concentrates nothing**. The share of actionable patients per neighbourhood sits between 36% and 43% across the ten neighbourhoods, and the three largest together are 32.7% of the queue, practically their weight in the base. A neighbourhood campaign has no advantage over a general one, which is why I produced no neighbourhood file. **The sex column should not be used for anything.** Besides barely separating (7.9% against 7.4%), it does not match the names: among the 2,514 patients whose first name ends in “a” or “o”, the column agrees with the ending in only 52.3% of cases, the same as tossing a coin. Either the field was filled in wrong at source, or it is swapped in part of the records. I left the column in the files only for you to check, but do not prioritise or write messages based on it.

File columns that went into no calculation

`data_cadastro` (inconsistent, explained in the limitations) and `nome` from the vaccination spreadsheets, used only for display in the lists. Every other column in `pacientes.csv` and `vacinacoes.csv` went into at least one cut above.

---

* **About this sample.** This is a real report from the Distilo pipeline, run over synthetic data for a fictional company: the patient-recall report over the same synthetic clinic as the other samples. That company is Brazilian: a vaccination clinic, so the amounts are in Brazilian reais, the neighbourhood and patient names are Brazilian, and part of the catalogue (Qdenga against dengue, for instance) reflects a Brazilian vaccination schedule. It is published as it was produced rather than adapted, because the analysis is what we are showing you and the analysis does not change with the country. Your own report is written in your language, about your business, in your currency.